Simultaneous confidence bands for derivatives of dependent functional data

Guanqun Cao

Department of Mathematics and Statistics
Auburn University, e-mail: gzc0009@auburn.edu

Abstract: In this work, consistent estimators and simultaneous confidence bands for the derivatives of mean functions are proposed when curves are repeatedly recorded for each subject. The within-curve correlation of trajectories has been considered while the proposed novel confidence bands still enjoys semiparametric efficiency. The proposed methods lead to a straightforward extension of the two-sample case in which we compare the derivatives of mean functions from two populations. We demonstrate in simulations that the proposed confidence bands are superior to existing approaches which ignore the within-curve dependence. The proposed methods are applied to investigate the derivatives of mortality rates from period lifetables that are repeatedly collected over many years for various countries.

Contents

1. Introduction

Advanced data collection technology has evolved over the last decade so as to permit observations densely sampled over time, space, and other continua. These observed data are often sets of functions represented in the form of curves, images or shapes; they meanwhile demand more powerful and flexible statistical analysis techniques. As a result, functional data analysis (FDA), which addresses emerging issues in the analysis of these complex data, has seen significant development in theory, methods, and computation. We refer to [15] and [11] for monographs on FDA. In the field of FDA, statistical tools are mainly investigated under the situations of independently sampled functions. In recent years, many longitudinal studies are collecting functional measurements at each visit, for example, mortality data [5] in which the age specific lifetables are recorded over years for various countries; longitudinal diffusion tensor imaging (DTI) [20] and repeated white matter tract data [18].

Our objective is to conduct estimation and simultaneous inference of derivatives of repeatedly observed random trajectories. In the fields of engineering and biomedical science, the inference of velocity or acceleration is also highly needed, for example, magnetic gradients and white matter tract. This work is motivated by characterizing the velocity of mean curves for mortality data in [5]. The derivative function reflects the overall trend or direct estimation of an underlying population progress and can be used as an important index for the population response. Hence, it is of particular interest in data analysis to construct simultaneous confidence bands (SCB) for the derivative functions instead of point-wise confidence intervals and to develop global test statistics for the general hypothesis testing problem on the derivative functions. In the previous works, the theoretical focus has mainly been on obtaining consistency and asymptotic normality of the nonparametric estimators, thereby providing the necessary ingredients to construct pointwise confidence intervals for the derivative functions [14]. Most SCB construction studied in the existing FDA literature only considers independent and no repeatedly observed trajectories cases. Particularly, [3] constructed asymptotic SCB for the derivative of mean curves in function data analysis without considering any within-curve dependence. [1, 8] and [4] considered asymptotic SCB for mean functions of the functional regression model when within-curve dependence is not present.

Recently, there have been some attempts to study such dependent functional data in various models. For instance, [9] emphasised a general hierarchical

model. [5] proposed a flexible longitudinally observed functional model and provided consistency results and asymptotic convergence rates for the estimated model components. [20] established the uniform convergence rate and confidence band for each estimated individual effect curve in multivariate varying coefficient models. [18] developed a functional mixed effects modeling framework to delineate the dynamic changes of diffusion properties along major fiber tracts and the structure of the variability of these white matter tract properties in various longitudinal studies. More recently, [2] studies polynomial spline confidence bands for mean curves of dependent functional data. However, we note that all currently available statistical nonparametric methods cannot be immediately used for constructing SCB for derivatives of mean functions when curves are repeatedly recorded.

To develop simultaneous inference for dependent functional responses, we encounter many new challenges. First, the greater technical difficulty to formulate SCB for infinite dimensional functional response and establish their theoretical properties. Second, unlike the scenarios considered in the classical FDA literature where the data consist of n independent units, in out settings, however, there is complex within-subject or spatial-temporal correlation structure. In this work we use polynomial splines to approximate the derivative functions. We show that the proposed spline confidence bands are asymptotically correct and satisfying semiparametric efficiency in the sense that they are asymptotically the same as if all random trajectories are fully recorded and without measurement errors as in [3]. In this context, we further extend the simultaneous inference to the two-sample case and evaluate the equality of derivative functions from two groups. Our Monte Carlo results show that the proposed bands are superior to existing methods which ignore the dependence within the repeatedly observed curves.

The paper is organized as follows. Section 2 states the model and introduces the spline estimates of the derivative functions for dependent functional data. Section 3.1 describes the asymptotic distribution of the spline derivative estimator in the framework of allowing unknown within-curve dependence of the trajectories. Using this asymptotic result, we construct SCB for derivative functions. Section 3.2 develops the confidence bands to study the difference of derivative functions from two populations. The actual steps to implement the confidence bands are provided in Section 4. A simulation study is presented in Section 5. We present applications to age-specific human mortality datasets in Section 6.

Technical proofs are collected in the Appendix.

2. The model and estimates

2.1. Background on dependent functional data

The sample of functions that give rise to the data are viewed as realizations of a smooth and square integrable random process $\{X_{ij}(s), s \in \mathcal{X}\}$ for $i = 1, \ldots, n,$

$j = 1, \ldots m_i$, where $\mathcal{X}$ is a compact interval, i is the subject index, and j is the repeated observation index for the i-th subject. Without loss of generality, let $\mathcal{X} = [0,1]$ and $m_i = m$ for $i = 1, \ldots, n$. If the repeated sizes are not equal, we do transformation of the data, which is discussed in details in Section 4.

The process $X_{ij}(s)$ is assumed to have covariance function $\mathrm{Cov}(X_{ij}(s)$, $X_{ij'}(t)) = G_{jj'}(s,t)$, where $j, j' = 1, \ldots, m$ and $s, t \in [0,1]$ and $\mathbf{G}(s,t) = \{G_{jj'}(s,t)\}_{j,j'=1}^{m}$ is an $m \times m$ matrix of symmetric, positive definite, and continuous function. Therefore, $\mathbf{G}(s,t)$ mainly characterizes within-subject correlation structure. For each (i,j), random process $X_{ij}(s)$ can be viewed as an L_2 process on $[0,1]$, and hence by the Karhunen-Loève expansion,

$$X_{ij}(s) = \mu(s) + \sum_{k=1}^{\infty} \xi_{ijk} \lambda_{jk}^{1/2} \psi_{jk}(s) = \mu(s) + \sum_{k=1}^{\infty} \xi_{ijk} \phi_{jk}(s),$$

where $\mu(s) = E\{X_{ij}(s)\}$, $\psi_{jk}(\cdot)$'s are orthonormal eigenfunctions of the covariance function $G_{jj}(s,t)$ and for each fixed (i,j), the ξ_{ijk}'s are uncorrelated random coefficients with mean 0 and variance 1. The random coefficients ξ_{ijk}'s are also referred to as the (jk)-th functional principal component (FPC) scores of the i-th subject. We assume that $\psi_{jk}(\cdot)$'s are kept in a descending order of λ_{jk}'s, i.e. $\lambda_{j1} \geq \lambda_{j2} \geq \cdots \geq 0$. The number of principal components is ∞ in our theories, but is often assumed to be finite for practical considerations.

In practice, one often observes measurements $Y_{ij}(s) = X_{ij}(s) + \varepsilon_{ij}(s)$, and $\varepsilon_i(s) = (\varepsilon_{i1}(s), \ldots, \varepsilon_{im}(s))^{\mathrm{T}}$ are mean zero measurement errors. Measurement errors $\varepsilon_i(s)$ are mutually independent of $\{X_{ij}(s)\}_{j=1}^{m}$ and $\varepsilon_i(s)$ are i.i.d. copies of stochastic processes with mean vector $\mathbf{0}$ and covariance functions $\mathbf{\Gamma}(s,t)$. Random processes $\varepsilon_i(s)$ and $\varepsilon_i(t)$ are assumed to be independent when $s \neq t$, and $\mathbf{\Gamma}(s,t) = \mathbf{\Gamma}(s)I(s = t)$, where $\mathbf{\Gamma}(s)$ is an $m \times m$ matrix of functions of s and $I(\cdot)$ is an indicator function. Let $\mathbf{Y}_i(s) = (Y_{i1}(s), Y_{i2}(s), \ldots, Y_{im}(s))^{\mathrm{T}}$, $i = 1, \ldots, n$. The covariance structure of $\mathbf{Y}_i(s)$ is characterized as $\mathbf{\Sigma}(s,t) = \mathbf{G}(s,t) + \mathbf{\Gamma}(s)I(s = t)$.

We consider a typical functional data setting where $\mathbf{Y}_i(s)$ is measured at the same grid of equidistant location points, i.e., $s = l/N$, $l = 1, 2, \ldots, N$, for each subject. The above leads to the following model for dependent functional measurements

$$Y_{ijl} = \mu(l/N) + \sum_{k=1}^{\infty} \xi_{ijk} \phi_{jk}(l/N) + \varepsilon_{ij}(l/N), \ 1 \leq i \leq n, 1 \leq j \leq m, 1 \leq l \leq N.$$

(2.1)

The sequences $\{\lambda_{jk}\}_{j,k=1}^{m,\infty}$, $\{\phi_{jk}(\cdot)\}_{j,k=1}^{m,\infty}$ and the random coefficients ξ_{ijk}'s only exist mathematically and they are unknown and unobservable respectively. The problem addressed here is estimation and inference of ν-th order derivative function $\mu^{(\nu)}(s)$. A similar model has been considered in [2].

2.2. Estimation procedure

We propose an estimation procedure that approximates the derivative of mean functions by polynomial splines. Let $\omega_0 = 0 < \omega_1 < \omega_2 \cdots < \omega_{N_\mu} < 1 =$

$\omega_{N_\mu+1}$ be equally-spaced points over $[0,1]$, called interior knots, in which $\omega_J = Jh_\mu$, $0 \le J \le N_\mu$, and $h_\mu = 1/(N_\mu + 1)$ is the distance between neighboring knots. Denote $\mathcal{H}^{(p-2)}$ the p-th order spline space, i.e., $p-2$ times continuously differentiable functions on $[0,1]$ that are polynomials of degree $p-1$ on $[t_J, t_{J+1}]$, $J = 0, \ldots, N_\mu$.

