

Article

Testing Equality of Several Distributions at High Dimensions: A Maximum-Mean-Discrepancy-Based Approach

Zhi Peng Ong ¹, Aixiang Andy Chen ², Tianming Zhu ³ and Jin-Ting Zhang ^{4,*}

¹ Department of Information Systems and Analytics, National University of Singapore, Singapore 117417, Singapore; zpodis@nus.edu.sg

² School of Statistics and Mathematics, Guangdong University of Finance and Economics, Guangzhou 510320, China; chenyh56@mail.sysu.edu.cn

³ National Institute of Education, Nanyang Technological University, Singapore 637616, Singapore; tianming.zhu@nie.edu.sg

⁴ Department of Statistics and Data Science, National University of Singapore, Singapore 117546, Singapore

* Correspondence: stazjt@nus.edu.sg

Abstract: With the development of modern data collection techniques, researchers often encounter high-dimensional data across various research fields. An important problem is to determine whether several groups of these high-dimensional data originate from the same population. To address this, this paper presents a novel k -sample test for equal distributions for high-dimensional data, utilizing the Maximum Mean Discrepancy (MMD). The test statistic is constructed using a V-statistic-based estimator of the squared MMD derived for several samples. The asymptotic null and alternative distributions of the test statistic are derived. To approximate the null distribution accurately, three simple methods are described. To evaluate the performance of the proposed test, two simulation studies and a real data example are presented, demonstrating the effectiveness and reliability of the test in practical applications.

Keywords: multi-sample test; hypothesis testing; parametric bootstrap; random permutation; Welch–Satterthwaite χ^2 -approximation; chi-squared-type mixtures

MSC: 62H15



Citation: Ong, Z.P.; Chen, A.A.; Zhu, T.; Zhang, J.-T. Testing Equality of Several Distributions at High Dimensions: A Maximum-Mean-Discrepancy-Based Approach. *Mathematics* **2023**, *11*, 4374. <https://doi.org/10.3390/math11204374>

Academic Editor: Ionut Florescu

Received: 25 September 2023

Revised: 18 October 2023

Accepted: 18 October 2023

Published: 21 October 2023



Copyright: © 2023 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/>).

1. Introduction

Testing whether multiple samples follow the same distribution is a fundamental challenge in data analysis, with wide-ranging applications across diverse fields. Traditional nonparametric tests designed for comparing the distributions of two samples, such as the Wald–Wolfowitz runs test, the Mann–Whitney–Wilcoxon test based on signed ranks, and the Kolmogorov–Smirnov test utilizing the Empirical Distribution Function (EDF), are well established for univariate data [1]. Extending these tests to the multivariate setting in $\mathcal{R}^p$ has been the focus of extensive research. This has led to the development of novel tests based on multivariate runs, ranks, EDF, distances, and projections, as pioneered by researchers like [2–5], among others.

In this paper, our primary focus is on addressing the multi-sample problem for equal distributions in high-dimensional data. In contemporary data analysis, high-dimensional datasets have become increasingly prevalent and easily accessible across various domains. For instance, Section 5 introduces a dataset derived from a keratoconus study involving corneal surfaces. This collaborative effort involves Ms. Nancy Tripoli and Dr. Kenneth L. Cohen from the Department of Ophthalmology at the University of North Carolina, Chapel Hill. The dataset comprises 150 corneal surfaces, each characterized by over 6000 height measurements. These surfaces are categorized into four distinct groups based on their corneal shapes, prompting the question of whether these groups share a common

underlying distribution. Consequently, there is a pressing need to develop tests that can assess distributional equality in the context of high-dimensional data.

