Statistica Sinica Preprint No: SS-2022-0358	
Title	Simultaneous Inference for Mean Curves of Functional and Longitudinal Data: A Unified Theory
Manuscript ID	SS-2022-0358
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0358
Complete List of Authors	Zhipeng Lou and Weibiao Wu
Corresponding Authors	Zhipeng Lou
E-mails	zhipengprob@gmail.com

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

Statistica Sinica

Simultaneous Inference for Mean Curves of Functional and Longitudinal Data: A Unified Theory

Zhipeng Lou[1] and Wei Biao Wu[2]

University of Pittsburgh[1] *and The University of Chicago*[2]

Abstract: The paper considers simultaneous statistical inference for mean curves of functional and longitudinal data in a unified framework. We establish the asymptotic distribution for the normalized maximum deviations of the local linear estimators from the true mean functions. The asymptotic distribution leads to simultaneous confidence bands with asymptotically correct coverage probabilities. A Gaussian multiplier bootstrap procedure is proposed to obtain the cutoff values and our fully data-driven approach has a good finite sample performance. All the results obtained in the present paper are unified with respect to the sampling schemes.

Key words and phrases: Functional data analysis; Gaussian approximation; Local linear smoothing; Longitudinal data analysis; Simultaneous confidence bands; Gaussian multiplier bootstrap.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

1. Introduction

Functional and longitudinal data arise in many disciplines and have received considerable attention in the statistics community over the last two decades. Notable work includes Ramsay and Silverman (2005), Ferraty and Vieu (2006), Wu and Zhang (2006), Ramsay and Silverman (2007), Horváth and Kokoszka (2012), Hsing and Eubank (2015); see also Cuevas (2014) and Wang et al. (2016) for a comprehensive overview. As a typical setting, one considers independent and identically distributed (i.i.d.) random functions $X_1(t), \ldots, X_n(t)$ defined on a compact interval $\mathcal{T}$, with mean function $\mu(t) = \mathbb{E}\{X_1(t)\}$ and covariance function $\gamma(t, s) = \mathrm{cov}\{X_1(t), X_1(s)\}$. In practice, the entire trajectory of $X_i(t)$ is not observable as data can only be collected discretely over time. Correspondingly, as a more realistic formulation, each process is supposed to be observed with additive measurement errors at $m_i \geq 1$ discrete random time points $t_{ij} \in \mathcal{T}$, $j = 1, \ldots, m_i$. The actual observations then follow

$$Y_{ij} = X_i(t_{ij}) + \varepsilon_{ij} = \mu(t_{ij}) + \nu_i(t_{ij}) + \varepsilon_{ij}, \qquad (1.1)$$

where $\{\varepsilon_{ij}\}_{i,j\in\mathbb{N}}$ are i.i.d. centered random errors with variance $\mathrm{var}(\varepsilon_{11}) = \sigma_\varepsilon^2$, $\{t_{ij}\}_{i,j\in\mathbb{N}}$ are i.i.d. random times with density function $f(t)$, $\{\nu_i(\cdot)\}_{i=1}^n$ are i.i.d. zero-mean random processes, and X_i's, t_{ij}'s and ε_{ij}'s are mutu-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

ally independent. In (1.1) above, $\nu_i(\cdot)$ can be interpreted as random effect and (1.1) is a special case of functional mixed effect model (Guo, 2002).

A fundamental problem associated with model (1.1) is the statistical inference of the mean function $\mu(\cdot)$. The latter problem has been well studied under different settings in the literature. One commonly focused on only one type (dense or sparse) of functional data depending on the magnitude of the m_i's relative to the sample size n. Conventionally, if all the m_i's are larger than some order of n, the observations are referred to as dense functional data (Ramsay and Silverman, 2005; Zhang and Chen, 2007), while the sparse case concerns uniformly bounded m_i's or that m_i's follow some fixed distribution with bounded moment (Yao et al., 2005a,b; Hall et al., 2006). For these two cases, different estimation procedures were adopted and different asymptotic properties may be derived for the estimators; see Wang et al. (2016) for a review of various nonparametric estimation methods. With the asymptotic distribution, one can perform statistical inference for the mean function; see Degras (2011), Choi and Reimherr (2018), Cao et al. (2012) for the dense case and Ma et al. (2012), Zheng et al. (2014), Cao et al. (2018) for the sparse case.

The primary goal of the paper is to construct simultaneous confidence bands (SCBs) for the mean function μ with asymptotically correct coverage

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

probabilities. Namely based on the observations from model (1.1), we shall find two functions $\mu_-(t)$ and $\mu_+(t)$ such that

$$\lim_{n\to\infty} \mathbb{P}\{\mu_-(t) \le \mu(t) \le \mu_+(t) \text{ for all } t \in \mathcal{T}\} = 1 - \alpha,$$

where $\alpha \in (0,1)$ is a prespecified significance level. Thus the confidence band $(\mu_-(t), \mu_+(t))_{t\in\mathcal{T}}$ is constructed with an asymptotically correct coverage probability, by which one can efficiently infer the shape of the mean curve μ such as monotonicity, convexity, linearity, unimodality or other forms. In particular, the confidence bands provide guidelines for parametric modelling of the mean function. On the other hand, simultaneous inference also enables us to perform uniform hypothesis testing of the mean curve such as testing whether $\mu = \mu_0$ for some given function μ_0 (Degras, 2017). Moreover, the constructed confidence bands can also be applied to two-sample mean function testing such as testing the equality of the ECG curves from the patients and control groups (Cuevas, 2014) and the equality of the fractional anisotropy curves between the control subjects and the multiple sclerosis patients (Degras, 2017), see also (Wang, 2021).

The SCBs is more informative than the point-wise version in terms of making statistical inference on patterns of the mean curve. See Degras (2017) for an excellent overview. As commented in the latter paper, the construction of SCBs with sparse functional data is markedly different from

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

the case of dense data since the sparsity of observations within curves leads to longitudinal dependence asymptotically negligible as $n \to \infty$. In practice, it can be nontrivial to predetermine a scenario since sparse and dense functional data are asymptotic concepts. Furthermore, a mixture of sparse and dense data may be collected in real applications, which can create extra difficulty in data analysis. Li and Hsing (2010) have made some progress towards a general estimation theory by proposing a unified local linear estimator; see also Cai and Yuan (2011) and Zhang and Wang (2016). Kim and Zhao (2013) proposed two unified approaches based on self-normalization to construct point-wise confidence intervals for $\mu(t)$ at a fixed t. It is still an open problem on simultaneous statistical inference for mean functions in a unified framework, which gives more insights by accounting for the entire curve. The nature that the repeated measurements on the same subject are correlated also imposes challenges on the construction of SCBs. Earlier researchers derived conservative confidence bands based on the Bonferroni correction procedure (Wu et al., 1998; Chiang et al., 2001; Huang et al., 2004). However, such confidence bands are known to be too wide and are therefore less informative about the overall shape of the mean curves.

In this paper, we will apply the recently developed high dimensional central limit theory in Chernozhukov et al. (2013, 2017) and provide an

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

answer to this open problem by providing a unified SCBs construction for both sparse and dense settings. In particular, we shall construct SCBs for the mean curves based on a weighted local linear estimator (cf. (2.1)), incorporating all types of functional data in a unified framework in which the magnitudes of $m_1, \ldots, m_n$ can be very flexible relative to the sample size n.

