

Article

Two-Sample Hypothesis Test for Functional Data

Jing Zhao ¹, Sanying Feng ²  and Yuping Hu ^{2,*}¹ China National Institute of Standardization, Beijing 100191, China² School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China

* Correspondence: hyp@zzu.edu.cn

Abstract: In this paper, we develop and study a novel testing procedure that has more a powerful ability to detect mean difference for functional data. In general, it includes two stages: first, splitting the sample into two parts and selecting principle components adaptively based on the first half-sample; then, constructing a test statistic based on another half-sample. An extensive simulation study is presented, which shows that the proposed test works very well in comparison with several other methods in a variety of alternative settings.

Keywords: functional data analysis; mean functions comparison; two sample testing; sample splitting

MSC: 62G10; 62G20

1. Introduction

In the recent literature, there has been an increasing interest in functional data analysis, with its extensive application in biometrics, chemometrics, econometrics, and medical research, as well as other fields. Functional data have intrinsically infinite dimensions and thus, classical methods for multivariate observations are not applicable. Therefore, it is necessary to develop special techniques for this type of data. There has intensive methodological and theoretical development in function data analysis; see [1–5] and so on.

In functional data analysis, a functional data set or curve can be modeled as independent realizations of an underlying stochastic process:

$$x_i(t) = \mu(t) + \alpha_i(t) + \epsilon_i(t), i = 1, 2, \dots, n, \quad (1)$$

where $\mu(t)$ is the mean function of the stochastic process, $\alpha_i(t)$ is the i th individual function variation from $\mu(t)$, and $\epsilon_i(t)$ is the i th measurement error process. In general, we assume $\alpha_i(t)$ and $\epsilon_i(t)$ are independent, and i.i.d. sample from $\alpha(t)$ and $\epsilon(t)$, respectively, where, $\alpha(t) \sim SP(0, \gamma)$, $\epsilon(t) \sim SP(0, \gamma_\epsilon)$, and SP denotes a stochastic process with mean function $\mu(t)$ and covariance function $\gamma(s, t)$.

The mean function $\mu(t)$ reflects the underlying trend and can be used as an important index for population response, such as in drug and biomedical purposes, among other. One important statistical inference other than estimation is related to testing of various hypotheses about the mean function. Therefore, we focus on the problem of testing the equality of mean functions in two random samples independently drawn from two functional random variables. There have been some approaches proposed so far to address this problem. For instance, Ref. [6] proposed an adaptive Neyman test, but in the case when the sampling information is in a “discrete” format. Ref. [7] discussed two methods: multivariate analysis-based and bootstrap-based testing.

However, these methods have only been applied in narrow fields and not available for a global testing result. Refs. [8,9] proposed an L^2 -norm based statistic to test the equality of mean functions. Ref. [10] proposed and studied a so-called Globalized Pointwise F-test, abbreviated as GPF test. The GPF test is in general comparable with the L^2 -norm-based test



Citation: Zhao, J.; Feng, S.; Hu, Y. Two-Sample Hypothesis Test for Functional Data. *Mathematics* **2022**, *10*, 4060. <https://doi.org/10.3390/math10214060>

Academic Editor: Mustapha Rachdi

Received: 26 September 2022

Accepted: 19 October 2022

Published: 1 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

and the F-type test adopted for the one-way ANOVA problem. Then, Ref. [11] proposed the F_{\max} -test; via some simulation studies, it was found that in terms of both level accuracy and power, the F_{\max} -test outperforms the Globalized Pointwise F (GPF) test of [10] when the functional data are highly or moderately correlated, and its performance is comparable with the latter otherwise. Ref. [5] proposed a statistic based on the functional principal component emi-distances. Furthermore, they gave a normalized version of the functional principal components that has a chi-square limit distribution. The statistic is scale-invariant. However, this method require pre-specifying a threshold to choose the leading principal components (PCs), where PCs are ranked based on eigenvalues. They chose the number of PCs, for example, d , based on the percentage of variance explained for the functional covariates. This method has two drawbacks: one is that it can only detect mean differences in this d -dimensional subspace; the other is that different thresholds often lead to different tests, whose power depends on the particular simple alternative hypothesis.

In this paper, we develop and study a novel testing procedure that overcomes the drawback that many tests can only detect mean difference in the d -dimensional subspace. Furthermore, the novel testing procedure is very powerful in the cases when there are differences in middle part and latter part of two function sequences. Additionally, we derived the asymptotic distribution of the new test statistics under the alternative hypothesis, which is the key difficulty in the current approach. In general, the novel testing procedure includes two stages: first, split the sample into two parts and select PCs adaptively based on the first half-sample; then, construct the test statistic based on another half-sample. Sample splitting is often used in high-dimension regression problems because most computationally efficient selection algorithms cannot guard against inclusion of noise variables. Asymptotically valid p -values are not available. Ref. [12] adopted this technology to reduce the number of variables to a manageable size using the first split sample and apply classical variable selection techniques with the remaining variables, using the data from the second split.

In our procedure, we mainly adopt two methods to select PCs in the first stage: the adaptive Neyman test [6] and the adaptive ordered Neyman test. At the same time, we also consider selecting PCs based on pre-specifying a threshold; however, this threshold is an association–projection index that combines both the variation and the projection along each direction. The purpose of splitting the sample is two-fold: (1) to decrease the random noise effect in the first stage; (2) to derive the asymptotic distribution of the test statistic. From simulation results, we can see that our testing procedure asymptotically achieves the pre-specified significance level, and enjoys certain optimality in terms of its power, even when the population is a non-Gaussian process.

This paper is organized as follows. In Section 2, we introduce the test problem and briefly review the existing global two-sample test methods. Section 3 proposes our testing procedure. Simulation studies are given in Section 4. A real-data example is analyzed in Section 5. Section 6 concludes the present work. The derivations are given in Appendix A.

2. The Testing Problem for Functional Data

2.1. Preliminary

Let $SP(\mu, \Gamma)$ denote a stochastic process with mean function $\mu(t), t \in T$ and covariance function $\Gamma(s, t), s, t \in T$, where $T \in [0, 1]$. Suppose we have the following two independent function samples:

$$X_1(t), \dots, X_{n_1}(t) \sim SP(\mu_1, \Gamma_1(s, t)), \tag{2}$$

$$Y_1(t), \dots, Y_{n_2}(t) \sim SP(\mu_2, \Gamma_2(s, t)), \tag{3}$$

where $\Gamma_1(s, t)$ denotes the covariance function of the function data $X(t)$ and $\Gamma_2(s, t)$ denotes the covariance function of the function data $Y(t)$. However, we do not know if $\Gamma_1(s, t)$ and $\Gamma_2(s, t)$ are equal. We want to test whether the two mean functions are equal:

$$H_0 : \mu_1(t) = \mu_2(t) \quad \text{vs.} \quad H_1 : \mu_1(t) \neq \mu_2(t). \tag{4}$$

Let $\bar{X}(t), \bar{Y}(t)$ denote the sample mean functions of the two samples, respectively. First, we have

$$\hat{\mu}_1(t) = \bar{X}(t), \hat{\mu}_2(t) = \bar{Y}(t), \tag{5}$$

and

$$\sqrt{n}(\hat{\mu}_1(t) - \hat{\mu}_2(t)) \sim \text{SP}(\sqrt{n}(\mu_1(t) - \mu_2(t)), \Gamma_{12}(s, t)), \tag{6}$$

where $\Gamma_{12}(s, t) = \frac{\Gamma_1(s, t)}{n_1/n} + \frac{\Gamma_2(s, t)}{n_2/n}, n = n_1 + n_2$. $\Gamma_{12}(s, t)$ can be written as $\Gamma_{12}(s, t) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(s) \varphi_k(t)$, where $\lambda_1 \geq \lambda_2 \cdots \geq 0$ are the eigenvalues and $\varphi_k(t), k = 1, \dots, \infty$ are eigenfunctions satisfying $\int_T \varphi_k(t)^2 dt = 1$ and $\int_T \varphi_k(t) \varphi_l(t) dt = 0, k \neq l$.

