

Simultaneous inference for the mean of repeated functional data

Guanqun Cao^a, Li Wang^{b,*}

^a Department of Mathematics and Statistics, Auburn University, United States

^b Department of Statistics and the Statistical Laboratory, Iowa State University, United States

ARTICLE INFO

Article history:

Received 29 May 2017

Available online 9 February 2018

AMS 2000 subject classifications:

primary 62M10

secondary 62G08

Keywords:

B-spline

Confidence band

Functional data

Repeated measures

Semiparametric efficiency

ABSTRACT

Motivated by recent works studying the longitudinal diffusion tensor imaging (DTI) studies, we develop a novel procedure to construct simultaneous confidence bands for mean functions of repeatedly observed functional data. A fully nonparametric method is proposed to estimate the mean function and variance–covariance function of the repeated trajectories via polynomial spline smoothing. The proposed confidence bands are shown to be asymptotically correct by taking into account the correlation of trajectories within subjects. The procedure is also extended to the two-sample case in which we focus on comparing the mean functions from two populations of functional data. We show the finite-sample properties of the proposed confidence bands by simulation studies, and compare the performance of our approach with the “naive” method that assumes the independence within the repeatedly observed trajectories. The proposed method is applied to the DTI study.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

With modern technological progress in measuring devices, sophisticated data are now easy to collect. These data are often sets of functions such as curves, images or shapes, whose high-dimensional and correlated features impose tremendous challenges on conventional statistical studies. Emerging as a promising field, functional data analysis (FDA), which deals with the analysis of curves, has recently undergone intense development. The interested reader is referred to Ramsay and Silverman [16] for a general introduction of FDA.

In this work, we focus on situations where curves are repeatedly recorded for each subject, e.g., mortality data [6] in which age-specific lifetables are collected over years for various countries, and electroencephalography (EEG) data [9] observed for patients at each visit. Such dependent types of curves or images now commonly arise in diverse fields including climatology, demography, economics, epidemiology, and finance.

Our work is motivated by a longitudinal neuroimaging study containing repeated functional measurements derived from diffusion tensor imaging (DTI); for a description, see [11,12]. DTI is a magnetic resonance imaging technique which provides different measures of water diffusivity along brain white matter tracts; its use is instrumental, especially in diseases that affect the brain white matter tissue such as multiple-sclerosis (MS); see, e.g., [1]. In this study, DTI brain scans are recorded for many multiple-sclerosis (MS) patients to assess the effect of neurodegeneration on disability. At each visit, fractional anisotropy (FA) was determined via DTI along the corpus callosum (CCA). One objective here is to better understand the

* Corresponding author.

E-mail address: lilywang@iastate.edu (L. Wang).

demyelination process via its FA proxy and investigate possible differences therein between female and male patients. Point-wise confidence intervals via estimation ± 2 point-wise standard errors are provided in [17]. It is unclear, however, what is the performance of global inference on the true underlying mean profile.

In this paper, we develop simultaneous inference for the mean of repeated functional data. Our approach can handle the within-subject correlation and provide global inference, which are the key advantages of our approach over available FDA methods. There have been some recent attempts to study such repeated functional data in various settings. For example, the importance of models for dependent functional data has been recognized in [7,9], in which the emphasis has been on a general hierarchical model. Chen and Müller [6] proposed a flexible longitudinally observed functional model and provided consistency results and asymptotic convergence rates for the estimated model components. Zhu et al. [23] established the uniform convergence rate and confidence band for each estimated individual effect curve in multivariate varying coefficient models.

Simultaneous confidence bands (SCBs) are an important tool to address the variability in the unknown function and to develop global test statistics for general hypothesis testing problems. In Wang et al. [19,20], smooth SCBs are developed for the cumulative distribution functions. Gu and Yang [14] constructed SCBs for the link function in a single-index model based on the oracally efficient kernel estimator. It is of particular interest in FDA to construct SCBs for mean functions. For example, Bunea et al. [2] proposed an asymptotically conservative confidence set for the mean function of Gaussian functional data. Song et al. [18] proposed an asymptotically correct SCBs for dense functional data using local linear smoothing. Recently, polynomial splines have found successful applications in SCB construction. Ma et al. [15] suggested spline SCBs for mean functions of sparse functional data based on polynomial spline smoothing. Gu et al. [13] investigated a varying coefficient regression model for sparse functional data and proposed simultaneous confidence corridors for the coefficient functions. Cao et al. [4,5] provided SCBs for mean and derivative functions of dense functional data, respectively.

In this paper, we derive SCBs for mean functions when curves are repeatedly recorded for each subject. Existing methodologies for constructing SCBs in FDA often assume the independence of trajectories within each subject. Thus, the within-subject effect is not reflected by the traditional covariance functions of the mean curve. We are unaware of any methodology that provides exact SCBs for mean curves of repeatedly observed functional data. In this work, we use polynomial splines to approximate the mean and covariance functions in the construction of the SCBs. We show that the proposed spline SCBs are asymptotically correct and semiparametrically efficient in the sense that they are asymptotically the same as if all random trajectories were observed entirely and without errors as in [5]. We further consider two-sample inference for dependent functional data and extend our SCB construction procedure to a two-sample problem to test whether the mean functions from two groups are different.

The dependence within the repeatedly observed curves adds extra difficulty for model implementation, e.g., the estimation of within-subjects correlation. Misspecification of the correlation structure may lead to some efficiency loss. To tackle this issue, it is desirable to make the structure as model-free as it can be, and nonparametric modeling is particularly useful in this sense. In this paper we propose to estimate the variance–covariance functions nonparametrically. Our Monte Carlo results show that the proposed bands have much more accurate coverage rates of the true function compared to the “naive” method that ignores the within-subject dependence.

The paper is organized as follows. Section 2 states the model and introduces the estimates of mean functions for repeated functional data. Section 3.1 describes the asymptotic distribution of the estimators in the framework of allowing unknown dependence of the trajectories within subjects. Using this asymptotic result, we construct SCBs for mean functions. Section 3.2 develops the SCBs to study the difference of mean functions from two populations. Section 4 discusses how to estimate the components in the proposed bands. A simulation study is presented in Section 5. Section 6 contains applications of our method to a diffusion tensor imaging data. Section 7 gives the concluding remarks. Further insights into the error structure of spline estimators and technical proofs are collected in the Appendix.

2. The model and estimates

2.1. Modeling repeated functional measurements

We consider data $\{X_{ij}(s) : s \in \mathcal{X}\}$, $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, where X_{ij} is a repeated random curve on the compact interval $\mathcal{X}$, i is the subject index, and j is the repeated trajectory index for the i th subject. Assume that for all $j \in \{1, \dots, m_i\}$, X_{ij} are iid copies of the L_2 process X_j defined on $[0, 1]$, with mean function defined, for all $s \in [0, 1]$, by $\mu(s) = E\{X_j(s)\}$.

For the i th subject one has the Karhunen–Loève representation of the process of $X_{ij}(s)$, i.e., $X_{ij}(s) = \mu(s) + \sum_{k=1}^{\infty} \xi_{ijk} \phi_{jk}(s)$, where the random coefficients ξ_{ijk} s are referred to as the (jk) th functional principal component (FPC) scores of the i th subject. For each fixed (i, j) , the ξ_{ijk} s are uncorrelated with mean 0 and variance 1. For notational convenience, let $\phi_{jk} = \sqrt{\lambda_{jk}} \psi_{jk}$; then λ_{jk} and ψ_{jk} are the eigenvalues and eigenfunctions of the covariance operator with kernel $G_{ij}(s, t) = \text{cov}\{X_{ij}(s), X_{ij}(t)\}$, respectively. Although the sequences $\{\lambda_{jk}\}_{j,k=1}^{m_i, \infty}$, $\{\phi_{jk}\}_{j,k=1}^{m_i, \infty}$ and the random coefficients ξ_{ijk} s exist; however, they are unknown or unobservable.

Let $\mathbf{Y}_i(s) = (Y_{i1}(s), Y_{i2}(s), \dots, Y_{im_i}(s))^T$ for all $i \in \{1, \dots, n\}$, and assume $Y_{ij}(s) = X_{ij}(s) + \varepsilon_{ij}(s)$, where $\varepsilon_{ij}(s)$ are mean zero measurement errors. Suppose $X_{ij}(s) = \mu(s) + \eta_{ij}(s)$, where $\eta_{ij}(s)$ characterizes individual curve variations from $\mu(s)$. Denote $\boldsymbol{\varepsilon}_i(s) = (\varepsilon_{i1}(s), \dots, \varepsilon_{im_i}(s))^T$ and $\boldsymbol{\eta}_i(s) = (\eta_{i1}(s), \dots, \eta_{im_i}(s))^T$. Suppose $\boldsymbol{\varepsilon}_i(s)$ and $\boldsymbol{\eta}_i(s)$ are mutually independent. Moreover, assume that $\boldsymbol{\eta}_i(s)$ and $\boldsymbol{\varepsilon}_i(s)$ are iid copies of stochastic processes with mean vector $\mathbf{0}$ and covariance functions

$\mathbf{G}_i(s, t)$ and $\Gamma_i(s, t)$, respectively. Here $\mathbf{G}_i(s, t) = \{G_{ij'}(s, t)\}_{j, j'=1}^{m_i}$ is an $m_i \times m_i$ matrix of functions $G_{ij'}(s, t) = \text{cov}\{X_{1j}(s), X_{1j'}(t)\}$. Measurement errors $\varepsilon_i(s)$ and $\varepsilon_i(t)$ are assumed to be independent when $s \neq t$, and $\Gamma_i(s, t) = \Gamma_i(s)\mathbf{1}(s = t)$, where $\Gamma_i(s)$ is an $m_i \times m_i$ matrix of functions of s and $\mathbf{1}(\cdot)$ is an indicator function. Hence, the covariance function $\Sigma_i(s, t) = \text{cov}\{Y_i(s), Y_i(t)\} = \mathbf{G}_i(s, t) + \Gamma_i(s)\mathbf{1}(s = t)$.

It is assumed throughout this paper that we examine the equally spaced dense design, i.e., $Y_i(s)$ is measured at the same N location points $s = \ell/N$ with $\ell \in \{1, \dots, N\}$, for all i . For the i th subject and j th repeated trajectory, its sample path $\{(\ell/N, Y_{ij\ell})\}_{\ell=1}^N$ is a noisy realization of the continuous time stochastic process $X_{ij}(s)$ in the sense that $Y_{ij\ell} = X_{ij}(\ell/N) + \varepsilon_{ij}(\ell/N)$.

Combining the above representation leads to the following model for repeated functional measurements:

$$Y_{ij\ell} = \mu(\ell/N) + \sum_{k=1}^{\infty} \xi_{ijk} \phi_{jk}(\ell/N) + \varepsilon_{ij}(\ell/N), \quad (1)$$

where $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$ and $\ell \in \{1, \dots, N\}$. The problem addressed here is the estimation of μ and its SCB.

2.2. Spline estimators

A polynomial spline of order $p \geq 0$ on $[0, 1]$ with knot sequence $\omega_0 = 0 < \omega_1 < \dots < \omega_{N_\mu} < 1 = \omega_{N_\mu+1}$, is a function that is a polynomial of degree $p - 1$ on each of the intervals $[\omega_j, \omega_{j+1}]$ with $j \in \{0, \dots, N_\mu\}$, and globally has $p - 2$ times continuous derivatives for $p \geq 1$. The collection of spline functions of a particular order p and knot sequence form a linear space denoted by $\mathcal{H}^{(p-2)}$. We propose to approximate the mean function μ via a B-spline basis expansion, viz.

$$\mu(s) \approx \sum_{j=1-p}^{N_\mu} \beta_j B_j(s), \quad (2)$$

where B_j is the j th B-spline basis function of order p defined in [8]. The approximation sign in (2) will be replaced by a strict equality with a fixed and known N_μ when μ belongs to the spline space $\mathcal{H}^{(p-2)}$. If μ is not restricted to $\mathcal{H}^{(p-2)}$, it is natural to let N_μ increase with the sample size, allowing a more accurate approximation when the sample size increases. Following (2), we can estimate β_j , hence $\mu(s)$ with any given N_μ by solving the following least squares problem:

$$(\hat{\beta}_{1-p}, \dots, \hat{\beta}_{N_\mu}) = \arg \min_{(\beta_{1-p}, \dots, \beta_{N_\mu}) \in \mathbb{R}^{N_\mu+p}} \sum_{i=1}^n \sum_{\ell=1}^N \left\{ \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij\ell} - \sum_{j=1-p}^{N_\mu} \beta_j B_j(\ell/N) \right\}^2. \quad (3)$$

Therefore, the spline estimator of $\mu(s)$ is

$$\hat{\mu}(s) = \sum_{j=1-p}^{N_\mu} \hat{\beta}_j B_j(s). \quad (4)$$

The idea of using basis expansions can be applied more generally to other basis systems for function approximation such as polynomial bases and Fourier bases. We focus in this paper on B-splines because of the good approximation properties of splines and computational simplicity of the B-spline basis.

