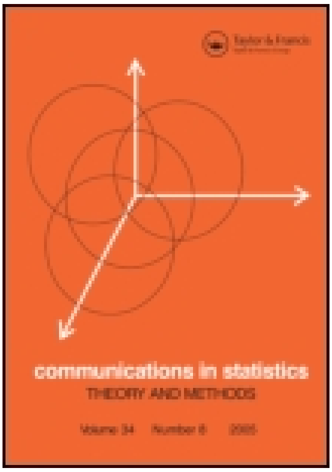

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

A Simultaneous Confidence Band for Dense Longitudinal Regression

Q. Song[a], R. Liu[b], Q. Shao[b] & L. Yang[cd]

[a] Department of Mathematical Sciences, The University of Texas at Dallas, Dallas, Texas, USA

[b] Department of Mathematics and Statistics, The University of Toledo, Toledo, Ohio, USA

[c] Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou, China

[d] Department of Statistics and Probability, Michigan State University, East Lansing, Michigan, USA

Accepted author version posted online: 14 Apr 2013.Published online: 24 Nov 2014.

To cite this article: Q. Song, R. Liu, Q. Shao & L. Yang (2014) A Simultaneous Confidence Band for Dense Longitudinal Regression, Communications in Statistics - Theory and Methods, 43:24, 5195-5210, DOI: 10.1080/03610926.2012.729643

To link to this article: http://dx.doi.org/10.1080/03610926.2012.729643

Communications in Statistics—Theory and Methods, 43: 5195–5210, 2014

ISSN: 0361-0926 print / 1532-415X online
DOI: 10.1080/03610926.2012.729643

A Simultaneous Confidence Band for Dense Longitudinal Regression

Q. SONG,[1] R. LIU,[2] Q. SHAO,[2] AND L. YANG[3,4]

[1]Department of Mathematical Sciences, The University of Texas at Dallas, Dallas, Texas, USA
[2]Department of Mathematics and Statistics, The University of Toledo, Toledo, Ohio, USA
[3]Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou, China
[4]Department of Statistics and Probability, Michigan State University, East Lansing, Michigan, USA

We present a method of using local linear smoothing to construct simultaneous confidence bands for the mean function of densely spaced functional data. Our approach works well under mild conditions. In addition, the local linear estimator and its accompanying confidence band enjoy semiparametric efficiency in the sense that they are asymptotically equivalent to the counterparts obtained from the random trajectories entirely observed without errors. We illustrate the performance of the proposed confidence band through a simulation study. Furthermore, an application in food science is presented.

Keywords Confidence band; Functional data; Karhunen-Loève L^2 representation; Local linear smoothing; Strong approximation.

Mathematics Subject Classification primary 62G08, 62G15; secondary 62G05, 62G20.

1. Introduction

Functional data with different designs receives increasing attention, see James, Hastie and Sugar (2000), James (2002), Cardot and Sarda (2005), James and Silverman (2005), Müller and Stadtmüller (2005), Ramsay and Silverman (2005), Hall et al. (2006), Müller and Yao (2008). Analysis of functional data is challenging since the basic units of information are entire observed curves. Two distinct types of functional data have been studied. Yao et al. (2005a, 2005b), and Yao (2007) studied sparse longitudinal data of which each trajectory is observed sparsely and randomly. Li and Hsing (2007, 2010), on the other hand, concerned

Received June 22, 2012; Accepted September 4, 2012.
Address correspondence to Qin Shao, University of Toledo, USA. E-mail: qin.shao@utoledo.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.

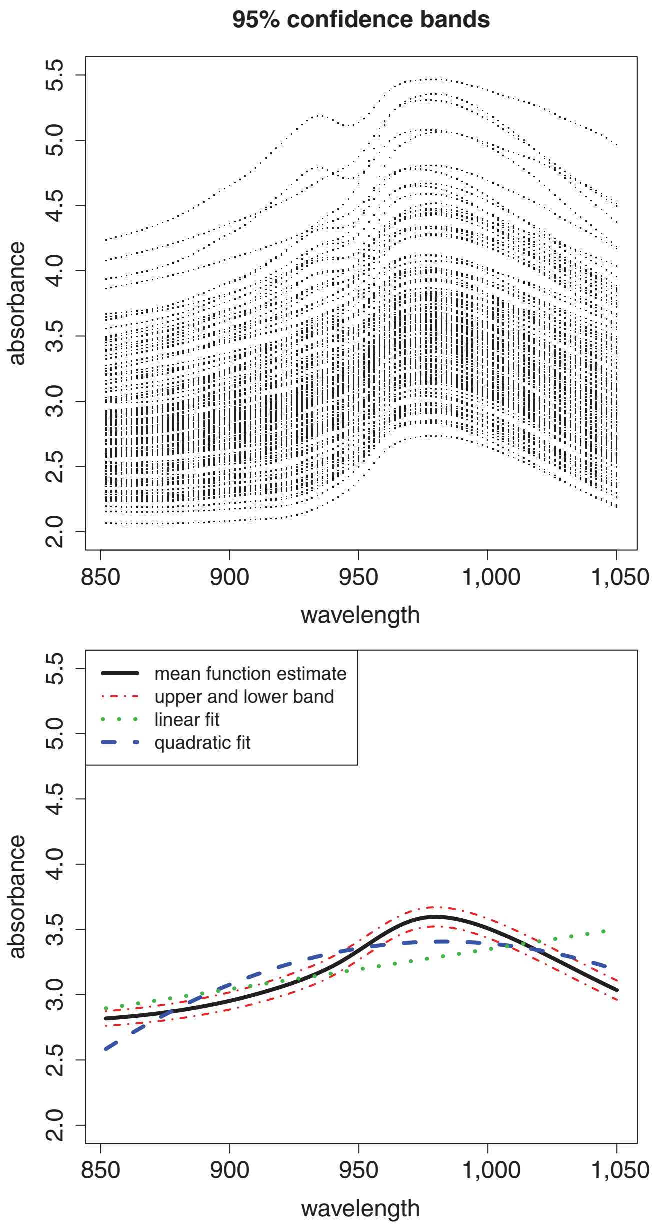

Figure 1. The upper plot shows 100 trajectories of absorbance in wavelength 850–1050 from Tecator data. The lower plot shows the mean estimate (solid line), confidence band (dashed lines), linear fit (dotted line), and quadratic fit (dash-dot line).

dense functional data and derived strong uniform convergence rates for local linear smooth estimators.

Our paper is motivated by a spectrometric data set from the food industry. In this study, 240 finely chopped pure meat samples with different fat contents were considered. Each meat sample provides a trajectory, consisting of 100 channel points of spectrum absorbance in the wavelength range of 850–1050 nm; see the upper panel of Fig. 1. These spectra are densely sampled at these 100 channel points and seem to be quite smooth. To detect any possible trend of the mean curve, we need to construct a simultaneous confidence band for it.

