Journal of Nonparametric Statistics
Vol. 00, No. 0, Month 2011, 1–19

Taylor & Francis
Taylor & Francis Group

Simultaneous inference for the mean function based on dense functional data

Guanqun Cao[b], Lijian Yang[a,b]* and David Todem[c]

[a] *Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou 215006, People's Republic of China;* [b] *Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA;* [c] *Division of Biostatistics, Department of Epidemiology, Michigan State University, East Lansing, MI 48824, USA*

(*Received 27 April 2011; final version received 26 October 2011*)

A polynomial spline estimator is proposed for the mean function of dense functional data together with a simultaneous confidence band which is asymptotically correct. In addition, the spline estimator and its accompanying confidence band enjoy oracle efficiency in the sense that they are asymptotically the same as if all random trajectories are observed entirely and without errors. The confidence band is also extended to the difference of mean functions of two populations of functional data. Simulation experiments provide strong evidence that corroborates the asymptotic theory while computing is efficient. The confidence band procedure is illustrated by analysing the near-infrared spectroscopy data.

Keywords: *B*-spline; confidence band; functional data; Karhunen–Loève L^2 representation; oracle efficiency

AMS Subject Classifications: Primary: 62M10; Secondary: 62G08

1. Introduction

In functional data analysis problems, estimation of mean functions is the fundamental first step (see, e.g. Cardot 2000; Rice and Wu 2001; Cuevas *et al.* 2006, Ferraty and Vieu 2006; Degras 2011; Ma *et al.* 2011). According to Ramsay and Silverman (2005), functional data consist of a collection of iid realisations $\{\eta_i(x)\}_{i=1}^n$ of a smooth random function $\eta(x)$, with unknown mean function $E\eta(x) = m(x)$ and covariance function $G(x, x') = \mathrm{cov}\{\eta(x), \eta(x')\}$. Although the domain of $\eta(\cdot)$ is an entire interval $\mathcal{X}$, the recording of each random curve $\eta_i(x)$ is only over a finite number N_i of points in $\mathcal{X}$, and contaminated with measurement errors. Without loss of generality, we take $\mathcal{X} = [0, 1]$.

Denote by Y_{ij} the jth observation of the random curve $\eta_i(\cdot)$ at time point X_{ij}, $1 \le i \le n, 1 \le j \le N_i$. Although we refer to variable X_{ij} as time, it could also be other numerical measures, such as wavelength in Section 6. In this paper, we examine the equally spaced dense design, in other words, $X_{ij} = j/N, 1 \le i \le n, 1 \le j \le N$, with N going to infinity. For the ith subject,

*Corresponding author. Email: yangli@stt.msu.edu

ISSN 1048-5252 print/ISSN 1029-0311 online

http://dx.doi.org/10.1080/10485252.2011.638071
http://www.tandfonline.com

$i = 1, 2, \ldots, n$, its sample path $\{j/N, Y_{ij}\}$ is the noisy realisation of the continuous time stochastic process $\eta_i(x)$ in the sense that $Y_{ij} = \eta_i(j/N) + \sigma(j/N)\varepsilon_{ij}$, with errors ε_{ij} satisfying $E(\varepsilon_{ij}) = 0$, $E(\varepsilon_{ij}^2) = 1$, and $\{\eta_i(x), x \in [0, 1]\}$ are iid copies of the process $\{\eta(x), x \in [0, 1]\}$ which is L^2, i.e. $E \int_{[0,1]} \eta^2(x)\,\mathrm{d}x < +\infty$.

For the standard process $\{\eta(x), x \in [0, 1]\}$, let sequences $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\psi_k(x)\}_{k=1}^{\infty}$ be the eigenvalues and eigenfunctions of $G(x, x')$. respectively, in which $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\sum_{k=1}^{\infty} \lambda_k < \infty$, $\{\psi_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L^2([0, 1])$ and $G(x, x') = \sum_{k=1}^{\infty} \lambda_k \psi_k(x)\psi_k(x')$, which implies that $\int G(x, x')\psi_k(x')\mathrm{d}x' = \lambda_k\psi_k(x)$. The process $\{\eta_i(x), x \in [0, 1]\}$ allows the Karhunen–Loève L^2 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 $\phi_k = \sqrt{\lambda_k}\psi_k$. In what follows, we assume that $\lambda_k = 0$, for $k > \kappa$, where κ is a positive integer or ∞, thus $G(x, x') = \sum_{k=1}^{\kappa} \phi_k(x)\phi_k(x')$ and the model that we consider is

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

Although the sequences $\{\lambda_k\}_{k=1}^{\kappa}$, $\{\phi_k(\cdot)\}_{k=1}^{\kappa}$ and the random coefficients ξ_{ik} exist mathematically, they are unknown or unobservable, respectively.

The existing literature focuses on two data types. Yao, Müller, and Wang (2005) studied sparse longitudinal data for which N_i, i.e. the number of observations for the ith curve, is bounded and follows a given distribution, in which case Ma *et al.* (2011) obtained asymptotically simultaneous confidence band for the mean function of the functional data, using piecewise constant spline estimation. Li and Hsing (2010a) established uniform convergence rate for local linear estimation of mean and covariance function of dense functional data, where $\min_{1\leq i\leq n} N_i \gg (n/\log n)^{1/4}$ as $n \to \infty$ similar to our Assumption (A3), but did not provide asymptotic distribution of maximal deviation or simultaneous confidence band. Degras (2011) built asymptotically correct simultaneous confidence band for dense functional data using local linear estimator. Bunea, Ivanescu, and Wegkamp (2011) proposed asymptotically conservative rather than correct confidence set for the mean function of Gaussian functional data.

In this paper, we propose polynomial spline confidence band for the mean function based on dense functional data. In function estimation problems, simultaneous confidence band is an important tool to address the variability in the mean curve; see Zhao and Wu (2008), Zhou, Shen, and wolfe (1998) and Zhou and Wu (2010) for related theory and applications. The fact that simultaneous confidence bands have not been widely used for functional data analysis is certainly not due to lack of interesting applications, but to the greater technical difficulty to formulate such bands for functional data and establish their theoretical properties. In this work, we have established asymptotic correctness of the proposed confidence band using various properties of spline smoothing. The spline estimator and the accompanying confidence band are asymptotically the same as if all the n random curves are recorded over the entire interval, without measurement errors. They are oracally efficient despite the use of spline smoothing (see Remark 2.2). This provides partial theoretical justification for treating functional data as perfectly recorded random curves over the entire data range, as in Ferraty and Vieu (2006). Theorem 3 of Hall, Müller, and Wang (2006) stated mean-square (rather than the stronger uniform) oracle efficiency for local linear estimation of eigenfunctions and eigenvalues (rather than the mean function), under assumptions similar to ours, but provided only an outline of proof. Among the existing works on functional data analysis, Ma *et al.* (2011) proposed the simultaneous confidence band for sparse functional data. However, their result does not enjoy the oracle efficiency stated in Theorem 2.1, since there are not enough observations for each subject to obtain a good estimate of the individual trajectories. As a result, it has the slow nonparametric convergence rate of $n^{-1/3}\log n$, instead of the parametric rate of $n^{-1/2}$ as this paper. This essential difference completely separates dense functional data from sparse ones.

The aforementioned confidence band is also extended to the difference of two regression functions. This is motivated by Li and Yu (2008), which applied functional segment discriminant analysis to a Tecator data set (Figure 3). In this data set, each observation (meat) consists of a 100-channel absorbance spectrum in the wavelength with different fat, water and protein percent. Li and Yu (2008) used the spectra to predict whether the fat percentage is greater than 20%. On the flip side, we are interested in building a $100(1-\alpha)\%$ confidence band for the difference between regression functions from the spectra of the less than 20% fat group and the higher than 20% fat group. If this $100(1-\alpha)\%$ confidence band covers the zero line, one accepts the null hypothesis of no difference between the two groups, with p-value no greater than α. Test for equality between two groups of curves based on the adaptive Neyman test and wavelet thresholding techniques were proposed in Fan and Lin (1998), which did not provide an estimator of the difference of the two mean functions nor a simultaneous confidence band for such estimator. As a result, their test did not extend to testing other important hypotheses on the difference of the two mean functions while our Theorem 2.4 provides a benchmark for all such testing. More recently, Benko, Härdle, and Kneip (2009) developed two-sample bootstrap tests for the equality of eigenfunctions, eigenvalues and mean functions by using common functional principal components and bootstrap tests.

The paper is organised as follows. Section 2 states main theoretical results on confidence bands constructed from polynomial splines. Section 3 provides further insights into the error structure of spline estimators. The actual steps to implement the confidence bands are provided in Section 4. A simulation study is presented in Section 5, and an empirical illustration on how to use the proposed spline confidence band for inference is reported in Section 6. Technical proofs are collected in the appendix.

2. Main results

For any Lebesgue measurable function ϕ on $[0,1]$, denote $\|\phi\|_\infty = \sup_{x\in[0,1]} |\phi(x)|$. For any $\nu \in (0,1]$ and non-negative integer q, let $C^{q,\nu}[0,1]$ be the space of functions with ν-Hölder continuous qth-order derivatives on $[0,1]$, i.e.

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

To describe the spline functions, we first introduce a sequence of equally spaced points $\{t_J\}_{J=1}^{N_m}$, called interior knots which divide the interval $[0,1]$ into (N_m+1) equal subintervals $I_J = [t_J, t_{J+1})$, $J = 0, \ldots, N_m - 1$, $I_{N_m} = [t_{N_m}, 1]$. For any positive integer p, introduce left boundary knots $t_{1-p}, \ldots, t_0$, and right boundary knots $t_{N_m+1}, \ldots, t_{N_m+p}$,

$$t_{1-p} = \cdots = t_0 = 0 < t_1 < \cdots < t_{N_m} < 1 = t_{N_m+1} = \cdots = t_{N_m+p},$$
$$t_J = Jh_m, \quad 0 \le J \le N_m + 1, h_m = \frac{1}{(N_m+1)},$$

in which h_m is the distance between neighbouring knots. Denote by $\mathcal{H}^{(p-2)}$ the space of pth-order spline space, i.e. $p-2$ times continuously differentiable functions on $[0,1]$ that are polynomials of degree $p-1$ on $[t_J, t_{J+1}]$, $J = 0, \ldots, N_m$. Then, $\mathcal{H}^{(p-2)} = \{\sum_{J=1-p}^{N_m} b_{J,p} B_{J,p}(x), b_{J,p} \in \mathcal{R}, x \in [0,1]\}$, where $B_{J,p}$ is the Jth B-spline basis of order p as defined in de Boor (2001).

We propose to estimate the mean function $m(x)$ by

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

The technical assumptions we need are as follows.

