UNIFORM CONVERGENCE RATES FOR NONPARAMETRIC REGRESSION AND PRINCIPAL COMPONENT ANALYSIS IN FUNCTIONAL/LONGITUDINAL DATA

BY YEHUA LI[1] AND TAILEN HSING[2]

University of Georgia and University of Michigan

We consider nonparametric estimation of the mean and covariance functions for functional/longitudinal data. Strong uniform convergence rates are developed for estimators that are local-linear smoothers. Our results are obtained in a unified framework in which the number of observations within each curve/cluster can be of any rate relative to the sample size. We show that the convergence rates for the procedures depend on both the number of sample curves and the number of observations on each curve. For sparse functional data, these rates are equivalent to the optimal rates in nonparametric regression. For dense functional data, root-n rates of convergence can be achieved with proper choices of bandwidths. We further derive almost sure rates of convergence for principal component analysis using the estimated covariance function. The results are illustrated with simulation studies.

1. Introduction. Estimating the mean and covariance functions are essential problems in longitudinal and functional data analysis. Many recent papers focused on nonparametric estimation so as to model the mean and covariance structures flexibly. A partial list of such work includes Ramsay and Silverman (2005), Lin and Carroll (2000), Wang (2003), Yao, Müller and Wang (2005a, 2005b), Yao and Lee (2006) and Hall, Müller and Wang (2006).

On the other hand, functional principal component analysis (FPCA) based on nonparametric covariance estimation has become one of the most common dimension reduction approaches in functional data analysis. Applications include temporal trajectory interpolation [Yao, Müller and Wang (2005a)], functional generalized linear models [Müller and Stadtmüller (2005) and Yao, Müller and Wang (2005b)] and functional sliced inverse regression [Férre and Yao (2005), Li and Hsing (2010)], to name a few. A number of algorithms have been proposed for FPCA, some of which are based on spline smoothing [James, Hastie and Sugar (2000), Zhou, Huang and Carroll (2008)] and others based on kernel smoothing

[Yao, Müller and Wang (2005a), Hall, Müller and Wang (2006)]. As usual, large-sample theories can provide a basis for understanding the properties of these estimators. So far, the asymptotic theories for estimators based on kernel smoothing or local-polynomial smoothing are better understood than those based on spline smoothing.

Some definitive theoretical findings on FPCA emerged in recent years. In particular, Hall and Hosseini-Nasab (2006) proved various asymptotic expansions for FPCA for densely recorded functional data, and Hall, Müller and Wang (2006) established the optimal L^2 convergence rate for FPCA in the sparse functional data setting. One of the most interesting findings in Hall, Müller and Wang (2006) was that the estimated eigenfunctions, although computed from an estimated two-dimensional surface, enjoy the convergence rate of one-dimensional smoothers, and under favorable conditions the estimated eigenvalues are root-n consistent. In contrast with the L^2 convergence rates of these nonparametric estimators, less is known in term of uniform convergence rates. Yao, Müller and Wang (2005a) studied the uniform consistency of the estimated mean, covariance and eigenfunctions, and demonstrated that such uniform convergence properties are useful in many settings; some other examples can also be found in Li et al. (2008).

In classical nonparametric regression where observations are independent, there are a number of well-known results concerning the uniform convergence rates of kernel-based estimators. Those include Bickel and Rosenblatt (1973), Härdle, Janssen and Serfling (1988) and Härdle (1989). More recently, Claeskens and Van Keilegom (2003) extended some of those results to local likelihood estimators and local estimating equations. However, as remarked in Yao, Müller and Wang (2005a), whether those optimal rates can be extended to functional data remains unknown.

In a typical functional data setting, a sample of n curves are observed at a set of discrete points; denote by m_i the number of observations for curve i. The existing literature focuses on two antithetical data types: the first one, referred to as dense functional data, is the case where each m_i is larger than some power of n; the second type, referred to as sparse functional data, is the situation where each m_i is bounded by a finite positive number or follows a fixed distribution. The methodologies used to treat the two situations have been different in the literature. For dense functional data, the conventional approach is to smooth each individual curve first before further analysis; see Ramsay and Silverman (2005), Hall, Müller and Wang (2006) and Zhang and Chen (2007). For sparse functional data, limited information is given by the sparsely sampled observations from each individual curve and hence it is essential to pool the data in order to conduct inference effectively; see Yao, Müller and Wang (2005a) and Hall, Müller and Wang (2006). However, in practice it is possible that some sample curves are densely observed while others are sparsely observed. Moreover, in dealing with real data, it may even be difficult to classify which scenario we are faced with and hence to decide which methodology to use.

This paper is aimed at resolving the issues raised in the previous two paragraphs. The precise goals will be stated after we introduce the notation in Section 2. In a nutshell, we will consider uniform rates of convergence of the mean and the covariance functions, as well as rates in the ensuing FPCA, using local-linear smoothers [Fan and Gijbels (1995)]. The rates that we obtain will address all possible scenarios of the m_i's, and we show that the optimal rates for dense and sparse functional data can be derived as special cases.

This paper is organized as follows. In Section 2, we introduce the model and data structure as well as all of the estimation procedures. We describe the asymptotic theory of the procedures in Section 3, where we also discuss the results and their connections to prominent results in the literature. Some simulation studies are provided in Section 4, and all proofs are included in Section 5.

2. Model and methodology. Let $\{X(t), t \in [a, b]\}$ be a stochastic process defined on a fixed interval $[a, b]$. Denote the mean and covariance function of the process by

$$\mu(t) = \mathbb{E}\{X(t)\}, \qquad R(s, t) = \mathrm{cov}\{X(s), X(t)\},$$

which are assumed to exist. Except for smoothness conditions on μ and R, we do not impose any parametric structure on the distribution of X. This is a commonly considered situation in functional data analysis.

Suppose we observe

$$Y_{ij} = X_i(T_{ij}) + U_{ij}, \qquad i = 1, \ldots, n, j = 1, \ldots, m_i,$$

where the X_i's are independent realizations of X, the T_{ij}'s are random observational points with density function $f_T(\cdot)$, and the U_{ij}'s are identically distributed random errors with mean zero and finite variance σ^2. Assume that the X_i's, T_{ij}'s and U_{ij}'s are all independent. Assume that $m_i \geq 2$ and let $N_i = m_i(m_i - 1)$.

Our approach is based on the local-linear smoother; see, for example, Fan and Gijbels (1995). Let $K(\cdot)$ be a symmetric probability density function on $[0, 1]$ and $K_h(t) = (1/h)K(t/h)$ where h is bandwidth. A local-linear estimator of the mean function is given by $\widehat{\mu}(t) = \widehat{a}_0$, where

$$(\widehat{a}_0, \widehat{a}_1) = \operatorname*{arg\,min}_{a_0, a_1} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \{Y_{ij} - a_0 - a_1(T_{ij} - t)\}^2 K_{h_\mu}(T_{ij} - t).$$

It is easy to see that

(2.1)
$$\widehat{\mu}(t) = \frac{R_0 S_2 - R_1 S_1}{S_0 S_2 - S_1^2},$$

where

$$S_r = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} K_{h_\mu}(T_{ij} - t)\{(T_{ij} - t)/h_\mu\}^r,$$

$$R_r = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} K_{h_\mu}(T_{ij} - t)\{(T_{ij} - t)/h_\mu\}^r Y_{ij}.$$

To estimate $R(s, t)$, we first estimate $C(s, t) := \mathbb{E}\{X(s)X(t)\}$. Let $\widehat{C}(s, t) = \widehat{a}_0$, where

$$
\begin{aligned}
(\widehat{a}_0, \widehat{a}_1, \widehat{a}_2) \\
= \underset{a_0, a_1, a_2}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \bigg[\frac{1}{N_i} \sum_{k \neq j} \{Y_{ij} Y_{ik} - a_0 \\
- a_1(T_{ij} - s) - a_2(T_{ik} - t)\}^2 \\
\times K_{h_R}(T_{ij} - s) K_{h_R}(T_{ik} - t) \bigg],
\end{aligned}
$$

(2.2)

with $\sum_{k \neq j}$ denoting sum over all $k, j = 1, \ldots, m_i$ such that $k \neq j$. It follows that

$$\widehat{C}(s, t) = (\mathscr{A}_1 R_{00} - \mathscr{A}_2 R_{10} - \mathscr{A}_3 R_{01})\mathscr{B}^{-1},$$

where

$$\mathscr{A}_1 = S_{20}S_{02} - S_{11}^2, \qquad \mathscr{A}_2 = S_{10}S_{02} - S_{01}S_{11}, \qquad \mathscr{A}_3 = S_{01}S_{20} - S_{10}S_{11},$$

$$\mathscr{B} = \mathscr{A}_1 S_{00} - \mathscr{A}_2 S_{10} - \mathscr{A}_3 S_{01},$$

$$S_{pq} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{k \neq j} \left(\frac{T_{ij} - s}{h_R}\right)^p \left(\frac{T_{ik} - t}{h_R}\right)^q K_{h_R}(T_{ij} - s) K_{h_R}(T_{ik} - t),$$

$$R_{pq} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{k \neq j} Y_{ij} Y_{ik} \left(\frac{T_{ij} - s}{h_R}\right)^p \left(\frac{T_{ik} - t}{h_R}\right)^q K_{h_R}(T_{ij} - s) K_{h_R}(T_{ik} - t).$$

We then estimate $R(s, t)$ by

(2.3) $$\widehat{R}(s, t) = \widehat{C}(s, t) - \widehat{\mu}(s)\widehat{\mu}(t).$$

To estimate σ^2, we first estimate $V(t) := C(t, t) + \sigma^2$ by $\widehat{V}(t) = \widehat{a}_0$, where

$$(\widehat{a}_0, \widehat{a}_1) = \underset{a_0, a_1}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \{Y_{ij}^2 - a_0 - a_1(T_{ij} - t)\}^2 K_{h_V}(T_{ij} - t).$$

As in (2.1),

(2.4) $$\widehat{V}(t) = \frac{Q_0 S_2 - Q_1 S_1}{S_0 S_2 - S_1^2},$$

where

$$Q_r = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} K_{h_V}(T_{ij} - t)\{(T_{ij} - t)/h_V\}^r Y_{ij}^2.$$

We then estimate σ^2 by

$$\widehat{\sigma}^2 = \frac{1}{b-a} \int_a^b \{\widehat{V}(t) - \widehat{C}(t,t)\}\, dt.$$

For the problem of mean and covariance estimation, the literature has focused on dense and sparse functional data. The sparse case roughly refers to the situation where each m_i is essentially bounded by some finite number M. Yao, Müller and Wang (2005a) and Hall, Müller and Wang (2006) considered this case and also used local-linear smoothers in their estimation procedures. The difference between the estimators in (2.1), (2.3) and those considered in Yao, Müller and Wang (2005a) and Hall, Müller and Wang (2006) is essentially that we attach weights, m_i^{-1} and N_i^{-1}, to each curve i in the optimizations [although Yao, Müller and Wang (2005a) smoothed the residuals in estimating R]. One of the purposes of those weights is to ensure that the effect that each curve has on the optimizers is not overly affected by the denseness of the observations.

Dense functional data roughly refer to data for which each $m_i \geq M_n \to \infty$ for some sequence M_n, where specific assumptions on the rate of increase of M_n are required for this case to have a distinguishable asymptotic theory in the estimation of the mean and covariance. Hall, Müller and Wang (2006) and Zhang and Chen (2007) considered the so-called "smooth-first-then-estimate" approach, namely, the approach that first preprocesses the discrete functional data by smoothing, and then adopts the empirical estimators of the mean and covariance based on the smoothed functional data. See also Ramsay and Silverman (2005).

