Statistica Sinica **33** (2023), 519-549
doi:https://doi.org/10.5705/ss.202021.0107

STATISTICAL INFERENCE FOR FUNCTIONAL TIME SERIES

Jie Li and Lijian Yang

Renmin University of China and Tsinghua University

Abstract: We investigate statistical inference for the mean function of stationary functional time series data with an infinite moving average structure. We propose a B-spline estimation for the temporally ordered trajectories of the functional moving average, which are used to construct a two-step estimator of the mean function. Under mild conditions, the B-spline mean estimator enjoys oracle efficiency in the sense that it is asymptotically equivalent to the infeasible estimator, that is, the sample mean of all trajectories observed entirely without errors. This oracle efficiency allows us to construct a simultaneous confidence band (SCB) for the mean function, which is asymptotically correct. Simulation results strongly corroborate the asymptotic theory. Using the SCB to analyze an electroencephalogram time series reveals strong evidence of a trigonometric form of the mean function.

Key words and phrases: B-spline, electroencephalogram, functional moving average, oracle efficiency, simultaneous confidence band.

1. Introduction

Functional data analysis (FDA) has garnered much research in the last two decades, extending the statistical analysis of multivariate data to more complicated and informative curve data; see Ferraty and Vieu (2006), Ramsay and Silverman (2002), Ramsay and Silverman (2005), Hsing and Eubank (2015), and Kokoszka and Reimherr (2017). Mathematically speaking, classical functional data consist of a collection of n trajectories $\{\eta_t(\cdot)\}_{t=1}^n$ corresponding to n subjects, where the tth trajectory $\eta_t(\cdot)$ for subject t is a continuous stochastic process equal in distribution to a standard process $\eta(\cdot)$. These trajectories $\{\eta_t(\cdot)\}_{t=1}^n$ play the role of univariate and multivariate random observations associated with individual subjects in most textbooks on introductory statistics. Thus one may be interested in predicting other numerical or categorical outcomes based on such random curves or, at a more basic level, measuring the location and scale of these curves. The latter consists of the mean and covariance functions $m(\cdot) = \mathbb{E}\{\eta(\cdot)\}$ and $G(x, x') = \text{Cov}\{\eta(x), \eta(x')\}$, respectively, of $\eta(\cdot)$, and has been studied in

Corresponding author: Lijian Yang, Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China. E-mail: yanglijian@tsinghua.edu.cn.

Cao, Yang and Todem (2012), Cao et al. (2016), and Zheng, Yang and Härdle (2014), who derive pointwise normal confidence intervals and simultaneous confidence bands (SCBs) have been derived for $m(\cdot)$ and $G(\cdot,\cdot)$ based on various limiting distributions.

Much of the above is done using "raw" functional data $\{Y_{tj}\}$, where Y_{tj} represents the discretely recorded value of the t-th trajectory $\eta_t(\cdot)$ at the j-th x-location X_{tj}, contaminated with measurement error $\sigma(X_{tj})\varepsilon_{tj}$,

$$Y_{tj} = \eta_t(X_{tj}) + \sigma(X_{tj})\varepsilon_{tj},\ 1 \le t \le n,\ 1 \le j \le N_t. \tag{1.1}$$

Therefore, the "raw data" are not a collection of curves $\{\eta_t(\cdot)\}_{t=1}^n$, which may be considered the "smooth data." In the case of densely recorded "raw" data, the trajectories $\{\eta_t(\cdot)\}_{t=1}^n$ can be estimated one at a time to produce something that resembles $\{\eta_t(\cdot)\}_{t=1}^n$, which might be best referred to as "smoothed pseudo data." To be precise, the "raw data" takes the form

$$Y_{tj} = \eta_t\left(\frac{j}{N}\right) + \sigma\left(\frac{j}{N}\right)\varepsilon_{tj},\ 1 \le t \le n,\ 1 \le j \le N, \tag{1.2}$$

where both N and n go to infinity. Spline estimates $\{\widehat{\eta}_t(\cdot)\}_{t=1}^n$ are obtained in Cao, Yang and Todem (2012), that is the "smoothed pseudo data" that can be used as substitutes for $\{\eta_t(\cdot)\}_{t=1}^n$ in data analysis. Without loss of generality, the functions $\eta(\cdot)$ and $\{\eta_t(\cdot)\}_{t=1}^n$ are defined on $[0,1]$, and $G(\cdot,\cdot)$ is defined on $[0,1]^2$. The trajectories $\{\eta_t(\cdot)\}_{t=1}^n$ are decomposed as $\eta_t(x) = m(x) + \xi_t(x)$, where $m(\cdot)$ is continuous on $[0,1]$, and $\xi_t(x)$ is a small-scale variation of x on the tth trajectory, a process with a continuous sample path $\mathbb{E}\xi_t(x) = 0, \mathbb{E}\max_{x\in[0,1]}\xi_t^2(x) < \infty$, and continuous covariance $G(x,x') = \text{Cov}\{\xi_t(x), \xi_t(x')\}$.

According to Hsing and Eubank (2015), there exist eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$, $\sum_{k=1}^\infty \lambda_k < \infty$, with corresponding eigenfunctions $\{\psi_k\}_{k=1}^\infty$ of $G(\cdot,\cdot)$, the latter being an orthonormal basis of $L^2[0,1]$, $G(x,x') = \sum_{k=1}^\infty \lambda_k\psi_k(x)\psi_k(x')$, $\int G(x,x')\ \psi_k(x')\,dx' = \lambda_k\psi_k(x)$. The process $\eta(x)$, for $x \in [0,1]$, then allows the well-known Karhunen–Loève L^2 representation $\eta(x) = m(x) + \sum_{k=1}^\infty \xi_k\phi_k(x)$, in which the random coefficients $\{\xi_k\}_{k=1}^\infty$, called functional principal component (FPC) scores, are uncorrelated, with mean zero and variance one. The rescaled eigenfunctions ϕ_k are called FPCs, $\phi_k = \sqrt{\lambda_k}\psi_k$, for $k \ge 1$.

Mean estimation is usually the essential first step in functional data analysis; see Ma, Yang and Carroll (2012) and Zheng, Yang and Härdle (2014) for theory and applications of sparse longitudinal data, and Cao, Yang and Todem (2012) for an SCB for the mean function based on dense functional data. One serious

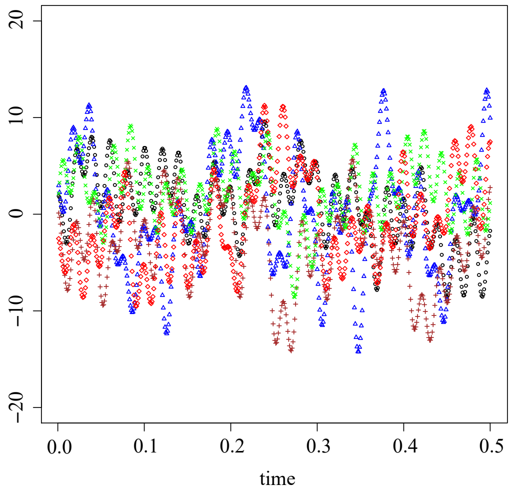

Figure 1. Five smoothed trajectories for the EEG data.

drawback in Cao et al. (2016) and Cao, Yang and Todem (2012) is the assumption that the number of records on each curve is dominated by the number of curves, that is $N = \mathcal{O}\left(n^{\theta}\right)$, for some $\theta > 0$. This constraint is unreasonable, because it prevents observing each subject more densely, whereas larger N is always more preferable, owing to the increased measurement precision, regardless of whether n is large or small (consider the limiting case $N = \infty$ when observations are made arbitrarily dense in the entire range). In our current work, we instead assume that the assumption $n = \mathcal{O}\left(N^{\theta}\right)$, with the more logical and natural understanding that the speed at which n asymptotics become significant is contingent upon the precision level set by N.

Existing works also restrict $\{\eta_t(\cdot)\}_{t=1}^{n}$ to be independent and identically distributed (i.i.d.) copies of the process $\eta(\cdot)$. As discussed in Bosq (2000), functional data do not always come in the form of i.i.d. replicates. One interesting example is the continued recording of an electroencephalogram (EEG) for a person in a resting eyes-closed state. The participant went through a five-minute test, and EEG signals were recorded at a 1,000 Hz sample rate from 32 scalp locations. Observations at the sixth location are divided into 400 consecutive segments, each consisting of $N = 500$ EEG signals recorded every 0.001 second. Figure 1 shows five randomly selected "smoothed pseudo data" from the 400 smooth curves $\{\widehat{\eta}_t(\cdot)\}_{t=1}^{400}$.

Horváth, Kokoszka and Reeder (2013) developed the asymptotic theory for testing the equality of two mean functions in functional samples exhibiting temporal dependence, assuming that the entire trajectories $\{\eta_t(\cdot)\}_{t=1}^n$ are fully observed. SCBs of mean functions of functional time series are constructed in Chen and Song (2015), but there are two major gaps in their theoretical development. First, they had required the number of positive eigenvalues to be finite, limiting the scope of applicability. Second, they do not specify clear assumptions on the FPC scores. In particular, the physical dependence condition of (2.1) does not ensure independence of all FPC scores ξ_k, for $k = 1, 2, \ldots$, which is a key condition if the strong Gaussian approximations in (A.3) of Lemma A.5 for all ξ_{tk}, $k = 1, 2, \ldots, k_n$; need to be jointly independent so that their linear combinations are Gaussian as well.

To set up an appropriate framework that accounts for dependence between trajectories $\{\eta_t(\cdot)\}_{t=1}^n$, the infinite moving average MA(∞) concept in classic time series analysis is extended to the functional setting. Specifically, the demeaned trajectories $\{\xi_t(\cdot)\}_{t=1}^n$ are regarded as a segment of zero mean processes $\{\xi_t(\cdot)\}_{t=-\infty}^{\infty}$ satisfying the functional moving average (FMA(∞)) equations

$$\xi_t(\cdot) = \sum_{t'=0}^{\infty} A_{t'} \zeta_{t-t'}(\cdot), \ t = 0, \pm 1, \pm 2, \ldots, \tag{1.3}$$

in which $A_{t'}$ are bounded linear operators $\mathcal{L}^2[0,1] \to \mathcal{L}^2[0,1]$ and $\{\zeta_t(\cdot)\}_{t=-\infty}^{\infty}$ are i.i.d. zero mean processes, that is, strong white noise according to Bosq (2000). Note that the classic MA(∞) is a rather broad category, and includes as a special case the widely used causal ARMA(p, q); see Theorem 3.1.1 of Brockwell and Davis (1991).

In the next section, under the above dependence structure, we propose B-spline estimator for the mean function $m(\cdot)$. B-splines are widely used in nonparametrics for their computational ease and conceptional simplicity; see de Boor (2001) and DeVore and Lorentz (1993). It is established in Theorem 2 that the B-spline estimator is as efficient as an infeasible "oracle" estimator, which is obtained as if all random trajectories are totally observed without measurement errors. Theorem 1 and Corollary 1 further establish an asymptotically correct SCB of the mean function $m(\cdot)$ under some mild conditions. SCB is a powerful tool for quantifying the variability of functions and making global inferences for functions; see Cao et al. (2016), Cao, Yang and Todem (2012), and Degras (2011) for SCBs of dense functional data, Gu et al. (2014), Ma, Yang and Carroll (2012), and Zheng, Yang and Härdle (2014) for SCBs of sparse functional data, Choi and Reimherr (2018) for the ghost region of functional parameters, and Gu and Yang (2015), Wang (2012), Wang and Yang (2009), Zheng et al. (2016), Wang et al.

(2020), and Yu et al. (2020) for applications and theory of SCB in other contexts. Applying the proposed SCB to the aforementioned EEG data set reveals strong evidence that the mean function is of a simple trigonometric form; see Section 6.

The reminder of the paper is organized as follows. Section 2 states the main theoretical results on an SCB constructed from a B-spline estimator. Section 3 contains a decomposition of the difference between the B-spline estimator and the infeasible one, so that each component can be easily bounded. Procedures to implement the proposed SCB are given in Section 4 with details. Section 5 documents our simulation findings, and an empirical study of the EEG functional time series using the proposed SCB is reported in Section 6. Section 7 concludes the paper. All technical proofs are collected in the Appendix.

2. Main Results

This section introduces the B-spline estimator for the mean function $m(\cdot)$ and studies the asymptotic properties of the proposed estimator.

The FMA(∞) operators A_t in (1.3) are of the form

$$A_t\left\{\sum_{k=1}^{\infty} c_k\phi_k(\cdot)\right\} = \sum_{k=1}^{\infty} a_{tk}c_k\phi_k(\cdot),\ a_{tk}\in\mathbb{R},\ k=1,2,\ldots,\ t=0,1,\ldots,$$

with geometrically decaying MA coefficients $|a_{tk}| < C_a\rho_a^t$ for constants $C_a > 0$, for $\rho_a \in (0,1)$, $k = 1,2,\ldots$, and $t = 0,1,\ldots$. Note that the geometric decay is not as restrictive as it might seem, because it holds for the MA coefficients of the causal ARMA model, according to equation (3.3.6) of Brockwell and Davis (1991).

