Statistica Sinica **22** (2012), 95-122
doi:http://dx.doi.org/10.5705/ss.2010.034

A SIMULTANEOUS CONFIDENCE BAND FOR SPARSE LONGITUDINAL REGRESSION

Shujie Ma[1], Lijian Yang[2,3] and Raymond J. Carroll[4]

[1]*University of California-Riverside,* [2]*Soochow University,*
[3]*Michigan State University and* [4]*Texas A&M University*

Abstract: Functional data analysis has received considerable recent attention and a number of successful applications have been reported. In this paper, asymptotically simultaneous confidence bands are obtained for the mean function of the functional regression model, using piecewise constant spline estimation. Simulation experiments corroborate the asymptotic theory. The confidence band procedure is illustrated by analyzing CD4 cell counts of HIV infected patients.

Key words and phrases: B spline, confidence band, functional data, Karhunen-Loève L^2 representation, knots, longitudinal data, strong approximation.

1. Introduction

Functional data analysis (FDA) has in recent years become a focal area in statistics research, and much has been published in this area. An incomplete list includes Cardot, Ferraty, and Sarda (2003), Cardot and Sarda (2005), Ferraty and Vieu (2006), Hall and Heckman (2002), Hall, Müller, and Wang (2006), Izem and Marron (2007), James, Hastie, and Sugar (2000), James (2002), James and Silverman (2005), James and Sugar (2003), Li and Hsing (2007, 2010), Morris and Carroll (2006), Müller and Stadtmüller (2005), Müller, Stadtmüller, and Yao (2006), Müller and Yao (2008), Ramsay and Silverman (2005), Wang, Carroll, and Lin (2005), Yao and Lee (2006), Yao, Müller, and Wang (2005a,b), Zhang and Chen (2007), Zhao, Marron, and Wells (2004), and Zhou, Huang, and Carroll (2008). According to Ferraty and Vieu (2006), a functional data set consists of iid realizations $\{\xi_i(x), x \in \mathcal{X}\}, 1 \le i \le n$, of a smooth stochastic process (random curve) $\{\xi(x), x \in \mathcal{X}\}$ over an entire interval $\mathcal{X}$. A more data oriented alternative in Ramsay and Silverman (2005) emphasizes smooth functional features inherent in discretely observed longitudinal data, so that the recording of each random curve $\xi_i(x)$ is over a finite number of points in $\mathcal{X}$, and contaminated with noise. This second view is taken in this paper.

A typical functional data set therefore has the form $\{X_{ij}, Y_{ij}\}, 1 \le i \le n, 1 \le j \le N_i$, in which N_i observations are taken for the i^{th} subject, with X_{ij}

and Y_{ij} the j^{th} predictor and response variables, respectively, for the i^{th} subject. Generally, the predictor X_{ij} takes values in a compact interval $\mathcal{X} = [a, b]$. For the i^{th} subject, its sample path $\{X_{ij}, Y_{ij}\}$ is the noisy realization of a continuous time stochastic process $\xi_i(x)$ in the sense that

$$Y_{ij} = \xi_i(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij}, \tag{1.1}$$

with errors ε_{ij} satisfying $E(\varepsilon_{ij}) = 0, E(\varepsilon_{ij}^2) = 1$, and $\{\xi_i(x), x \in \mathcal{X}\}$ are iid copies of a process $\{\xi(x), x \in \mathcal{X}\}$ which is L^2, i.e., $E \int_{\mathcal{X}} \xi^2(x)dx < +\infty$.

For the standard process $\{\xi(x), x \in \mathcal{X}\}$, one defines the mean function $m(x) = E\{\xi(x)\}$ and the covariance function $G(x, x') = \text{cov}\{\xi(x), \xi(x')\}$. Let sequences $\{\lambda_k\}_{k=1}^{\infty}, \{\psi_k(x)\}_{k=1}^{\infty}$ 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 $L^2(\mathcal{X})$ and $G(x, x') = \sum_{k=1}^{\infty} \lambda_k \psi_k(x)\psi_k(x')$, which implies that $\int G(x, x')\psi_k(x')dx' = \lambda_k \psi_k(x)$.

The process $\{\xi_i(x), x \in \mathcal{X}\}$ allows the Karhunen-Loève L^2 representation

$$\xi_i(x) = m(x) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(x),$$

where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variances 1, and the functions $\phi_k = \sqrt{\lambda_k}\psi_k$. In what follows, we assume that $\lambda_k = 0$, for $k > \kappa$, where κ is a positive integer, thus $G(x, x') = \sum_{k=1}^{\kappa} \phi_k(x)\phi_k(x')$ and the data generating process is now written as

$$Y_{ij} = m(X_{ij}) + \sum_{k=1}^{\kappa} \xi_{ik}\phi_k(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij}. \tag{1.2}$$

The sequences $\{\lambda_k\}_{k=1}^{\kappa}, \{\phi_k(x)\}_{k=1}^{\kappa}$ and the random coefficients ξ_{ik} exist mathematically, but are unknown and unobservable.

Two distinct types of functional data have been studied. Li and Hsing (2007), and Li and Hsing (2010) concern dense functional data, which in the context of model (1.1) means $\min_{1\leq i\leq n} N_i \to \infty$ as $n \to \infty$. On the other hand, Yao, Müller, and Wang (2005a,b), and Yao (2007) studied sparse longitudinal data for which N_i's are i.i.d. copies of an integer-valued positive random variable. Pointwise asymptotic distributions were obtained in Yao (2007) for local polynomial estimators of $m(x)$ based on sparse functional data, but without uniform confidence bands. Nonparametric simultaneous confidence bands are a powerful tool of global inference for functions, see Claeskens and Van Keilegom (2003), Fan and Zhang (2000), Hall and Titterington (1988), Härdle (1989), Härdle and Marron (1991) Huang et al. (2008), Ma and Yang (2011), Song and Yang

(2009), Wang and Yang (2009), Wu and Zhao (2007), Zhao and Wu (2008), and Zhou, Shen, and Wolfe (1998) for its theory and applications. The fact that a simultaneous confidence band has not been established for functional data analysis is certainly not due to lack of interesting applications, but to the greater technical difficulty in formulating such bands for functional data and establishing their theoretical properties. Specifically, the strong approximation results used to establish the asymptotic confidence level in nearly all published works on confidence bands, commonly known as “Hungarian embedding”, are unavailable for sparse functional data.

In this paper, we present simultaneous confidence bands for $m(x)$ in sparse functional data via a piecewise-constant spline smoothing approach. While there exist a number of smoothing methods for estimating $m(x)$ and $G(x, x')$ such as kernels (Yao, Müller, and Wang (2005a,b) and Yao (2007)), penalized splines (Cardot, Ferraty, and Sarda (2003); Cardot and Sarda (2005); Yao and Lee (2006)), wavelets Morris and Carroll (2006), and parametric splines James (2002), we choose B splines (Zhou, Huang, and Carroll (2008)) for simple implementation, fast computation and explicit expression, see Huang and Yang (2004), Wang and Yang (2007), and Xue and Yang (2006) for discussion of the relative merits of various smoothing methods.

We organize our paper as follows. In Section 2 we state our main results on confidence bands constructed from piecewise constant splines. In Section 3 we provide further insights into the error structure of spline estimators. Section 4 describes the actual steps to implement the confidence bands. Section 5 reports findings of a simulation study. An empirical example in Section 6 illustrates how to use the proposed confidence band for inference. Proofs of technical lemmas are in the Appendix.

2. Main Results

For convenience, we denote the supremum norm of a function r on $[a, b]$ by $\|r\|_\infty = \sup_{x\in[a,b]} |r(x)|$, and the modulus of continuity of a continuous function r on $[a, b]$ by $\omega(r, \delta) = \max_{x,x'\in[a,b],|x-x'|\le\delta} |r(x) - r(x')|$. Denote by $\|g\|_2$ the theoretical L^2 norm of a function g on $[a, b]$, $\|g\|_2^2 = E\{g^2(X)\} = \int_a^b g^2(x)f(x)dx$, where $f(x)$ is the density function of X, and the empirical L^2 norm as $\|g\|_{2,N_{\mathrm{T}}}^2 = N_{\mathrm{T}}^{-1}\sum_{i=1}^n\sum_{j=1}^{N_i} g^2(X_{ij})$, where we denote the total sample size by $N_{\mathrm{T}} = \sum_{i=1}^n N_i$. Without loss of generality, we take the range of X, $\mathcal{X} = [a, b]$, to be $[0, 1]$. For any $\beta \in (0, 1]$, we denote the collection of order β Hõlder continuous function on $[0, 1]$ by

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

in which $\|\phi\|_{0,\beta}$ is the $C^{0,\beta}$-seminorm of ϕ. Let $C[0,1]$ be the collection of continuous function on $[0,1]$. Clearly, $C^{0,\beta}[0,1] \subset C[0,1]$ and, if $\phi \in C^{0,\beta}[0,1]$, then $\omega(\phi,\delta) \le \|\phi\|_{0,\beta}\,\delta^{\beta}$.

To introduce the spline functions, divide the finite interval $[0,1]$ into $(N_{\rm s}+1)$ equal subintervals $\chi_J = [t_J, t_{J+1}), J = 0,\ldots,N_{\rm s}-1, \chi_{N_{\rm s}} = [t_{N_{\rm s}}, 1]$. A sequence of equally-spaced points $\{t_J\}_{J=1}^{N_{\rm s}}$, called interior knots, are given as

$$t_0 = 0 < t_1 < \cdots < t_{N_{\rm s}} < 1 = t_{N_{\rm s}+1}, t_J = Jh_{\rm s},\ 0 \le J \le N_{\rm s}+1, h_{\rm s} = \frac{1}{(N_{\rm s}+1)},$$

in which $h_{\rm s}$ is the distance between neighboring knots. We denote by $G^{(-1)} = G^{(-1)}[0,1]$ the space of functions that are constant on each χ_J. For any $x \in [0,1]$, define its location index as $J(x) = J_n(x) = \min\{[x/h_{\rm s}], N_{\rm s}\}$ so that $t_{J_n(x)} \le x < t_{J_n(x)+1}, \forall x \in [0,1]$. We propose to estimate the mean function $m(x)$ by

$$\widehat{m}(x) = \underset{g \in G^{(-1)}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \{Y_{ij} - g(X_{ij})\}^2. \tag{2.1}$$

The technical assumptions we need are as follows

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

(A2) *The functions* $f(x), \sigma(x)$, *and* $\phi_k(x) \in C^{0,\beta}[0,1]$ *for some* $\beta \in (2/3,1]$ *with* $f(x) \in [c_f, C_f], \sigma(x) \in [c_\sigma, C_\sigma], x \in [0,1]$, *for constants* $0 < c_f \le C_f < \infty, 0 < c_\sigma \le C_\sigma < \infty$.

(A3) *The set of random variables* $(N_i)_{i=1}^{n}$ *is a subset of* $(N_i)_{i=1}^{\infty}$ *consisting of independent variables* N_i, *the numbers of observations made for the* i*-th subject,* $i = 1,2,\ldots$, *with* $N_i \backsim N$, *where* $N > 0$ *is a positive integer-valued random variable with* $E\{N^{2r}\} \le r!c_N^r, r = 2,3,\ldots$ *for some constant* $c_N > 0$. *The set of random variables* $(X_{ij}, Y_{ij}, \varepsilon_{ij})_{i=1,j=1}^{n,N_i}$ *is a subset of* $(X_{ij}, Y_{ij}, \varepsilon_{ij})_{i=1,j=1}^{\infty,\infty}$ *in which* $(X_{ij}, \varepsilon_{ij})_{i=1,j=1}^{\infty,\infty}$ *are iid. The number* κ *of nonzero eigenvalues is finite and the random coefficients* $\xi_{ik}, k = 1,\ldots,\kappa, i = 1,\ldots,\infty$ *are iid* $N(0,1)$. *The variables* $(N_i)_{i=1}^{\infty}$, $(\xi_{ik})_{i=1,k=1}^{\infty,\kappa}$, $(X_{ij})_{i=1,j=1}^{\infty,\infty}$, $(\varepsilon_{ij})_{i=1,j=1}^{\infty,\infty}$ *are independent.*

(A4) *As* $n \to \infty$, *the number of interior knots* $N_{\rm s} = o(n^{\vartheta})$ *for some* $\vartheta \in (1/3, 2\beta-1)$ *while* $N_{\rm s}^{-1} = o\{n^{-1/3}(\log n)^{-1/3}\}$. *The subinterval length* $h_{\rm s} \sim N_{\rm s}^{-1}$.

(A5) *There exists* $r > 2/\{\beta - (1+\vartheta)/2\}$ *such that* $E|\varepsilon_{11}|^r < \infty$.

Assumptions (A1), (A2), (A4) and (A5) are similar to (A1)–(A4) in Wang and Yang (2009), with (A1) weaker than its counterpart. Assumption (A3) is the same as (A1.1), (A1.2), and (A5) in Yao, Müller, and Wang (2005b), without requiring joint normality of the measurement errors ε_{ij}.