As will be seen, our approach is suitable for both sparse and dense functional data. Thus, one particular advantage is that we do not have to discern data type—dense, sparse or mixed—and decide which methodology should be used accordingly. In Section 3, we will provide the convergence rates of $\widehat{\mu}(t)$, $\widehat{R}(s,t)$ and $\widehat{\sigma}^2$, and also those of the estimated eigenvalues and eigenfunctions of the covariance operator of X. The novelties of our results include:

(a) Almost-sure uniform rates of convergence for $\widehat{\mu}(t)$ and $\widehat{R}(s,t)$ over the entire range of s,t will be proved.
(b) The sample sizes m_i per curve will be completely flexible. For the special cases of dense and sparse functional data, these rates match the best known/conjectured rates.

3. Asymptotic theory. To prove a general asymptotic theory, assume that m_i may depend on n as well, namely, $m_i = m_{in}$. However, for simplicity we continue to use the notation m_i. Define

$$\gamma_{nk} = \left(n^{-1} \sum_{i=1}^n m_i^{-k}\right)^{-1}, \qquad k = 1, 2, \ldots,$$

which is the kth order harmonic mean of $\{m_i\}$, and for any bandwidth h,

$$\delta_{n1}(h) = [\{1 + (h\gamma_{n1})^{-1}\}\log n/n]^{1/2}$$

and

$$\delta_{n2}(h) = [\{1 + (h\gamma_{n1})^{-1} + (h^2\gamma_{n2})^{-1}\}\log n/n]^{1/2}.$$

We first state the assumptions. In the following h_μ, h_R and h_V are bandwidths, which are assumed to change with n.

(C1) For some constants $m_T > 0$ and $M_T < \infty$, $m_T \le f_T(t) \le M_T$ for all $t \in [a, b]$. Further, f_T is differentiable with a bounded derivative.

(C2) The kernel function $K(\cdot)$ is a symmetric probability density function on $[-1, 1]$, and is of bounded variation on $[-1, 1]$. Denote $\nu_2 = \int_{-1}^{1} t^2 K(t)\,dt$.

(C3) $\mu(\cdot)$ is twice differentiable and the second derivative is bounded on $[a, b]$.

(C4) All second-order partial derivatives of $R(s, t)$ exist and are bounded on $[a, b]^2$.

(C5) $\mathbb{E}(|U_{ij}|^{\lambda_\mu}) < \infty$ and $\mathbb{E}(\sup_{t \in [a,b]} |X(t)|^{\lambda_\mu}) < \infty$ for some $\lambda_\mu \in (2, \infty)$; $h_\mu \to 0$ and $(h_\mu^2 + h_\mu/\gamma_{n1})^{-1}(\log n/n)^{1-2/\lambda_\mu} \to 0$ as $n \to \infty$.

(C6) $\mathbb{E}(|U_{ij}|^{2\lambda_R}) < \infty$ and $\mathbb{E}(\sup_{t \in [a,b]} |X(t)|^{2\lambda_R}) < \infty$ for some $\lambda_R \in (2, \infty)$; $h_R \to 0$ and $(h_R^4 + h_R^3/\gamma_{n1} + h_R^2/\gamma_{n2})^{-1}(\log n/n)^{1-2/\lambda_R} \to 0$ as $n \to \infty$.

(C7) $\mathbb{E}(|U_{ij}|^{2\lambda_V}) < \infty$ and $\mathbb{E}(\sup_{t \in [a,b]} |X(t)|^{2\lambda_V}) < \infty$ for some $\lambda_V \in (2, \infty)$; $h_V \to 0$ and $(h_V^2 + h_V/\gamma_{n1})^{-1}(\log n/n)^{1-2/\lambda_V} \to 0$ as $n \to \infty$.

The moment condition $\mathbb{E}(\sup_{t \in [a,b]} |X(t)|^\lambda) < \infty$ in (C5)–(C7) hold rather generally; in particular, it holds for Gaussian processes with continuous sample paths [cf. Landau and Shepp (1970)] for all $\lambda > 0$. This condition was also adopted by Hall, Müller and Wang (2006).

3.1. *Convergence rates in mean estimation.* The convergence rate of $\widehat{\mu}(t)$ is given in the following result.

THEOREM 3.1. *Assume that* (C1)–(C3) *and* (C5) *hold. Then*

$$(3.1) \qquad \sup_{t \in [a,b]} |\widehat{\mu}(t) - \mu(t)| = O\big(h_\mu^2 + \delta_{n1}(h_\mu)\big) \qquad a.s.$$

The following corollary addresses the special cases of sparse and dense functional data. For convenience, we use the notation $a_n \lesssim b_n$ to mean $a_n = O(b_n)$.

COROLLARY 3.2. *Assume that* (C1)–(C3) *and* (C5) *hold.*

(a) *If* $\max_{1 \le i \le n} m_i \le M$ *for some fixed* M, *then*

$$(3.2) \qquad \sup_{t \in [a,b]} |\widehat{\mu}(t) - \mu(t)| = O\big(h_\mu^2 + \{\log n/(nh_\mu)\}^{1/2}\big) \qquad a.s.$$

(b) *If* $\min_{1 \leq i \leq n} m_i \geq M_n$ *for some sequence* M_n *where* $M_n^{-1} \lesssim h_\mu \lesssim (\log n / n)^{1/4}$ *is bounded away from* 0, *then*

$$\sup_{t \in [a,b]} |\widehat{\mu}(t) - \mu(t)| = O(\{\log n / n\}^{1/2}) \qquad a.s.$$

The proofs of Theorem 3.1, as the proofs of other results, will be given in Section 5. First, we give a few remarks on these results.

Discussion.

1. On the right-hand side of (3.1), $O(h_\mu^2)$ is a bound for bias while $\delta_{n1}(h_\mu)$ is a bound for $\sup_{t \in [a,b]} |\widehat{\mu}(t) - \mathbb{E}(\widehat{\mu}(t))|$. The derivation of the bias is easy to understand and is essentially the same as in classical nonparametric regression. The derivation of the second bound is more involved and represents our main contribution in this result. To obtain a uniform bound for $|\widehat{\mu}(t) - \mathbb{E}(\widehat{\mu}(t))|$ over $[a, b]$, we first obtained a uniform bound over a finite grid on $[a, b]$, where the grid grows increasingly dense with n, and then showed that the difference between the two uniform bounds is asymptotic negligible. This approach was inspired by Härdle, Janssen and Serfling (1988), which focused on nonparametric regression. One of the main difficulties in our result is that we need to deal within-curve dependence, which is not an issue in classical nonparametric regression. Note that the dependence between $X(t)$ and $X(t')$ typically becomes stronger as $|t - t'|$ becomes smaller. Thus, for dense functional data, the within-curve dependence constitutes an integral component of the overall rate derivation.

2. The sparse functional data setting in (a) of Corollary 3.2 was considered by Yao, Müller and Wang (2005a) and Hall, Müller and Wang (2006). Actually Yao, Müller and Wang (2005a) assumes that the m_i's are i.i.d. positive random variables with $\mathbb{E}(m_i) < \infty$, which implies that $0 < 1/\mathbb{E}(m_i) \leq \mathbb{E}(1/m_i) \leq 1$ by Jensen's inequality; this corresponds to the case where γ_{n1} is bounded away from 0 and also leads to (3.2). The rate in (3.2) is the classical nonparametric rate for estimating a univariate function. We will refer to this as a one-dimensional rate. The one-dimensional rate of $\widehat{\mu}(t)$ was eluded to in Yao, Müller and Wang (2005a) but was not specifically obtained there.

3. Hall, Müller and Wang (2006) and Zhang and Chen (2007) address the dense functional data setting in (b) of Corollary 3.2, where both papers take the approach of first fitting a smooth curve to $Y_{ij}, 1 \leq j \leq m_i$, for each i, and then estimating $\mu(t)$ and $R(s, t)$ by the sample mean and covariance functions, respectively, of the fitted curves. Two drawbacks are:
    (a) Differentiability of the sample curves is required. Thus, for instance, this approach will not be suitable for the Brownian motion, which has continuous but nondifferentiable sample paths.

(b) The sample curves that are included in the analysis need to be all densely observed; those that do not meet the denseness criterion are dropped even though they may contain useful information.

Our approach does not require sample-path differentiability and all of the data are used in the analysis. It is interesting to note that (b) of Corollary 3.2 shows that root-n rate of convergence for $\widehat{\mu}$ can be achieved if the number of observations per sample curve is at least of the order $(n/\log n)^{1/4}$ while a similar conclusion was also reached in Hall, Müller and Wang (2006) for the smooth-first-then-estimate approach.

4. Our nonparametric estimators $\widehat{\mu}$, $\widehat{R}$ and $\widehat{V}$ are based local-linear smoothers, but the methodology and theory can be easily generalized to higher-order local-polynomial smoothers. By the equivalent kernel theory for local-polynomial smoothing [Fan and Gijbels (1995)], higher-order local-polynomial smoothing is asymptotically equivalent to higher-order kernel smthing. Therefore, applying higher-order polynomial smoothing will result in improved rates for the bias under suitable smoothness assumptions. The rate for the variance, on the other hand, will remain the same. In our sparse setting, if pth order local polynomial smoothing is applied under suitable conditions, for some positive integer p, the uniform convergence rate of $\widehat{\mu}(t)$ will become

$$\sup_{t}|\widehat{\mu}(t) - \mu(t)| = O\big(h_{\mu}^{2([p/2]+1)} + \delta_{n1}(h_{\mu})\big) \qquad \text{a.s.,}$$

where $[a]$ denotes the integer part of a. See Claeskens and Van Keilegom (2003) and Masry (1996) for support of this claim in different but related contexts.

3.2. *Convergence rates in covariance estimation.* The following results give the convergence rates for $\widehat{R}(s,t)$ and $\widehat{\sigma}^2$.

THEOREM 3.3. *Assume that* (C1)–(C6) *hold. Then*

$$(3.3) \qquad \sup_{s,t\in[a,b]} |\widehat{R}(s,t) - R(s,t)| = O\big(h_{\mu}^2 + \delta_{n1}(h_{\mu}) + h_R^2 + \delta_{n2}(h_R)\big) \qquad a.s.$$

THEOREM 3.4. *Assume that* (C1), (C2), (C4), (C6) *and* (C7) *hold. Then*

$$(3.4) \quad \widehat{\sigma}^2 - \sigma^2 = O\big(h_R^2 + \delta_{n1}(h_R) + \delta_{n2}^2(h_R) + h_V^2 + \delta_{n1}^2(h_V)\big) \qquad a.s.$$

We again highlight the cases of sparse and dense functional data.

COROLLARY 3.5. *Assume that* (C1)–(C7) *hold.*

(a) *Suppose that* $\max_{1\leq i\leq n} m_i \leq M$ *for some fixed M. If* $h_R^2 \lesssim h_{\mu} \lesssim h_R$, *then*

$$(3.5) \qquad \sup_{s,t\in[a,b]} |\widehat{R}(s,t) - R(s,t)| = O\big(h_R^2 + \{\log n/(nh_R^2)\}^{1/2}\big) \qquad a.s.$$

If $h_V + (\log n/n)^{1/3} \lesssim h_R \lesssim h_V^2 n/\log n$, *then*

$$\widehat{\sigma}^2 - \sigma^2 = O\big(h_R^2 + \{\log n/(nh_R)\}^{1/2}\big) \qquad a.s.$$

(b) *If* $\min_{1 \leq i \leq n} m_i \geq M_n$ *for some sequence* M_n *where* $M_n^{-1} \lesssim h_\mu, h_R, h_V \lesssim$ $(\log n / n)^{1/4}$, *then both* $\sup_{s,t \in [a,b]} |\widehat{R}(s,t) - R(s,t)|$ *and* $\widehat{\sigma}^2 - \sigma^2$ *are* $O(\{\log n / n\}^{1/2})$ *a.s.*

Discussion.

