A SIMULTANEOUS CONFIDENCE BAND FOR SPARSE LONGITUDINAL REGRESSION

Shujie Ma\textsuperscript{1}, Lijian Yang\textsuperscript{2,1} and Raymond J. Carroll\textsuperscript{3}

\textsuperscript{1}Michigan State University, \textsuperscript{2}Soochow University and \textsuperscript{3}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


According to Ferraty and Vieu (2006), a functional data set consists of iid realizations $\{\xi_i(x), x \in \mathcal{X}\}, 1 \leq i \leq n$, 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
The sequence 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), Yao, Müller, and Wang (2005b), 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, Wang, Yang, and Kravchenko (2008), Ma and Yang (2010), 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); Yao, Müller, and Wang (2005b); 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 \( \omega(r, \delta) = \max_{x, x' \in [a, b], |x - x'| \leq \delta} |r(x) - r(x')| \). 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 \( ||g||_{2,N_T}^2 = N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} g^2(X_{ij}) \), where we denote the total sample size by \( N_T = \sum_{i=1}^{n} N_i \).
Without loss of generality, we take the range of \( X, X = [a, b] \), to be \([0, 1] \). For any \( \beta \in (0, 1) \), we denote the collection of order \( \beta \) Hölder 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 \( \phi \). 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 \( \omega(\phi, \delta) \leq \| \phi \|_{0,\beta} \delta^\beta \).

To introduce the spline functions, divide the finite interval \([0, 1] \) into \((N_s + 1)\) equal subintervals \( \chi_J = [t_J, t_{J+1}], J = 0, ..., N_s = 1, \chi_{N_s} = [t_{N_s}, 1] \). A sequence of equally-spaced points \( \{ t_J \}_{J=1}^{N_s} \), called interior knots, are given as
\[ t_0 = 0 < t_1 < \cdots < t_{N_s} < 1 = t_{N_s+1}, t_J = Jh_s, \ 0 \leq J \leq N_s + 1, h_s = 1/ (N_s + 1), \]
in which \( h_s \) is the distance between neighboring knots. We denote by \( G^{(-1)} = G^{(-1)}[0, 1] \) the space of functions that are constant on each \( \chi_J \). For any \( x \in [0, 1] \), define its location index as \( J(x) = J_n(x) = \min \{ \lfloor x/h_s \rfloor, N_s \} \) so that \( t_{J_n(x)} \leq x < t_{J_n(x)+1}, \forall x \in [0, 1] \). We propose to estimate the mean function \( m(x) \) by
\[ \hat{m}(x) = \arg\min_{g \in G^{(-1)}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} (Y_{ij} - g(X_{ij}))^2. \quad (2.1) \]

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 \( 0 < c_f \leq C_f < \infty, 0 < c_\sigma \leq C_\sigma < \infty \).
(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, \(i = 1, 2, ..., \) with \(N_i \sim N\), where \(N > 0\) is a positive integer-valued random variable with \(E\{N^{2r}\} \leq r!C_N^r, r = 2, 3, ...\) 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 \(\kappa\) of nonzero eigenvalues is finite and the random coefficients \(\xi_{ik}, k = 1, ..., \kappa, i = 1, ..., \infty\) are iid \(N(0, 1)\). The variables \((N_i)_{i=1}^\infty, (\xi_{ik})_{i=1,k=1}^{\kappa,\infty}, (X_{ij})_{i=1,j=1}^{\infty,\infty}\), \((\varepsilon_{ij})_{i=1,j=1}^{\infty,\infty}\) are independent.

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

(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 \(\varepsilon_{ij}\).

We now introduce the B-spline basis of \(G^{(-1)}\), the space of piecewise constant splines, as \(\{b_j(x)\}_{j=0}^{N_0}\), which are simply indicator functions of intervals \(\chi_J\), \(b_j(x) = I_{\chi_J}(x)\), \(J = 0, 1, ..., N_s\). Define

\[
c_{j,n} = \|b_j\|_2^2 = \int_0^1 b_j(x)f(x)dx, J = 0, ..., N_s, \quad (2.2)
\]

\[
\sigma^2_Y(x) = \text{var}(Y|X = x) = G(x,x) + \sigma^2(x), \forall x \in [0, 1],
\]

\[
\sigma^2_n(x) = c_{j,n}^2\left\{nE(N_1)\right\}^{-1} \left\{\frac{E\{N_1(N_1 - 1)\}}{EN_1} \sum_{k=1}^\kappa \left(\int_{\chi_{J(x)}} \phi_k(u) f(u) du\right)^2 + \int_{\chi_{J(x)}} \sigma^2_Y(u) f(u) du\right\}. \quad (2.3)
\]

In addition, define \(Q_{N_s+1}(\alpha) = b_{N_s+1} - a_{N_s+1}^{-1}\log(-1/2)\log(1 - \alpha)\),

\[
a_{N_s+1} = \left\{2\log (N_s + 1)\right\}^{1/2}, b_{N_s+1} = a_{N_s+1} - \frac{\log\left(2\pi a_{N_s+1}^2\right)}{2a_{N_s+1}}, \quad (2.4)
\]
for any \( \alpha \in (0, 1) \). We now state our main results.

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

\[
\lim_{n \to \infty} P \left\{ \sup_{x \in [0, 1]} \left| \hat{m}(x) - m(x) \right| / \sigma_n(x) \leq Q_{N_s+1}(\alpha) \right\} = 1 - \alpha,
\]

\[
\lim_{n \to \infty} P \left\{ \left| \hat{m}(x) - m(x) \right| / \sigma_n(x) \leq Z_{1-\alpha/2} \right\} = 1 - \alpha, \quad \forall x \in [0, 1],
\]

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 \( n \to \infty \),

\[
\sup_{x \in [0, 1]} \left\{ \left| \sigma_n^{-1}(x) \sigma_n_{\text{IID}}(x) - 1 \right| + \left| \sigma_n^{-1}(x) \sigma_n_{\text{LONG}}(x) - 1 \right| \right\} = O \left( h^3_s \right),
\]

in which for \( x \in [0, 1] \), \( \sigma_n_{\text{IID}}(x) \equiv \sigma_Y(x) \{ f(x) h_s n E(N_1) \}^{-1/2} \) and

\[
\sigma_n_{\text{LONG}}(x) \equiv \sigma_n_{\text{IID}}(x) \left\{ 1 + \frac{E \{ N_1 (N_1 - 1) \} h_s G(x,x) f(x)}{EN_1 \sigma_Y^2(x)} \right\}^{1/2}.
\]

Using \( \sigma_n_{\text{IID}}(x) \) instead of \( \sigma_n(x) \) means to treat the \((X_{ij}, Y_{ij})\) as iid data rather than as sparse longitudinal data, while using \( \sigma_n_{\text{LONG}}(x) \) 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

\[
\hat{m}(x) \pm \sigma_n(x) Q_{N_s+1}(\alpha),
\]

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

### 3. Decomposition
In this section, we decompose the estimation error \( \hat{m}(x) - m(x) \) 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 \( \{ B_j(x) \}_{J=0}^{N_s} \) for \( G(1-1) \), which is

\[
B_j(x) = b_j(x) \left\{ c_{J,n} \right\}^{-1/2}, \quad J = 0, \ldots, N_s.
\]

It is easily verified that \( \| B_j \|_2^2 = 1, J = 0,1, \ldots, N_s, \langle B_j, B_{j'} \rangle = 0, J \neq J'. \)