In nonparametric data analysis, an important theoretical and practical issue is detection of the global trend of a mean function. Many researchers attempted to solve this problem by using nonparametric simultaneous confidence band, as a powerful tool of global inference for functions; see Zhou et al. (1998), Fan and Zhang (2000), Claeskens and Van Keilegom (2003), Wu and Zhao (2007), Huang et al. (2008), Zhao and Wu (2008), Song and Yang (2009), Wang and Yang (2009) for its theory and applications. Simultaneous confidence bands have been proposed for functional data in Ma et al. (2012), Cao et al. (2012) based on B spline regression, and in Degras (2011) with local polynomial smoothing. Degras (2011), however, constructed confidence bands for the mean functions of dense functional data under high-level assumptions that are difficult to verify, see more discussions in Sec. 2.

In this paper, under some mild assumptions, we present simultaneous confidence bands for mean function in dense longitudinal data by local linear smoothing approach. Our simple yet flexible nonparametric method is not only theoretically well justified, but also easy to implement in practice. We establish the $\sqrt{n}$-consistency of the proposed estimator for the mean function. The local linear estimator and its accompanying confidence bands enjoy semiparametric efficiency in the sense that they are asymptotically equivalent to the counterparts obtained from the random trajectories entirely observed without errors.

We organize our paper as follows. In Sec. 2, we state our main results on confidence bands constructed from local linear smoothing. In Sec. 3, we provide further insights into the error structure of the estimator. In Sec. 4, we describe the procedure to implement the proposed confidence bands. In Sec. 5, we report the findings of simulation studies. The spectrometric data set, an empirical example in Sec. 6, illustrates how to use the proposed confidence band based on a local linear estimator for inference. Proofs of the technical lemmas are provided in the "Appendix."

2. Model and Main Results

A functional data set has the form $\{X_{ij}, Y_{ij}\}$, $1 \le i \le n$, $1 \le j \le N_i$, in which X_{ij} and Y_{ij} are the *j*th observations of the predictor and response variables for the *i*th subject. The sample path $\{X_{ij}, Y_{ij}\}$ is the noisy realization of a continuous time stochastic process $\xi_i(x)$ in the sense that $Y_{ij} = \xi_i(X_{ij}) + \sigma(j/N)\varepsilon_{ij}$, with the error ε_{ij} satisfying $E(\varepsilon_{ij}) = 0$, $E(\varepsilon_{ij}^2) = 1$, and $\{\xi_i(x), x \in \mathcal{X}\}$ independent and identical (iid) copies of a process $\{\xi(x), x \in \mathcal{X}\}$, which belongs to $L^2(\mathcal{X})$, i.e., $E\int_{\mathcal{X}} \xi^2(x)dx < \infty$. In this paper, we only deal with equally spaced designs and the same sample size for all subjects, i.e., $N_i \equiv N$, $1 \le i \le n$. Without loss of generality, let $\mathcal{X} = [0, 1]$ and $X_{ij} = j/N$, $1 \le j \le N$, for the *i*th subject.

For a standard process $\{\xi(x), x \in \mathcal{X}\}$, we define its mean function $m(x) = E\{\xi(x)\}$ and its covariance function $G(x, x') = \text{cov}\{\xi(x), \xi(x')\}$. Let sequences $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\psi_k(x)\}_{k=1}^{\infty}$ be eigenvalues and eigenfunctions of $G(x, x')$, respectively. The eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$ satisfy $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$, $\sum_{k=1}^{\infty} \lambda_k < \infty$, and the eigenfunctions $\{\psi_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L^2(\mathcal{X})$. Note that the covariance function $G(x, x') = \sum_{k=1}^{\infty} \lambda_k \psi_k(x)\psi_k(x')$, which implies that $\int G(x, x')\psi_k(x')dx' = \lambda_k \psi_k(x)$.

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

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

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

$$Y_{ij} = m(j/N) + \sum_{k=1}^{\kappa} \xi_{ik}\phi_k(j/N) + \sigma(j/N)\varepsilon_{ij}. \tag{2.1}$$

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.

For any Lebesgue measurable function ϕ on [0, 1], define $\|\phi\|_r = \{\int_0^1 |\phi(x)|^r dx\}^{1/r}$ $(1 \le r < \infty)$ and $\|\phi\|_\infty = \sup_{x\in[0,1]} |\phi(x)|$. For a continuous function ϕ on [0, 1], define the modulus of continuity as $\omega(\phi, \delta) = \max_{x,x'\in[0,1],|x-x'|\le\delta} |\phi(x) - \phi(x')|$. For any $\beta \in (0, 1]$, denote by $C^{0,\beta}[0, 1]$ the space of order β H ölder continuous functions on [0, 1], i.e.,

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

in which $\|\phi\|_{0,\beta}$ is called the $C^{0,\beta}$-norm of ϕ. Clearly, $C^{0,\beta}[0, 1] \subset C[0, 1]$. If $\phi \in C^{0,\beta}[0, 1]$, then $\omega(\phi, \delta) \le \|\phi\|_{0,\beta}\delta^\beta$. For any vector $\zeta = (\zeta_1, ..., \zeta_s)^T \in R^s$, denote $\|\zeta\|_r = (|\zeta_1|^r + \cdots + |\zeta_s|^r)^{1/r}$, $1 \le r < \infty$, and $\|\zeta\|_\infty = \max(|\zeta_1|, ..., |\zeta_s|)$.

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

$$E\zeta(x)\zeta(x') = G(x, x')\{G(x, x)G(x', x')\}^{-1/2}.$$

We denote by $Q_{1-\alpha}$ the $100(1 - \alpha)$-th percentile of the absolute maximum distribution of $\zeta(x)$, i.e., $P\{\sup_{x\in[0,1]} |\zeta(x)| \le Q_{1-\alpha}\} = 1 - \alpha$, $\forall\alpha \in (0, 1)$. Denote the $100(1 - \alpha/2)$-th percentile of the standard normal distribution by $z_{1-\alpha/2}$. Define the "infeasible estimator" for function m as

$$\bar{m}(x) = \bar{\xi}(x) = n^{-1} \sum_{i=1}^{n} \xi_i(x) = m(x) + \sum_{k=1}^{\kappa} \bar{\xi}_{.k}\phi_k(x). \tag{2.2}$$

where $\bar{\xi}_{.k} = n^{-1} \sum_{i=1}^{n} \xi_{ik}$, $1 \le k \le \kappa$. The term "infeasible" refers to the fact that $\bar{m}(x)$ relies on unknown quantity $\xi_i(x)$. If all the iid random curves $\xi_i(x)$ were observed, $\bar{m}(x)$ would be the natural estimator of $m(x)$, which is a view taken in Ferraty and Vieu (2006).