The strong functional white noise $\{\zeta_t(\cdot)\}_{t=-\infty}^{\infty}$ allows for its own Karhunen–Loève representation $\zeta_t(\cdot) = \sum_{k=1}^{\infty}\zeta_{t,k}\phi_k(\cdot)$, where $\{\zeta_{t,k}\}_{t=-\infty,k=1}^{\infty,\infty}$ are uncorrelated random variables with mean zero and variance one. Together with (1.3), we have

$$\begin{aligned}\xi_t(\cdot) &= \sum_{t'=0}^{\infty} A_{t'}\left\{\sum_{k=1}^{\infty}\zeta_{t-t',k}\phi_k(\cdot)\right\} = \sum_{t'=0}^{\infty}\sum_{k=1}^{\infty} a_{t',k}\zeta_{t-t',k}\phi_k(\cdot)\\ &= \sum_{k=1}^{\infty}\left(\sum_{t'=0}^{\infty} a_{t',k}\zeta_{t-t',k}\right)\phi_k(\cdot),\end{aligned}$$

which can be the Karhunen–Loève representation for $\{\xi_t(\cdot)\}_{t=1}^{n}$, as long as each random coefficient $\sum_{t'=0}^{\infty} a_{t',k}\zeta_{t-t',k}$ is white noise, that is, has variance one. This can be achieved by assuming $\sum_{t=0}^{\infty} a_{tk}^2 \equiv 1$, for $k = 1,2,\ldots$, which is reasonably analogous to what is assumed in numerical MA(∞).

With these in mind and (1.2), the FMA(∞) model is

$$\begin{aligned} Y_{tj} &= m\left(\frac{j}{N}\right) + \xi_t\left(\frac{j}{N}\right) + \sigma\left(\frac{j}{N}\right)\varepsilon_{tj} \\ &= m\left(\frac{j}{N}\right) + \sum_{k=1}^{\infty}\xi_{tk}\phi_k\left(\frac{j}{N}\right) + \sigma\left(\frac{j}{N}\right)\varepsilon_{tj},\ 1\le t\le n,\ 1\le j\le N, \end{aligned} \tag{2.1}$$

where for $1 \le t \le n, k = 1, 2, \ldots$,

$$\xi_t(\cdot) = \sum_{k=1}^{\infty}\xi_{tk}\phi_k(\cdot), \quad \xi_{tk} = \sum_{t'=0}^{\infty} a_{t'k}\zeta_{t-t',k} \quad a.s. \tag{2.2}$$

For any non-negative integer q and fraction $\mu \in (0, 1]$, denote by $\mathcal{C}^{(q,\mu)}[0, 1]$ the space of functions with a μ-Hölder continuous qth derivative, that is,

$$\mathcal{C}^{(q,\mu)}[0,1] = \left\{\varphi : [0,1] \to \mathbb{R} \left| \|\varphi\|_{q,\mu} = \sup_{x,y\in[0,1],x\neq y}\left|\frac{\varphi^{(q)}(x) - \varphi^{(q)}(y)}{|x-y|^{\mu}}\right| < +\infty \right.\right\}.$$

Because $m(\cdot)$ and $\phi_k(\cdot)$ both belong to $\mathcal{C}^{(q,\mu)}[0, 1]$ under Assumptions (A1) and (A3) below, $\eta(\cdot)$ can be regarded as $\mathcal{C}^{(q,\mu)}[0, 1]$-random variables. Had the trajectories $\eta_t(\cdot)$, for $1 \le t \le n$, been all observed over the interval $[0, 1]$ entirely, the population mean function $m(\cdot)$ could have been estimated by the sample mean of n random variables valued in $\mathcal{C}^{(q,\mu)}[0, 1]$,

$$\overline{m}(x) = n^{-1}\sum_{t=1}^{n}\eta_t(x),\ x \in [0,1]. \tag{2.3}$$

This "estimator" is infeasible because it uses unobservables. However, it serves as a benchmark.

To describe the spline functions, denote by $\{t_\ell\}_{\ell=1}^{J_s}$ a sequence of equally-spaced points, $t_\ell = \ell/(J_s + 1)$, for $0 \le \ell \le J_s + 1$, $0 = t_0 < t_1 < \cdots < t_{J_s} < 1 = t_{J_s+1}$, called interior knots, that divide the interval $[0, 1]$ into $(J_s + 1)$ equal subintervals $I_\ell = [t_\ell, t_{\ell+1})$, for $\ell = 0, \ldots, J_s - 1$, $I_{J_s} = [t_{J_s}, 1]$. Let $\mathcal{H}^{(p-2)} = \mathcal{H}^{(p-2)}[0, 1]$ be the polynomial spline space of order p on I_ℓ, for $\ell = 0, \ldots, J_s$, that consists of all $(p - 2)$ times continuously differentiable functions on $[0, 1]$ that are polynomials of degree $(p - 1)$ on subintervals I_ℓ, for $\ell = 0, \ldots, J_s$. Then, we denote by $\{B_{\ell,p}(\cdot), 1 \le \ell \le J_s + p\}$ the pth-order B-spline basis functions of $\mathcal{H}^{(p-2)}$; hence, $\mathcal{H}^{(p-2)} = \{\sum_{\ell=1}^{J_s+p}\lambda_{\ell,p}B_{\ell,p}(\cdot)|\lambda_{\ell,p} \in \mathbb{R}\}$.

The individual trajectories can be estimated via B-spline as

$$\widehat{\eta}_t(\cdot) \equiv \sum_{\ell=1}^{J_s+p} \widehat{\beta}_{\ell,p,t} B_{\ell,p}(\cdot),\ 1 \leq t \leq n, \tag{2.4}$$

with the coefficients $\{\widehat{\beta}_{1,p,t}, \ldots, \widehat{\beta}_{J_s+p,p,t}\}^\top$ solving the following least squares problem:

$$\left\{\widehat{\beta}_{1,p,t}, \cdots, \widehat{\beta}_{J_s+p,p,t}\right\}^\top = \underset{\{\beta_{1,p},\ldots,\beta_{J_s+p,p}\}\in R^{J_s+p}}{\text{argmin}} \sum_{j=1}^{N} \left\{Y_{tj} - \sum_{\ell=1}^{J_s+p} \beta_{\ell,p} B_{\ell,p}\left(\frac{j}{N}\right)\right\}^2. \tag{2.5}$$

The mean function $m(\cdot)$ is then estimated by an oracle estimator,

$$\widehat{m}(\cdot) = n^{-1} \sum_{t=1}^{n} \widehat{\eta}_t(\cdot), \tag{2.6}$$

which mimics the infeasible estimator $\overline{m}(\cdot)$ in (2.3).

Throughout this paper, $a_n \asymp b_n$ means that $a_n = \mathcal{O}(b_n)$ and $b_n = \mathcal{O}(a_n)$, as $n \to \infty$. For any measurable function $\phi(\cdot)$ defined on $[0,1]$, denote $\|\phi\|_\infty = \sup_{x\in[0,1]} |\phi(x)|$. Denote by I_n an integer-valued truncation index for the white noise sequence ζ_{tk}, for $-\infty < t \leq n$, which satisfies $I_n > -10 \log n / \log \rho_a$, $I_n \asymp \log n$. We next introduce some technical assumptions.

(A1) The mean function $m(\cdot) \in \mathcal{C}^{(q,\mu)}[0,1]$ for an integer $q > 0$ and a constant $\mu \in (0,1]$. We denote $p^* = q + \mu$ in what follows.

(A2) The standard deviation function $\sigma(\cdot) \in \mathcal{C}^{(0,\nu)}[0,1]$ for $\nu \in (0,1]$, and $c_\sigma \leq \sigma(x) \leq C_\sigma, \forall x \in [0,1]$, for constants $0 < c_\sigma < C_\sigma < \infty$.

(A3) There exists a constant $\theta > 0$ such that as $N \to \infty$, $n = n(N) \to \infty$, $n = \mathcal{O}\left(N^\theta\right)$.

(A4) There exists $C_G > 0$ such that $G_\varphi(x,x) \geq C_G$, $\forall x \in [0,1]$, with $G_\varphi(x,x)$ defined in (2.7). The FPCs $\phi_k(\cdot) \in \mathcal{C}^{(q,\mu)}[0,1]$, with $\sum_{k=1}^\infty \|\phi_k\|_{0,\mu} < +\infty$, $\sum_{k=1}^\infty \|\phi_k\|_{q,\mu} < +\infty$, and $\sum_{k=1}^\infty \|\phi_k\|_\infty < +\infty$; for increasing positive integers $\{k_n\}_{n=1}^\infty$, as $n \to \infty$, $\sum_{k_n+1}^\infty \|\phi_k\|_\infty = o(n^{-1/2})$ and $k_n = \mathcal{O}(n^\omega)$ for some $\omega > 0$.

(A5) There are constants $C_1,\ C_2 \in (0,+\infty)$, $\gamma_1,\ \gamma_2 \in (1,+\infty)$, and $\beta_1,\ \beta_2 \in (0,1/2)$, and i.i.d. $N(0,1)$ variables $\{Z_{tk,\zeta}\}_{t=-I_n+1,k=1}^{n,k_n}$, $\{Z_{tj,\varepsilon}\}_{t=1,j=1}^{n,N}$, where I_n is the truncation index such that

$$\mathbb{P}\left\{\max_{1\le k\le k_n}\max_{-I_n+1\le\tau\le n}\left|\sum_{t=-I_n+1}^{\tau}\zeta_{tk}-\sum_{t=-I_n+1}^{\tau}Z_{tk,\zeta}\right|>n^{\beta_1}\right\}<C_1n^{-\gamma_1},$$

$$\mathbb{P}\left\{\max_{1\le t\le n}\max_{1\le\tau\le N}\left|\sum_{j=1}^{\tau}\varepsilon_{tj}-\sum_{j=1}^{\tau}Z_{tj,\varepsilon}\right|>N^{\beta_2}\right\}<C_2N^{-\gamma_2}.$$

(A5′) The i.i.d. variables $\{\varepsilon_{tj}\}_{t\ge1,j\ge1}$ are independent of the FPC score white noise $\{\zeta_{tk}\}_{t\ge1,k\ge1}$. The number of distinct distributions for all FPC score white noise $\{\zeta_{tk}\}_{t\ge1,k\ge1}$ is finite. There exist constants $r_1>4+2\omega$ and $r_2>4+2\theta$, for ω in Assumption (A4) and θ in Assumption (A3), such that $\mathbb{E}|\zeta_{1k}|^{r_1}$ and $\mathbb{E}|\varepsilon_{11}|^{r_2}$ are finite.

(A6) The spline order $p\ge p^*$, the number of interior knots $J_s=N^{\gamma}d_N$, for some $\gamma>0$, with $d_N+d_N^{-1}=\mathcal{O}\left(\log^{\vartheta}N\right)$ for some $\vartheta>0$ as $N\to\infty$, and for p^* in Assumption (A1), ν in Assumption (A2), θ in Assumption (A3), β_2 in Assumption (A5), and r_1 in Assumption (A5′),

$$\max\left\{1-\nu,\frac{\theta}{2p^*}+\frac{2\theta}{r_1p^*}\right\}<\gamma<1-\theta/2-\beta_2.$$

Assumptions (A1) and (A2) are typical for spline smoothing. In particular, (A1) controls the size of the bias of the spline smoother for $m(\cdot)$, and (A2) requires that the variance function is uniformly bounded on its domain. Assumption (A3) restricts the sample size n to increase by a fractional power θ of N, the number of observations for each subject. The collective bounded smoothness of the principal components is stated in Assumption (A4). Assumption (A5) provides the strong Gaussian approximation of the estimation errors and the FPC score white noise $\{\zeta_{tk}\}_{t=-\infty,k=1}^{\infty,\infty}$. The high level of Assumption (A5) can be ensured by an elementary Assumption (A5′). The requirement for the number of knots of the splines is stipulated in Assumption (A6), which ensures the smoothness of the B-spline estimator.

Remark 1. The assumptions above are quite mild because they can easily be satisfied in various practical situations. We propose one simple and reasonable setting for the above parameters q, μ, θ, p, and γ, as follows: $q+\mu=p^*=4$, $\nu=1$, $\theta=1$, $p=4$ (cubic spline), $\gamma=1/4$ and $d_N\asymp\log\log N$. These constants are used as defaults in the implementation, see Section 4.

