Statistica Sinica Preprint No: SS-2025-0295	
Title	Inference on Two-Sample Covariance Difference for Large-Scale Functional Data
Manuscript ID	SS-2025-0295
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0295
Complete List of Authors	Kaijie Xue, Lan Xue and Riquan Zhang
Corresponding Authors	Riquan Zhang
E-mails	zhangriquan@163.com

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

Inference on two-sample covariance difference for large-scale functional data

Kaijie Xue[1], Lan Xue[2] and Riquan Zhang[1*]

[1]*Shanghai University of International Business and Economics*

[2]*Department of Statistics, Oregon State University, Corvallis Oregon, 97330*

**corresponding author: zhangriquan@163.com*

Abstract: In this study, we introduce an inferential procedure for assessing the covariance difference between two-samples of large-scale functional data, utilizing a computationally efficient multiplier bootstrap approach. In contrast to the existing method that focuses exclusively on a testing procedure, our approach starts by establishing a confidence region for the covariance difference under fairly flexible conditions. This leads not only to a more powerful test but also to an accessible estimated power function, which is shown consistent across a broad range of alternatives. A notable characteristic of the new procedure is that it requires less stringent distributional assumptions for large-scale functional data, and does not impose any structural constraints on covariances or correlations. Moreover, the proposed procedure benefits from the desirable properties of being "eigenvalue-decay-free" and "square-integrable-free", tailored for functional data. We conduct a simulation study and a real data application to assess the numerical performance of the proposed method.

Key words and phrases: Functional data; Ultra-high-dimensionality; Covariance function; Confidence region; Multiplier bootstrap.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

1. Introduction

Functional data are prevalent in various research areas, as noted by Ramsay and Dalzell (1991), Ramsay and Silverman (2005), Cardot et al. (1999), Ferraty and Vieu (2006), Wang et al. (2016), and Poldrack et al. (2024). Typically, functional data is represented by a random function $X(t)$ defined over a compact interval $\mathcal{T} \subset \mathbb{R}$, whose mean and covariance functions are expressed as $\mu(t) = EX(t)$ and $K(s,t) = cov(X(s), X(t))$, respectively. Detecting differences in covariance functions between two samples of functional data is of great interest in applied contexts. A considerable amount of literature (Fremdt et al., 2013; Panaretos et al., 2010; Paparoditis and Sapatinas, 2016; Guo et al., 2019; Pigoli et al., 2014) has focused on comparing two-sample covariance functions for a single functional predictor (i.e., $p = 1$).

However, in many practical scenarios, each individual in a sample may correspond to a large number of functional predictors, implying that the number of relevant functional predictors $p = p_n$ can grow beyond the total sample size. For instance, in neurophysiological research, electroencephalography (EEG) data are commonly gathered through numerous electrodes positioned on the scalp, with each electrode capturing activity from a distinct brain region over time. In many EEG studies, the number of electrodes (acting as functional predictors) is comparable to or even greater than the number of subjects, reflecting the high-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

dimensional nature of the data.

As in the alcoholism EEG data used in our real application, the motivation for comparing two-sample covariance differences, rather than focusing solely on mean differences, is fundamentally linked to the importance of functional connectivity. Functional connectivity refers to the statistical associations or correlations between different brain regions, capturing how different parts of the brain synchronize and work together during various states or tasks. While comparing mean differences between groups (e.g., alcoholics vs. controls) allows researchers to determine whether overall activity in certain brain regions is increased or decreased, this analysis alone is insufficient to fully characterize the complex interplay between regions. For example, alcoholics may exhibit similar mean levels of activity in certain brain regions compared to controls, yet their connectivity patterns could be significantly altered, as reflected by differences in covariance functions. By comparing covariance functions, researchers can identify altered connectivity patterns, network-level dysfunctions, and disruptions that would be missed by mean comparisons alone. Such insights are crucial for diagnosing neurological disorders, guiding targeted interventions, and understanding the mechanisms underlying cognitive and behavioral impairments.

Although there is extensive research on comparing two-sample large-scale covariance matrices (Srivastava and Yanagihara, 2010; Schott, 2007; Chang et al.,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

2017; He and Chen, 2018; Cai et al., 2013; Zou et al., 2021; Wu and Li, 2020, among others), covariance comparisons in functional data remains largely unexplored due to the intricate data structure arising from the inherent infinite-dimensionality of functional data combined with the large-scale setting where $p_n \gg n$. Fang et al. (2024) considers a large-scale multiple testing procedure for functional data; however, it is designed for one-sample scenario under fairly strong restrictions. Besides, multiple testing procedures typically aim to control the false discovery rate across many low-dimensional hypotheses, whereas our focus in the article is on testing a single hypothesis involving high-dimensional parameters, which is of fundamental difference. Xue et al. (2025) did propose a two-sample covariance test designed for large-scale functional data, but it requires rather stringent conditions on correlation structures and distributional types, making them difficult to verify and prone to violations in practical applications. Moreover, Xue et al. (2025) cannot be regarded as an inferential procedure in the strict sense, as it merely provides a testing method and lacks a theoretical framework for confidence region or power function assessment. This inspires us to develop an inferential procedure or confidence region for the two-sample covariance difference in large-scale functional data, accommodating more general correlation structures and distributional assumptions, ultimately resulting in a significantly more powerful testing method.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

Remarkably, multiplier bootstrap-based inferences (Xue, 2023; Xue and Yao, 2020, 2021, 2024; Chernozhukov et al., 2013, among others) have been demonstrated to considerably ease model constraints and enhance numerical performance in more complex scenarios. Motivated by this, in this article, we intend to develop a multiplier bootstrap-based inferential procedure that aims to construct a data-driven confidence region for the two-sample covariance difference of large-scale functional data, leading to a fairly effective inferential procedure involving both testing and power function assessment.

Our primary accomplishment in this study is to establish a confidence region for the covariance difference functions applicable to large-scale functional data. This leads to a fairly effective inferential procedure involving both a testing method and power function assessment for evaluating two-sample covariance differences in such data. In particular, although hypothesis testing and confidence region construction are closely related, they are not on equal footing. A confidence region naturally induces a testing procedure, so the information conveyed by the former subsumes that of the latter. In practice, a testing procedure can only be used to assess the validity of a specific null hypothesis, whereas a confidence region allows one to determine whether any parameter value of interest is compatible with the data at a prescribed confidence level. Nonetheless, the test induced by a confidence region occupies a central role in inference, much

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

as the relatively better-known classical t-test actually arises as the testing counterpart of the usual confidence interval for the mean. Notably, the proposed inferential procedure features several key advantages that set it apart from existing work. The first key advantage is that, unlike our previous work Xue et al. (2025), the theoretical validity of our newly proposed procedure does not demand any structural conditions on the population covariances. It remains valid under relatively weak moment assumptions. The second advantage arises from the more relaxed distributional assumptions on functional data, as opposed to the previously mentioned work. In general, stringent distributional assumptions are not only challenging to verify but are also frequently violated in real-world data. The third major advantage is that even under weaker moment and distributional assumptions, the new procedure still retains all the desirable properties from our previous work, such as being 'eigenvalue-decay-free', 'square-integrable-free', allowing exponentially growing p_n, and accommodating a diverging sample size ratio. As noted in Xue et al. (2025), the 'eigenvalue-decay-free' property signifies that there are no decay constraints imposed on the eigenvalues of the covariance functions. In contrast, most traditional functional data analysis (Hall and Hosseini-Nasab, 2006; Hall and Horowitz, 2007) typically requires some form of eigenvalue decay restrictions. Additionally, the 'square-integrable-free' property relaxes the conventional square-integrable condition on functional data.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

The rest of the article is structured as follows. Section 2 first introduces the mathematical notations and then discusses the inferential procedure based on Theorems 1–3. Specifically, Theorems 1 focuses on establishing a confidence region for the covariance difference, which leads to the new testing procedure. Theorems 2–3 then investigate the power assessment and analysis of the test. Section 3 compares the finite sample performance of the proposed test with its competitive method through a simulation study. Section 4 applies the proposed test to a real dataset. To save space, the auxiliary lemmas with proofs and the proofs of theorems are included in an online Supplement.