1. The rate in (3.5) is the classical nonparametric rate for estimating a surface (bivariate function), which will be referred to as a two-dimensional rate. Note $\widehat{\sigma}^2$ has a one-dimensional rate in the sparse setting, while both $\widehat{R}(s,t)$ and $\widehat{\sigma}^2$ have root-n rates in the dense setting. Most of the discussions in Section 3.1 obviously also apply here and will not be repeated.
2. Yao, Müller and Wang (2005a) smoothed the products of residuals instead of $Y_{ij} Y_{ik}$ in the local linear smoothing algorithm in (2.2). There is some evidence that a slightly better rate can be achieved in that procedure. However, we were not successful in establishing such a rate rigorously.

3.3. *Convergence rates in FPCA.* By (C5), the covariance function has the spectral decomposition

$$R(s,t) = \sum_{j=1}^{\infty} \omega_j \psi_j(s) \psi_j(t),$$

where $\omega_1 \geq \omega_2 \geq \cdots \geq 0$ are the eigenvalues of $R(\cdot, \cdot)$ and the ψ_j's are the corresponding eigenfunctions. The ψ_j's are also known as the functional principal components. Below, we assume that the nonzero ω_j's are distinct.

Suppose $\widehat{R}(s,t)$ is the covariance estimator given in Section 2, and it admits the following spectral decomposition:

$$\widehat{R}(s,t) = \sum_{j=1}^{\infty} \widehat{\omega}_j \widehat{\psi}_j(s) \widehat{\psi}_j(t),$$

where $\widehat{\omega}_1 > \widehat{\omega}_2 > \cdots$ are the estimated eigenvalues and the $\widehat{\psi}_j$'s are the corresponding estimated principal components. Computing the eigenvalues and eigenfunctions of an integral operator with a symmetric kernel is a well-studied problem in applied mathematics. We will not get into that aspect of FPCA in this paper.

Notice also that $\psi_j(t)$ and $\widehat{\psi}_j(t)$ are identifiable up to a sign change. As pointed out in Hall, Müller and Wang (2006), this causes no problem in practice, except when we discuss the convergence rate of $\widehat{\psi}_j$. Following the same convention as in Hall, Müller and Wang (2006), we let ψ_j take an arbitrary sign but choose $\widehat{\psi}_j$ such that $\|\widehat{\psi}_j - \psi_j\|$ is minimized over the two signs, where $\|f\| := \{\int f^2(t) \, dt\}^{1/2}$ denotes the usual L^2-norm of a function $f \in L^2[a,b]$.

Below let j_0 be a arbitrary fixed positive constant.

THEOREM 3.6. *Under conditions* (C1)–(C6), *for* $1 \leq j \leq j_0$:

(a) $\widehat{\omega}_j - \omega_j = O((\log n/n)^{1/2} + h_\mu^2 + h_R^2 + \delta_{n1}^2(h_\mu) + \delta_{n2}^2(h_R))$ a.s.;
(b) $\|\widehat{\psi}_j - \psi_j\| = O(h_\mu^2 + \delta_{n1}(h_\mu) + h_R^2 + \delta_{n1}(h_R) + \delta_{n2}^2(h_R))$ a.s.;
(c) $\sup_t |\widehat{\psi}_j(t) - \psi_j(t)| = O(h_\mu^2 + \delta_{n1}(h_\mu) + h_R^2 + \delta_{n1}(h_R) + \delta_{n2}^2(h_R))$ a.s.

Theorem 3.6 is proved by using the asymptotic expansions of eigenvalues and eigenfunctions of an estimated covariance function developed by Hall and Hosseini-Nasab (2006), and by applying the strong uniform convergence rate of $\widehat{R}(s,t)$ in Theorem 3.3. In the special case of sparse and dense functional data, we have the following corollary.

COROLLARY 3.7.   *Assume that* (C1)–(C6) *hold. Suppose that* $\max_{1 \le i \le n} m_i \le M$ *for some fixed* M. *Then the following hold for all* $1 \le j \le j_0$:

(a) *If* $(\log n/n)^{1/2} \lesssim h_\mu, h_R \lesssim (\log n/n)^{1/4}$ *then* $\widehat{\omega}_j - \omega_j = O(\{\log n/n\}^{1/2})$ *a.s.*
(b) *If* $h_\mu + (\log n/n)^{1/3} \lesssim h_R \lesssim h_\mu$ *then both of* $\|\widehat{\psi}_j - \psi_j\|$ *and* $\sup_t |\widehat{\psi}_j(t) - \psi_j(t)|$ *have the rate* $O(h_R^2 + \{\log n/(nh_R)\}^{1/2})$ *a.s.*

If $\min_{1 \le i \le n} m_i \ge M_n$ *for some sequence* M_n *where* $M_n^{-1} \lesssim h_\mu, h_R \lesssim (\log n/n)^{1/4}$, *then, for* $1 \le j \le j_0$, *all of* $\widehat{\omega}_j - \omega_j$, $\|\widehat{\psi}_j - \psi_j\|$ *and* $\sup_t |\widehat{\psi}_j(t) - \psi_j(t)|$ *have the rate* $O(\{\log n/n\}^{1/2})$.

Discussion.

1. Yao, Müller and Wang (2005a, 2005b) developed rate estimates for the quantities in Theorem 3.6. However, they are not optimal. Hall, Müller and Wang (2006) considered the rates of $\widehat{\omega}_j - \omega_j$ and $\|\widehat{\psi}_j - \psi_j\|$. The most striking insight of their results is that for sparse functional data, even though the estimated covariance operator has the two-dimensional nonparametric rate, $\widehat{\psi}_j$ converges at a one-dimensional rate while $\widehat{\omega}_j$ converges at a root-n rate if suitable smoothing parameters are used; remarkably they also established the asymptotic distribution of $\|\widehat{\psi}_j - \psi_j\|$. At first sight, it may seem counter-intuitive that the convergence rates of $\widehat{\omega}_j$ and $\widehat{\psi}_j$ are faster than that of $\widehat{R}$, since $\widehat{\omega}_j$ and $\widehat{\psi}_j$ are computed from $\widehat{R}$. However, this can be easily explained. For example, by (4.9) of Hall, Müller and Wang (2006), $\widehat{\omega}_j - \omega_j = \iint(\widehat{R}(s,t) - R(s,t))\psi_j(s)\psi_j(t)\,ds\,dt +$ lower-order terms; integrating $\widehat{R}(s,t) - R(s,t)$ in this expression results in extra smoothing, which leads to a faster convergence rate.
2. Our almost-sure convergence rates are new. However, for both dense and sparse functional data, the rates on $\widehat{\omega}_j - \omega_j$ and $\|\widehat{\psi}_j - \psi_j\|$ are slightly slower than the in-probability convergence rates obtained in Hall, Müller and Wang (2006), which do not contain the $\log n$ factor at various places of our rate bounds. This is due to the fact that our proofs are tailored to strong uniform convergence rate derivation. However, the general strategy in our proofs is amenable to deriving in-probability convergence rates that are comparable to those in Hall, Müller and Wang (2006).

3. A potential estimator the covariance function $R(s,t)$ is

$$\widetilde{R}(s,t) := \sum_{j=1}^{J_n} \widehat{\omega}_j \widehat{\psi}_j(s)\widehat{\psi}_j(t)$$

for some J_n. For the sparse case, in view of the one-dimensional uniform rate of $\widehat{\psi}_j(t)$ and the root-n rates of $\widehat{\omega}_j$, it might be possible to choose $J_n \to \infty$ so that $\widetilde{R}(s,t)$ has a faster rate of convergence than does $\widehat{R}(s,t)$. However, that requires the rates of $\widehat{\omega}_j$ and $\widehat{\psi}_j(t)$ for an unbounded number of j's, which we do not have at this point.

The proof of the theorems will be given in Section 5, whereas the proofs of the corollaries are straightforward and are omitted.

4. Simulation studies.

4.1. *Simulation* 1. To illustrate the finite sample performance of the method, we perform a simulation study. The data are generated from the following model:

$$Y_{ij} = X_i(T_{ij}) + U_{ij} \qquad \text{with } X_i(t) = \mu(t) + \sum_{k=1}^{3} \xi_{ik}\psi_j(t),$$

where $T_{ij} \sim \text{Uniform}[0,1]$, $\xi_{ik} \sim \text{Normal}(0, \omega_j)$ and $U_{ij} \sim \text{Normal}(0, \sigma^2)$ are independent variables. Let

$$\mu(t) = 5(t-0.6)^2, \qquad \psi_1(t) = 1,$$
$$\psi_2(t) = \sqrt{2}\sin(2\pi t), \qquad \psi_3(t) = \sqrt{2}\cos(2\pi t)$$

and $(\omega_1, \omega_2, \omega_3, \sigma^2) = (0.6, 0.3, 0.1, 0.2)$.

We let $n = 200$ and $m_i = m$ for all i. In each simulation run, we generated 200 trajectories from the model above, and then we compared the estimation results for $m = 5$, 10, 50 and ∞. When $m = \infty$, we assumed that we know the whole trajectory and so no measurement error was included. Note that the cases of $m = 5$ and $m = \infty$ may be viewed as representing sparse and complete functional data, respectively, whereas those of $m = 10$ and $m = 50$ represent scenarios between the two extremes. For each m value, we estimated the mean and covariance functions and used the estimated covariance function to conduct FPCA. The simulation was then repeated 200 times.

For $m = 5, 10, 50$, the estimation was carried out as described in Section 2. For $m = \infty$, the estimation procedure was different since no kernel smoothing is needed; in this case, we simply discretized each curve on a dense grid, then the mean and covariance functions were estimated using the gridded data.

Notice that $m = \infty$ is the ideal situation where we have the complete information of each curve, and the estimation results under this scenario represent the best

	h_μ	h_R	h_V
$m = 5$	0.153	0.116	0.138
$m = 10$	0.138	0.103	0.107
$m = 50$	0.107	0.077	0.084

we can do and all of the estimators have root-n rates. Our asymptotic theory shows that $m \to \infty$ as a function of n, and if m increases with a fast enough rate, the convergence rates for the estimators are also root-n. We intend to demonstrate this based on simulated data.

The performance of the estimators depends on the choice of bandwidths for $\mu(t)$, $C(s, t)$ and $V(t)$, and the best bandwidths vary with m. The bandwidth selection problem turns out to be very challenging. We have not come across a data-driven procedure that works satisfactorily and so this is an important problem for future research. For lack of a better approach, we tried picking the bandwidths by the integrated mean square error (IMSE); that is, for each m and for each function above, we calculated the IMSE over a range of h and selected the one that minimizes the IMSE. The bandwidths picked that way worked quite well for the inference of the mean, covariance and the leading principal components, but less well for σ^2 and the eigenvalues. After experimenting with a number of bandwidths, we decided to used bandwidths that are slightly smaller than the ones picked by IMSE. They are reported in Table 1. Note that undersmoothing in functional principal component analysis was also advocated by Hall, Müller and Wang (2006).

The estimation results for $\mu(\cdot)$ are summarized in Figure 1, where we plot the mean and the pointwise first and 99th percentiles of the estimator. To compare with standard nonparametric regression, we also provide the estimation results for μ when $m = 1$; note that in this case the covariance function is not estimable since there is no within-curve information. As can be seen, the estimation result for $m = 1$ is not very different from that of $m = 5$, reconfirming the nonparametric convergence rate of $\widehat{\mu}$ for sparse functional data. It is somewhat difficult to describe the estimation results of the covariance function directly. Instead, we summarize the results on $\psi_k(\cdot)$ and ω_k in Figure 2, where we plot the mean and the pointwise first and 99th percentiles of the estimated eigenfunctions. In Figure 3, we also show the empirical distributions of $\widehat{\omega}_k$ and $\widehat{\sigma}^2$. In all of the scenarios, the performance of the estimators improve with m; by $m = 50$, all of the the estimators perform almost as well as those for $m = \infty$.