Define a limiting covariance function

$$G_\varphi\left(x, x'\right) = \sum_{k=1}^{\infty} \phi_k(x)\phi_k(x') \left\{1 + 2\sum_{t=0}^{\infty} \sum_{t'=t+1}^{\infty} a_{tk}a_{t'k}\right\}, \quad x, x' \in [0,1], \qquad (2.7)$$

and for i.i.d. standard normal variables $\{U_k\}_{k=1}^{\infty}$, denote a Gaussian process

$$\varphi\left(x\right) = \frac{\sum_{k=1}^{\infty}\sum_{t=1}^{\infty} a_{tk}U_k\phi_k\left(x\right)}{G_\varphi\left(x,x\right)^{1/2}}, \quad x \in [0,1].$$

Then, $\varphi\left(x\right)$ satisfies $\mathbb{E}\varphi\left(x\right) \equiv 0$, $\mathbb{E}\varphi^2\left(x\right) \equiv 1$, and $x \in [0,1]$, with covariance function

$$\mathbb{E}\varphi\left(x\right)\varphi\left(x'\right) = G_\varphi\left(x, x'\right)\left\{G_\varphi\left(x,x\right)G_\varphi\left(x',x'\right)\right\}^{-1/2}, \quad x, x' \in [0,1].$$

For any $\alpha \in (0,1)$, define $z_{1-\alpha/2}$ as the $100\left(1-\alpha/2\right)$th percentile of the standard normal distribution. Denote by $Q_{1-\alpha}$ the $100\left(1-\alpha\right)$th percentile of the absolute maxima distribution of $\varphi\left(x\right)$ over $[0,1]$, that is,

$$\mathbb{P}\left[\sup_{x\in[0,1]} \left|\varphi\left(x\right)\right| \le Q_{1-\alpha}\right] = 1-\alpha. \qquad (2.8)$$

The following result establishes how well $m(\cdot)$ could be estimated had all trajectories $\eta_t\left(\cdot\right)$, for $1 \le t \le n$, been fully observed without error, and is used to compute the infeasible "oracle" estimator $\overline{m}(\cdot)$:

Theorem 1. *Under Assumptions* (A1), (A3)–(A5), *for* $\alpha \in (0,1)$, *as* $n \to \infty$, *the infeasible estimator* $\overline{m}(\cdot)$ *converges at the* $\sqrt{n}$ *rate to* $m(\cdot)$ *with asymptotic covariance function* $G_\varphi\left(x,x'\right)$, *and thus*

$$\mathbb{P}\left\{\sup_{x\in[0,1]} n^{1/2}\left|\overline{m}(x) - m(x)\right| G_\varphi\left(x,x\right)^{-1/2} \le Q_{1-\alpha}\right\} \to 1-\alpha,$$
$$\mathbb{P}\left\{n^{1/2}\left|\overline{m}(x) - m(x)\right| G_\varphi\left(x,x\right)^{-1/2} \le z_{1-\alpha/2}\right\} \to 1-\alpha, \quad x \in [0,1].$$

The next result enables one to construct an SCB based on $\widehat{m}\left(\cdot\right)$ in (2.6) by showing that $\widehat{m}\left(\cdot\right)$ has the same asymptotic property as $\overline{m}\left(\cdot\right)$ in (2.3), so there is no need to differentiate between the two.

Theorem 2. *Under Assumptions* (A1)–(A6), *as* $n \to \infty$, *the B-spline estimator* $\widehat{m}\left(\cdot\right)$ *is oracally efficient, that is, it is asymptotically equivalent to* $\overline{m}\left(\cdot\right)$ *up to order* $\mathcal{O}_p\left(n^{-1/2}\right)$

$$\sup_{x\in[0,1]} n^{1/2} \left|\overline{m}(x) - \widehat{m}(x)\right| = \mathcal{O}_p(1).$$

Corollary 1. *Under Assumptions* (A1)–(A6)*, for any* $\alpha \in (0,1)$*, as* $n \to \infty$*, an asymptotic* $100(1-\alpha)\%$ *correct confidence band for* $m(\cdot)$ *is*

$$\widehat{m}(x) \pm G_\varphi(x,x)^{1/2} Q_{1-\alpha} n^{-1/2}, \quad x \in [0,1], \tag{2.9}$$

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

$$\widehat{m}(x) \pm G_\varphi(x,x)^{1/2} z_{1-\alpha/2} n^{-1/2}, \quad x \in [0,1].$$

3. Decompositon

In this section, we decompose the estimation error $\widehat{\eta}_t(x) - \eta_t(x)$ into three convenient terms. For any L^2 integrable functions $\phi(x)$ and $\varphi(x)$, for $x \in [0,1]$, define their theoretical inner product as $\langle \phi, \varphi \rangle = \int_{[0,1]} \phi(x)\varphi(x)dx$, and the empirical inner product as $\langle \phi, \varphi \rangle_N = N^{-1}\sum_{j=1}^N \phi(j/N)\varphi(j/N)$. The related theoretical and empirical norms are $\|\phi\|_2^2 = \langle \phi, \phi \rangle$and $\|\phi\|_{2,N}^2 = \langle \phi, \phi \rangle_N$. For any function $\varphi(x)$ defined on $[0,1]$, denote its discretization by $\boldsymbol{\varphi} = \{\varphi(1/N), \ldots, \varphi(N/N)\}^\top$, that is, the vector of its values on the N measurement points. In particular,

$$\boldsymbol{\eta}_t = \left\{\eta_t\left(\frac{1}{N}\right), \ldots, \eta_t\left(\frac{N}{N}\right)\right\}^\top, \quad \mathbf{m} = \left\{m\left(\frac{1}{N}\right), \ldots, m\left(\frac{N}{N}\right)\right\}^\top,$$
$$\boldsymbol{\xi}_t = \left\{\xi_t\left(\frac{1}{N}\right), \ldots, \xi_t\left(\frac{N}{N}\right)\right\}^\top, \quad \boldsymbol{\eta}_t = \mathbf{m} + \boldsymbol{\xi}_t. \tag{3.1}$$

Matrix algebra represents the B-spline estimator $\widehat{\eta}_t(\cdot)$ in (2.4) as

$$\widehat{\eta}_t(x) = \{B_{1,p}(x), \ldots, B_{J_s+p,p}(x)\}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}_t, \tag{3.2}$$

where $Y_t = (Y_{t1}, \ldots, Y_{tN})^\top$ and the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} B_{1,p}\left(\frac{1}{N}\right) & \cdots & B_{J_s+p,p}\left(\frac{1}{N}\right) \\ \vdots & \cdots & \vdots \\ B_{1,p}\left(\frac{N}{N}\right) & \cdots & B_{J_s+p,p}\left(\frac{N}{N}\right) \end{pmatrix}_{N\times(J_s+p)}. \tag{3.3}$$

Define the empirical inner product matrix of the B-spline basis $\{B_{\ell,p}(x)\}_{\ell=1}^{J_s+p}$ as

$$\mathbf{V}_{n,p} = \left\{\langle B_{\ell,p}, B_{\ell',p}\rangle_N\right\}_{\ell,\ell'=1}^{J_s+p} = N^{-1}\mathbf{X}^\top\mathbf{X}, \tag{3.4}$$

and according to Lemma A.3 in Cao, Yang and Todem (2012), for some constant

$C_p > 0$,

$$\left\|\mathbf{V}_{n,p}^{-1}\right\|_\infty \leq C_p J_s. \tag{3.5}$$

Denote $\boldsymbol{\varepsilon}_t = \{\sigma(1/N)\varepsilon_{t1}, \ldots, \sigma(N/N)\varepsilon_{tN}\}^\top$ and $\mathbf{B}(x) = \{B_{1,p}(x), \ldots, B_{J_s+p,p}(x)\}^\top$. Then, the approximation error $\widehat{\eta}_t(x) - \eta_t(x)$ is decomposed according to (3.1) as

$$\widehat{\eta}_t(x) - \eta_t(x) = \widetilde{\eta}_t(x) - \eta_t(x) + \widetilde{\varepsilon}_t(x), \tag{3.6}$$

where

$$\widetilde{\eta}_t(x) = N^{-1}\mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1}\mathbf{X}^\top \boldsymbol{\eta}_t = \widetilde{m}(x) + \widetilde{\xi}_t(x), \tag{3.7}$$

$$\widetilde{m}(x) = N^{-1}\mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1}\mathbf{X}^\top \mathbf{m}, \quad \widetilde{\xi}_t(x) = N^{-1}\mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1}\mathbf{X}^\top \boldsymbol{\xi}_t, \tag{3.8}$$

$$\widetilde{\varepsilon}_t(x) = N^{-1}\mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1}\mathbf{X}^\top \boldsymbol{\varepsilon}_t. \tag{3.9}$$

Thus, one has $\widehat{\eta}_t(x) - \eta_t(x) = \widetilde{\xi}_t(x) - \xi_t(x) + \widetilde{m}(x) - m(x) + \widetilde{\varepsilon}_t(x)$. Therefore, by (2.3) and (2.6), the approximation error of $\widehat{m}(\cdot)$ in (2.6) to $\overline{m}(\cdot)$ is

$$\widehat{m}(x) - \overline{m}(x) = n^{-1}\sum_{t=1}^{n}\{\widetilde{\eta}_t(x) - \eta_t(x) + \widetilde{\varepsilon}_t(x)\}. \tag{3.10}$$

4. Implementation

This section describes procedures to implement the SCB in Corollary 1.

4.1. Knots selection

The number of knots is an important smoothing parameter, and is selected using the AIC.

According to Remark 1, $\gamma = 1/4$ and $d_N \asymp \log\log N$ meet Assumption (A6), for γ and d_N, with J_s being of order $N^{1/4}\log\log N$. Thus, we propose selecting a data-driven $\widehat{J}_s$ from the integers in $[0.8N_r, \min(10N_r, n/2)]$ using the AIC, with $N_r = N^{1/4}\log\log N$. Specifically, given any data set $(j/N, Y_{tj})_{j=1,t=1}^{N,n}$ from model (1.2), denote the estimator for the jth response Y_{tj} by $\widehat{Y}_{tj}(N_n) = \widehat{\eta}_t(j/N)$, for $j = 1, \ldots, N$. The trajectory estimate $\widehat{\eta}_t$ depends on the knot selection sequence, as given in (2.4). Then, $\widehat{J_{s,t}}$ for the tth curve is the one minimizing the AIC value

$$\widehat{J_{s,t}} = \underset{N_n \in [0.8N_r, \min(10N_r, n/2)]}{\operatorname{argmin}} \operatorname{AIC}(N_n), \quad t = 1, \ldots, n, \tag{4.1}$$

where $\operatorname{AIC}(N_n) = \log(\operatorname{RSS}/N) + 2(N_n + p)/N$, with the residual sum of squares $\operatorname{RSS} = \sum_{j=1}^{N}\{Y_{tj} - \widehat{Y}_{tj}(N_n)\}^2$. Then, $\widehat{J}_s$ is set as the median of $\{\widehat{J_{s,t}}\}_{t=1}^{n}$.

The spline estimator $\widehat{\eta}_t(\cdot)$ is obtained from (3.2) by using the selected number of knots $\widehat{J}_s$, and the estimator $\widehat{m}(\cdot)$ is computed from (2.6).

4.2. Covariance estimation

Denote $\widehat{\xi}_t\left(x\right) = \widehat{\eta}_t\left(x\right) - \widehat{m}\left(x\right)$, for $t = 1, \ldots, n$, and $x \in [0, 1]$. To estimate the covariance function $G_\varphi\left(x, x'\right)$, divide $\{\widehat{\xi}_t\left(\cdot\right)\}_{t=1}^n$ into l groups in order, where each group has B samples with $B = [n^{1/5}]$ and $l = [n/B]$, where $[a]$ denotes the integer part of a. Noting that $\widehat{G}_\varphi\left(\cdot, \cdot\right)$ is the limit of the covariance function of the process $\sqrt{n}\left(\overline{m}(\cdot) - \widehat{m}(\cdot)\right)$, we use $\widehat{m}\left(x\right)$ to mimic $m\left(x\right)$ and $\sqrt{B}\ \widehat{\delta}_j\left(x\right)$ to mimic the points from the process $\sqrt{n}\left(\overline{m}(\cdot) - \widehat{m}(\cdot)\right)$, where

$$\widehat{\delta}_j\left(x\right) = \frac{1}{B} \sum_{k=B(j-1)+1}^{Bj} \widehat{\xi}_k\left(x\right), \quad j = 1, \ldots, l, \quad x \in [0, 1].$$

The estimator $\widehat{G}_\varphi\left(x, x'\right)$ of $G_\varphi\left(x, x'\right)$ is defined as

$$\widehat{G}_\varphi\left(x, x'\right) = \frac{B}{l} \sum_{j=1}^{l} \left\{\widehat{\delta}_j\left(x\right)\widehat{\delta}_j\left(x'\right) - \overline{\widehat{\delta}}\left(x\right)\overline{\widehat{\delta}}\left(x'\right)\right\}, \quad x, x' \in [0, 1], \tag{4.2}$$

where $\overline{\widehat{\delta}}\left(x\right) = l^{-1}\sum_{j=1}^{l} \widehat{\delta}_j\left(x\right), x \in [0, 1]$. Because the consistency of $\widehat{G}_\varphi\left(x, x'\right)$ is straightforward, the proof is omitted to save space.