(A1) The regression function $m \in C^{p-1,1}[0,1]$, i.e. $m^{(p-1)} \in C^{0,1}[0,1]$.
(A2) The standard deviation function $\sigma(x) \in C^{0,\mu}[0,1]$ for some $\mu \in (0,1]$.
(A3) As $n \to \infty$, $N^{-1}n^{1/(2p)} \to 0$ and $N = O(n^{\theta})$ for some $\theta > 1/(2p)$, the number of interior knots N_m satisfies $NN_m^{-1} \to \infty$, $N_m^{-p}n^{1/2} \to 0$, $N^{-1/2}N_m^{1/2}\log n \to 0$ or equivalently $Nh_m \to \infty$, $h_m^p n^{1/2} \to 0$, $N^{-1/2}h_m^{-1/2}\log n \to 0$.
(A4) There exists $C_G > 0$ such that $G(x,x) \geq C_G$, $x \in [0,1]$, for $k \in \{1,\ldots,\kappa\}$, $\phi_k(x) \in C^{0,\mu}[0,1]$, $\sum_{k=1}^{\kappa} \|\phi_k\|_\infty < \infty$ and as $n \to \infty$, $h_m^{\mu} \sum_{k=1}^{\kappa_n} \|\phi_k\|_{0,\mu} = o(1)$ for a sequence $\{\kappa_n\}_{n=1}^{\infty}$ of increasing integers, with $\lim_{n\to\infty}\kappa_n = \kappa$ and the constant $\mu \in (0,1]$ as in Assumption (A2). In particular, $\sum_{k=\kappa_n+1}^{\kappa} \|\phi_k\|_\infty = o(1)$.
(A5) There are constants $C_1, C_2 \in (0,+\infty)$, $\gamma_1,\gamma_2 \in (1,+\infty)$, $\beta \in (0,1/2)$ and iid $N(0,1)$ variables $\{Z_{ik,\xi}\}_{i=1,k=1}^{n,\kappa}$, $\{Z_{ij,\varepsilon}\}_{i=1,j=1}^{n,N}$ such that

$$\max_{1\leq k\leq\kappa} P\left[\max_{1\leq t\leq n}\left|\sum_{i=1}^{t}\xi_{ik} - \sum_{i=1}^{t} Z_{ik,\xi}\right| > C_1 n^{\beta}\right] < C_2 n^{-\gamma_1}, \tag{3}$$

$$P\left\{\max_{1\leq j\leq N}\max_{1\leq t\leq n}\left|\sum_{i=1}^{t}\varepsilon_{ij} - \sum_{i=1}^{t} Z_{ij,\varepsilon}\right| > C_1 n^{\beta}\right\} < C_2 n^{-\gamma_2}. \tag{4}$$

Assumptions (A1) and (A2) are typical for spline smoothing (Huang and Yang 2004; Xue and Yang 2006; Wang and Yang 2009a; Li and Yang 2010; Ma and Yang 2011). Assumption (A3) concerns the number of observations for each subject and the number of knots of B-splines. Assumption (A4) ensures that the principal components have collectively bounded smoothness. Assumption (A5) provides the Gaussian approximation of estimation error process and is ensured by the following elementary assumption.

(A5') There exist $\eta_1 > 4$, $\eta_2 > 4 + 2\theta$ such that $E|\xi_{ik}|^{\eta_1} + E|\varepsilon_{ij}|^{\eta_2} < +\infty$, for $1 \leq i < \infty$, $1 \leq k \leq \kappa$, $1 \leq j < \infty$. The number κ of nonzero eigenvalues is finite or κ is infinite while the variables $\{\xi_{ik}\}_{1\leq i<\infty, 1\leq k<\infty}$ are iid.

Degras (2011) makes a restrictive assumption (A.2) on the Hölder continuity of the stochastic process $\eta(x) = m(x) + \sum_{k=1}^{\infty}\xi_k\phi_k(x)$. It is elementary to construct examples where our Assumptions (A4) and (A5) are satisfied while assumption (A.2) of Degras (2011) is not.

The part of Assumption (A4) on ϕ_k's holds trivially if κ is finite and all $\phi_k(x) \in C^{0,\mu}[0,1]$. Note also that by definition, $\phi_k = \sqrt{\lambda_k}\psi_k$, $\|\phi_k\|_\infty = \sqrt{\lambda_k}\|\psi_k\|_\infty$, $\|\phi_k\|_{0,\mu} = \sqrt{\lambda_k}\|\psi_k\|_{0,\mu}$, in which $\{\psi_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L^2([0,1])$, hence, Assumption (A4) is fulfilled for $\kappa = \infty$ as long as λ_k decreases to zero sufficiently fast. Following one referee's suggestion, we provide the following example. One takes $\lambda_k = \rho^{2[k/2]}$, $k = 1,2,\ldots$ for any $\rho \in (0,1)$, with $\{\psi_k\}_{k=1}^{\infty}$ the canonical orthonormal Fourier basis of $L^2([0,1])$

$$\psi_1(x) \equiv 1, \psi_{2k+1}(x) \equiv \sqrt{2}\cos(k\pi x)$$

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

In this case, $\sum_{k=1}^{\infty}\|\phi_k\|_\infty = 1 + \sum_{k=1}^{\infty}\rho^k(\sqrt{2}+\sqrt{2}) = 1 + 2\sqrt{2}\rho(1-\rho)^{-1} < \infty$, while for any $\{\kappa_n\}_{n=1}^{\infty}$ with κ_n increasing, odd and $\kappa_n \to \infty$, and Lipschitz order $\mu = 1$

$$\begin{aligned} h_m \sum_{k=1}^{\kappa_n} \|\phi_k\|_{0,1} &= h_m \sum_{k=1}^{(\kappa_n-1)/2} \rho^k(\sqrt{2}k\pi + \sqrt{2}k\pi) \\ &\le 2\sqrt{2}\pi h_m \rho \sum_{k=1}^{\infty} \rho^{k-1}k = 2\sqrt{2}\pi h_m (1-\rho)^{-2} \\ &= O(h_m) = o(1). \end{aligned}$$

Denote by $\zeta(x)$, $x \in [0,1]$ a standardised Gaussian process such that $E\zeta(x) \equiv 0$, $E\zeta^2(x) \equiv 1$, $x \in [0,1]$ with covariance function

$$E\zeta(x)\zeta(x') = G(x,x')\{G(x,x)G(x',x')\}^{-1/2}, x, x' \in [0,1]$$

and define the $100 \times (1-\alpha)$th percentile of the absolute maxima distribution of $\zeta(x)$, $\forall x \in [0,1]$, i.e. $P[\sup_{x\in[0,1]}|\zeta(x)| \le Q_{1-\alpha}] = 1-\alpha$, $\forall\alpha \in (0,1)$. Denote by $z_{1-\alpha/2}$ the $100(1-\alpha/2)$th percentile of the standard normal distribution. Define also the following 'infeasible estimator' of function m

$$\bar{m}(x) = \bar{\eta}(x) = n^{-1}\sum_{i=1}^{n}\eta_i(x), \quad x \in [0,1]. \tag{5}$$

The term 'infeasible' refers to the fact that $\bar{m}(x)$ is computed from an unknown quantity $\eta_i(x)$, $x \in [0,1]$, and it would be the natural estimator of $m(x)$ if all the iid random curves $\eta_i(x), x \in [0,1]$ were observed, a view taken in Ferraty and Vieu (2006).

We now state our main results in the following theorem.

THEOREM 2.1 *Under Assumptions (A1)–(A5), $\forall\alpha \in (0,1)$, as $n \to \infty$, the 'infeasible estimator' $\bar{m}(x)$ converges at the $\sqrt{n}$ rate*

$$P\left\{\sup_{x\in[0,1]} n^{1/2}|\bar{m}(x) - m(x)|G(x,x)^{-1/2} \le Q_{1-\alpha}\right\} \to 1-\alpha,$$
$$P\{n^{1/2}|\bar{m}(x) - m(x)|G(x,x)^{-1/2} \le z_{1-\alpha/2}\} \to 1-\alpha, \quad \forall x \in [0,1],$$

while the spline estimator $\hat{m}_p$ is asymptotically equivalent to $\bar{m}$ up to order $n^{1/2}$, i.e.

$$\sup_{x\in[0,1]} n^{1/2}|\bar{m}(x) - \hat{m}_p(x)| = o_P(1).$$

Remark 2.2 The significance of Theorem 2.1 lies in the fact that one does not need to distinguish between the spline estimator $\hat{m}_p$ and the 'infeasible estimator' $\bar{m}$ in Equation (5), which converges with $\sqrt{n}$ rate like a parametric estimator. We therefore have established oracle efficiency of the nonparametric estimator $\hat{m}_p$.

COROLLARY 2.3 *Under Assumptions (A1)–(A5), as $n \to \infty$, an asymptotic $100(1-\alpha)\%$ correct confidence band for $m(x), x \in [0,1]$ is*

$$\hat{m}_p(x) \pm G(x,x)^{1/2}Q_{1-\alpha}n^{-1/2}, \quad \forall\alpha \in (0,1)$$

while an asymptotic $100(1-\alpha)\%$ pointwise confidence interval for $m(x), x \in [0,1]$, is $\hat{m}_p(x) \pm G(x,x)^{1/2}z_{1-\alpha/2}n^{-1/2}$.

We next describe a two-sample extension of Theorem 2.1. Denote two samples indicated by $d = 1, 2$, which satisfy

$$Y_{dij} = m_d\left(\frac{j}{N}\right) + \sum_{k=1}^{\kappa_d} \xi_{dik}\phi_{dk}\left(\frac{j}{N}\right) + \sigma_d\left(\frac{j}{N}\right)\varepsilon_{dij}, \quad 1 \le i \le n_d, \quad 1 \le j \le N,$$

with covariance functions $G_d(x, x') = \sum_{k=1}^{\kappa_d} \phi_{dk}(x)\phi_{dk}(x')$, respectively. We denote the ratio of two-sample sizes as $\hat{r} = n_1/n_2$ and assume that $\lim_{n_1\to\infty} \hat{r} = r > 0$.

For both groups, let $\hat{m}_{1p}(x)$ and $\hat{m}_{2p}(x)$ be the order p spline estimates of mean functions $m_1(x)$ and $m_2(x)$ by Equation (2). Also denote by $\zeta_{12}(x), x \in [0, 1]$ a standardised Gaussian process such that $E\zeta_{12}(x) \equiv 0$, $E\zeta_{12}^2(x) \equiv 1$, $x \in [0, 1]$ with covariance function

$$E\zeta_{12}(x)\zeta_{12}(x') = \frac{G_1(x, x') + rG_2(x, x')}{\{G_1(x, x) + rG_2(x, x)\}^{1/2}\{G_1(x, x') + rG_2(x, x')\}^{1/2}}, \; x, x' \in [0, 1].$$

Denote by $Q_{12,1-\alpha}$ the $(1-\alpha)$th quantile of the absolute maxima deviation of $\zeta_{12}(x), x \in [0, 1]$ as above. We mimic the two-sample t-test and state the following theorem whose proof is analogous to that of Theorem 2.1.

THEOREM 2.4 *If Assumptions* (A1)–(A5) *are modified for each group accordingly, then for any* $\alpha \in (0, 1)$, *as* $n_1 \to \infty$, $\hat{r} \to r > 0$,

$$P\left\{\sup_{x\in[0,1]} \frac{n_1^{1/2}|(\hat{m}_{1p} - \hat{m}_{2p} - m_1 + m_2)(x)|}{\{(G_1 + rG_2)(x, x)\}^{1/2}} \le Q_{12,1-\alpha}\right\} \to 1 - \alpha.$$

Theorem 2.4 yields an uniform asymptotic confidence band for $m_1(x) - m_2(x)$, $x \in [0, 1]$.

COROLLARY 2.5 *If Assumptions* (A1)–(A5) *are modified for each group accordingly, as* $n_1 \to \infty$, $\hat{r} \to r > 0$, *a* $100 \times (1-\alpha)\%$ *asymptotically correct confidence band for* $m_1(x) - m_2(x)$, $x \in [0, 1]$ *is* $(\hat{m}_{1p} - \hat{m}_{2p})(x) \pm n_1^{-1/2} Q_{12,1-\alpha}\{(G_1 + rG_2)(x, x)\}^{1/2}$, $\forall\alpha \in (0, 1)$.

