

Statistica Sinica Preprint No: SS-14-182R1

Title	Oracle-efficient confidence envelopes for covariance functions in dense functional data
Manuscript ID	SS-14-182R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.2014-182
Complete List of Authors	Guanqun Cao Li Wang Yehua Li and Lijian Yang
Corresponding Author	Yehua Li
E-mail	yehuali@iastate.edu
Notice: Accepted version subject to English editing.	

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

Guanqun Cao¹, Li Wang², Yehua Li² and Lijian Yang³

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. Two real data examples on the near infrared spectroscopy data and speech recognition data are also 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 *et al.*, 2005; Fan *et al.*, 2007). In functional data analysis (Ramsay & Silverman, 2005), covariance estimation plays a critical role in functional principal component analysis (James *et al.*, 2000; Zhao *et al.*, 2004; Yao *et al.*, 2005b; Hall *et al.*, 2006; Yao & Lee, 2006; Zhou *et al.*, 2008; Li & Hsing, 2010b), functional generalized linear models (Cai & Hall, 2005; Yao *et al.*, 2005a; Li *et al.*, 2010), and other functional nonlinear models (Ramsay & Silverman, 2005; Li & Hsing, 2010a). Other related work on functional data analysis includes Bigot *et al.* (2010), Ferraty & Vieu (2006) and Morris & Carroll (2006).

There are some important recent works on nonparametric covariance estimation in functional data, which are mostly based on kernel smoothing, for example Yao *et al.* (2005b), Hall *et al.* (2006) and Li & Hsing (2010b). So far, all existing work 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 & Marron (1991), Claeskens & Van Keilegom (2003), Zhao & Wu (2008), Ma *et al.* (2012), Wang *et al.* (2014) and Zheng *et al.* (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 & Hsing, 2010b), recorded speeches for voice recognition (Hastie *et al.*, 1995), electroencephalogram (EEG) data (Crainiceanua *et al.*, 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 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, which 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 *et al.*, 1995; James & 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 struc-

ture 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 implementation 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. All proofs of the theorems and technical lemmas are provided in the appendix and the supplementary material.

2 Spline Covariance Estimation

2.1 Data structure and model assumptions

Following Ramsay & Silverman (2005), the data that 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 $\{\eta(x), x \in \mathcal{X}\}$ is a L^2 process, i.e. $E \int_{\mathcal{X}} \eta^2(x) dx < +\infty$, and define the mean and covariance functions as $m(x) = E\{\eta(x)\}$ and $G(x, x') = \text{cov}\{\eta(x), \eta(x')\}$. The covariance function is a symmetric nonnegative-definite function with a spectral decomposition, $G(x, x') = \sum_{k=1}^{\infty} \lambda_k \psi_k(x) \psi_k(x')$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, $\sum_{k=1}^{\infty} \lambda_k < +\infty$, are the eigenvalues, and $\{\psi_k(x)\}_{k=1}^{\infty}$ are the corresponding eigenfunctions and are 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 extensive further investigation.

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

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

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, such as kernel methods (Yao *et al.*, 2005a), regression B-splines, or penalized splines (Crainiceanu *et al.*, 2009). Asymptotically, these estimators are similar, and users can choose their own favorite smoother. In this paper, we consider the regression B-spline approach because of its good approximation properties and its computational convenience. We refer the readers to Huang & Yang (2004) for further discussion on the computational merits of regression splines.

To describe the tensor product spline estimator of the covariance functions, we begin by reviewing some basic facts on spline functions. Denote a sequence of equally-spaced points $\{t_J\}_{J=1}^{N_s}$, called interior knots which divide the interval $[0, 1]$ into $(N_s + 1)$ subintervals $I_J = [t_J, t_{J+1})$, $J = 0, \dots, N_s - 1$, $I_{N_s} = [t_{N_s}, 1]$. 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 , which consists all $p - 2$ times continuously differentiable functions on $[0, 1]$ that are polynomials of degree $p - 1$ on each subinterval. The J^{th} B-spline of order p is denoted by $B_{J,p}$ as in de Boor (2001). Thus we define the tensor product spline space as

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

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

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

for $1 \leq i \leq n$, $1 \leq j \leq N$. Denote $\bar{U}_{.jj'} = n^{-1} \sum_{i=1}^n U_{ij} U_{ij'}$, $1 \leq j \neq j' \leq N$, one

can then define the “oracle” estimator of the covariance function

$$\tilde{G}_{p_2}(\cdot, \cdot) = \underset{g(\cdot, \cdot) \in \mathcal{H}^{(p_2-2)} \otimes \mathcal{H}^{(p_2-2)}}{\operatorname{argmin}} \sum_{1 \leq j \neq j' \leq N} \{\bar{U}_{.jj'} - g(j/N, j'/N)\}^2, \quad (2.1)$$

using tensor product splines of order $p_2 \geq 2$. A similar covariance estimator was used by [Cao et al. \(2012\)](#) without theoretical justification to construct confidence bands for the mean function. Note that the diagonal terms $\bar{U}_{.jj}$ are biased estimations of the variance with the bias caused by the measurement errors, which is also referred 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 was used in kernel smoothing methods, see [Yao et al. \(2005b\)](#), [Hall et al. \(2006\)](#) and [Li & Hsing \(2010b\)](#). The same technique will be used in the covariance estimator defined in (2.2).