The rest of the paper is structured as follows. In Section 2, we shall introduce the weighted local linear estimate and establish the asymptotic distribution of the normalized maximum deviation of the mean estimator. We remark that our convergence result is new: neither $\sqrt{n}$-convergence using stochastic equicontinuity in the dense case nor Gumbel convergence in the sparse case as established in some existing work. We provide a detailed characterization of the asymptotic distributional approximations under different sampling schemes which can range from sparse cases to dense cases. In Section 3, we address implementation challenges in constructing the SCBs and tackle some key issues including asymptotic variance function estimation, bandwidth selection and bias correction. We conduct a simulation study in Section 4 to investigate the finite sample performance of the proposed method. All the technical proofs are given in online Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

We now introduce some notation. For $q > 0$, let $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$ if $\mathbb{E}|X|^q < \infty$. Denote by $\mathbb{I}\{\cdot\}$ the indicator function. For $a, b \in \mathbb{R}$, denote $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For two sequences of positive numbers $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$, we write $a_n \approx b_n$ if $a_n/b_n \to 1$ and write $a_n \lesssim b_n$ (resp. $a_n \asymp b_n$) if there exists some positive constant C such that $a_n \le Cb_n$ (resp. $C^{-1} \le a_n/b_n \le C$) for all sufficiently large n. For a sequence of random variables $\{\xi_n\}_{n\in\mathbb{N}}$, we write $\xi_n = O_{\mathbb{P}}(a_n)$ if $|\xi_n/a_n|$ is bounded in probability and write $\xi_n = o_{\mathbb{P}}(a_n)$ if ξ_n/a_n converges to zero in probability. Throughout the article, we shall use $C, C_1, C_2, \dots$ to denote generic positive constants whose values may vary from place to place.

2. Theoretical properties

In this section, we shall present the main theoretical results concerning the asymptotic properties of the local linear estimators. To quantify the impact of different sampling schemes, we shall firstly define

$$V_\phi = \frac{1}{n}\sum_{i=1}^{n} m_i^{-\phi}, \;\; \phi > 0.$$

Observe that smaller V_ϕ corresponds to denser observation time points. If $m_1, \dots, m_n$ are i.i.d. random variables with probability mass function $\mathbb{P}(m_1 = j) = \theta_j$, $j \ge 1$, then by law of large numbers $V_\phi \approx \sum_{j=1}^{\infty} j^{-\phi}\theta_j$. Note that $1/V_1$ is the harmonic mean of $\{m_1, \dots, m_n\}$ and it is related to the asymptotic variance of the local linear estimator; cf. (2.5).

2.1 Local linear smoothing

To estimate the mean curve μ, we shall use the local linear smoothing (Fan and Gijbels, 1996) due to its simple and explicit mathematical expression and good boundary performance. Let $K(u)$ be a symmetric probability density function and $K_b(u) = K(u/b)/b$, where $b > 0$ is the bandwidth satisfying $b \to 0$. Following Li and Hsing (2010), we apply the local linear smoother to the observations $\mathcal{D} = \{(t_{ij}, Y_{ij}), 1 \le j \le m_i, 1 \le i \le n\}$ and attach weight m_i^{-1} to each observation of the ith subject. Then the resulting mean estimator is $\widehat{\mu}_b(t) = \widehat{\omega}_1$, where

$$(\widehat{\omega}_1, \widehat{\omega}_2) = \operatorname*{arg\,min}_{\omega_1, \omega_2 \in \mathbb{R}} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \{Y_{ij} - \omega_1 - \omega_2(t_{ij} - t)\}^2 K_b(t_{ij} - t).$$

It follows that

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

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

$$S_k(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} K_b(t_{ij} - t)\{(t_{ij} - t)/b\}^k,$$

$$R_k(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} K_b(t_{ij} - t)\{(t_{ij} - t)/b\}^k Y_{ij}.$$

To study the theoretical properties of $\widehat{\mu}_b(t)$, we need the following regularity assumptions. Recall $\nu_i(t) = X_i(t) - \mu(t)$, $i = 1, \ldots, n$.

Assumption 1. The density function $f(t)$ is Lipschitz continuous on $\mathcal{T}$ and there exist positive constants m_f and M_f such that

$$0 < m_f \leq \inf_{t\in\mathcal{T}} f(t) \leq \sup_{t\in\mathcal{T}} f(t) \leq M_f < \infty.$$

Assumption 2. For some constant $q > 3$, we have $\|\sup_{t\in\mathcal{T}} |\nu_1(t)|\|_q < \infty$ and $\|\varepsilon_{11}\|_q < \infty$.

Assumption 3. The kernel function $K(u)$ is a symmetric probability density function with bounded support $[-1, 1]$. There exists some positive constant $L_K < \infty$ such that

$$|K(u) - K(v)| \leq L_K|u - v|, \;\; u, v \in \mathbb{R}.$$

Assumption 4. There exists a positive constant $\gamma_0 < \infty$ such that $\sigma_\varepsilon^2 \geq \gamma_0$ and $\inf_{t\in\mathcal{T}} \gamma(t, t) \geq \gamma_0$.

Assumption 5. There exist positive constants L_γ and α such that

$$|\gamma(t + h, s) - \gamma(t, s)| \leq L_\gamma|h|^\alpha, \;\; t, s \in \mathcal{T}. \tag{2.2}$$

Assumption 6. $\mu(t)$ is twice continuously differentiable on $\mathcal{T}$.

Remark 1. Assumptions 1 and 2 are similar to that in Li and Hsing (2010) and Zhang and Wang (2016). Assumption 3 is fairly mild and holds for many popular kernels. For example, it is satisfied by the popular Epanechnikov kernel $K(u) = 0.75(1 - u^2)\mathbf{1}\{|u| \leq 1\}$ and the triangle kernel $K(u) = (1-|u|)\mathbf{1}\{|u| \le 1\}$. Moreover, Assumption 3 implies $K(u)$ is uniformly upper bounded by some positive constant $M_K < \infty$. Assumption 4 ensures that the process $\nu(t)$ is nondegenerate for all $t \in \mathcal{T}$. Assumption 5 concerns the Hölder continuity of the covariance function $\gamma(t,s)$. Degras (2011) proposed a stochastic Hölder continuity condition on the process $\nu_i(t)$, and his condition implies $\|\nu_i(s) - \nu_i(t)\|^2 \le c_\nu |s-t|^\beta$ for some $c_\nu, \beta > 0$. Consequently, it follows that

$$|\gamma(t+h,s) - \gamma(t,s)| \le \sqrt{c_\nu} \sup_{s\in\mathcal{T}} \|\nu_i(s)\| |h|^{\beta/2}.$$

Then, (2.2) holds with $L_\gamma \ge \sqrt{c_\nu} \sup_{s\in\mathcal{T}} \|\nu_i(s)\|$ and $\alpha \le \beta/2$. The smoothness condition for μ in Assumption 6 is commonly used in the literature for nonparametric regression.

Remark 2. In the classical dense functional data analysis, it is often assumed that data are observed over a fine grid of deterministic points $t_{ij} = t_j$ and $m_i = m$ (Degras, 2011, 2017) for all $i = 1, \ldots, n$. Let $\mathcal{T} = [0,1]$, $t_j = F^{-1}(j/m)$ or $t_j = F^{-1}((j-1/2)/m)$, $j = 1, \ldots, m$, where $F' = f$ satisfies Assumption 1. Then a careful check of the proofs of our main results indicates that they are also valid for this deterministic design case.

2.2 Variance function

We now derive the asymptotic variance function of the local linear estimator $\widehat{\mu}_b(t)$ for each $t \in \mathcal{T}$. Define $\widetilde{f}_b(t) = S_0(t) - S_1(t)^2/S_2(t)$ and

$$R_k^*(t) = R_k(t) - S_k(t)\mu(t) - bS_{k+1}(t)\mu'(t), \quad k = 0, 1.$$

Note that $\widetilde{f}_b(t)S_2(t)$ is the denominator of $\widehat{\mu}_b(t)$ in (2.1). With this notation, we may write

$$\widetilde{f}_b(t)\{\widehat{\mu}_b(t) - \mu(t)\} = \bar{\eta}_n(t) + \frac{1}{n}\sum_{i=1}^{n} \varphi_i(t) + \frac{R_1^*(t)S_1(t)}{S_2(t)}, \qquad (2.3)$$

where $\bar{\eta}_n(t) = n^{-1}\sum_{i=1}^{n} \eta_i(t)$,

$$\eta_i(t) = \frac{1}{m_i}\sum_{j=1}^{m_i} K_b(t_{ij} - t)\{\nu_i(t_{ij}) + \varepsilon_{ij}\} \qquad (2.4)$$

and the bias term

$$\varphi_i(t) = \frac{1}{m_i}\sum_{j=1}^{m_i} K_b(t_{ij} - t)\{\mu(t_{ij}) - \mu(t) - (t_{ij} - t)\mu'(t)\}.$$