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

\[
\hat{m}(x) = \sum_{j=0}^{N_s} \lambda_j b_j(x),
\]

with coefficients \( \{ \lambda_0, \ldots, \lambda_{N_s} \}^T \) as solutions of the least squares problem

\[
\left\{ \lambda_0, \ldots, \lambda_{N_s} \right\}^T = \arg\min_{\lambda_0, \ldots, \lambda_{N_s} \in R^{N_s+1}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{ Y_{ij} - \sum_{j=0}^{N_s} \lambda_j b_j(X_{ij}) \right\}^2.
\]

Simple linear algebra shows that \( \hat{m}(x) = \sum_{j=0}^{N_s} \lambda_j B_j(x) \), where the coefficients \( \{ \lambda_0, \ldots, \lambda_{N_s} \}^T \) are solutions of the least squares problem

\[
\left\{ \lambda_0, \ldots, \lambda_{N_s} \right\}^T = \arg\min_{\lambda_0, \ldots, \lambda_{N_s} \in R^{N_s+1}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{ Y_{ij} - \sum_{j=0}^{N_s} \lambda_j B_j(X_{ij}) \right\}^2.
\]

Projecting the relationship in model (1.2) onto the linear subspace of \( R^{N_r} \) spanned by \( \{ B_j(X_{ij}) \}_{1 \leq j \leq N_1, 1 \leq i \leq n, 0 \leq j \leq N_s} \), we obtain the following crucial decomposition in the space \( G(-1) \) of spline functions:

\[
\hat{m}(x) = \tilde{m}(x) + \tilde{e}(x) + \hat{m}(x) + \tilde{\varepsilon}(x) + \sum_{k=1}^{\kappa} \tilde{\xi}_k(x),
\]

\[
\tilde{m}(x) = \sum_{J=0}^{N_s} \lambda_j \tilde{B}_j(x), \quad \tilde{\varepsilon}(x) = \sum_{J=0}^{N_s} \hat{a}_j \tilde{B}_j(x),
\]

\[
\tilde{\xi}_k(x) = \sum_{J=0}^{N_s} \tilde{\tau}_{k,j} \tilde{B}_j(x).
\]

The vectors \( \{ \tilde{\lambda}_0, \ldots, \tilde{\lambda}_{N_s} \}^T, \{ \tilde{\hat{a}}_0, \ldots, \tilde{\hat{a}}_{N_s} \}^T, \) and \( \{ \tilde{\tau}_{k,0}, \ldots, \tilde{\tau}_{k,N_s} \}^T \) 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 \( \tilde{m}(x) \). 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 $\|g - \phi\|_{\infty} \leq C_g \omega (\phi, h_s)$. In particular, if $\phi \in C^{0, \beta} [0, 1]$ for some $\beta \in (0, 1]$, then $\|g - \phi\|_{\infty} \leq C_g \|\phi\|_{0, \beta} h_s^\beta$. Under Assumptions (A1) and (A4), with probability approaching 1, the function $\tilde{m}(x)$ defined in (3.3) satisfies $\|\tilde{m}(x) - m(x)\|_{\infty} = O (h_s)$.

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

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

$$
\lim_{n \to \infty} P \left\{ \sup_{x \in [0,1]} |\sigma_n(x)^{-1} \tilde{c}(x)| \leq \tau a_{N_s+1} + b_{N_s+1} \right\} = \exp \left( -2e^{-\tau} \right).
$$

4. Implementation

In this section, we describe procedures to implement the confidence bands and intervals given in Corollary 1. Given any data set $(X_{ij}, Y_{ij})^{N_i}_{j=1,i=1}$ from model (1.2), the spline estimator $\hat{m}(x)$ is obtained by (3.2), and the number of interior knots in (3.2) is taken to be $N_s = \lceil cN_T^{1/3} \log n \rceil$, in which $\lceil a \rceil$ 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 $f(x)$, $\sigma^2(x)$, 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 $f(x)$, $\sigma^2(x)$, and $G(x, x)$ is taken to be $N_s^* = \lceil n^{1/3} \rceil$, and $h_s^* = 1 / (1 + N_s^*)$. The histogram pilot estimator of the density function $f(x)$ is

$$
\hat{f}(x) = \left\{ \sum_{i=1}^n \sum_{j=1}^{N_i} b_{j(x)} (X_{ij}) / (\sum_{i=1}^n N_i) h_s^* \right\}.
$$

Defining the vector $R = \{ R_{ij} \}_{1 \leq j \leq N_i, 1 \leq i \leq n} = \left\{ (Y_{ij} - \hat{m}(X_{ij}))^2 \right\}_{1 \leq j \leq N_i, 1 \leq i \leq n}^T$, the estimator of $\sigma_n^2(x)$ is $\hat{\sigma}_n^2(x) = \sum_{j=0}^{N_s^*} \hat{\rho}_j b_j(x)$, where the coefficients $\{ \hat{\rho}_0, ..., \hat{\rho}_{N_s^*} \}^T$ are solutions of the least squares problem:

$$
\left\{ \hat{\rho}_0, ..., \hat{\rho}_{N_s^*} \right\}^T = \arg \min_{\{ \hat{\rho}_0, ..., \hat{\rho}_{N_s^*} \} \in \mathbb{R}^{N_s+1}} \sum_{i=1}^n \sum_{j=1}^{N_i} \left\{ R_{ij} - \sum_{j=0}^{N_s^*} \hat{\rho}_j b_j (X_{ij}) \right\}^2.
$$

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

$$
\hat{G} (x, x') = \arg \min_{g \in G^{(-1)} \otimes G^{(-1)}} \sum_{i=1}^n \sum_{j,j'=1}^{N_i} \{ C_{ijj'} - g (X_{ij}, X_{ij'}) \}^2,
$$
where \( C_{ijj'} = \{ Y_{ij} - \hat{m}(X_{ij}) \} \{ Y_{ij'} - \hat{m}(X_{ij'}) \}, 1 \leq j, j' \leq N_i, 1 \leq i \leq n \). The function \( \sigma_n(x) \) is estimated by either \( \hat{\sigma}_{n,\text{IID}}(x) \equiv \hat{\sigma}_Y(x) \left\{ \hat{f}(x)h_s N_T \right\}^{-1/2} \) or

\[
\hat{\sigma}_{n,\text{LONG}}(x) \equiv \hat{\sigma}_{n,\text{IID}}(x) \left\{ 1 + \left( \sum_{i=1}^{n} N_i^2 / N_T - 1 \right) \frac{\hat{G}(x, x) - \hat{f}(x)h_s}{\sigma_Y^2(x)} \right\}^{1/2}.
\]

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 \( n \to \infty \)

\[
\sup_{x \in [0, 1]} \left\{ \left| \hat{\sigma}_{n,\text{IID}}(x) \sigma_{n,\text{IID}}^{-1}(x) - 1 \right| + \left| \hat{\sigma}_{n,\text{LONG}}(x) \sigma_{n,\text{LONG}}^{-1}(x) - 1 \right| \right\} = O_{a.s.} \left( h_s^3 + n^{-1/2}N^{-1} \right).
\]

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

\[
\hat{m}(x) \pm \hat{\sigma}_{n,\text{IID}}(x) Q_{N_n+1}(\alpha), \tag{4.1}
\]

\[
\hat{m}(x) \pm \hat{\sigma}_{n,\text{LONG}}(x) Q_{N_n+1}(\alpha), \tag{4.2}
\]

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

**5. Simulation**

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

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

with \( X \sim \text{Uniform}[0, 1], \xi_k \sim \text{Normal}(0, 1), k = 1, 2, \varepsilon \sim \text{Normal}(0, 1), \) \( N_i \) having a discrete uniform distribution from 25, \( \cdots, 35, \) for \( 1 \leq i \leq n, \) 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 \( N_n \) in Section 4 was taken to be...
We found that the confidence band (4.1) did not have good coverage rates for moderate sample sizes, and hence in Table [4] 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 \( \{k/100, k = 0, \ldots, 100\} \).

