Mean and covariance estimation for discretely observed high-dimensional functional data: rates of convergence and division of observational regimes

Alexander Petersen[a,*]

[a]*Department of Statistics*
2152 WVB
Brigham Young University
Provo, UT 84602

Abstract

Estimation of the mean and covariance parameters for functional data is a critical task, with local linear smoothing being a popular choice. In recent years, many scientific domains are producing multivariate functional data for which p, the number of curves per subject, is often much larger than the sample size n. In this setting of high-dimensional functional data, much of developed methodology relies on preliminary estimates of the unknown mean functions and the auto- and cross-covariance functions. This paper investigates the convergence rates of local linear estimators in terms of the maximal error across components and pairs of components for mean and covariance functions, respectively, in both L^2 and uniform metrics. The local linear estimators utilize a generic weighting scheme that can adjust for differing numbers of discrete observations N_{ij} across curves j and subjects i, where the N_{ij} vary with n. Particular attention is given to the equal weight per observation (OBS) and equal weight per subject (SUBJ) weighting schemes. The theoretical results utilize novel applications of concentration inequalities for functional data and demonstrate that, similar to univariate functional data, the order of the N_{ij} relative to p and n divides high-dimensional functional data into three regimes (sparse, dense, and ultra-dense), with the high-dimensional parametric convergence rate of $\{\log(p)/n\}^{1/2}$ being attainable in the latter two.

Keywords: Concentration inequalities, High-dimensional data, L^2 convergence, Local linear smoothing, Uniform Convergence
2020 MSC: Primary 62G20, Secondary 62G08, 62G05

1. Introduction

Over the last three decades, the foundational statistical principles underlying modern functional data analysis have been established [15, 20, 22, 25, 34]. The relevant literature in more recent years suggests that the field is undergoing a transition, compelled by real data applications, in which the traditional univariate setting (one curve per subject) is being replaced with settings of larger and more complex structure, as evidenced by several review papers [1, 2, 10, 17, 37]. The authors of [37] described these new data sets as "next-generation" functional data, including such cases as functions with non-linear constraints [6, 16, 21, 32] or functions that take values in non-Euclidean spaces [12, 13]. Similarly, [26] used the term "second generation" to refer to large functional data sets with complex dependencies, including longitudinal [5, 18, 31] and spatial [11, 19, 41] functional observations, as well as multivariate functional data. Developments for multivariate functional data, in which observations across multiple curves are available for each subject, extend back to early methodological work on dimension reduction [3, 7, 8], clustering [24], and regression models in which functions can appear as responses [9, 45] or predictors [38]. In the same way that classical multivariate data analysis led to modern developments in high-dimensional data analysis, some recent work in multivariate functional data methodology can best be described as high-dimensional functional

*Corresponding author. Email address: petersen@stat.byu.edu

data analysis, in which p, the number of curves for which observations are available per subject, is on the order of, or much larger than, the number of independent subjects n. Examples include interpretable dimension reduction [23] and discriminant analysis [39] for electroencephalography (EEG) data, and functional graphical models for both EEG [27, 33, 35, 36, 43, 46] and functional magnetic resonance imaging (fMRI) data [28, 40].

A common element of the existing work on high-dimensional functional data is the requirement of preliminary estimates of the p mean functions and the $p(p+1)/2$ distinct covariance functions. The theoretical properties of these preliminary estimates determine those of the given method. In terms of theoretical justification of their methodologies, previous works either assume that the curves are fully observed, or that they are observed on a regular, deterministic grid, with the number of observations per curve growing polynomially with the sample size. The former scenario does not reflect the practical realities of functional datasets, while the latter does not cover important cases of designs in which observations can be random, irregular across the domain, heterogeneous across subjects, or any combination of these; moreover, the number of observations per curve per subject may not diverge with n, or may do so at application-specific rates. For instance, [14, 44] illustrate methodologies using biomedical data that are described therein as "high-dimensional longitudinal data." As longitudinal data can be modeled as functional data with relatively few temporal observations compared to the number of independent sampling units, such data constitute an example of sparsely observed high-dimensional functional data for which crucial theory has not yet been developed. It should be emphasized that, while a multivariate functional data set with a small number of curves p, but a large number of observations per curve relative to the sample size n, may accurately be described as high-dimensional, this is not the sense in which this term is used in this paper. Rather, dimensionality refers to the number of curves per subject, and will always be assumed to be high, while the number of discrete observations per curve ranges from sparse to dense, and determines the convergence rate of the estimators.

The contribution of this paper is to derive rates of convergence for nonparametric estimators of the mean and covariance functions for high-dimensional functional data. Specifically, local linear estimators will be studied in terms of their accuracy in both the L^2 (in probability) and uniform (almost surely) metrics. This choice was made due to critical previous work in the setting of univariate or "first-generation" functional data [29, 42], corresponding to a special case of the developments herein when $p = 1$. Both of these works studied local linear estimators of mean and covariance functions in the manner described, the latter in a more general fashion. In particular, [29] applied a specific weighting scheme to the observations in the estimation criterion, whereas [42] considered a general weighting scheme and also demonstrated pointwise asymptotic normality of the estimators. Additionally, as the covariance estimates are slightly different in these two papers, that of [42] will be used in this paper; for a detailed and intuitive justification of this choice, the reader is referred to Section 6 of [42].

As the number of observations from a given curve can vary across subjects and across different components of the multivariate functional data, and since these observations are inherently correlated, it is reasonable to assign different weights for observations coming from distinct subjects or observational units in the data set. Two common choices are the subject (SUBJ) scheme, in which the total weight of all observations from a given subject is the same across subjects, and the observation (OBS) scheme, in which all observations share the same weight. The former choice was used in [29], while [42] studied a generic weighting scheme with specific focus on the SUBJ and OBS schemes. The following two critical insights were provided by [42]. First, if the number of observations for a given component curve is highly heterogeneous across subjects, the OBS scheme can perform poorly due to higher weight being granted to a group of potentially highly correlated observations; on the other hand, if they are homogeneous in an explicit sense, the OBS scheme will yield a rate no worse than the SUBJ scheme. Second, depending on the behavior of the average (for the OBS scheme) or hyperbolic average (for the SUBJ scheme) number of observations per subject relative to n, prescriptive bandwidth choices divide functional data sets into three observational regimes: sparse (or non-dense), dense, and ultra-dense. In the sparse scenario, the rate can be anywhere between the nonparametric (inclusive) rates of $n^{-2/5}$ and $n^{-1/3}$ for mean and covariance estimation, respectively, and the parametric rate of $n^{-1/2}$ (exclusive). For dense and ultra-dense data, the parametric rate is always attainable; the distinguishing feature of ultra-dense data is that the bias decays at a faster rate than the stochastic error, whereas these two are matched for dense data.

The results derived in this paper generalize the above findings to the high-dimensional setting where p diverges with n. Specifically, a comparable rate of convergence for the largest error across different mean or covariance estimates is derived in the high-dimensional setting, with the only differences being an inflation of the stochastic rate by $\{\log(p)\}^{1/2}$ and an additional term arising from the involvement of higher-order moments in the newly derived bounds; see Theorems 1–4. The OBS scheme is again found to be adversely affected by heterogeneous numbers of

observations, where the quantification of homogeneity (see (5) and (8)) is again explicit but more strict than that of [42]. On the other hand, if the homogeneity condition is satisfied, the OBS scheme is never worse than the SUBJ scheme; see Corollaries 1, 4, 7, and 10. For these two schemes, optimal bandwidth decay rates are also prescribed, yielding observational regime divisions analogous to those of [42] in the case of mean estimation. For covariance estimation, while the parametric rate $\{\log(p)/n\}^{1/2}$ is still attainable for dense and ultra-dense data, the results in this paper do not distinguish between these in terms of bias decay; see Remark 11 for a more detailed explanation.

The remainder of the paper is organized as follows. Section 2 provides definitions of the functional targets and estimators, as well as two particular classes of observational designs that will be considered. Sections 3 and 4 provide all technical assumptions and theoretical results corresponding to the mean and covariance estimators, respectively. The paper concludes with a brief discussion in Section 5, while all proofs are provided in Section 6. The supplementary material includes simulations that illustrate the performance of the OBS and SUBJ schemes for different observational designs in the high-dimensional regime $p > n$, including a discussion of computational aspects and challenges of constructing a large number of smoothing estimators.

2. Methodology

For $p \in \mathbb{N}$, let $\mathcal{T}_j$, $j \in \{1, \ldots, p\}$, be compact intervals of the real line and $\mathcal{T}^p = \bigtimes_{j=1}^p \mathcal{T}_j$ their Cartesian product. It will be assumed throughout that p diverges with n such that $\log(p)/n \to 0$. Let $\{X(\mathbf{t}) \in \mathbb{R}^p;\ \mathbf{t} = (t_1, \ldots, t_p) \in \mathcal{T}^p\}$ be a multivariate L^2 stochastic process, that is, $X(\mathbf{t}) = (X_1(t_1), \ldots, X_p(t_p))^\intercal$ satisfies $E(X_j^2(t)) < \infty$ for all $t \in \mathcal{T}_j$ and all $j \in \{1, \ldots, p\}$. The primary population targets for which estimates are typically sought in functional data analysis are the mean and covariance functions of X, which will be denoted in this paper by

$$\mu_j(s) = E[X_j(s)], \quad \gamma_{jk}(s,t) = \mathrm{Cov}[X_j(s), X_k(t)], \tag{1}$$

for $(s,t) \in \mathcal{T}_j \times \mathcal{T}_k$, $j,k \in \{1, \ldots, p\}$. When $j = k$, γ_{jj} is referred to as the j-th auto-covariance function of X, reflecting the intracurve dependence for a given component function, while γ_{jk} for $j \neq k$ are the cross-covariance functions corresponding to dependence between two distinct functions or curves. Note that generic arguments s and t are used for all of these functions regardless of the domain; these arguments will be referred to as time points, although the functional domains $\mathcal{T}_j$ need not correspond to time. The ultimate aim of this paper is to determine rates of convergence for estimators of these targets constructed from a suitable sample in the high-dimensional regime.

Suppose $X_1, \ldots, X_n$ are independently and identically distributed as X, and write each of these independent multivariate processes as $X_i(\mathbf{t}) = (X_{i1}(t_1), \ldots, X_{ip}(t_p))^\intercal$, $i \in \{1, \ldots, n\}$. A key challenge for functional data is that the processes X_i are never fully observed along the continua; rather, the collected data correspond to a finite number of measurements over a grid of points for each curve. The scheme by which the grid points arise and how these are modeled are referred to as the observational design. Two design settings, described in Section 2.1, will be considered in this paper as extensions of the random design setting treated in previous papers on the topic for univariate functional data [29, 42] in which $p = 1$. In both settings, data for each curve are collected at a random collection of points along the domain, where the number of points per curve may be bounded or diverge with the sample size n.

Prior to considering the particular nuances of the two designs, consider a general model and corresponding mean and covariance estimators. For each subject i and component j, the observed data are modeled as

$$Y_{ij\ell} = X_{ij}(T_{ij\ell}) + \epsilon_{ij\ell} = \mu_j(T_{ij\ell}) + U_{ij}(T_{ij\ell}) + \epsilon_{ij\ell}, \quad \ell \in \{1, \ldots, N_{ij}\}, \tag{2}$$

where $U_{ij} = X_{ij} - \mu_j$ is the centered process, N_{ij} are the number of observation points $T_{ij\ell}$ for this curve, and $\epsilon_{ij\ell}$ are error variables with mean zero, independent of the processes U_{ij}; interdependence of the errors will be left unspecified for the moment as it depends on the observational design. Throughout, the N_{ij} will be considered as deterministic, though varying with n, and their behavior as n diverges will heavily impact the convergence rates.

Following previous work [29, 42], local linear techniques will be used to estimate the functions in (1). Let K be a univariate probability density function and, for a bandwidth $b > 0$, define $K_b(\cdot) = b^{-1}K(\cdot/b)$. For each $j \in \{1, \ldots, p\}$, let $b_{\mu_j} > 0$ and consider positive weights w_{ij} that satisfy $\sum_{i=1}^n w_{ij}N_{ij} = 1$. For $t \in \mathcal{T}_j$, define $\hat{\mu}_j(t) = \hat{\beta}_0$, where

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{i=1}^n w_{ij} \sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell} - t) \left\{Y_{ij\ell} - \beta_0 - \beta_1(T_{ij\ell} - t)\right\}^2. \tag{3}$$

Covariance estimation then follows by defining the raw covariance terms $Z_{ijk\ell m} = \{Y_{ij\ell} - \hat{\mu}_j(T_{ij\ell})\}\{Y_{ikm} - \hat{\mu}_k(T_{ikm})\}$ for any $j, k \in \{1, \ldots, p\}$, $\ell \in \{1, \ldots N_{ij}\}$, and $m \in \{1, \ldots, N_{ik}\}$. To avoid inducing bias in the covariance estimation, raw covariances are typically removed from the estimation if they contain dependencies between the noise variables. To maintain generality for the moment, let $\mathcal{I}_{ijk} \subset \{1, \ldots, N_{ij}\} \times \{1, \ldots, N_{ik}\}$ denote a suitable subset of index pairs (ℓ, m) for which the raw covariances will be included in the estimation; further specification of these will be given in Section 2.1. Let $b_{\gamma_j}, b_{\gamma_k} > 0$ and consider positive weights v_{ijk} satisfying $\sum_{i=1}^{n} v_{ijk}|\mathcal{I}_{ijk}| = 1$. For $(s,t) \in \mathcal{T}_j \times \mathcal{T}_k$, define

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg\min_{\beta_0,\beta_1,\beta_2} \sum_{i=1}^{n} v_{ijk} \sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - s) K_{b_{\gamma_k}}(T_{ikm} - t) \left\{ Z_{ijk\ell m} - \beta_0 - \beta_1(T_{ij\ell} - s) - \beta_2(T_{ikm} - t) \right\}^2. \quad (4)$$

Then the estimate of $\gamma_{jk}(s,t)$ is $\hat{\gamma}_{jk}(s,t) = \hat{\beta}_0$. For simplicity, the same bandwidth b_{γ_j} is used for estimating the auto-covariance γ_{jj} and any cross-covariance γ_{jk}. For the kernel K, the following assumptions are required.

A1 K is a probability density function with support $[-1, 1]$, is symmetric about zero, and is of bounded variation.

A2 K is Lipschitz continuous.

Assumption A1 is ubiquitous in the kernel smoothing literature, while assumption A2 allows for simplification of the proofs of results involving the uniform metric as it governs the smoothness of the estimators [42]. While not strictly necessary, assumption A2 is not restrictive in practice, as it is satisfied for commonly used kernels, such as the Epanichnikov and Gaussian kernels. However, it can be omitted at the cost of more cumbersome arguments [29].

The estimators defined by (3) and (4) will be assessed in terms of their convergence rates in the L^2 and uniform norms, denoted by $\|f\|_j = \left\{\int_{\mathcal{T}_j} f^2(t)\mathrm{d}t\right\}^{1/2}$, $|||f|||_j = \sup_{t\in\mathcal{T}_j} |f(t)|$, $\|g\|_{j,k} = \left\{\int_{\mathcal{T}_j\times\mathcal{T}_k} g^2(s,t)\mathrm{d}s\mathrm{d}t\right\}^{1/2}$, and $|||g|||_{j,k} = \sup_{(s,t)\in\mathcal{T}_j\times\mathcal{T}_k} |g(s,t)|$. Specifically, rates of the convergence for the maximal of these norms across j or (j,k) will be determined that will allow for their consistent estimation so long as $\log(p)$ grows more slowly than the sample size. As in [42], rates will be determined for generic weighting schemes, and special attention will be given to the so-called observation (OBS) and subject (SUBJ) weighting schemes. Define $\overline{N}_j = n^{-1}\sum_{i=1}^{n} N_{ij}$ as the average number of observations of the j-th curve across subjects. Then the OBS weights are $w_{ij} = 1/(n\overline{N}_j)$ and $v_{ijk} = 1/(\sum_{i=1}^{n} |\mathcal{I}_{ijk}|)$. For the SUBJ scheme, one has $w_{ij} = 1/(nN_{ij})$ and $v_{ijk} = 1/(n|\mathcal{I}_{ijk}|)$. Write $\hat{\mu}_{\mathrm{obs},j}$ and $\hat{\gamma}_{\mathrm{obs},jk}$ for the mean and covariance estimators under the OBS scheme, and $\hat{\mu}_{\mathrm{subj},j}$ and $\hat{\gamma}_{\mathrm{subj},jk}$ for those under the SUBJ scheme.

2.1. Observational Designs

The first design setting treats the general case in which the observational time points may be different in number and location across curves. This will be referred to as the fully random design.

Definition 1. The observation times follow a fully random (FR) design if, for each $j \in \{1, \ldots, p\}$, $T_{ij\ell}$, $\ell \in \{1, \ldots, N_{ij}\}$ and $i \in \{1, \ldots, n\}$, are independently distributed on $\mathcal{T}_j$ with probability density f_j and are independent across j.

When the data come from an FR design, the errors $\epsilon_{ij\ell}$ will be assumed independent across i and ℓ and, for covariance estimation, they will also be assumed independent across j. When estimating covariance functions, raw covariances are thus only excluded from (4) when $j = k$ and $\ell = m$. Thus, when an FR design is assumed, one has

$$\mathcal{I}_{ijk} = \begin{cases} \left\{(\ell, m) : \ \ell \in \{1, \ldots, N_{ij}\},\ m \in \{1, \ldots, N_{ik}\}\right\}, & j \neq k, \\ \left\{(\ell, m) : \ \ell, m \in \{1, \ldots, N_{ij}\} \text{ and } \ell \neq m\right\}, & j = k. \end{cases}$$

While some multivariate functional data sets are most appropriately modelled with an FR design, many have a specialized structure, especially when considering examples where the multivariate dimension p is large. In these cases, all functions share the same domain and, for each subject, all p curves are observed simultaneously along a common set of timepoints. This will be referred to as the simultaneous random design.

Definition 2. The observation times follow a simultaneous random (SR) design if, for each $j \in \{1, \ldots, p\}$, $\mathcal{T}_j = \mathcal{T}$ for a compact interval $\mathcal{T}$, $N_{ij} = N_i$, and $T_{ij\ell} = T_{i\ell}$, where $T_{i\ell}$, $\ell \in \{1, \ldots, N_i\}$, and $i \in \{1, \ldots, n\}$, are independently distributed on $\mathcal{T}$ with probability density f.

When the data are assumed to come from an SR design, error variables $\epsilon_{ij\ell}$ and $\epsilon_{ik\ell}$ that are measured at a common timepoint $T_{i\ell}$ may be dependent. Thus, under this design, the index sets for raw covariances will take the form

$$\mathcal{I}_{ijk} = \mathcal{I}_i = \{(\ell, m) : \ell, m \in \{1, \ldots, N_i\} \text{ and } \ell \neq m\}.$$

The following notational simplifications will also be assumed under the SR design. The weights w_{ij} and v_{ijk} will be assumed independent of j and k, denoted as w_i and v_i when appropriate. In the OBS scheme, these take the form $w_i = 1/(n\overline{N})$ and $v_i = 1/(\sum_{i=1}^n N_i(N_i - 1))$, with $\overline{N} = n^{-1}\sum_{i=1}^n N_i$; in the SUBJ scheme, they are $w_i = 1/(nN_i)$ and $v_i = 1/(nN_i(N_i - 1))$. In addition, it will be assumed that a single bandwidth b_μ is used to estimate all p mean functions, and a common bandwidth b_γ is used for all $p(p+1)/2$ covariance functions.

2.2. Technical Assumptions on Model Parameters

The following assumptions on model (2) apply to all relevant theoretical results in Sections 3 and 4.