Since the mean function $m(x)$ is unknown in real data, one can replace it with a spline estimator

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

To mimic the above “oracle” smoother, we replace the unavailable errors U_{ij} by their estimates and define the spline covariance estimator for $G(x, x')$ as

$$\hat{G}_{p_1, p_2}(\cdot, \cdot) = \underset{g(\cdot, \cdot) \in \mathcal{H}^{(p_2-2)} \otimes \mathcal{H}^{(p_2-2)}}{\operatorname{argmin}} \sum_{1 \leq j \neq j' \leq N} \left\{ \hat{U}_{.jj', p_1} - g(j/N, j'/N) \right\}^2, \quad (2.2)$$

where $\hat{U}_{.jj', p_1} = n^{-1} \sum_{i=1}^n \hat{U}_{ijp_1} \hat{U}_{ij'p_1}$ 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. In other words, 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, denote tensor product spline basis function

$B_{JJ', p_2}(x, x') = B_{J, p_2}(x) B_{J', p_2}(x')$ and its matrix format

$$\begin{aligned} \mathbf{B}_{p_2}(x, x') &= (B_{1-p_2, 1-p_2, p_2}(x, x'), \dots, B_{N_{s_2}, 1-p_2, p_2}(x, x'), \\ &\quad \dots, B_{1-p_2, N_{s_2}, p_2}(x, x'), \dots, B_{N_{s_2}, N_{s_2}, p_2}(x, x'))^\top. \end{aligned}$$

Then tensor product spline estimator $\hat{G}_{p_1, p_2}(x, x')$ defined in (2.2) can be rewritten as

$$\hat{G}_{p_1, p_2}(x, x') \equiv \hat{\boldsymbol{\beta}}_{p_1, p_2}^\top \mathbf{B}_{p_2}(x, x'), \quad (2.3)$$

where $\hat{\beta}_{p_1, p_2}$ is the collector of the estimated spline coefficients by solving the following least squares

$$\hat{\beta}_{p_1, p_2} = \underset{\mathbf{b}_{p_2} \in R^{(N_{s_2} + p_2)^2}}{\operatorname{argmin}} \sum_{1 \leq j \neq j' \leq N} \left\{ \hat{U}_{\cdot jj', p_1} - \mathbf{b}_{p_2}^\top \mathbf{B}_{p_2}(j/N, j'/N) \right\}^2.$$

3 Asymptotic theory and simultaneous confidence envelopes

In this section, we first 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 desired asymptotic consistency of the proposed estimator is given in Propositions 2 below, while Theorems 2 and 3 provide simultaneous confidence envelopes for covariance functions and their two-sample cases.

3.1 Assumptions and the oracle property

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

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

We need the following 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 $n^{1/(4p_1)} \ll N_{s_1} \ll N$, $n^{1/(2p_2)} \ll N_{s_2} \ll \min(N^{1/2}, n^{1/3})$ 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 $\sum_{k=1}^{\infty} \|\phi_k^{(p_2-1)}\|_{0, 1} < \infty$.*
- (A5) *For any i, j and $k \geq 1$, $E|\xi_{ik}|^6 + E|\varepsilon_{ij}|^4 < +\infty$ and 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 the covariance function is a uniformly bounded function. Assumption (A3) implies the number of points on each curve N diverges to infinity as $n \rightarrow \infty$, which is a well-developed asymptotic scenario for dense functional data, see Theorem 3 of [Hall et al. \(2006\)](#), [Li & 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 gain deeper understanding on 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). Define

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

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

Proposition 1. *Under Assumptions (A2)-(A5), one has*

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

The proof of Proposition 1 is provided in the supplementary material. The next proposition shows that the difference between the tensor product spline estimator \hat{G}_{p_1, p_2} and the “oracle” smoother is uniformly bounded at an $o_p(n^{-1/2})$ rate, and hence \hat{G}_{p_1, p_2} is as efficient as the “oracle” estimator.

Proposition 2. *Under Assumptions (A1)-(A5), one has*

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

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

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

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 only need to study the limiting distribution of $\Delta(x,x')$.

3.2 Asymptotic confidence envelopes

The next theorem provides a pointwise approximation to the mean squared error of $\hat{G}_{p_1,p_2}(x,x')$.

Theorem 1. *Under Assumptions (A1)-(A5),*

$$nE[\hat{G}_{p_1,p_2}(x,x') - G(x,x')]^2 = V(x,x') + o(1),$$

where

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

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

$$\begin{aligned} \sup_{x,x' \in [0,1]} \sum_{k=1}^{\infty} \phi_k^2(x)\phi_k^2(x') &\leq C \sup_{x,x' \in [0,1]} \sum_{k=1}^{\infty} (\phi_k^4(x) + \phi_k^4(x')) \\ &\leq C \sup_{x,x' \in [0,1]} (G(x,x) + G(x',x')) < \infty. \end{aligned}$$

The following theorem addresses the simultaneous envelopes for the covariance function $G(x,x')$. Further discussion on how to evaluate $V(x,x')$ is provided in Section 4.

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

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

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

$$\begin{aligned} \Omega(x, x', y, y') &= cov(\zeta_Z(x, x'), \zeta_Z(y, y')) \\ &= \sum_{k \neq k'}^{\infty} \phi_k^2(x) \phi_{k'}^2(x') \phi_k^2(y) \phi_{k'}^2(y') \\ &\quad + \sum_{k, k'=1}^{\infty} (E\xi_{1k}^4 - 1) \phi_k^2(x) \phi_{k'}^2(x') \phi_k^2(y) \phi_{k'}^2(y') \end{aligned} \quad (3.3)$$

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. Theorems 1 and 2 shed new lights on constructing pointwise and simultaneous confidence envelopes for the covariance function. Specifically, under Assumptions (A1)-(A5), as $n \rightarrow \infty$, an asymptotic $100(1-\alpha)\%$ simultaneous confidence envelope for $G(x, x'), \forall(x, x') \in [0, 1]^2$ is

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

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

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

In practice, the percentile $Q_{1-\alpha}$ and the variance function $V(x, x')$ need to be estimated from the data. These practical issue will be 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 & 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 $\lim_{n_1, n_2 \rightarrow \infty} \hat{r} = r > 0$. 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). Further, denote by $\zeta_{12}(x, x')$, $\forall (x, x') \in [0, 1]^2$ 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 $\text{cov}(\zeta_{12}(x, x'), \zeta_{12}(y, y')) = \Omega_1(x, x', y, y') + r\Omega_2(x, x', y, y')$ 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). Denote $Q_{12, 1-\alpha}$ as the $(1 - \alpha)$ -th quantile of $\sup_{(x, x') \in [0, 1]^2} \zeta_{12}(x, x') \times (V_1(x, x') + rV_2(x, x'))^{-1/2}$.