We now introduce the B-spline basis of $G^{(-1)}$, the space of piecewise constant splines, as $\{b_J(x)\}_{J=0}^{N_s}$, which are simply indicator functions of intervals χ_J, $b_J(x) = I_{\chi_J}(x), J = 0, \ldots, N_s$. Define

$$
\begin{aligned}
c_{J,n} &= \|b_J\|_2^2 = \int_0^1 b_J(x) f(x) dx, J = 0, \ldots, N_s, && (2.2)\\
\sigma_Y^2(x) &= \mathrm{var}(Y\,|X = x\,) = G\,(x, x) + \sigma^2(x), \forall x \in [0, 1]\,,\\
\sigma_n^2(x) &= c_{J(x),n}^{-2}\{nE(N_1)\}^{-1}\Big\{\frac{E\,\{N_1\,(N_1-1)\}}{EN_1}\sum_{k=1}^{\kappa}\Big(\int_{\chi_{J(x)}}\phi_k\,(u)\,f\,(u)\,du\Big)^2\\
&\quad + \int_{\chi_{J(x)}}\sigma_Y^2\,(u)\,f\,(u)\,du\Big\}. && (2.3)
\end{aligned}
$$

In addition, define

$$
\begin{aligned}
Q_{N_s+1}\,(\alpha) &= b_{N_s+1} - a_{N_s+1}^{-1}\log\{-\frac{1}{2}\log(1-\alpha)\},\\
a_{N_s+1} &= \{2\log\,(N_s+1)\}^{1/2}\,, b_{N_s+1} = a_{N_s+1} - \frac{\log\left(2\pi a_{N_s+1}^2\right)}{2a_{N_s+1}}, \qquad (2.4)
\end{aligned}
$$

for any $\alpha \in (0, 1)$. We now state our main results.

Theorem 1. *Under Assumptions* (A1)$-$(A5), *for any* $\alpha \in (0,1)$,

$$
\begin{aligned}
&\lim_{n\to\infty} P\Big\{\sup_{x\in[0,1]}\frac{|\widehat{m}(x)-m(x)|}{\sigma_n(x)} \le Q_{N_s+1}\,(\alpha)\Big\} = 1-\alpha,\\
&\lim_{n\to\infty} P\left\{\frac{|\widehat{m}(x)-m(x)|}{\sigma_n(x)} \le Z_{1-\alpha/2}\right\} = 1-\alpha, \quad \forall\ x \in [0,1],
\end{aligned}
$$

where $\sigma_n(x)$ *and* $Q_{N_s+1}\,(\alpha)$ *are given in* (2.3) *and* (2.4), *respectively, while* $Z_{1-\alpha/2}$ *is the* $100\,(1-\alpha/2)^{th}$ *percentile of the standard normal distribution.*

The definition of $\sigma_n(x)$ in (2.3) does not allow for practical use. The next proposition provides two data-driven alternatives

Proposition 1. *Under Assumptions* (A2), (A3), *and* (A5), *as* $n \to \infty$,

$$
\sup_{x\in[0,1]}\left\{\left|\sigma_n^{-1}(x)\sigma_{n,\mathrm{IID}}(x)-1\right| + \left|\sigma_n^{-1}(x)\sigma_{n,\mathrm{LONG}}(x)-1\right|\right\} = O\left(h_s^{\beta}\right),
$$

in which for $x \in [0,1]$, $\sigma_{n,\mathrm{IID}}(x) \equiv \sigma_Y(x)\,\{f(x)h_s nE(N_1)\}^{-1/2}$ *and*

$$
\sigma_{n,\mathrm{LONG}}(x) \equiv \sigma_{n,\mathrm{IID}}(x)\left\{1 + \frac{E\,\{N_1\,(N_1-1)\}}{EN_1}h_s\frac{G\,(x,x)\,f(x)}{\sigma_Y^2(x)}\right\}^{1/2}.
$$

Using $\sigma_{n,\mathrm{IID}}(x)$ instead of $\sigma_n(x)$ means to treat the (X_{ij}, Y_{ij}) as iid data rather than as sparse longitudinal data, while using $\sigma_{n,\mathrm{LONG}}(x)$ means to correctly account for the longitudinal correlation structure. The difference of the two approaches, although asymptotically negligible uniformly for $x \in [0,1]$ according to Proposition 1, is significant in finite samples, as shown in the simulation results of Section 5. For similar phenomenon with kernel smoothing, see Wang, Carroll, and Lin (2005).

Corollary 1. *Under Assumptions* (A1)−(A5)*, for any* $\alpha \in (0,1)$*, as* $n \to \infty$*, an asymptotic* $100\,(1-\alpha)\,\%$ *simultaneous confidence band for* $m(x), x \in [0,1]$ *is*

$$\widehat{m}(x) \pm \sigma_n(x) Q_{N_\mathrm{s}+1}(\alpha),$$

while an asymptotic $100\,(1-\alpha)\,\%$ *pointwise confidence interval for* $m(x), x \in [0,1]$*, is* $\widehat{m}(x) \pm \sigma_n(x) Z_{1-\alpha/2}$.

3. Decomposition

In this section, we decompose the estimation error $\widehat{m}(x) - m(x)$ by the representation of Y_{ij} as the sum of $m(X_{ij})$, $\sum_{k=1}^{\kappa} \xi_{ik}\phi_k(X_{ij})$, and $\sigma(X_{ij})\varepsilon_{ij}$.

We introduce the rescaled B-spline basis $\{B_J(x)\}_{J=0}^{N_\mathrm{s}}$ for $G^{(-1)}$, which is $B_J(x) \equiv b_J(x)\,\|b_J\|_2^{-1}, J = 0,\ldots,N_\mathrm{s}$. Therefore,

$$B_J(x) \equiv b_J(x)\,\{c_{J,n}\}^{-1/2}, J = 0,\ldots,N_\mathrm{s}. \tag{3.1}$$

It is easily verified that $\|B_J\|_2^2 = 1, J = 0,\ldots,N_\mathrm{s}, \langle B_J, B_{J'}\rangle \equiv 0, J \neq J'$.

The definition of $\widehat{m}(x)$ in (2.1) means that

$$\widehat{m}(x) \equiv \sum_{J=0}^{N_\mathrm{s}} \widehat{\lambda}'_J b_J(x), \tag{3.2}$$

with coefficients $\left\{\widehat{\lambda}'_0,\ldots,\widehat{\lambda}'_{N_\mathrm{s}}\right\}^{\mathrm{T}}$ as solutions of the least squares problem

$$\left\{\widehat{\lambda}'_0,\ldots,\widehat{\lambda}'_{N_\mathrm{s}}\right\}^{\mathrm{T}} = \underset{\{\lambda_0,\ldots,\lambda_{N_\mathrm{s}}\}\in R^{N_\mathrm{s}+1}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{N_i}\left\{Y_{ij} - \sum_{J=0}^{N_\mathrm{s}} \lambda_J b_J(X_{ij})\right\}^2.$$

Simple linear algebra shows that $\widehat{m}(x) \equiv \sum_{J=0}^{N_\mathrm{s}} \widehat{\lambda}_J B_J(x)$, where the coefficients $\{\widehat{\lambda}_0,\ldots,\widehat{\lambda}_{N_\mathrm{s}}\}^{\mathrm{T}}$ are solutions of the least squares problem

$$\left\{\widehat{\lambda}_0,\ldots,\widehat{\lambda}_{N_\mathrm{s}}\right\}^{\mathrm{T}} = \underset{\{\lambda_0,\ldots,\lambda_{N_\mathrm{s}}\}\in R^{N_\mathrm{s}+1}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum\nolimits_{j=1}^{N_i}\left\{Y_{ij} - \sum_{J=0}^{N_\mathrm{s}} \lambda_J B_J(X_{ij})\right\}^2. \tag{3.3}$$

Projecting the relationship in model (1.2) onto the linear subspace of $R^{N_{\mathrm{T}}}$ spanned by $\{B_J(X_{ij})\}_{1\le j\le N_i, 1\le i\le n, 0\le J\le N_{\mathrm{s}}}$, we obtain the following crucial decomposition in the space $G^{(-1)}$ of spline functions:

$$\widehat{m}(x) = \widetilde{m}(x) + \widetilde{e}(x) = \widetilde{m}(x) + \widetilde{\varepsilon}(x) + \sum_{k=1}^{\kappa} \widetilde{\xi}_k(x), \tag{3.4}$$

$$\widetilde{m}(x) = \sum_{J=0}^{N_{\mathrm{s}}} \widetilde{\lambda}_J B_J(x),\ \widetilde{\varepsilon}(x) = \sum_{J=0}^{N_{\mathrm{s}}} \widetilde{a}_J B_J(x),$$

$$\widetilde{\xi}_k(x) = \sum_{J=0}^{N_{\mathrm{s}}} \widetilde{\tau}_{k,J} B_J(x). \tag{3.5}$$

The vectors $\{\widetilde{\lambda}_0, \ldots, \widetilde{\lambda}_{N_{\mathrm{s}}}\}^{\mathrm{T}}$, $\{\widetilde{a}_0, \ldots, \widetilde{a}_{N_{\mathrm{s}}}\}^{\mathrm{T}}$, and $\{\widetilde{\tau}_{k,0}, \ldots, \widetilde{\tau}_{k,N_{\mathrm{s}}}\}^{\mathrm{T}}$ are solutions to (3.3) with Y_{ij} replaced by $m(X_{ij})$, $\sigma(X_{ij})\varepsilon_{ij}$, and $\xi_{ik}\phi_k(X_{ij})$, respectively. We cite next an important result concerning the function $\widetilde{m}(x)$. The first part is from de Boor (2001, p.149), and the second is from Theorem 5.1 of Huang (2003).

Theorem 2. *There is an absolute constant $C_g > 0$ such that for every $\phi \in C[0,1]$, there exists a function $g \in G^{(-1)}[0,1]$ that satisfies $\|g-\phi\|_\infty \le C_g\omega(\phi, h_{\mathrm{s}})$. In particular, if $\phi \in C^{0,\beta}[0,1]$ for some $\beta \in (0,1]$, then $\|g-\phi\|_\infty \le C_g \|\phi\|_{0,\beta} h_{\mathrm{s}}^{\beta}$. Under Assumptions* (A1) *and* (A4)*, with probability approaching* 1*, the function $\widetilde{m}(x)$ defined in* (3.5) *satisfies $\|\widetilde{m}(x) - m(x)\|_\infty = O(h_{\mathrm{s}})$.*

The next proposition concerns the function $\widetilde{e}(x)$ given in (3.4).

Proposition 2. *Under Assumptions* (A2)−(A5)*, for any $\tau \in R$, and $\sigma_n(x)$, $a_{N_{\mathrm{s}}+1}$, and $b_{N_{\mathrm{s}}+1}$ as given in* (2.3) *and* (2.4)*,*

$$\lim_{n\to\infty} P\Big\{\sup_{x\in[0,1]} \left|\sigma_n(x)^{-1}\widetilde{e}(x)\right| \le \frac{\tau}{a_{N_{\mathrm{s}}+1}} + b_{N_{\mathrm{s}}+1}\Big\} = \exp\left(-2e^{-\tau}\right).$$

4. Implementation

In this section, we describe procedures to implement the confidence bands and intervals given in Corollary 1. Given any data set $(X_{ij}, Y_{ij})_{j=1,i=1}^{N_i,n}$ from model (1.2), the spline estimator $\widehat{m}(x)$ is obtained by (3.2), and the number of interior knots in (3.2) is taken to be $N_{\mathrm{s}} = [cN_{\mathrm{T}}^{1/3}(\log n)]$, in which $[a]$ denotes the integer part of a and c is a positive constant. When constructing the confidence bands, one needs to evaluate the function $\sigma_n^2(x)$ by estimating the unknown functions $f(x)$, $\sigma_Y^2(x)$, and $G(x,x)$, and then plugging in these estimators: the same approach is taken in Wang and Yang (2009).

