

STATISTICAL INFERENCE FOR FUNCTIONAL TIME SERIES: AUTOCOVARANCE FUNCTION

Chen Zhong and Lijian Yang

Wuhan University and Tsinghua University

Abstract: We investigate statistical inference for functional time series by extending the classic concept of an autocovariance function (ACF) to a functional ACF (FACF). We establish that for functional moving average (FMA) data, the FMA order can be determined as the highest nonvanishing order of an FACF, just as in classic time series analysis. We propose a two-step estimator for the FACF. The first step involves a simultaneous B-spline estimation of each time trajectory, and the second step is a plug-in estimation of the FACF, using the estimated trajectories in place of the latent true curves. Under simple and mild assumptions, the proposed tensor product spline FACF estimator is asymptotically equivalent to the oracle estimator with all known trajectories, leading to an asymptotically correct simultaneous confidence envelope (SCE) for the true FACF. Simulation experiments validate the asymptotic correctness of the SCE and the data-driven FMA order selection. The proposed SCEs are computed for the FACFs of an electroencephalogram (EEG) functional time series, yielding an interesting discovery of a finite FMA lag and Fourier-form functional principal components.

Key words and phrases: Confidence envelope, functional autocovariance function, functional Bartlett's formula, functional moving average, oracle efficiency.

1. Introduction

Functional data analysis (FDA) has become an important statistics research area, owing to its wide applications; see Ramsay and Silverman (2002), Ramsay and Silverman (2005), Ferraty and Vieu (2006), Hsing and Eubank (2015), and Guo, Zhou and Zhang (2019) for the development of FDA applications and theory.

A functional random variable is a square-integrable continuous stochastic process: that is, $\eta(\cdot) \in C[0, 1]$ almost surely, $E\{\sup_{x \in [0, 1]} \eta^2(x)\} < +\infty$. For such $\eta(\cdot)$, both the mean function $E\{\eta(x)\}$, $x \in [0, 1]$, and the covariance function $\text{Cov}\{\eta(x), \eta(x')\}$, $x, x' \in [0, 1]$, exist and are continuous. Functional data consist of a sequence $\{\eta_i(\cdot)\}_{i=1}^n$ of stochastic processes called trajectories, each having the same distribution as $\eta(\cdot)$, thus also the same mean and covariance

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

functions. These trajectories are decomposed as $\eta_i(\cdot) = m(\cdot) + \chi_i(\cdot)$, where the centered trajectories $\chi_i(\cdot)$ are small-scale variations of x on the i th trajectory, and are continuous stochastic processes with $E\{\chi_i(x)\} \equiv 0$ and covariance function $\text{Cov}\{\chi_i(x), \chi_i(x')\}, x, x' \in [0, 1]$.

Estimating the functional mean $E\{\eta(\cdot)\}$ using simultaneous confidence bands (SCBs) has been investigated by Degras (2011), Cao, Yang and Todem (2012), Ma, Yang and Carroll (2012), Gu et al. (2014), Zheng, Yang and Härdle (2014), Choi and Reimherr (2018), and Telschow and Schwartzman (2022). The simultaneous confidence envelope (SCE) of a bivariate covariance function was established in Cao et al. (2016) for dense functional data, and a sharper SCB was proposed by Wang et al. (2020a) for an univariate stationary covariance function. See also related works on covariance estimation for stationary stochastic processes over an infinite domain in Hall, Fisher and Hoffmann (1994), and for sparse longitudinal data, see Meyer (1998) and Zhou, Lin and Liang (2018).

All of the above are limited to independent and identically distributed (i.i.d.) observations $\{\eta_i(\cdot)\}_{i=1}^n$ of the functional random variable $\eta(\cdot)$. However, many functional data are collected order over time, and exhibit temporal dependence. These data can be regarded as functional dependent data. Estimating the functional mean $E\{\eta(\cdot)\}$ has been investigated by Chen and Song (2015), Guo and Chen (2019), Horváth Kokoszka and Reeder (2013), and Li and Yang (2021). For example, Li and Yang (2021) studied EEG data for human subjects in an eyes-closed resting state. Recorded at a sample rate of 1000 Hz (i.e., a recording at every 0.001 second), the EEG series is divided into $n = 300$ consecutive segments $\{\eta_t(\cdot)\}_{t=1}^{300}$, with 200 EEG signals in each segment. Figures 1 and 2 show estimates (middle surface) of the autocovariance function $\text{Cov}\{\eta_t(x), \eta_{t+1}(x')\}, x, x' \in [0, 1]$, which differs significantly from zero (flat surface in Figure 2), indicating strong dependence between $\eta_t(\cdot)$ and $\eta_{t+1}(\cdot)$, see Section 6.

To model time-ordered and time-dependent trajectories $\{\eta_t(\cdot)\}_{t=1}^n$, the centered trajectories $\chi_t(\cdot)$ are embedded into a strictly stationary functional infinite moving average series, or FMA(∞) series $\{\chi_t(\cdot)\}_{t=-\infty}^{\infty}$, as in Li and Yang (2021),

$$\chi_t(\cdot) = \sum_{t'=0}^{\infty} A_{t'} \zeta_{t-t'}(\cdot), t \in \mathbb{Z}, \quad (1.1)$$

where the bounded linear operators $A_{t'} : \mathcal{L}^2[0, 1] \rightarrow \mathcal{L}^2[0, 1]$ are scalar coefficients in a classic MA(∞) (Definition 3.2.1, Brockwell and Davis (1991)). The processes $\{\zeta_t(\cdot)\}_{t \in \mathbb{Z}}$ are strong functional white noises (Definition 3.1 of Bosq (2000)), corresponding to the classic white noises in Brockwell and Davis

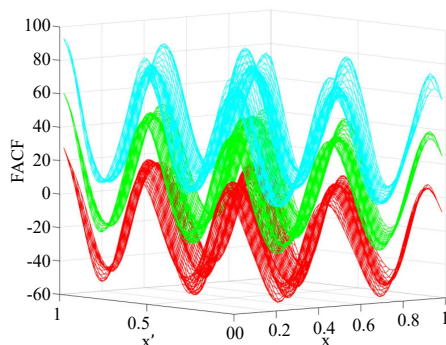


Figure 1. Two-step estimate $\hat{C}_1(\cdot, \cdot)$ of EEG data with $r = \lfloor n/4 \rfloor$ (middle surface) and its 95% SCE (upper and lower surfaces).

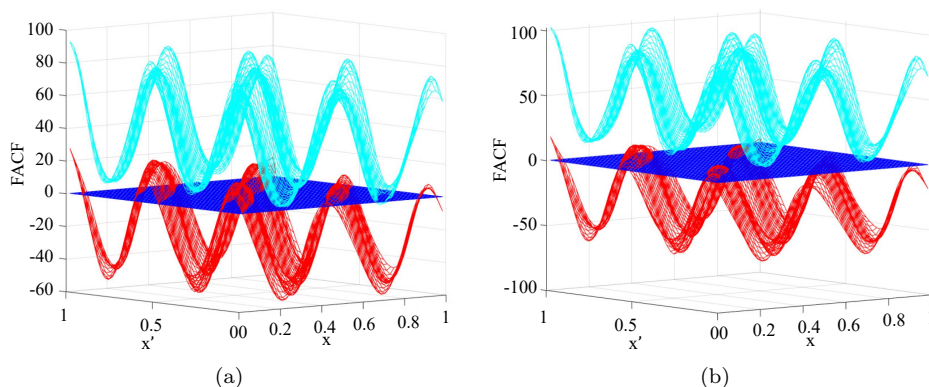


Figure 2. Null hypothesis zero plane and SCEs (upper and lower surfaces) for $C_1(\cdot, \cdot)$ of EEG data with (a) $\alpha = 0.05$, (b) $\alpha = 0.01$.

(1991). They are square-integrable continuous stochastic processes, i.i.d. over index $t \in \mathbb{Z}$, with $E\{\zeta_t(\cdot)\} \equiv 0$ and continuous covariance function $G(x, x') = E\{\zeta_t(x)\zeta_t(x')\}$. Mercer's lemma (Lemma 1.3, Bosq (2000)) states that $G(x, x') \equiv \sum_{k=1}^{\infty} \lambda_k \psi_k(x) \psi_k(x')$, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ and corresponding eigenfunctions $\{\psi_k\}_{k=1}^{\infty}$, the latter being an orthonormal basis of $\mathcal{L}^2[0, 1]$, such that $\sum_{k=1}^{\infty} \lambda_k < \infty$, $\{\psi_k\}_{k=1}^{\infty} \subset C[0, 1]$ and $\int G(x, x') \psi_k(x') dx' = \lambda_k \psi_k(x)$.

For any $t \in \mathbb{Z}$, $\zeta_t(\cdot)$ allows the well-known Karhunen–Loève \mathcal{L}^2 representation $\zeta_t(\cdot) = \sum_{k=1}^{\infty} \zeta_{tk} \phi_k(\cdot)$, in which the rescaled eigenfunctions, ϕ_k , called functional principal components (FPCs), satisfy that $\phi_k = \sqrt{\lambda_k} \psi_k$ and $\int \zeta_t(x) \phi_k(x) dx = \lambda_k \zeta_{tk}$, for $k \geq 1$. Therefore, the random coefficients ζ_{tk} are uncorrelated over $k \in \mathbb{N}_+$, and are i.i.d. over index $t \in \mathbb{Z}$, with mean zero and variance one. We assume that $\sum_{k=1}^{\infty} \|\phi_k\|_{\infty} < +\infty$ so that the Karhunen–Loève series converges

absolutely uniformly almost surely by the dominated convergence theorem.

The FMA(∞) coefficient operators $A_t, t \in \mathbb{N}$, are defined relative to the orthonormal basis $\{\psi_k\}_{k=1}^\infty$ of $\mathcal{L}^2[0, 1]$

$$A_t \left\{ \sum_{k=1}^{\infty} c_k \psi_k(\cdot) \right\} = \sum_{k=1}^{\infty} a_{tk} c_k \psi_k(\cdot), \quad a_{tk} \in \mathbb{R}, \quad t \in \mathbb{N}$$

$$|a_{tk}| < C_a \rho_a^t, \quad C_a \in (0, \infty), \quad \rho_a \in (0, 1), \quad t \in \mathbb{N}, \quad k \in \mathbb{N}_+. \quad (1.2)$$

Geometric decay of the MA coefficients $\{a_{tk}\}_{t \in \mathbb{N}}$ in (1.2) holds for many causal time series, such as the ARMA; see equation (3.3.6) of Brockwell and Davis (1991).

