Statistica Sinica Preprint No: SS-2023-0272	
Title	Functional Two-Sample Test Based on Projection
Manuscript ID	SS-2023-0272
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0272
Complete List of Authors	Yang Bai, Caihong Qin and Huichen Zhu
Corresponding Authors	Yang Bai
E-mails	statbyang@mail.shufe.edu.cn
Notice: Accepted version subject to English editing.	

Functional Two-Sample Test based on Projection

Yang Bai[1], Caihong Qin[1] and Huichen Zhu[2]

[1]*Shanghai University of Finance and Economics and*

[2]*The Chinese University of Hong Kong*

Abstract: Two-sample inference for population mean functions is a fundamental problem in functional data analysis. In recent years, projection-based testing has gained popularity, which constructs a test statistic by projecting functional observations into a finite-dimensional space. However, the criterion for selecting projection functions remains an open question, given the various types of functional spaces. In this paper, we introduce a novel measure of information loss caused by projection and provide the first theoretical analysis of the relationship between testing efficiency and the selection of projection functions. This analysis contributes to the understanding of projection-based testing and provides guidelines for selecting projection functions. Specifically, we derive the theoretical optimal projective space that achieves the best power and investigate three practical projective spaces. And the tests based on these three projective spaces exhibit superior performance in simulations and real data.

Key words and phrases: Information loss, optimal projection function, projection-based testing, selection of projection function, two-sample test

1. Introduction

As the application of functional data becomes increasingly diverse in various practical settings, the demand for efficient statistical inference methods has grown correspondingly. A key area of research in functional data analysis (FDA) focuses on the challenge of two-sample inference for population mean functions. A considerable amount of research has been dedicated to exploring methods for evaluating the differences between two samples.

Initially, the pointwise t-test (Ramsay and Silverman, 2005) simplified the problem by testing the null hypothesis against the alternative hypothesis at each time point separately. However, this approach is time-consuming and does not guarantee overall significance of the null hypothesis at a given significance level, even if the pointwise test is significant at each time point. To address these issues, global testing methods were developed by summarizing the statistics (Fan and Lin, 1998; Zhang et al., 2019) or the p-values (Cox and Lee, 2008) of the pointwise t-test. Later, the overall difference between mean functions was tested using the L_2-norm-based test (Faraway, 1997; Benko et al., 2009; Horváth et al., 2013) and the F-type test (Shen and Faraway, 2004; Laukaitis and Račkauskas, 2005; Estévez-Pérez and Vilar, 2013) by integrating the sample mean difference function over the entire time range. Further details on these methods can be found in Zhang (2013).

Additionally, the Maximum Mean Discrepancy (MMD) (Wynne and Duncan, 2022) has been used to detect differences in functional distributions, which can also be applied to functional mean testing problems under certain assumptions. These classic methods are typically straightforward to operate and implement, but they often lack sufficient testing efficiency.

In recent years, projection-based testing has gained popularity as a method to construct a test statistic by projecting functional observations into a finite-dimensional space. Initially, random projection (Cuesta-Albertos and Febrero-Bande, 2010; Meléndez et al., 2021) was proposed to convert functional data into scalars and then proceed as in the standard case. However, this approach has two main drawbacks: instability of testing results and loss of information. To address the instability issue, certain projection functions such as basis functions and eigenfunctions have been proposed. For instance, Górecki and Smaga (2015) proposed a test based on a basis function representation, while Horváth et al. (2013) and Pomann et al. (2016) projected observations into the linear space spanned by the leading eigenfunctions of the estimated sample covariance function, and the number of eigenfunctions used is determined by the fraction of variation explained, which is usually around 85%. Additionally, Pini et al. (2018) proposed a generalized Hotelling's T^2 test (GHT) in separable Hilbert spaces, which

is equivalent to using all eigenfunctions with non-zero eigenvalues as the projection functions. And Wang et al. (2022) proposed a testing method based on eigenfunctions chosen via hard thresholding and also introduced a power enhancement component to improve the power. However, the crucial problem of controlling information loss still remains unaddressed. To address this issue, it is necessary to provide a measure of useful information and analyze the relationship between testing efficiency and the selection of projection functions, which is the main focus of this paper.

In this paper, we make several key contributions to the field of projection-based testing. We begin with a comprehensive theoretical analysis that explores how the selection of projection functions impacts the testing power function, deepening the understanding of its relationship with test efficiency. Additionally, we introduce a novel measure of information loss due to projection and provide guidelines for selecting projection functions, thereby enhancing the power of projection-based testing. Furthermore, we derive the most powerful projection function and conduct a thorough investigation into three practical projection spaces, guided by our developed selection criteria. The effectiveness of these methods is validated through our simulation results. Overall, our research offers new insights into projection-based testing and proposes practical solutions to improve testing efficiency.

The rest of the paper is organized as follows. In Section 2, we consider general projection functions that are independent of observations, and analyze the relationship between testing efficiency and selection of projection functions. In Section 3, three types of practical projection functions are selected based on the information criteria proposed in Section 2.3. Simulation studies and illustrative examples are presented in Section 4 and 5, respectively. Discussion and conclusions are given in Section 6. The proofs are provided in the Supplementary Material.

2. Theoretical Projection Testing

2.1 Problem Formulation

We assume that functional observations are random functions sampled from an $L^2(\mathcal{T})$ space equipped with the inner product $\langle f, g\rangle = \int_{\mathcal{T}} f(t)g(t)dt$ and norm $\|f\| = \langle f, f\rangle^{1/2}$, where the domain $\mathcal{T} \subseteq \mathbb{R}$ is a compact interval, and for simplicity, we take $\mathcal{T} = [0, 1]$. We consider two independent random samples $\{\mathbf{y}_{ij}\}_{j=1}^{n_i}, i = 1, 2$. For each random sample, we assume that $\mathbf{y}_{i1}, \ldots, \mathbf{y}_{in_i}$ are independently and identically distributed (i.i.d.) from Gaussian processes $Y_i(t) = \mathrm{GP}(\boldsymbol{\mu}_i, \mathcal{K})$, characterized by a mean function $\boldsymbol{\mu}_i$ and a covariance operator $\mathcal{K}$. The sample sizes are n_1 and n_2. Although these two Gaussian processes share the same covariance opera-

tor, they may possess different mean functions. The covariance operator $\mathcal{K}$ is defined as the integral operator that $(\mathcal{K}f)(\cdot) = \int_{\mathcal{T}} \kappa(\cdot,t)f(t)dt$, with $\kappa(s,t) = \text{cov}\{Y_i(t), Y_i(s)\}$. For any function f and a vector of d functions denoted by $\Phi_d = (\phi_1, \ldots, \phi_d)^{\mathrm{T}}$, we define $\Phi_d f = (\langle\phi_1, f\rangle, \ldots, \langle\phi_d, f\rangle)^{\mathrm{T}}$, and $\Phi_d \mathcal{K} \Phi_d^{\mathrm{T}}$ as a $d \times d$ matrix, where the (k_1, k_2)th element is $\langle\phi_{k_1}, \mathcal{K}\phi_{k_2}\rangle$. With a slight abuse of notation, we represent Φ_d as Φ in the following text. Our objective is to test whether the two mean functions differ. Specifically, we aim to test the null hypothesis,

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \tag{2.1}$$

against the alternative $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are square-integrable. The null hypothesis implies that $\int (\boldsymbol{\mu}_1(t) - \boldsymbol{\mu}_2(t))^2 dt = 0$.