The number of interior knots for pilot estimation of $f(x)$, $\sigma_Y^2(x)$, and $G(x,x)$ is taken to be $N_{\mathrm{s}}^* = \left[n^{1/3}\right]$, and $h_{\mathrm{s}}^* = 1/(1+N_{\mathrm{s}}^*)$. The histogram pilot estimator of the density function $f(x)$ is

$$\widehat{f}(x) = \frac{\sum_{i=1}^n \sum_{j=1}^{N_i} b_{J(x)}(X_{ij})}{(\sum_{i=1}^n N_i) h_{\mathrm{s}}^*}.$$

Defining the vector $\mathbf{R} = \{R_{ij}\}_{1\le j\le N_i, 1\le i\le n}^{\mathrm{T}} = \left\{(Y_{ij} - \widehat{m}(X_{ij}))^2\right\}_{1\le j\le N_i, 1\le i\le n}^{\mathrm{T}}$, the estimator of $\sigma_Y^2(x)$ is $\widehat{\sigma}_Y^2(x) = \sum_{J=0}^{N_{\mathrm{s}}^*} \widehat{\rho}_J b_J(x)$, where the coefficients $\{\widehat{\rho}_0,\ldots, \widehat{\rho}_{N_{\mathrm{s}}^*}\}^{\mathrm{T}}$ are solutions of the least squares problem:

$$\left\{\widehat{\rho}_0,\ldots,\widehat{\rho}_{N_{\mathrm{s}}^*}\right\}^{\mathrm{T}} = \underset{\left\{\widehat{\rho}_0,\ldots,\widehat{\rho}_{N_{\mathrm{s}}^*}\right\}\in R^{N_{\mathrm{s}}+1}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^{N_i} \left\{R_{ij} - \sum_{J=0}^{N_{\mathrm{s}}^*} \rho_J b_J(X_{ij})\right\}^2.$$

The pilot estimator of covariance function $G(x,x')$ is

$$\widehat{G}(x,x') = \underset{g\in G^{(-1)}\otimes G^{(-1)}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j,j'=1, j\neq j'}^{N_i} \left\{C_{ijj'} - g\left(X_{ij}, X_{ij'}\right)\right\}^2,$$

where $C_{ijj'} = \{Y_{ij} - \widehat{m}(X_{ij})\}\left\{Y_{ij'} - \widehat{m}\left(X_{ij'}\right)\right\}, 1\le j,j'\le N_i, 1\le i\le n$. The function $\sigma_n(x)$ is estimated by either $\widehat{\sigma}_{n,\mathrm{IID}}(x) \equiv \widehat{\sigma}_Y(x)\left\{\widehat{f}(x) h_{\mathrm{s}} N_{\mathrm{T}}\right\}^{-1/2}$ or

$$\widehat{\sigma}_{n,\mathrm{LONG}}(x) \equiv \widehat{\sigma}_{n,\mathrm{IID}}(x)\left\{1 + \left(\sum_{i=1}^n \frac{N_i^2}{N_{\mathrm{T}}} - 1\right) \frac{\widehat{G}(x,x)}{\widehat{\sigma}_Y^2(x)} \widehat{f}(x) h_{\mathrm{s}}\right\}^{1/2}.$$

We now state a result. That is easily proved by standard theory of kernel and spline smoothing, as in Wang and Yang (2009).

Proposition 3. *Under Assumptions* (A1)−(A5), *as* $n\to\infty$

$$\sup_{x\in[0,1]} \left\{\left|\widehat{\sigma}_{n,\mathrm{IID}}(x)\sigma_{n,\mathrm{IID}}^{-1}(x) - 1\right| + \left|\widehat{\sigma}_{n,\mathrm{LONG}}(x)\sigma_{n,\mathrm{LONG}}^{-1}(x) - 1\right|\right\}$$
$$= O_{a.s.}\left(h_{\mathrm{s}}^\beta + n^{-1/2} N_{\mathrm{s}}^{-1} (\log n)^{1/2}\right).$$

Proposition 1, about how $\sigma_{n,\mathrm{IID}}(x)$ and $\sigma_{n,\mathrm{LONG}}(x)$ uniformly approximate $\sigma_n(x)$, and Proposition 3 together imply that both $\widehat{\sigma}_{n,\mathrm{IID}}(x)$ and $\widehat{\sigma}_{n,\mathrm{LONG}}(x)$ approximate $\sigma_n(x)$ uniformly at a rate faster than $(n^{-1/2+1/3}(\log n)^{1/2-1/3})$, according to Assumption (A5). Therefore as $n\to\infty$, the confidence bands

$$\widehat{m}(x) \pm \widehat{\sigma}_{n,\mathrm{IID}}(x) Q_{N_{\mathrm{s}}+1}(\alpha), \tag{4.1}$$

$$\widehat{m}(x) \pm \widehat{\sigma}_{n,\mathrm{LONG}}(x) Q_{N_{\mathrm{s}}+1}(\alpha), \tag{4.2}$$

with $Q_{N_{\mathrm{s}}+1}(\alpha)$ given in (2.4), and the pointwise intervals $\widehat{m}(x) \pm \widehat{\sigma}_{n,\mathrm{IID}}(x) Z_{1-\alpha/2}$, $\widehat{m}(x) \pm \widehat{\sigma}_{n,\mathrm{LONG}}(x) Z_{1-\alpha/2}$ have asymptotic confidence level $1-\alpha$.

5. Simulation

To illustrate the finite-sample performance of the spline approach, we generated data from the model

$$Y_{ij} = m(X_{ij}) + \sum_{k=1}^{2} \xi_{ik} \phi_k(X_{ij}) + \sigma \varepsilon_{ij}, \qquad 1 \le j \le N_i, 1 \le i \le n,$$

with $X \sim \mathrm{Uniform}[0,1]$, $\xi_k \sim \mathrm{Normal}(0,1), k = 1,2$, $\varepsilon \sim \mathrm{Normal}(0,1)$, N_i having a discrete uniform distribution from $25, \cdots, 35$, for $1 \le i \le n$, and $m(x) = \sin\{2\pi(x-1/2)\}$, $\phi_1(x) = -2\cos\{\pi(x-1/2)\}/\sqrt{5}$, $\phi_2(x) = \sin\{\pi(x-1/2)\}/\sqrt{5}$, thus $\lambda_1 = 2/5, \lambda_2 = 1/10$. The noise levels were $\sigma = 0.5, 1.0$, the number of subjects n was taken to be $20, 50, 100, 200$, the confidence levels were $1-\alpha = 0.95, 0.99$, and the constant c in the definition of N_{s} in Section 4 was taken to be $1, 2, 3$. We found that the confidence band (4.1) did not have good coverage rates for moderate sample sizes, and hence in Table 1 we report the coverage as the percentage out of the total 200 replications for which the true curve was covered by (4.2) at the 101 points $\{k/100, k = 0, \ldots, 100\}$.

At all noise levels, the coverage percentages for the confidence band (4.2) are very close to the nominal confidence levels 0.95 and 0.99 for $c = 1, 2$, but decline for $c = 3$ when $n = 20, 50$. The coverage percentages thus depend on the choice of N_{s}, and the dependency becomes stronger when sample sizes decrease. For large sample sizes $n = 100, 200$, the effect of the choice of N_{s} on the coverage percentages is insignificant. Because N_{s} varies with N_i, for $1 \le i \le n$, the data-driven selection of some "optimal" N_{s} remains an open problem.

We next examine two alternative methods to compute the confidence band, based on the observation that the estimated mean function $\widehat{m}(x)$ and the confidence intervals are step functions that remain the same on each subinterval χ_J, $0 \le J \le N_{\mathrm{s}}$. Follwing an associate editor's suggestion, locally weighted smoothing was applied to the upper and lower confidence limits to generate a smoothed confidence band. Following a referee's suggestion to treat the number $(N_{\mathrm{s}}+1)$ of subintervals as fixed instead of growing to infinity, a naive parametric confidence band was computed as

$$\widehat{m}(x) \pm \widehat{\sigma}_{n,\mathrm{LONG}}(x) Q_{1-\alpha.N_{\mathrm{s}}+1} \tag{5.1}$$

in which $Q_{1-\alpha.N_{\mathrm{s}}+1} = Z_{\left\{1+(1-\alpha)^{1/(N_{\mathrm{s}}+1)}\right\}/2}$ is the $(1-\alpha)$ quantile of the maximal absolute values of $(N_{\mathrm{s}}+1)$ iid $N(0,1)$ random variables. We compare the performance of the confidence band in (4.2), the smoothed band and naive parametric

Table 1. Uniform coverage rates from 200 replications using the confidence band (4.2). For each sample size n, the first row is the coverage of a nominal 95% confidence band, while the second row is for a 99% confidence band.

σ	n	$1-\alpha$	$c=1$	$c=2$	$c=3$
0.5	20	0.950	0.920	0.930	0.800
		0.990	0.990	0.990	0.900
	50	0.950	0.960	0.965	0.910
		0.990	0.995	0.995	0.965
	100	0.950	0.955	0.955	0.955
		0.990	1.000	1.000	0.985
	200	0.950	0.950	0.965	0.975
		0.990	0.985	0.985	0.990
1.0	20	0.950	0.935	0.930	0.735
		0.990	0.990	0.990	0.870
	50	0.950	0.975	0.960	0.895
		0.990	0.995	0.995	0.980
	100	0.950	0.950	0.940	0.935
		0.990	0.995	0.990	0.990
	200	0.950	0.940	0.965	0.960
		0.990	0.985	0.995	0.995

Table 2. Uniform coverage rates and average maximal widths of confidence intervals from 200 replications using the confidence bands (4.2), (5.1), and the smoothed bands respectively, for $1-\alpha = 0.99$.

n	σ	N_{s}	$\widehat{P}$	$\widehat{P}_{\mathrm{naive}}$	$\widehat{P}_{\mathrm{smooth}}$	W	W_{naive}	W_{smooth}
20	0.5	8	0.820	0.505	0.910	1.490	1.210	1.480
		12	0.930	0.765	0.955	1.644	1.363	1.628
	1.0	8	0.910	0.655	0.970	1.725	1.401	1.721
		12	0.960	0.820	0.985	1.937	1.606	1.928
50	0.5	44	0.990	0.960	0.990	1.651	1.522	1.609
	1.0	44	0.990	0.975	1.000	2.054	1.893	2.016

band in (5.1). Given $n = 20$ with $N_{\mathrm{s}} = 8, 12$, and $n = 50$ $N_{\mathrm{s}} = 44$ (by taking $c = 1$ in the definition of N_{s} in Section 4), $\sigma = 0.5, 1.0$, and $1-\alpha = 0.99$, Table 2 reports the coverage percentages $\widehat{P}$, $\widehat{P}_{\mathrm{naive}}, \widehat{P}_{\mathrm{smooth}}$ and the average maximal widths $W, W_{\mathrm{naive}}, W_{\mathrm{smooth}}$ of $N_{\mathrm{s}}+1$ intervals out of 200 replications calculated from confidence bands (4.2), (5.1), and the smoothed confidence bands, respectively.

In all experiments, one has $\widehat{P}_{\mathrm{smooth}} > \widehat{P} > \widehat{P}_{\mathrm{naive}}$ and $W > W_{\mathrm{smooth}} > W_{\mathrm{naive}}$. The coverage percentages for both the confidence bands in (4.2) and the smoothed bands are much closer to the nominal level than those of the naive bands in (5.1), while the smoothed bands perform slightly better than the constant spline bands in (4.2), with coverage percentages closer to the nominal and smaller widths.

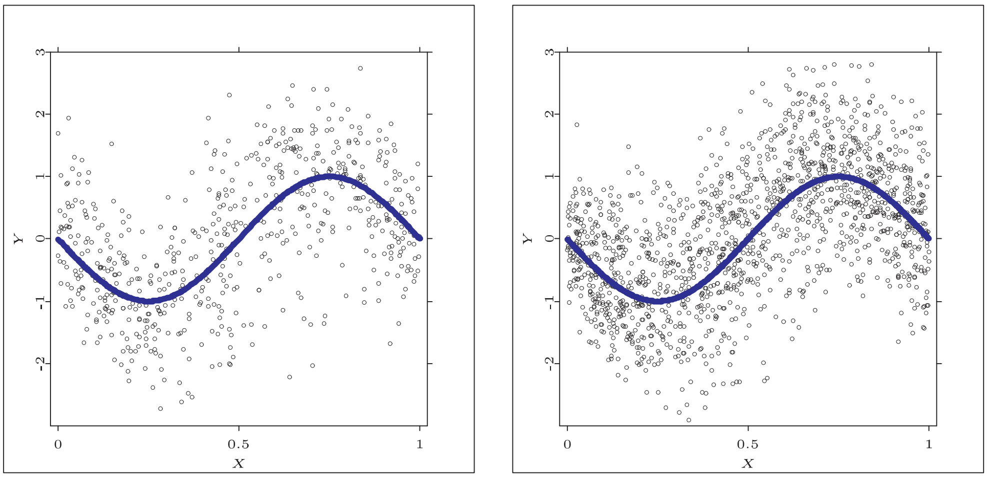

Figure 1. Plots of simulated data scatter points at $\sigma = 0.5$: (a) $n = 20$, (b) $n = 50$, and the true curve.

Based on these observations, the naive band is not recommended due to poor coverage. As for the smoothed band, although it has slightly better coverage than the constant spline band, its asymptotic property has yet to be established, and the second step smoothing adds to its conceptual complexity and computational burden. Therefore with everything considered, the constant spline band is recommended for its satisfactory theoretical property, fast computing, and conceptual simplicity.