The following holds absolutely uniformly almost surely according to (1.1):

$$\begin{aligned} \chi_t(\cdot) &= \sum_{t'=0}^{\infty} A_{t'} \left\{ \sum_{k=1}^{\infty} \zeta_{t-t',k} \phi_k(\cdot) \right\} = \sum_{t'=0}^{\infty} \sum_{k=1}^{\infty} a_{t',k} \zeta_{t-t',k} \phi_k(\cdot) \\ &= \sum_{k=1}^{\infty} \left(\sum_{t'=0}^{\infty} a_{t',k} \zeta_{t-t',k} \right) \phi_k(\cdot) = \sum_{k=1}^{\infty} \xi_{tk} \phi_k(\cdot), \quad t \in \mathbb{Z}, \end{aligned} \quad (1.3)$$

in which

$$\xi_{tk} = \sum_{t'=0}^{\infty} a_{t',k} \zeta_{t-t',k}, \quad t \in \mathbb{Z}, \quad k \in \mathbb{N}_+. \quad (1.4)$$

Thus, for each fixed $k \in \mathbb{N}_+$, the time series $\{\xi_{tk}\}_{t \in \mathbb{Z}}$ is a classic MA(∞) expressed in terms of the i.i.d. sequence $\{\zeta_{tk}\}_{t \in \mathbb{Z}}$. Note that the classic MA(∞) is sufficiently broad to include the causal ARMA(p, q), and consequently AR(p) and MA(q), as a special.

Assume for convenience that for each fixed $k \in \mathbb{N}_+$, $\sum_{t=0}^{\infty} a_{tk}^2 \equiv 1$. Then the mean and variance of ξ_{tk} equal zero and one, respectively, for $t \in \mathbb{Z}, k \in \mathbb{N}_+$, and are uncorrelated over $k \in \mathbb{N}_+$. The covariance function of $\{\chi_t(\cdot)\}_{t=-\infty}^\infty$ is equal to that of $\{\zeta_t(\cdot)\}_{t=-\infty}^\infty$,

$$\begin{aligned} E\{\chi_t(x) \chi_t(x')\} &= E \left\{ \sum_{k=1}^{\infty} \xi_{tk} \phi_k(x) \sum_{k=1}^{\infty} \xi_{tk} \phi_k(x') \right\} = \sum_{k=1}^{\infty} E(\xi_{tk}^2) \phi_k(x) \phi_k(x') \\ &= \sum_{k=1}^{\infty} \phi_k(x) \phi_k(x') = G(x, x') = E\{\zeta_t(x) \zeta_t(x')\}, \quad x, x' \in [0, 1]. \end{aligned}$$

Hence, $\chi_t(\cdot) = \sum_{k=1}^{\infty} \xi_{tk} \phi_k(\cdot)$ is the Karhunen–Loève expansion of the strictly stationary FMA(∞) series $\{\chi_t(\cdot)\}_{t=-\infty}^\infty$, where the convergence is absolutely uniformly almost sure according to (1.3), and the random coefficients $\{\xi_{tk}\}_{t \in \mathbb{Z}, k \in \mathbb{N}_+}$

are called FPC scores.

Raw data of the FMA(∞) are in the form

$$Y_{tj} = m\left(\frac{j}{N}\right) + \sum_{k=1}^{\infty} \xi_{tk} \phi_k\left(\frac{j}{N}\right) + \sigma\left(\frac{j}{N}\right) \varepsilon_{tj}, \quad 1 \leq t \leq n, \quad 1 \leq j \leq N, \quad (1.5)$$

with MA(∞) FPC scores ξ_{tk} given in (1.4). The terms $\sigma(j/N)\varepsilon_{tj}$ represent measurement errors which occur with data collection, $\{\varepsilon_{tj}\}_{t=-\infty, j=1}^{n, N}$ are i.i.d. with mean zero and variance one, and the standard deviation function $\sigma(\cdot)$ satisfies the Hölder continuity in Assumption (A2).

The autocovariance function (ACF) in classic time series extends naturally to the following functional autocovariance function (FACF):

$$C_h(x, x') = \mathbb{E}\{\chi_t(x) \chi_{t+h}(x')\}, \quad h \in \mathbb{N}, \quad t \in \mathbb{Z}, \quad x, x' \in [0, 1]. \quad (1.6)$$

Obviously, $C_0(x, x') \equiv G(x, x')$. It is shown in Proposition 1 that if the highest order h for which $C_h(x, x')$ is not identically zero is a finite integer $q \in \mathbb{N}_+$, then the FMA(∞) reduces to the simpler FMA(q).

This study extends the simultaneous confidence region for the i.i.d. functional data covariance function in Cao et al. (2016) and Wang et al. (2020a) to the much more complicated FACFs $C_h(x, x')$, for $h \in \mathbb{N}$. The FACFs are estimated in two steps. Step one estimates all n trajectories $\{\eta_t(\cdot)\}_{t=1}^n$ and their population mean $m(\cdot)$ using B-splines. Step two estimates the FACF $C_h(x, x')$ using the estimated trajectories and the population mean as if they were the true values. The greatest technical difficulty is to establish in Proposition 2 under the assumption of moving average trajectories $\{\eta_t(\cdot)\}_{t=1}^n$ that the proposed FACF estimator is asymptotically equivalent to the infeasible FACF estimator when all trajectories $\{\chi_t(\cdot)\}_{t=1}^n$ are entirely observed. This “oracle efficiency” allows us to construct an asymptotic SCE for each FACF $C_h(x, x')$, for $h \in \mathbb{N}$. The second technical difficulty is to obtain in Theorem 1 the limiting distribution of the FACF estimate under a moving average dependence of the trajectories $\{\chi_t(\cdot)\}_{t=1}^n$.

A nonparametric simultaneous confidence region is a versatile tool for global inference on functions; see, for instance, Wang and Yang (2009), Wang et al. (2014), Gu and Yang (2015), Zheng et al. (2016), Wang et al. (2020b), Gu, Wang and Yang (2021), Yu et al. (2021), and Zhong and Yang (2021) for the theory and applications of SCBs in diverse contexts. The SCE for $C_h(x, x')$ enables one to test against any null hypothesis, such as $H_0 : C_h(x, x') \equiv 0$, for any positive integers h . For the EEG time series, this null hypothesis is rejected when $h = 9$, but retained for $h = 10, 11, 12, 13, 14$, leading to the appropriate

decision of FMA(9) for the data; see Section 6, especially the SCEs in Figures 1 and 2.

The remainder of the paper is organized as follows. Section 2 examines the theoretical properties of FACF $C_h(x, x')$ and defines its infeasible and two-step estimators. Section 3 establishes the limiting distribution of the infeasible FACF estimator, and the asymptotic equivalence of the infeasible and two-step estimators. Asymptotic SCEs for the FACF are proposed in Section 4, with implementation details. Section 5 presents our simulation studies, and we discuss our analysis of EEG data in Section 6. All technical proofs are provided in the Supplementary Material.

2. FACF and Its Estimation

We begin by deriving the explicit formula of the FACF $C_h(x, x')$ in (1.6). Denote by $\gamma_k(h) = E(\xi_{tk}\xi_{t+h,k}) = \sum_{m=0}^{\infty} a_{mk}a_{m+h,k}$, $h \in \mathbb{N}$, $t \in \mathbb{Z}$, $k \in \mathbb{N}_+$ the classic autocovariance function of $\{\xi_{tk}\}_{t \in \mathbb{Z}}$, with the convention that $\gamma_k(-h) \equiv \gamma_k(h)$, for $h \in \mathbb{N}$. Noting the absolute almost sure convergence in (1.4), we can derive, with uniform absolute convergence,

$$\begin{aligned} C_h(x, x') &= \frac{1}{n-h} \sum_{t=1}^{n-h} \sum_{k,k'=1}^{\infty} E(\xi_{tk}\xi_{t+h,k'}) \phi_k(x) \phi_{k'}(x') \\ &= \frac{1}{n-h} \sum_{t=1}^{n-h} \sum_{k=1}^{\infty} \gamma_k(h) \phi_k(x) \phi_k(x') \\ &= \frac{1}{n-h} \sum_{t=1}^{n-h} \sum_{k,k'=1}^{\infty} \delta_{kk'} \gamma_k(h) \phi_k(x) \phi_{k'}(x'), \quad x, x' \in [0, 1], \end{aligned} \quad (2.1)$$

where $\delta_{kk'} = 1$ for $k = k'$, and is zero otherwise. The next proposition plays the same role as its classic analog, Proposition 3.2.1 in Brockwell and Davis (1991).

Proposition 1. *If $C_h(x, x') \equiv 0$, $x, x' \in [0, 1]$ for $h > q$, $C_q(x, x') \neq 0$ for some $q \in \mathbb{N}_+$, and the MA(∞) equation in (1.4) is invertible for all $k \in \mathbb{N}_+$ (i.e., there exist $b_{t',k} \in \mathbb{R}$ such that $\zeta_{tk} = \sum_{t'=0}^{\infty} b_{t',k} \xi_{t-t',k}$, $t \in \mathbb{Z}$, $k \in \mathbb{N}_+$), then $\chi_t(\cdot) = \sum_{t'=0}^q A_{t'} \zeta_{t-t'}(\cdot)$, $t \in \mathbb{Z}$, absolutely uniformly almost surely, and hence $\{\chi_t(\cdot)\}_{t=-\infty}^{\infty}$ is an FMA(q) series.*

According to Proposition 1, a finite moving average order q is determined if the FACF $C_h(\cdot, \cdot)$ vanishes exactly beyond order q . Testing whether an FACF is identically zero is conveniently done by estimating the FACF using an SCE.

If the t th centered trajectory $\chi_t(\cdot) = \eta_t(\cdot) - m(\cdot)$, $1 \leq t \leq n$, is fully observed, we can estimate the ACF $C_h(x, x')$ with the following sample ACF, as in (7.2.1) of Brockwell and Davis (1991):

$$\tilde{C}_h(x, x') = n^{-1} \sum_{t=1}^{n-h} \chi_t(x) \chi_{t+h}(x'), \quad x, x' \in [0, 1], \quad (2.2)$$

which can be written explicitly as

$$\tilde{C}_h(x, x') = n^{-1} \sum_{t=1}^{n-h} \sum_{k, k'=1}^{\infty} \xi_{tk} \xi_{t+h, k'} \phi_k(x) \phi_{k'}(x'). \quad (2.3)$$

Because $\{\chi_t(\cdot)\}_{t=1}^n$ are unobservable, the above estimator function $\tilde{C}_h(x, x')$ is infeasible. However, it does suggest the following plug-in sample covariance estimator:

$$\hat{C}_h(x, x') = n^{-1} \sum_{t=1}^{n-h} \hat{\chi}_t(x) \hat{\chi}_{t+h}(x'), \quad x, x' \in [0, 1], \quad (2.4)$$

in which

$$\hat{\chi}_t(\cdot) = \hat{\eta}_t(\cdot) - \hat{m}(\cdot), \quad 1 \leq t \leq n, \quad (2.5)$$

where $\hat{\eta}_t(\cdot)$ and $\hat{m}(\cdot)$ are some suitable estimators of $\eta_t(\cdot)$ and $m(\cdot)$, respectively.

