The Annals of Statistics
2016, Vol. 44, No. 5, 2281–2321
DOI: 10.1214/16-AOS1446

FROM SPARSE TO DENSE FUNCTIONAL DATA AND BEYOND[1]

BY XIAOKE ZHANG AND JANE-LING WANG

University of Delaware and University of California, Davis

Nonparametric estimation of mean and covariance functions is important in functional data analysis. We investigate the performance of local linear smoothers for both mean and covariance functions with a general weighing scheme, which includes two commonly used schemes, equal weight per observation (OBS), and equal weight per subject (SUBJ), as two special cases. We provide a comprehensive analysis of their asymptotic properties on a unified platform for all types of sampling plan, be it dense, sparse or neither. Three types of asymptotic properties are investigated in this paper: asymptotic normality, L^2 convergence and uniform convergence. The asymptotic theories are unified on two aspects: (1) the weighing scheme is very general; (2) the magnitude of the number N_i of measurements for the ith subject relative to the sample size n can vary freely. Based on the relative order of N_i to n, functional data are partitioned into three types: non-dense, dense and ultra-dense functional data for the OBS and SUBJ schemes. These two weighing schemes are compared both theoretically and numerically. We also propose a new class of weighing schemes in terms of a mixture of the OBS and SUBJ weights, of which theoretical and numerical performances are examined and compared.

1. Introduction. Functional data analysis (FDA) has gained increasing importance in modern data analysis due to the improved capability to record and store a vast amount of data and advances in scientific computing. In addition to the monographs by Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Bongiorno et al. (2014) and Hsing and Eubank (2015), recent developments in FDA are illustrated in the survey articles by Cuevas (2014), Marron and Alonso (2014), Shang (2014), Morris (2015) and Wang, Chiou and Müller (2016). A fundamental issue in FDA is the estimation of mean and covariance functions. It is not only of interest by itself but also involved in subsequent analyses, such as for dimension reduction and modeling of functional data.

Tribute: Life is short, all too short for Peter. But no doubt that Peter had gotten the most of it. Not only has he left a towering legacy of scientific discoveries, his kindness and generosity have touched the lives of many fortunate enough to have crossed path with him.

Peter's passing is a tremendous loss to us and the statistical community, but we are grateful to have had the privilege to share part of his life in Davis.

Received May 2015; revised January 2016.

[1]Supported by NSF Grant DMS-09-06813.

MSC2010 subject classifications. Primary 62G20; secondary 62G08, 62G05.

Key words and phrases. Local linear smoothing, asymptotic normality, L^2 convergence, uniform convergence, weighing schemes.

Nonparametric methods have been widely used in FDA, including smoothing splines [Cai and Yuan (2011), Rice and Silverman (1991)], B-splines [Cardot (2000), James, Hastie and Sugar (2000), Rice and Wu (2001)], penalized splines [Ruppert, Wand and Carroll (2003), Yao and Lee (2006)] and local polynomial smoothing [Yao and Li (2013), Zhang and Chen (2007)]. We focus on local linear smoothers due to its conceptual simplicity, attractive local features and well-known ability for automatic boundary correction [Fan and Gijbels (1992)]. Extensions of local polynomial smoothers from independent data to functional data include Yao, Müller and Wang (2005a), Hall, Müller and Wang (2006) and Li and Hsing (2010), who employed local linear smoothers to estimate the mean and covariance functions of functional data and demonstrated their excellent theoretical and numerical properties. Additional applications of local linear smoothers to the regression setting and their variants are illustrated in Baíllo and Grané (2009), Barrientos-Marin, Ferraty and Vieu (2010) and Boj, Delicado and Fortiana (2010).

In the context of FDA, we normally have n random functions or curves, representing n subjects, observed at discrete N_i time points for the ith subject. It has been realized that the magnitude of N_i should be carefully handled since it not only leads to distinct asymptotic properties but may also have an impact on the choice of estimation procedures. Although no official definition exists, generally, if all N_i are larger than some order of n, then the functional data are referred to as "dense" data. If all N_i are bounded, the data are commonly considered as "sparse." The words "sparsity" and "sparse" have multiple meanings in the FDA literature depending on the context, such as the sparsity of the model/function [Aneiros and Vieu (2014)], the sparsity of data related to the topological effects of small ball probabilities and the curse of dimensionality [Delaigle and Hall (2010), Ferraty and Vieu (2006), Geenens (2011)] and the sparsity of the time grid at which measurements are taken [James, Hastie and Sugar (2000), Yao, Müller and Wang (2005a)]. We adopt in this paper the last meaning of "sparsity" and "sparse" in terms of the amount of repeated measurements over time on experimental subjects. For dense data, the observations from each subject are often pre-smoothed, to remove noise in the data and to reconstruct the random curve of that subject, before subsequent analysis [Cardot (2000), Ramsay and Silverman (2005), Zhang and Chen (2007)]. For sparse data, such a pre-smoothing step is not viable so data from all subjects are pooled to borrow information from each other before subsequent analysis [Paul and Peng (2009), Rice and Wu (2001), Yao, Müller and Wang (2005a)]. Apart from the difference in estimation procedures, the estimators from dense and sparse data have distinct asymptotic properties. Consequently, it is common in the FDA literature that a paper focuses on one type of data, either dense or sparse. Data that are neither dense nor sparse were considered in Yao (2007), Li and Hsing (2010) and Cai and Yuan (2010, 2011) but have received much less attention in the literature.

We present in this paper a unified approach that handles both dense and sparse functional data plus other data that are of neither type. The unifying theoretical platform allows the magnitude of N_i relative to n to vary freely, and we derive both local and global asymptotic results. Distinct types of convergence rates emerge from the unified theory based on the relative order of N_i to n, which leads to a clear and natural criterion to define dense functional data: functional data such that the root-n convergence rate can be achieved in estimation; otherwise, the data are considered non-dense. In this sense, sparse data are just one special case of non-dense data and there are other non-dense data types with unbounded N_i, which one could term "semi-dense." We also define a new category of "ultra-dense" functional data such that not only the root-n convergence rate can be achieved but also the asymptotic bias, which typically present in nonparametric smoothing methods, disappears, so the theory for ultra-dense data falls squarely in the parametric paradigm.

Three types of asymptotic properties are often investigated for functional data: local asymptotic normality, L^2 convergence, and uniform convergence. In addition to deriving asymptotic normality of the mean estimator, Yao (2007) made the first attempt to prove the asymptotic normality for the covariance estimator. The asymptotic results there could accommodate large N_i but cannot achieve the root-n convergence rate. Zhang and Chen (2007) established the asymptotic normality results for both the mean and covariance estimators where a "smoothing first, then estimation" approach was applied to dense data and a root-n rate was achieved. For L^2 convergence, Hall, Müller and Wang (2006) developed the optimal rate of convergence for sparse data; however, their results focused on functional principal component analysis, rather than mean or covariance estimation. Cai and Yuan (2011) showed that a smoothing spline estimator of the mean function achieves the optimal minimax rate of L^2 convergence, and Cai and Yuan (2010) established the minimax rate for a regularized covariance estimator under a reproducing kernel Hilbert space framework. Uniform convergence was first explored in Yao, Müller and Wang (2005a) but the convergence rate there was sub-optimal. This was improved in Li and Hsing (2010) but using a different estimator, which will be discussed in the next paragraph. Each paper in the FDA literature typically focuses on one type of asymptotic properties, but in this paper we investigate all three types of large sample properties.

The difference between the estimation procedures in Yao, Müller and Wang (2005a) and Li and Hsing (2010) lies in the weighing scheme, the way weights are assigned to observations. Yao, Müller and Wang (2005a) employed a scatter plot smoother that assigns the same weight to each observation, which we term the "OBS" scheme, so a subject with a larger number of observations N_i receives more weights in total. In contrast, Li and Hsing (2010) assigned the same weight to each subject, which we term the "SUBJ" scheme. The follow-up work [Cai and Yuan (2011), Hall, Müller and Wang (2006), Kim and Zhao (2013), Yao (2007)] also uses either OBS or SUBJ scheme, rather than considering both or a more general one. We consider in this paper an approach that allows a general weighing scheme such that both the OBS and SUBJ schemes are special cases of it. Since the OBS and SUBJ schemes are the most commonly used ones in practice, we further compare them theoretically and numerically to provide guidance for practitioners.

We additionally propose a new class of weighing schemes in terms of a mixture of the OBS and SUBJ weights, of which theoretical and numerical performances are also compared with the OBS and SUBJ schemes.

To summarize, this paper provides a comprehensive and unifying analysis of three asymptotic properties (asymptotic normality, L^2 convergence, and uniform convergence) based on a general weighing scheme for functional data. The unified new theory also leads to new results under the OBS and SUBJ schemes, such as the asymptotic normality for SUBJ estimators as well as the L^2 and uniform convergence for OBS estimators. Moreover, we improve the asymptotic normality results of Yao (2007) such that root-n rate can now be attained for dense and ultra-dense data. We also ameliorate the covariance estimator in Li and Hsing (2010) to have better numerical performance.

A byproduct of the unified theory is a systematic partition of functional data into non-dense, dense and ultra-dense data. Another intriguing result is the "discontinuity of the asymptotic variance" phenomenon in the sense that the asymptotic variance expressions may be different for the variance and covariance estimators.

For the OBS and SUBJ schemes, they have the same asymptotic bias but different asymptotic variances. With some constraints on N_i, Jensen's inequality implies that the rate of convergence for the OBS approach is always as good as and, in some special cases, better than, the one from the SUBJ approach. The benefit of the OBS scheme could be substantial when the measurement schedule is dense for some subjects and sparse for others. When the aforementioned constraints on N_i are violated, the OBS scheme may be inferior to the SUBJ scheme in terms of mean integrated squared error and mean supremum absolute error. Perhaps not surprisingly, for ultra-dense data, the SUBJ estimators can achieve exactly the same asymptotic normality results as those based on the sample mean and sample covariance of the true individual functions. For either mean or covariance estimation, there exists a particular mixture of the OBS and SUBJ weights such that the global (either L^2 or uniform) convergence rate of the resulting estimator is at least as good as those of both OBS and SUBJ estimators.

The remainder of the paper is organized as follows: Section 2 introduces the model and estimators. The asymptotic results on the three types of convergence are shown in Sections 3–5, respectively. In Section 6, numerical performances of the OBS and SUBJ schemes together with their mixture scheme are compared via simulation. Section 7 contains the conclusion of this paper.

2. Methodology.

Let $\{X(t) : t \in I\}$ be an L^2 stochastic process on an interval I, which we assume for simplicity to be $[0, 1]$. The mean and covariance functions of $X(t)$ are, respectively, $\mu(t) = E(X(t))$ and $\gamma(s, t) = \mathrm{Cov}(X(s), X(t))$. With this notation, we can decompose $X(t)$ into

$$X(t) = \mu(t) + U(t), \tag{2.1}$$

where $U(t)$ is the stochastic part of $X(t)$ with $EU(t) = 0$ for any $t \in [0, 1]$ and $\mathrm{Cov}(U(s), U(t)) = \gamma(s, t)$ for all $s, t \in [0, 1]$.

Let X_i be i.i.d. copies of X, which are not observable in practice. Instead, for the ith subject, measurements are taken at N_i time points, T_{ij}, $j = 1, \dots, N_i$, and the observations at these time points are contaminated with additive random errors, so the actual observations follow

$$Y_{ij} = X_i(T_{ij}) + e_{ij} = \mu(T_{ij}) + U_i(T_{ij}) + e_{ij}, \tag{2.2}$$

where the e_{ij} are i.i.d. copies of e with $E(e) = 0$ and $\mathrm{Var}(e) = \sigma^2$. Thus, we observe $\{(T_{ij}, Y_{ij}) : i = 1, \dots, n, j = 1, \dots, N_i\}$, which are identically distributed as (T, Y) where $Y = \mu(T) + U(T) + e$. For convenience, we denote $U_{ij} = U_i(T_{ij})$ and $\delta_{ij} = U_{ij} + e_{ij}$.

To estimate $\mu(t)$ and $\gamma(s, t)$, local linear smoothing is applied due to its simplicity and attractive local features. In the FDA literature, the OBS [Yao, Müller and Wang (2005a)] and SUBJ [Li and Hsing (2010)] schemes are the most commonly used weighing schemes, and the OBS estimators are typically referred as the scatter-plot smoothers. Obviously, the two schemes coincide when all the N_i are equal. In this paper, we investigate a more general weighing scheme, of which OBS and SUBJ are two special cases. Hereafter, estimators with no subscript represent those using the general weighing scheme, while estimators with the subscript "obs" and "subj" represent the OBS and SUBJ estimators, respectively.

In what follows, denote $K_h(\cdot) = K(\cdot/h)/h$ for a one-dimensional kernel K and bandwidth h. We denote $a(n) \preceq b(n)$ to mean $\limsup_{n\to\infty} a(n)/b(n) < \infty$, and $a(n) \asymp b(n)$ to mean $a(n) \preceq b(n)$ and $b(n) \preceq a(n)$.

To estimate μ, a local linear smoother is applied to $\{(Y_{ij}, T_{ij}) : i = 1, \dots, n; j = 1, \dots, N_i\}$ and the weight w_i is attached to each observation for the ith subject such that $\sum_{i=1}^n N_i w_i = 1$. The mean estimator is $\hat{\mu}(t) = \hat{\beta}_0$ where

$$(\hat{\beta}_0, \hat{\beta}_1) = \operatorname*{argmin}_{\beta_0, \beta_1} \sum_{i=1}^n w_i \sum_{j=1}^{N_i} [Y_{ij} - \beta_0 - \beta_1(T_{ij} - t)]^2 K_{h_\mu}(T_{ij} - t). \tag{2.3}$$

To estimate γ, we assume $N_i \ge 2$. Once $\hat{\mu}(t)$ is obtained, a local linear smoother is applied to the "raw covariances," $C_{ijl} = [Y_{ij} - \hat{\mu}(T_{ij})][Y_{il} - \hat{\mu}(T_{il})]$, and the weight v_i is attached to each C_{ijl} for the ith subject such that $\sum_{i=1}^n N_i(N_i - 1)v_i = 1$. The resulting covariance estimator is $\hat{\gamma}(s, t) = \hat{\beta}_0$, where

$$\begin{aligned}(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \operatorname*{argmin}_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^n v_i \sum_{1 \le j \ne l \le N_i} [C_{ijl} - \beta_0 - \beta_1(T_{ij} - s) - \beta_2(T_{il} - t)]^2 \\ \times K_{h_\gamma}(T_{ij} - s) K_{h_\gamma}(T_{il} - t).\end{aligned} \tag{2.4}$$

For the OBS scheme, we let $w_i = 1/N_S$ where $N_S = \sum_{i=1}^n N_i$ and $v_i = 1/\sum_{i=1}^n N_i(N_i - 1)$; for the SUBJ scheme, we let $w_i = 1/(nN_i)$ and $v_i = 1/[nN_i(N_i - 1)]$. To establish asymptotic properties for the general weight estimators, we regard N_i, w_i and v_i as fixed quantities that are allowed to vary over n. When N_i are random, the theory can be regarded as conditional on the values of N_i.

3. Asymptotic normality. We provide a unified theory in this section on the asymptotic normality of the mean and covariance estimators in (2.3) and (2.4), regardless of the type of measurement schedule. For the special cases of OBS and SUBJ schemes, a byproduct of the asymptotic normality is the emergence of three types of asymptotic normality as presented in Corollaries 3.2 and 3.5.

3.1. *Mean function.* We first provide the asymptotic normality of the mean estimator based on the general weighing scheme.

THEOREM 3.1. *Under* (A1)–(A2), (B1)–(B4) *and* (C1a)–(C3a) *in Appendix* A *and for a fixed interior point* $t \in (0,1)$, *if* $\min\{h_\mu/\sum_{i=1}^n N_i w_i^2, 1/\sum_{i=1}^n N_i(N_i-1)w_i^2\}h_\mu^6 \to 0$ *and* $h_\mu \sum_{i=1}^n N_i(N_i-1)w_i^2/\sum_{i=1}^n N_i w_i^2 \to C_0 \in [0,\infty]$, *then*

$$\Gamma_\mu^{-1/2}\big[\hat{\mu}(t) - \mu(t) - \tfrac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t) + o_p(h_\mu^2)\big] \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\Gamma_\mu = \frac{\sum_{i=1}^n N_i w_i^2}{h_\mu}\|K\|^2\frac{\gamma(t,t)+\sigma^2}{f(t)} + \left[\sum_{i=1}^n N_i(N_i-1)w_i^2\right]\gamma(t,t).$$

By letting $w_i = 1/N_S$ and $w_i = 1/(nN_i)$, respectively, we obtain the corresponding results for the OBS and SUBJ estimators. We denote $N_{Sk} = \sum_{i=1}^n N_i^k$ for an integer $k \geq 2$, $\bar{N} = N_S/n$, $\bar{N}_{Sk} = N_{Sk}/n$, and $\bar{N}_H = (n^{-1}\sum_{i=1}^n N_i^{-1})^{-1}$, where the subscript "H" in $\bar{N}_H$ suggests that it is a harmonic mean.