2. Inference on the covariance difference

Before we delve into the inferential procedure, it is helpful to formulate the related testing problem in advance. Mathematically, we examine two independent random samples, $\mathcal{X}^{n_1} = \{X_i(\cdot) \in \mathbb{R}^{p_n} : i \leq n_1\}$ and $\mathcal{Y}^{n_2} = \{Y_i(\cdot) \in \mathbb{R}^{p_n} : i \leq n_2\}$, consisting of p_n-dimensional functional data defined over a common compact interval $\mathcal{T}$. Each individual data point is written as $X_i(\cdot) = (X_{i1}(\cdot), \ldots, X_{ip_n}(\cdot))'$ and $Y_i(\cdot) = (Y_{i1}(\cdot), \ldots, Y_{ip_n}(\cdot))'$, where the data dimension $p = p_n$ is permitted to vary with the total sample size $n = n_1 + n_2$. More specifically, the first sample $\mathcal{X}^{n_1}$ consists of n_1 independent and identically distributed ($i.i.d.$) copies with a shared mean function $\mu^X(t) = (\mu_1^X(t), \ldots, \mu_{p_n}^X(t))' \in$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

$\mathbb{R}^{p_n}$ and covariance functions $K^X_{j_1j_2}(s,t) = cov(X_{ij_1}(s), X_{ij_2}(t))$ for all $j_1, j_2 \in \{1, \ldots, p_n\}$. The second sample $\mathcal{Y}^{n_2}$ also contains $i.i.d.$ observations with a mean function $\mu^Y(t) = (\mu^Y_1(t), \ldots, \mu^Y_{p_n}(t))' \in \mathbb{R}^{p_n}$ and covariance functions $K^Y_{j_1j_2}(s,t) = cov(Y_{ij_1}(s), Y_{ij_2}(t))$. Remarkably, in this work, we adopt the dense design where the within-curve observation of functional data are densely observed over the grid points, ensuring a sufficiently accurate characterization of the functional data for subsequent inference.

Our primary goal is to develop a flexible inferential method for the covariance difference functions $K^X - K^Y = \{K^X_{j_1j_2} - K^Y_{j_1j_2} : 1 \le j_1, j_2 \le p_n\}$ and construct a superior test to evaluate the following hypotheses.

$$H_0 : \sup_{s\in\mathcal{T}} \sup_{t\in\mathcal{T}} \max_{j_1\le p_n} \max_{j_2\le p_n} |K^X_{j_1j_2}(s,t) - K^Y_{j_1j_2}(s,t)| = 0 \qquad \text{versus}$$

$$H_1 : \sup_{s\in\mathcal{T}} \sup_{t\in\mathcal{T}} \max_{j_1\le p_n} \max_{j_2\le p_n} |K^X_{j_1j_2}(s,t) - K^Y_{j_1j_2}(s,t)| \neq 0, \tag{2.1}$$

where the null hypothesis indicates that there is no difference in the covariance functions between the two samples of multivariate functional data.

Building on the rationale provided in Xue and Yao (2021) and Xue et al. (2024), we adopt a common predetermined basis $\{b_k : k \ge 1\}$ that is both orthonormal and complete in $L^2(\mathcal{T})$ to represent the functional data, rather than data-driven bases like eigenfunctions that require the estimation through p_n sep-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

arate univariate functional principal component analysis (FPCA) procedures–an approach that becomes computationally prohibitive when $p_n \gg n$. Consequently, every functional data and their associated mean functions are expressed with the basis representations as

$$X_{ij} = \sum_{k=1}^{\infty} \theta_{1,ijk} b_k, \quad \mu_j^X = \sum_{k=1}^{\infty} \eta_{1,jk} b_k, \quad Y_{ij} = \sum_{k=1}^{\infty} \theta_{2,ijk} b_k, \quad \mu_j^Y = \sum_{k=1}^{\infty} \eta_{2,jk} b_k,$$

with $\eta_{1,jk} = E(\theta_{1,ijk})$ and $\eta_{2,jk} = E(\theta_{2,ijk})$. Let $\theta^*_{1,ij} = (\theta_{1,ij1}, \theta_{1,ij2}, \dots)' \in \mathbb{R}^\infty$, and $\theta^*_{1,i} = (\theta^{*'}_{1,i1}, \dots, \theta^{*'}_{1,ip_n})'$ consist of all projected coefficients for the functional data observed on the i-th individual in the first sample. The means of these sequences are denoted by $\eta^*_{1,j} = E(\theta^*_{1,ij})$ and $\eta^*_1 = E(\theta^*_{1,i})$ respectively. Likewise, we use $\theta^*_{2,i}$, $\theta^*_{2,ij}$, η^*_2, $\eta^*_{2,j}$ to denote the corresponding terms for the second sample.

Accordingly, the covariance functions K^X and K^Y can be fully characterized by the infinite-dimensional covariance structures $\Sigma^{*X} = cov(\theta^*_{1,i})$ and $\Sigma^{*Y} = cov(\theta^*_{2,i})$, where testing (2.1) is equivalent to testing $H_0 : \Sigma^{*X} = \Sigma^{*Y}$. However, it is impossible to utilize the complete infinite-dimensional functional data $\{\theta^*_{1,i} \in \mathbb{R}^\infty : i \le n_1\}$ and $\{\theta^*_{2,i} \in \mathbb{R}^\infty : i \le n_2\}$ to construct an inferential procedure for covariance difference $\Sigma^{*X} - \Sigma^{*Y} \in \mathbb{R}^{\infty \times \infty}$. To address this limitation, it is typically supposed to utilize only the first $s = s_n$ truncated

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

terms $\{\theta_{1,ij1}, \ldots, \theta_{1,ijs_n}\}$. Such truncation practice has been widely adopted by many literature (Yao et al., 2005; Kong et al., 2016; Dai et al., 2017; Delaigle and Hall, 2012; Fan et al., 2015). Accordingly, the truncated form of the first sample $\mathcal{X}^{n_1}$ is represented as $\tilde{\mathcal{X}}^{n_1} = \{\theta_{1,i} \in \mathbb{R}^{p_n s_n} : i \leq n_1\}$, where the vector $\theta_{1,i} = (\theta'_{1,i1}, \ldots, \theta'_{1,ip_n})'$, with each sub-vector $\theta_{1,ij} = (\theta_{1,ij1}, \ldots, \theta_{1,ijs_n})' \in \mathbb{R}^{s_n}$. Likewise, the truncated version of the second sample $\mathcal{Y}^{n_2}$ is denoted by $\tilde{\mathcal{Y}}^{n_2} = \{\theta_{2,i} \in \mathbb{R}^{p_n s_n} : i \leq n_2\}$. We also express the truncated form of η_1^* as $\eta_1 = (\eta'_{1,1}, \ldots, \eta'_{1,p_n})'$, where each sub-vector $\eta_{1,j} = (\eta_{1,j1}, \ldots, \eta_{1,js_n})' \in \mathbb{R}^{s_n}$. A similar notation applies to η_2. Therefore, we designate a series of covariance matrices and variances for the truncated data as

$$\Sigma^X_{j_1 j_2} = cov(\theta_{1,ij_1}, \theta_{1,ij_2}) = (\sigma_{1,j_1 j_2 k_1 k_2})_{k_1 \leq s_n, k_2 \leq s_n} \in \mathbb{R}^{s_n \times s_n},$$

$$\Sigma^Y_{j_1 j_2} = cov(\theta_{2,ij_1}, \theta_{2,ij_2}) = (\sigma_{2,j_1 j_2 k_1 k_2})_{k_1 \leq s_n, k_2 \leq s_n} \in \mathbb{R}^{s_n \times s_n}.$$

In addition, $\Sigma^X = cov(\theta_{1,i}) = (\Sigma^X_{j_1 j_2})_{j_1 \leq p_n, j_2 \leq p_n} \in \mathbb{R}^{p_n s_n \times p_n s_n}$, $\Sigma^Y = cov(\theta_{2,i}) = (\Sigma^Y_{j_1 j_2})_{j_1 \leq p_n, j_2 \leq p_n} \in \mathbb{R}^{p_n s_n \times p_n s_n}$, $s_{1,j_1 j_2 k_1 k_2} = var\{(\theta_{1,ij_1 k_1} - \eta_{1,j_1 k_1})(\theta_{1,ij_2 k_2} - \eta_{1,j_2 k_2})\}$, and $s_{2,j_1 j_2 k_1 k_2} = var\{(\theta_{2,ij_1 k_1} - \eta_{2,j_1 k_1})(\theta_{2,ij_2 k_2} - \eta_{2,j_2 k_2})\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