We estimate the mean trend function $m(x)$ by local linear smoothing. It is commonly known that the choice of kernel functions is not as crucial as bandwidth selection regarding the asymptotic behavior of the estimator. The order of the bandwidth is described in Assumption (A2) below. We need the following technical assumptions:

(A1) *The regression function* $m \in C^2[0, 1]$, *and the standard deviation function* $\sigma(x) \in C^{0,\beta}[0, 1]$ *for some* $\beta \in (0, 1]$.
(A2) *As* $n \to \infty$, $Nn^{-1/4}(\log n)^{-1} \to \infty$ *and* $N = O(n^\theta)$ *for some* $\theta > 5/8$. *The bandwidth* h *satisfies* $Nh(\log n)^{-1} \to \infty$, $nh^4 \to 0$ *as* $n \to \infty$.
(A3) *The kernel function K is a probability density function with support* $[-1, 1]$, *symmetric about* 0 *and Lipschitz continuous.*
(A4) *There exists a constant* $C_G > 0$ *such that* $G(x, x) \ge C_G$, $x \in [0, 1]$; *for* $k \in \{1, \ldots, \kappa\}$, $\phi_k(x) \in C^{0,\beta}[0, 1]$, $\sum_{k=1}^{\kappa} \|\phi_k\|_\infty < \infty$ *and as* $n \to \infty$, $h^\beta \sum_{k=1}^{\kappa_n} \|\phi_k\|_{0,\beta} = o(1)$ *for a*

sequence $\{\kappa_n\}_{n=1}^{\infty}$ *of increasing integers, with* $\lim_{n\to\infty}\kappa_n = \kappa$ *and the constant* $\beta \in (0, 1]$ *as in Assumption (A2). In particular,* $\sum_{k=\kappa_n+1}^{\kappa}\|\phi_k\|_\infty = o(1)$.

(A5) *There are constants* $C_1, C_2 \in (0,\infty)$, $\gamma_1, \gamma_2 \in (1,\infty)$, $\rho \in (0, 1/2)$ *and* $\{Z_{ik,\xi}\}_{i=1,k=1}^{n,\kappa}$, $\{Z_{ij,\varepsilon}\}_{i=1,j=1}^{n,N}$ *are iid* $N(0, 1)$ *variables such that*

$$\max_{1\le k\le\kappa} P\left\{\max_{1\le t\le n}\left|\sum_{i=1}^{t}\xi_{ik} - \sum_{i=1}^{t} Z_{ik,\xi}\right| > C_1 n^{\rho}\right\} < C_2 n^{-\gamma_1}, \tag{2.3}$$

$$P\left\{\max_{1\le j\le N}\max_{1\le t\le n}\left|\sum_{i=1}^{t}\varepsilon_{ij} - \sum_{i=1}^{t} Z_{ij,\varepsilon}\right| > C_1 n^{\rho}\right\} < C_2 n^{-\gamma_2}. \tag{2.4}$$

Assumption (A4) guarantees that the principal components have collectively bounded smoothness. In practice, the selection of κ is a crucial issue, and we recommend using a cross-validation criterion for this issue; see, for instance, Yao et al. (2005b). Assumption (A5) provides Gaussian approximation to the estimated error process, and is implied by the following assumption:

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

We will prove that Assumption (A5′) ensures Assumption (A5) in the Appendix. As pointed out in Cao et al. (2012), the Assumption (A2) of Degras (2011) about the Hölder continuity of the stochastic process $\eta(x) = m(x) + \sum_{k=1}^{\infty}\xi_k\phi_k(x)$ is more restrictive than (A4) and (A5) above. It is easy to construct an example such that our Assumptions (A4) and (A5) are satisfied while Assumption (A2) of Degras (2011) is not. For instance, if $\kappa = \infty$, the Assumption (A2) of Degras (2011) would necessitate that $\sum_{k=1}^{\kappa}\|\phi_k\|_{0,\beta} < +\infty$, while our Assumption (A4) allows for the case of $\sum_{k=1}^{\kappa}\|\phi_k\|_{0,\beta} = +\infty$ as long as for a sequence of integers $\kappa_n \to \infty$, one has $\sum_{k=1}^{\kappa_n}\|\phi_k\|_{0,\beta} \to +\infty$ slower than $h^{-\beta}$. It is obvious that Assumption (A4) about ϕ_k's holds if κ is finite and all $\phi_k(x) \in C^{0,\beta}[0, 1]$.

We propose to estimate the mean function $m(x)$ by solving a local linear least squares

$$\left(\hat{a}, \hat{b}\right) = \arg\min_{(a,b)}\sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij} - a - b\left(\frac{j}{N} - x\right)\right\}^2 K_h\left(\frac{j}{N} - x\right)$$

with $K_h(u) = \frac{1}{h}K(\frac{u}{h})$, where $h = h_n \to 0$, as $n \to \infty$. For any $x \in [0, 1]$,

$$\hat{m}(x) = \hat{a} = \mathbf{e}_0^T\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y} \tag{2.5}$$

in which $\mathbf{Y} = (\bar{Y}_{\cdot 1}, \ldots, \bar{Y}_{\cdot N})^T$, $\bar{Y}_{\cdot j} = n^{-1}\sum_{i=1}^{n} Y_{ij}$ for $1 \le j \le N$, $\mathbf{e}_0^T = (1, 0)$, the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} 1 & \left(\frac{1}{N} - x\right) \\ \vdots & \vdots \\ 1 & \left(\frac{N}{N} - x\right) \end{pmatrix}_{N\times 2}, \tag{2.6}$$

and the weight matrix $\mathbf{W} = \mathrm{diag}(\{K_h(j/N - x)/N\}_{j=1}^{N})$.

Our confidence band is constructed based on $\hat{m}(x)$ in (2.5). We first present the asymptotic property of $\bar{m}(x)$ in (2.2), and then we show that $\hat{m}(x)$ has the same asymptotic property as $\bar{m}(x)$.

Theorem 1. *Under Assumptions (A1)–(A5), for any* $\alpha \in (0, 1)$*, as* $n \to \infty$*, the "infeasible estimator"* $\bar{m}(x)$ *converges at the rate of* $\sqrt{n}$ *both uniformly and pointwise with*

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

and for any $x \in [0, 1]$,

$$P\left\{n^{1/2}\left|\bar{m}(x) - m(x)\right| G(x, x)^{-1/2} \leq z_{1-\alpha/2}\right\} \to 1-\alpha.$$

Theorem 2. *Under Assumptions (A1)–(A5), the local linear estimator* $\hat{m}$ *is asymptotically equivalent to* $\bar{m}$*, that is*

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

Remark 1. The significance of Theorem 2 lies in the fact that one does not need to distinguish between the local linear estimator $\hat{m}$ and the "infeasible estimator" $\bar{m}$. Both converge at the rate of $\sqrt{n}$, the same rate as a parametric estimator.