For visualization of the actual function estimates, at $\sigma = 0.5$ with $n = 20, 50$, Figure 1 depicts the simulated data points and the true curve, and Figure 2 shows the true curve, the estimated curve, the uniform confidence band, and the pointwise confidence intervals.

6. Empirical Example

In this section, we apply the confidence band procedure of Section 4 to the data collected from a study by the AIDS Clinical Trials Group, ACTG 315 (Zhou, Huang, and Carroll (2008)). In this study, 46 HIV 1 infected patients were treated with potent antiviral therapy consisting of ritonavir, 3TC and AZT. After initiation of the treatment on day 0, patients were followed for up to 10 visits. Scheduled visit times common for all patients were 7, 14, 21, 28, 35, 42, 56, 70, 84, and 168 days. Since the patients did not follow exactly the scheduled times and/or missed some visits, the actual visit times T_{ij} were irregularly spaced and varied from day 0 to day 196. The CD4+ cell counts during HIV/AIDS treatments are taken as the response variable Y from day 0 to day 196. Figure 3 shows that the data points (dots) are extremely sparse between day 100 and 150,

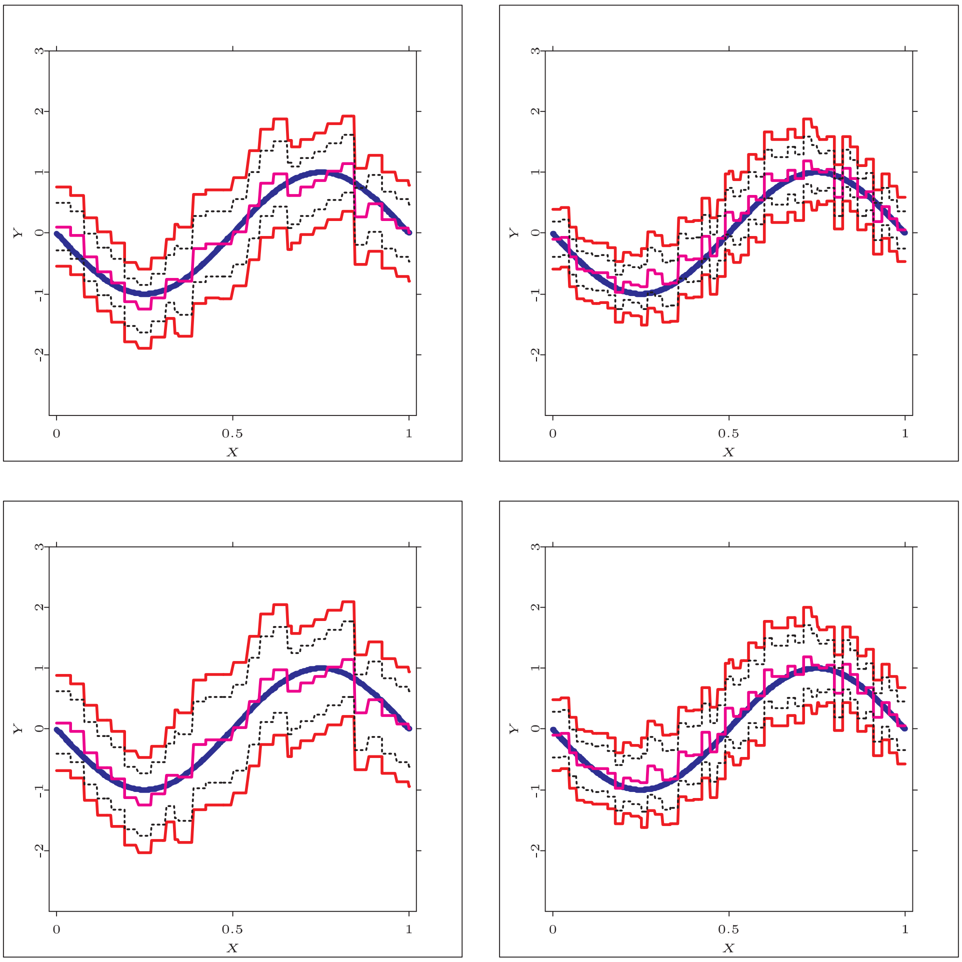

Figure 2. Plots of confidence bands (4.2) (upper and lower solid lines), pointwise confidence intervals (upper and lower dashed lines), the spline estimator (middle thin line), and the true function (middle thick line): (a) $1-\alpha = 0.95, n = 20$, (b) $1-\alpha = 0.95, n = 50$, (c) $1-\alpha = 0.99, n = 20$,(d) $1-\alpha = 0.99, n = 50$.

thus we first transform the data by $X_{ij} = T_{ij}^{1/3}$. A histogram (not shown) indicates that the X_{ij}-values are distributed fairly uniformly. The number of interior knots in (3.2) is taken to be $N_{\rm s} = 6$, so that the range for visit time T, which is $[0, 196]$, is divided into seven unequal subintervals, and in each subinterval, the mean CD4+ cell counts and the confidence bands remain the same. Table 3 gives the mean CD4+ cell counts and the confidence limits on each subinterval at simultaneous confidence level 0.95. For instance, from day 4 to 14, the mean CD4+ cell counts is 241.62 with lower and upper limits 171.81 and 311.43

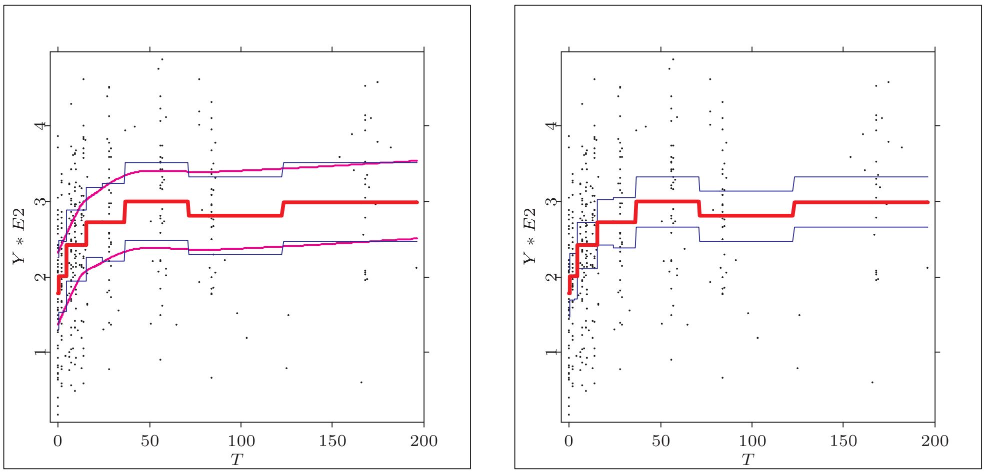

Figure 3. Plots of the piecewise-constant spline estimator (thick line), the data (dots), and (a) confidence band (4.2) (upper and lower solid lines), the smoothed band (upper and lower thin lines), (b) pointwise confidence intervals (upper and lower thin lines) at confidence level 0.95.

Table 3. The mean CD4+ cell counts and the confidence limits on each subinterval at simultaneous confidence level 0.95.

Days	Mean CD4+ cell counts	Confidence limits
$[0, 1)$	178.23	$[106.73, 249.72]$
$[1, 4)$	200.32	$[130.51, 270.13]$
$[4, 15)$	241.62	$[171.81, 311.43]$
$[15, 36)$	271.87	$[194.70, 349.04]$
$[36, 71)$	299.51	$[222.34, 376.68]$
$[71, 123)$	280.78	$[203.50, 358.06]$
$[123, 196]$	299.27	$[221.99, 376.55]$

respectively.

Figure 3 depicts (a) the 95% simultaneous (smoothed) confidence band according to (4.2) in (median) thin lines, and (b) the pointwise 95% confidence intervals in thin lines. The center thick line is the piecewise-constant spline fit $\widehat{m}(x)$. It can be seen that the pointwise confidence intervals are of course narrower than the uniform confidence band by the same ratio. Figure 3 is essentially a graphical representation of Table 3; both confirm that the mean CD4+ cell counts generally increases over time as Zhou, Huang, and Carroll (2008) pointed out. The advantage of the current method is that such inference on the overall trend is made with predetermined type I error probability, in this case 0.05.

7. Discussion

In this paper, we have constructed a simultaneous confidence band for the mean function $m(x)$ for sparse longitudinal data via piecewise-constant spline fitting. Our approach extends the asymptotic results in Wang and Yang (2009) for i.i.d. random designs to a much more complicated data structure by allowing dependence of measurements within each subject. The proposed estimator has good asymptotic behavior, and the confidence band had coverage very close to the nominal in our simulation study. An empirical study for the mean CD4+ cell counts illustrates the practical use of the confidence band.

Clearly the simultaneous confidence band in (4.2) can be improved in terms of both theoretical and numerical performance if higher order spline or local linear estimators are used. Constant piecewise spline estimators are less appealing and have sub-optimal convergence rates in the sense of Hall, Müller, and Wang (2006), which uses local linear approaches. Establishing the asymptotic confidence level for such extensions, however, requires highly sophisticated extreme value theory, for sequences of non-stationary Gaussian processes over intervals growing to infinity. That is much more difficult than the proofs of this paper. We consider the confidence band in (4.2) significant because it is the first of its kind for the longitudinal case with complete theoretical justification, and with satisfactory numerical performance for commonly encountered data sizes.

Our methodology can be applied to construct simultaneous confidence bands for other functional objects, such as the covariance function $G(x, x')$ and its eigenfunctions, see Yao (2007). It can also be adapted to the estimation of regression functions in the functional linear model, as in Li and Hsing (2007). We expect further research along these lines to yield deep theoretical results with interesting applications.

Acknowledgement

The authors thank Shuzhuan Zheng and the seminar participants at the University of Michigan, Georgia Institute of Technology, Georgia State University, University of Toledo, University of Georgia, Soochow University, University of Science and Technology of China, and Peking University for their comments on the paper. Ma and Yang's research was supported in part by NSF Awards DMS 0706518, DMS 1007594, an MSU Summer Support Fellowship and a grant from Risk Management Institute, National University of Singapore. Carroll's research was supported by a grant from the National Cancer Institute (CA57030) and by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST). The detailed and insightful comments from an associate editor and two referees are gratefully acknowledged.

Appendix

Throughout this section, $a_n \sim b_n$ means $\lim_{n\to\infty} b_n/a_n = c$, where c is some nonzero constant, and for functions $a_n(x), b_n(x)$, $a_n(x) = u\{b_n(x)\}$ means $a_n(x)/b_n(x) \to 0$ as $n \to \infty$ uniformly for $x \in [0,1]$.

A.1. Preliminaries

We first state some results on strong approximation, extreme value theory and the classic Bernstein inequality. These are used in the proofs of Lemma A.7, Theorem 1, and Lemma A.6.

Lemma A.1. (Theorem 2.6.7 of Csőrgő and Révész (1981)) *Suppose that* $\xi_i, 1 \le i \le n$ *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 exists a Wiener process* $\{W(t), 0 \le t < \infty\}$ *that is a Borel function of* $\xi_i, 1 \le i \le n$, *and constants* $C_1, C_2, a > 0$ *which depend only on the distribution of* ξ_1, *such that for any* $\{x_n\}_{n=1}^{\infty}$ *satisfying* $H^{-1}(n) < x_n < C_1(n\log n)^{1/2}$ *and* $S_k = \sum_{i=1}^{k}\xi_i$,

$$P\left\{\max_{1\le k\le n}|S_k - W(k)| > x_n\right\} \le C_2 n\{H(ax_n)\}^{-1}.$$

Lemma A.2. *Let* $\xi_i^{(n)}, 1 \le i \le n$, *be jointly normal with* $\xi_i^{(n)} \sim N(0,1)$. *Let* $r_{ij}^{(n)} = E\xi_i^{(n)}\xi_j^{(n)}$ *be such that for* $\gamma > 0, C_r > 0$, $\left|r_{ij}^{(n)}\right| < C_r/n^{\gamma}, i \ne j$. *Then for* $\tau \in R$, *as* $n \to \infty$, $P\{M_{n,\xi} \le \tau/a_n + b_n\} \to \exp(-2e^{-\tau})$, *in which* $M_{n,\xi} = \max\{|\xi_1^{(n)}|, \ldots, |\xi_n^{(n)}|\}$ *and* a_n, b_n *are as in* (2.4) *with* $N_{\mathrm{s}} + 1$ *replaced by* n.