We have the following theorem, the proof of which is analogous to that of Theorem 2 and therefore omitted.

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

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

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*

$$\hat{G}_{p_1, p_2}^{(1)}(x, x') - \hat{G}_{p_1, p_2}^{(2)}(x, x') \pm n_1^{-1/2} Q_{12, 1-\alpha} (V_1(x, x') + rV_2(x, x'))^{-1/2} \quad (3.5)$$

for all $(x, x') \in [0, 1]^2$. One can use the confidence envelopes to test hypotheses on $G^{(1)}(x, x') - G^{(2)}(x, x')$. It is also worth pointing out that confidence envelopes are graphical tools that provide more information than a test. They provide a vision on where these functions are different 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, estimating the variance function $V(x, x')$ and the percentile $Q_{1-\alpha}$.

4.1 Knots selection

The numbers of knots in spline smoothing are usually treated as unknown tuning parameters, which can affect the performance of the confidence envelopes in real 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. Instead, we find the following 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 be $N_{s_2} = [4n^{1/(2p_2)}\log\log n]$. These choices of knots satisfy condition (A3) in our theory. Similar empirical formulas were also used in [Cao *et al.* \(2012\)](#) to select the number of knots when constructing the confidence bands for the mean functions.

4.2 Estimating the variance of the measurement error

Define the variance of the response as $\sigma_Y^2(x) = \text{var}\{Y(x)\} = G(x, x) + \sigma^2(x)$. It is easy to see that $\hat{U}_{\cdot jj, p_1} = \frac{1}{n} \sum_{i=1}^n \hat{U}_{ijp_1}^2$ are moment estimators of $\sigma_Y^2(j/N)$, therefore we can estimate $\sigma_Y^2(x)$ by the following spline estimator

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

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 *et al.* \(2005a\)](#) and [Li & Hsing \(2010b\)](#) using kernel smoothing. Assuming $\sigma_Y^2(x)$ is a smooth function that satisfies the Hölder condition as for $m(x)$ and the knots for estimating $\sigma_Y^2(x)$ satisfies the same condition

as for N_{s_1} in assumption (A3), we can use the same argument as in Theorem 1 of [Cao et al. \(2012\)](#) 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 N_{s_1} 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 following eigen-equation

$$\int_0^1 \hat{G}_{p_1, p_2}(x, x') \hat{\phi}_k(x) dx = \hat{\lambda}_k \hat{\phi}_k(x'), \quad (4.1)$$

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 commonly used approach is to discretize \hat{G}_{p_1, p_2} and approximate the integrals by Riemann sums (see [Yao et al. \(2005a\)](#)). Since N is sufficiently large, we estimate the eigenfunctions and eigenvalues by discretizing the smoothed covariance, i.e. 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

$$N^{-1} \sum_{j=1}^N \hat{G}_{p_1, p_2}(j/N, j'/N) \hat{\phi}_k(j/N) = \hat{\lambda}_k \hat{\phi}_k(j'/N).$$

The k th principal component score of the i th 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

$$\hat{\xi}_{ik} = N^{-1} \sum_{j=1}^N \hat{\lambda}_k^{-1} (Y_{ij} - \hat{m}_{p_1}(j/N)) \hat{\phi}_k(j/N).$$

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

Further detailed calculation shows

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

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 $j \neq j'$,

$$\begin{aligned} E(U_{1j}^2 U_{1j'}^2) &= E\{(\eta_{1j} + \sigma(j/N)\varepsilon_{1j})^2 (\eta_{1j'} + \sigma(j'/N)\varepsilon_{1j'})^2\} \\ &= \mathcal{M}(j/N, j'/N) + G(j/N, j'/N)\sigma^2(j'/N) + G(j'/N, j'/N)\sigma^2(j/N) \\ &\quad + \sigma^2(j/N)\sigma^2(j'/N). \end{aligned}$$

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

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

with $\bar{w}_{\cdot, jj'} = \frac{1}{n} \sum_{i=1}^n \hat{U}_{ijp_1}^2 \hat{U}_{ij'p_1}^2$. 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 same argument as 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 same empirical formula as N_{s_2} 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 $\zeta_Z(x, x')$ in Section 3.2. Define

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

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 the 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 *et al.* (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 eigenfunctions for each group separately. The variance functions $\hat{V}_1(x, x')$ and $\hat{V}_2(x, x')$ are also estimated separately the same way 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 estimate $Q_{12,1-\alpha}$ 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) by 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 generate data from the model

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

where $\xi_{ik}, \varepsilon_{ij} \sim N(0, 1)$ are independent variables. Let $m(x) = \sin\{2\pi(x - 1/2)\} = \sin(2\pi x)$, $\phi_k(x) = \sqrt{\lambda_k} \psi_k(x)$, $\lambda_k = (1/4)^{\lfloor k/2 \rfloor}$, $\psi_{2k-1}(x) = \sqrt{2} \cos(2k\pi x)$, $\psi_{2k}(x) = \sqrt{2} \sin(2k\pi x)$, $k = 1, 2, \dots$. Notice that Assumption (A4) is satisfied as $\sum_{k=1}^{\infty} \|\phi_k\|_{0,1} \leq \sqrt{2}\pi \sum_{k=1}^{\infty} (1/4)^{\lfloor k/2 \rfloor / 2} k < \infty$. Since $\sum_{k=1}^{1000} \lambda_k / \sum_{k=1}^{\infty} \lambda_k >$

$1 - 10^{-30}$, we truncate the expansion in (5.1) at $\kappa = 1000$ which is a close approximation for infinite Karhunen-Loève expansion.

