Journal of Statistical Planning and Inference 142 (2012) 1557–1570

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Spline confidence bands for functional derivatives

Guanqun Cao [a], Jing Wang [b], Li Wang [c,*], David Todem [a]

[a] Michigan State University, United States
[b] University of Illinois at Chicago, United States
[c] University of Georgia, United States

ARTICLE INFO

Article history:
Received 17 April 2011
Received in revised form
31 October 2011
Accepted 4 January 2012
Available online 20 January 2012

Keywords:
B-spline
Eigenfunctions
Functional data
Karhunen–Loève representation
Semiparametric efficiency

ABSTRACT

We develop in this paper a new procedure to construct simultaneous confidence bands for derivatives of mean curves in functional data analysis. The technique involves polynomial splines that provide an approximation to the derivatives of the mean functions, the covariance functions and the associated eigenfunctions. We show that the proposed procedure has desirable statistical properties. In particular, we first show that the proposed estimators of derivatives of the mean curves are semiparametrically efficient. Second, we establish consistency results for derivatives of covariance functions and their eigenfunctions. Most importantly, we show that the proposed spline confidence bands are asymptotically efficient as if all random trajectories were observed with no error. Finally, the confidence band procedure is illustrated through numerical simulation studies and a real life example.

1. Introduction

Functional data analysis (FDA) is a relatively new field in statistics, where the data are typically curves or high-dimensional surfaces. In exploratory FDA, it is often of interest to estimate the mean functions; see for example, Ramsay and Silverman (2005), Yao et al. (2005a,b), Ferraty and Vieu (1996), Li and Hsing (2010) and Cao et al. (forthcoming). In some settings, however, estimation and inference of derivatives of the mean functions in FDA are of equal importance. For example, in economics, consistent and direct estimation of derivatives are essential for estimating elasticities, returns to scale, substitution rates and average derivatives. Often, these index (derivative) functions are as interesting as the mean functions themselves. Another example is in the fields of engineering and biomedical sciences, where the estimation of velocity and acceleration are of great importance in addition to obtaining a smooth curve of the measurements.

The problem of estimation and inference of derivatives for functional data is very challenging; see Ramsay and Silverman (2005), Liu and Müller (2009), and Hall et al. (2009) for some discussions. Existing methodologies for derivatives of the regression function in FDA often rely on a pointwise analysis. For example, in Liu and Müller (2009) the theoretical focus was primarily on obtaining consistency and asymptotic normality of the proposed estimators, thereby providing the necessary ingredients to construct pointwise confidence intervals. This approach is important but its usefulness in conducting global inferences is limited. To our knowledge, we are not aware of any methodology that provides simultaneous confidence bands for functional derivatives in FDA. In this paper, we develop such methodology with the primary aim to better understand the variability and shape of the mean curve.

* Corresponding author.
E-mail addresses: cao@stt.msu.edu (G. Cao), wangjing@math.uic.edu (J. Wang), lilywang@uga.edu (L. Wang), todem@msu.edu (D. Todem).

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2012.01.009

Nonparametric simultaneous confidence bands are powerful tools for global inference of functions. Some work has been conducted to study the simultaneous confidence band of the mean curves for FDA; see Degras (2011), Ma et al. (2012) and Cao et al. (forthcoming). The research work on confidence bands for functional derivatives is actually sparse. This is partially due to the technical difficulty to formulate such bands for FDA and establish the associated theoretical properties.

Some smoothing tools are necessary to construct the confidence bands. Popular smoothing methods include kernels (Gasser and Müller, 1984; Härdle, 1989; Xia, 1998; Claeskens and Van Keilegom, 2003), local polynomials (Fan and Gijbels, 1996), splines (Wahba, 1990; Stone, 1994) and series expansion methods (Morris and Carroll, 2006). In this paper, we use B-splines, which can be readily implemented due to their explicit expression, to construct the bands. B-spline approximation has also been employed to estimate the functional mixed-effect models in Shi et al. (1996) and Rice and Wu (2001), and to study functional data via principle components in Yao and Lee (2006) and Zhou et al. (2008). Other works include Zhou et al. (1998) and Wang and Yang (2009a) who have proposed B-spline confidence bands for regression functions.

The proposed confidence bands are asymptotically the same as if all the random trajectories are correctly recorded over the entire interval. As discussed in Section 2.3, the estimators are semiparametrically efficient thereby providing partial theoretical justification for treating functional data as perfectly recorded random curves over the entire data range, as in Cao et al. (forthcoming).

The rest of the paper is organized as follows. In Section 2, we introduce the model and the spline estimators for the mean curves and their derivatives. Section 3 presents the simultaneous confidence bands for the derivatives of the mean curves. Specifically, in Section 3.1, we show that the bands have asymptotically correct coverage probabilities; and in Section 3.2, we discuss how to estimate the unknown components involved in the band construction and other issues of the implementation. Section 4 reports findings from a simulation study and a real data set. Proofs of technical results are relegated to Appendix.

2. Models and spline estimators

2.1. Models

We consider a collection of trajectories $\{X_i(t)\}_{i=1}^n$ which are i.i.d. realizations of a smooth random function $X(t)$, defined on a continuous interval $\mathcal{T}$. Assume that $\{X(t), t\in\mathcal{T}\}$ is a $L^2(\mathcal{T})$ process, i.e. $E\int_{\mathcal{T}} X^2(t)\,dt<+\infty$, and define the mean and covariance functions as $\mu(t)=E\{X(t)\}$ and $G(t,s)=\mathrm{cov}\{X(t),X(s)\}$, $t,s\in\mathcal{T}$. The covariance function is a symmetric nonnegative-definite function with a spectral decomposition, $G(t,s)=\sum_{k=1}^{\infty}\lambda_k\phi_k(t)\phi_k(s)$, where $\lambda_1\ge\lambda_2\ge\cdots\ge 0$, are the eigenvalues satisfying $\sum_{k=1}^{\infty}\lambda_k<\infty$, and $\{\phi_k(t)\}_{k=1}^{\infty}$ are the corresponding eigenfunctions that form an orthonormal basis. By the standard Karhunen–Loève representation (Hall and Hosseini-Nasab, 2006), $X_i(t)=\mu(t)+\sum_{k=1}^{\infty}\xi_{ik}\phi_k(t)$, where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variance λ_k. In what follows, we assume that $\lambda_k=0$, for $k>\kappa$, where κ is a positive integer or ∞.

We consider a typical functional data setting where $X_i(\cdot)$ is recorded on a regular grid in $\mathcal{T}$, and assumed to be contaminated with measurement errors. Without loss of generality, we take $\mathcal{T}=[0,1]$. Then the observed data are $Y_{ij}=X_i(T_{ij})+\sigma(T_{ij})\varepsilon_{ij}$, for $1\le i\le n$, $1\le j\le N$, where $T_{ij}=j/N$, ε_{ij} are independent random errors with $E(\varepsilon_{ij})=0$ and $E(\varepsilon_{ij}^2)=1$, and $\sigma(\cdot)$ is the standard deviation of the measurement errors. By the Karhunen–Loève representation, the observed data can be written as

$$Y_{ij}=\mu(j/N)+\sum_{k=1}^{\kappa}\xi_{ik}\phi_k(j/N)+\sigma(j/N)\varepsilon_{ij}, \tag{1}$$

where $\mu(\cdot)$, $\sigma(\cdot)$ and $\{\phi_k(\cdot)\}_{k=1}^{\kappa}$ are smooth but unknown functions of t. In addition, $\{\phi_k(\cdot)\}_{k=1}^{\kappa}$ are further subject to constraints $\int_0^1\phi_k^2(t)\,dt=1$, and $\int_0^1\phi_k(t)\phi_{k'}(t)\,dt=0$, for $k'\ne k$.

2.2. Spline estimators

We first introduce some notation of the B-spline space. Divide the interval $\mathcal{T}=[0,1]$ into $(N_\mu+1)$ subintervals $I_J=[\omega_J,\omega_{J+1})$, $J=0,\ldots,N_\mu-1$, $I_{N_\mu}=[\omega_{N_\mu},1]$, where $\varpi_\mu:=\{\omega_J\}_{J=1}^{N_\mu}$ is a sequence of equally spaced points, called interior knots. Let $\mathcal{H}^{(p-2)}$ be the polynomial spline space of order p on [0,1]. This space consists of all $p-2$ times continuously differentiable functions on [0,1] that are polynomials of degree $p-1$ on each interval I_J. The J-th B-spline of order p is denoted by $B_{J,p}$. We augment the boundary and the number of interior knots as $\omega_{1-p}=\cdots=\omega_{-1}=\omega_0=0<\omega_1<\cdots<\omega_{N_\mu}<1=\omega_{N_\mu+1}=\cdots=\omega_{N_\mu+p}$, in which $\omega_J=Jh_\mu$, $J=0,1,\ldots,N_\mu+1$ and $h_\mu=1/(N_\mu+1)$ is the distance between neighboring knots.

Following Cao et al. (forthcoming), we estimate the mean function $\mu(\cdot)$ in (1) by

$$\hat{\mu}(\cdot)=\underset{g(\cdot)\in\mathcal{H}^{(p-2)}}{\arg\min}\sum_{i=1}^{n}\sum_{j=1}^{N}\{Y_{ij}-g(j/N)\}^2=\sum_{J=1-p}^{N_\mu}\hat{b}_{J,p}B_{J,p}(\cdot),$$

where the coefficients

$$\hat{\mathbf{b}}_p = \{\hat{b}_{1-p,p},\ldots,\hat{b}_{N_\mu,p}\}^{\mathrm{T}} = \underset{\mathbb{R}^{N_\mu+p}}{\arg\min} \sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij} - \sum_{J=1-p}^{N_\mu} b_{J,p}B_{J,p}(j/N)\right\}^2.$$

Let $\mathbf{Y} = (\overline{Y}_{\cdot 1},\ldots,\overline{Y}_{\cdot N})^{\mathrm{T}}$ and $\overline{Y}_{\cdot j} = n^{-1}\sum_{i=1}^{n} Y_{ij}$, $1 \le j \le N$. Applying elementary algebra, one obtains

$$\hat{\mu}(t) = \mathbf{B}_p(t)(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{Y}, \tag{2}$$

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

We denote by $\mu^{(\nu)}(t)$ the ν-th order derivative of $\mu(t)$ with respect to t. Since $\hat{\mu}(t)$ is an estimator of $\mu(t)$, it is natural to consider $\hat{\mu}^{(\nu)}(t)$ as the estimator of $\mu^{(\nu)}(t)$, for any $\nu = 1,\ldots,p-2$, i.e.

$$\hat{\mu}^{(\nu)}(t) = \mathbf{B}_p^{(\nu)}(t)(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{Y}, \tag{3}$$

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

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

Therefore, $\mathbf{B}_p^{(\nu)}(t) = \mathbf{B}_{p-\nu}(t)\mathbf{D}_{(\nu)}^{\mathrm{T}}$, in which $\mathbf{D}_{(\nu)} = \mathbf{D}_1 \cdots \mathbf{D}_{\nu-1}\mathbf{D}_\nu$, with matrix

$$\mathbf{D}_l = (p-l)\begin{pmatrix} \frac{-1}{\omega_1-\omega_{1-p+l}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\omega_1-\omega_{1-p+l}} & \frac{-1}{\omega_2-\omega_{2-p+l}} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\omega_2-\omega_{2-p+l}} & \frac{-1}{\omega_3-\omega_{3-p+l}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{\omega_{N_\mu+p-l}-\omega_{N_\mu}} \end{pmatrix}$$

for $1 \le l \le \nu \le p-2$, which is the same as Eq. (6) in Zhou and Wolfe (2000).

