Statistica Sinica Preprint No: SS-2023-0146	
Title	Statistical Inference for Multivariate Functional Panel Data
Manuscript ID	SS-2023-0146
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0146
Complete List of Authors	Shuang Sun and Lijian Yang
Corresponding Authors	Lijian Yang
E-mails	yanglijian@tsinghua.edu.cn
Notice: Accepted version subject to English editing.	

STATISTICAL INFERENCE FOR MULTIVARIATE FUNCTIONAL PANEL DATA

Shuang Sun and Lijian Yang

Tsinghua University

Abstract: Statistical inference is developed for vector-valued functional panel data which are i.i.d. with respect to subjects and infinite moving average in time. B-spline estimation is proposed for trajectories, which are used to construct a two-step estimator of the vector mean function. By using explicit Gaussian strong approximation in vector form, in the context of moving average panel, the proposed spline estimator is shown to be oracally efficient in the sense that it is asymptotically equivalent to the infeasible estimator with all trajectories known. This deep theoretical result points to a limiting Gaussian distribution of the vector mean estimator, which allows for the construction of various simultaneous confidence region (SCR) for the vector mean function itself and linear combination of its elements. Asymptotic correctness of the SCRs is both established in theory and validated in simulation experiments. The proposed SCRs are applied to an Electroencephalogram (EEG) multivariate functional panel data set, validating multiple scientific facts.

Key words and phrases: B-spline, simultaneous confidence region, ElectroEn-

cephalogram, moving average, oracle efficiency.

1. Introduction

Functional data analysis (FDA) has been an important area of statistics research for more than two decades. Comprehensive introduction to FDA can be found in Ramsay and Silverman (2005), Ferraty and Vieu (2006), Hsing and Eubank (2015) and Kokoszka and Reimherr (2017).

A functional random variable is a square-integrable continuous stochastic process: specifically, $\eta(\cdot) \in \mathcal{C}[0,1]$ almost surely, with $\mathbb{E}\sup_{x\in[0,1]}\eta^2(x) < \infty$. For such $\eta(\cdot)$, both mean function $\mathbb{E}\eta(\cdot)$ and covariance function $\mathrm{cov}\{\eta(x),\eta(x')\}$, $x, x' \in [0,1]$ exist and are continuous. A functional random vector is a vector-valued functional random variable $\boldsymbol{\eta}(\cdot) = \{\eta^{(1)}(\cdot),\ldots,\eta^{(L)}(\cdot)\}^\top$ where each element is a square-integrable continuous stochastic process. A functional data set in the abstract sense consists of repeated observations $\{\eta_i(\cdot)\}_{i=1}^n$ of a functional random variable $\eta(\cdot)$ or $\{\boldsymbol{\eta}_i(\cdot)\}_{i=1}^n$ of a functional random vector $\boldsymbol{\eta}(\cdot)$.

As the essential first step in functional data analysis, estimation of the population mean function $\mathbb{E}\eta(\cdot)$ was studied in Ma et al. (2012), Zheng et al. (2014) for sparse longitudinal data, and Cao et al. (2012), Cao and Wang (2018), Cai et al. (2020), Yu et al. (2021), Huang et al. (2022) for

dense functional data, all with simultaneous confidence band (SCB). All of these works are valid only for i.i.d. observations $\eta_i(\cdot), 1 \leq i \leq n$. More recently, Li and Yang (2023) has extended SCB methodology for functional mean to functional time series $\eta_t(\cdot), 1 \leq t \leq T$, while Zhong and Yang (2023) has established simultaneous confidence region for auto covariance function cov $\{\eta_t(x), \eta_{t+h}(x')\}$, $x, x' \in [0,1], h \in \mathbb{N}_+$ of functional time series. See also Horváth et al. (2013) for related work on functional time series.

In contrast to i.i.d. functional observations or functional time series, functional panel data combines both in a realistic and informative manner. A functional panel data consists of stochastic processes $\{\boldsymbol{\eta}_{it}(\cdot)\}_{i=1,t=1}^{n,T}$ called trajectories, where for each fixed $i \in \{1,\ldots,n\}$, $\{\boldsymbol{\eta}_{it}(\cdot)\}_{t=1}^{T}$ is a functional time series of length T, and each $\{\boldsymbol{\eta}_{it}(\cdot)\}_{t=1}^{T}, 1 \leq i \leq n$ has the same distribution of a standard functional time series $\{\boldsymbol{\eta}_t(\cdot)\}_{t=1}^{T}$. Within the functional time series $\{\boldsymbol{\eta}_t(\cdot)\}_{t=1}^{T}$, each $\boldsymbol{\eta}_t(\cdot), t \in 1,\ldots,T$ has the same distribution as $\boldsymbol{\eta}(\cdot)$. Instead of functional random variables from $\mathcal{C}[0,1]$, functional random vectors $\boldsymbol{\eta}(\cdot)$ taking values from $(\mathcal{C}[0,1])^L$ are investigated in this paper to accommodate the multivariate circumstances. Those multivariate trajectories are decomposed as $\boldsymbol{\eta}_{it}(\cdot) = \boldsymbol{m}(\cdot) + \boldsymbol{\xi}_{it}(\cdot)$, where $\boldsymbol{m}(\cdot) = \left\{m^{(1)}(\cdot),\ldots,m^{(L)}(\cdot)\right\}^\top$ is the vector mean function of $\boldsymbol{\eta}(\cdot)$, and centered trajectories $\boldsymbol{\xi}_{it}(\cdot) = \left\{\xi_{it}^{(1)}(\cdot),\ldots,\xi_{it}^{(L)}(\cdot)\right\}^\top$ are small-scale variations

of x on the t-th trajectory of the i-th subject, being $(\mathcal{C}[0,1])^L$ random variables with mean $\mathbb{E}\boldsymbol{\xi}_{it}(\cdot) = \mathbf{0}_L$ and matrix covariance function $\mathbb{E}\boldsymbol{\xi}_{it}(\cdot)\boldsymbol{\xi}_{it}(\cdot)^\top$ of size $L \times L$.

An example which motivates investigation of multivariate functional panel data is the Electroencephalogram (EEG) signals for human subjects during the task of surround suppression paradigm, studied in Langer et al. (2017). Recorded at sample rate 500Hz (i.e., one recording per 0.002 second), the multichannel EEG series of $n = 126$ participants during $T = 63$ consecutive trials constitute multivariate functional panel data, $\{\boldsymbol{\eta}_{it}(\cdot)\}_{i=1,t=1}^{126,63}$, see Figure 1. For more details of this example, see Section 5, and for EEG data as multivariate functional data, see Zhang et al. (2020).

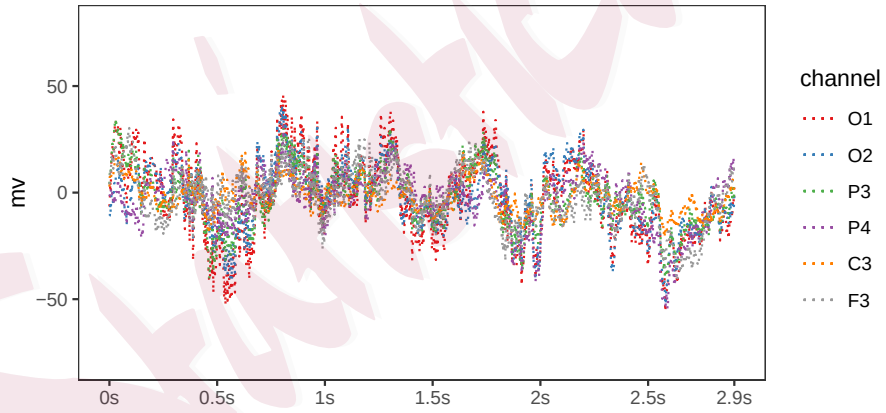


Figure 1: Plot of raw EEG signals on channels O_1, O_2, P_3, P_4, C_3, F_3 for one subject during one trial, which can be regarded as a sample of multivariate functional panel data.

To model the distribution of the vector-valued functional time series $\{\boldsymbol{\eta}_{it}(\cdot)\}_{t=1}^{T}$, the centered variations $\boldsymbol{\xi}_{it}(\cdot)$ are embedded into a strictly stationary functional vector infinite moving average FVMA(∞),

$$\boldsymbol{\xi}_{it}(\cdot) = \sum_{t'=0}^{\infty} \boldsymbol{A}_{t'} \boldsymbol{\zeta}_{i,t-t'}(\cdot),\ t = 0, \pm 1, \pm 2, \ldots, 1 \leq i \leq n, \tag{1.1}$$

where the operator $\boldsymbol{A}_{t'} : \{\mathcal{L}^2[0,1]\}^L \to \{\mathcal{L}^2[0,1]\}^L$ are bounded linear operators playing the roles of coefficient matrices in vector moving average models, see Lütkepohl (2005). The sequence of functional random vectors $\{\boldsymbol{\zeta}_{it}(\cdot)\}_{t\in\mathbb{Z}}$ with $\boldsymbol{\zeta}_{it}(\cdot)^\top = \left\{\zeta_{it}^{(1)}(\cdot), \ldots, \zeta_{it}^{(L)}(\cdot)\right\}$ are multivariate analogs of strong functional white noises in Definition 3.1 of Bosq (2000): they are functional random vectors of which each element is an $\mathcal{L}^2$ continuous process, i.i.d. over index $t \in \mathbb{Z}$, with vector mean function $\mathbb{E}\boldsymbol{\zeta}_{it}(\cdot) \equiv \mathbf{0}_L$ and matrix covariance function $\boldsymbol{G}_\zeta(x,x') = \left\{G_\zeta^{(l,l')}(x,x')\right\}_{l,l'=1}^{L} = \mathbb{E}\boldsymbol{\zeta}_{it}(x)\boldsymbol{\zeta}_{it}(x')^\top$. Model (1.1) extends the FMA(∞) proposed in Li and Yang (2023).

For the univariate functional random variable $\zeta_{it}^{(l)}(\cdot)$, $1 \leq l \leq L$, the l-th element of $\boldsymbol{\zeta}_{it}(\cdot)$, its covariance function $G_\zeta^{(l,l)}(x,x') = \mathbb{E}\zeta_{it}^{(l)}(x)\zeta_{it}^{(l)}(x')$ is continuous, and Mercer Lemma (Lemma 1.3, Bosq (2000)) entails that $G_\zeta^{(l,l)}(x,x') \equiv \sum_{k=1}^{\infty} \lambda_k^{(l)} \psi_k^{(l)}(x)\, \psi_k^{(l)}(x')$ with eigen values $\lambda_1^{(l)} \geq \lambda_2^{(l)} \geq \ldots \geq 0$ and corresponding eigenfunctions $\left\{\psi_k^{(l)}\right\}_{k=1}^{\infty}$, the latter an orthonormal ba-

sis of $\mathcal{L}^2[0,1]$, such that $\sum_{k=1}^{\infty}\lambda_k^{(l)}<\infty$, $\left\{\psi_k^{(l)}\right\}_{k=1}^{\infty}\subset\mathcal{C}[0,1]$ and $\int G_{\zeta}^{(l,l)}(x,x')$ $\psi_k^{(l)}(x')dx'=\lambda_k^{(l)}\psi_k^{(l)}(x)$. The well-known Karhunen-Loève expansion follows: $\zeta_{it}^{(l)}(\cdot)=\sum_{k=1}^{\infty}\zeta_{itk}^{(l)}\phi_k^{(l)}(\cdot)$, in which the rescaled eigenfunctions, $\phi_k^{(l)}$, called functional principle components (FPCs), satisfy $\phi_k^{(l)}=\left(\lambda_k^{(l)}\right)^{1/2}\psi_k^{(l)}$ and $\int\zeta_{it}^{(l)}(x)\phi_k^{(l)}(x)dx=\lambda_k^{(l)}\zeta_{itk}^{(l)}$, for $k\geq 1$; the random coefficients $\zeta_{itk}^{(l)}$'s, called FPC scores, are therefore uncorrelated over $k\in\mathbb{N}_+$ with mean 0 and variance 1. It is assumed that $\sum_{k=1}^{\infty}\|\phi_k^{(l)}\|_{\infty}<\infty$ (see Assumption (A4)), thus the Karhunen-Loève series converges absolutely uniformly almost surely by Dominated Convergence Theorem. Denote