Following [2], we propose to approximate the mean function $\mu(\cdot)$ by a linear combination of spline basis:

$$\hat{\mu}(s) = \sum_{J=1-p}^{N_\mu} \hat{\beta}_J B_J(s), \tag{2.2}$$

where B_J be the J-th B-spline basis of order p defined in [7] and coefficients $\{\hat{\beta}_{1-p}, \ldots, \hat{\beta}_{N_\mu}\}^{\mathrm{T}}$ are the solutions of the following least squares problem

$$\left\{\hat{\beta}_{1-p}, \ldots, \hat{\beta}_{N_\mu}\right\}^{\mathrm{T}}$$

$$= \operatorname{argmin}_{\{\beta_{1-p}, \ldots, \beta_{N_\mu}\} \in R^{N_\mu+p}} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{l=1}^{N} \left\{Y_{ijl} - \sum_{J=1-p}^{N_\mu} \beta_J B_J\left(l/N\right)\right\}^2. \tag{2.3}$$

Let $\mathbf{Y}(s) = (\bar{Y}_{\cdot 1}(s), \bar{Y}_{\cdot 2}(s), \ldots, \bar{Y}_{\cdot m}(s))^{\mathrm{T}}$. Applying elementary algebra, one obtains

$$\hat{\mu}(s) = \mathbf{B}_p(s) \left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1} \mathbf{B}^{\mathrm{T}}\mathbf{Y} \tag{2.4}$$

in which $\mathbf{B}_p(s) = (B_{1-p,p}(s), \ldots, B_{N_\mu,p}(s))$ and $\mathbf{B} = (\mathbf{B}_p^{\mathrm{T}}(1/N), \ldots, \mathbf{B}_p^{\mathrm{T}}(N/N))^{\mathrm{T}}$ is the design matrix.

Given the estimation of mean function $\hat{\mu}(s)$, we consider $\hat{\mu}^{(\nu)}(s)$ as the estimator of $\mu^{(\nu)}(s)$, for any $\nu = 1, \ldots, p-2$, i.e.

$$\hat{\mu}^{(\nu)}(s) = \mathbf{B}_p^{(\nu)}(s) \left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1} \mathbf{B}^{\mathrm{T}}\mathbf{Y}, \tag{2.5}$$

where $\mathbf{B}_p^{(\nu)}(s) = (B_{1-p,p}^{(\nu)}(s), \ldots, B_{N_\mu,p}^{(\nu)}(s))$. According to B-spline property in de Boor (2001), for $p > 2$ and $1 - p \le J \le N_\mu$,

$$\frac{d}{ds}B_{J,p}(s) = (p-1)\left(\frac{B_{J,p-1}(s)}{\omega_{J+p-1} - \omega_J} - \frac{B_{J+1,p-1}(s)}{\omega_{J+p} - \omega_{J+1}}\right).$$

Therefore, $\mathbf{B}_p^{(\nu)}(s) = \mathbf{B}_{p-\nu}(s)\mathbf{D}_{(\nu)}^{\mathrm{T}}$, in which $\mathbf{D}_{(\nu)} = \mathbf{D}_1 \cdots \mathbf{D}_{\nu-1}\mathbf{D}_\nu$, and matrix $\mathbf{D}_l$, $1 \le l \le \nu \le p-2$, is defined the same as equation (6) in [19].

3. Confidence bands

3.1. Asymptotic confidence bands

First we study the asymptotic properties of the spline estimator given in (2.2). Pretending $X_{ij}^{(\nu)}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ is known, the "ideal" or "infea-

sible" [3, 2] estimator of $\mu^{(\nu)}$ is

$$\bar{\mu}^{(\nu)}(s) = \bar{X}^{(\nu)}(s) = (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij}^{(\nu)}(s), \quad s \in [0, 1]. \tag{3.1}$$

Theorem 1 below shows that the spline estimator $\hat{\mu}^{(\nu)}$ in (2.2) is asymptotically equivalent to the "ideal" estimator $\bar{\mu}^{(\nu)}$ in (3.1) with $\sqrt{n}$-rate, which is the same rate of convergence in the parametric setting. Thus, we have the oracle efficiency of the nonparametric estimator $\hat{\mu}^{(\nu)}$.

We need more notations for Theorem 1. Define the derivative of covariance function $G_{jj'}(s, t)$ as $G_{jj'}^{(\nu)}(s, t) = V_{jj', k} \phi_{jk}^{(\nu)}(s) \phi_{j'k}^{(\nu)}(t)$, where $V_{jj', k} = \mathrm{Cov}(\xi_{1jk}, \xi_{1j'k})$, $j, j' = 1, \ldots, m$. Denote the average of all entries in derivative of covariance function $\mathbf{G}^{(\nu)}(s, t)$ by $\Omega(s, t) = m^{-2} \sum_{j, j'=1}^{m} G_{jj'}^{(\nu)}(s, t)$. Let $\zeta(s)$, $s \in [0, 1]$, be a standardized Gaussian process satisfying $E\zeta(s) = 0$, $E\zeta^2(s) = 1$ and $E\zeta(s)\zeta(t) = \Omega^{-1/2}(s, s)\Omega(s, t)\Omega^{-1/2}(t, t)$, $s, t \in [0, 1]$. For any $\alpha \in (0, 1)$, we denote $Q_{1-\alpha}$ the $100(1-\alpha)$-th percentile of the absolute maxima distribution of $\zeta(s)$, i.e., $P[\sup_{s \in [0,1]} |\zeta(s)| \le Q_{1-\alpha}] = 1 - \alpha$ and $z_{1-\alpha/2}$ the $100(1-\alpha/2)$-th percentile of the standard normal distribution.

THEOREM 1. *Suppose Assumptions (A1)–(A5) in Appendix hold. For any $\alpha \in (0, 1)$, as $n \to \infty$, the "infeasible estimator" $\bar{\mu}^{(\nu)}$ converges to $\mu^{(\nu)}$ uniformly at the $\sqrt{n}$ rate*

$$P \left\{ \sup_{s \in [0,1]} \sqrt{n} \left| \bar{\mu}^{(\nu)}(s) - \mu^{(\nu)}(s) \right| \Omega(s, s)^{-1/2} \le Q_{1-\alpha} \right\} \to 1 - \alpha,$$

$$P \left\{ \sqrt{n} \left| \bar{\mu}^{(\nu)}(s) - \mu^{(\nu)}(s) \right| \Omega(s, s)^{-1/2} \le z_{1-\alpha/2} \right\} \to 1 - \alpha, \quad \forall s \in [0, 1],$$

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

$$\sup_{s \in [0,1]} \sqrt{n} \left| \bar{\mu}^{(\nu)}(s) - \hat{\mu}^{(\nu)}(s) \right| = o_P(1).$$

The oracle efficiency in Theorem 1 immediately indicates the following result, which can be used to construct SCB or pointwise confidence intervals for $\mu^{(\nu)}(s)$, $s \in [0, 1]$.

COROLLARY 1. *Under Assumptions (A1)–(A5) in Appendix, for any $\alpha \in (0, 1)$, as $n \to \infty$, an asymptotic $100(1 - \alpha)\%$ correct confidence band for $\mu^{(\nu)}(s)$, $s \in [0, 1]$, is*

$$\hat{\mu}^{(\nu)}(s) \pm n^{-1/2} \Omega(s, s)^{1/2} Q_{1-\alpha}, \tag{3.2}$$

while an asymptotic $100(1 - \alpha)\%$ pointwise confidence interval for $\mu^{(\nu)}(s)$, $s \in [0, 1]$, is

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

3.2. Confidence bands for the difference between two derivative functions

In this section, we focus our review on the statistical inference for the derivative functions in the two-sample problems drawn from dependent function data sets. The aforementioned confidence bands for one sample derivative function can be extended to the two-sample case. Analogous to the previous notations, we denote two samples indicated by $d = 1, 2$, which satisfy

$$Y_{ijl,d} = \mu_d\left(l/N\right) + \sum_{k=1}^{\infty} \xi_{ijk,d}\phi_{jk,d}\left(l/N\right) + \varepsilon_{ij,d}\left(l/N\right),$$

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

For $d = 1, 2$, let $\hat{\mu}_d^{(\nu)}$ be the spline estimate of derivative function $\mu_d^{(\nu)}$ as given in (2.2), and define $\Omega_d(s,t) = m_d^{-2} \sum_{j,j'=1}^{m_d} G_{jj',d}^{(\nu)}(s,t)$. 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

$$E\zeta_{\text{diff}}(s)\zeta_{\text{diff}}(t) = \frac{\Omega_1\left(s,t\right) + \tau\Omega_2\left(s,t\right)}{\left\{\Omega_1(s,s) + \tau\Omega_2(s,s)\right\}^{1/2}\left\{\Omega_1\left(t,t\right) + \tau\Omega_2\left(t,t\right)\right\}^{1/2}},$$

for $s, t \in [0, 1]$. Next denote $Q_{\text{diff},1-\alpha}$ the $(1 - \alpha)$-th quantile of the absolute maxima deviation of $\zeta_{\text{diff}}(s)$, $s \in [0, 1]$ 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* Appendix *are modified for each group accordingly, then for any* $\alpha \in (0, 1)$, *as* $n_1 \to \infty$, $\hat{\tau} \to \tau > 0$,

$$P\left\{\sup_{s\in[0,1]} \frac{n_1^{1/2}\left|\left(\hat{\mu}_1^{(\nu)} - \hat{\mu}_2^{(\nu)} - \mu_1^{(\nu)} + \mu_2^{(\nu)}\right)(s)\right|}{\left\{\left(\Omega_1 + \tau\Omega_2\right)(s,s)\right\}^{1/2}} \le Q_{\text{diff},1-\alpha}\right\} \to 1 - \alpha.$$

Theorems 2 yields the following asymptotic SCB for $\mu_1^{(\nu)}(s) - \mu_2^{(\nu)}(s)$, $s \in [0, 1]$.

COROLLARY **2.** *If Assumptions (A1)–(A5) in* Appendix *are modified for each group accordingly and* $\lim_{n_1\to\infty} \hat{\tau} = \tau > 0$, *then for any* $\alpha \in (0, 1)$, *an* $100(1 - \alpha)\%$ *asymptotically correct confidence band for* $\mu_1^{(\nu)} - \mu_2^{(\nu)}$ *is* $(\hat{\mu}_1^{(\nu)} - \hat{\mu}_2^{(\nu)})(s) \pm n_1^{-1/2}Q_{\text{diff},1-\alpha}\{(\Omega_1 + \tau\Omega_2)(s,s)\}^{1/2}$, $s \in [0, 1]$.

4. Implementation

To apply the proposed confidence bands, there are a number of important function and parameter estimation issues that need to be addressed, such as the unknown function $\Omega(\cdot, \cdot)$ and the quantiles $Q_{1-\alpha}$ and $Q_{\text{diff},1-\alpha}$, as well as choosing the number of principal components.