For ease of illustration, we assume hereafter that the numbers of observations $m_1, \ldots, m_n$ are deterministic. Note that the first term $\bar{\eta}_n(t)$ on the right-hand side of (2.3) determines the asymptotic distribution of $\widehat{\mu}_b(t)$. Hence it is needed to deal with the variance function

$$\sigma^2(t) = \mathrm{var}\{\sqrt{n}\bar{\eta}_n(t)\}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

In the case where $m_1, \ldots, m_n$ are random, $\sigma^2(t)$ above could be similarly defined conditional on these m_i's. Under Assumptions 1–5, it follows that for each $i = 1, \ldots, n$,

$$\begin{aligned}\mathrm{var}\{\eta_i(t)\} &= \frac{\mathrm{var}\{K_b(t_{i1}-t)(\nu_i(t_{i1})+\varepsilon_{i1})\}}{m_i} \\ &\quad + \frac{\mathrm{cov}\{K_b(t_{i1}-t)(\nu_i(t_{i1})+\varepsilon_{i1}), K_b(t_{i2}-t)(\nu_i(t_{i2})+\varepsilon_{i2})\}}{m_i/(m_i-1)} \\ &\approx \frac{f(t)\lambda_K\{\gamma(t,t)+\sigma_\varepsilon^2\}}{m_i b} + \frac{f^2(t)\gamma(t,t)(m_i-1)}{m_i},\end{aligned}$$

where $\lambda_K = \int K^2(u)du$. Consequently, for any fixed interior point $t \in \mathcal{T}$,

$$\sigma^2(t) \approx \frac{f(t)V_1\lambda_K\{\gamma(t,t)+\sigma_\varepsilon^2\}}{b} + f^2(t)\gamma(t,t)(1-V_1). \qquad (2.5)$$

Note that the magnitude of $\sigma(t)$ can be quite different under different sampling schemes, which consequently can lead to different convergence rates of $\widehat{\mu}_b(t)$. In particular, under Assumptions 1–5, it satisfies uniformly that

$$\sigma^2(t) \asymp \frac{V_1}{b} + 1 =: \vartheta^2.$$

Let $\pi_i(t) = \eta_i(t)/\sigma(t)$ for each $i = 1, \ldots, n$ and $\bar{\pi}_n(t) = n^{-1}\sum_{i=1}^n \pi_i(t)$. Observe that $\pi_1(t), \ldots, \pi_n(t)$ are independent centered random variables and $\mathrm{var}\{\sqrt{n}\bar{\pi}_n(t)\} = 1$ for every $t \in \mathcal{T}$. Hence, under appropriate regularity conditions, the point-wise central limit theorem $\sqrt{n}\bar{\pi}_n(t) \Rightarrow N(0,1)$ holds for all types of functional data. This motivates us to consider the normalized

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

quantity

$$Z_b(t) = \frac{\widetilde{f}_b(t)\{\widehat{\mu}_b(t) - \mu(t)\} - b^2 r(t)}{\sigma(t)/\sqrt{n}}, \tag{2.6}$$

where $r(t) = \mu''(t) f(t) \int u^2 K(u) du/2$ and $b^2 r(t)$ is the asymptotic bias term of the local linear estimator $\widehat{\mu}_b(t)$ of the mean function $\mu(t)$. Consequently, it can be shown via standard arguments (Kim and Zhao, 2013; Zhang and Wang, 2016) that

$$Z_b(t) = \sqrt{n}\bar{\pi}_n(t) + o_{\mathbb{P}}(1) \Rightarrow N(0,1). \tag{2.7}$$

With the point-wise central limit theorem (2.7) and the Bonferroni correction; see, for instance, Knafl et al. (1985), Hall and Titterington (1988), Härdle and Marron (1991), Wu et al. (1998), Chiang et al. (2001) and Huang et al. (2004) among others, one can construct conservative confidence bands for μ which, however, can be quite wide and consequently is not informative about the overall pattern of the mean curve. In Section 2.3, we shall provide an asymptotic distributional approximation of $\sup_t |Z_b(t)|$ so that the SCBs can be constructed with asymptotically correct coverage probabilities.

2.3 Asymptotic distributions

In this section, we derive the asymptotic distributional approximation of the supremum of the studentized process $\{Z_b(t)\}_{t\in\mathcal{T}}$. To this end, we develop a Gaussian approximation result for $\sup_t |Z_b(t)|$ under quite mild conditions.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

In view of (2.7), we shall first investigate the asymptotic distribution of $\sup_t |\bar{\pi}_n(t)|$. Let $\{\mathcal{G}(t)\}_{t\in\mathcal{T}}$ be a centered Gaussian process with the same covariance function as that of $\{\sqrt{n}\bar{\pi}_n(t)\}_{t\in\mathcal{T}}$, namely,

$$\operatorname{cov}\{\mathcal{G}(t), \mathcal{G}(s)\} = \operatorname{cov}\{\sqrt{n}\bar{\pi}_n(t), \sqrt{n}\bar{\pi}_n(s)\} =: \mathcal{C}(t, s).$$

Then the Gaussian analogue of $\sup_t \sqrt{n}|\bar{\pi}_n(t)|$ is defined by $\sup_t |\mathcal{G}(t)|$. It is worth noting that the covariance function $\mathcal{C}(t, s)$ depends on the numbers $m_1, \ldots, m_n$, the bandwidth b and the variance of the random errors σ_ε^2. An asymptotic expression of $\mathcal{C}(t, s)$ for any fixed interior points $t, s \in \mathcal{T}$ is given in (2.13) below.

The following theorem establishes a bound for the Kolmogorov distance between the distribution functions of $\sup_{t\in\mathcal{T}} \sqrt{n}|\bar{\pi}_n(t)|$ and its Gaussian analogue.

Theorem 1. *Let* $m_\diamond = \min_{1\le i\le n} m_i$. *Define*

$$W = \frac{\{V_{q-1}(\log n/b)^q + 1\} \wedge b^{-q}}{\vartheta^q} \quad \text{and} \quad \bar{\sigma}^2 = \frac{m_\diamond^{-1} + b}{b\vartheta^2},$$

Assume that

$$\frac{\bar{\sigma}^2 \log^4 n}{n} + \frac{(V_2/b^2 + 1)^2 \log^7 n}{n\vartheta^6} + \frac{(\log n)^{3q/2-1} W}{n^{q/2-1}} \to 0. \tag{2.8}$$

Then, under Assumptions 1–5, we have

$$\rho_{\mathcal{T}} := \sup_{y\ge 0} \left| \mathbb{P}\left\{\sup_{t\in\mathcal{T}} \sqrt{n}|\bar{\pi}_n(t)| \le y\right\} - \mathbb{P}\left\{\sup_{t\in\mathcal{T}} |\mathcal{G}(t)| \le y\right\} \right| \to 0. \tag{2.9}$$

Remark 3. It is worth mentioning that the magnitude of $\{m_1, \ldots, m_n\}$ can be completely flexible relative to the sample size n and condition (2.8) is fairly mild. To illustrate this point, suppose now $m_i = m$ for all $i = 1, \ldots, n$. Then a simple sufficient condition for (2.8) is

$$\frac{(\log n)^{5-2/q}}{(nm)^{1-2/q}(mb \vee 1)b} \to 0. \tag{2.10}$$

Moreover, for any $m \geq 1$, (2.10) holds as long as $(\log n)^{5-2/q} = o(n^{1-2/q}b)$, a fairly mild condition on the bandwidth b.