- B1 The collections of functional data $\{X_i : i \in \{1, \ldots, n\}\}$, observation times $\{T_{ij\ell} : \ell \in \{1, \ldots, N_{ij}\}, j \in \{1, \ldots, p\}, i \in \{1, \ldots, n\}\}$, and errors $\{\epsilon_{ij\ell} : \ell \in \{1, \ldots, N_{ij}\}, j \in \{1, \ldots, p\}, i \in \{1, \ldots, n\}\}$ are independent of each other. Furthermore, the X_i are iid across i, and the data arise from either the FR design in Definition 1 or the SR design in Definition 2. Lastly, with $|\mathcal{T}_j|$ denoting the length of the interval $\mathcal{T}_j$, $0 < \lim_{n\to\infty} \min_{j\in\{1,\ldots,p\}} |\mathcal{T}_j| \leq \lim_{n\to\infty} \max_{j\in\{1,\ldots,p\}} |\mathcal{T}_j| < \infty$.
- B2 Under an FR design, for each $j \in \{1, \ldots, p\}$, f_j is a twice differentiable probability density function on $\mathcal{T}_j$. Moreover, defining $m = \lim_{n\to\infty} \min_{j\in\{1,\ldots,p\}} \inf_{t\in\mathcal{T}_j} f_j(t)$ and $M = \lim_{n\to\infty} \max_{j\in\{1,\ldots,p\}} \max\left\{ \left\| f_j \right\|_j, \left\| f_j'' \right\|_j \right\}$, $0 < m < M < \infty$. Under an SR design, the preceding holds with $f_j \equiv f$.
- B3 The μ_j are twice differentiable, and $\lim_{n\to\infty} \max_{j\in\{1,\ldots,p\}} \left\| \mu_j'' \right\|_j < \infty$.
- B4 The γ_{jk} are twice partially differentiable, and

$$\lim_{n\to\infty} \max_{j,k\in\{1,\ldots,p\}} \max\left(\left\| \frac{\partial^2 \gamma_{jk}}{\partial s^2} \right\|_{j,k}, \left\| \frac{\partial^2 \gamma_{jk}}{\partial s \partial t} \right\|_{j,k}, \left\| \frac{\partial^2 \gamma_{jk}}{\partial t^2} \right\|_{j,k} \right) < \infty.$$

Assumption B1 stipulates the independence of the different random components in the model and the nature of the observation times; to simplify later conditions on bandwidths, it also asserts that the size of the domains for the different functional data components are of the same order. Assumptions B2–B4 are regularly assumed in the case of univariate functional data, and are strengthened here in order for bounds to be uniformly controlled as p diverges.

3. Mean Estimation

In this section, rates of convergence will be provided for mean estimation in the high-dimensional regime for a generic weighting scheme under both FR and SR designs. These results demonstrate that the effects of high-dimensionality are as expected, in that the effective sample size is reduced to $n/\log(p)$. Hence, in determining the divisions between non-dense, dense, and ultra-dense functional data [42], the rate of $\{\log(p)/n\}^{1/2}$ will be critical.

3.1. L^2 Convergence

Consider the estimator defined in (3) for general sequences of weights w_{ij}. Let $\overline{w}_{nj} = \max_{i\in\{1,\ldots,n\}} w_{ij}N_{ij}$. The following assumptions on the distributional characteristics of the functional observations and the asymptotic behavior of bandwidths are required. Let $p \vee n$ denote the maximum of p and n.

- C1 For each $j \in \{1, \ldots, p\}$ and $t \in \mathcal{T}_j$, $U_{ij}(t)$ is sub-Exponential with parameter $\theta_j(t) > 0$, that is, $E[\exp\{\lambda U_{ij}(t)\}] \leq \exp\{\lambda^2\theta_j^2(t)/2\}$ for all $|\lambda| < \{\theta_j(t)\}^{-1}$. Furthermore, $\overline{\theta} = \lim_{n\to\infty} \max_{j\in\{1,\ldots,p\}} \sup_{t\in\mathcal{T}_j} \theta_j(t) < \infty$.

C2 For each j, the errors $\epsilon_{ij\ell}$ are sub-Exponential random variables with parameter σ_j, that is, $E[\exp\{\lambda\epsilon_{ijl}\}] \leq \exp\{\lambda^2\sigma_j^2/2\}$ for all $|\lambda| < \sigma_j^{-1}$. Furthermore, $\sigma = \lim_{n\to\infty} \max_{j\in\{1,\ldots,p\}} \sigma_j < \infty$. In addition, under either an FR or SR design, the $\epsilon_{ij\ell}$ are independent across i and ℓ for any fixed j, but may be dependent across j.

C3 The bandwidths satisfy $\log(p)\max_{j\in\{1,\ldots,p\}} \sum_{i=1}^n w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1) \to 0$, $\log(p)\max_{j\in\{1,\ldots,p\}} b_{\mu_j}^{-1}\overline{w}_{nj} \to 0$, and $\max_{j\in\{1,\ldots,p\}} b_{\mu_j} \to 0$. In addition, $\log(p)\max_{j\in\{1,\ldots,p\}} b_{\mu_j}^{-2} \sum_{i=1}^n w_{ij}^2 N_{ij} \to 0$ and $b_\mu^{-2} \sum_{i=1}^n w_i^2 N_i \to 0$ under the FR and SR designs, respectively.

C4 There exists $\alpha > 0$ such that

$$\frac{n^\alpha \max_{j\in\{1,\ldots,p\}} \left[\left\{\log(p\vee n)\sum_{i=1}^n w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1)\right\}^{1/2} + \log(p\vee n) b_{\mu_j}^{-1}\overline{w}_{nj}\right]}{\max_{j\in\{1,\ldots,p\}} b_{\mu_j}^{-2}} \to \infty.$$

Assumptions C1 and C2 are stronger tail conditions compared to previous work for the case $p = 1$, in which only moment bounds are used to derive concentration inequalities. In ordinary (non-functional) high-dimensional data analysis, two common classes of tail behavior for analyzing mean estimation are polynomial and sub-Exponential tails. The latter lead to probability bounds with exponential decay, and thus faster rates of convergence when p diverges. While acknowledging that rates and division of observational regimes for different classes of tail behavior are certainly of interest, this paper focuses on sub-Exponential tails as in assumptions C1 and C2, leaving other cases for future work. Assumptions C1 and C2, in conjunction with the stipulations on the bandwidths in assumption C3, thus strengthen the second moment assumptions made in [42] in order to obtain L^2 convergence results in Section 3.1 below that are comparable to previous work. In the case of uniform convergence, assumption C1 is not directly comparable to its counterpart in [29, 42], which require that $\left\|\left|X_{ij} - \mu_j\right|\right\|_j$ have bounded r-th moment for some $r > 2$. This latter condition can be viewed as a smoothness assumption on the sample paths of the X_{ij}, and does not imply, neither is it implied by, assumption C1 above. Nevertheless, as will be seen in Section 3.2, the same rates of convergence are also obtained for the uniform metric under tail conditions of assumptions C1 and C2 and under the conditions specified by assumptions C3 and C4 on the bandwidths. The proofs of all results in the section can be found in Section 6.2, with auxiliary lemmas and their proofs given in Section 6.1.

Theorem 1. *Under assumptions A1, B1–B3, and C1–C3,*

$$\max_{j\in\{1,\ldots,p\}} \left\|\hat{\mu}_j - \mu_j\right\|_j = O_P\left(\max_{j\in\{1,\ldots,p\}}\left[b_{\mu_j}^2 + \left\{\log(p)\sum_{i=1}^n w_{ij}^2 N_{ij}(b_{\gamma_j}^{-1} + N_{ij} - 1)\right\}^{1/2} + \log(p) b_{\mu_j}^{-1}\overline{w}_{nj}\right]\right).$$

Remark 1. The rate given in Theorem 1 applies to both FR and SR designs. Under the latter design, the maximum over j in the rate is redundant, requiring less stringent requirements on the bandwidth; see Corollary 3.

Remark 2. The rate in Theorem 1 is obtained by establishing an exponential tail bound on deviation probabilities for each mean estimate, followed by application of the union bound, which leads to the appearance of $\log(p)$ in last two terms making up the stochastic part of the rate; the bias is not affected. In order to compare this result with the rate given in [42], the arguments in the proof of Theorem 1 can be applied to a single mean estimate to yield

$$\left\|\hat{\mu}_j - \mu_j\right\|_j = O_P\left[b_{\mu_j}^2 + \left\{\sum_{i=1}^n w_{ij}^2 N_{ij}(b_{\gamma_j}^{-1} + N_{ij} - 1)\right\}^{1/2} + b_{\mu_j}^{-1}\overline{w}_{nj}\right],$$

while the rate of [42] omits the final term. Its appearance in the rate derived in this paper can be explained as follows. The result relies on an exponential tail bound obtained by applying Theorem 2.5 of [4] and involves all pointwise moments of the functional data as well as the error moments, as opposed to only the first two moments used by [42] that were sufficient to obtain the L^2 convergence rate in the setting $p = 1$ using Chebyshev's inequality. Specifically, to leverage Theorem 2.5 of [4], for any $\nu \geq 2$, one must obtain a moment bound of the form

$$\sum_{i=1}^n w_{ij}^\nu E\left[\left\{\sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell} - t)\right\}^\nu\right] \leq c_{n1}^2 c_{n2}^{\nu-2}.$$

As $\sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell}-t)$ is almost surely bounded above by a multiple of $b_{\mu_j}^{-1}N_{ij}$ and $E\left[\left\{\sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell}-t)\right\}^2\right]$ is bounded by a multiple of $N_{ij}(b_{\mu_j}^{-1}+N_{ij}-1)$, for some $C>0$,

$$\sum_{i=1}^{n} w_{ij}^{\nu} E\left[\left\{\sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell}-t)\right\}^{\nu}\right] \leq (Cb_{\mu_j}^{-1})^{\nu-2}\left\{C\sum_{i=1}^{n} w_{ij}^{\nu} N_{ij}^{\nu-1}(b_{\mu_j}^{-1}+N_{ij}-1)\right\}$$
$$\leq \left\{C\sum_{i=1}^{n} w_{ij}^{2} N_{ij}(b_{\mu_j}^{-1}+N_{ij}-1)\right\}(Cb_{\mu_j}^{-1}\overline{w}_{nj})^{\nu-2}$$

Hence, the final term in the rate cannot be eliminated under general bandwidth sequences. Nevertheless, when applied to the OBS and SUBJ weighting schemes, Corollaries 2 and 3 demonstrate that equivalent rates to those of [42] are obtained for proper bandwidth choices that force the extra term in Theorem 1 to be of smaller order than the others.

Remark 3. Assumption C3 stipulates that the errors $\epsilon_{ij\ell}$ are iid and sub-Exponential across i and ℓ for each fixed j, but dependence across j is arbitrary. In fact, the same rate obtained in Theorem 1 will hold under the weaker assumption $\epsilon_{ij\ell}=\epsilon_{ij}(T_{ij\ell})$ for iid processes ϵ_{ij} satisfying $\max_{j\in\{1,\ldots,p\}}\sup_{t\in\mathcal{T}_j} E(\exp\{\lambda\epsilon_{ij}(t)\})\leq\exp\{\lambda^2\sigma^2/2\}$ for some $0<\sigma^2<\infty$ and all $|\lambda|<(\sigma^2)^{-1}$. The reason that this does not affect the rate is that the error dependence does not dominate than the intracurve dependence present in the latent functional data X_i.

Explicit rates will now be presented when a common weighting scheme, OBS or SUBJ, is used for all curves, although Theorem 1 can still be applied if different weighting schemes are applied to different indices j. [42] demonstrated that the relative performance of OBS and SUBJ depends on the values of the N_{ij}, which may have different behaviors for different indices j. However, in practice, it can be difficult to determine which weighting scheme is best for a given curve, except for an extreme case where some indices j have much larger number of observations per curve compared to others. Thus, for clarity, the weights w_{ij} are hereafter assumed to be constructed according to either OBS or SUBJ for all j. Define $\overline{N}_{j^2}=n^{-1}\sum_{i=1}^{n}N_{ij}^2$, $N_j^+=\max_{i\in\{1,\ldots,n\}}N_{ij}$, and $N_j^H=n(\sum_{i=1}^{n}N_{ij}^{-1})^{-1}$.

Corollary 1. *Suppose the assumptions of Theorem 1 hold.*

(i) *OBS:*

$$\max_{j\in\{1,\ldots,p\}}\left\|\hat{\mu}_{obs,j}-\mu_j\right\|_j = O_P\left(\max_{j\in\{1,\ldots,p\}}\left[b_{\mu_j}^2+\left\{\frac{\log(p)}{n}\left(\frac{1}{\overline{N}_j b_{\mu_j}}+\frac{\overline{N}_{j^2}}{(\overline{N}_j)^2}\right)\right\}^{1/2}+\frac{N_j^+\log(p)}{n\overline{N}_j b_{\mu_j}}\right]\right).$$

(ii) *SUBJ:*

$$\max_{j\in\{1,\ldots,p\}}\left\|\hat{\mu}_{subj,j}-\mu_j\right\|_j = O_P\left(\max_{j\in\{1,\ldots,p\}}\left[b_{\mu_j}^2+\left\{\frac{\log(p)}{n}\left(\frac{1}{N_j^H b_{\mu_j}}+1\right)\right\}^{1/2}+\frac{\log(p)}{nb_{\mu_j}}\right]\right).$$

Remark 4. The rates of Corollary 1 allow one to distinguish between settings in which each weighting scheme is expected to outperform the other, in line with the findings of [42]. Specifically, if the N_{ij} are sufficiently homogeneous across i for each j, in the sense that

$$\limsup_{n\to\infty}\max_{j\in\{1,\ldots,p\}}\max\left\{\frac{\overline{N}_{j^2}}{\overline{N}_j^2},\frac{N_j^+}{\overline{N}_j}\right\}<\infty, \tag{5}$$

then the OBS scheme is never worse than the SUBJ scheme due to the fact that $\overline{N}_j\geq N_j^H$. Indeed, it is possible that the above homogeneity condition holds and that, in addition, $\max_{j\in\{1,\ldots,p\}}N_j^H=O(1)$ while $\min_{j\in\{1,\ldots,p\}}\overline{N}_j\to\infty$, in which case the OBS scheme is strictly better than SUBJ. However, if such homogeneity fails, the OBS scheme rate can suffer considerably, as it can place too much weight on a small proportion of curves. An interesting suggestion of [42] (see Remark 6 therein) for an alternative weighting scheme was to choose an optimal convex combination of the OBS and SUBJ weighting schemes, leading to an asymptotic rate that is better than either, although this theoretical appeal did not always lead to finite sample improvements in their numerical experiments.

In the case of diverging p, Corollary 1 suggests slightly inflated optimal bandwidth choices compared to the fixed p case. When the the curves for different indices j are a mixture of non-dense, dense, and ultra-dense, the rate is driven by the worst rate of any individual mean curve estimate. Under an FR design, an additional complication associated with the maximum rate over j for diverging p is that constants cannot be always ignored in the specification of optimal bandwidths. Thus, for clarity, the following results provide optimal bandwidth choices in accordance with the behavior of the most sparsely and most densely observed functions, represented by $\overline{N}_{\min} = \min_{j\in\{1,\ldots,p\}} \overline{N}_j$ and $\overline{N}_{\max} = \max_{j\in\{1,\ldots,p\}} \overline{N}_j$ for the OBS scheme or $N^H_{\min} = \min_{j\in\{1,\ldots,p\}} N^H_j$ and $N^H_{\max} = \max_{j\in\{1,\ldots,p\}} N^H_j$ for the SUBJ scheme. Remarkably, regardless of the type of functional data available, consistent estimation is possible whenever $\log(p)/n \to 0$, so that exponential growth of p is feasible for non-dense observations, including truly sparse ones in which the N_{ij} are bounded. For positive sequences a_{jn} and b_{jn}, denote by $a_{jn} \preceq b_{jn}$ and $a_{jn} \overset{u}{\preceq} b_{jn}$ the conditions $\limsup_{n\to\infty} a_{jn}b_{jn}^{-1} < \infty$ and $\limsup_{n\to\infty} \max_{j\in\{1,\ldots,p\}} a_{jn}b_{jn}^{-1} < \infty$, respectively. If $a_{jn} \preceq b_{jn}$ and $b_{jn} \preceq a_{jn}$ (respectively, $a_{jn} \overset{u}{\preceq} b_{jn}$ and $b_{jn} \overset{u}{\preceq} a_{jn}$), write $a_{jn} \asymp b_{jn}$ (resp., $a_{jn} \overset{u}{\asymp} b_{jn}$).

Corollary 2. *Assume an FR design and that assumptions A1, B1–B3, C1, and C2 hold.*

(i) *OBS: Assume that* (5) *holds.*

(a) *If* $\overline{N}_{max}\left\{\frac{\log(p)}{n}\right\}^{1/4} \to 0$ *and* $b_{\mu_j} \overset{u}{\asymp} \left\{\frac{\log(p)}{n\overline{N}_j}\right\}^{1/5}$, *then* $\max_{j\in\{1,\ldots,p\}} \left\|\hat{\mu}_{obs,j} - \mu_j\right\|_j = O_P\left[\left\{\frac{\log(p)}{n\overline{N}_{min}}\right\}^{2/5}\right]$.

(b) *If* $0 < \liminf_{n\to\infty} \overline{N}_{min}\left\{\frac{\log(p)}{n}\right\}^{1/4} \le \limsup_{n\to\infty} \overline{N}_{max}\left\{\frac{\log(p)}{n}\right\}^{1/4} < \infty$ *and* $b_{\mu_j} \overset{u}{\asymp} \left\{\frac{\log(p)}{n}\right\}^{1/4}$, *then* $\max_{j\in\{1,\ldots,p\}} \left\|\hat{\mu}_{obs,j} - \mu_j\right\|_j = O_P\left[\left\{\frac{\log(p)}{n}\right\}^{1/2}\right]$.

(c) *If* $\overline{N}_{min}\left\{\frac{\log(p)}{n}\right\}^{1/4} \to \infty$, $b_{\mu_j} \overset{u}{\asymp} b_n$, *where* $b_n = o\left[\left\{\frac{\log(p)}{n}\right\}^{1/4}\right]$, $b_n\overline{N}_{min} \to \infty$, *and* $b_n\left\{\frac{\log(p)}{n}\right\}^{-1/2} \to \infty$, *then* $\max_{j\in\{1,\ldots,p\}} \left\|\hat{\mu}_{obs,j} - \mu_j\right\|_j = O_P\left[\left\{\frac{\log(p)}{n}\right\}^{1/2}\right]$.

(ii) *SUBJ: Replacing* $\overline{N}_j$, $\overline{N}_{min}$, *and* $\overline{N}_{max}$ *with* N^H_j, N^H_{min}, *and* N^H_{max}, *respectively, in parts* (a)–(c) *of* (i) *leads to the corresponding results for* $\hat{\mu}_{subj,j}$.

Remark 5. Compared to the optimal bandwidth choices outlined in [42] for univariate functional data in the non-dense, dense, and ultra-dense regimes, those of Corollary 2 are nearly identical except that the division sample size n is replaced by $n/\log(p)$ in the above result. One minor exception is in the ultra-dense case (c); since the extra term $\log(p)/(nb_{\mu_j})$ in Corollary 1 is not affected by $\overline{N}_j$, under the FR design it is necessary to add an additional constraint to ensure that the smallest bandwidth does not decay more rapidly than $\{\log(p)/n\}^{1/2}$. In doing so, this tail term is always of a smaller order than the others, so does not affect the final rate. In the non-dense case (a), it is the bias and first stochastic term that dominate since $\overline{N}_j b_{\mu_j} \to 0$, and the rate can be anywhere between $\{\log(p)/n\}^{2/5}$ (inclusive) and $\{\log(p)/n\}^{1/2}$ (exclusive). In the dense case, the bias and stochastic terms are all balanced; only in the ultra-dense case do the terms involving the bandwidth become inconsequential, all converging at a rate faster than $\{\log(p)/n\}^{1/2}$.