COROLLARY 3.1. *Suppose that* (A1)–(A2), (B1)–(B4) *and* (C1a)–(C3a) *hold and* t *is a fixed interior point in* $(0,1)$.

(a) *OBS*: *If* $\min\{n\bar{N}h_\mu, n(\bar{N})^2/(\bar{N}_{S2}-\bar{N})\}h_\mu^6 \to 0$ *and* $h_\mu(\bar{N}_{S2}-\bar{N})/\bar{N} \to C_0 \in [0,\infty]$,

$$\Gamma_{\text{obs}}^{-1/2}\left[\hat{\mu}_{\text{obs}}(t) - \mu(t) - \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t) + o_p(h_\mu^2)\right] \xrightarrow{d} \mathcal{N}(0,1),$$

where $\Gamma_{\text{obs}} = \|K\|^2\dfrac{\gamma(t,t)+\sigma^2}{n\bar{N}h_\mu f(t)} + \dfrac{(\bar{N}_{S2}-\bar{N})}{n(\bar{N})^2}\gamma(t,t)$.

(b) *SUBJ*: *If* $\min\{n\bar{N}_H h_\mu, n/(1-1/\bar{N}_H)\}h_\mu^6 \to 0$ *and* $h_\mu(\bar{N}_H - 1) \to C_0 \in [0,\infty]$,

$$\Gamma_{\text{subj}}^{-1/2}\left[\hat{\mu}_{\text{subj}}(t) - \mu(t) - \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t) + o_p(h_\mu^2)\right] \xrightarrow{d} \mathcal{N}(0,1),$$

where $\Gamma_{\text{subj}} = \|K\|^2\dfrac{\gamma(t,t)+\sigma^2}{n\bar{N}_H h_\mu f(t)} + \dfrac{1}{n}\left(1-\dfrac{1}{\bar{N}_H}\right)\gamma(t,t)$.

REMARK 1. (i) The magnitude of N_i in Theorem 3.1 can be of any order of the sample size n. The result in Corollary 3.1 for the SUBJ estimator is new and the result for the OBS estimator improves that of Yao (2007), where $\{N_i : i = 1, \dots, n\}$ are assumed as i.i.d. copies of a random variable N and an additional assumption $ENh_\mu \to 0$ was used. Due to this additional assumption, root-n rate could not be attained in Yao (2007), while we established it in Corollary 3.2 below.

(ii) Likelihood based mixed-effects models have been investigated for longitudinal data [Rice and Wu (2001), Shi, Weiss and Taylor (1996)] to estimate fixed and random effects nonparametrically. These approaches involve the assumption of Gaussian process and have no theory involved. Wu and Zhang (2002) considered local polynomial mixed-effect models and developed asymptotic normality results for the local linear and local constant mixed-effects estimates, which also involves the Gaussian assumption. In contrast, the approach we consider in this paper does not rely on the Gaussian assumption and is thus more broadly applicable.

(iii) In Theorem 3.1, we assume that the errors e_{ij} are i.i.d. copies of e with zero mean $Ee = 0$ and constant variance $\mathrm{Var}(e) = \sigma^2$. One possible relaxation of this assumption is that $e_{ij} = e(T_{ij})$ where $Ee(t) = 0$, $\mathrm{Var}[e(t)] = \sigma^2(t)$, and $\sigma^2(\cdot)$ is twice-differentiable. The asymptotic result is the same except that σ^2 is replaced by $\sigma^2(t)$ and the proof is essentially similar. For simplicity and to avoid technical distraction, we assume the constant error variance throughout this paper.

For either the OBS or SUBJ scheme, three types of asymptotic normality emerge from Corollary 3.1 depending on the order of $\bar{N}$ and $\bar{N}_H$ relative to n and the order of h_μ.

COROLLARY 3.2. *Suppose that the assumptions for Corollary 3.1 hold and t is a fixed interior point in* $(0, 1)$.

(a) *OBS*: *Assume* $\limsup_n \bar{N}_{S_2}/(\bar{N})^2 < \infty$.

Case 1. *When* $\bar{N}/n^{1/4} \to 0$ *and* $h_\mu \asymp (n\bar{N})^{-1/5}$,

$$\sqrt{n\bar{N}h_\mu}\left[\hat{\mu}_{\mathrm{obs}}(t) - \mu(t) - \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)\right] \xrightarrow{d} \mathcal{N}\left(0, \|K\|^2\frac{\gamma(t,t)+\sigma^2}{f(t)}\right).$$

Case 2. *When* $\bar{N}/n^{1/4} \to C$ *and* $h_\mu\bar{N}_{S2}/\bar{N} \to C_1$ *where* $0 < C, C_1 < \infty$,

$$\sqrt{\frac{n(\bar{N})^2}{\bar{N}_{S2}}}\left[\hat{\mu}_{\mathrm{obs}}(t) - \mu(t) - \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)\right] \xrightarrow{d} \mathcal{N}\left(0, \|K\|^2\frac{\gamma(t,t)+\sigma^2}{f(t)C_1} + \gamma(t,t)\right).$$

Case 3. *When* $\bar{N}/n^{1/4} \to \infty$, $h_\mu = o(n^{-1/4})$, *and* $h_\mu\bar{N} \to \infty$,

$$\sqrt{\frac{n(\bar{N})^2}{\bar{N}_{S2}}}[\hat{\mu}_{\mathrm{obs}}(t) - \mu(t)] \xrightarrow{d} \mathcal{N}(0, \gamma(t,t)).$$

(b) *SUBJ*: *Similar three cases can be obtained for* $\hat{\mu}_{\text{subj}}$ *by replacing* $\bar{N}$, $\bar{N}_{S2}/\bar{N}$ *and* $(\bar{N})^2/\bar{N}_{S2}$ *in* (a) *by* $\bar{N}_H$, $\bar{N}_H$ *and* 1, *respectively.*

REMARK 2. (i) Corollary 3.2 leads to a systematic partition of functional data into three categories under either OBS or SUBJ scheme. The partition is based on the relative order of $\bar{N}$ and $\bar{N}_H$ to $n^{1/4}$, for $\hat{\mu}_{\text{obs}}$ and $\hat{\mu}_{\text{subj}}$, respectively. With either scheme, the result in case 1 is comparable to the one-dimensional local linear smoother for independent data. In both cases 2 and 3, root-n rate can be attained.

(ii) Corollary 3.2 also gives a clear and natural criterion to define dense functional data: Functional data such that root-n rate of convergence can be achieved are "dense"; otherwise, the data are "non-dense." Accordingly, cases 2 and 3 correspond to dense data while case 1 is for non-dense data since the optimal rate of convergence is $(n\bar{N})^{2/5}$ or $(n\bar{N}_H)^{2/5}$, both are of the order $o(n^{1/2})$. The rate in case 1 can be of any order between $n^{2/5}$ and $n^{1/2}$, but it can never reach root-n. Sparse data, where all N_i are bounded, is a special case of case 1. Although root-n rate can be attained in both cases 2 and 3, only case 3 falls in the parametric paradigm where the limiting normal distribution has zero mean. We term the data in case 3 as "ultra-dense." Therefore, functional data can be partitioned into three types: non-dense, dense and ultra-dense functional data.

Corollary 3.1 implies that $\hat{\mu}_{\text{obs}}$ and $\hat{\mu}_{\text{subj}}$ have identical asymptotic bias but different asymptotic variances. We decompose Γ_{obs} in Corollary 3.1 into $\Gamma_{\text{obs}} = \Gamma_{\text{obs}}^A + \Gamma_{\text{obs}}^B$ where

$$\Gamma_{\text{obs}}^A = \|K\|^2 \frac{\gamma(t,t)+\sigma^2}{n\bar{N}h_\mu f(t)}, \qquad \Gamma_{\text{obs}}^B(t) = \frac{(\bar{N}_{S2}-\bar{N})}{n(\bar{N})^2}\gamma(t,t).$$

For $\hat{\mu}_{\text{subj}}$, we can similarly decompose Γ_{subj} into two parts:

$$\Gamma_{\text{subj}}^A = \|K\|^2 \frac{\gamma(t,t)+\sigma^2}{n\bar{N}_H h_\mu f(t)}, \qquad \Gamma_{\text{subj}}^B = \frac{1}{n}\left(1-\frac{1}{\bar{N}_H}\right)\gamma(t,t).$$

We have the following two inequalities.

COROLLARY 3.3.

$$\Gamma_{\text{obs}}^A \le \Gamma_{\text{subj}}^A \quad \text{and} \quad \Gamma_{\text{obs}}^B \ge \Gamma_{\text{subj}}^B.$$

REMARK 3. Corollary 3.3 sheds lights on the different performance of the two weighing schemes for non-dense and ultra-dense data as shown in Corollary 3.2. For non-dense data, where Γ_{obs}^A and Γ_{subj}^A dominate Γ_{obs}^B and Γ_{subj}^B, respectively, the OBS scheme results in a more efficient mean estimator. In contrast, for ultra-dense data, where Γ_{obs}^B and Γ_{subj}^B dominate Γ_{obs}^A and Γ_{subj}^A, respectively, the SUBJ scheme outperforms the OBS scheme. In particular, for ultra-dense data, the SUBJ estimator is asymptotically equivalent to the sample mean when the true curves $X_i(t)$ are observable. Below we provide an intuitive explanation for these two scenarios.

For non-dense data the choice of the bandwidth h_μ in case 1 of Corollary 3.2 leads to $\bar{N}h_\mu, \bar{N}_H h_\mu \to 0$. Consider the simplest special case when $N_i h_\mu \to 0$ for all i, so each subject contributes at most one observation to the mean estimator $\hat{\mu}(t)$ asymptotically. Due to the independence between subjects, the data used for the mean estimation are essentially i.i.d. so the OBS scheme, which assigns the same weight to each observation, enables the most efficient estimator. For more general cases when only $\bar{N}h_\mu, \bar{N}_H h_\mu \to 0$ hold, the same arguments applies on the average, so even though $\hat{\mu}_{\text{obs}}$ may not be the most efficient estimator it is still better than the SUBJ estimator, which does not weigh in the extra information from subjects with larger N_i.

On the other hand, when data are ultra-dense the choice of the bandwidth h_μ in case 3 of Corollary 3.2 leads to $\bar{N}h_\mu, \bar{N}_H h_\mu \to \infty$, so an average subject contributes infinitely many observations to the mean estimator. Since observations from the same subject are correlated, the covariances within subjects now dominate the variances and it is unwise to allow any subject to have an undue influence as it may otherwise inflate the rate of the variance of the mean estimator. The SUBJ scheme achieves such a goal and is thus a safer and better strategy than the OBS scheme. Furthermore, the SUBJ scheme is equivalent to taking a sample mean after one pre-smoothes each individual data and it is well known [Hall, Müller and Wang (2006)] that for ultra-dense data pre-smoothing of individual data leads to a curve that is first order equivalent to the original signal $X_i(t)$. This explains why the SUBJ estimator is asymptotically equivalent to the sample mean based on the true curves.

3.2. *Covariance function.* Define

$$
\begin{aligned}
V_1(s,t) &= \operatorname{Var}\big[\big(Y_1-\mu(T_1)\big)\big(Y_2-\mu(T_2)\big)|T_1=s, T_2=t\big];\\
V_2(s,t) &= \operatorname{Cov}\big([Y_1-\mu(T_1)][Y_2-\mu(T_2)], [Y_1-\mu(T_1)][Y_3-\mu(T_3)]|\\
&\qquad T_1=s, T_2=t, T_3=t\big);\\
V_3(s,t) &= \operatorname{Cov}\big([Y_1-\mu(T_1)][Y_2-\mu(T_2)], [Y_3-\mu(T_3)][Y_4-\mu(T_4)]|\\
&\qquad T_1=s, T_2=t, T_3=s, T_4=t\big).
\end{aligned}
$$

The next theorem provides the asymptotic normality of the covariance estimator with the general weighing scheme. For simplicity and to illustrate the general concept of covariance estimation, we assume that the mean function μ is known so it is not entangled in the covariance estimation.

THEOREM 3.2. *Under* (A1)–(A2), (B1)–(B4) *and* (D1a)–(D3a), *for two fixed interior points* $s, t \in (0, 1)$, *if*

$$\min\left\{\frac{h_\gamma^2}{\sum_{i=1}^n N_i(N_i-1)v_i^2}, \frac{h_\gamma}{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2},\right.$$
$$\left.\frac{1}{\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2}\right\}h_\gamma^6 \to 0,$$

$$\frac{[\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2]\cdot[\sum_{i=1}^n N_i(N_i-1)v_i^2]}{[\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2]^2} \to C_0 \quad \textit{and}$$

$$\frac{h_\gamma^2\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2}{\sum_{i=1}^n N_i(N_i-1)v_i^2} \to C_0' \qquad \textit{where } C_0, C_0' \in [0,\infty],$$

then we have

$$\Gamma_\gamma^{-1/2}\left[\hat{\gamma}(s,t)-\gamma(s,t)-\frac{1}{2}h_\gamma^2\sigma_K^2\left(\frac{\partial^2\gamma}{\partial s^2}(s,t)+\frac{\partial^2\gamma}{\partial t^2}(s,t)\right)+o_p(h_\gamma^2)\right]$$
$$\xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\Gamma_\gamma = [1+I(s=t)]\left[\frac{\sum_{i=1}^n N_i(N_i-1)v_i^2}{h_\gamma^2}\|K\|^4\frac{V_1(s,t)}{f(s)f(t)}\right.$$
$$\left.+\frac{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2}{h_\gamma}\|K\|^2\frac{f(s)V_2(t,s)+f(t)V_2(s,t)}{f(s)f(t)}\right]$$
$$+\left[\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2\right]V_3(s,t),$$

and $I(\cdot)$ *is the indicator function.*

Letting $v_i = 1/\sum_{i=1}^n N_i(N_i-1)$ and $v_i = 1/[nN_i(N_i-1)]$ leads to the corresponding results for OBS and SUBJ estimators, respectively. Hereafter, we denote

$$P_1(N) = \frac{1}{n^{-1}\sum_{i=1}^n[N_i(N_i-1)]},$$
$$P_2(N) = \frac{n^{-1}\sum_{i=1}^n[N_i(N_i-1)(N_i-2)]}{(n^{-1}\sum_{i=1}^n[N_i(N_i-1)])^2},$$
$$P_3(N) = \frac{n^{-1}\sum_{i=1}^n[N_i(N_i-1)(N_i-2)(N_i-3)]}{(n^{-1}\sum_{i=1}^n[N_i(N_i-1)])^2},$$

$$P_4(N) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i(N_i-1)},$$

$$P_5(N) = \frac{1}{n}\sum_{i=1}^{n}\frac{N_i-2}{N_i(N_i-1)},$$

$$P_6(N) = \frac{1}{n}\sum_{i=1}^{n}\frac{(N_i-2)(N_i-3)}{N_i(N_i-1)}.$$

COROLLARY 3.4. *Suppose that* (A1)–(A2), (B1)–(B4) *and* (D1a)–(D3a) *hold*, *and* s *and* t *are two fixed interior points in* $(0,1)$.