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.

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>( n )</th>
<th>( 1 - \alpha )</th>
<th>( c = 1 )</th>
<th>( c = 2 )</th>
<th>( c = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>20</td>
<td>0.950</td>
<td>0.920</td>
<td>0.930</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.950</td>
<td>0.960</td>
<td>0.965</td>
<td>0.910</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>0.995</td>
<td>0.995</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.950</td>
<td>0.955</td>
<td>0.955</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>1.000</td>
<td>1.000</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.950</td>
<td>0.950</td>
<td>0.965</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>0.985</td>
<td>0.985</td>
<td>0.990</td>
</tr>
<tr>
<td>1.0</td>
<td>20</td>
<td>0.950</td>
<td>0.935</td>
<td>0.930</td>
<td>0.735</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.870</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.950</td>
<td>0.975</td>
<td>0.960</td>
<td>0.895</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>0.995</td>
<td>0.995</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.950</td>
<td>0.950</td>
<td>0.940</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>0.995</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.950</td>
<td>0.940</td>
<td>0.965</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.990</td>
<td>0.985</td>
<td>0.995</td>
<td>0.995</td>
</tr>
</tbody>
</table>

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 \( 1 \leq i \leq n \), 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 \( \hat{m}(x) \) and the confi-
dence intervals are step functions that remain the same on each subinterval \( \chi_J \), \( 0 \leq J \leq N_s \). Following 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 \((N_s + 1)\) of subintervals as fixed instead of growing to infinity, a naive parametric confidence band was computed as

\[
\hat{m}(x) \pm \hat{\sigma}_{n,\text{LONG}}(x) Q_{1-\alpha N_s+1}
\]  

in which \( Q_{1-\alpha N_s+1} = Z \{1+(1-\alpha)^{1/(N_s+1)}\}/2 \) is the \((1-\alpha)\) quantile of the maximal absolute values of \((N_s + 1)\) iid \( N(0,1) \) random variables. We compare the performance of the confidence band in (4.2), the smoothed band and naive parametric band in (5.1). Given \( n = 20 \) with \( N_s = 8, 12 \), and \( n = 50 \) \( N_s = 44 \) (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 \( \hat{P}, \hat{P}_{\text{naive}}, \hat{P}_{\text{smooth}} \) and the average maximal widths \( W, W_{\text{naive}}, W_{\text{smooth}} \) of \( N_s + 1 \) intervals out of 200 replications calculated from confidence bands (4.2), (5.1), and the smoothed confidence bands, respectively.

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 \( 1 - \alpha = 0.99 \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \sigma )</th>
<th>( N_s )</th>
<th>( P )</th>
<th>( P_{\text{naive}} )</th>
<th>( P_{\text{smooth}} )</th>
<th>( W )</th>
<th>( W_{\text{naive}} )</th>
<th>( W_{\text{smooth}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.5</td>
<td>8</td>
<td>0.820</td>
<td>0.505</td>
<td>0.910</td>
<td>1.490</td>
<td>1.210</td>
<td>1.480</td>
</tr>
<tr>
<td>20</td>
<td>0.5</td>
<td>12</td>
<td>0.930</td>
<td>0.765</td>
<td>0.955</td>
<td>1.644</td>
<td>1.363</td>
<td>1.628</td>
</tr>
<tr>
<td>1.0</td>
<td>8</td>
<td>0.910</td>
<td>0.655</td>
<td>0.970</td>
<td>1.725</td>
<td>1.401</td>
<td>1.721</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>12</td>
<td>0.960</td>
<td>0.820</td>
<td>0.985</td>
<td>1.937</td>
<td>1.606</td>
<td>1.928</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>44</td>
<td>0.990</td>
<td>0.960</td>
<td>0.990</td>
<td>1.651</td>
<td>1.522</td>
<td>1.609</td>
</tr>
<tr>
<td>1.0</td>
<td>44</td>
<td>0.990</td>
<td>0.975</td>
<td>1.000</td>
<td>2.054</td>
<td>1.893</td>
<td>2.016</td>
<td></td>
</tr>
</tbody>
</table>

In all experiments, one has \( \hat{P}_{\text{smooth}} > \hat{P} > \hat{P}_{\text{naive}} \) and \( W > W_{\text{smooth}} > W_{\text{naive}} \). 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.
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, 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 $N_s = 6$, 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 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 band.
Table 3: The mean CD4+ cell counts and the confidence limits on each subinterval at simultaneous confidence level 0.95.

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

intervals in thin lines. The center thick line is the piecewise-constant spline fit \( \hat{m}(x) \). 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 confi-
dence 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.

Acknowledgment

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 Csörgő and Révész (1981)) Suppose that \( \xi_i, 1 \leq i \leq n \) are iid with \( E(\xi_1) = 0, E(\xi_1^2) = 1 \), and \( H(x) > 0 \) for \( x \geq 0 \) 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 \( \{W(t), 0 \leq t < \infty\} \) that is a Borel function of \( \xi_i, 1 \leq i \leq n \), and constants \( C_1, C_2, a > 0 \) which depend only on the distribution of \( \xi_1 \), such that for any \( \{x_n\}_{n=1}^{\infty} \) satisfying \( H^{-1}(n) < x_n < C_1 (\log n)^{1/2} \) and \( S_k = \sum_{i=1}^{k} \xi_i \),

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

Lemma A.2. Let \( \xi_i^{(n)}, 1 \leq i \leq n, \) 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 \( \gamma > 0, C_\gamma > 0, \) \( |r_{ij}^{(n)}| < C_\gamma/n^\gamma, i \neq j \). Then for \( \tau \in \mathbb{R} \), as \( n \to \infty \), \( P\{M_{n,\xi} \leq \tau/a_n + b_n \} \to \exp(-2e^{-\tau}) \), in which \( M_{n,\xi} = \max \{\xi_1^{(n)}, \ldots, \xi_n^{(n)}\} \) and \( a_n, b_n \) are as in (2.7) with \( N_1 \) replaced by \( n \).