We consider the following moment estimators of the above quantities

$$\hat{\Sigma}^X_{j_1j_2} = n_1^{-1}\sum_{i=1}^{n_1}(\theta_{1,ij_1} - \hat{\eta}_{1,j_1})(\theta_{1,ij_2} - \hat{\eta}_{1,j_2})' = (\hat{\sigma}_{1,j_1j_2k_1k_2})_{k_1\le s_n,k_2\le s_n} \in \mathbb{R}^{s_n\times s_n},$$
$$\hat{\Sigma}^Y_{j_1j_2} = n_2^{-1}\sum_{i=1}^{n_2}(\theta_{2,ij_1} - \hat{\eta}_{2,j_1})(\theta_{2,ij_2} - \hat{\eta}_{2,j_2})' = (\hat{\sigma}_{2,j_1j_2k_1k_2})_{k_1\le s_n,k_2\le s_n} \in \mathbb{R}^{s_n\times s_n},$$

with elements $\hat{\sigma}_{1,j_1j_2k_1k_2} = n_1^{-1}\sum_{i=1}^{n_1}(\theta_{1,ij_1k_1} - \hat{\eta}_{1,j_1k_1})(\theta_{1,ij_2k_2} - \hat{\eta}_{1,j_2k_2})$ and $\hat{\sigma}_{2,j_1j_2k_1k_2} = n_2^{-1}\sum_{i=1}^{n_2}(\theta_{2,ij_1k_1} - \hat{\eta}_{2,j_1k_1})(\theta_{2,ij_2k_2} - \hat{\eta}_{2,j_2k_2})$, where $\hat{\eta}_1 = n_1^{-1}\sum_{i=1}^{n_1}\theta_{1,i}$ and $\hat{\eta}_2 = n_2^{-1}\sum_{i=1}^{n_2}\theta_{2,i}$ represent the sample means. Consequently, consider $\hat{\Sigma}^X = (\hat{\Sigma}^X_{j_1j_2})_{j_1\le p_n,j_2\le p_n} \in \mathbb{R}^{p_ns_n\times p_ns_n}$ and $\hat{\Sigma}^Y = (\hat{\Sigma}^Y_{j_1j_2})_{j_1\le p_n,j_2\le p_n} \in \mathbb{R}^{p_ns_n\times p_ns_n}$. In addition,

$$\hat{s}_{1,j_1j_2k_1k_2} = n_1^{-1}\sum_{i=1}^{n_1}\{(\theta_{1,ij_1k_1} - \hat{\eta}_{1,j_1k_1})(\theta_{1,ij_2k_2} - \hat{\eta}_{1,j_2k_2}) - \hat{\sigma}_{1,j_1j_2k_1k_2}\}^2,$$
$$\hat{s}_{2,j_1j_2k_1k_2} = n_2^{-1}\sum_{i=1}^{n_2}\{(\theta_{2,ij_1k_1} - \hat{\eta}_{2,j_1k_1})(\theta_{2,ij_2k_2} - \hat{\eta}_{2,j_2k_2}) - \hat{\sigma}_{2,j_1j_2k_1k_2}\}^2.$$

Recall that we represent the collection of covariance difference functions as $K^X - K^Y = \{K^X_{j_1j_2} - K^Y_{j_1j_2} : j_1 \le p_n, j_2 \le p_n\}$. Accordingly, we define

$$F_{\{b_k:k\le s_n\}}(K^X - K^Y) = \Sigma^X - \Sigma^Y,$$

as the function that projects $K^X - K^Y$ onto $\Sigma^X - \Sigma^Y$. Then, define a random

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

function $\hat{G}(\Sigma^X - \Sigma^Y)$ as

$$\hat{G}(\Sigma^X - \Sigma^Y) = \hat{G}(F_{\{b_k:k\le s_n\}}(K^X - K^Y))$$
$$= \max_{j_1\le p_n}\max_{j_2\le p_n}\max_{k_1\le s_n}\max_{k_2\le s_n}\frac{|(\hat{\sigma}_{1,j_1j_2k_1k_2} - \hat{\sigma}_{2,j_1j_2k_1k_2}) - (\sigma_{1,j_1j_2k_1k_2} - \sigma_{2,j_1j_2k_1k_2})|}{(n_1^{-1}\hat{s}_{1,j_1j_2k_1k_2} + n_2^{-1}\hat{s}_{2,j_1j_2k_1k_2})^{1/2}}, \tag{2.2}$$

which is employed to construct a confidence region for $K^X - K^Y$. It is worth to note that expression (2.2) essentially represents the ℓ_∞-norm of a $p_n^2 s_n^2$-dimensional standardized vector capturing the covariance difference. The ℓ_∞-statistic is commonly used in large-scale comparison scenarios, particularly when the underlying signal is relatively sparse (Cai et al., 2014, etc.). Due to essential differences in derivation approaches, we shall consider developing alternative types of test statistics, such as ℓ_2-statistics (Han and Wu, 2020, etc.) and threshold-type statistics (Donoho and Jin, 2004, etc.), in future work. To present the cutoff value for inference, we first introduce a random variable $\hat{G}_e$ as

$$\hat{G}_e = \max_{j_1\le p_n}\max_{j_2\le p_n}\max_{k_1\le s_n}\max_{k_2\le s_n}(n_1^{-1}\hat{s}_{1,j_1j_2k_1k_2} + n_2^{-1}\hat{s}_{2,j_1j_2k_1k_2})^{-1/2}\cdot$$
$$\Big|n_1^{-1}\sum_{i=1}^{n_1} e_i\{(\theta_{1,ij_1k_1} - \hat{\eta}_{1,j_1k_1})(\theta_{1,ij_2k_2} - \hat{\eta}_{1,j_2k_2}) - \hat{\sigma}_{1,j_1j_2k_1k_2}\} -$$
$$n_2^{-1}\sum_{i=1}^{n_2} e_{i+n_1}\{(\theta_{2,ij_1k_1} - \hat{\eta}_{2,j_1k_1})(\theta_{2,ij_2k_2} - \hat{\eta}_{2,j_2k_2}) - \hat{\sigma}_{2,j_1j_2k_1k_2}\}\Big|,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

with $e = \{e_1, \ldots, e_n\}$ standing for a set of $i.i.d.$ standard normals, independent of the data. The cutoff value is set as the $(1-\alpha)$th quantile of $\hat{G}_e$, defined by

$$c_B(\alpha) = \inf\{t \in \mathbb{R} : P_e(\hat{G}_e \leq t) \geq 1-\alpha\}, \qquad \alpha \in (0,1) \qquad (2.3)$$

where $P_e(\cdot)$ refers to the conditional probability concerning e alone. Remarkably, the value of $c_B(\alpha)$ can be rapidly obtained using a multiplier bootstrap procedure on e. Theorem 1 below provides the theoretical foundation for the proposed inferential procedure, based on more relaxed regularity conditions (A1)–(A4) outlined in the Theorem.

Theorem 1. *Suppose the following conditions (A1)–(A4) hold:*

(A1). $n_1/(n_1+n_2) \in (c_1, c_2)$, *for some universal constants* $0 < c_1 \leq c_2 < 1$.

(A2). $n^{-1}\log^9(np_ns_n) \to 0$, *as* $n \to \infty$.

(A3). *There exist universal constants* $K_1, K_2 > 0$ *such that*

$$\max_{j\leq p_n}\max_{k\leq s_n} E[\exp\{K_1\sigma_{1,jjkk}^{-1}(\theta_{1,ijk}-\eta_{1,jk})^2\}] \leq K_2,$$

$$\max_{j\leq p_n}\max_{k\leq s_n} E[\exp\{K_1\sigma_{2,jjkk}^{-1}(\theta_{2,ijk}-\eta_{2,jk})^2\}] \leq K_2.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

(A4). *There exists a universal constant* $c > 0$ *such that*

$$\min_{j_1\le p_n}\min_{j_2\le p_n}\min_{k_1\le s_n}\min_{k_2\le s_n}\min_{\tau_1\in\{1,2\}}\min_{\tau_2\in\{1,2\}} var\Big\{\Big(\frac{\theta_{1,ij_1k_1}-\eta_{1,j_1k_1}}{\sigma^{1/2}_{1,j_1j_1k_1k_1}}\Big)^{\tau_1}\Big(\frac{\theta_{1,ij_2k_2}-\eta_{1,j_2k_2}}{\sigma^{1/2}_{1,j_2j_2k_2k_2}}\Big)^{\tau_2}\Big\}\ge c,$$

$$\min_{j_1\le p_n}\min_{j_2\le p_n}\min_{k_1\le s_n}\min_{k_2\le s_n}\min_{\tau_1\in\{1,2\}}\min_{\tau_2\in\{1,2\}} var\Big\{\Big(\frac{\theta_{2,ij_1k_1}-\eta_{2,j_1k_1}}{\sigma^{1/2}_{2,j_1j_1k_1k_1}}\Big)^{\tau_1}\Big(\frac{\theta_{2,ij_2k_2}-\eta_{2,j_2k_2}}{\sigma^{1/2}_{2,j_2j_2k_2k_2}}\Big)^{\tau_2}\Big\}\ge c.$$

Then, the Kolmogorov distance between the distributions of $\hat{G}(F_{\{b_k:k\le s_n\}}(K^X - K^Y))$ *and* $\hat{G}_e$ *satisfies* $\lim_{n\to\infty}\sup_{t\ge 0}\big|P\{\hat{G}(F_{\{b_k:k\le s_n\}}(K^X-K^Y))\le t\} - P_e(\hat{G}_e\le t)\big| = 0$*, and consequently,*

$$\lim_{n\to\infty}\sup_{\alpha\in(0,1)}\big|P\{\hat{G}(F_{\{b_k:k\le s_n\}}(K^X-K^Y))\le c_B(\alpha)\} - (1-\alpha)\big| = 0.$$

Remember that the terms $\hat{G}(F_{\{b_k:k\le s_n\}}(K^X - K^Y))$ and $c_B(\alpha)$ are defined in (2.2) and (2.3), respectively. According to Theorem 1, an asymptotically valid $100(1-\alpha)\%$ confidence region for $K^X - K^Y$ can be expressed as

$$CR_{1-\alpha} = \{K^X - K^Y : \hat{G}(F_{\{b_k:k\le s_n\}}(K^X-K^Y))\le c_B(\alpha)\}.$$

This indicates that the associated testing procedure involves rejecting the null hypothesis H_0 defined by (2.1) at a nominal level $\alpha\in(0,1)$ if $\hat{G}(0) > c_B(\alpha)$.