$$\boldsymbol{\psi}_k=\left(\psi_k^{(1)},\ldots,\psi_k^{(L)}\right)^{\top},\ \boldsymbol{\phi}_k=\left(\phi_k^{(1)},\ldots,\phi_k^{(L)}\right)^{\top},$$
$$\boldsymbol{\zeta}_{itk}=\left(\zeta_{itk}^{(1)},\ldots,\zeta_{itk}^{(L)}\right)^{\top},\ \boldsymbol{\lambda}_k=\left(\lambda_k^{(1)},\ldots,\lambda_k^{(L)}\right)^{\top},$$

where $\psi_k^{(l)}$, $\phi_k^{(l)}$, $\zeta_{itk}^{(l)}$ and $\lambda_k^{(l)}$ are respectively the k-th eigenfunction, FPC, FPC score and eigenvalue for $\zeta_{it}^{(l)}(\cdot)$, the l-th element of $\boldsymbol{\zeta}_{it}(\cdot)$. Denote by "$\circ$" the element-wise product of two matrices of the same dimension, specifically $\{c_{ll'}\}_{l,l'=1}^{L,L'}\circ\{d_{ll'}\}_{l,l'=1}^{L,L'}=\{c_{ll'}d_{ll'}\}_{l,l'=1}^{L,L'}$, then the matrix covariance function $\boldsymbol{G}_{\zeta}(x,x')$ and the functional random vector $\boldsymbol{\zeta}_{it}(\cdot)$ are written as follows:

$$\boldsymbol{G}_{\zeta}(x,x')=\sum_{k=1}^{\infty}\boldsymbol{\Sigma}_{k,\zeta}\circ\left\{\boldsymbol{\phi}_k(x)\,\boldsymbol{\phi}_k(x')^{\top}\right\},\tag{1.2}$$

$$\boldsymbol{\zeta}_{it}(\cdot) = \sum_{k=1}^{\infty} \boldsymbol{\zeta}_{itk} \circ \boldsymbol{\phi}_k(\cdot), \tag{1.3}$$

where $\boldsymbol{\Sigma}_{k,\zeta} = \operatorname{cov}(\boldsymbol{\zeta}_{itk})$ is in fact a correlation matrix. Equations (1.2) and (1.3) are vector analogs for functional random vector $\boldsymbol{\zeta}_{it}(\cdot)$ of Mercer Lemma and the Karhunen-Loève expansion, with $\{\boldsymbol{\phi}_k(\cdot)\}_{k\in\mathbb{N}_+}$ called FPCs and random coefficients $\{\boldsymbol{\zeta}_{itk}\}_{k\in\mathbb{N}_+}$ called FPC scores for $\boldsymbol{\zeta}_{it}(\cdot)$, all vectors of dimension L.

As in Zhong and Yang (2023), elements of the FVMA(∞) coefficient operators $\boldsymbol{A}_t$ are defined relative to the orthonormal basis, then for any $t \in \mathbb{N}$, the operator $\boldsymbol{A}_t$ is defined via

$$\begin{aligned}
&\boldsymbol{A}_t \left\{ \sum_{k=1}^{\infty} \boldsymbol{c}_k \circ \boldsymbol{\psi}_k(\cdot) \right\} = \sum_{k=1}^{\infty} \boldsymbol{a}_{tk} \circ \boldsymbol{c_k} \circ \boldsymbol{\psi}_k(\cdot),\ \boldsymbol{c}_k \in \mathbb{R}^L,\ \sum_{k=1}^{\infty} \|\boldsymbol{c}_k\|_2^2 < \infty, \\
&\boldsymbol{a}_{tk} \in \mathbb{R}^L, \|\boldsymbol{a}_{tk}\|_\infty < C_a \rho_a^t,\ C_a \in (0,\infty),\ \rho_a \in (0,1), k \in \mathbb{N}_+,
\end{aligned} \tag{1.4}$$

where for any vector $\boldsymbol{c} = (c_1, \ldots, c_L)^\top \in \mathbb{R}^L$, $\|\boldsymbol{c}\|_r = (|c_1|^r + \ldots + |c_L|^r)^{1/r}$, $1 \le r < \infty$, $\|\boldsymbol{c}\|_\infty = \max_{1\le l\le L}(|c_1|, \ldots, |c_L|)$. The constraints in (1.4) on moving average coefficients $\boldsymbol{a}_{tk}$ are vector versions of constraints in Zhong and Yang (2023).

According to (1.1),

$$\boldsymbol{\xi}_{it}(\cdot)=\sum_{t'=0}^{\infty}\boldsymbol{A}_{t'}\left\{\sum_{k=1}^{\infty}\boldsymbol{\zeta}_{i,t-t',k}\circ\boldsymbol{\phi}_k(\cdot)\right\}=\sum_{k=1}^{\infty}\boldsymbol{\xi}_{itk}\circ\boldsymbol{\phi}_k(\cdot),$$

where

$$\boldsymbol{\xi}_{itk}=\sum_{t'=0}^{\infty}\boldsymbol{a}_{t'k}\circ\boldsymbol{\zeta}_{i,t-t',k},\ t\in\mathbb{Z},\ k\in\mathbb{N}_+. \tag{1.5}$$

Thus for each fixed $i\in\{1,\ldots,n\}, k\in\mathbb{N}_+$, the vector time series $\{\boldsymbol{\xi}_{itk}\}_{t\in\mathbb{Z}}$ with $\boldsymbol{\xi}_{itk}^{\top}=\left(\xi_{itk}^{(1)},\ldots,\xi_{itk}^{(L)}\right)$ is an L-dimensional VMA(∞) series of $\{\boldsymbol{\zeta}_{itk}\}_{t\in\mathbb{Z}}$. These VMA($\infty$) series are uncorrelated over $k\in\mathbb{N}_+$ and i.i.d. over $i\in\{1,\ldots,n\}$. One notes that VMA(∞) is a sufficiently broad class which includes the most common VARMA(p,q), see Lütkepohl (2005).

Assume for convenience that for each fixed $k\in\mathbb{N}_+$, $\sum_{t=0}^{\infty}\boldsymbol{a}_{tk}\circ\boldsymbol{a}_{tk}=\mathbf{1}_L$, then each element of $\boldsymbol{\xi}_{itk}$ are of mean 0 and variance 1 as well. Since $\{\boldsymbol{\xi}_{itk}\}_{k\in\mathbb{N}_+}$ are uncorrelated over $k\in\mathbb{N}_+$, $\boldsymbol{\xi}_{it}(\cdot)=\sum_{k=1}^{\infty}\boldsymbol{\xi}_{itk}\circ\boldsymbol{\phi}_k(\cdot)$ is also the Karhunen-Loève expansion of functional random vector $\boldsymbol{\xi}_{it}(\cdot)$, and random coefficients $\{\boldsymbol{\xi}_{itk}\}_{k\in\mathbb{N}_+}$ are the corresponding FPC scores. The matrix covariance function $\boldsymbol{G}(x,x')=\left\{G^{(l,l')}(x,x')\right\}_{1\le l,l'\le L}$ of $\boldsymbol{\xi}_{it}(\cdot)$ has the following expression similar to Mercer Lemma

$$\boldsymbol{G}(x,x')=\sum_{k=1}^{\infty}\boldsymbol{\Sigma}_{k,\xi}\circ\left\{\boldsymbol{\phi}_k(x)\,\boldsymbol{\phi}_k(x')^{\top}\right\}, \tag{1.6}$$

where $\boldsymbol{\Sigma}_{k,\xi} = \mathrm{cov}\left(\boldsymbol{\xi}_{itk}\right) = \boldsymbol{\Sigma}_{k,\zeta} \circ \sum_{t=0}^{\infty} \boldsymbol{a}_{tk}\boldsymbol{a}_{tk}^{\top}$ is actually a correlation matrix, $1 \leq k < \infty$. When $L = 1$, $\boldsymbol{\Sigma}_{k,\xi} \equiv \boldsymbol{\Sigma}_{k,\zeta} \equiv 1$.

In practice, however, trajectories $\{\boldsymbol{\eta}_{it}(\cdot)\}$ are unknown, and the observed multivariate functional panel data take the form of

$$\boldsymbol{Y}_{itj} = \boldsymbol{m}(j/N) + \sum_{k=1}^{\infty} \boldsymbol{\xi}_{itk} \circ \boldsymbol{\phi}_k(j/N) + \boldsymbol{\sigma}_{it}(j/N) \circ \boldsymbol{\varepsilon}_{itj}, \tag{1.7}$$

for $1 \leq i \leq n$, $1 \leq t \leq T$, $1 \leq j \leq N$, with $\boldsymbol{Y}_{itj} = \left(Y_{itj}^{(1)}, \ldots, Y_{itj}^{(L)}\right)^{\top}$, $\boldsymbol{\varepsilon}_{itj} = \left(\varepsilon_{itj}^{(1)}, \ldots, \varepsilon_{itj}^{(L)}\right)^{\top}$, $\boldsymbol{\sigma}_{it} = \left(\sigma_{it}^{(1)}, \ldots, \sigma_{it}^{(L)}\right)^{\top}$, all vectors of L-dimension. The terms $\boldsymbol{\sigma}_{it}(j/N) \circ \boldsymbol{\varepsilon}_{itj}$ represent measurement errors which occur with data collection, $\boldsymbol{\varepsilon}_{itj}$ independent over $i \geq 1$, $t \geq 1$ and i.i.d. over $j \geq 1$, with covariance matrix $\boldsymbol{\Sigma}_{it,\varepsilon}$, which is in fact a correlation matrix after normalization, and elements of the vector standard deviation function $\boldsymbol{\sigma}_{it}(\cdot)$ satisfying Hölder continuity in Assumption (A2). B-spline is used to approximate the vector-valued trajectories $\{\boldsymbol{\eta}_{it}(\cdot)\}$ from observed vectors $\{\boldsymbol{Y}_{itj}\}$, and further a two-step B-spline estimator for the vector mean function $\boldsymbol{m}(\cdot)$ is built.

Unlike existing works Huang et al. (2022), Li and Yang (2023) and Zhong and Yang (2023) using classic implicit Gaussian strong approximation of the kind in Einmahl (1989), this work uses explicit Gaussian strong approximation result of Götze and Zaitsev (2010). As a consequence, fi-

nite number of distinct distributions is no longer required of FPC scores or measurement errors (see Lemma S.1 in the Supplement), significantly broadening the applicability scope of the proposed method. Another noteworthy feature of the asymptotic results is that only length T of the time series needs to go to infinity, the number n of subjects can either be bounded or go to infinity, see also Remark 4.

The paper is organized as follows. Section 2 proposes SCR for the vector mean function and SCB for any linear combination of its elements built from B-spline estimation, and establishes theoretical properties of them. Implementation details of the proposed SCB and SCR are given in Section 3. Section 4 examines finite sample performance of the proposed methods in simulation settings. An EEG data is studied in Section 5 with the new SCB and SCR tools. All technical proofs are collected in the online Supplement.

2. Main results

2.1 Spline estimation of vector mean function

For simplicity of notations, one defines a bijective single rank over i and t, $r\,(i,t) = n(t-1)+i$, for $1 \le i \le n$ and $-\infty \le t \le T$, so that any double summation $\sum_{i=1}^{n}\sum_{t=1}^{T}$ is converted to $\sum_{r(i,t)=1}^{nT}$.