We therefore have established semiparametric efficiency of the nonparametric estimator $\hat{m}(x)$. The explicit expression of a confidence band for $m(x)$ is summarized in the following corollary, which is a direct result of the theorems above.

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

$$\hat{m}(x) \pm G(x, x)^{1/2} Q_{1-\alpha} n^{-1/2},\ x \in [0, 1], \tag{2.8}$$

and an asymptotic $100(1-\alpha)\%$ *pointwise confidence interval for* $m(x)$ *is*

$$\hat{m}(x) \pm G(x, x)^{1/2} z_{1-\alpha/2} n^{-1/2}.$$

3. Decomposition

In this section, we decompose the estimation error $\hat{m}(x) - m(x)$ by the representation of Y_{ij}. To understand this decomposition, we begin by discussing the representation of the local linear estimator $\hat{m}(x)$ in (2.5). We obtain the following crucial decomposition

$$\hat{m}(x) = \widetilde{m}(x) + \widetilde{\xi}(x) + \widetilde{e}(x), \tag{3.8}$$

in which

$$\widetilde{m}(x) = \mathbf{e}_0^T \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{W}\mathbf{m},$$

$$\tilde{\xi}(x) = \mathbf{e}_0^T \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \xi,$$
$$\widetilde{e}(x) = \mathbf{e}_0^T \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{e}, \tag{3.9}$$

with the signal vector $\mathbf{m} = (m(1/N), \ldots m(N/N))^T$, the eigenfunction vector $\xi = (\sum_{k=1}^{\kappa} \bar{\xi}_{.k}\phi_k(1/N), \ldots, \sum_{k=1}^{\kappa} \bar{\xi}_{.k}\phi_k(N/N))^T$ and the noise vector $\mathbf{e} = (\sigma(1/N)\bar{\varepsilon}_{.1}, \ldots, \sigma(N/N)\bar{\varepsilon}_{.N})^T$ with $\bar{\varepsilon}_{.j} = n^{-1}\sum_{i=1}^{n} \varepsilon_{ij}$, $1 \le j \le N$.

Now, $\hat{m}(x) - m(x) = \widetilde{m}(x) - m(x) + \widetilde{\xi}(x) + \widetilde{e}(x)$ with the bias term $\widetilde{m}(x) - m(x) + \tilde{\xi}(x)$ and the noise term $\widetilde{e}(x)$. The next three propositions concern $\widetilde{m}(x)$, $\tilde{\xi}(x)$, and $\widetilde{e}(x)$ given in (3.8), and Theorem 2 follows immediately.

Proposition 1. *Under Assumptions (A1)–(A3), as $n \to \infty$*

$$\sup_{x\in[0,1]} n^{1/2} \left|\widetilde{m}(x) - m(x)\right| G(x, x)^{-1/2} = o_p(1).$$

Proposition 2. *Under Assumptions (A2)–(A5), as $n \to \infty$*

$$\sup_{x\in[0,1]} n^{1/2} \left|\tilde{\xi}(x) - \{\bar{m}(x) - m(x)\}\right| = o_p(1), \tag{3.10}$$

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

$$P\left\{\sup_{x\in[0,1]} n^{1/2} \left|\tilde{\xi}(x)\right| G(x, x)^{-1/2} \le Q_{1-\alpha}\right\} \to 1 - \alpha. \tag{3.11}$$

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

$$\sup_{x\in[0,1]} n^{1/2} \left|\widetilde{e}(x)\right| G(x, x)^{-1/2} = o_p(1).$$

The Appendix contains the proofs for the above three propositions. These propositions together with (3.8) imply Theorem 1 .

4. Implementation

In this section, we describe the procedure to implement the confidence band and interval in Corollary 1. Given any data set $(j/N, Y_{ij})_{j=1,i=1}^{N,n}$ from model (2.1), we obtain the local linear estimator $\hat{m}(x)$ through (2.5). As described in Sec. 2, $Q_{1-\alpha}$ is the $100(1-\alpha)^{th}$ percentile of the absolute maximum distribution of Gaussian process $\zeta(x)$. Therefore, we generate a series of normally distributed Z_k, $1 \le k \le \kappa$, and we obtain $Q_{1-\alpha}$ through

$$P\left\{\sup_{x\in[0,1]} \frac{\left|\sum_{k=1}^{\kappa} \phi_k(x) Z_k\right|}{\sqrt{\phi_k^2(x)}} > Q_{1-\alpha}\right\} = 1 - \alpha.$$

That is the $Q_{1-\alpha}$ is the $100(1-\alpha)$th percentile of $\sup_{x\in[0,1]} \frac{|\sum_{k=1}^{\kappa} \phi_k(x) Z_k|}{\sqrt{\phi_k^2(x)}}$. The pilot estimator of the covariance function $G(x, x')$ is $\hat{G}(x, x') = \hat{a}(x, x')$ such that $\{\hat{a}(x, x'), \hat{b}_1(x, x'),$

$\hat{b}_2(x, x')\}$ minimize the following objective function:

$$\sum_{j,j'=1}^{N} \left\{ C_{.jj'} - a - b_1 \left(\frac{j}{N} - x \right) - b_2 \left(\frac{j'}{N} - x' \right) \right\}^2 K_h \left(\frac{j}{N} - x \right) K_h \left(\frac{j'}{N} - x' \right),$$

where $C_{.jj'} = n^{-1} \sum_{i=1}^{n} \{Y_{ij} - \hat{m}(j/N)\}\{Y_{ij} - \hat{m}(j'/N)\}$, $1 \leq j, j' \leq N$.

We use the quartic kernel $K(u) = 15(1 - u^2)^2 I_{\{|u| \leq 1\}}/16$ and the rule-of-thumb bandwidth h_{rot} of Silverman (1986) defined by

$$h_{\text{rot}} = (4\pi)^{1/10} \left(\frac{140}{3} \right)^{1/5} n^{-1/5} s_n,$$

where s_n is the sample standard deviation of $\{\bar{X}_{.j}\}_{j=1}^{N}$.