Lastly, under an SR design, similar rates are obtained under simpler assumptions. Let $\overline{N}$, N^H, $\overline{N}_{(2)}$, and N^+ be the common values of $\overline{N}_j$, N^H_j, $\overline{N}_{j^2}$, and N^+_j, respectively, across j.

Corollary 3. *Assume an SR design and that assumptions A1, B1–B3, C1, and C2 hold.*

(i) *OBS: Assume that* (5) *holds.*

(a) *If* $\overline{N}\left\{\frac{\log(p)}{n}\right\}^{1/4} \to 0$, $b_\mu \asymp \left\{\frac{\log(p)}{n\overline{N}}\right\}^{1/5}$, *then* $\max_{j\in\{1,\ldots,p\}} \left\|\hat{\mu}_{obs,j} - \mu_j\right\|_j = O_P\left[\left\{\frac{\log(p)}{n\overline{N}}\right\}^{2/5}\right]$.

(b) *If* $\overline{N}\left\{\frac{\log(p)}{n}\right\}^{1/4} \to C \in (0,\infty)$ *and* $b_\mu \asymp \left\{\frac{\log(p)}{n}\right\}^{1/4}$, *then* $\max_{j\in\{1,\ldots,p\}} \left\|\hat{\mu}_{obs,j} - \mu_j\right\|_j = O_P\left[\left\{\frac{\log(p)}{n}\right\}^{1/2}\right]$.

(c) *If* $\overline{N}\left\{\frac{\log(p)}{n}\right\}^{1/4} \to \infty$, $b_\mu = o\left[\left\{\frac{\log(p)}{n}\right\}^{1/4}\right]$ *and* $b_\mu\overline{N} \to \infty$, *then* $\max_{j\in\{1,\ldots,p\}} \left\|\hat{\mu}_{obs,j} - \mu_j\right\|_j = O_P\left[\left\{\frac{\log(p)}{n}\right\}^{1/2}\right]$.

(ii) *SUBJ: Replacing* $\overline{N}$ *with* N^H *in parts* (a)–(c) *of* (i) *leads to the corresponding results for* $\hat{\mu}_{subj,j}$.

3.2. Uniform Convergence

Next, consider rates in the uniform norms $\|\!|\cdot|\!\|_j$. Following [29] and [42], these rates will be given in the almost sure sense in order to allow for an easy comparison. Remarks 1–3 also apply to the following result, which provides the rate of convergence for a general weighting scheme.

Theorem 2. *Under assumptions A1, A2, B1–B3, and C1–C4, almost surely,*

$$\max_{j\in\{1,\ldots,p\}} \|\!|\hat{\mu}_j - \mu_j|\!\|_j = O\left(\max_{j\in\{1,\ldots,p\}}\left[b^2_{\mu_j} + \left\{\log(p\vee n)\sum_{i=1}^{n} w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1)\right\}^{1/2} + \log(p\vee n) b_{\mu_j}^{-1}\overline{w}_{nj}\right]\right).$$

Remark 6. As mentioned briefly at the beginning of this section, Theorem 2 does not require any moment bounds on $\|\!|X_{ij} - \mu_j|\!\|_j$, an assumption employed in [29, 42]; indeed, X_{ij} need not even have bounded sample paths in order for the above result to hold. In the proof, one approximates the supremum by a maximum over a grid, with the error in this approximation being bounded by the grid mesh size times a multiple of $\sum_{i=1}^{n} w_{ij}\sum_{\ell=1}^{N_{ij}} |U_{ij}(T_{ij\ell})|$. Under the sub-Exponential condition in assumption C1, this can be controlled directly without using the bound $|U_{ij}(T_{ij\ell})| \le \|\!|U_{ij}|\!\|_j$, regardless of the strength of intracurve dependence present in the latent functional data.

Remark 7. The rates derived in [29] and [42] for $p = 1$ are nearly the same rates as those for weak consistency in the L^2 norm, with the only difference being a slight inflation by $\{\log(n)\}^{1/2}$. Theorem 2 demonstrates a similar phenomenon in the high-dimensional setting, where the rates in Section 3.1 are augmented by $\{\log(p\vee n)\}$. Typically, the results derived in this paper will be of greatest interest when p is much larger than n, in which case the almost sure uniform rates matches exactly the weak consistency rates in the L^2 norm. However, in case $n/p \to \infty$, the new result demonstrates that the uniform rates are strictly worse than the L^2 rates, as should be expected.

The corresponding rates for OBS and SUBJ weighting schemes and the corresponding divisions into non-dense, dense, and ultra-dense data are immediate. Again, Remarks 4 and 5 also apply to the following corollaries.

Corollary 4. *Suppose the assumptions of Theorem 2 hold.*

(i) *OBS: Almost surely,*

$$\max_{j\in\{1,\ldots,p\}} \|\!|\hat{\mu}_{obs,j} - \mu_j|\!\|_j = O\left\{\max_{j\in\{1,\ldots,p\}}\left(b^2_{\mu_j} + \left[\frac{\log(p\vee n)}{n}\left\{\frac{1}{\overline{N}_j b_{\mu_j}} + \frac{\overline{N}_j}{(\overline{N}_j)^2}\right\}\right]^{1/2} + \frac{N_j^{+}\log(p\vee n)}{n\overline{N}_j b_{\mu_j}}\right)\right\}.$$

(ii) *SUBJ: Almost surely,*

$$\max_{j\in\{1,\ldots,p\}} \|\!|\hat{\mu}_{subj,j} - \mu_j|\!\|_j = O\left(\max_{j\in\{1,\ldots,p\}}\left[b^2_{\mu_j} + \left\{\frac{\log(p\vee n)}{n}\left(\frac{1}{N_j^H b_{\mu_j}} + 1\right)\right\}^{1/2} + \frac{\log(p\vee n)}{n b_{\mu_j}}\right]\right).$$

Corollary 5. *Assume an FR design and that assumptions A1, A2, B1–B3, C1, and C2 hold.*

(i) *OBS: Assume that* (5) *holds.*

(a) *If* $\overline{N}_{max}\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4} \to 0$ *and* $b_{\mu_j} \overset{u}{\asymp} \left\{\frac{\log(p\vee n)}{n\overline{N}_j}\right\}^{1/5}$, *then* $\max_{j\in\{1,\ldots,p\}} \|\!|\hat{\mu}_{obs,j} - \mu_j|\!\|_j = O\left[\left\{\frac{\log(p\vee n)}{n\overline{N}_{min}}\right\}^{2/5}\right]$ *almost surely.*

(b) *If* $0 < \liminf_{n\to\infty} \overline{N}_{min}\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4} \le \limsup_{n\to\infty} \overline{N}_{max}\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4} < \infty$ *and* $b_{\mu_j} \overset{u}{\asymp} \left\{\frac{\log(p\vee n)}{n}\right\}^{1/4}$, *then* $\max_{j\in\{1,\ldots,p\}} \|\!|\hat{\mu}_{obs,j} - \mu_j|\!\|_j = O\left[\left\{\frac{\log(p\vee n)}{n}\right\}^{1/2}\right]$ *almost surely.*

(c) *If* $\overline{N}_{min}\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4} \to \infty$, $b_{\mu_j} \overset{u}{\asymp} b_n$, *where* $b_n = o\left[\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4}\right]$, $b_n\overline{N}_{min} \to \infty$, *and* $b_n\left\{\frac{\log(p\vee n)}{n}\right\}^{-1/2} \to \infty$, *then* $\max_{j\in\{1,\ldots,p\}} \|\!|\hat{\mu}_{obs,j} - \mu_j|\!\|_j = O\left[\left\{\frac{\log(p\vee n)}{n}\right\}^{1/2}\right]$ *almost surely.*

(ii) *SUBJ: Replacing* $\overline{N}_j$, $\overline{N}_{min}$, *and* $\overline{N}_{max}$ *with* N_j^H, N_{min}^H, *and* N_{max}^H, *respectively, in parts* (a)–(c) *of* (i) *leads to the corresponding results for* $\hat{\mu}_{subj,j}$.

Corollary 6. *Assume an SR design and that assumptions A1, A2, B1–B3, C1, and C2 hold.*

(i) *OBS: Assume that* (5) *holds.*

(a) *If* $\overline{N}\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4} \to 0$ *and* $b_\mu \asymp \left\{\frac{\log(p\vee n)}{n\overline{N}}\right\}^{-1/5}$, *then* $\max_{j\in\{1,\dots,p\}} \left|\!\left|\!\left|\hat{\mu}_{obs,j} - \mu_j\right|\!\right|\!\right|_j = O\left[\left\{\frac{\log(p\vee n)}{n\overline{N}}\right\}^{2/5}\right]$ *almost surely.*

(b) *If* $\overline{N}\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4} \to C \in (0,\infty)$ *and* $b_\mu \asymp \left\{\frac{\log(p\vee n)}{n}\right\}^{1/4}$, *then, almost surely,* $\max_{j\in\{1,\dots,p\}} \left|\!\left|\!\left|\hat{\mu}_{obs,j} - \mu_j\right|\!\right|\!\right|_j = O\left[\left\{\frac{\log(p\vee n)}{n}\right\}^{1/2}\right]$.

(c) *If* $\overline{N}\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4} \to \infty$, $b_\mu = o\left[\left\{\frac{\log(p\vee n)}{n}\right\}^{1/4}\right]$ *and* $b_\mu\overline{N} \to \infty$, *then, almost surely,* $\max_{j\in\{1,\dots,p\}} \left|\!\left|\!\left|\hat{\mu}_{obs,j} - \mu_j\right|\!\right|\!\right|_j = O\left[\left\{\frac{\log(p\vee n)}{n}\right\}^{1/2}\right]$.

(ii) *SUBJ: Replacing* $\overline{N}$ *with* N^H *in parts* (a)–(c) *of* (i) *leads to the corresponding results for* $\hat{\mu}_{subj,j}$.

4. Covariance Estimation

Next, rates of convergence for the covariance estimators will be provided. In order to separate the effects of mean estimation from covariance estimation, the mean is assumed to be known so that the raw covariances Z_{ijklm} used in (4) are replaced by $Z_{ijk\ell m} = (Y_{ij\ell} - \mu_j(T_{ij\ell}))(Y_{ikm} - \mu_k(T_{ikm}))$. The ensuing estimates will still be denoted $\hat{\gamma}_{jk}$. The rates for the true empirical estimator are obtained by adding the uniform mean convergence rates, since raw covariances involve pointwise evaluations of the mean estimates.

4.1. L^2 Convergence

Under an FR design, define $\overline{v}_{njk} = \max_{i\in\{1,\dots,n\}} v_{ijk}|\mathcal{I}_{ijk}|$ and

$$q_{jkn} = \begin{cases} \sum_{i=1}^n v_{ijk}^2 N_{ij}N_{ik}(b_{\gamma_j}^{-1} + N_{ij} - 1)(b_{\gamma_k}^{-1} + N_{ik} - 1), & j \neq k, \\ \sum_{i=1}^n v_{ijj}^2 N_{ij}(N_{ij}-1)\left\{b_{\gamma_j}^{-2} + 2b_{\gamma_j}^{-1}(N_{ij}-2) + (N_{ij}-2)(N_{ij}-3)\right\}, & j = k. \end{cases} \tag{6}$$

Under an SR design, for all j,k, take $\overline{v}_{njk} = \max_{i\in\{1,\dots,n\}} v_i|\mathcal{I}_i|$ and q_{jkn} to be the common value of q_{jjn} across j in (6). In addition, define

$$\omega_{jk}^2 = \begin{cases} \max(\sum_{i=1}^n v_{ijk}^2 N_{ij}^2 N_{ik}, \sum_{i=1}^n v_{ijk}^2 N_{ij}N_{ik}^2), & j \neq k, \\ \sum_{i=1}^n v_{ijj}^2 N_{ij}(N_{ij}-1)^2, & j = k, \end{cases} \tag{7}$$

and let ω denote the common value of ω_{jj} under an SR design. The following assumptions will be used.

D1 For each $j \in \{1,\dots,p\}$ and $t \in \mathcal{T}_j$, $U_{ij}(t)$ is sub-Gaussian with parameter $\theta_j(t) > 0$, that is, $E[\exp\{\lambda U_{ij}(t)\}] \leq \exp\{\lambda^2\theta_j^2(t)/2\}$ for all $\lambda \in \mathbb{R}$. Furthermore, $\overline{\theta} = \lim_{n\to\infty} \max_{j\in\{1,\dots,p\}} \sup_{t\in\mathcal{T}_j} \theta_j(t) < \infty$.

D2 For each j, the errors $\epsilon_{ij\ell}$ are sub-Gaussian random variables with parameter σ_j, that is, $E[\exp\{\lambda\epsilon_{ij\ell}\}] \leq \exp\{\lambda^2\sigma_j^2/2\}$ for all $\lambda \in \mathbb{R}$. Furthermore, $\sigma = \lim_{n\to\infty} \max_{j\in\{1,\dots,p\}} \sigma_j < \infty$. In addition, under an FR design, the $\epsilon_{ij\ell}$ are independent across all indices; under an SR design, $\epsilon_{ij\ell}$ and $\epsilon_{i'j'\ell'}$ are independent whenever $(i,\ell) \neq (i',\ell')$.

D3 The bandwidths satisfy $\max_{j\in\{1,\dots,p\}} b_{\gamma_j} \to 0$ and $\log(p)\max_{j,k\in\{1,\dots,p\}} q_{jkn} \to 0$, with q_{jkn} defined in (6). In addition, $\log(p)\max_{j,k\in\{1,\dots,p\}} b_{\gamma_j}^{-1}b_{\gamma_k}^{-1}\omega_{jk}^2 \to 0$ under an FR design; under an SR design, $b_\gamma^{-2}\omega^2 \to 0$.

D4 There exists $\alpha > 0$ such that, with q_{jkn} as defined in (6),

$$\frac{n^{\alpha} \max_{j,k\in\{1,\ldots,p\}} \left[\left\{ \log(p \vee n) q_{jkn} \right\}^{1/2} + \log(p \vee n) b_{\gamma_j}^{-1} b_{\gamma_k}^{-1} \overline{v}_{njk} \right]}{\max_{j\in\{1,\ldots,p\}} b_{\gamma_j}^{-3}} \to \infty.$$

Assumptions D1–D4 correspond to assumptions C1–C4, but adapted for covariance estimation. Critically, assumptions D1 and D2 impose sub-Gaussian rather than sub-Exponential tails, due to the fact that covariance estimates involve averages of products, and products of sub-Gaussian random variables have sub-Exponential tails. Proofs of all results in this section can be found in Section 6.4, with auxiliary lemmas and their proofs given in Section 6.3.

Theorem 3. *Under assumptions A1, B1, B2, B4, and D1–D3,*

$$\max_{j,k\in\{1,\ldots,p\}} \left\| \hat{\gamma}_{jk} - \gamma_{jk} \right\|_{j,k} = O_P \left(\max_{j\in\{1,\ldots,p\}} b_{\gamma_j}^2 + \max_{j,k\in\{1,\ldots,p\}} \left[\left\{ \log(p) q_{jkn} \right\}^{1/2} + \log(p) b_{\gamma_j}^{-1} b_{\gamma_k}^{-1} \overline{v}_{njk} \right] \right).$$

Remark 8. As in Theorem 1, the rates in Theorem 3 are stated in the context of an FR design, but also hold under an SR design using the common values of the relevant quantities across j, k, including a common bandwidth b_γ, that have been defined previously. Assumption D3 and comparison of Corollaries 8 and 9 below illustrate the weaker bandwidth requirements under an SR design.

Remark 9. Similar to Remark 2, the rate in Theorem 3 can be compared with that of [42]. As the latter did not consider cross-covariance estimation, consider $j = k$. Once again, the arguments in the proof of Theorem 3, when applied to a single auto-covariance estimate $\hat{\gamma}_{jj}$, lead to the rate $b_{\gamma_j}^2 + q_{jkn}^{1/2} + b_{\gamma_j}^{-2} \overline{v}_{njj}$, the first two terms matching the rate of [42]. As for mean estimation, the additional term in the derived rate arises from the involvement of higher order moments needed for the exponential tail bound. This extra term has more impact on optimal bandwidth choice in covariance estimation than mean estimation, but still does not affect the overall rates.

The following corollaries translate the rate of Theorem 3 to the OBS and SUBJ weighting schemes under FR and SR designs, respectively. For $r, s \in \{1, 2\}$, define $\overline{N}_{j^r k^s} = n^{-1} \sum_{i=1}^n N_{ij}^r N_{ik}^s$, $\overline{N}_{jk} = \overline{N}_{j^1 k^1}$, $N_{jk}^+ = \max_{i\in\{1,\ldots,n\}} N_{ij} N_{ik}$, $N_{j^r k^s}^H = n \left(\sum_{i=1}^n N_{ij}^{-r} N_{ik}^{-s} \right)^{-1}$, and $N_{jk}^H = N_{j^1 k^1}^H$. Under an SR design, let $N_{(2)}^H$ denote the common value of N_{jk}^H across j, k.

Corollary 7. *Suppose the assumptions of Theorem 3 hold.*

(i) *OBS:*

$$\max_{j,k\in\{1,\ldots,p\}} \left\| \hat{\gamma}_{obs,jk} - \gamma_{jk} \right\|_{j,k} = O_P \left(\max_{j\in\{1,\ldots,p\}} b_{\gamma_j}^2 + \max_{j,k\in\{1,\ldots,p\}} \left[\left\{ \frac{\log(p)}{n} \left(\frac{1}{\overline{N}_{jk} b_{\gamma_j} b_{\gamma_k}} + \frac{\overline{N}_{jk^2}}{\overline{N}_{jk}^2 b_{\gamma_j}} + \frac{\overline{N}_{j^2k^2}}{\overline{N}_{jk}^2} \right) \right\}^{1/2} + \frac{N_{jk}^+ \log(p)}{n \overline{N}_{jk} b_{\gamma_j} b_{\gamma_k}} \right] \right).$$

(ii) *SUBJ:*

$$\max_{j,k\in\{1,\ldots,p\}} \left\| \hat{\gamma}_{obs,jk} - \gamma_{jk} \right\|_{j,k} = O_P \left(\max_{j\in\{1,\ldots,p\}} b_{\gamma_j}^2 + \max_{j,k\in\{1,\ldots,p\}} \left[\left\{ \frac{\log(p)}{n} \left(\frac{1}{N_{jk}^H b_{\gamma_j} b_{\gamma_k}} + \frac{1}{N_j^H b_{\gamma_j}} + 1 \right) \right\}^{1/2} + \frac{\log(p)}{n b_{\gamma_j} b_{\gamma_k}} \right] \right).$$

Remark 10. Corollary 7 allows one to compare the two specific weighting schemes. By strengthening the homogeneity condition on the observation numbers N_{ij} in (5) to the condition

$$\limsup_{n\to\infty} \max_{j,k\in\{1,\ldots,p\}} \max \left\{ \frac{\overline{N}_j \overline{N}_{jk^2}}{\overline{N}_{jk}^2}, \frac{\overline{N}_{j^2k^2}}{\overline{N}_{jk}^2}, \frac{N_{jk}^+}{\overline{N}_{jk}} \right\} < \infty, \tag{8}$$

then the OBS scheme is never worse than the SUBJ scheme due to the fact that $\overline{N}_j \geq N_j^H$ and $\overline{N}_{jk} \geq N_{jk}^H$. If (8) fails, however, the estimators may not be consistent under the OBS scheme.

Corollary 8. *Assume an FR design and that assumptions A1, B1, B2, B4, D1, and D2 hold.*

(i) *OBS: Assume that* (8) *holds and that* $\limsup_{n\to\infty} \max_{j,k\in\{1,\ldots,p\}}(\overline{N}_j\overline{N}_k)/\overline{N}_{jk} < \infty$.