4.2. *Simulation* 2.   To illustrate that the proposed methods are applicable even to the cases that the trajectory of X is not smooth, we now present a second simulation study where X is standard Brownian motion. Again, we set the time
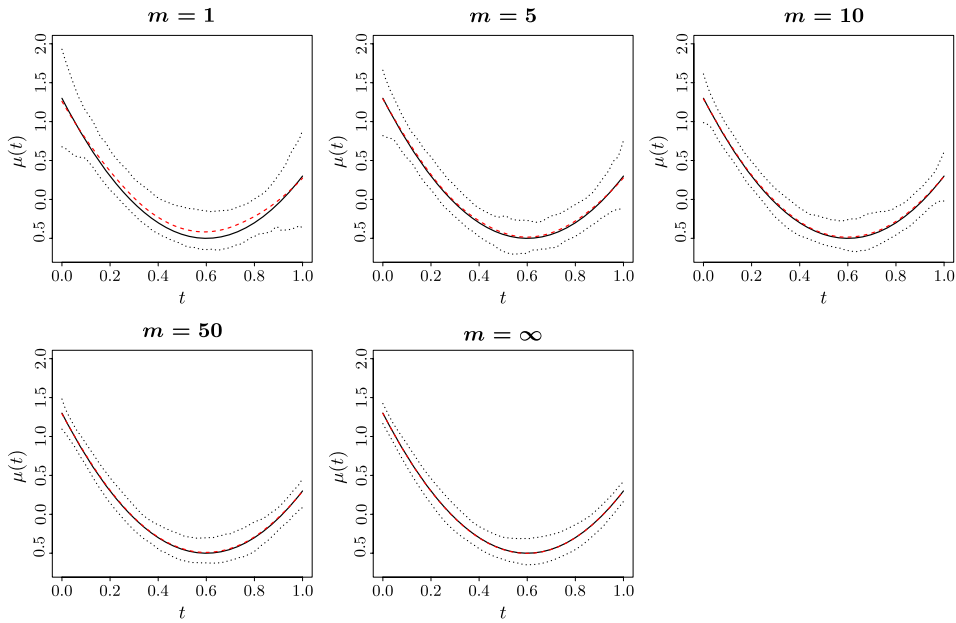
FIG. 1. *Estimated mean function in simulation* 1. *In each panel, the solid line is the true mean function, the dashed line is the pointwise mean and the two dotted lines are the pointwise* 1% *and* 99% *percentiles of the estimator of the mean function based on* 200 *runs.*

window $[a, b]$ to be $[0, 1]$. It is well known that the covariance function of X is $R(s, t) = \min(s, t)$, $s, t \in [0, 1]$, which has an infinite spectral decomposition with

$$\omega_k = 4/\{(2k - 1)^2 \pi^2\}, \qquad \psi_k(t) = \sqrt{2} \sin\{(k - 1/2)\pi t\}, \qquad k = 1, 2, \dots.$$

Again, let the observation times be $T_{ij} \sim \text{Uniform}[0, 1]$, $Y_{ij} = X_i(T_{ij}) + U_{ij}$, $U_{ij} \sim \text{Normal}(0, \sigma^2)$. We let $\sigma^2 = 0.1^2$, which is comparable to ω_3.

Since X is not differentiable with probability one, smoothing individual trajectories is not sensible even for large m values. Also, $R(s, t)$ is not differentiable on the diagonal $\{s = t\}$, and therefore the smoothness assumption in our theory is not satisfied. Nevertheless, as we will show below, the proposed method still works reasonably well. The reason is that the smoothness assumption on $R(s, t)$ in our theory is meant to guarantee the best convergence rate for the $\widehat{R}(s, t)$. When the assumption is mildly violated, the estimator may still perform well overall but may have a slower convergence rate at the nonsmooth points. A similar phenomenon was observed in Li et al. (2007), which studied kernel estimation of a stationary covariance function in a time-series setting.

We set $n = 200$ and $m = 5$, 10 or 50 in our simulations. The estimation results for the first three eigenfunctions are presented in Figure 4. Again, we plot the mean and the pointwise first and 99th percentiles of the estimated eigenfunctions. As can be seen, it is in general much harder to estimate the higher-order eigenfunctions,
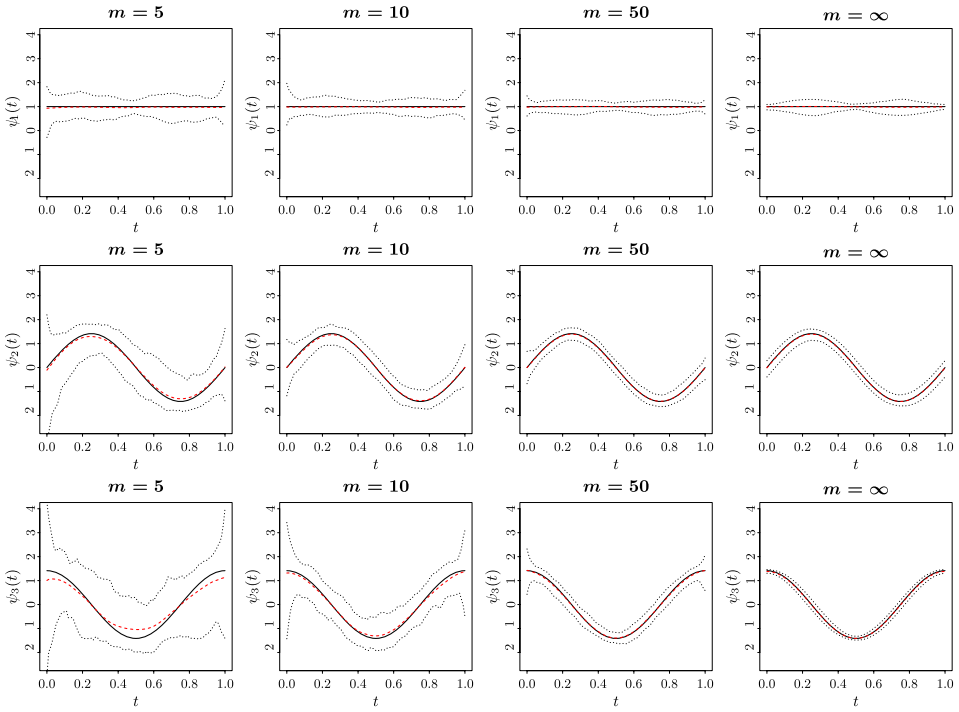
FIG. 2. *Estimated eigenfunctions in simulation* 1. *In each panel, the solid line is the eigenfunction, the dashed line is the pointwise mean and the two dotted lines are the pointwise* 1% *and* 99% *percentiles of the estimator of the eigenfunction in* 200 *runs. The three rows correspond to* ψ_1, ψ_2 *and* ψ_3; *different columns correspond to different m values.*

and the results improve as we increase m. The empirical distribution of the estimated eigenvalues as well as $\widehat{\sigma}^2$ are summarized in Figure 5. The estimated eigenvalues should be compared with the true ones, which are $(0.405, 0.045, 0.016)$. When m is large, the estimated eigenvalues are very close to the true values.

5. Proofs.

5.1. *Proof of Theorem* 3.1. The proof is an adaptation of familiar lines of proofs established in nonparametric function literature; see Claeskens and Van Keilegon (2003) and Härdle, Janssen and Serfling (1988). For simplicity, throughout this subsection, we abbreviate h_μ as h. Below, let $t_1 \wedge t_2 = \min(t_1, t_2)$ and $t_1 \vee t_2 = \max(t_1, t_2)$. Also define $K_{(\ell)}(t) = t^\ell K(t)$ and $K_{h,(\ell)}(v) = (1/h) K_{(\ell)}(v/h)$.

LEMMA 1. *Assume that*

$$(5.1) \quad \mathbb{E}\left(\sup_{t \in [a,b]} |X(t)|^\lambda \right) < \infty \quad and \quad \mathbb{E}|U|^\lambda < \infty \quad for\ some\ \lambda \in (2, \infty).$$
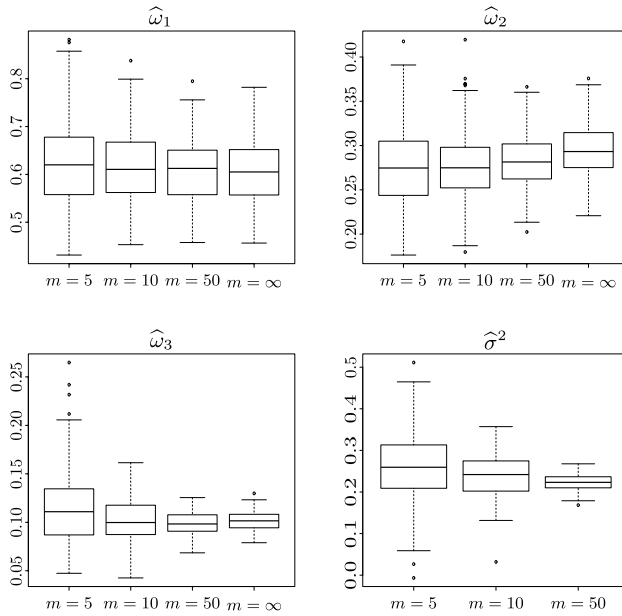
FIG. 3. *Box plots for $\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3$ and $\widehat{\sigma}^2$ in simulation 1.*

Let $\mathscr{Z}_{ij} = X_i(T_{ij})$ or U_{ij} for $1 \leq i \leq n, 1 \leq j \leq m_i$. Let c_n be any positive sequence tending to 0 and $\beta_n = c_n^2 + c_n/\gamma_{n1}$. Assume that $\beta_n^{-1}(\log n/n)^{1-2/\lambda} = o(1)$. Let

$$G_n(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{m_i} \sum_{j=1}^{m_i} \mathscr{Z}_{ij} I(T_{ij} \in [t_1 \wedge t_2, t_1 \vee t_2]) \right\},$$

(5.2)

$$G(t_1, t_2) = \mathbb{E}\{G_n(t_1, t_2)\}$$

and

$$V_n(t, c) = \sup_{|u| \leq c} |G_n(t, t+u) - G(t, t+u)|, \qquad c > 0.$$

Then

(5.3)
$$\sup_{t \in [a,b]} V_n(t, c_n) = O(n^{-1/2} \{\beta_n \log n\}^{1/2}) \qquad a.s.$$