In practical terms, the truncation size s_n for the aforementioned test is de-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

termined from a range of values to achieve the least possible maximum absolute deviation between the fitted and raw functional data across all grid points. The inferential procedure induced by Theorem 1 is more preferred due to its rather relaxed conditions (A1)–(A4). Condition (A1) demands comparable sample sizes, which is standard for two-sample high-dimensional functional data problems (e.g., condition (a) in Xue, 2023). Condition (A2) is typical in large-scale settings, where the data dimension p_n may either stay fixed or increase exponentially with $n = n_1 + n_2$. Condition (A3) pertains to the distribution type of the truncated samples $\tilde{\mathcal{X}}^{n_1}$ and $\tilde{\mathcal{Y}}^{n_2}$, where only some mild exponential-type tails are assumed. In comparison, the comparable method not only strictly requires sub-Gaussian observations, but also demands some difficult-to-verify parametric shape moment restrictions on the data distributions, like condition (d) in Xue et al. (2025). Condition (A4) guarantees that the products of certain related random variables are non-degenerate, which is a standard requirement for two-sample covariance testing. For example, this condition aligns with the principle of (7) in Cai et al. (2013). It is apparent from conditions (A1)–(A4) that no structural constraints on covariances or correlations are required for the new method, in contrast to condition (b) in Xue et al. (2025). In addition, the proposed new procedure permits either a fixed or diverging truncation size s_n, whereas the rival method strictly demands $s_n \to \infty$. Notably, due to

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

its more flexible conditions, the new method also inherits the desired advantages of being "eigenvalue-decay-free" and "square-integrable-free", as specified in Xue et al. (2025). More precisely, the "eigenvalue-decay-free" property signifies that there are no decay constraints imposed on the eigenvalues $\sigma_{1,jjkk}$ or $\sigma_{2,jjkk}$. In contrast, most traditional functional data analysis (Hall and Hosseini-Nasab, 2006; Hall and Horowitz, 2007) typically requires some form of eigenvalue decay restrictions, such as $\sigma_{1,jj11} \geq \sigma_{1,jj22} \geq \cdots \geq 0$ or $\sigma_{1,jjkk} - \sigma_{1,jjk+1,k+1} \gtrsim k^{-a-1}$. Additionally, the "square-integrable-free" property indicates that there is no requirement to adhere to the conventional square-integrable condition: $\int_{\mathcal{T}} E(X_j^2(t))dt < \infty$, or equivalently $\sum_{k=1}^{\infty} \sigma_{1,jjkk} < \infty$. This permits the extreme case where $\int_{\mathcal{T}} E(X_j^2(t))dt = \infty$ or $\sum_{k=1}^{\infty} \sigma_{1,jjkk} = \infty$.

Considering the assessment of power at the specified actual difference $K^X - K^Y$, the new testing method implies that the true power function is expressed as

$$power(K^X - K^Y) = P\{\|\hat{G}(0)\|_\infty > c_B(\alpha)|K^X - K^Y\}.$$

Nonetheless, evaluating the exact power is prohibitive because the distribution of $\|\hat{G}(0)\|_\infty$ is unknown. To tackle this problem, the primary strategy is to create a suitable approximation of $power(K^X - K^Y)$. Specifically, in line with the approach of Theorem 1, the proposed solution involves using a separate multiplier

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

bootstrap process to emulate the distribution of $\|\hat{G}(0)\|_\infty$, thereby estimating the real power. To this end, denoting a random function in $K^X - K^Y$ as

$$\begin{aligned}\hat{H}_e(K^X - K^Y) = \max_{j_1 \le p_n} \max_{j_2 \le p_n} \max_{k_1 \le s_n} \max_{k_2 \le s_n} (n_1^{-1}\hat{s}_{1,j_1j_2k_1k_2} + n_2^{-1}\hat{s}_{2,j_1j_2k_1k_2})^{-1/2} \times \\ \Big| n_1^{-1} \sum_{i=1}^{n_1} e_i \{(\theta_{1,ij_1k_1} - \hat{\eta}_{1,j_1k_1})(\theta_{1,ij_2k_2} - \hat{\eta}_{1,j_2k_2}) - \hat{\sigma}_{1,j_1j_2k_1k_2}\} - \\ n_2^{-1} \sum_{i=1}^{n_2} e_{i+n_1} \{(\theta_{2,ij_1k_1} - \hat{\eta}_{2,j_1k_1})(\theta_{2,ij_2k_2} - \hat{\eta}_{2,j_2k_2}) - \hat{\sigma}_{2,j_1j_2k_1k_2}\} \\ + (\sigma_{1,j_1j_2k_1k_2} - \sigma_{2,j_1j_2k_1k_2}) \Big|,\end{aligned}$$

we then define the estimated power function as

$$power^*(K^X - K^Y) = P_{e^*}\{\hat{H}_{e^*}(K^X - K^Y) > c_B(\alpha)\}, \tag{2.4}$$

where $e^* = \{e_1^*, \ldots, e_n^*\}$ represents an independent replication of the e that is utilized to generate $c_B(\alpha)$. Note that the calculation of (2.4) can be accomplished through a multiplier bootstrap procedure on e^*, wherein the distribution of $\hat{H}_{e^*}(K^X - K^Y)$ is employed to approximate that of $\|\hat{G}(0)\|_\infty$. Given the same conditions as in Theorem 1, Theorem 2 below demonstrates the asymptotic equivalence between $power(K^X - K^Y)$ and $power^*(K^X - K^Y)$.

Theorem 2. *Under conditions (A1)–(A4), given the genuine difference* $K^X -$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

K^Y, *it holds that:* $\lim_{n\to\infty} |power(K^X - K^Y) - power^*(K^X - K^Y)| = 0$.

For the analysis of power, the asymptotic consistency of $power^*(K^X - K^Y)$ is assured for a broad range of alternatives within the set $\mathcal{F}_n$ as detailed in Theorem 3 below.

Theorem 3. *Assume conditions (A1)–(A4), and define an alternative set* $\mathcal{F}_n$ *as*

$$\mathcal{F}_n = \Big\{K^X - K^Y : \max_{j_1\le p_n}\max_{j_2\le p_n}\max_{k_1\le s_n}\max_{k_2\le s_n}\frac{|\sigma_{1,j_1j_2k_1k_2}-\sigma_{2,j_1j_2k_1k_2}|}{(s_{1,j_1j_2k_1k_2}+n_1n_2^{-1}s_{2,j_1j_2k_1k_2})^{1/2}} \geq Kn^{-1/2}\log^{1/2}(np_ns_n)\Big\},$$

for a sufficiently large universal constant $K > 0$. *If the actual covariance difference* $K^X - K^Y \in \mathcal{F}_n$, *then it holds that* $power^*(K^X - K^Y) \to 1$, *as* $n \to \infty$.

Based on Theorem 3, it can be observed that the power function approaches 1 without requiring any sparsity assumptions on the differences $K^X - K^Y$. This feature significantly broadens the method's applicability.