4.3. Estimating the percentile

To estimate the percentile $Q_{1-\alpha}$, we first obtain the estimated eigenvalues $\widehat{\lambda}_{k,\varphi}$ and eigenfunctions $\widehat{\psi}_{k,\varphi}$ of $\widehat{G}_\varphi\left(x, x'\right)$ using $N^{-1}\sum_{j=1}^{N} \widehat{G}_\varphi(j/N, j'/N)\widehat{\psi}_{k,\varphi}(j/N) = \widehat{\lambda}_{k,\varphi}\widehat{\psi}_{k,\varphi}\left(j'/N\right)$. Next, we choose the number κ of eigenfunctions using the following standard criterion: $\kappa = \text{argmin}_{1\le l\le T}\{\sum_{k=1}^{l} \widehat{\lambda}_{k,\varphi}/\sum_{k=1}^{T} \widehat{\lambda}_{k,\varphi} > 0.95\}$, where $\{\lambda_{k,\varphi}\}_{k=1}^T$ are the first T estimated positive eigenvalues.

We then generate $\widehat{\zeta}_b\left(x\right) = \widehat{G}_\varphi\left(x, x\right)^{-1/2}\sum_{k=1}^{\kappa} Z_{k,b}\widehat{\phi}_{k,\varphi}\left(x\right)$, where $\widehat{\phi}_{k,\varphi} = \widehat{\lambda}_{k,\varphi}^{1/2}\widehat{\psi}_{k,\varphi}$ and $Z_{k,b}$ are i.i.d. standard normal variables with $1 \le k \le \kappa$ and $b = 1, \ldots, b_M$, where b_M is a preset large integer with default value 1,000. We take the maximal absolute value for each copy of $\widehat{\zeta}_b\left(x\right)$ and use the empirical quantile $\widehat{Q}_{1-\alpha}$ of these maximum values as an estimate of $Q_{1-\alpha}$.

Finally, the SCB for the mean function is computed as

$$\widehat{m}(x) \pm n^{-1/2}\widehat{G}_\varphi\left(x, x\right)^{1/2}\widehat{Q}_{1-\alpha}, \quad x \in [0, 1]. \tag{4.3}$$

5. Simulation

In this section, simulation studies are conducted to illustrate the finite-sample performance of the proposed method. The data are generated from the following model:

$$Y_{tj} = m\left(\frac{j}{N}\right) + \sum_{k=1}^{2} \xi_{tk}\phi_k\left(\frac{j}{N}\right) + \sigma\left(\frac{j}{N}\right)\varepsilon_{tj},\ 1 \le j \le N,\ 1 \le t \le n. \quad (5.1)$$

Case 1: $m(x) = 10 + \sin\{2\pi(x - 1/2)\}$, $\varepsilon_{tj} \sim N(0,1)$, for $1 \le t \le n$ and $1 \le j \le N$, $\phi_1(x) = -2\cos\{\pi(x - 1/2)\}$, and $\phi_2(x) = \sin\{\pi(x - 1/2)\}$. This setting implies $\lambda_1 = 2$ and $\lambda_2 = 0.5$. Here, $\{\xi_{tk}\}_{t=1,k=1}^{n,2}$ are generated from (2.2), where $\{\zeta_{tk}\}_{t=0,k=1}^{n,2}$ are i.i.d. $N(0,1)$ variables and

$$a_{0k} = 0.8, \quad a_{1k} = 0.6, \quad a_{tk} = 0,\ \forall t \ge 2, \quad k = 1, 2.$$

The number of curves n is taken to be 100, 400, 900, and 1600, and the number of observations per curve N is taken to be 120, 500, 1000, and 2000. The noise level is set to include both homoscedastic $\sigma(x) = 0.3$ and $\sigma(x) = 0.5$, and heteroscedastic $\sigma(x) = (\exp(x) - 0.9)/(\exp(x) + 0.9)$ and $\sigma(x) = 0.1\sin(2\pi x) + 0.2$.

Case 2: We set $m(x) = 0.4\sin\{50\pi(x - 1/2)\}$ to mimic the data example in Section 6, with ε_{tj}, $\phi_1(x)$, $\phi_2(x)$, and $\{\xi_{tk}\}_{t=1,k=1}^{n,2}$ the same as those in Case 1. The number of curves n is taken to be 100, 200, 300, and 400, and the number of observations per curve N is taken to be 500. The noise level is $\sigma = 0.005$.

Throughout this section, the mean function is estimated using cubic splines, that is, $p = 4$. Each simulation is repeated 1,000 times.

To visualize the SCBs for the mean function, Figure 2 shows the estimated mean functions and their 95% SCBs for the true curve $m(\cdot)$ in Case 1 when $\sigma = 0.3$ and $n = 100, 400, 900$, and 1600, with the true curve shown as solid, and the estimated curve and the SCBs shown as dashed. As expected, when n increases, the SCB becomes narrower and the cubic spline estimators are closer to the true curve. In all panels, the true mean function is entirely covered by the SCBs.

Tables 1 to 3 display the empirical coverage rate, that is, the percentage out of the 1,000 replications in which the true curve $m(\cdot)$ is covered by the cubic spline SCBs (4.3) at the N points $\{1/N, \ldots, (N-1)/N, 1\}$. It is shown in Tables 1 and 2 for Case 1 that regardless of the noise level and/or form, the coverage

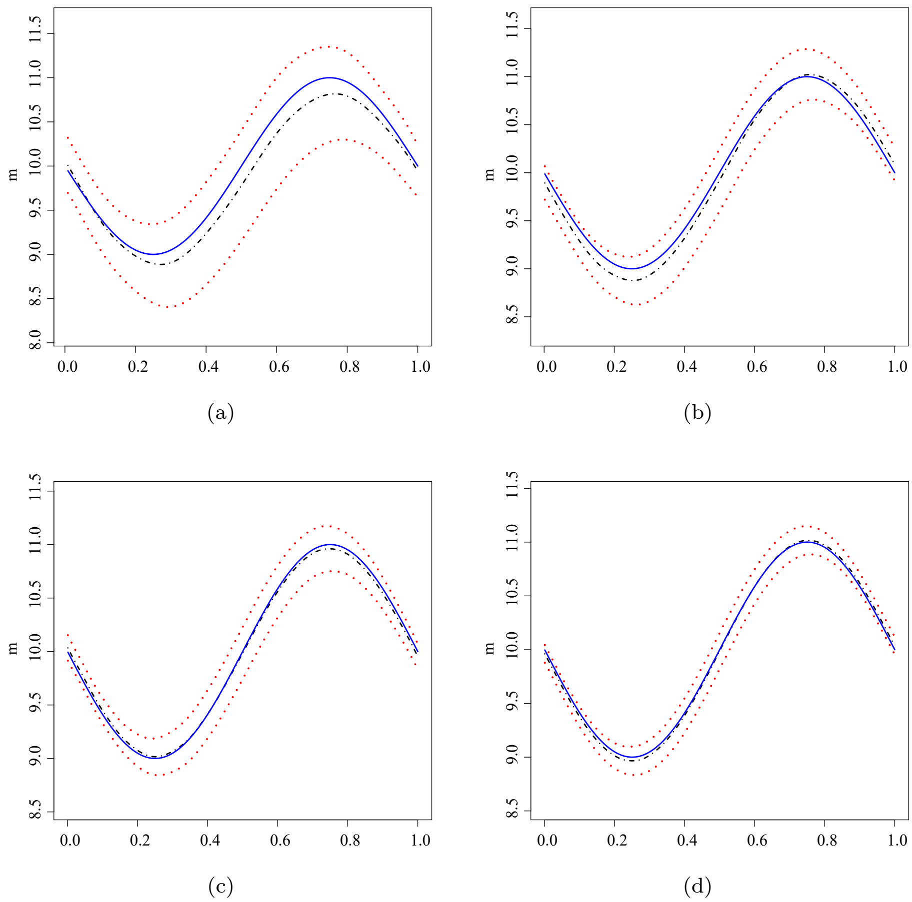

Figure 2. Plots of the cubic estimator in (2.6) for the simulated data (dash) and 95% SCBs in (4.3) (dotted), for $m(x)$ (solid). The number of observations N of (a)–(d) are 100, 400, 900, and 1600, respectively. In all panels, $\sigma = 0.3$.

Table 1. Coverage frequencies of the SCB in (4.3) with $p = 4$, Case 1.

	$\sigma = 0.3$				$\sigma = 0.5$			
(n, N)	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
(100,120)	0.962	0.897	0.840	0.708	0.957	0.896	0.831	0.692
(400,500)	0.974	0.921	0.871	0.770	0.971	0.919	0.867	0.763
(900,1000)	0.976	0.925	0.882	0.764	0.976	0.925	0.879	0.763
(1600,2000)	0.990	0.943	0.893	0.804	0.990	0.942	0.892	0.802

rate of the SCB becomes closer to the nominal confidence level as the sample size increases. The results in Table 3 for Case 2 are very similar, providing a positive confirmation of the asymptotic theory.

Table 2. Coverage frequencies of the SCB in (4.3) with $p = 4$, Case 1.

	$\sigma(x) = (\exp(x) - 0.9) / (\exp(x) + 0.9)$				$\sigma(x) = 0.1\sin(2\pi x) + 0.2$			
(n, N)	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
(100,120)	0.960	0.895	0.837	0.702	0.963	0.896	0.841	0.713
(400,500)	0.959	0.911	0.845	0.748	0.961	0.912	0.847	0.748
(900,1000)	0.984	0.930	0.882	0.752	0.984	0.912	0.862	0.740
(1600,2000)	0.986	0.942	0.896	0.792	0.986	0.940	0.882	0.784

Table 3. Coverage frequencies of the SCB in (4.3) with $p = 4$, Case 2.

(n, N)	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
(100,500)	0.966	0.882	0.826	0.722
(200,500)	0.972	0.912	0.844	0.714
(300,500)	0.976	0.920	0.854	0.766
(400,500)	0.980	0.938	0.872	0.761

6. Real Data Analysis

The SCB methodology is further illustrated using EEG data collected by the research group of Prof. Linhong Ji at Tsinghua University Department of Mechanical Engineering. EEG is known to provide rich information about brain function. For the study, 145 university students were recruited and EEG signals were recorded from 32 scalp locations based on the international 10/20 system of electrode placement. The experiment required participants to go through a five-minute closed-eye resting state while the EEG was being recorded at a sample rate of 1,000 Hz.

We have selected one person's EEG at the sixth scalp location, using the mid-portion 200,000 signals divided into 400 consecutive segments of 500 recordings each. Each piece can be regarded as an FMA trajectory, with $N = 500$ recordings and $n = 400$ trajectories. The data range is from -29.2 to 22.8, with an estimated noise level of 0.026; thus, the signal-to-noise ratio is around 2,000, close to that of Case 2 in Section 5. While conceding that there are other reasonable choices of n and N, we compute the coefficient of determination R^2 for the B-spline trajectory against the raw EEG data at each of the 400 segments. The sample minimum, 25th percentile, median, 75th percentile and maximum of the 400 R^2's are 0.9993, 0.9996, 0.9997, 0.9998, and 0.9999, respectively, showing very good fits for all times t. Four randomly selected trajectories are shown in Figure 3, together with the 500 corresponding raw EEG recordings.