(a) *OBS*: *If* $\min\{nP_1(N)^{-1}h_\gamma^2, nP_2(N)^{-1}h_\gamma, nP_3(N)^{-1}\}h_\gamma^6 \to 0$, $P_1(N)P_3(N)/P_2(N)^2 \to C_0$ *and* $h_\gamma^2 P_3(N)/P_1(N) \to C_0'$ *where* $C_0, C_0' \in [0,\infty]$,

$$\Gamma_{\mathrm{obs}}^{-1/2}\left[\hat{\gamma}_{\mathrm{obs}}(s,t) - \gamma(s,t) - \frac{1}{2}h_\gamma^2\sigma_K^2\left(\frac{\partial^2\gamma}{\partial s^2}(s,t) + \frac{\partial^2\gamma}{\partial t^2}(s,t)\right) + o_p(h_\gamma^2)\right] \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\Gamma_{\mathrm{obs}} = [1 + I(s=t)]\left[\frac{P_1(N)}{nh_\gamma^2}\|K\|^4\frac{V_1(s,t)}{f(s)f(t)} + \frac{P_2(N)}{nh_\gamma}\|K\|^2\frac{f(s)V_2(t,s) + f(t)V_2(s,t)}{f(s)f(t)}\right] + \frac{P_3(N)}{n}V_3(s,t).$$

(b) *SUBJ*: *If* $\min\{nP_4(N)^{-1}h_\gamma^2, nP_5(N)^{-1}h_\gamma, nP_6(N)^{-1}\}h_\gamma^6 \to 0$, $P_4(N) \times P_6(N)/P_5(N)^2 \to C_0$ *and* $h_\gamma^2 P_6(N)/P_4(N) \to C_0'$ *where* $C_0, C_0' \in [0,\infty]$,

$$\Gamma_{\mathrm{subj}}^{-1/2}\left[\hat{\gamma}_{\mathrm{subj}}(s,t) - \gamma(s,t) - \frac{1}{2}h_\gamma^2\sigma_K^2\left(\frac{\partial^2\gamma}{\partial s^2}(s,t) + \frac{\partial^2\gamma}{\partial t^2}(s,t)\right) + o_p(h_\gamma^2)\right] \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\Gamma_{\mathrm{subj}} = [1 + I(s=t)]\left[\frac{P_4(N)}{nh_\gamma^2}\|K\|^4\frac{V_1(s,t)}{f(s)f(t)} + \frac{P_5(N)}{nh_\gamma}\|K\|^2\frac{f(s)V_2(t,s) + f(t)V_2(s,t)}{f(s)f(t)}\right] + \frac{P_6(N)}{n}V_3(s,t).$$

Similar to Corollary 3.2, for either the OBS or SUBJ covariance estimator, we can also partition the functional data into three types: non-dense, dense and ultra-dense data, based on their asymptotic distributions.

COROLLARY 3.5. *Suppose that the assumptions for Corollary* 3.4 *hold and that s and t are two fixed interior points in* $(0, 1)$.

(a) *OBS*: *Assume* $\lim_n \bar{N}_{S2} \cdot \bar{N}_{S4}/(\bar{N}_{S3})^2 = C_0 \in [1, \infty]$, $\limsup_n (\bar{N}_{S2})/(\bar{N})^2$, $\limsup_n (\bar{N} \cdot \bar{N}_{S3})/(\bar{N}_{S2})^2$, $\limsup_n \bar{N}_{S4}/(\bar{N}_{S2})^2 < \infty$.

Case 1. *When* $P_1(N)^{-1}/n^{1/2} \to 0$ *and* $h_\gamma \asymp (nP_1(N)^{-1})^{-1/6}$,

$$\sqrt{nP_1(N)^{-1}h_\gamma^2}\Big[\hat{\gamma}_{\mathrm{obs}}(s,t) - \gamma(s,t) - \frac{1}{2}h_\gamma^2\sigma_K^2\Big(\frac{\partial^2\gamma}{\partial s^2}(s,t) + \frac{\partial^2\gamma}{\partial t^2}(s,t)\Big)\Big] \xrightarrow{d} \mathcal{N}\Big(0, [1 + I(s=t)]\|K\|^4\frac{V_1(s,t)}{f(s)f(t)}\Big).$$

Case 2. *When* $P_1(N)^{-1}/n^{1/2} \to C$ *and* $h_\gamma^2 P_3(n)/P_1(N) \to C_1^2$, *where* $0 < C, C_1 < \infty$,

$$\sqrt{\frac{n(\bar{N}_{S2})^2}{\bar{N}_{S4}}}\Big[\hat{\gamma}_{\mathrm{obs}}(s,t) - \gamma(s,t) - \frac{1}{2}h_\gamma^2\sigma_K^2\Big(\frac{\partial^2\gamma}{\partial s^2}(s,t) + \frac{\partial^2\gamma}{\partial t^2}(s,t)\Big)\Big]$$
$$\xrightarrow{d} \mathcal{N}\Big(0, [1 + I(s=t)]\Big[\frac{\|K\|^4 V_1(s,t)}{C_1^2 f(s)f(t)} + \frac{\|K\|^2}{C_1 C_0^{1/2}}\frac{f(s)V_2(t,s) + f(t)V_2(s,t)}{f(s)f(t)}\Big] + V_3(s,t)\Big).$$

Case 3. *When* $P_1(N)^{-1}/n^{1/2} \to \infty$, $h_\gamma = o(n^{-1/4})$, *and* $h_\gamma^2 P_1(N)^{-1} \to \infty$,

$$\sqrt{\frac{n(\bar{N}_{S2})^2}{\bar{N}_{S4}}}\big[\hat{\gamma}_{\mathrm{obs}}(s,t) - \gamma(s,t)\big] \xrightarrow{d} \mathcal{N}\big(0, V_1(s,t)\big).$$

(b) *SUBJ*: *Replacing* $P_1(N)$, $P_3(N)/P_1(N)$, *and* $(\bar{N}_{S2})^2/\bar{N}_{S4}$ *in cases* 1–3 *with* $P_4(N)$, $P_6(N)/P_4(N)$ *and* 1, *respectively*, *lead to the corresponding three cases*.

REMARK 4. (i) The results in Theorem 3.2 are intriguing as the asymptotic variance expressions are different for the variance ($s = t$) and covariance ($s \neq t$) estimators. We refer this fact as the "discontinuity of the asymptotic variance" of the covariance estimator. This difference results from the fact that $E[K_{h_\gamma}(T - t)K_{h_\gamma}(T - s)] = \|K\|^2 f(t)/h_\gamma + o(1/h_\gamma)$ for $s = t$, but $E[K_{h_\gamma}(T - s)K_{h_\gamma}(T - t)] = 0$ for $s \neq t$ when h_γ is sufficiently small. Details can be found in Appendix B.

Corollary 3.5 further reveals that the "discontinuity of the asymptotic variance" appears for non-dense and dense data but not for ultra-dense data. The explanation is as follows: The ith subject contributes to the variance of $\hat{\gamma}$ in terms of three sources of variation: $\mathrm{Var}(\delta_{ij}\delta_{il})$, for $j \neq l$ (Variation A), $\mathrm{Cov}(\delta_{ij}\delta_{il}, \delta_{ij}\delta_{ik})$, for pairwise unequal j, k, l (Variation B), and $\mathrm{Cov}(\delta_{ij}\delta_{il}, \delta_{im}\delta_{ik})$, for pairwise unequal j, k, l, m (Variation C). Due to the different expression of $E[K_{h_\gamma}(T-t)K_{h_\gamma}(T-s)]$ between $s = t$ and $s \neq t$, both Variation A and Variation B are doubled when $s = t$ compared to $s \neq t$, but Variation C is unaffected. Since Variation C dominates Variations A and B asymptotically when data are ultra-dense, the phenomenon "discontinuity of the asymptotic variance" does not appear there. In contrast, when data are either non-dense or dense, at least one of the Variations, A or B, is not negligible so this phenomenon occurs.

(ii) Remarks 1(i), 2 and 3 apply for the covariance estimators as well with all the arguments on one-dimensional smoothing changed to two-dimensional smoothing. The results for the OBS estimator improves that of Yao (2007) where an extra assumption $EN^3 h_\gamma \to 0$ was used under the random N_i scenario and only the case with $s \neq t$ was considered.

Based on the order of $P_1(N)^{-1}$ and $P_4(N)^{-1}$ relative to $n^{1/2}$, functional data are partitioned into three types in covariance estimation as in Corollary 3.5: non-dense, dense and ultra-dense. Similar to Corollary 3.3, we have $P_1(N) \leq P_4(N)$ and $P_3(N) \geq P_6(N)$. Therefore, the OBS scheme outperforms SUBJ scheme for non-dense data, while this is reversed for ultra-dense data. Moreover, for ultra-dense data, the SUBJ estimator is asymptotically equivalent to the sample covariance when the true curves $X_i(t)$ are observable.

4. L^2 convergence.

In this section, L^2 rates of convergence are provided for mean and covariance estimators with the general weighing scheme, for which the OBS and SUBJ schemes are two special cases.

For a univariate function $\phi(\cdot) \in [0, 1]$ and a bivariate function $\Phi(\cdot,\cdot) \in [0, 1]^2$, define the L^2 norm by $\|\phi\|_2 = [\int \phi(t)^2\,dt]^{1/2}$ and the Hilbert–Schmidt norm by $\|\Phi\|_{\mathrm{HS}} = [\iint \Phi(s,t)^2\,ds\,dt]^{1/2}$. The domains of the integrals are omitted unless otherwise specified.

4.1. *Mean function.*

We consider the mean estimator in (2.3) with a general weight w_i for subject i.

THEOREM 4.1. *Under* (A1)–(A2), (B1)–(B4) *and* (C1b)–(C3b),

$$\|\hat{\mu} - \mu\|_2 = O_p\left(h_\mu^2 + \left[\frac{\sum_{i=1}^n N_i w_i^2}{h_\mu} + \sum_{i=1}^n N_i(N_i - 1)w_i^2\right]^{1/2}\right).$$

COROLLARY 4.1. *Suppose that* (A1)–(A2), (B1)–(B4), (C1b) *and* (C3b) *hold.*

(a) *OBS*:

$$\|\hat{\mu}_{\text{obs}} - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\left(\frac{1}{\bar{N}h_\mu} + \frac{\bar{N}_{S2}}{(\bar{N})^2}\right)\frac{1}{n}}\right).$$

(b) *SUBJ*: *Under an additional assumption* (C2b),

$$\|\hat{\mu}_{\text{subj}} - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\left(\frac{1}{\bar{N}_H h_\mu} + 1\right)\frac{1}{n}}\right).$$

REMARK 5. (i) For the SUBJ scheme, if we let $h_\mu \asymp (n\bar{N}_H)^{-1/5}$, the L^2 rate for $\hat{\mu}_{\text{subj}}$ is the same as the minimax rate in Theorem 3.2 of Cai and Yuan (2011), who also considered the same weighing scheme but employed a smoothing spline estimator. Thus, the local polynomial smoother enjoys the same minimax optimality as this smoothing spline estimator.

(ii) By Jensen's inequality, $\bar{N} \geq \bar{N}_H$ and $\bar{N}_{S2} \geq (\bar{N})^2$, so Remark 3 also applies here. Moreover, if $\limsup_n \bar{N}_{S2}/(\bar{N})^2 < \infty$, the rate for the OBS mean estimator is always as good as the SUBJ estimator and in some special cases the former could be better. For example, if $N_i \leq m < \infty$, for $i = 1, \dots, [n/2]$, and $N_i \asymp n^\alpha$, for $i = [n/2]+1, \dots, n$, where $\alpha \geq 1/4$, then the rate of $\hat{\mu}_{\text{obs}}$ is strictly better than that of $\hat{\mu}_{\text{subj}}$. Explicitly, root-n rate can be attained for $\hat{\mu}_{\text{obs}}$ since $\bar{N} \asymp n^\alpha \geq n^{1/4}$, while the rate for $\hat{\mu}_{\text{subj}}$ is the same as the one for sparse data since $\bar{N}_H = O(1)$. This phenomenon is numerically confirmed in Section 6. The above statement needs to be taken with caution when the two schemes leads to the same $\sqrt{n}$-rate of convergence, where the SUBJ scheme is better for ultra-dense data as it leads to a smaller asymptotic variance.

COROLLARY 4.2. *Suppose that* (A1)–(A2), (B1)–(B4), (C1b) *and* (C3b) *hold*.

(a) *OBS*: *Assume* $\limsup_n \bar{N}_{S2}/(\bar{N})^2 < \infty$.

(1) *When* $\bar{N}/n^{1/4} \to 0$ *and* $h_\mu \asymp (n\bar{N})^{-1/5}$,

$$\|\hat{\mu}_{\text{obs}} - \mu\|_2 = O_p\left(h_\mu^2 + \frac{1}{\sqrt{n\bar{N}h_\mu}}\right).$$

(2) *When* $\bar{N}/n^{1/4} \to C$, *where* $0 < C < \infty$, *and* $h_\mu \asymp n^{-1/4}$,

$$\|\hat{\mu}_{\text{obs}} - \mu\|_2 = O_p\left(\frac{1}{\sqrt{n}}\right).$$

(3) *When* $\bar{N}/n^{1/4} \to \infty$, $h_\mu = o(n^{-1/4})$, *and* $h_\mu\bar{N} \to \infty$,

$$\|\hat{\mu}_{\text{obs}} - \mu\|_2 = O_p\left(\frac{1}{\sqrt{n}}\right).$$

(b) *SUBJ*: *Under* (C2b), *replacing* $\bar{N}$ *in* (a) *with* $\bar{N}_H$ *leads to the corresponding results*.

4.2. *Covariance function.* For the same reason as in Section 3.2, we assume that the mean function μ is known. The rate for $\hat{\gamma}$ in (2.4) is given in the next theorem.

THEOREM 4.2. *Under* (A1)–(A2), (B1)–(B4) *and* (D1b)–(D3b),

$$\|\hat{\gamma}-\gamma\|_{\mathrm{HS}} = O_p\Biggl(h_\gamma^2 + \biggl[\frac{\sum_{i=1}^n N_i(N_i-1)v_i^2}{h_\gamma^2} + \frac{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2}{h_\gamma} + \sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2\biggr]^{1/2}\Biggr).$$

We show below the corresponding results for $\hat{\gamma}_{\mathrm{obs}}$ and $\hat{\gamma}_{\mathrm{subj}}$. Denote $\bar{N}_{H2} = (n^{-1}\sum_{i=1}^n N_i^{-2})^{-1}$.

COROLLARY 4.3. *Suppose that* (A1)–(A2), (B1)–(B4), (D1b) *and* (D3b) *hold.*

(a) *OBS*:

$$\|\hat{\gamma}_{\mathrm{obs}}-\gamma\|_{\mathrm{HS}} = O_p\Biggl(h_\gamma^2 + \sqrt{\biggl(\frac{1}{\bar{N}_{S2}h_\gamma^2} + \frac{\bar{N}_{S3}}{(\bar{N}_{S2})^2 h_\gamma} + \frac{\bar{N}_{S4}}{(\bar{N}_{S2})^2}\biggr)\frac{1}{n}}\Biggr).$$

(b) *SUBJ*: *Under an additional assumption* (D2b),

$$\|\hat{\gamma}_{\mathrm{subj}}-\gamma\|_{\mathrm{HS}} = O_p\Biggl(h_\gamma^2 + \sqrt{\biggl(\frac{1}{\bar{N}_{H2}h_\gamma^2} + \frac{1}{\bar{N}_H h_\gamma} + 1\biggr)\frac{1}{n}}\Biggr).$$

Since $\bar{N} \geq \bar{N}_H$ and $\bar{N}_{S2} \geq \bar{N}_{H2}$, Jensen's inequality implies that the rate for $\hat{\gamma}_{\mathrm{obs}}$ is always as good as $\hat{\gamma}_{\mathrm{subj}}$ if $\limsup_n(\bar{N}_{S2})/(\bar{N})^2 < \infty$, $\limsup_n(\bar{N}\cdot\bar{N}_{S3})/(\bar{N}_{S2})^2 < \infty$, and $\limsup_n \bar{N}_{S4}/(\bar{N}_{S2})^2 < \infty$.

COROLLARY 4.4. *Suppose that* (A1)–(A2), (B1)–(B4), (D1b) *and* (D3b) *hold.*

(a) *OBS*: *Assume that* $\limsup_n(\bar{N}_{S2})/(\bar{N})^2, \limsup_n(\bar{N}\cdot\bar{N}_{S3})/(\bar{N}_{S2})^2, \limsup_n \bar{N}_{S4}/(\bar{N}_{S2})^2 < \infty$.