It is easy to see that $\hat{\Gamma}_1(s, t) = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i(t) - \hat{\mu}_1(t))(X_i(s) - \hat{\mu}_1(s)), \hat{\Gamma}_2(s, t) = \frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i(t) - \hat{\mu}_2(t))(Y_i(s) - \hat{\mu}_2(s))$ and $\hat{\Gamma}_{12}(s, t) = \frac{\hat{\Gamma}_1(s, t)}{n_1/n} + \frac{\hat{\Gamma}_2(s, t)}{n_2/n}$. $\hat{\Gamma}_{12}(s, t)$ can also be written in as an eigen-decomposition

$$\hat{\Gamma}_{12}(s, t) = \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\varphi}_k(s) \hat{\varphi}_k(t), \tag{7}$$

where the nonincreasing sequence $(\hat{\lambda}_k : k \geq 1)$ is the sample eigenvalues and $(\hat{\varphi}_k : k \geq 1)$ are the corresponding eigenfunctions forming an orthonormal basis of $L_2[0, 1]$.

To simplify notation, we use the symbol Γ for both the kernel and the operator. Now, the functional Mahalanobis semi-distance between $\hat{\mu}_1(t)$ and $\hat{\mu}_2(t)$ is defined as :

$$d_{FM}^2(\hat{\mu}_1, \hat{\mu}_2) = \langle \hat{\Gamma}_{12, p_n}^{-\frac{1}{2}} \sqrt{n}(\hat{\mu}_1 - \hat{\mu}_2), \hat{\Gamma}_{12, p_n}^{-\frac{1}{2}} \sqrt{n}(\hat{\mu}_1 - \hat{\mu}_2) \rangle. \tag{8}$$

Plugging (6) into (7), we have

$$d_{FM}^2(\hat{\mu}_1, \hat{\mu}_2) = \sum_{k=1}^{p_n} \frac{n \langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\hat{\lambda}_k}. \tag{9}$$

2.2. Existing Global Testing Methods

Although there is significant literature discussing the equality of means for two functional data sets, they can be roughly grouped into a few broad categories, as follows .

(1) L_2 -norm-based test

The test is based on the L_2 -norm of the difference between $\hat{\mu}_1(t)$ and $\hat{\mu}_2(t)$:

$$T_L = n \|\hat{\mu}_1(t) - \hat{\mu}_2(t)\|^2 = n \int (\hat{\mu}_1(t) - \hat{\mu}_2(t))^2 dt. \tag{10}$$

Ref. [8] proved that $T_L = n \int (\hat{\mu}_1(t) - \hat{\mu}_2(t))^2 dt \stackrel{d}{=} \sum_{k=1}^{\infty} \lambda_k A_k + o_p(1), A_k \sim \chi_1^2(\frac{nu_k^2}{\lambda_k})$,

where $X \stackrel{d}{=} Y$ denotes that X and Y have the same distribution, and $u_k = \int_0^1 (\mu_1(t) - \mu_2(t)) \varphi_k(t) dt, k = 1, 2, \dots$. Furthermore, they use the two-cumulant matched χ^2 approximation method and obtained an approximate distribution of T_L , which is, $\alpha \chi_d^2 + \beta$, where

$$\alpha = \frac{\sum_{k=1}^{\infty} \hat{\lambda}_k^3}{\sum_{k=1}^{\infty} \hat{\lambda}_k^2}, \quad d = \frac{\left\{ \sum_{k=1}^{\infty} \hat{\lambda}_k^2 \right\}^3}{\left\{ \sum_{k=1}^{\infty} \hat{\lambda}_k^3 \right\}^2}, \quad \beta = \sum_{k=1}^{\infty} \hat{\lambda}_k - \frac{\left\{ \sum_{k=1}^{\infty} \hat{\lambda}_k^2 \right\}^2}{\sum_{k=1}^{\infty} \hat{\lambda}_k^3}.$$

Then, they have $P(T_L > K) \approx P(\chi_d^2 > (K - \beta)/\alpha)$.

(2) Projection-based Test

Ref. [5] considered projecting the observed mean difference onto the space spanned by $\hat{\varphi}_1, \dots, \hat{\varphi}_d$, where d is determined based on the percentage of eigenvalues, constructed the following test statistic:

$$T_H = \sum_{k=1}^d \langle \sqrt{n}(\hat{\mu}_1 - \hat{\mu}_2), \hat{\varphi}_k \rangle^2. \tag{11}$$

Given d , the asymptotic distribution of T_H under the null hypothesis is the distribution of $\sum_{k=1}^d \lambda_k Z_k^2$, where $Z_k, k = 1, \dots, d$ are independent standard normal random variables. Alternately, they proposed a normalized version of T_H which is given by

$$NT_H = \sum_{k=1}^d \langle \sqrt{n}(\hat{\mu}_1 - \hat{\mu}_2), \hat{\varphi}_k \rangle^2 / \hat{\lambda}_k, \tag{12}$$

Then, under the null hypothesis, NT_H has an asymptotic χ_d^2 distribution.

(3) *F*-test

Finally, we describe the testing procedure proposed by [13]. Although they proposed a functional *F* test in a functional linear regression setting, the method can be specified for our two sample test as follows.

The *F*-test statistic for our setting is

$$T_F = \frac{RSS_0 - RSS_1}{RSS_1 / (n - 1)}, \tag{13}$$

where

$$\begin{aligned} RSS_1 &= \sum_{i=1}^{n_1} \int_0^1 (X_i(t) - \bar{X}(t))^2 dt + \sum_{i=1}^{n_2} \int_0^1 (Y_i(t) - \bar{Y}(t))^2 dt, \\ RSS_0 &= \sum_{i=1}^{n_1} \int_0^1 (X_i(t) - \bar{Z}(t))^2 dt + \sum_{i=1}^{n_2} \int_0^1 (Y_i(t) - \bar{Z}(t))^2 dt, \\ \bar{Z}(t) &= \frac{n_1 \bar{X}(t) + n_2 \bar{Y}(t)}{n_1 + n_2}. \end{aligned}$$

Ref. [13] also presented the distribution of the *F*-test as $\frac{\sum_{k=1}^{\infty} \lambda_k \chi_1^2}{\sum_{k=1}^{\infty} \lambda_k \chi_{n-2}^2}$. In practice, they also applied the idea of [14] approximation to derive approximate distribution of *F*-statistic, that is, $(\chi_{f_1}^2 / f_1) / (\chi_{f_2}^2 / f_2)$, which is an ordinary *F* distribution with degrees of freedom f_1 and f_2 , where $f_1 = (\sum_{k=1}^{\infty} \hat{\lambda}_k)^2 / \sum_{k=1}^{\infty} \hat{\lambda}_k^2$, $f_2 = (n - 2) (\sum_{k=1}^{\infty} \hat{\lambda}_k)^2 / \sum_{k=1}^{\infty} \hat{\lambda}_k^2$.