Mathematically, a multi-sample problem for high-dimensional data can be described as follows. Assume that we have the following k samples of observed random elements in $\mathcal{R}^p$:

$$\mathbf{y}_{\alpha 1}, \dots, \mathbf{y}_{\alpha n_{\alpha}} \stackrel{\text{i.i.d.}}{\sim} F_{\alpha}, \alpha = 1, \dots, k, \quad (1)$$

where the data dimension p can be very large and $F_{\alpha}, \alpha = 1, \dots, k$, are unknown cumulative distribution functions. Of interest is to test if the k distribution functions are the same:

$$H_0 : F_1 = \dots = F_k, \text{ vs. } H_1 : F_{\alpha} \neq F_{\beta} \text{ for some } \alpha \neq \beta. \quad (2)$$

When $k = 2$, a variety of distance-based tests designed for multivariate data, such as those proposed by [2–5], can potentially be applied to test (2) in the context of high-dimensional data. However, it has been demonstrated by [6] that these tests may lack the power to effectively detect differences in distribution scales in high dimensions. In response to this challenge, ref. [6] introduced a test based on interpoint distances, while [6] proposed a high-dimensional two-sample run test based on the shortest Hamiltonian path. Nevertheless, ref. [7] showed that these tests, along with a graph-based test by [8], are less effective in detecting differences in location. To address this limitation, ref. [7] introduced two asymptotic normal tests for both location and scale differences, based on interpoint distances. However, it is worth noting that these tests rely on a strong mixing condition and impose a natural ordering on the components of high-dimensional observations, limiting their applicability. Furthermore, these tests involve complex U-statistic estimators, making them computationally intensive. Several other approaches have also been proposed for high-dimensional distribution testing. Ref. [9] presented a high-dimensional permutation test using a symmetric measure of distance between data points, but it comes with high computational costs. Refs. [10,11] proposed tests based on projections, which are generally more effective in detecting low-dimensional distributional differences. Ref. [12] introduced a test based on energy distance and permutation, but it is less powerful in detecting scale differences. Refs. [13,14] proposed kernel two-sample tests based on the Maximum Mean Discrepancy (MMD). Ref. [14] demonstrated the equivalence between the energy-distance-based test and the kernel-based test, showing that the energy-distance-based test can be viewed as a kernel test utilizing a kernel induced by the interpoint distance. The MMD leverages the kernel trick to define a distance between the embeddings of distributions in Reproducing Kernel Hilbert Spaces (RKHSs). It is well suited for checking distribution differences among several high-dimensional samples and is applicable to various data types, such as vectors, strings, or graphs. Recently, there has been further investigation into unbiased and biased MMD-based two-sample tests for high-dimensional data by [15,16], respectively.

On the other hand, when $k > 2$, there are limited tests available for testing (2). One such test is the energy test developed by [12]. This test statistic is obtained by directly summing all pairwise energy distances, with its null distribution approximated through permutation. However, this test has some drawbacks, including being time-consuming and yielding p -values that may vary when applied multiple times to the same dataset, as supported by the evidence from Figure 3 in Section 5. Another approach is presented by [17], who extended the idea of MMD from the two-sample problem to the multi-sample problem for equal distributions. They constructed an MMD-based test statistic capable of detecting deviations from distribution homogeneity in several samples. However, the resulting test statistic and its null limit distribution are very complicated in form, which restricts its practical use. To address this limitation, ref. [18] developed a new MMD-based test for testing (2). This test statistic is constructed using U-statistics, making it easy to conduct and yielding accurate results.

In this paper, we maintain our focus on the MMD-based approach for testing (2). However, unlike [18], where a U-statistic technique is employed to construct the test statistic, we take a distinct path by constructing an L^2 -norm-based test statistic. To achieve

this, we first employ a canonical feature map to transform the k original samples (1) into k induced samples (3). Concurrently, we transform the k -sample equal distribution testing problem (2) into a mean vector testing problem (4). This transformation facilitates the straightforward construction of an L^2 -norm-based test (5) for assessing the mean vector testing problem (4). Leveraging a kernel trick, we derive a formula for computing the L^2 -norm-based test statistic using the k original samples (1).