(a) *If* $\overline{N}_{max}\left\{\frac{\log(p)}{n}\right\}^{1/4} \to 0$, $b_{\gamma_j} \overset{u}{\asymp} \left\{\frac{\log(p)}{n\overline{N}_j^2}\right\}^{1/6}$, *then* $\max_{j,k\in\{1,\ldots,p\}} \left\|\hat{\gamma}_{obs,jk} - \gamma_{jk}\right\|_{j,k} = O_P\left[\left\{\frac{\log(p)}{n\overline{N}_{min}^2}\right\}^{1/3}\right]$.

(b) *If* $\liminf_{n\to\infty} \overline{N}_{min}\left\{\frac{\log(p)}{n}\right\}^{1/4} > 0$ *and* $b_{\gamma_j} \overset{u}{\asymp} \left\{\frac{\log(p)}{n}\right\}^{1/4}$, *then* $\max_{j,k\in\{1,\ldots,p\}} \left\|\hat{\gamma}_{obs,jk} - \gamma_{jk}\right\|_{j,k} = O_P\left[\left\{\frac{\log(p)}{n}\right\}^{1/2}\right]$.

(ii) *SUBJ: Assume that* $\limsup_{n\to\infty} \max_{j,k\in\{1,\ldots,p\}}(N_j^H N_k^H)/N_{jk}^H < \infty$. *Replacing* $\overline{N}_j$, $\overline{N}_{min}$, *and* $\overline{N}_{max}$ *with* N_j^H, N_{min}^H, *and* N_{max}^H, *respectively, in parts* (a) *and* (b) *of* (i) *leads to the corresponding results for* $\hat{\gamma}_{subj,jk}$.

Corollary 9. *Assume an SR design and that assumptions A1, B1, B2, B4, D1, and D2 hold.*

(i) *OBS: Assume that* (8) *holds.*

(a) *If* $\overline{N}\left\{\frac{\log(p)}{n}\right\}^{1/4} \to 0$, $b_\gamma \asymp \left\{\frac{\log(p)}{n\overline{N}^2}\right\}^{1/6}$, *then* $\max_{j,k\in\{1,\ldots,p\}} \left\|\hat{\gamma}_{obs,jk} - \gamma_{jk}\right\|_{j,k} = O_P\left[\left\{\frac{\log(p)}{n\overline{N}^2}\right\}^{1/3}\right]$.

(b) *If* $\liminf_{n\to\infty} \overline{N}\left\{\frac{\log(p)}{n}\right\}^{1/4} > 0$ *and* $b_\gamma \asymp \left\{\frac{\log(p)}{n}\right\}^{1/4}$, *then* $\max_{j,k\in\{1,\ldots,p\}} \left\|\hat{\gamma}_{obs,jk} - \gamma_{jk}\right\|_{j,k} = O_P\left[\left\{\frac{\log(p)}{n}\right\}^{1/2}\right]$.

(ii) *SUBJ: Assume that* $\limsup_{n\to\infty}(N^H)^2/N_{(2)}^H < \infty$. *Replacing* $\overline{N}$ *with* N^H *in parts* (a) *and* (b) *of* (i) *leads to the corresponding results for* $\hat{\gamma}_{subj,jk}$.

Remark 11. Unlike Corollaries 2 and 3, Corollaries 8 and 9 contain only two cases for each of the OBS and SUBJ weighting schemes, with the dense and ultra-dense regimes being combined. This is due to the extra term in Corollary 7 discussed previously in Remark 9. For mean estimation, in the ultra-dense regime, the bandwidths b_{μ_j} are all allowed to decay more quickly than $\{\log(p)/n\}^{1/4}$. However, allowing the same behavior for covariance bandwidths b_{γ_j} would cause the last term in the rate to be slower than $\{\log(p)/n\}^{1/2}$. Hence, for high-dimensional functional data, while this result does not distinguish between dense and ultra-dense observation designs in terms of covariance estimation, in both regimes one is still able to obtain the appropriate rate. In the sparse regime, the rate can be anywhere between $\{\log(p)/n\}^{1/3}$ (inclusive) and $\{\log(p)/n\}^{1/2}$ (exclusive).

4.2. Uniform Convergence

For brevity, the strong uniform rates of convergence for covariance estimation will be stated without further discussion; the reader is referred to Remarks 6, 7, and 9–11 for relevant comments on these results.

Theorem 4. *Under assumptions A1, A2, B1, B2, B4, and D1–D4, almost surely,*

$$\max_{j,k\in\{1,\ldots,p\}} \left|\!\left|\!\left|\hat{\gamma}_{jk} - \gamma_{jk}\right|\!\right|\!\right|_{j,k} = O\left(\max_{j\in\{1,\ldots,p\}} b_{\gamma_j}^2 + \max_{j,k\in\{1,\ldots,p\}}\left[\left\{\log(p\vee n)q_{jkn}\right\}^{1/2} + \log(p\vee n)b_{\gamma_j}^{-1}b_{\gamma_k}^{-1}\overline{v}_{njk}\right]\right).$$

Corollary 10. *Suppose the assumptions of Theorem 3 hold.*

(i) *OBS:*

$$\max_{j,k\in\{1,\ldots,p\}} \left|\!\left|\!\left|\hat{\gamma}_{obs,jk} - \gamma_{jk}\right|\!\right|\!\right|_{j,k} = O_P\left(\max_{j\in\{1,\ldots,p\}} b_{\gamma_j}^2 + \max_{j,k\in\{1,\ldots,p\}}\left[\left\{\frac{\log(p\vee n)}{n}\left(\frac{1}{\overline{N}_{jk}b_{\gamma_j}b_{\gamma_k}} + \frac{\overline{N}_{jk^2}}{\overline{N}_{jk}^2 b_{\gamma_j}} + \frac{\overline{N}_{j^2k^2}}{\overline{N}_{jk}^2}\right)\right\}^{1/2} + \frac{N_{jk}^+\log(p\vee n)}{n\overline{N}_{jk}b_{\gamma_j}b_{\gamma_k}}\right]\right).$$

(ii) *SUBJ:*

$$\max_{j,k\in\{1,\ldots,p\}} \left|\!\left|\!\left| \hat{\gamma}_{obs,jk} - \gamma_{jk} \right|\!\right|\!\right|_{j,k} = O_P\left(\max_{j\in\{1,\ldots,p\}} b_{\gamma_j}^2 + \max_{j,k\in\{1,\ldots,p\}} \left[\left\{ \frac{\log(p\vee n)}{n} \left(\frac{1}{N_{jk}^H b_{\gamma_j} b_{\gamma_k}} + \frac{1}{N_j^H b_{\gamma_j}} + 1 \right) \right\}^{1/2} + \frac{\log(p\vee n)}{n b_{\gamma_j} b_{\gamma_k}} \right] \right).$$

Corollary 11. *Assume an FR design and that assumptions A1, B1, B2, B4, D1, and D2 hold.*

(i) *OBS: Assume that* (8) *holds and that* $\limsup_{n\to\infty} \max_{j,k\in\{1,\ldots,p\}} (\overline{N}_j \overline{N}_k)/\overline{N}_{jk} < \infty$.

(a) *If* $\overline{N}_{max} \left\{ \frac{\log(p\vee n)}{n} \right\}^{1/4} \to 0$, $b_{\gamma_j} \overset{u}{\asymp} \left\{ \frac{\log(p\vee n)}{n\overline{N}_j^2} \right\}^{1/6}$, *then* $\max_{j,k\in\{1,\ldots,p\}} \left|\!\left|\!\left| \hat{\gamma}_{obs,jk} - \gamma_{jk} \right|\!\right|\!\right|_{j,k} = O\left[\left\{ \frac{\log(p\vee n)}{n\overline{N}_{min}^2} \right\}^{1/3} \right]$ *almost surely.*

(b) *If* $\liminf_{n\to\infty} \overline{N}_{min} \left\{ \frac{\log(p\vee n)}{n} \right\}^{1/4} > 0$ *and* $b_{\gamma_j} \overset{u}{\asymp} \left\{ \frac{\log(p\vee n)}{n} \right\}^{1/4}$, *then, almost surely,* $\max_{j,k\in\{1,\ldots,p\}} \left|\!\left|\!\left| \hat{\gamma}_{obs,jk} - \gamma_{jk} \right|\!\right|\!\right|_{j,k} = O\left[\left\{ \frac{\log(p\vee n)}{n} \right\}^{1/2} \right]$.

(ii) *SUBJ: Assume that* $\limsup_{n\to\infty} \max_{j,k\in\{1,\ldots,p\}} (N_j^H N_k^H)/N_{jk}^H < \infty$. *Replacing* $\overline{N}_j$, $\overline{N}_{min}$, *and* $\overline{N}_{max}$ *with* N_j^H, N_{min}^H, *and* N_{max}^H, *respectively, in parts* (a) *and* (b) *of* (i) *leads to the corresponding results for* $\hat{\gamma}_{subj,jk}$.

Corollary 12. *Assume an SR design and that assumptions A1, B1, B2, B4, D1, and D2 hold.*

(i) *OBS: Assume that* (8) *holds.*

(a) *If* $\overline{N} \left\{ \frac{\log(p\vee n)}{n} \right\}^{1/4} \to 0$, *and* $b_\gamma \asymp \left\{ \frac{\log(p\vee n)}{n\overline{N}^2} \right\}^{1/6}$, *then* $\max_{j,k\in\{1,\ldots,p\}} \left\| \hat{\gamma}_{obs,jk} - \gamma_{jk} \right\|_{j,k} = O\left[\left\{ \frac{\log(p\vee n)}{n\overline{N}^2} \right\}^{1/3} \right]$ *almost surely.*

(b) *If* $\liminf_{n\to\infty} \overline{N} \left\{ \frac{\log(p\vee n)}{n} \right\}^{1/4} > 0$ *and* $b_\gamma \asymp \left\{ \frac{\log(p\vee n)}{n} \right\}^{1/4}$, *then, almost surely,* $\max_{j,k\in\{1,\ldots,p\}} \left\| \hat{\gamma}_{obs,jk} - \gamma_{jk} \right\|_{j,k} = O\left[\left\{ \frac{\log(p\vee n)}{n} \right\}^{1/2} \right]$.

(ii) *SUBJ: Assume that* $\limsup_{n\to\infty} (N^H)^2/N_{(2)}^H < \infty$. *Replacing* $\overline{N}$ *with* N^H *in parts* (a) *and* (b) *of* (i) *leads to the corresponding results for* $\hat{\gamma}_{subj,jk}$.

5. Discussion

The results derived in this paper provide an important foundation for high-dimensional functional data analysis by establishing sufficient conditions under which uniform consistency of a diverging number of mean and covariance estimates will hold. Importantly, the practical reality of discrete and noisy functional observations does not preclude consistent estimation in high dimensions; indeed, the results lead to a natural division of high-dimensional functional data into three regimes (sparse, dense, and ultra-dense) based on the behavior of the average number of observations available per component curve relative to $\{\log(p)/n\}^{1/4}$, providing the expected generalization of the regime divisions discovered in [42] for univariate functional data. By properly choosing the smoothing bandwidths, the worst case scenario for sparsely observed functions is the optimal nonparametric high-dimensional convergence rate for both mean and covariance estimation; for densely or ultra-densely observed curves, the parametric rate is always attainable.

The results utilize concentration inequalities in Hilbert spaces [4] that require stricter tail assumptions on the pointwise behavior of the functional data compared to previous results for the case $p = 1$. In this paper, sub-Exponential and sub-Gaussian tails were used for mean and covariance estimation, respectively. With these tail assumptions, the results were able to successfully distinguish between dense and ultra-dense functional data for both the OBS or SUBJ weighting scheme in mean estimation, but not so for covariance estimation, due to an additional term in the rate that was not present in previous work [29, 42] in the case $p = 1$; see Remarks 9 and 11. Thus, it is possible

that stronger concentration inequalities or alternative modes of analysis may lead to such a distinction in future work. For example, under the assumption of tails with only polynomial decay, similar arguments to those of [42] could be used to derive rates that, while slower than those obtained here, have the potential to effectively distinguish between dense and ultra-dense data. Besides different tail assumptions, there are many other choices that may lead to different divisions, including different learning tasks such as regression or classification, modes of convergence (e.g., pointwise asymptotic normality), or quantification of estimation error besides the L^2 and uniform metrics. Moreover, while the theorems and corrollaries present asymptotic properties, the lemmas in Sections 6.1 and 6.3 are non-asymptotic in nature, and may be useful for finite sample inference.

The use of local linear methodology in the analysis of high-dimensional functional data analysis will undoubtedly have many limitations, some of which are theoretical and others that are practical. This paper has addressed the former via asymptotic analyses as well as via simulations in the supplementary material. Important practical questions, such as bandwidth selection or approximations for computational speedup, will require adaptation of the usual tools for univariate functional data if they are to remain scalable. For a brief discussion of some of these issues, see Section 1.2 in the supplementary material.

Although the terms sparse and dense appear frequently in this paper, there has been no application of regularization or shrinkage in the estimation procedure. It is well-known that, for multivariate non-functional data of high dimension, such approaches can lead to improved rates of convergence and enhanced interpretability. For high-dimensional functional data, such regularization has been successfully applied in regression models and graphical model estimation, although theoretical guarantees have mostly been established in the oracle case of fully observed functions. The results in this paper will provide a path for theoretical investigation of these and other novel methodologies for high-dimensional functional data that are applicable under any observational design, requiring only initial mean and covariance plug-in estimates.

6. Technical Details

This section provides technical arguments for the theoretical results stated in Sections 3 and 4.

6.1. Auxiliary Lemmas for Mean Estimation

For $r \in \{0, 1, 2\}$, and $j \in \{1, \ldots, p\}$, define

$$S_{jr}(t) = \sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}\left(T_{ij\ell} - t\right)\left(\frac{T_{ij\ell} - t}{b_{\mu_j}}\right)^r, \quad R_{jr}(t) = \sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}\left(T_{ij\ell} - t\right)\left(\frac{T_{ij\ell} - t}{b_{\mu_j}}\right)^r Y_{ij\ell}.$$

Then the error in mean estimation can be expressed as

$$\hat{\mu}_j(t) - \mu_j(t) = \frac{\left\{R_{j0}(t) - \mu_j(t)S_{j0}(t) - b_{\mu_j}\mu_j'(t)S_{j1}(t)\right\}S_{j2}(t)}{S_{j0}(t)S_{j2}(t) - S_{j1}^2(t)} - \frac{\left\{R_{j1}(t) - \mu_j(t)S_{j1}(t) - b_{\mu_j}\mu_j'(t)S_{j2}(t)\right\}S_{j1}(t)}{S_{j0}(t)S_{j2}(t) - S_{j1}^2(t)}. \quad (9)$$

With $U_{ij\ell} = U_{ij}(T_{ij\ell}) + \epsilon_{ij\ell}$, the numerator in the first term on the right-hand side of (9) can be expressed as

$$\begin{aligned} R_{j0}(t) - \mu_j(t)S_{j0}(t) - b_{\mu_j}\mu_j'(t)S_{j1}(t) = &\sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell} - t)U_{ij\ell} \\ &+ \sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell} - t)(T_{ij\ell} - t)^2 \int_0^1 \mu_j''\{t + v(T_{ij\ell} - t)\}(1 - v)\mathrm{d}v, \end{aligned} \quad (10)$$

which follows from a Taylor expansion. Similarly,

$$\begin{aligned} R_{j1}(t) - \mu_j(t)S_{j1}(t) - b_{\mu_j}\mu_j'(t)S_{j2}(t) = &\sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell} - t)\left(\frac{T_{ij\ell} - t}{b_{\mu_j}}\right)U_{ij\ell} \\ &+ \sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell} - t)\left(\frac{T_{ij\ell} - t}{b_{\mu_j}}\right)(T_{ij\ell} - t)^2 \int_0^1 \mu_j''(t + v(T_{ij\ell} - t))(1 - v)\mathrm{d}v. \end{aligned} \quad (11)$$

6.1.1. Uniform Convergence of S_{jr}, $r \in \{0,1,2\}$

The results of this section allows for the derivation of lower bounds for the denominator of (9) that hold uniformly in t and across j with high probability. In this result and those that follow, constants that are uniform over all data generating mechanisms satisfying the mentioned assumptions will be denoted by C_a, $a \in \mathbb{N}$, and may take different values across different inequalities. Finally, these results do not depend on whether the observation times follow a fully or simultaneous random design.

Lemma 1. *Suppose that assumptions A1 and B1 hold. There exist $C_1, C_2 > 0$ such that, for any $\epsilon > 0$,*

$$\Pr\left(\left|\left|\left|S_{jr}(\cdot) - E[S_{jr}(\cdot)]\right|\right|\right|_j > \epsilon\right) \leq C_1 \exp\left\{\frac{-C_2 b_{\mu_j}^2 \epsilon^2}{\sum_{i=1}^n w_{ij}^2 N_{ij}}\right\}, \quad r \in \{0,1,2\},\ j \in \{1,\ldots,p\}.$$

Proof: Define the weighted empirical distribution function $\hat{F}_j(t) = \sum_{i=1}^n w_{ij} \sum_{\ell=1}^{N_{ij}} \mathbf{1}(T_{ij\ell} \leq t)$. Take $V_{K,r}$ to be the total variation of the kernel $K(u)u^r$ over $[-1,1]$ for $r \in \{0,1,2\}$, an set $V_K = \max_{r\in\{0,1,2\}} V_{K,r}$. Then standard arguments show that $\left|\left|\left|S_{jr}(\cdot) - E[S_{jr}(\cdot)]\right|\right|\right|_j \leq V_K b_{\mu_j}^{-1} \left|\left|\left|\hat{F}_j - F_j\right|\right|\right|_j$. Next, Lemma 1.1 of [30] implies that, for any $\tau > 0$,

$$\Pr\left\{\left|\left|\left|\hat{F}_j - F_j\right|\right|\right|_j > \tau\left(\sum_{i=1}^n w_{ij}^2 N_{ij}\right)^{1/2}\right\} \leq \left(1 + 2\sqrt{2\pi}\tau\right)\exp\left\{-\tau^2/8\right\}.$$

Since, for any polynomial $p(u)$, $\sup_{u>0} p(u)e^{-u^2} \leq a_1 e^{-a_2 u^2}$ for some constants $a_1, a_2 > 0$, it follows that, for any $\epsilon > 0$, there are constants $C_1, C_2 > 0$ such that, as claimed,

$$\Pr\left(\left|\left|\left|S_{jr}(\cdot) - E[S_{jr}(\cdot)]\right|\right|\right|_j > \epsilon\right) \leq \Pr\left(\left|\left|\left|\hat{F}_j - F_j\right|\right|\right|_j > V_K^{-1} b_{\mu_j}\epsilon\right) \leq C_1 \exp\left\{\frac{-C_2 b_{\mu_j}^2 \epsilon^2}{\sum_{i=1}^n w_{ij}^2 N_{ij}}\right\}.$$

□

Lemma 2. *Suppose assumptions A1, B1, B2, and C3 hold. Then there exist $C_1, C_2, \eta, N > 0$ such that, for any $n \geq N$,*

$$\Pr\left[\inf_{t\in\mathcal{T}_j}\left\{S_{j0}(t)S_{j2}(t) - S_{j1}^2(t)\right\} < \eta\right] \leq C_1 \exp\left\{-\frac{C_2 b_{\mu_j}^2 \eta^2}{\sum_{i=1}^n w_{ij}^2 N_{ij}}\right\}, \quad j \in \{1,\ldots,p\}.$$

Proof: Define $g(a,b,c) = ac - b^2$. By assumptions A1 and B2, Taylor expansions of each $E[S_{jr}(t)]$, $r \in \{0,1,2\}$, yield the bound

$$g\left\{E\left[S_{j0}(t)\right], E\left[S_{j1}(t)\right], E\left[S_{j2}(t)\right]\right\} \geq f_j^2(t) g\left\{\phi_{j0}(t), \phi_{j1}(t), \phi_{j2}(t)\right\} - 4Mb_{\mu_j}(1 - b_{\mu_j}),$$

where $\phi_{jr}(t) = \int_{\mathcal{U}_{tj}} u^r K(u)\mathrm{d}u$, $\mathcal{U}_{tj} = [-1,1] \cap \{b_{\mu_j}^{-1}(v-t); v \in \mathcal{T}_j\}$, and M is defined in assumption B2. Furthermore, suppose U and U^+ are random variables with densities K and K^+, respectively, where $K^+(u) = K(u)\left\{\int_0^1 K(u)\mathrm{d}u\right\}^{-1}$ for $u \in [0,1]$. Then it is straightforward to show that, for any $t \in \mathcal{T}_j$, $j \in \{1,\ldots,p\}$, if $b_{\mu_j} < |\mathcal{T}_j|/2$ for all j, then $g(\phi_{j0}(t), \phi_{j1}(t), \phi_{j2}(t)) \geq \tau > 0$, where τ is the minimum of $\mathrm{Var}(U)$ and $\left\{\int_0^1 K(u)\mathrm{d}u\right\}^2 \mathrm{Var}(U^+)$. By assumptions B1 and C3, there is N such that $n > N$ implies $\max_{j\in\{1,\ldots,p\}} b_{\mu_j}|\mathcal{T}_j| < 1/2$ and $\max_{j\in\{1,\ldots,p\}} 4Mb_{\mu_j}(1 - b_{\mu_j}) < m^2\tau/2$, where m is defined in assumption B2. Thus, for $n > N$, $\min_{j\in\{1,\ldots,p\}} \inf_{t\in\mathcal{T}_j} g\left\{E\left[S_{j0}(t)\right], E\left[S_{j1}(t)\right], E\left[S_{j2}(t)\right]\right\} \geq m^2\tau/2$.