Remark 4. Our asymptotic theory sheds new light on the classical SCBs construction based on nonparametric regression with $n = 1$ (Johnston, 1982; Neumann and Polzehl, 1998; Cummins et al., 2001). In this case for the sake of estimability we assume $\nu_1(\cdot) \equiv 0$ and denote $m = m_1$. Then $\eta_1(t)$ in (2.4) becomes $\eta_1(t) = m^{-1}\sum_{j=1}^{m} K_b(t_j - t)\varepsilon_{1j}$. Let $\zeta_1, \ldots, \zeta_m$ be i.i.d. $N(0,1)$ random variables, $t_j = j/m$ and $\mathcal{G}(t) = m^{-1}\sum_{j=1}^{m} K_b(t_j - t)\zeta_j$. Assume $\|\varepsilon_{11}\|_q < \infty$ for some constant $q > 3$ and

$$\varrho := (mb)^{-1/6}(\log m)^{7/6} + (m(mb)^{-q/2})^{1/(1+q)}(\log m)^{(3q-2)/(2+2q)} \to 0,$$

or equivalently, $m^{2/q-1}(\log m)^{3-2/q} = o(b)$. Then following the arguments in the proof of Theorem 1 and applying the high-dimensional central limit theorem in Chernozhukov et al. (2017), we obtain

$$\sup_{y \geq 0} \left| \mathbb{P}\left(\sup_{t \in \mathcal{T}} |\eta_1(t)| \leq y\right) - \mathbb{P}\left(\sup_{t \in \mathcal{T}} |\mathcal{G}(t)| \leq y\right) \right| \leq C\varrho \to 0,$$

for some positive constant $C < \infty$. Our asymptotic approach allows a unified treatment of SCBs constructions for nonparametric regression and functional data analysis. Degras (2017) commented that SCBs methods of those two scenarios are quite different in general.

Remark 5. In a related work, Cao et al. (2018) derived the asymptotic distribution of the weighted average $\widetilde{\pi}_n(t) = \sum_{i=1}^n w_i \eta_i(t)/\sigma_0(t)$, where

$$w_i = m_i \left(\frac{n}{V_1} \sum_{i=1}^n m_i \right)^{-1/2} \quad \text{and} \quad \sigma_0^2(t) = \frac{f(t) V_1 \lambda_K \{\gamma(t,t) + \sigma_\varepsilon^2\}}{b}.$$

Specifically, assuming that $\max_{1 \le i \le n} m_i \lesssim \min\{(nb)^{\delta/2}, b^{-\delta}\}$ for some constant $0 < \delta < 1$, Theorem 1 in Cao et al. (2018) implies that the distribution of $\sup_{t \in \mathcal{T}} \sqrt{n} |\widetilde{\pi}_n(t)|$, after appropriate centering and normalization, is asymptotically Gumbel. In comparison, Lemma 1 is substantially more general and $\sup_{t \in \mathcal{T}} |\mathcal{G}(t)|$ may or may not be asymptotically Gumbel. In particular, the convergence rate of $\rho_{\mathcal{T}}$ in (2.9) can be much faster than the Gumbel approximation. Specifically, a careful inspection of the proof of Lemma 1 implies that when $m_i = m$ for all $i = 1, \ldots, n$, and $mb \to 0$, we have

$$\rho_{\mathcal{T}} \le C_0 \left\{ \left(\frac{(\log n)^7}{nmb} \right)^{q/(6q+6)} + \frac{1}{(nb)^3} + \left(\frac{(\log n)^{5-2/q}}{(nm)^{1-2/q} b} \right)^{q/2(q+1)} \right\},$$

where $C_0 < \infty$ is a positive constant independent of n, m and b.

Theorem 2. *Suppose the conditions of Lemma 1 and Assumption 6 are satisfied. Assume that*

$$\sup_n \frac{nb^4 \log n}{\vartheta^2} < \infty \quad \text{and} \quad \frac{(\log n)^3}{nb} \to 0.$$

Then we have

$$\rho := \sup_{y \geq 0} \left| \mathbb{P}\left\{ \sup_{t \in \mathcal{T}_b} |Z_b(t)| \leq y \right\} - \mathbb{P}\left\{ \sup_{t \in \mathcal{T}_b} |\mathcal{G}(t)| \leq y \right\} \right| \to 0. \tag{2.11}$$

Remark 6. For any significance level $\alpha \in (0,1)$, let $Q_{1-\alpha}$ denote the $(1-\alpha)$th quantile of $\sup_{t \in \mathcal{T}_b} |\mathcal{G}(t)|$, namely,

$$Q_{1-\alpha} = \inf\left\{ y \geq 0 : \mathbb{P}\left(\sup_{t \in \mathcal{T}_b} |\mathcal{G}(t)| \leq y \right) \geq 1 - \alpha \right\}.$$

Motivated by Theorem 2, an asymptotic $(1-\alpha) \times 100\%$ SCBs for $\mu(t)$, $t \in \mathcal{T}_b$, can be constructed as

$$\left[\widehat{\mu}_b(t) - \frac{b^2 \widehat{r}(t)}{\widetilde{f}_b(t)} - \frac{Q_{1-\alpha}\widehat{\sigma}(t)}{\sqrt{n}\widetilde{f}_b(t)}, \widehat{\mu}_b(t) - \frac{b^2 \widehat{r}(t)}{\widetilde{f}_b(t)} + \frac{Q_{1-\alpha}\widehat{\sigma}(t)}{\sqrt{n}\widetilde{f}_b(t)} \right], \tag{2.12}$$

where $\widehat{r}(t)$ and $\widehat{\sigma}(t)$ are consistent estimators for $r(t)$ and $\sigma(t)$, respectively. In view of (2.11), this confidence band asymptotically achieves the correct coverage probability $1-\alpha$ if $\widehat{r}(t)$ and $\widehat{\sigma}(t)$ have reasonable convergence rates such that $\sup_{t \in \mathcal{T}_b} |\widehat{r}(t) - r(t)| = o_{\mathbb{P}}\{\vartheta(nb^4 \log n)^{-1/2}\}$ and $\sup_{t \in \mathcal{T}_b} |\widehat{\sigma}(t) - \sigma(t)| = o_{\mathbb{P}}\{\vartheta(\log n)^{-1}\}$ (cf. Corollary 2). However, estimation of $r(t)$ is quite nontrivial as it involves the unknown derivative function μ''. In Section 3.4, we propose a jackknife bias-corrected procedure which can avoid estimating $r(t)$.

For any fixed interior points $t, s \in \mathcal{T}$, it follows that

$$\mathcal{C}(t,s) \approx \frac{K^*(t-s)V_1 f(t)\{\gamma(t,t)+\sigma_\varepsilon^2\}}{\sigma(t)\sigma(s)b} + \frac{f(t)f(s)\gamma(t,s)(1-V_1)}{\sigma(t)\sigma(s)}, \quad (2.13)$$

where $K^*(u) = \int K(y)K(y+u/b)dy$. Note that for different orders of the quantity V_1/b, the leading term in the covariance function $\mathcal{C}(t,s)$ can be quite different, which consequently leads to different approximating distributions of $\sup_{t\in\mathcal{T}_b}|\mathcal{G}(t)|$. In the following corollary, we derive the asymptotic distributions of $\sup_{t\in\mathcal{T}_b}|Z_b(t)|$ for different orders of V_1/b explicitly.

Corollary 1. *Suppose the conditions of Theorem 2 are satisfied. Let $\mathcal{G}_\diamond(t)$ and $\mathcal{G}_\gamma(t)$, $t \in \mathcal{T}$, be two independent zero-mean Gaussian processes with covariance functions $K^*(t-s)/\lambda_K$ and $\gamma(t,s)$, respectively.*

(i) *Suppose $b(\log n)^2/V_1 \to 0$, then*

$$\sup_{y\geq 0}\left|\mathbb{P}\left(\sup_{t\in\mathcal{T}_b}|Z_b(t)|\leq y\right) - \mathbb{P}\left(\sup_{t\in\mathcal{T}_b}|\mathcal{G}_\diamond(t)|\leq y\right)\right| \to 0.$$

(ii) *Suppose $(V_1/b-\kappa_0)(\log n)^2 \to 0$ for some positive constant $\kappa_0 < \infty$, then*

$$\sup_{y\geq 0}\left|\mathbb{P}\left(\sup_{t\in\mathcal{T}_b}|Z_b(t)|\leq y\right) - \mathbb{P}\left(\sup_{t\in\mathcal{T}_b}|\widetilde{\mathcal{G}}(t)|\leq y\right)\right| \to 0,$$

where $\widetilde{\mathcal{G}}$ is a linear combination of the processes $\mathcal{G}_\diamond$ and $\mathcal{G}_\gamma$ given by

$$\widetilde{\mathcal{G}}(t) = \frac{[\lambda_K \kappa_0 f(t)\{\gamma(t,t)+\sigma_\varepsilon^2\}]^{1/2}\mathcal{G}_\diamond(t) + f(t)\mathcal{G}_\gamma(t)}{[\lambda_K \kappa_0 f(t)\{\gamma(t,t)+\sigma_\varepsilon^2\} + f^2(t)\gamma(t,t)]^{1/2}}.$$