Given the infinite dimension of the functional data, we adopt a testing procedure based on d-dimensional projections of the form $\mathbf{x}_{ij} = \Phi \mathbf{y}_{ij}$, where $1 \leq d < n = n_1 + n_2 - 2$. The projective space is a linear space spanned by $\phi_1, \ldots, \phi_d$. By projection, the infinite-dimensional functional data $\mathbf{y}_{ij}$ is reduced to a d-dimensional data $\mathbf{x}_{ij}$. Since $\mathbf{y}_{ij}$ is from a Gaussian process, $\mathbf{x}_{ij}$ is a d-dimensional random vector, which is normally distributed with mean $\boldsymbol{\nu}_i = \Phi \boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma} = \Phi \mathcal{K} \Phi^{\mathrm{T}}$ for $i = 1, 2$. Based on the

projective space, the null hypothesis (2.1) can be reformulated as follows:

$$H_0^{\Phi} : \boldsymbol{\nu}_1 = \boldsymbol{\nu}_2. \tag{2.2}$$

We will show in the following theorem that rejecting H_0^{Φ} leads to the rejection of H_0. To obtain a powerful test for H_0, the projective space Φ and test statistics should be carefully chosen. In the remaining of this section, we assume that the projective space spanned by the projection function does not depend on data.

Theorem 1. *If the projection functions Φ satisfy that $\Phi\boldsymbol{\mu} \neq 0$ if $\boldsymbol{\mu} \neq 0$, where $\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, then H_0^{Φ} holds if and only if H_0 holds. Moreover, the corresponding power functions are equivalent, i.e., $\beta_H(\cdot) = \beta_{H^{\Phi}}(\cdot)$, where $\beta_H(\cdot)$ represents the probability of rejecting H_0 given H is true.*

2.2 Test Statistics and Power Function

Through Φ, we project the infinite-dimensional functional data into a d-dimensional projective space. When d is fixed and both the independence and the normality assumptions of $\mathbf{x}_{ij}$ hold, the classical Hotelling's T^2-test is the most powerful invariant test (Anderson, 2003). Therefore, we apply Hotelling's T^2 test with equal variance-covariance matrix, and the

test statistic is defined as

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^{\mathrm{T}} \widehat{\mathbf{\Sigma}}^{-1} (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2), \tag{2.3}$$

where $\overline{\mathbf{x}}_i = \Phi \overline{\mathbf{y}}_i = \sum_{j=1}^{n_i} \Phi \mathbf{y}_{ij}/n_i$ is the projected sample mean, and $\widehat{\mathbf{\Sigma}} = \Phi \widehat{\mathcal{K}} \Phi^{\mathrm{T}}$ is the projected sample variance-covariance matrix with $\widehat{\mathcal{K}} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \overline{\mathbf{y}}_i)(\mathbf{y}_{ij} - \overline{\mathbf{y}}_i)^{\mathrm{T}}/(n_1 + n_2 - 2)$.

Under the alternative hypothesis $H_1^{\Phi} : \boldsymbol{\nu} = \boldsymbol{\nu}_1 - \boldsymbol{\nu}_2 \neq \mathbf{0}$, the exact distribution of the test statistics is,

$$\frac{n - d + 1}{nd} T^2 \sim F_{d, n-d+1}(\delta),$$

where $F_{d,n-d+1}(\delta)$ is a noncentral F-distribution with d and $n - d + 1$ degrees of freedom. The noncentrality parameter $\delta = n_1 n_2/(n_1 + n_2)\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$, where $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2 = \boldsymbol{\nu}^{\mathrm{T}} \mathbf{\Sigma}^{-1} \boldsymbol{\nu} = (\Phi \boldsymbol{\mu})^{\mathrm{T}} (\Phi \mathcal{K} \Phi^{\mathrm{T}})^{-1} \Phi \boldsymbol{\mu}$. It is obvious that the value of δ depends on the choice of the projective space. Based on the form of $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$, we can view it as the extracted key information in terms of the mean functions difference through projection. Under the null hypothesis (2.2), $(n - d + 1)(nd)^{-1} T^2 \sim F_{d,n-d+1}(0)$. The power function of the size-α

F-test can be calculated as:

$$\beta_{H^\Phi}\left(\|\Delta_{\boldsymbol{\nu}}^\Phi\|; d, n_1, n_2\right) = pr\left(\frac{n-d+1}{nd} T^2(\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2) > F_{d,n-d+1}^{\alpha} \mid H_1^\Phi\right), \quad (2.4)$$

where $F_{d,n-d+1}^{\alpha}$ is the upper 100α percentile of $F_{d,n-d+1}(0)$. Under the alternative hypothesis (2.2), the exact distribution of T^2 depends on $\|\Delta_{\boldsymbol{\nu}}^\Phi\|$. Thus, T^2 in the power function (2.4) is written with a slight abuse of notation. Equation (2.4) indicates that when α, n_1, n_2, and d are fixed, the power function $\beta_{H^\Phi}(\cdot)$ strictly increases with $\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2$. Moreover, Gupta and Perlman (1974) pointed out that when α, n_1, n_2, and $\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2$ are fixed, $\beta_{H^\Phi}(\cdot)$ strictly decreases with d. This highlights the importance of carefully choosing the projective space and the corresponding dimension. A theoretical example is presented in Section **??** to make this conclusion more intuitive and clearer.

In the preceding paragraph, we explored the changes in power with d and $\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2$ when n_1 and n_2 are fixed. As the sample sizes go to infinity, under the assumptions $d/n = c_n \to c \in (0,1)$, $n_1/(n_1+n_2) \to \kappa \in (0,1)$, and $\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2 = o(1)$, Bai and Saranadasa (1996) showed that the power function

of Hotelling's T^2-test satisfies

$$\beta_{H^\Phi}(\|\Delta_{\boldsymbol{\nu}}^\Phi\|; d, n_1, n_2) - G\left(-\eta_\alpha + \left(\frac{n(1-c)}{2c}\right)^{\frac{1}{2}} \kappa(1-\kappa)\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2\right) \xrightarrow{\mathbb{P}} 0, \quad (2.5)$$

where $G(\cdot)$ and η_α denote the cumulative distribution function and the upper $100(1-\alpha)$ percentile of the standard normal distribution $N(0,1)$, respectively. Based on (2.5), it is clear that the limiting value of $\beta_{H^\Phi}(\cdot)$ increases as c decreases or $\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2$ increases. However, c is strictly increasing, while $\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2$ is non-decreasing, with the dimension d increasing. Therefore, it is important to carefully consider the trade-off between c and $\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2$ when choosing the parameter d.

2.3 Selection of projection functions

Based on the discussion in Section 2.2, to obtain a more efficient test of H_0, we should choose a projective space, of which the corresponding extracted information $\|\Delta_{\boldsymbol{\nu}}^\Phi\|^2$ is as large as possible and the dimension of the projective space d is relatively small. In the following, we will delve into a more detailed discussion on the selection of projection functions.

2.3.1 Measure of information loss

The projection inevitably leads to information loss. Thus, we seek to investigate the upper bound of the extractable information. We notice that $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ can be written as $\boldsymbol{\mu}^{\mathrm{T}}\mathcal{K}^{-1/2}P_{\mathcal{K}^{1/2}\Phi^{\mathrm{T}}}\mathcal{K}^{-1/2}\boldsymbol{\mu}$, where $P_{\mathcal{K}^{1/2}\Phi^{\mathrm{T}}} = \mathcal{K}^{1/2}\Phi^{\mathrm{T}}(\Phi\mathcal{K}\Phi^{\mathrm{T}})^{-1}\Phi\mathcal{K}^{1/2}$. It is straightforward to see that,

$$\boldsymbol{\mu}^{\mathrm{T}}\mathcal{K}^{-\frac{1}{2}}(I - P_{\mathcal{K}^{\frac{1}{2}}\Phi^{\mathrm{T}}})\mathcal{K}^{-\frac{1}{2}}\boldsymbol{\mu} \triangleq \|\boldsymbol{\varepsilon}\|^2 \geq 0. \tag{2.6}$$

Let $\|\Delta_{\boldsymbol{\mu}}\|^2 = \boldsymbol{\mu}^{\mathrm{T}}\mathcal{K}^{-1}\boldsymbol{\mu}$, then (2.6) indicates that the upper bound of $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ is $\|\Delta_{\boldsymbol{\mu}}\|^2$. Thus the difference $\|\boldsymbol{\varepsilon}\|^2 = \|\Delta_{\boldsymbol{\mu}}\|^2 - \|\Delta_{\boldsymbol{\nu}}\|^2$ can be taken as a measure of information loss due to projection. It also should be noted that $\|\boldsymbol{\varepsilon}\|^2$ is equal to the sum of squared errors in fitting $\mathcal{K}^{-\frac{1}{2}}\boldsymbol{\mu}$ with $\mathcal{K}^{\frac{1}{2}}\Phi^{\mathrm{T}}$, i.e.,

$$\mathcal{K}^{-\frac{1}{2}}\boldsymbol{\mu} = \mathcal{K}^{\frac{1}{2}}\Phi^{\mathrm{T}}\boldsymbol{\theta}, \tag{2.7}$$

where the coefficient $\boldsymbol{\theta}$ takes the least squares estimator $\boldsymbol{\theta}_0 = (\Phi\mathcal{K}\Phi^{\mathrm{T}})^{-1}\Phi\boldsymbol{\mu}$.