5. Simulation

We carry out a simulation to illustrate finite sample behavior of the proposed confidence bands in (2.7). We generate the data from the model

$$Y_{ij} = m\,(j/N) + \sum_{k=1}^{2} \xi_{ik} \phi_k\,(j/N) + \sigma \varepsilon_{ij}, \tag{5}$$

with $\xi_{ik} \sim$ Normal(0, 1), $k = 1, 2$, $\varepsilon_{ij} \sim$ Normal(0, 1), and $m(x) = \sin\{2\pi(x - 1/2)\}$. We take the orthonormal functions $\phi_1(x) = -2\cos\{\pi(x - 1/2)\}$ and $\phi_2(x) = \sin\{\pi(x - 1/2)\}$ to be the eigenfunctions, thus $\lambda_1 = 2$, $\lambda_2 = 1/2$. Two different noise levels $\sigma = 0.2, 0.5$ are used to interpret the result, and the number of subjects n is taken to be 50, 100, 200, and 400. We use $N = [n^{0.8}\log n]$ to determine the number of grid for each subject.

Table 1 shows the coverage frequencies from 200 replications for the confidence levels $1 - \alpha = 0.95$ and 0.99. As we expect, the coverage percentages for the confidence bands, respectively, approach to the nominal confidence levels 0.95 and 0.99 at both noise levels as the sample size increases.

Table 1
Coverage frequencies from 200 replications

σ	n	$1 - \alpha = 0.95$	$1 - \alpha = 0.99$
	50	0.950	0.975
	100	0.950	0.995
0.2	200	0.950	0.985
	400	0.940	0.980
	50	0.900	0.970
	100	0.905	0.980
0.5	200	0.940	0.985
	400	0.940	0.990

6. Empirical Example

Here we apply the proposed method to the Tecator data set, which can be downloaded from http://lib.stat.cmu.edu/datasets/tecator. The data set contains measurements on $n = 240$ meat samples, where for each sample a $N = 100$ channel near-infrared spectrum of absorbance measurements is obtained; see the upper panel of Fig. 1. Each sample contains finely chopped pure meat with different moisture, fat, and protein contents, and the percentage of fat is determined by analytical chemistry. The aim of the study is to predict the fat content of the meat sample based on the near-infrared absorbance spectrum. Each spectrum was recorded on a Tecator Infratec Food and Feed Analyzer, and the Feed Analyzer worked in the wavelength range $850 - 1,050$ nm. Here, hypotheses of interest are:

$$H_0 : m_1(x) = a + bx,\ \forall x \in [850\,\text{nm}, 1050\,\text{nm}] \longleftrightarrow H_a : m_1(x) \neq a + bx.$$

The lower panel of Fig. 1 shows neither the linear nor quadratic estimates is covered entirely by the 95% confidence band. The linearity null hypothesis $H_0 : m(x) = a + bx$ is rejected with p-value less than 0.05, since the linear regression fit is not covered by the 95% confidence band entirely. Likewise, a quadratic null hypothesis $H_0 : m(x) = a + bx + cx^2$ is rejected with p-value < 0.05 as well. Thus our conclusion is that at the confidence level 95%, the mean function of absorbance is not linear or quadratic. The mean estimate increases slowly when wavelength is less than 950 nm. For wavelength between 950 and 970 nm, it increases much faster. After the turning point at around 970 nm, there is a decreasing trend for wavelength.

Funding

Lijian Yang's research was supported by NSF awards DMS 0706518 and DMS 1007594, and funding from the "Jiangsu-Specially Appointed Professor Program", Jiangsu Province, China.

References

Cao, G., Yang, L., Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *J. Nonparametric Stat.* 24:359–377.

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

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

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

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Stat. Sinica* 21:1735–1765.

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

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

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

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

Huang, X., Wang, L., Yang, L., Kravchenko, A. N. (2008). Management practice effects on relationships of grain yields with topography and precipitation. *Agronomy J.* 100:1463–1471.
James, G. M., Hastie, T., Sugar, C. (2000). Principal component models for sparse functional data. *Biometrika* 87:587–602.
James, G. M. (2002). Generalized linear models with functional predictors. *J. Royal Stat. Society Series B* 64:411–432.
James, G. M., Silverman, B. W. (2005). Functional adaptive model estimation. *J. Am. Stat. Assoc.* 100:565–576.
Li, Y., Hsing, T. (2007). On rates of convergence in functional linear regression. *J. Multivar. Anal.* 98:1782–1804.
Li, Y., Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Stat.* 38:3321–3351.
Ma, S., Yang, L., Carroll, R. J. (2012). Simultaneous confidence band for sparse longitudinal regression. *Stat. Sinica* 22:95–122.
Müller, H. G., Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Stat.* 33:774–805.
Müller, H. G., Yao, F. (2008). Functional additive models. *J. Am. Stat. Assoc.* 103:1534–1544.
Ramsay, J. O., Silverman, B. W. (2005). *Functional Data Analysis.* Second edition. Springer Series in Statistics. New York: Springer.
Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
Song, Q., Yang, L. (2009). Spline confidence bands for variance function. *J. Nonparametric Stat.* 21:589–609.
Wang, J., Yang, L. (2009). Polynomial spline confidence bands for regression curves. *Stat. Sinica* 19:325–342.
Wu, W., Zhao, Z. (2007). Inference of trends in time series. *J. Royal Stat. Society, Series B* 69:391–410.
Yao, F., Müller, H. G., Wang, J. L. (2005a). Functional linear regression analysis for longitudinal data. *Annals of Statistics* 33:2873–2903.
Yao, F., Müller, H. G., Wang, J. L. (2005b). Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* 100:577–590.
Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J. Multivar. Anal.* 98:40–56.
Zhao, Z., Wu, W. (2008). Confidence bands in nonparametric time series regression. *Ann. Stat.* 36:1854–1878.
Zhou, S., Shen, X., Wolfe, D. A. (1998). Local asymptotics of regression splines and confidence regions. *Ann. Stat.* 26:1760–1782.

Appendix

Throughout this section, C means some positive nonzero constant, and U and u denote uniformly O and o, respectively, for sequences of random variables.

A.1. Preliminaries

The following Lemma is needed to prove Lemma A.5.