If the confidence band in Corollary 2.3 is used to test the hypothesis

$$H_0 : m(x) = m_0(x), \quad \forall x \in [0, 1] \longleftrightarrow H_a : m(x) \ne m_0(x), \quad \text{for some } x \in [0, 1],$$

for some given function $m_0(x)$, as one referee pointed out, the asymptotic power of the test is α under H_0, 1 under H_1 due to Theorem 2.1. The same can be said for testing the hypothesis about $m_1(x) - m_2(x)$ using the confidence band in Corollary 2.5.

3. Error decomposition for the spline estimators

In this section, we break the estimation error $\hat{m}_p(x) - m(x)$ into three terms. We begin by discussing the representation of the spline estimator $\hat{m}_p(x)$ in Equation (2).

The definition of $\hat{m}_p(x)$ in Equation (2) means that

$$\hat{m}_p(x) \equiv \sum_{J=1-p}^{N_m} \hat{\beta}_{J,p} B_{J,p}(x),$$

with coefficients $\{\hat{\beta}_{1-p,p}, \ldots, \hat{\beta}_{N_m,p}\}^{\mathrm{T}}$ solving the following least-squares problem

$$\{\hat{\beta}_{1-p,p}, \ldots, \hat{\beta}_{N_m,p}\}^{\mathrm{T}} = \underset{\{\beta_{1-p,p},\ldots,\beta_{N_m,p}\}\in R^{N_m+p}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{N} \left\{ Y_{ij} - \sum_{J=1-p}^{N_m} \beta_{J,p} B_{J,p}\left(\frac{j}{N}\right) \right\}^2. \tag{6}$$

Applying elementary algebra, one obtains

$$\hat{m}_p(x) = \{B_{1-p,p}(x), \ldots, B_{N_m,p}(x)\}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}, \tag{7}$$

where $\mathbf{Y} = (\bar{Y}_{\cdot 1}, \ldots, \bar{Y}_{\cdot N})^{\mathrm{T}}$, $\bar{Y}_{\cdot j} = n^{-1}\sum_{i=1}^{n} Y_{ij}$, $1 \le j \le N$, and the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} B_{1-p,p}\left(\frac{1}{N}\right) & \cdots & B_{N_m,p}\left(\frac{1}{N}\right) \\ \vdots & \cdots & \vdots \\ B_{1-p,p}\left(\frac{N}{N}\right) & \cdots & B_{N_m,p}\left(\frac{N}{N}\right) \end{pmatrix}_{N\times(N_m+p)}.$$

Projecting via Equation (7) the relationship in model (1) onto the linear subspace of R^{N_m+p} spanned by $\{B_{J,p}(j/N)\}_{1\le j\le N, 1-p\le J\le N_m}$, we obtain the following crucial decomposition in the space $\mathcal{H}^{(p-2)}$ of spline functions:

$$\hat{m}_p(x) = \tilde{m}_p(x) + \tilde{e}_p(x) + \tilde{\xi}_p(x), \tag{8}$$

where

$$\tilde{m}_p(x) = \sum_{J=1-p}^{N_m} \tilde{\beta}_{J,p} B_{J,p}(x), \tilde{\varepsilon}_p(x) = \sum_{J=1-p}^{N_m} \tilde{a}_{J,p} B_{J,p}(x),$$

$$\tilde{\xi}_p(x) = \sum_{k=1}^{\kappa} \tilde{\xi}_{k,p}(x), \tilde{\xi}_{k,p}(x) = \sum_{J=1-p}^{N_m} \tilde{\tau}_{k,J,p} B_{J,p}(x). \tag{9}$$

The vectors $\{\tilde{\beta}_{1-p}, \ldots, \tilde{\beta}_{N_m}\}^{\mathrm{T}}$, $\{\tilde{a}_{1-p}, \ldots, \tilde{a}_{N_m}\}^{\mathrm{T}}$ and $\{\tilde{\tau}_{k,1-p}, \ldots, \tilde{\tau}_{k,N_m}\}^{\mathrm{T}}$ in Equation (9) are solutions to Equation (6) with Y_{ij} replaced by $m(j/N)$, $\sigma(j/N)\varepsilon_{ij}$ and $\xi_{ik}\phi_k(j/N)$, respectively. Alternatively,

$$\begin{aligned} \tilde{m}_p(x) &= \{B_{1-p,p}(x), \ldots, B_{N_m,p}(x)\}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{m}, \\ \tilde{e}_p(x) &= \{B_{1-p,p}(x), \ldots, B_{N_m,p}(x)\}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{e}, \\ \tilde{\xi}_{k,p}(x) &= \bar{\xi}_{\cdot k}\{B_{1-p,p}(x), \ldots, B_{N_m,p}(x)\}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\phi}_k, 1 \le k \le \kappa, \end{aligned}$$

in which $\mathbf{m} = (m(1/N), \ldots, m(N/N))^{\mathrm{T}}$ is the signal vector, $\mathbf{e} = (\sigma(1/N)\bar{\varepsilon}_{\cdot 1}, \ldots, \sigma(N/N)\bar{\varepsilon}_{\cdot N})^{\mathrm{T}}$, $\bar{\varepsilon}_{\cdot j} = n^{-1}\sum_{i=1}^{n}\varepsilon_{ij}$, $1 \le j \le N$ is the noise vector and $\boldsymbol{\phi}_k = (\phi_k(1/N), \ldots, \phi_k(N/N))^{\mathrm{T}}$ are the eigenfunction vectors, and $\bar{\xi}_{\cdot k} = n^{-1}\sum_{i=1}^{n}\xi_{ik}$, $1 \le k \le \kappa$.

We cite next an important result from de Boor (2001, p. 149).

THEOREM 3.1 *There is an absolute constant $C_{p-1,\mu} > 0$ such that for every $\phi \in C^{p-1,\mu}[0,1]$ for some $\mu \in (0,1]$, there exists a function $g \in \mathcal{H}^{(p-1)}[0,1]$ for which $\|g-\phi\|_\infty \leq C_{p-1,\mu}\|\phi^{(p-1)}\|_{0,\mu} h_m^{\mu+p-1}$.*

The next three propositions concern $\tilde{m}_p(x)$, $\tilde{e}_p(x)$ and $\tilde{\xi}_p(x)$ given in Equation (8).

PROPOSITION 3.2 *Under Assumptions (A1) and (A3), as $n \to \infty$*

$$\sup_{x\in[0,1]} n^{1/2}|\tilde{m}_p(x) - m(x)| = o(1). \tag{10}$$

PROPOSITION 3.3 *Under Assumptions (A2)–(A4), as $n \to \infty$*

$$\sup_{x\in[0,1]} n^{1/2}|\tilde{e}_p(x)| = o_P(1). \tag{11}$$

PROPOSITION 3.4 *Under Assumptions (A2)–(A4), as $n \to \infty$*

$$\sup_{x\in[0,1]} n^{1/2}|\tilde{\xi}_p(x) - (\bar{m}(x) - m(x))| = o_P(1) \tag{12}$$

also for any $\alpha \in (0,1)$

$$P\left\{\sup_{x\in[0,1]} n^{1/2}|\tilde{\xi}_p(x)|G(x,x)^{-1/2} \leq Q_{1-\alpha}\right\} \to 1-\alpha. \tag{13}$$

Equations (10)–(12) yield the asymptotic efficiency of the spline estimator $\hat{m}_p$, i.e. $\sup_{x\in[0,1]} n^{1/2}|\bar{m}(x) - \hat{m}_p(x)| = o_P(1)$. The appendix contains proofs for the above three propositions, which together with Equation (8), imply Theorem 2.1.

4. Implementation

This section describes procedures to implement the confidence band in Corollary 2.3.

Given any data set $(j/N, Y_{ij})_{j=1,i=1}^{N,n}$ from model (1), the spline estimator $\hat{m}_p(x)$ is obtained from Equation (7), the number of interior knots in estimating $m(x)$ is taken to be $N_m = [cn^{1/(2p)}\log(n)]$, in which $[a]$ denotes the integer part of a. Our experiences show that the choice of constant $c = 0.2, 0.3, 0.5, 1, 2$ seems quite adequate, and that is what we recommend. When constructing the confidence bands, one needs to estimate the unknown functions $G(\cdot,\cdot)$ and the quantile $Q_{1-\alpha}$ and then plug in these estimators: the same approach is taken in Ma *et al.* (2011) and Wang and Yang (2009a).

The pilot estimator $\hat{G}_p(x,x')$ of the covariance function $G(x,x')$ is

$$\hat{G}_p = \underset{g(\cdot,\cdot)\in\mathcal{H}^{(p-2),2}}{\operatorname{argmin}} \sum_{j\neq j'}^{N} \left\{C_{\cdot jj'} - g\left(\frac{j}{N},\frac{j'}{N}\right)\right\}^2,$$

with $C_{\cdot jj'} = n^{-1}\sum_{i=1}^{n}\{Y_{ij} - \hat{m}_p(j/N)\}\{Y_{ij'} - \hat{m}_p(j'/N)\}$, $1 \leq j \neq j' \leq N$ and the tensor product spline space $\mathcal{H}^{(p-2),2} = \{\sum_{J,J'=1-p}^{N_G} b_{JJ'}B_{J,p}(t)B_{J',p}(s), b_{JJ'} \in \mathcal{R}, t,s \in [0,1]\}$ in which $N_G = [n^{1/(2p)}\log(\log(n))]$.

In order to estimate $Q_{1-\alpha}$, one first does the eigenfunction decomposition of $\hat{G}_p(x,x')$, i.e. $N^{-1}\sum_{j=1}^{N}\hat{G}_p(j/N,j'/N)\hat{\psi}_k(j/N) = \hat{\lambda}_k\hat{\psi}_k(j'/N)$, to obtain the estimated eigenvalues $\hat{\lambda}_k$ and eigenfunctions $\hat{\psi}_k$. Next, one chooses the number κ of eigenfunctions by using the following standard and efficient criterion, i.e. $\kappa = \operatorname{argmin}_{1\le l\le T}\{\sum_{k=1}^{l}\hat{\lambda}_k/\sum_{k=1}^{T}\hat{\lambda}_k > 0.95\}$, where $\{\lambda_k\}_{k=1}^{T}$ are the first T estimated positive eigenvalues. Finally, one simulates $\hat{\zeta}_b(x) = \hat{G}_p(x,x)^{-1/2}\sum_{k=1}^{\kappa} Z_{k,b}\hat{\phi}_k(x)$, where $\hat{\phi}_k = \sqrt{\hat{\lambda}_k}\hat{\psi}_k$, $Z_{k,b}$ are iid standard normal variables with $1 \le k \le \kappa$ and $b = 1,\dots,b_M$, where b_M is a preset large integer, the default of which is 1000. One takes the maximal absolute value for each copy of $\hat{\zeta}_b(x)$ and estimates $Q_{1-\alpha}$ by the empirical quantile $\hat{Q}_{1-\alpha}$ of these maximum values. One then uses the following confidence band:

$$\hat{m}_p(x) \pm n^{-1/2}\hat{G}_p(x,x)^{1/2}\hat{Q}_{1-\alpha}, \quad x \in [0,1], \tag{14}$$

for the mean function. One estimates $Q_{12,1-\alpha}$ analogous to $\hat{Q}_{1-\alpha}$ and computes

$$(\hat{m}_{1p} - \hat{m}_{2p})(x) \pm n_1^{-1/2}\hat{Q}_{12,1-\alpha}\{(\hat{G}_{1p} + \hat{r}\hat{G}_{2p})(x,x)\}^{1/2}, \tag{15}$$

as confidence band for $m_1(x) - m_2(x)$. Although beyond the scope of this paper, as one referee pointed out, the confidence band in Equation (14) is expected to enjoy the same asymptotic coverage as if true values of $Q_{1-\alpha}$ and $G(x,x)$ were used instead, due to the consistency of $\hat{G}_p(x,x)$ estimating $G(x,x)$. The same holds for the band in Equation (15).