(iii) *Suppose $V_1(\log n)/b \to 0$ and $nb^4 \to 0$, then*

$$\sup_{y\ge 0}\left|\mathbb{P}\left(\sup_{t\in\mathcal{T}_b}\left|\frac{\sqrt{n}\{\widehat{\mu}_b(t)-\mu(t)\}}{\sqrt{\gamma(t,t)}}\right| \le y\right) - \mathbb{P}\left(\sup_{t\in\mathcal{T}_b}\left|\frac{\mathcal{G}_\gamma(t)}{\sqrt{\gamma(t,t)}}\right| \le y\right)\right| \to 0. \tag{2.14}$$

Corollary 1 depicts the dichotomous phenomenon and provides a detailed characterization of the asymptotic distributional approximations for local linear estimates in that three different approximating distributions can arise depending on the quantities V_1, b and n. It reveals the intrinsic difference between the point-wise inference based on (2.7) and the simultaneous inference in the context of functional data analysis.

For small b/V_1 with $b(\log n)^2/V_1 \to 0$, the variance function $\sigma^2(t) \approx f(t)V_1\lambda_K\{\gamma(t,t)+\sigma_\varepsilon^2\}/b \asymp V_1/b$ asymptotically. Hence the rate of convergence of $\widehat{\mu}_b(t)$ is $(nb/V_1)^{1/2}$, which is slower than the parametric rate $n^{1/2}$. Let $K_2 = \int\{K'(u)\}^2 du/(2\lambda_K)$. Applying Theorem A1 in Bickel and Rosenblatt (1973) yields the Gumbel approximation

$$\sup_y\left|\mathbb{P}\left\{(-2\log b)^{1/2}\left(\sup_{t\in\mathcal{T}_b}|Z_b(t)| - g_n\right) \le y\right\} - \exp\{-2\exp(-y)\}\right| \to 0, \tag{2.15}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

where the centering sequence

$$g_n = (-2\log b)^{1/2} + \frac{1}{(-2\log b)^{1/2}} \log \frac{K_2^{1/2}}{\sqrt{2}\pi}.$$

It is worth pointing out that the Gumbel approximation (2.15) above is of theoretical interest only. It is not recommended to use (2.15) for practical applications as the rate of convergence of Gumbel distribution can be extremely slow (Hall, 1991). In the special case where $m_1, \ldots, m_n$ are i.i.d. random variables with finite moment, Ma et al. (2012) and Zheng et al. (2014) derived similar Gumbel approximation as (2.15) for the spline and the local linear estimators of μ, respectively. Both of them require $X_i(\cdot)$ to be Gaussian process. In comparison, (2.15) is much more general in two aspects: first, we allow the m_i's to increase with n as long as $b(\log n)^2/V_1 \to 0$ is satisfied; second, instead of the Gaussian assumption of the process $X_i(\cdot)$, we only need the moment condition in Assumption 2, which is much weaker.

When $V_1(\log n)/b \to 0$, the variance function $\sigma^2(t) \approx f^2(t)\gamma(t,t) \asymp 1$. Hence the convergence rate is $\sqrt{n}$ and the effect of noise σ_ε is asymptotically negligible. In this case, the bias term $b^2 r(t)$ is asymptotically negligible as long as $nb^4 \to 0$. This is an appealing feature for practical applications. A simple sufficient condition for $V_1 \log n/b \to 0$ is

$$n^{1/4} V_1 \log n \to 0 \text{ and } nb^4 \to 0. \tag{2.16}$$

Under (2.16), Degras (2011) derived that

$$\sup_{t\in\mathcal{T}} \sqrt{n}|\widehat{\mu}_b(t) - \mu(t)| \Rightarrow \sup_{t\in\mathcal{T}} |\mathcal{G}_\gamma(t)|$$

for regularly observed and dense functional data with $m_i = m$ and $t_{ij} = t_j$, where $\int_0^{t_j} f(t)dt = (j - 0.5)/m$. Compared to (2.14), he assumed that the process $X_i(\cdot)$ satisfies a stochastic Hölder continuity condition, which is stronger than Assumption 5 used in Theorem 2; see the discussion in Remark 1. Moreover, his approach does not apply for the irregularly observed or sparse functional data.

Degras (2017) commented that constructions of SCBs can be quite different for sparse and dense functional data. Our Corollary 1 provides a more precise characterization through the magnitude of V_1/b; see Table 1 for a summary. Intuitively, V_1 is big (resp. small) when m_i are small (resp. big). However, to view the design of the data as sparse or dense in the context of SCBs construction, we also need to take into the consideration of the bandwidth b.

The particularly interesting case is the intermediate case in which $(V_1/b - \kappa_0)(\log n)^2 \to 0$. The parametric convergence rate is attained and the approximating distribution is the supremum of a linear combination of the other two cases which consists of two independent Gaussian processes $\mathcal{G}_\diamond$ and $\mathcal{G}_\gamma$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

Design		Details	Asymptotic Distribution
Sparse	(i)	b/V_1 small	$\mathcal{G}_\diamond$
Dense	(iii)	b/V_1 big	$\mathcal{G}_\gamma$
Borderline	(ii)	$b/V_1 \to$ a constant	$\widetilde{\mathcal{G}}$: a convolution of $\mathcal{G}_\diamond$ and $\mathcal{G}_\gamma$

Table 1: Types of designs for SCBs and the associated asymptotic distributions for the longitudinal and functional data in Corollary 1

On the other hand, while Corollary 1 presents an interesting theoretical characterization, it can cause unpleasant difficulties in application. It is nontrivial to determine which asymptotic approximation to use in practice. Section 3.2 proposes a unified Gaussian multiplier bootstrap approach which can automatically circumvent this issue.

3. Implementation

This section deals with various implementation issues in constructing the SCBs (2.12).

3.1 Estimation of variance function

Recall that $\sigma^2(t) = n^{-1}\sum_{i=1}^n \mathbb{E}\{\eta_i^2(t)\}$. Hence a natural moment estimator for $\sigma^2(t)$ is $n^{-1}\sum_{i=1}^n \eta_i^2(t)$. Since the η_i's are unobserved in practice, we shall substitute η_i with the consistent estimate $\widehat{\eta}_i$, which is

$$\widehat{\eta}_i(t) = \frac{1}{m_i}\sum_{j=1}^{m_i} K_b(t_{ij}-t)\{Y_{ij}-\widehat{\mu}_b(t_{ij})\}. \tag{3.1}$$

Consequently, we obtain the practically feasible estimator

$$\widehat{\sigma}^2(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\eta}_i(t)^2. \tag{3.2}$$

In the following theorem, we show that $\widehat{\sigma}(t)$ is uniformly consistent for $\sigma(t)$ for all types of functional data, with a uniform convergence rate.