3. Simultaneous confidence bands (SCBs)

In the following, for $1 \leq j \leq \max_{1 \leq i \leq n} m_i = m$, we define

$$\begin{cases} Y_{ij}^\dagger = Y_{ij}, & \xi_{ijk}^\dagger = \xi_{ijk}, & \varepsilon_{ij}^\dagger = \varepsilon_{ij}, & \phi_{jk}^\dagger = \phi_{jk} & \text{if } Y_{ij} \text{ is observed;} \\ Y_{ij}^\dagger = 0, & \xi_{ijk}^\dagger = 0, & \varepsilon_{ij}^\dagger = 0, & \phi_{jk}^\dagger = 0 & \text{if } Y_{ij} \text{ is missing.} \end{cases}$$

Hence, for notation simplicity, we do not distinguish Y_{ij} , ξ_{ijk} , ε_{ij} and ϕ_{jk} from $Y_{ij}^\dagger$, $\xi_{ijk}^\dagger$, $\varepsilon_{ij}^\dagger$ and $\phi_{jk}^\dagger$, respectively. Without loss of generality, we take $\mathcal{X} = [0, 1]$ and assume $m_i = m$ for all $i \in \{1, \dots, n\}$.

3.1. Asymptotic properties of the spline estimator

First we study the asymptotic properties of the spline estimator given in (4). Define the average of all entries in $\mathbf{G}(s, t)$ as

$$\bar{G}(s, t) = \frac{1}{m^2} \sum_{j, j'=1}^m G_{jj'}(s, t).$$

Let $\zeta(s)$, $s \in [0, 1]$, be a standardized Gaussian process satisfying

$$E\{\zeta(s)\} = 0, \quad E\{\zeta^2(s)\} = 1, \quad E\{\zeta(s)\zeta(t)\} = \bar{G}^{-1/2}(s, s)\bar{G}(s, t)\bar{G}^{-1/2}(t, t)$$

for all $s, t \in [0, 1]$. For any $\alpha \in (0, 1)$, we denote $Q_{1-\alpha}$ the $100 \times (1 - \alpha)$ th percentile of the absolute maxima distribution of $\zeta(s)$, i.e., $\Pr\{\sup_{s \in [0, 1]} |\zeta(s)| \leq Q_{1-\alpha}\} = 1 - \alpha$. Let $z_{1-\alpha}$ be the $100 \times (1 - \alpha)$ th percentile of the standard normal distribution.

If $Y_i(s)$ were observed without any measurement errors, one could estimate the mean function μ by setting, for all $s \in [0, 1]$,

$$\bar{\mu}(s) = \bar{X}(s) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m X_{ij}(s). \quad (5)$$

For the general case when $m_i \neq m$,

$$\bar{\mu}(s) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{m_i} \sum_{j=1}^{m_i} X_{ij}(s) \right\}.$$

Since $\bar{\mu}(s)$ is computed from an unknown quantity $X_{ij}(s)$, it is also called the “infeasible estimator” in Cao et al. [5]. **Theorem 1** shows that the difference between the spline estimator $\hat{\mu}$ in (4) and the “infeasible estimator” $\bar{\mu}$ in (5) converges to zero at the rate $\sqrt{n}$, which is the same rate of convergence in the parametric setting. Thus, we have the oracle efficiency of the nonparametric estimator $\hat{\mu}$.

Theorem 1. Suppose Assumptions (A1)–(A5) in the Appendix hold. For any $\alpha \in (0, 1)$, as $n \rightarrow \infty$, the “infeasible estimator” $\bar{\mu}$ converges to μ uniformly at the $\sqrt{n}$ rate, viz.

$$\Pr \left\{ \sup_{s \in [0, 1]} \sqrt{n} |\bar{\mu}(s) - \mu(s)| \bar{G}(s, s)^{-1/2} \leq Q_{1-\alpha} \right\} \rightarrow 1 - \alpha,$$

and for all $s \in [0, 1]$,

$$\Pr \left\{ \sqrt{n} |\bar{\mu}(s) - \mu(s)| \bar{G}(s, s)^{-1/2} \leq z_{1-\alpha/2} \right\} \rightarrow 1 - \alpha,$$

while the spline estimator $\hat{\mu}$ is asymptotically equivalent to $\bar{\mu}$ up to order $\sqrt{n}$, i.e.,

$$\sup_{s \in [0, 1]} \sqrt{n} |\bar{\mu}(s) - \hat{\mu}(s)| = o_p(1).$$

The oracle efficiency in **Theorem 1** immediately indicates the following result, which can be used to construct SCBs or point-wise confidence intervals for $\mu(s)$ at any $s \in [0, 1]$.

Corollary 1. Under Assumptions (A1)–(A5) in the Appendix, for any $\alpha \in (0, 1)$, as $n \rightarrow \infty$, an asymptotic $100 \times (1 - \alpha)\%$ SCB for $\mu(s)$, is given, for all $s \in [0, 1]$, by

$$\hat{\mu}(s) \pm n^{-1/2} \bar{G}(s, s)^{1/2} Q_{1-\alpha}, \quad (6)$$

and an asymptotic $100 \times (1 - \alpha)\%$ point-wise confidence interval for $\mu(s)$, is given, for all $s \in [0, 1]$, by

$$\hat{\mu}(s) \pm n^{-1/2} \bar{G}(s, s)^{1/2} z_{1-\alpha/2}.$$

3.2. SCBs for the difference of two mean functions

The aforementioned SCBs for one sample mean function can be extended to the two-sample case. In this section, we consider the inference of the difference of mean regression functions from two populations. Analogous to the previous notations, we denote two samples indicated by $d \in \{1, 2\}$, which satisfy

$$Y_{ij\ell}^{(d)} = \mu^{(d)}(\ell/N) + \sum_{k=1}^{\infty} \xi_{ijk}^{(d)} \phi_{jk}^{(d)}(\ell/N) + \varepsilon_{ij}^{(d)}(\ell/N),$$

where $i \in \{1, \dots, n_d\}$, $j \in \{1, \dots, m_d\}$ and $\ell \in \{1, \dots, N\}$. Define the ratio of two-sample sizes as $\hat{\tau} = n_1/n_2$ and assume that $\hat{\tau} \rightarrow \tau > 0$ as $n_1 \rightarrow \infty$.

For $d \in \{1, 2\}$, let $\hat{\mu}^{(d)}$ be the spline estimate of mean function $\mu^{(d)}$ as given in (4), and define

$$\bar{G}^{(d)}(s, t) = \frac{1}{m_d^2} \sum_{j, j'=1}^{m_d} G_{jj'}^{(d)}(s, t).$$

Next let $\zeta_{\text{diff}}(s)$ be a standardized Gaussian process on $[0, 1]$ such that $E\{\zeta_{\text{diff}}(s)\} = 0$, $E\{\zeta_{\text{diff}}^2(s)\} = 1$ and, for all $s, t \in [0, 1]$,

$$E\{\zeta_{\text{diff}}(s)\zeta_{\text{diff}}(t)\} = \frac{\bar{G}^{(1)}(s, t) + \tau \bar{G}^{(2)}(s, t)}{[\{\bar{G}^{(1)}(s, s) + \tau \bar{G}^{(2)}(s, s)\} \{\bar{G}^{(1)}(t, t) + \tau \bar{G}^{(2)}(t, t)\}]^{1/2}}.$$

Denote $Q_{\text{diff}, 1-\alpha}$ the $(1 - \alpha)$ th quantile of the absolute maximum deviation of $\zeta_{\text{diff}}(s)$ with $s \in [0, 1]$, as above. We mimic the two-sample t -test and state the following theorem whose proof is analogous to that of Theorem 1.

Theorem 2. If Assumptions (A1)–(A5) in the Appendix are modified for each group accordingly, then for any $\alpha \in (0, 1)$, as $n_1 \rightarrow \infty$, $\hat{\tau} \rightarrow \tau > 0$,

$$\Pr \left[\sup_{s \in [0, 1]} \frac{n_1^{1/2} |(\hat{\mu}^{(1)} - \hat{\mu}^{(2)} - \mu^{(1)} + \mu^{(2)})(s)|}{\{(\bar{G}^{(1)} + \tau \bar{G}^{(2)})(s, s)\}^{1/2}} \leq Q_{\text{diff}, 1-\alpha} \right] \rightarrow 1 - \alpha.$$

Theorem 2 yields the following simultaneous asymptotic SCB for $\mu^{(1)}(s) - \mu^{(2)}(s)$ with $s \in [0, 1]$.

Corollary 2. If Assumptions (A1)–(A5) in the Appendix are modified for each group accordingly and $\hat{\tau} \rightarrow \tau > 0$ as $n_1 \rightarrow \infty$, then for any $\alpha \in (0, 1)$, a $100 \times (1 - \alpha)\%$ asymptotically correct confidence band for $\mu^{(1)} - \mu^{(2)}$ is given, for all $s \in [0, 1]$, by

$$(\hat{\mu}^{(1)} - \hat{\mu}^{(2)})(s) \pm n_1^{-1/2} Q_{\text{diff}, 1-\alpha} \{(\bar{G}^{(1)} + \tau \bar{G}^{(2)})(s, s)\}^{1/2}.$$

4. Estimation of the components in SCBs

This section presents procedures to implement the SCBs given in Corollaries 1 and 2. Given any dataset $\{(\ell/N, Y_{ij\ell})\}_{i=1, j=1, \ell=1}^{n, m_i, N}$ from Model (1), the spline estimator $\hat{\mu}(s)$ can be easily obtained from (A.1) once the basis functions are determined. In the following we describe how to estimate the unknown function $\bar{G}$ and the quantiles $Q_{1-\alpha}$ and $Q_{\text{diff}, 1-\alpha}$.

4.1. Estimating the variance–covariance functions

We give estimating formulas based on general repeated dataset ($m_i \neq m$). Define $r_{ij} = \mathbf{1}\{Y_{ij} \text{ is observed}\}$, so that $\sum_{i=1}^n \sum_{j=1}^m r_{ij} = \sum_{i=1}^n m_i = n_m$. For all $\ell, \ell' \in \{1, \dots, N\}$ with $\ell \neq \ell'$, let

$$\bar{C}_{\ell\ell'} = \frac{1}{n_m} \sum_{i=1}^n \sum_{j=1}^{m_i} r_{ij} \{Y_{ij\ell} - \hat{\mu}(\ell/N)\} \{Y_{ij\ell'} - \hat{\mu}(\ell'/N)\}.$$

We pre-estimate the covariance function $G_{jj}(s, t)$ using the tensor product spline approach by Cao et al. [5]. The pilot spline estimator of $G_{jj}(s, t)$ is defined as

$$\hat{G}_{jj}(s, t) = \sum_{J, J'=1-p}^{N_G} \hat{b}_{JJ'} B_J(s) B_{J'}(t), \quad (7)$$

where N_G is the number of interior knots used to build the tensor product B-spline basis and the spline coefficients

$$\{\hat{b}_{JJ'}\}_{J, J'=1-p}^{N_G} = \arg \min_{\mathbb{R}^{N_G+p} \otimes \mathbb{R}^{N_G+p}} \sum_{\ell \neq \ell'}^N \left\{ \bar{C}_{\ell\ell'} - \sum_{1-p \leq J, J' \leq N_G} \hat{b}_{JJ'} B_J(\ell/N) B_{J'}(\ell'/N) \right\}^2.$$

Next we consider the eigenfunction decomposition of $\hat{G}_{jj}(s, t)$, i.e., for all $j \in \{1, \dots, m\}$,

$$\frac{1}{N} \sum_{\ell=1}^N \hat{G}_{jj}(\ell/N, \ell'/N) \hat{\psi}_{jk}(\ell/N) = \hat{\lambda}_{jk} \hat{\psi}_{jk}(\ell'/N),$$

so that we obtain the estimated eigenvalues $\hat{\lambda}_{jk}$ and eigenfunctions $\hat{\psi}_{jk}$. Following Yao et al. [22], the FPC scores ξ_{ijk} can be approximated by

$$\hat{\xi}_{ijk} = \frac{1}{N} \sum_{\ell=1}^N \hat{\lambda}_{jk}^{-1} \{Y_{ij\ell} - \hat{\mu}(\ell/N)\} \hat{\phi}_{jk}(\ell/N) r_{ij}.$$

Note that $G_{jj'}(s, t) = \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} E(\xi_{ijk} \xi_{ij'k'}) \phi_{jk}(s) \phi_{j'k'}(t)$ for $j \neq j'$, so we can estimate $G_{jj'}(s, t)$ by

$$\hat{G}_{jj'}(s, t) = \frac{1}{\sum_{i=1}^n r_{ij} r_{ij'}} \sum_{i=1}^n \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \hat{\xi}_{ijk} \hat{\xi}_{ij'k'} \hat{\phi}_{jk}(s) \hat{\phi}_{j'k'}(t).$$

Thus, $\bar{G}(s, s)$ can be estimated using $\hat{G}(s, s) = \sum_{j,j'=1}^m \hat{G}_{jj'}(s, s)/m^2$.