PROOF. We can obviously treat the positive and negative parts of $\mathscr{Z}_{ij}$ separately, and will assume below that $\mathscr{Z}_{ij}$ is nonnegative. Define an equally-spaced grid $\mathscr{G} := \{v_k\}$, with $v_k = a + kc_n$, for $k = 0, \ldots, [(b-a)/c_n]$, and $v_{[(b-a)/c_n]+1} = b$, where $[\cdot]$ denotes the greatest integer part. For any $t \in [a, b]$ and $|u| \leq c_n$, let v_k be a grid point that is within c_n of both t and $t+u$, which
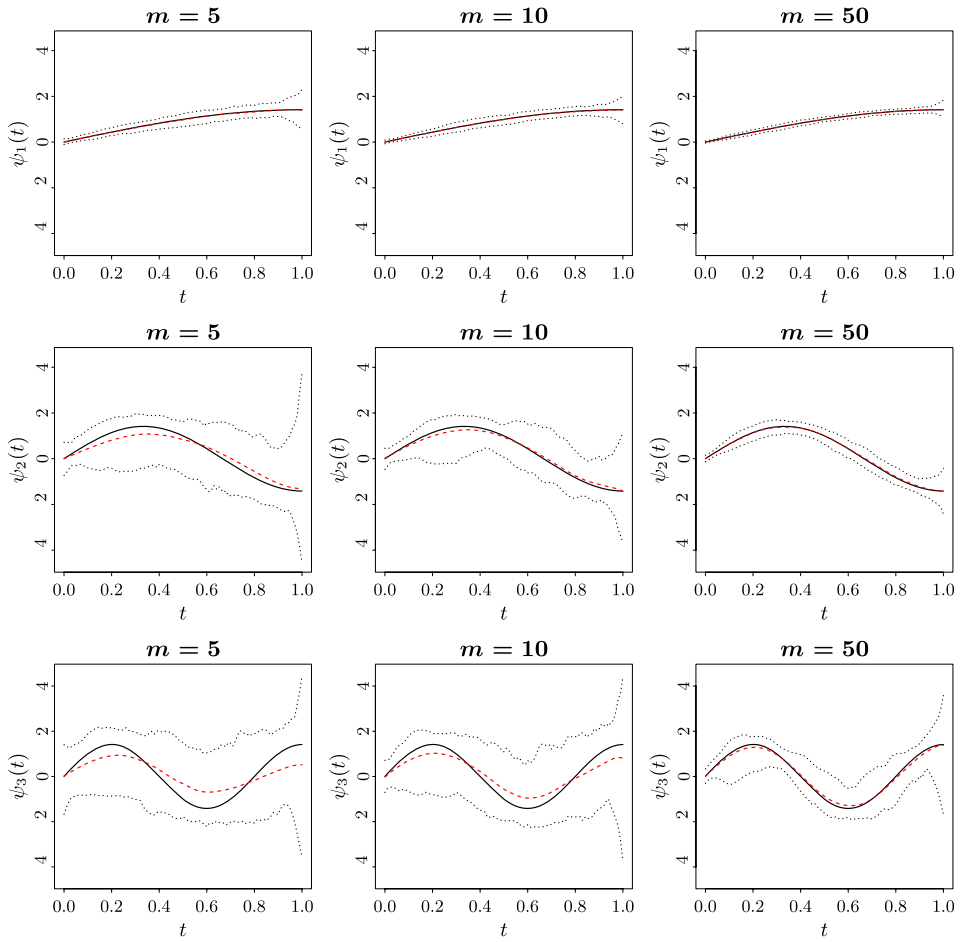
FIG. 4. *Estimated eigenfunctions in simulation 2. In each panel, the solid line is the eigenfunction, the dashed line is the pointwise mean and the two dotted lines are the pointwise 1% and 99% percentiles of the estimator of the eigenfunction in 200 runs. The three rows correspond to ψ_1, ψ_2 and ψ_3; different columns correspond to different m values.*

exists. Since

$$|G_n(t, t+u) - G(t, t+u)| \leq |G_n(v_k, t+u) - G(v_k, t+u)|$$
$$+ |G_n(v_k, t) - G(v_k, t)|,$$

we have

$$|G_n(t, t+u) - G(t, t+u)| \leq 2 \sup_{t \in \mathscr{G}} V_n(t, c_n).$$

Thus,

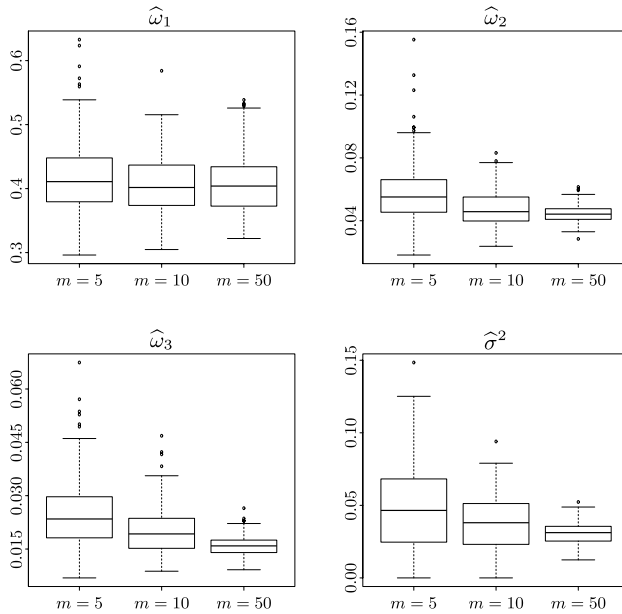(5.4) $$\sup_{t \in [a,b]} V_n(t, c_n) \leq 2 \sup_{t \in \mathscr{G}} V_n(t, c_n).$$

FIG. 5.   *Box plots for $\widehat{\omega}_1$, $\widehat{\omega}_2$, $\widehat{\omega}_3$ and $\widehat{\sigma}^2$ in simulation* 2.

From now on, we focus on the right-hand side of (5.4). Let

$$(5.5) \qquad a_n = n^{-1/2}\{\beta_n \log n\}^{1/2} \quad \text{and} \quad Q_n = \beta_n/a_n,$$

and define $G_n^*(t_1, t_2)$, $G^*(t_1, t_2)$ and $V_n^*(t, c_n)$ in the same way as $G_n(t_1, t_2)$, $G(t_1, t_2)$ and $V_n(t, c_n)$, respectively, except with $\mathscr{Z}_{ij} I(\mathscr{Z}_{ij} \leq Q_n)$ replacing $\mathscr{Z}_{ij}$. Then

$$(5.6) \qquad \sup_{t \in \mathscr{G}} V_n(t, c_n) \leq \sup_{t \in \mathscr{G}} V_n^*(t, c_n) + A_{n1} + A_{n2},$$

where

$$A_{n1} = \sup_{t \in \mathscr{G}} \sup_{|u| \leq c_n} \big(G_n(t, t+u) - G_n^*(t, t+u)\big),$$

$$A_{n2} = \sup_{t \in \mathscr{G}} \sup_{|u| \leq c_n} \big(G(t, t+u) - G^*(t, t+u)\big).$$

We first consider A_{n1} and A_{n2}. It follows that

$$(5.7) \qquad a_n^{-1} Q_n^{1-\lambda} = \{\beta_n^{-1}(\log n/n)^{1-2/\lambda}\}^{\lambda/2} = o(1).$$

For all t and u, by Markov's inequality,

$$a_n^{-1}\big(G_n(t, t+u) - G_n^*(t, t+u)\big)$$

$$\leq a_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i} \mathscr{Z}_{ij} I(\mathscr{Z}_{ij} > Q_n)\right\}$$

$$\leq a_n^{-1} Q_n^{1-\lambda} \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{m_i} \sum_{j=1}^{m_i} \mathscr{Z}_{ij}^{\lambda} I(\mathscr{Z}_{ij} > Q_n) \right\}$$

$$\leq a_n^{-1} Q_n^{1-\lambda} \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{m_i} \sum_{j=1}^{m_i} \mathscr{Z}_{ij}^{\lambda} \right\}.$$

Consider the case $\mathscr{Z}_{ij} = X_i(T_{ij})$, the other case being simpler. It follows that

$$\frac{1}{m_i} \sum_{j=1}^{m_i} \mathscr{Z}_{ij}^{\lambda} \leq W_i \qquad \text{where } W_i = \sup_{t \in [a,b]} |X_i(t)|^{\lambda}.$$

Thus,

(5.8) $$a_n^{-1}\big(G_n(t, t+u) - G_n^*(t, t+u)\big) \leq a_n^{-1} Q_n^{1-\lambda} \frac{1}{n} \sum_{i=1}^{n} W_i.$$

By the SLLN, $n^{-1} \sum_{i=1}^{n} W_i \xrightarrow{\text{a.s.}} \mathbb{E}(\sup_{t \in [a,b]} |X(t)|^{\lambda}) < \infty$. By (5.7) and (5.8), $a_n^{-1} A_{n1} \xrightarrow{\text{a.s.}} 0$. By (5.7) and (5.8) again, $a_n^{-1} A_{n2} = 0$, and so we have proved

(5.9) $$\lim_{n \to \infty} (A_{n1} + A_{n2}) = o(a_n) \qquad \text{a.s.}$$

To bound $V_n^*(t, c_n)$ for a fixed $t \in \mathscr{G}$, we perform a further partition. Define $w_n = [Q_n c_n / a_n + 1]$ and $u_r = r c_n / w_n$, for $r = -w_n, -w_n + 1, \ldots, w_n$. Note that $G_n^*(t, t+u)$ is monotone in $|u|$ since $\mathscr{Z}_{ij} \geq 0$. Suppose that $0 \leq u_r \leq u \leq u_{r+1}$. Then

$$G_n^*(t, t+u_r) - G^*(t, t+u_r) + G^*(t, t+u_r) - G^*(t, t+u_{r+1})$$
$$\leq G_n^*(t, t+u) - G^*(t, t+u)$$
$$\leq G_n^*(t, t+u_{r+1}) - G^*(t, t+u_{r+1}) + G^*(t, t+u_{r+1}) - G^*(t, t+u_r),$$

from which we conclude that

$$|G_n^*(t, t+u) - G^*(t, t+u)| \leq \max(\xi_{nr}, \xi_{n,r+1}) + G^*(t+u_r, t+u_{r+1}),$$

where

$$\xi_{nr} = |G_n^*(t, t+u_r) - G^*(t, t+u_r)|.$$

The same holds if $u_r \leq u \leq u_{r+1} \leq 0$. Thus,

$$V_n^*(t, c_n) \leq \max_{-w_n \leq r \leq w_n} \xi_{nr} + \max_{-w_n \leq r \leq w_n} G^*(t+u_r, t+u_{r+1}).$$

For all r,

$$G^*(t+u_r, t+u_{r+1}) \leq Q_n \mathbb{P}(t+u_r \leq T \leq t+u_{r+1})$$
$$\leq M_T Q_n(u_{r+1} - u_r) \leq M_T a_n.$$

Therefore, for any B,

$$(5.10) \qquad \mathbb{P}\{V_n^*(t, c_n) \geq Ba_n\} \leq \mathbb{P}\Big\{ \max_{-w_n \leq r \leq w_n} \xi_{nr} \geq (B - M_T)a_n \Big\}.$$

Now let $Z_i = m_i^{-1} \sum_{j=1}^{m_i} \mathscr{Z}_{ij} I(\mathscr{Z}_{ij} \leq Q_n) I(T_{ij} \in (t, t + u_r])$ so that $\xi_{nr} = |\frac{1}{n} \times \sum_{i=1}^{n}\{Z_i - \mathbb{E}(Z_i)\}|$. We have $|Z_i - \mathbb{E}(Z_i)| \leq Q_n$, and

$$\sum_{i=1}^{n} \mathrm{var}(Z_i) \leq \sum_{i=1}^{n} \mathbb{E}Z_i^2 \leq M \sum_{i=1}^{n} (c_n^2 + c_n/m_i) \leq Mn\beta_n$$

for some finite M. By Bernstein's inequality,

$$\mathbb{P}\{\xi_{nr} \geq (B - M_T)a_n\} \leq \exp\Big\{ -\frac{(B - M_T)^2 n^2 a_n^2}{2\sum_{i=1}^{n} \mathrm{var}(Z_i) + (2/3)(B - M_T)Q_n n a_n} \Big\}$$

$$\leq \exp\Big\{ -\frac{(B - M_T)^2 n^2 a_n^2}{2Mn\beta_n + (2/3)(B - M_T)n\beta_n} \Big\} \leq n^{-B^*},$$

where $B^* = \frac{(B - M_T)^2}{2M + (2/3)(B - M_T)}$. By (5.10) and Boole's inequality,

$$\mathbb{P}\Big\{ \sup_{t \in \mathscr{G}} V_n^*(t, c_n) \geq Ba_n \Big\} \leq \Big(\Big[\frac{b - a}{c_n}\Big] + 1 \Big)\Big(2\Big[\frac{Q_n c_n}{a_n} + 1\Big] + 1 \Big)n^{-B^*} \leq C\frac{Q_n}{a_n}n^{-B^*}$$

for some finite C. Now $Q_n/a_n = \beta_n/a_n^2 = n/\log n$. So $\mathbb{P}\{V_n^*(t, c_n) \geq Ba_n\}$ is summable in n if we select B large enough such that $B^* > 2$. By the Borel–Cantelli lemma,

$$(5.11) \qquad \sup_{t \in \mathscr{G}} V_n^*(t, c_n) = O(a_n) \qquad \text{a.s.}$$