Additionally, this paper makes several other significant contributions. Firstly, akin to the work of [17,18], we extend the concept of MMD from two-sample problems to the domain of multi-sample problems with equal distributions. Secondly, we derive the asymptotic null and alternative distributions of the proposed test statistic. Thirdly, we offer three distinct approaches for approximating the null distribution of the MMD-based test statistic, utilizing parametric bootstrap, random permutation, and the Welch–Satterthwaite (W–S) χ^2 -approximation methods. Lastly, we examine two specific scenarios in our comprehensive simulation studies. In the first scenario, the samples have the same mean vector but different covariance matrices, while in the second scenario, the samples exhibit both distinct mean vectors and covariance matrices. Our simulation results demonstrate that the tests we propose effectively maintain precise control over size in both scenarios. However, in terms of empirical power, they outperform (underperform) the energy test introduced by [12] in the first (second) scenario. In other words, when the primary difference in the distributions of the samples lies in their covariance matrices, the new tests are the preferred choice in terms of statistical power.

The remainder of this paper is organized as follows: Section 2 presents the main results and Section 3 introduces three methods to implement our test. In Section 4, we provide two simulation studies. Section 5 showcases an application to the corneal surface data mentioned earlier. Concluding remarks can be found in Section 6. Technical proofs of the main results are included in Appendix A.

2. Main Results

2.1. MMD for Several Distributions

In this section, we show how the MMD can be defined for several distributions. Let $\mathcal{H}$ be an RKHS associated with a characteristic reproducing kernel $K(\cdot, \cdot)$. For any u and v in $\mathcal{H}$, the inner product and L^2 -norm of $\mathcal{H}$ are defined as $\langle u, v \rangle$ and $\|u\| = \sqrt{\langle u, u \rangle}$, respectively. Let $\phi(\cdot)$ be the canonical feature mapping associated with $K(\cdot, \cdot)$, i.e., $\phi(y) = K(\cdot, y)$. Using this feature mapping and setting $x_{\alpha i} = \phi(y_{\alpha i})$ for $i = 1, \dots, n_\alpha$, we obtain the following k induced samples in the RKHS $\mathcal{H}$:

$$x_{\alpha 1}, \dots, x_{\alpha n_\alpha}, \alpha = 1, \dots, k, \quad (3)$$

which are derived from the k original samples in (1). Define $\mu_\alpha = E(x_{\alpha 1})$ for $\alpha = 1, \dots, k$, representing the mean embeddings of the k distributions F_α , where $\alpha = 1, \dots, k$.

Ref. [13] established the MMD for two distributions in a separable metric space (see, e.g., Theorem 5 of [13]). Here, we extend it naturally for multi-sample distributions in $\mathcal{R}^p$. According to the MMD of [13], for any $\alpha \neq \beta$, where $\alpha, \beta = 1, \dots, k$, “testing $H_0 : F_\alpha = F_\beta$ vs. $H_1 : F_\alpha \neq F_\beta$ ” based on the two samples $y_{\alpha i}, i = 1, \dots, n_\alpha$ and $y_{\beta i}, i = 1, \dots, n_\beta$ is equivalent to “testing $H_0 : \mu_\alpha = \mu_\beta$ vs. $H_1 : \mu_\alpha \neq \mu_\beta$ ” based on the two samples $x_{\alpha i}, i = 1, \dots, n_\alpha$ and $x_{\beta i}, i = 1, \dots, n_\beta$. Therefore, testing (2) using the k original samples in (1) is equivalent to testing the following hypothesis using the k induced samples in (3):

$$H_0 : \mu_1 = \dots = \mu_k, \text{ vs. } H_1 : \mu_\alpha \neq \mu_\beta \text{ for some } \alpha \neq \beta. \quad (4)$$