3. Our Testing Procedure

In order to determine the number of PCs p_n adaptively and find significant parts to construct a more powerful test statistic, we propose a two-stage procedure via a data-splitting technique. With the help of this technique, we can derive the distribution of the test statistic.

First, we assume that the sample size is even for simplicity and randomly split samples into two groups: $(X^{(1)}, Y^{(1)})$ and $(X^{(2)}, Y^{(2)})$. In the first stage, we choose p_n based on the adaptively truncated Hotelling T^2 order statistic via the first group sample $(X^{(1)}, Y^{(1)})$. In the second stage, we construct the test statistic via the second group sample $(X^{(2)}, Y^{(2)})$ and p_n .

Next, we show three methods of choosing p_n in a general case.

Denote $\hat{V}_k^{(1)} = \langle \sqrt{n/2}(\hat{\mu}_1^{(1)} - \hat{\mu}_2^{(1)}), \hat{\phi}_k^{(1)} \rangle^2 / \hat{\lambda}_k^{(1)}, k = 1, \dots, d, \dots$. In practice, since many of the trailing eigenvalues are close to zero and \hat{V}_k will be very large for a large k . Hence, we generally give the cutoff d_n a high threshold, for example, $d_n = \max\{k : \hat{\lambda}_k / \sum_{i=1}^k \hat{\lambda}_i > 0.001\}$. For most inference problems, there is no optimal test, but the adaptive Neyman tests have been shown to work well against a broad range of alternatives. Therefore, we choose p_n adaptively based on the following adaptive Neyman test methods. One method is maximizing normalized (8) (in the Appendix A we have proven \hat{V}_k has an approximate χ_1^2 distribution), that is,

$$p_{1n} = \operatorname{argmax}_{1 \leq d \leq d_n} \frac{\sum_{k=1}^d \hat{V}_k^{(1)} - d}{\sqrt{2d}}.$$

Considering some possible nonsignificant $\hat{V}_k^{(1)}$ terms, we propose another method that maximizes the sun of normalized order statistic $V_{(k)}$, that is,

$$p_{2n} = \operatorname{argmax}_{1 \leq d \leq d_n} \frac{\sum_{k=1}^d \hat{V}_{(k)}^{(1)} - E(\sum_{k=1}^d \hat{V}_{(k)}^{(1)})}{\sqrt{\operatorname{var}(\sum_{k=1}^d \hat{V}_{(k)}^{(1)})}},$$

where $\hat{V}_{(k)}^{(1)}$ is the k -th order statistic of $\hat{V}_k^{(1)}, k = 1, \dots, d$ in decreasing order. Unfortunately, there is no closed form since it involves order statistics. However, empirical approximations of this maximum value can be conducted by very fast Monte Carlo approximations. The third method is a hard threshold truncation method, but we truncate at the d -th term based on the percentage of \hat{V}_k , which combines both the variation and the projection along each direction. In our simulation, we set a truncation threshold as [5] projection-based test for comparison.

Remark 1. p_{1n} and p_{2n} are chosen adaptively based on the first group data. For the convenience of follow-up theoretical analysis, we denote them as p_n uniformly.

After we derive p_n in the first stage, we construct the following statistic based on the second group sample as follows:

$$\begin{aligned} T_{1A} &= \sum_{k=1}^{p_n} \hat{V}_k^{(2)}, \\ T_{2A} &= \sum_{k=1}^{p_n} \hat{V}_{(k)}^{(2)}, \\ NT_{1A} &= \frac{\sum_{k=1}^{p_n} \hat{V}_k^{(2)} - p_n}{\sqrt{2p_n}}, \\ NT_{2A} &= \frac{\sum_{k=1}^{p_n} \hat{V}_{(k)}^{(2)} - E(\sum_{k=1}^{p_n} \hat{V}_{(k)}^{(2)})}{\sqrt{\operatorname{var}(\sum_{k=1}^{p_n} \hat{V}_{(k)}^{(2)})}}, \end{aligned}$$

where

$$\hat{V}_k^{(2)} = \langle \sqrt{n/2}(\hat{\mu}_1^{(2)} - \hat{\mu}_2^{(2)}), \hat{\phi}_k^{(2)} \rangle^2 / \hat{\lambda}_k^{(2)}$$

To derive the asymptotic distribution of T_{1A} , T_{2A} , NT_{1A} , and NT_{2A} , we make the following assumptions:

Assumption 1. *There exist constants $a > 1$ and C such that $\lambda_k - \lambda_{k+1} \geq Ck^{-a-1}$ for $k \geq 1$ and $C^{-1}k^{-a} \leq \lambda_k \leq Ck^{-a}$.*

Assumption 2. $\int_I EX^4 < \infty, \int_I EY^4 < \infty$.

Assumption 3. $\tau = \lim_{n \rightarrow \infty} \frac{n_1}{n}, 0 < \tau < 1$.

Assumption 4. $p_n^{2a+3}m^{-1} = o_p(1)$, where $m = \min(n_1, n_2)$.

Assumptions 1 and 2 are regular in functional principle component analysis (FPCA). Assumption 1 implies that $\lambda_k \geq Ck^{-a}$. Because the covariance functions are bounded, one has $a > 1$. Assumption 1 essentially assumes that all the eigenvalues are positive, but decay exponentially. Assumption 3 requires that the two sample sizes n_1, n_2 tend to ∞ proportionally. Assumption 4 specifies the growing rate of p_n . In related literature e.g., [15], to guarantee estimation consistency, p_n is usually assumed to satisfy $p_n^{a+1}n^{-1} = o_p(1)$.

Theorem 1. *Under H_0 , $\sqrt{n}(\bar{X} - \bar{Y}) \xrightarrow{d} G$, where G is a Gaussian process with mean zero and covariance function Γ , where $\Gamma(s, t) = \frac{\Gamma_1(s,t)}{\tau} + \frac{\Gamma_2(s,t)}{1-\tau}$.*

The proof of Theorem 1 follows the trivial central limit of the stochastic process and we omit it.

Remark 2. *We can write G as $G = \sum_{j=1}^{\infty} \eta_j \sqrt{\lambda_j} \varphi_j$, where $\eta_j, j \geq 1$ are i.i.d centered real Gaussian random variables with variance 1.*

Theorem 2. *Under Assumptions 1–4, there exist some increasing sequences $(p_n)_n$ such that under H_0*

$$\lim_{p_n \rightarrow \infty} P(NT_{1A} \leq x) = \Phi(x), \tag{14}$$

where $\Phi(x)$ denotes the cumulative distribution function (cdf) of standard normal distribution.

Theorem 3. *Under Assumptions 1–4, the test statistic T_{2A} is approximately equivalent to $\sum_{k=1}^{p_n} \chi_{(k)}^2$ (1) under H_0 , where $\chi_{(1)}^2(1), \dots, \chi_{(p_n)}^2(1)$ are order statistics (in a decreasing order) of $p_n \chi^2(1)$ random variables.*

Remark 3. *The asymptotic null distribution of T_{2A} is affected not only by the values of $\hat{V}_k^{(2)}$, but also by the order of them. In practice, empirical approximations of quantiles and tail probability of the null distribution of T_{2A} can be deduced by very fast Monte Carlo approximations.*

Theorem 4. *Under Assumptions 1–4 and H_0 ,*

$$\lim_{(m, p_n) \rightarrow \infty} P(NT_{2A} \leq x) = \Phi(x), \tag{15}$$

where $\Phi(x)$ denotes the cdf of the standard normal distribution.