Theorem 3. *Let Assumptions 1–6 hold with $q > 4$. Define*

$$\begin{aligned}
\Delta_0 &= b^2 + \left(\frac{\vartheta^2 \log n}{n}\right)^{1/2} + \frac{W^{1/q}\vartheta \log n}{n^{1-1/q}},\\
\Delta_1 &= 1 + \frac{V_1 \log n}{b} + \frac{\{V_{2q-1}(\log n/b)^q + V_q\}^{1/q} \wedge (V_q^{1/q}/b)}{bn^{1-1/q}/(\log n)^2},\\
\Delta_2 &= \left(\frac{V_3 \log n}{nb^3} + \frac{\log n}{n}\right)^{1/2} + \frac{W^{2/q}\vartheta^2 \log n}{n^{1-2/q}}.
\end{aligned}$$

Assume that $\Delta_0 = o(1)$ and $\Delta_0\Delta_1 + \Delta_2 = o(\vartheta^2)$. Then we have

$$\sup_{t\in\mathcal{T}} \left|\frac{\widehat{\sigma}(t)}{\sigma(t)} - 1\right| = O_{\mathbb{P}}\left(\frac{\Delta_0\Delta_1 + \Delta_2}{\vartheta^2}\right) = o_{\mathbb{P}}(1).$$

Remark 7. Relation (2.5) suggests an alternative estimator

$$\frac{\widehat{f}(t)V_1\lambda_K\{\widehat{\gamma}(t,t)+\widehat{\sigma}_\varepsilon^2\}}{b} + \widehat{f}(t)\widehat{f}(t)\widehat{\gamma}(t,t)(1-V_1),$$

where $\widehat{f}(t)$, $\widehat{\gamma}(t,t)$ are $\widehat{\sigma}_\varepsilon^2$ are uniformly consistent estimators of $f(t)$, $\gamma(t,t)$ and σ_ε^2, respectively. For example, we can take $\widehat{f}(t) = S_0(t)$ and the unified

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

estimators $\widehat{\gamma}(t,t)$ and $\widehat{\sigma}_\varepsilon^2$ in Li and Hsing (2010). Although the uniform consistency of this estimator can be similarly established, it is not recommended for practical use as the nonparametric estimator $\widehat{\gamma}(t,t)$ above involves additional bandwidth selection, which is highly nontrivial in practice. Moreover, Kim and Zhao (2013) found that $\widehat{\gamma}(t,t)$ might be negative for some $t \in \mathcal{T}$, especially when the noise level σ_ε is high.

Remark 8. The conditions of Theorem 3 are quite mild and are easy to verify. For ease of illustration, suppose now $m_i = m$ for all $i = 1, \ldots, n$. Then a simple sufficient condition for $\Delta_0 = o(1)$ and $\Delta_0\Delta_1 + \Delta_2 = o(\vartheta^2)$ is

$$\frac{b^2 \log n}{mb \vee 1} + \frac{(\log n)^3}{(nm)^{1-2/q}(mb \vee 1)b} \to 0,$$

which is satisfied as long as $b^2 \log n = o(1)$ and $(\log n)^3 = o(n^{1-2/q}b)$.

Remark 9. By (2.7) and Theorem 3, we have the central limit theorem

$$\widehat{Z}_b(t) = \frac{\widetilde{f}_b(t)\{\widehat{\mu}_b(t) - \mu(t)\} - b^2 r(t)}{\widehat{\sigma}(t)/\sqrt{n}} \Rightarrow N(0,1).$$

Consequently, for any fixed interior point $t \in \mathcal{T}$, we can construct the $(1-\alpha) \times 100\%$ point-wise confidence interval for $\mu(t)$ as

$$\left[\widehat{\mu}_b(t) - \frac{b^2\widehat{r}(t)}{\widetilde{f}_b(t)} - \frac{\Phi^{-1}(1-\alpha/2)\widehat{\sigma}(t)}{\sqrt{n}\widetilde{f}_b(t)}, \widehat{\mu}_b(t) - \frac{b^2\widehat{r}(t)}{\widetilde{f}_b(t)} + \frac{\Phi^{-1}(1-\alpha/2)\widehat{\sigma}(t)}{\sqrt{n}\widetilde{f}_b(t)}\right],$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

Combining Theorem 2 and Theorem 3 yields the following Gaussian approximation for $\sup_{t\in\mathcal{T}_b}|\widehat{Z}_b(t)|$.

Corollary 2. *Suppose the conditions of Theorem 2 are satisfied and*

$$\Delta_0\Delta_1 + \Delta_2 = o\left(\frac{\vartheta^2}{\log n}\right).$$

Then we have

$$\widehat{\rho} := \sup_{y\geq 0}\left|\mathbb{P}\left\{\sup_{t\in\mathcal{T}_b}|\widehat{Z}_b(t)|\leq y\right\} - \mathbb{P}\left\{\sup_{t\in\mathcal{T}_b}|\mathcal{G}(t)|\leq y\right\}\right| \to 0.$$

3.2 Gaussian multiplier bootstrap

As discussed in Corollary 1, the asymptotic distributions of $\sup_{t\in\mathcal{T}_b}|Z_b(t)|$ take different forms under different sampling schemes and they depend on the magnitudes of the unknown quantities $\gamma(t,s)$, $f(t)$, $\sigma(t)$ and σ_ε^2. This phenomenon causes challenges for practical implementation as subjective choice between different scenarios may result in erroneous conclusions. It is therefore of both practical and theoretical importance to develop a unified approach that works for all types of functional data. To this end, we adopt the Gaussian multiplier bootstrap (Chernozhukov et al., 2013, 2014) to approximate the critical value $Q_{1-\alpha}$ and justify its validity. It is worth mentioning that the proposed bootstrap procedure is fully data-driven and easy to implement in practice.

Our procedure is summarized as follows.

1. Select the bandwidth b by the cross-validation procedure in Section 3.3. Obtain the nonparametric estimate $\widehat{\mu}_b(t)$ for $\mu(t)$ via (2.1). Compute $\widehat{\eta}_i(t)$ based on (3.1) and then calculate the estimate $\widehat{\sigma}(t)$ via (3.2).

2. Generate i.i.d. $N(0,1)$ random variables $z_1, \ldots, z_n$ and calculate
$$\widehat{G}(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(t)} \sum_{i=1}^{n} z_i \widehat{\eta}_i(t).$$

3. Repeat step 2 for B times to obtain $\widehat{G}_1(t), \ldots, \widehat{G}_B(t)$, and then calculate the $(1-\alpha)$th quantile $\widehat{Q}_{1-\alpha} = \inf\{y : \widehat{F}_B(y) \geq 1-\alpha\}$, where
$$\widehat{F}_B(y) = \frac{1}{B} \sum_{h=1}^{B} \mathbf{1}\left\{\sup_{t\in\mathcal{T}_b} |\widehat{G}_h(t)| \leq y\right\}.$$

The following theorem justifies the asymptotic validity of the bootstrap procedure.

Theorem 4. *Assume the conditions of Theorem 2 hold with $q > 4$ and*
$$\Delta_0\Delta_1 + \Delta_2 = o\left(\frac{\vartheta^2}{(\log n)^2}\right).$$
Then we have
$$\sup_{y\geq 0}\left|\mathbb{P}\left\{\sup_{t\in\mathcal{T}_b} |\widehat{Z}_b(t)| \leq y\right\} - \mathbb{P}_z\left\{\sup_{t\in\mathcal{T}_b} |\widehat{G}(t)| \leq y\right\}\right| = o_{\mathbb{P}}(1),$$
where $\mathbb{P}_z(\cdot) = \mathbb{P}(\cdot|\mathcal{D})$ stands for the conditional probability.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

Remark 10. Li and Hsing (2010) and Zhang and Wang (2016) considered an alternative simulation-based approach to estimate $Q_{1-\alpha}$. Let $\widehat{\gamma}(t,s)$ be a uniformly consistent estimator for $\gamma(t,s)$. In view of (2.13), a natural plug-in estimator for $\mathcal{C}(t,s)$ is

$$\widehat{\mathcal{C}}(t,s) = \frac{\mathcal{K}^*(t-s)V_1\widehat{f}(t)\{\widehat{\gamma}(t,t)+\widehat{\sigma}_\varepsilon^2\}}{\widehat{\sigma}(t)\widehat{\sigma}(s)b} + \frac{\widehat{f}(t)\widehat{f}(s)\widehat{\gamma}(t,s)(1-V_1)}{\widehat{\sigma}(t)\widehat{\sigma}(s)}.$$

Then one can estimate $Q_{1-\alpha}$ by the $(1-\alpha)$th quantile of $\sup_{t\in\mathcal{T}_b}|\widehat{\mathcal{G}}(t)|$, where $\mathcal{G}(t)$ is a centered Gaussian process with covariance function $\widehat{\mathcal{C}}(t,s)$. The drawback of this approach is that the nonparametric covariance function estimator $\widehat{\gamma}(t,s)$ is typically not guaranteed to be non-negative semidefinite and involves additional tuning parameter selection. In comparison, our estimate is always non-negative semidefinite and does not require additional tuning parameter selection.