5. Simulation

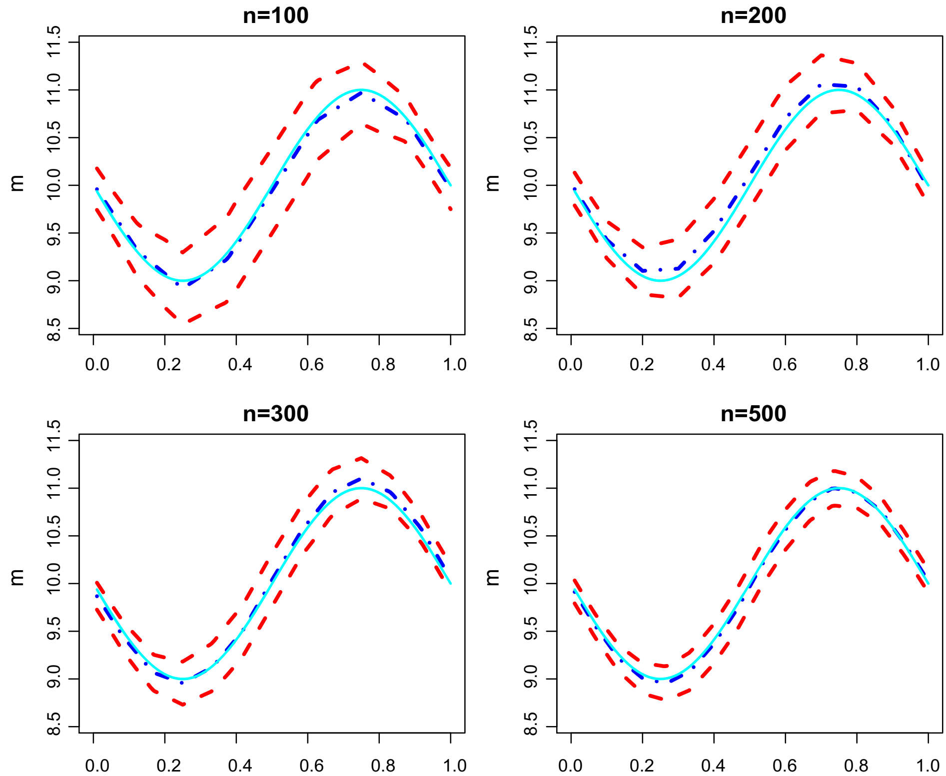
To demonstrate the practical performance of our theoretical results, we perform a set of simulation studies. Data are generated from model

$$Y_{ij} = m\left(\frac{j}{N}\right) + \sum_{k=1}^{2}\xi_{ik}\phi_k\left(\frac{j}{N}\right) + \sigma\varepsilon_{ij}, \quad 1 \le j \le N,\ 1 \le i \le n, \tag{16}$$

where $\xi_{ik} \sim N(0,1), k = 1,2$, $\varepsilon_{ij} \sim N(0,1)$, for $1 \le i \le n$, $1 \le j \le N$, $m(x) = 10 + \sin\{2\pi(x - 1/2)\}$, $\phi_1(x) = -2\cos\{\pi(x - 1/2)\}$ and $\phi_2(x) = \sin\{\pi(x - 1/2)\}$. This setting implies $\lambda_1 = 2$ and $\lambda_2 = 0.5$. The noise levels are set to be $\sigma = 0.5$ and 0.3. The number of subjects n is taken to be $60, 100, 200, 300$ and 500, and under each sample size, the number of observations per curve is assumed to be $N = [n^{0.25}\log^2(n)]$. This simulated process has a similar design as one of the simulation models in Yao *et al.* (2005), except that each subject is densely observed. We consider both linear and cubic spline estimators, and use confidence levels $1 - \alpha = 0.95$ and 0.99 for our simultaneous confidence bands. The constant c in the definition of N_m in Section 4 is taken to be 0.2, 0.3, 0.5, 1 and 2. Each simulation is repeated 500 times.

Figures 1 and 2 show the estimated mean functions and their 95% confidence bands for the true curve $m(\cdot)$ in model (16) with $\sigma = 0.3$ and $n = 100, 200, 300, 500$, respectively. As expected when n increases, the confidence band becomes narrower and the linear and cubic spline estimators are closer to the true curve.

Tables 1 and 2 show the empirical frequency that the true curve $m(\cdot)$ is covered by the linear and cubic spline confidence bands (14) at 100 points $\{1/100,\dots,99/100,1\}$, respectively. At all noise levels, the coverage percentages for the confidence band are close to the nominal confidence levels 0.95 and 0.99 for linear splines with $c = 0.5, 1$ (Table 1), and cubic splines with $c = 0.3, 0.5$ (Table 2) but decline slightly for $c = 2$ and markedly for $c = 0.2$. The coverage percentages thus depend on the choice of N_m, and the dependency becomes stronger when sample sizes decrease. For large sample sizes $n = 300$, 500, the effect of the choice of N_m on the coverage percentages is negligible. Although our theory indicates no optimal choice of c, we recommend using $c = 0.5$ for data analysis as its performance in simulation for both linear and cubic splines is either optimal or near optimal.



Colour online, B/W in print

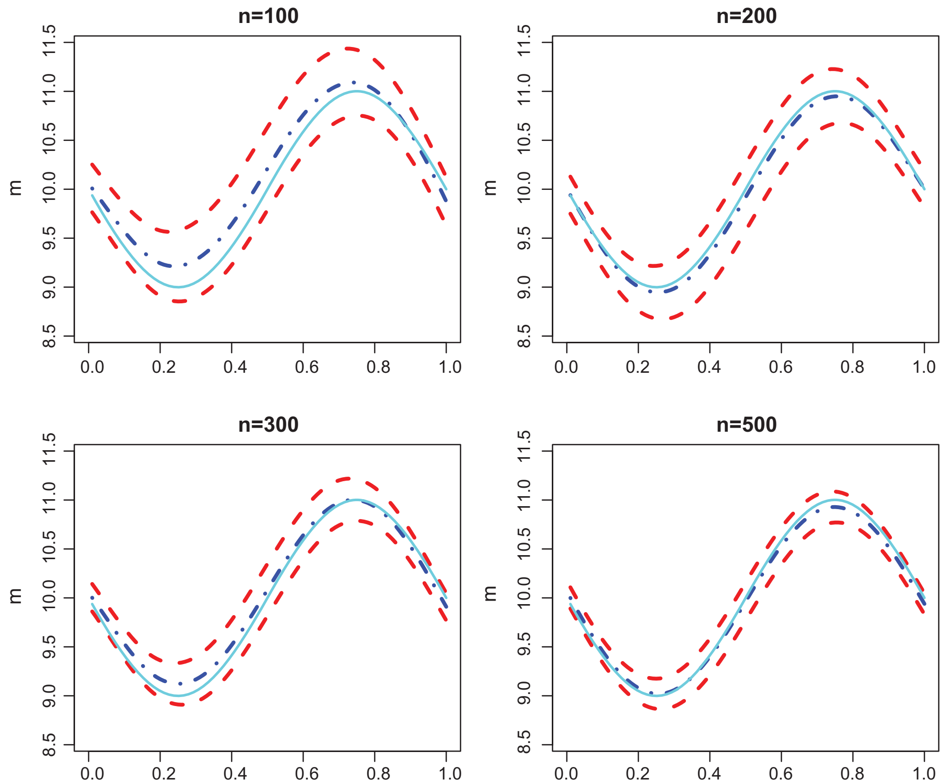
Figure 1. Plots of the linear spline estimator (2) for simulated data (dashed-dotted line) and 95% confidence bands (14) (upper and lower dashed lines) (14) for $m(x)$ (solid lines). In all panels, $\sigma = 0.3$.

Following the suggestion of one referee and the associate editor, we compare by simulation the proposed spline confidence band to the least-squares Bonferroni and least-squares bootstrap bands in Bunea *et al.* (2011) (BIW). Table 3 presents the empirical frequency that the true curve $m(\cdot)$ for model (16) is covered by these bands at $\{1/100, \ldots, 99/100, 1\}$, respectively, as Table 1. The coverage frequency of the BIW Bonferroni band is much higher than the nominal level, making it too conservative. The coverage frequency of the BIW bootstrap band is consistently lower than the nominal level by at least 10%, thus not recommended for practical use.

Q1

Following the suggestion of one referee and the associate editor, we also compare the widths of the three bands. For each replication, we calculate the ratios of widths of the two BIW bands against the spline band at $\{1/100, \ldots, 99/100, 1\}$ and then average these 100 ratios. Table 4 shows the five number summary of these 500 averaged ratios for $\sigma = 0.3$ and $p = 4$. The BIW Bonferroni band is much wider than cubic spline band, making it undesirable. While the BIW bootstrap band is narrower, we have mentioned previously that its coverage frequency is too low to be useful in practice. Simulation for other cases (e.g. $p = 2$, $\sigma = 0.5$) leads to the same conclusion.

To examine the performance of the two-sample test based on spline confidence band, Table 5 reports the empirical power and type-I error for the proposed two-sample test. The data were generated from Equation (16) with $\sigma = 0.5$ and $m_1(x) = 10 + \sin\{2\pi(x - 1/2)\} + \delta(x)$, $n = n_1$ for the first group, and $m_2(x) = 10 + \sin\{2\pi(x - 1/2)\}$, $n = n_2$ for the another group. The remaining parameters, ξ_{ik}, ε_{ij}, $\phi_1(x)$ and $\phi_2(x)$ were set to the same values for each group as in Equation (16). In order to mimic the real data in Section 6, we set $N = 50$, 100 and 200 when



Colour online, B/W in print

Figure 2. Plots of the cubic spline estimator (2) for simulated data (dashed-dotted line) and 95% confidence bands (14) (upper and lower dashed lines) (14) for $m(x)$ (solid lines). In all panels, $\sigma = 0.3$.

Table 1. Coverage frequencies from 500 replications using linear spline (14) with $p = 2$ and $N_m = [cn^{1/(2p)} \log(n)]$.

		Coverage frequency					Coverage frequency				
		$\sigma = 0.5$					$\sigma = 0.3$				
n	$1-\alpha$	$c = 0.2$	$c = 0.3$	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.2$	$c = 0.3$	$c = 0.5$	$c = 1$	$c = 2$
60	0.950	0.384	0.790	0.876	0.894	0.852	0.410	0.786	0.930	0.914	0.884
	0.990	0.692	0.938	0.970	0.976	0.942	0.702	0.950	0.972	0.966	0.954
100	0.950	0.184	0.826	0.886	0.884	0.838	0.198	0.822	0.916	0.916	0.896
	0.990	0.476	0.936	0.964	0.966	0.944	0.496	0.940	0.974	0.974	0.968
200	0.950	0.418	0.856	0.914	0.922	0.862	0.414	0.862	0.946	0.942	0.926
	0.990	0.712	0.966	0.976	0.990	0.972	0.720	0.966	0.984	0.984	0.980
300	0.950	0.600	0.888	0.920	0.932	0.874	0.602	0.896	0.940	0.934	0.926
	0.990	0.834	0.978	0.976	0.980	0.972	0.840	0.982	0.984	0.986	0.980
500	0.950	0.772	0.880	0.922	0.886	0.894	0.768	0.888	0.954	0.950	0.942
	0.990	0.902	0.964	0.984	0.976	0.976	0.906	0.968	0.992	0.994	0.988

$n_1 = 160$, 80 and 40 and $n_2 = 320$, 160 and 80 accordingly. The studied hypotheses are:

$$H_0 : m_1(x) = m_2(x),\ \forall x \in [0,1] \longleftrightarrow H_a : m_1(x) \neq m_2(x), \quad \text{for some } x \in [0,1].$$

Table 5 shows the empirical frequencies of rejecting H_0 in this simulation study with nominal test level equal to 0.05 and 0.01. If $\delta(x) \neq 0$, these empirical powers should be close to 1, and for $\delta(x) \equiv 0$, the nominal levels. Each set of simulations consists of 500 Monte Carlo runs. Asymptotic

Table 2. Coverage frequencies from 500 replications using cubic spline (14) with $p = 4$ and $N_m = [cn^{1/(2p)} \log(n)]$.