In practice, very often the repeated sizes are not equal, and we suggest the following specific data transformation. For $1 \le j \le \max_{1\le i\le n} m_i = m$, we

define $Y_{ij}^{\dagger}(\cdot) = Y_{ij}(\cdot)$, $\xi_{ijk}^{\dagger} = \xi_{ijk}$, $\varepsilon_{ij}^{\dagger}(\cdot) = \varepsilon_{ij}(\cdot)$ and $\phi_{jk}^{\dagger}(\cdot) = \phi_{jk}(\cdot)$, if $Y_{ij}(\cdot)$ is observed and $Y_{ij}^{\dagger}(\cdot) = 0$, $\xi_{ijk}^{\dagger} = 0$, $\varepsilon_{ij}^{\dagger}(\cdot) = 0$ and $\phi_{jk}^{\dagger}(\cdot) = 0$, if $Y_{ij}(\cdot)$ is missing. Further let $r_{ij} = I(Y_{ij}(\cdot)$ is observed), thus, $\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} = \sum_{i=1}^{n} m_i = n_m$. In the following, for simple notations we do not distinguish $Y_{ij}(\cdot)$, ξ_{ijk}, $\varepsilon_{ij}(\cdot)$ and $\phi_{jk}(\cdot)$ from $Y_{ij}^{\dagger}(\cdot)$, $\xi_{ijk}^{\dagger}$, $\varepsilon_{ij}^{\dagger}(\cdot)$ and $\phi_{jk}^{\dagger}(\cdot)$, respectively.

4.1. Estimating $\Omega(s,s)$

We first apply tensor product spline approach addressed in [3] to estimate the covariance function $G_{jj}(s,t)$. The pilot spline estimator of $G_{jj}(s,t)$ is

$$\tilde{G}_{jj}(s,t) = \sum_{J,J'=1-p}^{N_G} \tilde{b}_{JJ'} B_J(s) B_{J'}(s), \qquad (4.1)$$

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

$$\left\{\tilde{b}_{J,J'}\right\}_{J,J'=1-p}^{N_G}$$

$$= \operatorname{argmin}_{R^{N_G+p} \otimes R^{N_G+p}} \sum_{l \neq l'}^{N} \left\{ \bar{C}_{\cdot ll'} - \sum_{1-p \leq J,J' \leq N_G} b_{JJ'} B_J(l/N) B_{J'}(l'/N) \right\}^2,$$

and $\bar{C}_{\cdot ll'} = n_m^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} \{Y_{ijl} - \hat{\mu}(l/N)\}\{Y_{ijl'} - \hat{\mu}(l'/N)\}$, $1 \leq l \neq l' \leq N$.

The estimation procedure of eigenvalues $\hat{\lambda}_{jk}$, eigenfunctions $\hat{\psi}_{jk}$ and FPC $\hat{\xi}_{ijk}$ follows standard method discussed in [17]. We focus on estimation of the ν-th derivative of eigenfunctions $\hat{\psi}_{jk}^{(\nu)}$. According to [3], one has the eigenequation

$$\frac{d^\nu}{dt^\nu} \int_0^1 \tilde{G}_{jj}(s,t)\hat{\psi}_{jk}(s)\,ds = \int_0^1 \frac{\partial^\nu}{\partial t^\nu} \tilde{G}_{jj}(s,t)\hat{\psi}_{jk}(s)\,ds = \hat{\lambda}_{jk}\hat{\psi}_k^{(\nu)}(t), \quad (4.2)$$

where $\hat{\psi}_{jk}$ are subject to $\int_0^1 \hat{\psi}_{jk}^2(s)ds = 1$ and $\int_0^1 \hat{\psi}_{jk}(s)\hat{\psi}_{jk'}(s)ds = 0$ for $k' < k$. If N is sufficiently large, the left hand side of (4.2) can be approximated by $\frac{1}{N}\sum_{l=1}^{N} \tilde{G}_{jj}^{(\nu)}(\frac{l}{N}, \frac{l'}{N})\hat{\psi}_{jk}(\frac{l}{N})$. Then we estimate $G_{jj}^{(\nu)}(s,t)$ by $\tilde{G}_{jj}^{(\nu)}(s,t) = \sum_{k=1}^{\kappa} \hat{\lambda}_{jk}\hat{\psi}_{jk}^{(\nu)}(s)\hat{\psi}_{jk}^{(\nu)}(s)$.

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

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

Thus, $\Omega(s,s)$ can be estimated by

$$\tilde{\Omega}(s,s) = m^{-2} \sum_{j,j'=1}^{m} \tilde{G}_{jj'}^{(\nu)}(s,s). \qquad (4.3)$$

The following theorem shows that $\tilde{\Omega}(\cdot, \cdot)$ and $\Omega(\cdot, \cdot)$ are asymptomatically equivalent.

THEOREM **3.** *Under Assumptions (A1)–(A5), one has*

$$\sup_{(s,t)\in[0,1]^2} \left|\tilde{\Omega}(s,t) - \Omega(s,t)\right| = o_P(1).$$

The proof of Theorem 3 is given in Appendix A.3.

According to our experience, the aforementioned approach is less efficient than imposing a structure on the within-curve correlation. For example, exchangeable (EC), first-order autoregressive (AR-1) and toeplitz (TOEP) are some common options of the structure. Similar methods are introduced in the repeated measurement studies [13, 2]. In the following we assume the covariance structure between ξ_{1jk} and $\xi_{1j'k}$,

$$\mathrm{Cov}\left(\xi_{1jk}, \xi_{1j'k}\right) = \rho_{jj'} \lambda_{jk}^{1/2} \lambda_{j'k}^{1/2} I\left(j \neq j'\right) + \lambda_{jk}^{1/2} \lambda_{j'k}^{1/2} I\left(j = j'\right),$$

where $0 \leq \rho_{jj'} < 1$ for any $j \neq j'$. Let $\bar{Y}_{..l} = n_m^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} Y_{ijl}$, $1 \leq l \leq N$. Define the empirical version of with-subject covariance as

$$\hat{\Sigma}_{jj'} = \frac{1}{N^2 \sum_{i=1}^{n} r_{ij} r_{ij'}} \sum_{i=1}^{n} \sum_{l,l'=1}^{N} r_{ij} r_{ij'} (Y_{ijl} - \bar{Y}_{..l})(Y_{ij'l'} - \bar{Y}_{..l'}).$$

Without loss of generality and for notation simplicity, let us assume $\lambda_{jk} = \lambda_k$, $\psi_{jk} = \psi_k$, and estimate $\rho_{jj'}$ under three popular correlation structures: EC, AR-1 and TOEP. For any $j \neq j'$, the estimator of $\rho_{jj'}$ can be defined as

$$\hat{\rho}_{jj'}$$
$$= \begin{cases} \hat{\rho}^{|j-j'|} = \{(\mathbf{R}^{\mathsf{T}}\mathbf{R})^{-1}\mathbf{R}^{\mathsf{T}}\mathbf{D}\}^{|j-j'|}, & \text{AR-1} \\ \hat{\rho} = (m-1)^{-1} \sum_{j \neq j'}^{m} \hat{\Sigma}_{jj'} \left(\sum_{j=1}^{m} \hat{\Sigma}_{jj}\right)^{-1}, & \text{EC} \\ \hat{\rho}_{|j-j'|} = (m-|j-j'|)^{-1} m\left(\sum_{r=1}^{m-|j-j'|} \hat{\Sigma}_{r,r+|j-j'|}\right)\left(\sum_{r=1}^{m} \hat{\Sigma}_{rr}\right)^{-1}, & \text{TOEP} \end{cases}.$$

where $\mathbf{R}^{\mathsf{T}} = (\hat{\Sigma}_{j1}, \hat{\Sigma}_{j2}, \ldots, \hat{\Sigma}_{j,m-1}, j = 1, \ldots, m-1)$ and $\mathbf{D}^{\mathsf{T}} = (\hat{\Sigma}_{j2}, \hat{\Sigma}_{j3}, \ldots, \hat{\Sigma}_{j,m}, j = 1, \ldots, m-1)$. Therefore, we can estimate $G_{jj'}^{(\nu)}(s,t)$ by

$$\hat{G}_{jj'}^{(\nu)}(s,t) = \begin{cases} \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\psi}_k^{(\nu)}(s) \hat{\psi}_k^{(\nu)}(t), & \text{if } j = j' \\ \hat{\rho}_{jj'} \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\psi}_k^{(\nu)}(s) \hat{\psi}_k^{(\nu)}(t), & \text{if } j \neq j' \end{cases},$$

and define $\hat{\Omega}(s,s) = m^{-2} \sum_{j,j'=1}^{m} \hat{G}_{jj'}^{(\nu)}(s,s)$.

REMARK **1.** Note that under the assumption of the aforementioned with-subject correlation structures and simple notations, $\hat{\Omega}(s,s)$ is equivalent to $\tilde{\Omega}(s,s)$ defined in (4.3). Therefore, according to Theorem 3, under Assumptions (A1)–(A5), $\sup_{(s,t)\in[0,1]^2} |\hat{\Omega}(s,t) - \Omega(s,t)| = o_P(1)$.

4.2. Estimating the quantiles

We generate independent $\mathcal{R}^m$-valued Gaussian vectors, $\mathbf{Z}_{k,b} = (Z_{1k,b}, \ldots, Z_{mk,b})^{\mathrm{T}}$ satisfying that $\mathrm{Cov}(Z_{jk,b}, Z_{j'k',b}) = V_{jj',k}\delta_{kk'}$, where $\delta_{kk'} = I(k = k')$ and $b = 1, \ldots, b_M$. Here b_M is a preset large integer, the default of which is 1000. Let

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

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

In two-sample situation, for $d = 1, 2$, one generates independent $\mathcal{R}^{m_d}$-valued Gaussian vectors, $\mathbf{Z}_{k,b,d} = (Z_{1k,b,d}, \ldots, Z_{m_dk,b,d})^{\mathrm{T}}$ satisfying that $\mathrm{Cov}(Z_{jk,b,d}, Z_{j'k',b,d}) = V_{jj',k,d}\delta_{kk'}$. Similarly to $\hat{\zeta}_b(\cdot)$, let

$$\hat{\zeta}_{\mathrm{diff},b}(s)$$
$$= \left\{ (\hat{\Omega}_1 + \hat{\Omega}_2)(s,s) \right\}^{-1/2}$$
$$\times \left\{ \frac{1}{m_1} \sum_{j=1}^{m_1} \sum_{k=1}^{\infty} Z_{jk,b,1} \hat{\psi}_{k,1}^{(\nu)}(s) + \frac{n_1}{n_2 m_2} \sum_{j=1}^{m_2} \sum_{k=1}^{\infty} Z_{jk,b,2} \hat{\psi}_{k,2}^{(\nu)}(s) \right\}.$$

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

4.3. Selecting correlation/covariance structure

In the practice, an appropriate correlation matrix based on available data is needed. To choose correlation structure, it is natural to consider bootstrap method using the empirical coverage rates of inducing confidence bands as the choosing criterion. Following the suggestions in [2], we select the structure which provides the closest coverage rate to the nominal level. First, one computes the spline estimator $\hat{\mu}^{(\nu)}$ for observed data. Second, given a correlation structure, one constructs the corresponding confidence band based on the bootstrap sample data for 500 times and compute the coverage rate of the confidence band under each correlation structure from a given set of candidates. Here, $\hat{\mu}^{(\nu)}$ is assumed as the true underlying function when checking the empirical coverage rate. Finally, one chooses the covariance structure whose empirical coverage rate is the closest one to the nominal level.

