ORACLE-EFFICIENT CONFIDENCE ENVELOPES FOR COVARIANCE FUNCTIONS IN DENSE FUNCTIONAL DATA

Guanqun Cao[1], Li Wang[2], Yehua Li[2] and Lijian Yang[3]

[1]*Auburn University*, [2]*Iowa State University*
and [3]*Soochow University*

Abstract: We consider nonparametric estimation of the covariance function for dense functional data using computationally efficient tensor product B-splines. We develop both local and global asymptotic distributions for the proposed estimator, and show that our estimator is as efficient as an "oracle" estimator where the true mean function is known. Simultaneous confidence envelopes are developed based on asymptotic theory to quantify the variability in the covariance estimator and to make global inferences on the true covariance. Monte Carlo simulation experiments provide strong evidence that corroborates the asymptotic theory. Examples of near infrared spectroscopy data and speech recognition data are provided to illustrate the proposed method.

Key words and phrases: B-spline, confidence envelope, covariance function, functional data, Karhunen-Loève L^2 representation, longitudinal data.

1. Introduction

Covariance estimation is crucial in both functional and longitudinal data analysis. For longitudinal data, a good estimation of the covariance function improves the estimation efficiency of the mean parameters (Wang, Carroll and Lin (2005); Fan, Huang and Li (2007)). In functional data analysis (Ramsay and Silverman (2005)), covariance estimation plays a critical role in functional principal component analysis (James, Hastie and Sugar (2000); Zhao, Marron and Wells (2004); Yao, Müller and Wang (2005b); Hall, Müller and Wang (2006); Yao and Lee (2006); Zhou, Huang and Carroll (2008); Li and Hsing (2010b)), functional generalized linear models (Cai and Hall (2005); Yao, Müller and Wang (2005a); Li, Wang and Carroll (2010)), and other functional nonlinear models (Ramsay and Silverman (2005); Li and Hsing (2010a)). Other related work on functional data analysis includes Bigot et al. (2010), Ferraty and Vieu (2006) and Morris and Carroll (2006).

There are some important recent works on nonparametric covariance estimation in functional data that are mostly based on kernel smoothing, for

example Yao, Müller and Wang (2005b), Hall, Müller and Wang (2006) and Li and Hsing (2010b). So far, existing work has concentrated on estimation and the corresponding asymptotic convergence rate. There is no theoretical or methodological development for inference procedures on the covariance functions, such as simultaneous or uniform confidence envelopes. Nonparametric simultaneous confidence regions are powerful tools for making global inference on functions; see Härdle and Marron (1991), Claeskens and Van Keilegom (2003), Zhao and Wu (2008), Ma, Yang and Carroll (2012), Wang et al. (2014) and Zheng, Yang and Härdle (2014) for related theory and applications.

In this paper, we consider a typical functional data setting where the functions are recorded on a dense regular grid in an interval $\mathcal{X}$ and the measurements are contaminated with measurement errors. Some recent applications of this type of functional data include near infrared spectra (Li and Hsing (2010b)), recorded speeches for voice recognition (Hastie, Buja and Tibshirani (1995)), electroencephalogram (EEG) data (Crainiceanua, Staicu and Di (2009)). We propose to estimate the covariance function by tensor product B-splines. We show that the estimation error in the mean function is asymptotically negligible in estimating the covariance function, and thus our covariance estimator is as efficient as an "oracle" estimator where the true mean function is known. We derive both local and global asymptotic distribution for the proposed spline covariance estimator. Especially, based on the asymptotic distribution of the maximum deviation of the estimator, we propose a new simultaneous confidence envelope for the covariance function that can be used to quantify and visualize the variability of the covariance estimator and to make global inferences on the shape of the true covariance.

We apply the proposed confidence envelope method to a Tecator near infrared spectra data set to test the hypothesis that the covariance is stationary. In a speech recognition application, the classic functional linear discriminant analysis (Hastie, Buja and Tibshirani (1995); James and Hastie (2001)) assumes that the random curves from different classes share a common covariance function. We further extend our confidence envelope method to a two-sample problem, where one can test whether the covariance functions from two groups are different.

We organize our paper as follows. In Section 2 we describe the data structure and the proposed spline covariance estimator. In Section 3, we study the local and global asymptotic properties of the proposed estimator. Based on the theory, we propose a new confidence envelope approach and extend the method to two-sample hypothesis testing problems. More details of the proposed confidence envelopes are provided in Section 4. We present simulation studies in Section 5 and applications to the Tecator infrared spectroscopy and the speech recognition data set in Sections 6. Some concluding remarks are provided in Section 7. Proofs are provided in the Appendix and in the supplementary material.

2. Spline Covariance Estimation

2.1. Data structure and model assumptions

Following Ramsay and Silverman (2005), the data we consider are a collection of trajectories $\{\eta_i(x)\}_{i=1}^n$ which are i.i.d. realizations of a smooth random function $\eta(x)$, defined on a continuous interval $\mathcal{X}$. Assume that $\{\eta(x), x \in \mathcal{X}\}$ satisfies $E \int_{\mathcal{X}} \eta^2(x)dx < +\infty$, and let $m(x) = E\{\eta(x)\}$ and $G(x, x') = \text{Cov}\{\eta(x), \eta(x')\}$. The covariance function is a symmetric nonnegative-definite function with spectral decomposition, $G(x, x') = \sum_{k=1}^{\infty} \lambda_k \psi_k(x)\psi_k(x')$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\sum_{k=1}^{\infty} \lambda_k < +\infty$, are the eigenvalues, and $\{\psi_k(x)\}_{k=1}^{\infty}$ are the corresponding eigenfunctions, a set of orthonormal functions in $L^2(\mathcal{X})$. By the Karhunen-Loève representation, $\eta_i(x) = m(x) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(x)$, where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variance 1, and the rescaled eigenfunctions $\phi_k = \sqrt{\lambda_k}\psi_k$ converge to 0 in $L^2(\mathcal{X})$.

Without loss of generality, we take $\mathcal{X} = [0, 1]$. Then the observed data are $Y_{ij} = \eta_i(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij}$, for $1 \leq i \leq n$, $1 \leq j \leq N$, ε_{ij} are i.i.d. random errors with $E(\varepsilon_{11}) = 0, E(\varepsilon_{11}^2) = 1$, and $\sigma^2(x)$ is the variance function of the measurement errors. We take $X_{ij} = j/N$ as there is an abundance of interesting functional data sets of such form, see Section 6. The case of randomly observed X_{ij} requires further investigation.

By the Karhunen-Loève representation, the observed data can be written as

$$Y_{ij} = m\left(\frac{j}{N}\right) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k\left(\frac{j}{N}\right) + \sigma\left(\frac{j}{N}\right)\varepsilon_{ij}.$$

We model functions $m(\cdot)$ and $G(\cdot, \cdot)$ nonparametrically, and hence $\{\lambda_k\}_{k=1}^{\infty}$, $\{\phi_k(\cdot)\}_{k=1}^{\infty}$, and $\{\xi_{ik}\}_{k=1}^{\infty}$ are unknown and need to be estimated.

2.2. Spline covariance estimator

Many smoothing tools can be used to estimate the covariance function: kernel methods (Yao, Müller and Wang (2005a)), regression B-splines, penalized splines (Crainiceanua, Staicu and Di (2009)). Asymptotically, these estimators are similar, and users can choose their favorite smoother. Here, we consider the regression B-spline approach because of its good approximation properties and its computational convenience. See Huang and Yang (2004) for discussion of the computational merits of regression splines.

We review some basic facts on spline functions. Consider the equally-spaced points $\{t_J\}_{J=1}^{N_s}$ as interior knots dividing the interval $[0, 1]$ into $(N_s + 1)$ subintervals $I_J = [t_J, t_{J+1})$, $J = 0, \ldots, N_s - 1$, $I_{N_s} = [t_{N_s}, 1]$. Let $h_s = 1/(N_s + 1)$ be the distance between neighboring knots. Let $\mathcal{H}^{(p-2)} = \mathcal{H}^{(p-2)}[0, 1]$ be the polynomial spline space of order p, consisting of all $p - 2$ times continuously differentiable

functions on $[0, 1]$ that are polynomials of degree $p - 1$ on each subinterval. The Jth B-spline of order p is denoted by $B_{J,p}$ as in de Boor (2001). We define the tensor product spline space as

$$\mathcal{H}^{(p-2)} \otimes \mathcal{H}^{(p-2)} = \Big\{ \sum_{J,J'=1-p}^{N_s} b_{JJ'p} B_{J,p}(x) B_{J',p}(x'), b_{JJ'p} \in R, x, x' \in [0,1] \Big\}.$$

If the mean function $m(x)$ were known, one could compute the errors

$$U_{ij} \equiv Y_{ij} - m\left(\frac{j}{N}\right) = \sum_{k=1}^{\infty} \xi_{ik} \phi_k\left(\frac{j}{N}\right) + \sigma\left(\frac{j}{N}\right) \varepsilon_{ij},$$

for $1 \le i \le n$, $1 \le j \le N$. If $\bar{U}_{\cdot jj'} = n^{-1} \sum_{i=1}^{n} U_{ij} U_{ij'}$, $1 \le j \ne j' \le N$, one can define the "oracle" estimator of the covariance function

$$\tilde{G}_{p_2}(\cdot, \cdot) = \underset{g(\cdot,\cdot) \in \mathcal{H}^{(p_2-2)} \otimes \mathcal{H}^{(p_2-2)}}{\operatorname{argmin}} \sum_{1 \le j \ne j' \le N} \left\{ \bar{U}_{\cdot jj'} - g\left(\frac{j}{N}, \frac{j'}{N}\right) \right\}^2, \qquad (2.1)$$

using tensor product splines of order $p_2 \ge 2$. A similar covariance estimator was used by Cao, Yang and Todem (2012), without theoretical justification, to construct confidence bands for the mean function. The diagonal terms $\bar{U}_{\cdot jj}$ are biased estimations of the variance with the bias caused by the measurement errors also referred to as the nugget effect. We remove the nugget effect from the oracle estimator by excluding the $j = j'$ terms from the summation in (2.1). A similar approach has been used in kernel smoothing methods, see Yao, Müller and Wang (2005b), Hall, Müller and Wang (2006) and Li and Hsing (2010b).