		Coverage frequency $\sigma = 0.5$					Coverage frequency $\sigma = 0.3$				
n	$1-\alpha$	$c = 0.2$	$c = 0.3$	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.2$	$c = 0.3$	$c = 0.5$	$c = 1$	$c = 2$
60	0.950	0.644	0.916	0.902	0.890	0.738	0.672	0.922	0.940	0.940	0.916
	0.990	0.866	0.980	0.958	0.964	0.888	0.884	0.986	0.986	0.984	0.982
100	0.950	0.596	0.902	0.904	0.876	0.846	0.610	0.916	0.914	0.914	0.896
	0.990	0.786	0.970	0.968	0.956	0.952	0.798	0.980	0.974	0.970	0.964
200	0.950	0.928	0.942	0.932	0.936	0.904	0.938	0.952	0.950	0.948	0.934
	0.990	0.978	0.992	0.982	0.992	0.978	0.982	0.984	0.992	0.982	0.984
300	0.950	0.920	0.948	0.926	0.948	0.898	0.922	0.956	0.948	0.942	0.938
	0.990	0.976	0.986	0.986	0.988	0.980	0.982	0.984	0.988	0.984	0.982
500	0.950	0.928	0.922	0.954	0.902	0.898	0.928	0.928	0.936	0.932	0.916
	0.990	0.980	0.982	0.990	0.976	0.978	0.980	0.982	0.990	0.990	0.992

Table 3. Coverage frequencies from 500 replications using least-squares Bonferroni band and least-squares Bootstrap band.

		Coverage frequency: Least-squares Bonferroni		Least-squares bootstrap	
n	$1-\alpha$	$\sigma = 0.5$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.3$
60	0.950	0.990	0.988	0.742	0.744
	0.990	0.994	0.994	0.856	0.864
100	0.950	0.996	0.998	0.678	0.712
	0.990	0.998	1.000	0.860	0.870
200	0.950	0.988	0.992	0.710	0.734
	0.990	1.000	1.000	0.856	0.888
300	0.950	0.988	0.998	0.704	0.720
	0.990	1.000	1.000	0.868	0.870
500	0.950	0.996	0.998	0.718	0.732
	0.990	1.000	1.000	0.856	0.860

Table 4. Five number summary of ratios of confidence band widths.

		Least-squares Bonferroni/cubic spline					Least-squares bootstrap/cubic spline				
n	$1-\alpha$	Min.	$Q1$	Med.	$Q3$	Max.	Min.	$Q1$	Med.	$Q3$	Max.
60	0.950	0.964	1.219	1.299	1.397	1.845	0.522	0.667	0.716	0.770	0.967
	0.990	0.907	1.114	1.188	1.285	1.730	0.527	0.662	0.715	0.770	1.048
100	0.950	0.995	1.263	1.331	1.415	1.684	0.565	0.675	0.714	0.754	0.888
	0.990	0.910	1.148	1.219	1.295	1.603	0.536	0.665	0.708	0.752	0.925
200	0.950	1.169	1.326	1.383	1.433	1.653	0.600	0.683	0.715	0.743	0.855
	0.990	1.045	1.197	1.250	1.300	1.507	0.557	0.668	0.702	0.740	0.888
300	0.950	1.169	1.363	1.412	1.462	1.663	0.574	0.690	0.717	0.742	0.838
	0.990	1.067	1.228	1.277	1.322	1.509	0.587	0.676	0.707	0.739	0.850
500	0.950	1.273	1.395	1.432	1.476	1.601	0.620	0.691	0.714	0.737	0.818
	0.990	1.132	1.243	1.288	1.334	1.465	0.607	0.674	0.707	0.734	0.839

standard errors (as the number of Monte Carlo iterations tends to infinity) are reported in the last row of the table. Results are listed only for cubic spline confidence bands, as those of the linear spline are similar. Overall, the two-sample test performs well, even with a rather small difference ($\delta(x) = 0.7\sin(x)$), providing a reasonable empirical power. Moreover, the differences between nominal levels and empirical type-I error do diminish as the sample size increases.

Table 5. Empirical power and type-I error of two-sample test using cubic spline.

$\delta(x)$	$n_1 = 160, n_2 = 320$		$n_1 = 80, n_2 = 160$		$n_1 = 40, n_2 = 80$	
	Nominal test level		Nominal test level		Nominal test level	
	0.05	0.01	0.05	0.01	0.05	0.01
$0.6t$	1.000	1.000	0.980	0.918	0.794	0.574
$0.7\sin(x)$	1.000	1.000	0.978	0.910	0.788	0.566
0	0.058	0.010	0.068	0.010	0.096	0.028
Monte Carlo SE	0.001	0.004	0.001	0.004	0.001	0.004

6. Empirical example

In this section, we revisit the Tecator data mentioned in Section 1, which can be downloaded at http://lib.stat.cmu.edu/datasets/tecator. In this data set, there are measurements on $n = 240$ meat samples, where for each sample a $N = 100$ channel near-infrared spectrum of absorbance measurements was recorded, and contents of moisture (water), fat and protein were also obtained. The feed analyser worked in the wavelength range from 850 to 1050 nm. Figure 3 shows the scatter plot of this data set. The spectral data can be naturally considered as functional data, and we will perform a two-sample test to see whether absorbance from the spectrum differs significantly due to difference in fat content.

This data set has been used for comparing four classification methods (Li and Yu 2008), building a regression model to predict the fat content from the spectrum (Li and Hsing 2010b). Following Li and Yu (2008), we separate samples according to their fat contents being less than 20% or not. Figure 3 (right) shows 10 samples from each group. Here, hypothesis of interest is

$$H_0 : m_1(x) = m_2(x), \quad \forall x \in [850, 1050] \longleftrightarrow H_a : m_1(x) \neq m_2(x), \text{ for some } x \in [850, 1050],$$

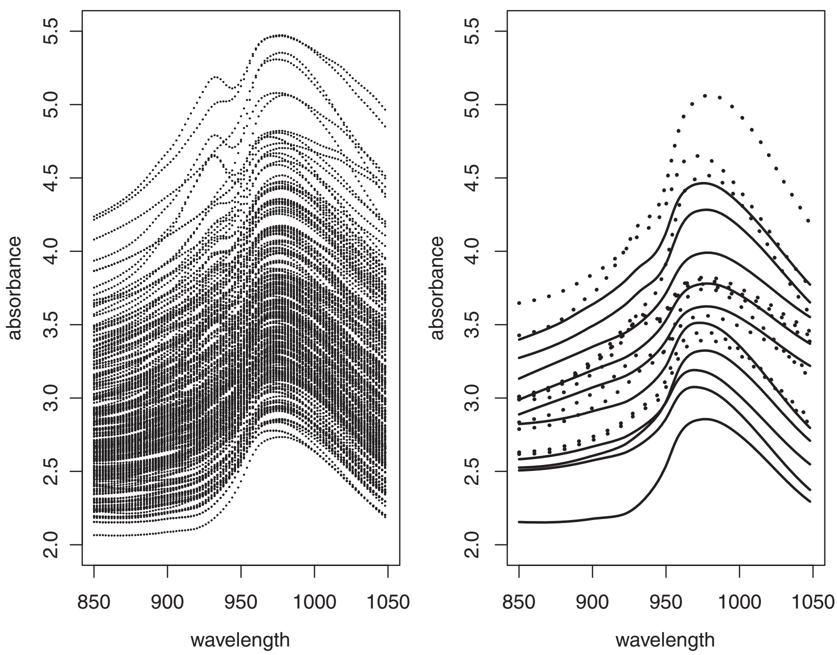
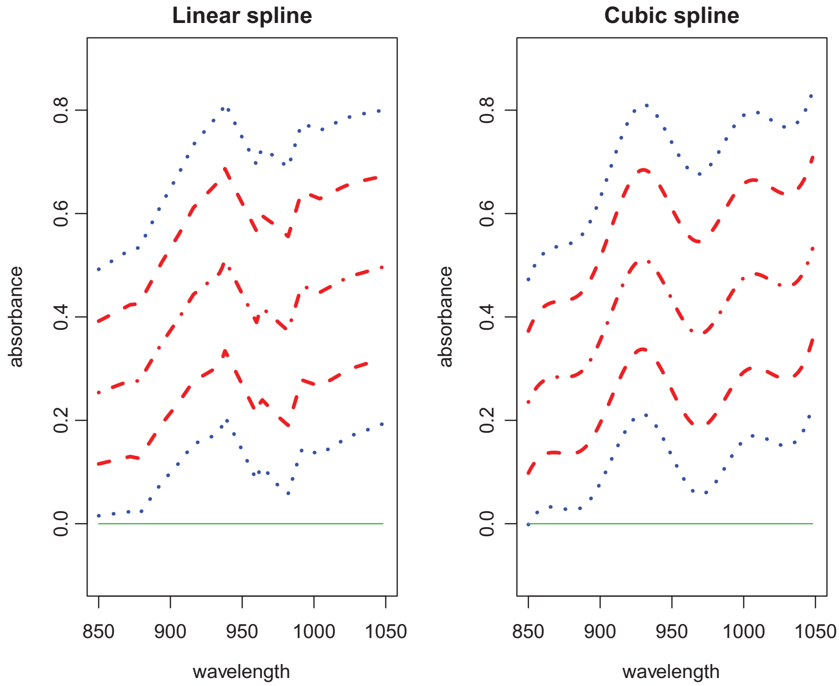


Figure 3. Left: Plot of Tecator data. Right: Sample curves for the Tecator data. Each class has 10 sample curves. Dashed lines represent spectra with fact >20% and solid lines represent spectra with fact <20%.



Colour online, B/W in print

Figure 4. Plots of the fitted linear and cubic spline regressions of $m_1(x) - m_2(x)$ for the Tecator data (dashed-dotted line), 99% confidence bands (15) (upper and lower dashed lines), 99.9995% confidence bands (15) (upper and lower dotted lines) and the zero line (solid line).

where $m_1(x)$ and $m_2(x)$ are the regression functions of absorbance on spectrum, for samples with fat content less than 20% and great than or equal to 20%, respectively. Among 240 samples, there are $n_1 = 155$ with fat content less than 20%, the rest $n_2 = 85$ no less than 20%. The numbers of interior knots in Equation (2) are computed as in Section 3 with $c = 0.5$ and are $N_{1m} = 4$ and $N_{2m} = 3$ for cubic spline fit and $N_{1m} = 8$ and $N_{2m} = 6$ for linear spline fit. Figure 4 depicts the linear and cubic spline confidence bands according to Equation (15) at confidence levels 0.99 (upper and lower dashed lines) and 0.999995 (upper and lower dotted lines), with the centre dashed-dotted line representing the spline estimator $\hat{m}_1(x) - \hat{m}_2(x)$ and a solid line representing zero. Since even the 99.9995% confidence band does not contain the zero line entirely, the difference of low fat and high fat populations' absorbance was extremely significant. In fact, Figure 4 clearly indicates that the less the fat contained, the higher the absorbance is.