In this work, we use the B-spline estimators $\hat{\eta}_t(x)$ and $\hat{m}(x)$. To describe a B-spline estimator, denote by $\{t_\ell\}_{\ell=1}^{J_s}$ a sequence of equally-spaced points. We call $t_\ell = \ell / (J_s + 1)$, $\ell \in \{1, \dots, J_s\}$, $0 < t_1 < \dots < t_{J_s} < 1$ interior knots, which divide the interval $[0, 1]$ into $(J_s + 1)$ equal subintervals $I_0 = [0, t_1]$, $I_\ell = [t_\ell, t_{\ell+1}]$, $\ell \in \{1, \dots, J_s - 1\}$, $I_{J_s} = [t_{J_s}, 1]$. For any positive integer p , let $t_{1-p} = \dots = t_0 = 0$ and $1 = t_{J_s+1} = \dots = t_{J_s+p}$ be auxiliary knots. Let $\mathcal{S}^{(p-2)} = \mathcal{S}^{(p-2)}[0, 1]$ be the polynomial spline space of order p on I_ℓ , $\ell \in \{0, \dots, J_s\}$, which consists of all $(p-2)$ times continuously differentiable functions on $[0, 1]$ that are polynomials of degree $(p-1)$ on subintervals I_ℓ , $0 \leq \ell \leq J_s$. Denote by $\{B_{\ell,p}(x), 1 \leq \ell \leq J_s + p\}$ the p -th order B-spline basis functions of $\mathcal{S}^{(p-2)}$ (de Boor (2001)), $\mathcal{S}^{(p-2)} = \{\sum_{\ell=1}^{J_s+p} \lambda_{\ell,p} B_{\ell,p}(\cdot) | \lambda_{\ell,p} \in \mathbb{R}\}$.

We estimate the trajectories $\eta_t(\cdot)$ and their population mean $m(\cdot)$ using the spline regression

$$\hat{m}(\cdot) = n^{-1} \sum_{t=1}^n \hat{\eta}_t(\cdot), \quad \hat{\eta}_t(\cdot) = \underset{g(\cdot) \in \mathcal{S}^{(p-2)}}{\operatorname{argmin}} \sum_{j=1}^N \left\{ Y_{tj} - g\left(\frac{j}{N}\right) \right\}^2. \quad (2.6)$$

3. Asymptotic Properties

This section studies the asymptotic properties of the proposed estimators.

3.1. Assumptions and the infeasible estimator

To study the asymptotic properties of the two-step spline estimator $\hat{C}_h(x, x')$, we require some mild assumptions. For sequences of real numbers a_n and b_n , denote $a_n \asymp b_n$ if $|a_n| \leq C|b_n|$, $|b_n| \leq C|a_n|$, $n \in \mathbb{N}_+$ for some constant $C > 0$. For any measurable function $\varphi(\cdot)$ defined on $[0, 1]$, denote $\|\varphi\|_\infty = \sup_{x \in [0, 1]} |\varphi(x)|$, and $\varphi^{(q)}(x)$ its q -th order derivative with respect to x , if it exists. For any functions $\phi(\cdot), \varphi(\cdot) \in \mathcal{L}^2[0, 1]$, define their theoretical and empirical inner products as $\langle \phi, \varphi \rangle = \int_{[0, 1]} \phi(x) \varphi(x) dx$, $\langle \phi, \varphi \rangle_N = N^{-1} \sum_{j=1}^N \phi(j/N) \varphi(j/N)$. The corresponding theoretical and empirical norms are $\|\phi\|_2^2 = \langle \phi, \phi \rangle$ and $\|\phi\|_{2,N}^2 = \langle \phi, \phi \rangle_N$.

For a nonnegative integer q and a real number $\mu \in (0, 1]$, write $\mathcal{H}^{(q, \mu)}[0, 1]$ as the space of (q, μ) -Hölder continuous functions, that is,

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

in which $\|\cdot\|_{q, \mu}$ denotes the (q, μ) -Hölder semi-norm. Denote by $I_n \in \mathbb{N}_+$ a truncation index such that the FMA(∞) $\chi_t(\cdot)$ is well approximated by $\sum_{t'=0}^{I_n} A_{t'} \zeta_{t-t'}(\cdot)$:

$$I_n > -10 \frac{\log n}{\log \rho_a}, \quad I_n \asymp \log n, \quad (3.1)$$

where ρ_a is the geometric decay parameter in (1.2).

The following are some technical assumptions.

- (A1) *There exist $q \in \mathbb{N}_+, \mu \in (0, 1]$, such that $m(\cdot) \in \mathcal{H}^{(q, \mu)}[0, 1]$. In the following, we denote $p^* = q + \mu$.*
- (A2) *The standard deviation function $\sigma(\cdot) \in \mathcal{H}^{(0, \nu)}[0, 1]$, for $\nu \in (0, 1]$, and for some constants $M_\sigma > 0$, $\sup_{x \in [0, 1]} \sigma(x) \leq M_\sigma$.*
- (A3) *There exists $\theta > 0$, such that as $N \rightarrow \infty$, $n = n(N) \rightarrow \infty$ and $n = O(N^\theta)$.*
- (A4) *The rescaled FPCs $\phi_k(\cdot) \in \mathcal{H}^{(q, \mu)}[0, 1]$, with $\sum_{k=1}^\infty \|\phi_k\|_{q, \mu} < +\infty$, $\sum_{k=1}^\infty \|\phi_k\|_\infty < +\infty$; for increasing positive integers $\{k_n\}_{n=1}^\infty$, as $n \rightarrow \infty$, $\sum_{k_n+1}^\infty \|\phi_k\|_\infty = O(n^{-1/2})$ and $k_n = O(n^\omega)$, for some $\omega > 0$.*
- (A5) *The i.i.d. variables $\{\varepsilon_{tj}\}_{t \geq 1, j \geq 1}$ are independent of $\{\zeta_{tk}\}_{t \in \mathbb{Z}, k \in \mathbb{N}_+}$. The random coefficients $\{\zeta_{tk}\}_{t \in \mathbb{Z}, k \in \mathbb{N}_+}$ are independent over $k \in \mathbb{N}_+$ and i.i.d. over*

$t \in \mathbb{Z}$. The number of distinct distributions for all $\{\zeta_{tk}\}_{t \in \mathbb{Z}, k \in \mathbb{N}_+}$ is finite and $\sup_{k \geq 1} \mathbb{E}(|\zeta_{1k}|^{r_0}) < \infty$, for some $r_0 \geq 4$. There exist $C_1, C_2 \in (0, +\infty)$, $\gamma_1, \gamma_2 \in (1, +\infty)$, $\beta_1, \beta_2 \in (0, 1/2)$, and i.i.d. $N(0, 1)$ variables $\{Z_{tj, \varepsilon}\}_{t=1, j=1}^{n, N}$, $\{Z_{tk, \zeta}\}_{t=-I_n+1, k=1}^{n, k_n}$, such that

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

$$\mathbb{P} \left\{ \max_{1 \leq t \leq n} \max_{1 \leq \tau \leq N} \left| \sum_{j=1}^{\tau} \varepsilon_{tj} - \sum_{j=1}^{\tau} Z_{tj, \varepsilon} \right| > N^{\beta_2} \right\} < C_2 N^{-\gamma_2}.$$

(A6) The spline order $p \geq p^*$, the number of interior knots $J_s \asymp N^\gamma d_N$, for some $\tau > 0$ with $d_N + d_N^{-1} = O(\log^\tau N)$ as $N \rightarrow \infty$, and for p^* in Assumption (A1), ν in Assumption (A2), θ in Assumption (A3), and r_0, β_1, β_2 in Assumption (A5),

$$\max \left\{ \frac{4\theta}{r_0 p^*} + \frac{\theta}{2p^*}, 1 - \nu \right\} < \gamma < 1 - \frac{\theta}{2} - \beta_2 - \theta\beta_1.$$

Assumptions (A1)–(A2) are regular conditions for the spline smoother. In particular, Assumption (A1) controls the size of the bias of the spline smoother for $m(\cdot)$, and Assumption (A2) ensures the variance function should be uniformly bounded. Assumption (A3) requires that sample size n grows in sync with the number N of observations per curve, and not faster than N^θ . Hence, all asymptotics in this section are stated with $N \rightarrow \infty$ only, not with $N \rightarrow \infty$ and $n \rightarrow \infty$. Assumption (A4) guarantees the bounded smoothness of the principal components. The independence of the latent FPC scores $\{\zeta_{tk}\}_{t \geq 1, k \geq 1}$ over $k \in \mathbb{N}_+$ in Assumption (A5) is common in existing works on functional data analysis; see Cao, Yang and Todem (2012), Ma, Yang and Carroll (2012), Gu et al. (2014), and Zheng, Yang and Härdle (2014). The probability inequalities in Assumption (A5) provide strong Gaussian approximations of the measurement errors $\{\varepsilon_{tj}\}_{t \geq 1, j \geq 1}$ and the latent FPC scores $\{\zeta_{tk}\}_{t \geq 1, k \geq 1}$. As a result, all the main results of this section hold without the data being Gaussian. Assumption (A5) is a high-level assumption that can be guaranteed by the elementary Assumption (A5') below, together with Assumption (A4); see Lemma S.3 in the Supplementary Material. Assumption (A6) on the choice of the knot number satisfies the requirements for B-spline smoothing.

(A5') The i.i.d. variables $\{\varepsilon_{tj}\}_{t \geq 1, j \geq 1}$ are independent of $\{\zeta_{tk}\}_{t \in \mathbb{Z}, k \in \mathbb{N}_+}$. The random coefficients $\{\zeta_{tk}\}_{t \in \mathbb{Z}, k \in \mathbb{N}_+}$ are independent over $k \in \mathbb{N}_+$, and are i.i.d. over $t \in \mathbb{Z}$. The number of distinct distributions for all $\{\zeta_{tk}\}_{t \in \mathbb{Z}, k \in \mathbb{N}_+}$ is finite. There exist constants $r_1 > 4 + 2\omega$, $r_2 > 4 + 2\theta$, for ω in Assumption (A4) and θ in Assumption (A3), such that $E(|\varepsilon_{11}|^{r_2})$ and $E(|\zeta_{1k}|^{r_1}), k \in \mathbb{N}_+$, are finite.