Since $m(x)$ is unknown, one can replace it with a spline estimator

$$\hat{m}_{p_1}(\cdot) = \underset{g(\cdot) \in \mathcal{H}^{(p_1-2)}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N} \left\{ Y_{ij} - g\left(\frac{j}{N}\right) \right\}^2, \ p_1 \ge 1.$$

To mimic the above "oracle" smoother, we replace the errors U_{ij} by their estimates and take the spline covariance estimator for $G(x, x')$ to be

$$\hat{G}_{p_1,p_2}(\cdot, \cdot) = \underset{g(\cdot,\cdot) \in \mathcal{H}^{(p_2-2)} \otimes \mathcal{H}^{(p_2-2)}}{\operatorname{argmin}} \sum_{1 \le j \ne j' \le N} \left\{ \hat{\bar{U}}_{\cdot jj',p_1} - g\left(\frac{j}{N}, \frac{j'}{N}\right) \right\}^2, \qquad (2.2)$$

where $\hat{\bar{U}}_{\cdot jj',p_1} = n^{-1} \sum_{i=1}^{n} \hat{U}_{ijp_1} \hat{U}_{ij'p_1}$ with $\hat{U}_{ijp_1} = Y_{ij} - \hat{m}_{p_1}(j/N)$.

Let N_{s_1} be the number of interior knots for mean estimation, and N_{s_2} be the number of interior knots for $\hat{G}_{p_1,p_2}(x, x')$ in each coordinate. Thus, we have $N_{s_2}^2$ interior knots for the tensor product spline space $\mathcal{H}^{(p_2-2)} \otimes \mathcal{H}^{(p_2-2)}$.

For simplicity, write the tensor product spline basis function as $B_{JJ',p_2}(x,x') = B_{J,p_2}(x)B_{J',p_2}(x')$ and in its matrix format,

$$\boldsymbol{B}_{p_2}(x,x') = \left(B_{1-p_2,1-p_2,p_2}(x,x'),\ldots,B_{N_{s_2},1-p_2,p_2}(x,x'),\right.$$
$$\left.\ldots,B_{1-p_2,N_{s_2},p_2}(x,x'),\ldots,B_{N_{s_2},N_{s_2},p_2}(x,x')\right)^{\mathrm{T}}.$$

Then $\hat{G}_{p_1,p_2}(x,x')$ at (2.2) can be rewritten as

$$\hat{G}_{p_1,p_2}(x,x') \equiv \hat{\beta}_{p_1,p_2}^{\mathrm{T}}\boldsymbol{B}_{p_2}(x,x'), \qquad (2.3)$$

where $\hat{\beta}_{p_1,p_2}$ is the collector of the estimated spline coefficients after solving

$$\hat{\beta}_{p_1,p_2} = \underset{\mathbf{b}_{p_2}\in R^{(N_{s_2}+p_2)^2}}{\operatorname{argmin}} \sum_{1\leq j\neq j'\leq N}\left\{\hat{U}_{\cdot jj',p_1} - \mathbf{b}_{p_2}^{\mathrm{T}}\boldsymbol{B}_{p_2}\left(\frac{j}{N},\frac{j'}{N}\right)\right\}^2.$$

3. Asymptotic Theory and Simultaneous Confidence Envelopes

In this section, we establish the oracle property of the tensor product spline estimator $\hat{G}_{p_1,p_2}$ in the sense that spline estimator can be shown uniformly close to the "oracle" smoother of covariance function. This asymptotic consistency of the proposed estimator is given in Propositions 2, while Theorems 2 and 3 provide simultaneous confidence envelopes for covariance functions and their two-sample cases.

3.1. Assumptions and the oracle property

Write $C^{q,\nu}[0,1]$ as the space of ν-Hölder continuous functions on $[0,1]$, $\nu \in (0,1]$,

$$C^{q,\nu}[0,1] = \left\{\phi : \|\phi\|_{q,\nu} = \sup_{x\neq x',x,x'\in[0,1]}\frac{|\phi^{(q)}(x) - \phi^{(q)}(x')|}{|x-x'|^\nu} < +\infty\right\}.$$

We need technical assumptions.

(A1) *The regression function* $m \in C^{p_1-1,1}[0,1]$.

(A2) *The standard deviation function* $\sigma(x) \in C^{0,\nu}[0,1]$. *Also* $\sup_{(x,x')\in[0,1]^2} G(x,x') < C$, *for some positive constant* C *and* $\min_{x\in[0,1]} G(x,x) > 0$.

(A3) *The number of knots* N_{s_1} *and* N_{s_2} *satisfy* $n^{1/(4p_1)} \ll N_{s_1} \ll N$, $n^{1/(2p_2)} \ll N_{s_2} \ll \min\left(N^{1/2},n^{1/3}\right)$ *and* $N_{s_2} \ll N_{s_1}^{p_1}$.

(A4) *The rescaled eigenfunctions* $\phi_k(x) \in C^{p_2-1,1}[0,1]$ *for any* $k \geq 1$ *and* $\sum_{k=1}^\infty \left\|\phi_k^{(p_2-1)}\right\|_{0,1} < \infty$.

(A5) *For any i, j and $k \geq 1$, $E |\xi_{ik}|^6 + E |\varepsilon_{ij}|^4 < +\infty$ and the random coefficients ξ_{ik} are independent.*

Assumptions (A1)−(A3) are standard in the spline smoothing literature; see Huang (2003), for instance. In particular, (A1) guarantee the orders of the bias terms of the spline smoothers for $m(x)$. Assumption (A2) ensures that the covariance function is a uniformly bounded function. Assumption (A3) implies that the number of points on each curve N diverges to infinity as $n \to \infty$, a well-developed asymptotic scenario for dense functional data, see Theorem 3 of Hall, Müller and Wang (2006), Li and Hsing (2010b). This assumption is practically relevant since curves or images measured using new technology are usually of much higher resolution than the previous generation. The smoothness of our estimator is controlled by the number of knots, which increases to infinity as specified in (A3). This increasing knots asymptotic framework guarantees the richness of the basis. Assumption (A4) concerns the bounded smoothness of principal components for bounding the bias terms in the spline covariance estimator and is satisfied by all the simulation examples in Section 5. Assumption (A5) is necessary for applying the uniform martingale difference central limit theorem.

To understand the behavior of the spline covariance estimator $\hat{G}_{p_1,p_2}$ in (2.2), we first study the asymptotic property of the "oracle" estimator $\tilde{G}_{p_2}$ in (2.1). Let

$$\Delta \left(x, x'\right) = \sum_{k,k'=1}^{\infty} \phi_k \left(x\right) \phi_{k'} \left(x'\right) \left(\bar{\xi}_{\cdot kk'} - \delta_{kk'}\right), \qquad (3.1)$$

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

Proposition 1. *Under* (A2)−(A5),

$$\sup_{(x,x')\in[0,1]^2} \left| \tilde{G}_{p_2}(x, x') - G(x, x') - \Delta \left(x, x'\right) \right| = o_p \left(n^{-1/2}\right).$$

The proof is provided in the supplementary material. We have that the difference between the tensor product spline estimator $\hat{G}_{p_1,p_2}$ and the "oracle" smoother is uniformly bounded at an $o_p \left(n^{-1/2}\right)$ rate; hence $\hat{G}_{p_1,p_2}$ is as efficient as the "oracle" estimator.

Proposition 2. *Under* (A1)−(A5),

$$\sup_{(x,x')\in[0,1]^2} \left| \hat{G}_{p_1,p_2}(x, x') - \tilde{G}_{p_2}(x, x') \right| = o_p \left(n^{-1/2}\right).$$

The proof is also in the supplementary material. By combining the results in Propositions 1 and 2, we have

$$\sup_{(x,x')\in[0,1]^2} \left| \hat{G}_{p_1,p_2}(x,x') - G(x,x') - \Delta\left(x,x'\right) \right| = o_p\left(n^{-1/2}\right).$$

This result shows that the random field $\Delta(x,x')$ is the leading term in the expansion of $\hat{G}_{p_1,p_2}(x,x') - G(x,x')$, and the remaining terms are uniformly bounded at an $o_p(n^{-1/2})$ rate. To understand the asymptotic properties of $\hat{G}_{p_1,p_2}(x,x')$ and build confidence envelopes for $G(x,x')$, we need only study the limiting distribution of $\Delta(x,x')$.

3.2. Asymptotic confidence envelopes

Theorem 1. *Under* $(A1)-(A5)$,
$$nE[\hat{G}_{p_1,p_2}(x,x') - G(x,x')]^2 = V\left(x,x'\right) + o(1),$$

where

$$V\left(x,x'\right) = G^2\left(x,x'\right) + G\left(x,x\right)G\left(x',x'\right) + \sum_{k=1}^{\infty} \phi_k^2\left(x\right)\phi_k^2\left(x'\right)\left(E\xi_{1k}^4 - 3\right). \quad (3.2)$$

Remark 1. The existence of $V(x,x')$ in the infinite-dimensional case is guaranteed by Mercer's Theorem and (A2),

$$\sup_{x,x'\in[0,1]}\sum_{k=1}^{\infty}\phi_k^2\left(x\right)\phi_k^2\left(x'\right) \leq C \sup_{x,x'\in[0,1]}\sum_{k=1}^{\infty}\left(\phi_k^4\left(x\right)+\phi_k^4\left(x'\right)\right)$$

$$\leq C \sup_{x,x'\in[0,1]}\left(G\left(x,x\right)+G\left(x',x'\right)\right) < \infty.$$

We address the simultaneous envelopes for the covariance function $G(x,x')$. Further discussion on how to evaluate $V(x,x')$ is in Section 4.