(1) *When* $\bar{N}_{S2}/n^{1/2} \to 0$, *and* $h_\gamma \asymp (n\bar{N}_{S2})^{-1/6}$,

$$\|\hat{\gamma}_{\mathrm{obs}}-\gamma\|_{\mathrm{HS}} = O_p\biggl(h_\gamma^2 + \frac{1}{\sqrt{n\bar{N}_{S2}h_\gamma^2}}\biggr).$$

(2) *When* $\bar{N}_{S2}/n^{1/2} \to C$, *and* $h_\gamma \asymp n^{-1/4}$,

$$\|\hat{\gamma}_{\text{obs}} - \gamma\|_{\text{HS}} = O_p\left(\frac{1}{\sqrt{n}}\right).$$

(3) *When* $\bar{N}_{S2}/n^{1/2} \to \infty$, $h_\gamma = o(n^{-1/4})$ *and* $h_\gamma \bar{N} \to \infty$,

$$\|\hat{\gamma}_{\text{obs}} - \gamma\|_{\text{HS}} = O_p\left(\frac{1}{\sqrt{n}}\right).$$

(b) *SUBJ*: *Under an additional assumption* (D2b), *replacing* $\bar{N}$ *and* $\bar{N}_{S2}$ *in* (a) *with* $\bar{N}_H$ *and* $\bar{N}_{H2}$, *respectively, leads to the corresponding results.*

REMARK 6. OBS and SUBJ are the most commonly used schemes, but for some cases, using a weighing scheme that is neither OBS nor SUBJ is likely to achieve a better rate. For instance, one can use a mixture of the OBS and SUBJ schemes, that is, $w_i = \alpha/(n\bar{N}) + (1-\alpha)/(nN_i)$ for some $0 \le \alpha \le 1$ to estimate μ, which satisfies both Assumptions (C1b) and (C2b). It can be easily shown that the corresponding mean estimator, denoted by $\hat{\mu}_\alpha$, can achieve the following rate:

$$\|\hat{\mu}_\alpha - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\alpha^2 c_{n1} + (1-\alpha)^2 c_{n2}}\right),$$

$$\text{where } c_{n1} = \left(\frac{1}{\bar{N}h_\mu} + \frac{\bar{N}_{S2}}{(\bar{N})^2}\right)\frac{1}{n}, \text{ and } c_{n2} = \left(\frac{1}{\bar{N}_H h_\mu} + 1\right)\frac{1}{n}.$$

In particular, if we let $\alpha^* = c_{n2}/(c_{n1} + c_{n2})$ such that $\alpha^2 c_{n1} + (1-\alpha)^2 c_{n2}$ is minimized, the rate becomes

$$\|\hat{\mu}_{\alpha^*} - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\frac{c_{n1}c_{n2}}{c_{n1} + c_{n2}}}\right).$$

Since $c_{n1}c_{n2}/(c_{n1} + c_{n2}) \le \min\{c_{n1}, c_{n2}\}$, the estimator $\hat{\mu}_{\alpha^*}$ always attains at least the better rate between the two estimators, $\hat{\mu}_{\text{obs}}$ and $\hat{\mu}_{\text{subj}}$. This is a very appealing feature since in reality it is difficulty to judge which of the two schemes is a better one but by using $\hat{\mu}_{\alpha^*}$ we will always do at least as well as, if not better than, both.

Similarly, to estimate γ, one could use $v_i = \theta/[\sum_{i=1}^n N_i(N_i - 1)] + (1 - \theta)/[nN_i(N_i - 1)]$ for some $0 \le \theta \le 1$ as an alternative weighing scheme, and the rate of the corresponding covariance estimator, denoted by $\hat{\gamma}_\theta$, is

$$\|\hat{\gamma}_\theta - \gamma\|_{\text{HS}} = O_p\left(h_\gamma^2 + \sqrt{\theta^2 d_{n1} + (1-\theta)^2 d_{n2}}\right),$$

where

$$d_{n1} = \left(\frac{1}{\bar{N}_{S2}h_\gamma^2} + \frac{\bar{N}_{S3}}{(\bar{N}_{S2})^2 h_\gamma} + \frac{\bar{N}_{S4}}{(\bar{N}_{S2})^2}\right)\frac{1}{n} \quad \text{and}$$

$$d_{n2} = \left(\frac{1}{\bar{N}_{H2}h_\gamma^2} + \frac{1}{\bar{N}_H h_\gamma} + 1\right)\frac{1}{n}.$$

A particular choice of θ is $\theta^* = d_{n2}/(d_{n1} + d_{n2})$ that minimizes this rate, and $\hat{\gamma}_{\theta^*}$ always attains at least the better rate between $\hat{\gamma}_{\text{obs}}$ and $\hat{\gamma}_{\text{subj}}$.

5. Uniform convergence. In this section, we focus on the uniform convergence for mean and covariance estimators. Again, for OBS and SUBJ schemes, there are three types of uniform rates for non-dense, dense and ultra-dense functional data, respectively.

5.1. *Mean function.*

THEOREM 5.1. *Under* (A1)–(A2), (B1)–(B4) *and* (C1c)–(C3c),

$$\sup_{t\in[0,1]} |\hat{\mu}(t) - \mu(t)| = O\left(h_\mu^2 + \left\{\log(n)\left[\frac{\sum_{i=1}^n N_i w_i^2}{h_\mu} + \sum_{i=1}^n N_i(N_i - 1)w_i^2\right]\right\}^{1/2}\right) \quad a.s.$$

The respective rates of convergence for $\hat{\mu}_{\text{obs}}$ and $\hat{\mu}_{\text{subj}}$ are as follows.

COROLLARY 5.1. *Suppose that* (A1)–(A2), (B1)–(B4) *and* (C1c)–(C2c) *hold*.

(a) *OBS*: *Under an additional assumption* (C3c),

$$\sup_{t\in[0,1]} |\hat{\mu}_{\text{obs}}(t) - \mu(t)| = O\left(h_\mu^2 + \sqrt{\left(\frac{1}{\bar{N}h_\mu} + \frac{\bar{N}_{S2}}{(\bar{N})^2}\right)\frac{\log(n)}{n}}\right) \quad a.s.$$

(b) *SUBJ*:

$$\sup_{t\in[0,1]} |\hat{\mu}_{\text{subj}}(t) - \mu(t)| = O\left(h_\mu^2 + \sqrt{\left(\frac{1}{\bar{N}_H h_\mu} + 1\right)\frac{\log(n)}{n}}\right) \quad a.s.$$

The result in Corollary 5.1 for the SUBJ scheme was first established in Li and Hsing (2010). Here, we restate it for completeness as it facilitates the next corollary.

COROLLARY 5.2. *Suppose that* (A1)–(A2), (B1)–(B4), *and* (C1c)–(C2c) *hold*.

(a) *OBS*: *Assume* (C3c).

(1) *When* $\bar{N}/(n/\log(n))^{1/4} \to 0$ *and* $h_\mu \asymp (n\bar{N}/\log(n))^{-1/5}$,

$$\sup_{t\in[0,1]} |\hat{\mu}_{\text{obs}}(t) - \mu(t)| = O\left(h_\mu^2 + \sqrt{\frac{\log(n)}{n\bar{N}h_\mu}}\right) \quad a.s.$$

(2) *When* $\bar{N}/(n/\log(n))^{1/4} \to C$, *where* $C > 0$, *and* $h_\mu \asymp (n/\log(n))^{-1/4}$,

$$\sup_{t\in[0,1]} \left|\hat{\mu}_{\text{obs}}(t) - \mu(t)\right| = O\left(\sqrt{\frac{\log(n)}{n}}\right) \qquad a.s.$$

(3) *When* $\bar{N}/(n/\log(n))^{1/4} \to \infty$, $h_\mu = o((n/\log(n))^{-1/4})$, *and* $h_\mu \bar{N} \to \infty$,

$$\sup_{t\in[0,1]} \left|\hat{\mu}_{\text{obs}}(t) - \mu(t)\right| = O\left(\sqrt{\frac{\log(n)}{n}}\right) \qquad a.s.$$

(b) *SUBJ*: *We can obtain the corresponding results by replacing* $\bar{N}$ *in* (a) *with* $\bar{N}_H$.

5.2. *Covariance function.*

THEOREM 5.2. *Under* (A1)–(A2), (B1)–(B4), (C1c)–(C3c) *and* (D1c)–(D3c),

$$\begin{aligned}
&\sup_{s,t\in[0,1]} \left|\hat{\gamma}(s,t) - \gamma(s,t)\right| \\
&\quad = O\Bigg(h_\mu^2 + \left\{\log(n)\left[\frac{\sum_{i=1}^n N_i w_i^2}{h_\mu} + \sum_{i=1}^n N_i(N_i-1)w_i^2\right]\right\}^{1/2} \\
&\qquad + h_\gamma^2 + \Bigg\{\log(n)\Bigg[\frac{\sum_{i=1}^n N_i(N_i-1)v_i^2}{h_\gamma^2} + \frac{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2}{h_\gamma} \\
&\qquad + \sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2\Bigg]\Bigg\}^{1/2}\Bigg) \qquad a.s.
\end{aligned}$$

Theorem 5.2 indicates that the rate for $\hat{\gamma}$ depends on the rate for $\hat{\mu}$. The uniform rates for the OBS and SUBJ estimators are as follows.

COROLLARY 5.3. *Assume* (A1)–(A2), (B1)–(B4), (C1c)–(C2c) *and* (D1c)–(D2c).

(a) *OBS*: *Under two additional assumptions* (C3c) *and* (D3c),

$$\begin{aligned}
&\sup_{s,t\in[0,1]} \left|\hat{\gamma}_{\text{obs}}(s,t) - \gamma(s,t)\right| \\
&\quad = O\Bigg(h_\mu^2 + \sqrt{\left(\frac{1}{\bar{N}h_\mu} + \frac{\bar{N}_{S2}}{(\bar{N})^2}\right)\frac{\log(n)}{n}} + h_\gamma^2 \\
&\qquad + \sqrt{\left(\frac{1}{\bar{N}_{S2}h_\gamma^2} + \frac{\bar{N}_{S3}}{(\bar{N}_{S2})^2 h_\gamma} + \frac{\bar{N}_{S4}}{(\bar{N}_{S2})^2}\right)\frac{\log(n)}{n}}\Bigg) \qquad a.s.
\end{aligned}$$

(b) *SUBJ*:

$$\sup_{s,t\in[0,1]} \left|\hat{\gamma}_{\text{subj}}(s,t) - \gamma(s,t)\right| = O\Bigg(h_\mu^2 + \sqrt{\left(\frac{1}{\bar{N}_H h_\mu} + 1\right)\frac{\log(n)}{n}} + h_\gamma^2 + \sqrt{\left(\frac{1}{\bar{N}_{H2} h_\gamma^2} + \frac{1}{\bar{N}_H h_\gamma} + 1\right)\frac{\log(n)}{n}}\Bigg) \qquad a.s.$$

REMARK 7. Li and Hsing (2010) also assigned equal weight for each subject to estimate γ, but used a different procedure. We denote their estimator by $\hat{\gamma}_{\text{LH}}$ which is as follows: $\hat{\gamma}_{\text{LH}}(s,t) = \hat{a}_0 - \hat{\mu}_{\text{subj}}(s)\hat{\mu}_{\text{subj}}(t)$ where

$$\begin{aligned}
&(\hat{a}_0, \hat{a}_1, \hat{a}_2) \\
&\quad = \underset{a_0,a_1,a_2}{\operatorname{argmin}} \sum_{i=1}^n \frac{1}{nN_i(N_i-1)} \\
&\qquad \times \sum_{1\le j\ne l\le N_i} \left[Y_{ij}Y_{il} - a_0 - a_1(T_{ij}-s) - a_2(T_{il}-t)\right]^2 \qquad (5.1)\\
&\qquad \times K_{h_\gamma}(T_{ij}-s)K_{h_\gamma}(T_{il}-t).
\end{aligned}$$

Obviously, $\hat{\gamma}_{\text{subj}}$ is obtained by smoothing residuals from the mean estimation while $\hat{\gamma}_{\text{LH}}$ is obtained by first estimating $EX(s)X(t)$ and then centering by $\hat{\mu}_{\text{subj}}$. Theoretically, $\hat{\gamma}_{\text{subj}}$ and $\hat{\gamma}_{\text{LH}}$ have the same rate of convergence but they differ in numerical performance as shown in Section 6, where $\hat{\gamma}_{\text{subj}}$ outperforms $\hat{\gamma}_{\text{LH}}$.

COROLLARY 5.4. *Assume* (A1)–(A2), (B1)–(B4), (C1c)–(C2c) *and* (D1c)–(D2c).

(a) *OBS*: *Assume* (C3c) *and* (D3c).

(1) *When* $\bar{N}_{S2}/(n/\log(n))^{1/2} \to 0$, $\bar{N}/(n/\log(n))^{1/4} \to 0$, $h_\mu \asymp (n\bar{N}/\log(n))^{-1/5}$, *and* $h_\gamma \asymp (n\bar{N}_{S2}/\log(n))^{-1/6}$,

$$\sup_{s,t\in[0,1]} \left|\hat{\gamma}_{\text{obs}}(s,t) - \gamma(s,t)\right| = O\left(h_\gamma^2 + \sqrt{\frac{\log(n)}{n\bar{N}_{S2}h_\gamma^2}}\right) \qquad a.s.$$

(2) *When* $\bar{N}_{S2}/(n/\log(n))^{1/2} \to C_1$, $\bar{N}/(n/\log(n))^{1/4} \to C_2$, *where* $0 < C_1, C_2 < \infty$, *and* $h_\mu \asymp h_\gamma \asymp (n/\log(n))^{-1/4}$,

$$\sup_{s,t\in[0,1]} \left|\hat{\gamma}_{\text{obs}}(s,t) - \gamma(s,t)\right| = O\left(\sqrt{\frac{\log(n)}{n}}\right) \qquad a.s.$$

(3) *When* $\bar{N}_{S2}/(n/\log(n))^{1/2} \to \infty$, $\bar{N}/(n/\log(n))^{1/4} \to \infty$, $h_\mu, h_\gamma = o((n/\log(n))^{-1/4})$, *and* $h_\mu\bar{N}, h_\gamma\bar{N} \to \infty$,

$$\sup_{s,t\in[0,1]} \left|\hat{\gamma}_{\text{obs}}(s,t) - \gamma(s,t)\right| = O\left(\sqrt{\frac{\log(n)}{n}}\right) \qquad a.s.$$

(b) *SUBJ*: *Replacing* $\bar{N}$ *and* $\bar{N}_{S2}$ *in* (a) *with* $\bar{N}_H$ *and* $\bar{N}_{H2}$, *respectively*, *leads to the corresponding results*.

Following similar arguments in Remark 6, the uniform rate of $\hat{\mu}_{\alpha^*}$ is at least as good as those of $\hat{\mu}_{\text{obs}}$ and $\hat{\mu}_{\text{subj}}$. Given the same effect of $\hat{\mu}$ on the rate of $\hat{\gamma}$, $\hat{\gamma}_{\theta^*}$ always performs at least as well as, if not better than, both of $\hat{\gamma}_{\text{obs}}$ and $\hat{\gamma}_{\text{subj}}$.

6. Simulation. In this section, we evaluate the numerical performances of the OBS and SUBJ schemes, together with the weighing schemes based on their mixture where we only consider $\hat{\mu}_{\alpha^*}$ and $\hat{\gamma}_{\theta^*}$ as defined in Remark 6. We also compare $\hat{\gamma}_{\text{subj}}$ with $\hat{\gamma}_{\text{LH}}$ to investigate which estimation procedure is better as mentioned in Remark 7.

We considered four settings for N_i:

- *Setting* 1: N_i are i.i.d. from a discrete uniform distribution on the set $\{2, 3, 4, 5\}$.
- *Setting* 2: $N_i = n/4$ with probability 0.5, and follow Setting 1 with probability 0.5.
- *Setting* 3: $N_i = n/4$ with probability $n^{-1/2}$, and follow Setting 1 with probability $1 - n^{-1/2}$.
- *Setting* 4: N_i are i.i.d. from a discrete uniform distribution on the interval $[\lfloor n/8 \rfloor, \lfloor 3n/8 \rfloor]$.

The notation $\lfloor x \rfloor$ represents the integer part of x.

Setting 1, where all N_i are bounded, corresponds to sparse data for both OBS and SUBJ schemes. In Setting 2, N_i are bi-partitioned with approximately half bounded and half proportional to n. In Setting 3, a majority of the subjects are sparsely measured but a small portion have a large amount of observations. In Setting 4, the data are purely ultra-dense for both OBS and SUBJ schemes. Settings 1–4 are designed to confirm Remarks 3 and 5. By Remark 3, we expect the OBS scheme to outperform the SUBJ scheme in Setting 1, while the SUBJ scheme would be superior in Setting 4. By Remark 5(ii), since $\bar{N} \asymp n$ and $\bar{N}_H \asymp 1$ in Setting 2, the OBS estimator is expected to perform better. In Setting 3, since (C3c) and (D3c) are violated for the OBS scheme, and $\bar{N}_{S2}/(\bar{N})^2, \bar{N}_{S4}/(\bar{N}_{S2})^2 \asymp n^{1/2}$, the SUBJ scheme is expected to be better.