To mimic the real data examples in Section 6, we set the noise level to be $\sigma = 0.1$ or 0.2 and the number of curves n to be 200, 300, 500, 800 or 1200. Different noise levels represent different signal to noise level. For example, $\sigma^2 \approx \lambda_6$ when $\sigma = 0.1, 0.2$. Under each sample size, the number of observations per curve is assumed to be $N = 4\lceil n^{0.3} \log(n) \rceil$. 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 are estimated by cubic splines, i.e. $p_1 = p_2 = 4$. 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 is repeated 1000 times.

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 becomes closer to the nominal confidence level as the sample size increases, which shows 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, as can be seen from Table 1, the confidence envelopes based on two spline estimators behave rather similarly in terms of coverage probability. We have also tried other estimation schemes such as $p_1 = 4, p_2 = 2$ and $p_1 = 2, p_2 = 4$, the coverage rates are not shown here because they are similar to the those presented in Table 1.

We show in Figure 1 the true covariance and the 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 $\sigma = 0.1$. As we can see, 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 conduct further simulation studies to evaluate the size and power of a hypothesis test based on the proposed spline confidence envelopes. The hypotheses

Table 1: Simulation 1: uniform coverage rates from 1000 replications.

σ	n	Nominal level	Linear Spline $p_1 = p_2 = 2$	Cubic Spline $p_1 = p_2 = 4$
0.10	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
		0.990	0.990	0.991
	800	0.950	0.944	0.951
		0.990	0.992	0.992
	1200	0.950	0.952	0.956
		0.990	0.993	0.995
0.20	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
		0.990	0.988	0.989
	800	0.950	0.938	0.952
		0.990	0.991	0.994
	1200	0.950	0.945	0.954
		0.990	0.994	0.995

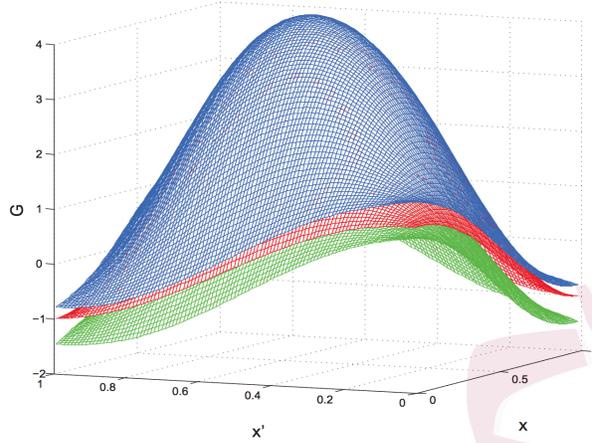


Figure 1: Plot of the true covariance function (middle surface) of the simulated data and its 95% confidence envelopes (3.4) (upper and lower surfaces). The plot is based on a typical simulation run when $n=200$, $N=100$, and $\sigma=0.02$. The confidence envelopes are based on cubic splines, i.e. $p_1 = p_2 = 4$.

under study are:

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

where $G_0(x, x')$ is the same 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 generate 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 are generated exactly the same way as 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_3 to $(1/4)^{[5/2]} + 2\delta$. Therefore, data under the local alternative hypothesis specified above can be simulated similarly as Simulation 1. We set sample size $n = 200, 300, 500, 800$ and 1200 , $N = 4[n^{0.3} \log(n)]$ and $\sigma = 0.1$.

Table 2 shows the empirical frequencies of rejecting H_0 based on 1000 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. As we can see 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,

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 1000 replications using both linear ($p_1 = p_2 = 2$) and cubic splines ($p_1 = p_2 = 4$). The noise level is $\sigma = 0.1$, $\xi_{ik}, \varepsilon_{ij} \sim N(0, 1)$ and $\lambda_k = (1/4)^{\lfloor k/2 \rfloor}$, $k = 1, \dots, 1000$.

n	Nominal Level	$p_1 = p_2 = 2$				$p_1 = p_2 = 4$			
		δ				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
1200	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

indicating the proposed test is very powerful. The performance of the test is consistent for both linear and cubic splines.

6 Empirical examples

6.1 Tecator near infrared spectra data

We first apply 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 are recorded in the wavelength range from 850 to 1050 nm.

Figure 2 shows the scatter plot of the spectra. As we can see, 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

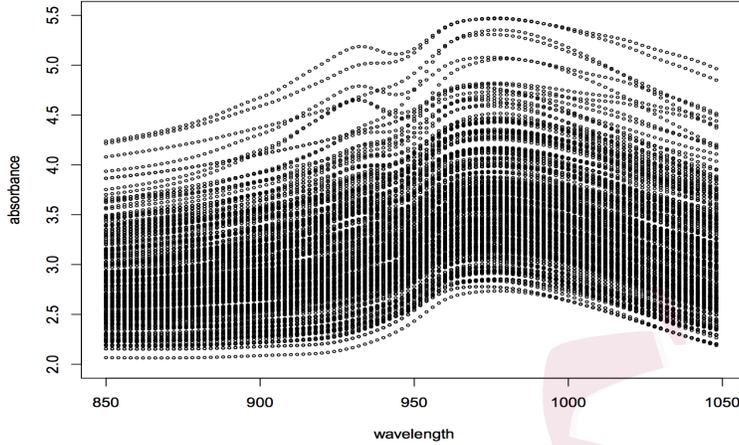


Figure 2: Plot of the Tecator data.