Theorem 2. *Under* $(A1)-(A5)$, *for any* $\alpha \in (0,1)$ *and* $(x,x') \in [0,1]^2$

$$\lim_{n\to\infty} P\left\{ \sup_{(x,x')\in[0,1]^2} n^{1/2}\left|\hat{G}_{p_1,p_2}(x,x') - G(x,x')\right| V^{-1/2}\left(x,x'\right) \leq Q_{1-\alpha}\right\}$$
$$= 1 - \alpha,$$

where $Q_{1-\alpha}$ *is the* $100\left(1-\alpha\right)^{th}$ *percentile of the absolute maxima distribution of* $\zeta_Z\left(x,x'\right)V^{-1/2}\left(x,x'\right)$, *and* $\zeta_Z\left(x,x'\right)$ *is a gaussian random field with mean* $E\zeta_Z\left(x,x'\right) = 0$, *variance* $E\zeta_Z^2\left(x,x'\right) = V\left(x,x'\right)$, *and covariance*

$$\Omega\left(x,x',y,y'\right) = \text{Cov}\left(\zeta_Z\left(x,x'\right),\zeta_Z\left(y,y'\right)\right)$$

$$= \sum_{k\neq k'}^{\infty}\phi_k^2\left(x\right)\phi_{k'}^2\left(x'\right)\phi_k^2\left(y\right)\phi_{k'}^2\left(y'\right)$$

$$+ \sum_{k,k'=1}^{\infty}\left(E\xi_{1k}^4 - 1\right)\phi_k^2\left(x\right)\phi_{k'}^2\left(x'\right)\phi_k^2\left(y\right)\phi_{k'}^2\left(y'\right) \quad (3.3)$$

for any (x, x'), $(y, y') \in [0, 1]^2$.

Remark 2. Although the proposed spline covariance estimator is not guaranteed to be positive semi-definite, it tends to the true positive semi-definite covariance function in probability.

Remark 3. Under $(A1)-(A5)$, as $n \to \infty$, an asymptotic $100\,(1-\alpha)\,\%$ simultaneous confidence envelope for $G(x, x')$, $\forall (x, x') \in [0, 1]^2$ is

$$\hat{G}_{p_1, p_2}(x, x') \pm n^{-1/2} Q_{1-\alpha} V^{1/2}\left(x, x'\right), \tag{3.4}$$

while an asymptotic $100\,(1-\alpha)\,\%$ pointwise confidence envelope for $G(x, x')$ is

$$\hat{G}_{p_1, p_2}(x, x') \pm n^{-1/2} Z_{1-\alpha/2} V^{1/2}\left(x, x'\right), \quad \forall (x, x') \in [0, 1]^2.$$

In practice, the percentile $Q_{1-\alpha}$ and the variance function $V(x, x')$ need to be estimated from the data. These issues are addressed in Section 4.

3.3. Extension to two-sample problems

In functional analysis of variance and linear discriminant analysis, it is commonly assumed that the covariance functions are the same across different treatment groups. More recently, Delaigle and Hall (2012) strongly advocate linear discriminant analysis for classification of functional data. One of their fundamental assumptions is that the random curves from different classes share a common covariance function. It is therefore important to extend our method to two-sample problems, and to construct confidence envelopes for the difference between the covariances functions from two independent groups, similar to a two-sample t-test. Applying the two-sample simultaneous confidence envelope, one can test the common covariance function assumption with quantified uncertainty.

Suppose we have two independent groups of curves with sample sizes n_1 and n_2, respectively. We denote the ratio of two sample sizes as $\hat{r} = n_1/n_2$ and assume that $\lim_{n_1, n_2 \to \infty} \hat{r} = r > 0$. Let $\hat{G}_{p_1, p_2}^{(1)}(x, x')$ and $\hat{G}_{p_1, p_2}^{(2)}(x, x')$ be the spline estimates of covariance functions $G^{(1)}(x, x')$ and $G^{(2)}(x, x')$ by (2.2). Let $\zeta_{12}(x, x')$, $\forall (x, x') \in [0, 1]^2$ be a Gaussian process such that $E\zeta_{12}(x, x') \equiv 0$, $E\zeta_{12}^2(x, x') \equiv V_1(x, x') + rV_2(x, x')$, and $\text{Cov}\left(\zeta_{12}(x, x'), \zeta_{12}(y, y')\right) = \Omega_1(x, x', y, y') + r\Omega_2(x, x', y, y')$ for any $x, x', y, y' \in [0, 1]$, where $\Omega_1(x, x', y, y')$ and $\Omega_2(x, x', y, y')$ are the covariance functions for the two groups as defined in (3.3). Take $Q_{12, 1-\alpha}$ as the $(1-\alpha)$-th quantile of $\sup_{(x, x') \in [0, 1]^2} \zeta_{12}(x, x') \times (V_1(x, x') + rV_2(x, x'))^{-1/2}$.

Theorem 3. *Under* $(A1)-(A5)$, *modified for each group accordingly, for any* $\alpha \in (0, 1)$, *as* $n_1 \to \infty$, $\hat{r} \to r > 0$,

$$P\left\{\sup_{(x, x') \in [0, 1]^2} n_1^{1/2} \frac{\left|\hat{G}_{p_1, p_2}^{(1)}(x, x') - \hat{G}_{p_1, p_2}^{(2)}(x, x') - G^{(1)}(x, x') + G^{(2)}(x, x')\right|}{\left(V_1(x, x') + rV_2(x, x')\right)^{1/2}} \leq Q_{12, 1-\alpha}\right\}$$
$$= 1 - \alpha.$$

Remark 4. Theorem 3 suggests that a $100(1 - \alpha)\%$ simultaneous confidence envelope for $G^{(1)}(x, x') - G^{(2)}(x, x')$ can be constructed as

$$\hat{G}^{(1)}_{p_1,p_2}(x, x') - \hat{G}^{(2)}_{p_1,p_2}(x, x') \pm n_1^{-1/2} Q_{12,1-\alpha} \left(V_1\left(x, x'\right) + r V_2\left(x, x'\right) \right)^{-1/2} \quad (3.5)$$

for all $(x, x') \in [0, 1]^2$. One can use the confidence envelopes to test hypotheses on $G^{(1)}(x, x') - G^{(2)}(x, x')$. Confidence envelops are graphical tools that provide more information than a test, providing a picture of where the functions differ and what the difference looks like.

4. Implementation

In order to implement the confidence envelopes, there are a number of issues that need to be addressed, including selecting the number of knots for spline smoothing, estimating the error variance $\sigma^2(x)$, functional principal component analysis, and estimating the variance function $V(x, x')$ and the percentile $Q_{1-\alpha}$.

4.1. Knot selection

The numbers of knots in spline smoothing are usually treated as unknown tuning parameters, that affect the performance of the confidence envelopes in data applications. Data-driven methods, such as the leave-one-curve-out cross validation, can be used to select the numbers of knots. However, these methods tend to be extremely time consuming for large data sets. We find empirical formulas for setting the number of knots to work quite well in our numerical studies. Given the data set $(j/N, Y_{ij})_{j=1,i=1}^{N,n}$, the number of interior knots N_{s_1} for $\hat{m}_{p_1}(x)$ is taken to be $[2n^{1/(4p_1)} \log n]$, where $[a]$ denotes the integer part of a. The number of interior knots for the spline estimator $\hat{G}_{p_1,p_2}(x, x')$ is set to $N_{\text{s}_2} = [4n^{1/(2p_2)} \log \log n]$. These choices of knots satisfy condition (A3) in our theory. Similar empirical formulas were also used in Cao, Yang and Todem (2012) to select the number of knots when constructing the confidence bands for the mean functions.

4.2. Estimating the variance of the measurement error

Let $\sigma_Y^2(x) = \text{Var}\{Y(x)\} = G(x, x) + \sigma^2(x)$. As $\hat{\bar{U}}_{\cdot jj,p_1} = (1/n) \sum_{i=1}^n \hat{U}_{ijp_1}^2$ are moment estimators of $\sigma_Y^2(j/N)$, we can estimate $\sigma_Y^2(x)$ by

$$\widehat{\sigma}_Y^2(x) = \operatorname*{argmin}_{g(x) \in \mathcal{H}^{(p_1-2)}} \sum_{j=1}^N \left\{ \hat{\bar{U}}_{\cdot jj,p_1} - g\left(\frac{j}{N}\right) \right\}^2.$$

We can then estimate $\sigma^2(x)$ by $\hat{\sigma}^2(x) = \hat{\sigma}_Y^2(x) - \hat{G}_{p_1,p_2}(x, x)$. Similar estimators of $\sigma^2(x)$ are given in Yao, Müller and Wang (2005a) and Li and Hsing (2010b)

using kernel smoothing. Assuming $\sigma_Y^2(x)$ is a smooth function that satisfies the Hölder condition as for $m(x)$, and that the knots for estimating $\sigma_Y^2(x)$ satisfy the same condition as for N_{s_1} in (A3), we can use the argument for Theorem 1 of Cao, Yang and Todem (2012) to show that $\sup_x |\hat\sigma_Y^2(x) - \sigma_Y^2(x)| = O_p(n^{-1/2})$ and hence $\sup_x |\hat\sigma^2(x) - \sigma^2(x)| = O_p(n^{-1/2})$. In practice, the number of knots for $\hat\sigma_Y^2(x)$ can be determined by cross-validation or the same empirical formula as given in Section 4.1.