In each setting, we generated $Q = 500$ simulation runs with $n = 200$ subjects in each run. The true mean and covariance functions are respectively

$$\mu(t) = 1.5 \sin\bigl(3\pi(t + 0.5)\bigr) + 2t^3, \quad \gamma(s,t) = \sum_{k=1}^{4} \lambda_k \phi_k(s)\phi_k(t), \qquad s,t \in [0,1],$$

where $\lambda_k = 1/(k+1)^2, k = 1, \ldots, 4$ and

$$\phi_1(t) = \sqrt{2}\cos(2\pi t), \qquad \phi_2(t) = \sqrt{2}\sin(2\pi t),$$
$$\phi_3(t) = \sqrt{2}\cos(4\pi t), \qquad \phi_4(t) = \sqrt{2}\sin(4\pi t).$$

TABLE 1

Minimum MISE and MSAE of $\hat{\mu}_{\text{obs}}$, $\hat{\mu}_{\text{subj}}$ and $\hat{\mu}_{\alpha^*}$ in the four settings for N_i's. The corresponding standard deviations are in the parentheses

		Setting 1	**Setting 2**	**Setting 3**	**Setting 4**
MISE (1e-2)	$\hat{\mu}_{\text{obs}}$	1.1304 (0.4832)	0.5793 (0.3910)	1.4149 (0.9514)	0.3387 (0.2117)
	$\hat{\mu}_{\text{subj}}$	1.1875 (0.5046)	0.7857 (0.3540)	1.1280 (0.4841)	0.3304 (0.1920)
	$\hat{\mu}_{\alpha^*}$	1.1331 (0.4827)	0.5328 (0.3025)	0.9966 (0.5065)	0.3277 (0.1967)
MSAE	$\hat{\mu}_{\text{obs}}$	0.2972 (0.1137)	0.1715 (0.0574)	0.2717 (0.0915)	0.1326 (0.0401)
	$\hat{\mu}_{\text{subj}}$	0.3059 (0.1203)	0.2339 (0.0776)	0.2899 (0.1061)	0.1331 (0.0403)
	$\hat{\mu}_{\alpha^*}$	0.2980 (0.1159)	0.1780 (0.0604)	0.2589 (0.0925)	0.1314 (0.0397)

The time points T_{ij} are i.i.d. from the uniform distribution on $[0, 1]$ and $U_{ij} = \sum_{k=1}^{4} A_{ik}\phi_k(T_{ij})$ where A_{ik} are independent from $N(0, \lambda_k)$. The response was generated by $Y_{ij} = \mu(T_{ij}) + U_{ij} + e_{ij}$ where e_{ij} are i.i.d. from $N(0, 0.01)$. Epanechnikov kernel was used and the values of h_μ and h_γ varied on a dense grid. To evaluate the performance of $\hat{\mu}_{\text{obs}}$, given a specific bandwidth h_μ, we calculated the mean integrated squared error (MISE) and mean supremum absolute error (MSAE) defined as follows:

$$\text{MISE}(\hat{\mu}_{\text{obs}}, h_\mu) = \frac{1}{Q}\sum_{q=1}^{Q}\int_0^1 \big[\hat{\mu}_{\text{obs}}(t)^{[q]} - \mu(t)\big]^2\,dt,$$

$$\text{MSAE}(\hat{\mu}_{\text{obs}}, h_\mu) = \frac{1}{Q}\sum_{q=1}^{Q}\sup_{t\in[0,1]}\big|\hat{\mu}_{\text{obs}}(t)^{[q]} - \mu(t)\big|,$$

where $\{\hat{\mu}_{\text{obs}}^{[q]}, q = 1, \ldots, Q\}$ are mean estimators obtained from the $Q = 500$ datasets. We can similarly define $\text{MISE}(\hat{\mu}_{\text{subj}}, h_\mu)$, $\text{MSAE}(\hat{\mu}_{\text{subj}}, h_\mu)$, $\text{MISE}(\hat{\mu}_{\alpha^*}, h_\mu)$, and $\text{MSAE}(\hat{\mu}_{\alpha^*}, h_\mu)$. The minimum MISE and MSAE over the grid of bandwidths are presented in Table 1.

To evaluate the covariance estimators, we assume that the mean μ is known so that the covariance estimation is distangled from the mean estimation. We defined MISE and MSAE as follows:

$$\text{MISE}(\hat{\gamma}_{\text{obs}}, h_\gamma) = Q^{-1}\sum_{q=1}^{Q}\iint_{[0,1]^2}\big[\hat{\gamma}_{\text{obs}}(s,t)^{[q]} - \gamma(s,t)\big]^2\,ds\,dt,$$

$$\text{MSAE}(\hat{\gamma}_{\text{obs}}, h_\gamma) = Q^{-1}\sum_{q=1}^{Q}\sup_{s,t\in[0,1]}\big|\hat{\gamma}_{\text{obs}}(s,t)^{[q]} - \gamma(s,t)\big|,$$

where $\{\hat{\gamma}_{\text{obs}}^{[q]}, q = 1, \ldots, Q\}$ are the estimators from the $Q = 500$ datasets. $\text{MISE}(\hat{\gamma}_{\text{subj}}, h_\gamma)$, $\text{MSAE}(\hat{\gamma}_{\text{subj}}, h_\gamma)$, $\text{MISE}(\hat{\gamma}_{\theta^*}, h_\gamma)$, $\text{MSAE}(\hat{\gamma}_{\theta^*}, h_\gamma)$, $\text{MISE}(\hat{\gamma}_{\text{LH}},$

TABLE 2

Minimum MISE and MSAE of $\hat{\gamma}_{\mathrm{obs}}$, $\hat{\gamma}_{\mathrm{subj}}$, $\hat{\gamma}_{\theta^*}$ and $\hat{\gamma}_{\mathrm{LH}}$ in the four settings for N_i's. The corresponding standard deviations are in the parentheses

		Setting 1	**Setting 2**	**Setting 3**	**Setting 4**
MISE (1e-2)	$\hat{\gamma}_{\mathrm{obs}}$	1.3489 (0.5249)	0.3970 (0.2159)	1.9106 (1.3814)	0.2487 (0.1297)
	$\hat{\gamma}_{\mathrm{subj}}$	1.6796 (0.7680)	0.9485 (0.3602)	1.5385 (0.5715)	0.2184 (0.0968)
	$\hat{\gamma}_{\theta^*}$	1.3840 (0.5592)	0.3913 (0.1797)	1.1089 (0.4028)	0.2171 (0.1036)
	$\hat{\gamma}_{\mathrm{LH}}$	26.8078 (12.7777)	12.5926 (4.5385)	23.9794 (10.6328)	2.2406 (0.9320)
MSAE	$\hat{\gamma}_{\mathrm{obs}}$	0.4611 (0.1451)	0.2375 (0.0949)	0.4523 (0.2086)	0.1898 (0.0640)
	$\hat{\gamma}_{\mathrm{subj}}$	0.5138 (0.1867)	0.4049 (0.1717)	0.4975 (0.1834)	0.1875 (0.0644)
	$\hat{\gamma}_{\theta^*}$	0.4784 (0.1214)	0.2805 (0.1311)	0.4547 (0.1294)	0.1828 (0.0647)
	$\hat{\gamma}_{\mathrm{LH}}$	3.2164 (1.4695)	2.2682 (1.0607)	3.1272 (1.2488)	1.0148 (0.4434)

h_γ) and MSAE($\hat{\gamma}_{\mathrm{LH}}, h_\gamma$) can be similarly defined. The minimal MISE and MSAE over the grid of bandwidths are presented in Table 2.

By comparing $\hat{\mu}_{\mathrm{obs}}$ and $\hat{\mu}_{\mathrm{subj}}$ in Table 1, and $\hat{\gamma}_{\mathrm{obs}}$ and $\hat{\gamma}_{\mathrm{subj}}$ in Table 2, one can see that the OBS scheme leads to smaller MSAE values for both the mean and covariance estimation in all the four settings of N_i except for the covariance estimation in Setting 4. With respect to MISE, $\hat{\mu}_{\mathrm{obs}}$ and $\hat{\gamma}_{\mathrm{obs}}$ outperform their counterparts in Settings 1 and 2 while $\hat{\mu}_{\mathrm{subj}}$ and $\hat{\gamma}_{\mathrm{subj}}$ are superior in Settings 3 and 4, which conforms with our expectations. In consideration of Remark 3 on ultra-dense data and the violation of design assumptions for OBS scheme in Setting 3, it is interesting that $\hat{\mu}_{\mathrm{obs}}$ and $\hat{\gamma}_{\mathrm{obs}}$ outperform their competitors in terms of MSAE even in Setting 4 for mean estimation and in Setting 3 for both mean and covariance estimation. This suggests that the design conditions might be further relaxed and could be an interesting future project. As shown in Tables 1 and 2, both $\hat{\mu}_{\alpha^*}$ and $\hat{\gamma}_{\theta^*}$ performed equally well (in terms of MISE and MSAE) and in some settings better than both the corresponding OBS and SUBJ estimators. This provides the numerical evidence for the benefit of using a mixture of the OBS and SUBJ schemes as discussed in Remark 6.

Table 2 demonstrates that $\hat{\gamma}_{\mathrm{subj}}$ always outperforms $\hat{\gamma}_{\mathrm{LH}}$ with respect to both MISE and MSAE in all the four settings for N_i. This is perhaps not surprising as smoothing over residuals can reduce the error involved in the estimation and can be heuristically explained as follows: In (5.1),

$$Y_{ij}Y_{il} = X_i(T_{ij})X_i(T_{il}) + e_{ij}e_{il} + X_i(T_{ij})e_{il} + X_i(T_{il})e_{ij},$$

where $X_i(T_{ij})X_i(T_{il})$ targets $EX(s)X(t)$ and $e_{ij}e_{il} + X_i(T_{ij})e_{il} + X_i(T_{il})e_{ij}$ can be viewed as “noise.” In contrast, the raw covariance is

$$C_{ijl} \approx \delta_{ij}\delta_{il} = U_{ij}U_{il} + e_{ij}e_{il} + U_{ij}e_{il} + U_{il}e_{ij},$$

where $U_{ij}U_{il}$ targets $\gamma(s,t)$ and $e_{ij}e_{il} + U_{ij}e_{il} + U_{il}e_{ij}$ can be viewed as “noise.” Therefore, $\hat{\gamma}_{\mathrm{LH}}$ involves higher levels of “noise” than $\hat{\gamma}_{\mathrm{subj}}$ making the former less efficient as reflected in this simulation.

The optimal bandwidths in this simulation study were the best ones over a prefixed grid of bandwidths. It is useful to understand the numerical performance of the competing procedures but data-driven methods will be needed in practice and this remains an open and challenging issue in FDA. Leave-one-curve-out cross-validation (CV), as suggested by Rice and Silverman (1991), is commonly used. This method has two major advantages: First, it does not depend on the within-subject correlation structure; second, there is no need to specify whether data are non-dense, dense or ultra-dense. Its consistency property was established by Hart and Wehrly (1993) for a different mean function estimator from (2.3) where the entire trajectory of each X_i is observable. To reduce computational cost, K-fold CV is a feasible alternative to leave-one-curve-out CV, especially for covariance estimation. See Jiang and Wang (2010) among others. Generalized cross-validation is another computationally simpler method [Liu and Müller (2009), Zhang and Chen (2007)], which could alleviate the tendency of CV to undersmooth, but its consistency property has not been established. Since there exists no consensus on which bandwidth selection method is preferred, practitioners may rely on their own judgments. For example, one could obtain multiple bandwidths selected by different methods and choose the one with the best fit (e.g., the smallest residual sum of squares) or that is most interpretable.

7. Conclusion. In this paper, we focused on local linear smoothers when estimating the mean and covariance functions. We considered a general weighing scheme which incorporates the commonly used OBS and SUBJ schemes. Three types of convergence are investigated: asymptotic normality, L^2 convergence and uniform convergence. For each type, unified results were established in the sense that (1) the weighing scheme is general; (2) N_i is allowed to be of any order relative to n. The unified theory sheds light on the partition of functional data into three categories: non-dense, dense, and ultra-dense, when either the OBS or SUBJ weighing scheme is employed. The partition depends on the target (mean or covariance) and the weighing scheme (OBS or SUBJ). Theoretical and numerical performances of OBS and SUBJ schemes are also systematically compared. We additionally consider and compare a new class of weighing schemes based on a mixture of the OBS and SUBJ weights.

To summarize, functional data are considered as dense if root-n rate can be attained for the corresponding estimator. Otherwise, data are non-dense (or at least not dense enough to achieve the root-n rate). For non-dense data, asymptotic results are comparable to those nonparametric smoothing methods when all the measurements are independent. When $N_i = O(1)$ for all i, the data are often called "sparse data," a special case of non-dense data with the lowest rate of convergence. We caution here that the technical meaning of "sparse," "non-dense," versus "dense" data may not conform with their literary meanings since it is possible to achieve the same rate as conventional sparse data even if some of N_i's are unbounded. We lean toward a technical viewpoint to define sparse functional data as those with the lowest convergence rate and term "semi-dense" functional data as those that are neither sparse nor dense, that is, those with rates of convergence faster than the rate for sparse data but slower than root-n. For dense data, we propose a further partition through the asymptotic bias. Functional data can be regarded as falling in the parametric paradigm only when both the parametric rate (root-n) of convergence is attained and the asymptotic bias is of the order $o(n^{-1/2})$, which corresponds to our definition of ultra-dense data. This reopens the discussion on what should be considered dense functional data. We adopt the convention that defines dense data as those that can achieve the parametric (root-n) rate of convergence but introduce a new category of "ultra-dense functional data" when additionally the asymptotic bias is of the order $o(n^{-1/2})$.

For the two global convergence criteria, under additional assumptions such as (C3c), (D3c) and $\limsup_n \bar{N}_{S2}/(\bar{N})^2 < \infty$ as in Remark 5, the convergence rates of the OBS estimators are at least as good as, and sometimes better than, the SUBJ estimators. These additional assumptions essentially imply that the distribution of N_i cannot have a heavy tail. Pointwise asymptotic results suggest that the OBS scheme is more efficient for non-dense data but the SUBJ scheme is more efficient for ultra-dense data, under which the SUBJ estimators are asymptotically equivalent to the sample mean or sample covariance function when the true individual curves are observed without errors. Although the theory in this paper helps to understand when and why one weighing scheme might be better than the other, the actual choice between the OBS and SUBJ schemes is still challenging in reality since it may not be obvious which type of functional data one has in hand. A general guideline is to adopt the OBS scheme unless one believes that the data are ultra-dense or if the distribution of N_i suggests a heavy upper tail. An ad-hoc means to detect ultra-dense data was proposed in Kim and Zhao (2013), who considered ultra-dense data as those when n is about 30–200 and $\min_{1\le i\le n} N_i$ is about 10–30. If the consideration is not limited to the OBS and SUBJ schemes, one could alternatively use a mixture of OBS and SUBJ weights as in Remark 6 which can guarantee at least the better rate of convergence between the OBS and SUBJ estimators.

The asymptotic properties developed in this paper will play a direct role in statistical inference, for instance, in the construction of a simultaneous confidence band for the mean function. When data are sufficiently dense, the mean process $\sqrt{n}(\hat{\mu}(t) - \mu(t))$ is likely to weakly converge to a Gaussian process $W(t)$, which can facilitate the confidence band based on $\sup_t |W(t)|$ due to the continuous mapping theorem. See Degras (2011) for the special case of dense data with measurements sampled according to the same time schedule for all subjects. However, when data are not dense enough, this process is not tight, so the construction needs to follow the line of work by Bickel and Rosenblatt (1973), and this has been illustrated in Zheng, Yang and Härdle (2014) for sparse functional data. The general approach developed in this paper can facilitate the construction of simultaneous confidence band for any design once we realize whether a parametric or nonparametric convergence rate emerges for the corresponding estimate.