For sequence of real numbers a_n and b_n, denote $a_n \asymp b_n$ if $a_n = \mathcal{O}(b_n)$

and $b_n = \mathcal{O}(a_n)$, as $n \to \infty$. For any sequence $\boldsymbol{c} = (c_n)_{n\in\mathbb{N}_+} \in \ell^p$, denote the norm $\|\boldsymbol{c}\|_p = (\sum_{n\in\mathbb{N}_+} |c_n|^p)^{1/p}$. For any positive definite matrix $\boldsymbol{H}$, denote by $\lambda_{\max}(\boldsymbol{H})$ and $\lambda_{\min}(\boldsymbol{H})$ the maximal and minimal eigenvalues of $\boldsymbol{H}$. For any function $f(\cdot) \in \mathcal{C}[0,1]$, denote the norm $\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$. For functions $f, g \in \mathcal{L}^2[0,1]$, denote the inner product $\langle f, g\rangle = \int_{[0,1]} f(x)g(x)dx$ with norm $\|f\|_2 = \{\langle f, f\rangle\}^{1/2}$. 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 μ-Hölder continuous q-th derivative, i.e.,

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

Besides, for vector-valued function $\boldsymbol{f} = \left(f^{(1)}, \ldots, f^{(L)}\right)^\top \in (\mathcal{C}[0,1])^L$, denote

$$\|\boldsymbol{f}\|_\infty = \max_{1\le l\le L} \left\|f^{(l)}\right\|_\infty, \quad \|\boldsymbol{f}\|_{q,\mu} = \max_{1\le l\le L} \left\|f^{(l)}\right\|_{q,\mu}, \tag{2.8}$$

and for vector-valued functions $\boldsymbol{f}, \boldsymbol{g} \in (\mathcal{L}^2[0,1])^L$ with $\boldsymbol{f} = \left(f^{(1)}, \ldots, f^{(L)}\right)^\top$, $\boldsymbol{g} = \left(g^{(1)}, \ldots, g^{(L)}\right)^\top$, denote $\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \sum_{l=1}^L \langle f^{(l)}, g^{(l)}\rangle$, $\|\boldsymbol{f}\|_2 = \{\langle \boldsymbol{f}, \boldsymbol{f}\rangle\}^{1/2}$.

Since $\boldsymbol{m}(\cdot)$ and $\boldsymbol{\phi}_k(\cdot)$ both belong to $\left(\mathcal{C}^{(q,\mu)}[0,1]\right)^L$ under Assumptions (A1) and (A4) below, $\boldsymbol{\eta}_{it}(\cdot)$ can be regarded as a functional random vector taking values from $\left(\mathcal{C}^{(q,\mu)}[0,1]\right)^L$. Had trajectories $\{\boldsymbol{\eta}_{it}(\cdot)\}_{r(i,t)=1}^{nT}$ been all observed over the entire interval $[0,1]$, the population vector mean function

$\boldsymbol{m}(\cdot)$ can be estimated by the sample mean

$$\bar{\boldsymbol{m}}(\cdot) = (nT)^{-1} \sum_{r(i,t)=1}^{nT} \boldsymbol{\eta}_{it}(\cdot). \tag{2.9}$$

This "estimator" is infeasible as it makes use of unobservables. It however serves as a benchmark.

To describe the spline functions, one denotes by $\{t_\ell\}_{\ell=1}^{J_s}$ a sequence of equally-spaced points, $t_\ell = \ell/(J_s+1)$, $0 \le \ell \le J_s+1$, $0 = t_0 < t_1 < \ldots < t_{J_s} < 1 = t_{J_s+1}$, called interior knots, which divide the interval $[0,1]$ into (J_s+1) equal subintervals $I_\ell = [t_\ell, t_{\ell+1})$, $\ell = 0, \ldots, J_s - 1$ and $I_{J_s} = [t_{J_s}, 1]$. Let $\mathcal{H}^{(p-2)} = \mathcal{H}^{(p-2)}[0,1]$ be the polynomial spline space of order p on I_ℓ, $\ell = 0, \ldots, J_s$, which consists of all $(p-2)$ times continuously differentiable functions on $[0,1]$ that are polynomials of degree $(p-1)$ on subintervals I_ℓ, $\ell = 0, \ldots, J_s$. Then, we denote by $\{B_{\ell,p}(\cdot), 1 \le \ell \le J_s + p\}$ the p-th order B-spline basis functions of $\mathcal{H}^{(p-2)}$ (de Boor (2001)), hence $\mathcal{H}^{(p-2)} = \left\{ \sum_{\ell=1}^{J_s+p} \lambda_{\ell,p} B_{\ell,p}(\cdot) \middle| \lambda_{\ell,p} \in \mathbb{R} \right\}$.

Trajectories and their mean are estimated by spline regression

$$\hat{\boldsymbol{m}}(\cdot) = (nT)^{-1} \sum_{r(i,t)=1}^{nT} \hat{\boldsymbol{\eta}}_{it}(\cdot), \ \hat{\boldsymbol{\eta}}_{it}(\cdot) = \underset{\boldsymbol{g}(\cdot) \in \left(\mathcal{H}^{(p-2)}\right)^L}{\arg\min} \sum_{j=1}^{N} \|\boldsymbol{Y}_{itj} - \boldsymbol{g}(j/N)\|_2^2. \tag{2.10}$$

2.2 Assumptions

We first list in order some constraints on constants

$$\mu \in (0,1],\ q \in \mathbb{N},\ p^* = q + \mu,\ \nu \in (0,1], \tag{2.11}$$

$$\theta \in \left(0, \min\left\{2\nu, \frac{2p^*}{p^*+1}\right\}\right), \tag{2.12}$$

$$\beta_2 \in \left(0, \min\left\{\frac{1}{2}, \nu - \frac{\theta}{2}, 1 - \frac{\theta}{2} - \frac{\theta}{2p^*}\right\}\right), \tag{2.13}$$

$$r_0 > \max\left\{4, \frac{4\theta}{p^*(2-2\beta_2-\theta)-\theta}\right\}, \tag{2.14}$$

$$\max\left\{1-\nu, \frac{\theta}{p^*}\left(\frac{1}{2}+\frac{2}{r_0}\right)\right\} < \gamma < 1 - \frac{\theta}{2} - \beta_2. \tag{2.15}$$

Elementary algebra shows that (2.12) is needed for (2.13) to be solvable for β_2, (2.12) and (2.13) for (2.14) to be solvable for r_0, and that (2.12), (2.13) and (2.14) together ensure the existence of γ that satisfies (2.15).

The above constraints allow the following technical assumptions.

(A1) The vector mean function $\boldsymbol{m}(\cdot) \in \left(\mathcal{C}^{(q,\mu)}[0,1]\right)^L$ for the integer q, constant μ and $p^* = q + \mu$ in (2.11).

(A2) The vector standard deviation functions $\boldsymbol{\sigma}_{it}(\cdot) \in \left(\mathcal{C}^{(0,\nu)}[0,1]\right)^L$ for ν in (2.11), and $\max_{1\le r(i,t)\le nT}\left(\|\boldsymbol{\sigma}_{it}\|_\infty + \|\boldsymbol{\sigma}_{it}\|_{0,\nu}\right) \le C_\sigma$ for $0 < C_\sigma < \infty$.

(A3) As $T \to \infty$, $N = N(T) \to \infty$, $n \times T = \mathcal{O}\left(N^\theta\right)$ for the θ in (2.12).

(A4) There exists $c_\varphi > 0$ such that $\boldsymbol{G}_\varphi(x,x) \geq c_\varphi \boldsymbol{I}_L$, $\forall x \in [0,1]$ with $\boldsymbol{G}_\varphi(x,x)$ defined in (2.16). The rescaled FPCs $\boldsymbol{\phi}_k(\cdot) \in \left(\mathcal{C}^{(q,\mu)}[0,1]\right)^L$ with $\sum_{k=1}^{\infty} \|\boldsymbol{\phi}_k\|_{0,\mu} + \sum_{k=1}^{\infty} \|\boldsymbol{\phi}_k\|_{q,\mu} + \sum_{k=1}^{\infty} \|\boldsymbol{\phi}_k\|_{\infty} < \infty$ with the norm defined in (2.8).

(A5) On the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ are FPC scores $\{\boldsymbol{\zeta}_{itk}\}_{r(i,t)\in\mathbb{Z},k\geq 1}$ independent over $k \geq 1$ and i.i.d. over $r(i,t) \in \mathbb{Z}$ with $\sup_{k\geq 1} \mathbb{E}\|\boldsymbol{\zeta}_{11k}\|_2^{r_0} < \infty$ for r_0 in (2.14), measurement errors $\{\boldsymbol{\varepsilon}_{itj}\}_{r(i,t)\geq 1, j\geq 1}$ independent over $r(i,t) \geq 1$ and i.i.d. over $j \geq 1$, and $\{\boldsymbol{\zeta}_{itk}\}_{r(i,t)\in\mathbb{Z},k\geq 1}$ independent of $\{\boldsymbol{\varepsilon}_{itj}\}_{r(i,t)\geq 1, j\geq 1}$, with covariance matrices $\boldsymbol{\Sigma}_{k,\zeta} = \operatorname{cov}\left(\boldsymbol{\zeta}_{11,k}\right)$, $\boldsymbol{\Sigma}_{it,\varepsilon} = \operatorname{cov}\left(\boldsymbol{\varepsilon}_{it,1}\right)$. There exist i.i.d. $N(\mathbf{0}_L, \boldsymbol{\Sigma}_{k,\zeta})$ random vectors $\{\boldsymbol{Z}_{itk,\zeta}\}_{r(i,t)=1-nI_{nT}}^{nT}$ for $1 \leq k \leq k_{nT}$, and i.i.d $N(\mathbf{0}_L, \boldsymbol{\Sigma}_{it,\varepsilon})$ random vectors $\{\boldsymbol{Z}_{itj,\varepsilon}\}_{j=1}^{N}$ for $1 \leq r(i,t) \leq nT$ on a new probability space $\left(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}\right)$, C_1, $C_2 \in (0,+\infty)$, γ_1, $\gamma_2 \in (1,+\infty)$, $\beta_1 \in (0,1/2)$, β_2 in (2.13), such that

$$\mathbb{P}\left\{\max_{1\leq k\leq k_{nT}} \max_{1-nI_{nT}\leq\tau\leq nT} \left\| \sum_{r(i,t)=1-nI_{nT}}^{\tau} \left(\boldsymbol{\zeta}_{itk} - \boldsymbol{Z}_{itk,\zeta}\right) \right\|_2 > (nT)^{\beta_1} \right\} < C_1 (nT)^{-\gamma_1},$$

$$\mathbb{P}\left\{\max_{1\leq r(i,t)\leq nT} \max_{1\leq\tau\leq N} \left\| \sum_{j=1}^{\tau} \left(\boldsymbol{\varepsilon}_{itj} - \boldsymbol{Z}_{itj,\varepsilon}\right) \right\|_2 > N^{\beta_2} \right\} < C_2 N^{-\gamma_2},$$

where positive integers $k_{nT} = \mathcal{O}\{(nT)^{\omega}\}$ for some $\omega > 0$, $I_{nT} > -10\log(nT)/\log\rho_a$, $I_{nT} \asymp \log(nT)$ and ρ_a the geometric decay rate

in (1.4). To avoid complex notations, $\{\boldsymbol{\zeta}_{itk}\}_{r(i,t)=1-nI_{nT},k=1}^{nT,k_{nT}}$, $\{\boldsymbol{\varepsilon}_{itj}\}_{r(i,t)=1,j=1}^{nT,N}$ are used to represent random variables on $\left(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{\mathbb{P}}\right)$ with the same joint distribution, and $\mathbb{P}$ to represent $\tilde{\mathbb{P}}$.

(A6) The spline order $p \geq p^*$, the number of interior knots $J_s = N^{\gamma} d_N$ for γ in (2.15), $d_N + d_N^{-1} = \mathcal{O}\left(\log^{\theta} N\right)$ for θ in (2.12), as $N \to \infty$.