Proof. Let $\{\eta_i\}_{i=1}^n$ be i.i.d. standard normal r.v.'s, $\mathbf{u} = \{u_i\}_{i=1}^n, \mathbf{v} = \{v_i\}_{i=1}^n$ be vectors of real numbers, and $w = \min(|u_1|, \ldots, |u_n|, |v_1|, \ldots, |v_n|)$. By the Normal Comparison Lemma (Leadbetter, Lindgren, and Rootzén (1983, Lemma 11.1.2)),

$$\left|P\left\{-v_j < \xi_j^{(n)} \le u_j \text{ for } j = 1,\ldots,n\right\} - P\{-v_j < \eta_j \le u_j \text{ for } j = 1,\ldots,n\}\right|$$
$$\le \frac{4}{2\pi}\sum_{1\le i<j\le n}\left|r_{ij}^{(n)}\right|\left(1 - \left|r_{ij}^{(n)}\right|^2\right)^{-1/2}\exp\left(\frac{-w^2}{1 + r_{ij}^{(n)}}\right).$$

If $u_1 = \cdots = u_n = v_1 = \cdots = v_n = \tau/a_n + b_n = \tau_n$, it is clear that $\tau_n^2/(2\log n) \to 1$, as $n \to \infty$. Then $\tau_n^2 > (2-\varepsilon)\log n$, for any $\varepsilon > 0$ and large n. Since $1 - r_{ij}^{(n)2} \ge 1 - (C_r/n^{\gamma})^2 \to 1$ as $n \to \infty, i \ne j$, for $i \ne j$, $\exists C_{r2} > 0$ such that $1 - r_{ij}^{(n)2} \ge C_{r2} > 0$ and $1 + r_{ij}^{(n)} < 1 + \epsilon$ for any $\epsilon > 0$ and large n.

Let $M_{n,\eta} = \max\{|\eta_1|,\ldots,|\eta_n|\}$. By Leadbetter, Lindgren, and Rootzén (1983, Thm. 1.5.3), $P\{M_{n,\eta} \le \tau_n\} \to \exp(-2e^{-\tau})$ as $n \to \infty$, while the above results entail

$$
\begin{aligned}
&|P(M_{n,\xi} \le \tau_n) - P(M_{n,\eta} \le \tau_n)| \\
&\le \frac{4}{2\pi} \sum_{1\le i<j\le n} \left|r_{ij}^{(n)}\right| \left(1 - \left|r_{ij}^{(n)}\right|^2\right)^{-1/2} \exp\left(\frac{-w^2}{1 + r_{ij}^{(n)}}\right) \\
&\le \frac{4}{2\pi} \sum_{1\le i<j\le n} C_r n^{-\gamma} C_{r2}^{-1/2} \exp\left\{\frac{-(2-\varepsilon)\log n}{1+\epsilon}\right\} \\
&\le C_r' n^{2-\gamma-(2-\varepsilon)(1+\epsilon)^{-1}} \to 0
\end{aligned}
$$

as $n \to \infty$. Hence $P\{M_{n,\xi} \le \tau_n\} \to \exp(-2e^{-\tau})$, as $n \to \infty$.

Lemma A.3. (Theorem 1.2 of Bosq (1998)) *Suppose that* $\{\xi_i\}_{i=1}^n$ *are iid with* $E(\xi_1) = 0, \sigma^2 = E\xi_1^2$, *and there exists* $c > 0$ *such that for* $r = 3, 4, \ldots, E|\xi_1|^r \le c^{r-2} r! E\xi_1^2 < +\infty$. *Then for each* $n > 1$, $t > 0$, $P(|S_n| \ge \sqrt{n}\sigma t) \le 2\exp(-t^2(4 + 2ct/\sqrt{n}\sigma)^{-1})$, *in which* $S_n = \sum_{i=1}^n \xi_i$.

Lemma A.4. *Under Assumption* (A2), *as* $n \to \infty$ *for* $c_{J,n}$ *defined in* (2.2), $c_{J,n} = f(t_J) h_{\mathrm{s}} (1 + r_{J,n}), \langle b_J, b_{J'}\rangle \equiv 0, J \ne J'$, *where* $\max_{0\le J\le N_{\mathrm{s}}} |r_{J,n}| \le C\omega(f, h_{\mathrm{s}})$. *There exist constants* $C_B > c_B > 0$ *such that* $c_B h_{\mathrm{s}}^{1-r/2} \le E\{B_J(X_{ij})\}^r \le C_B h_{\mathrm{s}}^{1-r/2}$ *for* $r = 1, 2, \ldots$ *and* $1 \le J \le N_{\mathrm{s}} + 1, 1 \le j \le N_i, 1 \le i \le n$.

Proof. By the definition of $c_{J,n}$ in (2.2),

$$
c_{J,n} = \int b_J(x) f(x) dx = \int_{[t_J, t_{J+1}]} f(x) dx = f(t_J) h_{\mathrm{s}} + \int_{[t_J, t_{J+1}]} \{f(x) - f(t_J)\} dx.
$$

Hence for all $J = 0, \ldots, N_{\mathrm{s}}, |c_{J,n} - f(t_J) h_{\mathrm{s}}| \le \int_{[t_J,t_{J+1}]} |f(x) - f(t_J)| dx \le \omega(f, h_{\mathrm{s}}) h_{\mathrm{s}}$, or $|r_{J,n}| = |c_{J,n} - f(t_J) h_{\mathrm{s}}| \{f(t_J) h_{\mathrm{s}}\}^{-1} \le C\omega(f, h_{\mathrm{s}}), J = 0, \ldots, N_{\mathrm{s}}$. By (3.1), $E\{B_J(X_{ij})\}^r = (c_{J,n})^{-r/2} \int b_J(x) f(x) dx = (c_{J,n})^{1-r/2} \sim h_{\mathrm{s}}^{1-r/2}$.

Proof of Proposition 1. By Lemma A.4 and Assumption (A2) on the continuity of functions $\phi_k^2(x), \sigma^2(x)$ and $f(x)$ on $[0, 1]$, for any $x \in [0, 1]$

$$
\left|\int_{\chi_{J(x)}} \phi_k(x) f(x) du - \int_{\chi_{J(x)}} \phi_k(u) f(u) du\right| \le \omega(\phi_k f, h_{\mathrm{s}}) h_{\mathrm{s}} = O\left(h_{\mathrm{s}}^{1+\beta}\right),
$$

$$
\left|\int_{J(x)} \left\{\sigma_Y^2(x) f(x) - \sigma_Y^2(u) f(u)\right\} du\right| \le \omega\left(\sigma_Y^2 f, h_{\mathrm{s}}\right) h_{\mathrm{s}} = O\left(h_{\mathrm{s}}^{1+\beta}\right).
$$

Hence,

$$\sigma_n^2(x) = c_{J(x),n}^{-2} (nEN_1)^{-1} \int_{J(x)} \sigma_Y^2(u) f(u)\, du$$
$$\times \left\{ 1 + \frac{E\{N_1(N_1-1)\}}{EN_1} \sum_{k=1}^{\kappa} \Big(\int_{\chi_{J(x)}} \phi_k(u) f(u)\, du \Big)^2 \right.$$
$$\left. \times \Big\{ \int_{J(x)} \sigma_Y^2(u) f(u)\, du \Big\}^{-1} \right\}$$
$$= \left\{ f(x)h_{\rm s} + U\left(h_{\rm s}^{1+\beta}\right) \right\}^{-2} (nEN_1)^{-1} \left\{ \sigma_Y^2(x) f(x) h_{\rm s} + U\left(h_{\rm s}^{1+\beta}\right) \right\}$$
$$\times \left\{ 1 + \frac{E\{N_1(N_1-1)\}}{EN_1} \sum_{k=1}^{\kappa} \left\{ \phi_k(x) f(x) h_{\rm s} + U\left(h_{\rm s}^{1+\beta}\right) \right\}^2 \right.$$
$$\left. \times \left\{ \sigma_Y^2(x) f(x) h_{\rm s} + U\left(h_{\rm s}^{1+\beta}\right) \right\}^{-1} \right\}$$
$$= (f(x)h_{\rm s} nEN_1)^{-1} \sigma_Y^2(x) \left\{ 1 + \frac{E\{N_1(N_1-1)\}}{EN_1} \frac{\sum_{k=1}^{\kappa} \phi_k^2(x) f(x) h_{\rm s}}{\sigma_Y^2(x)} \right\}$$
$$\times \left\{ 1 + U\left(h_{\rm s}^{\beta}\right) \right\}$$
$$= \sigma_{n,\rm LONG}^2(x) \left\{ 1 + U\left(h_{\rm s}^{\beta}\right) \right\} = \sigma_{n,\rm IID}^2(x) \left\{ 1 + U\left(h_{\rm s}^{\beta}\right) \right\}.$$

A.2. Proof of Theorem 1

Note that $B_{J(x)}(x) \equiv c_{J(x),n}^{-1/2}$, $x \in [0,1]$, so the terms $\widetilde{\xi}_k(x)$ and $\widetilde{\varepsilon}(x)$ defined in (3.5) are

$$\widetilde{\xi}_k(x) = \sum_{J=0}^{N_{\rm S}} N_{\rm T}^{-1} B_J(x) \|B_J\|_{2,N_{\rm T}}^{-2} \sum_{i=1}^{n} \sum_{j=1}^{N_i} B_J(X_{ij}) \phi_k(X_{ij}) \xi_{ik}$$
$$= c_{J(x),n}^{-1/2} \left\|B_{J(x)}\right\|_{2,N_{\rm T}}^{-2} N_{\rm T}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} B_{J(x)}(X_{ij}) \phi_k(X_{ij}) \xi_{ik},$$
$$\widetilde{\varepsilon}(x) = c_{J(x),n}^{-1/2} \left\|B_{J(x)}\right\|_{2,N_{\rm T}}^{-2} N_{\rm T}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} B_{J(x)}(X_{ij}) \sigma(X_{ij}) \varepsilon_{ij}.$$

Let

$$\widehat{\xi}_k(x) = \left\|B_{J(x)}\right\|_{2,N_{\rm T}}^{2} \widetilde{\xi}_k(x) = c_{J(x),n}^{-1/2} N_{\rm T}^{-1} \sum_{i=1}^{n} R_{ik,\xi,J(x)} \xi_{ik},$$
$$\widehat{\varepsilon}(x) = \left\|B_{J(x)}\right\|_{2,N_{\rm T}}^{2} \widetilde{\varepsilon}(x) = c_{J(x),n}^{-1/2} N_{\rm T}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} R_{ij,\varepsilon,J(x)} \varepsilon_{ij}, \qquad \text{(A.1)}$$

where

$$R_{ik,\xi,J} = \sum_{j=1}^{N_i} B_J(X_{ij})\,\phi_k(X_{ij}),\, R_{ij,\varepsilon,J} = B_J(X_{ij})\,\sigma(X_{ij}),\, 0 \le J \le N_{\mathrm{s}}. \quad \text{(A.2)}$$

Lemma A.5. *Under Assumption* (A3), *for* $\widetilde{e}(x)$ *given in* (3.4) *and* $\widehat{\xi}_k(x)$, $\widehat{\varepsilon}(x)$ *given in* (A.1), *we have*

$$\left|\widetilde{e}(x) - \left\{\sum_{k=1}^{\kappa}\widehat{\xi}_k(x) + \widehat{\varepsilon}(x)\right\}\right| \le A_n(1-A_n)^{-1}\left|\sum_{k=1}^{\kappa}\widehat{\xi}_k(x) + \widehat{\varepsilon}(x)\right|, x \in [0,1],$$

where $A_n = \sup_{0\le J\le N_{\mathrm{S}}}\left|\|B_J\|_{2,N_{\mathrm{T}}}^2 - 1\right|$. *There exists* $C_A > 0$, *such that for large* n, $P\left(A_n \ge C_A\sqrt{\log(n)/(nh_{\mathrm{s}})}\right) \le 2n^{-3}$. $A_n = O_{a.s.}\left(\sqrt{\log(n)/(nh_{\mathrm{s}})}\right)$ *as* $n \to \infty$.

See the Supplement of Wang and Yang (2009) for a detailed proof.

Lemma A.6. *Under Assumptions* (A2) *and* (A3), *for* $R_{1k,\xi,J}$, $R_{11,\varepsilon,J}$ *in* (A.2),

$$ER_{1k,\xi,J}^2 = c_{J,n}^{-1}\Bigg[E(N_1)\int b_J(u)\,\phi_k^2(u)\,f(u)\,du + E\{N_1(N_1-1)\}\left(\int b_J(u)\,\phi_k(u)\,f(u)\,du\right)^2\Bigg],$$

$$ER_{11,\varepsilon,J}^2 = c_{J,n}^{-1}\int b_J(u)\,\sigma^2(u)\,f(u)\,du, 0 \le J \le N_{\mathrm{s}},$$

there exist $0 < c_R < C_R < \infty$, *such that* $ER_{1k,\xi,J}^2, ER_{11,\varepsilon,J}^2 \in [c_R, C_R]$ *for* $0 \le J \le N_{\mathrm{s}}$, $\sup_{0\le J\le N_{\mathrm{s}}}\left|n^{-1}\sum_{i=1}^n R_{ik,\xi,J}^2 - ER_{1k,\xi,J}^2\right| = O_{a.s.}\left(\sqrt{\log n/(nh_{\mathrm{s}})}\right)$, $1 \le k \le \kappa$, $\sup_{0\le J\le N_{\mathrm{s}}}\left|N_{\mathrm{T}}^{-1}\sum_{i=1}^n\sum_{j=1}^{N_i} R_{ij,\varepsilon,J}^2 - ER_{11,\varepsilon,J}^2\right| = O_{a.s.}\left(\sqrt{\log n/(nh_{\mathrm{s}})}\right)$ *as* $n \to \infty$.