4.4. Selecting spline knots and the number of eigenfunctions

The numbers of knots serve as the smoothing parameters, playing the same role as bandwidths in the local linear method. According to Assumption (A3)

in Appendix, the number of interior knots in estimating $\mu^{(\nu)}(s)$ is taken to be $N_\mu = [n^{1/2(p-\nu)} \log(n)]$, in which $[a]$ denotes the integer part of a. Meanwhile, the spline estimator $\tilde{G}_{jj}(s,t)$ in (4.1) is calculated with the number of interior knots $N_G = [2n^{1/(2p)} \log(\log(n))]$. Similar knots selections are suggested in [3]. These choices of knots also satisfy Assumption (A3) in Appendix.

We choose the number of eigenfunctions using the following standard and efficient criterion: $\kappa = \mathrm{argmin}_{1 \leq l \leq T} \{ \sum_{k=1}^{l} \hat{\lambda}_k / \sum_{k=1}^{T} \hat{\lambda}_k > 0.95 \}$, where $\{\lambda_k, k = 1, \ldots, T\}$ are the first T estimated positive eigenvalues.

5. Simulation results

To access the practical behavior of the proposed methodology, we conduct two simulation examples to explore the performance of the proposed asymptotic SCB.

5.1. Simulation 1: Empirical coverage rates

We take the observations to be generated from the following model

$$Y_{ijl} = \mu(l/N) + \sum_{k=1}^{2} \xi_{ijk} \phi_k(l/N) + \varepsilon_{ijl}, \ \ 1 \leq i \leq n, 1 \leq j \leq m, 1 \leq l \leq N, \ (5.1)$$

where $\mu(s) = 5s + 4\sin\{2\pi(s - 1/2)\}$, $\phi_1(s) = \cos(\pi s)$ and $\phi_2(s) = \sin(\pi s)$. We generate the FPC scores $\xi_{ik} = (\xi_{i1k}, \ldots, \xi_{imk})^{\mathrm{T}} \sim N(\mathbf{0}, \lambda_k \mathbf{\Omega})$, $\lambda_1 = 2/3$, $\lambda_2 = 1/2$ and $\varepsilon_{il} = (\varepsilon_{i1l}, \ldots, \varepsilon_{iml})^{\mathrm{T}} \sim N(\mathbf{0}, 0.1^2 \mathbf{\Omega})$. We consider the following settings for the correlation matrix $\mathbf{\Omega} = \{\Omega_{jj'}\}_{j,j'=1}^{m}$.

Case 1. IND: $\Omega_{jj'} = I(j = j')$;
Case 2. EC: $\Omega_{jj'} = I(j = j') + \rho I(j \neq j')$, where $\rho = 0.05, 0.1$;
Case 3. AR-1: $\Omega_{jj'} = I(j = j') + \rho^{|j-j'|} I(j \neq j')$, where $\rho = 0.2, 0.4$;
Case 4. TOEP: $\Omega_{jj'} = I(j = j') + \rho_{|j-j'|} I(j \neq j')$, where $\rho_{|j-j'|} = \rho_r$ with $\rho_r = 0.2^r + 0.05$ and $0.3^r + 0.1$, $r = 1, \ldots, m-1$.

We use the proposed method in (3.2) and its "oracle" version $\tilde{\Omega}(s,s)$ in (4.3) to construct the confidence bands for $\mu^{(1)}(\cdot)$ under different correlation structures: "IND", "EC", "AR-1" and "TOEP". We consider two confidence levels: $1 - \alpha = 0.95, 0.99$. The number of trajectories n is taken to be 30, 50, 100, and for each n, the number of observations on the trajectory is $N = n$. Each simulation is repeated 1000 times.

For each setting, we adopt cubic spline ($p = 4$) smoothing and calculate the percentage of coverage of the true function on regular grid with increments 0.01 by the bands constructed using four different correlation structures. Tables 1 and 2 summarize the percentage that the true curve $\mu^{(1)}(\cdot)$ is covered by the cubic spline confidence bands respectively.

In Table 1, the true underlying within-curve correlation structure is chosen as independent ("IND"), i.e. $\mathbf{\Omega}$ is identity matrix. Hence, one would expect the bands constructed using the "IND" structure to be the best performer. In

G. Cao

TABLE 1
Empirical coverage rates of the bands using various correlation structures (Case 1)

m	n	α	Est.				Oracle			
			IND	EC	AR-1	TOEP	IND	EC	AR-1	TOEP
10	30	0.05	0.808	0.786	0.807	0.788	0.849	–	–	–
		0.01	0.943	0.933	0.941	0.931	0.964	–	–	–
	50	0.05	0.906	0.897	0.902	0.896	0.917	–	–	–
		0.01	0.983	0.978	0.983	0.974	0.987	–	–	–
	100	0.05	0.946	0.939	0.949	0.948	0.948	–	–	–
		0.01	0.993	0.991	0.993	0.991	0.996	–	–	–
20	30	0.05	0.743	0.718	0.739	0.716	0.766	–	–	–
		0.01	0.925	0.900	0.929	0.902	0.945	–	–	–
	50	0.05	0.894	0.874	0.888	0.870	0.899	–	–	–
		0.01	0.984	0.981	0.984	0.975	0.981	–	–	–
	100	0.05	0.939	0.926	0.939	0.932	0.950	–	–	–
		0.01	0.990	0.991	0.992	0.991	0.994	–	–	–

fact, Table 1 shows that the bands constructed using correlated structures are very competitive, too. In several cases, the coverage rates of the bands with correlated structures ("EC", "AR-1" and "TOEP") are very close to or even better than the bands with "IND" structure. As we expected coverage rate of "oracle" bands approach the nominal levels is faster than the estimated one, but both of them converge to the nominal level clearly, which show positive confirmation of Theorems 1 and 3. To summarize, in many correlated structure settings, our method can lead to substantially better coverage rates than the "naive" ("IND")-type band, while the loss in the case of an underlying "IND" structure is quiet small when sample sizes increase.

In Tables 2 and 3, we investigate the coverage percentages when the true correlation structures are "EC", "AR-1" or "TOEP". For each true correlation structure, we construct spline confidence bands based on all four correlation assumptions including "IND". As expected, the coverage percentages for the confidence bands are close to the nominal levels when applying the correct correlation structure. The bands obtained by "IND" have significantly smaller coverage rates than the nominal levels regardless of the sample sizes and correlation structures. Particularly, when the correlation structure becomes stronger, for example, increasing $\rho_{jj'}$ in $\Omega_{jj'}$, the "IND"-type of band performs much worse than its counterparts. While the advantage of applying our proposed bands over the "naive" band seems to persist across the table, it is more evident in the situation with stronger correlation structure, i.e., increasing $\rho_{jj'}$ in $\Omega_{jj'}$. From Tables 2 and 3, we can see that in most of the scenarios the "oracle" confidence bands outperform the estimated bands, and the "oracle" bands approach the nominal coverage with sample sizes increasing. In Tables 1 to 3, when nominal level is 0.05, the standard error for each empirical coverage rate ranges from 0.0001 to 0.00025 and the average standard error is around 0.0002. When nominal level is 0.01, the standard errors of empirical coverage rates are less than 0.0001.

We depict a typical comparison between two 95% confidence bands constructed using "IND" and "TOEP" correlation structures in Figure 1. The true correlation structure is "TOEP" (with $\rho = 0.3^r + 0.1$). This plot is based on

TABLE 2
Empirical coverage rates of the bands using various correlation structures (Cases 2, 3 and 4) and each trajectory is repeatedly observed $m_1 = m_2 = 10$ times

True Structure (true ρ)	n	α	Est.				Oracle			
			IND	EC	AR-1	TOEP	IND	EC	AR-1	TOEP
EC(0.05)	30	0.05	0.693	0.812	0.736	0.811	–	0.866	–	–
		0.01	0.887	0.938	0.899	0.942	–	0.976	–	–
	50	0.05	0.805	0.887	0.835	0.881	–	0.911	–	–
		0.01	0.934	0.981	0.948	0.978	–	0.987	–	–
	100	0.05	0.857	0.944	0.882	0.944	–	0.947	–	–
		0.01	0.970	0.994	0.976	0.992	–	0.994	–	–
EC(0.10)	30	0.05	0.611	0.826	0.699	0.821	–	0.881	–	–
		0.01	0.813	0.951	0.869	0.947	–	0.979	–	–
	50	0.05	0.715	0.922	0.797	0.922	–	0.930	–	–
		0.01	0.904	0.988	0.941	0.986	–	0.995	–	–
	100	0.05	0.762	0.954	0.841	0.956	–	0.962	–	–
		0.01	0.934	0.992	0.973	0.993	–	0.994	–	–
AR-1(0.2)	30	0.05	0.693	0.794	0.821	0.807	–	–	0.859	–
		0.01	0.879	0.936	0.952	0.937	–	–	0.970	–
	50	0.05	0.807	0.898	0.906	0.896	–	–	0.919	–
		0.01	0.932	0.979	0.983	0.980	–	–	0.988	–
	100	0.05	0.857	0.950	0.946	0.943	–	–	0.950	–
		0.01	0.972	0.997	0.997	0.994	–	–	0.997	–
AR-1(0.4)	30	0.05	0.554	0.840	0.840	0.839	–	–	0.881	–
		0.01	0.780	0.950	0.956	0.951	–	–	0.976	–
	50	0.05	0.673	0.914	0.921	0.918	–	–	0.938	–
		0.01	0.858	0.990	0.991	0.992	–	–	0.997	–
	100	0.05	0.725	0.946	0.954	0.954	–	–	0.955	–
		0.01	0.907	0.992	0.995	0.992	–	–	0.997	–
TOEP $(0.2^r + 0.05)$	30	0.05	0.627	0.850	0.813	0.851	–	–	–	0.901
		0.01	0.841	0.950	0.937	0.953	–	–	–	0.976
	50	0.05	0.712	0.919	0.894	0.921	–	–	–	0.935
		0.01	0.902	0.983	0.977	0.981	–	–	–	0.992
	100	0.05	0.763	0.944	0.912	0.946	–	–	–	0.955
		0.01	0.930	0.994	0.983	0.992	–	–	–	0.994
TOEP $(0.3^r + 0.1)$	30	0.05	0.496	0.863	0.810	0.867	–	–	–	0.908
		0.01	0.703	0.962	0.946	0.961	–	–	–	0.982
	50	0.05	0.572	0.882	0.864	0.882	–	–	–	0.945
		0.01	0.798	0.970	0.952	0.968	–	–	–	0.993
	100	0.05	0.596	0.930	0.884	0.926	–	–	–	0.947
		0.01	0.818	0.986	0.984	0.986	–	–	–	0.998