(A5') The FPC scores $\{\boldsymbol{\zeta}_{itk}\}_{r(i,t)\in\mathbb{Z},k\geq 1}$ are independent over $k \geq 1$ and i.i.d. over $r(i,t) \in \mathbb{Z}$, the measurement errors $\{\boldsymbol{\varepsilon}_{itj}\}_{r(i,t)\geq 1,j\geq 1}$ are independent over $r(i,t) \geq 1$ and i.i.d. over $j \geq 1$, and $\{\boldsymbol{\varepsilon}_{itj}\}_{r(i,t)\geq 1,j\geq 1}$ are independent of $\{\boldsymbol{\zeta}_{itk}\}_{r(i,t)\in\mathbb{Z},k\geq 1}$. There exist constants $r_1 > 4+2\omega$, $r_2 > (2+\theta)/\beta_2$, for some $\omega > 0$, θ in (2.12) and β_2 in (2.13), such that $\sup_{k\geq 1}\mathbb{E}\|\boldsymbol{\zeta}_{11,k}\|_2^{r_1} + \sup_{r(i,t)\geq 1}\mathbb{E}\|\boldsymbol{\varepsilon}_{it,1}\|_2^{r_2} < \infty$. For covariance matrices $\boldsymbol{\Sigma}_{k,\zeta} = \operatorname{cov}\left(\boldsymbol{\zeta}_{11,k}\right), \boldsymbol{\Sigma}_{it,\varepsilon} = \operatorname{cov}\left(\boldsymbol{\varepsilon}_{it,1}\right)$, there exists a constant $c_\lambda > 0$, such that $\lambda_{\max}\left(\boldsymbol{\Sigma}_{k,\zeta}\right)/\lambda_{\min}\left(\boldsymbol{\Sigma}_{k,\zeta}\right) < c_\lambda, \lambda_{\max}\left(\boldsymbol{\Sigma}_{it,\varepsilon}\right)/\lambda_{\min}\left(\boldsymbol{\Sigma}_{it,\varepsilon}\right) < c_\lambda, \forall k,i,t \in \mathbb{N}_+$.

Assumptions (A1) and (A2) are typical for spline smoothing. In particular, (A1) controls the size of the bias of the spline smoother for $\boldsymbol{m}(\cdot)$ and (A2) requires that the variance function is uniformly bounded on its domain. Assumption (A3) restricts that total sample size $n \times T$ increases by a fractional power θ of N, the number of observations for each tra-

jectory. The collective bounded smoothness of the principal components is stated in Assumption (A4). Assumption (A5) provides the Gaussian strong approximation of estimation errors as well as the FPC score innovations $\{\boldsymbol{\zeta}_{itk}\}_{i=1,t=-\infty,k=1}^{\infty,\infty,\infty}$. The "high level" Assumption (A5) can be ensured by an elementary Assumption (A5'). Assumption (A6) makes constraints on the number of interior knots J_s and the measurement times N.

Remark 1. The assumptions above are quite mild and easily satisfied. Default values for q, μ, θ, p, γ are $q + \mu = p^* = 4$, $\nu = 1$, $\theta = 1$, $p = 4$ (cubic spline), $\gamma = 3/8$, $d_N \asymp \log\log N$ are used for simulation in Section 4. The choice of $q = 0$, $\mu = 1/2$, $p^* = 1/2$, $\nu = 1/2$, $\theta = 3/5$, $p = 1$ (constant spline), $\gamma = 16/25, d_N \asymp \log\log N$ also satisfies all assumptions and allows for non differentiable trajectories, as in Mohammadi and Panaretos (2023).

Remark 2. The Gaussian process $\boldsymbol{\varphi}(\cdot)$ of Theorem 1 is L-dimensional according to definition in (2.19). This rules out the possibility of dimension $L \to \infty$, since as a weak limit, $\boldsymbol{\varphi}(\cdot)$ is fixed.

Remark 3. We concur with one Referee that it is feasible to allow time dependent measurement errors $\{\boldsymbol{\varepsilon}_{itj}\}_{r(i,t)\geq 1, j\geq 1}$, making use of techniques in Huang et al. (2022). This has been omitted due to space constraint.

Remark 4. A Referee has observed that the number n of subjects does

not have to diverge hence it is feasible to make inference on each subject's vector mean function as long as T goes to infinity.

Remark 5. The independence of FPC scores $\{\boldsymbol{\zeta}_{itk}\}_{r(i,t)\in\mathbb{Z},k\geq 1}$ over $k \geq 1$ is presumed in most existing literature, and is needed to combine strong Gaussian approximation of $\{\boldsymbol{\zeta}_{itk}\}_{r(i,t)\in\mathbb{Z}}$ for all $k \geq 1$ in one common probability space, see Lemma S.5 in the Supplement.

2.3 Asymptotic properties of the infeasible estimator

The infeasible estimator $\bar{\boldsymbol{m}}(\cdot)$ is examined in this subsection.

Define a limiting matrix covariance function

$$\boldsymbol{G}_{\varphi}(x,x') = \sum_{k=1}^{\infty} \boldsymbol{\Delta}_k \circ \left\{ \boldsymbol{\phi}_k(x) \boldsymbol{\phi}_k(x')^{\top} \right\}, \tag{2.16}$$

where $\boldsymbol{\Delta}_k$ is the long-run covariance matrix of the vector series $\{\boldsymbol{\xi}_{itk}\}_{t\in\mathbb{Z}}$:

$$\boldsymbol{\Delta}_k = \sum_{h=-\infty}^{+\infty} \mathbb{E}\boldsymbol{\xi}_{itk}\boldsymbol{\xi}_{i,t+h,k}^{\top} = \left(\sum_{t=0}^{\infty} \boldsymbol{a}_{tk}\boldsymbol{a}_{tk}^{\top} + 2\sum_{t=0}^{\infty}\sum_{h=1}^{\infty} \boldsymbol{a}_{tk}\boldsymbol{a}_{t+h,k}^{\top} \right) \circ \boldsymbol{\Sigma}_{k,\zeta}.$$

Noting that all elements of $\boldsymbol{\Delta}_k$ are uniformly bounded for $k \in \mathbb{N}_+$, one can find a sequence of independent Gaussian vectors $\{\boldsymbol{U}_k\}_{k\in\mathbb{N}_+}$ of L-dimension

with $\mathbb{E}\boldsymbol{U}_k = \mathbf{0}_L$ and $\mathbb{E}\boldsymbol{U}_k\boldsymbol{U}_k^\top = \boldsymbol{\Delta}_k$. Define

$$U = \left\{ \left(\boldsymbol{\lambda}_1^{1/2} \circ \boldsymbol{U}_1\right)^\top, \left(\boldsymbol{\lambda}_2^{1/2} \circ \boldsymbol{U}_2\right)^\top, \ldots \right\}^\top, \tag{2.17}$$

where $\boldsymbol{\lambda}_k^{1/2}, k \geq 1$ denote the element-wise square-root of the vector, then $\mathbb{E}\left\|U\right\|_2^2 \leq C\sum_{k=1}^{\infty}\left\|\boldsymbol{\lambda}_k\right\|_1 < \infty$, thus $U \in \ell^2$ *a.s.*. In fact, U is a Gaussian ℓ^2-random variable since $u^*(U)$ is Gaussian for all $u^* \in (\ell^2)^*$. For any $u = \left(\boldsymbol{u}_1^\top, \boldsymbol{u}_2^\top, \ldots\right)^\top \in \ell^2$, where $\{\boldsymbol{u}_k\}_{k\in\mathbb{N}_+}$ are vectors of length L, consider a map $\Pi : \ell^2 \to \left(\mathcal{L}^2[0,1]\right)^L$ defined as $\Pi(u)(\cdot) = \sum_{k=1}^{\infty}\boldsymbol{u}_k \circ \boldsymbol{\psi}_k(\cdot)$, then Π is isometric, which yields a Gaussian $\left(\mathcal{L}^2[0,1]\right)^L$-random variable

$$\Pi(U)(\cdot) = \sum_{k=1}^{\infty}\boldsymbol{\lambda}_k^{1/2} \circ \boldsymbol{U}_k \circ \boldsymbol{\psi}_k(\cdot) = \sum_{k=1}^{\infty}\boldsymbol{U}_k \circ \boldsymbol{\phi}_k(\cdot) \tag{2.18}$$

with matrix covariance function $\boldsymbol{G}_\varphi$.

One defines next several rescaled Gaussian processes derived from $\Pi(U)(\cdot)$

$$\boldsymbol{\varphi}(\cdot) = \boldsymbol{G}_\varphi^{-1/2}(\cdot,\cdot)\,\Pi(U)(\cdot), \tag{2.19}$$

$$\boldsymbol{\varphi}_L(\cdot) = \left[\operatorname{diag}\left\{\boldsymbol{G}_\varphi(\cdot,\cdot)\right\}\right]^{-1/2}\boldsymbol{G}_\varphi^{1/2}(\cdot,\cdot)\,\boldsymbol{\varphi}(\cdot), \tag{2.20}$$

$$\varphi_{\boldsymbol{b}}(\cdot) = \left\{\boldsymbol{b}^\top\boldsymbol{G}_\varphi(\cdot,\cdot)\boldsymbol{b}\right\}^{-1/2}\boldsymbol{b}^\top\boldsymbol{G}_\varphi^{1/2}(\cdot,\cdot)\,\boldsymbol{\varphi}(\cdot), \tag{2.21}$$

where $\boldsymbol{b} \in \mathbb{R}^L/\{\mathbf{0}_L\}$ and the $\operatorname{diag}(\boldsymbol{A})$ sets all off-diagonal elements of a

square matrix $\boldsymbol{A}$ to zero. They are respectively $(\mathcal{L}^2[0,1])^L$, $(\mathcal{L}^2[0,1])^L$ and $\mathcal{L}^2[0,1]$ Gaussian random variables, with matrix covariance functions

$$\mathbb{E}\boldsymbol{\varphi}(x)\,\boldsymbol{\varphi}(x')^\top = \boldsymbol{G}_\varphi^{-1/2}(x,x)\,\boldsymbol{G}_\varphi(x,x')\,\boldsymbol{G}_\varphi^{-1/2}(x',x'),$$

$$\mathbb{E}\boldsymbol{\varphi}_L(x)\boldsymbol{\varphi}_L(x')^\top = \left[\operatorname{diag}\left\{\boldsymbol{G}_{\boldsymbol{\varphi}}(x,x)\right\}\right]^{-1/2}\boldsymbol{G}_{\boldsymbol{\varphi}}(x,x')\left[\operatorname{diag}\left\{\boldsymbol{G}_{\boldsymbol{\varphi}}(x',x')\right\}\right]^{-1/2},$$

$$\mathbb{E}\varphi_{\boldsymbol{b}}(x)\varphi_{\boldsymbol{b}}(x') = \boldsymbol{b}^\top\boldsymbol{G}_\varphi(x,x')\boldsymbol{b}\left\{\boldsymbol{b}^\top\boldsymbol{G}_\varphi(x,x)\boldsymbol{b}\boldsymbol{b}^\top\boldsymbol{G}_\varphi(x',x')\boldsymbol{b}\right\}^{-1/2},\ x,x'\in[0,1].$$

Weak convergence of the infeasible estimator $\bar{\boldsymbol{m}}(\cdot)$ follows.