2.3. Asymptotic convergence rate

Define the following "infeasible estimator" of function $\mu^{(\nu)}$

$$\overline{\mu}^{(\nu)}(t) = \overline{X}^{(\nu)}(t) = n^{-1}\sum_{i=1}^{n} X_i^{(\nu)}(t), \quad t \in [0,1]. \tag{4}$$

The term "infeasible", borrowed from Cao et al. (forthcoming), refers to the fact that $\overline{\mu}^{(\nu)}(\cdot)$ would be a natural estimator of $\mu^{(\nu)}(\cdot)$ if all random curves $X_i^{(\nu)}(\cdot)$ were observed. In the following, we want to show that the spline estimator $\hat{\mu}^{(\nu)}(\cdot)$ in (3) is asymptotically equivalent to $\overline{\mu}^{(\nu)}(\cdot)$.

We break the error $\hat{\mu}^{(\nu)}(\cdot) - \mu^{(\nu)}(\cdot)$ into three terms. Let $\overline{\varepsilon}_{\cdot j} = n^{-1}\sum_{i=1}^{n}\varepsilon_{ij}$, $1 \le j \le N$. Denote the signal vector, the noise vector and the eigenfunction vectors by $\boldsymbol{\mu} = (\mu(1/N),\ldots,\mu(N/N))^{\mathrm{T}}$, $\mathbf{e} = (\sigma(1/N)\overline{\varepsilon}_{\cdot 1},\ldots,\sigma(N/N)\overline{\varepsilon}_{\cdot N})^{\mathrm{T}}$ and $\boldsymbol{\phi}_k = (\phi_k(1/N),\ldots,\phi_k(N/N))^{\mathrm{T}}$. Projecting the relationship in model (2) onto the linear subspace of $\mathbb{R}^{N_\mu+p}$ spanned by $\{B_{J,p}(j/N)\}_{1\le j\le N, 1-p\le J\le N_\mu}$, we obtain the following crucial decomposition:

$$\hat{\mu}^{(\nu)}(t) = \tilde{\mu}^{(\nu)}(t) + \tilde{e}^{(\nu)}(t) + \tilde{\xi}^{(\nu)}(t), \tag{5}$$

where $\tilde{\mu}^{(\nu)}(t) = \Gamma^{(\nu)}(t)\boldsymbol{\mu}$, $\tilde{e}^{(\nu)}(t) = \Gamma^{(\nu)}(t)\mathbf{e}$ and $\tilde{\xi}^{(\nu)}(t) = \sum_{k=1}^{\kappa}\overline{\xi}_{\cdot k}\Gamma^{(\nu)}(t)\boldsymbol{\phi}_k$ with $\Gamma^{(\nu)}(t) = \mathbf{B}_p^{(\nu)}(t)(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}$ and $\overline{\xi}_{\cdot k} = n^{-1}\sum_{i=1}^{n}\xi_{ik}$, $1 \le k \le \kappa$.

The following proposition provides asymptotic properties of the three terms.

Proposition 1. *Under Assumptions* (A1)–(A6) *in Appendix, one has*

$$\sup_{t\in[0,1]}\left|\tilde{\mu}^{(\nu)}(t) - \mu^{(\nu)}(t)\right| = o(n^{-1/2}), \tag{6}$$

$$\sup_{t\in[0,1]}\left|\tilde{\xi}^{(\nu)}(t) - (\overline{\mu}^{(\nu)}(t) - \mu^{(\nu)}(t))\right| = o_P(n^{-1/2}), \tag{7}$$

$$\sup_{t\in[0,1]}\left|\tilde{e}^{(\nu)}(t)\right| = o_P(n^{-1/2}). \tag{8}$$

Appendix A.2 contains proofs for the above proposition, which together with (5), leads to the following semiparametric efficiency result.

Theorem 1. *Under Assumptions* (A1)–(A6) *in Appendix, the B-spline estimator* $\hat{\mu}^{(\nu)}$ *is asymptotically equivalent to* $\overline{\mu}^{(\nu)}$ *with the* $\sqrt{n}$ *approximation power, i.e.*

$$\sup_{t\in[0,1]}\left|\hat{\mu}^{(\nu)}(t)-\overline{\mu}^{(\nu)}(t)\right| = o_P(n^{-1/2}).$$

Remark 1. Since the "infeasible estimator" $\overline{\mu}^{(\nu)}(t)$ is the sample average of i.i.d. trajectories $\{X_i(t)\}_{i=1}^n$, an application of the central limit theorem gives $\sup_{t\in[0,1]}\left|\overline{\mu}^{(\nu)}(t)-\mu^{(\nu)}(t)\right| = O_P(n^{-1/2})$. Thus combining with the results in Theorem 1, one has

$$\sup_{t\in[0,1]}\left|\hat{\mu}^{(\nu)}(t)-\mu^{(\nu)}(t)\right| = O_P(n^{-1/2}).$$

3. Confidence bands

In this section, we develop the simultaneous confidence bands for the derivative function $\mu^{(\nu)}(\cdot)$.

3.1. Asymptotic confidence bands

Let $\Sigma(\cdot,\cdot)$ be a positive definite function, and defined as $\Sigma(t,s) = \sum_{k=1}^{\kappa}\lambda_k\phi_k^{(\nu)}(t)\phi_k^{(\nu)}(s)$, $t,s\in[0,1]$. Denote by $\zeta(t)$, $t\in[0,1]$ a standardized Gaussian process such that $E\zeta(t)\equiv 0$, $E\zeta^2(t)\equiv 1$ with covariance function $E\zeta(t)\zeta(s) = \Sigma(t,s)\{\Sigma(t,t)\Sigma(s,s)\}^{-1/2}$, $t,s\in[0,1]$. Denote by $q_{1-\alpha}$ the $100(1-\alpha)$th percentile of the absolute maxima distribution of $\zeta(t)$, $t\in[0,1]$, i.e. $P\{\sup_{t\in[0,1]}\left|\zeta(t)\right| \le q_{1-\alpha}\} = 1-\alpha, \forall\alpha\in(0,1)$.

Theorem 2. *Under Assumptions* (A1)–(A6) *in Appendix*, $\forall\alpha\in(0,1)$, *as* $n\to\infty$,

$$P\left\{\sup_{t\in[0,1]} n^{1/2}\left|\overline{\mu}^{(\nu)}(t)-\mu^{(\nu)}(t)\right|\Sigma(t,t)^{-1/2} \le q_{1-\alpha}\right\}\to 1-\alpha.$$

Applying Theorems 1 and 2 gives asymptotic confidence bands for $\mu^{(\nu)}(t)$, $t\in[0,1]$.

Corollary 1. *Under Assumptions* (A1)–(A6) *in Appendix*, $\forall\alpha\in(0,1)$, *as* $n\to\infty$, *an asymptotic* $100\times(1-\alpha)\%$ *exact confidence band for* $\mu^{(\nu)}(t)$ *is*

$$P\{\mu^{(\nu)}(t)\in\hat{\mu}^{(\nu)}(t)\pm n^{-1/2}q_{1-\alpha}\Sigma(t,t)^{1/2},\quad t\in[0,1]\}\to 1-\alpha.$$

3.2. Implementation

When constructing the confidence bands, one needs to estimate the unknown function $\Sigma(t,s)$. Note that $\Sigma(t,s) = G(t,s)$, when $\nu = 0$. Following Liu and Müller (2009), we estimate $\phi_k^{(\nu)}$ through the derivatives of $G(t,s)$. According to Cao et al. (2011), $G(t,s)$ is estimated by

$$\hat{G}(t,s) = \sum_{J,J'=1-p}^{N_G}\hat{b}_{JJ'}B_{J,p}(t)B_{J',p}(s), \tag{9}$$

where $\hat{R}_{\cdot jj'} = n^{-1}\sum_{i=1}^n\{Y_{ij}-\hat{\mu}(j/N)\}\{Y_{ij'}-\hat{\mu}(j'/N)\}$, $1\le j\ne j'\le N$, N_G is the number of interior knots for B-spline, and the coefficients

$$\{\hat{b}_{JJ'}\}_{J,J'=1-p}^{N_G} = \mathop{\arg\min}_{\mathbb{R}^{N_G+p}\otimes\mathbb{R}^{N_G+p}}\sum_{j\ne j'}^{N}\left\{\hat{R}_{\cdot jj'} - \sum_{1-p\le J,J'\le N_G} b_{JJ'}B_{J,p}(j/N)B_{J',p}(j'/N)\right\}^2$$

with "$\otimes$" being the tensor product of two spaces. They showed that $\hat{G}$ converges to G as n goes to ∞. In this section, we further show that $\hat{G}$ and G are asymptomatically equivalent up to the ν-th partial derivative. We define the ν-th derivative with respect to s for $G(t,s)$ and $\hat{G}(t,s)$ as

$$G^{(0,\nu)}(t,s) = \frac{\partial^\nu}{\partial s^\nu}G(t,s),\quad \hat{G}^{(0,\nu)}(t,s) = \frac{\partial^\nu}{\partial s^\nu}\hat{G}(t,s) = \sum_{J,J'}\hat{b}_{JJ'}B_{J,p}(t)B_{J',p}^{(\nu)}(s). \tag{10}$$

Theorem 3. *Under Assumptions* (A1)–(A6), *one has*

$$\sup_{(t,s)\in[0,1]^2}\left|\hat{G}^{(0,\nu)}(t,s)-G^{(0,\nu)}(t,s)\right| = o_P(1),\quad 1\le\nu\le p-2.$$

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

According to Liu and Müller (2009), we estimate the v-th derivative of eigenfunctions $\hat{\phi}_k^{(v)}$ using the following eigenequations:

$$\frac{d^v}{ds^v}\int_0^1 \hat{G}(t,s)\hat{\phi}_k(t)\,dt = \int_0^1 \frac{\partial^v}{\partial s^v}\hat{G}(t,s)\hat{\phi}_k(t)\,dt = \hat{\lambda}_k\hat{\phi}_k^{(v)}(s), \tag{11}$$

where $\hat{\phi}_k$ are subject to $\int_0^1 \hat{\phi}_k^2(t)\,dt = 1$ and $\int_0^1 \hat{\phi}_k(t)\hat{\phi}_{k'}(t)\,dt = 0$ for $k' < k$. If N is sufficiently large, the left hand side of (11) can be approximated by $(1/N)\sum_{j=1}^N \hat{G}^{(0,v)}(j/N,j'/N)\hat{\phi}_k(j/N)$. Then we estimate $\Sigma(t,s)$ by $\hat{\Sigma}(t,s) = \sum_{k=1}^{\kappa}\hat{\lambda}_k\hat{\phi}_k^{(v)}(t)\hat{\phi}_k^{(v)}(s)$.