Remark 1. The aforementioned assumptions are easily satisfied in practice. One simple and reasonable setup for the above parameters q , μ , θ , p and γ is the following: $q + \mu = p^* = 4$, $\nu = 1, r_0 > 4$, $\theta = 1$, $p = 4$ (cubic spline), $\gamma = 3/8$, $d_N \asymp \log \log N$. These constants are used as defaults for the implementation in Section 4.

We first examine the infeasible estimator $\tilde{C}_h(x, x')$. Denote $\Delta_h(x, x') = \tilde{C}_h(x, x') - \{(n-h)/n\}C_h(x, x')$, $x, x' \in [0, 1]$. Then, (2.1) and (2.3) imply that

$$\Delta_h(x, x') = n^{-1} \sum_{t=1}^{n-h} \sum_{k, k'=1}^{\infty} \{\xi_{tk} \xi_{t+h, k'} - \delta_{kk'} \gamma_k(h)\} \phi_k(x) \phi_{k'}(x').$$

Let $\varphi_h(x, x')$ be a zero-mean Gaussian random field defined on $[0, 1]^2$, with covariance function $\Omega_h(x, x', y, y')$, defined as

$$\begin{aligned} \Omega_h(x, x', y, y') &= \text{Cov} \{ \varphi_h(x, x'), \varphi_h(y, y') \} \\ &= \sum_{l=-\infty}^{\infty} C_l(x, y) C_l(x', y') + \sum_{l=-\infty}^{\infty} C_{l+h}(x, x') C_{l-h}(y, y') \\ &\quad + \sum_{m=1}^{\infty} \{E(\zeta_{0m}^4) - 3\} \gamma_m^2(h) \phi_m(x) \phi_m(x') \phi_m(y) \phi_m(y'), \end{aligned}$$

for $x, x', y, y' \in [0, 1]$. Then $E\{\varphi_h^2(x, x')\} = \Omega_h(x, x', x, x') = \Xi_h(x, x')$ satisfies a functional version of Bartlett's formula

$$\begin{aligned} \Xi_h(x, x') &= \sum_{l=-\infty}^{\infty} \{C_l(x, x) C_l(x', x') + C_{l+h}(x, x') C_{l-h}(x, x')\} \\ &\quad + \sum_{m=1}^{\infty} \{E(\zeta_{0m}^4) - 3\} \gamma_m^2(h) \phi_m^2(x) \phi_m^2(x'). \end{aligned} \quad (3.2)$$

Theorem 1 below gives the asymptotics of $\Delta_h(x, x')$.

Theorem 1. *Under Assumptions (A3)–(A5), for $h \in \mathbb{N}$, as $N \rightarrow \infty$,*

$$\begin{aligned} & \|n\mathbb{E}\{\Delta_h(x, x')\Delta_h(y, y')\} - \Omega_h(x, x', y, y')\|_\infty \rightarrow 0, \\ & \sqrt{n}\Delta_h(x, x') \rightarrow_D \varphi_h(x, x'), \quad \sqrt{n}\{\tilde{C}_h(x, x') - C_h(x, x')\} \rightarrow_D \varphi_h(x, x'). \end{aligned}$$

Remark 2. Bartlett's formula (Chapter 7, Brockwell and Davis (1991)) for the sample ACF of a numerical time series is extended to a functional Bartlett's formula (3.2) for an infeasible FACF $\tilde{C}_h(x, x')$. If we use the degenerate FMA(∞) with $\phi_1(x) \equiv 1, \phi_k(x) \equiv 0, k \geq 2$, the infeasible ACF $\tilde{C}_h(x, x')$ simplifies to the sample ACF $\gamma^*(h) = n^{-1} \sum_{t=1}^{n-h} \xi_{t1} \xi_{t+h,1}$ of the numerical time series $\{\xi_{t1}\}_{t=-\infty}^\infty$, and $\Xi_h(x, x')$ becomes the asymptotic covariance of $\gamma^*(h)$, $\{\mathbb{E}(\zeta_{01}^4) - 3\}\gamma_1^2(h) + \sum_{l=-\infty}^\infty \{\gamma_1^2(l) + \gamma_1(l+h)\gamma_1(l-h)\}$, as in equation (7.3.3) in Brockwell and Davis (1991), by setting $p = q = h$.

Theorem 1 is proved in Section S3 of the Supplementary Material. While it provides desirable asymptotics of the infeasible estimator $\tilde{C}_h(x, x')$, it is not a statistic because the centered trajectories $\chi_t(\cdot)$ are unobservable.

3.2. Oracle efficiency

Proposition 2 states that the proposed two-step FACF estimator $\hat{C}_h(x, x')$ in (2.4) is oracally efficient: up to order $n^{1/2}$, it is asymptotically equivalent to, or as efficient as $\tilde{C}_h(x, x')$, the infeasible FACF estimator if all centered trajectories $\chi_t(\cdot)$ are fully known by “oracle.” Thus, $\hat{C}_h(x, x')$ enjoys all the same asymptotic properties as $\tilde{C}_h(x, x')$.

Proposition 2. *Under Assumptions (A1)–(A6), for $h \in \mathbb{N}$, as $N \rightarrow \infty$,*

$$\sup_{(x, x') \in [0, 1]^2} \left| \hat{C}_h(x, x') - \tilde{C}_h(x, x') \right| = o_p\left(n^{-1/2}\right).$$

Proposition 2 and Theorem 1 lead to the following.

Theorem 2. *Under Assumptions (A1)–(A6), for $h \in \mathbb{N}$, as $N \rightarrow \infty$,*

$$\begin{aligned} \sup_{(x, x') \in [0, 1]^2} \left| \hat{C}_h(x, x') - C_h(x, x') - \Delta_h(x, x') \right| &= o_p\left(n^{-1/2}\right), \\ \sqrt{n} \left\{ \hat{C}_h(x, x') - C_h(x, x') \right\} &\rightarrow_D \varphi_h(x, x'). \end{aligned} \quad (3.3)$$

Therefore, the limiting distribution of $\hat{C}_h(x, x') - C_h(x, x')$ is the same as $\tilde{C}_h(x, x') - C_h(x, x')$, and hence the term “oracle efficiency.”

4. SCEs

In this section, we construct an SCE for the functional autocovariance function $C_h(x, x')$.

4.1. Asymptotic SCE

The main theorem in this section establishes the asymptotic behavior of the normalized maximal deviation of the FCAF estimator $\hat{C}_h(x, x')$. Using (3.3) in Theorem 2, we obtain the following standardized limiting distribution:

$$\sqrt{n} \left\{ \hat{C}_h(x, x') - C_h(x, x') \right\} \Xi_h^{-1/2}(x, x') \rightarrow_D \varphi_h(x, x') \Xi_h^{-1/2}(x, x'), \quad (4.1)$$

in which $\varphi_h(x, x')$ is the mean-zero Gaussian random field defined in Theorem 1 with the pointwise variance function $\Xi_h(x, x') = E\{\varphi_h^2(x, x')\}$. Therefore, the standardized random field $\varphi_h(x, x') \Xi_h^{-1/2}(x, x')$ has mean zero and variance one. Denote by $Q_{1-\alpha}$ the $100(1 - \alpha)$ th percentile of the absolute maxima distribution of $\varphi_h(x, x') \Xi_h^{-1/2}(x, x')$, that is, for $\alpha \in (0, 1)$, $h \in \mathbb{N}$,

$$\mathbb{P} \left\{ \sup_{(x, x') \in [0, 1]^2} |\varphi_h(x, x')| \Xi_h^{-1/2}(x, x') \leq Q_{1-\alpha} \right\} = 1 - \alpha. \quad (4.2)$$

The next theorem follows directly from (4.1) and (4.2).

Theorem 3. *Under Assumptions (A1)–(A6), for $\alpha \in (0, 1)$, $h \in \mathbb{N}$,*

$$\lim_{N \rightarrow \infty} \mathbb{P} \left\{ \sup_{(x, x') \in [0, 1]^2} n^{1/2} \left| \hat{C}_h(x, x') - C_h(x, x') \right| \Xi_h^{-1/2}(x, x') \leq Q_{1-\alpha} \right\} = 1 - \alpha.$$

Corollary 1. *Under Assumptions (A1)–(A6), as $N \rightarrow \infty$, an asymptotic $100(1 - \alpha)\%$ SCE for $C_h(x, x')$ is $\hat{C}_h(x, x') \pm n^{-1/2} Q_{1-\alpha} \Xi_h^{1/2}(x, x')$, $x, x' \in [0, 1]$.*

4.2. Knot selection

The number of knots J_s for the spline smoothing is selected subject to the constraints of Assumption (A6). The smoothness order (q, μ) of the mean function $m(\cdot)$ and eigenfunctions $\phi_k(\cdot)$ is taken as $(3, 1)$ or $(4, 0)$, by default, with a matching spline order $p = 4$ (cubic spline). Therefore, $J_s = \lceil cN^\gamma \log \log(N) \rceil$ is recommended with a constant c , where $[a]$ denotes the integer part of a . The default values of the parameters $\gamma = 3/8$ and $c = 0.8$ are adequate in our simulations.

4.3. FPC analysis

We now describe the covariance function estimator $\hat{C}_0(\cdot, \cdot) = \hat{G}(\cdot, \cdot)$, eigenfunction estimators $\hat{\phi}_k(\cdot)$, and eigenvalue estimators $\hat{\lambda}_k$ in the FPC analysis. We can estimate $C_0(\cdot, \cdot)$ by

$$\hat{C}_0(x, x') = n^{-1} \sum_{t=1}^n \hat{\chi}_t(x) \hat{\chi}_t(x') = \sum_{s=1}^{J_s+p} \sum_{s'=1}^{J_s+p} \hat{\beta}_{ss'} B_{s,p}(x) B_{s',p}(x'), \quad (4.3)$$

where $\hat{\chi}_t(\cdot)$ is defined in (2.5) and $\hat{\beta}_{ss'}$ are the coefficients.