$n = 30$, $m = 10, 20$. In Figure 1, one sees that the "IND" and "TOEP" produce the same estimator for the true curve $\mu^{(1)}(\cdot)$, however, the confidence bands obtained are very different. From Figure 1, one sees that the true curve $\mu^{(1)}(\cdot)$ lies completely inside the "TOEP"-type of band, but it can not be covered entirely by the "IND"-type of band.

5.2. Simulation 2: Empirical sizes and power

We conduct a numerical study to evaluate the empirical size and power of hypothesis tests for the derivative functions based on the proposed confidence bands. We compare the results of the bands developed under the "IND" as-

TABLE 3

Empirical coverage rates of the bands using various correlation structures (Cases 2, 3 and 4) and each trajectory is repeatedly observed $m_1 = m_2 = 20$ times

True Structure (true ρ)	n	α	Est.				Oracle			
			IND	EC	AR-1	TOEP	IND	EC	AR-1	TOEP
EC(0.05)	30	0.05	0.493	0.782	0.547	0.788	–	0.841	–	–
		0.01	0.757	0.923	0.806	0.928	–	0.961	–	–
	50	0.05	0.675	0.899	0.728	0.905	–	0.915	–	–
		0.01	0.880	0.982	0.906	0.983	–	0.992	–	–
	100	0.05	0.709	0.930	0.775	0.934	–	0.942	–	–
		0.01	0.911	0.992	0.935	0.992	–	0.995	–	–
EC(0.10)	30	0.05	0.386	0.849	0.518	0.852	–	0.900	–	–
		0.01	0.658	0.975	0.789	0.975	–	0.984	–	–
	50	0.05	0.495	0.913	0.640	0.909	–	0.931	–	–
		0.01	0.751	0.982	0.856	0.981	–	0.987	–	–
	100	0.05	0.555	0.951	0.723	0.955	–	0.961	–	–
		0.01	0.810	0.998	0.907	0.999	–	0.998	–	–
AR-1(0.2)	30	0.05	0.592	0.752	0.762	0.750	–	–	0.804	–
		0.01	0.834	0.909	0.929	0.903	–	–	0.952	–
	50	0.05	0.798	0.890	0.905	0.893	–	–	0.912	–
		0.01	0.938	0.980	0.985	0.982	–	–	0.989	–
	100	0.05	0.815	0.932	0.938	0.934	–	–	0.943	–
		0.01	0.970	0.993	0.995	0.991	–	–	0.994	–
AR-1(0.4)	30	0.05	0.466	0.830	0.859	0.835	–	–	0.881	–
		0.01	0.749	0.960	0.971	0.959	–	–	0.980	–
	50	0.05	0.625	0.913	0.918	0.910	–	–	0.930	–
		0.01	0.848	0.980	0.985	0.984	–	–	0.992	–
	100	0.05	0.685	0.943	0.947	0.944	–	–	0.958	–
		0.01	0.896	0.997	0.997	0.997	–	–	0.999	–
TOEP $(0.2^r + 0.05)$	30	0.05	0.481	0.836	0.726	0.842	–	–	–	0.893
		0.01	0.729	0.950	0.900	0.945	–	–	–	0.968
	50	0.05	0.564	0.899	0.805	0.900	–	–	–	0.915
		0.01	0.902	0.983	0.977	0.981	–	–	–	0.984
	100	0.05	0.650	0.935	0.853	0.937	–	–	–	0.945
		0.01	0.858	0.992	0.964	0.994	–	–	–	0.992
TOEP $(0.3^r + 0.1)$	30	0.05	0.338	0.877	0.736	0.875	–	–	–	0.906
		0.01	0.576	0.963	0.905	0.961	–	–	–	0.979
	50	0.05	0.398	0.922	0.817	0.928	–	–	–	0.955
		0.01	0.672	0.989	0.957	0.985	–	–	–	0.994
	100	0.05	0.476	0.953	0.851	0.955	–	–	–	0.960
		0.01	0.724	0.993	0.972	0.994	–	–	–	0.996

sumption with those taking into account the within-curve correlations. In order to mimic the two-sample testing problems, we consider the following hypothesis test:

$$H_0 : \mu_1^{(1)}(s) = \mu_2^{(1)}(s), \ \forall s \in [0,1] \longleftrightarrow H_a : \mu_1^{(1)}(s) \neq \mu_2^{(1)}(s), \ \exists s \in [0,1]. \quad (5.2)$$

We generate two groups of data from the model given in (5.1) with $\mu_1(s) = 5s + \delta s + 4\sin\{2\pi(s - 1/2)\}$ and $\mu_2(s) = 5s + 4\sin\{2\pi(s - 1/2)\}$, and all the remaining settings are the same as in Example 1 under the same parameter values. The constant δ takes 6 values $\{0, 0.25, 0.5, \dots, 1.25\}$. We choose the sample sizes $n_1 = n_2 = 100$, and the number of observation points of each curve $N = 100$. For each subject, the noisy trajectory is repeatedly observed
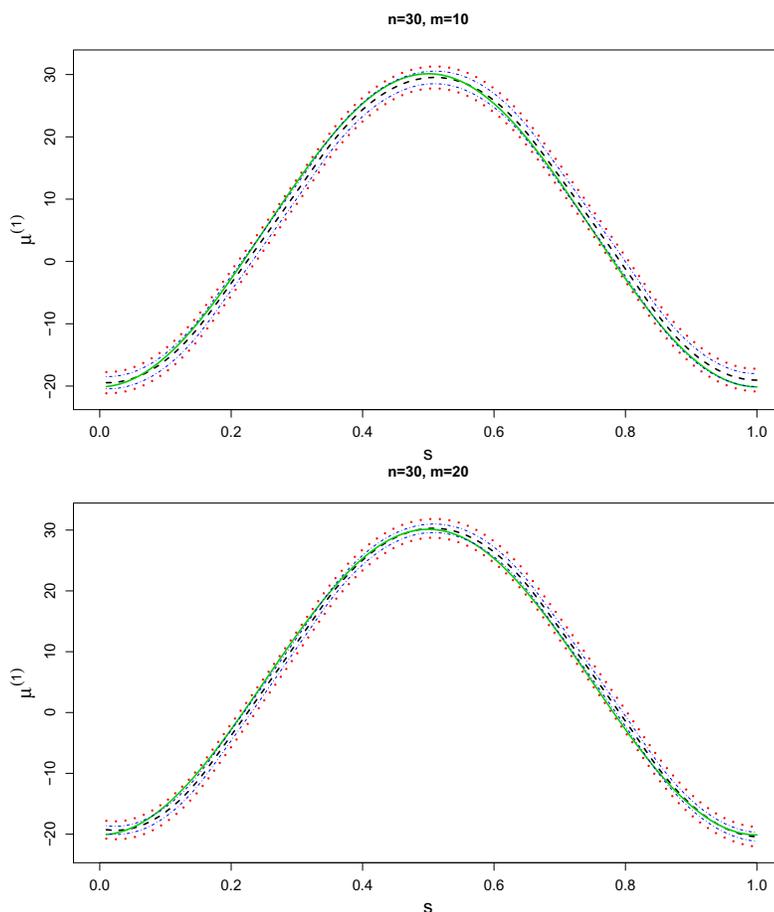
n=30, m=10

n=30, m=20

FIG 1. *The estimated derivative function (middle dashed line), the true derivative function (middle solid line) and its* 95% *confidence bands (3.2) using "TOEP" structure (upper and lower dotted lines) and using "IND" structure (upper and lower dot-dashed lines).*

$m_1 = m_2 = 10$ times under three true correlation structures: "AR-1" ($\rho = 0.2$), "EC" ($\rho = 0.05$) and "TOEP" ($\rho_r = 0.2^r + 0.05$, $r = 1, \ldots, m-1$). For each simulation, the significance level was set at $\alpha = 0.05$, and 500 replications were used to estimate the rejection rates. Indeed, large values of δ shift the derivative of mean function for the first group data further away from second one, therefore the rejection rates increase with the values of δ.

Table 4 summarizes the empirical type I errors and powers of the hypothesis test in (5.2) under different true correlation structures respectively. For each correlation structure, we compare the performance of the "IND"-type band and the band using the true structure at 5% nominal level. It can be seen from Table 4, when $\delta = 0$, the "IND" does not maintain its sizes regardless of the true correlation structures and sample sizes. This finding also correspond with its

TABLE 4

Empirical type I errors and powers for hypothesis testing (5.2) at 5% nominal levels with various correlation structures. For each case, sample sizes are $n_1 = n_2 = N = 100$ and each trajectory is repeatedly observed $m_1 = m_2 = 10$ times

True Structure (true ρ)	Applied Structure	δ					
		0.00	0.25	0.50	0.75	1.00	1.25
TOEP($\rho_r = 0.2^r + 0.05$)	TOEP	0.056	0.124	0.332	0.642	0.966	1.000
	IND	0.241	0.422	0.790	0.981	1.000	1.000
EC(0.05)	EC	0.061	0.153	0.424	0.796	0.995	1.000
	IND	0.129	0.318	0.723	0.973	1.000	1.000
AR-1(0.20)	AR-1	0.057	0.148	0.426	0.808	0.998	1.000
	IND	0.135	0.309	0.699	0.978	1.000	1.000

low coverage rates in Tables 2 and 3. Therefore, although the power of "IND" is larger than our proposed method, it is not recommended for practical applications. Overall, the tests based on our proposed method maintain their sizes at 5% nominal level, while increasing values of δ and sample sizes improve the power of detecting the alternatives.