To obtain the asymptotic distribution of NT_{1A} under the alternative in (3), we choose the local alternative, as defined in the following assumption:

Assumption 5.

$$H_{1n} : \mu_1(t) - \mu_2(t) = n^{-\frac{1}{2}}u(t), \tag{16}$$

where $u(t)$ is any fixed real function such that $0 < \| u \| < \infty$.

Then, we have the following asymptotic power of NT_{1A} :

Theorem 5. Under Assumptions 1–6, the asymptotic distribution of the NT_{1A} is given by

$$\lim_{(m,p_n) \rightarrow \infty} P(NT_{1A} > z_{1-\alpha}) = \Phi(-z_{1-\alpha} + \|\Gamma_{12}^{-1}u(t)\|^2),$$

where P denotes that the distribution have been obtained under the alternative, and $z_{1-\alpha}$ is the upper $(1 - \alpha)$ 100% point of the standard normal distribution.

Theorem 6. Under Assumptions 1–5, the distribution of the T_{2A} is approximately equivalent to a noncentral χ^2 distribution $\chi_{p_n}^{2*}(\zeta_0)$, where

$$\zeta_0 = \sum_{k=1}^{p_n} V_k^2,$$

and $V_k^2 = \frac{\langle \sqrt{n/2}(\mu_1 - \mu_2), \varphi_k^{(2)} \rangle^2}{\lambda_k^{(2)}} = \frac{\langle u(t), \varphi_k^{(2)} \rangle^2}{\lambda_k^{(2)}}.$

4. Simulations

In this section, we report some Monte Carlo simulation results to compare the finite samples performance of the classical and proposed methods on the two sample mean testing problem under different settings, including a fixed simple alternative and sparse signals with varying locations.

4.1. Fixed Simple Alternative

In this subsection, we first look at a simple setting where the alternatives are fixed. We generate curves from two populations that are generated by 40 Fourier bases as

$$X(t) = \sum_{k=1}^{40} (\theta_k^{1/2}z_{1k} + \mu_{1k})\phi_k(t), \quad Y(t) = \sum_{k=1}^{40} (\theta_k^{1/2}z_{2k} + \mu_{2k})\phi_k(t).$$

Here, z_{1k}, z_{2k} are independent standard normal random variables. In each case, we take $\phi_k(t) = \sqrt{2}\sin((k - 0.5)\pi t), t \in [0, 1]$ for $k = 1, 2, \dots$ and generate the data on a discrete grid of 100 equispaced points in $[0, 1]$. We took $\theta_k = 1/(\pi(k - 0.5))^2$. We choose μ_{1k} s and μ_{2k} s depending on the property that we want to illustrate (see below). We compare power and size under three different methods: Ref. [8] L_2 -norm based test (T_L); Ref. [13]’s F-test T_F ; Ref. [5]’s projection-based test (T_H) with fixed truncation and our two methods. We choose the commonly used threshold 99% to determine the truncated term in [5]’s projection-based test (T_H). The results are based on 1000 Monte Carlo replications. In all scenarios, we set the nominal size $\alpha = 0.05$.

To cover as many different scenes as possible, we set five different settings referring to the mean difference: (1) the mean differences arise early in the sequence μ_{1k} s and μ_{2k} s, that is, $(\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}, \mu_{15}, \mu_{16}) = (0.5, -0.5, 1.5, -0.5, 1.5, -0.5)$ and $\mu_{1k} = 0$ for $k > 6$, $\mu_{2k} = \mu_{1k} + \delta$ for $k \leq 6$, $\mu_{2k} = 0$ for $k > 6$; (2) the mean differences arise in the middle of the sequence μ_{1k} s and μ_{2k} s, that is, $(\mu_{1,11}, \mu_{1,12}, \mu_{1,13}, \mu_{1,14}, \mu_{1,15}, \mu_{1,16}) = (0.5, -0.5, 1.5, -0.5, 1.5, -0.5)$ and $\mu_{1k} = 0$ for other k , $\mu_{2k} = \mu_{1k} + \delta$ for $11 \leq k \leq 16$, $\mu_{2k} = 0$ for other k ; (3) the mean differences arise in the latter part of the sequence μ_{1k} s and μ_{2k} s, that is, $(\mu_{1,21}, \mu_{1,22}, \mu_{1,23}, \mu_{1,24}, \mu_{1,25}, \mu_{1,26}) = (0.5, -0.5, 1.5, -0.5, 1.5, -0.5)$ and $\mu_{1k} = 0$ for other k , $\mu_{2k} = \mu_{1k} + \delta$ for $21 \leq k \leq 26$, $\mu_{2k} = 0$ for other k ; (4) the mean differ-

ences are scattered in the first, middle and latter part, that is, $(\mu_{1,11}, \mu_{1,12}, \mu_{1,13}, \mu_{1,21}, \mu_{1,22}, \mu_{1,23}) = (0.5, -0.5, 1.5, -0.5, 1.5, -0.5)$ and $\mu_{1k} = 0$ for other k , $\mu_{2k} = \mu_{1k} + \delta$ for $k \in \{1, 2, 11, 12, 21, 22\}$, $\mu_{2k} = 0$ for other k ; (5) the tiny differences appear in all the principal components. In this case, we set μ_{1k} as independent $N(0, 1)$ random variables, and $\mu_{2k} = \mu_{1k} + \delta, 1 \leq k \leq 40$.

From Tables 1–5, we can see that these are obvious different performances in different settings. From Table 1, we can see that when the mean difference lies in early part in the sequence, Ref. [5]’s projection-based test (T_H) has most powerful performance. This should not be surprising, because their method just chooses projection space spanned by the first few eigenfunctions, where the mean difference lies. From Table 2 we observe that when mean difference lies in the middle part in the sequence, our method has very high power compared to T_H, T_F and T_L . Particularly, we notice that Ref. [5]’s projection-based test (T_H) has a dramatic power loss. From Table 3, we can see that when the mean difference lies in the latter part of the sequence, our method still has the best performance. At the same time, we can find that T_F and T_L have higher power than T_H in this case. This illustrates that T_F and T_L are sensitive to divergence degree and T_H is more sensitive to location of mean difference. Furthermore, we notice that the power of T_L and T_F outperform our method only on large sample sizes and large discrepancies between the null hypothesis and alternative hypothesis. This is understandable, because our method also depends on mean difference projection on space spanned by the eigenfunction, excluding last few eigenfunctions. Tables 4 and 5 illustrate more general cases. Table 4 demonstrates that when there are tiny differences in all directions, our method is still the most powerful, while T_L and T_F are useless. Table 5 demonstrates the performance of each method when there are differences in the early part, middle part, and latter part. From the simulation results, we can see that in this general case, our method has the most satisfactory performance.

Table 1. Size and power of five methods in Setting 1 (mean difference lies in early part $\mu_2[1 : 6] = \mu_1[1 : 6] + \delta$).