The following proposition states that, as $d \to \infty$, the information loss caused by projection functions Φ tends to zero under certain conditions.

Proposition 1. *When taking projection functions* $\Phi = (\phi_1, \ldots, \phi_d)^{\mathrm{T}}$ *with* ϕ_k, $k = 1, 2, \ldots$ *spanning the range of* $\mathcal{K}$, *the value of* $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ *converges to*

$\|\Delta_{\boldsymbol{\mu}}\|^2$ *as* $d \to \infty$*, resulting in no loss of information asymptotically.*

2.3.2 The most powerful projection function

Although Proposition 1 is established, it is important to recall from Section 2.2 that an excessively large d can negatively impact testing efficiency. In other words, increasing d to achieve less information loss does not necessarily lead to a more powerful test. In this section, we aim to identify projection spaces that have high informativeness but also low dimensions.

For the case when $d=1$, i.e., $\Phi=(\phi_1)^{\mathrm{T}}$, both $\Phi\boldsymbol{\mu}$ and $\Phi\mathcal{K}\Phi^{\mathrm{T}}$ are scalars. If $\|\mathcal{K}^{-1/2}\boldsymbol{\mu}\|<\infty$, by the Cauchy-Schwarz inequality, we have

$$\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2=\frac{(\Phi\boldsymbol{\mu})^2}{\Phi\mathcal{K}\Phi^{\mathrm{T}}}\leq\frac{\|\mathcal{K}^{1/2}\Phi^{\mathrm{T}}\|^2\|\mathcal{K}^{-1/2}\boldsymbol{\mu}\|^2}{\Phi\mathcal{K}\Phi^{\mathrm{T}}}=\|\mathcal{K}^{-1/2}\boldsymbol{\mu}\|^2=\|\Delta_{\boldsymbol{\mu}}\|^2.$$

If $\|\mathcal{K}^{-1}\boldsymbol{\mu}\|<\infty$, the equality can be achieved for $\Phi^{\mathrm{T}}=\mathcal{K}^{-1}\boldsymbol{\mu}\triangleq\phi^*$. It means the information loss with ϕ^* equals to 0. Considering that ϕ^* has $d=1$, it is the **most powerful projection function** among all Φ with $d\geq 1$.

When $d>1$, for given projection functions $\Phi=(\phi_1,\ldots,\phi_d)^{\mathrm{T}}$, it is natural to consider their linear combination, i.e., $\Phi^{\mathrm{T}}\boldsymbol{\vartheta}=\sum_{k=1}^{d}\vartheta_k\phi_k\triangleq\phi^{\dagger}$, as the projection function. Now let's compare the testing efficiency of Φ and $\phi^{\dagger}$.

By the Cauchy-Schwarz inequality, if $\|(\Phi\mathcal{K}\Phi^{\mathrm{T}})^{-1/2}\Phi\boldsymbol{\mu}\| < \infty$, we have

$$\|\Delta_{\boldsymbol{\nu}}^{\phi^\dagger}\|^2 = \frac{(\boldsymbol{\vartheta}^{\mathrm{T}}\Phi\boldsymbol{\mu})^2}{\boldsymbol{\vartheta}^{\mathrm{T}}\Phi\mathcal{K}\Phi^{\mathrm{T}}\boldsymbol{\vartheta}} \leq \|(\Phi\mathcal{K}\Phi^{\mathrm{T}})^{-1/2}\Phi\boldsymbol{\mu}\|^2 = \|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2.$$

If $\|(\Phi\mathcal{K}\Phi^{\mathrm{T}})^{-1}\Phi\boldsymbol{\mu}\| < \infty$, $\|\Delta_{\boldsymbol{\nu}}^{\phi^\dagger}\|^2$ can reach its upper bound $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ with $\boldsymbol{\vartheta} = (\Phi\mathcal{K}\Phi^{\mathrm{T}})^{-1}\Phi\boldsymbol{\mu} \triangleq \boldsymbol{\theta}_0$. It means the extracted information of Φ and $\Phi^{\mathrm{T}}\boldsymbol{\theta}_0 \triangleq \phi^{\dagger *}$ are the same. Thus, $\phi^{\dagger *}$, having a lower dimension, is more efficient than Φ, and is **the most powerful projection function** among all $\phi^\dagger$. What's more, $\|\Delta_{\boldsymbol{\nu}}^{\phi^{\dagger *}}\|^2 = \|\Delta_{\boldsymbol{\nu}}^{\phi^*}\|^2 = \|\Delta_{\boldsymbol{\mu}}\|^2$ if there exists a constant c such that $\phi^{\dagger *} = c\phi^*$, which means $\phi^{\dagger *}$ is equivalent to ϕ^* if ϕ^* is in the span of Φ.

Remark 1. *In practice, both $\mathcal{K}$ and $\boldsymbol{\mu}$ are unknown, and $\boldsymbol{\theta}_0$ can be estimated by $\widehat{\boldsymbol{\theta}}_0 = (\Phi\widehat{\mathcal{K}}\Phi^{\mathrm{T}})^{-1}\Phi\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\nu}}$. It can be verified that the statistic T^2 in (2.3) with the projection function $\Phi^{\mathrm{T}}\widehat{\boldsymbol{\theta}}_0 \triangleq \widehat{\phi^{\dagger *}}$ is equal to the one with Φ. Therefore, the tests based on Φ or $\widehat{\phi^{\dagger *}}$ are equivalent in practice.*

2.3.3 Selection Criteria

Combining the analyses from the previous sections, we propose three perspectives for identifying efficient projection spaces with finite dimensions, which we then apply practically in Section 3. Specifically,

1. Minimize information loss. In Section 3.1, we employ an optimization

algorithm within the Krylov subspace to approximate the optimal projection function that achieves zero information loss.

2. Maximize the extracted information. In Section 3.2, we assess the useful information that each eigenfunction can extract and propose to select eigenfunctions based on this assessment.

3. Accurately fit Equation (2.7). In Section 3.3, we utilize B-Spline basis functions as projection functions. These are data-independent and are particularly suitable when the sample size is small.

As previously discussed, a large d can adversely affect testing efficiency. Moreover, we know that when fitting functional observations with basis functions, the optimal rate of d increases very slowly with the number of observed discrete time points (Eubank, 1999). Therefore, if the projection space is appropriately chosen, projection functions Φ with a finite d can satisfactorily fit Equation (2.7). Consequently, we recommend conducting projection-based testing with a finite number of functions. In practice, the specific value of d can be determined by consulting the fitting performance of Equation (2.7). And for the projection spaces in Section 3, we will provide specific methods for determining the value of d.