3. Simulation Study

In this section, we evaluate the finite sample performance of the new testing procedure (hereafter referred to as XXZ) in comparison to the existing rival procedure introduced by Xue et al. (2025) (referred to as XYZ hereafter) across various settings. The nominal significance level is set as $\alpha = 0.05$, and the proposed XXZ test is executed using a multiplier bootstrap with a resampling size

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

of $N = 10^3$. Specifically, we examine the large-scale settings: $(n_1, n_2, p_n) = (80, 120, 240)$, $(n_1, n_2, p_n) = (70, 105, 240)$, and $(n_1, n_2, p_n) = (60, 90, 240)$. These settings encompass a wide range of smaller sample sizes relative to the data dimension p_n, with n_1 and n_2 being notably different in each case. The two random samples $\mathcal{X}^{n_1} = \{X_i(\cdot) \in \mathbb{R}^{p_n} : i \leq n_1\}$ and $\mathcal{Y}^{n_2} = \{Y_i(\cdot) \in \mathbb{R}^{p_n} : i \leq n_2\}$ are generated over the same interval $\mathcal{T} = [0, 1]$. Assuming without affecting generality, we set both mean functions as zero, i.e., $\mu^X(\cdot) = \mu^Y(\cdot) = 0 \in \mathbb{R}^{p_n}$. To this end, we define $\{\phi_k(\cdot) : k \geq 1\}$ as a complete orthonormal Fourier basis on $\mathcal{T}$, where $\phi_1 = 1$, $\phi_{2\ell} = 2^{1/2}\cos\{\ell\pi(2t-1)\}$ for $\ell = 1, 2, \ldots, \infty$, and $\phi_{2\ell-1} = 2^{1/2}\sin\{(\ell-1)\pi(2t-1)\}$ for $\ell = 2, 3, \ldots, \infty$. To simulate the functional data $\{X_{ij}(\cdot) : i \leq n_1, j \leq p_n\}$ from $\mathcal{X}^{n_1}$, the initial step involves generating a collection of $i.i.d.$ random processes $\{U_{ij}(\cdot) : i \leq n_1, j \leq p_n\}$ over $\mathcal{T}$ as $U_{ij}(t) = \sum_{k=1}^{50} \xi_{1,ijk}\phi_k(t)$, where $\{\xi_{1,ijk} : i \leq n_1, j \leq p_n, k \leq 50\}$ represents a collection of independent random variables, distributed as

$$\xi_{1,ijk} \sim \begin{cases} (8\nu_k^{1/2})^{-1/2}\{\chi^2(4) - 4\} & \text{if } 1 \leq j \leq \lfloor p_n/4 \rfloor, \\ (5\nu_k^{1/2}/3)^{-1/2}t(5) & \text{if } \lfloor p_n/4 \rfloor + 1 \leq j \leq \lfloor p_n/2 \rfloor, \\ (8\nu_k^{1/2})^{-1/2}\{\chi^2(4) - 4\} & \text{if } \lfloor p_n/2 \rfloor + 1 \leq j \leq \lfloor 3p_n/4 \rfloor, \\ (5\nu_k^{1/2}/3)^{-1/2}t(5) & \text{if } \lfloor 3p_n/4 \rfloor + 1 \leq j \leq p_n, \end{cases}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

with mean $E(\xi_{1,ijk}) = 0$ and variance $var(\xi_{1,ijk}) = \nu_k^{-1/2}$. The sequence $\nu_1, \ldots, \nu_{50}$ signifies a predetermined random permutation of the numbers from 1 to 50, and it is kept unchanged throughout all the simulations. We further create a set $\{\psi_j : j \leq 50\}$ of independent random variables distributed as uniform$(1, 2)$, which stay unchanged for subsequent simulations. The functional data $\{X_{ij}(\cdot) : i \leq n_1, j \leq p_n\}$ are then determined by the autoregressive relationship among $\{U_{ij}(\cdot) : i \leq n_1, j \leq p_n\}$, exhibiting a polynomial decay pattern. More precisely, we define

$$X_{ij}(t) = \sum_{j'=1}^{p_n} \psi_{j'}(1 + |j - j'|)^{-3} U_{ij'}(t) = \sum_{k=1}^{50} \sum_{j'=1}^{p_n} \psi_{j'}(1 + |j - j'|)^{-3} \xi_{1,ij'k} \phi_k(t)$$
$$= \sum_{k=1}^{50} \theta_{1,ijk} \phi_k(t),$$

where the projected scores are given by $\theta_{1,ijk} = \sum_{j'=1}^{p_n} \psi_{j'}(1 + |j - j'|)^{-3} \xi_{1,ij'k}$. Denoting the following matrices:

$$\Sigma = (\Sigma_{j_1 j_2})_{j_1 \leq p_n, j_2 \leq p_n} \quad \text{with each } \Sigma_{j_1 j_2} = (1 + |j_1 - j_2|)^{-3} I_{50\times 50},$$

$$\Omega = diag\{\Omega_j : j \leq p_n\} \quad \text{with each } \Omega_j = diag\{\nu_1^{-1/2}, \ldots, \nu_{50}^{-1/2}\},$$

$$\Phi = diag\{\Phi_j : j \leq p_n\} \quad \text{with each } \Phi_j = \psi_j I_{50\times 50},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

it is straightforward to verify that the covariance structure $\Sigma^{*X} = cov(\theta^*_{1,i})$ can be equivalently represented as $\Sigma^{*X} = \Sigma\Phi\Omega\Phi\Sigma$.

Similarly, to generate the second sample $\{Y_{ij}(\cdot) : i \leq n_2, j \leq p_n\}$, we begin by simulating a set of $i.i.d.$ random processes $\{V_{ij}(\cdot) : i \leq n_2, j \leq p_n\}$ over $\mathcal{T}$ as $V_{ij}(t) = \sum_{k=1}^{50} \xi_{2,ijk}\phi_k(t)$, with $\{\xi_{2,ijk} : i \leq n_2, j \leq p_n, k \leq 50\}$ representing a collection of independent random variables, distributed as

$$\xi_{2,ijk} \sim \begin{cases} (8\nu_k^{1/2})^{-1/2}\{\chi^2(4) - 4\} & \text{if } 1 \leq j \leq \lfloor p_n/4 \rfloor, \\ (5\nu_k^{1/2}/3)^{-1/2}t(5) & \text{if } \lfloor p_n/4 \rfloor + 1 \leq j \leq \lfloor p_n/2 \rfloor, \\ (8\nu_k^{1/2})^{-1/2}\{\chi^2(4) - 4\} & \text{if } \lfloor p_n/2 \rfloor + 1 \leq j \leq \lfloor 3p_n/4 \rfloor, \\ (5\nu_k^{1/2}/3)^{-1/2}t(5) & \text{if } \lfloor 3p_n/4 \rfloor + 1 \leq j \leq p_n. \end{cases}$$

Denoting $\{g_{ijk} : i \leq n_2, j \leq p_n, k \leq 50\}$ as a collection of $i.i.d.$ standard normal random variables, we then define

$$\begin{aligned} Y_{ij}(t) &= \sum_{j'=1}^{p_n} \psi_{j'}(1 + |j - j'|)^{-3}\{V_{ij'}(t) + \delta^{1/2}1_{\{j \leq \lfloor \beta p_n \rfloor\}} \sum_{k=1}^{50} g_{ij'k}\phi_k(t)\} \\ &= \sum_{k=1}^{50}\sum_{j'=1}^{p_n} \psi_{j'}(1 + |j - j'|)^{-3}(\xi_{2,ij'k} + \delta^{1/2}1_{\{j \leq \lfloor \beta p_n \rfloor\}}g_{ij'k})\phi_k(t) \\ &= \sum_{k=1}^{50} \theta_{2,ijk}\phi_k(t), \end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

with each $\theta_{2,ijk} = \sum_{j'=1}^{p_n} \psi_{j'}(1+|j-j'|)^{-3}(\xi_{2,ij'k} + \delta^{1/2} 1_{\{j \le \lfloor \beta p_n \rfloor\}} g_{ij'k})$. It is easy to verify that the covariance structure $\Sigma^{*Y} = cov(\theta^*_{2,i})$ can be expressed as

$$\Sigma^{*Y} = \Sigma^{*X} + \delta diag\{I_{50\lfloor \beta p_n \rfloor}, 0\} \Sigma \Phi \Phi \Sigma diag\{I_{50\lfloor \beta p_n \rfloor}, 0\}.$$

Observe that the parameter $\delta \ge 0$ regulates the signal intensity level of the covariance difference. An increase in $\delta \ge 0$ signifies a stronger signal intensity, where $\delta = 0$ means the null. The parameter $\beta \in [0, 1]$ quantifies the sparseness level of the difference, with $\lfloor \beta p_n \rfloor$ representing the largest integer less than or equal to βp_n. As β increases, the signal becomes denser, ranging from the null hypothesis ($\beta = 0$) to the fully dense alternative ($\beta = 1$). In the simulation, we consider a diverse set of sparseness levels $\beta \in \{0, 0.1, 0.2, 0.4, 0.6, 0.8, 1\}$, along with various signal strength values $\delta \in \{0.2, 0.25, 0.3\}$. The functional data $\{X_{ij}(\cdot) : i \le n_1, j \le p_n\}$ and $\{Y_{ij}(\cdot) : i \le n_2, j \le p_n\}$ are observed at m uniformly distributed points $\{t_{ijl}, l = 1, \ldots, m\} \in \mathcal{T}$. Note that the within-curve observation density increases in m. To model these observations, an orthonormal cubic spline basis is utilized to fit each curve.