4.3. Functional principal component analysis

Estimates of the eigenfunctions and eigenvalues, $\hat\phi_k$ and $\hat\lambda_k$, are obtained by solving the eigen-equation

$$\int_0^1 \hat{G}_{p_1,p_2}(x, x')\hat\phi_k(x)\, dx = \hat\lambda_k \hat\phi_k(x'),\tag{4.1}$$

where the $\hat\phi_k$ are subject to $\int_0^1 \hat\phi_k^2(t)\, dt = \hat\lambda_k$ and $\int_0^1 \hat\phi_k(t)\hat\phi_{k'}(t)\, dt = 0$ for $k' < k$. Our spline covariance estimator provides a continuous functional estimate for $G(x, x')$; however, to solve the integral equation in principal components decomposition, a common approach is to discretize $\hat{G}_{p_1,p_2}$ and approximate the integrals by Riemann sums (see Yao, Müller and Wang (2005a)). Since N is sufficiently large, we estimate the eigenfunctions and eigenvalues by decomposing the smoothed covariance matrix $\{\hat{G}_{p_1,p_2}(j/N, j'/N)\}_{j,j'=1}^N$. In particular, (4.1) can be approximated by

$$N^{-1}\sum_{j=1}^N \hat{G}_{p_1,p_2}\left(\frac{j}{N}, \frac{j'}{N}\right)\hat\phi_k\left(\frac{j}{N}\right) = \hat\lambda_k \hat\phi_k\left(\frac{j'}{N}\right).$$

The kth principal component score of the ith curve, which by definition is $\xi_{ik} = \lambda_k^{-1}\int\{\eta_i(x) - m(x)\}\phi_k(x)dx$, can be estimated by a numerical integration

$$\hat\xi_{ik} = N^{-1}\sum_{j=1}^N \hat\lambda_k^{-1}\left(Y_{ij} - \hat{m}_{p_1}\left(\frac{j}{N}\right)\right)\hat\phi_k\left(\frac{j}{N}\right).$$

4.4. Estimating the variance function $V(x, x')$

Further detailed calculation shows

$$V(x, x') = \mathcal{M}(x, x') - G^2(x, x'),\tag{4.2}$$

where $\mathcal{M}(x, x') = E\{\eta^2(x)\eta^2(x')\}$. A proof of (4.2) is provided in Section S4 in the supplementary material. Since the covariance function $G(x, x')$ is already estimated by tensor product splines, we only need to estimate the function $\mathcal{M}$.

Notice that, for any $j \neq j'$,

$$
E(U_{1j}^2 U_{1j'}^2) = E\{(\eta_{1j} + \sigma(\frac{j}{N})\varepsilon_{1j})^2(\eta_{1j'} + \sigma(\frac{j'}{N})\varepsilon_{1j'})^2\}
$$

$$
= \mathcal{M}(\frac{j}{N}, \frac{j'}{N}) + G(\frac{j}{N}, \frac{j}{N})\sigma^2(\frac{j'}{N}) + G(\frac{j'}{N}, \frac{j'}{N})\sigma^2(\frac{j}{N}) + \sigma^2(\frac{j}{N})\sigma^2(\frac{j'}{N}).
$$

We can therefore construct a tensor product spline estimator for $\mathcal{M}(x, x')$ that is similar in spirit to $\hat{G}(x, x')$. Let

$$
\hat{\mathcal{M}}_{p_1,p_2}(\cdot, \cdot) = \operatorname*{argmin}_{g(\cdot,\cdot) \in \mathcal{H}^{(p_2-2)} \otimes \mathcal{H}^{(p_2-2)}} \sum_{1 \leq j \neq j' \leq N} \left\{ \bar{w}_{\cdot jj'} - \hat{G}_{p_1,p_2}(\frac{j}{N}, \frac{j}{N})\hat{\sigma}^2(\frac{j'}{N}) \right.
$$

$$
\left. - \hat{G}_{p_1,p_2}(\frac{j'}{N}, \frac{j'}{N})\hat{\sigma}^2(\frac{j}{N}) - \hat{\sigma}^2(\frac{j}{N})\hat{\sigma}^2(\frac{j'}{N}) - g\left(\frac{j}{N}, \frac{j'}{N}\right) \right\}^2,
$$

with $\bar{w}_{\cdot jj'} = (1/n) \sum_{i=1}^n \hat{U}_{ijp_1}^2 \hat{U}_{ij'p_1}^2$. Then, we estimate $V(x, x')$ by

$$
\hat{V}_{p_1,p_2}(x, x') = \hat{\mathcal{M}}_{p_1,p_2}(x, x') - \hat{G}_{p_1,p_2}^2(x, x').
$$

Assuming $\sigma^2(x)$ and $\mathcal{M}(x, x')$ are smooth functions that satisfy the Hölder continuous conditions similar to $m(x)$ and $G(x, x')$, we can show $\sup_{x,x'} |\hat{\mathcal{M}}_{p_1,p_2}(x, x') - \mathcal{M}(x, x')| = O_p(n^{-1/2})$ using the argument for Theorem 2. This in turn leads to

$$
\sup_{(x,x') \in [0,1]^2} |\hat{V}_{p_1,p_2}(x, x') - V(x, x')| = O_p(n^{-1/2}).
$$

The number of knots for $\hat{\mathcal{M}}_{p_1,p_2}(x, x')$ can be determined by cross-validation or the empirical formula as given in Section 4.1.

4.5. Estimating the percentile $Q_{1-\alpha}$

To evaluate $Q_{1-\alpha}$, we need to simulate the Gaussian random field $\zeta_Z(x, x')$ of Section 3.2. Let

$$
\hat{\zeta}(x, x') = \left\{ \sum_{k \neq k'}^{\infty} Z_{kk'} \phi_k(x) \phi_{k'}(x') + \sum_{k=1}^{\infty} \phi_k(x) \phi_k(x') Z_k \left(E\xi_{1k}^4 - 1\right)^{1/2} \right\},
$$

where $Z_{kk'} = Z_{k'k}$ and Z_k are i.i.d. standard gaussian random variables. Hence, $\hat{\zeta}(x, x')$ is a Gaussian field such that $E\hat{\zeta}(x, x') = 0$, $E\hat{\zeta}^2(x, x') = V(x, x')$, and $\operatorname{Cov}\{\hat{\zeta}(x, x'), \hat{\zeta}(y, y')\} = \operatorname{Cov}\{\zeta_Z(x, x'), \zeta_Z(y, y')\}$ for any (x, x'), $(y, y') \in [0,1]^2$. In practice, we truncate the infinite expansion of $\hat{\zeta}(x, x')$ at a chosen order κ, and replace ϕ_k's by their estimators described in Section 4.3. The number of principal component κ can be chosen using the Akaike information criterion proposed by Li, Wang and Carroll (2013). For a small or moderate

sample size, we find that the simple "fraction of variation explained" method (Müller (2009)) is often satisfactory. For example, we can select the number of eigenvalues that can explain 95% of the variation in the data. The fourth moment of ξ_{1k} is replaced by the empirical fourth moment of $\hat{\xi}_{ik}$. We simulate a large number of independent realizations of $\hat{\zeta}(x, x')$, and take the maximal absolute deviation for each copy of $\hat{\zeta}(x, x')\hat{V}^{-1/2}(x, x')$. Then $Q_{1-\alpha}$ is estimated by the empirical percentiles of these maximum values.

4.6. Additional implementation issues in the two sample problem

For the two-sample problem, we estimate the mean, covariance, and eigen-functions for each group separately. The variance functions $\hat{V}_1(x, x')$ and $\hat{V}_2(x, x')$ are estimated separately as in a one-sample problem. To evaluate $Q_{12,1-\alpha}$, we need to simulate two random fields $\hat{\zeta}_1(x, x')$ and $\hat{\zeta}_2(x, x')$ separately, and esti-mate $Q_{12,1-\alpha}$ by the empirical percentile of $\sup_{(x,x')\in[0,1]^2} |\hat{\zeta}_1(x, x') - \hat{\zeta}_2(x, x')|$. The confidence envelopes for $G^{(1)}(x, x') - G^{(2)}(x, x')$ are constructed as described in (3.5), substituting the unknown quantities with their estimates.

5. Simulation Studies

5.1. Simulation 1: coverage rate of the confidence envelopes

To illustrate the finite-sample performance of the proposed methods, we generated data from the model

$$Y_{ij} = m\left(\frac{j}{N}\right) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k\left(\frac{j}{N}\right) + \sigma\varepsilon_{ij}, 1 \leq j \leq N, 1 \leq i \leq n, \qquad (5.1)$$

where $\xi_{ik}, \varepsilon_{ij} \sim N(0, 1)$ are independent variables. Let $m(x) = \sin\{2\pi(x - 1/2)\}$ $= \sin(2\pi x)$, $\phi_k(x) = \sqrt{\lambda_k}\psi_k(x)$, $\lambda_k = (1/4)^{[k/2]}$, $\psi_{2k-1}(x) = \sqrt{2}\cos(2k\pi x)$, $\psi_{2k}(x) = \sqrt{2}\sin(2k\pi x)$, $k = 1, 2, \ldots$. Assumption (A4) is satisfied as $\sum_{k=1}^{\infty} \|\phi_k\|_{0,1}$ $\leq \sqrt{2}\pi \sum_{k=1}^{\infty} (1/4)^{[k/2]/2} k < \infty$. Since $\sum_{k=1}^{1,000} \lambda_k / \sum_{k=1}^{\infty} \lambda_k > 1 - 10^{-30}$, we trun-cated the expansion in (5.1) at $\kappa = 1,000$.