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

Theorem 4. *Under Assumptions* (A1)–(A6), *one has*

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

The proof of Theorem 4 is given in Appendix A.4.

In practice, we choose the first L positive eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_L > 0$ by eigenvalue decomposition of $\hat{G}(t,s)$. Then we apply a standard criterion in Müller (2009), to choose the number of eigenfunctions, i.e. $\kappa = \arg\min_{1\leq l\leq L}\{\sum_{k=1}^{l}\hat{\lambda}_k/\sum_{k=1}^{L}\hat{\lambda}_k > 0.95\}$. Müller (2009) suggests the "pseudo-AIC" and this simple method of counting the percentage of variation explained can be used to choose the number of principal components. The simple method performed well in our simulations and is used for our numerical studies.

To construct the confidence bands, we use cubic splines to estimate the mean and covariance functions and their first order derivatives. Generalized cross-validation is used to choose the number of knots N_μ (from 2 to 20), to smooth out the mean function. According to Assumption (A3), the number of knots for smoothing the covariance function is taken to be $N_G = [n^{1/(2p)}\log(n)]$, where $[a]$ denotes the integer part of a.

Finally, in order to estimate $q_{1-\alpha}$, we generate i.i.d. standard normal variables $Z_{k,b}$, $1 \leq k \leq \kappa$, $b = 1,\ldots,5000$. Let $\hat{\zeta}_b(t) = \hat{\Sigma}(t,t)^{-1/2}\sum_{k=1}^{\kappa}\sqrt{\hat{\lambda}_k}Z_{k,b}\hat{\phi}_k^{(v)}(t)$, $t \in [0,1]$, $q_{1-\alpha}$ can be estimated by $100(1-\alpha)$-th percentile of $\{\sup_{t\in[0,1]}|\hat{\zeta}_b(t)|\}_{b=1}^{5000}$. Therefore, in application we recommend the following band:

$$\hat{\mu}^{(v)}(t) \pm n^{-1/2}\hat{\Sigma}(t,t)^{1/2}\hat{q}_{1-\alpha}, \quad t \in [0,1]. \tag{12}$$

4. Numerical studies

4.1. Simulated examples

To illustrate the finite-sample performance of the confidence band in (12), we generate data from the following:

$$Y_{ij} = \mu(j/N) + \sum_{k=1}^{\kappa}\xi_{ik}\phi_k(j/N) + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{i.i.d}{\sim} N(0,0.1^2)$$

for $1 \leq i \leq n$, $1 \leq j \leq N$. We consider two scenarios.

Model I: $\mu(t) = 5t + 4\sin(2\pi(t-0.5))$, $\phi_1(t) = -\sqrt{2}\cos(2\pi(t-0.5))$, $\phi_2(t) = \sqrt{2}\sin(4\pi(t-0.5))$, $\xi_{ik} \sim N(0,\lambda_k)$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\kappa = 2$;

Model II: $\mu(t) = 4t + (1/\sqrt{2\pi}0.1)\exp(-(t-0.5)^2/2(0.1)^2)$, $\phi_k(t) = \sqrt{2}\sin(\pi kt)$, $\xi_{ik} \sim N(0,\lambda_k)$, $\lambda_k = 2^{-(k-1)}$, $k = 1,2,\ldots,\kappa = 8$.

The second case has similar design as in Simulation C of Liu and Müller (2009). We use the proposed method in (12) and its "oracle" version with true $\Sigma(t,t)$ to construct the confidence bands for $\mu^{(1)}(\cdot)$, respectively, in both studies. We consider

Table 1
Coverage rates of the spline confidence bands in Model I.

n	N	95%		99%	
		Est.	Oracle	Est.	Oracle
30	30	0.756	0.831	0.901	0.936
	60	0.833	0.900	0.926	0.969
50	50	0.824	0.863	0.932	0.933
	100	0.856	0.904	0.947	0.979
100	100	0.851	0.897	0.943	0.971
	200	0.856	0.907	0.949	0.971
200	200	0.866	0.910	0.950	0.972
	400	0.869	0.944	0.959	0.987

two confidence levels: $1-\alpha=0.95, 0.99$. The number of trajectories n is taken to be 30, 50, 100, 200, and for each n, we try different numbers of observations on the trajectory. Each simulation consists of 1000 Monte Carlo samples.

We evaluate the coverage of the bands over 200 equally spaced points on [0,1] and test whether the true functions are covered by the confidence bands at these points. Tables 1 and 2 show the empirical coverage probabilities out of 1000 replications for Models I and II, respectively. From Tables 1 and 2, we observe that coverage probabilities for both estimated bands and "oracle" bands approach the nominal levels, which show positive confirmation of Theorem 2. In most of the scenarios the "oracle" confidence bands outperform the estimated bands, and the "oracle" bands arrive at about the

Table 2
Coverage rates of the spline confidence bands in Model II.

n	N	95%		99%	
		Est.	Oracle	Est.	Oracle
30	30	0.650	0.730	0.830	0.854
	60	0.695	0.839	0.860	0.941
50	50	0.817	0.911	0.930	0.982
	100	0.830	0.929	0.933	0.986
100	100	0.858	0.940	0.948	0.986
	200	0.876	0.939	0.960	0.986
200	200	0.874	0.939	0.949	0.980
	400	0.889	0.946	0.963	0.991

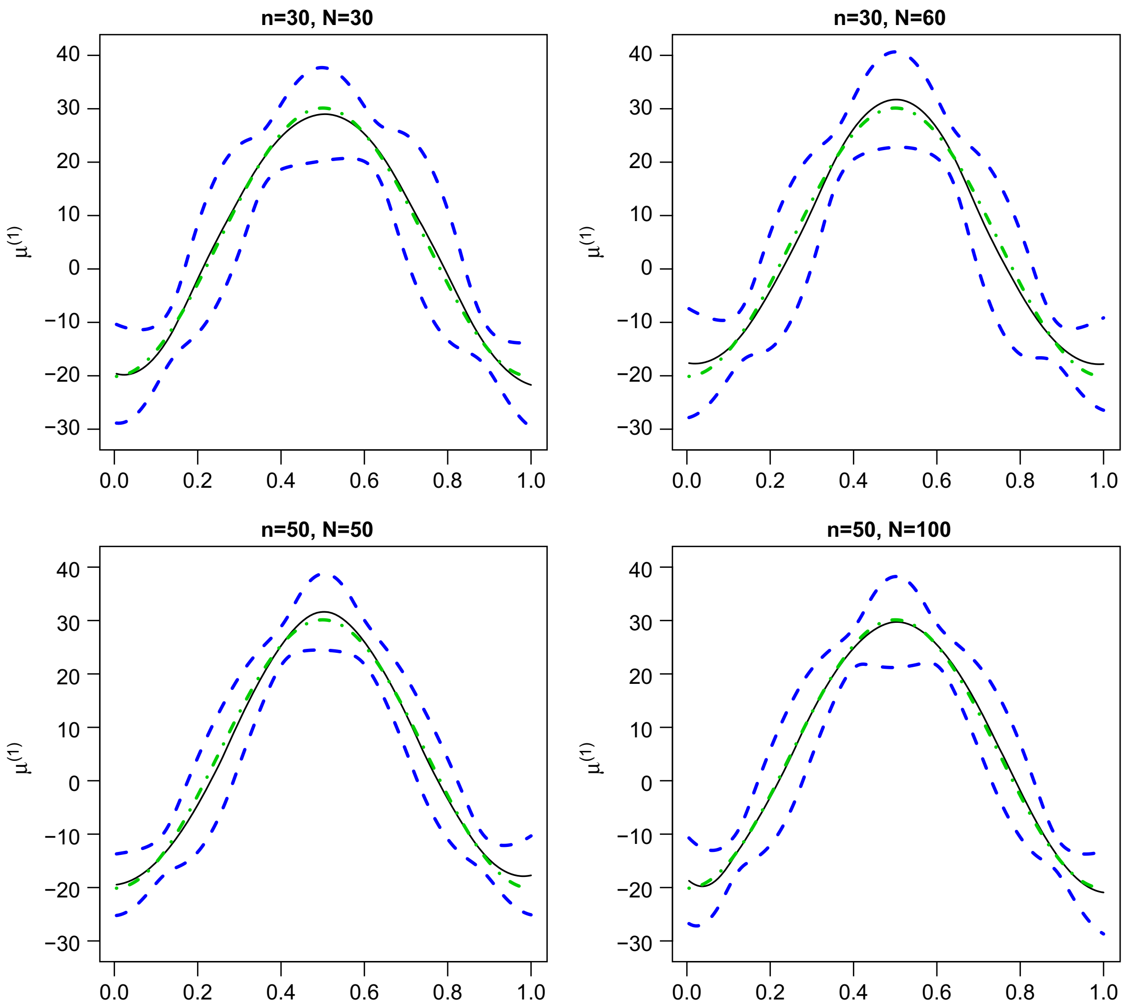

Fig. 1. Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $\mu^{(1)}(t)$ (solid line) in Model I.

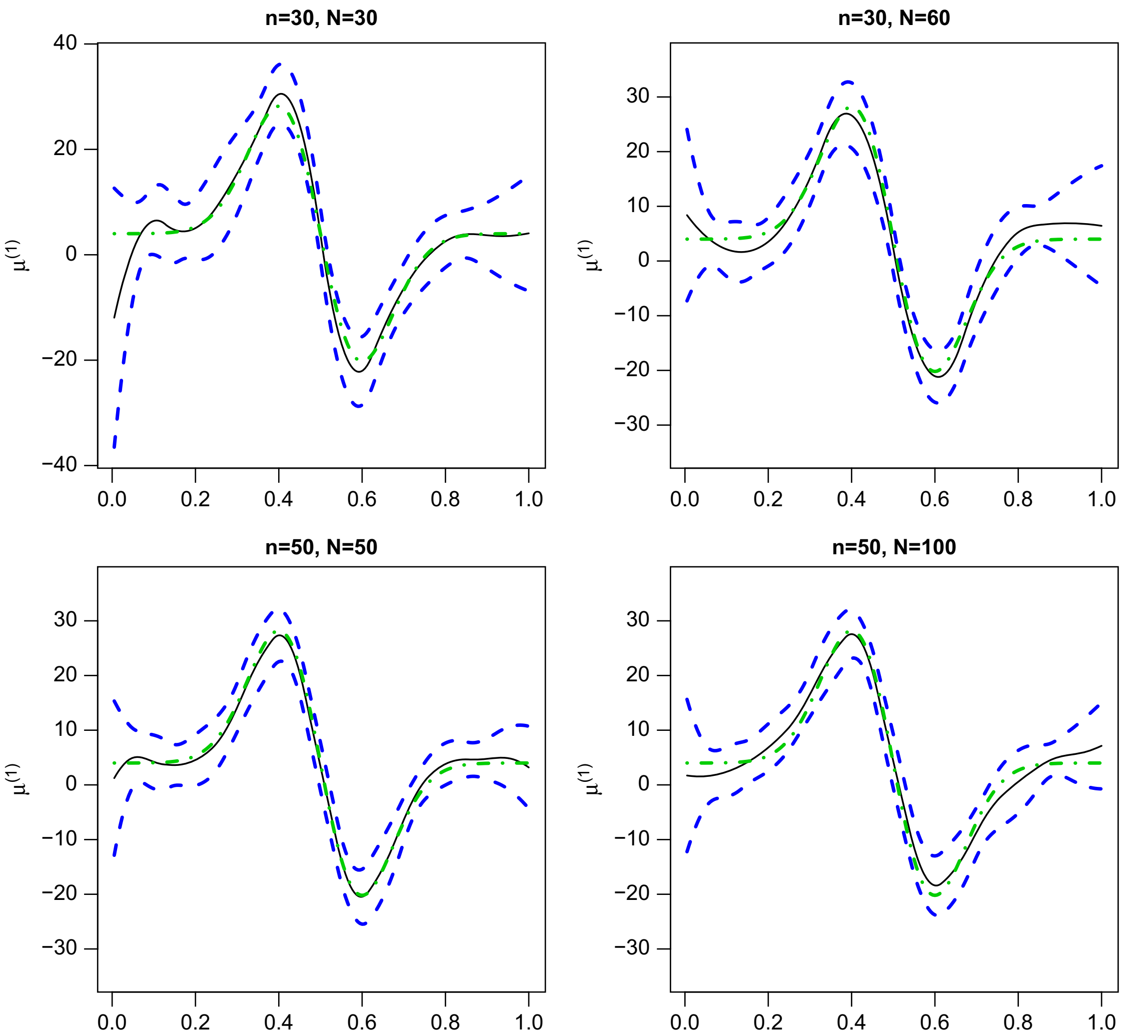

Fig. 2. Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $\mu^{(1)}(t)$ (solid line) in Model II.

nominal coverage for large n and N. The convergence rates of estimated bands are slower than those "oracle" bands, but the convergence trend to nominal level is clearly.