Lemma A.1. *[Theorem 2.6.7 of Csőrgő and Révész (1981)] Suppose that random variables* ξ_i, $1 \leq i \leq n$ *are iid with* $E(\xi_1) = 0$ *and* $E(\xi_1^2) = 1$, *and function* $H(x) > 0$ $(x \geq 0)$ *is increasing and continuous such that* $x^{-2-\gamma}H(x)$ *is increasing for some* $\gamma > 0$ *and* $x^{-1}logH(x)$ *is decreasing with* $EH(|\xi_1|) < \infty$. *Then there exist constants* C_1, C_2 *and* $a > 0$, *which depend only on the distribution of* ξ_1 *such that for any* $\{x_n\}_{n=1}^{\infty}$ *satisfying*

$H^{-1}(n) < x_n < C_1(nlogn)^{1/2}$,

$$P\left\{\max_{1\le t\le n}|S_t - W(t)| > x_n\right\} \le C_2 n\left\{H(ax_n)\right\}^{-1},$$

where $S_t = \sum_{i=1}^{t}\xi_i$ *and* $\{W(t)\}$ *is a Brownian motion.*

A.2. Proof of Proposition 1

We have $\widetilde{m}(x) = \mathbf{e}_0^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{m}$, so the dispersion matrix is

$$\mathbf{X}^T\mathbf{W}\mathbf{X} = \mathrm{diag}(1, h)\,\mathbf{D}_{N,x}\,\mathrm{diag}(1, h),$$

where

$$\mathbf{D}_{N,x} = \begin{pmatrix} s_{N,0}(x) & s_{N,1}(x) \\ s_{N,1}(x) & s_{N,2}(x) \end{pmatrix}$$

with $s_{N,l}(x) = N^{-1}\sum_{j=1}^{N} K_h(j/N - x)\{(j/N - x)/h\}^l$, $l = 0, 1, 2$. Denote

$$\mathbf{D}_x = \begin{pmatrix} \mu_{0,x}(K) & \mu_{1,x}(K) \\ \mu_{1,x}(K) & \mu_{2,x}(K) \end{pmatrix} \text{ and } D_x(K) = \mu_{2,x}(K)\,\mu_{0,x}(K) - \mu_{1,x}^2(K),$$

where

$$\mu_{l,x}(K) = \begin{cases} \int_{-x/h}^{1} v^l K(v)\,dv, & x \in [0, h); \\ \int_{-1}^{1} v^l K(v)\,dv, & x \in [h, 1-h]; \\ \int_{-1}^{(1-x)/h} v^l K(v)\,dv, & x \in (1-h, 1]. \end{cases}$$

Then it is obvious that

$$\mathbf{D}_x^{-1} = D_x^{-1}(K)\begin{pmatrix} \mu_{2,x}(K) & -\mu_{1,x}(K) \\ -\mu_{1,x}(K) & \mu_{0,x}(K) \end{pmatrix}.$$

Lemma A.2. *Under Assumption (A3), for* $x \in [0, 1]$

$$0 < D_0(K) \le D_x(K) \le D_{1/2}(K) = \mu_{2,1/2}(K) < \infty, \tag{A.1}$$

and hence $\sup_{x\in[0,1]}|\mathbf{D}_x^{-1}| < \infty$.

Proof. It is obvious that $D_x(K) = \mu_{2,x}(K) > 0$ for any $x \in [h, 1-h]$. For $x \in [0, h)$,

$$D_x(K) = \frac{1}{2}\int_{-x/h}^{1} u^2 K(u)\,\mathrm{d}u \int_{-x/h}^{1} K(v)\,\mathrm{d}v + \frac{1}{2}\int_{-x/h}^{1} K(u)\,\mathrm{d}u \int_{-x/h}^{1} v^2 K(v)\,\mathrm{d}v$$
$$- \int_{-x/h}^{1} uK(u)\,\mathrm{d}u \int_{-x/h}^{1} vK(v)\,\mathrm{d}v$$

$$= \int_{-x/h}^{1} \int_{-x/h}^{1} \frac{(u-v)^2}{2} K(u) K(v) \, \mathrm{d}u \mathrm{d}v$$
$$\geq \int_{0}^{1} \int_{0}^{1} \frac{(u-v)^2}{2} K(u) K(v) \, \mathrm{d}u \mathrm{d}v = D_0(K).$$

Hence $D_x(K) \geq D_0(K) > 0$. Similarly, we can show that it is true for $x \in [1-h, 1]$. The upper bound for $D_x(K)$ follows likewise. The proof is complete. □

Proof of Proposition 1. It is easy to get $\mathbf{D}_{N,x} = \mathbf{D}_x + U(h)$, so

$$\left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} = \mathrm{diag}(1, h^{-1})\{\mathbf{D}_x^{-1} + U(h)\}\, \mathrm{diag}(1, h^{-1}).$$

Without loss of generality, let $x \in [h, 1-h]$. We have

$$\mathrm{diag}\left(1, h^{-1}\right) \mathbf{X}^T \mathbf{W} \left\{\mathbf{m} - m(x) \mathbf{X}\mathbf{e}_0 - m'(x) \mathbf{X}\mathbf{e}_1\right\}$$
$$= \begin{pmatrix} N^{-1} \sum_{j=1}^{N} K_h(j/N - x) \left\{m(j/N) - m(x) - m'(x)(j/N - x)\right\} \\ N^{-1} \sum_{j=1}^{N} K_h(j/N - x) \{(j/N - x)/h\} \left\{m(j/N) - m(x) - m'(x)(j/N - x)\right\} \end{pmatrix}$$
$$= \begin{pmatrix} N^{-1} \sum_{j=1}^{N} K_h(j/N - x) \left\{\frac{1}{2} m''(x)(j/N - x)^2 + u\left(h^2\right)\right\} \\ N^{-1} \sum_{j=1}^{N} K_h(j/N - x) \{(j/N - x)/h\} \left\{\frac{1}{2} m''(x)(j/N - x)^2 + u\left(h^2\right)\right\} \end{pmatrix}$$
$$= \frac{1}{2} m''(x) h^2 \begin{pmatrix} \mu_{2,x}(K) + u(1) \\ \mu_{3,x}(K) + u(1) \end{pmatrix}. \tag{A.2}$$

Therefore, according to (A.2) and Lemma A.2, after some manipulation, we obtain

$$\widetilde{m}(x) - m(x)$$
$$= \mathbf{e}_0^T \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \left\{\mathbf{m} - m(x) \mathbf{X}\mathbf{e}_0 - m'(x) \mathbf{X}\mathbf{e}_1\right\}$$
$$= \mathbf{e}_0^T \, \mathrm{diag}\left(1, h^{-1}\right) \left\{\mathbf{D}_x^{-1} + U(h)\right\} \mathrm{diag}\left(1, h^{-1}\right) \mathbf{X}^T \mathbf{W} \left\{\mathbf{m} - m(x) \mathbf{X}\mathbf{e}_0 - m'(x) \mathbf{X}\mathbf{e}_1\right\}$$
$$= \frac{1}{2} m''(x) h^2 \mathbf{e}_0^T \, \mathrm{diag}\left(1, h^{-1}\right) \mathbf{D}_x^{-1} \begin{pmatrix} \mu_{2,x}(K) + u(1) \\ \mu_{3,x}(K) + u(1) \end{pmatrix} + U\left(h^3\right)$$
$$= U\left(h^2\right)$$

with $\mathbf{e}_1^T = (0, 1)$. By Assumption (A2), $\sqrt{n}\{\widetilde{m}(x) - m(x)\} = U(n^{1/2} h^2) = o(1)$. The proof is complete. □