To mimic the data examples in Section 6, we set the noise level to $\sigma = 0.1$ or 0.2 and the number of curves n to 200, 300, 500, 800 or 1,200. Different noise levels represent different signal to noise level and for example, $\sigma^2 \approx \lambda_6$ when $\sigma = 0.1, 0.2$. Under each sample size, the number of observations per curve was $N = 4[n^{0.3}\log(n)]$. We present two estimation schemes: a) both mean and covariance functions are estimated by linear splines, i.e., $p_1 = p_2 = 2$; b) both estimated by cubic splines, $p_1 = p_2 = 4$. We use confidence levels $1 - \alpha = 0.95$ and 0.99 for our simultaneous confidence envelopes. To check the true coverage rate of the proposed confidence envelopes, each simulation was repeated 1,000 times.

Table 1. Simulation 1: uniform coverage rates from 1,000 replications.

σ	n	Nominal level	Linear Spline $p_1 = p_2 = 2$	Cubic Spline $p_1 = p_2 = 4$
	200	0.950	0.902	0.910
		0.990	0.974	0.984
	300	0.950	0.934	0.932
		0.990	0.984	0.985
	500	0.950	0.943	0.948
0.10		0.990	0.990	0.991
	800	0.950	0.944	0.951
		0.990	0.992	0.992
	1,200	0.950	0.952	0.956
		0.990	0.993	0.995
	200	0.950	0.910	0.914
		0.990	0.971	0.979
	300	0.950	0.926	0.922
		0.990	0.987	0.987
	500	0.950	0.936	0.945
0.20		0.990	0.988	0.989
	800	0.950	0.938	0.952
		0.990	0.991	0.994
	1,200	0.950	0.945	0.954
		0.990	0.994	0.995

Table 1 shows the empirical frequency that the true surface $G(x, x')$ is entirely covered by the confidence envelopes. At all noise levels, the true coverage probability of the confidence envelopes grows closer to the nominal confidence level as the sample size increases, a positive confirmation of Theorem 2. Since the true covariance function is smooth in our simulation, the cubic spline estimator provides better estimate of the covariance function. However, see Table 1, the confidence envelopes based on two spline estimators behave similarly in terms of coverage probability. We also tried such other estimation schemes as $p_1 = 4$, $p_2 = 2$ and $p_1 = 2$, $p_2 = 4$, the coverage rates are not shown because they are similar to the those presented in Table 1.

Figure 1 shows the true covariance and its 95% confidence envelopes based on a cubic spline covariance estimator. The plot is based on a typical run under the setting $n = 200$, $N = 100$, and $\sigma = 0.1$. Here, the true covariance function is entirely covered by the upper and lower envelopes.

5.2. Simulation 2: power of the test based on the confidence envelopes

We conducted further simulation studies to evaluate the size and power of a hypothesis test based on the proposed spline confidence envelopes. The hypothe-
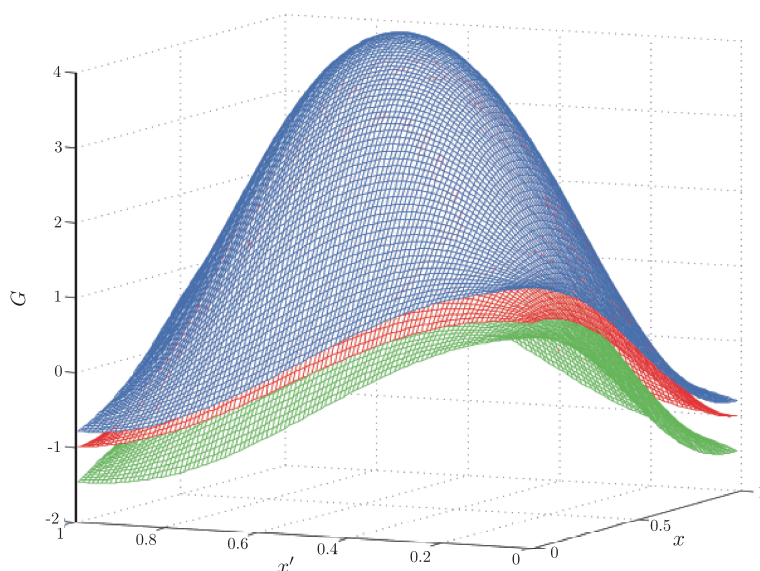
Figure 1. Plot of the true covariance function (middle surface) of the simulated data and its 95% confidence envelopes (3.4) (upper and lower surfaces). The plot is based on a typical simulation run when $n = 200$, $N = 100$, and $\sigma = 0.02$. The confidence envelopes are based on cubic splines, $p_1 = p_2 = 4$.

ses under study were:

$$H_0 : G(x, x') = G_0(x, x'), \quad v.s. \quad H_a : G(x, x') = G_0(x, x') + \delta\Lambda(x, x'),$$

where $G_0(x, x')$ is the covariance function used in Simulation 1, and $\Lambda(x, x')$ $= 4\cos(6\pi x) \times \cos(6\pi x')$. To check the size and power of the test, we generated data under the alternative hypothesis for $\delta = 0$, 0.3, 0.4 and 0.5. When $\delta = 0$, the null hypothesis is true and the data were generated exactly the same way as in Simulation 1. Since $\sqrt{2}\cos(6\pi x)$ is the third eigenfunction of G_0, adding $\delta\Lambda(x, x')$ to $G_0(x, x')$ is equivalent to increasing λ_5 to $(1/4)^{[5/2]} + 2\delta$. Therefore, data under the local alternative hypothesis specified above can be simulated similarly as in Simulation 1. We took $n = 200$, 300, 500, 800, and 1,200, $N = 4[n^{0.3}\log(n)]$, and $\sigma = 0.1$.

Table 2 shows the empirical frequencies of rejecting H_0 based on 1,000 simulation runs, with nominal test level equal to 0.05 and 0.01. When $\delta = 0$, these frequencies represent the size of the test; and when $\delta \neq 0$, the reported frequencies represent the power of the test. From the table, when the sample size is moderate or large, the size of the test is very close to the nominal one. The power of the test also increases to 1 very quickly if n or δ is large, suggesting that the proposed test is quite powerful. The performance of the test was consistent for both linear and cubic splines.

Table 2. Simulation 2: empirical size and power of the test on the covariance function based on the proposed simultaneous confidence envelopes. The reported numbers are based on 1,000 replications using both linear ($p_1 = p_2 = 2$) and cubic splines ($p_1 = p_2 = 4$). The noise level is $\sigma = 0.1$, ξ_{ik}, $\varepsilon_{ij} \sim N(0,1)$ and $\lambda_k = (1/4)^{[k/2]}$, $k = 1,\ldots, 1,000$.

n	Nominal Level	$p_1 = p_2 = 2$ δ				$p_1 = p_2 = 4$ δ			
		0	0.3	0.4	0.5	0	0.3	0.4	0.5
200	0.05	0.098	0.523	0.693	0.809	0.090	0.517	0.718	0.855
	0.01	0.026	0.166	0.269	0.386	0.016	0.163	0.305	0.470
300	0.05	0.066	0.783	0.912	0.977	0.068	0.769	0.923	0.971
	0.01	0.016	0.321	0.528	0.721	0.015	0.326	0.559	0.727
500	0.05	0.057	0.986	0.997	1.000	0.052	0.982	1.000	1.000
	0.01	0.010	0.753	0.932	0.987	0.009	0.744	0.927	0.983
800	0.05	0.056	1.000	1.000	1.000	0.049	1.000	1.000	1.000
	0.01	0.008	0.987	1.000	1.000	0.008	0.978	0.999	1.000
1,200	0.05	0.048	1.000	1.000	1.000	0.044	1.000	1.000	1.000
	0.01	0.007	1.000	1.000	1.000	0.005	1.000	1.000	1.000

6. Empirical Examples

6.1. Tecator near infrared spectra data

We applied our methodology to the Tecator data, which can be downloaded from `http://lib.stat.cmu.edu/datasets/tecator`. This data set contains measurements on $n = 240$ meat samples. There is a $N = 100$ channel near-infrared spectrum of absorbance for each sample. The spectra were recorded in the wavelength range from 850 to 1,050 nm.

Figure 2 shows the scatter plot of the spectra. The spectra can be naturally considered as functional data, since they are recorded on a dense grid of points with little measurement error. On the other hand, there is a lot of variation among different curves. We show the estimated covariance function and the 95% confidence envelope in Figures 3. These results were obtained by applying cubic spline smoothing ($p_1 = p_2 = 4$) to both the mean and covariance functions, with the number of knots $N_{s_1} = 15$ and $N_{s_2} = 13$, respectively. We tried other combinations of knots numbers and linear spline estimators ($p_1 = p_2 = 2$), and the results were similar. From Figure 3, the within curve covariance is positive and quite significant, since the zero hyperplane is far below the lower bound of the confidence envelope.