Proof. Let \( \{\eta_i\}_{i=1}^{n} \) be i.i.d. standard normal r.v.’s, \( u = \{u_i\}_{i=1}^{n}, v = \{v_i\}_{i=1}^{n} \) 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),

\[
P\left( \max_{1 \leq i < j \leq n} |v_j - \eta_j| \leq u_j \right)
\leq \frac{4}{2\pi} \sum_{1 \leq i < j \leq n} \left| r_{ij}^{(n)} \right| \left( 1 - \left| r_{ij}^{(n)} \right|^2 \right)^{-1/2} \exp \left( \frac{-u^2}{1 + r_{ij}^{(n)}^2} \right).
\]

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 \( 1 - r_{ij}^{(n)} \geq 1 - (C_\gamma/n^\gamma)^2 \to 1 \) as \( n \to \infty \), \( i \neq j \), \( \exists C_{r,2} > 0 \) such that \( 1 - r_{ij}^{(n)^2} \geq C_{r,2} > 0 \) and \( 1 + r_{ij}^{(n)} < 1 + \varepsilon \) for any \( \varepsilon > 0 \) and large \( n \). Let \( M_{n,\xi} = \max \{|\eta_1|, \ldots, |\eta_n|\} \). By Leadbetter, Lindgren and Rootzén (1983), Theorem 1.5.3, \( P\{M_{n,\xi} \leq \tau_n \} \to \exp(-2e^{-\tau}) \) as \( n \to \infty \), while the above results entail

\[
P\left( \max_{1 \leq i \leq n} |S_n - W(n)| \leq \tau_n \right)
\leq \frac{4}{2\pi} \sum_{1 \leq i < j \leq n} C_{r,2} \gamma n^{-\gamma} C_{r,2}^{1/2} \exp \left( \frac{-\gamma^2}{1 + \varepsilon} \right) \leq C_{r,2}^2 n^{-\gamma - (2-\varepsilon)(1+\varepsilon)^{-1}} \to 0
\]
as \( n \to \infty \). Hence \( P\left( \max_{1 \leq i \leq n} |S_n - W(n)| \leq \tau_n \right) \to \exp(-2e^{-\tau}) \), 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 \( r = 3, 4, \ldots, E|\xi_1|^r \leq c e^{-2r} E\xi_1^2 \to \infty \). Then for each \( n > 1, t > 0 \), \( P(|S_n| \geq \sqrt{n} \sigma t) \leq 2 \exp \left( -t^2 \left( 4 + 2c t / \sqrt{n} \sigma \right)^{-1} \right) \), 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), \( c_{J,n} = f(t_j) h_s(1 + r_{J,n}) \), \( \langle b_j, b_{J'} \rangle \equiv 0, J \neq J' \), where \( \max_{0 \leq j \leq N_s} |r_{J,n}| \leq C \omega(f, h_s) \). There exist constants \( C_B > c_B > 0 \) such that \( c_B h_s^{1-r/2} \leq E \{ B_J(X_{ij}) \}^r \leq C_B h_s^{1-r/2} \) for \( r = 1, 2, \ldots \) and \( 1 \leq J \leq N_s + 1, 1 \leq j \leq N_i, 1 \leq i \leq n \).

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

\[
c_{J,n} = \int b_j(x)f(x)dx = \int_{[t_j, t_{j+1}]} f(x)dx = f(t_j) h_s + \int_{[t_j, t_{j+1}]} \{ f(x) - f(t_j) \} dx.
\]

Hence for all \( J = 0, \ldots, N_s, c_{J,n} - f(t_j) h_s | \leq \int_{[t_j, t_{j+1}]} \{ f(x) - f(t_j) \} dx \leq \omega(f, h_s) h_s \), or \( |r_{J,n}| = |c_{J,n} - f(t_j) h_s| / \{ f(t_j) h_s \} \) \( \omega(f, h_s) \), \( J = 0, \ldots, N_s \). By (3.1),

\[
E \{ B_J(X_{ij}) \}^r = (c_{J,n})^{-r/2} \int b_j(x)f(x)dx = (c_{J,n})^{1-r/2} \sim h_s^{1-r/2}.
\]

**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 \( x \in [0, 1] \)

\[
\left| \int_{\chi_{J(x)}} \phi_k(x)f(x)du - \int_{\chi_{J(x)}} \phi_k(u)f(u)du \right| \leq \omega(\phi_k f, h_s) = O(h_s^{1+\beta}),
\]

\[
\left| \int_{J(x)} \{ \sigma_Y^2(x)f(x) - \sigma_Y^2(u)f(u) \} du \right| \leq \omega(\sigma_Y^2 f, h_s) = O(h_s^{1+\beta}).
\]

Hence,

\[
\sigma_n^2(x) = c_{J(x),n}(nEN_1)^{-1} \int_{J(x)} \sigma_Y^2(u)f(u)du \times
\]

\[
\left\{ 1 + E \{ N_1 (N_1 - 1) \} \sum_{k=1}^{\infty} \left( \int_{\chi_{J(x)}} \phi_k(u)f(u)du \right)^2 \left\{ \int_{J(x)} \sigma_Y^2(u)f(u)du \right\}^{-1} \right\}
\]

\[
= \left\{ f(x)h_s + U(h_s^{1+\beta}) \right\}^{-2} (nEN_1)^{-1} \left\{ \sigma_Y^2(x)f(x)h_s + U(h_s^{1+\beta}) \right\} \times
\]

\[
\left\{ 1 + E \{ N_1 (N_1 - 1) \} \sum_{k=1}^{\infty} \phi_k(x)f(x)h_s + U(h_s^{1+\beta}) \right\}^2 \left\{ \sigma_Y^2(x)f(x)h_s + U(h_s^{1+\beta}) \right\}^{-1}
\]

\[
= (f(x)h_n nEN_1)^{-1} \sigma_Y^2(x) \left\{ 1 + E \{ N_1 (N_1 - 1) \} \sum_{k=1}^{\infty} \phi_k^2(x)f(x)h_s \right\} \sigma_Y^2(x) \{ 1 + U(h_s^{1+\beta}) \}
\]

\[
= \sigma_n^2 \text{, LONG}(x) \left\{ 1 + U(h_s^{1+\beta}) \right\} = \sigma_n^2 \text{, IID}(x) \left\{ 1 + U(h_s^{1+\beta}) \right\} \Box
\]

A.2. Proof of Theorem 1
Note that $B_{J}(x) \equiv c_{J(x),n}^{-1/2}$, $x \in [0, 1]$, so the terms $\tilde{\xi}_{k}(x)$ and $\tilde{\varepsilon}(x)$ defined in (3.5) are

\[
\tilde{\xi}_{k}(x) = \sum_{j=0}^{N_{k}} N_{T}^{-1} B_{J}(x) \| B_{J} \|_{2,N_{T}}^{-2} \sum_{i=1}^{N_{i}} B_{J}(X_{ij}) \phi_{k}(X_{ij}) \xi_{ik}
\]

\[
= c_{J(x),n}^{-1/2} \| B_{J} \|_{2,N_{T}}^{-2} N_{T}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_{i}} B_{J}(X_{ij}) \phi_{k}(X_{ij}) \xi_{ik},
\]

\[
\tilde{\varepsilon}(x) = c_{J(x),n}^{-1/2} \| B_{J} \|_{2,N_{T}}^{-2} N_{T}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_{i}} B_{J}(X_{ij}) \sigma(X_{ij}) \varepsilon_{ij}.
\]

Let

\[
\tilde{\tilde{\xi}}_{k}(x) = \| B_{J} \|_{2,N_{T}}^{-2} \tilde{\xi}_{k}(x) = c_{J(x),n}^{-1/2} N_{T}^{-1} \sum_{i=1}^{n} R_{ik,\xi,J(x)}(x) \xi_{ik},
\]

\[
\tilde{\varepsilon}(x) = \| B_{J} \|_{2,N_{T}}^{-2} \tilde{\varepsilon}(x) = c_{J(x),n}^{-1/2} N_{T}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_{i}} R_{ij,\varepsilon,J} \varepsilon_{ij},
\]