The mean function reflects the overall trend of the EEG series, and serves as a preliminary step for further data analysis. The mean function of the EEG

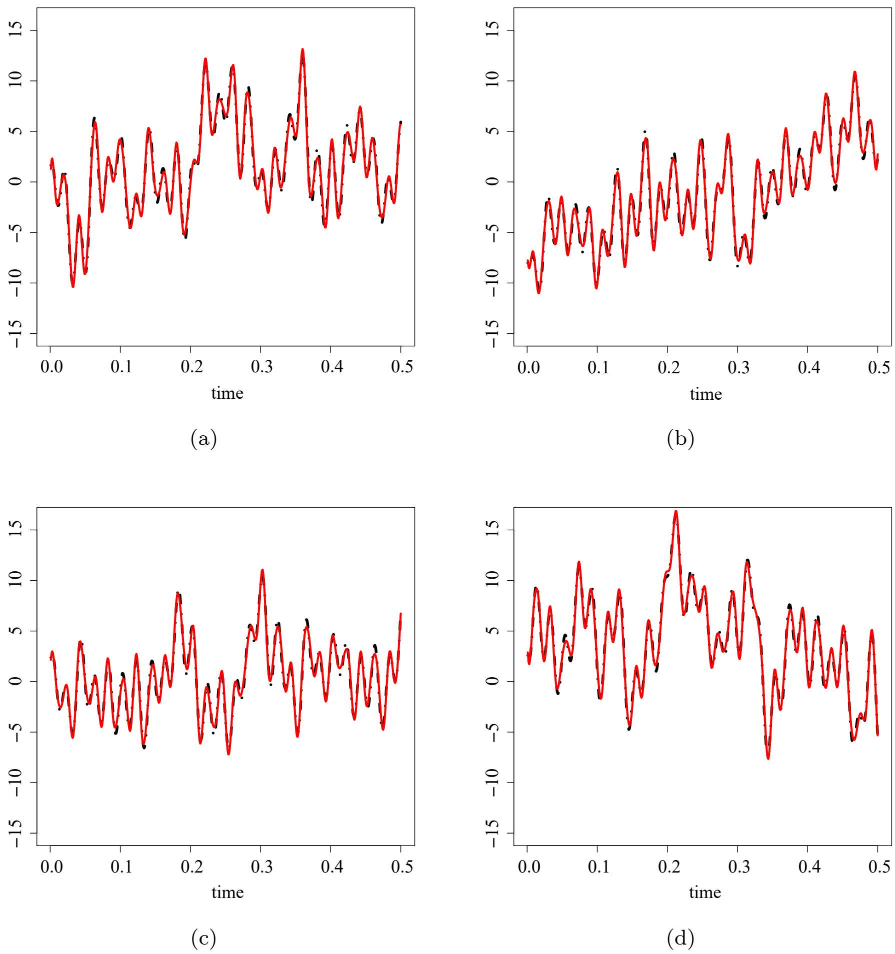

Figure 3. Plots of four randomly selected B-spline trajectories (solid), together with their raw EEG data (dash).

is estimated using (4.3) with a cubic spline ($p = 4$) and the number of knots chosen using the AIC, as in Section 4. The accompanying SCB enables us to test hypotheses on the mean function, such as a certain parametric form. Figure 4 shows that the estimated mean function looks trigonometric. Hence, we test the null hypothesis $H_0 : m(x) \equiv m_0(x) \equiv a_0 + a_1 \sin(100\pi x) + b_1 \cos(100\pi x)$, with parameters a_0, a_1, and b_1 estimated using the least squares method as $\hat{a}_0 = -0.0148$, $\hat{a}_1 = -0.632$, and $\hat{b}_1 = 0.157$. Because the lowest confidence level at which the SCB contains the entire null curve is 2.8% (see Figure 4), we cannot reject the null hypothesis with a large p-value of $1 - 0.028 = 0.972$. This suggests

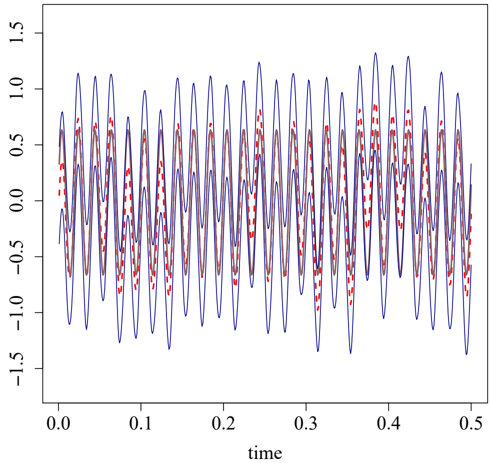

Figure 4. Plots of the null hypothesis curve $m_0(x) = -0.0148 + 0.632\sin(100\pi x) + 0.157\cos(100\pi x)$ (thick), spline estimator $\widehat{m}(x)$ (dashed), and $100(1-\alpha)\% = 100(1-0.972)\%$ SCB (solid) for $m(x)$.

strongly that the mean function of the EEG data is trigonometric in form. We also carried out SCB testing for other participants in the study, and reached similar conclusions.

7. Conclusion

We have proposed a computationally efficient B-spline estimator for the mean estimation in functional time series. We establish asymptotic properties of the estimator, with an SCB as a theoretical byproduct, which proves to be a versatile tool for inference on the true mean function. The SCB performs well numerically, and is illustrated by testing against a hypothesis on the possible form of the mean function. The FMA(∞) model can be extended to functional panel data, which promises more interesting discovery of both the mean function and the functional autocovariance function. The methodology is expected to find wide application in studies involving physiological data such as EEG and ECG data.

Appendix

A.1. Preliminaries

Throughout this section, $\mathcal{O}_p$ (or $\scriptstyle{\mathcal{O}}_p$) denotes a sequence of random variables of certain order in probability. For instance, $\scriptstyle{\mathcal{O}}_p\left(n^{-1/2}\right)$ means a smaller order than $n^{-1/2}$ in probability, and by $\mathcal{O}_{a.s.}$ (or $o_{a.s.}$) almost surely $\mathcal{O}$ (or o). In addition, $\mathcal{U}_p$ denotes a sequence of random functions which are $\mathcal{O}_p$ uniformly defined in the domain.

For any vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{R}^n$, denote the norm $\|\mathbf{a}\|_r = (|a_1|^r + \cdots + |a_n|^r)^{1/r}$, $1 \le r < +\infty$, $\|\mathbf{a}\|_\infty = \max(|a_1|, \ldots, |a_n|)$. For any matrix $\mathbf{A} = (a_{ij})_{i=1,j=1}^{m,n}$, denote its L_r norm as $\|\mathbf{A}\|_r = \max_{\mathbf{a}\in\mathcal{R}^n,\mathbf{a}\neq\mathbf{0}} \|\mathbf{A}\mathbf{a}\|_r \|\mathbf{a}\|_r^{-1}$, for $r < +\infty$ and $\|\mathbf{A}\|_r = \max_{1\le i\le m} \sum_{j=1}^n |a_{ij}|$, for $r = \infty$. For any random variable $\mathbf{X}$, if it is L_p-integrable, denotes its L_p norm as $\|\mathbf{X}\|_p = (\mathbb{E}|\mathbf{X}|^p)^{1/p}$.

Lemma A.1. *(Theorem 2.6.7, Csörgö and Révész (1981)) Suppose that ξ_i, $1 \le i \le n$ are iid with $\mathbb{E}(\xi_1) = 0$, $\mathbb{E}(\xi_1^2) = 1$ and $H(x) > 0$ $(x \ge 0)$ is an increasing continuous function such that $x^{-2-\gamma}H(x)$ is increasing for some $\gamma > 0$ and $x^{-1}\log H(x)$ is decreasing with $\mathbb{E}H(|\xi_1|) < \infty$. Then there exist constants C_1, C_2, $a > 0$ which depend only on the distribution of ξ_1 and a sequence of Brownian motions $\{W_n(m)\}_{n=1}^\infty$, such that for any $\{x_n\}_{n=1}^\infty$ satisfying $H^{-1}(n) < x_n < C_1(n\log n)^{1/2}$ and $S_m = \sum_{i=1}^m \xi_i$, then $\mathbb{P}\{\max_{1\le m\le n}|S_m - W_n(m)| > x_n\} \le C_2 n\{H(ax_n)\}^{-1}$.*

Lemma A.2. *(Theorem 7.5, Billingsley (1999)) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X map Ω into $C[0,1]$: $X(\omega)$ is an element of $C[0,1]$ with value $X_t(\omega) = X(t,\omega)$ at t. For $F \in C[0,1]$, denote $\omega(F,h) = \sup_{x,x'\in[0,1],|x-x'|\le h} |F(x') - F(x)|$ as the modulus of continuity. Suppose that $X, X^1, X^2, \ldots$ are random functions. If $\left(X_{t_1}^n, \ldots, X_{t_k}^n\right) \to_D (X_{t_1}, \ldots, X_{t_k})$ holds for all $t_1, \ldots, t_k$, and if*

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}\left[\omega(X^n, \delta) \ge \epsilon\right] = 0 \tag{A.1}$$

for each positive ϵ, then $X^n \to_D X$.

Lemma A.3. *For $n > 2$, $a > 2$, $W_i \sim N\left(0, \sigma_i^2\right), \sigma_i > 0, i = 1, \ldots, n$*

$$\mathbb{P}\left(\max_{1\le i\le n}\left|\frac{W_i}{\sigma_i}\right| > a\sqrt{\log n}\right) < \sqrt{\frac{2}{\pi}}n^{1-a^2/2}. \tag{A.2}$$

As $n \to \infty$, $(\max_{1\le i\le n}|W_i|)/(\max_{1\le i\le n}\sigma_i) \le \max_{1\le i\le n}|W_i/\sigma_i| = \mathcal{O}_{a.s.}\left(\sqrt{\log n}\right)$.

Proof. Note that

$$\mathbb{P}\left(\max_{1\le i\le n}\left|\frac{W_i}{\sigma_i}\right| > a\sqrt{\log n}\right) \le \sum_{i=1}^n \mathbb{P}\left(\left|\frac{W_i}{\sigma_i}\right| > a\sqrt{\log n}\right)$$
$$\le 2n\left\{1 - \Phi\left(a\sqrt{\log n}\right)\right\} < 2n\frac{\phi\left(a\sqrt{\log n}\right)}{a\sqrt{\log n}}$$

$$\leq 2n\phi\left(a\sqrt{\log n}\right) = \sqrt{\frac{2}{\pi}}n^{1-a^2/2},$$

for $n > 2$, $a > 2$, which proves (A.2). The lemma follows by applying Borel-Cantelli Lemma with choice of $a > 2$.

Lemma A.4. *Assumption* (A5) *holds under Assumptions* (A3), (A4) *and* (A5′).

Proof. Under Assumption (A5′), $\mathbb{E}\left|\zeta_{tk}\right|^{r_1} < \infty$, $r_1 > 4 + 2\omega$, $\mathbb{E}\left|\varepsilon_{tj}\right|^{r_2} < \infty$, $r_2 > 4+2\theta$, where ω is defined in Assumption (A4) and θ is defined in Assumption (A3), so there exists some $\beta_0, \beta_1, \beta_2 \in (0, 1/2)$, such that $r_1 > (2+\omega)/\beta_0$, $r_2 > (2+\theta)/\beta_2$.

Let $H(x) = x^{r_1}$. Lemma A.1 entails that there exist constants c_{1k} and a_k depending on the distribution of ζ_{tk}, such that for $x_n = (n+I_n)^{\beta_0}$, $(n+I_n)/H(a_k x_n) = a_k^{-r_1}(n+I_n)^{1-r_1\beta_0}$ and i.i.d. $N(0,1)$ variables $Z_{tk,\zeta}$,

$$\mathbb{P}\left\{\max_{-I_n+1\leq\tau\leq n}\left|\sum_{t=-I_n+1}^{\tau}\zeta_{tk} - \sum_{t=-I_n+1}^{\tau}Z_{tk,\zeta}\right| > (n+I_n)^{\beta_0}\right\} < c_{1k}a_k^{-r_1}(n+I_n)^{1-r_1\beta_0},$$

Since there are only a finite number of distinct distributions for $\{\zeta_{tk}\}_{t=-I_n+1,k=1}^{n,k_n}$ by Assumption (A5′), there exists a common $c_1 > 0$, such that

$$\max_{1\leq k\leq k_n}\mathbb{P}\left\{\max_{-I_n+1\leq\tau\leq n}\left|\sum_{t=-I_n+1}^{\tau}\zeta_{t,k} - \sum_{t=-I_n+1}^{\tau}Z_{tk,\zeta}\right| > (n+I_n)^{\beta_0}\right\} < c_1(n+I_n)^{1-r_1\beta_0}.$$

Since $I_n \asymp \log n$ by definition, there exists $\epsilon > \beta_0\log\{(n+I_n)/n\}/\log n$, such that $n^{\beta_0+\epsilon} > (n+I_n)^{\beta_0}$. Noting that $\beta_0\log\{(n+I_n)/n\}/\log n \to 0$ as $n \to \infty$, one can choose $\epsilon < 1/2 - \beta_0$. Denote $\beta_1 = \beta_0 + \epsilon$, then $\beta_1 < 1/2$ and $n^{\beta_1} > (n+I_n)^{\beta_0}$. Because $1 - r_1\beta_0 < 0$, it is clear that $(n+I_n)^{1-r_1\beta_0} < n^{1-r_1\beta_0}$. Thus one has

$$\max_{1\leq k\leq k_n}\mathbb{P}\left\{\max_{-I_n+1\leq\tau\leq n}\left|\sum_{t=-I_n+1}^{\tau}\zeta_{t,k} - \sum_{t=-I_n+1}^{\tau}Z_{tk,\zeta}\right| > n^{\beta_1}\right\} < c_1 n^{1-r_1\beta_0}.$$

Recalling that $r_1 > (2+\omega)/\beta_0$, one can let $\gamma_1 = r_1\beta_0 - 1 - \omega > 1$, and there exists a $C_1 > 0$ such that

$$\mathbb{P}\left\{\max_{1\leq k\leq k_n}\max_{-I_n+1\leq\tau\leq n}\left|\sum_{t=-I_n+1}^{\tau}\zeta_{t,k} - \sum_{t=-I_n+1}^{\tau}Z_{tk,\zeta}\right| > n^{\beta_1}\right\} < k_n c_1 n^{1-r_1\beta_0} \leq C_1 n^{-\gamma_1}.$$

Likewise, under Assumption (A5′), taking $H(x) = x^{r_2}$, Lemma A.1 implies that there exists constants c_2 and b depending on the distribution of ε_{ij}, such that for $x_N = N^{\beta_2}$, $N/H(bx_N) = b^{-r_2}N^{1-r_2\beta_2}$ and i.i.d. standard normal random variables $\{Z_{tj,\varepsilon}\}_{t=1,j=1}^{n,N}$ such that

$$\max_{1\le t\le n} \mathbb{P}\left\{\max_{1\le\tau\le N}\left|\sum_{j=1}^{\tau}\varepsilon_{tj} - \sum_{j=1}^{\tau} Z_{tj,\varepsilon}\right| > N^{\beta_2}\right\} < c_2 b^{-r_2}N^{1-r_2\beta_2}.$$

Assumption (A3) states that $n = \mathcal{O}\left(N^{\theta}\right)$, so there is a $C_2 > 0$ such that

$$\mathbb{P}\left\{\max_{1\le t\le n}\max_{1\le\tau\le N}\left|\sum_{j=1}^{\tau}\varepsilon_{tj} - \sum_{j=1}^{\tau} Z_{tj,\varepsilon}\right| > N^{\beta_2}\right\} < c_2 b^{-r_2} n \times N^{1-r_2\beta_2} \le C_2 N^{\theta+1-r_2\beta_2}.$$

Since $r_2\beta_2 > (2+\theta)$, there is $\gamma_2 = r_2\beta_2 - 1 - \theta > 1$ and Assumption (A5) follows.