Figs. 1 and 2 show the estimated functions and their 99% confidence bands for the first order derivative curve $\mu^{(1)}(\cdot)$ for Models I and II, respectively. As expected when n and N increase, the confidence band is narrower and the cubic spline estimator is closer to the true derivative curve. For Model I, the boundary effects in all four panels are almost unnoticeable. For Model II, there seems to be some boundary effects for small n and N, which are attenuated as n and N increase.

4.2. Tecator data

Here we apply the proposed method to the Tecator dataset, which can be downloaded from http://lib.stat.cmu.edu/datasets/tecator. This data set contains measurements on $n=240$ meat samples and for each sample $N=100$ channel near-infrared spectrum of absorbance measurements were obtained. The Feed Analyzer worked in the wavelength range 850–1050 nm. Fig. 3 shows the estimated mean absorbance measurements $\mu(\cdot)$ in the upper panel and its estimated first order derivative $\mu^{(1)}(\cdot)$ in the lower panel. Their 99% confidence bands (dashed lines) are also included in the figure, both bands have similar band width around 0.1–0.2 even though the bands for $\mu^{(1)}(\cdot)$ looks much narrower in the figure.

As shown in Fig. 3, in the region of 850–950 nm the derivative estimate of mean absorbance is increasing gradually above 0, which corroborates with the convex behavior of the corresponding estimated mean function. For wavelength between 950 and 970 nm, the big bump in the derivative graph is consistent with the changing pattern of the mean estimate before it reaches the turning point at around 970 nm. When wavelength is larger than 970 nm, its estimated derivative turns negative and is relatively flat after 1000 nm, and it is in accordance with the quick dip of the mean absorbance and a linear decreasing trend for wavelength larger than 1000 nm.

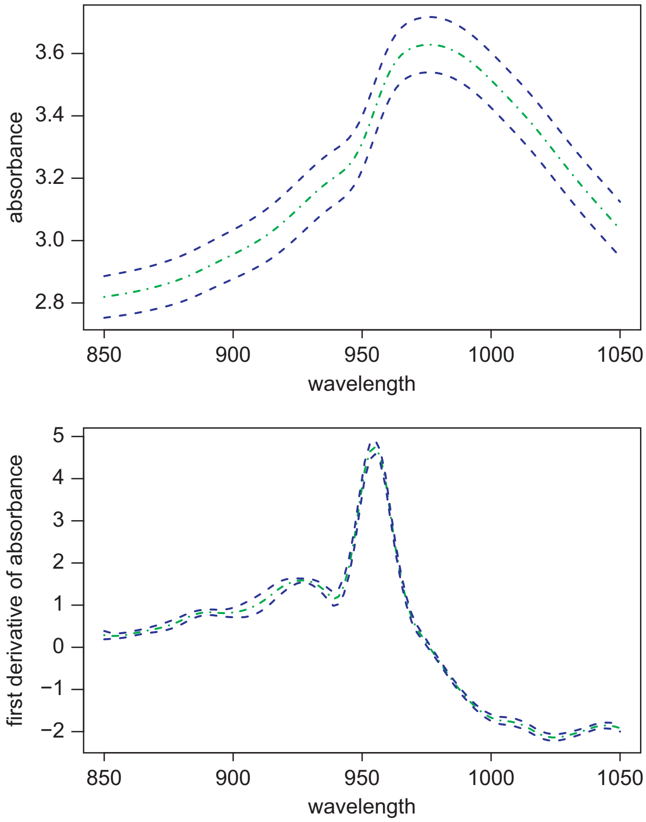

Fig. 3. Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of the mean function and its first order derivative.

Acknowledgment

The first and fourth authors' work was supported in part by NCI/NIH K-award, 1K01 CA131259 and its supplement from the 2009 American Recovery and Reinvestment Act funding mechanism; in addition the first author received support from the NSF awards DMS 0706518, 1007594; the second author's research was supported in part by NSF award DMS 1107017 and NSF ADVANCED IT WISEST award at University of Illinois at Chicago; the third author's research was supported in part by NSF award DMS 0905730 and the ASA/NSF/BLS Research Fellow program. The views expressed here are those of the authors and do not necessarily reflect the views of the BLS. The authors appreciate the helpful comments from two referees and the Associate Editor, which have led to significant improvement of the paper.

Appendix A

Throughout this section, we use C to denote some positive constant. For any vector $\mathbf{a}=(a_1,\ldots,a_n)\in\mathbb{R}^n$, denote $\|\mathbf{a}\|_r=(|a_1|^r+\cdots+|a_n|^r)^{1/r}$, $1\le r<+\infty$, $\|\mathbf{a}\|_\infty=\max(|a_1|,\ldots,|a_n|)$. For any $n\times n$ symmetric matrix $\mathbf{A}$, denote its L_∞ norm as $\|\mathbf{A}\|_\infty=\max_{1\le i\le n}\sum_{j=1}^n|\mathbf{a}_{ij}|$.

In this paper, $\mathcal{T}=[0,1]$ or $[0,1]^2$. For any function ϕ on a domain $\mathcal{T}$, denote $\|\phi\|_\infty=\sup_{\mathbf{t}\in\mathcal{T}}|\phi(\mathbf{t})|$, and define $\|\phi\|_{q,r}=\sup_{t\ne s,t,s\in[0,1]}|\phi^{(q)}(t)-\phi^{(q)}(s)|/|t-s|^r$. For any Lebesgue measurable L_2-integrable functions ϕ and φ on [0,1], define their theoretical and empirical inner products as $\langle\phi,\varphi\rangle=\int_{\mathcal{T}}\phi(t)\varphi(t)\,dt$ and $\langle\phi,\varphi\rangle_{2,N}=N^{-1}\sum_{j=1}^N\phi(j/N)\varphi(j/N)$, with the corresponding theoretical and empirical inner product norms defined as $\|\phi\|_2^2=\int_0^1\phi^2(t)\,dt$ and $\|\phi\|_{2,N}^2=N^{-1}\sum_{j=1}^N\phi^2(j/N)$, respectively.

For any $r\in(0,1]$, we denote $C^{q,r}[0,1]$ as the space of Hölder continuous functions on [0,1], $C^{q,r}[0,1]=\{\phi:\|\phi\|_{q,r}<+\infty\}$. The technical assumptions we need are as follows:

(A1) *The regression function* $\mu\in C^{p-1,1}[0,1]$;
(A2) *The standard deviation function* $\sigma(t)\in C^{0,\delta}[0,1]$ *for some* $\delta\in(0,1]$;
(A3) *The number of observations for each trajectory* $N\gg n^\theta$ *for some* $\theta>(1+2\nu)/2(p-\nu)$; *the number of interior knots satisfies* $n^{1/2(p-\nu)}\ll N_\mu\ll(N/\log(n))^{1/(1+2\nu)}$, $n^{1/2p}\ll N_G\ll n^{1/(2+2\nu)}$;

(A4) *There exists a constant* $C>0$ *such that* $\Sigma(t,t)>C$, *for any* $t \in [0,1]$;

(A5) *For* $k \in \{1,\ldots,\kappa\}$, $\iota = 0,1,\ldots,p-2$, $\phi_k^{(\iota)}(t) \in C^{0,\delta}[0,1]$, *for some* $\delta \in (0,1]$, $\sum_{k=1}^{\kappa} \sqrt{\lambda_k} \|\phi_k^{(\iota)}\|_\infty < \infty$; *and for a sequence* $\{\kappa_n\}_{n=1}^{\infty}$ *of increasing integers with* $\lim_{n\to\infty}\kappa_n = \kappa$, $N_\mu^{-\delta}\sum_{k=1}^{\kappa_n}\sqrt{\lambda_k}\|\phi_k^{(\iota)}\|_{0,\delta} = o(1)$;

(A6) *There exist* $\delta_1>0$, $\delta_2>0$ *such that* $E|\xi_{ik}|^{4+\delta_1} + E|\varepsilon_{ij}|^{4+\delta_2} < +\infty$, *for* $1 \le i < \infty$, $1 \le k \le \kappa$, $1 \le j < \infty$. *The number* κ *of nonzero eigenvalues is finite or* κ *is infinite while the variables* $\{\xi_{ik}\}_{1\le i<\infty, 1\le k<\infty}$ *are i.i.d.*

Assumptions (A1)–(A6) are standard in the spline smoothing literature; see Huang (2003) and Cao et al. (forthcoming), for instance. In particular, (A1) and (A2) guarantee the convergence rates of $\hat{\mu}(t)$ and its derivatives. Assumption (A3) states the requirement of the number of observations within each curve to the sample size, and the order of the number of knots of splines. Assumptions (A4) and (A5) concern that the derivatives of principal components have collectively bounded smoothness. When $\iota=0$, Assumption (A5) is the same as (A4) in Cao et al. (forthcoming). Assumption (A6) is necessary when using strong approximation result in Lemma A4.