Next, by continuity of g', if $\max(|a|,|b|,|c|) < L$ and $\max(|a-a'|,|b-b'|,|c-c'|) < \epsilon < L$, then $|g(a,b,c) - g(a',b',c')| \leq 8L\epsilon$. From assumption B2, $\left|\left|\left|E[S_{jr}(\cdot)]\right|\right|\right|_j < M$, uniformly in n, $j \in \{1,\ldots,p\}$, and $r \in \{0,1,2\}$. Hence, take $\eta = m^2\tau/4$, $L = \max(\eta, M)$, and $\epsilon = \eta/8L$. The results then follows from Lemma 1 since, for $n \geq N$,

$$\Pr\left[\inf_{t\in\mathcal{T}_j}\left\{S_{j0}(t)S_{j2}(t) - S_{j1}^2(t)\right\} < \eta\right] \leq \Pr\left\{\max_{r\in\{0,1,2\}} \left|\left|\left|S_{jr} - E[S_{jr}(\cdot)]\right|\right|\right|_j > \frac{\eta}{8L}\right\}.$$

□

6.1.2. L^2 and Uniform Convergence of Numerator Terms

The next two results provide exponential tail bounds related to each of the terms in (10) and (11) in the L^2 and uniform metrics, respectively. Define $w^*_{nj} = \max_{i\in\{1,\dots,n\}} w_{ij}$, $\mathbb{T}_{ij} = \{T_{ij\ell}\}_{\ell=1}^{N_{ij}}$, and

$$W^r_{ij}(t) = w_{ij}\sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell}-t)\left(\frac{T_{ij\ell}-t}{b_{\mu_j}}\right)^r U_{ij\ell}, \quad r\in\{0,1\}. \tag{12}$$

Lemma 3. *Suppose that assumptions A1, B1–B3, and C1–C3 hold. Then there exist constants C_1, C_2, and C_3 such that, for any $\epsilon > 0$, $j \in \{1,\dots,p\}$, and $r \in \{0,1\}$,*

$$\begin{aligned}\Pr\left(\left\|\sum_{i=1}^n W^r_{ij}\right\|_j > \epsilon\right) &\le C_1\exp\left\{\frac{-C_2\epsilon^2}{\sum_{i=1}^n w^2_{ij}N_{ij}(b^{-1}_{\mu_j}+N_{ij}-1)+b^{-1}_{\mu_j}\overline{w}_{nj}\epsilon}\right\},\\ \Pr\left(\left\|S_{j0}\right\|_j - C_3 > \epsilon\right) &\le C_1\exp\left\{\frac{-C_2\epsilon^2}{b^{-1}_{\mu_j}\sum_{i=1}^n w^2_{ij}N_{ij}+b^{-1/2}_{\mu_j}w^*_{nj}\epsilon}\right\}.\end{aligned} \tag{13}$$

Proof: Conditional on $\mathbb{T}_{ij}$, $W^r_{ij}(t)$ is, for each $t\in\mathcal{T}_j$, a sub-Exponential random variable with parameter at most $w_{ij}\sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell}-t)\theta^*$, where $\theta^* = \overline{\theta}+\sigma$. Thus, for any $\nu\ge 2$, $E\left[\left|W^r_{ij}(t)\right|^\nu \mid \mathbb{T}_{ij}\right] \le 2\nu!(2\theta^* w_{ij})^\nu\left\{\sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell}-t)\right\}^\nu$. Letting $K_\infty = \sup_{|u|\le 1}K(u)$, $K_2 = \int_0^1 K^2(u)\mathrm{d}u$, and M as in assumption B2, it follows that

$$E\left[\left|W^r_{ij}(t)\right|^\nu\right] \le 2M\nu!(2\theta^* w_{ij})^\nu(K_\infty b^{-1}_{\mu_j}N_{ij})^{\nu-2}\left\{K_2N_{ij}b^{-1}_{\mu_j}+MN_{ij}(N_{ij}-1)\right\}. \tag{14}$$

By Jensen's inequality, $\sum_{i=1}^n E\left[\left\|W^r_{ij}\right\|_j^\nu\right] \le \nu L^2_{1nj}L^{\nu-2}_{2nj}/2$, where $L^2_{1nj} = 16M\theta^{*2}|\mathcal{T}_j|\max(K_2,M)\sum_{i=1}^n w^2_{ij}N_{ij}(b^{-1}_{\mu_j}+N_{ij}-1)$ and $L_{2nj} = 2K_\infty\theta^*|\mathcal{T}_j|^{1/2}b^{-1}_{\mu_j}\overline{w}_{nj}$. Hence, Theroem 2.5 of [4] implies the first line of (13).

Finally, direct calculations show that $\left\|\left|E[S_{j0}(\cdot)]\right|\right\|_j \le M$ and $\left\|\left|\mathrm{Var}[S_{j0}(\cdot)]\right|\right\|_j \le MK_2b^{-1}_{\mu_j}\sum_{i=1}^n w^2_{ij}N_{ij}$, whence $E\left[\left\|S_{j0}\right\|_j^2\right] \le |\mathcal{T}_j|\left(K_2Mb^{-1}_{\mu_j}\sum_{i=1}^n w^2_{ij}N_{ij}+M^2\right) \le C_3^2$ for some fixed constant C_3 independent of j, p and n. Now, the functions $w_{ij}K_{b_{\mu_j}}(T_{ij\ell}-\cdot)$ are independent across both i and ℓ. Moreover, $\left\|w_{ij}K_{b_{\mu_j}}(T_{ij\ell}-\cdot)\right\|_j \le K_2^{1/2}b^{-1/2}_{\mu_j}w^*_{nj}$ for any i and ℓ, and $\sum_{i=1}^n\sum_{\ell=1}^{N_{ij}} E\left(\left\|w_{ij}K_{b_{\mu_j}}(T_{ij\ell}-\cdot)\right\|_j^2\right) \le K_2M|\mathcal{T}_j|b^{-1}_{\mu_j}\sum_{i=1}^n w^2_{ij}N_{ij}$ follow from assumptions A1 and B2. Thus, the second line in (13) follows by applying Theorem 2.6 of [4]. □

Lemma 4. *Suppose that assumptions A1, B1, B2, and C1–C3 hold. Then there exist constants $C_1, C_2, C_3 > 0$ such that, for any $\epsilon > 0$ and $j\in\{1,\dots,p\}$,*

$$\begin{aligned}\Pr\left(\left|\sum_{i=1}^n w_{ij}\sum_{\ell=1}^{N_{ij}}\left\{|U_{ij\ell}|-E\left[|U_{ij\ell}|\right]\right\}\right| > \epsilon\right) &\le C_1\exp\left\{\frac{-C_2\epsilon^2}{\sum_{i=1}^n w^2_{ij}N^2_{ij}+\overline{w}_{nj}\epsilon}\right\},\\ \Pr\left(\left\|\left|S_{j0}\right|\right\|_j - C_3 > \epsilon\right) &\le C_1\exp\left\{\frac{-C_2b^2_{\mu_j}\epsilon^2}{\sum_{i=1}^n w^2_{ij}N_{ij}}\right\}.\end{aligned} \tag{15}$$

Proof: Conditional on $\mathbb{T}_{ij}$, $U_{ij\ell}$ are sub-Exponential random variables with parameter at most $\theta^* = \overline{\theta}+\sigma$. Hence, for any $\nu\ge 2$, by applying Jensen's inequality,

$$\sum_{i=1}^n E\left[\left(w_{ij}\sum_{\ell=1}^{N_{ij}}|U_{ij\ell}|\right)^\nu\right] \le \sum_{i=1}^n w^\nu_{ij}N^\nu_{ij}\left(\frac{1}{N_{ij}}\sum_{\ell=1}^{N_{ij}}E\left[|U_{ij\ell}|^\nu\right]\right) \le \frac{\nu!}{2}\left(2\theta^*\overline{w}_{nj}\right)^{\nu-2}\left(16\theta^{*2}\sum_{i=1}^n w^2_{ij}N^2_{ij}\right).$$

Then, by Theorem 2.5 of [4], the first line of (15) is established. Finally, as previously observed, $E\left[S_{j0}(t)\right] \le M$, so one may take $C_3 = M$. The third line of (15) then follows by applying the result of Lemma 1. □

6.2. Proofs of Results in Section 3

Proof of Theorem 1: The proof will be given for an FR design. For an SR design, the proof is simplified by the fact that S_{jr} does not depend on j, thus necessitating less stringent requirements on the bandwidth as outlined in assumption C3. From (9),

$$\left\|\mu_j - \hat{\mu}_j\right\|_j \leq \left[\inf_{t\in\mathcal{T}_j}\left\{S_{j0}(t)S_{j2}(t) - S_{j1}^2(t)\right\}\right]^{-1}\left(\left|\!\left|\!\left|S_{j2}\right|\!\right|\!\right|_j\left\|R_{j0} - \mu_j S_{j0} - b_{\mu_j}\mu_j' S_{j1}\right\|_j + \left|\!\left|\!\left|S_{j1}\right|\!\right|\!\right|_j\left\|R_{j1} - \mu_j S_{j1} - b_{\mu_j}\mu_j' S_{j2}\right\|_j\right).$$

Let C_1, C_2, C_3, N, and η satisfy the results of Lemmas 2 and 3 simultaneously. Then Lemma 2 implies

$$\begin{aligned}\Pr\left[\min_{j\in\{1,\ldots,p\}}\inf_{t\in\mathcal{T}_j}\left\{S_{j0}(t)S_{j2}(t) - S_{j1}^2(t)\right\} < \eta\right] &\leq p\max_{j\in\{1,\ldots,p\}}\Pr\left(\inf_{t\in\mathcal{T}_j}\left\{S_{j0}(t)S_{j2}(t) - S_{j1}^2(t)\right\} < \eta\right)\\ &\leq C_1 p\max_{j\in\{1,\ldots,p\}}\exp\left\{-\frac{C_2 b_{\mu_j}^2\eta^2}{\sum_{i=1}^n w_{ij}^2 N_{ij}}\right\} = C_1 p^{1-C_2\eta^2\left\{\log(p)\max_{j\in\{1,\ldots,p\}} b_{\mu_j}^{-2}\sum_{i=1}^n w_{ij}^2 N_{ij}\right\}^{-1}},\end{aligned}$$

which converges to 0 as $n\to\infty$ by assumption C3. Similarly, since $\max_{j\in\{1,\ldots,p\}}\left|\!\left|\!\left|E[S_{jr}(\cdot)]\right|\!\right|\!\right|_j < \infty$ for $r\in\{0,1,2\}$, Lemma 1 ensures the existence of some $R>0$ such that

$$\Pr\left(\max_{j\in\{1,\ldots,p\}}\left|\!\left|\!\left|S_{jr}\right|\!\right|\!\right|_j > R\right) \leq C_1 p^{1-C_2R^2\left\{\log(p)\max_{j=1,\ldots,p} b_{\mu_j}^{-2}\sum_{i=1}^n w_{ij}^2 N_{ij}\right\}^{-1}},$$

which again converges to 0 as $n\to\infty$ by (C3). Hence, $\max_{j\in\{1,\ldots,p\}}\left[\inf_{t\in\mathcal{T}_j}\left\{S_{j0}(t)S_{j2}(t) - S_{j1}^2(t)\right\}\right]^{-1} = O_P(1)$ and $\max_{j\in\{1,\ldots,p\}}\left|\!\left|\!\left|S_{jr}\right|\!\right|\!\right|_j = O_P(1)$, $r\in\{0,1,2\}$.

Next, let $R>0$ again be arbitrary and define

$$\begin{aligned}a_{n1} &= \max_{j\in\{1,\ldots,p\}}\left[\left\{\sum_{i=1}^n w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1)\right\}^{1/2} + \{\log(p)\}^{1/2} b_{\mu_j}^{-1}\overline{w}_{nj}\right],\\ a_{n2} &= \max_{j\in\{1,\ldots,p\}}\left[\left(b_{\mu_j}^{-1}\sum_{i=1}^n w_{ij}^2 N_{ij}\right)^{1/2} + \{\log(p)\}^{1/2} b_{\mu_j}^{-1/2} w_{nj}^*\right].\end{aligned}$$

Then the union bound and Lemma 3 imply that

$$\begin{aligned}\Pr\left(\max_{j\in\{1,\ldots,p\}}\left\|\sum_{i=1}^n W_{ij}^0\right\|_j > Ra_{n1}\{\log(p)\}^{1/2}\right) &\leq p\max_{j\in\{1,\ldots,p\}}\Pr\left(\left\|\sum_{i=1}^n W_{ij}^0\right\|_j > Ra_{n1}\{\log(p)\}^{1/2}\right)\\ &\leq C_1 p\max_{j\in\{1,\ldots,p\}}\exp\left\{-\frac{R^2 a_{n1}^2\log(p)}{\sum_{i=1}^n w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1) + b_{\mu_j}^{-1}\overline{w}_{nj}Ra_{n1}\{\log(p)\}^{1/2}}\right\} \leq C_1 p^{1-\frac{C_2^2R^2}{1+R}}.\end{aligned} \tag{16}$$

By similar reasoning,

$$\Pr\left(\max_{j\in\{1,\ldots,p\}}\left\|S_{j0}\right\|_j - C_3 > Ra_{n2}\{\log(p)\}^{1/2}\right) \leq C_1 p^{1-\frac{C_2^2R^2}{1+R}}.$$

In addition, using assumptions A1 and B3, there is a constant C_4 such that

$$\left|\sum_{i=1}^n w_{ij}\sum_{\ell=1}^{N_{ij}} K_{b_{\mu_j}}(T_{ij\ell} - t)(T_{ij\ell} - t)^2\int_0^1 \mu_j''(t_v(T_{ij\ell} - t))(1-v)\mathrm{d}v\right| \leq C_4 b_{\mu_j}^2 S_{j0}(t).$$

Thus, by the definition of W_{ij}^0 and (16),

$$\begin{aligned}\max_{j\in\{1,\ldots,p\}}\left\|R_{j0} - \mu_j S_{j0} - b_{\mu_j}\mu_j' S_{j1}\right\|_j &\leq \max_{j\in\{1,\ldots,p\}}\left\|W_{ij}^0\right\|_j + \max_{j\in\{1,\ldots,p\}} b_{\mu_j}^2\left\|S_{j0}\right\|_j\\ &= O_P\left(\{\log(p)\}^{1/2}a_{n1} + \left[C_3 + \{\log(p)\}^{1/2}a_{n2}\right]\max_{j\in\{1,\ldots,p\}} b_{\mu_j}^2\right).\end{aligned} \tag{17}$$

Since $\max_{j\in\{1,\dots,p\}} b_{\mu_j} \to 0$ by assumption C3 and $w^*_{nj} \leq \overline{w}_{nj}$, it follows that

$$\{\log(p)\}^{1/2} a_{n2} \leq \left(\log(p) \max_{j\in\{1,\dots,p\}} b_{\mu_j}^{-1} \sum_{i=1}^{n} w_{ij}^2 N_{ij} \right)^{1/2} + \log(p) \max_{j\in\{1,\dots,p\}} b_{\mu_j}^{-1} \overline{w}_{nj} \to 0.$$

Thus, the rate in (17) matches the one given in the theorem statement. The same argument shows that $\max_{j\in\{1,\dots,p\}} \left\| R_{j1} - \mu_j S_{j1} - b_{\mu_j} \mu'_j S_{j2} \right\|_j$ has the same rate, completing the proof. □

Proof of Corollary 1: For the OBS scheme, $\sum_{i=1}^{n} w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1) = (n\overline{N}_j b_{\mu_j})^{-1} + (\overline{N}_{j^2} - \overline{N}_j) n^{-1} \left(\overline{N}_j\right)^{-2}$ and $b_{\mu_j}^{-1} \overline{w}_{nj} = N_j^{+} (n\overline{N}_j b_{\mu_j})^{-1}$. For the SUBJ scheme, $\sum_{i=1}^{n} w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1) = (nN_j^H b_{\mu_j})^{-1} + n^{-1} \left(1 - \left\{N_j^H\right\}^{-1}\right)$ and $b_{\mu_j}^{-1} \overline{w}_{nj} = (nb_{\mu_j})^{-1}$. Hence, the rates given in the corollary statement follow immediately from the general rate of Theorem 1. □

Proof of Corollary 2: The result is given for part (a) of (i). The remaining rates can be derived in a similar manner, and the details are omitted. Under the conditions given in part (a), $\max_{j\in\{1,\dots,p\}} b_{\mu_j}^2 \leq \left\{\log(p)(n\overline{N}_{\min})^{-1}\right\}^{2/5}$, $\min_{j\in\{1,\dots,p\}} b_{\mu_j} \overline{N}_j \geq \left\{\overline{N}_{\min}^4 \log(p) n^{-1}\right\}^{1/5}$, and $\min_{j\in\{1,\dots,p\}} b_{\mu_j} \geq \left\{\log(p)(n\overline{N}_{\max})^{-1}\right\}^{1/5}$. The given rate then follows by the condition on $\overline{N}_{\max}$. □

Proof of Corollary 3: The proof follows the same logic as the proof of Corollary 1 and the details are omitted. □

Proof of Theorem 2: Again, the proof is only given for the case of an FR design for the same reason stated at the beginning of the proof of Theorem 1. From (9),

$$\begin{aligned} \left\| \mu_j - \hat{\mu}_j \right\|_j \leq \left[\inf_{t\in\mathcal{T}_j} \left\{ S_{j0}(t) S_{j2}(t) - S_{j1}^2(t) \right\} \right]^{-1} \Bigg(& \left\| S_{j2} \right\|_j \left\| R_{j0} - \mu_j S_{j0} - b_{\mu_j} \mu'_j S_{j1} \right\|_j \\ & + \left\| S_{j1} \right\|_j \left\| R_{j1} - \mu_j S_{j1} - b_{\mu_j} \mu'_j S_{j2} \right\|_j \Bigg). \end{aligned}$$

The same arguments used in the proof of Theorem 1 imply that $\max_{j\in\{1,\dots,p\}} \left[\inf_{t\in\mathcal{T}_j} \left\{ S_{j0}(t) S_{j2}(t) - S_{j1}^2(t) \right\} \right]^{-1} = O_P(1)$ and $\max_{j\in\{1,\dots,p\}} \left\| S_{jr} \right\|_j = O_P(1)$, $r \in \{0, 1, 2\}$. Thus, the uniform rates of the terms in (10) and (11) are sufficient.