$n_2 = n_1$	δ	T_L	T_F	T_H	NT_{1A}	NT_{2A}
$n_1 = 120$	0.00	0.067	0.065	0.068	0.045	0.055
	0.10	0.151	0.105	0.461	0.357	0.365
	0.20	0.245	0.258	0.786	0.483	0.471
	0.30	0.405	0.413	0.998	0.872	0.805
$n_1 = 220$	0.00	0.045	0.053	0.066	0.051	0.065
	0.10	0.265	0.253	0.643	0.482	0.471
	0.20	0.385	0.393	0.881	0.564	0.585
	0.30	0.553	0.512	0.999	0.976	0.918

Table 2. Size and power of five methods in Setting 2 (mean difference lies in middle part $\mu_2[11 : 16] = \mu_1[11 : 16] + \delta$).

$n_2 = n_1$	δ	T_L	T_F	T_H	NT_{1A}	NT_{2A}
$n_1 = 120$	0.00	0.052	0.063	0.055	0.067	0.066
	0.10	0.352	0.355	0.384	0.876	0.923
	0.20	0.445	0.466	0.448	0.998	0.999
	0.30	0.581	0.592	0.544	0.999	0.999
$n_1 = 220$	0.00	0.034	0.045	0.066	0.068	0.063
	0.10	0.442	0.461	0.583	0.975	0.923
	0.20	0.565	0.543	0.654	0.999	0.999
	0.30	0.675	0.691	0.685	0.999	0.999

Table 3. Size and power of five methods in Setting 3 (mean difference lies in latter part $\mu_2[21 : 26] = \mu_1[21 : 26] + \delta$).

$n_2 = n_1$	δ	T_L	T_F	T_H	NT_{1A}	NT_{2A}
$n_1 = 120$	0.00	0.035	0.052	0.054	0.055	0.042
	0.10	0.342	0.355	0.372	0.486	0.497
	0.20	0.565	0.552	0.593	0.744	0.795
	0.30	0.725	0.753	0.765	0.996	0.985
$n_1 = 220$	0.00	0.065	0.061	0.045	0.052	0.055
	0.10	0.431	0.425	0.456	0.527	0.588
	0.20	0.628	0.635	0.681	0.823	0.885
	0.30	0.864	0.875	0.824	0.999	0.999

Table 4. Size and power of five methods in Setting 4 (mean difference lies in separable part $\mu_2[1 : 2] = \mu_1[1 : 2] + \delta, \mu_2[11 : 12] = \mu_1[11 : 12] + \delta, \mu_2[21 : 22] = \mu_1[21 : 22] + \delta$).

$n_2 = n_1$	δ	T_L	T_F	T_H	NT_{1A}	NT_{2A}
$n_1 = 120$	0.00	0.045	0.053	0.067	0.068	0.062
	0.10	0.162	0.165	0.326	0.463	0.554
	0.20	0.205	0.272	0.466	0.963	0.955
	0.30	0.361	0.32	0.565	0.999	0.999
$n_1 = 220$	0.00	0.045	0.038	0.067	0.045	0.068
	0.10	0.192	0.215	0.393	0.497	0.605
	0.20	0.282	0.314	0.516	0.999	0.999
	0.30	0.461	0.465	0.689	1.000	1.000

Table 5. Size and power of five methods in Setting 5 (difference averaged in total vector $\mu_2 = \mu_1 + \delta \mathbf{1}_{40}$).

$n_2 = n_1$	δ	T_L	T_F	T_H	NT_{1A}	NT_{2A}
$n_1 = 120$	0.00	0.045	0.052	0.068	0.065	0.053
	0.10	0.182	0.179	0.366	0.794	0.836
	0.20	0.372	0.364	0.854	0.999	0.999
	0.30	0.577	0.564	0.896	1.000	1.000
$n_1 = 220$	0.00	0.036	0.045	0.062	0.063	0.065
	0.10	0.212	0.249	0.905	0.826	0.887
	0.20	0.413	0.424	0.935	0.999	0.999
	0.30	0.625	0.636	0.943	1.000	1.000

We also conducted simulation studies under other similar scenarios. As they demonstrated similar patterns to those discussed above, we omit them here to save space.

It is worth noting that the proposed method has a first stage with randomly split data. There could be a potential limitation with large randomness. In order to understand the robustness for this splitting, we perform some supporting simulation studies, again considering multi-fold cross validation (CV), including two-fold CV, five-fold CV, and ten-fold CV. For convenience, we use the same data setting of Table 5. The results are shown in Table 6. From Table 6, we can see that in most cases, the hypothesis is robust for this splitting.

Table 6. Size and power of NT_{2A} in Setting 5 (difference averaged in total vector $\mu_2 = \mu_1 + \delta \mathbf{1}_{40}$).

$n_2 = n_1 = 120$	δ			
	0.00	0.1	0.2	0.3
two-fold CV	0.058	0.468	0.786	0.947
five-fold CV	0.056	0.459	0.772	0.943
ten-fold CV	0.053	0.465	0.765	0.978

4.2. Sparse Signals with Varying Locations

In this section, we demonstrate the performance of different tests under signals with varying locations. We set $\mu_{1k}(1 \leq k \leq 40)$ as independent $U(0,1)$ random variables, and $\mu_{2k} = \mu_{1k} + \delta$, where k is six random locations out of 40. In Setting 1, the signal of difference appears randomly in six of the first twenty principal components; we denote $M(20,6)$. In Setting 2, the signal of difference appears in six of forty principal components; we denote $M(40,6)$. The simulation results are illustrated in Tables 7 and 8. From the simulation results, we can see that in these cases, our method also have the most satisfactory performance.

We also compare our method with the full-sample adaptive methods, including the adaptive Neyman–Pearson test (T_A) and the ordered adaptive test T_{OA} . In the full sample, the distribution of adaptive Neyman–Pearson test (T_A) and ordered adaptive test T_{OA} is notable, that is,

$$T_A = \max_{1 \leq d \leq p_n} \frac{\sum_{k=1}^d \hat{V}_k^{(1)} - d}{\sqrt{2d}},$$

$$T_{OA} = \max_{1 \leq d \leq p_n} \frac{\sum_{k=1}^d \hat{V}_{(k)}^{(1)} - E(\sum_{k=1}^d \hat{V}_{(k)}^{(1)})}{\sqrt{\text{var}(\sum_{k=1}^d \hat{V}_{(k)}^{(1)})}}.$$

We use permutation method to calculate size and power. Table 9 illustrates the simulation results, where Time1 is the time of running once T_A and T_{OA} and Time2 is the time to run NT_{1A} and NT_{2A} once. The unit is seconds. From Table 9, we can see that our splitting sample methods have a slight power loss compared to the adaptive Neyman–Pearson test T_A and the ordered adaptive test T_{OA} . However, we can save significant time in real computing.

Table 7. Size and power of five methods under randomized signal in Setting 1 $M(20,6)$.

$n_2 = n_1$	δ	T_L	T_F	T_H	NT_{1A}	NT_{2A}
$n_1 = 120$	0.00	0.034	0.035	0.046	0.057	0.065
	0.10	0.262	0.275	0.366	0.594	0.606
	0.20	0.375	0.384	0.589	0.926	0.935
	0.30	0.534	0.548	0.825	0.999	0.999
$n_1 = 220$	0.00	0.067	0.056	0.063	0.066	0.068
	0.10	0.309	0.315	0.417	0.628	0.733
	0.20	0.423	0.439	0.615	0.986	0.995
	0.30	0.674	0.695	0.923	0.999	0.999

Table 8. Size and power of five methods under randomized signal in Setting 2 $M(40,6)$.