Theorem 1. *Under Assumptions (A1), (A4) and (A5), as $T\to\infty$, the infeasible estimator $\bar{\boldsymbol{m}}(\cdot)$ converges at the $(nT)^{1/2}$ rate to $\boldsymbol{m}(\cdot)$ with asymptotic matrix covariance function $\boldsymbol{G}_\varphi(x,x')$, i.e.,*

$$(nT)^{1/2}\,\boldsymbol{G}_\varphi^{-1/2}(\cdot,\cdot)\left\{\bar{\boldsymbol{m}}(\cdot)-\boldsymbol{m}(\cdot)\right\}\Rightarrow\boldsymbol{\varphi}(\cdot).$$

Consequently, (2.20) and (2.21) imply that

$$(nT)^{1/2}\left[\operatorname{diag}\left\{\boldsymbol{G}_\varphi(\cdot,\cdot)\right\}\right]^{-1/2}\left\{\bar{\boldsymbol{m}}(\cdot)-\boldsymbol{m}(\cdot)\right\}\Rightarrow\boldsymbol{\varphi}_L(\cdot),$$

$$(nT)^{1/2}\left\{\boldsymbol{b}^\top\boldsymbol{G}_\varphi(\cdot,\cdot)\boldsymbol{b}\right\}^{-1/2}\boldsymbol{b}^\top\left\{\bar{\boldsymbol{m}}(\cdot)-\boldsymbol{m}(\cdot)\right\}\Rightarrow\boldsymbol{\varphi}_{\boldsymbol{b}}(\cdot).$$

For any $\alpha\in(0,1)$, denote by $Q_{\boldsymbol{b},1-\alpha}$ and $Q_{L,1-\alpha}$ respectively the $100(1-\alpha)$-th percentile of the absolute maxima distribution of $\varphi_{\boldsymbol{b}}(x)$ and $\|\boldsymbol{\varphi}_L(x)\|_\infty$

over $x \in [0,1]$, i.e.,

$$\begin{aligned} \mathbb{P}\left\{\sup_{x\in[0,1]} |\varphi_{\boldsymbol{b}}(x)| \leq Q_{\boldsymbol{b},1-\alpha}\right\} &= 1-\alpha,\ \boldsymbol{b} \in \mathbb{R}^L / \{\mathbf{0}_L\}, \\ \mathbb{P}\left\{\sup_{x\in[0,1]} \|\boldsymbol{\varphi}_L(x)\|_\infty \leq Q_{L,1-\alpha}\right\} &= 1-\alpha. \end{aligned} \tag{2.22}$$

Corollary 1. *Under Assumptions (A1), (A3)-(A5), as $T \to \infty$, one has*

$$\begin{aligned} &\mathbb{P}\left[\sup_{x\in[0,1]} (nT)^{1/2} \left|\left\{\boldsymbol{b}^\top \boldsymbol{G}_\varphi(x,x)\boldsymbol{b}\right\}^{-1/2} \boldsymbol{b}^\top \{\bar{\boldsymbol{m}}(x) - \boldsymbol{m}(x)\}\right| \leq Q_{\boldsymbol{b},1-\alpha}\right] \to 1-\alpha, \\ &\mathbb{P}\left[\sup_{x\in[0,1]} (nT)^{1/2} \left\|\left[\operatorname{diag}\{\boldsymbol{G}_\varphi(x,x)\}\right]^{-1/2} \{\bar{\boldsymbol{m}}(x) - \boldsymbol{m}(x)\}\right\|_\infty \leq Q_{L,1-\alpha}\right] \to 1-\alpha, \end{aligned} \tag{2.23}$$

where $\boldsymbol{b} \in \mathbb{R}^L / \{\mathbf{0}_L\}$.

To construct SCBs for $\boldsymbol{m}(\cdot)$, one shows next that the spline estimator $\hat{\boldsymbol{m}}(\cdot)$ in (2.10) is a good substitute of the infeasible estimator $\bar{\boldsymbol{m}}(\cdot)$.

2.4 Oracle efficiency

The next Theorem states that up to order $o_p\left\{(nT)^{-1/2}\right\}$, the proposed two-step spline estimator $\hat{\boldsymbol{m}}(\cdot)$ is oracally efficient, i.e., it is asymptotically equivalent to, or as efficient as the infeasible estimator $\bar{\boldsymbol{m}}(\cdot)$ with all trajectories $\boldsymbol{\eta}_{it}(\cdot)$ fully known by "oracle". Thus $\hat{\boldsymbol{m}}(\cdot)$ enjoys all the same asymptotic properties as $\bar{\boldsymbol{m}}(\cdot)$.

Theorem 2. *Under Assumptions (A1)–(A6), the B-spline estimator* $\hat{\boldsymbol{m}}(\cdot)$ *is oracally efficient, i.e. as* $T \to \infty$,

$$\sup_{x\in[0,1]} (nT)^{1/2} \left\|\bar{\boldsymbol{m}}(x) - \hat{\boldsymbol{m}}(x)\right\|_\infty = \mathcal{O}_p(1).$$

Corollary 2. *Under Assumptions (A1)-(A6), for any* $\alpha \in (0,1)$, *as* $T \to \infty$, *an asymptotic* $100(1-\alpha)\%$ *correct SCB for any linear combination* $\boldsymbol{b}^\top \boldsymbol{m}(\cdot)$ *of* $\boldsymbol{m}(\cdot)$, *is given by*

$$\boldsymbol{b}^\top \hat{\boldsymbol{m}}(\cdot) \pm \left\{\boldsymbol{b}^\top \boldsymbol{G}_\varphi(\cdot,\cdot)\boldsymbol{b}\right\}^{1/2} Q_{\boldsymbol{b},1-\alpha} (nT)^{-1/2}, \ \boldsymbol{b} \in \mathbb{R}^L / \{\boldsymbol{0}_L\},$$

and an asymptotic $100(1-\alpha)\%$ *correct SCR for* $\boldsymbol{m}(\cdot)$ *is given by*

$$\hat{\boldsymbol{m}}(\cdot) \pm \left[\operatorname{diag}\left\{\boldsymbol{G}_\varphi(\cdot,\cdot)\right\}\right]^{1/2} \boldsymbol{1}_L Q_{L,1-\alpha} (nT)^{-1/2}.$$

2.5 Multiple comparison

This subsection concerns simultaneous comparison of vector mean functions from multiple samples.

Suppose for each $s = 1, \ldots, S$ an independent set of functional panel data of size n_s and time length T is recorded, for which Assumptions (A1)–

(A6) are satisfied, then as $T \to \infty$,

$$(n_s T)^{1/2} \left[\operatorname{diag}\left\{\boldsymbol{G}_{\varphi,s}(\cdot,\cdot)\right\}\right]^{-1/2} \left\{\hat{\boldsymbol{m}}_s(\cdot) - \boldsymbol{m}_s(\cdot)\right\} \Rightarrow \boldsymbol{\varphi}_{L,s}(\cdot),$$

where $\boldsymbol{\varphi}_{L,s}(\cdot)$, $s = 1, \ldots, S$ are independent $(\mathcal{L}^2[0,1])^L$-Gaussian processes.

For $1 \le s < s' \le S$, denote sample ratios $\tau_{ss'} = \lim_{T\to\infty} n_s/n_{s'}$, which satisfy $\tau_{ss'} \in [c_\tau, C_\tau]$ for some constants $0 < c_\tau < C_\tau < \infty$. For $1 \le s < s' \le S$, define $(\mathcal{L}^2[0,1])^L$-Gaussian processes

$$\boldsymbol{\varphi}_{L,ss'}(\cdot) = \left[\operatorname{diag}\left\{\boldsymbol{G}_{\varphi,s}(\cdot,\cdot)\right\}\right]^{1/2} \boldsymbol{\varphi}_{L,s}(\cdot) + \left[\tau_{ss'}\operatorname{diag}\left\{\boldsymbol{G}_{\varphi,s'}(\cdot,\cdot)\right\}\right]^{1/2} \boldsymbol{\varphi}_{L,s'}(\cdot),$$

which is of zero mean and covariance function $\boldsymbol{G}_{ss'}(x,x') = \boldsymbol{G}_{\varphi,s}(x,x') + \tau_{ss'}\boldsymbol{G}_{\varphi,s'}(x,x')$, and percentiles $Q_{1-\alpha}^{(s,s')}$, $Q_{1-\alpha}^{(\max)}$ satisfying

$$\mathbb{P}\left\{\sup_{x\in[0,1]} \left\|\left[\operatorname{diag}\left\{\boldsymbol{G}_{ss'}(x,x)\right\}\right]^{-1/2} \boldsymbol{\varphi}_{L,ss'}(x)\right\|_\infty \le Q_{1-\alpha}^{(s,s')}\right\} = 1-\alpha,$$

$$\mathbb{P}\left\{\sup_{x\in[0,1]} \max_{1\le s<s'\le S} \left\|\left[\operatorname{diag}\left\{\boldsymbol{G}_{ss'}(x,x)\right\}\right]^{-1/2} \boldsymbol{\varphi}_{L,ss'}(x)\right\|_\infty \le Q_{1-\alpha}^{(\max)}\right\} = 1-\alpha.$$

Denote difference of vector mean functions $\boldsymbol{d}_{ss'}(\cdot) = \boldsymbol{m}_s(\cdot) - \boldsymbol{m}_{s'}(\cdot)$ with estimate $\hat{\boldsymbol{d}}_{ss'}(\cdot) = \hat{\boldsymbol{m}}_s(\cdot) - \hat{\boldsymbol{m}}_{s'}(\cdot)$.

Corollary 3. *As* $T \to \infty$,

$$\mathbb{P}\left[\sup_{x\in[0,1]} (n_s T)^{1/2} \left\|[\operatorname{diag}\{\boldsymbol{G}_{ss'}(x,x)\}]^{-1/2}\left\{\hat{\boldsymbol{d}}_{ss'}(x) - \boldsymbol{d}_{ss'}(x)\right\}\right\|_\infty \le Q_{1-\alpha}^{(s,s')}\right] \to 1-\alpha,$$

$$\mathbb{P}\left[\sup_{x\in[0,1]} \max_{1\le s<s'\le S} (n_s T)^{1/2} \left\|[\operatorname{diag}\{\boldsymbol{G}_{ss'}(x,x)\}]^{-1/2}\left\{\hat{\boldsymbol{d}}_{ss'}(x) - \boldsymbol{d}_{ss'}(x)\right\}\right\|_\infty \le Q_{1-\alpha}^{(\max)}\right]$$

$$\to 1-\alpha.$$

Consequently, a multiple SCR for $\boldsymbol{d}_{ss'}(\cdot), 1 \le s < s' \le S$ *is*

$$\hat{\boldsymbol{d}}_{ss'}(x) \pm [\operatorname{diag}\{\boldsymbol{G}_{ss'}(x,x)\}]^{1/2} \mathbf{1}_L Q_{1-\alpha}^{(\max)} (n_s T)^{-1/2}, 1 \le s < s' \le S.$$

3. Implementation

This section describes implementation of the SCB and SCR in Corollary 2.

3.1 Knot selection

The number of knots J_s for spline smoothing is selected subject to the constraints of Assumption (A6). The smoothness order (q, μ) of vector mean function $\boldsymbol{m}(\cdot)$ and eigenfunctions $\{\boldsymbol{\phi}_k\}_{k\in\mathbb{N}_+}$ is taken as $(3,1)$ or $(4,0)$ by default with a matching spline order $p = 4$ (cubic spline). Therefore, $J_s = [cN^\gamma \log\log N]$ is recommended with constant c, where $[a]$ denotes the

integer part of a. The default values of parameters $\gamma = 3/8$ and $c = 1$ are found to be adequate in the simulation study.