The following theorem shows that $\hat{G}(s, t)$ and $\bar{G}(s, t)$ are asymptotically equivalent. Hence, we replace $\bar{G}(s, s)$ in (6) by $\hat{G}(s, s)$ when constructing SCB.

Theorem 3. If Assumptions (A1)–(A5) in the Appendix hold, then

$$\sup_{s,t \in [0,1]} |\hat{G}(s,t) - \bar{G}(s,t)| = o_p(1).$$

The proof of Theorem 3 is given in the Appendix.

4.2. Estimating the quantiles

We generate independent $\mathbb{R}^m$ -valued Gaussian vectors, $\mathbf{Z}_{k,b} = (Z_{1k,b}, \dots, Z_{mk,b})^\top$ such that $\text{cov}(Z_{jk,b}, Z_{j'k,b}) = \text{cov}(\xi_{1jk}, \xi_{1j'k}) \equiv V_{jj',k}$, $\text{cov}(Z_{jk,b}, Z_{j'k',b}) = 0$ if $k \neq k'$ for all $b \in \{1, \dots, b_M\}$. Here b_M is a preset large integer, the default of which is 1000. Let

$$\hat{\zeta}_b(s) = \hat{G}(s,s)^{-1/2} \left\{ \frac{1}{m} \sum_{j=1}^m \sum_{k=1}^\infty Z_{jk,b} \hat{\psi}_{jk}(s) \right\}.$$

One takes the maximal absolute value for each copy of $\hat{\zeta}_b(s)$ and estimates $Q_{1-\alpha}$ by the empirical quantile $\hat{Q}_{1-\alpha}$ of these maximum values.

In the two-sample case, we generate independent $\mathbb{R}^{m_d}$ -valued Gaussian vectors, $\mathbf{Z}_{k,b}^{(d)} = (Z_{1k,b}^{(d)}, \dots, Z_{m_d k,b}^{(d)})^\top$, $d \in \{1, 2\}$, satisfying

$$\text{cov}(Z_{jk,b}^{(d)}, Z_{j'k,b}^{(d)}) = V_{jj',k}^{(d)}, \quad \text{cov}(Z_{jk,b}^{(d)}, Z_{j'k',b}^{(d)}) = 0$$

if $k \neq k'$ for all $b \in \{1, \dots, b_M\}$. Similarly to $\hat{\zeta}_b$, we define

$$\hat{\zeta}_{\text{diff},b}(s) = \left\{ (\hat{G}^{(1)} + \hat{\tau} \hat{G}^{(2)})(s,s) \right\}^{-1/2} \times \left\{ \frac{1}{m_1} \sum_{j=1}^{m_1} \sum_{k=1}^\infty Z_{jk,b}^{(1)} \hat{\psi}_k^{(1)}(s) + \frac{\hat{\tau}}{m_2} \sum_{j=1}^{m_2} \sum_{k=1}^\infty Z_{jk,b}^{(2)} \hat{\psi}_k^{(2)}(s) \right\}.$$

We obtain $\hat{Q}_{\text{diff},1-\alpha}$ by taking the empirical quantile of these maximum values of $\hat{\zeta}_{\text{diff},b}(s)$.

4.3. Selecting κ and spline knots

To take into account the truncation error in generating the critical value, we choose the number of eigenfunctions using the following criterion. For $j \in \{1, \dots, m\}$, let

$$\kappa_j^* = \arg \min_{\ell \in \{1, \dots, T_j\}} \left\{ \sum_{k=1}^{\ell} \hat{\lambda}_{jk} / \sum_{k=1}^{T_j} \hat{\lambda}_{jk} > 0.95 \right\},$$

where $\hat{\lambda}_{j1}, \dots, \hat{\lambda}_{jT_j}$ are the first T_j estimated positive eigenvalues. Denote $\kappa = (\kappa_1, \dots, \kappa_m)$, where $\kappa_j^* \leq \kappa_j \leq T_j$, for $j \in \{1, \dots, m\}$. For a given κ , let

$$\hat{\zeta}_{b,\kappa}(s) = \hat{G}(s,s)^{-1/2} \left\{ \frac{1}{m} \sum_{j=1}^m \sum_{k=1}^{\kappa_j} Z_{jk,b} \hat{\psi}_{jk}(s) \right\},$$

and we take the maximal absolute value for each copy of $\hat{\zeta}_{b,\kappa}(s)$, then define $Q_{1-\alpha,\kappa}$ by the empirical quantile $\hat{Q}_{1-\alpha,\kappa}$ of these maximum values. Finally, we estimate $Q_{1-\alpha}$ by $\hat{Q}_{1-\alpha} = \max_{\kappa} \hat{Q}_{1-\alpha,\kappa}$ to reduce the effect of the truncation and provide a better coverage.

For knot selection in our procedure, we use equally spaced knots and select only the number of interior knots. According to Assumption (A3) in the Appendix, the number of knots in estimating $\mu(s)$ is taken to be $N_\mu = \lfloor n^{1/(2p)} \ln(n) \rfloor$, in which $\lfloor a \rfloor$ denotes the integer part of a . Meanwhile, the estimator $\tilde{G}_{ij}(s,t)$ in (7) is calculated with the number of knots $N_G = \lfloor 2n^{1/(2p)} \ln(\ln(n)) \rfloor$. Similar knots selections are suggested in Cao et al. [5]. These choices of knots also satisfy Assumption (A3) in the Appendix.

5. Simulation

5.1. Example 1: empirical coverage rates of the SCBs

In this example, we conduct a study to investigate the performance of the proposed SCBs under various correlation structure settings. Our data are generated from the following model

$$Y_{ij\ell} = \mu(\ell/N) + \sum_{k=1}^6 \xi_{ijk} \phi_k(\ell/N) + \varepsilon_{ij\ell}, \quad (8)$$

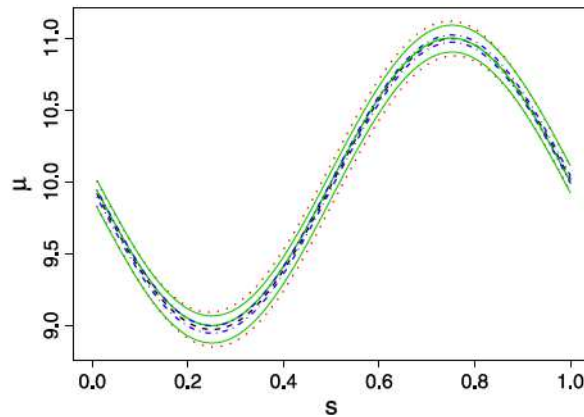


Fig. 1. The estimated mean function (middle dashed line), the true mean function (middle solid line) and its 95% “ORA” band constructed using the true correlation structure (upper and lower solid lines), the 95% “NP” band via the nonparametrically estimated variance–covariance function (upper and lower dotted lines), and the 95% “naive” band assuming independence within the repeatedly observed curves (upper and lower dot-dashed lines).

where $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$, $\ell \in \{1, \dots, N\}$, $\mu(s) = 10 + \sin\{2\pi(s - 1/2)\}$, $\phi_{2k-1}(s) = \cos(k\pi s)$ and $\phi_{2k}(s) = \sin(k\pi s)$ for $k \in \{1, 2\}$. We generate the FPC scores $\xi_{ik} = (\xi_{i1k}, \dots, \xi_{im_ik})^\top \sim \mathcal{N}(\mathbf{0}, \lambda_k \Omega)$, $\lambda_{2k-1} = \lambda_{2k} = 0.5^{2k}$ for $k \in \{1, 2\}$, and $\varepsilon_{i\ell} = (\varepsilon_{i1\ell}, \dots, \varepsilon_{im_i\ell})^\top \sim \mathcal{N}(\mathbf{0}, 0.1^2 \Omega)$. We consider the following settings for the correlation matrix $\Omega = \{\Omega_{jj'}\}_{j,j'=1}^{m_i}$:

Case 1. Independent (IND): $\Omega_{jj'} = \mathbf{1}(j = j')$.

Case 2. Autoregressive (AR-1): $\Omega_{jj'} = \mathbf{1}(j = j') + \rho^{|j-j'|} \mathbf{1}(j \neq j')$, $\rho = 0.2$.

Case 3. Toeplitz (TOEP): $\Omega_{jj'} = \mathbf{1}(j = j') + \rho_{|j-j'|} \mathbf{1}(j \neq j')$, where $\rho_{|j-j'|} = \rho_r$ with $0.1^r + 0.05$ for all $r \in \{1, \dots, m_i - 1\}$.

The number of subjects n is taken to be 30, 60 and 120, and the number of observations per curve is assumed to be $N = n$. The number of repeatedly observed curves for each subject is $m_i \in \{3, 5, 7\}$ with two different missing rates: 0 (fully observed) and $\{0, 0.2/m_i, \dots, 0.2(m_i - 1)/m_i\}$, i.e., j th repeatedly observed curve is missing with probability $0.2(j - 1)/m_i$. These settings mimic the real data structure in Section 6. We consider two nominal levels, namely $1 - \alpha = 0.95$ and 0.99 , and carry out 500 simulation replications.

For each setting, we construct the spline SCBs in three different ways: (i) using the “naive” independent correlation structure (IND); (ii) using the proposed nonparametric method (NP); (iii) using the “oracle” method with the true correlation structure (ORA). We adopt quadratic spline ($p = 3$) and cubic spline ($p = 4$) smoothing for each case, respectively. The coverage of the “IND”, “NP” and “ORA” bands is evaluated on grids $\{1/N, \dots, 99/N, 1\}$ and is checked whether the true function is covered entirely by the SCBs at these points. The “ORA” bands using the true correlation structure is expected to be the best among the three type of bands and serve as a benchmark.

Tables 1–2 summarize the empirical coverage probabilities of three different types of bands given different missing rates and spline orders. In Tables 1–2, one sees that in most of the cases the coverage probabilities of the bands improve with sample size. In addition, as expected, the coverage percentages of the “oracle” bands are the closest to the nominal levels since they are constructed using the correct correlation structure. For different missing mechanisms and spline orders, the “NP” bands constructed using the nonparametrically estimated variance–covariance functions are comparable to the “oracle” bands, and are much better than the “IND” bands using independence correlation structure when the true correlation is not independent. For the “NP” bands, the differences between the coverage probabilities and the nominal levels are acceptable when n is large enough, which implies that the proposed nonparametric method is very robust when the true correlation structure is unknown. In contrast, when the true correlation is not independent, the “naive” bands obtained using the independent correlation structure have significantly smaller coverage rates than the nominal levels regardless of the sample sizes and the number of repeatedly observed curves.

Fig. 1 displays a typical comparison between two 95% SCBs constructed using the independence structure, fully nonparametric structure and the true “TOEP” correlation structure. This plot is based on one sampled replication with $n = 120$, $p = 4$ and $m_i = 3$. In Fig. 1, one sees that the “IND”, “ORA” and “NP” produce the same estimator for the true curve μ , however, the “IND” band is very different from the “ORA” and “NP” bands. From Fig. 1, one can see that the true curve μ lies completely inside the “ORA” and “NP” bands, but it cannot be covered entirely by the “IND” bands.

5.2. Example 2: two-sample testing

We conduct a numerical study to evaluate the empirical size and power of hypothesis tests for the mean functions from two groups of repeatedly observed trajectories. We compare the SCBs constructed using the proposed nonparametric

Table 1

Empirical coverage rates of the bands under various correlation structures with missing rate = 0.