Existing papers have only investigated testing based on a specific pro-

jection space and have not considered the choice of projection functions. Their performance can be evaluated by analyzing the information they extract. For instance, the random projection function (Cuesta-Albertos and Febrero-Bande, 2010) may only capture useful information by chance and is inherently unstable. The commonly used eigenfunctions with the largest eigenvalues (Horváth et al., 2013; Pomann et al., 2016) do not necessarily align with the direction of mean difference. Moreover, employing all eigenfunctions with non-zero eigenvalues (Pini et al., 2018) can introduce significant noise into the statistic, thereby disrupting the testing. Additionally, the classical L_2-norm-based test (Benko et al., 2009) can be interpreted as a test based on the projection function $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}_1 - \widehat{\boldsymbol{\mu}}_2$, which does not take the covariance function $\mathcal{K}$ into consideration, thus making it less sensitive in testing. These analyses are consistent with their performance in the simulation studies.

3. Practical Projection Functions and Power Analysis

3.1 Krylov Subspace

According to the discussion in Section 2.3, the theoretical most powerful projection function is $\phi^* = \mathcal{K}^{-1}\boldsymbol{\mu}$, which results in information loss $\|\boldsymbol{\varepsilon}\|^2 = 0$. However, the direct estimation of ϕ^* is difficult. Through some calculation,

we find that minimizing the loss of information $\|\boldsymbol{\varepsilon}\|^2$ with respect to Φ is equivalent to minimizing a quadratic function $\langle\phi,\mathcal{K}\phi\rangle/2-\langle\boldsymbol{\mu},\phi\rangle$ with respect to ϕ. The projection function ϕ^* is the minimizer of both optimization problems. If the minimizer exists, we propose to apply the conjugate gradient (CG) algorithm in Kraus and Stefanucci (2019) to approximate ϕ^*. The CG algorithm solves the minimization problem with ϕ restricted to the d-dimensional Krylov subspace $K_d(\mathcal{K},\boldsymbol{\mu})$, spanned by $\boldsymbol{\mu},\mathcal{K}\boldsymbol{\mu},\ldots,\mathcal{K}^{d-1}\boldsymbol{\mu}$, in the first d steps. Namely, the CG algorithm minimize $\langle\phi,\mathcal{K}\phi\rangle/2-\langle\boldsymbol{\mu},\phi\rangle$ subject to $\phi\in K_d(\mathcal{K},\boldsymbol{\mu})$ and the solution obtained by the CG algorithm ϕ_d^{CG} is a linear combination of $\boldsymbol{\mu},\mathcal{K}\boldsymbol{\mu},\ldots,\mathcal{K}^{d-1}\boldsymbol{\mu}$.

We denote by $\widehat{\phi}_d^{\mathrm{CG}}$ the solution with $\boldsymbol{\mu},\mathcal{K}$ substituted with $\widehat{\boldsymbol{\mu}},\widehat{\mathcal{K}}$. Since the projected difference $\boldsymbol{\nu}$ is a scalar, we can calculate the test statistic as follows,

$$T_{\mathrm{CG}}=\left(\frac{n_1n_2}{n_1+n_2}\right)^{\frac{1}{2}}\frac{\widehat{\boldsymbol{\mu}}^{\mathrm{T}}\widehat{\phi}_d^{\mathrm{CG}}}{\langle\widehat{\phi}_d^{\mathrm{CG}},\widehat{\mathcal{K}}\widehat{\phi}_d^{\mathrm{CG}}\rangle^{1/2}}.$$

Under certain assumptions, we establish the convergence of the statistics T_{CG} under both the original null hypothesis and alternative hypothesis.

Assumption 1. *The proportion of the sample sizes converges to a constant,* $n_1/(n_1+n_2)\to\kappa\in(0,1)$ *as* $n_0=\min(n_1,n_2)\to\infty$.

Assumption 2.

$$n_0^{-1/2}\omega_d^{-1}\|\gamma_d\| + n_0^{-1}\omega_d^{-3} \to 0 \text{ as } n_0 \to \infty,$$

where ω_d *is the smallest eigenvalue of the* $d \times d$ *matrix* $H = \Phi_{CG}\mathcal{K}\Phi_{CG}^{\mathrm{T}}$ *with* $\Phi_{CG} = (\boldsymbol{\mu}, \mathcal{K}\boldsymbol{\mu}, \ldots, \mathcal{K}^{d-1}\boldsymbol{\mu})^{\mathrm{T}}$ *and the* d*-vector* $\gamma_d = H^{-1}\Phi_{CG}\boldsymbol{\mu}$.

Theorem 2. *If Assumptions 1 and 2 hold,*

1. *Under* H_0,

$$T_{CG} \xrightarrow{\mathbb{D}} N(0,1).$$

2. *Under* H_1,

$$\frac{n_1+n_2}{n_1 n_2} T_{CG}^2 = \|\widehat{\Delta}_{\boldsymbol{\nu}}^{CG,d}\|^2 \xrightarrow{\mathbb{P}} \frac{(\boldsymbol{\mu}^{\mathrm{T}}\phi_d^{CG})^2}{\langle \phi_d^{CG}, \mathcal{K}\phi_d^{CG}\rangle} = \|\Delta_{\boldsymbol{\nu}}^{CG,d}\|^2,$$

which means T_{CG}^2 *grows linearly with* n_0*, and the test is consistent, i.e.* $T_{CG}^2 \xrightarrow{\mathbb{P}} \infty$. *And*

$$T_{CG} - \left(\frac{n_1 n_2}{n_1+n_2}\right)^{\frac{1}{2}} \|\Delta_{\boldsymbol{\nu}}^{CG,d}\| \xrightarrow{\mathbb{D}} N(0,1).$$

Remark 2. *In practice, we use the sample-splitting method (Huang, 2015), which randomly divides the data into two parts, one part for estimating the*

conjugate gradient solution, and the other part for calculating the projection test statistic. The purpose of random splitting is to preserve the Type I error rate. To mitigate the impact of randomness and less power issues of sample-splitting, we repeat this procedure J times and take the average of statistics as our final test statistics. Taking a larger J helps to obtain more stable results, and we set $J = 100$ in our simulation. In addition, the value of d can be determined using the stopping rule of the CG algorithm, such as when the relative difference between successive iterations of the solution falls below a specified threshold.

3.2 Principle Components

Functional principle component analysis (FPCA) is commonly used in FDA. Galeano et al. (2015) proposed a Mahalanobis distance for functional observations, which shares the same form as $\|\Delta_{\boldsymbol{\mu}}\|$. They also used a finite number of eigenvalues and eigenfunctions to approximate the functional Mahalanobis distance. Building upon this idea, we consider the projective space spanned by the eigenfunctions of $\mathcal{K}$, which is denoted as $E_d(\mathcal{K})$ and the corresponding projection funciton is $\Phi_{\mathrm{PC}}(B) = (\psi_{r_1}, \ldots, \psi_{r_d})^{\mathrm{T}}$. Here, $B = \{r_1, \ldots, r_d\}$ represents a subset of the indices of eigenfunctions. The spectral decomposition of $\mathcal{K}$ is $\mathcal{K} = \sum_{r=1}^{\infty} \lambda_r \psi_r \psi_r^{\mathrm{T}}$, where $\lambda_1 > \lambda_2 > \cdots \geq 0$

are the eigenvalues and $\psi_{r_1}, \ldots, \psi_{r_d}$ are the eigenfunctions. Once $\Phi_{\mathrm{PC}}(B)$ is determined, the extracted information is,

$$\|\Delta_{\boldsymbol{\nu}}^{\mathrm{PC},d}\|^2 = \sum_{k=1}^{d} \frac{(\boldsymbol{\mu}^{\mathrm{T}}\psi_{r_k})^2}{\lambda_{r_k}} \triangleq \sum_{k=1}^{d} Q_{r_k},\ r_k \in B. \tag{3.8}$$

In FDA, one common approach is to select the eigenfunctions corresponding to the leading d eigenvalues (Horváth et al., 2013), i.e., $B_\lambda = \{1, \ldots, d\}$. This choice provides a good approximation of the variation of functional observations. However, $\|\Delta_{\boldsymbol{\nu}}^{\mathrm{PC},d}\|^2$ based on $\Phi_{\mathrm{PC}}(B_\lambda)$ is not necessarily the largest among all possible choices. In order to obtain a more efficient test after projection, we instead choose the subset of indices B_Q, where the corresponding Q_{r_k} are the d largest among all $\{Q_r = (\boldsymbol{\mu}^{\mathrm{T}}\psi_r)^2/\lambda_r \mid \lambda_r > 0\}$. The projective space spanned by $\Phi_{PC}(B_Q)$ results in the largest extracted information among all possible combinations of the eigenfunctions of $\mathcal{K}$ for a given d.