Denote $\mathbf{B}(x) = \{B_{1,p}(x), \dots, B_{J_s+p,p}(x)\}^\top$, and the $N \times (J_s + p)$ design matrix \mathbf{B} for spline regression is $\mathbf{B} = \{\mathbf{B}(1/N), \dots, \mathbf{B}(N/N)\}^\top$. Then, for any $k \in \{1, \dots, \kappa\}$, consider the following spline approximation for $\psi_k(\cdot)$: $\hat{\psi}_k(x') = \sum_{\ell=1}^{J_s+p} \hat{\gamma}_{\ell k} B_{\ell,p}(x')$, where $\hat{\gamma}_{\ell k}$ are the coefficients of the B-spline estimator, subject to $\hat{\gamma}_k^\top \mathbf{B}^\top \mathbf{B} \hat{\gamma}_k = N$ with $\hat{\gamma}_k = (\hat{\gamma}_{1,k}, \dots, \hat{\gamma}_{J_s+p,k})^\top$. The estimates of the eigenfunctions and eigenvalues, corresponding to ψ_k and λ_k , respectively, can be obtained by solving the eigenequations

$$\int \hat{C}_0(x, x') \hat{\psi}_k(x') dx' = \hat{\lambda}_k \hat{\psi}_k(x), \quad 1 \leq k \leq J_s + p. \quad (4.4)$$

According to (4.3), solving (4.4) is equivalent to solving $N^{-1} \mathbf{B}^\top(x) \hat{\beta} \mathbf{B}^\top \mathbf{B} \hat{\gamma}_k = \hat{\lambda}_k \mathbf{B}^\top(x) \hat{\gamma}_k$, $k \in \{1, \dots, \kappa\}$, where $\hat{\beta}^\top = (\hat{\beta}_{ss'})_{s,s'=1}^{J_s+p}$. The above is equivalent to solving $N^{-1} \hat{\beta} \mathbf{B}^\top \mathbf{B} \hat{\gamma}_k = \hat{\lambda}_k \hat{\gamma}_k$, for $1 \leq k \leq J_s + p$. Consider the following Cholesky decomposition: $\mathbf{B}^\top \mathbf{B} = \mathbf{L}_B \mathbf{L}_B^\top$. Solving (4.4) is equivalent to solving $\hat{\lambda}_k \mathbf{L}_B^\top \hat{\gamma}_k = N^{-1} \mathbf{L}_B^\top \hat{\beta} \mathbf{L}_B \mathbf{L}_B^\top \hat{\gamma}_k$, subject to $\hat{\gamma}_k^\top \mathbf{L}_B \mathbf{L}_B^\top \hat{\gamma}_k = N$, that is, $\hat{\lambda}_k$ and $\mathbf{L}_B^\top \hat{\gamma}_k$, $1 \leq k \leq J_s + p$, are the eigenvalues and unit eigenvectors, respectively, of $N^{-1} \mathbf{L}_B^\top \hat{\beta} \mathbf{L}_B$. In other words, $\hat{\gamma}_k$ is obtained by left multiplying $N^{1/2} (\mathbf{L}_B^\top)^{-1}$ to the unit eigenvectors of $N^{-1} \mathbf{L}_B^\top \hat{\beta} \mathbf{L}_B$, yielding $\hat{\psi}_k(\cdot)$. Consequently, $\hat{\phi}_k(x') = \hat{\lambda}_k^{1/2} \hat{\psi}_k(x')$. The k -th FPC score of the t -th curve is estimated by numerical integration:

$$\hat{\xi}_{tk} = N^{-1} \sum_{j=1}^N \hat{\lambda}_k^{-1} \hat{\chi}_t \left(\frac{j}{N} \right) \hat{\phi}_k \left(\frac{j}{N} \right), \quad 1 \leq t \leq n, 1 \leq k \leq J_s + p.$$

Instead of computing all $J_s + p$ eigenvalues, it is typical to truncate the spectral decomposition at a much smaller number of eigenvalues that account for 95% of the variation in the data, $\kappa = \operatorname{argmin}_{1 \leq l \leq J_s+p} \{ \sum_{k=1}^l \hat{\lambda}_k / \sum_{k=1}^{J_s+p} \hat{\lambda}_k \geq 0.95 \}$; see Cao et al. (2016) and Wang et al. (2020a).

4.4. Estimating the variance function Ξ_h and the percentile $Q_{1-\alpha}$

The strong approximation property of ξ_{tk} , guaranteed by Assumption (A5) and Lemma S.4, makes a procedure for approximating Ξ_h and $Q_{1-\alpha}$ possible. Denote the matrices $\Gamma_k^{(n)} = \{\gamma_k(|i-j|)\}_{i,j=1}^n, k = 1, \dots, \kappa$, where $\gamma_k(|i-j|)$ is the (i, j) th entry of $\Gamma_k^{(n)}$. It is easily shown that $\Gamma_k^{(n)}$ is the covariance matrix of random vector $(\xi_{1k}, \xi_{2k}, \dots, \xi_{nk})$. Useful estimates of $\gamma_k(h)$ can only be made if $n \geq 50$ and $0 \leq h \leq n/4$; see p. 221, Chapter 7, in Brockwell and Davis (1991). Therefore we consider the $r \times r$ submatrix $\Gamma_k^{(r)}$ of $\Gamma_k^{(n)}$, where r can be chosen as $\lfloor \sqrt{n} \rfloor$ or $\lfloor n/4 \rfloor$. Then, according to the definition of $\gamma_k(h), h \in \mathbb{N}$, we can construct the estimation $\hat{\Gamma}_k^{(r)}$ of $\Gamma_k^{(r)}$ as $\hat{\gamma}_k(h) = n^{-1} \sum_{t=1}^{n-h} (\hat{\xi}_{tk} - \hat{\xi}_{\cdot k})(\hat{\xi}_{t+h,k} - \hat{\xi}_{\cdot k})$, for $0 \leq h \leq r$, where $\hat{\xi}_{\cdot k} = n^{-1} \sum_{t=1}^{n-h} \hat{\xi}_{tk}, 1 \leq k \leq \kappa$.

For $1 \leq k \leq \kappa$, we can generate B independent replications of r -dimensional normal random vectors $\{(Z_{1k}^b, \dots, Z_{rk}^b)\}_{b=1}^B$ with mean zero and covariance matrix $\hat{\Gamma}_k^{(r)}$, where B is a preset large integer. Let $\hat{C}_h^b(x, x') = r^{-1} \sum_{t=1}^{r-h} \{\sum_{k=1}^{\kappa} Z_{tk}^b \hat{\phi}_k(x)\} \{\sum_{k=1}^{\kappa} Z_{t+h,k}^b \hat{\phi}_k(x')\}$. Then we can estimate $\Xi_h(x, x')$ by

$$\hat{\Xi}_h^B(x, x') = \frac{r}{B-1} \sum_{b=1}^B \left\{ \hat{C}_h^b(x, x') - \bar{C}_h^*(x, x') \right\}^2,$$

where $\bar{C}_h^*(x, x') = B^{-1} \sum_{b=1}^B \hat{C}_h^b(x, x')$.

Lastly, define the empirical quantile

$$\hat{Q}_{1-\alpha}^B \equiv (1-\alpha)\text{-th quantile of } \left\{ \sup_{x, x' \in [0,1]} \frac{n^{1/2} |\hat{C}_h^b(x, x') - \bar{C}_h^*(x, x')|}{\hat{\Xi}_h^B(x, x')^{1/2}}, b = 1, \dots, B \right\},$$

where the supremum is computed over $N \times N$ equal distance grid points on $[0, 1]^2$, with $B = 500$. The SCEs are then computed as

$$\hat{C}_h(x, x') \pm n^{-1/2} \hat{Q}_{1-\alpha}^B \hat{\Xi}_h^B(x, x')^{1/2}, \quad x, x' \in [0, 1]. \quad (4.5)$$

While the consistency of $\hat{Q}_{1-\alpha}^B \hat{\Xi}_h^B(x, x')$ as an estimate of $Q_{1-\alpha} \Xi_h(x, x')$ is conceivable, its full theoretical proof is beyond the scope of this work.

5. Simulation Studies

In this section, we examine the finite-sample performance of the SCEs. The data are generated from the following model:

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

Case 1: $m(x) = 10 + \sin\{2\pi(x - 1/2)\}$, $\varepsilon_{tj} \sim N(0, 1)$, $1 \leq t \leq n$, $1 \leq j \leq N$, $\phi_1(x) = -2 \cos\{\pi(x - 1/2)\}$, and $\phi_2(x) = \sin\{\pi(x - 1/2)\}$. $\{\xi_{tk}\}_{t=1, k=1}^{n, 2}$ are generated from (1.4), where $\{\zeta_{tk}\}_{t=1, k=1}^{n, 2}$ are i.i.d. $N(0, 1)$

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

so $\{\xi_{tk}\}_{t=1}^n$ is an MA(1) process. The number of trajectories n is taken to be 160, 400, 900, and 1600, and the number of observations per trajectory N is taken as 200, 500, 1000, and 2000, respectively, with the noise level $\sigma = 0.1$.

Case 2: $\{\xi_{tk}\}_{t=1, k=1}^{n, 2}$ are generated from (1.4), where $\{\zeta_{tk}\}_{t=1, k=1}^{n, 2}$ are i.i.d. $N(0, 1)$ variables and

$$a_{0k} = 0.3, \quad a_{1k} = 0.6, \quad a_{2k} = 0.741, \quad a_{tk} = 0, \quad \forall t \geq 3, \quad k = 1, 2,$$

so $\{\xi_{tk}\}_{t=1}^n$ is an MA(2) process; $m(x)$, ε_{tj} , $\phi_1(x)$, and $\phi_2(x)$ are the same as those in Case 1. The number of curves n is taken to be 400 and the number N of observations per curve is taken to be 500.

Case 3: $\{\xi_{tk}\}_{t=1, k=1}^{n, 2}$ are generated from (1.4), where $\{\zeta_{tk}\}_{t=1, k=1}^{n, 2}$ are i.i.d. $N(0, 1)$ variables and

$$a_{tk} = 2^{-(t+1)/2}, \quad \forall t \in \mathbb{N}, \quad k = 1, 2,$$

so $\{\xi_{tk}\}_{t=1}^n$ is an MA(∞) process; $m(x)$, $\phi_1(x)$, and $\phi_2(x)$ are the same as those in Case 1, and ε_{tj} follows a standardized t -distribution with degrees of freedom 40, $\varepsilon_{tj} \sim \sqrt{19/20}t_{40}$. The number of trajectories n is taken to be 160, 400, 900, and 1600, and the number of observations per trajectory N is taken as 200, 500, 1000, and 2000, respectively, with the noise level $\sigma = 0.1$.