Correlation	p	n	α	$m_i = 3$			$m_i = 5$			$m_i = 7$		
				IND	NP	ORA	IND	NP	ORA	IND	NP	ORA
IND	3	30	0.05	0.934	0.930	0.934	0.926	0.920	0.916	0.944	0.938	0.942
			0.01	0.978	0.978	0.978	0.994	0.976	0.994	0.994	0.986	0.994
		60	0.05	0.946	0.938	0.948	0.960	0.954	0.958	0.970	0.966	0.970
			0.01	0.998	1.000	0.994	0.996	0.996	0.996	0.994	0.990	0.992
		120	0.05	0.950	0.954	0.948	0.948	0.946	0.948	0.944	0.946	0.946
			0.01	0.992	0.992	0.990	0.994	0.992	0.992	0.992	0.992	0.990
	4	30	0.05	0.956	0.946	0.960	0.930	0.938	0.928	0.944	0.938	0.946
			0.01	0.992	0.986	0.992	0.990	0.988	0.990	0.986	0.984	0.988
		60	0.05	0.952	0.952	0.954	0.946	0.942	0.952	0.948	0.936	0.944
			0.01	0.990	0.996	0.990	0.992	0.986	0.990	0.994	0.994	0.996
		120	0.05	0.966	0.960	0.964	0.930	0.930	0.930	0.948	0.952	0.954
			0.01	0.996	0.996	0.996	0.990	0.984	0.990	0.986	0.986	0.984
AR-1	3	30	0.05	0.884	0.924	0.926	0.864	0.916	0.934	0.864	0.920	0.952
			0.01	0.958	0.966	0.972	0.966	0.988	0.984	0.974	0.992	0.998
		60	0.05	0.892	0.922	0.946	0.876	0.928	0.948	0.870	0.922	0.940
			0.01	0.968	0.980	0.984	0.960	0.986	0.994	0.958	0.976	0.994
		120	0.05	0.926	0.964	0.970	0.874	0.934	0.948	0.894	0.946	0.958
			0.01	0.992	0.998	0.998	0.960	0.986	0.990	0.966	0.992	0.986
	4	30	0.05	0.884	0.918	0.932	0.856	0.896	0.920	0.870	0.916	0.950
			0.01	0.952	0.980	0.976	0.956	0.968	0.988	0.954	0.982	0.990
		60	0.05	0.882	0.930	0.950	0.888	0.928	0.950	0.888	0.940	0.954
			0.01	0.980	0.988	0.994	0.970	0.980	0.988	0.964	0.984	0.996
		120	0.05	0.900	0.940	0.954	0.886	0.940	0.962	0.854	0.934	0.966
			0.01	0.976	0.986	0.986	0.972	0.984	0.968	0.968	0.984	0.990
TOEP	3	30	0.05	0.926	0.934	0.934	0.864	0.918	0.928	0.826	0.900	0.914
			0.01	0.980	0.984	0.990	0.960	0.972	0.978	0.944	0.974	0.978
		60	0.05	0.886	0.920	0.924	0.846	0.908	0.922	0.850	0.912	0.940
			0.01	0.960	0.976	0.976	0.948	0.976	0.980	0.966	0.982	0.988
		120	0.05	0.904	0.930	0.944	0.876	0.924	0.952	0.852	0.930	0.954
			0.01	0.974	0.988	0.992	0.964	0.982	0.994	0.966	0.990	0.992
	4	30	0.05	0.890	0.910	0.920	0.870	0.916	0.920	0.856	0.920	0.940
			0.01	0.958	0.978	0.978	0.956	0.978	0.980	0.966	0.976	0.976
		60	0.05	0.892	0.912	0.926	0.852	0.914	0.928	0.858	0.930	0.950
			0.01	0.964	0.970	0.980	0.956	0.982	0.976	0.948	0.976	0.980
		120	0.05	0.900	0.934	0.942	0.874	0.942	0.946	0.866	0.938	0.962
			0.01	0.970	0.990	0.990	0.968	0.984	0.994	0.960	0.990	0.996

variance–covariance functions with the “naive” bands developed under the independence assumption. To mimic the two-sample testing problem, we consider the following hypotheses:

$$\mathcal{H}_0 : \forall_{s \in [0,1]} \mu^{(1)}(s) = \mu^{(2)}(s) \Leftrightarrow \mathcal{H}_a : \exists_{s \in [0,1]} \mu^{(1)}(s) \neq \mu^{(2)}(s). \quad (9)$$

We generate two groups of data from the model given in (8) with $\mu^{(1)}(s) = 10 + \sin\{2\pi(s - 1/2)\} + \delta$ and $\mu^{(2)}(s) = 10 + \sin\{2\pi(s - 1/2)\}$, and all the remaining settings are the same as in Example 1. The constant δ takes 10 values, namely $\delta \in \{0, 0.03, 0.06, \dots, 0.27\}$. Indeed, large values of δ shift the mean of the first group data further away from the second, therefore making the difference between $\mu^{(1)}$ and $\mu^{(2)}$ more easily detectable. We choose $n_1 = n_2 = 60$, take a cubic spline ($p = 4$), and $N = 60$ observation points on each curve. For each subject, the noisy trajectory is repeatedly observed three times with the true correlation structures: “AR-1” ($\rho = 0.2$) and “TOEP” ($\rho_r = 0.1^r + 0.05$ for all $r \in \{1, 2, 3\}$). The missing rate is $\{0, 0.2/3, 0.4/3\}$ for each individual at repeated time 1, 2, 3. We run 500 Monte Carlo simulations.

Fig. 2 illustrates the empirical frequencies of rejecting $\mathcal{H}_0$ against δ using our method and the naive “IND” method. For each correlation structure, we compare the performance of the “IND” bands and the “NP” bands. When $\delta = 0$, these relative frequencies represent the size of the test. The relative frequency for the proposed method is around 7% regardless of the true correlation structures, which is fairly close to the set significant level of 5%. This shows that our estimate of the null distribution is approximately correct. However, when $\delta = 0$, the relative frequency for “IND” is about 12% under both AR-1 and TOEP correlation structures, which is much larger than the set significance level. This implies that the “IND” is too liberal to maintain its sizes at the nominal level. In addition, one sees that increasing values of δ improves the power of detecting the alternatives for both methods. In fact, the power of our method increases more rapidly than that of the “IND” method,

Table 2

Empirical coverage rates of the bands under various correlation structures with missing rate $\{0, 0.2/m_i, \dots, 0.2(m_i - 1)/m_i\}$.

Correlation	p	n	α	$m_i = 3$			$m_i = 5$			$m_i = 7$		
				IND	NP	ORA	IND	NP	ORA	IND	NP	ORA
IND	3	30	0.05	0.914	0.910	0.914	0.932	0.922	0.934	0.930	0.900	0.932
			0.01	0.986	0.978	0.984	0.984	0.986	0.986	0.984	0.980	0.982
		60	0.05	0.928	0.936	0.932	0.932	0.930	0.928	0.922	0.914	0.920
			0.01	0.984	0.982	0.982	0.978	0.974	0.978	0.988	0.982	0.984
		120	0.05	0.942	0.936	0.940	0.940	0.934	0.936	0.940	0.938	0.940
			0.01	0.980	0.980	0.982	0.984	0.986	0.984	0.990	0.990	0.996
	4	30	0.05	0.920	0.912	0.922	0.940	0.916	0.942	0.912	0.908	0.918
			0.01	0.976	0.972	0.978	0.972	0.986	0.972	0.992	0.984	0.992
		60	0.05	0.938	0.926	0.932	0.932	0.926	0.934	0.932	0.912	0.922
			0.01	0.986	0.988	0.988	0.994	0.984	0.990	0.980	0.982	0.978
		120	0.05	0.952	0.958	0.954	0.942	0.936	0.944	0.926	0.922	0.926
			0.01	0.994	0.990	0.994	0.988	0.988	0.990	0.984	0.992	0.988
AR-1	3	30	0.05	0.868	0.904	0.926	0.858	0.896	0.922	0.844	0.908	0.928
			0.01	0.956	0.962	0.976	0.970	0.976	0.988	0.946	0.976	0.988
		60	0.05	0.894	0.928	0.938	0.842	0.900	0.944	0.848	0.902	0.920
			0.01	0.958	0.970	0.982	0.964	0.984	0.988	0.938	0.970	0.980
		120	0.05	0.892	0.922	0.942	0.890	0.936	0.948	0.846	0.920	0.942
			0.01	0.972	0.984	0.986	0.956	0.980	0.982	0.958	0.972	0.988
	4	30	0.05	0.882	0.912	0.922	0.826	0.866	0.896	0.840	0.880	0.912
			0.01	0.960	0.970	0.976	0.940	0.960	0.974	0.946	0.956	0.976
		60	0.05	0.868	0.904	0.932	0.848	0.892	0.916	0.844	0.910	0.932
			0.01	0.868	0.982	0.932	0.934	0.964	0.982	0.954	0.964	0.984
		120	0.05	0.918	0.940	0.950	0.876	0.926	0.940	0.864	0.926	0.940
			0.01	0.974	0.990	0.990	0.962	0.986	0.994	0.950	0.974	0.974
TOEP	3	30	0.05	0.852	0.896	0.896	0.932	0.922	0.934	0.826	0.890	0.904
			0.01	0.952	0.974	0.978	0.984	0.986	0.986	0.934	0.974	0.980
		60	0.05	0.888	0.914	0.920	0.932	0.930	0.928	0.836	0.914	0.932
			0.01	0.954	0.964	0.976	0.978	0.974	0.978	0.944	0.986	0.982
		120	0.05	0.906	0.932	0.942	0.940	0.934	0.936	0.842	0.922	0.942
			0.01	0.964	0.984	0.990	0.984	0.986	0.984	0.948	0.978	0.986
	4	30	0.05	0.888	0.914	0.926	0.862	0.900	0.918	0.814	0.880	0.898
			0.01	0.966	0.976	0.980	0.944	0.970	0.972	0.934	0.970	0.972
		60	0.05	0.900	0.922	0.942	0.860	0.922	0.948	0.874	0.920	0.934
			0.01	0.976	0.986	0.994	0.958	0.968	0.978	0.952	0.976	0.984
		120	0.05	0.888	0.926	0.944	0.890	0.932	0.946	0.848	0.914	0.942
			0.01	0.966	0.988	0.988	0.966	0.984	0.990	0.948	0.974	0.984

and the power of both methods reaches 1 simultaneously at around $\delta = 0.27$ regardless of the true correlation structures. In summary, given that Type I errors are generally considered more serious than Type II errors, our method is preferable to the “IND” method.

6. Empirical examples

6.1. Estimation and SCBs for the mean function of DTI data

In this section, we apply our proposed method to the DTI study of Goldsmith et al. [11,12]. In this study, DTI brain scans are recorded for many MS patients at several visits with the goal of assessing the effect of neurodegeneration on disability. The dataset is publicly available from the `refund` package in the R software, which consists of 100 subjects. The number of visits per subject ranged from 2 to 8, with a median of 3, and was approximately annual. In this dataset, we consider the DTI measure called FA along CCA for case group. After deleting incomplete data, we use the remaining 332 observations in our analysis. For illustration, the left panel in Fig. 3 shows 50 subjects' trajectories for the FA profiles.

Let $Y_{ij}(s)$ denote the FA value observed for the i th patient, the j th visit at the s th location, with $s \in [0, 93]$, where $i \in \{1, \dots, 100\}$ and $j \in \{1, \dots, m_i\}$. Applying cubic spline smoothing to the dataset, we obtain the estimated overall mean function of FA values $\hat{\mu}$. To construct the SCB, we first estimate the variance–covariance functions using the proposed nonparametric method in Section 4. To explore the comparison of the bands under different structure assumptions, in the right panel of Fig. 3 we show the proposed 95% SCB (upper and lower dashed lines).

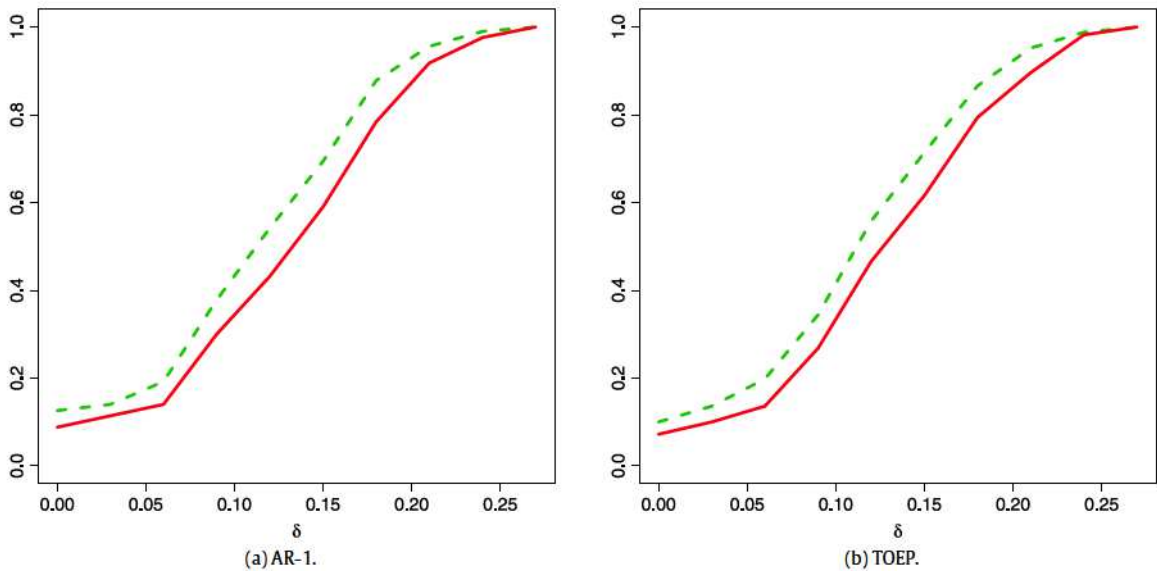


Fig. 2. Type I errors and powers for hypothesis testing (9) at 5% nominal levels. For each case, sample sizes are $n_1 = n_2 = N = 60$ and each trajectory is repeatedly observed three times with missing rate $\{0, 0.2/3, 0.4/3\}$. The dotted line represents the performance of the naive method using the “IND” assumption and the solid line represents the performance of the proposed method using the nonparametrically estimation variance–covariance function.