If κ is finite and all $\phi_k^{(\nu)}(t) \in C^{0,\delta}[0,1]$, then Assumption (A5) on $\phi_k^{(\nu)}$'s holds trivially. For $\kappa=\infty$, Assumption (A5) is fulfilled as long as λ_k decreases to zero sufficiently fast. For example, considering the following canonical orthonormal Fourier basis of $L^2([0,1])$:

$$\phi_1(t) \equiv 1, \phi_{2k+1}(t) \equiv \sqrt{2}\cos(k\pi t),$$

$$\phi_{2k}(t) \equiv \sqrt{2}\sin(k\pi t), \quad k=1,2,\ldots, t \in [0,1],$$

we can take $\lambda_1 = 1$ and $\lambda_k = ([k/2]\pi)^{-2\nu}\rho^{2[k/2]}$, $k=2,\ldots$, for any $\rho \in (0,1)$, then $\sum_{k=1}^{\infty}\sqrt{\lambda_k}\|\phi_k^{(\nu)}\|_\infty = 1+\sum_{k=1}^{\infty}\rho^k(\sqrt{2}+\sqrt{2}) = 1+2\sqrt{2}\rho(1-\rho)^{-1}<\infty$. While for any $\delta=1$ and $\{\kappa_n\}_{n=1}^{\infty}$ with κ_n increasing, odd and $\kappa_n \to \infty$, one has

$$N_\mu^{-1}\sum_{k=1}^{\kappa_n}\sqrt{\lambda_k}\|\phi_k^{(\nu)}\|_{0,1} = N_\mu^{-1}\sum_{k=1}^{(\kappa_n-1)/2}\rho^k(\sqrt{2}k\pi+\sqrt{2}k\pi) \le 2\sqrt{2}\pi N_\mu^{-1}\rho\sum_{k=1}^{\infty}\rho^{k-1}k = 2\sqrt{2}\pi N_\mu^{-1}(1-\rho)^{-2} = O(N_\mu^{-1}) = o(1).$$

In the following, define the theoretical and empirical inner product matrices of $\{B_{J,p}(t)\}_{J=1-p}^{N_\mu}$ as

$$\mathbf{V}_p = (\langle B_{J,p}, B_{J',p}\rangle)_{J,J'=1-p}^{N_\mu} = (v_{JJ',p})_{J,J'=1-p}^{N_\mu},$$

$$\hat{\mathbf{V}}_p = (\langle B_{J,p}, B_{J',p}\rangle_{2,N})_{J,J'=1-p}^{N_\mu} = (\hat{v}_{JJ',p})_{JJ'=1-p}^{N_\mu}. \tag{A.1}$$

We establish next that $\hat{\mathbf{V}}_p$ has an inverse with bounded L_∞ norm.

Lemma A1 (*Cao et al., forthcoming*). *Under Assumption* (A3), *for* $\mathbf{V}_p$ *and* $\hat{\mathbf{V}}_p$ *defined in* (A.1), $\|\mathbf{V}_p - \hat{\mathbf{V}}_p\|_\infty = O(N^{-1})$ *and* $\|\hat{\mathbf{V}}_p^{-1}\|_\infty = O(N_\mu)$.

A.1. Proof of Proposition 1

Following Wang and Yang (2009b), we introduce the p-th order quasi-interpolant of μ corresponding to the knots ϖ, denoted by $Q_\varpi(\mu)$; see DeVore and Lorentz (1993, Eq. (4.12) on p. 146) for details. According to DeVore and Lorentz (1993, Theorem 7.7.4), the following lemma holds.

Lemma A2. *There exists a constant* $C>0$, *such that for* $0 \le \nu \le p-2$ *and* $\mu \in C^{p,1}[0,1]$,

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

Lemma A3. *Under Assumptions* (A2), (A3) *and* (A6), *one has*

$$\frac{1}{N}\sum_{j=1}^{N} B_{J,p}(j/N)\sigma(j/N)\bar{\varepsilon}_{\cdot j} = O_P\left(\sqrt{\frac{\log(n)}{nNN_\mu}}\right).$$

Proof. We first truncate the random error ε_{ij} by $u_n = (nN)^\gamma$ $(2/9<\gamma<1/3)$ and write $\varepsilon_{ij} = \varepsilon_{ij,1}+\varepsilon_{ij,2}+a_{ij}$, where $\varepsilon_{ij,1} = \varepsilon_{ij}I\{|\varepsilon_{ij}|>u_n\}$, $\varepsilon_{ij,2} = \varepsilon_{ij}I\{|\varepsilon_{ij}| \le u_n\} - a_{ij}$ and $a_{ij} = E[\varepsilon_{ij}I\{|\varepsilon_{ij}|\le u_n\}]$. It is easy to see that $|a_{ij}| = |-E[\varepsilon_{ij}I\{|\varepsilon_{ij}|> u_n\}]| \le E(|\varepsilon_{ij}|^{4+\delta_2})u_n^{-(3+\delta_2)}$. It is straightforward from the boundedness of spline basis that

$$\left|\frac{1}{nN}\sum_{j=1}^{N}B_{J,p}(j/N)\sigma(j/N)\sum_{i=1}^{n}a_{ij}\right| = O(u_n^{-(3+\delta_2)}).$$

Next we show that the tail part vanishes almost surely. Note that

$$\sum_{n=1}^{\infty}P\{|\varepsilon_{ij}|>u_n\} \le \sum_{n=1}^{\infty}\frac{E|\varepsilon_{ij}|^{4+\delta_2}}{u_n^{4+\delta_2}} \le M_\delta\sum_{n=1}^{\infty}u_n^{-(4+\delta_2)} < \infty.$$

By the Borel–Cantelli lemma, $P\{\omega|\exists N(\omega), |\varepsilon_{ij}(\omega)| \le u_n \text{ for } n > N(\omega)\} = 1$. Let $\nu_\varepsilon = \max\{|\varepsilon_{ij}|, 1 \le i,j \le N(\omega)\}$ and there exists $N_1(\omega) > N(\omega)$, $u_{N_1(\omega)} > \nu_\varepsilon$. Since $u_n = (nN)^\gamma$ is an increasing function, $u_n > u_{N_1(\omega)} > \nu_\varepsilon$ and $n > N_1(\omega)$.

Therefore $P\{\omega|\exists N(\omega), |\varepsilon_{ij}(\omega)| \le u_n,\ 1 \le i \le n,\ 1 \le j \le N, \text{ for } \min(n,N) > N(\omega)\} = 1$, which implies $P\{\omega|\exists N(\omega), |\varepsilon_{ij,1}| = 0,\ 1 \le i \le n,\ 1 \le j \le N \text{ for } \min(n,N) > N(\omega)\} = 1$. Thus

$$\left|\frac{1}{nN}\sum_{j=1}^{N} B_{J,p}(j/N)\sigma(j/N)\sum_{i=1}^{n}\varepsilon_{ij,1}\right| = O_{a.s.}(nN)^{-k}, \quad \text{for any } k > 0.$$

Next denote $D_j = (nN)^{-1}B_{J,p}(j/N)\sigma(j/N)\sum_{i=1}^{n}\varepsilon_{ij,2}$. Since

$$\mathrm{Var}(\varepsilon_{ij,2}) = 1 - E[\varepsilon_{ij}^2 I\{|\varepsilon_{ij}| > u_n\}] - a_{ij}^2 = 1 + O_P\{u_n^{-\delta_2} + u_n^{-2(1+\delta_2)}\},$$

one has $V_n^2 = \mathrm{Var}(\sum_{j=1}^{N} D_j) = c(nNN_\mu)^{-1}$ for a constant $c > 0$. Now Minkowski's inequality implies that

$$E|\varepsilon_{ij,2}|^k \le 2^{k-1}E\{\varepsilon_{ij}^k I\{|\varepsilon_{ij}| \le u_n\} + a_{ij}^k\} \le 2^{k-2}u_n^{k-2}E|\varepsilon_{ij,2}|^2 k!, \quad k \ge 2.$$

Thus $E|D_j|^k \le (2n^{-1}N^{-1}u_n)^{k-2}k!E(D_j^2) < \infty$ with the Cramer constant $c^* = 2u_n/nN$. For any δ, let $\delta_n = \delta\sqrt{\log(n)/nNN_\mu}$. By the Bernstein inequality, for any large enough $\delta > 0$,

$$P\left\{\left|\sum_{j=1}^{N} D_j\right| \ge \delta_n\right\} \le 2\exp\left\{\frac{-\delta_n^2}{4V_n^2 + 2c^*\delta_n}\right\} = 2\exp\left\{\frac{-\delta_n^2}{\dfrac{4c}{nNN_\mu} + 4\dfrac{u_n}{nN}\delta_n}\right\} \le 2n^{-3}.$$

Hence $\sum_{n=1}^{\infty} P(|(1/nN)\sum_{j=1}^{N} B_{J,p}(j/N)\sigma(j/N)\sum_{i=1}^{n}\varepsilon_{ij,2}| \ge \delta_n) \le 2\sum_{n=1}^{\infty} n^{-3} < \infty$, for such $\delta > 0$. Thus Borel–Cantelli's lemma implies the desired result. □

Lemma A4 (Csőrgő and Révész, 1981, Theorem 2.6.7). *Suppose that ξ_i, $1 \le i \le n$ are i.i.d. with $E(\xi_1) = 0$, $E(\xi_1^2) = 1$ and $H(x) > 0$ ($x \ge 0$) is an increasing continuous function such that $x^{-2-\gamma}H(x)$ is increasing for some $\gamma > 0$ and $x^{-1}\log H(x)$ is decreasing with $EH(|\xi_1|) < \infty$. Then there exist constants C_1, C_2, $a > 0$ which depend only on the distribution of ξ_1 and a sequence of Brownian motions $\{W_n(m)\}_{n=1}^{\infty}$, such that for any $\{x_n\}_{n=1}^{\infty}$ satisfying $H^{-1}(n) < x_n < C_1(n\log(n))^{1/2}$ and $S_m = \sum_{i=1}^{m}\xi_i$, then $P\{\max_{1\le m\le n}|S_m - W_n(m)| > x_n\} \le C_2 n\{H(ax_n)\}^{-1}$.*

Proof of Proposition 1. We first show (6). According to Theorem A.1 of Huang (2003), there exists an absolute constant $C > 0$, such that

$$\|\tilde{\mu} - \mu\|_\infty \le C\inf_{g\in C^{p,1}}\|g - \mu\|_\infty \le C\|\mu^{(p)}\|_\infty h_\mu^p. \tag{A.2}$$

Applying Lemma A2, for $0 \le \nu \le p-2$,

$$\|\{Q_\varpi(\mu) - \mu\}^{(\nu)}\|_\infty \le C\|\mu^{(p)}\|_\infty h_\mu^{p-\nu}. \tag{A.3}$$

As a consequence of (A.2) and (A.3) if $\nu = 0$, one has $\|Q_\varpi(\mu) - \tilde{\mu}\|_\infty \le C\|\mu^{(p)}\|_\infty h_\mu^p$, which, according to the differentiation of B-spline given in de Boor (2001), entails that

$$\|\{Q_\varpi(\mu) - \tilde{\mu}\}^{(\nu)}\|_\infty \le C\|\mu^{(p)}\|_\infty h_\mu^{p-\nu}, \tag{A.4}$$

for $0 \le \nu \le p-2$. Combining (A.3) and (A.4) proves (6) for $\nu = 1, \ldots, p-2$.

Next we prove (7). Similar to the definition of $\tilde{\mu}^{(\nu)}(t)$ and $\tilde{\xi}_k^{(\nu)}(t)$, in the following we denote $\tilde{\phi}_k^{(\nu)}(t) = \Gamma^{(\nu)}(t)\boldsymbol{\phi}_k$, for any $k \ge 1$. Using the similar arguments as in the proof of (6), we can show that, for any $k \ge 1$,

$$\|\phi_k^{(\nu)} - \tilde{\phi}_k^{(\nu)}\|_\infty \le C_\phi h_\mu^{p-\nu}. \tag{A.5}$$

Also, according to triangle inequality one has that $\|\tilde{\phi}_k^{(\nu)}\|_\infty \le c_\phi\|\phi_k^{(\nu)}\|_\infty = O(1)$.