Lemma A.5. *Under Assumptions* (A5) *and* (A5′), *as* $n \to \infty$, *there exist* $C_3, C_4 \in (0,+\infty)$, $\gamma_3 \in (1,+\infty)$, $\beta_3 \in (0,1/2)$ *and* $N(0,1)$ *variables* $Z_{tk,\xi} = \sum_{t'=0}^{\infty} a_{t'k} Z_{t-t',k,\zeta}$, $t = 1,\ldots,n$, $k = 1,\ldots,k_n$, *with* $Z_{tk,\zeta}$*'s as in Lemma* A.4. *Consequently for* $1 \le j \le n$, $1 \le h \le n-j$, $\mathrm{Cov}\left(Z_{jk,\xi}, Z_{j+h,k,\xi}\right) = \sum_{m=0}^{\infty} a_{mk}$ $a_{m+h,k}$ *and one has*

$$\mathbb{P}\left\{\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\xi_{tk} - \sum_{t=1}^{\tau} Z_{tk,\xi}\right| > C_3 n^{\beta_3}\right\} < C_4 n^{-\gamma_3}. \tag{A.3}$$

Proof. Since $\sum_{t=0}^{\infty} a_{tk}^2 = 1$ and $|a_{tk}| < C_a\rho_a^t$, for $t = 0,\ldots,n$, $k = 1,\ldots,k_n$, together with $I_n > -10\log n/\log\rho_a$ in Assumption (A5), then $\rho_a^{I_n} < n^{-10}$, $\rho_a^{t'-I_n}$, when $t' > I_n$ and $\sum_{t=0}^{I_n}|a_{tk}| < M$ for some constant $M > 0$. It is clear that

$$\xi_{tk} = \sum_{t'=0}^{I_n} a_{t'k}\zeta_{t-t',k} + \sum_{t'=I_n+1}^{\infty} a_{t'k}\zeta_{t-t',k},$$

$$\left|\xi_{tk} - \sum_{t'=0}^{I_n} a_{t'k}\zeta_{t-t',k}\right| \le \sum_{t'=I_n+1}^{\infty} C_a n^{-10}\rho_a^{t'-I_n}\left|\zeta_{t-t',k}\right|,$$

$$\left|\xi_{tk} - \sum_{t'=0}^{I_n} a_{t'k}\zeta_{t-t',k}\right| \le C_a n^{-10}\sum_{t'=1}^{\infty}\rho_a^{t'}\left|\zeta_{t-I_n-t',k}\right|.$$

Hence,

$$\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\xi_{tk} - \sum_{t=1}^{\tau}\sum_{t'=0}^{I_n} a_{t'k}\zeta_{t-t',k}\right| \le \max_{1\le k\le k_n}\max_{1\le t\le n} C_a n^{-9}\sum_{t'=1}^{\infty}\rho_a^{t'}\left|\zeta_{t-I_n-t',k}\right|$$

Denote $W_{tk} = \sum_{t'=1}^{\infty} \rho_a^{t'} |\zeta_{t-I_n-t',k}|$ and note that $\sup_{t,k} \mathbb{E} |\zeta_{t,k}|^{r_1} < \infty$, where $r_1 > 4 + 2\omega$,

$$\|W_{tk}\|_{r_1} \le \sum_{t'=1}^{\infty} \rho_a^{t'} \|\zeta_{t-I_n-t',k}\|_{r_1} < \infty.$$

Therefore, $\mathbb{E} W_{tk}^{r_1} < K$ for some $K > 0$, $t = 1, \ldots, n$, $k = 1, \ldots, k_n$. Note that $k_n = \mathcal{O}(n^{\omega})$ in Assumption (A4), thus

$$\mathbb{P}\left(C_a n^{-9} \max_{1 \le k \le k_n} \max_{1 \le t \le n} W_{tk} > M n^{\beta_3} \right) < n k_n \frac{C_a^{r_1} K}{M^{r_1}} n^{-(\beta_3+9) r_1} < \frac{C_a^{r_1} K}{M^{r_1}} n^{-(\beta_3+9) r_1 + 1 + \omega}.$$

So,

$$\mathbb{P}\left[\max_{1 \le k \le k_n} \max_{1 \le \tau \le n} \left| \sum_{t=1}^{\tau} \xi_{tk} - \sum_{t=1}^{\tau} \sum_{t'=0}^{I_n} a_{t'k} \zeta_{t-t',k} \right| > M n^{\beta_3} \right] < \frac{C_a^{r_1} K}{M^{r_1}} n^{-(\beta_3+9) r_1 + 1 + \omega}.$$

Next, define $U_{tk} = \sum_{t'=I_n+1}^{\infty} a_{t'k} Z_{t-t',k,\zeta}$, then $U_{tk} \sim N\left(0, \sum_{t'=I_n+1}^{\infty} a_{t'k}^2\right)$, $k = 1, \ldots, k_n$. It is obvious that

$$\max_{1 \le k \le k_n} \max_{1 \le \tau \le n} \left| \sum_{t=1}^{\tau} \sum_{t'=I_n+1}^{\infty} a_{t'k} Z_{t-t',k,\zeta} \right| \le n \max_{1 \le k \le k_n} \max_{1 \le t \le n} |U_{tk}|.$$

Note that $\sum_{t'=I_n+1}^{\infty} a_{t'k}^2 < C n^{-20}$ for some $C > 0$, $k = 1, \ldots, k_n$ and $k_n = \mathcal{O}(n^{\omega})$ for some $\omega > 0$, one has

$$\mathbb{P}\left(n \max_{1 \le k \le k_n} \max_{1 \le t \le n} |U_{tk}| > M n^{\beta_3} \right) < n k_n \frac{C n^{-20}}{M^2 \left(n^{\beta_3 - 1}\right)^2} < \frac{C}{M^2} n^{-17 - 2\beta_3 + \omega},$$

which leads to

$$\mathbb{P}\left[\max_{1 \le k \le k_n} \max_{1 \le \tau \le n} \left| \sum_{t=1}^{\tau} \sum_{t'=I_n+1}^{\infty} a_{t'k} Z_{t-t',k,\zeta} \right| > M n^{\beta_3} \right] < n k_n \frac{C n^{-20}}{M^2 \left(n^{\beta_3 - 1}\right)^2} < \frac{C}{M^2} n^{-17 - 2\beta_3 + \omega}.$$

Now Assumption (A5) entails that for $0 \le t' \le I_n$, $1 \le t \le n$, $-I_n + 1 \le t - t' \le n$

$$\mathbb{P}\left\{ \max_{1 \le k \le k_n} \max_{-I_n+1 \le \tau \le n} \left| \sum_{t=-I_n+1}^{\tau} \zeta_{tk} - \sum_{t=-I_n+1}^{\tau} Z_{tk,\zeta} \right| > n^{\beta_3} \right\} < C_1 n^{-\gamma_1}.$$

Then,

$$
\begin{aligned}
&\mathbb{P}\left[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}(\zeta_{t-t',k}-Z_{t-t',k,\zeta})\right|>2Mn^{\beta_3}\right]\\
&=\mathbb{P}\left[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t'=0}^{I_n}a_{t'k}\sum_{t=1}^{\tau}(\zeta_{t-t',k}-Z_{t-t',k,\zeta})\right|>2Mn^{\beta_3}\right]\\
&\le\mathbb{P}\left[\max_{1\le k\le k_n}\left\{\sum_{t'=0}^{I_n}|a_{t'k}|\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\zeta_{t-t',k}-\sum_{t=1}^{\tau}Z_{t-t',k,\zeta}\right|\right\}>2Mn^{\beta_3}\right]\\
&\le\mathbb{P}\left[M\max_{1\le k\le k_n}\max_{1\le\tau\le n}\max_{0\le t'\le I_n}\left|\sum_{t=1}^{\tau}\zeta_{t-t',k}-\sum_{t=1}^{\tau}Z_{t-t',k,\zeta}\right|>2Mn^{\beta_3}\right]\\
&\le\mathbb{P}\left\{2\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=-I_n+1}^{\tau}\zeta_{tk}-\sum_{t=-I_n+1}^{\tau}Z_{tk,\zeta}\right|>2n^{\beta_3}\right\}<C_1n^{-\gamma_1}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\mathbb{P}\left[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\xi_{tk}-\sum_{t=1}^{\tau}\left(\sum_{t'=0}^{\infty}a_{t'k}Z_{t-t',k,\zeta}\right)\right|>4Mn^{\beta_3}\right]\\
&=\mathbb{P}\Bigg[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\Bigg|\sum_{t=1}^{\tau}\xi_{tk}-\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}\zeta_{t-t',k}+\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}\zeta_{t-t',k}\\
&\quad-\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}Z_{t-t',k,\zeta}-\sum_{t=1}^{\tau}\sum_{t'=I_n+1}^{\infty}a_{t'k}Z_{t-t',k,\zeta}\Bigg|>4Mn^{\beta_3}\Bigg]\\
&\le\mathbb{P}\Bigg[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\Bigg\{\left|\sum_{t=1}^{\tau}\xi_{tk}-\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}\zeta_{t-t',k}\right|+\Bigg|\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}\zeta_{t-t',k}\\
&\quad-\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}Z_{t-t',k,\zeta}\Bigg|+\left|\sum_{t=1}^{\tau}\sum_{t'=I_n+1}^{\infty}a_{t'k}Z_{t-t',k,\zeta}\right|\Bigg\}>4Mn^{\beta_3}\Bigg]\\
&\le\mathbb{P}\left[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\xi_{tk}-\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}\zeta_{t-t',k}\right|>Mn^{\beta_3}\right]\\
&\quad+\mathbb{P}\left[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}\zeta_{t-t',k}-\sum_{t=1}^{\tau}\sum_{t'=0}^{I_n}a_{t'k}Z_{t-t',k,\zeta}\right|>2Mn^{\beta_3}\right]\\
&\quad+\mathbb{P}\left[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\sum_{t'=I_n+1}^{\infty}a_{t'k}Z_{t-t',k,\zeta}\right|>Mn^{\beta_3}\right]\\
&\le\frac{C_a^{r_1}K}{M^{r_1}}n^{-(\beta_3+9)r_1+1+\omega}+\frac{C}{M^2}n^{-17-2\beta_3+\omega}+C_1n^{-\gamma_1}<C_4n^{-\gamma_3}
\end{aligned}
$$

Denote $C_3 = 4M$ and $Z_{tk,\xi} = \sum_{t'=0}^{\infty} a_{t'k} Z_{t-t',k,\zeta}$, $t = 1, \ldots, n$, $k = 1, \ldots, k_n$, then $\{Z_{tk,\xi}\}_{t=1,k=1}^{n,k_n}$ are $N(0,1)$ variables and $\mathrm{Cov}\,(Z_{j,k,\xi}, Z_{j+h,k,\xi}) = \sum_{m=0}^{\infty} a_{mk} a_{m+h,k}$, $1 \le j \le n,\ 1 \le h \le n-j$, thus

$$\mathbb{P}\left[\max_{1\le k\le k_n}\max_{1\le\tau\le n}\left|\sum_{t=1}^{\tau}\xi_{tk} - \sum_{t=1}^{\tau} Z_{tk,\xi}\right| > C_3 n^{\beta_3}\right] < C_4 n^{-\gamma_3}.$$

The proof is completed.