the 95% confidence envelope in Figures 3. These results are obtained by applying cubic spline smoothing ($p_1 = p_2 = 4$) to both the mean and covariance functions, with the number of knots $N_{s_1} = 15$, $N_{s_2} = 13$, respectively. We also tried other combinations of knots numbers and linear spline estimators ($p_1 = p_2 = 2$), and the results are very similar. From Figure 3, we can see that 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 interesting hypotheses on the true covariance function, such as the true covariance being stationary. Specifically, we are interested in the following hypothesis,

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

where $g(\cdot)$ is a stationary covariance function, and $[a, b]$ is the range of wavelength.

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. Letting $\hat{G}(x, x')$ be the tensor product B-spline covariance estimator, we define $\hat{G}_S(u) = (b - a - u)^{-1} \int_a^{b-u} \hat{G}(x, x + u) dx$ 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} , the new estimator \hat{G}_S is not guaranteed to be positive semi-definite, but it is sufficient

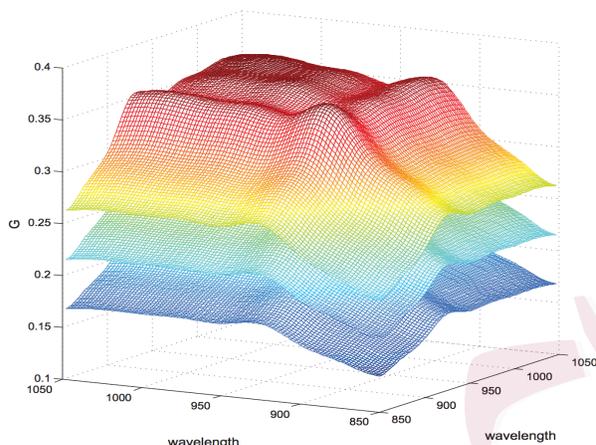


Figure 3: Plots of the cubic tensor product spline covariance estimator (2.3) for the Tecator data (middle surface) and the 95% simultaneous confidence envelope (3.4) (upper and lower surfaces).

for our purpose. Under the stationarity assumption, \hat{G}_S is a better estimator of the true covariance. We will 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 is $\hat{G}_S(x - x')$ as a two-dimensional function. As we can see, even for such a high confidence level, the estimator under the stationarity assumption is still 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 linear tensor spline method.

6.2 Speech recognition data

The data were extracted from the TIMIT database (TIMIT Acoustic-Phonetic Continuous Speech Corpus, NTIS, US Dept of Commerce) which is a widely used resource for research in speech recognition. The data set we use was formed by selecting five phonemes for classification based on digitized speech from this database. From continuous speech of 50 male speakers, 4509 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

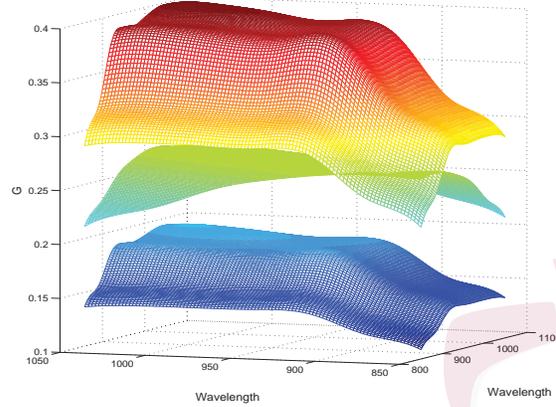


Figure 4: Plot for testing hypothesis (6.1) for the Tecator data. The upper and lower surfaces are the 99.95% confidence envelopes for the covariance function, and the middle surface is the covariance estimator under stationarity assumption, $\hat{G}_S(x - x')$.

recognition. The five phonemes in this data set are transcribed as follows: “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 purpose, 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 $n_2 = 1022$ log-periodograms in the “ao” group. Each log-periodogram consists $N = 256$ 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 et al. \(1995\)](#) 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 first obtain the cubic spline ($p_1 = p_2 = 4$) covariance estimators for the two phoneme classes separately, which are shown in Figure 6. These results are obtained by using $N_{s_1} = 20$, $N_{s_2} = 17$ number of knots for the “sh” class, and $N_{s_1} = 21$ and $N_{s_2} = 18$ for the “ao” class. Different number of knots between the two groups reflects that the sample sizes are different.

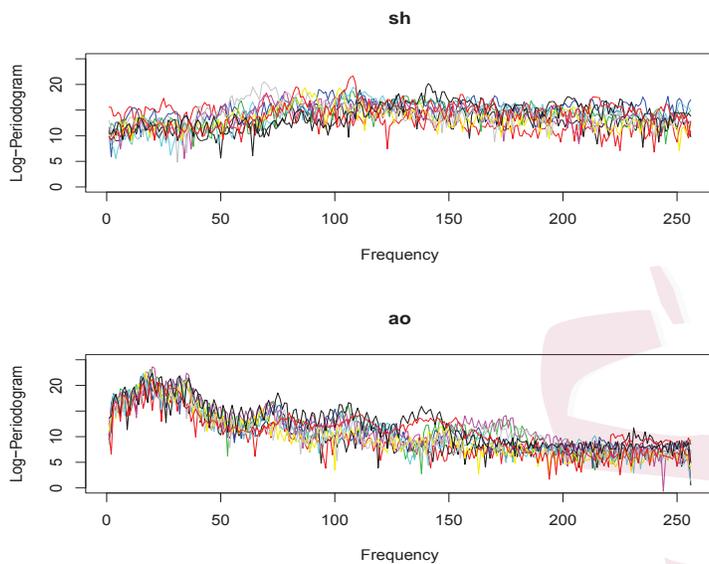


Figure 5: Plots of the speech recognition data.