Proof. By independence of $X_{1j}, 1 \le j \le N_1$ and N_1 and (3.1),

$$ER_{1k,\xi,J}^2 = E\Bigg\{\sum_{j,j'=1}^{N_1} E\left\{B_J(X_{1j})\,B_J(X_{1j'})\,\phi_k(X_{1j})\,\phi_k(X_{1j'})\,|N_1\right\}\Bigg\}$$
$$= E\Bigg\{\sum_{j=1}^{N_1} E\left\{B_J^2(X_{1j})\,\phi_k^2(X_{1j})\,|N_1\right\}\Bigg\}$$
$$+E\Bigg\{\sum_{j\ne j'}^{N_1} E\left\{B_J(X_{1j})\,B_J(X_{1j'})\,\phi_k(X_{1j})\,\phi_k(X_{1j'})\,|N_1\right\}\Bigg\}$$

$$= c_{J(x),n}^{-1}\Big\{E(N_1)\int b_J(u)\,\phi_k^2(u)\,f(u)\,du$$
$$+E\{N_1(N_1-1)\}\Big(\int b_J(u)\,\phi_k(u)\,f(u)\,du\Big)^2\Big\}.$$

It is easily shown that $\exists\ 0 < c_R < C_R < \infty$ such that $c_R \leq ER_{1k,\xi,J}^2 \leq C_R, 0 \leq J \leq N_{\mathrm{s}}$. Let $\zeta_{i,J} = \zeta_{i,k,J} = R_{ik,\xi,J}^2, \zeta_{i,J}^* = \zeta_{i,J} - E(\zeta_{1,J})$ for $r \geq 1$ and large n,

$$E(\zeta_{i,J})^r = E\Big\{\sum_{j=1}^{N_i} B_J(X_{ij})\,\phi_k(X_{ij})\Big\}^{2r} \leq C_\phi^{2r}E\Big\{\sum_{j=1}^{N_i} B_J(X_{ij})\Big\}^{2r}$$
$$= C_\phi^{2r}E\Big\{\sum_{0\leq\nu_1\cdots\nu_{N_i}\leq 2r}^{\nu_1+\cdots+\nu_{N_i}=2r}\binom{2r}{\nu_1\cdots\nu_{N_i}}\prod_{j=1}^{N_i}E\{B_J(X_{ij})\}^{\nu_j}\Big\}$$
$$\leq C_\phi^{2r}E\Big\{N_1^{2r}\max\Big\{\prod_{j=1}^{N_i}E\{B_J(X_{ij})\}^{\nu_j}\Big\}\Big\} \leq C_\phi^{2r}\left(EN_1^{2r}\right)C_B h_{\mathrm{s}}^{1-r}$$
$$\leq C_\phi^{2r}C_B c_N^r r! h_{\mathrm{s}}^{1-r} = C_\zeta r! h_{\mathrm{s}}^{1-r},$$
$$E(\zeta_{i,J})^r \geq c_\phi^{2r}E\Big\{\sum_{j=1}^{N_i}E\{B_J(X_{ij})\}^{2r}\Big\} \geq c_\phi^{2r}(EN_1)\,c_B h_{\mathrm{s}}^{1-r},$$

by Lemma A.4. So $\{E(\zeta_{1,J})\}^r \sim 1$, $E(\zeta_{i,J})^r \gg \{E(\zeta_{1,J})\}^r$ for $r \geq 2$, and $\exists C_\zeta' > c_\zeta' > 0$ such that $C_\zeta' h_{\mathrm{s}}^{-1} \geq \sigma_{\zeta^*}^2 \geq c_\zeta' h_{\mathrm{s}}^{-1}$, for $\sigma_{\zeta^*} = \{E(\zeta_{i,J}^*)^2\}^{1/2}$. We obtain $E|\zeta_{i,J}^*|^r \leq c_*^{r-2}r!E(\zeta_{i,J}^*)^2$ with $c_* = (C_\zeta/c_\zeta')^{1/(r-2)}h_{\mathrm{s}}^{-1}$, which implies that $\{\zeta_{i,J}^*\}_{i=1}^n$ satisfies Cramér's condition. Applying Lemma A.3 to $\sum_{i=1}^n \zeta_{i,J}^*$, for $r > 2$ and any large enough $\delta > 0$, $P\{n^{-1}|\sum_{i=1}^n \zeta_{i,J}^*| \geq \delta\sqrt{\log n/(nh_{\mathrm{s}})}\}$ is bounded by

$$2\exp\left\{\frac{-\delta^2\left(C_\zeta'\right)^{-1}(\log n)}{4+2\left(C_\zeta/c_\zeta'\right)^{\frac{1}{r-2}}\delta\left(c_\zeta'\right)^{-1}h_{\mathrm{s}}^{1/2}(\log n)^{1/2}n^{-1/2}}\right\} \leq 2\exp\left\{\frac{-\delta^2(\log n)}{4C_\zeta'}\right\}$$
$$\leq 2n^{-3}.$$

Hence

$$\sum_{n=1}^\infty P\left\{\sup_{0\leq J\leq N_{\mathrm{s}}}\left|\frac{1}{n}\sum_{i=1}^n R_{ik,\xi,J}^2 - ER_{1k,\xi,J}^2\right| \geq \delta\sqrt{\log\frac{n}{(nh_{\mathrm{s}})}}\right\} \leq \sum_{n=1}^\infty \frac{2N_{\mathrm{s}}}{n^3} < \infty.$$

Thus, $\sup_{0\leq J\leq N_{\mathrm{s}}}\left|n^{-1}\sum_{i=1}^n R_{ik,\xi,J}^2 - ER_{1k,\xi,J}^2\right| = O_{\mathrm{a.s.}}\left(\sqrt{\log n/(nh_{\mathrm{s}})}\right)$ as $n \to \infty$ by Borel-Cantelli Lemma. The properties of $R_{ij,\varepsilon,J}$ are obtained similarly.

Order all $X_{ij}, 1 \le j \le N_i, 1 \le i \le n$ from large to small as $X_{(t)}$, $X_{(1)} \ge \cdots \ge X_{(N_{\mathrm{T}})}$, and denote the ε_{ij} corresponding to $X_{(t)}$ as $\varepsilon_{(t)}$. By (A.1),

$$\begin{aligned}\widehat{\varepsilon}(x) &= c_{J(x),n}^{-1} N_{\mathrm{T}}^{-1} \sum_{t=1}^{N_{\mathrm{T}}} b_{J(x)}\left(X_{(t)}\right) \sigma\left(X_{(t)}\right) \varepsilon_{(t)} \\ &= c_{J(x),n}^{-1} N_{\mathrm{T}}^{-1} \sum_{t=1}^{N_{\mathrm{T}}} b_{J(x)}\left(X_{(t)}\right) \sigma\left(X_{(t)}\right) \left\{S_t - S_{t-1}\right\},\end{aligned}$$

where $S_q = \sum_{t=1}^{q} \varepsilon_{(t)}, q \ge 1$ and $S_0 = 0$.

Lemma A.7. *Under Assumptions* (A2)−(A5)*, there is a Wiener process* $\{W(t), 0 \le t < \infty\}$ *independent of* $\{N_i, X_{ij}, 1 \le j \le N_i, \xi_{ik}, 1 \le k \le \kappa, 1 \le i \le n\}$*, such that as* $n \to \infty$, $\sup_{x\in[0,1]} \left|\widehat{\varepsilon}^{(0)}(x) - \widehat{\varepsilon}(x)\right| = o_{a.s.}\left(n^t\right)$ *for some* $t < -(1-\vartheta)/2 < 0$*, where* $\widehat{\varepsilon}^{(0)}(x)$ *is*

$$\left(c_{J(x),n} N_{\mathrm{T}}\right)^{-1} \sum_{t=1}^{N_{\mathrm{T}}} b_{J(x)}\left(X_{(t)}\right) \sigma\left(X_{(t)}\right) \left\{W(t) - W(t-1)\right\}, x \in [0,1]. \tag{A.3}$$

Proof. Define $M_{N_{\mathrm{T}}} = \max_{1\le q\le N_{\mathrm{T}}} |S_q - W(q)|$, in which $\{W(t), 0 \le t < \infty\}$ is the Wiener process as in Lemma A.1 that as a Borel function of the set of variables $\left\{\varepsilon_{(t)} 1 \le t \le N_{\mathrm{T}}\right\}$ is independent of $\{N_i, X_{ij}, 1 \le j \le N_i, \xi_{ik}, 1 \le k \le \kappa, 1 \le i \le n\}$ since $\left\{\varepsilon_{(t)}\ 1 \le t \le N_{\mathrm{T}}\right\}$ is. Further,

$$\begin{aligned}&\sup_{x\in[0,1]} \left|\widehat{\varepsilon}^{(0)}(x) - \widehat{\varepsilon}(x)\right| \\ &= \sup_{x\in[0,1]} c_{J(x),n}^{-1} N_{\mathrm{T}}^{-1} \Big| b_{J(x)}\left(X_{(N_{\mathrm{T}})}\right) \sigma\left(X_{(N_{\mathrm{T}})}\right) \left\{W(N_{\mathrm{T}}) - S_{N_{\mathrm{T}}}\right\} \\ &\quad + \sum_{t=1}^{N_{\mathrm{T}}-1} \left\{b_{J(x)}\left(X_{(t)}\right) \sigma\left(X_{(t)}\right) - b_{J(x)}\left(X_{(t+1)}\right) \sigma\left(X_{(t+1)}\right)\right\} \left\{W(t) - S_t\right\} \Big| \\ &\le \max_{0\le J\le N_{\mathrm{s}}+1} c_{Jn}^{-1} N_{\mathrm{T}}^{-1} \Big\{ b_J\left(X_{(N_{\mathrm{T}})}\right) \sigma\left(X_{(N_{\mathrm{T}})}\right) \\ &\quad + \sum_{t=1}^{N_{\mathrm{T}}-1} \left|b_J\left(X_{(t)}\right) \sigma\left(X_{(t)}\right) - b_J\left(X_{(t+1)}\right) \sigma\left(X_{(t+1)}\right)\right| \Big\} M_{N_{\mathrm{T}}} \\ &\le \max_{0\le J\le N_{\mathrm{s}}+1} c_{J,n}^{-1} N_{\mathrm{T}}^{-1} M_{N_{\mathrm{T}}} \Big\{ 3C_\sigma + \sum_{1\le t\le N_{\mathrm{T}}-1, X_{(t)}\in b_J} \left|\sigma\left(X_{(t)}\right) - \sigma\left(X_{(t+1)}\right)\right| \Big\}\end{aligned}$$

which, by the Hölder continuity of σ in Assumption (A2), is bounded by

$$N_{\mathrm{T}}^{-1} M_{N_{\mathrm{T}}} \max_{0\le J\le N_{\mathrm{s}}+1} c_{J,n}^{-1} \Big\{ 3C_\sigma + \|\sigma\|_{0,\beta} \sum_{1\le t\le N_{\mathrm{T}}-1, X_{(t)}\in b_J} \left|X_{(t)} - X_{(t+1)}\right|^\beta \Big\}$$

$$\leq N_{\mathrm{T}}^{-1} M_{N_{\mathrm{T}}} \max_{0\leq J\leq N_{\mathrm{s}}+1} c_{J,n}^{-1}\left\{3C_{\sigma}+\|\sigma\|_{0,\beta}\, n_J^{1-\beta}\left(\sum_{1\leq t\leq N_{\mathrm{T}}-1, X_{(t)}\in b_J}\left|X_{(t)}-X_{(t+1)}\right|\right)^{\beta}\right\}$$

$$\leq N_{\mathrm{T}}^{-1} M_{N_{\mathrm{T}}}\left(\max_{0\leq J\leq N_{\mathrm{s}}+1} c_{J,n}^{-1}\right)\left\{3C_{\sigma}+\|\sigma\|_{0,\beta}\, h_{\mathrm{s}}^{\beta}\left(\max_{0\leq J\leq N_{\mathrm{s}}+1} n_J\right)^{1-\beta}\right\}$$