Lemma A.6. *Under Assumptions* (A2), (A5) *and* (A6), *as* $n \to \infty$

$$\max_{1\le \ell\le J_s+p}\left|(nN)^{-1}\sum_{t=1}^{n}\sum_{j=1}^{N} B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right) Z_{tj,\varepsilon}\right|$$
$$= \mathcal{O}_{a.s.}\left(n^{-1/2}N^{-1/2}J_s^{-1/2}\log^{1/2} N\right).$$

Proof. Note that $(nN)^{-1}\sum_{t=1}^{n}\sum_{j=1}^{N} B_{\ell,p}(j/N)\sigma\,(j/N)\, Z_{tj,\varepsilon} = N^{-1}\sum_{j=1}^{N} B_{\ell,p}(j/N)\sigma\,(j/N)\, Z_{\cdot j,\varepsilon}$, where $Z_{\cdot j,\varepsilon} = n^{-1}\sum_{t=1}^{n} Z_{tj,\varepsilon}$, one can apply Lemma A.3 to obtain uniform bound for zero mean Gaussian variables $N^{-1}\sum_{j=1}^{N} B_{\ell,p}(j/N)\sigma\,(j/N)$ $Z_{\cdot j,\varepsilon}$, $1 \le \ell \le J_s + p$ with variance

$$\mathbb{E}\left\{N^{-1}\sum_{j=1}^{N} B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right) Z_{\cdot j,\varepsilon}\right\}^2 = n^{-1}N^{-2}\sum_{j=1}^{N} B_{\ell,p}^2\left(\frac{j}{N}\right)\sigma^2\left(\frac{j}{N}\right)$$
$$= n^{-1}N^{-1}\left\|B_{\ell,p}\sigma\right\|_{2,N}^2 \asymp J_s^{-1}N^{-1}n^{-1}.$$

It follows from Lemma A.3 that

$$\max_{1\le \ell\le J_s+p}\left|N^{-1}\sum_{j=1}^{N} B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right) Z_{\cdot j,\varepsilon}\right|$$
$$= \mathcal{O}_{a.s.}\left\{n^{-1/2}N^{-1/2}J_s^{-1/2}\log^{1/2}(J_s+p)\right\}$$
$$= \mathcal{O}_{a.s.}\left(n^{-1/2}N^{-1/2}J_s^{-1/2}\log^{1/2} N\right), \tag{A.4}$$

where the last step follows from Assumption (A6) on the order of J_s relative to N. Thus the lemma holds.

Lemma A.7. *Under Assumptions* (A2), (A5) *and* (A6), *as* $n \to \infty$

$$\sup_{x\in[0,1]} n^{-1}\left|\sum_{t=1}^{n}\widetilde{\varepsilon}_t(x)\right| = \mathcal{O}_{a.s.}\left(n^{-1/2}N^{-1/2}J_s^{1/2}\log^{1/2} N + N^{\beta_2-1}J_s\right).$$

Proof. According to Assumption (A5), it is trivial that

$$\max_{1\leq t\leq n}\max_{1\leq j\leq N}\left|N^{-1}\sum_{i=1}^{j}(\varepsilon_{ti}-Z_{ti,\varepsilon})\right| = \mathcal{O}_{a.s.}(N^{\beta_2-1}).$$

Next, the B spline basis satisfies

$$\left|B_{\ell,p}\left(\frac{j}{N}\right)-B_{\ell,p}\left(\frac{j+1}{N}\right)\right|\leq N^{-1}\left\|B_{\ell,p}\right\|_{0,1}\leq CJ_sN^{-1}$$

uniformly over $1\leq j\leq N$ and $1\leq \ell\leq J_s+p$, while Assumptions (A2) and (A6) imply that $J_sN^{-1}\sim N^{\gamma}d_NN^{-1}\sim N^{\gamma-1}d_N\gg N^{-\nu}$, hence

$$\left|\sigma\left(\frac{j}{N}\right)-\sigma\left(\frac{j+1}{N}\right)\right|\leq N^{-\nu}\left\|\sigma\right\|_{0,\nu}\leq CJ_sN^{-1}$$

uniformly over $1\leq j\leq N$. Note that for $1\leq \ell\leq J_s+p$, both $B_{\ell,p}(\cdot)$ and $\sigma(\cdot)$ are bounded on $[0,1]$, then

$$\begin{aligned}
&\left|B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right)-B_{\ell,p}\left(\frac{j+1}{N}\right)\sigma\left(\frac{j+1}{N}\right)\right|\\
&=\left|\left\{B_{\ell,p}\left(\frac{j}{N}\right)-B_{\ell,p}\left(\frac{j+1}{N}\right)+B_{\ell,p}\left(\frac{j+1}{N}\right)\right\}\sigma\left(\frac{j}{N}\right)\right.\\
&\quad\left.-B_{\ell,p}\left(\frac{j+1}{N}\right)\sigma\left(\frac{j+1}{N}\right)\right|\\
&\leq\left|B_{\ell,p}\left(\frac{j}{N}\right)-B_{\ell,p}\left(\frac{j+1}{N}\right)\right|\sigma\left(\frac{j}{N}\right)+\left|\sigma\left(\frac{j}{N}\right)-\sigma\left(\frac{j+1}{N}\right)\right|B_{\ell,p}\left(\frac{j+1}{N}\right)\\
&\leq CJ_sN^{-1}.
\end{aligned}$$

Since the support of $B_{\ell,p}(\cdot)$ has length at most $p/\left(J_s+1\right)$, one obtains that

$$\sum_{j=1}^{N-1}\left|B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right)-B_{\ell,p}\left(\frac{j+1}{N}\right)\sigma\left(\frac{j+1}{N}\right)\right|\leq C$$

Hence,

$$\begin{aligned}
&\left|(nN)^{-1}\sum_{t=1}^{n}\sum_{j=1}^{N}B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right)(\varepsilon_{tj}-Z_{tj,\varepsilon})\right|\\
&=\left|n^{-1}\sum_{t=1}^{n}\left[\sum_{j=1}^{N-1}\left\{B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right)-B_{\ell,p}\left(\frac{j+1}{N}\right)\sigma\left(\frac{j+1}{N}\right)\right\}\right.\right.
\end{aligned}$$

$$N^{-1}\sum_{i=1}^{j}(\varepsilon_{ti}-Z_{ti,\varepsilon})\Bigg]+n^{-1}\sum_{t=1}^{n}\left\{B_{\ell,p}(1)\,\sigma(1)\,N^{-1}\sum_{j=1}^{N}(\varepsilon_{tj}-Z_{tj,\varepsilon})\right\}\Bigg|$$

$$\leq\left\{\max_{1\leq t\leq n}\max_{1\leq j\leq N}\left|N^{-1}\sum_{i=1}^{j}(\varepsilon_{ti}-Z_{ti,\varepsilon})\right|\right\}\left\{\sum_{j=1}^{N-1}\left|B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right)\right.\right.$$

$$\left.\left.-B_{\ell,p}\left(\frac{j+1}{N}\right)\sigma\left(\frac{j+1}{N}\right)\right|\right\}+C\left\{\max_{1\leq t\leq n}\max_{1\leq j\leq N}\left|N^{-1}\sum_{i=1}^{j}(\varepsilon_{ti}-Z_{ti,\varepsilon})\right|\right\}$$

$$=\mathcal{O}_{a.s}\left\{N^{\beta_2-1}+N^{\beta_2-1}\right\}=\mathcal{O}_{a.s}\left(N^{\beta_2-1}\right).$$

Hence,

$$\max_{1\leq\ell\leq J_s+p}\left|(nN)^{-1}\sum_{t=1}^{n}\sum_{j=1}^{N}B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right)(\varepsilon_{tj}-Z_{tj,\varepsilon})\right|=\mathcal{O}_{a.s}\left(N^{\beta_2-1}\right).$$

The above inequality and Lemma A.6 together imply that

$$\max_{1\leq\ell\leq J_s+p}\left|(nN)^{-1}\sum_{t=1}^{n}\sum_{j=1}^{N}B_{\ell,p}\left(\frac{j}{N}\right)\sigma\left(\frac{j}{N}\right)\varepsilon_{tj}\right|$$
$$=\mathcal{O}_{a.s.}\left(n^{-1/2}N^{-1/2}J_s^{-1/2}\log^{1/2}N+N^{\beta_2-1}\right).$$

Clearly $(nN)^{-1}\mathbf{X}^\top\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t=\{(nN)^{-1}\sum_{t=1}^{n}\sum_{j=1}^{N}B_{\ell,p}(j/N)\sigma(j/N)\varepsilon_{tj}\}_{\ell=1}^{J_s+p}$, thus

$$\left\|(nN)^{-1}\mathbf{X}^\top\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\right\|_\infty=\mathcal{O}_{a.s.}\left(n^{-1/2}N^{-1/2}J_s^{-1/2}\log^{1/2}N+N^{\beta_2-1}\right).$$

Recalling the definition of $\widetilde{\varepsilon}_i(x)$ in (3.9) and equation (3.5), the proof is completed by

$$\sup_{x\in[0,1]}n^{-1}\left|\sum_{t=1}^{n}\widetilde{\varepsilon}_t(x)\right|=\left\|n^{-1}N^{-1}\mathbf{B}(x)^\top\mathbf{V}_{n,p}^{-1}\mathbf{X}^\top\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\right\|_\infty$$
$$=\mathcal{O}_{a.s.}\left(n^{-1/2}N^{-1/2}J_s^{1/2}\log^{1/2}N+N^{\beta_2-1}J_s\right).$$

A.2. Proof of Theorem 2

For any $k=1,2,\ldots$, let $\boldsymbol{\phi}_k=(\phi_k(1/N),\ldots,\phi_k(N/N))^\top$, and denote $\widetilde{\phi}_k(x)=N^{-1}\mathbf{B}(x)^\top\mathbf{V}_{n,p}^{-1}\mathbf{X}^\top\boldsymbol{\phi}_k$. According to (3.7), $\widetilde{\eta}_t(x)=\widetilde{m}(x)+\sum_{k=1}^{\infty}\xi_{tk}\widetilde{\phi}_k(x)$, therefore,

$$\widetilde{\eta}_t(x) - \eta_t(x) = \widetilde{m}(x) - m(x) + \sum_{k=1}^{\infty} \xi_{tk} \left\{ \widetilde{\phi}_k(x) - \phi_k(x) \right\}.$$

Lemma A.4 of Cao, Yang and Todem (2012) provides a constant $C_{q,\mu} > 0$, such that

$$\|\widetilde{m} - m\|_\infty \leq C_{q,\mu} \|m\|_{q,\mu} J_s^{-p^*}, \tag{A.5}$$

$$\|\widetilde{\phi}_k - \phi_k\|_\infty \leq C_{q,\mu} \|\phi_k\|_{q,\mu} J_s^{-p^*}, \quad k \geq 1 \tag{A.6}$$

Thus (A.5) and (A.6), and Assumption (A4) lead to

$$\|\widetilde{\eta}_t - \eta_t\|_\infty \leq \|\widetilde{m} - m\|_\infty + \sum_{k=1}^{\infty} |\xi_{tk}| \|\widetilde{\phi}_k - \phi_k\|_\infty \leq C_{q,\mu} W_t J_s^{-p^*},$$

where $W_t = \|m\|_{q,\mu} + \sum_{k=1}^{\infty} |\xi_{tk}| \|\phi_k\|_{q,\mu}$, $t = 1, \ldots, n$, are identically distributed nonnegative random variables with finite mean. Assumption (A6) then implies that

$$\mathbb{P}\left\{ \max_{1\leq t\leq n} W_t > (n \log n)^{2/r_1} \right\} \leq n \frac{\mathrm{E}W_t^{r_1}}{(n \log n)^2} = \mathrm{E}W_t^{r_1} n^{-1} (\log n)^{-2},$$

thus,

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{ \max_{1\leq t\leq n} W_t > n \log n \right\} \leq \mathrm{E}W_t^{r_1} \sum_{n=1}^{\infty} n^{-1} (\log n)^{-2} < +\infty,$$

so $\max_{1\leq t\leq n} W_t = \mathcal{O}_{a.s.}\{(n \log n)^{2/r_1}\}$, $\max_{1\leq t\leq n} \|\widetilde{\eta}_t - \eta_t\|_\infty = \mathcal{O}_{a.s.}\{J_s^{-p^*} (n \log n$ $)^{2/r_1}\}$, thus,

$$\left\| n^{-1} \sum_{t=1}^{n} \{\widetilde{\eta}_t(x) - \eta_t(x)\} \right\|_\infty = \mathcal{O}_{a.s.}\{J_s^{-p^*} (n \log n)^{2/r_1}\}.$$

By noticing that $\left\| n^{-1} \sum_{t=1}^{n} \widetilde{\varepsilon}_t(x) \right\|_\infty = \mathcal{O}_{a.s.}(n^{-1/2} N^{-1/2} J_s^{1/2} \log^{1/2} N + N^{\beta_2 - 1} J_s)$ in Lemma A.7 and 3.10, one obtaions that

$$\sup_{x\in[0,1]} |\overline{m}(x) - \widehat{m}(x)|$$
$$= \mathcal{O}_{a.s.} \left\{ J_s^{-p^*} (n \log n)^{2/r_1} + n^{-1/2} N^{-1/2} J_s^{1/2} \log^{1/2} N + N^{\beta_2 - 1} J_s \right\}$$

The orders of J_s and n relative to N in Assumptions (A3) and (A6) imply that

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

The proof is completed.