Acknowledgements

This work has been supported in part by NSF awards DMS 0706518, 1007594, NCI/NIH K-award, 1K01 CA131259, a Dissertation Continuation Fellowship from Michigan State University, and funding from the Jiangsu Specially-Appointed Professor Programme, Jiangsu Province, China. The helpful comments by two referees and the Associate Editor have led to significant improvement of the paper.

References

Benko, M., Härdle, W., and Kneip, A. (2009), 'Common Functional Principal Components', *The Annals of Statistics*, 37, 1–34.

Bunea, F., Ivanescu, A.E., and Wegkamp, M. (2011), 'Adaptive Inference for the Mean of a Gaussian Process in Functional Data', *Journal of the Royal Statistical Society. Series B*, Forthcoming. Q2

Cardot, H. (2000), 'Nonparametric Estimation of Smoothed Principal Components Analysis of Sampled Noisy Functions', *Journal of Nonparametric Statistics*, 12, 503–538.

Csőrgő, M., and Révész, P. (1981), *Strong Approximations in Probability and Statistics*, New York: Academic Press.

Cuevas, A., Febrero, M., and Fraiman, R. (2006), 'On the Use of the Bootstrap for Estimating Functions with Functional Data', *Computational Statistics and Data Analysis*, 51, 1063–1074.

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

Degras, D.A. (2011), 'Simultaneous Confidence Bands for Nonparametric Regression with Functional Data', *Statistica Sinica*, 21, 1735–1765.

DeVore, R., and Lorentz, G. (1993), *Constructive Approximation: Polynomials and Splines Approximation*, Berlin: Springer-Verlag.

Fan, J., and Lin, S.-K. (1998), 'Tests of Significance When Data are Curves', *Journal of the American Statistical Association*, 93, 1007–1021.

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

Hall, P., Müller, H.G., and Wang, J.L. (2006), 'Properties of Principal Component Methods for Functional and Longitudinal Data Analysis', *The Annals of Statistics*, 34, 1493–1517.

Huang, J., and Yang, L. (2004), 'Identification of Nonlinear Additive Autoregressive Models', *Journal of the Royal Statistical Society. Series B*, 66, 463–477.

Li, Y., and Hsing, T. (2010a), 'Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data', *The Annals of Statistics*, 38, 3321–3351.

Li, Y., and Hsing, T. (2010b), 'Deciding the Dimension of Effective Dimension Reduction Space for Functional and High-Dimensional Data', *The Annals of Statistics*, 38, 3028–3062.

Li, B., and Yu, Q. (2008), 'Classification of Functional Data: A Segmentation Approach', *Computational Statistics and Data Analysis*, 52, 4790–4800.

Liu, R., and Yang, L. (2010), 'Spline-Backfitted Kernel Smoothing of Additive Coefficient Model', *Econometric Theory*, 26, 29–59.

Ma, S., and Yang, L. (2011), 'Spline-Backfitted Kernel Smoothing of Partially Linear Additive Model', *Journal of Statistical Planning and Inference*, 141, 204–219.

Ma, S., Yang, L., and Carroll, R.J. (2011), 'A Simultaneous Confidence Band for Sparse Longitudinal Data', *Statistica Sinica*, in press. **Q2**

Ramsay, J.O., and Silverman, B.W. (2005), *Functional Data Analysis* (2nd ed.), Springer Series in Statistics, New York: Springer.

Rice, J.A., and Wu, C.O. (2001), 'Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves', *Biometrics*, 57, 253–259.

Wang, J., and Yang, L. (2009a), 'Polynomial Spline Confidence Bands for Regression Curves', *Statistica Sinica*, 19, 325–342.

Wang, L., and Yang, L. (2009b), 'Spline Estimation of Single Index Model', *Statistica Sinica*, 19, 765–783.

Xue, L., and Yang, L. (2006), 'Additive Coefficient Modelling via Polynomial Spline', *Statistica Sinica*, 16, 1423–1446.

Yao, F., Müller, H.G., and Wang, J.L. (2005), 'Functional Data Analysis for Sparse Longitudinal Data', *Journal of the American Statistical Association*, 100, 577–590.

Zhao, Z., and Wu, W. (2008), 'Confidence Bands in Nonparametric Time Series Regression', *The Annals of Statistics*, 36, 1854–1878.

Zhou, Z., and Wu, W. (2010), 'Simultaneous Inference of Linear Models with Time Varying Coefficients', *Journal of the Royal Statistical Society. Series B*, 72, 513–531.

Zhou, S., Shen, X., and Wolfe, D.A. (1998), 'Local Asymptotics of Regression Splines and Confidence Regions', *The Annals of Statistics*, 26, 1760–1782.

Appendix

In this appendix, we use C to denote a generic positive constant unless otherwise stated.

A.1. Preliminaries

For any vector $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_s) \in R^s$, denote the norm $\|\boldsymbol{\zeta}\|_r = (|\zeta_1|^r + \cdots + |\zeta_s|^r)^{1/r}$, $1 \le r < +\infty$, $\|\boldsymbol{\zeta}\|_\infty = \max(|\zeta_1|, \dots, |\zeta_s|)$. For any $s \times s$ symmetric matrix $\mathbf{A}$, we define $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ as its smallest and largest eigenvalues, and its L_r norm as $\|\mathbf{A}\|_r = \max_{\zeta \in R^s, \zeta \neq \mathbf{0}} \|\mathbf{A}\zeta\|_r \|\boldsymbol{\zeta}\|_r^{-1}$. In particular, $\|\mathbf{A}\|_2 = \lambda_{\max}(\mathbf{A})$, and if $\mathbf{A}$ is also nonsingular, $\|\mathbf{A}^{-1}\|_2 = \lambda_{\min}^{-1}(\mathbf{A})$.

For functions $\phi, \varphi \in L_2[0,1]$, one denotes the theoretical and empirical inner products as $\langle \phi, \varphi \rangle = \int_0^1 \phi(x)\varphi(x)\,\mathrm{d}x$ and $\langle \phi, \varphi \rangle_{2,N} = N^{-1} \sum_{j=1}^{N} \phi(j/N)\varphi(j/N)$. The corresponding norms are $\|\phi\|_2^2 = \langle \phi, \phi \rangle$, $\|\phi\|_{2,N}^2 = \langle \phi, \phi \rangle_{2,N}$.

We state a strong approximation result, which is used in the proof of Lemma A.6.

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

$$P\left\{\max_{1\le l\le n}|S_l - W_n(l)| > x_n\right\} \le C_2 n\{H(ax_n)\}^{-1}.$$

The next lemma is a special case of Theorem 13.4.3, p. 404, of DeVore and Lorentz (1993). Let p be a positive integer, a matrix $\mathbf{A} = (a_{ij})$ is said to have bandwidth p if $a_{ij} = 0$ when $|i-j| \ge p$, and p is the smallest integer with this property.

Lemma A.2 *If a matrix* $\mathbf{A}$ *with bandwidth p has an inverse* $\mathbf{A}^{-1}$ *and* $d = \|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2$ *is the condition number of* $\mathbf{A}$, *then* $\|\mathbf{A}^{-1}\|_\infty \le 2c_0(1-\eta)^{-1}$, *with* $c_0 = \nu^{-2p}\|\mathbf{A}^{-1}\|_2$, $\eta = ((d^2-1)/(d^2+1))^{1/(4p)}$.

One writes $\mathbf{X}^{\mathrm{T}}\mathbf{X} = N\hat{\mathbf{V}}_p$, $\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \{\sum_{j=1}^{N} B_{J,p}(j/N)\bar{Y}_{\cdot j}\}_{J=1-p}^{N_m}$, where the theoretical and empirical inner product matrices of $\{B_{J,p}(x)\}_{J=1-p}^{N_m}$ are denoted as

$$\mathbf{V}_p = (\langle B_{J,p}, B_{J',p}\rangle)_{J,J'=1-p}^{N_m}, \hat{\mathbf{V}}_p = (\langle B_{J,p}, B_{J',p}\rangle_{2,N})_{J,J'=1-p}^{N_m}. \tag{A1}$$

We establish next that the theoretical inner product matrix $\mathbf{V}_p$ defined in Equation (A1) has an inverse with bounded L_∞ norm.

Lemma A.3 *For any positive integer p, there exists a constant* $M_p > 0$ *depending only on p, such that* $\|\mathbf{V}_p^{-1}\|_\infty \le M_p h_m^{-1}$, *where* $h_m = (N_m+1)^{-1}$.

Proof According to Lemma A.1 in Wang and Yang (2009b), $\mathbf{V}_p$ is invertible since it is a symmetric matrix with all eigenvalues positive, i.e. $0 < c_p N_m^{-1} \le \lambda_{\min}(\mathbf{V}_p) \le \lambda_{\max}(\mathbf{V}_p) \le C_p N_m^{-1} < \infty$, where c_p and C_p are positive real numbers. The compact support of B-spline basis makes $\mathbf{V}_p$ of bandwidth p; hence, one can apply Lemma A.2. Since $d_p = \lambda_{\max}(\mathbf{V}_p)/\lambda_{\min}(\mathbf{V}_p) \le C_p/c_p$; hence

$$\eta_p = (d_p^2-1)^{1/4p}(d_p^2+1)^{-1/4p} \le (C_p^2c_p^{-2}-1)^{1/4p}(C_p^2c_p^{-2}+1)^{-1/4p} < 1.$$

If $p=1$, then $\mathbf{V}_p^{-1} = h_m^{-1}\mathbf{I}_{N_m+p}$, the lemma holds with $M_p = 1$. If $p > 1$, let $\mathbf{u}_{1-p} = (1, \mathbf{0}_{N_m+p-1}^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{u}_0 = (\mathbf{0}_{p-1}^{\mathrm{T}}, 1, \mathbf{0}_{N_m}^{\mathrm{T}})^{\mathrm{T}}$, then $\|\mathbf{u}_{1-p}\|_2 = \|\mathbf{u}_0\|_2 = 1$. Also Lemma A.1 in Wang and Yang (2009b) implies that

$$\lambda_{\min}(\mathbf{V}_p) = \lambda_{\min}(\mathbf{V}_p)\|\mathbf{u}_{1-p}\|_2^2 \le \mathbf{u}_{1-p}^{\mathrm{T}}\mathbf{V}_p\mathbf{u}_{1-p} = \|B_{1-p,p}\|_2^2,$$
$$\mathbf{u}_0^{\mathrm{T}}\mathbf{V}_p\mathbf{u}_0 = \|B_{0,p}\|_2^2 \le \lambda_{\max}(\mathbf{V}_p)\|\mathbf{u}_0\|_2^2 = \lambda_{\max}(\mathbf{V}_p),$$

hence $d_p = \lambda_{\max}(\mathbf{V}_p)/\lambda_{\min}(\mathbf{V}_p) \ge \|B_{0,p}\|_2^2\|B_{1-p,p}\|_2^{-2} = r_p > 1$, where r_p is an absolute constant depending only on p. Thus, $\eta_p = (d_p^2-1)^{1/(4p)}(d_p^2+1)^{-1/(4p)} \ge (r_p^2-1)^{1/(4p)}(r_p^2+1)^{-1/(4p)} > 0$. Applying Lemma A.2 and putting the above bounds together, one obtains

$$\begin{aligned}\|\mathbf{V}_p^{-1}\|_\infty h_m &\le 2\eta_p^{-2p}\|\mathbf{V}_p^{-1}\|_2(1-\eta_p)^{-1}h_m \\ &\le 2\left(\frac{r_p^2+1}{r_p^2-1}\right)^{1/2}\lambda_{\min}^{-1}(\mathbf{V}_p) \times \left\{1 - \left(\frac{C_p^2c_p^{-2}-1}{C_p^2c_p^{-2}+1}\right)^{1/4p}\right\}^{-1} h_m \\ &\le 2\left(\frac{r_p^2+1}{r_p^2-1}\right)^{1/2} c_p^{-1}\left\{1 - \left(\frac{C_p^2c_p^{-2}-1}{C_p^2c_p^{-2}+1}\right)^{1/4p}\right\}^{-1} \equiv M_p.\end{aligned}$$