A.3. Proof of Proposition 2 and Theorem 1

Lemma A.3. *Under Assumption (A5), for $C_0 = C_1(1 + \rho C_2 \sum_{t=1}^{\infty} t^{\rho-1-\gamma_1})$ and $n \geq 1$*

$$\max_{1 \leq k \leq \kappa} E\left|\bar{\xi}_{.k} - \bar{Z}_{.k,\xi}\right| \leq C_0 n^{\rho-1}, \tag{A.3}$$

$$\max_{1\le j\le N}\left|\bar{\varepsilon}_{.j}-\bar{Z}_{.j,\varepsilon}\right| = O_{a.s.}\left(n^{\rho-1}\right), \tag{A.4}$$

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

$$\max_{1\le k\le\kappa} E\left|\bar{\xi}_{.k}\right| \le n^{-1/2}(2/\pi)^{1/2} + C_0 n^{\rho-1}. \tag{A.5}$$

Proof. See Cao et al. (2012), Lemma A.5. □

Lemma A.4. *For any* $\phi \in C[0,1]$, $\|\tilde{\phi}\|_\infty \le \|\phi\|_\infty$. *Furthermore, if* $\phi \in C^{0,\beta}[0,1]$ *for some* $\beta \in (0,1]$, $\|\tilde{\phi}-\phi\|_\infty \le \|\phi\|_{0,\beta} h^\beta$.

Proof. The first statement is trivial by the definition of $\|\phi\|_\infty$. By the definition of $\|\phi\|_{0,\beta}$, for any $x \in [0,1]$,

$$\begin{aligned}\left|\tilde{\phi}(x)-\phi(x)\right| &= \left|\frac{\sum_{j=1}^{N} K(j/N-x)\{\phi(j/N)-\phi(x)\}}{\sum_{j=i}^{N} K(j/N-x)}\right| \\ &\le \sup_{x\in[0,1]} |\phi(j/N)-\phi(x)| \le h^\beta \|\phi\|_{0,\beta}.\end{aligned}$$

The proof is complete. □

Lemma A.5. *Assumption (A5′) implies Assumption (A5).*

Proof. See Cao et al. (2012), Lemma A.6. □

Denote $\widetilde{\xi}(x) = \sum_{k=1}^{\kappa}\widetilde{\xi}_k(x)$, where $\widetilde{\xi}_k(x) = \bar{\xi}_{.k}\mathbf{e}_0^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\phi_k$ and $\phi_k = (\phi_k(1/N), \ldots, \phi_k(N/N))^T$. Let $\widetilde{\phi}_k(x)$ be the solution to the least square problem

$$\arg\min_{(a,b)}\sum_{j=1}^{N}\{\phi_k(j/N)-a-b(j/N-x)\}^2 K_h(j/N-x),$$

i.e., $\widetilde{\phi}_k(x) = \mathbf{e}_0^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\phi_k$, so $\widetilde{\xi}_k(x) = \bar{\xi}_{.k}\widetilde{\phi}_k(x)$. Denote $\widetilde{Z}_k(x) = \bar{Z}_{.k,\xi}\widetilde{\phi}_k(x)$, $k = 1, \ldots, \kappa$, similar to the definition of $\widetilde{m}(x)$ and $\widetilde{\xi}_k(x)$ in (3.9). Also denote $\widetilde{\zeta}_k(x) = \bar{Z}_{.k,\xi}\phi_k(x)$, $k = 1, \ldots, \kappa$ and define

$$\widetilde{\zeta}(x) = n^{1/2}\left\{\sum_{k=1}^{\kappa}\phi_k^2(x)\right\}^{-1/2}\sum_{k=1}^{\kappa}\widetilde{\zeta}_k(x) = n^{1/2}G(x,x)^{-1/2}\sum_{k=1}^{\kappa}\widetilde{\zeta}_k(x). \tag{A.6}$$

Proof of Proposition 2 and Theorem 1. Applying the above Lemma A.4 and Assumptions (A4), (A5), one has

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

$$\leq C\left\{\sum_{k=1}^{\kappa_n}\|\phi_k\|_{0,\beta}\,h^{\beta}+\sum_{k=\kappa_n+1}^{\kappa}\|\phi_k\|_{\infty}\right\}=o\,(1)\,.$$

In addition, (A.3) and Assumptions (A3)–(A4) entail that

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

$$\leq\max_{1\leq k\leq\kappa}E|\bar{\xi}_{.k}-\bar{Z}_{.k,\xi}|\sum_{k=1}^{\kappa}\|\phi_k\|_{\infty}$$

$$\leq Cn^{\rho-1/2}\sum_{k=1}^{\kappa}\|\phi_k\|_{\infty}=o\,(1)\,.$$

Hence

$$n^{1/2}\sup_{x\in[0,1]}G\,(x,x)^{-1/2}\left|\sum_{k=1}^{\kappa}\bar{\xi}_{.k}\left\{\phi_k(x)-\tilde{\phi}_k(x)\right\}\right|=o_p\,(1)\,, \tag{A.7}$$

$$n^{1/2}\sup_{x\in[0,1]}G\,(x,x)^{-1/2}\left|\sum_{k=1}^{\kappa}\left(\bar{Z}_{.k,\xi}-\bar{\xi}_{.k}\right)\phi_k(x)\right|=o_p\,(1)\,. \tag{A.8}$$

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

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

It is clear that $\tilde{\zeta}(x)$ is a Gaussian process with mean 0, variance 1, and covariance $E\tilde{\zeta}(x)\tilde{\zeta}(x')=G(x,x)^{-1/2}G(x,x')^{-1/2}G(x,x')$, for any $x,x'\in[0,1]$. Thus $\tilde{\zeta}(x)$ has the same distribution as $\zeta(x)$. Note that by (2.2),

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

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

Hence

$$n^{1/2}\sup_{x\in[0,1]}G\,(x,x)^{-1/2}\left|\bar{m}(x)-m\,(x)-\tilde{\xi}(x)\right|=o_P\,(1)\,, \tag{A.9}$$

$$\sup_{x\in[0,1]}\left|\tilde{\zeta}\,(x)-n^{1/2}G\,(x,x)^{-1/2}\{\bar{m}(x)-m\,(x)\}\right|=o_P\,(1)\,, \tag{A.10}$$

according to (A.7) and (A.8). Then Theorem 1 follows from (A.10). Theorem 1 and (A.9) lead to both (3.10) and (3.11). □