where $n_J = \sum_{t=1}^{N_{\mathrm{T}}} I\left(X_{(t)} \in \chi_J\right), 0 \leq J \leq N_{\mathrm{s}}+1$, has a binomial distribution with parameters $(N_{\mathrm{T}}, p_{J,n})$, where $p_{J,n} = \int_{\chi_J} f(x)dx$. Simple application of Lemma A.3 entails $\max_{0\leq J\leq N_{\mathrm{s}}+1} n_J = O_{\mathrm{a.s.}}\left(N_{\mathrm{T}}N_{\mathrm{s}}^{-1}\right)$. Meanwhile, by letting $H(x) = x^r, x_n = n^{t'}$, $t' \in (2/r, \beta-(1+\vartheta)/2)$, the existence of which is due to the Assumption (A4) that $r > 2/\{\beta-(1+\vartheta)/2\}$. It is clear that $\left\{\varepsilon_{(t)}\right\}_{t=1}^{N_{\mathrm{T}}}$ satisfies the conditions in Lemma A.1. Since $n/H(ax_n) = a^{-r}n^{1-rt'} = O\left(n^{-\gamma_1}\right)$ for some $\gamma_1 > 1$, one can use the probability inequality in Lemma A.1 and the Borel-Cantelli Lemma to obtain $M_{N_{\mathrm{T}}} = O_{\mathrm{a.s.}}(x_n) = O_{\mathrm{a.s.}}\left(n^{t'}\right)$. Hence Lemma A.4 and the above imply

$$\begin{aligned}\sup_{x\in[0,1]}\left|\widehat{\varepsilon}^{(0)}(x)-\widehat{\varepsilon}(x)\right| &= O_{\mathrm{a.s.}}\left(N_{\mathrm{s}}n^{t'-1}\right)\left\{1+N_{\mathrm{s}}^{-\beta}\left(N_{\mathrm{T}}N_{\mathrm{s}}^{-1}\right)^{1-\beta}\right\}\\ &= O_{\mathrm{a.s.}}\left(N_{\mathrm{s}}n^{t'-1}+N_{\mathrm{s}}n^{t'-1}\times N_{\mathrm{s}}^{-1}n^{1-\beta}\right)\\ &= O_{\mathrm{a.s.}}\left(N_{\mathrm{s}}n^{t'-1}+N_{\mathrm{s}}n^{t'-\beta}\right) = o_{\mathrm{a.s.}}\left(n^{t'-\beta+\vartheta}\right)\end{aligned}$$

since $t' < \beta-(1+\vartheta)/2$ by definition, implying $t'-1 \leq t'-\beta < -(1+\vartheta)/2$. The Lemma follows by setting $t = t'-\beta+\vartheta$.

Now

$$\begin{aligned}\widehat{\varepsilon}^{(0)}(x) &= c_{J(x),n}^{-1}N_{\mathrm{T}}^{-1}\sum_{t=1}^{N_{\mathrm{T}}} b_{J(x)}\left(X_{(t)}\right)\sigma\left(X_{(t)}\right)Z_{(t)}\\ &= c_{J(x),n}^{-1}N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} b_{J(x)}\left(X_{ij}\right)\sigma\left(X_{ij}\right)Z_{ij},\end{aligned} \tag{A.4}$$

where $Z_{(t)} = W(t) - W(t-1), 1 \leq t \leq N_{\mathrm{T}}$, are i.i.d. $N(0,1)$, $\xi_{ik}, Z_{ij}, X_{ij}, N_i$ are independent, for $1 \leq k \leq \kappa, 1 \leq j \leq N_i, 1 \leq i \leq n$, and $\widehat{\xi}_k(x), \widehat{\varepsilon}^{(0)}(x)$ are conditional independent of $X_{ij}, N_i, 1 \leq j \leq N_i, 1 \leq i \leq n$. If the conditional variances of $\widehat{\xi}_k(x), \widehat{\varepsilon}^{(0)}(x)$ on $(X_{ij}, N_i)_{1\leq j\leq N_i, 1\leq i\leq n}$ are $\sigma_{\xi_k,n}^2(x), \sigma_{\varepsilon,n}^2(x)$, we have

$$\sigma_{\xi_k,n}(x) = \left\{c_{J(x),n}^{-1}N_{\mathrm{T}}^{-2}\sum_{i=1}^{n} R_{ik,\xi,J(x)}^2\right\}^{1/2} \tag{A.5}$$

$$\sigma_{\varepsilon,n}(x) = \Big\{ c_{J(x),n}^{-1} N_{\mathrm{T}}^{-2} \sum_{i=1}^{n} \sum_{j=1}^{N_i} R_{ij,\varepsilon,J(x)}^2 \Big\}^{1/2},$$

where $R_{ik,\xi,J(x)}, R_{ij,\varepsilon,J(x)}$, and $c_{J(x),n}$ are given in (A.2) and (2.2).

Lemma A.8. *Under Assumptions* (A2) *and* (A3)*, let*

$$\eta(x) = \Big\{ \sum_{k=1}^{\kappa} \sigma_{\xi_k,n}^2(x) + \sigma_{\varepsilon,n}^2(x) \Big\}^{-1/2} \Big\{ \sum_{k=1}^{\kappa} \widehat{\xi}_k(x) + \widehat{\varepsilon}^{(0)}(x) \Big\}, \tag{A.6}$$

with $\sigma_{\xi_k,n}(x), \sigma_{\varepsilon,n}(x), \widehat{\xi}_k(x), \widehat{\varepsilon}^{(0)}(x)$*, and* $c_{J(x),n}$ *given in* (A.5), (A.1), (A.3)*, and* (2.2)*. Then* $\eta(x)$ *is a Gaussian process consisting of* $(N_{\mathrm{s}}+1)$ *standard normal variables* $\{\eta_J\}_{J=0}^{N_{\mathrm{s}}}$ *such that* $\eta(x) = \eta_{J(x)}$ *for* $x \in [0,1]$*, and there exists a constant* $C > 0$ *such that for large* n*,* $\sup_{0 \le J \ne J' \le N_{\mathrm{s}}} |E\eta_J \eta_{J'}| \le C h_{\mathrm{s}}$.

Proof. It is apparent that $\mathcal{L}\{\eta_J \,|(X_{ij}, N_i), 1 \le j \le N_i, 1 \le i \le n\} = N(0,1)$ for $0 \le J \le N_{\mathrm{s}}$, so $\mathcal{L}\{\eta_J\} = N(0,1)$, for $0 \le J \le N_{\mathrm{s}}$. For $J \ne J'$, by (A.2) and (3.1), $R_{ij,\varepsilon,J} R_{ij,\varepsilon,J'} = B_J(X_{ij}) B_{J'}(X_{ij}) \sigma^2(X_{ij}) = 0$, along with (A.4), (A.3), the conditional independence of $\widehat{\xi}_k(x), \widehat{\varepsilon}^{(0)}(x)$ on $X_{ij}, N_i, 1 \le j \le N_i, 1 \le i \le n$, and independence of $\xi_{ik}, Z_{ij}, X_{ij}, N_i, 1 \le k \le \kappa, 1 \le j \le N_i, 1 \le i \le n$, $E(\eta_J \eta_{J'})$ is

$$\begin{aligned}
&E\Bigg\{\Big\{ \sum_{i=1}^{n} \Big\{ \sum_{k=1}^{\kappa} R_{ik,\xi,J}^2 + \sum_{j=1}^{N_i} R_{ij,\varepsilon,J}^2 \Big\}\Big\}^{-1/2} \Big\{ \sum_{i=1}^{n} \Big\{ \sum_{k=1}^{\kappa} R_{ik,\xi,J'}^2 + \sum_{j=1}^{N_i} R_{ij,\varepsilon,J'}^2 \Big\}\Big\}^{-1/2} \\
&\quad E\Big\{ \sum_{k=1}^{\kappa} \Big\{ \sum_{i=1}^{n} R_{ik,\xi,J} \xi_{ik} \Big\}\Big\{ \sum_{i=1}^{n} R_{ik,\xi,J'} \xi_{ik} \Big\} \\
&\quad + \Big\{ \sum_{i=1}^{n} \sum_{j=1}^{N_i} R_{ij,\varepsilon,J} Z_{ij} \Big\}\Big\{ \sum_{i=1}^{n} \sum_{j=1}^{N_i} R_{ij,\varepsilon,J'} Z_{ij} \Big\} \Big| \Big(X_{ij}, N_i\Big)_{1 \le j \le N_i, 1 \le i \le n} \Big\}\Bigg\} \\
&= E C_{n,J,J'}
\end{aligned}$$

in which

$$\begin{aligned}
C_{n,J,J'} = {} & \Big\{ N_{\mathrm{T}}^{-1} \sum_{i=1}^{n} \Big\{ \sum_{k=1}^{\kappa} R_{ik,\xi,J}^2 + \sum_{j=1}^{N_i} R_{ij,\varepsilon,J}^2 \Big\}\Big\}^{-1/2} \\
& \times \Big\{ N_{\mathrm{T}}^{-1} \sum_{i=1}^{n} \Big\{ \sum_{k=1}^{\kappa} R_{ik,\xi,J'}^2 + \sum_{j=1}^{N_i} R_{ij,\varepsilon,J'}^2 \Big\}\Big\}^{-1/2} \\
& \times \Big\{ N_{\mathrm{T}}^{-1} \sum_{k=1}^{\kappa} \sum_{i=1}^{n} R_{ik,\xi,J} R_{ik,\xi,J'} \Big\}.
\end{aligned}$$

Note that according to definitions of $R_{ik,\xi,J}, R_{ij,\varepsilon,J}$, and Lemma A.5,

$$N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\Big\{\sum_{k=1}^{\kappa}R_{ik,\xi,J}^{2}+\sum_{j=1}^{N_i}R_{ij,\varepsilon,J}^{2}\Big\}$$
$$\geq c_\sigma^2 N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}B_J^2\left(X_{ij}\right)=c_\sigma^2\left\|B_J\right\|_{2,N_{\mathrm{T}}}^2\geq c_\sigma^2\left(1-A_n\right),\ \text{for } 0\leq J\leq N_{\mathrm{s}},$$

$$P\Bigg\{\inf_{0\leq J\neq J'\leq N_{\mathrm{s}}}\Big\{N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\Big(\sum_{k=1}^{\kappa}R_{ik,\xi,J}^{2}+\sum_{j=1}^{N_i}R_{ij,\varepsilon,J}^{2}\Big)\Big\}$$
$$\times\Big\{N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\Big(\sum_{k=1}^{\kappa}R_{ik,\xi,J'}^{2}+\sum_{j=1}^{N_i}R_{ij,\varepsilon,J'}^{2}\Big)\Big\}\geq c_\sigma^4\Big(1-C_A\sqrt{\frac{\log(n)}{nh_{\mathrm{s}}}}\Big)^2\Bigg\}$$
$$\geq 1-2n^{-3},$$

by Lemma A.5. Thus for large n, with probability $\geq 1-2n^{-3}$, the numerator of $C_{n,J,J'}$ is uniformly greater than $c_\sigma^2/2$. Applying Bernstein's inequality to $N_{\mathrm{T}}^{-1}\left\{\sum_{k=1}^{\kappa}\sum_{i=1}^{n}R_{ik,\xi,J}R_{ik,\xi,J'}\right\}$, there exists $C_0>0$ such that, for large n,

$$P\Big(\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\Big|N_{\mathrm{T}}^{-1}\sum_{k=1}^{\kappa}\sum_{i=1}^{n}R_{ik,\xi,J}R_{ik,\xi,J'}\Big|\leq C_0h_{\mathrm{s}}\Big)\geq 1-2n^{-3}.$$

Putting the above together, for large n, $C_1=C_0\left(c_\sigma^2/2\right)^{-1}$,

$$P\Big(\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\Big|C_{n,J,J'}\Big|\leq C_1h_{\mathrm{s}}\Big)\geq 1-4n^{-3}.$$

Note that as a continuous random variable, $\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\left|C_{n,J,J'}\right|\in[0,1]$, thus

$$E\Big(\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\left|C_{n,J,J'}\right|\Big)=\int_0^1 P\Big(\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\left|C_{n,J,J'}\right|>t\Big)dt.$$

For large n, $C_1h_{\mathrm{s}}<1$ and then $E\left(\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\left|C_{n,J,J'}\right|\right)$ is

$$\int_0^{C_1h_{\mathrm{s}}}P\Big\{\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\left|C_{n,J,J'}\right|>t\Big\}dt+\int_{C_1h_{\mathrm{s}}}^1 P\Big\{\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\left|C_{n,J,J'}\right|>t\Big\}dt$$
$$\leq\int_0^{C_1h_{\mathrm{s}}}1dt+\int_{C_1h_{\mathrm{s}}}^1 4n^{-3}dt\leq C_1h_{\mathrm{s}}+4n^{-3}\leq Ch_{\mathrm{s}}$$

for some $C>0$ and large enough n. The lemma now follows from

$$\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\left|E\left(C_{n,J,J'}\right)\right|\leq E\Big(\sup_{0\leq J\neq J'\leq N_{\mathrm{s}}}\left|C_{n,J,J'}\right|\Big)\leq Ch_{\mathrm{s}}.$$