Tables 1–3 display the empirical coverage frequencies out of 1,000 replications of the true surface $C_h(\cdot, \cdot)$, for $h = 0, 1, 2$, in Case 1 being covered by the SCE in (4.5) at all $N \times N$ grid points $\{(j/N, j'/N), 1 \leq j, j' \leq N\}$ on $[0, 1]^2$. Overall, the empirical coverage frequency approaches the nominal level as n and N increase, and is slightly higher than the nominal level in Table 3 because the true FAF $C_2(\cdot, \cdot) \equiv 0$. For Case 3, with genuine FMA(∞) trajectories, Tables 4–6 exhibit similar patterns.

Figure 3 depicts the FAF estimate $\hat{C}_1(\cdot, \cdot)$ and 95% SCEs for the true FAF for one replication of Case 1, with $\sigma = 0.1$, $r = \lfloor n/4 \rfloor$, and various n, N

Table 1. Coverage frequencies for $C_0(\cdot, \cdot)$ in Case 1 by SCE (4.5) with $p = 4$.

(n, N)	r	$1 - \alpha = 0.95$	$1 - \alpha = 0.99$
(160, 200)	$[\sqrt{n}]$	0.912	0.986
	$[n/4]$	0.912	0.975
(400, 500)	$[\sqrt{n}]$	0.953	0.991
	$[n/4]$	0.952	0.991
(900, 1000)	$[\sqrt{n}]$	0.946	0.992
	$[n/4]$	0.952	0.990
(1600, 2000)	$[\sqrt{n}]$	0.952	0.992
	$[n/4]$	0.956	0.992

Table 2. Coverage frequencies for $C_1(\cdot, \cdot)$ in Case 1 by SCE (4.5) with $p = 4$.

(n, N)	r	$1 - \alpha = 0.95$	$1 - \alpha = 0.99$
(160, 200)	$[\sqrt{n}]$	0.914	0.988
	$[n/4]$	0.906	0.982
(400, 500)	$[\sqrt{n}]$	0.939	0.995
	$[n/4]$	0.946	0.992
(900, 1000)	$[\sqrt{n}]$	0.954	0.994
	$[n/4]$	0.954	0.987
(1600, 2000)	$[\sqrt{n}]$	0.954	0.994
	$[n/4]$	0.950	0.994

Table 3. Coverage frequencies for $C_2(\cdot, \cdot)$ in Case 1 by SCE (4.5) with $p = 4$.

(n, N)	r	$1 - \alpha = 0.95$	$1 - \alpha = 0.99$
(160, 200)	$[\sqrt{n}]$	0.954	0.997
	$[n/4]$	0.966	0.997
(400, 500)	$[\sqrt{n}]$	0.958	0.998
	$[n/4]$	0.963	0.995
(900, 1000)	$[\sqrt{n}]$	0.952	0.994
	$[n/4]$	0.956	0.991
(1600, 2000)	$[\sqrt{n}]$	0.950	0.994
	$[n/4]$	0.963	0.995

combinations. As expected, the SCEs become narrower as the sample size n increases. Figure 4, with a different lag h , shows similar patterns to those in Figure 3, conforming our asymptotic theory.

Figure 5 shows the FACF estimate and 95% SCE of the true FACF for one replication of Case 2 with $\sigma = 0.1$ and $(n, N) = (400, 500)$, for $h = 0, 1, 2, 3, p = 4$. Here, the 95% SCE fails to cover the zero plane entirely for lag $h = 0, 1, 2$,

Table 4. Coverage frequencies for $C_0(\cdot, \cdot)$ in Case 3 by SCE (4.5) with $p = 4$.

(n, N)	r	$1 - \alpha = 0.95$	$1 - \alpha = 0.99$
(160, 200)	$[\sqrt{n}]$	0.898	0.954
	$[n/4]$	0.904	0.951
(400, 500)	$[\sqrt{n}]$	0.939	0.981
	$[n/4]$	0.942	0.977
(900, 1000)	$[\sqrt{n}]$	0.939	0.993
	$[n/4]$	0.945	0.982
(1600, 2000)	$[\sqrt{n}]$	0.953	0.995
	$[n/4]$	0.951	0.992

Table 5. Coverage frequencies for $C_1(\cdot, \cdot)$ in Case 3 by SCE (4.5) with $p = 4$.

(n, N)	r	$1 - \alpha = 0.95$	$1 - \alpha = 0.99$
(160, 200)	$[\sqrt{n}]$	0.902	0.963
	$[n/4]$	0.909	0.950
(400, 500)	$[\sqrt{n}]$	0.948	0.986
	$[n/4]$	0.941	0.982
(900, 1000)	$[\sqrt{n}]$	0.949	0.996
	$[n/4]$	0.944	0.986
(1600, 2000)	$[\sqrt{n}]$	0.951	0.995
	$[n/4]$	0.944	0.986

Table 6. Coverage frequencies for $C_2(\cdot, \cdot)$ in Case 3 by SCE (4.5) with $p = 4$.

(n, N)	r	$1 - \alpha = 0.95$	$1 - \alpha = 0.99$
(160, 200)	$[\sqrt{n}]$	0.921	0.979
	$[n/4]$	0.913	0.982
(400, 500)	$[\sqrt{n}]$	0.946	0.985
	$[n/4]$	0.942	0.982
(900, 1000)	$[\sqrt{n}]$	0.940	0.992
	$[n/4]$	0.945	0.986
(1600, 2000)	$[\sqrt{n}]$	0.943	0.993
	$[n/4]$	0.955	0.994

whereas it does for $h = 3$. When testing the null hypothesis $H_0 : C_h(x, x') \equiv 0$ for $h = 0, 1, 2, 3$, we retain the null hypothesis for $h = 3$ with p -value > 0.05 , and reject the null hypothesis for $h = 0, 1, 2$ at the significance level $\alpha = 0.05$. Thus, the $\text{FMA}(\infty)$ is actually $\text{FMA}(2)$ according to Proposition 1, which is the model of Case 2.

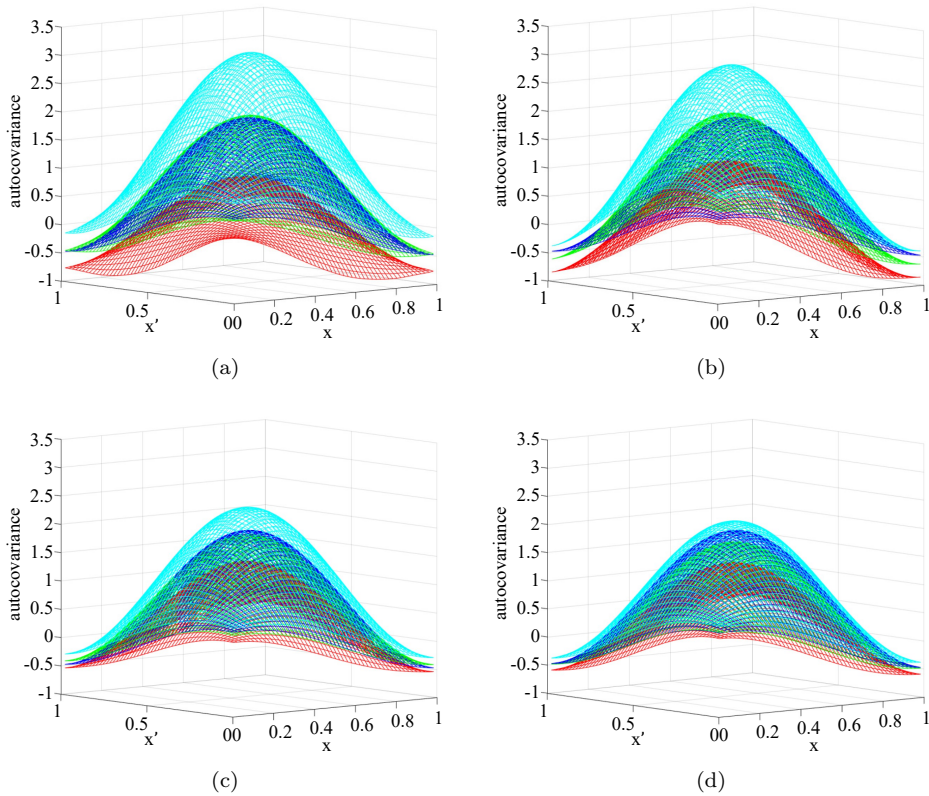


Figure 3. Plots of the true $C_1(\cdot, \cdot)$ in Case 1 and two-step estimate $\hat{C}_1(\cdot, \cdot)$ (middle surfaces), with 95% SCEs (upper and lower surfaces), $r = \lfloor n/4 \rfloor$. (a)–(d) correspond to $(n, N) = (160, 200), (400, 500), (900, 1000), (1600, 2000)$, respectively.

However, for the genuine $\text{FMA}(\infty)$ Case 3, Figure 6 shows that for one replication, the 95% SCE fails to cover the zero plane entirely for lag $h = 0, 1, 2, 3, 4$, whereas it does for $h = 5$. Then, under the same testing procedure, we can infer that the $\text{FMA}(\infty)$ is actually $\text{FMA}(4)$, according to Proposition 1. It is unclear what the asymptotics of the FMA order by Proposition 1 should look like when the true model is a genuine $\text{FMA}(\infty)$. Further research is also needed on how Proposition 1 should be applied to multiple h , simultaneously or sequentially.

6. Real-Data Analysis

In this section, we study the EEG data discussed in Section 1. The data were collected by the research group of Prof. Linhong Ji at Tsinghua University Department of Mechanical Engineering, from 142 university student participants.