For any $\delta > 0$ and for each j, let $\chi_j(\delta)$ be a discrete uniform grid for $\mathcal{T}_j$ with spacing at most δ, and let L be a constant, guaranteed by assumption B1, so that $\max_{j\in\{1,\dots,p\}} |\chi_j(\delta)| \leq L\delta^{-1}$, with $|\chi_j(\delta)|$ denoting the number of elements in the grid. Given the decomposition in (10), consider the uniform convergence of the each term on the right-hand side of this equation. Beginning with the first term,

$$\left\| \sum_{i=1}^{n} W_{ij}^r \right\|_j \leq \sup_{t\in\chi_j(\delta)} \left| \sum_{i=1}^{n} W_{ij}^r(t) \right| + \sup_{|s-t|\leq\delta} \left| \sum_{i=1}^{n} \left\{ W_{ij}^r(s) - W_{ij}^r(t) \right\} \right|, \tag{18}$$

where W_{ij}^r is defined in (12) for $r \in \{0, 1\}$. By (14), for any $\nu \geq 2$, $j \in \{1, \dots, p\}$ and $t \in \mathcal{T}_j$, $\sum_{i=1}^{n} E\left[|W_{ij}^r(t)|^\nu \right] \leq \frac{\nu!}{2} L_{1nj}^2 L_{2nj}^{\nu-2}$, where $L_{1nj}^2 = 16M\theta^{*2} \max(K_2, M) \sum_{i=1}^{n} w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1)$ and $L_{2nj} = 2K_\infty \theta^* b_{\mu_j}^{-1} \overline{w}_{nj}$. Thus, there exist constants C_1 and C_2 such that, for any $\epsilon, \delta > 0$,

$$\Pr\left\{ \max_{t\in\chi_j(\delta)} \left| \sum_{i=1}^{n} W_{ij}^r(t) \right| > \epsilon \right\} \leq L\delta^{-1} C_1 \exp\left\{ -\frac{C_2 \epsilon^2}{\sum_{i=1}^{n} w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1) + b_{\mu_j}^{-1} \overline{w}_{nj} \epsilon} \right\}. \tag{19}$$

Putting $a_{n1}(\delta) = \max_{j\in\{1,\dots,p\}} \left[\left\{ \sum_{i=1}^{n} w_{ij}^2 N_{ij}(b_{\mu_j}^{-1} + N_{ij} - 1) \right\}^{1/2} + \left\{ \log(p\delta^{-1}) \right\}^{1/2} b_{\mu_j}^{-1} \overline{w}_{nj} \right]$, for any $R > 0$, (19) implies

$$\Pr\left[\max_{j\in\{1,\dots,p\}} \max_{t\in\chi_j(\delta)} \left| \sum_{i=1}^{n} W_{ij}^r(t) \right| > R a_{n1}(\delta) \{\log(p\delta^{-1})\}^{1/2} \right] \leq C_1 L \left(p\delta^{-1}\right)^{1 - \frac{C_2 R^2}{1+R}}. \tag{20}$$

With $\delta = \delta_n = n^{-\alpha}$, the Borel-Cantelli lemma and (20), imply that, almost surely,

$$\max_{j\in\{1,\ldots,p\}} \max_{t\in\chi_j(\delta_n)} \left| \sum_{i=1}^{n} W_{ij}^r(t) \right| = O\left(\max_{j\in\{1,\ldots,p\}} \left[\left\{ \log(p\vee n) \sum_{i=1}^{n} w_{ij}^2 N_{ij} (b_{\mu_j}^{-1} + N_{ij} - 1) \right\}^{1/2} + \log(p\vee n) b_{\mu_j}^{-1} \overline{w}_{nj} \right] \right). \tag{21}$$

Next, with L_K being the Lipschitz constant in assumption A2 and $C_4 = \limsup_{n\to\infty} \max_{j\in\{1,\ldots,p\}} E\left[|U_{ij\ell}|\right] < \infty$, it follows that

$$\sup_{|s-t|\le\delta} \left| \sum_{i=1}^{n} \left\{ W_{ij}^r(s) - W_{ij}^r(t) \right\} \right| \le L_K \delta b_{\mu_j}^{-2} \sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} \left| U_{ij\ell} \right| \le L_K \delta b_{\mu_j}^{-2} \left\{ C_4 + \sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} \left(\left| U_{ij\ell} \right| - E\left[\left| U_{ij\ell} \right| \right] \right) \right\}.$$

Set $a_{n2} = \max_{j\in\{1,\ldots,p\}} \left[\left(\sum_{i=1}^{n} w_{ij}^2 N_{ij}^2 \right)^{1/2} + \{\log(p)\}^{1/2} \overline{w}_{nj} \right]$. Then Lemma 4 implies that there are constants $C_1, C_2 > 0$ such that, for any $R > 0$,

$$\Pr\left[\max_{j\in\{1,\ldots,p\}} \left| \sum_{i=1}^{n} w_{ij} \sum_{\ell=1}^{N_{ij}} \left(|U_{ij\ell}| - E\left[\left| U_{ij\ell} \right| \right] \right) \right| > R a_{n2} \{\log(p)\}^{1/2} \right] \le C_1 p^{1 - \frac{C_2 R^2}{1+R}}.$$

By the Borel-Cantelli Lemma, assumption C4, and again taking $\delta = \delta_n = n^{-\alpha}$, it follows that, almost surely,

$$\begin{aligned} \max_{j\in\{1,\ldots,p\}} \sup_{|s-t|<\delta_n} \left| \sum_{i=1}^{n} \{ W_{ij}^r(s) - W_{ij}^r(t) \} \right| &= O\left(n^{-\alpha} \max_{j\in\{1,\ldots,p\}} b_{\mu_j}^{-2} \right) \\ &= o\left(\max_{j\in\{1,\ldots,p\}} \left[\left\{ \log(p\vee n) \sum_{i=1}^{n} w_{ij}^2 N_{ij} (b_{\mu_j}^{-1} + N_{ij} - 1) \right\}^{1/2} + \log(p\vee n) b_{\mu_j}^{-1} \overline{w}_{nj} \right] \right). \end{aligned} \tag{22}$$

Then (18), (21) and (22) together imply the result since, almost surely,

$$\max_{j\in\{1,\ldots,p\}} \left\| \sum_{i=1}^{n} W_{ij}^r \right\|_j = O\left(\max_{j\in\{1,\ldots,p\}} \left[\left\{ \log(p\vee n) \sum_{i=1}^{n} w_{ij}^2 N_{ij} (b_{\mu_j}^{-1} + N_{ij} - 1) \right\}^{1/2} + \log(p\vee n) b_{\mu_j}^{-1} \overline{w}_{nj} \right] \right).$$

□

Proofs of Corollaries 4–6: The proofs are similar to those of Corollaries 1–3, and are omitted. □

6.3. Auxiliary Lemmas For Covariance Estimation

For $q, r \in \{0, 1, 2\}$, and $j, k \in \{1, \ldots, p\}$, define

$$\begin{aligned} S_{jkqr}(s,t) &= \sum_{i=1}^{n} v_{ijk} \sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}\left(T_{ij\ell} - s\right) K_{b_{\gamma_k}}\left(T_{ikm} - t\right) \left(\frac{T_{ij\ell} - s}{b_{\gamma_j}} \right)^q \left(\frac{T_{ikm} - t}{b_{\gamma_k}} \right)^r, \\ R_{jkqr}(s,t) &= \sum_{i=1}^{n} v_{ijk} \sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}\left(T_{ij\ell} - s\right) K_{b_{\gamma_k}}\left(T_{ikm} - t\right) \left(\frac{T_{ij\ell} - s}{b_{\gamma_j}} \right)^q \left(\frac{T_{ikm} - t}{b_{\gamma_k}} \right)^r Z_{ijk\ell m}. \end{aligned} \tag{23}$$

Additionally, dropping the functional arguments s and t for the component elements defined above in (23), set

$$Q_{jk0} = S_{jk20} S_{jk02} - S_{jk11}^2, \quad Q_{jk1} = S_{jk10} S_{jk02} - S_{jk01} S_{jk11}, \quad Q_{jk2} = S_{jk10} S_{jk11} - S_{jk01} S_{jk20}. \tag{24}$$

As mentioned in Section 4, the mean is assumed to be known; that is, $\hat{\gamma}_{jk}(s,t) = \hat{\beta}_0$, where $\hat{\beta}_0$ is computed as in (4) with $Z_{ijk\ell m} = \{Y_{ij\ell} - \mu_j(T_{ij\ell})\}\{Y_{ikm} - \mu_k(T_{ikm})\}$ instead of the true empirical version in which the $Y_{ij\ell}$ and Y_{ikm} are centered with respect to their estimated means. Then the error in covariance estimation can be expressed as

$$\begin{aligned}\hat{\gamma}_{jk}(s,t) - \gamma_{jk}(s,t) = &\left(Q_{jk0}S_{jk00} - Q_{jk1}S_{jk10} + Q_{jk2}S_{jk01}\right)^{-1} \\ &\times \Bigg[Q_{jk0}\left\{R_{jk00} - \gamma_{jk}(s,t)S_{jk00} - b_{\gamma_j}\frac{\partial\gamma_{jk}}{\partial s}(s,t)S_{jk10} - b_{\gamma_k}\frac{\partial\gamma_{jk}}{\partial t}(s,t)S_{jk01}\right\} \\ &- Q_{jk1}\left\{R_{jk10} - \gamma_{jk}(s,t)S_{jk10} - b_{\gamma_j}\frac{\partial\gamma_{jk}}{\partial s}(s,t)S_{jk20} - b_{\gamma_k}\frac{\partial\gamma_{jk}}{\partial t}(s,t)S_{jk11}\right\} \\ &+ Q_{jk2}\left\{R_{jk01} - \gamma_{jk}(s,t)S_{jk01} - b_{\gamma_j}\frac{\partial\gamma_{jk}}{\partial s}(s,t)S_{jk11} - b_{\gamma_k}\frac{\partial\gamma_{jk}}{\partial t}(s,t)S_{jk02}\right\}\Bigg].\end{aligned} \tag{25}$$

The terms in square brackets in the last three lines of (25) can be broken down further as follows. Define

$$U_{ijk\ell m} = U_{ij}(T_{ij\ell})U_{ik}(T_{ikm}) - \gamma_{jk}(s,t) + U_{ij}(T_{ij\ell})\epsilon_{ikm} + U_{ik}(T_{ikm})\epsilon_{ij\ell} + \epsilon_{ij\ell}\epsilon_{ikm}$$

and define $\tilde{K}_b(u) = K_b(u)(u/b)$. Furthermore, set

$$B_{jk\ell m}(s,t) = \gamma_{jk}(T_{ij\ell}, T_{ikm}) - \gamma_{jk}(s,t) - b_{\gamma_j}(T_{ij\ell} - s)\frac{\partial\gamma_{jk}}{\partial s}(s,t) - b_{\gamma_k}(T_{ij\ell} - t)\frac{\partial\gamma_{jk}}{\partial t}(s,t).$$

Then

$$\begin{aligned}R_{jk00} - \gamma_{jk}(s,t)S_{jk00} - b_{\gamma_j}\frac{\partial\gamma_{jk}}{\partial s}S_{jk10} - b_{\gamma_k}\frac{\partial\gamma_{jk}}{\partial t}S_{jk01} = &\sum_{i=1}^{n} v_{ijk}\sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - s)K_{b_{\gamma_k}}(T_{ikm} - t)U_{ijk\ell m} \\ &+ \sum_{i=1}^{n} v_{ijk}\sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - s)K_{b_{\gamma_k}}(T_{ikm} - t)B_{jk\ell m}(s,t),\end{aligned} \tag{26}$$

$$\begin{aligned}R_{jk10} - \gamma_{jk}(s,t)S_{jk10} - b_{\gamma_j}\frac{\partial\gamma_{jk}}{\partial s}S_{jk20} - b_{\gamma_k}\frac{\partial\gamma_{jk}}{\partial t}S_{jk11} = &\sum_{i=1}^{n} v_{ijk}\sum_{(\ell,m)\in\mathcal{I}_{ijk}} \tilde{K}_{b_{\gamma_j}}(T_{ij\ell} - s)K_{b_{\gamma_k}}(T_{ikm} - t)U_{ijk\ell m} \\ &+ \sum_{i=1}^{n} v_{ijk}\sum_{(\ell,m)\in\mathcal{I}_{ijk}} \tilde{K}_{b_{\gamma_j}}(T_{ij\ell} - s)K_{b_{\gamma_k}}(T_{ikm} - t)B_{jk\ell m}(s,t),\end{aligned} \tag{27}$$

$$\begin{aligned}R_{jk01} - \gamma_{jk}(s,t)S_{jk01} - b_{\gamma_j}\frac{\partial\gamma_{jk}}{\partial s}S_{jk11} - b_{\gamma_k}\frac{\partial\gamma_{jk}}{\partial t}S_{jk02} = &\sum_{i=1}^{n} v_{ijk}\sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - s)\tilde{K}_{b_{\gamma_k}}(T_{ikm} - t)U_{ijk\ell m} \\ &+ \sum_{i=1}^{n} v_{ijk}\sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - s)\tilde{K}_{b_{\gamma_k}}(T_{ikm} - t)B_{jk\ell m}(s,t).\end{aligned} \tag{28}$$

6.3.1. Uniform Convergence of S_{jkqr}, $q, r \in \{0,1,2\}$ and $0 \le q + r \le 2$

To begin, a result similar to Lemma 1 will be established, however the proof is much more involved. Specifically, an analog of Lemma 1.1 of [30] is proved that is suitable for the quantities S_{jkqr} in (23) that are not simply weighted kernel density estimators.

Lemma 5. *Let G_1 and G_2 be arbitrary cumulative distribution functions. For any $n \in \mathbb{N}$, let $N_{ij} \in \mathbb{N}$, $i \in \{1,\ldots,n\}$, $j \in \{1,2\}$, be arbitrary and consider independent arrays of random variables $\mathbb{T}_j = \left\{T_{ij\ell} : \ell \in \{1,\ldots,N_{ij}\},\ i \in \{1,\ldots,n\}\right\}$,*

where the elements of $\mathbb{T}_j$ are independent and identically distributed according to G_j, $j \in \{1,2\}$. Furthermore, let v_i, $i \in \{1,\ldots,n\}$, be arbitrary constants and define $\omega_1^2 = \sum_{i=1}^n v_i^2 N_{i1}^2 N_{i2}$, $\omega_2^2 = \sum_{i=1}^n v_i^2 N_{i1} N_{i2}^2$, and $\overline{\omega} = \max_{j\in\{1,2\}} \omega_j$. Set $\mathbb{A}_n(s,t) = \sum_{i=1}^n v_i \sum_{\ell=1}^{N_{i1}} \sum_{m=1}^{N_{i2}} \{\mathbf{1}(T_{i1\ell} \le s)\mathbf{1}(T_{i2m} \le t) - G_1(s)G_2(t)\}$. Then, for any $\epsilon > 0$,

$$\Pr\left\{\sup_{s,t\in\mathbb{R}} |\mathbb{A}_n(s,t)| > \epsilon\overline{\omega}\right\} \le \left(3 + \frac{3\sqrt{2\pi}\epsilon}{2} + \frac{2\pi\epsilon^2}{9}\right) e^{-\epsilon^2/288}.$$

Proof: First, note that $\mathbb{A}_n(s,t) = \mathbb{A}_{n0}(s,t) + \mathbb{A}_{n1}(s,t) + \mathbb{A}_{n2}(s,t)$, where

$$\begin{aligned}
\mathbb{A}_{n0}(s,t) &= \sum_{i=1}^n v_i \sum_{\ell=1}^{N_{i1}} \sum_{m=1}^{N_{i2}} \{\mathbf{1}(T_{i1\ell} \le s) - G_1(s)\}\{\mathbf{1}(T_{i2m} \le s) - G_2(t)\}, \\
\mathbb{A}_{n1}(s,t) &= G_2(t) \sum_{i=1}^n v_i N_{i2} \sum_{\ell=1}^{N_{i1}} \{\mathbf{1}(T_{i1\ell} \le s) - G_1(s)\}, \\
\mathbb{A}_{n2}(s,t) &= G_1(s) \sum_{i=1}^n v_i N_{i1} \sum_{m=1}^{N_{i2}} \{\mathbf{1}(T_{i2m} \le t) - G_2(t)\}.
\end{aligned}$$

As $\sup_s |G_j(s)| = 1$ for $j \in \{1,2\}$, it follows immediately from Lemma 1.1 of [30] that, for any $\epsilon > 0$,

$$\Pr\left\{\sup_{s,t} \left|\mathbb{A}_{nj}(s,t)\right| > \epsilon\omega_j\right\} \le \left(1 + 2\sqrt{2\pi}\epsilon\right) e^{-\epsilon^2/8}, \quad j \in \{1,2\}. \tag{29}$$

To uniformly bound $\mathbb{A}_{n0}$, it will be established that, for any $\lambda > 0$ and $\underline{\omega} = \min_{j\in\{1,2\}} \omega_j$,

$$E\left[\exp\left\{\lambda \sup_{s,t} |\mathbb{A}_{n0}(s,t)|\right\}\right] \le 1 + 64\sqrt{2\pi}\lambda\underline{\omega} + 16\sqrt{2\pi}\lambda\underline{\omega}\left(1 + 32\sqrt{2\pi}\lambda\underline{\omega}\right) e^{32\lambda^2\underline{\omega}^2}. \tag{30}$$

Once established, (30) implies that, for any $\epsilon > 0$ and $\lambda = \epsilon(64\underline{\omega})^{-1}$,

$$\Pr\left\{\sup_{s,t} |\mathbb{A}_{n0}(s,t)| > \epsilon\underline{\omega}\right\} \le e^{-\lambda\underline{\omega}\epsilon} E\left[\exp\left\{\lambda \sup_{s,t} \mathbb{A}_{n0}(s,t)\right\}\right] \le \left(1 + \frac{5\sqrt{2\pi}\epsilon}{4} + \frac{\pi\epsilon^2}{2}\right) e^{-\epsilon^2/128}. \tag{31}$$

Together, (29) and (31) imply that

$$\begin{aligned}
\Pr\left\{\sup_{s,t\in\mathbb{R}} |\mathbb{A}_n(s,t)| > \epsilon\overline{\omega}\right\} &\le \Pr\left\{\sup_{s,t} |\mathbb{A}_{n0}(s,t)| > \frac{2\epsilon\underline{\omega}}{3}\right\} + \sum_{j=1}^2 \Pr\left\{\sup_{s,t} \left|\mathbb{A}_{nj}(s,t)\right| > \frac{\epsilon\omega_j}{6}\right\} \\
&\le \left(1 + \frac{5\sqrt{2\pi}\epsilon}{6} + \frac{2\pi\epsilon^2}{9}\right) e^{-\epsilon^2/288} + 2\left(1 + \frac{\sqrt{2\pi}\epsilon}{3}\right) e^{-\epsilon^2/288} = \left(3 + \frac{3\sqrt{2\pi}\epsilon}{2} + \frac{2\pi\epsilon^2}{9}\right) e^{-\epsilon^2/288}.
\end{aligned}$$

Thus, it remains only to prove (30).