Hence, (5.3) follows from combining (5.4), (5.6), (5.9) and (5.11). $\quad\square$

LEMMA 2. *Let $\mathscr{Z}_{ij}$ be as in Lemma 1 and assume that (5.1) holds. Let $h = h_n$ be a bandwidth and let $\beta_n = h^2 + h/\gamma_{n1}$. Assume that $h \to 0$ and $\beta_n^{-1}(\log n/n)^{1-2/\lambda} = o(1)$ For any nonnegative integer p, let*

$$D_{p,n}(t) = \frac{1}{n} \sum_{i=1}^{n} \Big[\frac{1}{m_i} \sum_{j=1}^{m_i} K_{h,(p)}(T_{ij} - t)\mathscr{Z}_{ij} \Big].$$

Then we have

$$\sup_{t \in [a,b]} \sqrt{nh^2/(\beta_n \log n)}|D_{p,n}(t) - \mathbb{E}\{D_{p,n}(t)\}| = O(1) \qquad \text{a.s.}$$

PROOF. Since both K and t^p are bounded variations, $K_{(p)}$ is also a bounded variation. Thus, we can write $K_{(p)} = K_{(p),1} - K_{(p),2}$ where $K_{(p),1}$ and $K_{(p),2}$ are both increasing functions; without loss of generality, assume that $K_{(p),1}(-1) =$

$K_{(p),2}(-1) = 0$. Below, we apply Lemma 1 by letting $c_n = 2h$. It is clear that the assumptions of Lemma 1 hold here. Write

$$D_n(t) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{m_i} \sum_{j=1}^{m_i} K_{h,(p)}(T_{ij} - t) \mathscr{Z}_{ij} \right\}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{m_i} \sum_{j=1}^{m_i} \mathscr{Z}_{ij} I(-h \le T_{ij} - t \le h) \int_{-h}^{T_{ij}-t} dK_{h,(p)}(v) \right\}$$

$$= \int_{-h}^{h} \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{m_i} \sum_{j=1}^{m_i} \mathscr{Z}_{ij} I(v \le T_{ij} - t \le h) \right\} dK_{h,(p)}(v)$$

$$= \int_{-h}^{h} G_n(t + v, t + h) \, dK_{h,(p)}(v),$$

where G_n is as defined in (5.2). We have

$$\sup_{t \in [a,b]} |D_{p,n}(t) - \mathbb{E}\{D_{p,n}(t)\}|$$

(5.12)
$$\le \sup_{t \in [a,b]} V_n(t, 2h) \int_{-h}^{h} |dK_{h,(p)}|$$

$$\le \{K_{(p),1}(1) + K_{(p),2}(1)\} h^{-1} \sup_{t \in [a,b]} V_n(t, 2h),$$

and the conclusion of the lemma follows from Lemma 1. $\square$

PROOF OF THEOREM 3.1.   Define

$$R_r^* = R_r - \mu(t) S_r - h\mu^{(1)}(t) S_{r+1}.$$

By straightforward calculations, we have

(5.13)
$$\widehat{\mu}(t) - \mu(t) = \frac{R_0^* S_2 - R_1^* S_1}{S_0 S_2 - S_1^2},$$

where S_0, S_1, S_2 are defined as in (2.1). Write

$$R_r^* = \frac{1}{n} \sum_i \left[\frac{1}{m_i} \sum_{j=1}^{m_i} K_h(T_{ij} - t)\{(T_{ij} - t)/h\}^r \{Y_{ij} - \mu(t) - \mu^{(1)}(t)(T_{ij} - t)\} \right]$$

$$= \frac{1}{n} \sum_i \left[\frac{1}{m_i} \sum_{j=1}^{m_i} K_h(T_{ij} - t)\{(T_{ij} - t)/h\}^r \right.$$

$$\left. \times \{\varepsilon_{ij} + \mu(T_{ij}) - \mu(t) - \mu^{(1)}(t)(T_{ij} - t)\} \right].$$

By Taylor's expansion and Lemma 2, uniformly in t,

$$(5.14) \qquad R_r^* = \frac{1}{n} \sum_i \frac{1}{m_i} \sum_j K_h(T_{ij} - t)\{(T_{ij} - t)/h\}^r \varepsilon_{ij} + O(h^2),$$

and it follows from Lemma 2 that

$$(5.15) \qquad\qquad R_i^* = O\big(h^2 + \delta_{n1}(h)\big) \qquad \text{a.s.}$$

Now, at any interior point $t \in [a + h, b - h]$, since f has a bounded derivative,

$$\mathbb{E}\{S_0\} = \int_{-1}^{1} K(v) f(t + hv)\, dv = f(t) + O(h),$$

$$\mathbb{E}\{S_1\} = O(h), \qquad \mathbb{E}\{S_2\} = f(t)v_2 + O(h),$$

where $v_2 = \int v^2 K(v)\, dv$. By Lemma 2, we conclude that, uniformly for $t \in [a + h, b - h]$,

$$(5.16) \qquad \begin{aligned} S_0 &= f(t) + O\big(h + \delta_{n1}(h)\big), \qquad S_1 = O\big(h + \delta_{n1}(h)\big), \\ S_2 &= f(t)v_2 + O\big(h + \delta_{n1}(h)\big). \end{aligned}$$

Thus, the rate in the theorem is established by applying (5.13). The same rate can also be similarly seen to hold for boundary points. $\square$

5.2. *Proofs of Theorems 3.3 and 3.4.*

LEMMA 3. *Assume that*

$$(5.17) \quad \mathbb{E}\Big(\sup_{t \in [a,b]} |X(t)|^{2\lambda} \Big) < \infty \quad \text{and} \quad \mathbb{E}|U|^{2\lambda} < \infty \qquad \text{for some } \lambda \in (2, \infty).$$

Let $\mathscr{Z}_{ijk}$ be $X(T_{ij})X(T_{ik})$, $X(T_{ij})U_{ik}$ or $U_{ij}U_{ik}$. Let c_n be any positive sequence tending to 0 and $\beta_n = c_n^4 + c_n^3/\gamma_{n1} + c_n^2/\gamma_{n2}$. Assume that $\beta_n^{-1}(\log n/n)^{1-2/\lambda} = o(1)$. Let

$$(5.18) \qquad \begin{aligned} & G_n(s_1, t_1, s_2, t_2) \\ &\quad = \frac{1}{n} \sum_{i=1}^{n} \Big\{ \frac{1}{N_i} \sum_{k \neq j} \mathscr{Z}_{ijk} I(T_{ij} \in [s_1 \wedge s_2, s_1 \vee s_2], \\ &\qquad\qquad\qquad\qquad\qquad T_{ik} \in [t_1 \wedge t_2, t_1 \vee t_2]) \Big\}, \end{aligned}$$

$G(s_1, t_1, s_2, t_2) = \mathbb{E}\{G_n(s_1, t_1, s_2, t_2)\}$ *and*

$$V_n(s, t, \delta) = \sup_{|u_1|, |u_2| \leq \delta} |G_n(s, t, s + u_1, t + u_2) - G(s, t, s + u_1, t + u_2)|.$$

Then

$$\sup_{s, t \in [a,b]} V_n(s, t, c_n) = O(n^{-1/2}\{\beta_n \log n\}^{1/2}) \qquad \text{a.s.}$$

PROOF.   The proof is similar to that of Lemma 1, and so we only outline the main differences. Let a_n, Q_n be as in (5.5). Let $\mathscr{G}$ be a two-dimensional grid on $[a, b]^2$ with mesh c_n, that is, $\mathscr{G} = \{(v_{k_1}, v_{k_2})\}$ where v_k is defined as in the proof of Lemma 1. Then we have

$$(5.19) \qquad \sup_{s,t\in[a,b]} V_n(s, t, c_n) \leq 4 \sup_{(s,t)\in\mathscr{G}} V_n(s, t, c_n).$$

Define $G_n^*(s_1, t_1, s_2, t_2), G^*(s_1, t_1, s_2, t_2)$ and $V_n^*(s, t, \delta)$ in the same way as $G_n(s_1, t_1, s_2, t_2), G(s_1, t_1, s_2, t_2)$ and $V_n(s, t, \delta)$ except with $\mathscr{Z}_{ijk}I(\mathscr{Z}_{ijk} \leq Q_n)$ replacing $\mathscr{Z}_{ijk}$. Then

$$(5.20) \qquad \sup_{(s,t)\in\mathscr{G}} V_n(s, t, c_n) \leq \sup_{(s,t)\in\mathscr{G}} V_n^*(s, t, c_n) + A_{n1} + A_{n2},$$

where

$$A_{n1} = \sup_{(s,t)\in\mathscr{G}} \sup_{|u_1|,|u_2|\leq c_n} |G_n(s, t, s + u_1, t + u_2) - G_n^*(s, t, s + u_1, t + u_2)|,$$

$$A_{n2} = \sup_{(s,t)\in\mathscr{G}} \sup_{|u_1|,|u_2|\leq c_n} |G(s, t, s + u_1, t + u_2) - G^*(s, t, s + u_1, t + u_2)|.$$

Using the technique similar to that in the proof of Lemma 1, we can show A_{n1} and A_{n2} is $o(a_n)$ almost surely. To bound $V_n^*(s, t, c_n)$ for fixed (s, t), we create a further partition. Put $w_n = [Q_n c_n/a_n + 1]$ and $u_r = rc_n/w_n, r = -w_n, \ldots, w_n$. Then

$$V_n^*(s, t, c_n) \leq \max_{-w_n\leq r_1,r_2\leq w_n} \xi_{n,r_1,r_2}$$
$$+ \max_{-w_n\leq r_1,r_2\leq w_n} \{G^*(s, t, s + u_{r_1+1}, t + u_{r_2+1})$$
$$- G^*(s, t, s + u_{r_1}, t + u_{r_2})\},$$

where

$$\xi_{n,r_1,r_2} = |G_n^*(s, t, s + u_{r_1}, t + u_{r_2}) - G^*(s, t, s + u_{r_1}, t + u_{r_2})|.$$

It is easy to see that $\text{var}(\xi_{n,r_1,r_2}) \leq Mn\beta_n$ for some finite M, and the rest of the proof completely mirrors that of Lemma 1 and is omitted. □

LEMMA 4.   Let $\mathscr{Z}_{ijk}$ be as in Lemma 3 and assume that (5.17) holds. Let $h = h_n$ be a bandwidth and let $\beta_n = h^4 + h^3/\gamma_{n1} + h^2/\gamma_{n2}$. Assume that $h \to 0$ and $\beta_n^{-1}(\log n/n)^{1-2/\lambda} = o(1)$. For any nonnegative integers p, q, let

$$D_{p,q,n}(s, t) = \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{N_j}\sum_{k\neq j} \mathscr{Z}_{ijk}K_{h,(p)}(T_{ij} - s)K_{h,(q)}(T_{ik} - t)\right].$$

Then, for any p, q,

$$\sup_{s,t\in[a,b]} \sqrt{nh^4/(\beta_n \log n)}|D_{p,q,n}(s, t) - \mathbb{E}\{D_{p,q,n}(s, t)\}| = O(1) \qquad a.s.$$