The lemma is proved. ∎

For any function $\phi \in C[0,1]$, denote the vector $\boldsymbol{\phi} = (\phi(1/N), \ldots, \phi(N/N))^{\mathrm{T}}$ and function

$$\tilde{\phi}(x) \equiv \{B_{1-p,p}(x), \ldots, B_{N_m,p}(x)\}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\phi}.$$

Lemma A.4 *Under Assumption* (A3), *for* $\mathbf{V}_p$ *and* $\hat{\mathbf{V}}_p$ *defined in Equation* (A1), $\|\mathbf{V}_p - \hat{\mathbf{V}}_p\|_\infty = O(N^{-1})$ *and* $\|\hat{\mathbf{V}}_p^{-1}\|_\infty \le 2h_m^{-1}$. *There exists* $c_{\phi,p} \in (0,\infty)$ *such that when n is large enough,* $\|\tilde{\phi}\|_\infty \le c_{\phi,p}\|\phi\|_\infty$ *for any* $\phi \in C[0,1]$. *Furthermore, if* $\phi \in C^{p-1,\mu}[0,1]$ *for some* $\mu \in (0,1]$, *then for* $\tilde{C}_{p-1,\mu} = (c_{\phi,p}+1)C_{p-1,\mu}$

$$\|\tilde{\phi} - \phi\|_\infty \le \tilde{C}_{p-1,\mu}\|\phi^{(p-1)}\|_{0,\mu}h_m^{\mu+p-1}. \tag{A2}$$

Proof We first show that $\|\mathbf{V}_p - \hat{\mathbf{V}}_p\|_\infty = O(N^{-1})$. In the case of $p = 1$, define for any $0 \le J \le N_m$, the number of design points j/N in the Jth interval I_J as N_J, then

$$N_J = \begin{cases} \#\left\{j : j \in \left[\dfrac{NJ}{(N_m+1)}, \dfrac{N(J+1)}{(N_m+1)}\right)\right\}, & 0 \le J < N_m, \\ \#\left\{j : j \in \left[\dfrac{NJ}{(N_m+1)}, \dfrac{N(J+1)}{(N_m+1)}\right]\right\}, & J = N_m. \end{cases}$$

Clearly, $\max_{0\le J\le N_m} |N_J - Nh_m| \le 1$ and hence

$$\|\mathbf{V}_1 - \hat{\mathbf{V}}_1\|_\infty = \max_{0\le J\le N_m} |\|B_{J,1}\|_{2,N}^2 - \|B_{J,1}\|_2^2| = \max_{0\le J\le N_m} \left| N^{-1}\sum_{j=1}^N B_{J,1}^2\left(\frac{j}{N}\right) - h_m \right|$$

$$= \max_{0\le J\le N_m} |N^{-1}N_J - h_m| = N^{-1} \max_{0\le J\le N_m} |N_J - Nh_m| \le N^{-1}.$$

For $p > 1$, de Boor (2001, p. 96) B-spline property ensures that there exists a constant $C_{1,p} > 0$ such that

$$\max_{1-p\le J,J'\le N_m} \max_{1\le j\le N} \sup_{x\in[(j-1)/N, j/N]} \left| B_{J,p}\left(\frac{j}{N}\right) B_{J',p}\left(\frac{j}{N}\right) - B_{J,p}(x)B_{J',p}(x) \right| \le C_{1,p}N^{-1}h_m^{-1},$$

while there exists a constant $C_{2,p} > 0$ such that $\max_{1-p\le J,J'\le N_m} N_{J,J'} \le C_{2,p}Nh_m$, where $N_{J,J'} = \#\{j : 1 \le j \le N, B_{J,p}(j/N)B_{J',p}(j/N) > 0\}$. Hence,

$$\|\mathbf{V}_p - \hat{\mathbf{V}}_p\|_\infty = \max_{1-p\le J,J'\le N_m} \left| N^{-1}\sum_{j=1}^N B_{J,p}\left(\frac{j}{N}\right) B_{J',p}\left(\frac{j}{N}\right) - \int_0^1 B_{J,p}(x)B_{J',p}(x)\,\mathrm{d}x \right|$$

$$\le \max_{1-p\le J,J'\le N_m} \sum_{j=1}^N \int_{(j-1)/N}^{j/N} \left| B_{J,p}\left(\frac{j}{N}\right) B_{J',p}\left(\frac{j}{N}\right) - B_{J,p}(x)B_{J',p}(x) \right| \mathrm{d}x$$

$$\le C_{2,p}Nh_m \times N^{-1} \times C_{1,p}N^{-1}h_m^{-1} \le CN^{-1}.$$

According to Lemma A.3, for any $(N_m + p)$ vector $\boldsymbol{\gamma}$, $\|\mathbf{V}_p^{-1}\boldsymbol{\gamma}\|_\infty \le h_m^{-1}\|\boldsymbol{\gamma}\|_\infty$. Hence, $\|\mathbf{V}_p\boldsymbol{\gamma}\|_\infty \ge h_m\|\boldsymbol{\gamma}\|_\infty$. By Assumption (A3), $N^{-1} = o(h_m)$ so if n is large enough, for any $\boldsymbol{\gamma}$, one has

$$\|\hat{\mathbf{V}}_p\boldsymbol{\gamma}\|_\infty \ge \|\mathbf{V}_p\boldsymbol{\gamma}\|_\infty - \|\mathbf{V}_p\boldsymbol{\gamma} - \hat{\mathbf{V}}_p\boldsymbol{\gamma}\|_\infty \ge h_m\|\boldsymbol{\gamma}\|_\infty - O(N^{-1})\|\boldsymbol{\gamma}\|_\infty = \frac{h_m}{2}\|\boldsymbol{\gamma}\|_\infty.$$

Hence, $\|\hat{\mathbf{V}}_p^{-1}\|_\infty \le 2h_m^{-1}$.

To prove the last statement of the lemma, note that for any $x \in [0, 1]$ at most $(p + 1)$ of the numbers $B_{1-p,p}(x), \ldots, B_{N_m,p}(x)$ are between 0 and 1, others being 0, so

$$|\tilde{\phi}(x)| \le (p+1)|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\phi}| = (p+1)|\hat{\mathbf{V}}_p^{-1}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\phi}N^{-1})|$$

$$\le (p+1)\|\hat{\mathbf{V}}_p^{-1}\|_\infty|\mathbf{X}^{\mathrm{T}}\boldsymbol{\phi}N^{-1}| \le 2(p+1)h_m^{-1}|\mathbf{X}^{\mathrm{T}}\mathbf{I}_N N^{-1}|\|\phi\|_\infty,$$

in which $\mathbf{I}_N = (1, \ldots, 1)^{\mathrm{T}}$. Clearly, $|\mathbf{X}^{\mathrm{T}}\mathbf{I}_N N^{-1}| \le Ch_m$ for some $C > 0$, hence $|\tilde{\phi}(x)| \le 2(p+1)C\|\phi\|_\infty = c_{\phi,p}\|\phi\|_\infty$. Now if $\phi \in C^{p-1,\mu}[0, 1]$ for some $\mu \in (0, 1]$, let $g \in \mathcal{H}^{(p-1)}[0, 1]$ be such that $\|g - \phi\|_\infty \le C_{p-1,\mu}\|\phi^{(p-1)}\|_{0,\mu}h_m^{\mu+p-1}$ according to Theorem 3.1, then $\tilde{g} \equiv g$ as $g \in \mathcal{H}^{(p-1)}[0, 1]$; hence,

$$\|\tilde{\phi} - \phi\|_\infty = \|\tilde{\phi} - \tilde{g} - (\phi - g)\|_\infty \le \|\tilde{\phi} - \tilde{g}\|_\infty + \|\phi - g\|_\infty$$

$$\le (c_{\phi,p} + 1)\|\phi - g\|_\infty \le (c_{\phi,p} + 1)C_{p-1,\mu}\|\phi^{(p-1)}\|_{0,\mu}h_m^{\mu+p-1},$$

proving Equation (A2). ∎

LEMMA A.5 *Under Assumption (A5), for $C_0 = C_1(1 + \beta C_2 \sum_{s=1}^\infty s^{\beta-1-\gamma_1})$ and $n \ge 1$*

$$\max_{1\le k\le\kappa} E|\bar{\xi}_{\cdot,k} - \bar{Z}_{\cdot k,\xi}| \le C_0 n^{\beta-1}, \tag{A3}$$

$$\max_{1\le j\le N} |\bar{\varepsilon}_{\cdot j} - \bar{Z}_{\cdot j,\varepsilon}| = O_{\mathrm{a.s.}}(n^{\beta-1}), \tag{A4}$$

where $\bar{Z}_{\cdot k,\xi} = n^{-1}\sum_{i=1}^n Z_{ik,\xi}$, $\bar{Z}_{\cdot j,\varepsilon} = n^{-1}\sum_{i=1}^n Z_{ij,\varepsilon}$, $1 \le j \le N$, $1 \le k \le \kappa$. Also

$$\max_{1\le k\le\kappa} E|\bar{\xi}_{\cdot,k}| \le n^{-1/2}\left(\frac{2}{\pi}\right)^{1/2} + C_0 n^{\beta-1}. \tag{A5}$$

Proof The proof of Equation (A4) is trivial. Assumption (A5) entails that $\bar{F}_{n+t,k} < C_2(n+t)^{-\gamma_1}$, $k = 1, \ldots, \kappa$, $t = 0, 1, \ldots, \infty$, in which $\bar{F}_{n+t,k} = P[|\sum_{i=1}^{n} \xi_{ik} - \sum_{i=1}^{n} Z_{ik,\xi}| > C_1(n+t)^{\beta}]$. Taking expectation, one has

$$
\begin{aligned}
E\left|\sum_{i=1}^{n} \xi_{ik} - \sum_{i=1}^{n} Z_{ik,\xi}\right| &\leq C_1(n+0)^{\beta} + \sum_{t=1}^{\infty} C_1(n+t)^{\beta}(\bar{F}_{n+t-1,k} - \bar{F}_{n+t,k}) \\
&\leq C_1 n^{\beta} + \sum_{t=0}^{\infty} C_1 C_2 (n+t)^{-\gamma_1} \beta (n+t)^{\beta-1} \leq C_1 \left\{ n^{\beta} + \beta C_2 \sum_{t=0}^{\infty} (n+t)^{\beta-1-\gamma_1} \right\} \\
&\leq n^{\beta} C_1 \left[1 + \beta C_2 n^{-1-\gamma_1} \sum_{s=1}^{\infty} \sum_{t=sn-n}^{sn-1} \left(1 + \frac{t}{n}\right)^{\beta-1-\gamma_1} \right] \\
&\leq n^{\beta} C_1 \left[1 + \beta C_2 n^{-1-\gamma_1} \times n \sum_{t=1}^{\infty} t^{\beta-1-\gamma_1} \right] \leq C_0 n^{\beta},
\end{aligned}
$$

which proves Equation (A3) if one divides the above inequalities by n. The fact that $\bar{Z}_{.k,\xi} \sim N(0, 1/n)$ entails that $E|\bar{Z}_{.k,\xi}| = n^{-1/2}(2/\pi)^{1/2}$ and thus $\max_{1\leq k\leq\kappa} E|\bar{\xi}_{\cdot,k}| \leq n^{-1/2}(2/\pi)^{1/2} + C_0 n^{\beta-1}$. ■

LEMMA A.6 *Assumption* (A5) *holds under Assumption* (A5′).