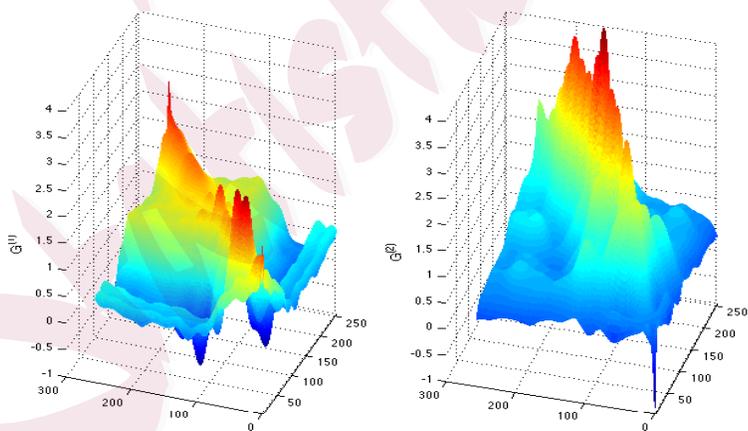


Figure 6: Plots of tensor spline estimators for “sh” and “ao” data sets. Right: “sh”;
Left: “ao”.

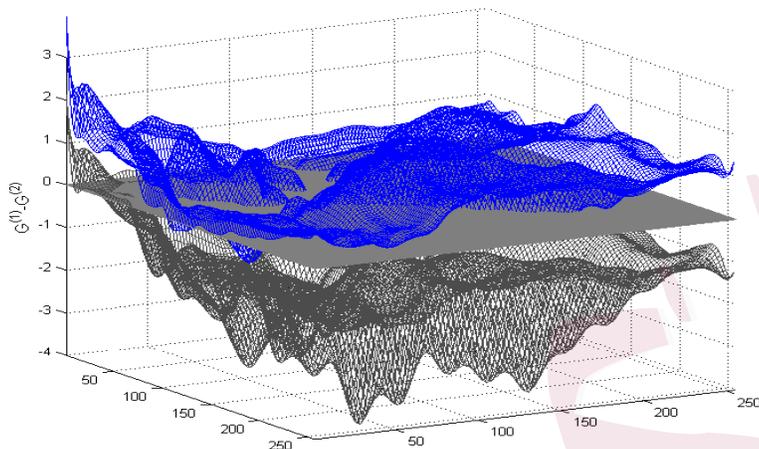


Figure 7: Plot of hypothesis test (6.2) for the speech recognition data. The upper and lower surfaces are the 99.95% confidence envelopes based on cubic tensor product spline covariance estimators and the middle flat surface is the zero plane.

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

$$\begin{aligned}
 H_0 : G^{(1)}(x, x') &\equiv G^{(2)}(x, x'), \quad \forall (x, x') \in [0, 1]^2 \\
 \text{v.s. } H_a : G^{(1)}(x, x') &\neq G^{(2)}(x, x'), \quad \exists (x, x') \in [0, 1]^2.
 \end{aligned} \tag{6.2}$$

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 also tried different numbers of knots and the result is not sensitive to this choice.

7 Summary

In this paper, we consider covariance estimation in functional data and propose a new computationally efficient tensor-product B-spline estimator. The proposed estimator can be used as a building block for further data analysis, such as princi-

pal 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 enjoy superior theoretical properties which have not been well studied in the literature, especially on simultaneous inference for the covariance functions in functional data analysis. The procedure is also 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, the method is applied to a Tecator near-infrared spectra data to test the stationarity assumption on the covariance. In a classification problem, we further extend our method to a two-sample problem to test the equal covariance assumption between different treatment groups.

8 Appendix: Technical Lemmas and Proofs of Theorems

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

The detailed proofs of the following lemmas, Proposition 1 and Proposition 2 can be found in the supplement.

8.1 Preliminaries

For any Lebesgue measurable function $\phi(\mathbf{x})$ on a domain \mathcal{D} , where $\mathcal{D} = [0, 1]$ or $[0, 1]^2$, denote $\|\phi\|_\infty = \sup_{\mathbf{x} \in \mathcal{D}} |\phi(\mathbf{x})|$. For any L^2 integrable functions $\phi(\mathbf{x})$ and $\varphi(\mathbf{x})$, $\mathbf{x} \in \mathcal{D}$, define their theoretical inner product $\langle \phi, \varphi \rangle = \int_{\mathcal{D}} \phi(\mathbf{x}) \varphi(\mathbf{x}) d\mathbf{x}$, with the theoretical L^2 norm $\|\phi\|_2^2 = \langle \phi, \phi \rangle$. Define the empirical inner product of ϕ and φ as $\langle \phi, \varphi \rangle_N = N^{-1} \sum_{1 \leq j \leq N} \phi\left(\frac{j}{N}\right) \varphi\left(\frac{j}{N}\right)$ for $\mathcal{D} = [0, 1]$ and

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

L^2 semi-norm $\|\phi\|_{2,N}^2 = \langle \phi, \phi \rangle_{2,N}$.

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

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

The following lemma is from [Cao et al. \(2012\)](#), which established the upper bound of $\|\mathbf{V}_p^{-1}\|_\infty$.

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

Denote by “ \otimes ” the Kronecker product of two matrices. Note that $\|(\mathbf{A} \otimes \mathbf{A})^{-1}\|_\infty = \|\mathbf{A}^{-1} \otimes \mathbf{A}^{-1}\|_\infty \leq \|\mathbf{A}^{-1}\|_\infty^2$, for any invertible matrix \mathbf{A} , which, together with Lemma 1, leads to the following result.