According to Assumption (A6), $E|\xi_{ik}|^{4+\delta_1} < +\infty$, $\delta_1 > 0$, so there exists some $\beta \in (0,1/2)$, such that $4+\delta_1 > 2/\beta$. Let $H(x) = x^{4+\delta_1}$, $x_n = n^\beta$, then $n/H(ax_n) = a^{-4-\delta_1}n^{1-(4+\delta_1)\beta} = O(n^{-\gamma_1})$ for some $\gamma_1 > 1$. Applying Lemma A4 and Borel–Cantelli Lemma, one finds i.i.d. variables $Z_{ik,\xi} \sim N(0,1)$ such that

$$\max_{k\ge 1}|\overline{\xi}_{\cdot k} - \overline{Z}_{\cdot k,\xi}\sqrt{\lambda_k}| = O_{a.s.}(n^{\beta-1}), \tag{A.6}$$

where $\overline{Z}_{\cdot k,\xi} = n^{-1}\sum_{i=1}^{n} Z_{ik,\xi}$, $k \ge 1$.

If κ is finite, according to (A.6) note that $|\overline{\xi}_{\cdot k}| \le |\overline{Z}_{\cdot k,\xi}|\sqrt{\lambda_k} + |\overline{\xi}_{\cdot k} - \overline{Z}_{\cdot k,\xi}\sqrt{\lambda_k}|$, $1 \le k \le \kappa$, so $\max_{1\le k\le\kappa}|\overline{\xi}_{\cdot k}| = O_P(n^{-1/2} + n^{\beta-1})$. Then the definition of $\overline{\mu}(t)$ in (4) and (A.5) entail that

$$\|\overline{\mu}^{(\nu)} - \mu^{(\nu)} - \overline{\xi}^{(\nu)}\|_\infty \le \kappa\max_{1\le k\le\kappa}|\overline{\xi}_{\cdot k}|\max_{1\le k\le\kappa}\|\phi_k^{(\nu)} - \tilde{\phi}_k^{(\nu)}\|_\infty = o_P(n^{-1/2}).$$

Thus (8) holds according to Assumption (A3).

If $\kappa=\infty$, using similar arguments in Cao et al. (forthcoming), by (A.6) one can show that $|\overline{\xi}_{\cdot k}|\lambda_k^{-1/2}\le |\overline{Z}_{\cdot k,\xi}|+|\overline{\xi}_{\cdot k}\lambda_k^{-1/2}-\overline{Z}_{\cdot k,\xi}|$, for any $k\ge 1$, so $|\overline{\xi}_{\cdot k}|\lambda_k^{-1/2}=O_P(n^{-1/2}+n^{\beta-1})$. Also following Assumption (A5) one has

$$\begin{aligned}
&E\sup_{t\in[0,1]}|\overline{\mu}^{(v)}(t)-\mu^{(v)}(t)-\tilde{\xi}^{(v)}(t)|\\
&\le \sum_{k=1}^{\kappa}E|\overline{\xi}_{\cdot k}|\sup_{t\in[0,1]}|\phi_k^{(v)}(t)-\tilde{\phi}_k^{(v)}(t)|\le C\sum_{k=1}^{\kappa}E|\overline{\xi}_{\cdot k}|\sup_{t\in[0,1]}|\phi_k^{(v)}(t)|\\
&\le C\left\{\sum_{k=1}^{\kappa_n}(E|\overline{\xi}_{\cdot k}|\lambda_k^{-1/2})\lambda_k^{1/2}\|\phi_k^{(v)}\|_{0,\delta}h_\mu^{\delta}+\sum_{k=\kappa_n+1}^{\kappa}(E|\overline{\xi}_{\cdot k}|\lambda_k^{-1/2})\lambda_k^{1/2}\|\phi_k^{(v)}\|_\infty\right\}\\
&\le Cn^{-1/2}\left\{\sum_{k=1}^{\kappa_n}\lambda_k^{1/2}\|\phi_k^{(v)}\|_{0,\delta}h_\mu^{\delta}+\sum_{k=\kappa_n+1}^{\kappa}\lambda_k^{1/2}\|\phi_k^{(v)}\|_\infty\right\}=o(n^{-1/2}).
\end{aligned}$$

Hence, $\|\overline{\mu}^{(v)}-\mu^{(v)}-\tilde{\xi}^{(v)}\|_\infty=o_P(n^{-1/2})$. According to Lemmas A.1 and A.3, finally we have

$$\begin{aligned}
\|\tilde{e}^{(v)}\|_\infty&=\|\mathbf{B}_p^{(v)}(t)\hat{\mathbf{V}}_p^{-1}N^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{e}\|_\infty\le Ch_\mu^{-v}\|\hat{\mathbf{V}}_p^{-1}\|_\infty\|N^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{e}\|_\infty\\
&=O_{a.s.}(n^{-1/2}N^{-1/2}h_\mu^{-1/2-v}\sqrt{\log(n)}).
\end{aligned}$$

Thus (8) holds according to Assumption (A3). □

A.2. *Proof of Theorem 2*

We denote $\tilde{\zeta}_k(t)=\sqrt{\lambda_k}Z_{\cdot k,\xi}\phi_k^{(v)}(t)$, $k=1,\ldots,\kappa$ and define

$$\tilde{\zeta}(t)=n^{1/2}\left[\sum_{k=1}^{\kappa}\lambda_k\{\phi_k^{(v)}(t)\}^2\right]^{-1/2}\sum_{k=1}^{\kappa}\tilde{\zeta}_k(t)=n^{1/2}\Sigma(t,t)^{-1/2}\sum_{k=1}^{\kappa}\tilde{\zeta}_k(t).$$

It is clear that, for any $t\in[0,1]$, $\tilde{\zeta}(t)$ is Gaussian with mean 0 and variance 1, and the covariance $E\tilde{\zeta}(t)\tilde{\zeta}(s)=\Sigma(t,t)^{-1/2}\Sigma(s,s)^{-1/2}\Sigma(t,s)$, for any $t,s\in[0,1]$. That is, the distribution of $\tilde{\zeta}(t)$, $t\in[0,1]$ and the distribution of $\zeta(t)$, $t\in[0,1]$ in Section 3.1 are identical. Similar to the proof of (7), Note that

$$\begin{aligned}
&E\sup_{t\in[0,1]}|\tilde{\zeta}(t)-n^{1/2}\Sigma(t,t)^{-1/2}\tilde{\xi}^{(v)}(t)|\\
&=n^{1/2}E\sup_{t\in[0,1]}\Sigma(t,t)^{-1/2}\left|\sum_{k=1}^{\kappa}\tilde{\zeta}_k(t)-\tilde{\xi}^{(v)}(t)\right|\\
&\le n^{1/2}E\sup_{t\in[0,1]}\Sigma(t,t)^{-1/2}\sum_{k=1}^{\kappa}(|\overline{Z}_{\cdot k,\xi}\sqrt{\lambda_k}-\overline{\xi}_{\cdot k}||\phi_k^{(v)}(t)|+|\overline{\xi}_{\cdot k}||\phi_k^{(v)}(t)-\tilde{\phi}_k^{(v)}(t)|).
\end{aligned}$$

If κ is finite, then $E\sup_{t\in[0,1]}|\tilde{\zeta}(t)-n^{1/2}\Sigma(t,t)^{-1/2}\tilde{\xi}^{(v)}(t)|=O(n^{\beta-1/2}+h_\mu^{p-v})=o(1)$. If $\kappa=\infty$, by Assumption (A5) one has

$$\begin{aligned}
&n^{1/2}E\sup_{t\in[0,1]}\Sigma(t,t)^{-1/2}\sum_{k=1}^{\kappa}(|\sqrt{\lambda_k}Z_{\cdot k,\xi}-\overline{\xi}_{\cdot k}||\phi_k^{(v)}(t)|+|\overline{\xi}_{\cdot k}||\phi_k^{(v)}(t)-\tilde{\phi}_k^{(v)}(t)|)\\
&\le n^{1/2}\sup_{t\in[0,1]}\Sigma(t,t)^{-1/2}\left\{n^{\beta-1}\sum_{k=1}^{\kappa}\sqrt{\lambda_k}|\phi_k^{(v)}(t)|+n^{-1/2}\sum_{k=1}^{\kappa}\sqrt{\lambda_k}|\phi_k^{(v)}(t)-\tilde{\phi}_k^{(v)}(t)|\right\}\\
&=o(1).
\end{aligned}$$

Theorem 2 follows directly.

A.3. *Proof of Theorem 3*

Following Cao et al. (2011), we define the tensor product spline space as

$$\mathcal{H}^{(p-2),2}[0,1]^2\equiv\mathcal{H}^{(p-2),2}=\mathcal{H}^{(p-2)}\otimes\mathcal{H}^{(p-2)}=\left\{\sum_{J,J'=1-p}^{N_G}b_{JJ'}B_{J,p}(t)B_{J',p}(s),t,s\in[0,1]\right\}.$$

Let $R_{ij}\equiv Y_{ij}-\mu(j/N)$, $1\le i\le n$, $1\le j\le N$ and $\overline{R}_{\cdot jj'}=n^{-1}\sum_{i=1}^{n}R_{ij}R_{ij'}$, $1\le j,j'\le N$. Note that the spline estimator in (9) is

$$\hat{G}(\cdot,\cdot)=\underset{g(\cdot,\cdot)\in\mathcal{H}^{(p-2),2}}{\arg\min}\sum_{j\ne j'}^{N}\{\hat{R}_{\cdot jj'}-g(j/N,j'/N)\}^2,$$

so we define the "infeasible estimator" of the covariance function

$$\tilde{G}(\cdot,\cdot)=\underset{g(\cdot,\cdot)\in\mathcal{H}^{(p-2),2}}{\arg\min}\sum_{1\le j\ne j'\le N}\{\overline{R}_{\cdot jj'}-g(j/N,j'/N)\}^2. \tag{A.7}$$

Denote $\mathbf{X}=\mathbf{B}\otimes\mathbf{B}$, $\mathbf{B}_p(t,s)=\mathbf{B}_p(t)\otimes\mathbf{B}_p(s)$ and $\mathbf{B}_p^{(0,\nu)}(t,s)=\mathbf{B}_p(t)\otimes\mathbf{B}_p^{(\nu)}(s)$. Let vector $\overline{\mathbf{R}}=\{\overline{R}_{\cdot jj'}\}_{1\le j,j'\le N}$, then we have

$$\tilde{G}^{(0,\nu)}(t,s)\equiv\frac{\partial^\nu}{\partial s^\nu}\hat{G}(t,s)=\mathbf{B}_p^{(0,\nu)}(t,s)(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\overline{\mathbf{R}}. \tag{A.8}$$

In the following we write $\phi_{kk'}(t,s)=\phi_k(t)\phi_{k'}(s)$, $\phi_{kk'}^{(0,\nu)}(t,s)=\phi_k(t)\phi_{k'}^{(\nu)}(s)$. Let $\boldsymbol{\phi}_{kk'}=\boldsymbol{\phi}_k\otimes\boldsymbol{\phi}_{k'}$, where $\boldsymbol{\phi}_k$ is a N-dimensional vector defined in Section 2.3, $\tilde{\phi}_{kk'}(t,s)=\mathbf{B}_p(t,s)(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\phi}_{kk'}$. Further we denote $\tilde{\phi}_{kk'}^{(0,\nu)}(t,s)=\mathbf{B}_p^{(0,\nu)}(t,s)(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\phi}_{kk'}$.