It is important to highlight that our simulation framework presents significant challenges for testing the hypotheses (2.1), primarily attributed to the following five key factors. The first factor arises from the combination of the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

large-scale framework and the significantly unequal sample sizes. The second factor is attributed to the "eigenvalue-decay-free" characteristic linked with the simulated functional data, which is a result of the random permutation process used to generate the sequence $\nu_1, \ldots, \nu_{50}$. The third factor stems from the fact that the functional data are produced in such a way that they asymptotically fail to meet the square-integrable condition, as evidenced by $\int_{\mathcal{T}} E(X_{ij}^2(t))dt \asymp \int_{\mathcal{T}} E(Y_{ij}^2(t))dt \asymp \sum_{k=1}^{50} \nu_k^{-1/2} \asymp \sum_{k=1} k^{-1/2} = \infty$. The fourth factor is due to the highly complex nature of the functional data's distribution, which is a blend of heavy-tailed (t-distribution) and skewed (chi-square distribution) distributions. The fifth factor comes from the stronger and more intricate correlation structures among the functional predictors, within the framework of an autoregressive relation that exhibits a polynomial decay pattern.

In the numerical study, we evaluate and compare the performance of the two tests (XXZ and XYZ) by examining the corresponding rejection proportions (empirical size or power) across 1000 Monte Carlo runs for each distinct pairing of levels δ and β. Regarding the Setting I with $(n_1, n_2, p_n, m) = (80, 120, 240, 100)$, the numerical outcomes are summarized in the top portion of Table 1. The row corresponding to $\beta = 0$ presents the empirical sizes for both the XXZ and XYZ tests. It is evident that the proposed XXZ test closely aligns with the nominal level $\alpha = 0.05$, whereas the XYZ test shows a tendency

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

to underestimate the type I error rate. When evaluating the power performance against alternative hypotheses, given any signal strength level of δ, it is seen that the empirical power of the XXZ test consistently outperforms that of the XYZ test across all values of β. Thus, the results from Setting I not only affirm the theoretical validity of our proposed test but also underscore its superiority over the competing test. To further investigate the impact of different levels of ultra-high-dimensionality, we present the numerical results for smaller sample sizes with higher data dimension to sample size ratios. Specifically, these results pertain to the configurations of $(n_1, n_2, p_n, m) = (70, 105, 240, 100)$ and $(n_1, n_2, p_n, m) = (60, 90, 240, 100)$ in Settings II and III, respectively, as detailed in Table 1. Notably, the observed patterns in Settings II and III closely resemble those found in Setting I, suggesting that the proposed test consistently holds its validity and surpasses the competitor method across various degrees of ultra-high-dimensionality. To examine the effect of within-curve observation density, we additionally include Setting IV in Table 1, which matches Setting III in all respects except that the number of grid points is reduced to $m = 50$. A comparison of Settings III and IV in Table 1 indicates that the associated tests exhibit improved performance under denser within-curve observation, that is, for larger m. This pattern is expected, as increasing the number of observed points per subject provides richer functional information and thereby enhances

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

inferential accuracy. To further investigate the sensitivity of the test to varying truncation s_n, we consider Setting V in Table 1. In Setting V, the data-generating mechanism is identical to that of Setting III with $\delta = 0.3$, but the tests are implemented using pre-specified values $s_n \in \{8, 10, 12\}$ rather than the data-driven choice of s_n employed in previous settings. As evidenced by the results presented in Setting V of Table 1, the empirical sizes and powers exhibit minimal variation across different pre-specified values of the truncation parameter s_n. This indicates that the proposed method is relatively insensitive to fluctuations in s_n. With respect to computational efficiency, a single Monte Carlo replication of the proposed test requires, on average, approximately 60 seconds of wall-clock time when executed on a Victus by HP Laptop 16-d0xxx equipped with an 11th Gen Intel(R) Core(TM) i7-11800H 2.30 GHz CPU and 32 GB of RAM. To visualize how each scenario performs across various parameter combinations, we present the empirical power plots for both tests in Figure 1. These plots reaffirm the proposed test's validity and superiority, even in the presence of fairly complex data structures such as ultra-high-dimensionality with unbalanced sample sizes, highly correlated functional data, and heavy-tailed or skewed distributions.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

Table 1: Rejection proportions (%) derived from 1000 Monte Carlos for both the proposed test (XXZ) and its competitor (XYZ), at varying levels of signal intensity (δ) and sparsity (β). Here, $\beta = 0$ represents the null hypothesis scenario, while $\beta = 1$ indicates a fully dense alternative. Settings I-III denote varying extents of ultra-high-dimensionality. Setting IV explores a sparser grid points density than Setting III. Setting V examines the sensitivity to changes in s_n.

Setting I: $(n_1, n_2, p_n, m) = (80, 120, 240, 100)$						
	$\delta = 0.2$		$\delta = 0.25$		$\delta = 0.3$	
Test	XXZ	XYZ	XXZ	XYZ	XXZ	XYZ
$\beta = 0$	4.40	3.20	4.50	3.40	4.80	3.30
$\beta = 0.1$	16.9	14.0	43.7	39.3	77.8	73.7
$\beta = 0.2$	30.3	26.4	68.2	62.1	96.1	94.6
$\beta = 0.4$	52.1	47.4	89.8	86.1	99.8	99.5
$\beta = 0.6$	62.8	58.2	96.2	94.5	100	100
$\beta = 0.8$	73.1	68.0	99.2	98.1	100	100
$\beta = 1$	78.6	74.4	99.5	98.8	100	100
Setting II: $(n_1, n_2, p_n, m) = (70, 105, 240, 100)$						
	$\delta = 0.2$		$\delta = 0.25$		$\delta = 0.3$	
Test	XXZ	XYZ	XXZ	XYZ	XXZ	XYZ
$\beta = 0$	4.70	3.50	5.00	3.70	4.90	3.90
$\beta = 0.1$	10.1	8.25	28.5	23.9	56.3	48.8
$\beta = 0.2$	15.4	13.2	42.9	35.7	74.5	70.1
$\beta = 0.4$	29.7	25.8	64.7	58.3	93.5	92.2
$\beta = 0.6$	39.2	35.0	81.9	75.1	98.4	98.0
$\beta = 0.8$	50.7	44.8	91.2	86.5	99.8	99.3
$\beta = 1$	56.1	51.8	95.6	92.9	100	100
Setting III: $(n_1, n_2, p_n, m) = (60, 90, 240, 100)$						
	$\delta = 0.2$		$\delta = 0.25$		$\delta = 0.3$	
Test	XXZ	XYZ	XXZ	XYZ	XXZ	XYZ
$\beta = 0$	5.20	3.70	5.10	3.80	5.30	3.60
$\beta = 0.1$	10.0	7.60	17.6	14.8	34.6	28.8
$\beta = 0.2$	13.5	10.9	28.2	24.2	53.2	45.9
$\beta = 0.4$	22.4	18.6	47.5	41.9	76.1	68.4
$\beta = 0.6$	24.5	19.8	58.2	50.3	90.1	84.5
$\beta = 0.8$	32.5	26.7	66.9	60.2	94.3	90.5
$\beta = 1$	41.7	33.3	75.6	67.8	98.4	94.8
Setting IV: $(n_1, n_2, p_n, m) = (60, 90, 240, 50)$						
	$\delta = 0.2$		$\delta = 0.25$		$\delta = 0.3$	
Test	XXZ	XYZ	XXZ	XYZ	XXZ	XYZ
$\beta = 0$	4.40	3.10	4.60	3.30	4.50	3.40
$\beta = 0.1$	8.90	6.70	17.3	13.9	34.0	28.7
$\beta = 0.2$	12.0	9.30	28.1	22.4	53.0	45.4
$\beta = 0.4$	20.6	17.0	44.0	37.0	74.2	65.6
$\beta = 0.6$	23.9	18.7	56.2	48.7	84.9	79.4
$\beta = 0.8$	32.3	25.8	66.0	58.2	91.9	88.5
$\beta = 1$	38.9	33.1	74.2	66.6	97.1	94.6
Setting V: $(n_1, n_2, p_n, m, \delta) = (60, 90, 240, 100, 0.3)$						
	$s_n = 8$		$s_n = 10$		$s_n = 12$	
Test	XXZ	XYZ	XXZ	XYZ	XXZ	XYZ
$\beta = 0$	4.60	3.10	5.40	3.60	5.10	3.50
$\beta = 0.1$	33.9	27.3	34.6	28.9	32.4	28.4
$\beta = 0.2$	52.9	43.1	53.1	45.0	51.3	47.0
$\beta = 0.4$	74.8	68.3	75.9	66.4	75.8	68.5
$\beta = 0.6$	87.8	83.5	89.8	84.8	88.8	85.5
$\beta = 0.8$	93.6	90.3	94.1	90.0	94.2	90.5
$\beta = 1$	96.8	94.7	97.8	94.6	97.7	94.7