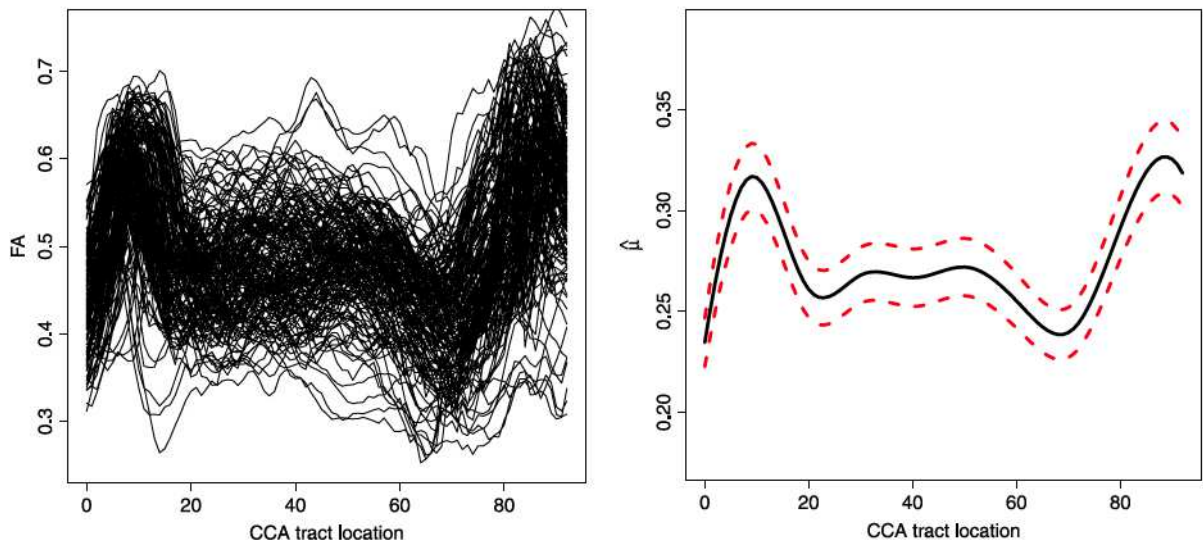


Fig. 3. Left: Sample FA profiles along CCA. Right: Cubic spline estimator (middle solid line) $\hat{\mu}(s)$ and its 95% “NP” bands (upper and lower dashed lines) using the nonparametrically estimated variance–covariance function.

6.2. Two-sample test for the difference between male and female FA

We further compare the FA values of female and male subjects. We focus on the same 100 subjects analyzed in Section 6.1 and consider male and female subjects as two different populations. Our hypotheses of interest are

$$\mathcal{H}_0 : \forall_{s \in [0, 93]} \mu_M(s) = \mu_F(s) \Leftrightarrow \mathcal{H}_a : \exists_{s \in [0, 93]} \mu_M(s) \neq \mu_F(s),$$

where $\mu_M(s)$ and $\mu_F(s)$ are the mean functions of mortality rates for males and females with sample sizes $n_1 = 66$, $n_2 = 34$.

Fig. 4 depicts the cubic spline SCBs constructed using the nonparametrically estimated variance–covariance function at 0.95 (upper and lower dotted lines) and 0.999 confidence level (upper and lower dashed lines), with the center dashed-dotted line representing the spline estimator $\hat{\mu}_M(s) - \hat{\mu}_F(s)$ and a solid line representing zero. The 99.9% SCB goes below the zero line except around CCA tract locations 0, 15 and 80, the difference between male and female FA is significant at

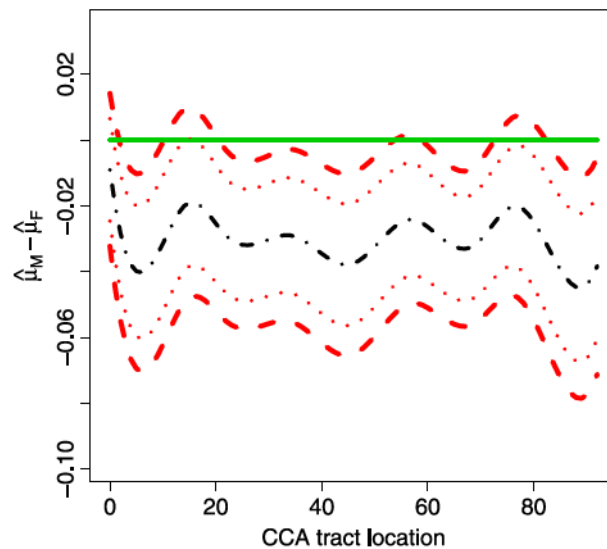


Fig. 4. Cubic spline estimator $\hat{\mu}_M(s) - \hat{\mu}_F(s)$ (dashed-dotted line) and the corresponding 95% (upper and lower dotted lines) and 99.9% “NP” band (upper and lower dashed lines) using the nonparametrically estimated variance-covariance function. The solid line represents zero reference.

level 0.001. It is fairly clear from Fig. 4 that females have a lower FA measurement than males at the same location. Fig. 4 also suggests that the variation of the discrepancy is constant over the range of tract locations. Similar findings on the discrepancy of FA profiles for difference gender patients have also been reported in Scheipl et al. [17].

7. Conclusions

In this paper, we target the inference of mean functions for repeatedly observed functional data. We allow for dependence of the trajectories within subjects and obtain both the global and local asymptotic distributions of the spline estimator. SCBs are developed to quantify and visualize the variability of the estimator and to make global inferences on the shape of the population mean function. In longitudinal analysis, one primary hypothesis is to test if the pattern of observations over time is constant within a time interval. Our paper provides a method to test this hypothesis. Note that the hypothesis of no time effect states that all the J mean functions are identical. Under this assumption, the asymptotic null distribution is obtained. The proposed SCBs based on the asymptotic distribution can be used to test this hypothesis. If this $100 \times (1 - \alpha)\%$ SCB covers the zero line, one cannot reject the null hypothesis of no time effect, with p -value not greater than α .

In the development of the bands for repeated functional data, the presence of within-subjects dependence poses new challenges for us, and we propose a fully nonparametric method to estimate the covariance function to avoid misspecification. The advantages of our approach are: first, compared with the “naive” band that assumes iid trajectories, the proposed SCB has excellent coverage rate of the true mean function regardless of the within-subjects structures, and thus it can be applied flexibly to both independent and dependent functional data; second, it provides valuable insights into the correlation structures of dependent functional data as in practice the true correlation structure is usually unknown; third, this approach can be easily extended from the one-/two-sample case to a multi-sample testing problem; fourth, when estimating covariance function and within-subjects correlation, it does not require complex fitting algorithms, and the bands can be usually constructed within fractions of a second.

Acknowledgments

Cao's research was supported in part by the Simons Foundation under grant #354917 and the National Science Foundation under grant DMS 1736470. Wang's research was supported in part by the National Science Foundation under grants DMS 1106816 and DMS 1542332. The MRI/DTI data were collected at Johns Hopkins University and the Kennedy-Krieger Institute. We thank the Editor-in-Chief, the Associate Editor and anonymous reviewers for their helpful and constructive comments, which led to significant improvement in this paper.

Appendix

Let κ be the number of eigenfunctions and κ can be a positive integer or infinity. We use C to denote a generic positive constant unless otherwise stated. For any vector $\mathbf{a} = (a_1, \dots, a_k) \in \mathbb{R}^k$, set $\|\mathbf{a}\|_\infty = \max(|a_1|, \dots, |a_k|)$ and $\|\mathbf{a}\|_r = (|a_1|^r + \dots + |a_k|^r)^{1/r}$ for all $r \in [1, \infty)$. For any function ϕ on $[0, 1]$, denote $\|\phi\|_\infty = \sup_{s \in [0, 1]} |\phi(s)|$. For

functions $\phi, \varphi \in L_2[0, 1]$, write $\langle \phi, \varphi \rangle_{2,N} = \sum_{\ell=1}^N \phi(\ell/N) \varphi(\ell/N) / N$ the empirical inner product with corresponding norm $\|\phi\|_{2,N}^2 = \langle \phi, \phi \rangle_{2,N}$.

A.1. Assumptions

In this paper we restrict our attention to splines with equally spaced knots. Denote $\omega_j = jh_\mu$, $0 \leq j \leq N_\mu$, and let $h_\mu = 1/(N_\mu + 1)$ be the distance between neighboring knots.

For any $v \in (0, 1]$ and nonnegative integer q , let $C^{q,v}[0, 1]$ be the space of functions with v -Hölder continuous q th order derivatives on $[0, 1]$, i.e.,

$$C^{q,v}[0, 1] = \left\{ \phi : \|\phi\|_{q,v} = \sup_{t \neq s, t, s \in [0, 1]} |t - s|^{-v} |\phi^{(q)}(t) - \phi^{(q)}(s)| < \infty \right\}.$$

The technical assumptions we need are as follows:

- (A1) The regression function μ belongs to $C^{p-1,1}[0, 1]$, i.e., $\mu^{(p-1)} \in C^{0,1}[0, 1]$.
- (A2) If $\Gamma_{jj'}(s)$ is the (j, j') th entry of $\Gamma(s)$, then $\Gamma_{jj'}(s) \in C^{0,v}[0, 1]$ for some $v \in (0, 1]$.
- (A3) As $n \rightarrow \infty$, $N^{-1}n^{1/(2p)} \rightarrow 0$ and $N = O(n^\theta)$ for some $\theta > 1/(2p)$; the number of interior knots N_μ satisfies $N_\mu^{-1}N \rightarrow \infty$, $N_\mu^{-p}n^{1/2} \rightarrow 0$, $N^{-1/2}N_\mu^{1/2} \ln(n) \rightarrow 0$.
- (A4) There exists $C_G > 0$ such that, for $j, j' \in \{1, \dots, m\}$, $G_{jj'}(s, s) \geq C_G$ for all $s \in [0, 1]$; for $k \in \{1, \dots, \kappa\}$ and $j \in \{1, \dots, m\}$, $\phi_{jk}(s) \in C^{0,v}[0, 1]$, $\sum_{k=1}^\kappa \|\phi_{jk}\|_\infty < \infty$ and as $n \rightarrow \infty$, $h_\mu^\nu \sum_{k=1}^{\kappa_n} \|\phi_{jk}\|_{0,v} = o(1)$ for a sequence $\{\kappa_n\}_{n=1}^\infty$ of increasing integers, with $\kappa_n \rightarrow \kappa$ as $n \rightarrow \infty$ and the constant $v \in (0, 1]$ as in Assumption (A2). In particular, $\sum_{k=\kappa_n+1}^\kappa \|\phi_{jk}\|_\infty = o(1)$ for all $j \in \{1, \dots, m\}$. The eigenvalue sequence $\{\lambda_{jk}\}_{j=1, k=1}^{m, \infty}$ has a finite sum, i.e., $\sum_{k=1}^\infty \lambda_{jk} < \infty$.
- (A5) There exist $\eta_1, \eta_2 > 4$, such that $E|\xi_{ijk}|^{\eta_1} + E|\varepsilon_{ij\ell}|^{\eta_2} < \infty$ for all $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m\}$, $k \in \{1, \dots, \kappa\}$, and $\ell \in \{1, \dots, N\}$. The number κ of nonzero eigenvalues is finite or κ is infinite while the random vectors $\xi_{ik} = (\xi_{i1k}, \dots, \xi_{imk})^\top$ are iid for $i \in \mathbb{N}$ and $k \in \{1, \dots, \kappa\}$.