$n_2 = n_1$	δ	T_L	T_F	T_H	NT_{1A}	NT_{2A}
$n_1 = 120$	0.000	0.063	0.038	0.067	0.065	0.063
	0.1	0.275	0.286	0.447	0.457	0.478
	0.2	0.355	0.372	0.743	0.785	0.799
	0.3	0.426	0.465	0.878	0.943	0.966
$n_1 = 220$	0.00	0.042	0.053	0.064	0.066	0.065
	0.1	0.323	0.376	0.524	0.539	0.567
	0.2	0.465	0.486	0.874	0.923	0.995
	0.3	0.549	0.563	0.925	0.999	0.999

Table 9. Size and power of seven methods under randomized signal in Setting 2 $M(40, 6)$.

$n_2 = n_1$	δ	T_L	T_F	T_H	T_A	T_{OA}	NT_{1A}	NT_{2A}	Time1	Time2
120	0.00	0.045	0.052	0.045	0.063	0.065	0.064	0.061	41.4575	1.1953
	0.10	0.221	0.236	0.252	0.315	0.403	0.224	0.367	41.6208	1.1964
	0.20	0.415	0.454	0.621	0.883	0.966	0.749	0.851	37.4070	1.0789
	0.30	0.626	0.684	0.827	0.994	0.975	0.891	0.896	36.4592	1.0681
180	0.00	0.051	0.062	0.054	0.057	0.054	0.063	0.067	34.6782	1.4470
	0.10	0.293	0.296	0.305	0.483	0.491	0.317	0.428	41.6208	1.1964
	0.20	0.495	0.521	0.715	0.936	0.995	0.836	0.922	37.4070	1.0789
	0.30	0.721	0.784	0.935	0.999	0.999	0.966	0.975	36.4592	1.0681

5. Application

In this section, we apply our proposed hypothesis testing procedures to a real PM 2.5 dataset in Beijing, Tianjin, and Shijiazhuang between January 2017 and December 2019. The dataset was downloaded from the website <http://www.tianqihoubao.com/aqi/>, accessed on 10 June 2021. The data readings were taken every day, so the total data size is 1085 for every city. Beijing is surrounded by Tianjin and Shijiazhuang. Therefore, we want to know more about the average PM 2.5 difference in these three areas. The following Figures 1 and 2 show the mean PM 2.5 ($\mu\text{g}/\text{m}^3$) in Beijing, Tianjin, and Hebei Province in different time periods. There are some missing days in some cycles. Note that Figure 1 shows negative values at the beginning for a measure that is always greater than zero because of B-spline approximation.

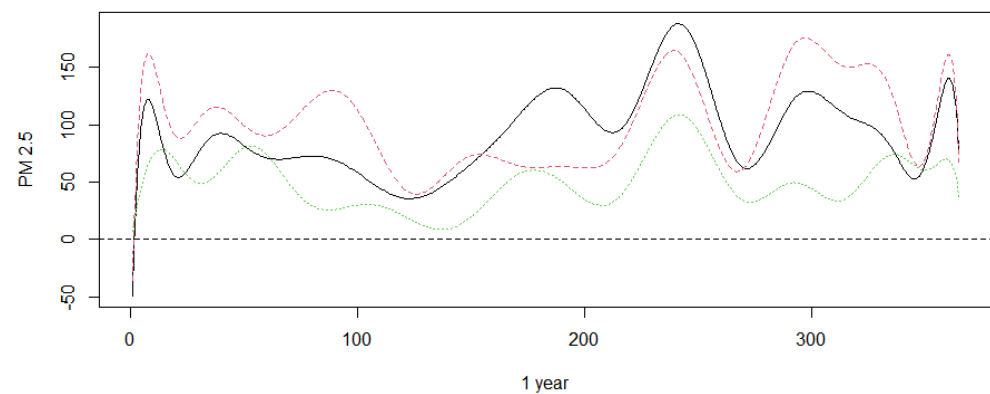


Figure 1. The mean PM 2.5 ($\mu\text{g}/\text{m}^3$) of Beijing, Tianjin, and Shijiazhuang from December 2019 to January 2019. The black line stands for Beijing. The red dashed line stands for Tianjin. The green dotted line stands for Shijiazhuang.

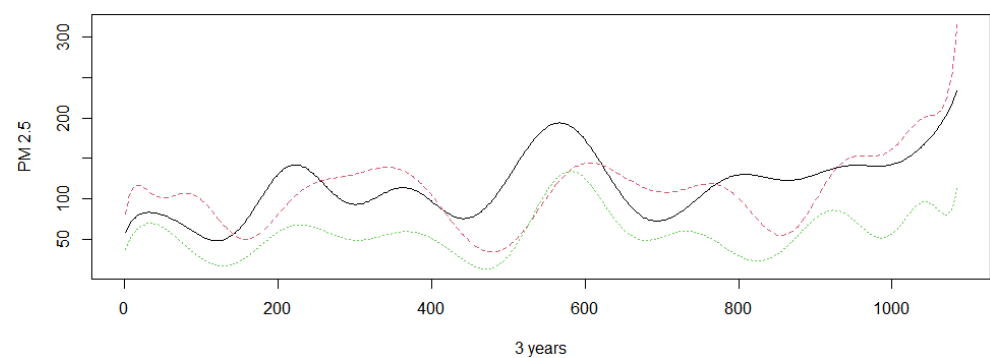


Figure 2. The mean PM 2.5 ($\mu\text{g}/\text{m}^3$) of Beijing, Tianjin, and Shijiazhuang from December 2019 to January 2017. The black line stands for Beijing. The red dashed line stands for Tianjin. The green dotted line stands for Shijiazhuang.

It is obvious that PM 2.5 changes over individual periods. Here, we test whether there is a significant difference in PM 2.5 among the three cities using the method we proposed. First, the sample is divided into two data sets: the training sample is the dataset in 2017; the test sample is the dataset in 2018–2019. The principle components are adaptively based on the training sample. Then, the test statistic is constructed via the test sample and the principle components are selected by the training sample. To test whether there is significant difference in PM 2.5 between the three cities, we carry out permutations 1000 times within each group to calculate the rejection proportions; then, we obtain the p -value of the test. The results are shown in Table 10.

Table 10. p -value of two tests.

	Beijing vs. Tianjin		Beijing vs. Shijiazhuang		Tianjin vs. Shijiazhuang	
Test	NT_{1A}	NT_{2A}	NT_{1A}	NT_{2A}	NT_{1A}	NT_{2A}
p -value	0.035	0.034	0.025	0.032	0.039	0.045

From Table 10, we can see that all p -values are less than 0.05. The tests are statistically significant and suggest that the average PM 2.5 in these three areas are different from each other at a 0.05 level of significance.

6. Conclusions and Discussions

In this paper, we consider the problem of testing the equality of mean functions in two random samples independently drawn from two functional random variables. We develop and study a novel testing procedure that has a more powerful ability to detect mean difference. In general, it includes two stages: first, splitting the sample into two parts and selecting principle components adaptively based on the first half-sample; then, constructing the test statistic based on another half-sample. An extensive simulation study is presented, which shows that the proposed test works very well in comparison with several other methods in a variety of settings. Our future project is to detect differences in the covariance functions of independent sample curves. There have been some approaches proposed so far to address this problem, for instance, the factor-based test proposed by [4] and the regularized M -test introduced by [16].