To test (4), following [19], a natural L^2 -norm-based test statistic using (3) is given by

$$T_n = \sum_{\alpha=1}^k n_\alpha \|\bar{x}_\alpha - \bar{x}\|^2, \quad (5)$$

where $\bar{x}_\alpha = n_\alpha^{-1} \sum_{i=1}^{n_\alpha} x_{\alpha i}$ and $\bar{x} = n^{-1} \sum_{\alpha=1}^k \sum_{i=1}^{n_\alpha} x_{\alpha i}$ denote the group and grand sample means, respectively. Through some simple algebra, as given in Appendix A, we can express T_n as

$$T_n = \sum_{\alpha=1}^k \frac{n_\alpha(n - n_\alpha)}{n} \|\bar{x}_\alpha\|^2 - 2 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_\alpha n_\beta}{n} \langle \bar{x}_\alpha, \bar{x}_\beta \rangle = \sum_{1 \leq \alpha < \beta \leq k} \frac{n_\alpha n_\beta}{n} \|\bar{x}_\alpha - \bar{x}_\beta\|^2. \quad (6)$$

Let $\pi = (\pi_1, \dots, \pi_k)^\top$ denote the weights of the k distributions $F_1, \dots, F_k$ such that $\pi_1, \dots, \pi_k \in (0, 1)$ and $\sum_{\alpha=1}^k \pi_\alpha = 1$. Then, when we estimate π using $(n_1, \dots, n_k)^\top / n$, the test statistic T_n / n estimates the following quantity:

$$\text{MMD}^2(F_1, \dots, F_k | \pi) = \sum_{1 \leq \alpha < \beta \leq k} \pi_\alpha \pi_\beta \|\mu_\alpha - \mu_\beta\|^2,$$

which can be naturally defined as the MMD of the k distributions $F_1, \dots, F_k$ with the weight vector π . It is worth noting that the MMD for multiple distributions presented above is equivalent to the one derived by [18], offering a much simpler alternative compared to the formulation proposed by [17].

It is easy to justify that $\text{MMD}^2(F_1, \dots, F_k | \pi)$ is indeed an MMD of the k distributions $F_1, \dots, F_k$ in the sense that $\text{MMD}^2(F_1, \dots, F_k | \pi) = 0$ if and only if $F_1 = \dots = F_k$. On the one hand, when $\text{MMD}^2(F_1, \dots, F_k | \pi) = 0$, we have $\mu_1 = \dots = \mu_k$ so that for any $1 \leq \alpha \neq \beta \leq k$, we have $\mu_\alpha = \mu_\beta$, implying that for any $1 \leq \alpha \neq \beta \leq k$, we have $F_\alpha = F_\beta$ and hence $F_1 = \dots = F_k$. On the other hand, when $F_1 = \dots = F_k$, for any $1 \leq \alpha \neq \beta \leq k$, we have $F_\alpha = F_\beta$ so that for any $1 \leq \alpha \neq \beta \leq k$, we have $\mu_\alpha = \mu_\beta$ and hence $\mu_1 = \dots = \mu_k$. Therefore, we can use T_n to test the equality of k distributions based on the k induced samples (3).

2.2. Computation of the Test Statistic

Notice that the k induced samples (3) are not directly computable, as the canonical feature mapping $\phi(y)$ is explicitly defined through the reproducing kernel. Fortunately, the reproducing kernel $K(\cdot, \cdot)$ and its canonical feature mapping $\phi(\cdot)$ can be utilized with the following useful kernel trick: $K(y, y') = \langle \phi(y), \phi(y') \rangle$. Using this, we can express the inner product as follows:

$$\langle x_{\alpha i}, x_{\beta j} \rangle = \langle \phi(y_{\alpha i}), \phi(y_{\beta j}) \rangle = K(y_{\alpha i}, y_{\beta j}). \quad (7)$$