3.3 Bandwidth selection

The performance of both the local linear estimates and the proposed SCBs of the mean function depends on the bandwidth b. In this section, we discuss how to choose the bandwidth b in practice. Since the measurements from the same subject are correlated, the conventional leave-one-observation-out cross-validation is not suitable in the current context. Instead, we shall adopt the leave-one-subject-out cross-validation procedure (Rice and

Silverman, 1991). Specifically, for any bandwidth $b > 0$, the cross-validation score is defined by

$$\mathrm{cv}(b) = \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \left\{ Y_{ij} - \widehat{\mu}_b^{-i}(t_{ij}) \right\}^2,$$

where $\widehat{\mu}_b^{-i}(t)$ is the local linear estimator for $\mu(t)$ computed following (2.1) using all the observations except those of ith subject. Consequently the cross-validation bandwidth b_{cv} is obtained by minimizing $\mathrm{cv}(b)$ with respect to b, that is,

$$b_{\mathrm{cv}} = \arg\min_{b>0} \mathrm{cv}(b).$$

The main advantage of this approach is that it can preserve the dependence structure of the process $\nu_i(\cdot)$ for each $i = 1, \ldots, n$. Similar ideas have been widely used in the context of functional and longitudinal data analysis; see, for instance, Hoover et al. (1998), Chiang et al. (2001), Wu and Zhang (2002), Huang et al. (2004), Kim and Zhao (2013) and Cao et al. (2018) among others.

Following the arguments in Hoover et al. (1998), we now provide an intuitive justification of the above cross-validation procedure. Define the weighted average square error for the estimator $\widehat{\mu}_b(t)$ as

$$\mathrm{ASE}(b) = \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \{\mu(t_{ij}) - \widehat{\mu}_b(t_{ij})\}^2.$$

Intuitively, the bandwidth $b_{\rm cv}$ should minimize ASE(b) asymptotically. Indeed, observe that

$$\begin{aligned}\mathrm{cv}(b) &= \sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\left\{\mu(t_{ij})-\widehat{\mu}_b^{-i}(t_{ij})\right\}^2+\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}e_{ij}^2\\ &\quad+\sum_{i=1}^{n}\frac{2}{m_i}\sum_{j=1}^{m_i}\left\{\mu(t_{ij})-\widehat{\mu}_b^{-i}(t_{ij})\right\}e_{ij}, \qquad (3.3)\end{aligned}$$

where $e_{ij}=\nu_i(t_{ij})+\varepsilon_{ij}$. The second term on the right-hand side of (3.3) is independent of the bandwidth b. And the third term is stochastically dominated by the first term when n is sufficiently large as $\mu(t_{ij})-\widehat{\mu}_b^{-i}(t_{ij})$ and e_{ij} are independent. Therefore the bandwidth $b_{\rm cv}$ asymptotically minimizes the first term on the right hand side of (3.3), which is an approximation of ASE(b).

3.4 Bias correction

To apply Theorem 2 in practice, we need to deal with the bias term $b^2r(t)$ in (2.6), which is the asymptotic mean of $\widetilde{f}_b(t)\{\widehat{\mu}_b(t)-\mu(t)\}$ in view of (2.7). The estimation for $r(t)$, in particular for $\mu''(t)$, can be highly nontrivial. Here we shall use the idea of jackknife and avoid estimating $r(t)$. Noting that $(\sqrt{2}b)^2r(t)=2b^2r(t)$ is the asymptotic mean of $\widetilde{f}_{\sqrt{2}b}(t)\{\widehat{\mu}_{\sqrt{2}b}(t)-\mu(t)\}$, we can propose the jackknife estimate

$$\widetilde{r}(t)=\frac{\widetilde{f}_b(t)\widetilde{f}_{\sqrt{2}b}(t)\{\widehat{\mu}_{\sqrt{2}b}(t)-\widehat{\mu}_b(t)\}}{\{2\widetilde{f}_b(t)-\widetilde{f}_{\sqrt{2}b}(t)\}b^2}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

Then our bias-corrected estimator for $\mu(t)$ is defined as

$$\widetilde{\mu}_b(t) = \widehat{\mu}_b(t) - \frac{b^2\widetilde{r}(t)}{\widetilde{f}_b(t)} = \frac{2\widetilde{f}_b(t)\widehat{\mu}_b(t) - \widetilde{f}_{\sqrt{2}b}(t)\widehat{\mu}_{\sqrt{2}b}(t)}{2\widetilde{f}_b(t) - \widetilde{f}_{\sqrt{2}b}(t)}.$$

Under conditions of Theorem 2, it can be shown that the bias of $\widetilde{\mu}_b(t)$ is of order $o(b^2)$ and it is asymptotically negligible. Note that implementing estimator $\widetilde{\mu}_b(t)$ is asymptotically equivalent to using the higher order kernel $\widetilde{K}(u) = 2K(u) - K(u/\sqrt{2})/\sqrt{2}$. Denote $\widetilde{K}_b(u) = \widetilde{K}(u/b)/b$. Performing the bootstrap procedure in Section 3.2, with $\widehat{\eta}_i(t)$ and $\widehat{\sigma}^2(t)$ replaced by

$$\widetilde{\eta}_i(t) = \frac{1}{m_i}\sum_{j=1}^{m_i} \widetilde{K}_b(t_{ij} - t)\{Y_{ij} - \widehat{\mu}_b(t_{ij})\} \text{ and } \widetilde{\sigma}^2(t) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{\eta}_i(t)^2,$$

respectively, we obtain the estimated critical value $\widetilde{Q}_{1-\alpha}$. Consequently, we construct the SCBs for $\mu(t)$ as

$$\left[\widetilde{\mu}_b(t) - \frac{\widetilde{Q}_{1-\alpha}\widetilde{\sigma}(t)}{\sqrt{n}\{2\widetilde{f}_b(t) - \widetilde{f}_{\sqrt{2}b}(t)\}}, \widetilde{\mu}_b(t) + \frac{\widetilde{Q}_{1-\alpha}\widetilde{\sigma}(t)}{\sqrt{n}\{2\widetilde{f}_b(t) - \widetilde{f}_{\sqrt{2}b}(t)\}}\right]. \quad (3.4)$$

4. Simulation Study

In this section, we conduct a Monte Carlo simulation study to assess the finite sample performance of the proposed SCBs in (3.4). For comparison, we also implement the point-wise confidence intervals (with $\widetilde{Q}_{1-\alpha}$ replaced by $\Phi^{-1}(1-\alpha/2)$ in (3.4)), the Bonferroni corrected confidence bands based on $\mathcal{N} = 199$ equispaced points of the interval $[0, 1]$ (with $\widetilde{Q}_{1-\alpha}$ replaced by

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

$\Phi^{-1}(1-\alpha/(2\mathcal{N}))$ in (3.4)) and the SCBs constructed based on the Gumbel approximation in (2.15). The explicit expression of the SCBs is given in (S7.1) in which we plug in the true bias function $b^2 r(t)$ for ease of illustration. For convenience, we denote these four different methods in the tables below as Gmb, Point, Bonf and Gumbel respectively. Throughout this section, we use the Epanechnikov kernel $K(u) = 0.75(1-u^2)\mathbf{1}\{|u| \leq 1\}$.

We generate data from the model

$$Y_{ij} = \mu(t_{ij}) + \sum_{l=1}^{4} \omega_l \zeta_{il} \phi_l(t_{ij}) + \varepsilon_{ij},$$

where $\{\zeta_{il}\}_{i,l\in\mathbb{N}}$ and $\{\varepsilon_{ij}\}_{i,j\in\mathbb{N}}$ are independent copies of ζ in (4.1) and are scaled such that $\mathrm{var}(\zeta_{11}) = 1$ and $\mathrm{var}(\varepsilon_{11}) = 0.1$,

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

$\omega_l = 0.4/(l+1)$ for $l \geq 1$, $\{t_{ij}\}_{i,j\in\mathbb{N}}$ are i.i.d. Uniform$[0,1]$ random variables and

$$\mu(t) = \sin(\pi t) + t + \frac{\cos(2\pi t) + \sin(2\pi t)}{4}.$$

We consider three different distributions of ζ as follows,

$$(1)\;\; \zeta \sim N(0,1); \quad (2)\;\; \zeta \sim t_5; \quad (3)\;\; \zeta \sim (\chi^2_5 - 5). \tag{4.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

To investigate the impact of different sampling schemes on the performance of the proposed method, we consider three different setups of the m_i's as follows, ranging from sparse cases to dense cases,

$$\text{case 1} : m_i \sim \text{Uniform}\{4, 5, 6\};$$

$$\text{case 2} : m_i \sim \text{Uniform}\{\lfloor n^{1/4} \rfloor \times 4, \lfloor n^{1/4} \rfloor \times 6\};$$

$$\text{case 3} : m_i = \lfloor n/4 \rfloor,$$

where $\lfloor u \rfloor$ represents the integer part of u. For each case, we take the sample size $n = 200$ and the bootstrap size $B = 1000$. The bandwidth b_{cv} is taken as the median of 100 cross-validation bandwidths selected via the cross-validation procedure (cf. Section 3.3) based on 100 independent replications. To demonstrate the robustness of our method against the bandwidth selection, we consider three different bandwidths $c_0 \times b_{\text{cv}}$ with $c_0 \in \{0.9, 1.0, 1.1\}$.