(8.1)

where

\[
R_{ik,\xi,J} = \sum_{j=1}^{N_{i}} B_{J}(X_{ij}) \phi_{k}(X_{ij}), R_{ij,\varepsilon,J} = B_{J}(X_{ij}) \sigma(X_{ij}), 0 \leq J \leq N_{a}.
\]

(8.2)

Lemma A.5. Under Assumption (A3), for $\tilde{\varepsilon}(x)$ given in (3.4) and $\tilde{\xi}_{k}(x)$, $\tilde{\varepsilon}(x)$ given in (8.1), we have

\[
\left| \tilde{\varepsilon}(x) - \left( \sum_{k=1}^{K} \tilde{\xi}_{k}(x) + \tilde{\varepsilon}(x) \right) \right| \leq A_{n} (1 - A_{n})^{-1} \left| \sum_{k=1}^{K} \tilde{\xi}_{k}(x) + \tilde{\varepsilon}(x) \right|, x \in [0, 1],
\]

where $A_{n} = \sup_{0 \leq J \leq N_{a}} \| B_{J} \|_{2,N_{T}}^{-2}$. There exists $C_{A} > 0$, such that for large $n$,

$P \left( A_{n} \geq C_{A} \sqrt{\log(n) / (n b_{3})} \right) \leq 2n^{-3}$. $A_{n} = O_{a.s.} \left( \sqrt{\log(n) / (n b_{3})} \right)$ 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 $R_{ik,\xi,J}$, $R_{11,\varepsilon,J}$ in (8.2),

\[
ER_{1k,\xi,J}^{2} = c_{J,n}^{-1} \left[ E \left( N_{i} \right) \int b_{J}(u) \phi_{k}^{2}(u) f(u) du + E \left( N_{i} \right) \left( \int b_{J}(u) \phi_{k}(u) f(u) du \right)^{2} \right],
\]

\[
ER_{11,\varepsilon,J}^{2} = c_{J,n}^{-1} \int b_{J}(u) \sigma^{2}(u) f(u) du, 0 \leq J \leq N_{a},
\]
there exist $0 < c_R < C_R < \infty$, such that $ER_{1k,\xi,J}^2, ER_{11,\xi,J}^2 \in [c_R, C_R]$ for $0 \leq J \leq N$, $\sup_{0 \leq J \leq N} \left| R_{1k,\xi,J}^2 - ER_{1k,\xi,J}^2 \right| = O_{a.s.} \left( \sqrt{\log n / (nh_n)} \right)$, $1 \leq k \leq \kappa$, $\sup_{0 \leq J \leq N} \left| N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} R_{ij,\xi,J}^2 - ER_{11,\xi,J}^2 \right| = O_{a.s.} \left( \sqrt{\log n / (nh_n)} \right)$ as $n \to \infty$.

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

\[
ER_{1k,\xi,J}^2 = E \left\{ \sum_{j', j'' = 1}^{N_1} E \left\{ B_J (X_{ij}) B_J (X_{ij'}) \phi_k (X_{ij}) \phi_k (X_{ij'}) \right\} \right\} \\
= E \left\{ \sum_{j = 1}^{N_1} E \left\{ B_J^2 (X_{ij}) \phi_k^2 (X_{ij}) \right\} \right\} + \sum_{j \neq j'} E \left\{ B_J (X_{ij}) B_J (X_{ij'}) \phi_k (X_{ij}) \phi_k (X_{ij'}) \right\} \\
= c_{J(x),n}^{-1} \left\{ E (N_1) \int b_J (u) \phi_k^2 (u) f (u) du + \right\} \left\{ E (N_1 (N_1 - 1)) \left( \int b_J (u) \phi_k (u) f (u) du \right)^2 \right\}.
\]

It is easily shown that $\exists 0 < c_R < C_R < \infty$ such that $c_R \leq ER_{1k,\xi,J}^2 \leq C_R, 0 \leq J \leq N$.

Let $\zeta_{i,J} = \zeta_{i,1,J} = R_{1k,\xi,J}^2, \zeta_{i,J} = \zeta_{i,1,J} - E (\zeta_{i,1,J})$ for $r \geq 1$ and large $n$,

\[
E (\zeta_{i,J})^r = E \left\{ \sum_{j = 1}^{N_1} B_J (X_{ij}) \phi_k (X_{ij}) \right\}^{2r} \leq C_2^{2r} E \left\{ \sum_{j = 1}^{N_1} B_J (X_{ij}) \right\}^{2r} \\
= C_2^{2r} E \left\{ \sum_{\nu \leq \nu_1, \cdots, \nu_{N_1} \leq 2r} \left( \prod_{j = 1}^{N_1} E \left\{ B_J (X_{ij}) \right\} \right)^{\nu_j} \right\} \\
\leq C_2^{2r} E \left\{ N_1^{r \nu} \prod_{j = 1}^{N_1} E \left\{ B_J (X_{ij}) \right\}^{\nu_j} \right\} \leq C_2^{2r} \left( EN_1^{r} \right) C_B h_n^{1-r} \\
\leq C_2^{2r} C_B C_N^{r} h_n^{1-r} = C_C^{r \nu} h_n^{1-r},
\]

\[
E (\zeta_{i,J})^r \geq c_2^{2r} E \left\{ \sum_{j = 1}^{N_1} E \left\{ B_J (X_{ij}) \right\}^{2r} \right\} \geq c_2^{2r} \left( EN_1 \right) c_B h_n^{1-r},
\]

by Lemma A.4. So $\left\{ E (\zeta_{1,J}) \right\}^r \sim 1, E (\zeta_{i,J})^r \gg \left\{ E (\zeta_{1,J}) \right\}^2$ for $r \geq 2$, and $\exists C'_r > c'_r > 0$ such that $C'_r h_n^{1-r} \geq \sigma_r^{2r} \geq c'_r h_n^{1-r},$ for $\sigma_{C_r} = \left\{ E (\zeta_{1,J})^2 \right\}^{1/2}$. We obtain $E |\zeta_{1,J}|^r \leq c_2^{2r - 2r!} E (\zeta_{1,J})^2$ with $c_2 = \left( C_C c_{\xi} \right)^{-1} h_n^{1-r},$ which implies that $\left\{ \zeta_{i,J} \right\}^n_{i=1}$ satisfies Cramér’s condition. Applying Lemma A.3 to $\sum_{i=1}^{n} \zeta_{i,J},$ for $r > 2$ and any large
enough \( \delta > 0 \), \( P \left\{ n^{-1} | \sum_{i=1}^{n} \zeta_{i,j}^* | \geq \delta \sqrt{\log n / (n h_n)} \right\} \) is bounded by

\[
2 \exp \left\{ \frac{-\delta^2 \left( c'_\zeta \right)^{-1} \left( \log n \right)}{4 + 2 \left( c'_\zeta / c'_\zeta \right)^{1/2} \delta \left( c'_\zeta \right)^{-1} h_n^{1/2} \left( \log n \right)^{1/2} n^{-1/2}} \right\} 
\leq 2 \exp \left\{ \frac{-\delta^2 \left( \log n \right)}{4 C''_\zeta} \right\} \leq 2n^{-3}.
\]

Hence

\[
\sum_{n=1}^{\infty} P \left\{ \sup_{0 \leq j \leq N_n} \left| \frac{1}{n} \sum_{i=1}^{n} R^2_{i k, \xi, J} - ER^2_{i k, \xi, J} \right| \geq \delta \sqrt{\log n / (n h_n)} \right\} \leq \sum_{n=1}^{\infty} \frac{2N_n}{n^3} < \infty.
\]