Lemma A5. *Under Assumptions* (A5), *for any* $1\le\nu\le p-2$, *and* $\kappa=\infty$, *one has*

$$\sum_{k,k'\ge1}\sqrt{\lambda_k\lambda_{k'}}\|\phi_{kk'}^{(0,\nu)}-\tilde{\phi}_{kk'}^{(0,\nu)}\|_\infty=o(1).$$

Proof. When $\kappa=\infty$, one has $\lambda_k>0$ for any $k\ge1$. Note that

$$\sum_{k,k'\ge1}\sqrt{\lambda_k\lambda_{k'}}\|\phi_{kk'}^{(0,\nu)}\|_\infty\le\sum_{k\ge1}\sqrt{\lambda_k}\|\phi_k\|_\infty\sum_{k'\ge1}\sqrt{\lambda_{k'}}\|\phi_{k'}^{(\nu)}\|_\infty=o(1).$$

Also similarly,

$$\sum_{k,k'\ge1}\sqrt{\lambda_k\lambda_{k'}}\|\tilde{\phi}_{kk'}^{(0,\nu)}\|_\infty\le\sum_{k\ge1}\sqrt{\lambda_k}\|\tilde{\phi}_k\|_\infty\sum_{k'\ge1}\sqrt{\lambda_{k'}}\|\tilde{\phi}_{k'}^{(\nu)}\|_\infty\le C\sum_{k\ge1}\sqrt{\lambda_k}\|\phi_k\|_\infty\sum_{k'\ge1}\sqrt{\lambda_{k'}}\|\phi_{k'}^{(\nu)}\|_\infty=o(1).$$

Lemma A6. *Under Assumption* (A5), *for any* $0\le\nu\le p-2$ *and* $k,k'\ge1$, *one has* $\|\phi_{kk'}^{(0,\nu)}-\tilde{\phi}_{kk'}^{(0,\nu)}\|_\infty=O(h_G^{p-\nu})$.

Proof. According to Theorem 12.8 of Schumaker (2007), there exists an absolute constant $C>0$, such that

$$\|\tilde{\phi}_{kk'}-\phi_{kk'}\|_\infty\le C\|\phi_k^{(p)}\phi_{k'}+\phi_k\phi_{k'}^{(p)}\|_\infty h_G^p,$$

which proves (6) for the case $\upsilon=0$. Let $Q(f)$ the p-th order quasi-interpolant of a function f; see the definition in (12.29) of Schumaker (2007), for $0\le\nu\le p-2$,

$$\|\{Q(\phi_{kk'})-\phi_{kk'}\}^{(0,\nu)}\|_\infty\le C\|\phi_k^{(p)}\phi_{k'}+\phi_k\phi_{k'}^{(p)}\|_\infty h_G^{p-\nu}.$$

For the case $\nu=0$, one has

$$\|Q(\phi_{kk'})-\tilde{\phi}_{kk'}\|_\infty=\|Q(\phi_{kk'})-Q(\tilde{\phi}_{kk'})\|_\infty\le C\|\tilde{\phi}_{kk'}-\phi_{kk'}\|_\infty\le C\|\phi_k^{(p)}\phi_{k'}+\phi_k\phi_{k'}^{(p)}\|_\infty h_G^p,$$

which, according to the differentiation of B-spline given in de Boor (2001), entails that $\|\{Q(\phi_{kk'})-\tilde{\phi}_{kk'}\}^{(0,\nu)}\|_\infty\le C\|\phi_k^{(p)}\phi_{k'}+\phi_k\phi_{k'}^{(p)}\|_\infty h_G^{p-\nu}$, for $0\le\nu\le p-2$. □

Lemma A7. *Under Assumptions* (A1)–(A6), *for any* $1\le\nu\le p-2$

$$\|\tilde{G}^{(0,\nu)}-G^{(0,\nu)}\|_\infty=O_P(n^{-1/2}+h_G^{p-\nu})+o_P(N^{-1}n^{-1/2}h_G^{-1-\nu}\log^{1/2}n), \tag{A.9}$$

$$\|\hat{G}^{(0,\nu)}-\tilde{G}^{(0,\nu)}\|_\infty=O_P(n^{-1}h_G^{-3/2-\nu}+n^{-1/2}h_G^{-1-\nu}h_\mu^p), \tag{A.10}$$

where $\hat{G}^{(0,\nu)}$ *and* $\hat{G}^{(0,\nu)}$ *are given in* (10) and (A.8).

Proof. Let $\overline{\xi}_{\cdot kk'}=n^{-1}\sum_{i=1}^n\xi_{ik}\xi_{ik'}$ and $\overline{\varepsilon}_{\cdot jj'}=n^{-1}\sum_{i=1}^n\varepsilon_{ij}\varepsilon_{ij'}$. To show (A.9), we decompose $\overline{R}_{\cdot jj'}$ in (A.7) into

$$\overline{R}_{1jj'}=\sum_{k\ne k'}^{\kappa}\overline{\xi}_{\cdot kk'}\phi_{kk'}\left(\frac{j}{N},\frac{j'}{N}\right),\quad\overline{R}_{2jj'}=\sum_{k=1}^{\kappa}\overline{\xi}_{\cdot kk}\phi_{kk}\left(\frac{j}{N},\frac{j'}{N}\right),\quad\overline{R}_{3jj'}=\sigma\left(\frac{j}{N}\right)\sigma\left(\frac{j'}{N}\right)\overline{\varepsilon}_{\cdot jj'},$$

$$\overline{R}_{4jj'}=n^{-1}\sum_{i=1}^n\left\{\sum_{k=1}^{\kappa}\xi_{ik}\phi_k\left(\frac{j}{N}\right)\sigma\left(\frac{j'}{N}\right)\varepsilon_{ij'}+\sum_{k=1}^{\kappa}\xi_{ik}\phi_k\left(\frac{j'}{N}\right)\sigma\left(\frac{j}{N}\right)\varepsilon_{ij}\right\}.$$

Denote $\overline{\mathbf{R}}_i=\{\overline{R}_{ijj'}\}_{1\le j,j'\le N}$, $\tilde{\mathcal{R}}_i^{(0,\nu)}(t,s)=\mathbf{B}_p^{(0,\nu)}(t,s)(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\overline{\mathbf{R}}_i$, $i=1,2,3,4$. Then $\tilde{G}^{(0,\nu)}(t,s)=\tilde{\mathcal{R}}_1^{(0,\nu)}(t,s)+\tilde{\mathcal{R}}_2^{(0,\nu)}(t,s)+\tilde{\mathcal{R}}_3^{(0,\nu)}(t,s)+\tilde{\mathcal{R}}_4^{(0,\nu)}(t,s)$. Next we define

$$\mathcal{R}_1^{(0,\nu)}(t,s)=\sum_{k\ne k'}^{\kappa}\overline{\xi}_{\cdot kk'}\phi_{kk'}^{(0,\nu)}(t,s),$$

$$\mathcal{R}_2^{(0,\nu)}(t,s)=G^{(0,\nu)}(t,s)+\sum_{k=1}^{\kappa}\{\phi_{kk}^{(0,\nu)}(t,s)(\overline{\xi}_{\cdot kk}-\lambda_k)\}.$$

Note that $\tilde{\mathcal{R}}_1^{(0,\nu)}(t,s)=\sum_{k\neq k'}^{\kappa}\overline{\xi}_{\cdot kk'}\tilde{\phi}_{kk'}^{(0,\nu)}(t,s)$, then Lemma A5 and Assumption (A5) imply that if κ is ∞, $\lambda_k\lambda_{k'}>0$ and one has

$$\sup_{(t,s)\in[0,1]^2} E\left|\tilde{\mathcal{R}}_1^{(0,\nu)}(t,s)-\mathcal{R}_1^{(0,\nu)}(t,s)\right| \leq \sum_{k\neq k'}^{\kappa} E\left|\overline{\xi}_{\cdot kk'}(\lambda_k\lambda_{k'})^{-1/2}\right|\sqrt{\lambda_k\lambda_{k'}}\|\phi_{kk'}^{(0,\nu)}-\tilde{\phi}_{kk'}^{(0,\nu)}\|_\infty = o(1).$$

Similarly, one has

$$\sup_{(t,s)\in[0,1]^2} E\left|\tilde{\mathcal{R}}_2^{(0,\nu)}(t,s)-\mathcal{R}_2^{(0,\nu)}(t,s)\right| \leq \sum_{k=1}^{\kappa} E\left|\overline{\xi}_{\cdot kk}\lambda_k^{-1}-1\right|\lambda_k\|\phi_{kk}^{(0,\nu)}-\tilde{\phi}_{kk}^{(0,\nu)}\|_\infty = o(1).$$

If κ is finite, then Lemma A.6 and Assumption (A3) imply that

$$\|\tilde{\mathcal{R}}_1^{(0,\nu)}-\mathcal{R}_1^{(0,\nu)}\|_\infty \leq \kappa^2 \max_{1\leq k\neq k'\leq\kappa}\left|\overline{\xi}_{\cdot kk'}\right|\|\phi_{kk'}^{(0,\nu)}-\tilde{\phi}_{kk'}^{(0,\nu)}\|_\infty = O_P(h_G^{p-\nu}n^{-1/2}).$$

Similarly,

$$\|\tilde{\mathcal{R}}_2^{(0,\nu)}-\mathcal{R}_2^{(0,\nu)}\|_\infty \leq \kappa \max_{1\leq k\leq\kappa}\left|\overline{\xi}_{\cdot kk}-\lambda_k\right|\|\phi_{kk}^{(0,\nu)}-\tilde{\phi}_{kk}^{(0,\nu)}\|_\infty = O_P(h_G^{p-\nu}).$$

Hence, $\|\tilde{\mathcal{R}}_2^{(0,\nu)}-\mathcal{R}_2^{(0,\nu)}\|_\infty+\|\tilde{\mathcal{R}}_1^{(0,\nu)}-\mathcal{R}_1^{(0,\nu)}\|_\infty=o(1)$. By Proposition 3.1 in Cao et al. (2011), one has

$$\|N^{-2}\mathbf{X}^{\mathrm{T}}\overline{\mathbf{R}}_3\|_\infty = \|N^{-2}\mathbf{X}^{\mathrm{T}}\overline{\mathbf{R}}_4\|_\infty = o_P(N^{-1}n^{-1/2}h_G\log^{1/2}(n)).$$