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

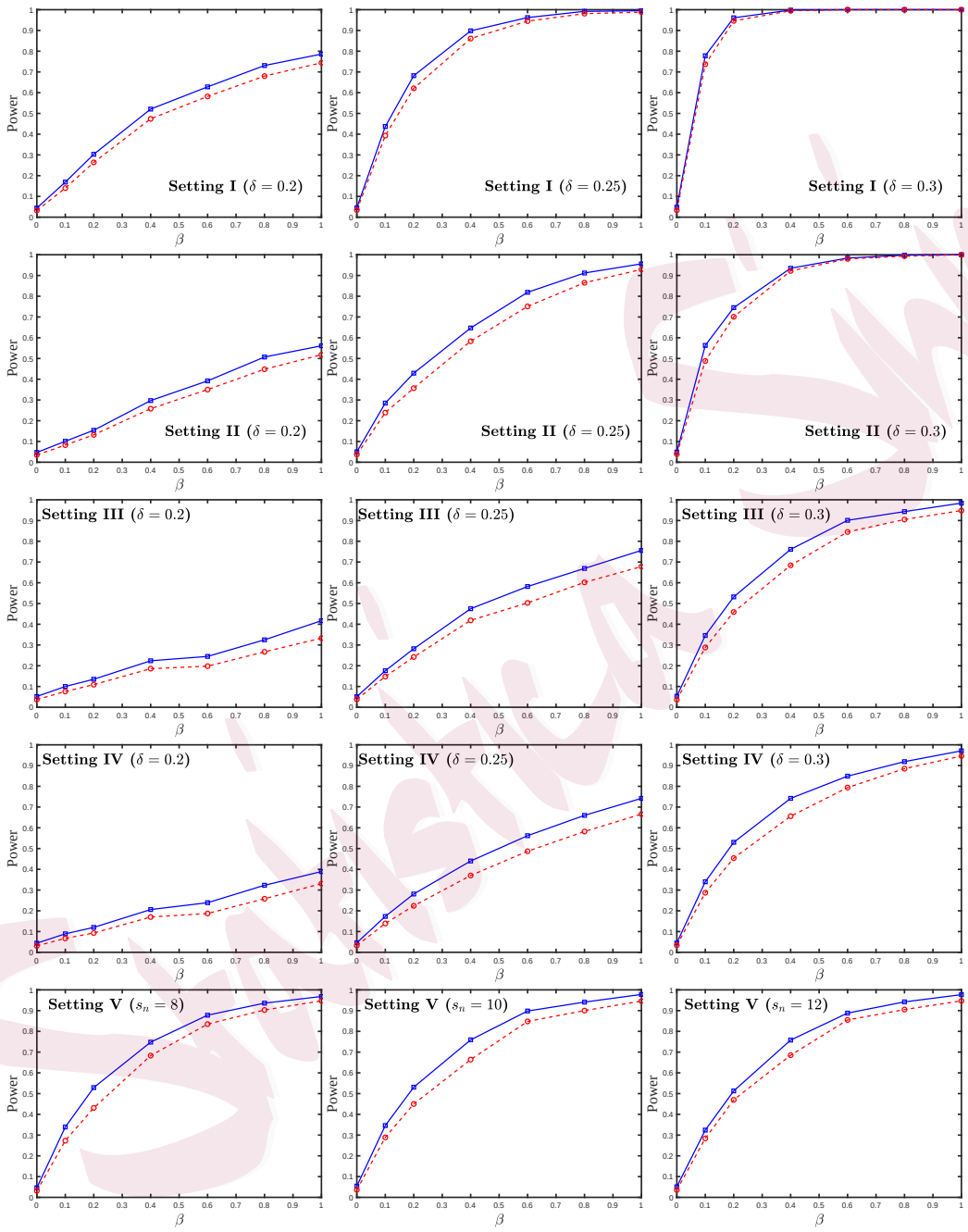

Figure 1: Depicted are the empirical power curves for the two methods: the proposed XXZ test (squares), the XYZ test (circles) based on 1000 Monte Carlos under Settings I–V across varying levels of signal intensity (δ) and sparsity (β).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

4. Real data analysis

To demonstrate the proposed test, we apply it to the task of analyzing a real dataset related to electroencephalography (EEG), which can be accessed at the following link: `https://kdd.ics.uci.edu/databases/eeg/eeg.html`. The EEG dataset comprises $n = 122$ participants, including $n_1 = 45$ control subjects and $n_2 = 77$ individuals identified as alcoholics. Each individual has 64 electrodes positioned on their scalp, with EEG readings from a specific electrode being captured at a frequency of 256 Hz (equivalent to a 3.9-ms epoch) per second. Consequently, there are $p_n = 64$ functional predictors associated with these electrodes for each participant. Within the control group, the covariance function between electrodes j_1 and j_2 is represented by $K^X_{j_1 j_2}(s,t)$, whereas in the alcoholic group, it is expressed as $K^Y_{j_1 j_2}(s,t)$. In practical application, our primary focus is on testing the hypotheses:

$$H_0 : \sup_{s\in[0,1]} \sup_{t\in[0,1]} \max_{j_1\le p_n} \max_{j_2\le p_n} |K^X_{j_1 j_2}(s,t) - K^Y_{j_1 j_2}(s,t)| = 0 \qquad \text{versus}$$

$$H_1 : \sup_{s\in[0,1]} \sup_{t\in[0,1]} \max_{j_1\le p_n} \max_{j_2\le p_n} |K^X_{j_1 j_2}(s,t) - K^Y_{j_1 j_2}(s,t)| \neq 0,$$

with the goal of identifying any covariance differences between the two sets of functional data. When the proposed test is implemented on the dataset with resampling size $N = 10^4$, it is found that the null hypothesis H_0 is re-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

jected at a nominal significance level of $\alpha = 0.01$, with an empirical p-value of 2×10^{-4}, revealing a statistically significant difference in covariance between the two samples. This finding is corroborated by extensive studies on EEG analysis, including the work of Neeraj et al. (2021). To gain more insight into the differences in covariance, we consider two electrodes: the 8th electrode, located in the frontal region, and the 14th electrode, positioned in the temporal region. For the control group, we present the moment estimates of the auto-covariance surfaces $\hat{K}^X_{8,8}(s,t)$, $\hat{K}^X_{14,14}(s,t)$, as well as the inter-covariance surface $\hat{K}^X_{8,14}(s,t)$, depicted in the upper section of Figure 2. Correspondingly, for the alcoholic group, the estimated covariance surfaces $\hat{K}^Y_{8,8}(s,t)$, $\hat{K}^Y_{14,14}(s,t)$, and $\hat{K}^X_{8,14}(s,t)$ are displayed in the lower part of Figure 2. As seen in Figure 2, the auto-covariance and inter-covariance surfaces for the control group differ considerably from those of the alcoholic group. This observation indicates that the functional connectivity between the frontal and temporal regions is altered in individuals with alcohol dependence, as supported by various studies (Rzepecki-Smith et al., 2010; Maleki et al., 2022). As depicted in Figure 2, the covariances (i.e., brain functional connectivities) in the control group, which exhibit more intricate grooves, tend, in general, to be more active than those in the alcoholic group. This phenomenon for the alcoholic is consistent with various studies (Weiland et al., 2014). Similarly, many cognitive disorders, such as Alzheimer's disease, have

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

also been reported to show impaired functional connectivity in several brain regions relative to healthy individuals (e.g., Engels et al., 2015). Taken together, these accumulating findings suggest that chronic alcohol misuse leads to adverse neurophysiological consequences that may, from a biological standpoint, share features with other cognitive disorders, through mechanisms that ultimately results in the attenuation of activity of multiple brain regions (Chang et al., 2025). Accordingly, the practical value of the proposed test is confirmed through its application in a real-world scenario.