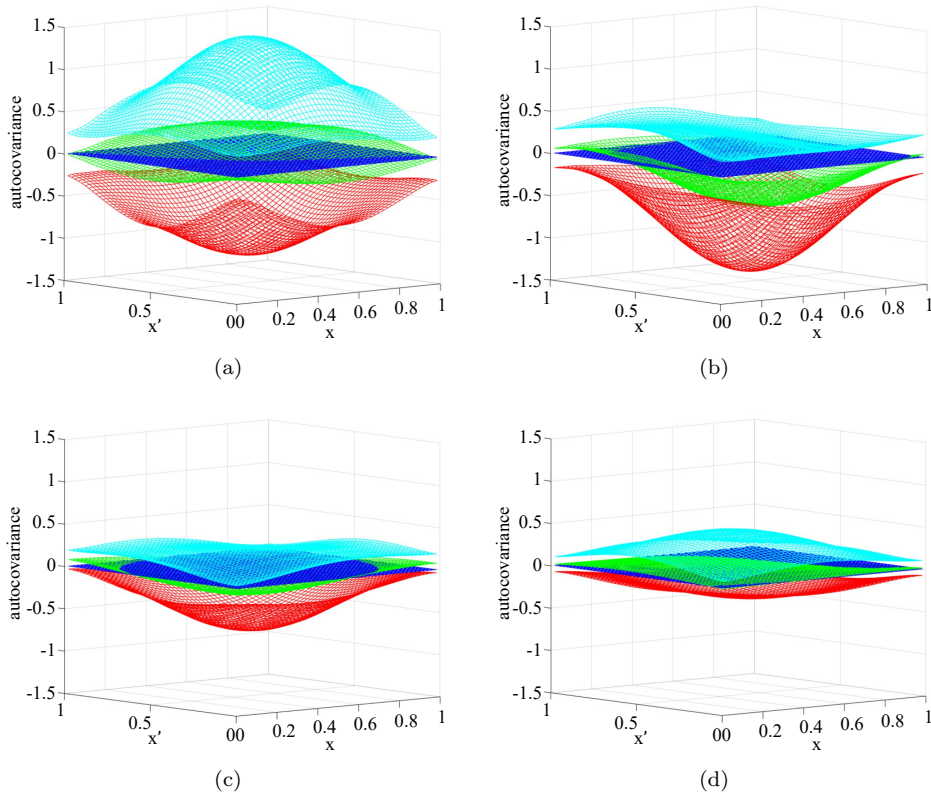


Figure 4. Plots of the true $C_2(\cdot, \cdot)$ in Case 1 and two-step estimate $\hat{C}_2(\cdot, \cdot)$ (middle surfaces), with 95% SCEs (upper and lower surfaces), $r = \lceil n/4 \rceil$. (a)–(d) correspond to $(n, N) = (160, 200), (400, 500), (900, 1000), (1600, 2000)$, respectively.

The data are recorded by electrodes placed on the scalp surface of an individual subject A at the sixth of 32 scalp locations, at a 1000 Hz sample rate. The mid-portion of 60,000 EEG recordings are used as 300 consecutive segments, each consisting of 200 EEG signals. Hence, there are $n = 300$ unobserved curves, with $N = 200$ signals recorded in each curve (see Figure 7 for the raw data). The data range is from -22 to 22. The data are available upon request.

Multiple SCEs are used to test the null hypothesis $H_0 : C_h(\cdot, \cdot) \equiv 0$, with preset lag h . For $h = 1$, the estimated FACF and 95% SCE computed by (4.5) are shown in Figure 1. Figure 2(a) shows the 95% SCEs (upper and lower surfaces) and the null hypothesis zero plane. Here, we reject the null hypothesis at the significance level $\alpha = 0.05$ because the zero plane is not entirely covered by the 95% SCE. Moreover, Figure 2(b) shows that even the 99% SCE does not contain the zero plane. Hence, we reject the null hypothesis $H_0 : C_1(\cdot, \cdot) \equiv 0$ with p -

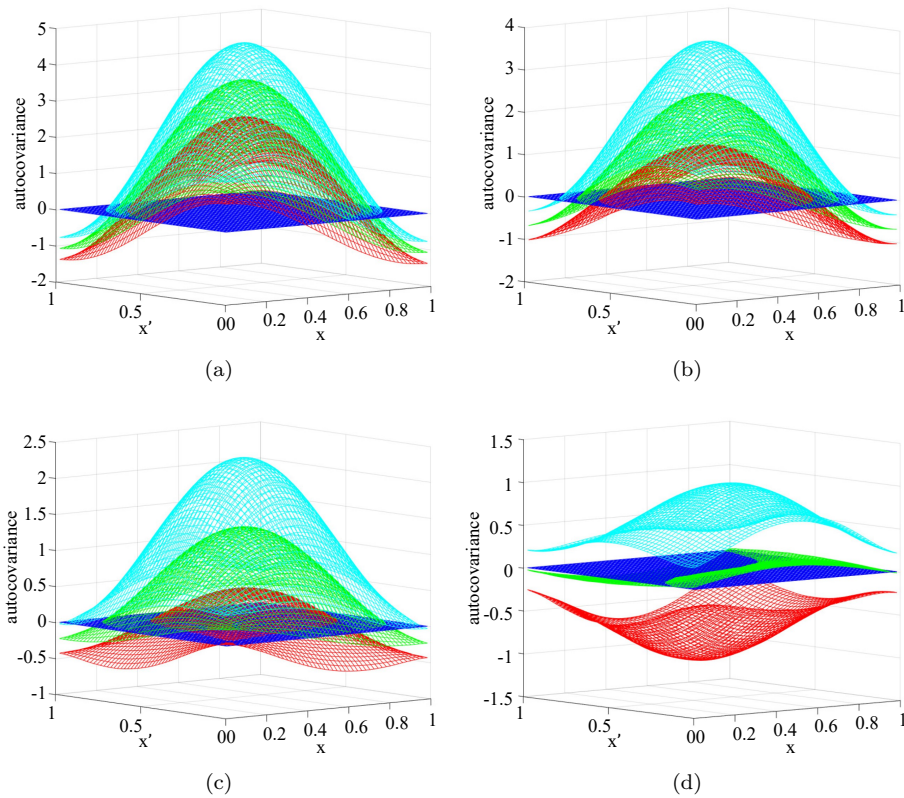


Figure 5. Plots of the true $C_h(\cdot, \cdot)$ in Case 2 and two-step estimate $\hat{C}_h(\cdot, \cdot)$ (middle surfaces), with 95% SCEs (upper and lower surfaces), $r = \lfloor n/4 \rfloor$. (a)–(d) correspond to $h = 0, 1, 2, 3$, respectively.

value < 0.01 . Similarly for $h = 2, 3, \dots, 9$, we reject the null hypothesis with p -values less than 0.05. For $h = 10$, the 95% SCE covers the null hypothesis zero plane, and the lowest confidence level of the SCE covering the entire zero plane is 0.896. Hence, we retain the null hypothesis $H_0 : C_{10}(x, x') \equiv 0$ with p -value 0.104. Similar null hypotheses are retained for $h = 11, 12, 13, 14$, with p -values = 0.150, 0.220, 0.204, 0.192, respectively. Thus, Proposition 1 points to an FMA(9) model for the data.

We also investigate the specific form of the covariance function $C_0(x, x')$ for the EEG data. Figure 8(a) depicts the estimated FACF, which strongly suggests a trigonometric form. We test the null hypothesis $H_0 : C_0(x, x') = a_0 + \sum_{k=1}^2 \{2a_k \cos(2k\pi x) \cos(2k\pi x') + 2b_k \sin(2k\pi x) \sin(2k\pi x')\}$, in which the parameters $\{a_k\}_{k=0}^2, \{b_k\}_{k=1}^2$ are estimated by linear least squares. Figure 8(a) shows that the 95% SCE (upper and lower surfaces) cover the null surface com-

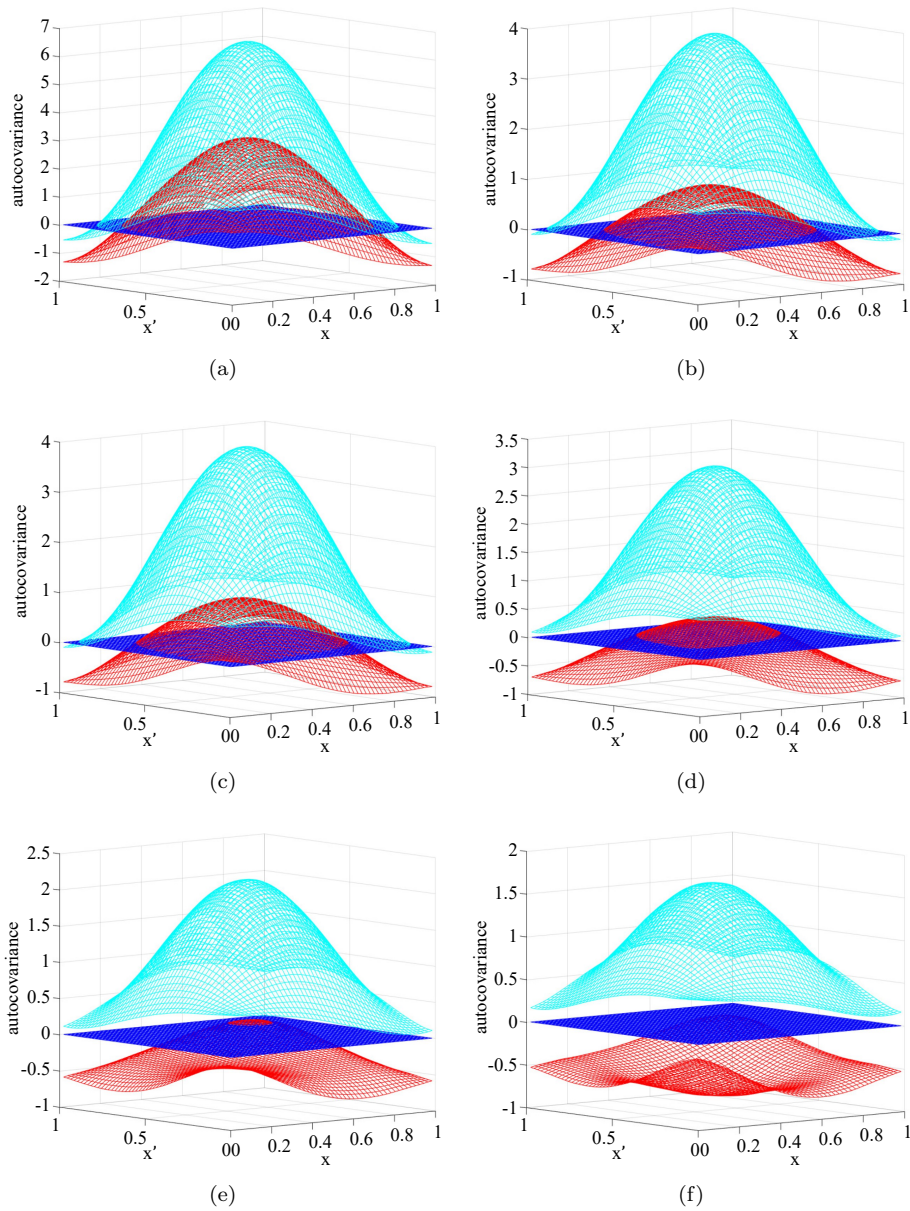


Figure 6. Plots of the true $C_h(\cdot, \cdot)$ in Case 3 and two-step estimate $\hat{C}_h(\cdot, \cdot)$ (middle surfaces), with 95% SCEs (upper and lower surfaces), $r = \lfloor n/4 \rfloor$. (a)–(f) correspond to $h = 0, 1, 2, 3, 4, 5$, respectively.