Using the simultaneous confidence envelopes, one can test other hypotheses on the true covariance function, such as whether the true covariance is stationary.
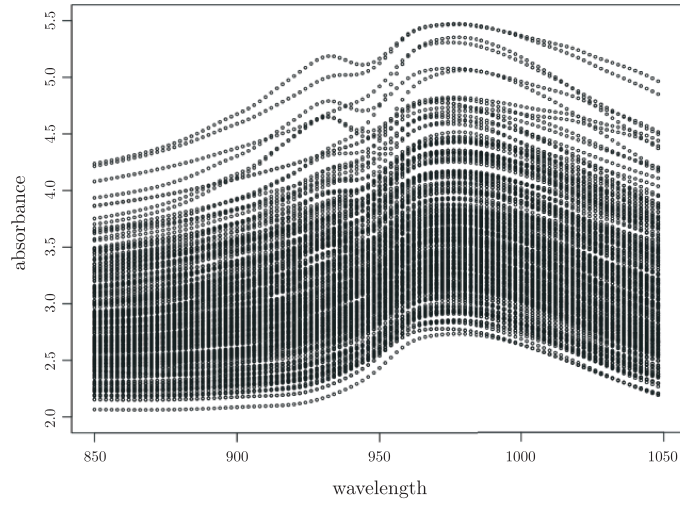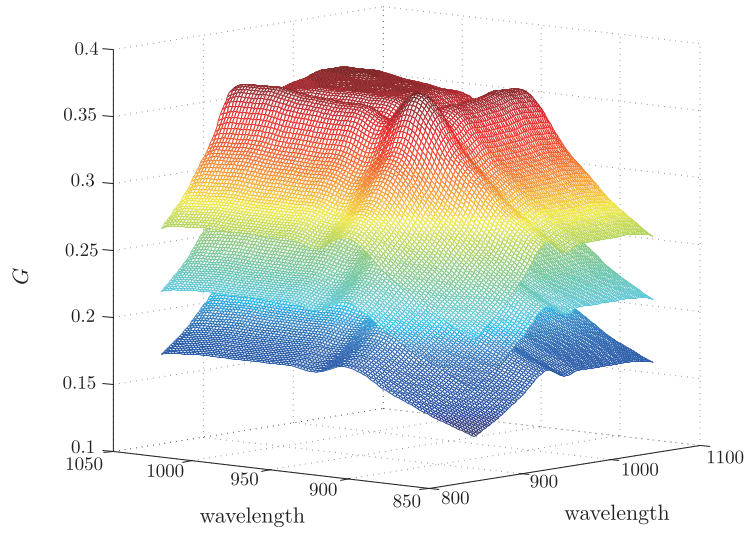
Figure 2. Plot of the Tecator data.

Figure 3. Plots of the cubic tensor product spline covariance estimator (2.3) for the Tecator data (middle surface) and the 95% simultaneous confidence envelope (3.4) (upper and lower surfaces).

Specifically,

$$
\begin{aligned}
H_0 &: G(x, x') \equiv g(|x - x'|), \ \forall \left(x, x'\right) \in [a, b]^2 \\
\text{v.s.} \qquad H_a &: G(x, x') \neq g(|x - x'|), \ \exists \left(x, x'\right) \in [a, b]^2,
\end{aligned}
\tag{6.1}
$$

where $g(\cdot)$ is a stationary covariance function, and $[a, b]$ is the range of wavelength.
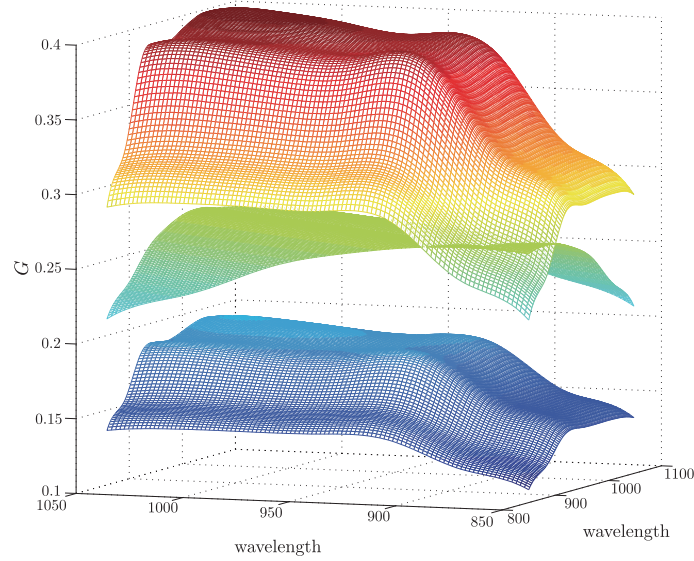
Figure 4. Plot for testing hypothesis (6.1) for the Tecator data. The upper and lower surfaces are the 99.95% confidence envelopes for the covariance function, and the middle surface is the covariance estimator under stationarity assumption, $\hat{G}_S(x - x')$.

To test the hypothesis in (6.1), we need to generate a new estimator under the stationarity assumption and check if this estimator can be covered by the simultaneous confidence envelope. With $\hat{G}(x, x')$ the tensor product B-spline covariance estimator, we take $\hat{G}_S(u) = (b - a - u)^{-1} \int_a^{b-u} \hat{G}(x, x + u)dx$ for $0 \leq u \leq b - a$ and $\hat{G}_S(u) = \hat{G}_S(-u)$ for $a - b < u < 0$. Similar to $\hat{G}$, $\hat{G}_S$ is not guaranteed to be positive semi-definite, but it is sufficient for our purpose. Under the stationarity assumption, $\hat{G}_S$ is a better estimator of the true covariance. We pretend that $\hat{G}_S$ is the true covariance and reject the null hypothesis if this function is not covered by the confidence envelope.

Figure 4 shows cubic tensor spline envelopes with 0.9995 confidence level, and the center surface $\hat{G}_S(x - x')$ as a two-dimensional function. Even for such a high confidence level, the estimator under the stationarity assumption is not fully covered in the envelopes. We conclude that the covariance structure in these Tecator spectra is non-stationary. The same conclusion can be drawn using the linear tensor spline method.

6.2. Speech recognition data

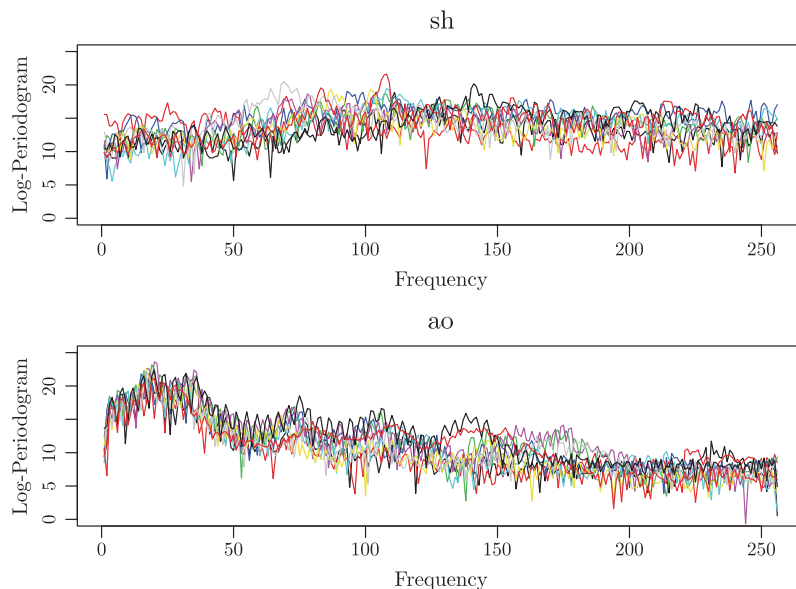The data were extracted from the TIMIT database (TIMIT Acoustic-Phonetic

Figure 5. Plots of the speech recognition data.

Continuous Speech Corpus, NTIS, US Dept of Commerce), a widely used resource for research in speech recognition. The data set we used was formed by selecting five phonemes for classification based on digitized speech. From continuous speech of 50 male speakers, 4,509 speech frames of 32 msec duration were selected. From each speech frame, a log-periodogram was used as transformation for casting speech data in a form suitable for speech recognition. The five phonemes in this data set were transcribed as: "sh" as in "she", "dcl" as in "dark", "iy" as the vowel in "she", "aa" as the vowel in "dark", and "ao" as the first vowel in "water". For illustration, we focus on the "sh" and "ao" phoneme classes as representatives of consonants and vowels. There are $n_1 = 872$ log-periodograms in the "sh" class, and $n_2 = 1,022$ log-periodograms in the "ao" group. Each log-periodogram consists of $N = 256$ equally spaced points. Figure 5 shows a sample 10 log-periodograms from each of the two phoneme classes.

This data set was first analyzed by Hastie, Buja and Tibshirani (1995) using penalized linear discriminant analysis. One of the basic assumptions is that the covariance functions are the same for different classes. Judging from the scatter plot of the data in Figure 5, despite the clear difference between the mean functions of the two groups, there is no obvious indication of difference in covariance structures.

We obtained the cubic spline ($p_1 = p_2 = 4$) covariance estimators for the two phoneme classes separately, see in Figure 6. These results were obtained by using $N_{s_1} = 20$ and $N_{s_2} = 17$ knots for the "sh" class, and $N_{s_1} = 21$ and $N_{s_2} = 18$ for
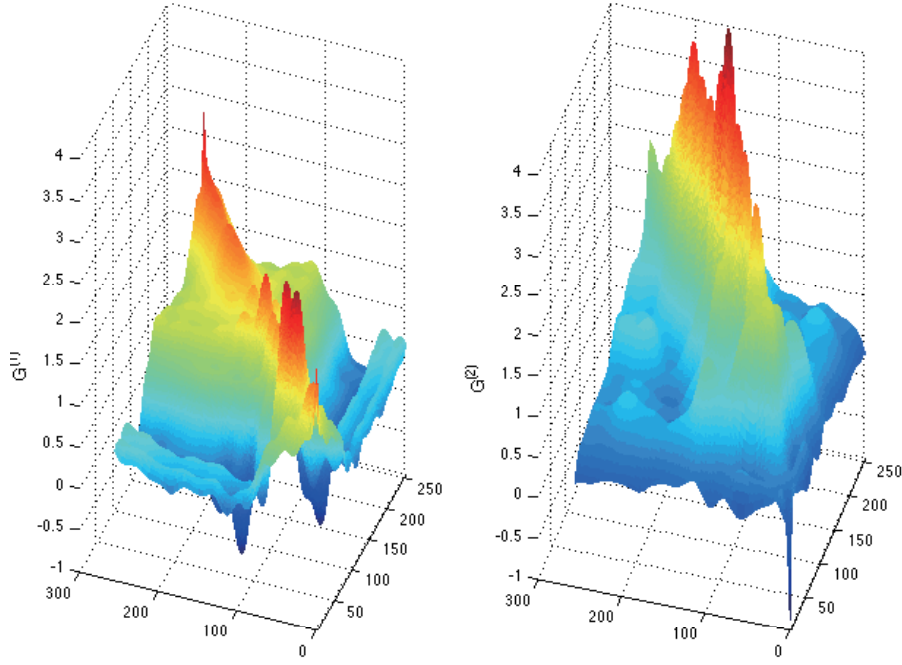
Figure 6. Plots of tensor spline estimators for "sh" and "ao" data sets. Right: "sh"; Left: "ao".

the "ao "class. Different number of knots between the two groups reflects that the sample sizes differed.