3.2 Estimation of matrix covariance function

Recall that the matrix covariance function of $\boldsymbol{\xi}_{it}(\cdot)$ is $\boldsymbol{G}(x, x') = \mathbb{E}\boldsymbol{\xi}_{it}(x)\boldsymbol{\xi}_{it}^{\top}(x')$, which is well approximated by the sample version $\bar{\boldsymbol{G}}(x, x') = (nT)^{-1}\sum_{r(i,t)=1}^{nT} \boldsymbol{\xi}_{it}(x)\boldsymbol{\xi}_{it}^{\top}(x')$. Since trajectories $\{\boldsymbol{\xi}_{it}(\cdot)\}$ are unobservable, $\bar{\boldsymbol{G}}(x, x')$ is infeasible. Yet it suggests to use the plug-in sample covariance

$$\hat{\boldsymbol{G}}(x, x') = \left\{\hat{G}^{(l,l')}(x, x')\right\}_{1\le l,l'\le L} = (nT)^{-1}\sum_{r(i,t)=1}^{nT} \hat{\boldsymbol{\xi}}_{it}(x)\hat{\boldsymbol{\xi}}_{it}^{\top}(x'), \qquad (3.24)$$

where $\hat{\boldsymbol{\xi}}_{it}(\cdot) = \hat{\boldsymbol{\eta}}_{it}(\cdot) - \hat{\boldsymbol{m}}(\cdot)$. It is feasible to show that

$$\max_{1\le l,l'\le L}\sup_{(x,x')\in[0,1]^2}(nT)^{1/2}\left|\hat{G}^{(l,l')}(x, x') - G^{(l,l')}(x, x')\right| = \mathcal{O}_p(1).$$

The limiting matrix long-run covariance function $\boldsymbol{G}_{\varphi}(\cdot, \cdot)$ of $\boldsymbol{\xi}_{it}(\cdot)$ is essential to construct the SCB and SCR. Note that

$$\begin{aligned}\boldsymbol{G}_{\varphi}(x, x') &= \lim_{T\to\infty} nT\,\mathbb{E}\left\{\bar{\boldsymbol{m}}(x) - \boldsymbol{m}(x)\right\}\left\{\bar{\boldsymbol{m}}(x') - \boldsymbol{m}(x')\right\}^{\top} \\ &= \lim_{T\to\infty}\mathbb{E}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\boldsymbol{\eta}_{1t}(x) - \boldsymbol{m}(x)\right\}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\boldsymbol{\eta}_{1t}(x') - \boldsymbol{m}(x')\right\}^{\top}.\end{aligned}$$

One defines

$$\hat{\boldsymbol{\delta}}_{ij}(\cdot)=\frac{1}{\sqrt{B}}\sum_{t=(j-1)B+1}^{jB}\left\{\hat{\boldsymbol{\eta}}_{it}(\cdot)-\hat{\boldsymbol{m}}(\cdot)\right\},$$

for $i=1,\ldots,n$, $j=1,\ldots,[T/B]$, where $B\to\infty$ and $B=o(T)$ as $T\to\infty$, one would then use the following estimator of $\boldsymbol{G}_{\varphi}(x,x')$

$$\hat{\boldsymbol{G}}_{\varphi}(x,x')=\frac{1}{n[T/B]}\sum_{i=1}^{n}\sum_{j=1}^{[T/B]}\hat{\boldsymbol{\delta}}_{ij}(x)\,\hat{\boldsymbol{\delta}}_{ij}(x')^{\top},(x,x')\in[0,1]^2. \tag{3.25}$$

3.3 FPC analysis

For functional random variable $\xi_{it}^{(l)}(\cdot)$, the l-th element of functional random vector $\boldsymbol{\xi}_{it}(\cdot)$, its covariance function is the l-th diagonal element of the matrix covariance function $\boldsymbol{G}(x,x)$, with

$$\int_{[0,1]}G^{(l,l)}(x,x')\,\psi_k^{(l)}(x')dx'=\lambda_k^{(l)}\psi_k^{(l)}(x),\quad\int_{[0,1]}\left\{\psi_k^{(l)}(x)\right\}^2dx=1. \tag{3.26}$$

The eigenvalues and eigenfunctions for $G^{(l,l)}(\cdot,\cdot)$ are approximated as

$$\left\{\hat{\psi}_k^{(l)}(\cdot)\in\mathcal{H}^{(p-2)},\ \lambda_1^{(l)}\geq\lambda_2^{(l)}\geq\ldots\geq0,\ k=1,\ldots,J_s+p:\right.$$
$$\left.\int\hat{G}^{(l,l)}(x,x')\,\hat{\psi}_k^{(l)}(x')\,dx'=\hat{\lambda}_k^{(l)}\hat{\psi}_k^{(l)}(x),\quad\int\left[\hat{\psi}_k^{(l)}(x)\right]^2dx=1\right\}.$$

Linear algebra shows that $\hat{\psi}_k^{(l)}$ and $\hat{\lambda}_k^{(l)}$ are well defined.

Next, the number of eigenvalues to select is $\kappa = \max\left\{\kappa^{(l)}, l = 1, \ldots, L\right\}$, where $\kappa^{(l)} = \arg\min_{1\le\kappa'\le J_s+p}\left\{\sum_{k=1}^{\kappa'}\hat{\lambda}_k^{(l)} / \sum_{k=1}^{J_s+p}\hat{\lambda}_k^{(l)} > 0.95\right\}$. Note that $\left\{\hat{\lambda}_k^{(l)}, k = 1, \ldots, J_s + p\right\}$ are the complete set of eigenvalues of $\hat{G}^{(l,l)}(\cdot,\cdot)$, which are non-negative since the integral operator defined by function $\hat{G}^{(l,l)}(\cdot,\cdot)$ is non-negative definite.

3.4 Estimating the percentiles

Define $\hat{\Pi}(u)(\cdot) = \sum_{k=1}^{J_s+p}\boldsymbol{u}_k \circ \hat{\boldsymbol{\psi}}_k(\cdot)$ for $u = \left(\boldsymbol{u}_1^\top, \ldots, \boldsymbol{u}_{J_s+p}^\top\right)^\top$, where $\boldsymbol{u}_k \in \mathbb{R}^L$, $\hat{\boldsymbol{\psi}}_k = \left\{\hat{\psi}_k^{(1)}, \ldots, \hat{\psi}_k^{(L)}\right\}^\top$. Noting that $\left\{\hat{\psi}_k^{(l)}, 1 \le k \le J_s + p\right\}$ is orthonormal basis of $\mathcal{H}^{(p-2)}$, $\hat{\Pi} : \mathbb{R}^{L(J_s+p)} \to \left\{\mathcal{H}^{(p-2)}\right\}^L$ is isometric. Define $\hat{U}$ as an $L(J_s + p)$-dimensional random vector by

$$\hat{U} = \left\{\left(\hat{\boldsymbol{\lambda}}_1^{1/2} \circ \hat{\boldsymbol{U}}_1\right)^\top, \ldots, \left(\hat{\boldsymbol{\lambda}}_{J_s+p}^{1/2} \circ \hat{\boldsymbol{U}}_{J_s+p}\right)^\top\right\}^\top, \ \hat{\boldsymbol{U}}_k \sim N(\boldsymbol{0}_L, \hat{\boldsymbol{\Delta}}_k),$$

where

$$\hat{\boldsymbol{\Delta}}_k = \int_{[0,1]}\int_{[0,1]}\left\{\hat{\boldsymbol{\lambda}}_k^{-1/2} \circ \hat{\boldsymbol{\psi}}_k(x)\right\}\left\{\hat{\boldsymbol{\lambda}}_k^{-1/2} \circ \hat{\boldsymbol{\psi}}_k(x)\right\}^\top \circ \hat{\boldsymbol{G}}_\varphi(x, x')\, dxdx',$$

$\hat{\boldsymbol{\lambda}}_k = \left(\hat{\lambda}_k^{(1)}, \ldots, \hat{\lambda}_k^{(L)}\right)^\top, 1 \le k \le J_s + p$, and $\hat{\boldsymbol{\lambda}}_k^{1/2}, \hat{\boldsymbol{\lambda}}_k^{-1/2}$ denote the element-wise square-root and inverse of square-root of vector $\hat{\boldsymbol{\lambda}}_k$, then $\hat{U}$ is a data version of the ℓ^2-random variable U in (2.17). For a large integer s_M, gen-

erate Gaussian random vectors $\left\{\hat{U}_s\right\}_{1\leq s\leq s_M}$ distributed as $\hat{U}$, and denote Gaussian functional random vector $\hat{\boldsymbol{\varphi}}_{L,s}(\cdot)$ and Gaussian functional random variable $\hat{\varphi}_{\boldsymbol{b},s}(\cdot)$ by

$$\hat{\boldsymbol{\varphi}}_{L,s}(\cdot) = \left[\operatorname{diag}\left\{\hat{\boldsymbol{G}}_{\varphi}(\cdot,\cdot)\right\}\right]^{-1/2}\hat{\Pi}(\hat{U}_s)(\cdot),$$
$$\hat{\varphi}_{\boldsymbol{b},s}(\cdot) = \left\{\boldsymbol{b}^{\top}\hat{\boldsymbol{G}}_{\varphi}(\cdot,\cdot)\boldsymbol{b}\right\}^{-1/2}\boldsymbol{b}^{\top}\hat{\Pi}(\hat{U}_s)(\cdot),\ s=1,\ldots,s_M.$$

One takes respectively the empirical quantiles $\hat{Q}_{\boldsymbol{b},1-\alpha}$ of $\left\{\sup_{x\in[0,1]}|\hat{\varphi}_{\boldsymbol{b},s}(x)|\right\}_{s=1}^{s_M}$ and $\hat{Q}_{L,1-\alpha}$ of $\left\{\sup_{x\in[0,1]}\left\|\hat{\boldsymbol{\varphi}}_{L,s}(x)\right\|_{\infty}\right\}_{s=1}^{s_M}$ as estimates of $Q_{\boldsymbol{b},1-\alpha}$ and $Q_{L,1-\alpha}$.

For any vector $\boldsymbol{b}\in\mathbb{R}^L/\{\boldsymbol{0}_L\}$, the SCB for $\boldsymbol{b}^{\top}\boldsymbol{m}(\cdot)$ is given by

$$\boldsymbol{b}^{\top}\hat{\boldsymbol{m}}(\cdot)\pm\left\{\boldsymbol{b}^{\top}\hat{\boldsymbol{G}}_{\varphi}(\cdot,\cdot)\boldsymbol{b}\right\}^{1/2}\hat{Q}_{\boldsymbol{b},1-\alpha}\left(nT\right)^{-1/2}, \tag{3.27}$$

and the SCR for $\boldsymbol{m}(\cdot)=\left\{m^{(1)}(\cdot),\ldots,m^{(L)}(\cdot)\right\}^{\top}$ is given by

$$\hat{\boldsymbol{m}}(\cdot)\pm\left[\operatorname{diag}\left\{\hat{\boldsymbol{G}}_{\varphi}(\cdot,\cdot)\right\}\right]^{1/2}\boldsymbol{1}_L\hat{Q}_{L,1-\alpha}\left(nT\right)^{-1/2}. \tag{3.28}$$

4. Simulation

In this section, one examines finite-sample performance of the proposed SCB and SCR. The data are generated from the following model: $\boldsymbol{Y}_{itj} = \boldsymbol{m}\left(j/N\right)+\sum_{k=1}^{7}\boldsymbol{\xi}_{itk}\circ\boldsymbol{\phi}_k\left(j/N\right)+\boldsymbol{\sigma}_{it}\left(j/N\right)\circ\boldsymbol{\varepsilon}_{itj}$, for $1\leq i\leq n$, $1\leq t\leq T$,

$1 \leq j \leq N$, with $L = 3$. Mean functions $m^{(1)}(x) = 5 + 2\sin\{4\pi(0.5 - x)\}$, $m^{(2)}(x) = 5 + 2\sin\{4\pi(0.5 - x)\}$, $m^{(3)}(x) = 5 + \sin\{4\pi(0.5 - x)\}$, $x \in [0, 1]$, and FPC components

$$\boldsymbol{\phi}_1(x) = \sqrt{2}\sin(2\pi x)(\sqrt{4}, \sqrt{3}, \sqrt{2})^\top, \boldsymbol{\phi}_2(x) = \sqrt{2}\cos(2\pi x)(\sqrt{2}, \sqrt{2}, \sqrt{1})^\top,$$

$$\boldsymbol{\phi}_3(x) = \sqrt{2}\sin(4\pi x)(\sqrt{0.5}, \sqrt{1}, \sqrt{1})^\top, \boldsymbol{\phi}_4(x) = \sqrt{2}\cos(4\pi x)(\sqrt{0.5}, \sqrt{0.25}, \sqrt{0.5})^\top,$$

$$\boldsymbol{\phi}_5(x) = \sqrt{2}\sin(6\pi x)(\sqrt{0.25}, \sqrt{0.5}, \sqrt{0.25})^\top,$$

$$\boldsymbol{\phi}_6(x) = \sqrt{2}\cos(6\pi x)(\sqrt{0.25}, \sqrt{0.16}, \sqrt{0.25})^\top,$$

$$\boldsymbol{\phi}_7(x) = \sqrt{2}\sin(8\pi x)(\sqrt{0.09}, \sqrt{0.09}, \sqrt{0.16})^\top,$$