Let $(\mathbb{T}_1', \mathbb{T}_2')$ represent an independent copy of $(\mathbb{T}_1, \mathbb{T}_2)$, leading to $\mathbb{A}_{n0}'(s,t)$ in analogy to $\mathbb{A}_{n0}(s,t)$. Let $e_{i1\ell}$ and e_{i2m}, $i \in \{1,\ldots,n\}$, $\ell \in \{1,\ldots,N_{i1}\}$, $m \in \{1,\ldots,N_{i2}\}$ denote independent and identically distributed Radamacher variables, that is, $\Pr(e_{i1\ell} = 1) = \Pr(e_{i1\ell} = -1) = 1/2$. By symmetry, simple calculations show that $\mathbb{A}_{n0}(s,t) - \mathbb{A}_{n0}'(s,t) \overset{D}{=} Y_1(s,t) + Y_2(s,t)$, where equality of distribution holds at the process level and

$$\begin{aligned}
Y_1(s,t) &= \sum_{i=1}^n v_i \sum_{\ell=1}^{N_{i1}} \sum_{m=1}^{N_{i2}} \left\{\mathbf{1}(T_{i1\ell} \le s) - \mathbf{1}(T_{i1\ell}' \le s)\right\}\{\mathbf{1}(T_{i2m} \le t) - G_2(t)\} e_{i1\ell} \\
Y_2(s,t) &= \sum_{i=1}^n v_i \sum_{\ell=1}^{N_{i1}} \sum_{m=1}^{N_{i2}} \{\mathbf{1}(T_{i1\ell} \le s) - G_1(s)\}\left\{\mathbf{1}(T_{i2m} \le t) - \mathbf{1}(T_{i2m}' \le t)\right\} e_{i2m}.
\end{aligned}$$

Then, for $\lambda > 0$, by applying Jensen's inequality, the triangle inequality, and Cauchy-Schwarz,

$$E\left[\exp\left\{\lambda \sup_{s,t}|\mathbb{A}_{n0}(s,t)|\right\}\right] \leq E\left(\exp\left\{\sup_{s,t}|\mathbb{A}_{n0}(s,t)-\mathbb{A}'_{n0}(s,t)|\right\}\right) \leq E\left(\exp\{\lambda \sup_{s,t}|Y_1(s,t)|\}\exp\{\lambda \sup_{s,t}|Y_2(s,t)|\}\right)$$
$$\leq \left(E\left[\exp\left\{2\lambda \sup_{s,t}|Y_1(s,t)|\right\}\right]E\left[\exp\left\{2\lambda \sup_{s,t}|Y_2(s,t)|\right\}\right]\right)^{1/2}. \tag{32}$$

Next, define $v_{i\ell}(t) = v_i \sum_{m=1}^{N_{i2}}\{\mathbf{1}(T_{i2m} \leq t) - G_2(t)\}$, let $N_1^S = \sum_{i=1}^n N_{i1}$ and suppose that $\{T_{1\ell'}\}_{\ell'=1}^{N_1^S}$ is a non-decreasing ordering of $\mathbb{T}_1$. If ℓ', i and ℓ are such that $T_{1\ell'} = T_{i1\ell}$, set $v_{\ell'}(t) = v_{i\ell}(t)$ and $e_{1\ell'} = e_{i1\ell}$. Then applications of Cauchy-Schwarz and the triangle inequality, along with the fact that $\mathbb{T}_1 \overset{D}{=} \mathbb{T}'_1$, yield

$$E\left[\exp\left\{2\lambda \sup_{s,t}|Y_1(s,t)|\right\}\right] \leq E\left[\exp\left\{4\lambda \sup_{s,t}\left|\sum_{i=1}^n\sum_{\ell=1}^{N_{i1}} v_{i\ell}(t)\mathbf{1}(T_{i1\ell} \leq s)e_{i1\ell}\right|\right\}\right] = E\left[\exp\left\{4\lambda \max_{\ell''\in\{1,\ldots,N_1^S\}}\sup_t\left|\sum_{\ell'=1}^{\ell''} v_{\ell'}(t)e_{1\ell'}\right|\right\}\right]$$
$$= 1 + \int_1^\infty \Pr\left(\max_{\ell''\in\{1,\ldots,N_1^S\}}\sup_t\left|\sum_{\ell'=1}^{\ell''} v_{\ell'}(t)e_{1\ell'}\right| > \frac{\log(\gamma)}{4\lambda}\right)\mathrm{d}\gamma. \tag{33}$$

Let $\eta > 0$. Then, by Levy's inequality, applied conditionally on $\mathbb{T}_2$,

$$\Pr\left\{\max_{\ell''\in\{1,\ldots,N_1^S\}}\sup_t\left|\sum_{\ell'=1}^{\ell''} v_{\ell'}(t)e_{1\ell'}\right| > \eta\right\} \leq 2\Pr\left[\sup_t\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}} e_{i1\ell}\{\mathbf{1}(T_{i2m} \leq t) - G_2(t)\}\right| > \eta\right]. \tag{34}$$

To bound this probability, the moment generating function will again be bounded by symmetrization. Specifically, for any $\delta > 0$, apply Jensen's (conditionally on the $e_{i1\ell}$) to obtain

$$E\left[\exp\left\{\delta\sup_t\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}} e_{i1\ell}\{\mathbf{1}(T_{i2m} \leq t) - G_2(t)\}\right|\right\}\right] \leq E\left[\exp\left\{2\delta\sup_t\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}}\mathbf{1}(T_{i2m} \leq t)e_{i1\ell}e_{i2m}\right|\right\}\right]. \tag{35}$$

Let $N_2^S = \sum_{i=1}^n N_{i2}$ and let $\{T_{2m'}\}_{m'=1}^{N_2^S}$ be a non-decreasing ordering of $\mathbb{T}_2$. Define $v_{im} = v_i\sum_{\ell=1}^{N_{i1}} e_{i1\ell}$ and, if $T_{2m'} = T_{i2m}$, set $v'_{m'} = v_{im}$ and $e_{2m'} = e_{i2m}$. Then, for any $\tau > 0$, another application of Levy's inequality combined with Hoeffding's inequality yields

$$\Pr\left\{\sup_t\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}}\mathbf{1}(T_{i2m} \leq t)e_{i1\ell}e_{i2m}\right| > \tau\right\} = \Pr\left(\max_{m''\in\{1,\ldots,N_2^S\}}\left|\sum_{m'=1}^{m''} v'_{m'}e_{2m'}\right| > \tau\right)$$
$$\leq 2\Pr\left(\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}} e_{i1\ell}e_{i2m}\right| > \tau\right) \leq 4\exp\left\{-\frac{\tau^2}{2\underline{\omega}}\right\}. \tag{36}$$

Combining (35) and (36), it can be concluded that

$$E\left[\exp\left\{\delta\sup_t\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}} e_{i1\ell}\{\mathbf{1}(T_{i2m} \leq t) - G_2(t)\}\right|\right\}\right] = 1 + \int_1^\infty \Pr\left(\sup_t\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}}\mathbf{1}(T_{i2m} \leq t)e_{i1\ell}e_{i2m}\right| > \frac{\log(\gamma)}{2\delta}\right)\mathrm{d}\gamma$$
$$\leq 1 + 8\delta\underline{\omega}\int_0^\infty e^{2\delta\underline{\omega}\tau}e^{-\tau^2/2}\mathrm{d}\tau \leq 1 + 8\delta\underline{\omega}\sqrt{2\pi}e^{2\delta^2\underline{\omega}^2},$$

where the change of variables $\log(\gamma) = 2\delta\underline{\omega}\tau$ and the integral bound $\int_0^\infty e^{rx}e^{-x^2/2}\mathrm{d}x \leq \sqrt{2\pi}e^{r^2/2}$ for $r > 0$ have been used. Hence, for any $\eta > 0$, setting $\delta = \eta/4\underline{\omega}$ yields the bound

$$\Pr\left[\sup_t\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}} e_{i1\ell}\{\mathbf{1}(T_{i2m} \leq t) - G_2(t)\}\right| > \eta\underline{\omega}\right] \leq e^{-\delta\underline{\omega}\eta}E\left(\exp\left\{\delta\underline{\omega}\sup_t\left|\sum_{i=1}^n v_i\sum_{\ell=1}^{N_{i1}}\sum_{m=1}^{N_{i2}} e_{i1\ell}[\mathbf{1}(T_{i2m} \leq t) - G_2(t)]\right|\right\}\right)$$
$$\leq e^{-\eta^2/4}\left(1 + 2\eta\sqrt{2\pi}e^{\eta^2/8}\right) \leq (1 + 2\eta\sqrt{2\pi})e^{-\eta^2/8}. \tag{37}$$

Returning to a bound for the first expectation in (33), (34) and (37) imply

$$E\left[\exp\left\{2\lambda \sup_{s,t}|Y_1(s,t)|\right\}\right] \leq 1 + 2\int_1^\infty \Pr\left(\sup_t \left|\sum_{i=1}^n v_i \sum_{\ell=1}^{N_{i1}} \sum_{m=1}^{N_{i2}} e_{i1\ell}\left[\mathbf{1}(T_{i2m} \leq t) - G_2(t)\right]\right| > \frac{\log(\gamma)}{4\lambda}\right) \mathrm{d}\gamma$$
$$\leq 1 + 16\lambda\underline{\omega}\int_1^\infty (1 + 4\sqrt{2\pi}\eta)e^{8\lambda\underline{\omega}\eta}e^{-\eta^2/2}\mathrm{d}\eta \leq 1 + 16\lambda\underline{\omega}\left[4\sqrt{2\pi} + \sqrt{2\pi}\left(1 + 32\lambda\underline{\omega}\sqrt{2\pi}\right)e^{32\lambda^2\underline{\omega}^2}\right]$$
$$\leq 1 + 64\sqrt{2\pi}\lambda\underline{\omega} + 16\sqrt{2\pi}\lambda\underline{\omega}\left(1 + 32\sqrt{2\pi}\lambda\underline{\omega}\right)e^{32\lambda^2\underline{\omega}^2},$$

where the change of variable $\log(\gamma) = 8\lambda\underline{\omega}\tau$ and the integral bound $\int_0^\infty xe^{rx}e^{-x^2/2}\mathrm{d}x \leq 1 + r\sqrt{2\pi}e^{r^2/2}$ for $r > 0$ have been used. By symmetry, the same bound applies to the term involving Y_2 in (32), so (30) follows. □

Lemma 6. *Let G be an arbitrary cumulative distribution function. For any $n \in \mathbb{N}$, let $N_i \in \mathbb{N}$, $i \in \{1,\ldots,n\}$ be arbitrary and consider an independent array of random variables $\mathbb{T} = \{T_{i\ell} : \ell \in \{1,\ldots,N_i\},\ i \in \{1,\ldots,n\}\}$, whose elements are independent and identically distributed according to G. Furthermore, let v_i, $i \in \{1,\ldots,n\}$, be arbitrary constants and define $\omega^2 = \sum_{i=1}^n v_i^2 N_i(N_i-1)^2$. Set $\mathbb{B}_n(s,t) = \sum_{i=1}^n v_i \sum_{\ell=1}^{N_i}\sum_{m\neq\ell}\{\mathbf{1}(T_{i\ell} \leq s)\mathbf{1}(T_{im} \leq t) - G(s)G(t)\}$ Then, for any $\epsilon > 0$,*

$$\Pr\left(\sup_{s,t\in\mathbb{R}}|\mathbb{B}_n(s,t)| > \epsilon\omega\right) \leq \left(1 + 2\sqrt{2\pi}\epsilon\right)e^{-\epsilon^2/8}$$

Proof: The proof is similar to that of Lemma 5 and the details are omitted. □

Lemma 7. *Suppose that assumptions A1 and B1 hold, and recall the definition of ω_{jk} in (7). Then, there exist $C_1, C_2 > 0$ such that, for any $\epsilon > 0$, integers $q, r \in \{0,1,2\}$ such that $0 \leq q + r \leq 2$ and $j,k \in \{1,\ldots,p\}$,*

$$\Pr\left\{\left|\left|\left|S_{jkqr}(\cdot,*) - E[S_{jkqr}(\cdot,*)]\right|\right|\right|_{j,k} > \epsilon\right\} \leq C_1\exp\left\{\frac{-C_2 b_{\gamma_j}^2 b_{\gamma_k}^2\epsilon^2}{\omega_{jk}^2}\right\}$$

under an FR design and, under an SR design,

$$\Pr\left\{\left|\left|\left|S_{jkqr}(\cdot,*) - E[S_{jkqr}(\cdot,*)]\right|\right|\right|_{j,k} > \epsilon\right\} \leq C_1\exp\left\{\frac{-C_2 b_\gamma^4\epsilon^2}{\omega^2}\right\}.$$

Proof: The proof under both designs will be shown for $j \neq k$; the case $j = k$ follows similar, but simpler arguments. Begin with the FR design. Standard derivations arising from application of Riemann-Stiltjes integration by parts yield $\left|\left|\left|S_{jkqr}(\cdot,*) - E\left[S_{jkqr}(\cdot,*)\right]\right|\right|\right|_{j,k} \leq V_K b_{\gamma_j}^{-1} b_{\gamma_k}^{-1}\left|\left|\left|\mathbb{A}_{jk}\right|\right|\right|_{j,k}$, where V_K is a constant depending only on the kernel K, $\mathbb{A}_{jk}(s,t) = \sum_{i=1}^n v_{ijk}\sum_{\ell=1}^{N_{ij}}\sum_{m=1}^{N_{ik}}\left\{\mathbf{1}(T_{ij\ell} \leq s)\mathbf{1}(T_{ijk} \leq t) - F_j(s)F_k(t)\right\}$, and F_j is the cdf corresponding to f_j. Thus, the first inequality of the lemma follows from an application of Lemma 5.

In the case of an SR design, write S_{qr} and v_i for the common values of S_{jkqr} and v_{ijk}. Then $\left|\left|\left|S_{qr}(\cdot,*) - E(S_{qr}(\cdot,*))\right|\right|\right| \leq V_K b_\gamma^{-2}\left|\left|\left|\mathbb{B}\right|\right|\right|$, where $\mathbb{B}(s,t) = \sum_{i=1}^n v_i\sum_{\ell=1}^{N_i}\sum_{m\neq l}\left[\mathbf{1}(T_{i\ell} - s)\mathbf{1}(T_{im} - t) - F(s)F(t)\right]$. The result then follows by applying Lemma 6. □

Lemma 8. *Suppose assumptions A1, B1, B2, and D3 hold. Define*

$$S_{jk}^*(s,t) = Q_{jk0}S_{jk00} - Q_{jk1}S_{jk10} + Q_{jk2}S_{jk01},$$

and define ω_{jk} and ω as in (7). Then there exist $C_1, C_2, \eta, N > 0$ such that, for any $n \geq N$ and $j,k \in \{1,\ldots,p\}$,

$$\Pr\left\{\inf_{(s,t)\in\mathcal{T}_j\times\mathcal{T}_k} S_{jk}^*(s,t) < \eta\right\} \leq C_1\exp\left\{\frac{-C_2 b_{\gamma_j}^2 b_{\gamma_k}^2\epsilon^2}{\omega_{jk}^2}\right\}$$

under an FR design and, under an SR design,

$$\Pr\left\{\inf_{(s,t)\in\mathcal{T}^2} S_{jk}^*(s,t) < \eta\right\} \leq C_1\exp\left\{\frac{-C_2 b_\gamma^4\epsilon^2}{\omega^2}\right\}.$$

Proof: The proof follows the same lines as the proof of Lemma 2, so only a sketch will be provided. Define

$$\overline{S}^*_{jk}(s,t) = \left(E\left[S_{jk20}\right]E\left[S_{jk02}\right] - E\left[S_{jk11}\right]^2\right)E\left[S_{jk00}\right] - \left(E\left[S_{jk10}\right]E\left[S_{jk02}\right] - E\left[S_{jk01}\right]E\left[S_{jk11}\right]\right)E\left[S_{jk10}\right] + \left(E\left[S_{jk10}\right]E\left[S_{jk11}\right] - E\left[S_{jk01}\right]E\left[S_{jk20}\right]\right)E\left[S_{jk01}\right].$$

First, one establishes that $\overline{S}^*_{jk}(s,t) \geq \tau f_j^3(s) f_k^3(t) - o(1)$ uniformly in j, k, s, and t, where τ depends only on K. Then, by assumptions B2 and D3, there exists N such that, when $n > N$, $\min_{j,k\in\{1,\ldots,p\}} \inf_{(s,t)\in\mathcal{T}_j\times\mathcal{T}_k} \overline{S}^*_{jk}(s,t) \geq \frac{m^4\tau}{2}$. The remainder of the argument follows from uniform continuity and Lemma 7. □

Lemma 9. *Suppose assumptions A1, B1, B2, and D3 hold. Let Q_{jkr} be as defined in* (24)*, $r \in \{0,1,2\}$, and let ω_{jk} be as in* (7)*. Then there exist $C_1, C_2, C_3 > 0$ such that, for any $\epsilon > 0$, $j,k \in \{1,\ldots,p\}$, and $r \in \{0,1,2\}$,*

$$\Pr\left(\left|\left|\left|Q_{jkr}\right|\right|\right|_{j,k} - C_3 > \epsilon\right) \leq C_1 \exp\left\{\frac{-C_2 b_{\gamma_j}^2 b_{\gamma_k}^2 \epsilon^2}{\omega_{jk}^2}\right\}$$

under an FR design and, under an SR design,

$$\Pr\left(\left|\left|\left|Q_{jkr}\right|\right|\right|_{j,k} - C_3 > \epsilon\right) \leq C_1 \exp\left\{\frac{-C_2 b_{\gamma}^4 \epsilon^2}{\omega^2}\right\}.$$

Proof: Assumptions B2 and D3 imply that there is C_3 such that $\left|\left|\left|E(Q_{jkr})\right|\right|\right|_{j,k} \leq C_3$. Apply Lemma 7. □

6.3.2. L^2 and Uniform Convergence of Numerator Terms

To derive the L^2 convergence rates in Section 4, preliminary concentration inequalities will be provided for the L^2 norms of the quantities in (26)–(28). For $(q,r) \in \{(0,0),(1,0),(0,1)\}$, define $\mathbb{T}_{ijk} = \left\{(T_{ij\ell}, T_{ikm}) : (\ell,m) \in \mathcal{I}_{ijk}\right\}$ and

$$W_{ijk}^{qr}(s,t) = v_{ijk} \sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - s) K_{b_{\gamma_k}}(T_{ikm} - t) \left(\frac{T_{ij\ell} - s}{b_{\gamma_j}}\right)^q \left(\frac{T_{ikm} - t}{b_{\gamma_k}}\right)^r U_{ijk\ell m} \tag{38}$$

Recall the definitions of the rates q_{jkn} in (6) and define $v^*_{njk} = \max_{i\in\{1,\ldots,n\}} v_{ijk}$. As $|B_{jk}(s,t)| \leq C b_{\gamma_j}^2 b_{\gamma_k}^2$ for a universal constant C by assumption B4, the term on the last line of each of (26)–(28) are bounded by a multiple of S_{jk00}.