Assumptions (A1)–(A2) are typical for spline smoothing; see [4,5,21]. Assumption (A3) concerns the relationship among the number of subjects, the number of observations for each subject and the number of knots of B-splines. Assumption (A4) ensures that the principal components have collectively bounded smoothness. Assumption (A5) is necessary for applying Gaussian approximation of the error process.

A.2. Error decomposition for spline estimators

In this section, we break the estimation error $\hat{\mu}(s) - \mu(s)$ into three terms. We begin by discussing the representation of the spline estimator $\hat{\mu}(s)$ in (4). Let $\mathbf{B}(s) = (B_{1-p}(s), \dots, B_{N_\mu}(s))^\top$, $\mathbf{X}_{N \times (N_\mu+p)} = (\mathbf{B}(1/N), \dots, \mathbf{B}(N/N))^\top$ and

$$\hat{\mathbf{V}} = \mathbf{X}^\top \mathbf{X} / N = (\langle B_j, B_{j'} \rangle_{2,N})_{j,j'=1-p}^{N_\mu}.$$

Applying elementary algebra to the spline estimator $\hat{\mu}(s)$ defined in (4), one finds

$$\hat{\mu}(s) = N^{-1} \mathbf{B}(s) \hat{\mathbf{V}}^{-1} \mathbf{X}^\top \mathbf{Y}, \quad (\text{A.1})$$

where $\mathbf{Y} = (\bar{Y}_{\cdot,1}, \dots, \bar{Y}_{\cdot,N})^\top$ and $\bar{Y}_{\cdot,\ell} = \sum_{i=1}^n \sum_{j=1}^m Y_{ij\ell} / (nm)$ for all $\ell \in \{1, \dots, N\}$.

Projecting the relationship in model (1) onto the linear subspace of $\mathbb{R}^{N_\mu+p}$ spanned by $\{\mathbf{B}(\ell/N)\}_{\ell=1}^N$, we obtain the following decomposition:

$$\hat{\mu}(s) = \tilde{\mu}(s) + \tilde{e}(s) + \tilde{\xi}(s), \quad (\text{A.2})$$

where

$$\tilde{\mu}(s) = \sum_{j=1-p}^{N_\mu} \tilde{\beta}_j B_j(s), \quad \tilde{e}(s) = \sum_{j=1-p}^{N_\mu} \tilde{a}_j B_j(s), \quad \tilde{\xi}(s) = \sum_{k=1}^\kappa \tilde{\xi}_k(s), \quad \tilde{\xi}_k(s) = \sum_{j=1-p}^{N_\mu} \tilde{\tau}_{kj} B_j(s), \quad (\text{A.3})$$

and vectors $(\tilde{\beta}_{1-p}, \dots, \tilde{\beta}_{N_\mu})^\top$, $(\tilde{a}_{1-p}, \dots, \tilde{a}_{N_\mu})^\top$ and $(\tilde{\tau}_{k,1-p}, \dots, \tilde{\tau}_{k,N_\mu})^\top$ in (A.3) are solutions to (3) with $Y_{ij\ell}$ replaced by $\mu(\ell/N)$, $\varepsilon_{ij}(\ell/N)$ and $\phi_{jk}(\ell/N)$, respectively.

Alternatively, $\tilde{\mu}(s) = N^{-1} \mathbf{B}(s) \hat{\mathbf{V}}^{-1} \mathbf{X}^\top \boldsymbol{\mu}$, $\tilde{e}(s) = \mathbf{B}(s) \hat{\mathbf{V}}^{-1} \mathbf{X}^\top \mathbf{e} / N$, $\tilde{\xi}_k(s) = \mathbf{B}(s) \hat{\mathbf{V}}^{-1} \mathbf{X}^\top \boldsymbol{\phi}_k / N$, for all $k \in \{1, \dots, \kappa\}$, where $\boldsymbol{\mu} = (\mu(1/N), \dots, \mu(N/N))^\top$ is the signal vector, $\mathbf{e} = (\bar{\varepsilon}_{\cdot,1}, \dots, \bar{\varepsilon}_{\cdot,N})^\top$ where, for all $\ell \in \{1, \dots, N\}$,

$$\bar{\varepsilon}_{\cdot,\ell} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \varepsilon_{ij}(\ell/N)$$

is the noise vector, and vector $\boldsymbol{\phi}_k = \sum_{i=1}^n \sum_{j=1}^m \xi_{ijk} \boldsymbol{\phi}_{jk} / (nm)$ with $\boldsymbol{\phi}_{jk} = (\phi_{jk}(1/N), \dots, \phi_{jk}(N/N))^\top$.

The next three lemmas concern $\tilde{\mu}(s)$, $\tilde{e}(s)$ and $\tilde{\xi}(s)$ given in (A.2).

Lemma A.1. Under Assumptions (A1) and (A3), $\sup_{s \in [0,1]} n^{1/2} |\tilde{\mu}(s) - \mu(s)| = o(1)$.

Proof. Applying Lemma A.4, $\|\tilde{\mu} - \mu\|_\infty \leq C_{p-1,1} h_\mu^p$. Since Assumption (A3) implies that $O(h_\mu^p n^{1/2}) = o(1)$, the desired result is obtained. $\square$

Lemma A.2. Under Assumptions (A2)–(A5), $\sup_{s \in [0,1]} n^{1/2} |\tilde{e}(s)| = o_P(1)$.

Proof. Let $\tilde{Z}_\varepsilon(s) = \mathbf{B}(s) \hat{\mathbf{V}}^{-1} \mathbf{X}^\top \mathbf{Z} / N$, where $\mathbf{Z} = (\bar{Z}_{\cdot,1,\varepsilon}, \dots, \bar{Z}_{\cdot,N,\varepsilon})^\top$ with $\bar{Z}_{\cdot,\ell,\varepsilon} = \sum_{i=1}^n \sum_{j=1}^m Z_{ij\ell,\varepsilon} / (nm)$. By (A.9), $\|\mathbf{Z} - \mathbf{e}\|_\infty = O_{a.s.}(n^{\beta-1})$, while

$$\begin{aligned} \|N^{-1} \mathbf{X}^\top (\mathbf{Z} - \mathbf{e})\|_\infty &\leq \|\mathbf{Z} - \mathbf{e}\|_\infty \max_{1 \leq j \leq N_\mu} \langle B_j, 1 \rangle_{2,N} \\ &\leq C \|\mathbf{Z} - \mathbf{e}\|_\infty \max_{1 \leq j \leq N_\mu} \#\{\ell : B_j(\ell/N) > 0\} N^{-1} \leq C \|\mathbf{Z} - \mathbf{e}\|_\infty h_\mu. \end{aligned}$$

Note that $\|\tilde{Z}_\varepsilon - \tilde{e}\|_\infty = \sup_{s \in [0,1]} \|\mathbf{B}(s) \hat{\mathbf{V}}^{-1} N^{-1} \mathbf{X}^\top (\mathbf{Z} - \mathbf{e})\|_\infty$. According to Lemma A.3 in Cao et al. [5], $\|\hat{\mathbf{V}}^{-1}\|_\infty = O_{a.s.}(N_\mu)$. Thus, $\|\tilde{Z}_\varepsilon - \tilde{e}\|_\infty \leq C \|\hat{\mathbf{V}}^{-1}\|_\infty \|\mathbf{Z} - \mathbf{e}\|_\infty h_\mu = O_{a.s.}(n^{\beta-1})$. Observe next that $\hat{\mathbf{V}}^{-1} N^{-1} \mathbf{X}^\top \mathbf{Z}$ is $(N_\mu + p)$ -dimensional normal with covariance matrix

$$N^{-2} \hat{\mathbf{V}}^{-1} \mathbf{X}^\top \text{var}(\mathbf{Z}) \mathbf{X} \hat{\mathbf{V}}^{-1} = N^{-2} \hat{\mathbf{V}}^{-1} \mathbf{X}^\top \text{diag}\{\text{var}(\bar{Z}_{\cdot,1,\varepsilon}), \dots, \text{var}(\bar{Z}_{\cdot,N,\varepsilon})\} \mathbf{X} \hat{\mathbf{V}}^{-1},$$

bounded above by $\max_{1 \leq \ell \leq N} \text{var}(\bar{Z}_{\cdot,\ell,\varepsilon}) \|N^{-1} N^{-1} \hat{\mathbf{V}}^{-1} \hat{\mathbf{V}}^{-1}\|_\infty \leq C / (N n h_\mu)$.

To bound the tail probabilities of entries of $\hat{\mathbf{V}}^{-1} N^{-1} \mathbf{X}^\top \mathbf{Z}$, applying the Borel–Cantelli Lemma, one obtains $\|\hat{\mathbf{V}}^{-1} N^{-1} \mathbf{X}^\top \mathbf{Z}\|_\infty = O_{a.s.}\{(N n h_\mu / \ln n)^{-1/2}\}$. Hence, we have $\|n^{1/2} \tilde{Z}_\varepsilon\|_\infty = O_{a.s.}(N^{-1/2} h_\mu^{-1/2} \ln^{1/2} n)$ and $\|n^{1/2} \tilde{e}\|_\infty = O_{a.s.}(n^{\beta-1/2} + N^{-1/2} h_\mu^{-1/2} \ln^{1/2} n) = o_{a.s.}(1)$. Thus, Lemma A.2 follows from Assumption (A3). $\square$

Lemma A.3. Under Assumptions (A2)–(A5), $\sup_{s \in [0,1]} n^{1/2} |\tilde{\xi}(s) - \bar{\mu}(s) - \mu(s)| = o_P(1)$ and, for all $\alpha \in (0, 1)$,

$$\Pr \left\{ \sup_{s \in [0,1]} n^{1/2} |\tilde{\xi}(s)| \bar{G}(s, s)^{-1/2} \leq Q_{1-\alpha} \right\} \rightarrow 1 - \alpha.$$

Proof. Let $\tilde{\zeta}_k(s) = n^{1/2} (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m Z_{ijk,\varepsilon} \phi_{jk}(s)$ for all $k \in \{1, \dots, \kappa\}$, and define

$$\tilde{\zeta}(s) = \left\{ \frac{1}{nm^2} \sum_{k,k'=1}^\kappa \sum_{i=1}^n \sum_{j,j'=1}^m E(\xi_{ijk,\varepsilon} \xi_{ij'k',\varepsilon}) \phi_{jk}(s) \phi_{j'k'}(s) \right\}^{-1/2} \sum_{k=1}^\kappa \tilde{\zeta}_k(s) = \bar{G}(s, s)^{-1/2} \sum_{k=1}^\kappa \tilde{\zeta}_k(s).$$

It is clear that $\tilde{\zeta}(s)$ is a Gaussian process with mean 0, variance 1 and covariance function given, for all $s, t \in [0, 1]$, by

$$E\tilde{\zeta}(s)\tilde{\zeta}(t) = \bar{G}(s, s)^{-1/2} \bar{G}(t, t)^{-1/2} \bar{G}(s, t).$$

Thus, $\tilde{\zeta}(s)$ has the same distribution as $\zeta(s)$ over $s \in [0, 1]$.