The partition on functional data may shed light on the asymptotic properties in a broader context. An immediate application is functional principal component analysis (FPCA) by invoking the perturbation theory [Bosq (2000), Dauxois, Pousse and Romain (1982), Hall and Hosseini-Nasab (2006), Kato (1980)]. However, as demonstrated in Hall, Müller and Wang (2006) for sparse functional data, this may not lead to optimal results since a better rate, same as for one-dimensional nonparametric smoothing, can be attained for the estimated eigenfunction if the covariance function was under-smoothed before the eigenfunctions were constructed. In other words, with proper care the results in this paper will be useful to explore the corresponding properties derived from FPCA.

Another application is functional regression. The results in this paper will be applicable to concurrent models [Ramsay and Silverman (2005)], such as varying-coefficient models [Hoover et al. (1998), Huang, Wu and Zhou (2002), Wu and Chiang (2000)], functional index models [Jiang and Wang (2011)], and functional additive models [Zhang, Park and Wang (2013)]. For more complicated regression models that involve the inversion of a covariance operator [Cai and Hall (2006), Cardot, Ferraty and Sarda (1999), Hall and Horowitz (2007), Yao, Müller and Wang (2005b)], which is an ill-posed problem, the impact of N_i on the asymptotic properties is unknown but the approach in this paper might be useful for future explorations.

APPENDIX A: ASSUMPTIONS

A.1. Kernel function. We assume the one-dimensional kernel K to follow

(A1) $K(\cdot)$ is a symmetric probability density function on $[-1, 1]$ and

$$\sigma_K^2 = \int u^2 K(u)\,du < \infty, \qquad \|K\|^2 = \int K(u)^2\,du < \infty.$$

(A2) $K(\cdot)$ is Lipschitz continuous: There exists $0 < L < \infty$ such that

$$\left|K(u) - K(v)\right| \le L|u - v|, \qquad \text{for any } u, v \in [0, 1].$$

This implies $K(\cdot) \le M_K$ for a constant M_K.

A.2. Time points and true functions.

(B1) $\{T_{ij} : i = 1, \dots, n, j = 1, \dots, N_i\}$, are i.i.d. copies of a random variable T defined on $[0, 1]$. The density $f(\cdot)$ of T is bounded from below and above:

$$0 < m_f \le \min_{t\in[0,1]} f(t) \le \max_{t\in[0,1]} f(t) \le M_f < \infty.$$

Furthermore, $f^{(2)}(\cdot)$, the second derivative of $f(\cdot)$, is bounded.

(B2) X is independent of T and e is independent of T and U.

(B3) $\mu^{(2)}(t)$, the second derivative of $\mu(t)$, is bounded on $[0, 1]$.

(B4) $\partial^2\gamma(s,t)/\partial s^2$, $\partial^2\gamma(s,t)/\partial s\,\partial t$ and $\partial^2\gamma(s,t)/\partial t^2$ are bounded on $[0, 1]^2$.

A.3. Assumptions for mean estimation.

- Asymptotic normality:

(C1a) $h_\mu \to 0$, $\sum_{i=1}^n N_i w_i^2/h_\mu \to 0$, $\sum_{i=1}^n N_i(N_i-1)w_i^2 \to 0$.

(C2a) $\max\{\sum_{i=1}^n N_i w_i^3/h_\mu^2, \sum_{i=1}^n N_i(N_i-1)w_i^3/h_\mu, \sum_{i=1}^n N_i(N_i-1)\times(N_i-2)w_i^3\}/[\sum_{i=1}^n N_i w_i^2/h_\mu + \sum_{i=1}^n N_i(N_i-1)w_i^2]^{3/2} \to 0$.

(C3a) $\sup_{t\in[0,1]} E|U(t)|^3 < \infty$ and $E|e|^3 < \infty$.

- L^2 convergence:

(C1b) $h_\mu \to 0$, $\log(1/\sum_{i=1}^n N_i w_i^2)\sum_{i=1}^n N_i w_i^2/h_\mu \to 0$, $\sum_{i=1}^n N_i(N_i-1)w_i^2 \to 0$.

(C2b) $\max_i w_i^2 \log(1/\sum_{i=1}^n N_i w_i^2)/[h_\mu \sum_{i=1}^n N_i w_i^2] \to 0$.

(C3b) $\sup_{t\in[0,1]} E|U(t)|^2 < \infty$ and $E|e|^2 < \infty$.

- Uniform convergence:

(C1c) $h_\mu \to 0$, $\log(n)\sum_{i=1}^n N_i w_i^2/h_\mu \to 0$, $\log(n)\sum_{i=1}^n N_i(N_i-1)w_i^2 \to 0$.

(C2c) For some $\alpha > 2$, $E\sup_{t\in[0,1]}|U(t)|^\alpha < \infty$, $E|e|^\alpha < \infty$ and

$$n\left[\sum_{i=1}^n N_i w_i^2 h_\mu + \sum_{i=1}^n N_i(N_i-1)w_i^2 h_\mu^2\right]\left[\frac{\log(n)}{n}\right]^{2/\alpha-1} \to \infty.$$

(C3c) $\sup_n(n\max_i N_i w_i) \le B < \infty$.

A.4. Assumptions for covariance estimation.

- Asymptotic normality:

(D1a) $h_\gamma \to 0$, $\sum_{i=1}^n N_i(N_i-1)v_i^2/h_\gamma^2 \to 0$, $\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2/h_\gamma \to 0$, $\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2 \to 0$.

(D2a) $\max\{\sum_{i=1}^n N_i(N_i-1)v_i^3/h_\gamma^4, \sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^3/h_\gamma^3, \sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^3/h_\gamma^2, \sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)(N_i-4)v_i^3/h_\gamma, \sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)(N_i-4)(N_i-5)v_i^3\}/[\sum_{i=1}^n N_i(N_i-1)v_i^2/h_\gamma^2 + \sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2/h_\gamma + \sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2]^{3/2} \to 0$.

(D3a) $\sup_{t\in[0,1]} E|U(t)|^6 < \infty$ and $E|e|^6 < \infty$.

- L^2 convergence:

(D1b) $h_\gamma \to 0$, $\log\{1/[\sum_{i=1}^n N_i(N_i-1)v_i^2 \max_i(N_i(N_i-1))]\}\times\max_i(N_i(N_i-1))\sum_{i=1}^n N_i(N_i-1)v_i^2/h_\gamma^2 \to 0$, $\log\{1/[\sum_{i=1}^n N_i(N_i-1)v_i^2\times\max_i(N_i(N_i-1))]\}\max_i(N_i(N_i-1))\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2/h_\gamma \to 0$, $\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2 \to 0$.

(D2b) $\max_i[N_i(N_i-1)v_i^2]\log\{1/[\sum_{i=1}^n N_i(N_i-1)v_i^2 \max_i(N_i(N_i-1))]\}/[h_\gamma^2 \sum_{i=1}^n N_i(N_i-1)v_i^2 + h_\gamma \sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2] \to 0$.

(D3b) $\sup_{t\in[0,1]} E|U(t)|^4 < \infty$ and $E|e|^4 < \infty$.

- Uniform convergence:

(D1c) $h_\gamma \to 0$, $\log(n)\sum_{i=1}^n N_i(N_i-1)v_i^2/h_\gamma^2 \to 0$, $\log(n)\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2/h_\gamma \to 0$, $\log(n)\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2 \to 0$.

(D2c) For some $\beta > 2$, $E\sup_{t\in[0,1]}|U(t)|^{2\beta} < \infty$, $E|e|^{2\beta} < \infty$, and

$$n\Bigg[\sum_{i=1}^n N_i(N_i-1)v_i^2h_\gamma^2 + \sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2h_\gamma^3 + \sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2h_\gamma^4\Bigg]\bigg[\frac{\log(n)}{n}\bigg]^{2/\beta-1} \to \infty.$$

(D3c) $\sup_n(n\max_i N_i(N_i-1)v_i) \le B' < \infty$.

Assumptions (A1) and (A2) are standard in the context of smoothing but (A2) could be replaced by another assumption. For instance, instead of (A2), Li and Hsing (2010) assumed that K is of bounded variation. The slightly stronger assumption (A2) just makes the proof of uniform convergence simpler. Assumptions (B1)–(B4) are typical in the context of FDA and local polynomial smoothing, which are assumed in all asymptotic results.

Assumptions (C1a), (C1b), (C1c), (D1a), (D1b) and (D1c) are used to guarantee the consistency of the estimators. Their counterparts for the OBS and SUBJ schemes are standard in local polynomial smoothing. Assumptions (C2a) and (D2a) are used to check the Lyapunov condition for asymptotic normality, and (C2b) and (D2b) are useful to prove Lemmas 2 and 4, respectively, for L^2 convergence. Similar versions of (C2c) and (D2c) were adopted by Li and Hsing (2010). For the OBS estimators, (C2b) and (D2b) are implied by (C1b) and (D1b), respectively. For the SUBJ estimators, (C3c) and (D3c) are automatically satisfied.

APPENDIX B: ASYMPTOTIC NORMALITY

We only give the proofs of Theorems 3.1 and 3.2, and Corollary 3.3. The proofs of Corollaries 3.1, 3.2, 3.4 and 3.5 are straightforward and thus omitted.

B.1. Mean function.

PROOF OF THEOREM 3.1. Easy calculation results in

$$\hat{\mu}(t) = \frac{R_0S_2 - R_1S_1}{S_0S_2 - S_1^2}, \tag{B.1}$$

where for $r = 0, 1, 2$,

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

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

By (B.1)

$$\hat{\mu}(t) = \frac{R_0}{\tilde{S}_0} - \frac{\tilde{S}_1}{\tilde{S}_0}\hat{\mu}^{(1)}(t) + t\hat{\mu}^{(1)}(t) \qquad \text{where } \hat{\mu}^{(1)}(t) = \frac{1}{h_\mu}\frac{R_1 - R_0 S_1/S_0}{S_2 - S_1^2/S_0}$$

and

$$\tilde{S}_r = \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K_{h_\mu}(T_{ij} - t) T_{ij}^r, \qquad r = 0, 1.$$

We first prove that

$$\left[\min\left\{\frac{h_\mu}{\sum_{i=1}^n N_i w_i^2}, \frac{1}{\sum_{i=1}^n N_i (N_i - 1) w_i^2}\right\}\right]^{1/2} \big(\hat{\mu}(t) - \tilde{\mu}(t)\big) = o_p(1), \tag{B.2}$$

$$\text{where } \tilde{\mu}(t) = \frac{R_0}{\tilde{S}_0} - \frac{\tilde{S}_1}{\tilde{S}_0}\mu^{(1)}(t) + t\mu^{(1)}(t),$$

and $\mu^{(1)}(\cdot)$ is the first derivative of $\mu(\cdot)$. To see it, easy calculation shows that

$$\hat{\mu} - \tilde{\mu} = -\frac{S_1}{S_0}\frac{S_0(R_1 - \mu S_1 - h_\mu \mu^{(1)} S_2) - S_1(R_0 - \mu S_0 - h_\mu \mu^{(1)} S_1)}{S_0 S_2 - S_1^2}.$$

It is straightforward to show that both S_0 and $S_0 S_2 - S_1^2$ are positive and bounded away from 0 with probability tending to one, $S_1 = O_p[h_\mu + (\sum_{i=1}^n N_i w_i^2 / h_\mu)^{1/2}]$, and $R_1 - \mu S_1 - h_\mu \mu^{(1)} S_2$, $R_0 - \mu S_0 - h_\mu \mu^{(1)} S_1 = O_p[h_\mu^2 + (\sum_{i=1}^n N_i w_i^2 / h_\mu)^{1/2} + (\sum_{i=1}^n N_i (N_i - 1) w_i^2)^{1/2}]$. Hence, (B.2) holds and it suffices to show the asymptotic normality of $\tilde{\mu}(t)$.

In analogy to the proof of Theorem 4.1 in Bhattacharya and Müller (1993) for independent data and by the Cramér–Wold device and Lyapunov condition due to (C2a), we can achieve the asymptotic joint normality of $(R_0 - ER_0, \tilde{S}_1 - E\tilde{S}_1, \tilde{S}_0 - E\tilde{S}_0)$, where the rate of convergence is $[\min\{h_\mu / \sum_{i=1}^n N_i w_i^2, 1/\sum_{i=1}^n N_i (N_i - 1) w_i^2\}]^{1/2}$. Explicitly for $r, r' = 0, 1$,

$$E\tilde{S}_r = t^r f(t) + \frac{h_\mu^2}{2}\sigma_K^2 [2r f^{(1)}(t) + t^r f^{(2)}(t)] + o(h_\mu^2);$$

$$
\begin{aligned}
ER_0 &= \mu(t)f(t) + \frac{h_\mu^2}{2}\sigma_K^2\big[\mu(t)f^{(2)}(t) + 2\mu^{(1)}(t)f^{(1)}(t) + \mu^{(2)}(t)f(t)\big] \\
&\quad + o(h_\mu^2); \\
\operatorname{Cov}(\tilde S_r, \tilde S_{r'}) &= \frac{\sum_{i=1}^n N_i w_i^2}{h_\mu}\|K\|^2 t^{r+r'} f(t) + o\biggl(\frac{\sum_{i=1}^n N_i w_i^2}{h_\mu}\biggr); \\
\operatorname{Var}(R_0) &= \frac{\sum_{i=1}^n N_i w_i^2}{h_\mu}\|K\|^2\big(\mu(t)^2 + \gamma(t,t) + \sigma^2\big)f(t) \\
&\quad + \biggl(\sum_{i=1}^n N_i(N_i-1)w_i^2\biggr)\gamma(t,t)f(t)^2 \\
&\quad + o\biggl(\frac{\sum_{i=1}^n N_i w_i^2}{h_\mu}\biggr) + o\biggl(\sum_{i=1}^n N_i(N_i-1)w_i^2\biggr); \\
\operatorname{Cov}(R_0, \tilde S_r) &= \frac{\sum_{i=1}^n N_i w_i^2}{h_\mu}\|K\|^2 t^r \mu(t) f(t) + o\biggl(\frac{\sum_{i=1}^n N_i w_i^2}{h_\mu}\biggr).
\end{aligned}
$$

From here, the asymptotic normality of $\tilde\mu(t)$ follows from the delta method. □

PROOF OF COROLLARY 3.3. By Cauchy–Schwarz inequality, $\bar N \ge \bar N_H$ so $\Gamma^A_{\mathrm{obs}}(t) \le \Gamma^A_{\mathrm{subj}}(t)$. Since $\{N_i^2\}_{i=1}^n$ and $\{1 - N_i^{-1}\}_{i=1}^n$ can be concomitantly rearranged to be monotone, Chebyshev's sum inequality implies:

$$
\begin{aligned}
\frac{1}{n}\sum_i N_i^2\biggl[1 - \frac{1}{N_i}\biggr] &\ge \biggl(\frac{1}{n}\sum_i N_i^2\biggr)\biggl(\frac{1}{n}\sum_i\biggl[1 - \frac{1}{N_i}\biggr]\biggr) \\
&\ge \biggl(\frac{1}{n}\sum_i N_i\biggr)^2\biggl(\frac{1}{n}\sum_i\biggl[1 - \frac{1}{N_i}\biggr]\biggr),
\end{aligned}
$$

where the last inequality follows from Cauchy–Schwarz inequality. Therefore,

$$
\frac{(\bar N_{S2} - \bar N)}{(\bar N)^2} \ge 1 - \frac{1}{\bar N_H} \quad \text{and} \quad \Gamma^B_{\mathrm{obs}}(t) \ge \Gamma^B_{\mathrm{subj}}(t). \qquad \square
$$

B.2. Covariance function.