To evaluate the four different confidence bands, we compute the empirical coverage probabilities based on 2000 independent replications for each case. The complete results are summarized in Tables 3–5 for the three different distributions of ζ respectively in online Supplementary Material. Here we only present the case with $c_0 = 1$ in Table 2. It is demonstrated that our proposed method performs well for both Gaussian and non-Gaussian func-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

		Gmb		Bonf		Point		Gumbel	
ζ	$1-\alpha$	90%	95%	90%	95%	90%	95%	90%	95%
$N(0,1)$	case 1	87.45	93.50	98.05	98.75	19.15	43.45	91.20	96.35
	case 2	88.80	94.20	98.70	99.25	19.90	42.10	93.40	97.55
	case 3	89.40	94.80	98.55	99.20	27.45	50.00	95.55	98.55
t_5	case 1	88.55	93.80	98.40	99.20	21.90	45.95	91.15	96.75
	case 2	89.50	94.65	98.35	99.15	19.45	43.05	93.65	97.70
	case 3	89.25	94.60	98.75	99.15	27.65	50.15	94.70	98.20
(χ_5^2-5)	case 1	87.40	93.35	98.45	99.00	21.40	43.10	91.35	97.05
	case 2	87.75	93.50	98.50	99.20	21.65	43.10	93.50	97.60
	case 3	90.05	95.15	98.60	99.25	26.05	51.60	95.95	98.35

Table 2: Empirical coverage probabilities for the four different confidence bands of the mean curve.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

tional data. In particular, our method is robust to bandwidth selection in that the empirical coverage probabilities of our SCBs are close to the nominal level $1-\alpha$. In comparison, the SCBs constructed based on the Gumbel approximation works relatively well in the sparse case which matches the theoretical result for sparse scenario (i) in Corollary 1. However, when the observations become denser, as we can see from Tables 2, the Gumbel SCBs no longer works well and becomes quite conservative even with the true bias function plugged in. In contrast, our method works quite well for all the three different sampling schemes, ranging from sparse to dense cases.

5. Conclusions

This paper studies the simultaneous inference of the mean functions for longitudinal and functional data. Unlike most of the existing methods which focused on only one type (sparse or dense) of functional data, we propose a unified approach to constructing the simultaneous confidence bands for the mean curves and develop a unified distribution theory for the local linear estimates. In particular, our result depicts the dichotomous phenomenon and provides a detailed characterization of the asymptotic distributional approximations, ranging from sparse to dense cases. For practical implementation, we propose a Gaussian multiplier bootstrap procedure to esti-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0358

mate the critical value which is easy to implement and works well for all types of functional data. Our framework of simultaneous inference is quite general and can be easily extended to study more complicated scenarios in which the longitudinal/functional data are adaptively collected or are temporal dependent over time.

6. Supplementary Material

The online supplementary material contains the technical proofs for all the theoretical results presented in the previous sections.

References

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

Cai, T. T. and M. Yuan (2011). Optimal estimation of the mean function based on discretely sampled functional data: phase transition. *Ann. Statist. 39*(5), 2330–2355.

Cao, G., L. Yang, and D. Todem (2012). Simultaneous inference for the mean function based on dense functional data. *J. Nonparametr. Stat. 24*(2), 359–377.

Cao, H., W. Liu, and Z. Zhou (2018). Simultaneous nonparametric regression analysis of sparse longitudinal data. *Bernoulli 24*(4A), 3013–3038.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist. 41*(6), 2786–2819.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist. 42*(5), 1787–1818.

Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab. 45*(4), 2309–2352.

Chiang, C.-T., J. A. Rice, and C. O. Wu (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc. 96*(454), 605–619.

Choi, H. and M. Reimherr (2018). A geometric approach to confidence regions and bands for functional parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 80*(1), 239–260.

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

Cummins, D. J., T. G. Filloon, and D. Nychka (2001). Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage. *J. Amer. Statist. Assoc. 96*(453), 233–246.

Degras, D. (2017). Simultaneous confidence bands for the mean of functional data. *Wiley Interdiscip. Rev. Comput. Stat. 9*(3), e1397, 15.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica 21*(4), 1735–1765.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Ferraty, F. and P. Vieu (2006). *Nonparametric functional data analysis: theory and practice.* Springer Science & Business Media.

Guo, W. (2002). Functional mixed effects models. *Biometrics 58*(1), 121–128.

Hall, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields 89*(4), 447–455.

Hall, P., H.-G. Müller, and J.-L. Wang (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist. 34*(3), 1493–1517.

Hall, P. and D. M. Titterington (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal. 27*(1), 228–254.

Härdle, W. and J. S. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist. 19*(2), 778–796.

Hoover, D. R., J. A. Rice, C. O. Wu, and L.-P. Yang (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika 85*(4), 809–822.

Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.

Hsing, T. and R. Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators.* John Wiley & Sons.

Huang, J. Z., C. O. Wu, and L. Zhou (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica 14*(3), 763–788.

Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multivariate Anal. 12*(3), 402–414.

Kim, S. and Z. Zhao (2013). Unified inference for sparse and dense longitudinal models. *Biometrika 100*(1), 203–212.

Knafl, G., J. Sacks, and D. Ylvisaker (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc. 80*(391), 683–691.

Li, Y. and T. Hsing (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist. 38*(6), 3321–3351.

Ma, S., L. Yang, and R. J. Carroll (2012). A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica 22*(1), 95–122.

Neumann, M. H. and J. Polzehl (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparametr. Statist. 9*(4), 307–333.

Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis.* Springer-Verlag New York.

Ramsay, J. O. and B. W. Silverman (2007). *Applied functional data analysis: methods and case studies.* Springer.

Rice, J. A. and B. W. Silverman (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B 53*(1), 233–243.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). Functional data analysis. *Annual Review of Statistics and Its Application 3*, 257–295.

Wang, Q. (2021). Two-sample inference for sparse functional data. *Electron. J. Stat. 15*(1), 1395–1423.

Wu, C. O., C.-T. Chiang, and D. R. Hoover (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc. 93*(444), 1388–1402.

Wu, H. and J.-T. Zhang (2002). Local polynomial mixed-effects models for longitudinal data. *J. Amer. Statist. Assoc. 97*(459), 883–897.

Wu, H. and J.-T. Zhang (2006). *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*, Volume 515. John Wiley & Sons.

Yao, F., H.-G. Müller, and J.-L. Wang (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc. 100*(470), 577–590.

Yao, F., H.-G. Müller, and J.-L. Wang (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist. 33*(6), 2873–2903.

Zhang, J.-T. and J. Chen (2007). Statistical inferences for functional data. *Ann. Statist. 35*(3), 1052–1079.

Zhang, X. and J.-L. Wang (2016). From sparse to dense functional data and beyond. *Ann.*

Statist. *44*(5), 2281–2321.

Zheng, S., L. Yang, and W. K. Härdle (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *J. Amer. Statist. Assoc.* *109*(506), 661–673.

Department of Statistics, University of Pittsburgh, Pittsburgh, Pittsburgh, PA 15260

E-mail: zhl318@pitt.edu

Department of Statistics, University of Chicago, Chicago, IL 60637.

E-mail: wbwu@galton.uchicago.edu