A.4. Proof of Proposition 3

Since $G(x,x)$ is bounded, we only need to consider $\sup_{x\in[0,1]}|\widetilde{e}(x)|$. Notice that

$$\begin{aligned}\widetilde{e}(x) &= \mathbf{e}_0^T\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{e}\\ &= \mathbf{e}_0^T\,\mathrm{diag}\left(1,h^{-1}\right)\left\{\mathbf{D}_x^{-1}+U(h)\right\}\mathrm{diag}\left(1,h^{-1}\right)\mathbf{X}^T\mathbf{W}\mathbf{e}\\ &= V_{N,h}(x)\{C_0+U(h)\},\end{aligned}$$

where $V_{N,h}(x)=N^{-1}\sum_{j=1}^N K_h(j/N-x)\bar{\varepsilon}_{.j}$. Define

$$R_{N,h}(x)=N^{-1}\sum_{j=1}^N K_h\left(j/N-x\right)\bar{Z}_{.j,\varepsilon},$$

where $\bar{Z}_{.j,\varepsilon}$ is defined in Lemma A.2. According to Lemma A.2, $\{\bar{Z}_{.j,\varepsilon}, 1\le j\le N\}$ are independent and identically distributed as the normal distribution $N(0,1/n)$. Hence,

$$\sup_{x\in[0,1]}|\widetilde{e}(x)|\le\left(\sup_{x\in[0,1]}\left|V_{N,h}(x)-R_{N,h}(x)\right|+\sup_{x\in[0,1]}\left|R_{N,h}(x)\right|\right)\{C_0+U(h)\}.$$

We will establish that $\sup_{x\in[0,1]}|R_{N,h}(x)|=o_p(n^{-1/2})$ and

$$\sup_{x\in[0,1]}\left|V_{N,h}(x)-R_{N,h}(x)\right|=O_p(n^{\nu-1})=o_p(n^{-1/2}).$$

In fact for any $x\in[0,1]$,

$$\begin{aligned}\left|V_{N,h}(x)-R_{N,h}(x)\right| &\le \max_{1\le j\le N}\left|\bar{\varepsilon}_{.j}-\bar{Z}_{.j,\varepsilon}\right|(Nh)^{-1}\|K\|_\infty\sum_{j=1}^N I\left(|j/N-x|\le h\right)\\ &\le C\max_{1\le j\le N}\left|\bar{\varepsilon}_{.j}-\bar{Z}_{.j,\varepsilon}\right|\|K\|_\infty=O_p(n^{\nu-1})=o_p(n^{-1/2}).\end{aligned}$$

We discretize the interval [0, 1] and partition it into $N^*=\sqrt{N/h^4}$ subintervals $\{I_k\}$ of equal length. Let x_k be the center of I_k.

$$\begin{aligned}\sup_{x\in[0,1]}\left|R_{N,h}(x)\right| &\le \max_{1\le k\le N^*}\left|R_{N,h}(x_k)\right|+\max_{1\le k\le N^*}\sup_{x\in I_k}\left|R_{N,h}(x)-R_{N,h}(x_k)\right|\\ &\le \max_{1\le k\le N^*}\left|R_{N,h}(x_k)\right|+\max_{1\le j\le N}\left|\bar{Z}_{.j,\varepsilon}\right|\frac{C}{N^*h^2}.\end{aligned}\tag{A.11}$$

Let Φ denote the cumulative distribution function of the standard normal distribution $N(0,1)$, then there exists a constant $C_\Phi>0$ such that for any sufficiently large $t>0$,

$$1-\Phi(t)\le C_\Phi e^{-t^2/2}.$$

Hence

$$\max_{1\le j\le N}P\left\{\left|n^{1/2}\bar{Z}_{.j,\varepsilon}\right|>4\sqrt{\log N}\right\}\le C_\Phi e^{-\left(4\sqrt{\log N}\right)^2/2}=C_\Phi N^{-8},$$

which implies

$$P\left\{\max_{1\le j\le N}\left|n^{1/2}\bar{Z}_{.j,\varepsilon}\right| > 4\sqrt{\log N}\right\} \le N \times C_\Phi N^{-8} = C_\Phi N^{-7}.$$

Thus,

$$\sum_{n=1}^{\infty} P\left\{\max_{1\le j\le N}\left|n^{1/2}\bar{Z}_{.j,\varepsilon}\right| > 4\sqrt{\log N}\right\} < C_\Phi \sum_{n=1}^{\infty} n^{-7/4} < \infty.$$

Under Assumption (A3),

$$\begin{aligned}\frac{1}{N^* h^2}\max_{1\le j\le N}\left|\bar{Z}_{.j,\varepsilon}\right| &= O_{a.s.}\left(\sqrt{\log N}/(n^{-1/2}h^2 N^*)\right)\\ &= o_{a.s}(n^{-1/2}). \qquad \text{(A.12)}\end{aligned}$$

Note that for any $k = 1, \ldots, N^*$, $\frac{R_{N,h}(x_k)}{\sqrt{N^{-2}n^{-1}\sum_{j=1}^{N}\{K_h(j/N - x_k)\}^2}} \sim N(0, 1)$. Thus,

$$\max_{1\le k\le N^*}\left|\frac{R_{N,h}(x_k)}{\sqrt{N^{-2}n^{-1}\sum_{j=1}^{N}\{K_h(j/N - x_k)\}^2}}\right| = O_{a.s}\left(\sqrt{\log N^*}\right),$$

and thus under Assumption (A3),

$$\max_{1\le k\le N^*}\left|R_{N,h}(x_k)\right| \le O_{a.s}\left(\sqrt{N^{-1}h^{-1}n^{-1}\log N^*}\right) = o_{a.s}\left(n^{-1/2}\right). \qquad \text{(A.13)}$$

From (A.11), (A.12), and (A.13), $\sup_{x\in[0,1]}|R_{N,h}(x)| = o_p(n^{-1/2})$. The proof is complete.

A.5. Proof of Theorem 2

According to decomposition of $\hat{m}(x)$, we have

$$\begin{aligned}\bar{m}(x) - \hat{m}(x) &= \{\bar{m}(x) - m(x)\} - \left\{\widetilde{m}(x) - m(x) + \widetilde{\xi}(x) + \widetilde{e}(x)\right\}\\ &= \left\{\bar{m}(x) - m(x) - \widetilde{\xi}(x)\right\} - \{\widetilde{m}(x) - m(x)\} - \widetilde{e}(x).\end{aligned}$$

From Proposition 1, Proposition 3, and (A.9), $\sup_{x\in[0,1]}|\bar{m}(x) - \hat{m}(x)| = o_p(n^{-1/2})$ follows immediately.