PROOF. Write

$$D_{p,q,n}(s,t)$$

$$= \sum_{i=1}^{n}\left[\frac{1}{N_i}\sum_{k\neq j}\mathscr{Z}_{ijk}I(T_{ij}\leq s+h)I(T_{ik}\leq t+h)\right.$$

$$\left.\times K_{h,(p)}(T_{ij}-s)K_{h,(q)}(T_{ik}-t)\right]$$

$$= \int\int_{(u,v)\in[-h,h]^2}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N_i}\sum_{k\neq j}\mathscr{Z}_{ijk}\right.$$

$$\times I(T_{ij}\in[s+u,s+h])$$

$$\left.\times I(T_{ik}\in[t+v,t+h])\right]dK_{h,(p)}(u)\,dK_{h,(q)}(v)$$

$$= \int\int_{(u,v)\in[-h,h]^2}G_n(s+u,t+v,s+h,t+h)\,dK_{h,(p)}(u)dK_{h,(q)}(v),$$

where G_n is as in (5.18). Now,

$$\sup_{(s,t)\in[a,b]^2}|D_{p,q,n}(s,t)-\mathbb{E}\{D_{p,q,n}(s,t)\}|$$

$$\leq \sup_{s,t\in[a,b]}V_n(s,t,2h)\int\int_{(u,v)\in[-h,h]^2}\left|d\{K_{h,(p)}(u)\}\right|\left|d\{K_{h,(q)}(v)\}\right|$$

$$= O[\{\beta_n\log n/(nh^4)\}^{1/2}]\qquad\text{a.s.}$$

by Lemma 3, using the same argument as in (5.12). □

PROOF OF THEOREM 3.3. Let S_{pq}, R_{pq}, $\mathscr{A}_i$ and $\mathscr{B}$ be defined as in (2.3). Also, for $p,q\geq 0$, define

$$R_{pq}^* = R_{pq} - C(s,t)S_{pq} - h_R C^{(1,0)}(s,t)S_{p+1,q} - h_R C^{(0,1)}(s,t)S_{p,q+1}.$$

By straightforward algebra, we have

(5.21) $$(\widehat{C}-C)(s,t) = (\mathscr{A}_1 R_{00}^* - \mathscr{A}_2 R_{10}^* - \mathscr{A}_3 R_{01}^*)\mathscr{B}^{-1}.$$

By standard calculations, we have the following rates uniformly on $[a+h_R,b-h_R]^2$:

$$\mathbb{E}(S_{00}) = f(s)f(t)+O(h_R),\qquad \mathbb{E}(S_{01})=O(h_R),$$

$$\mathbb{E}(S_{10})=O(h_R),\qquad \mathbb{E}(S_{02})=f(s)f(t)v_2+O(h_R),$$

$$\mathbb{E}(S_{20})=f(s)f(t)v_2+O(h_R),\qquad \mathbb{E}(S_{11})=O(h_R).$$

By these and Lemma 4, we have the following almost sure uniform rates:

$$\mathscr{A}_1 = f^2(s) f^2(t) v_2^2 + O(h_R + \delta_{n2}(h_R)),$$

$$\mathscr{A}_2 = O(h_R + \delta_{n2}(h_R)),$$

(5.22)

$$\mathscr{A}_3 = O(h_R + \delta_{n2}(h_R)),$$

$$\mathscr{B} = f^3(s) f^3(t) v_2^2 + O(h_R + \delta_{n2}(h_R)).$$

To analyze the behavior of the components of (5.21), it suffices now to analyze R_{pq}^*. Write

$$R_{00}^* = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{N_i} \sum_{k \neq j} \{Y_{ij} Y_{ik} - C(s,t) \right.$$

$$- C^{(1,0)}(s,t)(T_{ij} - s)$$

$$\left. - C^{(0,1)}(s,t)(T_{ik} - t)\} \times K_{h_R}(T_{ij} - s) K_{h_R}(T_{ik} - t) \right].$$

Let $\varepsilon_{ijk}^* = Y_{ij} Y_{ik} - C(T_{ij}, T_{ik})$. By Taylor's expansion,

$$Y_{ij} Y_{ik} - C(s,t) - C^{(1,0)}(s,t)(T_{ij} - s) - C^{(0,1)}(s,t)(T_{ik} - t)$$

$$= Y_{ij} Y_{ik} - C(s,t) - C(T_{ij}, T_{ik}) + C(T_{ij}, T_{ik})$$

$$- C^{(1,0)}(s,t)(T_{ij} - s) - C^{(0,1)}(s,t)(T_{ik} - t)$$

$$= \varepsilon_{ijk}^* + O(h_R^2) \qquad \text{a.s.}$$

It follows that

(5.23) $$R_{00}^* = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{k \neq j} \varepsilon_{ijk}^* K_{h_R}(T_{ij} - s) K_{h_R}(T_{ik} - t) + O(h_R^2).$$

Applying Lemma 4, we obtain, uniformly in s, t,

(5.24) $$R_{00}^* = O(\delta_{n2}(h_R) + h_R^2) \qquad \text{a.s.}$$

By (5.22),

(5.25) $$\mathscr{A}_1 \mathscr{B}^{-1} = [f(s) f(t)]^{-1} + O(h_R + \delta_{n2}(h_R)).$$

Thus, $R_{00}^* \mathscr{A}_1 \mathscr{B}^{-1} = O(\delta_{n2}(h_R) + h_R^2)$ a.s. Similar derivations show that $R_{10}^* \mathscr{A}_2 \times \mathscr{B}^{-1}$ and $R_{01}^* \mathscr{A}_3 \mathscr{B}^{-1}$ are both of lower order. Thus, the rate in (3.3) is obtained for $s, t \in [a + h_R, b - h_R]$. As for s and/or t in $[a, a + h) \cup (b - h, b]$, similar calculations show that the same rate also holds. The result follows by taking into account of the rate of $\widehat{\mu}$. □

PROOF OF THEOREM 3.4.   Note that

$$\widehat{\sigma}^2 - \sigma^2 = \frac{1}{b-a} \int_a^b \{\widehat{V}(t) - V(t)\}\,dt - \frac{1}{b-a} \int_a^b \{\widehat{C}(t,t) - C(t,t)\}\,dt.$$

To consider $\widehat{V}(t) - V(t)$ we follow the development in the proof of Theorem 3.1. Recall (2.4) and let $Q_r^* = Q_r - V(t)S_r - hV^{(1)}(t)S_{r+1}$. Then, as in (5.13), we obtain

$$\widehat{V}(t) - V(t) = \frac{Q_0^* S_2 - Q_1^* S_1}{S_0 S_2 - S_1^2}.$$

Write

$$Q_r^* = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} K_{h_V}(T_{ij} - t)\{(T_{ij} - t)/h_V\}^r \{Y_{ij}^2 - V(t) - V^{(1)}(t)(T_{ij} - t)\}$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} K_{h_V}(T_{ij} - t)\{(T_{ij} - t)/h\}^r \{Y_{ij}^2 - V(T_{ij})\} + O(h_V^2),$$

which, by Lemma 1, has the uniformly rate $O(h_V^2 + \delta_{n1}(h_V))$ a.s. By (5.16), we have

$$\widehat{V}(t) - V(t) = \frac{1}{f(t)n} \sum_i \frac{1}{m_i} \sum_{j=1}^{m_i} K_h(T_{ij} - t)\{Y_{ij}^2 - V(T_{ij})\}$$

$$+ O(h_V^2 + \delta_{n1}^2(h_V)) \qquad \text{a.s.}$$

Thus,

$$\int_a^b \{\widehat{V}(t) - V(t)\}\,dt = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} \{Y_{ij}^2 - V(T_{ij})\} \int_a^b K_{h_V}(T_{ij} - t) f^{-1}(t)\,dt$$

$$+ O(h_V^2 + \delta_{n1}^2(h_V)) \qquad \text{a.s.}$$

Note that

$$\left| \int_a^b K_{h_V}(T_{ij} - t) f^{-1}(t)\,dt \right| \leq \sup_t f^{-1}(t).$$

By Lemma 5 below in this subsection,

$$(5.26) \qquad \int_a^b \{\widehat{V}(t) - V(t)\}\,dt = O((\log n/n)^{1/2} + h_V^2 + \delta_{n1}^2(h_V)) \qquad \text{a.s.}$$

Next, we consider $\widehat{C}(t,t) - C(t,t)$. We apply (5.21) but will focus on $R_{00}^* \mathscr{A}_1 \mathscr{B}^{-1}$ since the other two terms are dealt with similarly. By (5.23)–(5.25),

$$R_{00}^* \mathscr{A}_1 \mathscr{B}^{-1} = \frac{1}{f(s)f(t)n} \sum_{i=1}^n \frac{1}{N_i} \sum_{k \neq j} \varepsilon_{ijk}^* K_{h_R}(T_{ij} - s) K_{h_R}(T_{ik} - t)$$

$$(5.27)$$

$$+ O(h_R^2 + \delta_{n2}^2(h_R)) \qquad \text{a.s.}$$

Thus,

$$\int_a^b \{\widehat{C}(t,t) - C(t,t)\} \, dt$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{1}{N_i} \sum_{k \neq j} \varepsilon_{ijk}^* \int_a^b K_{h_R}(T_{ij} - t) K_{h_R}(T_{ik} - t) f^{-2}(t) \, dt$$

$$+ O\big(h_R^2 + \delta_{n2}^2(h_R)\big) \qquad \text{a.s.}$$

Write

$$\int_a^b K_{h_R}(T_{ij} - t) K_{h_R}(T_{ik} - t) f^{-2}(t) \, dt$$

$$= \int_{-1}^1 K(u) K_{h_R}\big((T_{ik} - T_{ij}) + u h_R\big) f^{-2}(T_{ij} - u h_R) \, du.$$

A slightly modified version of Lemma 1 leads to the "one-dimensional" rate:

$$\sup_{u \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{N_i} \sum_{k \neq j} \varepsilon_{ijk}^* K_{h_R}\big((T_{ik} - T_{ij}) + u h_R\big) f^{-2}(T_{ij} - u h_R) \right|$$

$$= O(\delta_{n1}(h_R)) \qquad \text{a.s.}$$

It follows that

$$(5.28) \quad \int_a^b \{\widehat{C}(t,t) - C(t,t)\} \, dt = O\big(h_R^2 + \delta_{n1}(h_R) + \delta_{n2}^2(h_R)\big) \qquad \text{a.s.}$$

The theorem follows from (5.26) and (5.28). □

LEMMA 5. *Assume that $\xi_{ni}, 1 \leq i \leq n$, are independent random variables with mean zero and finite variance. Also assume that there exist i.i.d. random variables ξ_i with mean zero and finite δth moment for some $\delta > 2$ such that $|\xi_{ni}| \leq |\xi_i|$. Then*

$$\frac{1}{n} \sum_{i=1}^n \xi_{ni} = O\big((\log n/n)^{1/2}\big) \qquad \text{a.s.}$$

PROOF.   Let $a_n = (\log n/n)^{1/2}$. Assume that $\xi_{ni} \geq 0$. Write

$$\xi_{ni} = \xi_{ni\succ} + \xi_{ni\prec} := \xi_{ni} I(|\xi_{ni}| > a_n^{-1}) + \xi_{ni} I(|\xi_{ni}| \leq a_n^{-1}).$$

Then

$$\left| \frac{1}{a_n n} \sum_{i=1}^n \xi_{ni\succ} \right| \leq \frac{1}{a_n n} \sum_{i=1}^n |\xi_{ni\succ}|^\delta |\xi_{ni\succ}|^{1-\delta} \leq a_n^{\delta-2} \frac{1}{n} \sum_{i=1}^n |\xi_i|^\delta \to 0 \qquad \text{a.s.}$$

by the law of large numbers. The mean of the left-hand side is also tending to zero by the same argument. Thus, $n^{-1}\sum_{i=1}^{n}(\xi_{ni\succ} - \mathbb{E}\{\xi_{ni\succ}\}) = o(a_n)$. Next, by Bernstein's inequality,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}(\xi_{ni\prec} - \mathbb{E}\{\xi_{ni\prec}\}) > Ba_n\right) \leq \exp\left\{-\frac{B^2 n^2 a_n^2}{2n\sigma^2 + (2/3)Bn}\right\}$$

$$= \exp\left\{-\frac{B^2 \log n}{2\sigma^2 + (2/3)B}\right\},$$

which is summable for large enough B. The result follows from the Borel–Cantelli lemma. $\square$

5.3. *Proof of Theorem* 3.6. Let Δ be the integral operator with kernel $\widehat{R} - R$.