Lemma 10. *Suppose that assumptions A1, B1, B2, B4, and D1–D3 hold. Then there exist constants C_1 and C_2 such that, for any $\epsilon > 0$, $(q,r) \in \{(0,0),(0,1),(1,0)\}$, and $j,k \in \{1,\ldots,p\}$,*

$$\Pr\left(\left|\left|\sum_{i=1}^n W_{ijk}^{qr}\right|\right|_{j,k} > \epsilon\right) \leq C_1 \exp\left\{\frac{-C_2\epsilon^2}{q_{jkn} + b_{\gamma_j}^{-1} b_{\gamma_k}^{-1} \overline{v}_{njk}\epsilon}\right\}.$$

Proof: As the proof follows the same logic as that of Lemma 3, the arguments will be sketched for the case $q = r = 0$. For simplicity, write W_{ijk} for W_{ijk}^{00}. First, under assumptions D1 and D2, it can be deduced that, conditionally on $\mathbb{T}_{ijk}$, the random variables $U_{ijk\ell m}$ are sub-Exponential random variables with parameters bounded by some universal constant $\rho^2 < \infty$ depending only on $\overline{\theta}$ and σ in assumptions D1 and D2, respectively. Hence, using moment bounds for sub-Exponential random variables, it can be concluded that, for any j,k, $(s,t) \in \mathcal{T}_j \times \mathcal{T}_k$, and $\nu \geq 2$,

$$E\left[\left|W_{ijk}(s,t)\right|^\nu \mid \mathbb{T}_{ijk}\right] \leq 2\nu!(2\rho^2 v_{ijk})^\nu \left\{\sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - s) K_{b_{\gamma_k}}(T_{ikm} - t)\right\}^\nu.$$

Then assumptions A1, B1, and B2 imply that $E\left[\left\{\sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - s) K_{b_{\gamma_k}}(T_{ikm} - t)\right\}^2\right] \leq C a_{ijk}$, for some C, with

$$a_{ijk} = \begin{cases} N_{ij}N_{ik}(b_{\gamma_j}^{-1} + N_{ij} - 1)(b_{\gamma_k}^{-1} + N_{ik} - 1), & j \neq k, \\ N_{ij}(N_{ij} - 1)\left\{b_{\gamma_j}^{-1}(b_{\gamma_j}^{-1} + N_{ij} - 2) + (N_{ij} - 2)(b_{\gamma_j}^{-1} + N_{ij} - 3)\right\}, & j = k, \end{cases}$$

under an FR design, and $a_{ijk} \equiv a_i$ under an SR design, where a_i is the common value of a_{ijj} across j in the above display. Hence, applying Jensen's inequality, there are universal constants B_1 and B_2 such that

$$\sum_{i=1}^{n} E\left[\left\|W_{ijk}\right\|_{j,k}^{\nu}\right] \leq \frac{\nu!}{2}\left(B_2 \sum_{i=1}^{n} v_{ijk}^2 a_{ijk}\right)(B_1 b_{\gamma_j}^{-1} b_{\gamma_k}^{-1} \overline{v}_{njk})^{\nu-2},$$

where $\overline{v}_{njk} = \max_{i\in\{1,\ldots,n\}} v_{ijk}|\mathcal{I}_{ijk}|$. Using the definitions of q_{jkn} and applying Theorem 2.5 of [4], the inequality of the lemma is established for $q = r = 0$. For the other values, apply the same arguments to the kernel $\tilde{K}(u) = uK(u)$. □

Lemma 11. *Suppose that assumptions A1, B1, B2, and D1–D3 hold. Then there exist constants $C_1, C_2, C_3 > 0$ such that, for any $\epsilon > 0$, and $j, k \in \{1, \ldots, p\}$,*

$$\Pr\left\{\left|\sum_{i=1}^{n} v_{ijk} \sum_{(\ell,m)\in\mathcal{I}_{ijk}} \left(|U_{ijk\ell m}| - E\left[|U_{ijk\ell m}|\right]\right)\right| > \epsilon\right\} < C_1 \exp\left\{\frac{-C_2\epsilon^2}{\sum_{i=1}^{n} v_{ijk}^2|\mathcal{I}_{ijk}|^2 + \overline{v}_{njk}\epsilon}\right\}.$$

Proof. The proof follows the same lines as the proof of Lemma 4, using the fact that each of $U_{ijk\ell m}$ are, conditionally on $\mathbb{T}_{ijk}$, sub-Exponential random variables due to assumptions D1 and D2. The details are omitted. □

6.4. Proofs of Results in Section 4

Proof of Theorem 3: The proof applies the same logic as the proof of Theorem 1, but using Lemmas 7–10 that are relevant for covariance estimation rather than the corresponding Lemmas 1–3 that are for mean estimation. Due to the similarities, step-by-step derivations of the bounds obtained below will not be provided. As for mean estimation, the proof will be given for an FR design, noting that the same arguments can be followed under an SR design with weaker conditions on the bandwidth as in assumption D3 since S_{jkqr} do not depend on j, k. Let C_1, C_2, C_3 be sufficiently large so that $\limsup_{n\to\infty} \max_{j,k\in\{1,\ldots,p\}} \left\|\left\|E[S_{jk00}]\right\|\right\|_{j,k} \leq C_3$ and Lemmas 8–10 hold simultaneously. Let ω_{jk} be as defined in (7), and $S_{jk}^*(s,t)$, η be as in Lemma 8. Recalling the expression of $\hat{\gamma}_{jk}(s,t) - \gamma_{jk}(s,t)$ in (25), first apply Lemma 8 to conclude that, for $n \geq N$,

$$\Pr\left\{\min_{j,k\in\{1,\ldots,p\}} S_{jk}^*(s,t) < \eta\right\} \leq C_1 p^{2-C_2\eta^2\left\{\log(p)\max_{j,k\in\{1,\ldots,p\}} b_{\gamma_j}^{-1} b_{\gamma_k}^{-1}\omega_{jk}^2\right\}}$$

so that, by assumption D3, $\max_{j,k\in\{1,\ldots,p\}}\left\{\inf_{(s,t)\in\mathcal{T}_j\times\mathcal{T}_k} S_{jk}^*(s,t)\right\}^{-1} = O_P(1)$. Similarly, Lemma 9 implies that, with Q_{jkr} as in (24), $\max_{j,k\in\{1,\ldots,p\}} \left\|\left\|Q_{jkr}\right\|\right\|_{j,k} = O_P(1)$, $r \in \{0, 1, 2\}$. Hence, the rate is determined by that of the terms in (26)–(28).

Applying Lemma 10, for any nonnegative integers q and r such that $0 \leq q + r \leq 1$, and any $R > 0$,

$$\Pr\left(\max_{j,k\in\{1,\ldots,p\}} \left\|W_{ijk}^{qr}\right\|_{j,k} > R\{\log(p)\}^{1/2} \max_{j,k\in\{1,\ldots,p\}} \left[q_{jkn} + \{\log(p)\}^{1/2} b_{\gamma_j}^{-1} b_{\gamma_k}^{-1}\overline{v}_{njk}\right]\right) \leq C_1 p^{2-\frac{C_2R^2}{1+R}}.$$

Lemma 7 along with the bound $\left\|S_{jk00}\right\|_{j,k} \leq B\left\|\left\|S_{jk00}\right\|\right\|_{j,k}$ for a universal constant B implies that, for any $R > 0$,

$$\Pr\left[\max_{j,k\in\{1,\ldots,p\}} \left(\left\|S_{jk00}\right\|_{j,k} - |\mathcal{T}_j|^{1/2}|\mathcal{T}_k|^{1/2}C_3\right) > R\{\log(p)\}^{1/2} \max_{j,k} b_{\gamma_j}^{-1} b_{\gamma_k}^{-1}\omega_{jk}^2\right] \leq C_1 p^{2-C_2R^2}.$$

Since $\left\|\left\|B_{jk\ell m}\right\|\right\|_{j,k} \leq C b_{\gamma_j} b_{\gamma_k}$ for a universal constant C, the proof is complete upon observing that

$$\begin{aligned}
&\max_{j,k\in\{1,\ldots,p\}} \left\| \sum_{i=1}^{n} W_{ijk}^{qr}(\cdot,*) + \sum_{i=1}^{n} v_{ijk} \sum_{(\ell,m)\in\mathcal{I}_{ijk}} K_{b_{\gamma_j}}(T_{ij\ell} - \cdot) K_{b_{\gamma_k}}(T_{ikm} - *)\left(\frac{T_{ij\ell} - \cdot}{b_{\gamma_j}}\right)^q \left(\frac{T_{ikm} - *}{b_{\gamma_k}}\right)^r B_{jklm}(\cdot,*)\right\|_{j,k} \\
&\quad = O_P\left(\max_{j\in\{1,\ldots,p\}} b_{\gamma_j}^2 + \max_{j,k\in\{1,\ldots,p\}} \left[\left\{\log(p) q_{jkn}\right\}^{1/2} + \log(p) b_{\gamma_j}^{-1} b_{\gamma_k}^{-1}\overline{v}_{njk}\right]\right).
\end{aligned}$$

□

Proofs of Corollary 7–9: The proofs are similar to those of Corollaries 1–3 and are omitted. □

Proof of Theorem 4: The proof is similar to that of Theorem 2, so only the details will be sketched. First, as in the proof of Theorem 3, it can be established that, with S^*_{jk} as defined in Lemma 8 and Q_{jkr} as defined in (24) for $r \in \{0, 1, 2\}$, almost surely, $\max_{j,k\in\{1,\ldots,p\}} \left\{\inf_{(s,t)\in\mathcal{T}_j\times\mathcal{T}_k} S^*_{jk}(s,t)\right\}^{-1} = O(1)$ and $\max_{j,k\in\{1,\ldots,p\}} \left|\!\left|\!\left| Q_{jkr} \right|\!\right|\!\right|_{j,k} = O(1)$. It then follows from (25)–(28) that, almost surely,

$$\max_{j,k\in\{1,\ldots,p\}} \left|\!\left|\!\left| \hat{\gamma}_{jk} - \gamma_{jk} \right|\!\right|\!\right|_{j,k} = O\left\{ \max_{j\in\{1,\ldots,p\}} b^2_{\gamma_j} + \max_{j,k\in\{1,\ldots,p\}} \left(\left|\!\left|\!\left| \sum_{i=1}^n W^{00}_{ijk} \right|\!\right|\!\right|_{j,k} + \left|\!\left|\!\left| \sum_{i=1}^n W^{01}_{ijk} \right|\!\right|\!\right|_{j,k} + \left|\!\left|\!\left| \sum_{i=1}^n W^{10}_{ijk} \right|\!\right|\!\right|_{j,k} \right) \right\}. \tag{39}$$

Next, letting $\chi_j(\delta)$ represent an equally spaced grid of $\mathcal{T}_j$ with spacing no larger than $\delta > 0$, set $\chi_{jk}(\delta) = \chi_j(\delta) \times \chi_k(\delta)$ such that $|\chi_{jk}(\delta)| \leq L\delta^{-2}$ for some universal constant L. Then, for any nonnegative integers q, r with $q + r \leq 1$,

$$\left|\!\left|\!\left| \sum_{i=1}^n W^{qr}_{ijk} \right|\!\right|\!\right|_{j,k} \leq \max_{(s,t)\in\chi_{jk}(\delta)} \left| \sum_{i=1}^n W^{qr}_{ijk}(s,t) \right| + \sup_{|s-s'|,|t-t'|<\delta} \left| \sum_{i=1}^n \left\{ W^{qr}_{ijk}(s,t) - W^{qr}_{ijk}(s',t') \right\} \right|$$

Then, with $a_n(\delta) = \max_{j,k\in\{1,\ldots,p\}} \left[q^{1/2}_{jkn} + \{\log(p\delta^{-2})\}^{1/2} b^{-1}_{\gamma_j} b^{-1}_{\gamma_k} \overline{v}_{njk} \right]$, using the arguments of Lemma 10, it can be shown that, for any $R > 0$, there exist universal constants C_1 and C_2 such that

$$\Pr\left[\max_{j,k\in\{1,\ldots,p\}} \max_{(s,t)\in\chi_{jk}(\delta)} \left| \sum_{i=1}^n W^{qr}_{ijk}(s,t) \right| > R \left\{ \log(p\delta^{-2}) \right\}^{1/2} a_n(\delta) \right] \leq C_1 (p\delta^{-2})^{1-\frac{C_2 R^2}{1+R}},$$

whence

$$\max_{j,k\in\{1,\ldots,p\}} \max_{(s,t)\in\chi_{jk}(n^{-\alpha})} \left| \sum_{i=1}^n W^{qr}_{ijk}(s,t) \right| = O\left[\left\{ \log(p \vee n) q_{jkn} \right\}^{1/2} + \log(p \vee n) b^{-1}_{\gamma_j} b^{-1}_{\gamma_k} \overline{v}_{njk} \right] \tag{40}$$

almost surely by Borel-Cantelli, where α satisfies assumption D4. In addition, using assumption A2 and Lemma 11, it follows that

$$\max_{j,k\in\{1,\ldots,p\}} \max_{|s-s'|,|t-t'|<\delta} \left| \sum_{i=1}^n \left\{ W^{qr}_{ijk}(s,t) - W^{qr}_{ijk}(s',t') \right\} \right| = O\left(\delta \max_{j\in\{1,\ldots,p\}} b^{-3}_{\gamma_j} \right) \tag{41}$$

almost surely. Applying $\delta = n^{-\alpha}$, α satisfying assumption D4, to (39)–(41) proves the result. □

Proofs of Corollaries 10–12: The proofs are similar to those of Corollaries 7–9, and are omitted. □

Acknowledgments

This research was supported by NSF Award DMS-2310943.

Supplementary Materials

The supplementary materials include comparisons of the performance of the OBS and SUBJ schemes on simulated data sets, along with corresponding code.

References

[1] G. Aneiros, R. Cao, R. Fraiman, C. Genest, P. Vieu, Recent advances in functional data analysis and high-dimensional statistics, J. Multivariate Anal. 170 (2019) 3–9.

[2] G. Aneiros, I. Horová, M. Hušková, P. Vieu, On functional data analysis and related topics, Journal of Multivariate Analysis 189 (2022) 104861.

[3] J. R. Berrendero, A. Justel, M. Svarc, Principal components for multivariate functional data, Comput. Statist. Data Anal. 55 (2011) 2619–2634.

[4] D. Bosq, Linear processes in function spaces, volume 149 of Lecture Notes in Statistics, Springer-Verlag, New York, 2000. Theory and applications.
[5] K. Chen, H.-G. Müller, Modeling repeated functional observations, J. Amer. Statist. Assoc. 107 (2012) 1599–1609.
[6] Y. Chen, Z. Lin, H.-G. Müller, Wasserstein regression, J. Amer. Statist. Assoc. 118 (2023) 869–882.
[7] J.-M. Chiou, Y.-T. Chen, Y.-F. Yang, Multivariate functional principal component analysis: a normalization approach, Statist. Sinica 24 (2014) 1571–1596.
[8] J.-M. Chiou, H.-G. Müller, Linear manifold modelling of multivariate functional data, J. R. Stat. Soc. Ser. B. Stat. Methodol. 76 (2014) 605–626.
[9] J.-M. Chiou, Y.-F. Yang, Y.-T. Chen, Multivariate functional linear regression and prediction, J. Multivariate Anal. 146 (2016) 301–312.
[10] A. Cuevas, A partial overview of the theory of statistics with functional data, Journal of Statistical Planning and Inference 147 (2014) 1–23.
[11] P. Delicado, R. Giraldo, C. Comas, J. Mateu, Statistics for spatial functional data: some recent contributions, Environmetrics 21 (2010) 224–239.
[12] P. Dubey, H.-G. Müller, Functional models for time-varying random objects, J. R. Stat. Soc. Ser. B. Stat. Methodol. 82 (2020) 275–327.
[13] P. Dubey, H.-G. Müller, Modeling time-varying random objects and dynamic networks, J. Amer. Statist. Assoc. 117 (2022) 2252–2267.
[14] E. X. Fang, Y. Ning, R. Li, Test of significance for high-dimensional longitudinal data, Ann. Statist. 48 (2020) 2622–2645.
[15] F. Ferraty, P. Vieu, Nonparametric functional data analysis, Springer Series in Statistics, Springer, New York, 2006. Theory and practice.
[16] L. Ghodrati, V. M. Panaretos, Distribution-on-distribution regression via optimal transport maps, Biometrika 109 (2022) 957–974.
[17] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics [Editorial], J. Multivariate Anal. 146 (2016) 1–6.
[18] S. Greven, C. Crainiceanu, B. Caffo, D. Reich, Longitudinal functional principal component analysis, in: Recent advances in functional data analysis and related topics, Contrib. Statist., Physica-Verlag/Springer, Heidelberg, 2011, pp. 149–154.
[19] S. Hörmann, P. Kokoszka, Consistency of the mean and the principal components of spatially distributed functional data, in: Recent advances in functional data analysis and related topics, Contrib. Statist., Physica-Verlag/Springer, Heidelberg, 2011, pp. 169–175.
[20] L. Horváth, P. Kokoszka, Inference for functional data with applications, Springer Series in Statistics, Springer, New York, 2012.
[21] K. Hron, A. Menafoglio, M. Templ, K. Hrůzová, P. Filzmoser, Simplicial principal component analysis for density functions in Bayes spaces, Comput. Statist. Data Anal. 94 (2016) 330–350.
[22] T. Hsing, R. Eubank, Theoretical foundations of functional data analysis, with an introduction to linear operators, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 2015.
[23] X. Hu, F. Yao, Sparse functional principal component analysis in high dimensions, Statist. Sinica 32 (2022) 1939–1960.
[24] J. Jacques, C. Preda, Model-based clustering for multivariate functional data, Comput. Statist. Data Anal. 71 (2014) 92–106.
[25] P. Kokoszka, M. Reimherr, Introduction to functional data analysis, CRC press, 2017.
[26] S. Koner, A.-M. Staicu, Second-generation functional data, Annu. Rev. Stat. Appl. 10 (2023) 547–572.
[27] K.-Y. Lee, L. Li, B. Li, H. Zhao, Nonparametric functional graphical modeling through functional additive regression operator, J. Amer. Statist. Assoc. 118 (2023) 1718–1732.
[28] B. Li, E. Solea, A nonparametric graphical model for functional data with application to brain networks based on fMRI, J. Amer. Statist. Assoc. 113 (2018) 1637–1655.
[29] Y. Li, T. Hsing, Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data, Ann. Statist. 38 (2010) 3321–3351.
[30] M. B. Marcus, J. Zinn, The bounded law of the iterated logarithm for the weighted empirical distribution process in the non-i.i.d. case, Ann. Probab. 12 (1984) 335–360.
[31] S. Y. Park, A.-M. Staicu, Longitudinal functional data analysis, Stat 4 (2015) 212–226.
[32] A. Petersen, H.-G. Müller, Functional data analysis for density functions by transformation to a Hilbert space, Ann. Statist. 44 (2016) 183–218.
[33] X. Qiao, S. Guo, G. M. James, Functional graphical models, J. Amer. Statist. Assoc. 114 (2019) 211–222.
[34] J. O. Ramsay, B. W. Silverman, Functional data analysis, Springer Series in Statistics, Springer, New York, second edition, 2005.
[35] E. Solea, H. Dette, Nonparametric and high-dimensional functional graphical models, Electron. J. Stat. 16 (2022) 6175–6231.
[36] E. Solea, B. Li, Copula Gaussian graphical models for functional data, J. Amer. Statist. Assoc. 117 (2022) 781–793.
[37] J.-L. Wang, J.-M. Chiou, H.-G. Müller, Functional data analysis, Annual Review of Statistics and its Application 3 (2016) 257–295.
[38] R. K. W. Wong, Y. Li, Z. Zhu, Partially linear functional additive models for multivariate functional data, J. Amer. Statist. Assoc. 114 (2019) 406–418.
[39] K. Xue, J. Yang, F. Yao, Optimal linear discriminant analysis for high-dimensional functional data, J. Amer. Statist. Assoc. 119 (2024) 1055–1064.
[40] J. Zapata, S. Y. Oh, A. Petersen, Partial separability and functional graphical models for multivariate Gaussian processes, Biometrika 109 (2022) 665–681.
[41] H. Zhang, Y. Li, Unified principal component analysis for sparse and dense functional data under spatial dependency, J. Bus. Econom. Statist. 40 (2022) 1523–1537.
[42] X. Zhang, J.-L. Wang, From sparse to dense functional data and beyond, Ann. Statist. 44 (2016) 2281–2321.
[43] B. Zhao, Y. S. Wang, M. Kolar, FuDGE: a method to estimate a functional differential graph in a high-dimensional setting, J. Mach. Learn. Res. 23 (2022) Paper No. [82], 82.
[44] P.-S. Zhong, R. Li, S. Santo, Homogeneity tests of covariance matrices with high-dimensional longitudinal data, Biometrika 106 (2019) 619–634.
[45] H. Zhu, J. S. Morris, F. Wei, D. D. Cox, Multivariate functional response regression, with application to fluorescence spectroscopy in a cervical pre-cancer study, Comput. Statist. Data Anal. 111 (2017) 88–101.
[46] H. Zhu, N. Strawn, D. B. Dunson, Bayesian graphical models for multivariate functional data, J. Mach. Learn. Res. 17 (2016) Paper No. 204, 27.