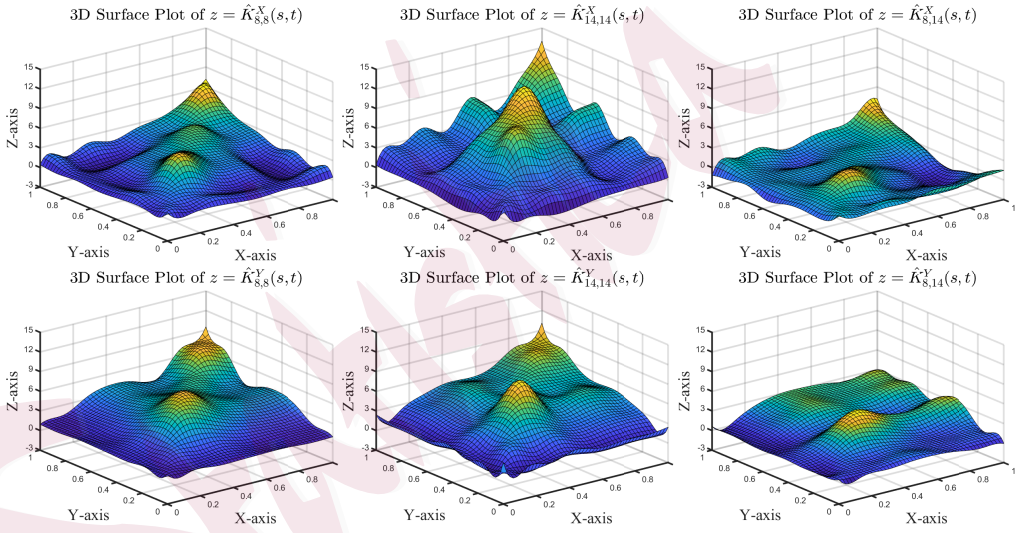

Figure 2: Shown are the estimated covariances surfaces for electrodes 8 and 14. $\hat{K}^X_{8,8}(s,t)$, $\hat{K}^X_{14,14}(s,t)$, $\hat{K}^X_{8,14}(s,t)$ correspond to the control group, while $\hat{K}^Y_{8,8}(s,t)$, $\hat{K}^Y_{14,14}(s,t)$, $\hat{K}^Y_{8,14}(s,t)$ depict the the alcoholic group.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

5. Discussion

The present work still has several potential gaps that merit further investigation. For instance, our procedure currently relies on a loss criterion via minimizing a maximum absolute deviation (chebyshev distance) to select the truncation level s_n, while investigating the test performance under alternative loss criteria induced by other norms or distance measures constitutes an important direction for future research. In addition, another interesting avenue is to extend the proposed framework to settings involving more than two populations.

Supplementary Material

The auxiliary lemmas with their proofs, along with the proofs of the main theorems, are relegated to an online Supplementary Material to save space.

Acknowledgments

Kaijie Xue and Lan Xue are the co-first authors. Riquan Zhang is the corresponding author: zhangriquan@163.com. The authors are listed in alphabetical order, and all authors have made equal contributions. This research was supported by the National Natural Science Foundation of China (12371268, 12531013, 12371272), the Youth Project of Shanghai Eastern Talent Program (QNJY2024152), the Basic Research Project of Shanghai Science and Technol-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

ogy Commission (22JC1400800), and awards from the National Institutes of Diabetes, Digestive, and Kidney Disease Award numbers R01DK132385.

References

Cai, T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association 108*(501), 265–277.

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society. Series B: Statistical Methodology 76*(2), 349–372.

Cardot, H., F. Ferraty, and P. Sarda (1999). Functional linear model. *Statistics & Probability Letters 45*(1), 11–22.

Chang, J., W. Zhou, W.-X. Zhou, and L. Wang (2017). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics 73*(1), 31–41.

Chang, J.-S., H.-Z. Huang, M. Yuan, Y. Zhou, D. Liu, K.-B. Zhan, and L.-Q. Zhu (2025). Alcohol addiction and alzheimers disease: a molecular collision course. *Translational Psychiatry 15*(1), 410.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

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

Dai, X., H.-G. Muller, and F. Yao (2017). Optimal bayes classifiers for functional data and density ratios. *Biometrika 104*, 545–560.

Delaigle, A. and P. Hall (2012). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society. Series B 74*(2), 267–286.

Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics 32*(3), 962–994.

Engels, M. M., C. J. Stam, W. M. Van Der Flier, P. Scheltens, H. de Waal, and E. C. van Straaten (2015). Declining functional connectivity and changing hub locations in alzheimers disease: an eeg study. *BMC neurology 15*(1), 145.

Fan, Y., G. M. James, and P. Radchenko (2015). Functional additive regression. *Annals of Statistics 43*(5), 2296–2325.

Fang, Q., Q. Jiang, and X. Qiao (2024). Large-scale multiple testing of cross-covariance functions with applications to functional network models. *arXiv preprint arXiv:2407.19399*.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

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

Fremdt, S., J. G. Steinebach, L. Horvath, and P. Kokoszka (2013). Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics 40*(1), 138–152.

Guo, J., B. Zhou, and J.-T. Zhang (2019). New tests for equality of several covariance functions for functional data. *Journal of the American Statistical Association 114*(527), 1251–1263.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics 35*(1), 70–91.

Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B 68*(1), 109–126.

Han, Y. and W. B. Wu (2020). Test for high dimensional covariance matrices. *The Annals of Statistics 48*(6), 3565–3588.

He, J. and S. Chen (2018). High-dimensional two-sample covariance matrix testing via super-diagonals. *Statistica Sinica 28*(4), 2671–2696.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

Kong, D., K. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika 103*(1), 147–159.

Maleki, N., K. S. Sawyer, S. Levy, G. J. Harris, and M. Oscar-Berman (2022). Intrinsic brain functional connectivity patterns in alcohol use disorder. *Brain Communications 4*(6), fcac290.

Neeraj, V. Singhal, J. Mathew, and R. K. Behera (2021). Detection of alcoholism using eeg signals and a cnn-lstm-attn network. *Computers in Biology and Medicine 138*, 104940.

Panaretos, V. M., D. Kraus, and J. H. Maddocks (2010). Second-order comparison of gaussian random functions and the geometry of dna minicircles. *Journal of the American Statistical Association 105*(490), 670–682.

Paparoditis, E. and T. Sapatinas (2016). Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika 103*(3), 727–733.

Pigoli, D., J. A. D. Aston, I. L. Dryden, and P. Secchi (2014). Distances and inference for covariance operators. *Biometrika 101*(2), 409–422.

Poldrack, R. A., J. A. Mumford, and T. E. Nichols (2024). *Handbook of functional MRI data analysis*. Cambridge University Press.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

Ramsay, J. O. and C. J. Dalzell (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 53*(3), 539–572.

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

Rzepecki-Smith, C. I., S. A. Meda, V. D. Calhoun, M. C. Stevens, M. J. Jafri, R. S. Astur, and G. D. Pearlson (2010). Disruptions in functional network connectivity during alcohol intoxicated driving. *Alcoholism: clinical and experimental research 34*(3), 479–487.

Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics and Data Analysis 51*(12), 6535–6542.

Srivastava, M. S. and H. Yanagihara (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis 101*(6), 1319–1329.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). Functional data analysis. *Annual Review of Statistics and its application 3*(1), 257–295.

Weiland, B. J., A. Sabbineni, V. D. Calhoun, R. C. Welsh, A. D. Bryan, R. E.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

Jung, A. R. Mayer, and K. E. Hutchison (2014). Reduced left executive control network functional connectivity is associated with alcohol use disorders. *Alcoholism: Clinical and Experimental Research 38*(9), 2445–2453.

Wu, T.-L. and P. Li (2020). Projected tests for high-dimensional covariance matrices. *Journal of Statistical Planning and Inference 207*, 73–85.

Xue, K. (2023). Distribution/correlation-free test for two-sample means in high-dimensional functional data with eigenvalue decay relaxed. *SCIENCE CHINA Mathematics 66*, 2337–2346.

Xue, K., J. Yang, and F. Yao (2024). Optimal linear discriminant analysis for high-dimensional functional data. *Journal of the American Statistical Association 119*(546), 1055–1064.

Xue, K., J. Yang, and R. Zhang (2025). Testing the equality of covariances for large-scale functional data. *SCIENCE CHINA Mathematics*, DOI: 10.1007/s11425–023–2441–0.

Xue, K. and F. Yao (2020). Distribution and correlation-free two-sample test of high-dimensional means. *The Annals of Statistics 48*(3), 1304–1328.

Xue, K. and F. Yao (2021). Hypothesis testing in large-scale functional linear regression. *Statistica Sinica 31*, 1101–1123.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0295

Xue, K. and F. Yao (2024). Inference on large-scale partially functional linear model with heterogeneous errors. *Statistica Sinica 34*, 679–697.

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

Zou, T., R. Lin, S. rong Zheng, and G.-L. Tian (2021). Two-sample tests for high-dimensional covariance matrices using both difference and ratio. *Electronic Journal of Statistics 15*, 135–210.