In practice, let $\widehat{\lambda}_r$ and $\widehat{\psi}_r$ denote the estimated eigenvalues and eigenfunctions of $\widehat{\mathcal{K}}$, respectively. The test statistic is

$$T^2_{\mathrm{PC}}(B) = \frac{n_1 n_2}{n_1 + n_2} \sum_{k=1}^{d} \frac{(\widehat{\boldsymbol{\mu}}^{\mathrm{T}}\widehat{\psi}_{r_k})^2}{\widehat{\lambda}_{r_k}},\ r_k \in B.$$

Assumption 3. *No ties exist in* $\lambda_{r_1}, \ldots, \lambda_{r_d}$ *and* $\min_{1 \le k \le d} \lambda_{r_k} > 0$.

Assumption 4. *As* $n_0 \to \infty$, $\lambda_{d^*}^4 n_0 d^{-2} \to \infty$ *and* $\lambda_{d^*}^2 n_0 (d a_{d^*})^{-2} \to \infty$, *where* $a_1 = 2^{3/2}(\lambda_1 - \lambda_2)^{-1}$, $a_k = 2^{3/2}\max\{(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}\}$ *for* $k = 2, 3, \ldots,$ *and* $d^* = \max_{1 \le k \le d} r_k$.

Theorem 3. *If Assumptions 1 and 3 hold,*

1. *Under* H_0,

$$T_{PC}^2(B) \xrightarrow{\mathbb{D}} \chi_d^2,$$

where χ_d^2 *is a chi-square random variable with* d *degrees of freedom.*

2. *Under* H_1, *if Assumption 4 also holds , then*

$$\frac{n_1 + n_2}{n_1 n_2} T_{PC}^2(B) = \|\widehat{\Delta}_{\boldsymbol{\nu}}^{PC,d}\|^2 \xrightarrow{\mathbb{P}} \sum_{k=1}^{d} \frac{(\boldsymbol{\mu}^{\mathrm{T}} \psi_{r_k})^2}{\lambda_{r_k}} = \|\Delta_{\boldsymbol{\nu}}^{PC,d}\|^2,$$

which means T_{PC}^2 *grows linearly with* n_0. *In particular, if* $\boldsymbol{\mu}$ *is not orthogonal to the linear span of* $\psi_{r_1}, \ldots, \psi_{r_d}$, *then* $T_{PC}^2(B) \xrightarrow{\mathbb{P}} \infty$. *And*

$$T_{PC}^2 - \frac{n_1 n_2}{n_1 + n_2} \|\Delta_{\boldsymbol{\nu}}^{PC,d}\|^2 \xrightarrow{\mathbb{D}} \chi_d^2.$$

Note that under H_1 in Theorem 3, the consistency of the test based on the eigenfunctions of $\widehat{\mathcal{K}}$ relies on the assumption that $\boldsymbol{\mu}$ is not orthogonal to the linear span of $\psi_{r_1}, \ldots, \psi_{r_d}$. However, this assumption is not guaranteed

to hold in practice, even when d is large enough. This prompts us to consider the test based on the eigenfunctions of $\widehat{\mathcal{M}}$, where

$$\widehat{\mathcal{M}} = \frac{1}{n_1+n_2-1}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\mathbf{y}_{ij}-\overline{\mathbf{y}})(\mathbf{y}_{ij}-\overline{\mathbf{y}})^{\mathrm{T}},$$

and $\overline{\mathbf{y}} = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\mathbf{y}_{ij}/(n_1+n_2)$. We know that $\widehat{\mathcal{K}}$ is a measure of the within-population covariation and will converge to $\mathcal{K}$ regardless of the truth or falsity of H_0, while $\widehat{\mathcal{M}}$ is a measure of the overall covariation and will converge to $\mathcal{K}$ only under H_0. Under H_1, $\widehat{\mathcal{M}}$ converges to $\mathcal{M} = \mathcal{K} + \kappa(1-\kappa)\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}$, which is proved to be a covariance function (Horváth and Kokoszka, 2012) and thus has orthonormal eigenfunctions ψ_k^* and eigenvalues λ_k^* satisfying the spectral decomposition $\mathcal{M} = \sum_{k=1}^{\infty}\lambda_k^*\psi_k^*\psi_k^{*\mathrm{T}}$. It is obvious that $\boldsymbol{\mu}(\neq 0)$ is not orthogonal to the range of $\mathcal{M}$ as $\boldsymbol{\mu}^{\mathrm{T}}\mathcal{M}\boldsymbol{\mu} \geq \kappa(1-\kappa)(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu})^2 > 0$.

When selecting projection functions $\Phi_{\mathrm{PC}^*}(B) = (\psi_{r_1}^*, \ldots, \psi_{r_d}^*)^{\mathrm{T}}$, it should be noted that $\Phi_{\mathrm{PC}^*}\mathcal{K}\Phi_{\mathrm{PC}^*}^{\mathrm{T}}$ is not a diagonal matrix. Therefore, $\|\Delta_{\boldsymbol{\nu}}^{\mathrm{PC}^*,d}\|^2 = (\Phi_{\mathrm{PC}^*}\boldsymbol{\mu})^{\mathrm{T}}(\Phi_{\mathrm{PC}^*}\mathcal{K}\Phi_{\mathrm{PC}^*}^{\mathrm{T}})^{-1}\Phi_{\mathrm{PC}^*}\boldsymbol{\mu}$ cannot be simplified as a sum of d-terms, as in (3.8). Specifically, the $r_k, r_{k'}$th element of $\Phi_{\mathrm{PC}^*}\mathcal{K}\Phi_{\mathrm{PC}^*}^{\mathrm{T}}$ is $\lambda_{r_k}^{*\mathrm{T}}\mathcal{K}\lambda_{r_{k'}}^*$, which is equal to $-\kappa(1-\kappa)\boldsymbol{\mu}^{\mathrm{T}}\psi_{r_k}^*\boldsymbol{\mu}^{\mathrm{T}}\psi_{r_{k'}}^*$ when $r_k \neq r_{k'}$, and to $\lambda_{r_k}^* - \kappa(1-\kappa)(\boldsymbol{\mu}^{\mathrm{T}}\psi_{r_k}^*)^2$ when $r_k = r_{k'}$, for $r_k, r_k' = 1, \ldots, d$. However, selecting the eigenfunction $\psi_{r_k}^*$ with larger $(\boldsymbol{\mu}^{\mathrm{T}}\psi_{r_k}^*)^2/\lambda_{r_k}^* \triangleq Q_{r_k}^*$ is still reasonable. By ig-

noring the off-diagonal elements, the information extracted by $\psi^*_{r_k}$, which is $(\boldsymbol{\mu}^{\mathrm{T}}\psi^*_{r_k})^2/\{\lambda^*_{r_k}-\kappa(1-\kappa)(\boldsymbol{\mu}^{\mathrm{T}}\psi^*_{r_k})^2\} = 1/\{1/Q^*_{r_k}-\kappa(1-\kappa)\}$, is positively correlated with $Q^*_{r_k}$. Thus, we select the subset B_{Q^*} with r_k satisfying $Q^*_{r_1} \geq \cdots \geq Q^*_{r_d} > 0$ for a given d. After projecting the observations onto $\Phi_{\mathrm{PC}^*}(B_{Q^*})$, we use the Hotelling's T^2 statistic, which will take the covariance into consideration.