LEMMA 6. *For any bounded measurable function ψ on $[a, b]$,*

$$\sup_{t \in [a,b]} |(\Delta\psi)(t)| = O\left(h_\mu^2 + \delta_{n1}(h_\mu) + h_R^2 + \delta_{n1}(h_R) + \delta_{n2}^2(h_R)\right) \qquad a.s.$$

PROOF. It follows that

$$(\Delta\psi)(t) = \int_{s=a}^{b}(\widehat{C} - C)(s,t)\psi(s)\,ds - \int_{s=a}^{b}\{\widehat{\mu}(s)\widehat{\mu}(t) - \mu(s)\mu(t)\}\psi(s)\,ds$$

$$=: A_{n1} - A_{n2}.$$

By (5.21),

$$A_{n1} = \int_{s=a}^{b}(\mathscr{A}_1 R_{00}^* - \mathscr{A}_2 R_{10}^* - \mathscr{A}_3 R_{01}^*)\mathscr{B}^{-1}\psi(s)\,ds.$$

We focus on $\int_{s=a}^{b}\mathscr{A}_1 R_{00}^*\mathscr{B}^{-1}\psi(s)\,ds$ since the other two terms are of lower order and can be dealt with similarly. By (5.23) and (5.25),

$$\int_{s=a}^{b}\mathscr{A}_1 R_{00}^*\mathscr{B}^{-1}\psi(s)\,ds$$

$$= \frac{1}{f(t)n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{k\neq j}\varepsilon_{ijk}^* K_{h_R}(T_{ik} - t)\int_{s=a}^{b}K_{h_R}(T_{ij} - s)\psi(s)f(s)^{-1}\,ds$$

$$+ O\left(h_R^2 + \delta_{n2}^2(h_R)\right).$$

Note that

$$\left|\int_{s=a}^{b}K_{h_R}(T_{ij} - s)\psi(s)f(s)^{-1}\,ds\right| \leq \sup_{s \in [a,b]}(|\psi(s)|f(s)^{-1})\int_{u=-1}^{1}K(u)\,du.$$

Thus, Lemma 1 can be easily improvised to give the following uniform rate over t:

$$\frac{1}{f(t)n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{k\neq j}\varepsilon_{ijk}^{*}K_{h_R}(T_{ik}-t)\int_{s=a}^{b}K_{h_R}(T_{ij}-s)\psi(s)f(s)^{-1}\,ds$$

$$= O(\delta_{n1}(h_R))\qquad\text{a.s.}$$

Thus,

$$\int_{s=a}^{b}\mathscr{A}_1 R_{00}^{*}\mathscr{B}^{-1}\psi(s)\,ds = O(\delta_{n1}(h_R)+h_R^2+\delta_{n2}^2(h_R))\qquad\text{a.s.,}$$

which is also the rate of A_{n1}. Next, we write

$$A_{n2} = \widehat{\mu}(t)\int_{s=a}^{b}\{\widehat{\mu}(s)-\mu(s)\}\psi(s)\,ds - \{\widehat{\mu}(t)-\mu(t)\}\int_{s=a}^{b}\mu(s)\psi(s)\,ds,$$

which has the rate $O(h_\mu^2+\delta_{n1}(h_\mu))$ by Theorem 3.1. $\square$

PROOF OF THEOREM 3.6. We prove (b) first. Hall and Hosseini-Nasab (2006) give the L^2 expansion

$$\widehat{\psi}_j - \psi_j = \sum_{k\neq j}(\lambda_j-\lambda_k)^{-1}\langle\Delta\psi_j,\psi_k\rangle\phi_k + O(\|\Delta\|^2),$$

where $\|\Delta\| = (\iint\{\widehat{R}(s,t)-R(s,t)\}^2\,ds\,dt)^{1/2}$, the Hilbert–Schmidt norm of Δ. By Bessel's inequality, this leads to

$$\|\widehat{\psi}_j - \psi_j\| \leq C(\|\Delta\psi_j\| + \|\Delta\|^2).$$

By Lemma 6 and Theorem 3.3,

$$\|\Delta\psi_j\| = O(h_\mu^2+\delta_{n1}(h_\mu)+h_R^2+\delta_{n1}(h_R)+\delta_{n2}^2(h_R)),$$

$$\|\Delta\|^2 = O(h_\mu^4+\delta_{n1}^2(h_\mu)+h_R^4+\delta_{n2}^2(h_R))\qquad\text{a.s.}$$

Thus,

$$\|\widehat{\psi}_j - \psi_j\| = O(h_\mu^2+\delta_{n1}(h_\mu)+h_R^2+\delta_{n1}(h_R)+\delta_{n2}^2(h_R))\qquad\text{a.s.,}$$

proving (b).

Next, we consider (a). By (4.9) in Hall, Müller and Wang (2006),

$$\widehat{\omega}_j - \omega_j = \iint(\widehat{R}-R)(s,t)\psi_j(s)\psi_j(t)\,ds\,dt + O(\|\Delta\psi_j\|^2)$$

$$= \iint(\widehat{C}-C)(s,t)\psi_j(s)\psi_j(t)\,ds\,dt$$

$$\quad - \iint\{\widehat{\mu}(s)\widehat{\mu}(t)-\mu(s)\mu(t)\}\psi_j(s)\psi_j(t)\,ds\,dt + O(\|\Delta\psi_j\|^2)$$

$$=: A_{n1} - A_{n2} + O(\|\Delta\psi_j\|^2).$$

Now,

$$A_{n1} = \iint (\mathscr{A}_1 R_{00}^* - \mathscr{A}_2 R_{10}^* - \mathscr{A}_3 R_{01}^*) \mathscr{B}^{-1} \psi_j(s) \psi_j(t) \, ds \, dt.$$

Again it suffices to focus on $\iint \mathscr{A}_1 R_{00}^* \mathscr{B}^{-1} \psi_j(s) \psi_j(t) \, ds \, dt$. By (5.23) and (5.25),

$$\iint \mathscr{A}_1 R_{00}^* \mathscr{B}^{-1} \psi_j(s) \psi_j(t) \, ds \, dt$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i(m_i - 1)} \sum_{k \neq j} \varepsilon_{ijk}^* \iint K_{h_R}(T_{ij} - s) K_{h_R}(T_{ik} - t)$$

$$\times \psi_j(s) \psi_j(t) \{f(s) f(t)\}^{-1} \, ds \, dt$$

$$+ O(h_R^2 + \delta_{n2}^2(h_R)) \qquad \text{a.s.,}$$

where the first term on the right-hand side can be shown to be $O((\log /n)^{1/2})$ a.s. by Lemma 5. Thus,

$$A_{n1} = O((\log /n)^{1/2} + h_R^2 + \delta_{n2}^2(h_R)).$$

Next, write

$$A_{n2} = \int \{\widehat{\mu}(s) - \mu(s)\} \psi_j(s) \, ds \int \widehat{\mu}(t) \psi_j(t) \, dt$$

$$+ \int \mu(s) \psi_j(s) \, ds \int \{\widehat{\mu}(t) - \mu(t)\} \psi_j(t) \, dt,$$

and it can be similarly shown that

$$A_{n2} = O((\log /n)^{1/2} + h_\mu^2 + \delta_{n1}^2(h_\mu)) \qquad \text{a.s.}$$

This establishes (a).

Finally, we consider (c). For any $t \in [a, b]$,

$$\widehat{\omega}_j \widehat{\psi}_j(t) - \omega_j \psi_j(t)$$

$$= \int \widehat{R}(s, t) \widehat{\psi}_j(s) \, ds - \int R(s, t) \psi_j(s) \, ds$$

$$= \int \{\widehat{R}(s, t) - R(s, t)\} \psi_j(s) \, ds + \int \widehat{R}(s, t) \{\widehat{\psi}_j(s) - \psi_j(s)\} \, ds.$$

By the Cauchy–Schwarz inequality, uniformly for all $t \in [a, b]$,

$$\left| \int \widehat{R}(s, t) \{\widehat{\psi}_j(s) - \psi_j(s)\} \, ds \right| \leq \left\{ \int \widehat{R}^2(s, t) \, ds \right\}^{1/2} \|\widehat{\psi}_j - \psi_j\|$$

$$\leq |b - a|^{1/2} \sup_{s,t} |\widehat{R}(s, t)| \times \|\widehat{\psi}_j - \psi_j\|$$

$$= O(\|\widehat{\psi}_j - \psi_j\|) \qquad \text{a.s.}$$

Thus,

$$\widehat{\omega}_j \widehat{\psi}_j(t) - \omega_j \psi_j(t) = O\big(h_\mu^2 + \delta_{n1}(h_\mu) + h_R^2 + \delta_{n1}(h_R) + \delta_{n2}^2(h_R)\big) \qquad \text{a.s.}$$

By the triangle inequality and (b),

$$\begin{aligned}
\omega_j |\widehat{\psi}_j(t) - \psi_j(t)| \\
&= |\widehat{\omega}_j \widehat{\psi}_j(t) - \omega_j \psi_j(t) - (\widehat{\omega}_j - \omega_j)\widehat{\psi}_j(t)| \\
&\leq |\widehat{\omega}_j \widehat{\psi}_j(t) - \omega_j \psi_j(t)| + |\widehat{\omega}_j - \omega_j| \sup_t |\widehat{\psi}_j(t)| \\
&= O\big((\log n/n)^{1/2} + h_\mu^2 + \delta_{n1}(h_\mu) + h_R^2 + \delta_{n1}(h_R) + \delta_{n2}^2(h_R)\big) \qquad \text{a.s.}
\end{aligned}$$

Note that $(\log n/n)^{1/2} = o(\delta_{n1}(h_\mu))$. This completes the proof of (c). $\quad\square$

REFERENCES

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. MR0348906

CLAESKENS, G. and VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884. MR2036392

FAN, J. and GIJBELS, I. (1995). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, New York. MR1383587

FERRÉ, L. and YAO, A. (2005). Smoothed functional sliced inverse regression. *Statist. Sinica* **15** 665–685.

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 109–126. MR2212577

HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365

HÄRDLE, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.* **29** 163–179. MR1004333

HÄRDLE, W., JANSSEN, P. and SERFLING, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16** 1428–1449. MR0964932

JAMES, G. M., HASTIE, T. J. and SUGAR, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602. MR1789811

LANDAU, H. J. and SHEPP, L. A. (1970). On the supremum of Gaussian processes. *Sankhyā Ser. A* **32** 369–378. MR0286167

LI, Y. and HSING, T. (2010). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *Ann. Statist.* To appear.

LI, E., LI, Y., WANG, N.-Y. and WANG, N. (2008). Functional latent feature models for data with longitudinal covariate processes. Unpublished manuscript, Dept. Statistics, Texas A&M Univ., College Station, TX.

LI, Y., WANG, N., HONG, M., TURNER, N., LUPTON, J. and CARROLL, R. J. (2007). Nonparametric estimation of correlation functions in spatial and longitudinal data, with application to colon carcinogenesis experiments. *Ann. Statist.* **35** 1608–1643. MR2351099

LIN, X. and CARROLL, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J. Amer. Statist. Assoc.* **95** 520–534. MR1803170

MASRY, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Ser. Anal.* **17** 571–599. MR1424907

MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805. MR2163159

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

WANG, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* **90** 43–52. MR1966549

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

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106

ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. MR2341698

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

DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602-7952
USA
E-MAIL: yehuali@uga.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109-1107
USA
E-MAIL: thsing@umich.edu