Proof Under Assumption (A5'), $E|\xi_{ik}|^{\eta_1} < +\infty$, $\eta_1 > 4$, $E|\varepsilon_{ij}|^{\eta_2} < +\infty$, $\eta_2 > 4 + 2\theta$, so there exists some $\beta \in (0, 1/2)$ such that $\eta_1 > 2/\beta$, $\eta_2 > (2+\theta)/\beta$.

Now, let $H(x) = x^{\eta_1}$, then Lemma A.1 entails that there exists constants C_{1k}, C_{2k}, a_k which depend on the distribution of ξ_{ik}, such that for $x_n = C_{1k} n^{\beta}$, $(n/H(a_k x_n)) = a_k^{-\eta_1} C_{1k}^{-\eta_1} n^{1-\eta_1\beta}$ and iid $N(0,1)$ variables $Z_{ik,\xi}$ such that

$$
P\left[\max_{1\leq t\leq n} \left|\sum_{i=1}^{t} \xi_{ik} - \sum_{i=1}^{t} Z_{ik,\xi}\right| > C_{1k} n^{\beta}\right] < C_{2k} a_k^{-\eta_1} C_{1k}^{-\eta_1} n^{1-\eta_1\beta}.
$$

Since $\eta_1 > 2/\beta$, $\gamma_1 = \eta_1\beta - 1 > 1$. If the number κ of k is finite, so there are common constants $C_1, C_2 > 0$ such that $P[\max_{1\leq t\leq n} |\sum_{i=1}^{t} \xi_{ik} - \sum_{i=1}^{t} Z_{ik,\xi}| > C_1 n^{\beta}] < C_2 n^{-\gamma_1}$ which entails Equation (3) since κ is finite. If κ is infinite but all the ξ_{ik}'s are iid, then C_{1k}, C_{2k}, a_k are the same for all k, so the above is again true.

Likewise, under Assumption (A5′), if one lets $H(x) = x^{\eta_2}$, Lemma A.1 entails that there exists constants C_1, C_2, a which depend on the distribution of ε_{ij}, such that for $x_n = C_1 n^{\beta}$, $(n/H(a_k x_n)) = a^{-\eta_2} C_1^{-\eta_2} n^{1-\eta_2\beta}$ and iid $N(0,1)$ variables $Z_{ij,\varepsilon}$ such that

$$
\max_{1\leq j\leq N} P\left\{\max_{1\leq t\leq n} \left|\sum_{i=1}^{t} \varepsilon_{ij} - \sum_{i=1}^{t} Z_{ij,\varepsilon}\right| > C_1 n^{\beta}\right\} \leq C_2 a^{-\eta_2} C_1^{-\eta_2} n^{1-\eta_2\beta},
$$

now $\eta_2\beta > 2 + \theta$ implies that there is $\gamma_2 > 1$ such that $\eta_2\beta - 1 > \gamma_2 + \theta$ and Equation (4) follows. ■

Proof of Proposition 3.2 Applying Equation (A2), $\|\tilde{m}_p - m\|_{\infty} \leq C_{p-1,1} h_m^p$. Since Assumption (A3) implies that $O(h_m^p n^{1/2}) = o(1)$, Equation (10) is proved. ■

Proof of Proposition 3.3 Denote by $\tilde{\mathbf{Z}}_{p,\varepsilon}(x) = \{B_{1-p,p}(x), \ldots, B_{N_m,p}(x)\}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}$, where $\mathbf{Z} = (\sigma(1/N)\bar{Z}_{\cdot 1,\varepsilon}, \ldots, \sigma(N/N)\bar{Z}_{\cdot N,\varepsilon})^{\mathrm{T}}$. By Equation (A4), one has $\|\mathbf{Z} - \mathbf{e}\|_{\infty} = O_{\mathrm{a.s.}}(n^{\beta-1})$, while

$$
\begin{aligned}
\|N^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{Z} - \mathbf{e})\|_{\infty} &\leq \|\mathbf{Z} - \mathbf{e}\|_{\infty} \max_{1-p\leq J\leq N_m} \langle B_{J,p}, 1\rangle_{2,N} \\
&\leq C\|\mathbf{Z} - \mathbf{e}\|_{\infty} \max_{1-p\leq J\leq N_m} \#\left\{j : B_{J,p}\left(\frac{j}{N}\right) > 0\right\} N^{-1} \leq C\|\mathbf{Z} - \mathbf{e}\|_{\infty} h_m.
\end{aligned}
$$

Also for any fixed $x \in [0, 1]$, one has

$$
\begin{aligned}
\|\tilde{\mathbf{Z}}_{p,\varepsilon}(x) - \tilde{e}_p(x)\|_{\infty} &= \|\{B_{1-p,p}(x), \ldots, B_{N_m,p}(x)\}\hat{\mathbf{V}}_p^{-1} N^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{Z} - \mathbf{e})\|_{\infty} \\
&\leq C\|\hat{\mathbf{V}}_p^{-1}\|_{\infty} \|\mathbf{Z} - \mathbf{e}\|_{\infty} h_m = O_{\mathrm{a.s.}}(n^{\beta-1}).
\end{aligned}
$$

Note next that the random vector $\hat{\mathbf{V}}_p^{-1} N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}$ is $(N_m + p)$-dimensional normal with covariance matrix $N^{-2}\hat{\mathbf{V}}_p^{-1}\mathbf{X}^{\mathrm{T}}\,\mathrm{var}(\mathbf{Z})\mathbf{X}\hat{\mathbf{V}}_p^{-1}$, bounded above by

$$
\max_{x\in[0,1]} \sigma^2(x) \|n^{-1} N^{-1} \hat{\mathbf{V}}_p^{-1} \hat{\mathbf{V}}_p \hat{\mathbf{V}}_p^{-1}\|_{\infty} \leq C N^{-1} n^{-1} \|\hat{\mathbf{V}}_p^{-1}\|_{\infty} \leq C N^{-1} n^{-1} h_m^{-1},
$$

bounding the tail probabilities of entries of $\hat{\mathbf{V}}_p^{-1}N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}$ and applying the Borel–Cantelli lemma leads to

$$\begin{aligned}\|\hat{\mathbf{V}}_p^{-1}N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}\|_\infty &= O_{\mathrm{a.s.}}(N^{-1/2}n^{-1/2}h_m^{-1/2}\log^{1/2}(N_m+p)) \\ &= O_{\mathrm{a.s.}}(N^{-1/2}n^{-1/2}h_m^{-1/2}\log^{1/2}n).\end{aligned}$$

Hence, $\sup_{x\in[0,1]}|n^{1/2}\tilde{\mathbf{Z}}_{p,\varepsilon}(x)| = O_{\mathrm{a.s.}}(N^{-1/2}h_m^{-1/2}\log^{1/2}n)$ and

$$\sup_{x\in[0,1]}|n^{1/2}\tilde{e}_p(x)| = O_{\mathrm{a.s.}}(n^{\beta-1/2}+N^{-1/2}h_m^{-1/2}\log^{1/2}n) = o_{\mathrm{a.s.}}(1).$$

Thus, Equation (11) holds according to Assumption (A3). ■

Proof of Proposition 3.4 We denote $\tilde{\zeta}_k(x) = \bar{Z}_{\cdot k,\xi}\phi_k(x)$, $k=1,\ldots,\kappa$ and define

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

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

Using Lemma A.4, one obtains that

$$\|\tilde{\phi}_k\|_\infty \le c_{\phi,p}\|\phi_k\|_\infty,\ \|\tilde{\phi}_k-\phi_k\|_\infty \le \tilde{C}_{0,\mu}\|\phi_k\|_{0,\mu}h_m^\mu,\ 1\le k\le\kappa. \tag{A6}$$

Applying the above Equations (A6) and (A5) and Assumptions (A3) and (A4), one has

$$\begin{aligned}&En^{1/2}\sup_{x\in[0,1]}G(x,x)^{-1/2}\left|\sum_{k=1}^{\kappa}\bar{\xi}_{\cdot k}\{\phi_k(x)-\tilde{\phi}_k(x)\}\right| \\ &\le Cn^{1/2}\left\{\sum_{k=1}^{\kappa_n}E|\bar{\xi}_{\cdot k}|\|\phi_k\|_{0,\mu}h_m^\mu + \sum_{k=\kappa_n+1}^{\kappa}E|\bar{\xi}_{\cdot k}|\|\phi_k\|_\infty\right\} \\ &\le C\left\{\sum_{k=1}^{\kappa_n}\|\phi_k\|_{0,\mu}h_m^\mu + \sum_{k=\kappa_n+1}^{\kappa}\|\phi_k\|_\infty\right\} = o(1),\end{aligned}$$

hence

$$n^{1/2}\sup_{x\in[0,1]}G(x,x)^{-1/2}\left|\sum_{k=1}^{\kappa}\bar{\xi}_{\cdot k}\{\phi_k(x)-\tilde{\phi}_k(x)\}\right| = o_P(1). \tag{A7}$$

In addition, Equation (A3) and Assumptions (A3) and (A4) entail that

$$\begin{aligned}&En^{1/2}\sup_{x\in[0,1]}G(x,x)^{-1/2}\left|\sum_{k=1}^{\kappa}(\bar{Z}_{\cdot k,\xi}-\bar{\xi}_{\cdot k})\phi_k(x)\right| \\ &\le Cn^{\beta-1/2}\sum_{k=1}^{\kappa}\|\phi_k\|_\infty = o(1),\end{aligned}$$

hence

$$n^{1/2}\sup_{x\in[0,1]}G(x,x)^{-1/2}\left|\sum_{k=1}^{\kappa}(\bar{Z}_{\cdot k,\xi}-\bar{\xi}_{\cdot k})\phi_k(x)\right| = o_P(1). \tag{A8}$$

Note that

$$\bar{m}(x)-m(x)-\tilde{\xi}_p(x) = \sum_{k=1}^{\kappa}\bar{\xi}_{\cdot k}\{\phi_k(x)-\tilde{\phi}_k(x)\},$$

$$n^{-1/2}G(x,x)^{1/2}\tilde{\zeta}(x)-\{\bar{m}(x)-m(x)\} = \sum_{k=1}^{\kappa}(\bar{Z}_{\cdot k,\xi}-\bar{\xi}_{\cdot k})\phi_k(x),$$

hence

$$n^{1/2}\sup_{x\in[0,1]}G(x,x)^{-1/2}|\bar{m}(x)-m(x)-\tilde{\xi}_p(x)| = o_P(1),$$

$$\sup_{x\in[0,1]}|\tilde{\zeta}(x)-n^{1/2}G(x,x)^{-1/2}\{\bar{m}(x)-m(x)\}| = o_P(1).$$

according to Equations (A7) and (A8), which leads to both Equations (12) and (13). ■