Let $V_{\alpha\alpha} = \|\bar{x}_\alpha\|^2 = \langle \bar{x}_\alpha, \bar{x}_\alpha \rangle$ and $V_{\alpha\beta} = \langle \bar{x}_\alpha, \bar{x}_\beta \rangle$. Then, using (7), we have

$$V_{\alpha\alpha} = \frac{1}{n_\alpha^2} \sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\alpha} K(y_{\alpha i}, y_{\alpha j}), \quad \text{and} \quad V_{\alpha\beta} = \frac{1}{n_\alpha n_\beta} \sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} K(y_{\alpha i}, y_{\beta j}).$$

Therefore, using (6), we can compute T_n using any of the following useful expressions:

$$T_n = \sum_{\alpha=1}^k \frac{n_\alpha(n - n_\alpha)}{n} V_{\alpha\alpha} - 2 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_\alpha n_\beta}{n} V_{\alpha\beta} = \sum_{1 \leq \alpha < \beta \leq k} \frac{n_\alpha n_\beta}{n} (V_{\alpha\alpha} + V_{\beta\beta} - 2V_{\alpha\beta}). \quad (8)$$

In other words, you can compute the value of T_n using the original k samples (1) and the above expressions.

2.3. Asymptotic Null Distribution

To explore the null distribution of T_n , we can rewrite T_n as

$$T_n = \tilde{T}_n + 2S_n + \sum_{\alpha=1}^k n_\alpha \|\mu_\alpha - \bar{\mu}\|^2, \quad (9)$$

where $\bar{\mu} = n^{-1} \sum_{\alpha=1}^k n_{\alpha} \mu_{\alpha}$ represents the weighted average of the mean embeddings of the k distributions. Additionally, we have

$$\tilde{T}_n = \sum_{\alpha=1}^k n_{\alpha} \|(\bar{x}_{\alpha} - \mu_{\alpha}) - (\bar{x} - \bar{\mu})\|^2, \text{ and } S_n = \sum_{\alpha=1}^k n_{\alpha} \langle (\bar{x}_{\alpha} - \mu_{\alpha}) - (\bar{x} - \bar{\mu}), \mu_{\alpha} - \bar{\mu} \rangle. \quad (10)$$

In the expression for $\tilde{T}_n$, you can observe that the mean embeddings of the k distributions have been subtracted. Consequently, $\tilde{T}_n$ follows the same distribution as that of T_n under the null hypothesis. Thus, studying the null distribution of T_n is equivalent to studying the distribution of $\tilde{T}_n$.

Similar to the proof of (6), we can express $\tilde{T}_n$ as follows:

$$\begin{aligned} \tilde{T}_n &= \sum_{\alpha=1}^k \frac{n_{\alpha}(n-n_{\alpha})}{n} \|\bar{x}_{\alpha} - \mu_{\alpha}\|^2 - 2 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_{\alpha} n_{\beta}}{n} \langle \bar{x}_{\alpha} - \mu_{\alpha}, \bar{x}_{\beta} - \mu_{\beta} \rangle \\ &= \sum_{1 \leq \alpha < \beta \leq k} \frac{n_{\alpha} n_{\beta}}{n} \|(\bar{x}_{\alpha} - \mu_{\alpha}) - (\bar{x}_{\beta} - \mu_{\beta})\|^2. \end{aligned} \quad (11)$$

Let $\tilde{K}(u, v)$ denote the centered version of $K(u, v)$, defined as

$$\begin{aligned} \tilde{K}(u, v) &= \langle \phi(u) - \mu_u, \phi(v) - \mu_v \rangle \\ &= K(u, v) - E_{v'}[K(u, v')] - E_{u'}[K(u', v)] + E_{u', v'}[K(u', v')], \end{aligned} \quad (12)$$

where $\mu_u = E[\phi(u)]$, $\mu_v = E[\phi(v)]$, and u' and v' are independent copies of u and v , respectively. We can observe two useful properties: when $u = v$, we have

$$E_u[\tilde{K}(u, u)] = E_u \|\phi(u) - \mu_u\|^2 > 0. \quad (13)$$

When u and v are independent, we have

$$E_u[\tilde{K}(u, v)] = E_v[\tilde{K}(u, v)] = E_{u,v}[\tilde{K}(u, v)] = 0. \quad (14)$$