Thus, \( \sup_{0 \leq j \leq N_n} \left| n^{-1} \sum_{i=1}^{n} R^2_{i k, \xi, J} - ER^2_{i k, \xi, J} \right| = O_{a.s.} \left( \sqrt{\log n / (n h_n)} \right) \) as \( n \to \infty \) by Borel-Cantelli Lemma. The properties of \( R_{i j, \xi, J} \) are obtained similarly. \( \square \)

Order all \( X_{ij}, 1 \leq j \leq N_i, 1 \leq i \leq n \) from large to small as \( X_{(t)}, X_{(1)} \geq \cdots \geq X_{(N_i)} \), and denote the \( \varepsilon_{ij} \) corresponding to \( X_{(t)} \) as \( \varepsilon_{(t)} \) By (8.1),

\[
\hat{\varepsilon}(x) = c_{j(x), n}^{-1} N_T^{-1} \sum_{t=1}^{N_T} b_{j(x)} \left( X_{(t)} \right) \sigma \left( X_{(t)} \right) \varepsilon_{(t)}
= c_{j(x), n}^{-1} N_T^{-1} \sum_{t=1}^{N_T} b_{j(x)} \left( X_{(t)} \right) \sigma \left( X_{(t)} \right) \left\{ S_t - S_{t-1} \right\},
\]

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

**Lemma A.7.** Under Assumptions (A2)-(A5), there is a Wiener process \( \{W(t) : 0 \leq t < \infty\} \)

independent of \( \{N_i, X_{ij}, 1 \leq j \leq N_i, \xi_{ik}, 1 \leq k \leq \kappa, 1 \leq i \leq n\} \), such that as \( n \to \infty \),

\[
\sup_{x \in [0,1]} \left| \hat{\varepsilon}(0)(x) - \hat{\varepsilon}(x) \right| = o_{a.s.} \left( n^q \right) \text{ for some } t < -(1-\theta)/2 < 0, \text{ where } \hat{\varepsilon}(0)(x) \text{ is}
\]

\[
\left( c_{j(x), n} N_T \right)^{-1} \sum_{t=1}^{N_T} b_{j(x)} \left( X_{(t)} \right) \sigma \left( X_{(t)} \right) \left\{ W(t) - W(t-1) \right\}, x \in [0,1]. \quad (8.3)
\]

**Proof.** Define \( M_{N_T} = \max_{1 \leq \xi \leq N_T} |S_q - W(q)| \), in which \( \{W(t), 0 \leq t < \infty\} \)

is the Wiener process as in Lemma A.1 that as a Borel function of the set of variables \( \{\varepsilon_{(t)} : 1 \leq t \leq N_T\} \) is independent of \( \{N_i, X_{ij}, 1 \leq j \leq N_i, \xi_{ik}, 1 \leq k \leq \kappa, 1 \leq i \leq n\} \) since \( \{\varepsilon_{(t)} : 1 \leq t \leq N_T\} \) is. Further,

\[
\sup_{x \in [0,1]} \left| \hat{\varepsilon}(0)(x) - \hat{\varepsilon}(x) \right| = \sup_{x \in [0,1]} c_{j(x), n} N_T^{-1} \left| b_{j(x)} \left( X_{(N_T)} \right) \sigma \left( X_{(N_T)} \right) \left\{ W(N_T) - S_{N_T} \right\} \right.
\]

\[
+ \sum_{t=1}^{N_T-1} \left\{ b_{j(x)} \left( X_{(t)} \right) \sigma \left( X_{(t)} \right) - b_{j(x)} \left( X_{(t+1)} \right) \sigma \left( X_{(t+1)} \right) \right\} \left\{ W(t) - S_t \right\}
\leq \max_{0 \leq j \leq N_n+1} c_{j(x)} N_T^{-1} \left\{ b_{j} \left( X_{(N_T)} \right) \sigma \left( X_{(N_T)} \right) + \sum_{t=1}^{N_T-1} \left| b_{j} \left( X_{(t)} \right) \sigma \left( X_{(t)} \right) - b_{j} \left( X_{(t+1)} \right) \sigma \left( X_{(t+1)} \right) \right| \right\} M_{N_T}
\]
\[
\leq \max_{0 \leq J \leq N_n + 1} c_{J,n}^{-1} N_T^{-1} M_{N_T} \left\{ 3C_{\sigma} + \sum_{1 \leq t \leq N_T-1, X(t) \in \mathbb{B}_j} |\sigma(X(t)) - \sigma(X(t+1))| \right\}
\]

which, by the H"older continuity of \( \sigma \) in Assumption (A2), is bounded by

\[
N_T^{-1} M_{N_T} \max_{0 \leq J \leq N_n + 1} c_{J,n}^{-1} \left\{ 3C_{\sigma} + \|\sigma\|_{0,\beta} \left( \sum_{1 \leq t \leq N_T-1, X(t) \in \mathbb{B}_j} |X(t) - X(t+1)|^\beta \right) \right\} \leq N_T^{-1} M_{N_T} \max_{0 \leq J \leq N_n + 1} c_{J,n}^{-1} \left\{ 3C_{\sigma} + \|\sigma\|_{0,\beta} \left( \sum_{1 \leq t \leq N_T-1, X(t) \in \mathbb{B}_j} |X(t) - X(t+1)| \right)^\beta \right\}
\]

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

\[
\sup_{x \in [0,1]} |\tilde{\xi}(x) - \tilde{\xi}(x)| = O_{a.s.} \left( N_{n^{r'-1}} \left( 1 + N_n^{-\beta} \left( N_T N_n^{-1}\right)^{-1-\beta} \right) \right)
\]

\[
= O_{a.s.} \left( N_{n^{r'-1}} + N_{n^{r'-1}} N_n^{-1} n^{1-\beta} \right)
\]

\[
= O_{a.s.} \left( N_{n^{r'-1}} + N_{n^{r'-\beta}} \right) = o_{a.s.} \left( n^{\beta - \beta + \vartheta} \right)
\]

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

Now

\[
\tilde{\xi}(0)(x) = c_{J(x),n}^{-1} N_T^{-1} \sum_{t=1}^{N_T} b_{J(x)}(X(t)) \sigma(X(t)) Z(t)
\]

\[
= c_{J(x),n}^{-1} N_T^{-1} \sum_{i=1}^{N_i} \sum_{j=1}^{N_i} b_{J(x)}(X_{ij}) \sigma(X_{ij}) Z_{ij}, \tag{8.4}
\]

where \( Z(t) = W(t) - W(t-1), 1 \leq t \leq N_T, \) are i.i.d \( \mathcal{N}(0,1) \), \( \xi_k, \) \( Z_{ij}, X_{ij}, N_i \) are independent, for \( 1 \leq k \leq \kappa, 1 \leq j \leq N_i, 1 \leq i \leq n, \) and \( \tilde{\xi}_k(x), \tilde{\xi}_0(x) \) are conditional independent of \( X_{ij}, N_i, 1 \leq j \leq N_i, 1 \leq i \leq n. \) If the conditional variances of \( \tilde{\xi}_k(x), \tilde{\xi}_0(x) \)
on \((X_{ij}, N_i)_{1 \leq j \leq N_i, 1 \leq i \leq n}\) are \(\sigma_{\epsilon, n}^2(x), \sigma_{\varepsilon, n}^2(x)\), we have

\[
\sigma_{\varepsilon, n}(x) = \left\{ \epsilon_{J(x), n}^{-1} N^{-2} \sum_{i=1}^{n} R_{ik, \xi, J(x)} \right\}^{1/2},
\]

where \(R_{ik, \xi, J(x)}, R_{ij, \varepsilon, J(x)}, \) and \(c_{J(x), n}\) are given in (8.2) and (2.2).