6. Real data analysis

6.1. Confidence bands for derivative function of human mortality rates

Assessing demographic trends in mortality is of growing interest due to the demographic impacts of aging and extended longevity on the future viability of social security and retirement systems [6]. Such trends not only raise basic biodemographic questions about the nature and the limits of human lifespan but also about the demographic future of aging societies. Our dataset contains the lifetables for 41 countries and areas from the calendar year 1992 to year 2005 obtained from the Human Mortality Database [12] (www.mortality.org). Figure 2 depicts mortality rates within age 60 to age 100 in two countries (Australia and Iceland) from year 1992 to year 2005. It can be found that the mortality rate decreases slightly over the years in Australia, while there are no clear trends happening in Iceland during these years. Various approaches have been proposed for the modeling of mortality data including i.i.d. functional approach [6] and two-dimensional functional principal analysis approach [5]. There is one interesting issue: the analysis of recurring mortality patterns and their structure. In our analysis, we focus on the estimation and inference of the first order derivative of mortality rate over calendar years for elder population. We also apply the proposed confidence bands to period lifetables to recur mortality trends.

Let $Y_{ij}(s)$ denote the mortality rate observed for the ith country, jth calendar year at age s, where $i = 1, \ldots, 41$, $j = 1992, \ldots, 2005$. We focus on the older individuals within the age range $60 \leq s \leq 100$.

Applying cubic spline smoothing to the dataset, we obtain the estimated overall first order derivative function of mortality rates $\hat{\mu}^{(1)}(\cdot)$ for age from 60 to 100. To construct the confidence band, we first select the correlation structure
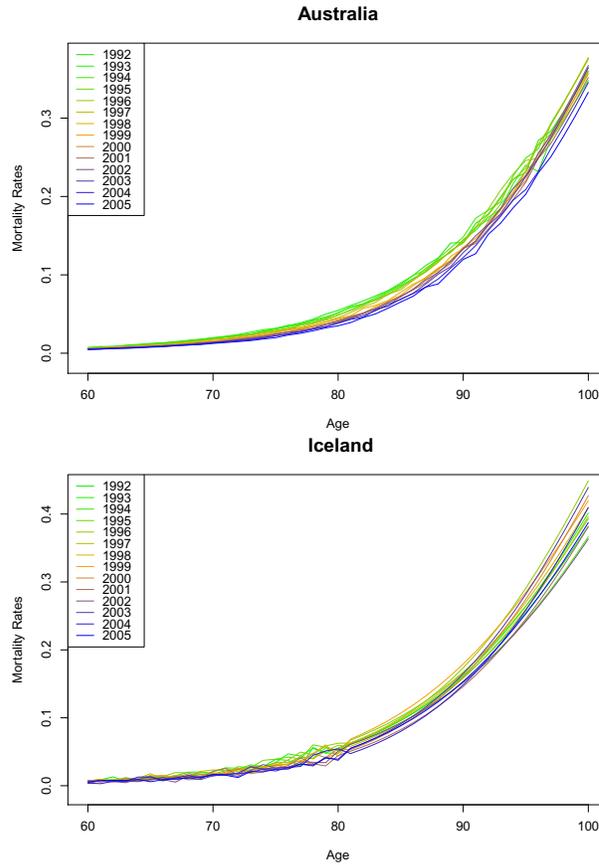
Australia

Iceland

FIG 2. *Plot of mortality rates in Australia and Iceland from year 1992 to year 2005.*

using the resampling method in Section 4, and "AR-1" gives the closest coverage rate to the nominal level based on 1000 resampling data. "IND" and "AR-1" supply the same estimated derivative function $\mu^{(1)}$; see the middle solid line in Figure 3. To explore the comparison of the bands under different structure assumptions, in Figure 3 we present both the "IND"-type of band (upper and lower dot-dashed lines) and the "EC"-type of band (upper and lower dashed lines) at 95% confidence level. As shown in Figure 3, the estimated derivative function of mortality rate is increasing quickly from 0, which corroborates with the convex trends in Figure 2.

6.2. *Two-sample test for the difference between male and female mortality rates*

We further compare the mortality velocity of female and male subjects. From the Human Mortality Database (www.mortality.org), we obtain the lifetables
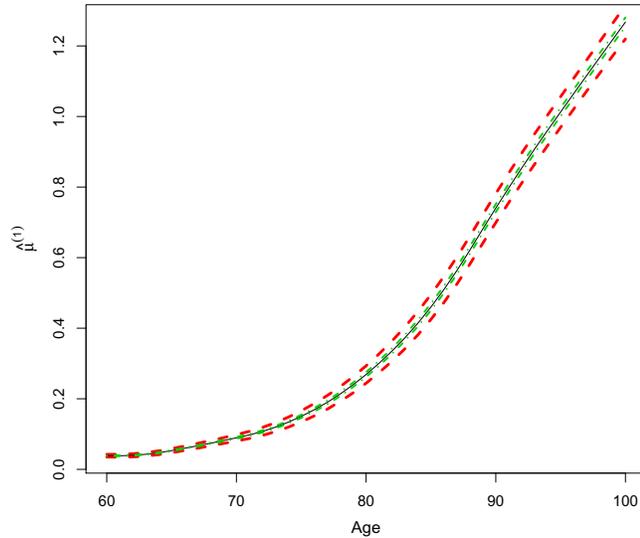
FIG 3. *Cubic spline estimator (middle solid line)* $\hat{\mu}^{(1)}(s)$ *and its* 95% *confidence bands (3.2) using "AR-1" structure (upper and lower dashed lines) and "IND" structure (upper and lower dot-dashed lines).*

for female and male subjects respectively. We focus on the same 41 countries and areas analyzed in Section 6.1 from the calendar year 1992 to year 2005. Our hypothesis of interest is:

$$H_0 : \mu_M^{(1)}(s) = \mu_F^{(1)}(s), \ \forall s \in [60, 100] \longleftrightarrow H_a : \mu_M^{(1)}(s) \neq \mu_F^{(1)}(s), \ \exists s \in [60, 100],$$

where $\mu_M^{(1)}(s)$ and $\mu_F^{(1)}(s)$ are the first derivative functions of mortality rates for males and females in these $n_1 = n_2 = 41$ countries and areas respectively.

Figure 4 depicts the cubic spline confidence bands at confidence level 0.99 (upper and lower dashed lines), with the center dashed-dotted line representing the spline estimator $\hat{\mu}_M^{(1)}(s) - \hat{\mu}_F^{(1)}(s)$ and a solid line representing zero. Resampling method in Section 4 chooses "AR-1" and "EC" as proper correlation structures for male and female groups respectively. The 99% confidence band increases slightly above the zero line and is relatively flat. After the single tuning point around age 90, it decreases under zero fairly fast. Due to few data collected for aged population, Figure 4 also indicates the variation of the discrepancy increases when people become aged. By comparing the 99% confidence band and zero reference line, the difference between male and female death rates velocity is extremely significant and female elders has lower mortality velocity than male elders at the same age before age around 90 and after this tuning point male has lower mortality velocity. To our best knowledge, there has been no demographic literature discussing the similar findings. Hence, a potentially interesting exten-
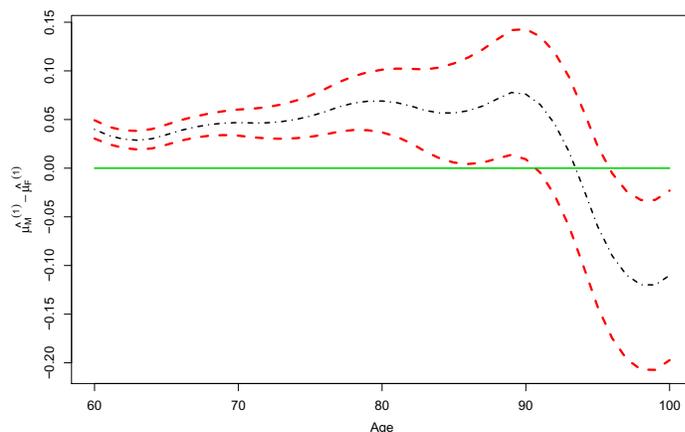
FIG 4. *Cubic spline estimator* $\hat{\mu}_M^{(1)}(s) - \hat{\mu}_F^{(1)}(s)$ *(dashed-dotted line) and the corresponding 99% confidence band (dashed lines). The solid line represents zero reference.*

sion of this work is to consider including some demographic variables and study the relationship of the mortality rate and these demographic regressors.

Acknowledgements

Appendix

For any vector $\mathbf{a} = (a_1, \ldots, a_k) \in \mathcal{R}^k$, denote the norm $\|\mathbf{a}\|_r = (|a_1|^r + \cdots + |a_k|^r)^{1/r}$, $1 \le r < +\infty$, $\|\mathbf{a}\|_\infty = \max(|a_1|, \ldots, |a_k|)$. For any function ϕ on $[0, 1]$, denote $\|\phi\|_\infty = \sup_{s \in [0,1]} |\phi(s)|$. We use C to denote a generic positive constant unless otherwise stated.

A.1. Technical assumptions

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

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

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

The technical assumptions we need are as follows:

(A1) The regression function $\mu \in \mathcal{C}^{p-1,1}[0,1]$, i.e., $\mu^{(p-1)} \in \mathcal{C}^{0,1}[0,1]$.

(A2) Let $\Gamma_{jj'}(s)$ be the (j,j')-th entry of $\mathbf{\Gamma}(s)$, and $\Gamma_{jj'}(s) \in \mathcal{C}^{0,\delta}[0,1]$ for some $\delta \in (0,1]$.