PROOF OF THEOREM 3.2. We can obtain

$$
\begin{aligned}
&\hat\gamma(s,t) \\
&\quad = \frac{(S_{20}S_{02} - S_{11}^2)R_{00} - (S_{10}S_{02} - S_{01}S_{11})R_{10} + (S_{10}S_{11} - S_{01}S_{20})R_{01}}{(S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}},
\end{aligned} \tag{B.3}
$$

where for $p, q = 0, 1, 2$,

$$S_{p,q} = \sum_{i=1}^{n} v_i \sum_{1\le j\ne l\le N_i} K_{h_\gamma}(T_{ij}-s)K_{h_\gamma}(T_{il}-t)\left(\frac{T_{ij}-s}{h_\gamma}\right)^p\left(\frac{T_{il}-t}{h_\gamma}\right)^q,$$

$$R_{p,q} = \sum_{i=1}^{n} v_i \sum_{1\le j\ne l\le N_i} K_{h_\gamma}(T_{ij}-s)K_{h_\gamma}(T_{il}-t)\left(\frac{T_{ij}-s}{h_\gamma}\right)^p\left(\frac{T_{il}-t}{h_\gamma}\right)^q C_{ijl}.$$

By (B.3), $\hat{\gamma}(s,t) = (R_{00} - \hat{\beta}_1 S_{10} - \hat{\beta}_2 S_{01})/S_{00}$. We now consider $\tilde{\gamma}(s,t)$ where

$$\tilde{\gamma}(s,t) = \left(\left(\tilde{R}_{00} - \frac{\partial\gamma}{\partial s}(s,t)\tilde{S}_{10} - \frac{\partial\gamma}{\partial t}(s,t)\tilde{S}_{01}\right)\Big/\tilde{S}_{00}\right) + s\frac{\partial\gamma}{\partial s}(s,t) + t\frac{\partial\gamma}{\partial t}(s,t),$$

$$\tilde{R}_{00} = \sum_{i=1}^{n} v_i \sum_{1\le j\ne l\le N_i} K_{h_\gamma}(T_{ij}-s)K_{h_\gamma}(T_{il}-t)\delta_{ij}\delta_{il},$$

$$\tilde{S}_{p,q} = \sum_{i=1}^{n} v_i \sum_{1\le j\ne l\le N_i} K_{h_\gamma}(T_{ij}-s)K_{h_\gamma}(T_{il}-t)T_{ij}^p T_{il}^q, \qquad 0\le p+q\le 1.$$

Note that $\hat{\mu} = \mu$, so $C_{ijl} = \delta_{ij}\delta_{il}$ and $\tilde{R}_{00} = R_{00}$. Following similar arguments in the proof of Theorem 3.1, we can prove that $[\min\{h_\gamma^2/\sum_{i=1}^n N_i(N_i-1)v_i^2, h_\gamma/\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2, 1/\sum_{i=1}^n N_i(N_i-1)(N_i-2)\times(N_i-3)v_i^2\}]^{1/2}(\hat{\gamma}(s,t)-\tilde{\gamma}(s,t)) = o_p(1)$. It thus suffices to prove the asymptotic normality of $\tilde{\gamma}$.

By (D2a), we can establish the asymptotic joint normality of $(\tilde{R}_{00} - E\tilde{R}_{00}, \tilde{S}_{00} - E\tilde{S}_{00}, \tilde{S}_{10} - E\tilde{S}_{10}, \tilde{S}_{01} - E\tilde{S}_{01})$ through the Cramér–Wold device and Lyapunov condition. By easy calculation, we have

$$\begin{aligned} E\tilde{S}_{p,q} &= s^p t^q f(s)f(t) + \frac{h_\gamma^2}{2}\sigma_K^2\big[s^p t^q f^{(2)}(s)f(t) + s^p t^q f(s)f^{(2)}(t)\big] \\ &\quad + \frac{h_\gamma^2}{2}\sigma_K^2\big[2pf^{(1)}(s)f(t) + 2qf^{(1)}(t)f(s)\big] + o(h_\gamma^2), \\ E\tilde{R}_{00} &= \gamma(s,t)f(s)f(t) \\ &\quad + \frac{h_\gamma^2}{2}\sigma_K^2\bigg[\frac{\partial^2 G}{\partial s^2}(s,t)f(s)f(t) \\ &\quad + 2\frac{\partial\gamma}{\partial s}(s,t)f^{(1)}(s)f(t) + \gamma(s,t)f^{(2)}(s)f(t) \\ &\quad + \frac{\partial^2\gamma}{\partial t^2}(s,t)f(t)f(s) \\ &\quad + 2\frac{\partial\gamma}{\partial t}(s,t)f^{(1)}(t)f(s) + \gamma(s,t)f^{(2)}(t)f(s)\bigg] + o(h_\gamma^2). \end{aligned}$$

To express the asymptotic variances concisely, we introduce the following terms in addition to $V_1(s,t)$, $V_2(s,t)$ and $V_3(s,t)$ in Section 3.2:

$$E_1(s,t) = E\big[(Y_1 - \mu(T_1))^2 (Y_2 - \mu(T_2))^2 | T_1 = s, T_2 = t\big];$$

$$E_2(s,t) = E\big[(Y_1 - \mu(T_1))^2 (Y_2 - \mu(T_2))(Y_3 - \mu(T_3))| \\ T_1 = s, T_2 = t, T_3 = t\big].$$

Computing the asymptotic variances involves $E[K_{h_\gamma}(T-t)K_{h_\gamma}(T-s)]$, which equals $\|K\|^2 f(t)/h_\gamma + o(1/h_\gamma)$ for $s = t$ and 0 for $s \neq t$ as $h_\gamma \to 0$, since $K(\cdot) = 0$ outside $[-1, 1]$. Therefore the asymptotic variances are different for $s = t$ and $s \neq t$. For $0 \le p + q \le 1$ and $0 \le p' + q' \le 1$,

$$\begin{aligned}
\operatorname{Var}(\tilde{R}_{00}) = \big[1 + I(s=t)\big]\Bigg[& \frac{\sum_{i=1}^n N_i(N_i-1)v_i^2}{h_\gamma^2} \|K\|^4 E_1(s,t) f(s) f(t) \\
& + \frac{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2}{h_\gamma} \\
& \times \|K\|^2 \big[f^2(s) f(t) E_2(t,s) + f(s) f^2(t) E_2(s,t)\big]\Bigg] \\
& + \left[\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2\right] \\
& \times V_3(s,t) f^2(s) f^2(t) + o\left(\frac{\sum_{i=1}^n N_i(N_i-1)v_i^2}{h_\gamma^2}\right) \\
& + o\left(\frac{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2}{h_\gamma}\right) \\
& + o\left(\sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2\right);
\end{aligned}$$

$$\begin{aligned}
& \operatorname{Cov}(\tilde{S}_{p,q}, \tilde{S}_{p',q'}) \\
& \quad = \big[1 + I(s=t)\big] s^{p+p'} t^{q+q'} \Bigg[\frac{\sum_{i=1}^n N_i(N_i-1)v_i^2}{h_\gamma^2} \|K\|^4 f(s) f(t) \\
& \quad + \frac{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2}{h_\gamma} \|K\|^2 \big[f^2(s) f(t) + f(s) f^2(t)\big]\Bigg] \\
& \quad + o\left(\frac{\sum_{i=1}^n N_i(N_i-1)v_i^2}{h_\gamma^2}\right) + o\left(\frac{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2}{h_\gamma}\right);
\end{aligned}$$

$$
\begin{aligned}
&\mathrm{Cov}(\tilde{S}_{p,q}, \tilde{R}_{00}) \\
&\quad = [1 + I(s = t)] s^p t^q \biggl[\frac{\sum_{i=1}^n N_i(N_i - 1) v_i^2}{h_\gamma^2} \|K\|^4 \gamma(s,t) f(s) f(t) \\
&\qquad + \frac{\sum_{i=1}^n N_i(N_i - 1)(N_i - 2) v_i^2}{h_\gamma} \|K\|^2 \gamma(s,t) [f^2(s) f(t) + f(s) f^2(t)] \biggr] \\
&\qquad + o\biggl(\frac{\sum_{i=1}^n N_i(N_i - 1) v_i^2}{h_\gamma^2} \biggr) + o\biggl(\frac{\sum_{i=1}^n N_i(N_i - 1)(N_i - 2) v_i^2}{h_\gamma} \biggr).
\end{aligned}
$$

Now apply the delta method to complete the proof. □

APPENDIX C: L^2 CONVERGENCE

We only give the proofs of Theorems 4.1 and 4.2. The proofs of Corollaries 4.1–4.4 are straightforward, and thus omitted.

C.1. Mean function. We first give the rate for $\sum_{i=1}^n w_i \sum_{j=1}^{N_i} K((T_{ij} - t)/h_\mu)\delta_{ij}$.

LEMMA 1. *Under the assumptions for Theorem* 4.1,

$$
\begin{aligned}
&\biggl\| \sum_{i=1}^n w_i \sum_{j=1}^{N_i} K\biggl(\frac{T_{ij} - t}{h_\mu} \biggr) \delta_{ij} \biggr\|_2 \\
&\qquad = O_p\biggl(\biggl[\sum_{i=1}^n N_i w_i^2 h_\mu + \sum_{i=1}^n N_i(N_i - 1) w_i^2 h_\mu^2 \biggr]^{1/2} \biggr).
\end{aligned}
$$

PROOF. Denote $a_n = (\sum_{i=1}^n N_i w_i^2 h_\mu + \sum_{i=1}^n N_i(N_i - 1) w_i^2 h_\mu^2)^{1/2}$. For any $M > 0$, by Markov inequality,

$$
\begin{aligned}
&P\biggl(\biggl\| \sum_{i=1}^n w_i \sum_{j=1}^{N_i} K\biggl(\frac{T_{ij} - t}{h_\mu} \biggr) \delta_{ij} \biggr\|_2 > M a_n \biggr) \\
&\qquad = P\biggl(\int \biggl[\sum_{i=1}^n w_i \sum_{j=1}^{N_i} K\biggl(\frac{T_{ij} - t}{h_\mu} \biggr) \delta_{ij} \biggr]^2 dt > M^2 a_n^2 \biggr) \\
&\qquad \le \frac{E \int [\sum_{i=1}^n w_i \sum_{j=1}^{N_i} K((T_{ij} - t)/h_\mu)\delta_{ij}]^2 \, dt}{M^2 a_n^2}.
\end{aligned}
$$

Note that

$$E\int\left[\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K\left(\frac{T_{ij}-t}{h_\mu}\right)\delta_{ij}\right]^2 dt$$
$$= \sum_{i=1}^{n} w_i^2 \left\{ N_i \int E\left[K^2\left(\frac{T-t}{h_\mu}\right)\left(\gamma(T,T)+\sigma^2\right)\right] dt \right.$$
$$\left. + (N_i^2 - N_i)\int EK\left(\frac{T_1-t}{h_\mu}\right)K\left(\frac{T_2-t}{h_\mu}\right)\gamma(T_1,T_2)\,dt\right\}$$
$$\le M'\left(\sum_{i=1}^{n} N_i w_i^2 h_\mu + \sum_{i=1}^{n} N_i(N_i-1) w_i^2 h_\mu^2\right) = M' a_n^2,$$

for some constant M'. Therefore, as $M \to \infty$,

$$P\left(\left\|\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K\left(\frac{T_{ij}-t}{h_\mu}\right)\delta_{ij}\right\|_2 > M a_n\right) \le \frac{M''}{M^2} \to 0.$$

□

We next give the uniform rate of S_r, $r = 0, 1, 2$ as defined in (B.1).

LEMMA 2. *Under the assumptions for Theorem* 4.1,

$$\sup_{t\in[0,1]} \left|S_r(t) - ES_r(t)\right| = o_p(1).$$

PROOF. The proof is similar to that of Lemma 1 in Zhang, Park and Wang (2013) and details are provided in the supplementary material [Zhang and Wang (2016)]. □

PROOF OF THEOREM 4.1. From (B.1) we have,

$$\hat{\mu}(t) - \mu(t)$$
$$= \frac{(R_0 - \mu(t)S_0 - h_\mu \mu^{(1)}(t)S_1)S_2 - (R_1 - \mu(t)S_1 - h_\mu \mu^{(1)}(t)S_2)S_1}{S_0 S_2 - S_1^2}. \tag{C.1}$$

By Taylor expansion and Lemma 1,

$$\left\|R_0 - \mu(t)S_0 - h_\mu \mu^{(1)}(t)S_1\right\|_2$$
$$= \left\|\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K_{h_\mu}(T_{ij}-t)\right.$$
$$\left. \times \left(\delta_{ij} + (T_{ij}-t)^2 \int_0^1 \mu^{(2)}\left(t + v(T_{ij}-t)\right)(1-v)\,dv\right)\right\|_2$$

$$\leq \left\| \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K_{h_\mu}(T_{ij} - t)\delta_{ij} \right\|_2 + O_p(h_\mu^2)$$

$$= O_p\left(h_\mu^2 + \left[\frac{\sum_{i=1}^{n} N_i w_i^2}{h_\mu} + \sum_{i=1}^{n} N_i(N_i - 1)w_i^2\right]^{1/2}\right).$$

Similarly,

$$\|R_1 - \mu(t)S_1 - h_\mu \mu^{(1)}(t)S_2\|_2$$

$$= O_p\left(h_\mu^2 + \left[\frac{\sum_{i=1}^{n} N_i w_i^2}{h_\mu} + \sum_{i=1}^{n} N_i(N_i - 1)w_i^2\right]^{1/2}\right).$$

Lemma 2 implies that $\sup_{t\in[0,1]} |S_1(t)|, \sup_{t\in[0,1]} |S_2(t)| = O_p(1)$ and $S_0S_2 - S_1^2$ is positive and bounded away from 0 on [0, 1] with probability approaching one. The proof is complete by (C.1). □

C.2. Covariance function. Define

$$W(s,t) = \sum_{i=1}^{n} v_i \sum_{1\leq j\neq l\leq N_i} K\left(\frac{T_{ij} - s}{h_\gamma}\right)K\left(\frac{T_{il} - t}{h_\gamma}\right)\delta_{ij}\delta_{il}.$$

The Hilbert–Schmidt rate of convergence for $W(s,t)$ is given below.

LEMMA 3. *Under the assumptions for Theorem 4.2,*

$$\|W\|_{\mathrm{HS}} = O_p\left(\left[\sum_{i=1}^{n} N_i(N_i - 1)v_i^2h_\gamma^2 + \sum_{i=1}^{n} N_i(N_i - 1)(N_i - 2)v_i^2h_\gamma^3 + \sum_{i=1}^{n} N_i(N_i - 1)(N_i - 2)(N_i - 3)v_i^2h_\gamma^4\right]^{1/2}\right).$$

PROOF. The proof is very similar to that of Lemma 1. See details in the supplementary material [Zhang and Wang (2016)]. □

For $S_{p,q}(s,t), 0 \leq p + q \leq 2$ defined in (B.3), their uniform rates are given below.

LEMMA 4. *Under the assumptions for Theorem 4.2,*

$$\sup_{s,t\in[0,1]} |S_{p,q}(s,t) - ES_{p,q}(s,t)| = o_p(1).$$

PROOF. See the proof in the supplementary material [Zhang and Wang (2016)], which is similar to that of Lemma 2. □

PROOF OF THEOREM 4.2. We have the decomposition below by (B.3)

$$
\begin{aligned}
&\hat{\gamma}(s,t) - \gamma(s,t) \\
&\quad = (S_{20}S_{02} - S_{11}^2)\left[R_{00} - \gamma(s,t)S_{00} - h_\gamma \frac{\partial G}{\partial s}(s,t)S_{10} - h_\gamma \frac{\partial \gamma}{\partial t}(s,t)S_{01}\right] \\
&\qquad \Big/\big((S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}\big) \\
&\qquad - (S_{10}S_{02} - S_{01}S_{11}) \\
&\qquad \times \left[R_{10} - \gamma(s,t)S_{10} - h_\gamma \frac{\partial G}{\partial s}(s,t)S_{20} - h_\gamma \frac{\partial \gamma}{\partial t}(s,t)S_{11}\right] \qquad \text{(C.2)} \\
&\qquad \Big/\big((S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}\big) \\
&\quad + (S_{10}S_{11} - S_{01}S_{20})\left[R_{01} - \gamma(s,t)S_{01} - h_\gamma \frac{\partial G}{\partial s}(s,t)S_{11} - h_\gamma \frac{\partial \gamma}{\partial t}(s,t)S_{02}\right] \\
&\qquad \Big/\big((S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}\big).
\end{aligned}
$$

Consequently, by Lemma 3 and Taylor expansion,

$$
\begin{aligned}
&\left\| R_{00} - \gamma(s,t)S_{00} - h_\gamma \frac{\partial \gamma}{\partial s}(s,t)S_{10} - h_\gamma \frac{\partial \gamma}{\partial t}(s,t)S_{01} \right\|_{\mathrm{HS}} \\
&\quad = O_p\Bigg(h_\gamma^2 + \Bigg[\frac{\sum_{i=1}^n N_i(N_i-1)v_i^2}{h_\gamma^2} + \frac{\sum_{i=1}^n N_i(N_i-1)(N_i-2)v_i^2}{h_\gamma} \\
&\qquad + \sum_{i=1}^n N_i(N_i-1)(N_i-2)(N_i-3)v_i^2\Bigg]^{1/2}\Bigg).
\end{aligned}
$$

Similarly, $R_{10} - \gamma(s,t)S_{10} - h_\gamma \frac{\partial G}{\partial s}(s,t)S_{20} - h_\gamma \frac{\partial \gamma}{\partial t}(s,t)S_{11}$ and $R_{01} - \gamma(s,t) \times S_{01} - h_\gamma \frac{\partial G}{\partial s}(s,t)S_{11} - h_\gamma \frac{\partial \gamma}{\partial t}(s,t)S_{02}$ both have this rate. Moreover, by Lemma 4, each denominator in (C.2) is positive and bounded away from 0 on $[0,1]^2$ with probability approaching one and that $\sup_{s,t\in[0,1]}|S_{20}S_{02} - S_{11}^2|$, $\sup_{s,t\in[0,1]}|S_{10}S_{02} - S_{01}S_{11}|$, $\sup_{s,t\in[0,1]}|S_{10}S_{02} - S_{01}S_{11}| = O_p(1)$. Thus, the proof is complete by (C.2). □

APPENDIX D: UNIFORM CONVERGENCE

We only give the proofs of Theorems 5.1 and 5.2. The proofs of Corollaries 5.1–5.4 are straightforward, and thus omitted.