By comparing the covariance estimators in Figure 6, there seems to be a visible difference between the two classes. We want to test the equal covariance assumption formally. The hypotheses of interest are

$$H_0 : G^{(1)}(x, x') \equiv G^{(2)}(x, x'), \ \forall \left(x, x'\right) \in [0, 1]^2$$
$$\text{v.s.} \qquad H_a : G^{(1)}(x, x') \neq G^{(2)}(x, x'), \ \exists \left(x, x'\right) \in [0, 1]^2. \tag{6.2}$$

The 99.95% confidence envelopes for the difference of the two covariance functions are provided in Figure 7, and the zero hyperplane is used as a reference. Since the zero hyperplane is not covered by the envelopes, the equal covariance hypothesis is rejected with p-value < 0.0005. We tried different numbers of knots and the result was not sensitive to this.

7. Summary

In this paper, we consider covariance estimation in functional data and propose a new computationally efficient tensor-product B-spline estimator. The
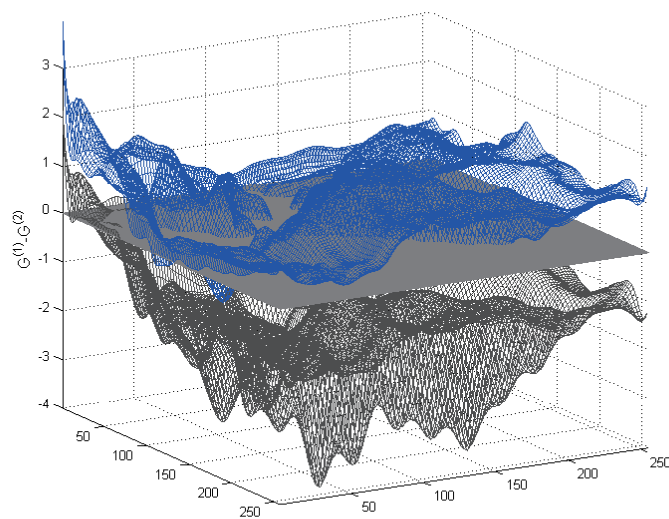
Figure 7. Plot of hypothesis test (6.2) for the speech recognition data. The upper and lower surfaces are the 99.95% confidence envelopes based on cubic tensor product spline covariance estimators and the middle flat surface is the zero plane.

proposed estimator can be used as a building block for further data analysis, such as principal component analysis, linear discriminant analysis and analysis of variance. We study both global and local asymptotic properties of our estimator and propose a simultaneous envelope approach to make inference on the true covariance function. It is shown that the proposed method enjoys superior theoretical properties that have not been well studied in the literature, especially on simultaneous inference for the covariance functions in functional data analysis. The procedure is easily implementable. As demonstrated by our simulation results, the method works well for the case of densely sampled and regularly-spaced functional data. The study of the covariance function inference given in this paper has the potential to provide valuable insights into a number of application problems. For instance, to test the stationarity assumption on the covariance. In a classification problem, we extended our method to test the equal covariance assumption of different treatment groups.

Supplementary Materials

The online supplementary material contains the proofs of Lemmas A.3, A.4 and A.6, Propositions 1 and 2 and equation (4.2).

Acknowledgement

Appendix: Technical Lemmas and Proofs of Theorems

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

Detailed proofs can be found in the supplement.

A.1. Preliminaries

For any Lebesgue measurable function $\phi(\mathbf{x})$ on a domain $\mathcal{D}$, $\mathcal{D} = [0,1]$ or $[0,1]^2$, let $\|\phi\|_\infty = \sup_{\mathbf{x} \in \mathcal{D}} |\phi(\mathbf{x})|$. For any L^2 integrable functions $\phi(\mathbf{x})$ and $\varphi(\mathbf{x})$, $\mathbf{x} \in \mathcal{D}$, take $\langle \phi, \varphi \rangle = \int_{\mathcal{D}} \phi(\mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x}$, with $\|\phi\|_2^2 = \langle \phi, \phi \rangle$. We then set $\langle \phi, \varphi \rangle_N = N^{-1} \sum_{1 \leq j \leq N} \phi\left(\frac{j}{N}\right) \varphi\left(\frac{j}{N}\right)$ for $\mathcal{D} = [0,1]$ and

$$\langle \phi, \varphi \rangle_{2,N} = N^{-2} \sum_{1 \leq j \neq j' \leq N} \phi\left(\frac{j}{N}, \frac{j'}{N}\right) \varphi\left(\frac{j}{N}, \frac{j'}{N}\right) \text{ for } \mathcal{D} = [0,1]^2, \text{ with}$$

$\|\phi\|_{2,N}^2 = \langle \phi, \phi \rangle_{2,N}$.

For any positive integer p, take the theoretical and empirical inner product matrices of $\{B_{J,p}(x)\}_{J=1-p}^{N_s}$ as

$$\mathbf{V}_p = \left(\langle B_{J,p}, B_{J',p} \rangle\right)_{J,J'=1-p}^{N_s}, \quad \hat{\mathbf{V}}_p = \left(\langle B_{J,p}, B_{J',p} \rangle_{2,N}\right)_{J,J'=1-p}^{N_s}.$$

The following lemma is from Cao, Yang and Todem (2012).

Lemma A.1. *For any positive integer p there exists a constant $M_p > 0$, depending only on p, such that $\left\|\mathbf{V}_p^{-1}\right\|_\infty \leq M_p h_s^{-1}$ for large enough n, where $h_s = (N_s + 1)^{-1}$.*

With "$\otimes$" as the Kronecker product of two matrices, $\left\|(\mathbf{A} \otimes \mathbf{A})^{-1}\right\|_\infty = \left\|\mathbf{A}^{-1} \otimes \mathbf{A}^{-1}\right\|_\infty \leq \left\|\mathbf{A}^{-1}\right\|_\infty^2$, for any invertible matrix $\mathbf{A}$ which, together with Lemma A.1, leads to the following.

Lemma A.2. *For any positive integer p, there exists a constant $M_p > 0$, depending only on p, such that $\left\| (\mathbf{V}_p \otimes \mathbf{V}_p)^{-1} \right\|_\infty \leq M_p^2 h_s^{-2}$.*

Take

$$\mathbf{X} = \left(\boldsymbol{B}_{p_2} \left(\frac{2}{N}, \frac{1}{N} \right), \ldots, \boldsymbol{B}_{p_2} \left(1, \frac{1}{N} \right), \ldots, \boldsymbol{B}_{p_2} \left(\frac{1}{N}, 1 \right), \ldots, \boldsymbol{B}_{p_2} \left(1 - \frac{1}{N}, 1 \right) \right)^{\mathrm{T}}.$$

By elementary algebra and the least square algorithm,

$$\tilde{G}_{p_2}(x, x') = \boldsymbol{B}_{p_2}^{\mathrm{T}}(x, x') \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{X}^{\mathrm{T}} \bar{\mathbf{U}},$$

$$\hat{G}_{p_1, p_2}(x, x') = \boldsymbol{B}_{p_2}^{\mathrm{T}}(x, x') \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{X}^{\mathrm{T}} \hat{\bar{\mathbf{U}}}_{p_1},$$

where $\hat{\bar{\mathbf{U}}}_{p_1} = \left(\hat{\bar{U}}_{\cdot 21, p_1}, \ldots, \hat{\bar{U}}_{\cdot N1, p_1}, \ldots, \hat{\bar{U}}_{\cdot 1N, p_1}, \ldots, \hat{\bar{U}}_{\cdot (N-1)N, p_1} \right)^{\mathrm{T}}$,
$\bar{U} = (\bar{U}_{\cdot 21}, \ldots, \bar{U}_{\cdot N1}, \ldots, \bar{U}_{\cdot 1N}, \ldots, \bar{U}_{\cdot (N-1)N})^{\mathrm{T}}$.

Next we define the theoretical and empirical inner product matrices of tensor product spline basis $\left\{ B_{JJ', p_2}(x, x') \right\}_{J, J'=1-p_2}^{N_{s_2}}$ as

$$\mathbf{V}_{p_2, 2} = \left(\left\langle B_{JJ', p_2}, B_{J'' J''', p_2} \right\rangle \right)_{J, J', J'', J'''=1-p_2}^{N_{s_2}},$$

$$\hat{\mathbf{V}}_{p_2, 2} = \left(\left\langle B_{JJ', p_2}, B_{J'' J''', p_2} \right\rangle_{2, N} \right)_{J, J', J'', J'''=1-p_2}^{N_{s_2}}. \tag{A.1}$$

It is easy to obtain $\hat{\mathbf{V}}_{p_2, 2} = N^{-2} \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)$, so we study the properties of $\hat{\mathbf{V}}_{p_2, 2}$ and $\mathbf{V}_{p_2, 2}$.