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

\[
\eta(x) = \left\{ \sum_{k=1}^{K} \sigma_{\varepsilon, n}^2(x) + \sigma_{\varepsilon, n}^2(x) \right\}^{-1/2} \left\{ \sum_{k=1}^{K} \tilde{z}_k(x) + \tilde{z}^{(0)}(x) \right\},
\]

with \(\sigma_{\varepsilon, n}(x), \sigma_{\varepsilon, n}(x), \tilde{z}_k(x), \tilde{z}^{(0)}(x)\), and \(c_{J(x), n}\) given in (5.5), (6.4), (6.3), and (2.2). Then \(\eta(x)\) is a Gaussian process consisting of \((N_n + 1)\) standard normal variables \(\{\eta_j\}_{j=0}^{N_n}\) such that \(\eta(x) = \eta_j(x)\) for \(x \in [0, 1]\), and there exists a constant \(C > 0\) such that for large \(n\), \(\sup_{0 \leq t \neq t' \leq N_n} |E\eta_j \eta_{j'}| \leq C n^2\).

**Proof.** It is apparent that \(\mathcal{L}\{\eta_j | (X_{ij}, N_i), 1 \leq j \leq N_i, 1 \leq i \leq n\} = \mathcal{N}(0, 1)\), for \(0 \leq J \leq N_n\), so \(\mathcal{L}\{\eta_j\} = \mathcal{N}(0, 1)\), for \(0 \leq J \leq N_n\). For \(J \neq J'\), by (8.2) and (3.1), \(R_{ij, \varepsilon, J} R_{ij, \varepsilon, J'} = B_{fJ} (X_{ij}) B_{fJ'} (X_{ij}) \sigma^2 (X_{ij}) = 0\), along with (8.4), (8.3), the conditional independence of \(\tilde{z}_k(x), \tilde{z}^{(0)}(x)\) on \(X_{ij}, N_i, 1 \leq j \leq N_i, 1 \leq i \leq n\), and independence of \(\xi_{ik}, Z_{ij}, X_{ij}, N_i, 1 \leq k \leq \kappa, 1 \leq j \leq N_i, 1 \leq i \leq n\), \(E(\eta_j \eta_{j'})\) is

\[
E \left\{ \left\{ \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K} R_{ik, \xi, J} + \sum_{j=1}^{n} R_{ij, \varepsilon, J}^2 \right\} \right\}^{-1/2} \left\{ \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K} R_{ik, \xi, J'}^2 \right\} \right\}
\]

\[
= EC_{n, J, J'}
\]

in which

\[
C_{n, J, J'} = \left\{ \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K} R_{ik, \xi, J} + \sum_{j=1}^{n} R_{ij, \varepsilon, J} \right\} \right\}^{-1/2} \times
\]

\[
N^{-1} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K} R_{ik, \xi, J} + \sum_{j=1}^{n} R_{ij, \varepsilon, J} \right\}
\]

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

\[
N^{-1} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K} R_{ik, \xi, J} + \sum_{j=1}^{n} R_{ij, \varepsilon, J} \right\}
\]
\[ c_\sigma^2 N^{-1}_T \sum_{i=1}^{n} \sum_{j=1}^{N_i} B_{ij}^2 (X_{ij}) = c_\sigma^2 \| B_J \|_{2, N_T}^2 \geq c_\sigma^2 (1 - A_n), \quad \text{for } 0 \leq J \leq N_s, \]

\[
P \left\{ \inf_{0 \leq J \neq J' \leq N_s} \left\{ N^{-1}_T \sum_{i=1}^{n} \left( \sum_{k=1}^{K} R_{ik, J}^2 + \sum_{k=1}^{K} R_{kj, J'}^2 \right) \right\} \right\} \times \]

\[
\left\{ N^{-1}_T \sum_{i=1}^{n} \left( \sum_{k=1}^{K} R_{ik, J}^2 + \sum_{k=1}^{K} R_{ij, J'}^2 \right) \right\} \geq c_\sigma^2 \left( 1 - C_A \sqrt{\frac{\log(n)}{nh_s}} \right)^2 \geq 1 - 2n^{-3},
\]

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

\[
N^{-1}_T \sum_{i=1}^{n} \left( \sum_{k=1}^{K} R_{ik, J}^2 + \sum_{k=1}^{K} R_{kj, J'}^2 \right)
\]

there exists \( C_0 > 0 \) such that, for large \( n \),

\[
P \left( \sup_{0 \leq J \neq J' \leq N_s} \left| N^{-1}_T \sum_{i=1}^{n} \sum_{k=1}^{K} R_{ik, J} R_{ij, J'} \right| \leq C_0 h_s \right) \geq 1 - 2n^{-3}.
\]

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

\[
P \left( \sup_{0 \leq J \neq J' \leq N_s} |C_{n, J, J'}| \leq C_1 h_s \right) \geq 1 - 4n^{-3}.
\]

Note that as a continuous random variable, \( \sup_{0 \leq J \neq J' \leq N_s} |C_{n, J, J'}| \in \{0, 1\} \), thus

\[
E \left( \sup_{0 \leq J \neq J' \leq N_s} |C_{n, J, J'}| \right) = \int_0^1 P \left( \sup_{0 \leq J \neq J' \leq N_s} |C_{n, J, J'}| > t \right) \, dt.
\]

For large \( n \), \( C_1 h_s < 1 \) and then \( E \left( \sup_{0 \leq J \neq J' \leq N_s} |C_{n, J, J'}| \right) \) is

\[
\int_0^{C_1 h_s} P \left( \sup_{0 \leq J \neq J' \leq N_s} |C_{n, J, J'}| > t \right) \, dt + \int_{C_1 h_s}^{1} P \left( \sup_{0 \leq J \neq J' \leq N_s} |C_{n, J, J'}| > t \right) \, dt
\]

\[
\leq \int_0^{C_1 h_s} 1 \, dt + \int_{C_1 h_s}^{1} 4n^{-3} \, dt \leq C_1 h_s + 4n^{-3} \leq C h_s
\]

for some \( C > 0 \) and large enough \( n \). The lemma now follows from

\[
\sup_{0 \leq J \neq J' \leq N_s} E \left( C_{n, J, J'} \right) \leq E \left( \sup_{0 \leq J \neq J' \leq N_s} |C_{n, J, J'}| \right) \leq C h_s.
\]