(A3) *The number of observations for each trajectory $N \gg n^\theta$ for some $\theta > \frac{1+2\nu}{2(p-\nu)}$; the number of interior knots satisfies $n^{\frac{1}{2(p-\nu)}} \ll N_\mu \ll (N/\log n)^{\frac{1}{1+2\nu}}$, $n^{\frac{1}{2p}} \ll N_G \ll n^{\frac{1}{2+2\nu}}$.*

(A4) There exists $C_G > 0$, for $j,j' = 1,\ldots,m$, such that $G_{jj'}(s,s) \geq C_G$, $s \in [0,1]$; for $k \in \{1,\ldots,\kappa\}$, $\nu = 0,1,\ldots,p-2$ and $j = 1,\ldots,m$, $\phi_{jk}^{(\nu)}(s) \in \mathcal{C}^{0,\delta}[0,1]$, $\sum_{k=1}^{\kappa} \|\phi_{jk}^{(\nu)}\|_\infty < \infty$ and as $n \to \infty$, $h_\mu^\delta \sum_{k=1}^{\kappa_n} \|\phi_{jk}^{(\nu)}\|_{0,\delta} = o(1)$ for a sequence $\{\kappa_n\}_{n=1}^\infty$ of increasing integers, with $\lim_{n\to\infty} \kappa_n = \kappa$ and the constant $\delta \in (0,1]$ as in Assumption (A2).

(A5) There exist $\eta_1,\eta_2 > 4$, such that $E|\xi_{ijk}|^{\eta_1} + E|\varepsilon_{ijl}|^{\eta_2} < +\infty$, for $1 \leq i < \infty$, $1 \leq j \leq m$, $1 \leq k \leq \kappa$, $1 \leq l < \infty$. The number κ of nonzero eigenvalues is finite or κ is infinite while the random vectors $\xi_{ik} = (\xi_{i1k},\ldots,\xi_{imk})^{\mathrm{T}}$ are i.i.d. for $1 \leq i < \infty$, $1 \leq k \leq \kappa$.

Assumptions (A1)–(A2) are typical conditions for spline smoothing; see [16, 3, 2]. In fact, (A1) and (A2) ensures the convergence rates of $\hat{\mu}^{(\nu)}$. Assumption (A3) concerns the relationship among the number of subjects, the number of observations with each trajectory and the number of knots of B-splines. Assumption (A4) guarantees that the derivatives of principal components have collectively bounded smoothness. If $\nu = 0$, (A4) is the same as Assumption (A4) in [2]. Assumption (A5) is necessary for applying Gaussian approximation of estimation error process.

A.2. Error decomposition for the spline estimators

In this section, we break the estimation error $\hat{\mu}^{(\nu)}(s) - \mu^{(\nu)}(s)$ into three terms. We begin by discussing the representation of the spline estimator $\hat{\mu}^{(\nu)}(s)$ in (2.2).

Define $\mathbf{B}(s) = (B_{1-p}(s),\ldots,B_{N_\mu}(s))^{\mathrm{T}}$, $\mathbf{X}_{N\times(N_\mu+p)} = (\mathbf{B}(1/N),\ldots,\mathbf{B}(N/N))^{\mathrm{T}}$ and $\hat{\mathbf{V}} = N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X} = (\langle B_J, B_{J'} \rangle_{2,N})_{J,J'=1-p}^{N_\mu}$. Projecting the relationship in model (2.1) onto the linear subspace of $R^{N_\mu+p}$ spanned by $\{\mathbf{B}(l/N)\}_{1\leq l \leq N}$, we obtain the following decomposition:

$$\hat{\mu}^{(\nu)}(s) = \tilde{\mu}^{(\nu)}(s) + \tilde{e}^{(\nu)}(s) + \tilde{\xi}^{(\nu)}(s), \tag{A.1}$$

where

$$\tilde{\mu}^{(\nu)}(s) = \sum_{J=1-p}^{N_\mu} \tilde{\beta}_J B_J^{(\nu)}(s), \quad \tilde{\varepsilon}^{(\nu)}(s) = \sum_{J=1-p}^{N_\mu} \tilde{a}_J B_J^{(\nu)}(s),$$

$$\tilde{\xi}^{(\nu)}(s) = \sum_{k=1}^{\kappa} \tilde{\xi}_k^{(\nu)}(s), \quad \tilde{\xi}_k^{(\nu)}(s) = \sum_{J=1-p}^{N_\mu} \tilde{\tau}_{k,J} B_J^{(\nu)}(s). \tag{A.2}$$

and the vectors $\{\tilde{\beta}_{1-p},\ldots,\tilde{\beta}_{N_\mu}\}^{\mathrm{T}}$, $\{\tilde{a}_{1-p},\ldots,\tilde{a}_{N_\mu}\}^{\mathrm{T}}$ and $\{\tilde{\tau}_{k,1-p},\ldots,\tilde{\tau}_{k,N_\mu}\}^{\mathrm{T}}$ in (A.2) are solutions to (2.3) with Y_{ijl} replaced by $\mu(l/N)$, $\varepsilon_{ij}(l/N)$ and

$\phi_{jk}(l/N)$ respectively. Alternatively, $\tilde{\mu}^{(\nu)}(s) = N^{-1}\mathbf{B}^{(\nu)}(s)\hat{\mathbf{V}}^{-1}\mathbf{X}^{\mathrm{T}}\mu$, $\tilde{e}^{(\nu)}(s) = N^{-1}\mathbf{B}^{(\nu)}(s)\hat{\mathbf{V}}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{e}$, $\tilde{\xi}_k^{(\nu)}(s) = N^{-1}\mathbf{B}^{(\nu)}(s)\hat{\mathbf{V}}^{-1}\mathbf{X}^{\mathrm{T}}\phi_k$, for $1 \leq k \leq \kappa$, where $\mu^{(\nu)} = (\mu^{(\nu)}(1/N), \ldots, \mu^{(\nu)}(N/N))^{\mathrm{T}}$ is the signal vector, $\mathbf{e} = (\bar{\varepsilon}_{..1}, \ldots, \bar{\varepsilon}_{..N})^{\mathrm{T}}$ with $\bar{\varepsilon}_{..l} = (nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\varepsilon_{ij}(l/N)$, $1 \leq l \leq N$, is the noise vector, and vector $\phi_k = (nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\xi_{ijk}\phi_{jk}$ with $\phi_{jk} = (\phi_{jk}(1/N), \ldots, \phi_{jk}(N/N))^{\mathrm{T}}$.

In the following, we denote by $Q_\varpi(\mu)$ the p-th order quasi-interpolant of μ corresponding to the knots ϖ, see equation (4.12), page 146 of [10]. According to Theorem 7.7.4 in [10], the following lemma holds.

LEMMA **1.** *There exists a constant $C > 0$, such that for $0 \leq \nu \leq p - 2$ and $\mu \in C^{p,1}[0,1]$,*

$$\left\| (\mu - Q_\varpi(\mu))^{(\nu)} \right\|_\infty \leq C \left\| \mu^{(p)} \right\|_\infty h_\mu^{p-\nu}.$$

The next lemma concerns the bias order of $\tilde{\mu}^{(\nu)}(s)$ given in (A.1).

LEMMA **2** ([3], Proposition 1). *Under Assumptions (A1) and (A3),* $\sup_{s \in [0,1]} n^{1/2}|\tilde{\mu}^{(\nu)}(s) - \mu^{(\nu)}(s)| = o(1)$.

LEMMA **3.** *Under Assumptions (A2), (A3) and (A5), $\sup_{s \in [0,1]} n^{1/2}|\tilde{e}^{(\nu)}(s)| = o_P(1)$.*

PROOF: Under Assumptions (A2), (A3) and (A5), Proposition 1 in [3] entails that $\|N^{-1}\mathbf{B}(s)\mathbf{X}^{\mathrm{T}}\mathbf{e}\|_\infty = O_{a.s.}(n^{-1/2}N^{-1/2}h_\mu^{1/2}\log^{1/2} n)$. Also under Assumption (A3), Lemma A.3 in [3] leads to $\|\hat{\mathbf{V}}^{-1}\|_\infty = O(h_\mu^{-1})$. Hence, one has

$$
\begin{aligned}
\sup_{s \in [0,1]} n^{1/2}\left|\tilde{e}^{(\nu)}(s)\right| &= \sup_{s \in [0,1]} n^{1/2}\left|N^{-1}\mathbf{B}^{(\nu)}(s)\hat{\mathbf{V}}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{e}\right| \\
&= n^{1/2}h_\mu^{-\nu}\left\|\hat{\mathbf{V}}^{-1}\right\|_\infty \left\|N^{-1}\mathbf{B}(s)\mathbf{X}^{\mathrm{T}}\mathbf{e}\right\|_\infty \\
&= O_{a.s.}\left(N^{-1/2}h_\mu^{-1/2-\nu}\log^{1/2} n\right) = o_{a.s.}(1).
\end{aligned}
$$

Thus, Lemma 3 follows from Assumption (A3).

LEMMA **4** ([2], Lemma A.7). *When Assumption (A5) holds and $\beta \in (0, 1/2)$, one has*

$$\max_{1 \leq k \leq \kappa} E\left|(nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(\xi_{ijk} - Z_{ijk,\xi})\right| = O\left(n^{\beta-1}\right), \qquad \text{(A.3)}$$

$$\max_{1 \leq l \leq N} E\left|(nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(\varepsilon_{ijl} - Z_{ijl,\varepsilon})\right| = O_{a.s.}\left(n^{\beta-1}\right). \qquad \text{(A.4)}$$

Also

$$\max_{1 \leq k \leq \kappa}(nm)^{-1}\sum_{j=1}^{m}\sum_{i=1}^{n} E\left|\xi_{ijk}\right| = O\left(n^{-1/2} + n^{\beta-1}\right). \qquad \text{(A.5)}$$

LEMMA **5.** *Under Assumptions (A2)–(A5),* $\sup_{s \in [0,1]} n^{1/2} |\tilde{\xi}^{(\nu)}(s) - \bar{\mu}^{(\nu)}(s) + \mu^{(\nu)}(s)| = o_P(1)$ *and* $P\{\sup_{s \in [0,1]} n^{1/2} |\tilde{\xi}^{(\nu)}(s)| \Omega(s,s)^{-1/2} \le Q_{1-\alpha}\} \to 1 - \alpha$ *for any* $\alpha \in (0,1)$.

PROOF: Let $\tilde{\zeta}_{\cdot k}(s) = n^{1/2}(nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m Z_{ijk,\xi} \phi_{jk}^{(\nu)}(s)$, $k = 1, \ldots, \kappa$, and define

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

$$
= \Omega(s,s)^{-1/2} \sum_{k=1}^{\kappa} \tilde{\zeta}_{\cdot k}(s).
$$

It is clear that $\tilde{\zeta}(s)$ is a Gaussian process with mean 0, variance 1 and covariance $E\tilde{\zeta}(s)\tilde{\zeta}(t) = \Omega(s,s)^{-1/2}\Omega(t,t)^{-1/2}\Omega(s,t)$, for any $s, t \in [0,1]$. Thus, $\tilde{\zeta}(s)$ has the same distribution as $\zeta(s)$ over $s \in [0,1]$.