Remark 3. *In practice, we determine d by identifying the jump point in $\widehat{Q}_{r_k}$, i.e. $d = \arg\max_k \widehat{Q}_{r_k}/\widehat{Q}_{r_{k+1}}$. To prevent selecting eigenfunctions with eigenvalues close to 0, we advise choosing the subset B within $\{k : \widehat{\lambda}_k > c_1\}$ or taking $\widehat{Q}_{r_k} = (\widehat{\boldsymbol{\mu}}^{\mathrm{T}}\widehat{\psi}_{r_k})^2/(\widehat{\lambda}_{r_k}+c_2)$, where c_1, c_2 are constants close to 0.*

Remark 4. *The main difference between our test in this section and the one proposed by Wang et al. (2022) is the selection criterion for the eigenfunctions. While Wang et al. (2022) defines Q_k directly and uses a hard threshold to sum up Q_k with two manually specified parameters, we select the eigenfunctions that can extract the most useful information, which is determined by Q_k as derived from our analysis. Moreover, to cope with the situation when μ is orthogonal to the linear span of the eigenfunctions, Wang et al. (2022) adds a power enhancement component to their test statistic. In contrast, we propose the test based on the eigenfunctions of $\mathcal{M}$, which are inherently not orthogonal to μ.*

3.3 Basis Functions

Basis functions are a common tool in FDA. For instance, Shen and Faraway (2004) and Górecki and Smaga (2015) proposed functional two-sample tests based on B-Spline basis and orthonormal basis, respectively. These tests use an F-test statistic, calculated from the coefficients of the functional observation represented by the selected basis functions. In contrast, our test considers a projective space $L_d(\cdot)$ spanned by basis functions, which can accurately fit the two-sample difference $\mathcal{K}^{-1/2}\boldsymbol{\mu}$ but not the functional observations. The test statistic is the Mahalanobis distance between the projected two-sample data. Notably, the projective space $L_d(\cdot)$ is data independent, thus it inherits the theoretical analysis in Section 2.

For B-Spline basis, the projection function is $\Phi_{\mathrm{BS}} = (\varphi_1, \ldots, \varphi_d)^{\mathrm{T}}$, where φ_k is the cubic spline basis with $d-2$ inner knots. The test statistic is

$$T^2_{\mathrm{BS}} = \frac{n_1 n_2}{n_1 + n_2}(\Phi_{\mathrm{BS}}\widehat{\boldsymbol{\mu}})^{\mathrm{T}}(\Phi_{\mathrm{BS}}\widehat{\mathcal{K}}\Phi^{\mathrm{T}}_{\mathrm{BS}})^{-1}\Phi_{\mathrm{BS}}\widehat{\boldsymbol{\mu}}.$$

Theorem 4. *If Assumption 1 holds*

1. *Under H_0,*

$$T^2_{BS} \xrightarrow{\mathbb{D}} \chi^2_d.$$

2. *Under* H_1,

$$\frac{n_1+n_2}{n_1 n_2} T_{BS}^2 = \|\widehat{\Delta}_{\boldsymbol{\nu}}^{BS,d}\|^2 \xrightarrow{\mathbb{P}} (\Phi_{BS}\boldsymbol{\mu})^{\mathrm{T}} (\Phi_{BS}\mathcal{K}\Phi_{BS}^{\mathrm{T}})^{-1}\Phi_{BS}\boldsymbol{\mu} = \|\Delta_{\boldsymbol{\nu}}^{BS,d}\|^2,$$

which means T_{BS}^2 *grows linearly with* n_0. *In particular, if* $\boldsymbol{\mu}$ *is not orthogonal to the linear span of* $\varphi_1, \ldots, \varphi_d$, *then* $T_{BS}^2 \xrightarrow{\mathbb{P}} \infty$. *And*

$$T_{BS}^2 - \frac{n_1 n_2}{n_1+n_2}\|\Delta_{\boldsymbol{\nu}}^{BS,d}\|^2 \xrightarrow{\mathbb{D}} \chi_d^2.$$

Remark 5. *The selection of the type and number of basis functions is critical and should be customized for each specific case. For periodic functional data, Fourier basis functions are preferred because they match the periodic features. For non-periodic data, B-spline basis functions are more suitable, given their flexibility in modeling diverse data structures. Furthermore, we can determine the value of* d *by fitting Equation* (2.7) *with a penalty.*

3.4 Comparison

To summarize the above, the selection of ϕ_d^{CG} in the Krylov subspace aims to minimize the information loss. The projective space $\Phi_{\mathrm{PC}}(B_Q)$ and $\Phi_{\mathrm{PC}^*}(B_{Q^*})$, chosen in the eigenspace of $\mathcal{K}$ and $\mathcal{M}$, respectively, intend to maximize the extracted information. The projective space Φ_{BS}, selected

based on the basis function family, is to enhance performance in fitting (2.7). Although these information criteria are equivalent, the testing efficiency of these projection functions may not be the same for a given d, as stated in the following proposition.

Proposition 2. *For a given d, the testing power function of H_0 based on $\Phi_{PC}(B_Q)$ is higher than that based on ϕ_d^{CG} when there are infinite eigenvalues of $\mathcal{K}$ close to 0, and the latter is higher than or equal to that based on $\Phi_{PC}(B_\lambda)$.*

Furthermore, besides the performance in testing efficiency, projection functions selected from different subspace require different assumptions to achieve the same properties. For example, the consistency of the test based on $\Phi_{\mathrm{PC}}(B_Q)$ and Φ_{BS} relies on the assumption that $\boldsymbol{\mu}$ is not orthogonal to the linear span of $\psi_{r_1}, \ldots, \psi_{r_d}$ and $\varphi_1, \ldots, \varphi_d$, respectively. In contrast, this assumption is naturally satisfied for ϕ_d^{CG} and $\Phi_{\mathrm{PC}^*}(B_{Q^*})$. Additionally, the advantage of Φ_{BS} is that there is no need to estimate projection functions, which avoids estimation errors and results in better performance under finite sample size.

4. Simulation

In this section, we compare the performance of our projection tests with several existing methods. The function of individual j in sample i is generated by the following function:

$$y_{ij}(t) = \mu_i(t) + \sqrt{2}\xi_{ij1}sin(2\pi t) + \sqrt{2}\xi_{ij2}cos(2\pi t) + \varepsilon_{ijt}, \tag{4.9}$$

where $\mu_1(t) = t, \mu_2(t) = t + at^3$, $\xi_{ij1} \overset{\text{iid}}{\sim} N(0,10)$, $\xi_{ij2} \overset{\text{iid}}{\sim} N(0,5)$, and observation noise $\varepsilon_{ijt} \overset{\text{iid}}{\sim} N(0,0.25)$. The functions are observed on a grid of $p = 100$ equispaced points in $[0,1]$.

As discussed in Section 2, the dimension of the projective space can impact the testing power. In the Supplementary Material **??**, we provide a comprehensive analysis of the finite sample behavior of the projection tests under various values of d. In this section, we set $d = 5$ for the conjugate gradient solution and $d = 6$ for the B-Spline basis with order $\min(d, 4)$, and the tests based on these two types of projective spaces are recorded as CG and BS, respectively. For principal components, we denote PC_K_Q and PC_M_Q as the tests based on the selected eigenfunctions of $\mathcal{K}$ and $\mathcal{M}$, respectively, where the selection criterion is described in Section 3.2. We also consider seven other existing testing methods. The first four meth-

ods are based on principal components, HT (Wang et al., 2022) based on permutation, GHT (Pini et al., 2018), PC_K_λ (Horváth et al., 2013) and PC_M_λ (Pomann et al., 2016). Furthermore, we include the method based on random projection function (RP, Cuesta-Albertos and Febrero-Bande, 2010), as well as the methods based on distance, L_2 (Benko et al., 2009) and MMD (Wynne and Duncan, 2022). Further details about these methods are provided in the preceding sections. We implement FPCA and MMD using the R packages `refund` and `fdahotelling`, respectively. In Figures 1 - 3, we present the results with $n_1 = n_2 = 20, 50, 100$, respectively, which correspond to total sample sizes smaller than, equal to, and greater than p.