pletely, and the coefficient of determination for the two-step estimate $\hat{C}_0(x, x')$ by the trigonometric form is $R^2 = 0.9661$. The lowest confidence level by which the SCE covers the entire null surface (Figure 8(b)) is 67.4%, so we retain the

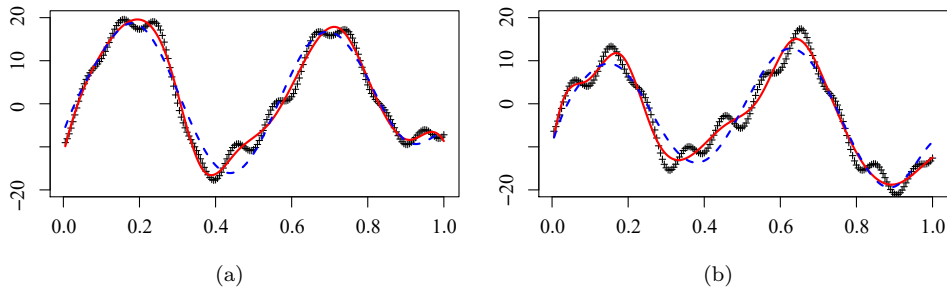


Figure 7. Randomly selected segments of raw EEG data (crosses), spline trajectories (solid), and trajectories based on trigonometric FPCs (dash).

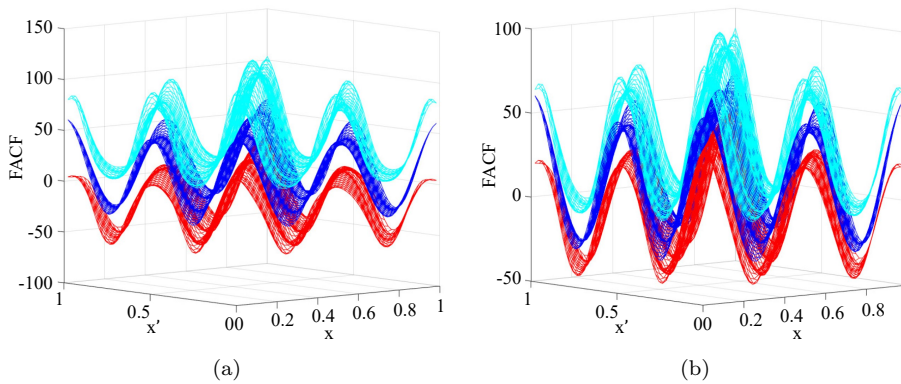


Figure 8. Null hypothesis middle surface $a_0 + \sum_{k=1}^2 \{2a_k \cos(2k\pi x) \cos(2k\pi x') + 2b_k \sin(2k\pi x) \sin(2k\pi x')\}$ and SCEs (upper and lower surfaces) of $C_0(x, x')$ of EEG data with (a) $\alpha = 0.05$, (b) $\alpha = 0.326$.

null hypothesis with p -value = 0.326.

Because all of the estimated parameters $\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2$ are significantly positive, the above null hypothesis is a Mercer's lemma expansion of the FADF with positive eigenvalues in descending order:

$$\begin{aligned} \lambda_{0,1} = \hat{b}_2 = 19.6, \quad \lambda_{0,2} = \hat{a}_2 = 17.1, \quad \lambda_{0,3} = \hat{a}_0 = 14.6, \\ \lambda_{0,4} = \hat{a}_1 = 3.5, \quad \lambda_{0,5} = \hat{b}_1 = 2.9, \end{aligned}$$

and the orthonormal eigenfunctions $\psi_k(x)$ are a Fourier basis:

$$\begin{aligned} \psi_{0,1}(x) &\equiv \sqrt{2} \cos(4\pi x), \quad \psi_{0,2}(x) \equiv \sqrt{2} \sin(4\pi x), \quad \psi_{0,3}(x) \equiv 1, \\ \psi_{0,4}(x) &\equiv \sqrt{2} \sin(2\pi x), \quad \psi_{0,5}(x) \equiv \sqrt{2} \cos(2\pi x). \end{aligned} \quad (6.1)$$

This strongly suggests a Karhunen–Loève expansion with Fourier form FPCs

$$\begin{aligned}\phi_{0,1}(x) &\equiv (2\hat{b}_2)^{1/2} \cos(4\pi x), \quad \phi_{0,2}(x) \equiv (2\hat{a}_2)^{1/2} \sin(4\pi x), \quad \phi_{0,3}(x) \equiv \hat{a}_0^{1/2}, \\ \phi_{0,4}(x) &\equiv (2\hat{a}_1)^{1/2} \sin(2\pi x), \quad \phi_{0,5}(x) \equiv (2\hat{b}_1)^{1/2} \cos(2\pi x).\end{aligned}$$

The centered trajectories $\hat{\chi}_t(\cdot)$ in (2.5) are then well approximated by

$$\sum_{k=1}^5 \hat{\xi}_{tk} \phi_{0,k}(\cdot), \quad \hat{\xi}_{tk} = \lambda_{0,k}^{-1} \int_t \hat{\chi}_t(x) \phi_{0,k}(x) dx, \quad 1 \leq k \leq 5, \quad 1 \leq t \leq 300.$$

For two randomly chosen integers t between 1 and 300, Figure 7 shows both the spline trajectory $\hat{\eta}_t(\cdot) = \hat{m}(\cdot) + \hat{\chi}_t(\cdot)$ (solid), and the trajectory with hypothesized FPCs $\{\phi_{0,k}(\cdot)\}_{k=1}^5: \hat{m}(\cdot) + \sum_{k=1}^5 \hat{\xi}_{tk} \phi_{0,k}(\cdot)$ (dash). Both appear to be faithful representations of the raw EEG data (crosses), with R^2 values (0.985, 0.959) and (0.969, 0.948), respectively, for the spline and Fourier basis trajectories and the two raw data segments. This further corroborates that for this EEG data, the Fourier FPCs in (6.1) are appropriate.

7. Conclusion

We investigate the properties of the FACF for $\text{FMA}(\infty)$, and propose a two-step tensor-product spline estimator for the FACF, which is asymptotically equivalent to an infeasible estimator at the rate of $\mathcal{O}_p(n^{-1/2})$. We establish an asymptotic SCE that can be used to test any hypothesis on the FACF, such as equality to zero. For EEG time series, strong evidence points to an FMA of finite lag and Fourier-form FPCs. The theoretically justified SCE is a powerful inference tool, and is expected to find wide application in various scientific fields.

Supplementary Material

The online Supplementary Material contains detailed proofs for the main results.

Acknowledgments

This research was supported by the National Natural Science Foundation of China awards 12171269, 12026242, and 11771240. The authors are grateful to Professor Linhong Ji and his research group for providing the EEG data, and to the two referees for their helpful comments and suggestions.

References

- 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. Springer, 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.* **12**, 503–538.
- 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.
- de Boor, C. (2001). *A Practical Guide to Splines*. Revised Edition. Springer-Verlag, New York.
- Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21**, 1735–1765.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York.
- Gu, L., Wang, L., Härdle, W. K. and Yang, L. (2014). A simultaneous confidence corridor for varying coefficient regression with sparse functional data. *TEST* **23**, 806–843.
- Gu, L., Wang, S. and Yang, L. (2021). Smooth simultaneous confidence band for the error distribution function in nonparametric regression. *Comput. Stat. Data. Anal* **155**, 107106.
- Gu, L. and Yang, L. (2015). Oracally efficient estimation for single-index link function with simultaneous confidence band. *Electron. J. Stat* **9**, 1540–1561.
- Guo, J. and Chen, Y. (2019). An L^2 -norm based ANOVA test for the equality of weakly dependent functional time series. *Stat. Interface* **12**, 167–180.
- Guo, J., Zhou, B. and Zhang, J. (2019). New tests for equality of several covariance functions for functional data. *J. Am. Stat. Assoc* **114**, 1251–1263.
- Hall, P., Fisher I. N. and Hoffmann, B. (1994). On the nonparametric estimation of covariance function. *Ann. Statist* **22**, 2115–2334.
- Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *J. Roy. Stat. Soc. B* **75**, 103–122.
- Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, Chichester.
- Li, J. and Yang, L. (2021). Statistical inference for functional time series. *Statist. Sinica* **33**, 519–549.
- Ma, S., Yang, L. and Carroll, R. J. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica* **22**, 95–122.
- Meyer, K. (1998). Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel.* **30**, 221–240.
- Ramsay, J. O. and Sliverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.
- Ramsay, J. O. and Sliverman, B. W. (2005). *Functional Data Analysis*. 2nd Edition. Springer, New York.

- Telschow F. J. E. and Schwartzman A. (2022). Simultaneous confidence bands for functional data using the Gaussian Kinematic formula. *J. Statist. Plann. Inference* **216**, 70–94.
- Wang, J. (2012). Modelling time trend via spline confidence band. *Ann. Inst. Statist. Math.* **64**, 275–301.
- Wang, J., Cao, G., Wang, L. and Yang, L. (2020a). Simultaneous confidence band for stationary covariance function of dense functional data. *J. Multivariate Anal.* **176**, 104584.
- Wang, J., Liu, R., Cheng, F. and Yang, L. (2014). Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. *Ann. Statist.* **42**, 654–668.
- Wang, Y., Wang, G., Wang, L. and Ogden, T. (2020b). 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, G., Wang, L. and Yang, L. (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statist. Sinica* **31**, 1463–1487.
- Zheng, S., Liu, R., Yang, L. and Härdle, W. K. (2016). Statistical inference for generalized additive models: Simultaneous confidence corridors and variable selection. *TEST* **25**, 607–626.
- Zheng, S., Yang, L. and Härdle, W. K. (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *J. Am. Stat. Assoc.* **109**, 661–673.
- Zhong, C. and Yang, L. (2021). Simultaneous confidence bands for comparing variance functions of two samples based on deterministic designs. *Comput. Stat.* **36**, 1197–1218.
- Zhou, L., Lin, H. and Liang, H. (2018). Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *J. Am. Stat. Assoc.* **113**, 1550–1564.

Chen Zhong

School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China.

E-mail: zhchzhlm@163.com

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 February 2022)