Author Contributions: Data curation, S.F.; Funding acquisition, S.F.; Investigation, J.Z.; Methodology, J.Z.; Project administration, Y.H.; Validation, Y.H.; Writing—original draft, J.Z.; Writing—review & editing, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China National Institute of Standardization through the “Special funds for basic R&D undertakings by welfare research institutions”(522022Y-9402), the Humanities and Social Science Project of the Ministry of Education of China (21YJC910003), the Foundation of Henan Educational Committee (No. 21A910004), and the Natural Science Foundation of Henan Province (No. 212300410412).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs of the Main Results

Before proving the main results, we introduce the following useful lemmas. Furthermore, for the convenience of notation, we give proofs in the full sample.

Lemma A1. *Under Assumptions 1–4, we have*

$$\frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n \langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\hat{\lambda}_k} = \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n \langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\lambda_k} + o_p(1). \tag{A1}$$

Proof. Denote $\varepsilon_p(n) = \{\hat{\Delta} = \|\Gamma_{12} - \hat{\Gamma}_{12}\| \leq \frac{1}{2}\lambda_{p_n}\}$; note that provided $\varepsilon_p(n)$ holds, we have

$$\begin{aligned} & \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\hat{\lambda}_k} - \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\lambda_k} \\ &= \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n(\lambda_k - \hat{\lambda}_k)}{\hat{\lambda}_k \lambda_k} \langle \bar{f} - \bar{g}, \hat{\varphi}_k \rangle^2 \\ &\leq \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \bar{X} - \bar{Y}, \hat{\varphi}_k \rangle^2}{\hat{\lambda}_k \lambda_k} \sup |\hat{\lambda}_k - \lambda_k| \\ &\leq \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \bar{X} - \bar{Y}, \hat{\varphi}_k \rangle^2}{\hat{\lambda}_k^2} \|\Gamma_{12} - \hat{\Gamma}_{12}\|. \end{aligned}$$

It can be proven easily that $E\|\Gamma_1 - \hat{\Gamma}_1\|^2 = O(n_1^{-1})$, $E\|\Gamma_2 - \hat{\Gamma}_2\|^2 = O(n_2^{-1})$; then, $\|\Gamma_{12} - \hat{\Gamma}_{12}\| = O_p(m^{-1/2})$.

According to central limit theory, we have

$$\frac{n\langle \bar{X} - \bar{Y}, \hat{\varphi}_k \rangle}{\sqrt{\hat{\lambda}_k}} \xrightarrow{d} N(0, 1), \tag{A2}$$

Then, $\frac{n\langle \bar{X} - \bar{Y}, \hat{\varphi}_k \rangle^2}{\hat{\lambda}_k} \xrightarrow{d} \chi_1^2$, which means $\frac{n\langle \bar{X} - \bar{Y}, \hat{\varphi}_k \rangle^2}{\hat{\lambda}_k}$ is bounded in probability.

Notice that $|\hat{\lambda}_k - \lambda_k| = O_p(m^{-1/2})$; therefore

$$\begin{aligned} & \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\hat{\lambda}_k} - \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\lambda_k} \\ &\leq O_p\left(\frac{1}{\sqrt{2p_n}} m^{-1/2} p_n^{a+1}\right) = o_p(1). \end{aligned}$$

Ref. [15] has proven that $P(\varepsilon_p) \rightarrow 1$ as $m \rightarrow \infty$; thus, Lemma A1 holds. \square

Lemma A2. Under Assumptions 1–4, we have

$$\frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\lambda_k} = \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle^2}{\lambda_k} + o_p(1). \tag{A3}$$

Proof. First,

$$\begin{aligned} & \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\lambda_k} \\ &= \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k - \varphi_k + \varphi_k \rangle^2}{\lambda_k} \\ &= \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \left\{ \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k - \varphi_k \rangle^2}{\lambda_k} \right\} + \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \left\{ \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle^2}{\lambda_k} \right\} \\ & \quad + \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \left\{ 2 \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k - \varphi_k \rangle \langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle}{\lambda_k} \right\} \end{aligned}$$

Then,

$$\begin{aligned} & \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle^2}{\lambda_k} - \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle^2}{\lambda_k} \\ = & \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \left\{ \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k - \varphi_k \rangle^2}{\lambda_k} \right\} + \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \left\{ 2 \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k - \varphi_k \rangle \langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle}{\lambda_k} \right\} \\ = & I_1 + I_2. \end{aligned}$$

where

$$\begin{aligned} I_1 &= \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \left\{ \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k - \varphi_k \rangle^2}{\lambda_k} \right\}, \\ I_2 &= \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \left\{ 2 \frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k - \varphi_k \rangle \langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle}{\lambda_k} \right\}. \end{aligned}$$

It is obvious that

$$\sum_{k=1}^{p_n} \langle \sqrt{n}(\bar{X} - \bar{Y}), \hat{\varphi}_k - \varphi_k \rangle^2 \lambda_k^{-1} \leq \sum_{k=1}^{p_n} \|\sqrt{n}(\bar{f} - \bar{g})\|^2 \cdot \|\hat{\varphi}_k - \varphi_k\|^2 \lambda_k^{-1}.$$

It can also be easily proven that $\|\sqrt{n}(\bar{X} - \bar{Y})\|^2 = O_p(1)$. According to the result of [17], we have $\|\hat{\varphi}_k - \varphi_k\| = O_p(km^{-1/2})$ under corresponding conditions. Then, we have

$$I_1 = O_p\left(\frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} k^2 m^{-1} k^a\right) = O_p\left(\frac{1}{\sqrt{2p_n}} p_n^3 m^{-1}\right) = o_p(1).$$

Similarly, we can prove $I_2 = o_p(1)$. □

Lemma A3. Under Assumptions 1–4, we have

$$T^* = \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \left[\frac{n\langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle^2}{\lambda_k} - 1 \right]$$

which converges in distribution to a centered Gaussian random variable g with variance 1.

The proof of Lemma A3 is similar to the techniques used by [18], so we omit it here.

Proof of Theorem 2. Combine Lemmas A1 and A2 with Lemma A3; then, can proof Theorem 2. □

Proof of Theorem 3. According to Theorem 1, Lemmas A1 and A2, we have $\hat{V}_k \sim N(0, 1)$ and $\hat{V}_k^2 \sim \chi^2(1)$. Then, the conclusion is obvious. □

Proof of Theorem 5. We note that

$$\begin{aligned} n\langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle^2 &= n\langle \hat{\mu}_1 - \hat{\mu}_2 - \mu_1 + \mu_2, \varphi_k \rangle^2 \\ &\quad + 2n\langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle \langle \mu_1 - \mu_2, \varphi_k \rangle - n\langle \mu_1 - \mu_2, \varphi_k \rangle^2, \\ &= J_{k1} + 2J_{k2} - J_{k3}. \end{aligned}$$

where $J_{k1} = n\langle \hat{\mu}_1 - \hat{\mu}_2 - \mu_1 + \mu_2, \varphi_k \rangle^2$, $J_{k2} = n\langle \hat{\mu}_1 - \hat{\mu}_2, \varphi_k \rangle \langle \mu_1 - \mu_2, \varphi_k \rangle$, $J_{k3} = n\langle \mu_1 - \mu_2, \varphi_k \rangle^2$. Then

$$\frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \hat{V}_k = \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} (J_{k1} + 2J_{k2} - J_{k3}) / \hat{\lambda}_k.$$