A.3. Proof of Theorem 1

Noting that $\operatorname{Cov}(Z_{jk,\xi}, Z_{j+h,k,\xi}) = \sum_{m=0}^{\infty} a_{mk}a_{m+h,k},\ 1 \le j \le n,\ 1 \le h \le n-j$ in Lemma A.5, it is easy to compute that

$$\operatorname{Var}\left(\overline{Z}_{\cdot k,\xi}\right) = n^{-1} + 2n^{-2}\left\{\sum_{m=1}^{n-1}\sum_{t=0}^{\infty}(n-m)a_{tk}a_{t+m,k}\right\}.$$

Denote $\widetilde{\varphi}_k(x) = \overline{Z}_{\cdot k,\xi}\phi_k(x),\ k = 1,\ldots,\infty$ and define $\varphi_n(x) = n^{1/2}G_{\varphi}(x,x)^{-1/2}\sum_{k=1}^{\infty}\widetilde{\varphi}_k(x)$, then for any $(x_1,\ldots,x_l) \in [0,1]^l$ and $(b_1,\ldots,b_l) \in \mathbb{R}^l$,

$$\begin{aligned}
&\lim_{n\to\infty}\operatorname{Var}\left(\sum_{i=1}^{l} b_i\varphi_n(x_i)\right)\\
&= \lim_{n\to\infty}\operatorname{Var}\left(n^{1/2}\sum_{i=1}^{l} b_i G_{\varphi}(x_i,x_i)^{-1/2}\sum_{k=1}^{\infty}\overline{Z}_{\cdot k,\xi}\phi_k(x_i)\right)\\
&= \sum_{i=1}^{l} b_i^2 + 2\sum_{1\le i<j\le l} b_i b_j G_{\varphi}(x_i,x_i)^{-1/2}G_{\varphi}(x_j,x_j)^{-1/2}G_{\varphi}(x_i,x_j)\\
&= \operatorname{Var}\left(\sum_{i=1}^{l} b_i\varphi(x_i)\right).
\end{aligned}$$

Hence

$$\{\varphi_n(x_1),\ldots,\varphi_n(x_l)\} \to_D \{\varphi(x_1),\ldots,\varphi(x_l)\}. \tag{A.7}$$

Assumption (A4) states that $G_{\varphi}(x,x) \ge C_{\varphi} > 0, x \in [0,1]$, so $\omega(\varphi_n,\delta)$ satisfies

$$\begin{aligned}
\omega(\varphi_n,\delta) &= \sup_{x,x'\in[0,1],|x-x'|\le\delta}\left|\varphi_n(x)-\varphi_n(x')\right|\\
&\le \sup_{x,x'\in[0,1],|x-x'|\le\delta} n^{1/2}C_{\varphi}^{-1/2}\sum_{k=1}^{\infty}\left|\phi_k(x)-\phi_k(x')\right|\left|\overline{Z}_{\cdot k,\xi}\right|\\
&\le n^{1/2}C_{\varphi}^{-1/2}\delta^{\mu}\sum_{k=1}^{\infty}\|\phi_k\|_{0,\mu}\left|\overline{Z}_{\cdot k,\xi}\right|.
\end{aligned}$$

Since $\mathbb{E}\left|\overline{Z}_{\cdot k,\xi}\right| = (2/\pi)^{1/2}\operatorname{Var}\left(\overline{Z}_{\cdot k,\xi}\right)^{1/2}$, thus

$$\mathbb{P}\left[\omega(\varphi_n,\delta) \ge \epsilon\right] \le \mathbb{P}\left(n^{1/2}\delta^{\mu}C_{\varphi}^{-1/2}\sum_{k=1}^{\infty}\|\phi_k\|_{0,\mu}\left|\overline{Z}_{\cdot k,\xi}\right| \ge \epsilon\right)$$

$$\leq \frac{(2/\pi)^{1/2}\delta^{\mu}C_{\varphi}^{-1/2}\sum_{k=1}^{\infty}\|\phi_k\|_{0,\mu}\left\{n\mathrm{Var}\left(\overline{Z}_{\cdot k,\xi}\right)\right\}^{1/2}}{\epsilon}.$$

Since $\sum_{k=1}^{\infty}\|\phi_k\|_{0,\mu} < +\infty$ in Assumption (A4) and $n\mathrm{Var}\left(\overline{Z}_{\cdot k,\xi}\right) \to 1 + 2\sum_{t=0}^{\infty}\sum_{t'=t+1}^{\infty} a_{tk}a_{t'k}$ as $n \to \infty$, it is clear that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{(2/\pi)^{1/2}\delta^{\mu}C_{\varphi}^{-1/2}\sum_{k=1}^{\infty}\|\phi_k\|_{0,\mu}\left\{n\mathrm{Var}\left(\overline{Z}_{\cdot k,\xi}\right)\right\}^{1/2}}{\epsilon} = 0.$$

Thus

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left[\omega\left(\varphi_n,\delta\right)\geq\epsilon\right] = 0.$$

According to (A.7) and Lemma A.2, $\varphi_n \to_D \varphi$.

Note that

$$\begin{aligned}
&n^{1/2}\sup_{x\in[0,1]} G_{\varphi}\left(x,x\right)^{-1/2}\left|\sum_{k=1}^{\infty}\left(\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right)\phi_k(x)\right| \\
&\leq n^{1/2}\sup_{x\in[0,1]} G_{\varphi}\left(x,x\right)^{-1/2}\sum_{k=1}^{k_n}\left|\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right|\left|\phi_k(x)\right| \\
&\quad + n^{1/2}\sup_{x\in[0,1]} G_{\varphi}\left(x,x\right)^{-1/2}\sum_{k=k_n+1}^{\infty}\left|\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right|\left|\phi_k(x)\right|.
\end{aligned}$$

Applying (A.3) leads to that

$$\mathbb{P}\left\{\max_{1\leq k\leq k_n}\left|\overline{\xi}_{\cdot k}-\overline{Z}_{\cdot k,\xi}\right| > C_3 n^{\beta_3-1}\right\} < C_4 n^{-\gamma_3}.$$

By Borel Cantelli lemma, one has

$$\max_{1\leq k\leq k_n}\left|\overline{\xi}_{\cdot k}-\overline{Z}_{\cdot k,\xi}\right| = \mathcal{O}_{a.s.}\left(n^{\beta_3-1}\right). \tag{A.8}$$

By Assumption (A4), $\sum_{k=1}^{\infty}\|\phi_k\|_{\infty} < +\infty$, thus $\sum_{k=1}^{k_n}\|\phi_k\|_{\infty} < C$ for some constant C. Together with (A.8) and Assumption (A3), one obtains that

$$\begin{aligned}
&n^{1/2}\sup_{x\in[0,1]} G_{\varphi}\left(x,x\right)^{-1/2}\sum_{k=1}^{k_n}\left|\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right|\left|\phi_k(x)\right| \\
&\leq n^{1/2}C_G^{-1/2}\sup_{x\in[0,1]}\sum_{k=1}^{k_n}\left|\phi_k(x)\right|\max_{1\leq k\leq k_n}\left|\overline{\xi}_{\cdot k}-\overline{Z}_{\cdot k,\xi}\right|
\end{aligned}$$

$$\leq n^{1/2}C_G^{-1/2}\sum_{k=1}^{k_n}\|\phi_k\|_\infty \max_{1\leq k\leq k_n}\left|\overline{\xi}_{\cdot k}-\overline{Z}_{\cdot k,\xi}\right|$$
$$\leq n^{1/2}C_G^{-1/2}C\mathcal{O}_{a.s.}\left(n^{\beta_3-1}\right)=\mathcal{O}_{a.s.}\left(n^{\beta_3-1/2}\right)=\mathcal{O}_{a.s.}(1) \qquad \text{(A.9)}$$

Note that

$$\left(\mathbb{E}\left|\overline{\xi}_{\cdot k}\right|\right)^2=\left(\mathbb{E}\left|\overline{Z}_{\cdot k,\xi}\right|\right)^2\leq \mathbb{E}\overline{Z}^2_{\cdot k,\xi}=n^{-1}+2n^{-2}\left\{\sum_{m=1}^{n-1}\sum_{t=0}^{\infty}(n-m)a_{tk}a_{t+m,k}\right\},$$

thus $\mathbb{E}\left|\overline{\xi}_{\cdot k}\right|=\mathbb{E}\left|\overline{Z}_{\cdot k,\xi}\right|=\mathcal{O}\left(n^{-1/2}\right)$. In addition, Assumption (A4) states that $\sum_{k=k_n+1}^{\infty}\|\phi_k\|_\infty=\mathcal{O}\left(n^{-1/2}\right)$, then there exists

$$\mathbb{E}n^{1/2}\sup_{x\in[0,1]}G_\varphi(x,x)^{-1/2}\sum_{k=k_n+1}^{\infty}\left|\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right||\phi_k(x)|$$
$$\leq n^{1/2}C_G^{-1/2}\sum_{k=k_n+1}^{\infty}\|\phi_k\|_\infty\mathbb{E}\left|\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right|$$
$$\leq n^{1/2}C_G^{-1/2}\mathcal{O}\left(n^{-1/2}\right)\mathcal{O}\left(n^{-1/2}\right)=\mathcal{O}(1) \qquad \text{(A.10)}$$

Combining (A.9) and (A.10), one has

$$\mathbb{E}n^{1/2}\sup_{x\in[0,1]}G_\varphi(x,x)^{-1/2}\left|\sum_{k=1}^{\infty}\left(\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right)\phi_k(x)\right|=\mathcal{O}(1),$$

hence

$$n^{1/2}\sup_{x\in[0,1]}G_\varphi(x,x)^{-1/2}\left|\sum_{k=1}^{\infty}\left(\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right)\phi_k(x)\right|=\mathcal{O}_p(1).$$

Note that

$$\varphi_n(x)-n^{1/2}G_\varphi(x,x)^{-1/2}\left\{\overline{m}(x)-m(x)\right\}$$
$$=n^{1/2}G_\varphi(x,x)^{-1/2}\sum_{k=1}^{\infty}\left(\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}\right)\phi_k(x),$$

hence

$$\sup_{x\in[0,1]}\left|\varphi_n(x)-n^{1/2}G_\varphi(x,x)^{-1/2}\left\{\overline{m}(x)-m(x)\right\}\right|=\mathcal{O}_p(1).$$

The proof is completed by applying Slutsky's theorem.

Acknowledgments

This research was supported by the National Natural Science Foundation of China, awards 11771240, 12026242 and 12171269. The authors thank the associate editor and two referees for their helpful comments and suggestions.

References

Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley, New York.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications.* Springer-Verlag, New York.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* 2nd Edition. Springe-Verlag, New York.

Cao, G., Wang, L., Li, Y. and Yang, L. (2016). Oracle-efficient confidence envelopes for covariance functions in dense functional data. *Statist. Sinica* **26**, 359–383.

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

Chen, M. and Song, Q. (2015). Simultaneous inference of the mean of functional time series. *Electron. J. Stat.* **9**, 1779–1798.

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

Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

de Boor, C. (2001). *A Practical Guide to Splines.* Springer-Verlag, New York.

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

DeVore, R. and Lorentz, G. (1993). *Constructive Approximation: Polynomials and Splines Approximation.* Springer-Verlag, Berlin.

Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice.* Springer, New York.

Gu, L., Wang, L., Härdle, W. and Yang, L. (2014). A simultaneous confidence corridor for varying coefficient regression with sparse functional data. *TEST* **23**, 806–843.

Gu, L. and Yang, L. (2015). Oracally efficient estimation for single-index link function with simultaneous confidence band. *Electron. J. Stat.* **9**, 1540–1561.

Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **75**, 103–122.

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

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis.* Chapman & Hall/CRC, Boca Raton.

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

Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies.* Springer, New York.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis.* Springer, New York.

Wang J. (2012). Modelling time trend via spline confidence band. *Ann. Inst. Statist. Math.* **64**, 275–301.

Wang, Y., Wang, G., Wang, L. and Ogden, T. (2020). Simultaneous confidence corridors for mean functions in functional data analysis of imaging data. *Biometrics* **76**, 427–437.

Wang, J. and Yang, L. (2009). Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19**, 325–342.

Yu, S., Wang, L., Wang, G., Liu, C. and Yang, L. (2020). Estimation and inference for generalized geoadditive models. *J. Amer. Statist. Assoc.* **115**, 761–774.

Zheng, S., Liu, R., Yang, L. and Härdle, W. (2016). Statistical inference for generalized additive models: Simultaneous confidence corridors and variable selection. *TEST* **25**, 607–626.

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

Jie Li

School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: li-j17@tsinghua.org.cn

Lijian Yang

Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China.

E-mail: yanglijian@tsinghua.edu.cn

(Received March 2021; accepted July 2021)