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

Define

$$\mathbf{X} = \left(\mathbf{B}_{p_2} \left(\frac{2}{N}, \frac{1}{N} \right), \dots, \mathbf{B}_{p_2} \left(1, \frac{1}{N} \right), \dots, \mathbf{B}_{p_2} \left(\frac{1}{N}, 1 \right), \dots, \mathbf{B}_{p_2} \left(1 - \frac{1}{N}, 1 \right) \right)^\top$$

. By elementary algebra and least square algorithm, one obtains

$$\begin{aligned} \tilde{\mathbf{G}}_{p_2}(x, x') &= \mathbf{B}_{p_2}^\top(x, x') (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \bar{\mathbf{U}}, \\ \hat{\mathbf{G}}_{p_1, p_2}(x, x') &= \mathbf{B}_{p_2}^\top(x, x') (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{U}}_{p_1}, \end{aligned}$$

where $\hat{\mathbf{U}}_{p_1} = (\hat{U}_{\cdot 21, p_1}, \dots, \hat{U}_{\cdot N1, p_1}, \dots, \hat{U}_{\cdot 1N, p_1}, \dots, \hat{U}_{\cdot (N-1)N, p_1})^\top$,
 $\bar{\mathbf{U}} = (\bar{U}_{\cdot 21}, \dots, \bar{U}_{\cdot N1}, \dots, \bar{U}_{\cdot 1N}, \dots, \bar{U}_{\cdot (N-1)N})^\top$.

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

$$\begin{aligned} \mathbf{V}_{p_2, 2} &= (\langle B_{JJ', p_2}, B_{J''J''', p_2} \rangle)_{J, J', J'', J'''=1-p_2}^{N_{s_2}}, \\ \hat{\mathbf{V}}_{p_2, 2} &= (\langle B_{JJ', p_2}, B_{J''J''', p_2} \rangle_{2,N})_{J, J', J'', J'''=1-p_2}^{N_{s_2}}. \end{aligned} \quad (8.1)$$

It is easy to obtain $\hat{\mathbf{V}}_{p_2, 2} = N^{-2} (\mathbf{X}^\top \mathbf{X})$, so we now study the properties of $\hat{\mathbf{V}}_{p_2, 2}$ and $\mathbf{V}_{p_2, 2}$. The following result shows that the difference of $\mathbf{V}_{p_2, 2}$ and $\hat{\mathbf{V}}_{p_2, 2}$ is

negligible. Using the results from Lemma 2, we can obtain the upper bound of the norm of $\hat{\mathbf{V}}_{p_2,2}^{-1}$.

Lemma 3. *Under Assumption (A3), for $\mathbf{V}_{p_2,2}$ and $\hat{\mathbf{V}}_{p_2,2}$ defined in (8.1), $\|\mathbf{V}_{p_2,2} - \hat{\mathbf{V}}_{p_2,2}\|_\infty = O(N^{-1})$ and $\|\hat{\mathbf{V}}_{p_2,2}^{-1}\|_\infty = O(h_{s_2}^{-2})$.*

Lemma 4. *For $\hat{\mathbf{V}}_{p_2,2}$ defined in (8.1) and any $N(N-1)$ vector $\boldsymbol{\rho} = (\rho_{jj'})$, there exists a constant $C > 0$, such that*

$$\sup_{(x,x') \in [0,1]^2} \left\| N^{-2} \mathbf{B}_{p_2}^\top(x, x') \hat{\mathbf{V}}_{p_2,2}^{-1} \mathbf{X}^\top \boldsymbol{\rho} \right\|_\infty \leq C \|\boldsymbol{\rho}\|_\infty.$$

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

$$\tilde{\phi}_{kk'}(x, x') = \mathbf{B}_{p_2}^\top(x, x') (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \phi_{kk'}, \quad (8.2)$$

and $\boldsymbol{\phi}_{kk'} = \{\phi_k(2/N)\phi_{k'}(1/N), \dots, \phi_k(1)\phi_{k'}(1/N), \dots, \phi_k(1/N)\phi_{k'}(1), \dots, \phi_k(1-1/N)\phi_{k'}(1)\}^\top$.

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

Lemma 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 Assumption (A2) holds,*

$$\begin{aligned} & \sup_{(x,x') \in [0,1]^2} \left| \phi_{kk'}(x, x') - \tilde{\phi}_{kk'}(x, x') \right| \\ & \leq C_\phi h_{s_2}^{p-1+\mu} \left(\|\phi_{k'}\|_\infty \left\| \phi_k^{(p-1)} \right\|_{0,\mu} + \|\phi_k\|_\infty \left\| \phi_{k'}^{(p-1)} \right\|_{0,\mu} \right), \end{aligned}$$

where $\tilde{\phi}_{kk'}(x, x')$ is defined in (8.2). 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 $\phi \in C[0,1]$.

8.2 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 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

$$\begin{aligned}
 nE [\Delta(x, x')]^2 &= \sum_{k, k'=1}^{\infty} \phi_k^2(x) \phi_{k'}^2(x') + \sum_{k, k'=1}^{\infty} \phi_k(x') \phi_k(x) \phi_{k'}(x) \phi_{k'}(x') \\
 &\quad + \sum_{k=1}^{\infty} \phi_k^2(x) \phi_k^2(x') (E\xi_{1k}^4 - 3) \\
 &= G^2(x, x') + G(x, x) G(x', x') + \sum_{k=1}^{\infty} \phi_k^2(x) \phi_k^2(x') (E\xi_{1k}^4 - 3) \\
 &\equiv V(x, x'),
 \end{aligned}$$

and the desired result follows from Proposition 2.