Observe that

$$\frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} (J_{k2} - J_{k3}) / \hat{\lambda}_k = \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} n \langle \mu_1 - \mu_2, \varphi_k \rangle \langle \hat{\mu}_1 - \hat{\mu}_2 - \mu_1 + \mu_2, \varphi_k \rangle / \hat{\lambda}_k.$$

According to (A2), we have

$$n \langle \hat{\mu}_1 - \hat{\mu}_2 - \mu_1 + \mu_2, \varphi_k \rangle / \sqrt{\hat{\lambda}_k} = O_p(1).$$

By Assumptions 1–5 and Lemma A1, we have that

$$\begin{aligned} & \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \langle \mu_1 - \mu_2, \varphi_k \rangle / \hat{\lambda}_k \\ \leq & \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} n^{-\frac{1}{2}} \|u(t)\| / \lambda_k = O_p \left(\frac{n^{-\frac{1}{2}+b} \sum_{k=1}^{p_n} k^a}{\sqrt{2p_n}} \right) \\ = & O_p \left(n^{-\frac{1}{2}} p_n^{a+1} \right) = o_p(1). \end{aligned}$$

Under Assumptions 4 and 5, we have

$$\begin{aligned} \lim_{p_n \rightarrow \infty} \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} J_{k2} / \hat{\lambda}_k &= \lim_{p_n \rightarrow \infty} \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} n \langle \hat{\mu}_1 - \hat{\mu}_2, \hat{\varphi}_k \rangle \langle \mu_1 - \mu_2, \hat{\varphi}_k \rangle / \hat{\lambda}_k \\ &= \lim_{p_n \rightarrow \infty} \frac{1}{\sqrt{2p_n}} \sum_{k=1}^{p_n} \frac{n \langle \mu_1 - \mu_2, \varphi_k \rangle^2}{\lambda_k} \\ &= \lim_{p_n \rightarrow \infty} \sum_{k=1}^{p_n} \langle u(t), \varphi_k \rangle^2 / \lambda_k \\ &= \|\Gamma_{12}^{-1} u(t)\|^2. \end{aligned} \tag{A4}$$

From Theorem 2 and the above results, we have

$$\begin{aligned} P(NT_{1A} \geq z_{1-\alpha}) &= P \left(\frac{\sum_{k=1}^{p_n} \hat{V}_k - p_n}{\sqrt{2p_n}} \geq z_{1-\alpha} \right) \\ &= P \left(\frac{\sum_{k=1}^{p_n} J_{k1} / \hat{\lambda}_k - p_n}{\sqrt{2p_n}} \geq z_{1-\alpha} - \frac{\sum_{k=1}^{p_n} J_{k2} / \hat{\lambda}_k}{\sqrt{2p_n}} \right) \end{aligned}$$

From (A2) we can obtain $\frac{J_{k1}}{\hat{\lambda}_k} \xrightarrow{d} \chi_{1^2}^2$; then,

$$\lim_{(m, p_n) \rightarrow \infty} P \left(\frac{\sum_{k=1}^{p_n} J_{k1} / \hat{\lambda}_k - p_n}{\sqrt{2p_n}} \leq x \right) = \Phi(x), \tag{A5}$$

Combined with (A4), we have

$$\lim_{(m, p_n) \rightarrow \infty} P(NT_{1A} > z_{1-\alpha}) = \Phi(-z_{1-\alpha} + \|\Gamma_{12}^{-1}u(t)\|^2).$$

□

Proof of Theorem 6. By Lemma A2, we have $(\sum_{k=1}^{p_n} \hat{V}_k - \sum_{k=1}^{p_n} V_k) \xrightarrow{P} 0$ as $n \rightarrow \infty$. Define $(k_1^*, \dots, k_{p_n}^*)$ as decreasing orders of V_1, \dots, V_{p_n} and (k_1, \dots, k_{p_n}) as decreasing orders of $\hat{V}_1, \dots, \hat{V}_{p_n}$. Ref. [19] have proven that the random orders in the selection procedure $\hat{V}_{(1)}, \dots, \hat{V}_{(p_n)}$ are asymptotically equivalent to fixed orders $V_{(1)}, \dots, V_{(p_n)}$. □

References

1. Besse, P.; Ramsay, J. Principal components analysis of sampled functions. *Psychometrika* **1986**, *51*, 285–311. [\[CrossRef\]](#)
2. Rice, J.A.; Silverman, B.W. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1991**, *53*, 233–243. [\[CrossRef\]](#)
3. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*; Springer Press: Berlin/Heidelberg, Germany, 2005.
4. Ferraty, F.; Vieu, P.; Viguier-Pla, S. Factor-based comparison of groups of curves. *Comput. Stat. Data Anal.* **2007**, *51*, 4903–4910. [\[CrossRef\]](#)
5. Haróath, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer Press: Berlin/Heidelberg, Germany, 2012.
6. Fan, J.Q.; Lin, S.K. Test of significance when data are curves. *J. Am. Stat. Assoc.* **1998**, *93*, 1007–1021. [\[CrossRef\]](#)
7. Faraway, J.J. Regression analysis for a functional response. *Technometrics* **1997**, *39*, 254–261. [\[CrossRef\]](#)
8. Zhang, C.Q.; Peng, H.; Zhang, J.T. Two samples tests for functional data. *Commun. Stat.-Methods* **2010**, *39*, 559–578. [\[CrossRef\]](#)
9. Zhang, J.T. Statistical inferences for linear models with functional responses. *Stat. Sin.* **2011**, *21*, 1431–1451. [\[CrossRef\]](#)
10. Zhang, J.T.; Liang, X. One-way ANOVA for functional data via globalizing the pointwise F-test. *Scand. J. Stat.* **2014**, *45*, 51–71. [\[CrossRef\]](#)
11. Zhang, J.T.; Cheng, M.Y.; Wu, H.T.; Zhou, B. A new test for functional one-way ANOVA with applications to ischemic heart screening. *Comput. Stat. Data Anal.* **2019**, *132*, 3–17. [\[CrossRef\]](#)
12. Wasserman, L.; Roeder, K. High dimensional variable selection. *Ann. Stat.* **2009**, *37*, 2178–2201. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Shen, Q.; Faraway, J.L. An F test for linear models with functional responses. *Stat. Sin.* **2004**, *14*, 1239–1257.
14. Satterthwaites, F.E. Synthesis of variance. *Psychometrika*. **1941**, *6*, 309–3167. [\[CrossRef\]](#)
15. Hall, P.; Hosseini-Nasab, M. On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2006**, *68*, 109–126. [\[CrossRef\]](#)
16. Kraus, D.; Panaretos, V.M. Dispersion operators and resistant second-order functional data analysis. *Biometrika* **2012**, *99*, 813–832. [\[CrossRef\]](#)
17. Kong, D.; Xue, K.; Yao, F.; Zhang, H.H. Partially functional linear regression in high dimensions (supplementary material). *Biometrika* **2016**, *103*, 147–159. [\[CrossRef\]](#)
18. Shang, Y.L. A central limit theorem for randomly indexed m-dependent random variables. *Filomat* **2012**, *26*, 713–717. [\[CrossRef\]](#)
19. Su, Y.R.; Di, C.Z.; Li, H. Hypothesis testing in functional linear models. *Biometrics* **2017**, *73*, 551–561. [\[CrossRef\]](#) [\[PubMed\]](#)