In Figure 1 - 3, it is obvious that the empirical size of each methods is around 0.05, and the empirical power of each methods tends to 1 as a and n_1, n_2 increases. Our four methods, HT, and GHT exhibit increasing power as a increases from 0 to 0.5, while the other five methods are less sensitive when $a = 0.5$, which is in accordance with our analysis in Section 2.3.3. Furthermore, the slightly lower performance of HT might be due to the less-than-ideal choice of parameters for the hard threshold. The empirical power of our four methods and HT increase significantly with sample size, while GHT performs similarly for $n_1 = n_2 = 20, 50$ and only shows improvement when $n_1 = n_2 = 100$, which can be seen from the Figure 4.

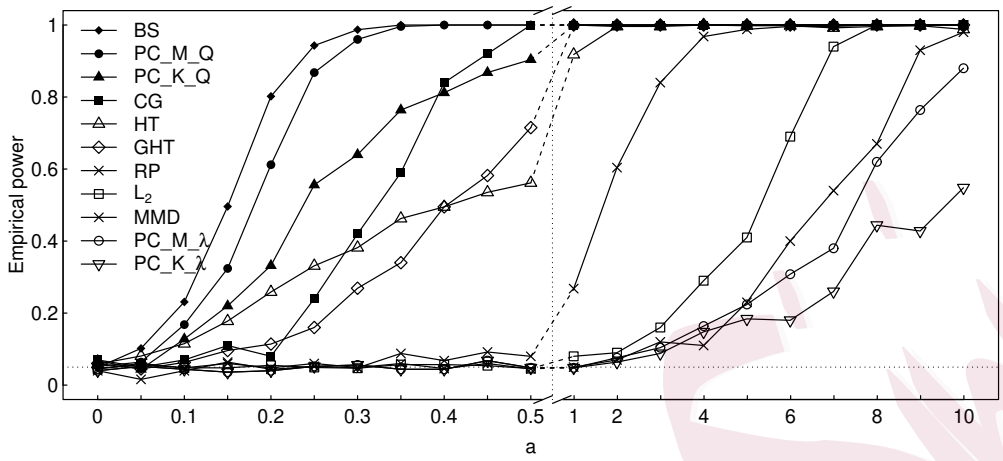

Figure 1: The empirical power function of different methods with sample size $n_1 = n_2 = 20$. The horizontal dotted line represents the confidence level of 0.05. Different scales are used on both sides of the vertical dotted line, with the left side ranging from 0 to 0.5 and the right side ranging from 1 to 10.

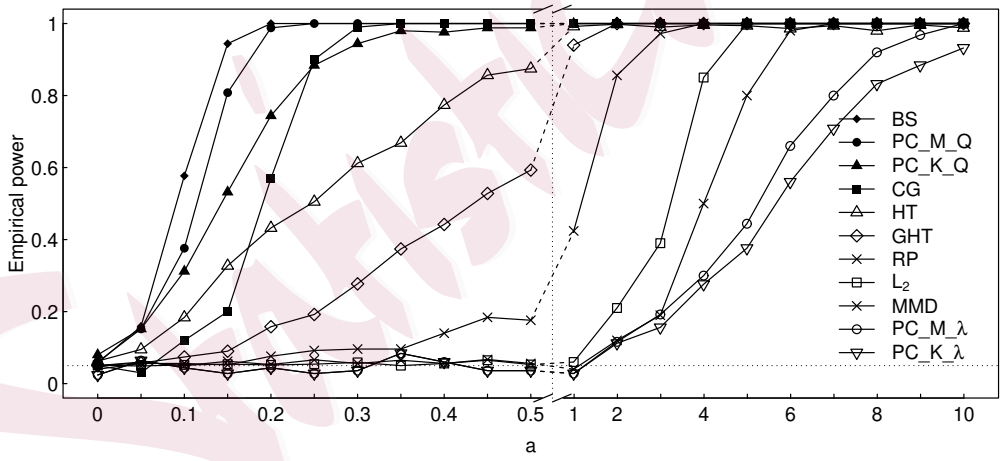

Figure 2: The empirical power function of different methods with sample size $n_1 = n_2 = 50$. The horizontal dotted line represents the confidence level of 0.05. Different scales are used on both sides of the vertical dotted line, with the left side ranging from 0 to 0.5 and the right side ranging from 1 to 10.

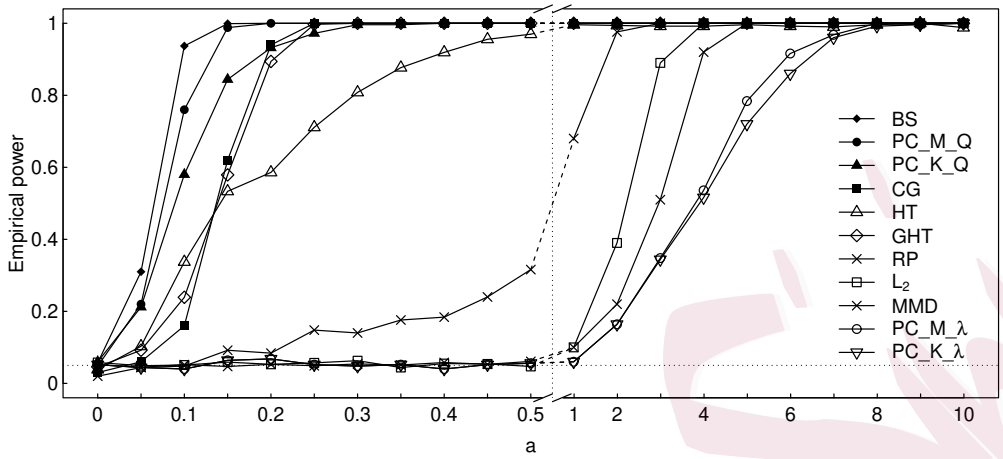

Figure 3: The empirical power function of different methods with sample size $n_1 = n_2 = 100$. The horizontal dotted line represents the confidence level of 0.05. Different scales are used on both sides of the vertical dotted line, with the left side ranging from 0 to 0.5 and the right side ranging from 1 to 10.

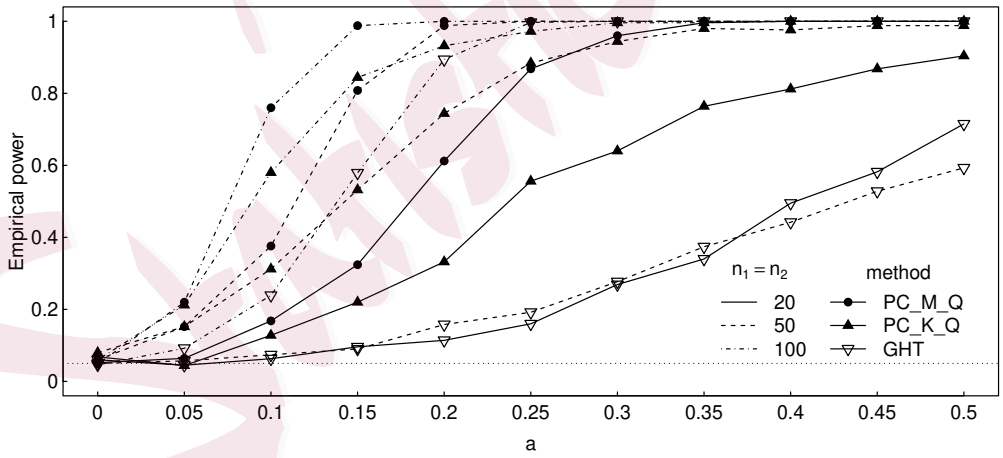

Figure 4: The empirical power function of different methods with sample size $n_1 = n_2 = 20, 50, 100$. The horizontal dotted line represents the confidence level of 0.05.