According to Lemma A.4, $\|\tilde{\phi}_{jk}\|_\infty \leq c_{\phi,p} \|\phi_{jk}\|_\infty$, $\|\tilde{\phi}_{jk} - \phi_{jk}\|_\infty \leq \tilde{C}_{0,v} \|\phi_{jk}\|_{0,v} h_\mu^v$ for all $k \in \{1, \dots, \kappa\}$ and $j \in \{1, \dots, m\}$. Applying (A.10) and Assumptions (A3), (A4), one finds

$$\begin{aligned} E \left[n^{1/2} \sup_{s \in [0,1]} \bar{G}(s, s)^{-1/2} \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^\kappa \xi_{ijk,\varepsilon} \{\phi_{jk}(s) - \tilde{\phi}_{jk}(s)\} \right| \right] \\ \leq C n^{1/2} \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left(\sum_{k=1}^{\kappa_n} E|\xi_{ijk,\varepsilon}| \|\phi_{jk}\|_{0,v} h_\mu^v + \sum_{k=\kappa_n+1}^\kappa E|\xi_{ijk,\varepsilon}| \|\phi_{jk}\|_\infty \right) \\ \leq C \left(\sum_{j=1}^m \sum_{k=1}^{\kappa_n} \|\phi_{jk}\|_{0,v} h_\mu^v + \sum_{j=1}^m \sum_{k=\kappa_n+1}^\kappa \|\phi_{jk}\|_\infty \right) = o(1). \end{aligned}$$

Hence,

$$n^{1/2} \sup_{s \in [0,1]} \bar{G}(s, s)^{-1/2} \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^\kappa \xi_{ijk,\varepsilon} \{\phi_{jk}(s) - \tilde{\phi}_{jk}(s)\} \right| = o_P(1). \quad (\text{A.4})$$

In addition, (A.8) and Assumptions (A3), (A4) entail that

$$\mathbb{E} n^{1/2} \sup_{s \in [0,1]} \bar{G}(s, s)^{-1/2} \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{\kappa} (Z_{ijk, \xi} - \xi_{ijk}) \phi_{jk}(s) \right| = C n^{\beta-1/2} \sum_{j=1}^m \sum_{k=1}^{\kappa} \|\phi_{jk}\|_{\infty} = o(1).$$

Therefore,

$$n^{1/2} \sup_{s \in [0,1]} \bar{G}(s, s)^{-1/2} \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{\kappa} (Z_{ijk, \xi} - \xi_{ijk}) \phi_{jk}(s) \right| = o_P(1). \quad (\text{A.5})$$

Note that $\bar{\mu}(s) - \mu(s) - \tilde{\xi}(s) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{\kappa} \xi_{ijk} \{\phi_{jk}(s) - \tilde{\phi}_{jk}(s)\} / (nm)$,

$$n^{-1/2} \bar{G}(s, s)^{1/2} \tilde{\zeta}(s) - \{\bar{\mu}(s) - \mu(s)\} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{\kappa} (Z_{ijk, \xi} - \xi_{ijk}) \phi_{jk}(s).$$

Hence, according to (A.4) and (A.5),

$$\begin{aligned} \sqrt{n} \sup_{s \in [0,1]} \bar{G}(s, s)^{-1/2} |\bar{\mu}(s) - \mu(s) - \tilde{\xi}(s)| &= o_P(1), \\ \sup_{s \in [0,1]} |\tilde{\zeta}(s) - n^{1/2} \bar{G}(s, s)^{-1/2} \{\bar{\mu}(s) - \mu(s)\}| &= o_P(1), \end{aligned}$$

which leads to the desired results. $\square$

Proof of Theorem 1. Theorem 1 follows directly from the decomposition in (A.2) and Lemmas A.1–A.3.

Proof of Theorem 3. Theorem 3 can be derived directly from Lemmas A.9 and A.10.

A.3. Technical lemmas

This section contains some technical details used in the proofs of Lemmas A.1–A.3.

Lemma A.4 ([8], p. 149). *There is an absolute constant $C_{p-1, \nu} > 0$, $\nu \in (0, 1]$, such that for every $\phi \in C^{p-1, \nu}[0, 1]$, there exists a function $g \in \mathcal{H}^{(p-2)}[0, 1]$ satisfying $\|g - \phi\|_{\infty} \leq C_{p-1, \nu} \|\phi^{(p-1)}\|_{0, \mu} h_{\mu}^{\nu+p-1}$.*

Lemma A.5 ([10], Theorem 4). *Let $H : [0, \infty) \rightarrow [0, \infty)$ be a continuous function such that the function $x^{-3-\gamma} H(x)$ is nondecreasing for some $\gamma > 0$ and $x^{-1/2} \ln\{H(x)\}$ is nonincreasing. Suppose that $\xi : \Omega \rightarrow \mathcal{R}^m$ is a random vector defined in p -space $(\Omega, \mathcal{A}, \Pr)$ with $\mathbb{E}(\xi) = 0$, $\text{cov}(\xi) = \Sigma$. Assume $\mathbb{E}\{H(\|\xi\|_2)\} < \infty$. Then one can construct a p -space $(\Omega_0, \mathcal{A}_0, \Pr_0)$ and two sequences of independent random vectors $\{\xi_i\}$ and $\{\mathbf{W}_i\}$ with $\Pr_0 \circ \xi_i = \Pr \circ \xi$, $\Pr_0 \circ \mathbf{W}_i = \mathcal{N}(\mathbf{0}, \Sigma)$ for all $i \in \{1, \dots, n\}$, such that for any $z > 0$,*

$$\Pr_0 \left(\max_{1 \leq u \leq n} \left\| \sum_{i=1}^u \xi_i - \sum_{i=1}^u \mathbf{W}_i \right\|_2 > z \right) \leq \frac{C_1 n}{H(C_2 z)},$$

where positive constants C_1, C_2 depend on the distribution of ξ .

Let $\xi_{ik} = (\xi_{i1k}, \dots, \xi_{imk})^\top$ and $\varepsilon_{i\ell} = (\varepsilon_{i1\ell}, \dots, \varepsilon_{im\ell})^\top$ for all $i \in \{1, \dots, n\}$, $k \in \{1, \dots, \kappa\}$, and $\ell \in \{1, \dots, N\}$.

Lemma A.6. *Under Assumption (A5), there are constants $C_1, C_2 > 0$, $\gamma_1, \gamma_2 > 1$, $\beta \in (0, 1/2)$ and independent Gaussian random $\mathcal{R}^m$ -valued vectors $\mathbf{Z}_{ik, \xi} = (Z_{i1k, \xi}, \dots, Z_{imk, \xi})^\top$, $\mathbf{Z}_{i\ell, \varepsilon} = (Z_{i1\ell, \varepsilon}, \dots, Z_{im\ell, \varepsilon})^\top$ for all $i \in \{1, \dots, n\}$, $k \in \{1, \dots, \kappa\}$, and $\ell \in \{1, \dots, N\}$, such that $\mathbb{E}(\mathbf{Z}_{ik, \xi}) = \mathbf{0}$, $\text{cov}(\mathbf{Z}_{ik, \xi}) = \text{cov}(\xi_{ik})$, $\mathbb{E}(\mathbf{Z}_{i\ell, \varepsilon}) = \mathbf{0}$, $\text{cov}(\mathbf{Z}_{i\ell, \varepsilon}) = \text{cov}(\varepsilon_{i\ell})$ and*

$$\max_{1 \leq k \leq \kappa} \Pr \left(\max_{1 \leq u \leq n} \left\| \sum_{i=1}^u \xi_{ik} - \sum_{i=1}^u \mathbf{Z}_{ik, \xi} \right\|_2 > C_1 n^\beta \right) < C_2 n^{-\gamma_1}, \quad (\text{A.6})$$

$$\Pr \left(\max_{1 \leq \ell \leq N} \max_{1 \leq u \leq n} \left\| \sum_{i=1}^u \varepsilon_{i\ell} - \sum_{i=1}^u \mathbf{Z}_{i\ell, \varepsilon} \right\|_2 > C_1 n^\beta \right) < C_2 n^{-\gamma_2}. \quad (\text{A.7})$$

Proof. Under Assumption (A5), $\mathbb{E}|\xi_{ijk}|^{\eta_1} < +\infty$, $\eta_1 > 4$, $\mathbb{E}|\varepsilon_{ij\ell}|^{\eta_2} < +\infty$, $\eta_2 > 4$, so there exists some $\beta \in (0, 1/2)$ such that $\eta_1, \eta_2 > 2/\beta$. Now let $H(x) = x^{\eta_1}$, then $\mathbb{E}\{H(\|\xi_{1k}\|_2)\} < \infty$. One constructs a p -space $(\Omega_0, \mathcal{A}_0, \Pr_0)$ and two sequences of independent random vectors $\{\check{\xi}_{ik}\}$ and $\{\check{\mathbf{Z}}_{ik, \xi}\}$ with $\Pr_0 \circ \check{\xi}_{ik} = P \circ \xi_{ik}$, $\Pr_0 \circ \check{\mathbf{Z}}_{ik, \xi} = \mathcal{N}[\mathbf{0}, \text{cov}(\xi_{1k})]$ for all $i \in \{1, \dots, n\}$

and $k \in \{1, \dots, \kappa\}$. One also defines random vectors $\{\mathbf{Z}_{ik,\xi}\}$ with $\Pr \circ \mathbf{Z}_{ik,\xi} = \Pr_0 \circ \check{\mathbf{Z}}_{ik,\xi} = \mathcal{N}[\mathbf{0}, \text{cov}(\xi_{1k})]$ and $\{\mathbf{Z}_{ik,\xi}\}$ is independent from $\{\xi_{ik}\}$, for all $i \in \{1, \dots, n\}$ and $k \in \{1, \dots, \kappa\}$. Lemma A.5 entails that there exist constants C_{1k} and C_{2k} which depend on the distribution of ξ_{1k} , such that for $z = C_{1k}n^\beta$,

$$\Pr \left(\max_{1 \leq u \leq n} \left\| \sum_{i=1}^u \xi_{ik} - \sum_{i=1}^u \mathbf{Z}_{ik,\xi} \right\|_2 > C_{1k}n^\beta \right) = \Pr_0 \left(\max_{1 \leq u \leq n} \left\| \sum_{i=1}^u \check{\xi}_{ik} - \sum_{i=1}^u \check{\mathbf{Z}}_{ik,\xi} \right\|_2 > C_{1k}n^\beta \right) < C_{2k}n^{1-\eta_1\beta}.$$

Note that $\eta_1 > 2/\beta$, $\gamma_1 = \eta_1\beta - 1 > 1$. If κ is finite, then there are common constants $C_1, C_2 > 0$ such that

$$\Pr \left(\max_{1 \leq u \leq n} \left\| \sum_{i=1}^u \xi_{ik} - \sum_{i=1}^u \mathbf{Z}_{ik,\xi} \right\|_2 > C_{1k}n^\beta \right) < C_2n^{-\gamma_1},$$

which entails (A.6) because κ is finite. If κ is infinite but all the ξ_{ik} s are iid, then C_{1k}, C_{2k}, a_k are the same for all k , so the above is again true.

Likewise, under Assumption (A5), if one lets $H(x) = x^{\eta_2}$, then $E\{H(\|\varepsilon_{1\ell}\|_2)\} < \infty$. One can construct a p -space $(\Omega_0, \mathcal{A}_0, \Pr_0)$ and two sequences of independent random vectors $\{\check{\varepsilon}_{i\ell}\}$ and $\{\check{\mathbf{Z}}_{i\ell,\varepsilon}\}$, with $\Pr_0 \circ \check{\varepsilon}_{i\ell} = \Pr_0 \circ \varepsilon_{1\ell}$, $\Pr_0 \circ \check{\mathbf{Z}}_{i\ell,\varepsilon} = \mathcal{N}[\mathbf{0}, \text{cov}(\varepsilon_{1\ell})]$ for all $i \in \{1, \dots, n\}$ and $\ell \in \{1, \dots, N\}$. One also define random vectors $\{\mathbf{Z}_{i\ell,\varepsilon}\}$ with $\Pr \circ \mathbf{Z}_{i\ell,\varepsilon} = \Pr_0 \circ \check{\mathbf{Z}}_{i\ell,\varepsilon} = \mathcal{N}[\mathbf{0}, \text{cov}(\varepsilon_{1\ell})]$ and $\{\mathbf{Z}_{i\ell,\varepsilon}\}$ is independent from $\{\varepsilon_{i\ell}\}$, for all $i \in \{1, \dots, n\}$ and $\ell \in \{1, \dots, N\}$. Lemma A.5 entails that there exist constants C_1 and C_2 , which depend on the distribution of $\varepsilon_{1\ell}$, such that for $z = C_1n^\beta$,

$$\max_{1 \leq \ell \leq N} \Pr \left(\max_{1 \leq u \leq n} \left\| \sum_{i=1}^u \varepsilon_{i\ell} - \sum_{i=1}^u \mathbf{Z}_{i\ell,\varepsilon} \right\|_2 > C_1n^\beta \right) = \max_{1 \leq \ell \leq N} \Pr_0 \left(\max_{1 \leq u \leq n} \left\| \sum_{i=1}^u \check{\varepsilon}_{i\ell} - \sum_{i=1}^u \check{\mathbf{Z}}_{i\ell,\varepsilon} \right\|_2 > C_1n^\beta \right) \leq C_2n^{1-\eta_2\beta},$$

now $\eta_2\beta > 2$ implies that there is $\gamma_2 > 1$ such that $\eta_2\beta - 1 > \gamma_2$ and (A.7) follows. $\square$

Lemma A.7. In Lemma A.6, for $C_0 = C_1(1 + \beta C_2 \sum_{s=1}^{\infty} s^{\beta-1-\gamma_1})$, one has

$$\max_{1 \leq k \leq \kappa} E \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (\xi_{ijk} - Z_{ijk,\xi}) \right| = O(n^{\beta-1}), \quad (\text{A.8})$$

$$\max_{1 \leq \ell \leq N} E \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (\varepsilon_{ij\ell} - Z_{ij\ell,\varepsilon}) \right| = O_{a.s.}(n^{\beta-1}). \quad (\text{A.9})$$

Also

$$\max_{1 \leq k \leq \kappa} \frac{1}{nm} \sum_{j=1}^m \sum_{i=1}^n E |\xi_{ijk}| = O(n^{-1/2} + n^{\beta-1}). \quad (\text{A.10})$$