$\boldsymbol{\phi}_k(\cdot) \equiv \mathbf{0}$ for all $k \geq 8$. FPC scores $\boldsymbol{\xi}_{itk} = \sum_{t'=0'}^{\infty} \boldsymbol{a}_{t'k} \circ \boldsymbol{\zeta}_{i,t-t',k}$ with $\boldsymbol{a}_{0k} = 0.8\mathbf{1}_3$, $\boldsymbol{a}_{1k} = \boldsymbol{a}_{2k} = 0.4\mathbf{1}_3$, $\boldsymbol{a}_{3k} = -0.2\mathbf{1}_3$, $\boldsymbol{a}_{tk} = \mathbf{0}_3$, $\forall t \geq 4$, $k = 1, \ldots, 7$. For all $1 \leq i \leq n$, $1 \leq t \leq T$, $1 \leq k \leq 7$, $1 \leq j \leq N$, $\boldsymbol{\zeta}_{itk}$ and $\boldsymbol{\varepsilon}_{itj}$ are mutually independent and identically distributed. Elements of $\boldsymbol{\varepsilon}_{itj} = \left(\varepsilon_{itj}^{(1)}, \ldots, \varepsilon_{itj}^{(L)}\right)^\top$ are independent and distributed as $N(0, 1)$ or $U(-\sqrt{3}, \sqrt{3})$. Multivariate normal distribution and t-distribution are considered for $\boldsymbol{\zeta}_{itk}$, both distributions with two forms of covariance matrix $\boldsymbol{\Sigma}_{k,\zeta} = \left\{\boldsymbol{\Sigma}_{k,\zeta}^{(l,l')}\right\}_{l,l'=1}^{L}$:

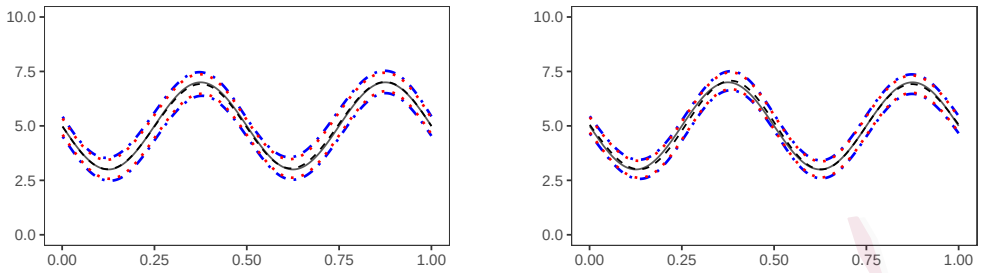
Case 1. Autoregressive (AR-1): $\boldsymbol{\Sigma}_{k,\zeta}^{(l,l')} = 0.15^{|l-l'|}$

Case 2. Toeplitz (TOEP): $\boldsymbol{\Sigma}_{k,\zeta}^{(l,l')} = 1(l = l') + \left(0.05 + 0.1^{|l-l'|}\right)\mathbf{1}(l \neq l')$.

The homoscedastic and strongly heteroscedastic deviation function, set respectively as $\boldsymbol{\sigma}_{it}(\cdot) \equiv \boldsymbol{\sigma}_{\text{homo}}(\cdot) \equiv 0.5\mathbf{1}_3$ and $\boldsymbol{\sigma}_{it}(x) \equiv \boldsymbol{\sigma}_{\text{hetero}}(x) \equiv 0.5\{5-\exp(-x)\}/\{5+\exp(-x)\}\mathbf{1}_3$ are both considered. The number N of observations per curve is taken to be 900 and 1600. The population size and the number of curve segments for per individual are respectively $n = [4\log N]$ and $T = [0.25N^{0.8}]$. Cubic spline is used, i.e. $p = 4$, and number of knots is taken as $J_s = \left[N^{3/8}\log\log N\right]$.

Table 1 displays the empirical coverage frequencies of proposed SCB of $\boldsymbol{b}^\top\boldsymbol{m}(\cdot)$, namely the percentage out of 1000 replications of the true curve $\boldsymbol{b}^\top\boldsymbol{m}(\cdot)$ being covered by the cubic spline SCB (3.27) at the N points $\{1/N,\ldots,N/N\}$, in which constant vector $\boldsymbol{b}$ is taken as $(1,-1,0)^\top$. Similar results for $\boldsymbol{b} = (1,0,0)^\top$ are given in Table S.1 of the Supplement. Table 2 displays the empirical coverage frequencies of the SCR of $\boldsymbol{m}(\cdot)$. Overall, the empirical coverage frequency approaches the nominal confidence level as N increases.

Figure 2 depicts the SCB for $m^{(1)}(\cdot) = (1,0,0)^\top\boldsymbol{m}(\cdot)$, with SCR of $\boldsymbol{m}(\cdot) = \left\{m^{(l)}(\cdot)\right\}_{l=1}^{L}$ for $l = 1$ also displayed. The SCRs are wider than SCBs which are based on linear projection of $\boldsymbol{m}(\cdot)$, and as expected, the SCBs and SCRs become narrower as N increases (thus $n(N)$ and $T(N)$ increase). Figure S.1 of the Supplement shows the cubic spline estimator



(a) (b)

Figure 2: Plots of the true $m^{(1)}(\cdot)$ (solid) and cubic spline estimator $\hat{m}^{(1)}(\cdot)$ (dashed), with 95% SCB (dotted) and SCR (dot-dashed). The number N of observations in (a) and (b) are 900 and 1600 respectively.

and SCB for $(1,-1,0)^\top \boldsymbol{m}(\cdot) = m^{(1)}(\cdot) - m^{(2)}(\cdot) \equiv 0$, by which one can test $H_0 : m^{(1)}(\cdot) \equiv m^{(2)}(\cdot)$. Since the constant function 0 is entirely covered by the SCB, one retains the null hypothesis, which is consistent with the fact.

5. Real data analysis

We apply the proposed method to a data set under *Multimodal Resource for Studying Information Processing in the Developing Brain (MIPDB)* project of Child Mind Institute. The data were presented as a resource for the investigation of information processing in the developing brain, including EEG, eye tracking and cognitive and behavior data collected from 126 individuals aged 6-44. EEG data were recorded at a sampling rate of 500Hz on 128 channels of HydrocCel Geodesic Sensor Net. 126 individuals were

divided into 6 groups according to their ages and accordingly cognitive developing stages, groups 1-6 indicating young to old. We focus on analyzing EEG during the task of surround suppression paradigm, see Langer et al. (2017) for more details.

In the task of surround suppression paradigm, each trial began with the presentation of the fixation spot for 500 ms, after which four circular stimuli were flickered on-and-off at 25Hz for 2,400 ms on a plain background. Then the following trial was initiated after an inter-trial interval of 500 ms. The task consisted of 63 complete trials. In our study, multichannel EEG signals during each trial are extracted as trajectories, which constitute multivariate functional panel data, with $N = 2.9 \times 500 = 1450$, $T = 63$ for $n_1 = 18$, $n_2 = 17$, $n_3 = 18$, $n_4 = 22$, $n_5 = 13$, and $n_6 = 17$ respectively in $S = 6$ groups (some individuals with incomplete data records are deleted).

Our study focuses on signals from 6 channels, namely O1, O2 over occipital, P3, P4 over parietal, C3 over central scalp and F3 over frontal (named by E70, E83, E52, E92, E36, E24 in 128-HydrocCel Geodesic Sensor Net respectively). Hence in fact $L = 6$ and $\boldsymbol{m}(\cdot) = \left\{m^{(O_1)}(\cdot), m^{(O_2)}(\cdot), m^{(P_3)}(\cdot), m^{(P_4)}(\cdot), m^{(C_3)}(\cdot), m^{(F_3)}(\cdot)\right\}^\top$ is the vector mean function on 6 channels.

Cubic spline estimators for mean functions of group 1 with 95% SCB of the each selected channels and 95% SCR of all selected channels are depicted

in Figures S.2 and S.3 of the Supplement respectively. Mean functions at O1, O2, P3 and P4 show similar fluctuating patterns, which is quite reasonable since occipital O1 and O2 reflect visual processing and parietal P3 and P4 sensory and visual functions. Figure 3 displays the SCBs for testing

$$H_0: \; m^{(l)}(\cdot) \equiv m^{(l')}(\cdot) \quad \text{versus} \quad H_1: \; m^{(l)}(x) \neq m^{(l')}(x) \;\text{ for some } x \in [0,1],$$

with significance level $\alpha = 0.05$ for $(l, l') = (1,2), (3,4), (1,3), (1,5), (1,6), (3,5)$ respectively, in which $l \neq l' \in \{1,2,3,4,5,6\}$ correspond to $\{O_1, O_2, P_3, P_4, C_3, F_3\}$, so for instance, $m^{(O_1)}(\cdot) \equiv m^{(1)}(\cdot), m^{(C_3)}(\cdot) \equiv m^{(5)}(\cdot)$. The constant function 0 is entirely covered by the SCB in subfigures (a) and (b), indicating that difference between O1 and O2, as well as between P3 and P4 are negligible. Subfigures (d)-(f) show however, that constant function 0 is not entirely covered by SCB, rejecting hypotheses, $H_0: \; m^{(O_1)}(\cdot) \equiv m^{(C_3)}(\cdot)$, $H_0: \; m^{(O_1)}(\cdot) \equiv m^{(F_3)}(\cdot)$ and $H_0: \; m^{(P_3)}(\cdot) \equiv m^{(C_3)}(\cdot)$, thus quantitatively validates that the flickering stimuli elicit visual evoked potentials of higher amplitudes over posterior scalp (where O_1, O_2, P_3, P_4 lie), which is also found by Langer et al. (2017). Subfigure (c) implies the rejection of $H_0: \; m^{(O_1)}(\cdot) \equiv m^{(P_3)}(\cdot)$, suggesting that O_1 is evoked of higher potential

amplitudes and more sensible to visual stimuli than P_3.

Based on Corollary 3, vector mean functions of the six groups are compared by testing the hypothesis that they are all equal, i.e.,

$$H_0 : \boldsymbol{m}_1(\cdot) \equiv \boldsymbol{m}_2(\cdot) = \boldsymbol{m}_3(\cdot) \equiv \boldsymbol{m}_4(\cdot) = \boldsymbol{m}_5(\cdot) \equiv \boldsymbol{m}_6(\cdot), \tag{5.29}$$
$$\text{vs. } H_1 : \boldsymbol{m}_s(x) \neq \boldsymbol{m}_{s'}(x), \text{ for some } 1 \leq s < s' \leq S = 6, \ x \in [0,1].$$

The p-value is $< 10^{-4}$ (i.e., a confidence level of $> 99.99\%$ is needed for the multiple SCR in Corollary 3 to contain $\boldsymbol{d}_{ss'} \equiv 0$ for all pairs $1 \leq s < s' \leq S = 6$), providing strong evidence against H_0. Thus one tests

$$H_0 : \boldsymbol{m}_s(\cdot) \equiv \boldsymbol{m}_{s'}(\cdot), \text{ vs. } H_1 : \boldsymbol{m}_s(x) \neq \boldsymbol{m}_{s'}(x), \text{ for some } x \in [0,1]. \tag{5.30}$$

for all pairs (s, s') to further probe. Figure 4 depicts the p-values, showing no strong evidence against the null for $(s, s') = (1,2), (2,3), (4,5), (5,6)$, while for other pairs (s, s') the p-values are all < 0.019. These findings suggest that significance of EEG mean difference between groups is positively related to age gap of the groups. This fact points to further investigation into the basic sensory excitation by visual stimuli in the brain development.