Hence, $\|\tilde{\mathcal{R}}_3^{(0,\nu)}\|_\infty=\|\tilde{\mathcal{R}}_4^{(0,\nu)}\|_\infty=o_P(N^{-1}n^{-1/2}h_G^{-1-\nu}\log^{1/2}n)$. Therefore,

$$\begin{aligned}\|\tilde{G}^{(0,\nu)}-G^{(0,\nu)}\|_\infty &\leq \sum_{k\neq k'}^{\kappa}\overline{\xi}_{\cdot kk'}\|\phi_{kk'}^{(0,\nu)}\|_\infty+\sum_{k=1}^{\kappa}(\overline{\xi}_{\cdot kk}-\lambda_k)\|\phi_{kk}^{(0,\nu)}\|_\infty+\|\tilde{\mathcal{R}}_1^{(0,\nu)}-\mathcal{R}_1^{(0,\nu)}\|_\infty+\|\tilde{\mathcal{R}}_2^{(0,\nu)}-\mathcal{R}_2^{(0,\nu)}\|_\infty+\|\tilde{\mathcal{R}}_3^{(0,\nu)}\|_\infty+\|\tilde{\mathcal{R}}_4^{(0,\nu)}\|_\infty\\ &= O_P(n^{-1/2}+h_G^{p-\nu})+o_P(N^{-1}n^{-1/2}h_G^{-1-\nu}\log^{1/2}n).\end{aligned}$$

The proof of (A.10) is similar to Proposition 2.1 in Cao et al. (2011), thus omitted. □

Proof of Theorem 3. According to Lemma A7, one has

$$\|\hat{G}^{(0,\nu)}-G^{(0,\nu)}\|_\infty \leq \|\tilde{G}^{(0,\nu)}-G^{(0,\nu)}\|_\infty+\|\hat{G}^{(0,\nu)}-\tilde{G}^{(0,\nu)}\|_\infty = o_P(1). \quad \square$$

A.4. Proof of Theorem 4

We first show asymptotic consistency of $\hat{\lambda}_k$ and $\hat{\phi}_k(\cdot)$, for $k\geq 1$, in the following lemma.

Lemma A8. *Under Assumptions* (A1)–(A6), *one has*

$$|\hat{\lambda}_k-\lambda_k| = o_P(1), \quad \|\hat{\phi}_k-\phi_k\|_\infty = o_P(1), \quad k\geq 1.$$

Proof. We first want to show that for any $k\geq 1$, $\|\Delta\phi_k\|_\infty=o_P(1)$, in which Δ is the integral operator with kernel $\hat{G}-G$. Note that $(\Delta\phi_k)(t)=\int(\hat{G}-G)(s,t)\phi_k(s)\,ds$. By Theorem 3, when $\nu=0$, $\|\hat{G}-G\|_\infty=o_P(1)$. Thus, for any $k\geq 1$, $\|\Delta\phi_k\|_\infty=o_P(1)$.

Hall and Hosseini-Nasab (2006) gives the L^2 expansion:

$$\hat{\phi}_k-\phi_k=\sum_{k'\neq k}(\lambda_k-\lambda_{k'})^{-1}\langle\Delta\phi_k,\phi_{k'}\rangle\phi_{k'}+O(\|\Delta\|_2^2),$$

where $\|\Delta\|_2=[\iint\{\hat{G}(s,t)-G(s,t)\}^2\,ds\,dt]^{1/2}$. By Bessel's inequality, one has $\|\hat{\phi}_k-\phi_k\|_2\leq C(\|\Delta\phi_k\|_\infty^2+\|\Delta\|_2^2)=o_P(1)$. By (4.9) in Hall et al. (2006) and Theorem 3

$$\hat{\lambda}_k-\lambda_k=\iint(\hat{G}-G)(s,t)\phi_k(s)\phi_k(t)\,ds\,dt+O(\|\Delta\phi_k\|_2^2)=o_P(1).$$

Thus $|\hat{\lambda}_k-\lambda_k|=o_P(1)$ for any $k\geq 1$. Next note that

$$\begin{aligned}\hat{\lambda}_k\hat{\phi}_k(t)-\lambda_k\phi_k(t) &= \int\hat{G}(s,t)\hat{\phi}_k(s)\,ds-\int G(s,t)\phi_k(s)\,ds\\ &= \int(\hat{G}-G)(s,t)(\hat{\phi}_k(s)-\phi_k(s))\,ds+\int(\hat{G}-G)(s,t)\phi_k(s)\,ds+\int G(s,t)\{\hat{\phi}_k(s)-\phi_k(s)\}\,ds.\end{aligned}$$

By the Cauchy–Schwarz inequality, uniformly for all $t\in[0,1]$

$$\int G(s,t)(\hat{\phi}_k(s)-\phi_k(s))\,ds \leq \left(\int G^2(s,t)\,ds\right)^{1/2}\|\hat{\phi}_k-\phi_k\|_2=o_P(1).$$

Similar arguments and Theorem 3 imply that $\int(\hat{G}-G)(s,t)(\hat{\phi}_k(s)-\phi_k(s))\,ds = o_P(1)$ and $\int(\hat{G}-G)(s,t)\phi_k(s)\,ds = o_P(1)$. All the above together yield $\|\hat{\lambda}_k\hat{\phi}_k-\lambda_k\phi_k\|_\infty = o_P(1)$. By the triangle inequality and

$$\lambda_k\|\hat{\phi}_k-\phi_k\|_\infty \le \|\hat{\lambda}_k\hat{\phi}_k-\lambda_k\phi_k\|_\infty + |\hat{\lambda}_k-\lambda_k|\|\hat{\phi}_k\|_\infty = o_P(1),$$

the second result in Lemma A8 follows directly. □

Proof of Theorem 4. According to Lemma A8 and Theorem 3, one has

$$\begin{aligned}|\hat{\phi}_k^{(\nu)}(s)-\phi_k^{(\nu)}(s)| &= \left|\hat{\lambda}_k^{-1}\int_0^1 \hat{G}^{(0,\nu)}(t,s)\hat{\phi}_k(t)\,dt - \lambda_k^{-1}\int_0^1 G^{(0,\nu)}(t,s)\phi_k(t)\,dt\right| \\ &\le |\hat{\lambda}_k^{-1}-\lambda_k^{-1}| \times \int_0^1 |\hat{G}^{(0,\nu)}(t,s)\hat{\phi}_k(t)|\,dt + \lambda_k^{-1}\int_0^1 |\hat{G}^{(0,\nu)}(t,s)\hat{\phi}_k(t)-G^{(0,\nu)}(t,s)\phi_k(t)|\,dt.\end{aligned}$$

Hence, $\sup_{s\in[0,1]}\{\int_0^1 |(\hat{G}^{(0,\nu)}-G^{(0,\nu)})(t,s)\hat{\phi}_k(t)+G(t,s)(\hat{\phi}_k-\phi_k)(t)|\,dt\} = o_P(1)$ and $\|\hat{\phi}_k^{(\nu)}-\phi_k^{(\nu)}\|_\infty = o_P(1)$. Theorem 4 follows from the definition of $\Sigma(t,s)$. □

References

Cao, G., Wang, L., Li, Y., Yang, L., 2011. Spline confidence envelopes for covariance function in dense functional/longitudinal data. Manuscript.
Cao, G., Yang, L., Todem, D. Simultaneous inference for the mean function of dense functional data. Journal of Nonparametric Statistics, forthcoming.
Claeskens, G., Van Keilegom, I., 2003. Bootstrap confidence bands for regression curves and their derivatives. Annals of Statistics 31, 1852–1884.
Csőrgő, M., Révész, P., 1981. Strong Approximations in Probability and Statistics. Academic Press, New York, London.
de Boor, C., 2001. A Practical Guide to Splines. Springer-Verlag, New York.
Degras, D.A., 2011. Simultaneous confidence bands for nonparametric regression with functional data. Statistica Sinica 21, 1735–1765.
DeVore, R., Lorentz, G., 1993. Constructive Approximation: Polynomials and Splines Approximation. Springer-Verlag, Berlin.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman and Hall, London.
Ferraty, F., Vieu, P., 1996. Nonparametric Functional Data Analysis: Theory and Practice. Springer, New York.
Gasser, T., Müller, H.G., 1984. Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics 11, 171–185.
Hall, P., Hosseini-Nasab, M., 2006. On properties of functional principal components analysis. Journal of the Royal Statistical Society: Series B 68, 109–126.
Hall, P., Müller, H.G., Wang, J.L., 2006. Properties of principal component methods for functional and longitudinal data analysis. Annals of Statistics 34, 1493–1517.
Hall, P., Müller, H.G., Yao, F., 2009. Estimation of functional derivatives. Annals of Statistics 37, 3307–3329.
Härdle, W., 1989. Asymptotic maximal deviation of M-smoothers. Journal of Multivariate Analysis 29, 163–179.
Huang, J., 2003. Local asymptotics for polynomial spline regression. Annals of Statistics 31, 1600–1635.
Li, Y., Hsing, T., 2010. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. Annals of Statistics 38, 3321–3351.
Liu, B., Müller, H.G., 2009. Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. Journal of the American Statistical Association 104, 704–717.
Ma, S., Yang, L., Carroll, R.J., 2012. A simultaneous confidence band for sparse longitudinal regression. Statistica Sinica 22, 95–122.
Morris, J.S., Carroll, R.J., 2006. Wavelet-based functional mixed models. Journal of the Royal Statistical Society, Series B 68, 179–199.
Müller, H.G., 2009. Functional modeling of longitudinal data. In: Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. (Eds.), Longitudinal Data Analysis (Handbooks of Modern Statistical Methods), Wiley, New York, pp. 223–252.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis. Springer, New York.
Rice, J.A., Wu, C.O., 2001. Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics 57, 253–259.
Schumaker, L., 2007. Spline Functions Basic Theory, Third ed. Cambridge University Press, Cambridge.
Shi, M., Weiss, R.E., Taylor, J.M.G., 1996. An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. Journal of the Royal Statistical Society, Series B 45, 151–163.
Stone, C.J., 1994. The use of polynomial splines and their tensor products in multivariate function estimation. Annals of Statistics 22, 118–184.
Wahba, G., 1990. Spline Models for Observational Data, vol. 59. SIAM, Philadelphia, PA.
Wang, J., Yang, L., 2009a. Polynomial spline confidence bands for regression curves. Statistica Sinica 19, 325–342.
Wang, L., Yang, L., 2009b. Spline estimation of single index model. Statistica Sinica 19, 765–783.
Xia, Y., 1998. Bias-corrected confidence bands in nonparametric regression. Journal of the Royal Statistical Society, Series B 60, 797–811.
Yao, F., Lee, T.C.M., 2006. Penalized spline models for functional principal component analysis. Journal of the Royal Statistical Society, Series B 68, 3–25.
Yao, F., Müller, H.G., Wang, J.L., 2005a. Functional linear regression analysis for longitudinal data. Annals of Statistics 33, 2873–2903.
Yao, F., Müller, H.G., Wang, J.L., 2005b. Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association 100, 577–590.
Zhou, L., Huang, J., Carroll, R.J., 2008. Joint modelling of paired sparse functional data using principal components. Biometrika 95, 601–619.
Zhou, S., Shen, X., Wolfe, D.A., 1998. Local asymptotics of regression splines and confidence regions. Annals of Statistics 26, 1760–1782.
Zhou, S., Wolfe, D.A., 2000. On derivative estimation in spline regression. Statistica Sinica 10, 93–108.