Lemma A.3. *Under Assumption* (A3), *for $\mathbf{V}_{p_2, 2}$ and $\hat{\mathbf{V}}_{p_2, 2}$ at* (A.1),
$\left\| \mathbf{V}_{p_2, 2} - \hat{\mathbf{V}}_{p_2, 2} \right\|_\infty = O\left(N^{-1} \right)$ *and* $\left\| \hat{\mathbf{V}}_{p_2, 2}^{-1} \right\|_\infty = O\left(h_{s_2}^{-2} \right)$.

Lemma A.4. *For $\hat{\mathbf{V}}_{p_2, 2}$ at* (A.1) *and any $N(N-1)$ vector $\rho = (\rho_{jj'})$, there exists a constant $C > 0$ such that*

$$\sup_{(x, x') \in [0, 1]^2} \left\| N^{-2} \boldsymbol{B}_{p_2}^{\mathrm{T}}(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} \mathbf{X}^{\mathrm{T}} \rho \right\|_\infty \leq C \left\| \rho \right\|_\infty.$$

Let $\quad \phi_{kk'}(x, x') = \phi_k(x) \phi_{k'}(x')$,

$$\widetilde{\phi}_{kk'}(x, x') = \boldsymbol{B}_{p_2}^{\mathrm{T}}(x, x') \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{X}^{\mathrm{T}} \phi_{kk'}, \tag{A.2}$$

and $\phi_{kk'} = \{ \phi_k(2/N) \phi_{k'}(1/N), \ldots, \phi_k(1) \phi_{k'}(1/N), \ldots, \phi_k(1/N) \phi_{k'}(1), \ldots, \phi_k(1 - 1/N) \phi_{k'}(1) \}^{\mathrm{T}}$.

The following lemma is a direct result from de Boor (2001, p.149), , and Theorem 5.1 of Huang (2003). The proof is omitted.

Lemma A.5. *There is an absolute constant $C_g > 0$ such that for every $g \in C^{p-1,\mu}[0,1]$, there exists a function $g^* \in \mathcal{H}^{(p-1)}[0,1]$ and some $\mu \in (0,1]$ such that $\sup_{x \in [0,1]} |g(x) - g^*(x)| \leq C_g h_s^{p-1+\mu} \|g^{(p-1)}\|_{0,\mu}$. If (A2) holds,*

$$\sup_{(x,x') \in [0,1]^2} \left| \phi_{kk'}(x,x') - \widetilde{\phi}_{kk'}(x,x') \right|$$

$$\leq C_\phi h_{s_2}^{p-1+\mu} \left(\|\phi_{k'}\|_\infty \left\| \phi_k^{(p-1)} \right\|_{0,\mu} + \|\phi_k\|_\infty \left\| \phi_{k'}^{(p-1)} \right\|_{0,\mu} \right),$$

for $\widetilde{\phi}_{kk'}(x,x')$ at (A.2). There exists a $c_{\phi,p} \in (0,\infty)$ such that, when n is large enough, $\|\tilde{\phi}\|_\infty \leq c_{\phi,p} \|\phi\|_\infty$ for any $\phi \in C[0,1]$.

A.2. Proofs of Theorems 1 and 2

Proof of Theorem 1. By Propositions 1,

$$E[\tilde{G}_{p_1,p_2}(x,x') - G(x,x')]^2 = E\Delta^2(x,x') + o(1),$$

where $\Delta(x,x')$ is given in (3.1). Let $\bar{\xi}_{\cdot kk'} = n^{-1} \sum_{i=1}^n \xi_{ik}\xi_{ik'}$, $1 \leq k, k'$. It is a straightforward consequence of its definition that

$$\Delta(x,x') = \sum_{k \neq k'}^{\infty} \bar{\xi}_{\cdot kk'} \phi_k(x) \phi_{k'}(x') + \sum_{k=1}^{\infty} (\bar{\xi}_{\cdot kk} - 1) \phi_k(x) \phi_k(x').$$

Since

$$nE\left[\Delta(x,x')\right]^2 = \sum_{k,k'=1}^{\infty} \phi_k^2(x) \phi_{k'}^2(x') + \sum_{k,k'=1}^{\infty} \phi_k(x') \phi_k(x) \phi_{k'}(x) \phi_{k'}(x')$$

$$+ \sum_{k=1}^{\infty} \phi_k^2(x) \phi_k^2(x') \left(E\xi_{1k}^4 - 3\right)$$

$$= G^2(x,x') + G(x,x) G(x',x') + \sum_{k=1}^{\infty} \phi_k^2(x) \phi_k^2(x') \left(E\xi_{1k}^4 - 3\right)$$

$$\equiv V(x,x'),$$

the desired result follows from Proposition 2.

Next let $\zeta(x,x') = n^{1/2} \Delta(x,x')$.

Lemma A.6. *Under (A2)−(A5), as $n \to \infty$, $\mathcal{L}(\zeta(x,x'), (x,x') \in [0,1]^2)$ converges to $\mathcal{L}(\zeta_Z(x,x'), (x,x') \in [0,1]^2)$, where $\zeta_Z(x,x')$ has mean 0, variance $V(x,x')$ at (3.2) and covariance function $\Omega(x,x',y,y')$ at (3.3).*

Proof of Theorem 2. According to Lemma A.6, Propositions 1 and 2, and

Theorem 1, for $\forall \alpha \in (0,1)$ as $n \to \infty$,

$$
\lim_{n \to \infty} P\Big\{ \sup_{(x,x') \in [0,1]^2} \sqrt{n} \left| \hat{G}_{p_1,p_2}(x,x') - G(x,x') \right| V^{-1/2}(x,x') \leq Q_{1-\alpha} \Big\}
$$

$$
= \lim_{n \to \infty} P\Big\{ \sup_{(x,x') \in [0,1]^2} \sqrt{n} \left| \tilde{G}_{p_1,p_2}(x,x') - G(x,x') \right| V^{-1/2}(x,x') \leq Q_{1-\alpha} \Big\}
$$

$$
= \lim_{n \to \infty} P\Big\{ \sup_{(x,x') \in [0,1]^2} \left| \zeta(x,x') \right| V^{-1/2}(x,x') \leq Q_{1-\alpha} \Big\}
$$

$$
= \lim_{n \to \infty} P\Big\{ \sup_{(x,x') \in [0,1]^2} \left| \zeta_Z(x,x') \right| V^{-1/2}(x,x') \leq Q_{1-\alpha} \Big\}.
$$

References

Bigot, J., Biscay, R., Loubes, J. and Muniz-Alvarez, L. (2010). Nonparametric estimation of covariance functions by model selection. *Electronic J. Statist.* **4**, 822-855.

Cai, T. and Hall, P. (2005). Prediction in functional linear regression. *Ann. Statist.* **34**, 2159-2179.

Cao, G., Yang, L. and Todem, D. (2012). Simultaneous inference for the mean function of dense functional data. *J. Nonparametr. Statist.* **24**, 359-377.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852-1884.

Crainiceanua, C. M., Staicu, A.-M. and Di, C.-Z. (2009). Generalized multilevel functional regression. *J. Amer. Statist. Assoc.* **104**, 1550-1561.

de Boor, C. (2001). *A Practical Guide to Splines.* Springer-Verlag, New York.

Delaigle, A. and Hall, P. (2012). Achieving near-perfect classification for functional data. *J. Roy. Statist. Soc. Ser. B* **74**, 267-286.

Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *J. Amer. Statist. Assoc.* **102**, 632-642.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer, Berlin.

Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493-1517.

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778-796.

Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized discriminant analysis. *Ann. Statist.* **23**, 73-102.

Huang, J. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31**, 1600-1635.

Huang, J. and Yang, L. (2004). Identification of nonlinear additive autoregressive models. *J. Roy. Statist. Soc. Ser. B* **66**, 463-477.

James, G. and Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *J. Roy. Statist. Soc. Ser. B* **63**, 533-550.

James, G. M., Hastie, T. and Sugar, C. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587-602.

Li, Y. and Hsing, T. (2010a). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *Ann. Statist.* **38**, 3028-3062.

Li, Y. and Hsing, T. (2010b). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38**, 3321-3351.

Li, Y., Wang, N. and Carroll, R. J. (2010). Generalized functional linear models with semiparametric single-index interactions. *J. Amer. Statist. Assoc.* **105**, 621-633.

Li, Y., Wang, N. and Carroll, R. J. (2013). Selecting the number of principal components in functional data. *J. Amer. Statist. Assoc.* **108**, 1284-1294.

Ma, S., Yang, L. and Carroll, R. J. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica* **22**, 95-122.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *J. Roy. Statist. Soc. Ser. B* **68**, 179-199.

Müller, H. G. (2009). *Functional modeling of longitudinal data. In: Longitudinal Data Analysis.* Wiley, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis. Second Edition.* Springer, New York.

Wang, J., Liu, R., Cheng, F. and Yang, L. (2014). Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. *Ann. Statist.* **42**, 654-668.

Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Amer. Statist. Assoc.* **100**, 147-157.

Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. Roy. Statist. Soc. Ser. B* **68**, 3-25.

Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577-590.

Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873-2903.

Zhao, X., Marron, J. S. and Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statist. Sinica* **14**, 789-808.

Zhao, Z. and Wu, W. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36**, 1854-1878.

Zheng, S., Yang, L. and Härdle, W. (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *J. Amer. Statist. Assoc.* **109**, 661-673.

Zhou, L., Huang, J. and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**, 601-619.

Department of Mathematics and Statistics, Auburn University, Auburn, Alabama 36849, USA,

E-mail: gzc0009@auburn.edu

Department of Statistics and Statistical Laboratory, Iowa State University, Ames, Iowa 50011, USA

E-mail: lilywang@iastate.edu

Department of Statistics and Statistical Laboratory, Iowa State University, Ames, Iowa 50011, USA

E-mail: yehuali@iastate.edu

Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou, Jiangsu, China.

E-mail: yanglijian@suda.edu.cn