6. Conclusions

A computationally efficient B-spline estimator is proposed for the mean estimation in multivariate functional panel data. Asymptotic properties of the estimator are established with simultaneous confidence band (SCB) and simultaneous confidence region (SCR) as byproducts, which are versatile inference tools. The SCBs and SCRs have satisfactory performance in simulation studies, and have uncovered illuminating phenomena about Electroencephalogram (EEG) data. The methodology is expected to find wide applications on other physiological data such as event related potential (ERP) and Electrocardiogram (ECG).

Supplementary Materials

The online Supplement provides technical lemmas, detailed proofs of the theorems, and additional figures and tables.

Acknowledgements

This research has been partially supported by National Natural Science Foundation of China award 12171269. The insightful comments of two Referees have led to significant improvement of this work.

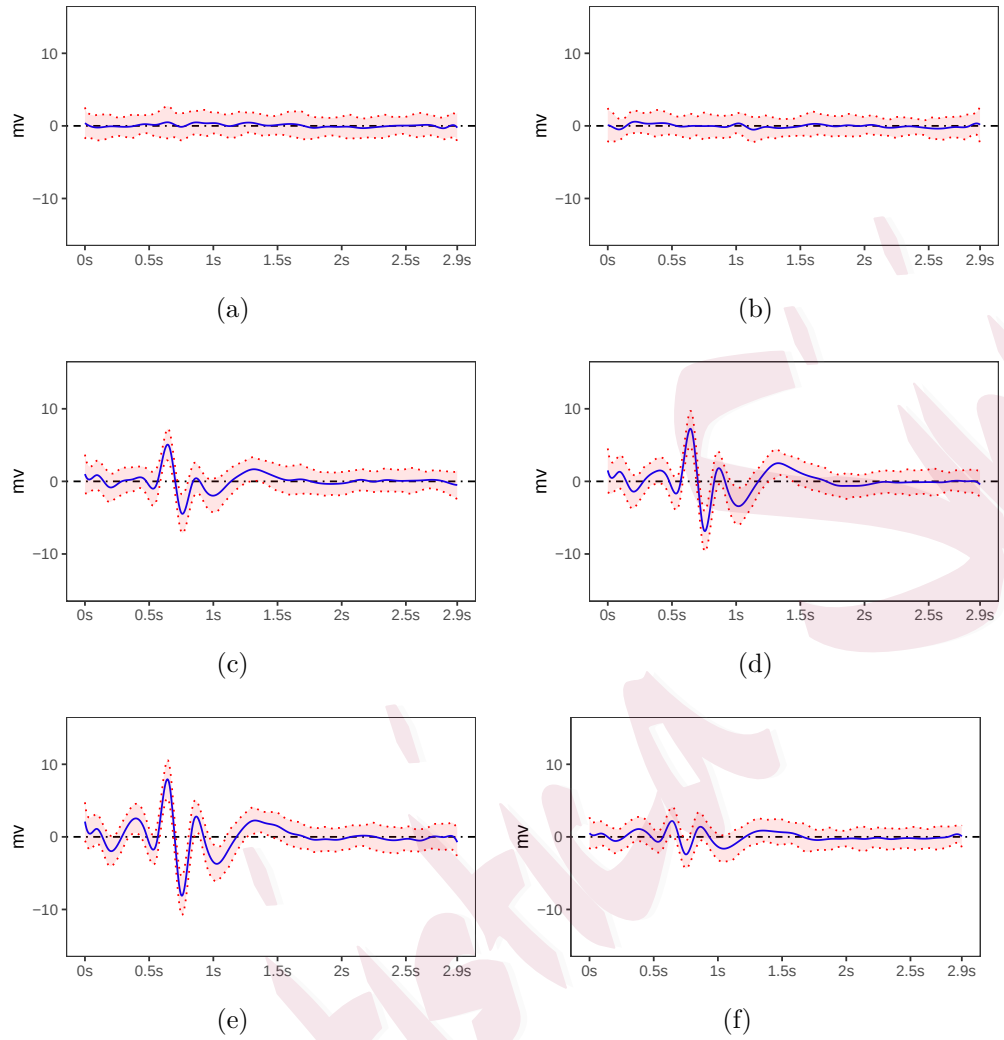


Figure 3: Plots of cubic spline estimator (solid) and 95% SCB (dotted) for $\boldsymbol{b}^{\top}\boldsymbol{m}(\cdot)$ of group 1, with reference constant function 0 (dot-dashed), (a)-(f) correspond to $\boldsymbol{b}^{\top}\boldsymbol{m}(\cdot) = m^{(O_1)}(\cdot) - m^{(O_2)}(\cdot)$, $m^{(P_3)}(\cdot) - m^{(P_4)}(\cdot)$, $m^{(O_1)}(\cdot) - m^{(P_3)}(\cdot)$, $m^{(O_1)}(\cdot) - m^{(C_3)}(\cdot)$, $m^{(O_1)}(\cdot) - m^{(F_3)}(\cdot)$ and $m^{(P_3)}(\cdot) - m^{(C_3)}(\cdot)$ respectively.

References

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

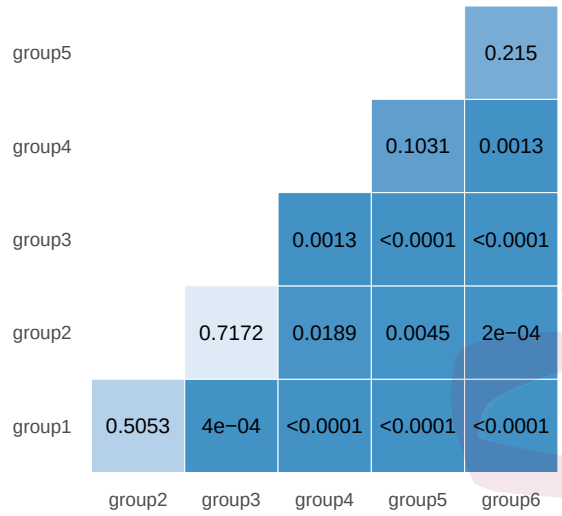


Figure 4: Heat map for p-values of two-sample tests described in (5.30).

Cai, L., L. Li, S. Huang, L. Ma, and L. Yang (2020). Oracally efficient estimation for dense functional data with holiday effects. *TEST* **29** 282–306.

Cao, G. and L. Wang (2018). Simultaneous inference for the mean of repeated functional data. *Journal of Multivariate Analysis* **165**, 279–295.

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

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

Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of Multivariate Analysis* **28**, 20–68.

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

Table 1: Coverage frequencies for $m^{(1)}(\cdot) - m^{(2)}(\cdot) \equiv (1, -1, 0)\boldsymbol{m}(\cdot)$ by SCB (3.27).

$\boldsymbol{\zeta}_{itk}$	$\boldsymbol{\Sigma}_{\zeta,k}$	N	$1-\alpha$	$\varepsilon_{itj}^{(l)} \sim N(0,1)$		$\varepsilon_{itj}^{(l)} \sim \mathrm{U}(-\sqrt{3},\sqrt{3})$	
				$\boldsymbol{\sigma}_{\mathrm{homo}}$	$\boldsymbol{\sigma}_{\mathrm{hetero}}$	$\boldsymbol{\sigma}_{\mathrm{homo}}$	$\boldsymbol{\sigma}_{\mathrm{hetero}}$
Normal	AR-1	900	0.95	0.942	0.945	0.947	0.944
			0.99	0.984	0.985	0.986	0.983
		1600	0.95	0.954	0.953	0.955	0.951
			0.99	0.992	0.993	0.991	0.992
	TOEP	900	0.95	0.935	0.935	0.936	0.935
			0.99	0.984	0.987	0.979	0.985
		1600	0.95	0.946	0.942	0.941	0.943
			0.99	0.988	0.991	0.99	0.99
Student's t	AR-1	900	0.95	0.946	0.944	0.945	0.944
			0.99	0.986	0.986	0.983	0.984
		1600	0.95	0.954	0.955	0.95	0.948
			0.99	0.993	0.99	0.991	0.991
	TOEP	900	0.95	0.934	0.932	0.932	0.933
			0.99	0.979	0.977	0.983	0.982
		1600	0.95	0.943	0.946	0.944	0.946
			0.99	0.994	0.993	0.993	0.992

Götze, F. and A.Y. Zaitsev (2010). Rates of approximation in the multidimensional invariance principle for sums of i.i.d. random vectors with finite moments. *Journal of Mathematical Sciences* **167**, 495–500.

Horváth, L., P. Kokoszka, and R. Reeder (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society Series B* **75**, 103–122.

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

Table 2: Coverage frequencies for $\boldsymbol{m}(\cdot)$ by SCR (3.28).

$\boldsymbol{\zeta}_{itk}$	$\boldsymbol{\Sigma}_{\zeta,k}$	N	$1-\alpha$	$\varepsilon_{itj}^{(l)} \sim N(0,1)$		$\varepsilon_{itj}^{(l)} \sim \mathrm{U}(-\sqrt{3},\sqrt{3})$	
				$\boldsymbol{\sigma}_{\text{homo}}$	$\boldsymbol{\sigma}_{\text{hetero}}$	$\boldsymbol{\sigma}_{\text{homo}}$	$\boldsymbol{\sigma}_{\text{hetero}}$
Normal	AR-1	900	0.95	0.932	0.935	0.928	0.932
			0.99	0.98	0.979	0.98	0.98
		1600	0.95	0.939	0.943	0.939	0.943
			0.99	0.993	0.991	0.992	0.993
	TOEP	900	0.95	0.926	0.93	0.924	0.926
			0.99	0.981	0.98	0.981	0.981
		1600	0.95	0.925	0.928	0.932	0.926
			0.99	0.987	0.987	0.983	0.984
Student's t	AR-1	900	0.95	0.919	0.92	0.921	0.918
			0.99	0.976	0.972	0.971	0.973
		1600	0.95	0.924	0.924	0.921	0.92
			0.99	0.99	0.988	0.991	0.991
	TOEP	900	0.95	0.934	0.932	0.932	0.936
			0.99	0.987	0.987	0.985	0.985
		1600	0.95	0.94	0.939	0.937	0.946
			0.99	0.982	0.982	0.984	0.982

Huang, K., S. Zheng, and L. Yang (2022). Inference for dependent error functional data with application to event-related potentials. *TEST* **31**, 1100-1120.

Kokoszka, P. and M. Reimherr (2017). *Introduction to Functional Data Analysis*. CRC Press, Boca Raton, FL.

Langer, N., E. J. Ho, L. M. Alexander, H. Y. Xu, R. K. Jozanovic, S. Henin, A. Petroni, S. Cohen, E. T. Marcelle, L. C. Parra, M. P. Milham, and S. P. Kelly (2017). A resource for assessing information processing in the developing brain using eeg and eye tracking. *Scientific Data* **4**, 170040.

Li, J. and L. Yang (2023). Statistical inference for functional time series. *Statistica Sinica* **33**, 519–549.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

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

Mohammadi, N. and V. M. Panaretos (2023). Functional data analysis with rough sample paths? *Journal of Nonparametric Statistics*, DOI: 10.1080/10485252.2023.2215347.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*, 2nd edn. Springer, New York.

Yu, S., G. Wang, L. Wang, and L. Yang (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statistica Sinica* **31**, 1463–1487.

Zhang, Y., C. Wang, F. Wu, K. Huang, L. Yang, and L. Ji (2020). Prediction of working memory ability based on eeg by functional data analysis. *Journal of Neuroscience Methods* **333**, 108552.

Zheng, S., L. Yang, and W. K. Härdle (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association* **109**, 661–673.

Zhong, C. and L. Yang (2023). Statistical inference for functional time series: autocovariance

function. *Statistica Sinica* **33**, 2519–2543.

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

E-mail: ss19@mails.tsinghua.edu.cn

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

E-mail: yanglijian@tsinghua.edu.cn