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

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

PROOF OF THEOREM 2. According to Lemma 6, Propositions 1 and 2 and Theorem 1, for $\forall \alpha \in (0, 1)$, as $n \rightarrow \infty$,

$$\begin{aligned}
 &\lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} \sqrt{n} \left| \hat{G}_{p_1, p_2}(x, x') - G(x, x') \right| V^{-1/2}(x, x') \leq Q_{1-\alpha} \right\} \\
 &= \lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} \sqrt{n} \left| \tilde{G}_{p_1, p_2}(x, x') - G(x, x') \right| V^{-1/2}(x, x') \leq Q_{1-\alpha} \right\} \\
 &= \lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} |\zeta(x, x')| V^{-1/2}(x, x') \leq Q_{1-\alpha} \right\} \\
 &= \lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} |\zeta_Z(x, x')| V^{-1/2}(x, x') \leq Q_{1-\alpha} \right\}.
 \end{aligned}$$

Acknowledgment Wang's research was partially supported by NSF awards DMS 1106816 and DMS 1309800. Li's research was partially supported by NSF awards DMS 1105634 and NSF CAREER award DMS 1317118. Yang's

research was partially supported by DMS 0706518, 1007594, Jiangsu Specially-Appointed Professor Program SR10700111, Jiangsu Province Key-Discipline Program (Statistics) ZY107002, ZY107992, National Natural Science Foundation of China award 11371272, and Research Fund for the Doctoral Program of Higher Education of China award 20133201110002. We are grateful to the Editor, the associate editor and the referees for their valuable comments and suggestions which led to significant improvement in this paper.

References

- BIGOT, J., BISCAY, R., LOUBES, J. & MUNIZ-ALVAREZ, L. (2010). Non-parametric estimation of covariance functions by model selection. *Electronic Journal of Statistics* **4**, 822–855.
- CAI, T. & HALL, P. (2005). Prediction in functional linear regression. *The Annals of Statistics* **34**, 2159–2179.
- CAO, G., YANG, L. & TODEM, D. (2012). Simultaneous inference for the mean function of dense functional data. *Journal of Nonparametric Statistics* **24**, 359–377.
- CLAESKENS, G. & VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics* **31**, 1852–1884.
- CRAINICEANUA, C. M., STAIUCU, A.-M. & DI, C.-Z. (2009). Generalized multilevel functional regression. *Journal of the American Statistical Association* **104**, 1550–1561.
- DE BOOR, C. (2001). *A Practical Guide to Splines*. New York: Springer-Verlag.
- DELAIGLE, A. & HALL, P. (2012). Achieving near-perfect classification for functional data. *Journal of the Royal Statistical Society, Series B* **74**, 267–286.
- FAN, J., HUANG, T. & LI, R. (2007). Analysis of longitudinal data with semi-parametric estimation of covariance function. *Journal of the American Statistical Association* **102**, 632–642.

- FERRATY, F. & VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Berlin: Springer Series in Statistics, Springer.
- HALL, P., MÜLLER, H. G. & WANG, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34**, 1493–1517.
- HÄRDLE, W. & MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *The Annals of Statistics* **19**, 778–796.
- HASTIE, T., BUJA, A. & TIBSHIRANI, R. (1995). Penalized discriminant analysis. *The Annals of Statistics* **23**, 73–102.
- HUANG, J. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics* **31**, 1600–1635.
- HUANG, J. & YANG, L. (2004). Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society, Series B* **66**, 463–477.
- JAMES, G. & HASTIE, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, Series B* **63**, 533–550.
- JAMES, G. M., HASTIE, T. & SUGAR, C. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587–602.
- LI, Y. & HSING, T. (2010a). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *The Annals of Statistics* **38**, 3028–3062.
- LI, Y. & HSING, T. (2010b). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **38**, 3321–3351.
- LI, Y., WANG, N. & CARROLL, R. J. (2010). Generalized functional linear models with semiparametric single-index interactions. *Journal of the American Statistical Association* **105**, 621–633.

- LI, Y., WANG, N. & CARROLL, R. J. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association* **108**, 1284–1294.
- 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: *Longitudinal Data Analysis (Handbooks of Modern Statistical Methods)*. New York: Wiley.
- RAMSAY, J. O. & SILVERMAN, B. W. (2005). *Functional Data Analysis. Second Edition*. New York: Springer Series in Statistics. Springer.
- WANG, J., LIU, R., CHENG, F. & YANG, L. (2014). Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. *The Annals of Statistics* **42**, 654–668.
- WANG, N., CARROLL, R. J. & LIN, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustering data. *Journal of the American Statistical Association* **100**, 147–157.
- 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 data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.
- YAO, F., MÜLLER, H. G. & WANG, J. L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33**, 2873–2903.
- ZHAO, X., MARRON, J. S. & WELLS, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica* **14**, 789–808.

ZHAO, Z. & WU, W. (2008). Confidence bands in nonparametric time series regression. *The Annals of Statistics* **36**, 1854–1878.

ZHENG, S., YANG, L. & HÄRDLE, W. (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association* **109**, 661–673.

ZHOU, L., HUANG, J. & CARROLL, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**, 601–619.

Guanqun Cao

Department of Mathematics and Statistics, Auburn University

E-mail: (gzc0009@auburn.edu)

Lily Wang

Department of Statistics and Statistical Laboratory, Iowa State University

E-mail: (lilywang@iastate.edu)

Yehua Li

Department of Statistics and Statistical Laboratory, Iowa State University

E-mail: (yehuali@iastate.edu)

Lijian Yang

Center for Advanced Statistics and Econometrics Research, Soochow University

E-mail: (yanglijian@suda.edu.cn)