Proof. Under Lemma A.6, the proof of (A.9) is trivial. Lemma A.6 entails that $\bar{F}_{n+u,k} < C_2(n+u)^{-\gamma_1}$ for all $k \in \{1, \dots, \kappa\}$ and $t \in \{0, 1, \dots\}$, in which

$$\bar{F}_{n+u,k} = \Pr \left\{ \left\| \sum_{i=1}^n \xi_{ik} - \sum_{i=1}^n \mathbf{Z}_{ik,\xi} \right\|_2 > C_1(n+u)^\beta \right\},$$

where Gaussian random $\mathbb{R}^m$ -valued vector $\mathbf{Z}_{ik,\xi} = (Z_{i1k,\xi}, \dots, Z_{imk,\xi})^\top$. Taking expectation, one has

$$\begin{aligned} E \left\| \sum_{i=1}^n \xi_{ik} - \sum_{i=1}^n \mathbf{Z}_{ik,\xi} \right\|_2 &\leq C_1(n+0)^\beta + \sum_{u=1}^{\infty} C_1(n+u)^\beta (\bar{F}_{n+u-1,k} - \bar{F}_{n+u,k}) \\ &\leq C_1n^\beta + \sum_{t=0}^{\infty} C_1C_2(n+u)^{-\gamma_1} \beta(n+u)^{\beta-1} \leq n^\beta C_1 \left\{ 1 + \beta C_2 n^{1-\gamma_1} \sum_{v=1}^{\infty} \sum_{u=vn-n}^{vn-1} (1+u/n)^{\beta-1-\gamma_1} \right\} \\ &\leq n^\beta C_1 \left(1 + \beta C_2 n^{1-\gamma_1} \times n \sum_{u=1}^{\infty} u^{\beta-1-\gamma_1} \right) \leq Cn^\beta, \end{aligned}$$

which proves $E|\sum_{i=1}^n (\xi_{ijk} - Z_{ijk,\xi})| = n^\beta$ for all $j \in \{1, \dots, m\}$ and $k \in \{1, \dots, \kappa\}$. Hence, $E|\sum_{i=1}^n (\xi_{ijk} - Z_{ijk,\xi})/(nm)| = O(n^{\beta-1})$ for all $j \in \{1, \dots, m\}$ and $k \in \{1, \dots, \kappa\}$, which entails (A.8).

Observe that $\max_{1 \leq j \leq m} C(nm)^{-2} \text{var}(\sum_{i=1}^n Z_{ijk,\xi}) = O(n^{-1})$. Therefore, there exists a positive constant C_m such that $\sum_{i=1}^n E|Z_{ijk,\xi}|/(nm) = C_m n^{-1/2}$ for all $j \in \{1, \dots, m\}$, which entails (A.10). $\square$

Lemma A.8. Under Assumptions (A1)–(A5), for any $j, j' \in \{1, \dots, m\}$ and $k, k' \in \{1, \dots, \kappa\}$, one has

$$|\hat{\lambda}_{jk} - \lambda_{jk}| = o_p(1), \quad \|\hat{\phi}_{jk}\hat{\phi}_{j'k'} - \phi_{jk}\phi_{j'k'}\|_\infty = o_p(1). \quad (\text{A.11})$$

Proof. Given each $j \in \{1, \dots, m\}$ and $k \in \{1, \dots, \kappa\}$, by Lemma A8 in Cao et al. [4] one has $|\hat{\lambda}_{jk} - \lambda_{jk}| = o_p(1)$ and $\|\hat{\phi}_{jk} - \phi_{jk}\|_\infty = o_p(1)$. Hence, $\|\hat{\phi}_{jk}\hat{\phi}_{j'k'} - \phi_{jk}\phi_{j'k'}\|_\infty \leq \|\hat{\phi}_{jk} - \phi_{jk}\|_\infty \|\phi_{j'k'}\|_\infty + \|\hat{\phi}_{j'k'} - \phi_{j'k'}\|_\infty \|\phi_{jk}\|_\infty = o_p(1)$. $\square$

Lemma A.9. Under Assumptions (A1)–(A5), for any $j \in \{1, \dots, m\}$, $\sup_{(s,t) \in [0,1]^2} |\hat{G}_{jj}(s,t) - G_{jj}(s,t)| = o_p(1)$.

Proof. We first define “oracle” smoother $\tilde{G}_{jj}(s,t) = \sum_{J,J'=1-p}^{N_G} \tilde{b}_{JJ'} B_J(s) B_{J'}(t)$, where N_G is the number of interior knots used to build the tensor product B-spline basis and the spline coefficients

$$\{\tilde{b}_{JJ'}\}_{J,J'=1-p}^{N_G} = \arg \min_{\mathbb{R}^{N_G+p} \otimes \mathbb{R}^{N_G+p}} \sum_{\ell \neq \ell'}^N \left\{ \tilde{C}_{\ell\ell'} - \sum_{1-p \leq J' \leq N_G} b_{JJ'} B_J(\ell/N) B_{J'}(\ell'/N) \right\}^2,$$

and

$$\tilde{C}_{\ell\ell'} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \{Y_{ij\ell} - \mu(\ell/N)\} \{Y_{ij\ell'} - \mu(\ell'/N)\}$$

for all $\ell, \ell' \in \{1, \dots, N\}$ such that $\ell \neq \ell'$. According to Proposition 2 in [3], we have that the difference between the tensor product spline estimator $\hat{G}_{jj}(s,t)$ and the “oracle” smoother is uniformly bounded at an $o_p(n^{-1/2})$ rate, i.e., $\sup_{(s,t) \in [0,1]^2} |\hat{G}_{jj}(s,t) - \tilde{G}_{jj}(s,t)| = o_p(n^{-1/2})$, for any $j \in \{1, \dots, m\}$. Hence, the lemma is proved. $\square$

Lemma A.10. Under Assumptions (A1)–(A5), for any $1 \leq j \neq j' \leq m$, $\sup_{(s,t) \in [0,1]^2} |\hat{G}_{jj'}(s,t) - G_{jj'}(s,t)| = o_p(1)$.

Proof. According to the definitions of $G_{jj'}(s,t)$ and $\hat{G}_{jj'}$, one has

$$\begin{aligned} \sup_{(s,t) \in [0,1]^2} |G_{jj'}(s,t) - \hat{G}_{jj'}(s,t)| &\leq \sup_{(s,t) \in [0,1]^2} \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \left| E(\xi_{1jk} \xi_{1j'k'}) - \frac{1}{n} \sum_{i=1}^n \hat{\xi}_{ijk} \hat{\xi}_{ij'k'} \right| \phi_{jk}(s) \phi_{j'k'}(t) \\ &\quad + \sup_{(s,t) \in [0,1]^2} \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} |E(\xi_{1jk} \xi_{1j'k'})| \times |\hat{\phi}_{jk}(s) \hat{\phi}_{j'k'}(t) - \phi_{jk}(s) \phi_{j'k'}(t)| \\ &\leq \max_{k,k'} \left| E(\xi_{1jk} \xi_{1j'k'}) - \frac{1}{n} \sum_{i=1}^n \hat{\xi}_{ijk} \hat{\xi}_{ij'k'} \right| \sum_{k,k'=1}^{\infty} \|\phi_{jk}\|_\infty \|\phi_{j'k'}\|_\infty + \max_{k,k'} E(\xi_{1jk} \xi_{1j'k'}) \sum_{k,k'=1}^{\infty} \|\hat{\phi}_{jk} \hat{\phi}_{j'k'} - \phi_{jk} \phi_{j'k'}\|_\infty. \end{aligned}$$

By Assumption (A4), $\sum_{k,k'=1}^{\infty} \|\phi_{jk}\|_\infty \|\phi_{j'k'}\|_\infty < \infty$, and Lemma A.8 leads to $\sum_{k,k'=1}^{\infty} \|\hat{\phi}_{jk} \hat{\phi}_{j'k'} - \phi_{jk} \phi_{j'k'}\|_\infty = o_p(1)$. Noting that $\hat{\lambda}_{jk} \rightarrow \lambda_{jk}$, $\|\hat{\mu} - \mu\|_\infty = o_p(1)$ and $\|\hat{\phi}_{jk} - \phi_{jk}\|_\infty = o_p(1)$,

$$\hat{\xi}_{ijk} = \frac{1}{N} \sum_{\ell=1}^N \hat{\lambda}_{jk}^{-1} \{Y_{ij\ell} - \hat{\mu}(\ell/N)\} \hat{\phi}_{jk}(\ell/N),$$

one has $|\hat{\xi}_{ijk} - \xi_{ijk}| = o_p(1)$. Hence, $|E(\xi_{1jk} \xi_{1j'k'}) - \sum_{i=1}^n \hat{\xi}_{ijk} \hat{\xi}_{ij'k'} / n| = o_p(1)$. Lemma A.9 is proved. $\square$

References

- [1] P.J. Basser, J. Mattiello, D. LeBihan, MR diffusion tensor spectroscopy and imaging, *Biophys. J.* 66 (1) (1994) 259–267.
- [2] F. Bunea, A.E. Ivanescu, M.H. Wegkamp, Adaptive inference for the mean of a Gaussian process in functional data, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73 (4) (2011) 531–558.
- [3] G. Cao, L. Wang, Y. Li, L. Yang, Oracle-efficient confidence envelopes for covariance functions in dense functional data, *Statist. Sinica* 26 (1) (2016) 359–383.
- [4] G. Cao, J. Wang, L. Wang, D. Todem, Spline confidence bands for functional derivatives, *J. Statist. Plann. Inference* 142 (6) (2012) 1557–1570.
- [5] G. Cao, L. Yang, D. Todem, Simultaneous inference for the mean function based on dense functional data, *J. Nonparametr. Stat.* 24 (2) (2012) 359–377.
- [6] K. Chen, H.-G. Müller, Modeling repeated functional observations, *J. Amer. Statist. Assoc.* 107 (500) (2012) 1599–1609.
- [7] C.M. Crainiceanu, A.-M. Staicu, C.-Z. Di, Generalized multilevel functional regression, *J. Amer. Statist. Assoc.* 104 (488) (2009) 1550–1561.
- [8] C. de Boor, *A Practical Guide to Splines*, revised ed., Springer, New York, 2001.
- [9] C.-Z. Di, C.M. Crainiceanu, B.S. Caffo, N.M. Punjabi, Multilevel functional principal component analysis, *Ann. Appl. Stat.* 3 (1) (2009) 458–488.
- [10] U. Einmahl, Extensions of results of Komlós, Major, and Tusnády to the multivariate case, *J. Multivariate Anal.* 28 (1) (1989) 20–68.
- [11] J. Goldsmith, J. Bobb, C.M. Crainiceanu, B.S. Caffo, D. Reich, Penalized functional regression, *J. Comput. Graph. Statist.* 24 (4) (2011) 830–851.

- [12] J. Goldsmith, C.M. Crainiceanu, B.S. Caffo, D. Reich, Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements, *J. R. Stat. Soc. Ser. C. Appl. Stat.* 61 (3) (2012) 453–469.
- [13] L. Gu, L. Wang, W.K. Härdle, L. Yang, A simultaneous confidence corridor for varying coefficient regression with sparse functional data, *TEST* 23 (4) (2014) 806–843.
- [14] L. Gu, L. Yang, Oracally efficient estimation for single-index link function with simultaneous confidence band, *Electron. J. Stat.* 9 (1) (2015) 1540–1561.
- [15] S. Ma, L. Yang, R.J. Carroll, A simultaneous confidence band for sparse longitudinal regression, *Statist. Sinica* 22 (2012) 95–122.
- [16] J.O. Ramsay, B.W. Silverman, *Functional Data Analysis*, Springer, New York, 2005.
- [17] F. Scheipl, A.-M. Staicu, S. Greven, Functional additive mixed models, *J. Comput. Graph. Statist.* 24 (2) (2015) 477–501.
- [18] Q. Song, R. Liu, Q. Shao, L. Yang, A simultaneous confidence band for dense longitudinal regression, *Comm. Statist. Theory Methods* 43 (24) (2014) 5195–5210.
- [19] J. Wang, F. Cheng, L. Yang, Smooth simultaneous confidence bands for cumulative distribution functions, *J. Nonparametr. Stat.* 25 (2) (2013) 395–407.
- [20] J. Wang, R. Liu, F. Cheng, L. Yang, Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band, *Ann. Statist.* 42 (2) (2014) 654–668.
- [21] J. Wang, L. Yang, Polynomial spline confidence bands for regression curves, *Statist. Sinica* 19 (2009) 325–342.
- [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] H. Zhu, R. Li, L. Kong, Multivariate varying coefficient model for functional responses, *Ann. Statist.* 40 (5) (2012) 2634–2666.