By Lemma A.8, the $(N_{\mathrm{s}}+1)$ standard normal variables $\eta_0,\dots,\eta_{N_{\mathrm{s}}}$ satisfy the conditions of Lemma A.2 Hence for any $\tau \in R$,

$$\lim_{n\to\infty} P\Big(\sup_{x\in[0,1]} |\eta(x)| \le \frac{\tau}{a_{N_{\mathrm{s}}+1}} + b_{N_{\mathrm{s}}+1}\Big) = \exp\left(-2e^{-\tau}\right). \tag{A.7}$$

For $x \in [0,1]$, $R_{ik,\xi,J}, R_{ij,\varepsilon,J}$ given in (A.2), define the ratio of population and sample quantities as $r_n(x) = \{nE(N_1)/N_{\mathrm{T}}\}^{1/2}\{\bar{R}_n(x)/\bar{R}(x)\}^{1/2}$, with

$$\bar{R}_n(x) = N_{\mathrm{T}}^{-1}\Big\{\sum_{i=1}^{n}\Big(\sum_{k=1}^{\kappa} R^2_{ik,\xi,J(x)} + \sum_{j=1}^{N_i} R^2_{ij,\varepsilon,J(x)}\Big)\Big\},$$
$$\bar{R}(x) = (EN_1)^{-1}\sum_{k=1}^{\kappa} ER^2_{1k,\xi,J(x)} + ER^2_{11,\varepsilon,J(x)}.$$

Lemma A.9. *Under Assumptions* (A2), (A3), *for* $\eta(x), \sigma_n(x)$ *in* (A.6), (2.3),

$$\left|\sigma_n(x)^{-1}\Big\{\sum_{k=1}^{\kappa}\widehat{\xi}_k(x) + \widehat{\varepsilon}^{(0)}(x)\Big\} - \eta(x)\right| = |r_n(x)-1|\,|\eta(x)| \tag{A.8}$$

as $n\to\infty$, $\sup_{x\in[0,1]}\{a_{N_{\mathrm{s}}+1}|r_n(x)-1|\} = O_{a.s.}\Big(\sqrt{\{\log(N_{\mathrm{s}}+1)\}(\log n)/(nh_{\mathrm{s}})}\Big)$.

Proof. Equation (A.8) follows from the definitions of $\eta(x)$ and $\sigma_n(x)$. By Lemma A.6, $\sup_{x\in[0,1]}|N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} R^2_{ij,\varepsilon,J(x)} - ER^2_{11,\varepsilon,J(x)}| = O_{\mathrm{a.s.}}(\sqrt{\log n/(nh_{\mathrm{s}})})$,

$$\begin{aligned}
&\sup_{x\in[0,1]}\left|N_{\mathrm{T}}^{-1}\sum_{k=1}^{\kappa}\sum_{i=1}^{n} R^2_{ik,\xi,J(x)} - (EN_1)^{-1}\sum_{k=1}^{\kappa} ER^2_{1k,\xi,J(x)}\right| \\
&\le \sup_{x\in[0,1]}(EN_1)^{-1}\sum_{k=1}^{\kappa}\left|n^{-1}\sum_{i=1}^{n} R^2_{ik,\xi,J(x)} - ER^2_{1k,\xi,J(x)}\right| \\
&\quad + \sup_{x\in[0,1]}(EN_1)^{-1}\sum_{k=1}^{\kappa}\left|n(EN_1)N_{\mathrm{T}}^{-1}-1\right|\left|n^{-1}\sum_{i=1}^{n} R^2_{ik,\xi,J(x)}\right| \\
&= O_{\mathrm{a.s.}}\left(\sqrt{\frac{\log n}{nh_{\mathrm{s}}}}\right) + O_{\mathrm{a.s.}}\left(n^{-1/2}\right) = O_{\mathrm{a.s.}}\left(\sqrt{\frac{\log n}{nh_{\mathrm{s}}}}\right),
\end{aligned}$$

and there exist constants $0 < c_{\bar{R}} < C_{\bar{R}} < \infty$ such that for all $x \in [0,1]$, $c_{\bar{R}} < \bar{R}(x) < C_{\bar{R}}$. Thus, $\sup_{x\in[0,1]}\left|\bar{R}_n(x) - \bar{R}(x)\right|$ is bounded by

$$\sup_{x\in[0,1]}\left|N_{\mathrm{T}}^{-1}\sum_{k=1}^{\kappa}\sum_{i=1}^{n} R^2_{ik,\xi,J(x)} - (EN_1)^{-1}\sum_{k=1}^{\kappa} ER^2_{1k,\xi,J(x)}\right|$$

$$+\sup_{x\in[0,1]}\left|N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}R_{ij,\varepsilon,J(x)}^2-ER_{11,\varepsilon,J(x)}^2\right|=O_{\mathrm{a.s.}}\left(\sqrt{\frac{\log n}{nh_{\mathrm{s}}}}\right).$$

Thus $\sup_{x\in[0,1]}\left|\left\{\bar{R}_n(x)\right\}^{1/2}-\left\{\bar{R}(x)\right\}^{1/2}\right|\le\sup_{x\in[0,1]}\left|\bar{R}_n(x)-\bar{R}(x)\right|\sup_{x\in[0,1]}$ $\left\{\bar{R}(x)\right\}^{-1/2}=O_{\mathrm{a.s.}}\left(\sqrt{\log n/(nh_{\mathrm{s}})}\right)$. Then $\sup_{x\in[0,1]}\left\{a_{N_{\mathrm{s}}+1}\left|r_n(x)-1\right|\right\}$ is bounded by

$$\begin{aligned}
&a_{N_{\mathrm{s}}+1}\left\{\left\{\frac{nE(N_1)}{N_{\mathrm{T}}}\right\}^{1/2}\sup_{x\in[0,1]}\left|\left\{\frac{\bar{R}_n(x)}{\bar{R}(x)}\right\}^{1/2}-1\right|+\left|1-\left\{\frac{nE\left(N_1\right)}{N_{\mathrm{T}}}\right\}^{1/2}\right|\right\}\\
&\le a_{N_{\mathrm{s}}+1}\left\{\left\{\frac{nE\left(N_1\right)}{N_{\mathrm{T}}}\right\}^{1/2}\sup_{x\in[0,1]}\left\{\bar{R}(x)\right\}^{-1/2}\sup_{x\in[0,1]}\left|\left\{\bar{R}_n(x)\right\}^{1/2}-\left\{\bar{R}(x)\right\}^{1/2}\right|\right.\\
&\qquad\left.+\left|1-\left\{\frac{nE\left(N_1\right)}{N_{\mathrm{T}}}\right\}^{1/2}\right|\right\}=O_{\mathrm{a.s.}}\left(\sqrt{\left\{\log\left(N_{\mathrm{s}}+1\right)\right\}\frac{\log n}{nh_{\mathrm{s}}}}\right).
\end{aligned}$$

Proof of Proposition 2. The proof follows from Lemmas A.5, A.7, A.9, (A.7), and Slutsky's Theorem.

Proof of Theorem 1. By Theorem 2, $\|\widetilde{m}(x)-m(x)\|_\infty=O_p\left(h_{\mathrm{s}}\right)$, so

$$a_{N_{\mathrm{s}}+1}\Big(\sup_{x\in[0,1]}\sigma_n^{-1}(x)\left|\widetilde{m}(x)-m(x)\right|\Big)=O_p\left\{(nh_{\mathrm{s}})^{1/2}\sqrt{\log\left(N_{\mathrm{s}}+1\right)}h_{\mathrm{s}}\right\}=o_p\left(1\right),$$

$$a_{N_{\mathrm{s}}+1}\Big(\sup_{x\in[0,1]}\sigma_n^{-1}(x)\left|\widehat{m}(x)-m(x)\right|-\sup_{x\in[0,1]}\sigma_n^{-1}(x)\Big|\sum_{k=1}^{\kappa}\widetilde{\xi}_k(x)+\widetilde{\varepsilon}(x)\Big|\Big)=o_p\left(1\right).$$

Meanwhile, (3.4) and Proposition 2 entail that, for any $\tau\in R$,

$$\lim_{n\to\infty}P\Big\{a_{N_{\mathrm{s}}+1}\Big(\sup_{x\in[0,1]}\sigma_n^{-1}(x)\Big|\sum_{k=1}^{\kappa}\widetilde{\xi}_k(x)+\widetilde{\varepsilon}(x)\Big|-b_{N_{\mathrm{s}}+1}\Big)\le\tau\Big\}=\exp\left(-2e^{-\tau}\right).$$

Thus Slutsky's Theorem implies that

$$\lim_{n\to\infty}P\Big\{a_{N_{\mathrm{s}}+1}\Big(\sup_{x\in[0,1]}\sigma_n^{-1}(x)\left|\widehat{m}(x)-m(x)\right|-b_{N_{\mathrm{s}}+1}\Big)\le\tau\Big\}=\exp\left(-2e^{-\tau}\right).$$

Let $\tau=-\log\left\{-\log\left(1-\alpha\right)/2\right\}$, definitions of $a_{N_{\mathrm{s}}+1},b_{N_{\mathrm{s}}+1}$, and $Q_{N_{\mathrm{s}}+1}\left(\alpha\right)$ in (2.4) entail

$$\begin{aligned}
&\lim_{n\to\infty}P\left\{m(x)\in\widehat{m}(x)\pm\sigma_n(x)Q_{N_{\mathrm{s}}+1}\left(\alpha\right),\forall x\in[0,1]\right\}\\
&\quad=\lim_{n\to\infty}P\Big\{Q_{N_{\mathrm{s}}+1}^{-1}\left(\alpha\right)\sup_{x\in[0,1]}\sigma_n^{-1}(x)\left|\widetilde{e}(x)+\widetilde{m}(x)-m(x)\right|\le1\Big\}=1-\alpha.
\end{aligned}$$

by (3.4). That $\sigma_n(x)^{-1}\left\{\widehat{m}(x)-m(x)\right\}\to_d N\left(0,1\right)$ for any $x\in[0,1]$ follows by directly using $\eta(x)\sim N\left(0,1\right)$, without reference to $\sup_{x\in[0,1]}\left|\eta(x)\right|$.

References

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer-Verlag, New York.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13**, 571-591.

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *J. Multivariate Anal.* **92**, 24-41.

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

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

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

Fan, J. and Zhang, W. Y. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27**, 715-731.

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

Hall, P. and Heckman, N. (2002). Estimating and depicting the structure of a distribution of random functions. *Biometrika* **89**, 145-158.

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

Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27**, 228-254.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.* **29**, 163-179.

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

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

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

Huang, X., Wang, L., Yang, L. and Kravchenko, A. N. (2008). Management practice effects on relationships of grain yields with topography and precipitation. *Agronomy J.* **100**, 1463-1471.

Izem, R. and Marron, J. S. (2007). Analysis of nonlinear modes of variation for functional data. *Electronic J. Statist.* **1**, 641-676.

James, G. M. (2002). Generalized linear models with functional predictors. *J. Roy. Statist. Soc. Ser. B* **64**, 411-432.

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

James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *J. Amer. Statist. Assoc.* **100**, 565-576.

James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* **98**, 397-408.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York.

Li, Y. and Hsing, T. (2007). On rates of convergence in functional linear regression. *J. Multivariate Anal.* **98**, 1782-1804.

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

Ma, S. and Yang, L. (2011). A jump-detecting procedure based on spline estimation. *J. Nonparametric Statist.* **23**, 67-81.

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

Müller, H. G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33**, 774-805.

Müller, H. G., Stadtmüller, U. and Yao, F. (2006). Functional variance processes. *J. Amer. Statist. Assoc.* **101**, 1007-1018.

Müller, H. G. and Yao, F. (2008). Functional additive models. *J. Amer. Statist. Assoc.* **103**, 1534-1544.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* 2nd edition. Springer; New York.

Song, Q. and Yang, L. (2009). Spline confidence bands for variance function. *J. Nonparametric Statist.* **21**, 589-609.

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

Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Statist.* **35**, 2474-2503.

Wang, L. and Yang, L. (2009). Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19**, 325-342.

Wu, W. and Zhao, Z. (2007). Inference of trends in time series. *J. Roy. Statist. Soc. Ser. B* **69**, 391-410.

Xue, L. and Yang, L. (2006). Additive coefficient modelling via polynomial spline. *Statist. Sinica* **16**, 1423-1446.

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

Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J. Multivariate Anal.* **98**, 40-56.

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

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

Zhang, J. T. and Chen, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35**, 1052-1079.

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

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

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

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics of regression splines and confidence regions. *Ann. Statist.* **26**, 1760-1782.

Department of Statistics, University of California, Riverside, 1418 Olmsted Hall, 900 University Avenue, Riverside, CA 92521, USA.

E-mail: shujie.ma@ucr.edu

Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou 215006, People's Republic of China.

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA.

E-mail: yanglijian@suda.edu.cn

Department of Statistics, Texas A&M University, College Station, TX 77843, USA.

E-mail: carroll@stat.tamu.edu

(Received February 2010; accepted August 2010)