According to Lemma 1, $\|\tilde{\phi}_{jk}^{(\nu)}\|_\infty \le c_{\phi,p}\|\phi_{jk}^{(\nu)}\|_\infty$ and $\|\tilde{\phi}_{jk}^{(\nu)} - \phi_{jk}^{(\nu)}\|_\infty \le \tilde{C}_{0,\nu}\|\phi_{jk}^{(\delta)}\|_{0,\delta}h_\mu^\nu$, $1 \le k \le \kappa$, $1 \le j \le m$. Applying (A.5) and Assumptions (A3), (A4), one has

$$
En^{1/2} \sup_{s \in [0,1]} \left| (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{\kappa} \xi_{ijk} \left\{ \phi_{jk}^{(\nu)}(s) - \tilde{\phi}_{jk}^{(\nu)}(s) \right\} \right|
$$

$$
\le Cn^{1/2}(nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m \left\{ \sum_{k=1}^{\kappa_n} E|\xi_{ijk}| \sup_{s \in [0,1]} \left| \phi_{jk}^{(\nu)}(s) \right| \right\}
$$

$$
\le Cn^{1/2}(nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m \left\{ \sum_{k=1}^{\kappa_n} E|\xi_{ijk}| \left\| \phi_{jk}^{(\nu)} \right\|_{0,\delta} h_\mu^\delta + \sum_{k=\kappa_n+1}^{\kappa} E|\xi_{ijk}| \left\| \phi_{jk}^{(\nu)} \right\|_\infty \right\}
$$

$$
\le C \left\{ \sum_{j=1}^m \sum_{k=1}^{\kappa_n} \|\phi_{jk}^{(\nu)}\|_{0,\delta} h_\mu^\delta + \sum_{j=1}^m \sum_{k=\kappa_n+1}^{\kappa} \|\phi_{jk}^{(\nu)}\|_\infty \right\} = o(1),
$$

hence

$$
n^{1/2} \sup_{s \in [0,1]} \Omega(s,s)^{-1/2} \left| (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{\kappa} \xi_{ijk}\{\phi_{jk}^{(\nu)}(s) - \tilde{\phi}_{jk}^{(\nu)}(s)\} \right| = o_P(1).
$$

$$\text{(A.6)}$$

In addition, (A.3) and Assumptions (A3), (A4) entail that

$$
En^{1/2} \sup_{s \in [0,1]} \Omega(s,s)^{-1/2} \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{\kappa} (Z_{ijk,\xi} - \xi_{ijk}) \phi_{jk}^{(\nu)}(s) \right|
$$

$$
\le Cn^{\beta-1/2} \sum_{j=1}^m \sum_{k=1}^{\kappa} \|\phi_{jk}^{(\nu)}\|_\infty = o(1).
$$

Therefore,

$$n^{1/2} \sup_{s\in[0,1]} \Omega(s,s)^{-1/2} \left| (nm)^{-1} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{\kappa} (Z_{ijk,\xi} - \xi_{ijk})\phi_{jk}^{(\nu)}(s) \right| = o_P(1).$$
(A.7)

Note that

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

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

Hence according to (A.6) and (A.7),

$$n^{1/2}\sup_{s\in[0,1]}\Omega(s,s)^{-1/2}|\bar{\mu}^{(\nu)}(s) - \mu^{(\nu)}(s) - \tilde{\xi}^{(\nu)}(s)| = o_P(1),$$

$$\sup_{s\in[0,1]}|\tilde{\zeta}^{(\nu)}(s) - n^{1/2}\Omega(s,s)^{-1/2}\{\bar{\mu}^{(\nu)}(s) - \mu^{(\nu)}(s)\}| = o_P(1),$$

which leads to the desired results.

PROOF OF THEOREM 1: Theorem 1 follows directly from the decomposition in (A.1) and Lemmas 2–5.

A.3. Proof of Theorem 3

We first show asymptotic consistency of $\hat{\phi}_{jk}^{(\nu)}$ and $\hat{\xi}_{ijk}$, for $k \geq 1$, in the following lemma.

LEMMA **6.** *Under Assumptions (A1)–(A5), one has* $\|\hat{\phi}_{jk} - \phi_{jk}\|_\infty = \|\hat{\phi}_{jk}^{(\nu)} - \phi_{jk}^{(\nu)}\|_\infty = o_P(1)$, $k \geq 1$, $j = 1, 2, \ldots m$ *and* $\nu = 1, \ldots p - 2$.

PROOF: For any fixed $j = 1, 2, \ldots m$, the proof is analogue of Lemma A8 and Theorem 4 in [3], thus omitted.

LEMMA **7.** *Under Assumptions (A1)–(A5), one has* $|\hat{\xi}_{ijk} - \xi_{ijk}| = o_P(1)$, $k \geq 1$, $j = 1, 2, \ldots m$.

PROOF: Note that

$$\begin{aligned}
\left|\hat{\xi}_{ijk} - \xi_{ijk}\right| &= \left|\int (Y_{ij}(s) - \hat{\mu}(s))\,\hat{\phi}_{jk}(s)\,ds - \int (X_{ij}(s) - \mu(s))\,\phi_{jk}(s)\,ds\right| \\
&\leq \left|\int (Y_{ij}(s) - \hat{\mu}(s))\left(\hat{\phi}_{jk}(s) - \phi_{jk}(s)\right)ds\right| \\
&\quad + \left|\int (Y_{ij}(s) - \hat{\mu}(s))\,\phi_{jk}(s)\,ds - \int (X_{ij}(s) - \mu(s))\,\phi_{jk}(s)\,ds\right|
\end{aligned}$$

$$\leq \quad \left(\|\hat{\mu} - \mu\|_\infty + \|\mathbf{e}\|_\infty + \int \left| \sum_{k=1}^{\infty} \xi_{ijk} \phi_{jk}(s) \right| ds \right) \left\| \hat{\phi}_{jk} - \phi_{jk} \right\|_\infty$$
$$- \quad (\|\hat{\mu} - \mu\|_\infty + \|\mathbf{e}\|_\infty) \|\phi_{jk}\|_\infty .$$

Since $X_{ij}(s)$ is an L_2 process, one has $\int |\sum_{k=1}^{\infty} \xi_{ijk} \phi_{jk}(s)| \, ds < \infty$. According to Theorem 2.1 and Proposition 3.3 in [4], one has $\|\hat{\mu} - \mu\|_\infty = o_P(n^{-1/2})$ and $\|\mathbf{e}\|_\infty = o_P(n^{-1/2})$. Hence, Lemma 6 leads to Lemma 7 directly.

PROOF OF THEOREM 3: Theorem 3 follows directly from Lemmas 6 and 7.

References

[1] BUNEA, F., IVANESCU, A.E. AND WEGKAMP, M. (2011), "Adaptive inference for the mean of a Gaussian process in functional data," *Journal of the Royal Statistical Society, Series B*, 73, 531–558. MR2853729

[2] CAO, G. AND WANG, L. (2014), "Simultaneous inference for repeated functional data," *Manuscript*.

[3] CAO, G., WANG, L., WANG, J. AND TODEM, D. (2012), "Spline confidence bands for functional derivatives," *Journal of Statistical Planning and Inference*, 142, 1557–1570. MR2891507

[4] CAO, G., YANG, L. AND TODEM, D. (2012), "Simultaneous inference for the mean function of dense functional data," *Journal of Nonparametric Statistics*, 24, 359–377. MR2921141

[5] CHEN, K. AND MÜLLER, H.G. (2012), "Modeling repeated functional observations," *Journal of the American Statistical Association*, 107, 1599–1609. MR3036419

[6] CHIOU, J.M. AND MÜLLER, H.G. (2009), "Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting," *Journal of the American Statistical Association*, 104, 572–585. MR2751439

[7] DE BOOR, C. (2001), *A Practical Guide to Splines*, New York: Springer-Verlag. MR1900298

[8] DEGRAS, D.A. (2011), "Simultaneous confidence bands for nonparametric regression with functional data," *Statistica Sinica*, 21, 1735–1765. MR2895997

[9] DI, C., CRAINICEANU, C.M., CAFFO, B.S. AND PUNJABI, N.M. (2009), "Multilevel functional principal component Models," *Annals of Applied Statistics*, 3, 458–488. MR2668715

[10] DEVORE, R. AND LORENTZ, G. (1993), *Constructive Approximation: Polynomials and Splines Approximation*, Berlin: Springer-Verlag. MR1261635

[11] FERRATY, F. AND VIEU, P. (2006), *NonParametric Functional Data Analysis: Theory and Practice*, New York: Springer Series in Statistics, Springer. MR2229687

[12] *Human Mortality Database*, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at

www.mortality.org or www.humanmortality.de (data download on January 24, 2013 and May 9, 2013).

[13] LIANG, K.Y. AND ZEGER, S.L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22. MR0836430

[14] LIU, B. AND MÜLLER, H.G. (2009), "Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics," *Journal of American Statistical Association*, 104, 704–717. MR2541589

[15] RAMSAY, J.O. AND SILVERMAN, B.W. (2005), *Functional Data Analysis*, Second Edition, New York: Springer Series in Statistics, Springer. MR2168993

[16] WANG, J. AND YANG, L. (2009), "Polynomial spline confidence bands for regression curves," *Statistica Sinica*, 19, 325–342. MR2487893

[17] YAO, F., MÜLLER, H.G. AND WANG, J.L. (2005), "Functional data analysis for sparse longitudinal data," *Journal of the American Statistical Association*, 100, 577–590. MR2160561

[18] YUAN, Y., GILMORE, J.H., GENG, X., MARTIN, S., CHEN, K., WANG, J.L. AND ZHU, H. (2014), "FMEM: functional mixed effects modeling for the analysis of longitudinal white matter tract data," *Neuroimage*, 84, 753–764.

[19] ZHOU, S. AND WOLFE, D. A. (2000), "On derivative estimation in spline regression," *Statistica Sinica* 10, 93–108. MR1742102

[20] ZHU, H., LI, R. AND KONG, L. (2012), "Multivariate varying coefficient model for functional responses," *Annals of Statistics*, 40, 2634–2666. MR3097615