By Lemma [A.8], the \((N_s + 1)\) standard normal variables \( \eta_0, ..., \eta_{N_s} \) satisfy the conditions of Lemma [A.2]. Hence for any \( \tau \in R \),

\[
\lim_{n \to \infty} P \left( \sup_{x \in [0, 1]} |\eta(x)| \leq \tau / a_{N_s+1} + b_{N_s+1} \right) = \exp \left( -2e^{-\tau} \right) . \quad (8.7)
\]
For $x \in [0, 1]$, $R_{ik,\xi,J}$, $R_{ij,\xi,J}$ given in (8.2), define the ratio of population and sample quantities as $r_n(x) = \{n E(N_1)/N_x \}^{1/2} \{\bar{R}_n(x)/\bar{R}(x)\}^{1/2}$, with

$$\bar{R}_n(x) = \frac{N_x^{-1} \left( \sum_{i=1}^{N_x} \sum_{k=1}^{\kappa} R_{ik,\xi,J}^2 + \sum_{j=1}^{N_x} R_{ij,\xi,J}^2 \right)}{N_x^{-1} \sum_{i=1}^{N_x} ER_{ik,\xi,J}^2 + ER_{11,\xi,J}^2},$$

$$\bar{R}(x) = \frac{(EN_1)^{-1} \sum_{i=1}^{\kappa} ER_{ik,\xi,J}^2}{(EN_1)^{-1} \sum_{k=1}^{\kappa} ER_{ik,\xi,J}^2 + ER_{11,\xi,J}^2}.$$

**Lemma A.9.** Under Assumptions (A2), (A3), for $\eta(x), \sigma_n(x)$ in (8.6), (2.3),

$$|\sigma_n(x)^{-1} \left( \sum_{k=1}^{\kappa} \hat{\xi}_k(x) + \hat{\xi}(0)(x) \right) - \eta(x) - \{r_n(x) - 1\} | \eta(x) | = O_{a.s.} \left( \sqrt{\{\log(N_n + 1)\} (\log n) / (nh_3)} \right).$$

**Proof.** Equation (8.8) follows from the definitions of $\eta(x)$ and $\sigma_n(x)$. By Lemma A.6, $\sup_{x \in [0,1]} \left| N_x^{-1} \sum_{i=1}^{N_x} \sum_{j=1}^{\kappa} R_{ij,\xi,J}^2 - ER_{11,\xi,J}^2 \right| = O_{a.s.} \left( \sqrt{\log n / (nh_3)} \right)$,

$$\sup_{x \in [0,1]} \left| N_x^{-1} \sum_{k=1}^{\kappa} \sum_{i=1}^{\kappa} R_{ik,\xi,J}^2 - (EN_1)^{-1} \sum_{k=1}^{\kappa} ER_{ik,\xi,J}^2 \right| \leq \sup_{x \in [0,1]} \left| (EN_1)^{-1} \sum_{k=1}^{\kappa} n^{-1} \sum_{i=1}^{\kappa} R_{ik,\xi,J}^2 + ER_{ik,\xi,J}^2 \right|

+ \sup_{x \in [0,1]} \left| n (EN_1)^{-1} \left( N_x^{-1} n - 1 \right) \sum_{i=1}^{\kappa} R_{ik,\xi,J}^2 \right|

= O_{a.s.} \left( \sqrt{\log n / (nh_3)} \right) + O_{a.s.} \left( \sqrt{\log n / (nh_3)} \right),$$

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

$$\sup_{x \in [0,1]} \left| N_x^{-1} \sum_{k=1}^{\kappa} \sum_{i=1}^{\kappa} R_{ik,\xi,J}^2 - (EN_1)^{-1} \sum_{k=1}^{\kappa} ER_{ik,\xi,J}^2 \right| + \sup_{x \in [0,1]} \left| N_x^{-1} \sum_{i=1}^{\kappa} \sum_{j=1}^{N_x} R_{ij,\xi,J}^2 - ER_{11,\xi,J}^2 \right| = O_{a.s.} \left( \sqrt{\log n / (nh_3)} \right).$$

Thus $\sup_{x \in [0,1]} \left( \{\bar{R}_n(x)\}^{1/2} - \{\bar{R}(x)\}^{1/2} \right) \leq \sup_{x \in [0,1]} |\bar{R}_n(x) - \bar{R}(x)| \sup_{x \in [0,1]} \{\bar{R}(x)\}^{-1/2} = O_{a.s.} \left( \sqrt{\log n / (nh_3)} \right)$. Then $\sup_{x \in [0,1]} \{a_{N_n+1} |r_n(x) - 1\}$ is bounded by

$$a_{N_n+1} \left\{ \frac{\{n E(N_1)/N_x \}^{1/2} \sup_{x \in [0,1]} \{\bar{R}_n(x)/\bar{R}(x)\}^{1/2} - 1}{\{\bar{R}_n(x)/\bar{R}(x)\}^{1/2}} \right\},$$

$$\leq a_{N_n+1} \left\{ \frac{\{n E(N_1)/N_x \}^{1/2} \sup_{x \in [0,1]} \{\bar{R}(x)\}^{-1/2} \sup_{x \in [0,1]} \{\bar{R}_n(x)\}^{1/2} - \{\bar{R}(x)\}^{1/2}}{\{\bar{R}(x)\}^{1/2}} \right\} = O_{a.s.} \left( \sqrt{\log (N_n + 1) (\log n) / (nh_3)} \right). \quad \Box
Proof of Proposition 2. The proof follows from Lemmas (A.3) (A.7) (A.9) (8.7), and Slutsky’s Theorem. □

Proof of Theorem 1. By Theorem 2, \( \| \hat{m}(x) - m(x) \|_\infty = O_p(h_n) \), so

\[
a_{N_n+1} \left( \sup_{x \in [0,1]} \sigma_n^{-1}(x) |\hat{m}(x) - m(x)| \right) = O_p \left((nh_n)^{1/2} \sqrt{\log (N_n+1)h_n} \right) = o_p(1),
\]

\[
a_{N_n+1} \left( \sup_{x \in [0,1]} \sigma_n^{-1}(x) |\hat{m}(x) - m(x)| - \sup_{x \in [0,1]} \sigma_n^{-1}(x) \left| \sum_{k=1}^\kappa \xi_k(x) + \tilde{\eta}(x) \right| \right) = o_p(1).
\]

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

\[
\lim_{n \to \infty} P \left\{ a_{N_n+1} \left( \sup_{x \in [0,1]} \sigma_n^{-1}(x) \left| \sum_{k=1}^\kappa \xi_k(x) + \tilde{\eta}(x) \right| - b_{N_n+1} \right) \leq \tau \right\} = \exp (-2e^{-\tau}).
\]

Thus Slutsky’s Theorem implies that

\[
\lim_{n \to \infty} P \left\{ a_{N_n+1} \left( \sup_{x \in [0,1]} \sigma_n^{-1}(x) |\hat{m}(x) - m(x)| - b_{N_n+1} \right) \leq \tau \right\} = \exp (-2e^{-\tau}).
\]

Let \( \tau = -\log \left( \frac{1}{2} \log (1 - \alpha) \right) \), definitions of \( a_{N_n+1}, b_{N_n+1} \), and \( Q_{N_n+1} (\alpha) \) in (2.4) entail

\[
\lim_{n \to \infty} P \left\{ m(x) \in \hat{m}(x) \pm \sigma_n(x)Q_{N_n+1} (\alpha), \forall x \in [0,1] \right\} = 1 - \alpha.
\]

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

References


Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824
E-mail: (mashujie@stt.msu.edu)

Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou 215006, People’s Republic of China; and Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824
E-mail: (yanglijian@suda.edu.cn)

Department of Statistics, Texas A&M University, College Station, TX 77843
E-mail: (carroll@stat.tamu.edu)
Figure 1: Plots of simulated data scatter points at $\sigma = 0.5$: (a) $n = 20$, (b) $n = 50$, and the true curve.
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) \(1 - \alpha = 0.99, n = 50\).
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.