5. Case Study

To demonstrate the practical application of the proposed tests, we conduct an experiment on the BeetleFly data set (Dau et al., 2018). This data set involves distinguishing between the outlines of a beetle and a fly, with 20 instances in each class. Each instance is represented by a functional data observed at 512 points, obtained by mapping the outline into a 1-D series of distances to the center.

We apply the eleven testing methods in simulation to examine whether the outlines of a beetle and a fly are significantly different. With the exception of the method PC_K_λ, all other methods conclude that the outlines are significantly different. Furthermore, we investigate the sensitivity of these methods to the sparsity of functional observations in practice. We partition the observations of each instance into K parts, creating K new data sets. We apply the ten testing methods on each of the K data sets and record whether the null hypothesis is rejected or not. The rejection percentage is the average of all the K data sets. We selected $K = 2, \ldots, 10$, where a larger value of K indicates sparser functional observations in the new data set. Table 1 presents the rejection percentage of 4 methods, and the other 6 methods have a rejection percentage of 1 for all K values.

Additionally, we examine the sensitivity of these methods to the sam-

Table 1: The stability of different methods to the sparsity of functional observations, where a larger value of K indicates sparser observations

method	$K=2$	3	4	5	6	7	8	9	10
HT	1	1	1	1	1	1	1	1	0.6
GHT	1	1	1	1	1	1	1	1	0.9
RP	0.5	0.33	0.75	0.8	0.83	0.86	0.38	0.78	0.5
PC_K_λ	0	0	0	0	0	0	0	0	0

Table 2: The stability of different methods to the sample size, where a smaller value of n_0 indicates a smaller sample size

method	$n_0=18$	16	14
BS	1	1	1
PC_M_Q	1	1	1
PC_K_Q	1	1	1
CG	1	1	1
HT	1	1	0.98
GHT	1	1	1
RP	0.52	0.44	0.48
L_2	0.92	0.88	0.64
MMD	1	1	0.92
PC_M_λ	0.52	0.48	0.40
PC_K_λ	0.20	0.08	0.16

ple size in practice. Specifically, for a given n_0, we randomly sampled n_0 instances without replacement from each class to form a new data set. We repeat this process 50 times and run the test. We consider $n_0 = 18, 16, 14$, where a smaller value of n_0 indicates a smaller sample size. Table 2 presents the rejection percentage of different methods under different n_0.

Given that beetles and flies are two distinct creatures, it is reasonable to assume that their outlines are different. Therefore, a higher rejection

percentage in Tables 1 and 2 indicates a better method. From the results, it is evident that the four methods proposed in this paper demonstrate greater stability in handling observation sparsity and small sample size.

6. Conclusion and Future Work

Projection-based methods have gained significant popularity in FDA, and this paper tackles a crucial aspect of such methods: the selection of projection functions that minimize information loss and enhance the efficiency of two-sample tests. By conducting an in-depth analysis of the relationship between testing power and projection function selection, we offer valuable guidelines for choosing practical projection functions that perform well. The methods we propose, based on our investigation of three practical projection spaces, have shown promising results in our simulations.

Furthermore, we emphasize the flexibility and adaptability of the proposed projection-based test, particularly in scenarios with relaxed assumptions or more complex data structures. Notably, the assumption of equal covariance matrices can be circumvented by utilizing an appropriately estimated pooled covariance matrix. While our paper does not explicitly tackle the non-Gaussian scenario, this challenge can be adeptly managed by acquiring critical values through permutation techniques. Additionally, the

analytical approach we have developed for selecting projection functions extends beyond the realm of two-sample testing; it holds applicability in various other FDA domains, including projection-based clustering, classification, and regression.

Supplementary Materials

The online Supplementary Material contains additional simulations and the technical proofs.

Acknowledgements

Bai's research was supported by the Innovative Research Team of Shanghai University of Finance and Economics #2022110918. Zhu's research was supported by the General Research Fund #14308823 and Direct Grants for the Chinese University of Hong Kong #4053590. Yang Bai would also like to acknowledge the support of the Shanghai Research Center for Data Science and Decision Technology. The authors thank the editor, associate editor, and two anonymous referees for their constructive comments and suggestions.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis (3rd ed.)*. John Wiley & Sons.

Bai, Z. and H. Saranadasa (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica 6*(2), 311–329.

Benko, M., W. Härdle, A. Kneip, et al. (2009). Common functional principal components. *The Annals of Statistics 37*(1), 1–34.

Cox, D. D. and J. S. Lee (2008). Pointwise testing with functional data using the westfall–young randomization method. *Biometrika 95*(3), 621–634.

Cuesta-Albertos, J. and M. Febrero-Bande (2010). A simple multiway anova for functional data. *Test 19*(3), 537–557.

Dau, H. A., E. Keogh, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, Yanping, B. Hu, N. Begum, A. Bagnall, A. Mueen, G. Batista, and Hexagon-ML (2018, October). The ucr time series classification archive. `https://www.cs.ucr.edu/~eamonn/time_series_data_2018/`.

Estévez-Pérez, G. and J. A. Vilar (2013). Functional anova starting from discrete data: an application to air quality data. *Environmental and ecological statistics 20*(3), 495–517.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.

Fan, J. and S.-K. Lin (1998). Test of significance when data are curves. *Journal of the American*

Statistical Association 93(443), 1007–1021.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics 39*(3), 254–261.

Galeano, P., E. Joseph, and R. E. Lillo (2015). The mahalanobis distance for functional data with applications to classification. *Technometrics 57*(2), 281–291.

Górecki, T. and Ł. Smaga (2015). A comparison of tests for the one-way anova problem for functional data. *Computational Statistics 30*(4), 987–1010.

Gupta, S. D. and M. D. Perlman (1974). Power of the noncentral f-test: effect of additional variates on hotelling's t2-test. *Journal of the American Statistical Association 69*(345), 174–180.

Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.

Horváth, L., P. Kokoszka, and R. Reeder (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(1), 103–122.

Huang, Y. (2015). *Projection test for high-dimensional mean vectors with optimal direction*. PhD dissertation, Department of Statistics, The Pennsylvania State University at University Park.

Kraus, D. and M. Stefanucci (2019). Classification of functional fragments by regularized linear

classifiers with domain selection. *Biometrika 106*(1), 161–180.

Laukaitis, A. and A. Račkauskas (2005). Functional data analysis for clients segmentation tasks. *European journal of operational research 163*(1), 210–216.

Meléndez, R., R. Giraldo, and V. Leiva (2021). Sign, wilcoxon and mann-whitney tests for functional data: An approach based on random projections. *Mathematics 9*(1), 44.

Pini, A., A. Stamm, and S. Vantini (2018). Hotelling's t2 in separable hilbert spaces. *Journal of Multivariate Analysis 167*, 284–305.

Pomann, G.-M., A.-M. Staicu, and S. Ghosh (2016). A two sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *Journal of the Royal Statistical Society. Series C, Applied statistics 65*(3), 395.

Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis*. Springer, New York.

Shen, Q. and J. Faraway (2004). An f test for linear models with functional responses. *Statistica Sinica 14*(4), 1239–1257.

Wang, Q., S. Guo, F. Yao, and C. Zou (2022). Thresholding mean test for functional data with power enhancement. *Stat 11*(1), e509.

Wynne, G. and A. B. Duncan (2022). A kernel two-sample test for functional data. *Journal of Machine Learning Research 23*(73), 1–51.

Zhang, J.-T. (2013). *Analysis of Variance for Functional Data*. Chapman and Hall/CRC.

Zhang, J.-T., M.-Y. Cheng, H.-T. Wu, and B. Zhou (2019). A new test for functional one-

way anova with applications to ischemic heart screening. *Computational Statistics & Data Analysis 132*, 3–17.

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

E-mail: statbyang@mail.shufe.edu.cn

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

E-mail: caihong.qin@163.sufe.edu.cn

Department of Statistics, The Chinese University of Hong Kong, Hong Kong 999077, China

E-mail: huichenzhu@cuhk.edu.hk