D.1. Mean function. We first show the rate of convergence for

$$L(t) = \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K\left(\frac{T_{ij} - t}{h_\mu}\right) U_{ij}^+,$$

where U_{ij}^+ is the positive part of U_{ij}.

LEMMA 5. *Under the assumptions for Theorem 5.1,*

$$\sup_{t\in[0,1]} \left|L(t) - EL(t)\right| = O\left(\left\{\log(n)\left[\sum_{i=1}^{n} N_i w_i^2 h_\mu + \sum_{i=1}^{n} N_i(N_i - 1) w_i^2 h_\mu^2\right]\right\}^{1/2}\right) \qquad a.s.$$

PROOF. Denote $a_n = \{\log(n)[\sum_{i=1}^{n} N_i w_i^2 h_\mu + \sum_{i=1}^{n} N_i(N_i - 1) w_i^2 h_\mu^2]\}^{1/2}$ and $A_n = a_n[n/\log(n)]$. By (C1c), we can find a constant $\gamma > 0$ such that $n^\gamma h_\mu a_n \to \infty$. Let $\chi(\gamma)$ be an equidistant partition on [0, 1] with grid length $n^{-\gamma}$. Therefore,

$$\sup_{t\in[0,1]} \left|L(t) - EL(t)\right| \le \sup_{t\in\chi(\gamma)} \left|L(t) - EL(t)\right| + D_1 + D_2$$

where

$$D_1 = \sup_{t,s\in[0,1];|t-s|\le n^{-\gamma}} \left|L(t) - L(s)\right|,$$

$$D_2 = \sup_{t,s\in[0,1];|t-s|\le n^{-\gamma}} \left|EL(t) - EL(s)\right|.$$

By (A2) and Hölder inequality,

$$\begin{aligned}
D_1 &\le \sup_{t,s\in[0,1];|t-s|\le n^{-\gamma}} \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} U_{ij}^+ \left|K\left(\frac{T_{ij} - t}{h_\mu}\right) - K\left(\frac{T_{ij} - s}{h_\mu}\right)\right| \\
&\le \left(\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} U_{ij}^+\right) L n^{-\gamma}/h_\mu \\
&\le \left(\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} (U_{ij}^+)^\alpha\right)^{1/\alpha} L n^{-\gamma}/h_\mu \\
&\le \left(\sum_{i=1}^{n} N_i w_i \sup_{t\in[0,1]} |U_i(t)|^\alpha\right)^{1/\alpha} L n^{-\gamma}/h_\mu.
\end{aligned}$$

By (C2c), (C3c) and the strong law of large numbers,

$$\begin{aligned}\sum_{i=1}^{n} N_i w_i \sup_{t\in[0,1]} \big|U_i(t)\big|^{\alpha} &\le \Big(n \max_i N_i w_i\Big)\cdot \frac{1}{n}\sum_{i=1}^{n} \sup_{t\in[0,1]} \big|U_i(t)\big|^{\alpha} \\ &\le B\cdot \frac{1}{n}\sum_{i=1}^{n} \sup_{t\in[0,1]} \big|U_i(t)\big|^{\alpha} \\ &\to B\cdot E \sup_{t\in[0,1]} \big|U(t)\big|^{\alpha} < \infty, \qquad \text{a.s.}\end{aligned}$$

By $n^{\gamma} h_{\mu} a_n \to \infty$, $D_1, D_2 = o(a_n)$, a.s.

Let the truncated $L(t)$ be $L(t)^* = \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K((T_{ij}-t)/h_{\mu}) U_{ij}^{+} I(U_{ij}^{+} \le A_n)$, where $I(\cdot)$ is the indicator function. Then

$$\sup_{t\in\chi(\gamma)} \big|L(t) - EL(t)\big| \le \sup_{t\in\chi(\gamma)} \big|L(t)^* - EL(t)^*\big| + E_1 + E_2,$$

where

$$E_1 = \sup_{t\in\chi(\gamma)} \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K\big((T_{ij}-t)/h_{\mu}\big) U_{ij}^{+} I\big(U_{ij}^{+} > A_n\big),$$

$$E_2 = \sup_{t\in\chi(\gamma)} \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} E\big[K\big((T_{ij}-t)/h_{\mu}\big) U_{ij}^{+} I\big(U_{ij}^{+} > A_n\big)\big].$$

We have $E_1, E_2 = o(a_n)$, a.s. since by (A2), (C2c) and (C3c), for every $t \in \chi(\gamma)$,

$$\begin{aligned}&\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} K\big((T_{ij}-t)/h_{\mu}\big) U_{ij}^{+} I\big(U_{ij}^{+} > A_n\big) \\ &\quad\le M_K A_n^{1-\alpha} \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} |U_{ij}|^{\alpha} \le B M_K A_n^{1-\alpha} \Bigg(n^{-1} \sum_{i=1}^{n} \sup_{t\in[0,1]} \big|U_i(t)\big|^{\alpha}\Bigg).\end{aligned}$$

We can rewrite $L(t)^* - EL(t)^* = \sum_{i=1}^{n} (V_i - EV_i)$ where $V_i = w_i \times \sum_{j=1}^{N_i} K((T_{ij}-t)/h_{\mu}) U_{ij}^{+} I(U_{ij}^{+} \le A_n)$. By (C3c), $|V_i - EV_i| \le 2N_i w_i M_K A_n \le 2BM_K A_n/n$. Additionally,

$$\begin{aligned}E(V_i - EV_i)^2 \le EV_i^2 &\le N_i w_i^2 E\bigg[K\bigg(\frac{T-t}{h_{\mu}}\bigg) U^{+}(T)\bigg]^2 + N_i(N_i-1) w_i^2 \\ &\quad\times E\bigg[K\bigg(\frac{T_1-t}{h_{\mu}}\bigg) K\bigg(\frac{T_2-t}{h_{\mu}}\bigg) U^{+}(T_1) U^{+}(T_2)\bigg] \\ &\le M_U\big[N_i w_i^2 h_{\mu} + N_i(N_i-1) w_i^2 h_{\mu}^2\big]\end{aligned}$$

for some constant $M_U > 0$. By Bernstein inequality, for any $M > 0$,

$$
\begin{aligned}
&P\Big(\sup_{t\in\chi(\gamma)} \big|L(t)^* - EL(t)^*\big| > Ma_n\Big)\\
&\quad \le n^\gamma P\left(\left|\sum_{i=1}^n (V_i - EV_i)\right| > Ma_n\right)\\
&\quad \le 2n^\gamma \exp\left(-\frac{M^2a_n^2/2}{\sum_{i=1}^n M_U[N_i w_i^2 h_\mu + N_i(N_i-1)w_i^2h_\mu^2] + 2BM_K A_n Ma_n/3n}\right)\\
&\quad = 2n^{\gamma - M^*},
\end{aligned}
$$

where $M^* = M^2/(2M_U + 4BM_K M/3)$. By Borel–Cantelli's lemma,

$$
\sup_{t\in\chi(\gamma)} \big|L(t)^* - EL(t)^*\big| = O(a_n) \qquad \text{a.s.}
$$

and the proof is complete. □

PROOF OF THEOREM 5.1. By Taylor expansion and the fact that $K_{h_\mu}(T_{ij} - t) = 0$ if $|T_{ij} - t| > h_\mu$,

$$
\begin{aligned}
&R_0 - \mu(t)S_0 - h_\mu\mu^{(1)}(t)S_1\\
&\quad = \sum_{i=1}^n w_i \sum_{j=1}^{N_i} K_{h_\mu}(T_{ij} - t)\big(\delta_{ij} + \mu(T_{ij}) - \mu(t) - \mu^{(1)}(t)(T_{ij} - t)\big) \qquad \text{(D.1)}\\
&\quad = \sum_{i=1}^n w_i \sum_{j=1}^{N_i} K_{h_\mu}(T_{ij} - t)\delta_{ij} + O\big(h_\mu^2\big).
\end{aligned}
$$

By Lemma 5, $\sup_{t\in[0,1]} |\sum_{i=1}^n w_i \sum_{j=1}^{N_i} K_{h_\mu}(T_{ij} - t)\delta_{ij}| = O(a_n/h_\mu)$, a.s. so $\sup_{t\in[0,1]} |R_0 - \mu(t)S_0 - h_\mu\mu^{(1)}(t)S_1| = O(h_\mu^2 + a_n/h_\mu)$, a.s. Similarly, $\sup_{t\in[0,1]} |R_1 - \mu(t)S_1 - h_\mu\mu^{(1)}(t)S_2| = O(h_\mu^2 + a_n/h_\mu)$, a.s.

For $t \in [h_\mu, 1 - h_\mu]$, $ES_0 = f(t) + O(h_\mu^2)$, $ES_1 = O(h_\mu)$ and $ES_2 = f(t)\sigma_K^2 + O(h_\mu)$. For $t \in [0, h_\mu]$, $ES_0 = f(t)\int_{[-t/h_\mu,1]} K(u)\,du + O(h_\mu)$, $ES_1 = f(t)\int_{[-t/h_\mu,1]} uK(u)\,du + O(h_\mu)$ and $ES_2 = f(t)\int_{-t/h_\mu}^1 u^2K(u)\,du + O(h_\mu)$. For $t \in [1 - h_\mu, 1]$, $ES_0 = f(t)\int_{[-1,1-h_\mu]} K(u)\,du + O(h_\mu)$, $ES_1 = f(t)\int_{[-1,1-h_\mu]} uK(u)\,du + O(h_\mu)$ and $ES_2 = f(t)\int_{[-1,1-h_\mu]} u^2K(u)\,du + O(h_\mu)$. Therefore, $S_0S_2 - S_1^2$ is positive and bounded away from 0 on [0, 1] a.s. and S_1 and S_2 are both bounded on [0, 1] a.s. The proof is thus complete by (C.1). □

D.2. Covariance function.

PROOF OF THEOREM 5.2. The proof of Theorem 5.1 can be easily extended to the two-dimensional case for covariance. See details in the supplementary material [Zhang and Wang (2016)]. □

Acknowledgements. The authors are grateful for the Editor, Associate Editor and two referees for insightful comments. The authors also thank Yang Zhou, a Ph.D. candidate in Statistics at University of California, Davis, for careful reading and valuable suggestions.

SUPPLEMENTARY MATERIAL

Supplement to "From sparse to dense functional data and beyond" (DOI: 10.1214/16-AOS1446SUPP; .pdf). In the supplementary material, we provide the proofs of Lemmas 2–4 and the detailed proof of Theorem 5.2.

REFERENCES

ANEIROS, G. and VIEU, P. (2014). Variable selection in infinite-dimensional problems. *Statist. Probab. Lett.* **94** 12–20. MR3257356

BAÍLLO, A. and GRANÉ, A. (2009). Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.* **100** 102–111. MR2460480

BARRIENTOS-MARIN, J., FERRATY, F. and VIEU, P. (2010). Locally modelled regression and functional data. *J. Nonparametr. Stat.* **22** 617–632. MR2682211

BHATTACHARYA, P. K. and MÜLLER, H. (1993). Asymptotics for nonparametric regression. *Sankhyā Ser. A* **55** 420–441. MR1323398

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

BOJ, E., DELICADO, P. and FORTIANA, J. (2010). Distance-based local linear regression for functional predictors. *Comput. Statist. Data Anal.* **54** 429–437. MR2756437

BONGIORNO, E. G., SALINELLI, E., GOIA, A. and VIEU, P. (2014). *Contributions in Infinite-Dimensional Statistics and Related Topics*. Societa Editrice Esculapio, Bologna.

BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications*. *Lecture Notes in Statistics* **149**. Springer, New York. MR1783138

CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496

CAI, T. T. and YUAN, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. Technical report, Univ. Pennsylvania, Philadelphia, PA.

CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355. MR2906870

CARDOT, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonparametr. Stat.* **12** 503–538. MR1785396

CARDOT, H., FERRATY, F. and SARDA, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45** 11–22. MR1718346

CUEVAS, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference* **147** 1–23. MR3151843

DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934

DEGRAS, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21** 1735–1765. MR2895997

DELAIGLE, A. and HALL, P. (2010). Defining probability density for a distribution of random functions. *Ann. Statist.* **38** 1171–1193. MR2604709

FAN, J. and GIJBELS, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20** 2008–2036. MR1193323

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*: *Theory and Practice*. Springer, New York. MR2229687

GEENENS, G. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Stat. Surv.* **5** 30–43. MR3018508

HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269

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. and WANG, J. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365

HART, J. D. and WEHRLY, T. E. (1993). Consistency of cross-validation when the data are curves. *Stochastic Process. Appl.* **45** 351–361. MR1208879

HOOVER, D. R., RICE, J. A., WU, C. O. and YANG, L. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, New York. MR2920735

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. Wiley, Chichester. MR3379106

HUANG, J. Z., WU, C. O. and ZHOU, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89** 111–128. MR1888349

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

JIANG, C. and WANG, J. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *Ann. Statist.* **38** 1194–1226. MR2604710

JIANG, C. and WANG, J. (2011). Functional single index models for longitudinal data. *Ann. Statist.* **39** 362–388. MR2797850

KATO, T. (1980). *Perturbation Theory for Linear Operators*. Springer, Berlin.

KIM, S. and ZHAO, Z. (2013). Unified inference for sparse and dense longitudinal models. *Biometrika* **100** 203–212. MR3034333

LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. MR2766854

LIU, B. and MÜLLER, H. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *J. Amer. Statist. Assoc.* **104** 704–717. MR2541589

MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. MR3258083

MORRIS, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application* **2** 321–359.

PAUL, D. and PENG, J. (2009). Consistency of restricted maximum likelihood estimators of principal components. *Ann. Statist.* **37** 1229–1271. MR2509073

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

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **53** 233–243. MR1094283

RICE, J. A. and WU, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57** 253–259. MR1833314

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720

SHANG, H. L. (2014). A survey of functional principal component analysis. *AStA Adv. Stat. Anal.* **98** 121–142. MR3254025

SHI, M., WEISS, R. E. and TAYLOR, J. M. (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Appl. Stat.* **45** 151–163.

WANG, J., CHIOU, J. and MÜLLER, H. (2015). Functional data analysis. *Annual Review of Statistics and Its Application* **3** 257–295.

WU, C. O. and CHIANG, C. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statist. Sinica* **10** 433–456. MR1769751

WU, H. and ZHANG, J. (2002). Local polynomial mixed-effects models for longitudinal data. *J. Amer. Statist. Assoc.* **97** 883–897. MR1941417

YAO, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J. Multivariate Anal.* **98** 40–56. MR2292916

YAO, F. and LEE, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 3–25. MR2212572

YAO, W. and LI, R. (2013). New local estimation procedure for a non-parametric regression function for longitudinal data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 123–138. MR3008274

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

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

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

ZHANG, X., PARK, B. U. and WANG, J. (2013). Time-varying additive models for longitudinal data. *J. Amer. Statist. Assoc.* **108** 983–998. MR3174678

ZHANG, X. and WANG, J. (2016). Supplement to "From sparse to dense functional data and beyond." DOI:10.1214/16-AOS1446SUPP.

ZHENG, S., YANG, L. and HÄRDLE, W. K. (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *J. Amer. Statist. Assoc.* **109** 661–673. MR3223741

DEPARTMENT OF APPLIED ECONOMICS AND STATISTICS
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19716
USA
E-MAIL: xkzhang@udel.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, DAVIS
DAVIS, CALIFORNIA 95616
USA
E-MAIL: janelwang@ucdavis.edu