Using (12), we can express

$$\tilde{K}(y_{\alpha i}, y_{\beta j}) = \langle x_{\alpha i} - \mu_{\alpha}, x_{\beta j} - \mu_{\beta} \rangle. \quad (15)$$

Let

$$\tilde{V}_{\alpha\alpha} = \|\bar{x}_{\alpha} - \mu_{\alpha}\|^2, \text{ and } \tilde{V}_{\alpha\beta} = \langle \bar{x}_{\alpha} - \mu_{\alpha}, \bar{x}_{\beta} - \mu_{\beta} \rangle. \quad (16)$$

It is evident that $\tilde{V}_{\alpha\alpha}$ and $\tilde{V}_{\alpha\beta}$ can be considered as centered versions of $V_{\alpha\alpha}$ and $V_{\alpha\beta}$, respectively. Consequently, by using (11) and (16), we can express

$$\tilde{T}_n = \sum_{\alpha=1}^k \frac{n_{\alpha}(n-n_{\alpha})}{n} \tilde{V}_{\alpha\alpha} - 2 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_{\alpha} n_{\beta}}{n} \tilde{V}_{\alpha\beta}. \quad (17)$$

Utilizing (15) and performing some straightforward algebraic manipulations, we can express

$$\begin{aligned} \tilde{V}_{\alpha\alpha} &= \frac{1}{n_{\alpha}^2} \sum_{i=1}^{n_{\alpha}} \sum_{j=1}^{n_{\alpha}} \tilde{K}(y_{\alpha i}, y_{\alpha j}) = \frac{1}{n_{\alpha}^2} \left[\sum_{i=1}^{n_{\alpha}} \tilde{K}(y_{\alpha i}, y_{\alpha i}) + 2 \sum_{1 \leq i < j \leq n_{\alpha}} \tilde{K}(y_{\alpha i}, y_{\alpha j}) \right], \text{ and} \\ \tilde{V}_{\alpha\beta} &= \frac{1}{n_{\alpha} n_{\beta}} \sum_{i=1}^{n_{\alpha}} \sum_{j=1}^{n_{\beta}} \tilde{K}(y_{\alpha i}, y_{\beta j}), \alpha \neq \beta. \end{aligned} \quad (18)$$

Assuming that $\tilde{K}(y, y')$ is square-integrable, i.e., $E[\tilde{K}^2(y, y')] < \infty$, we can express $\tilde{K}(y, y')$ using Mercer's expansion:

$$\tilde{K}(y, y') = \sum_{r=1}^{\infty} \lambda_r \psi_r(y) \psi_r(y'), \quad (19)$$

where $\lambda_1, \lambda_2, \dots$ represent the eigenvalues of $\tilde{K}(\mathbf{y}, \mathbf{y}')$, and $\psi_1(\mathbf{y}), \psi_2(\mathbf{y}), \dots$ are the corresponding orthonormal eigenelements, satisfying

$$\int \tilde{K}(\mathbf{y}, \mathbf{y}') \psi_r(\mathbf{y}) dF(\mathbf{y}) = \lambda_r \psi_r(\mathbf{y}'), \text{ and } \int \psi_r(\mathbf{y}) \psi_s(\mathbf{y}) dF(\mathbf{y}) = \delta_{rs}, \quad r, s = 1, 2, \dots, \quad (20)$$

where δ_{rs} equals 1 when $r = s$ and 0 otherwise. Now, let us introduce the following conditions:

- C1. We have $\mathbf{y}_{\alpha 1}, \dots, \mathbf{y}_{\alpha, n_\alpha}, \alpha = 1, \dots, k \stackrel{\text{i.i.d.}}{\sim} F$.
- C2. As $n \rightarrow \infty$, we have $n_\alpha / n \rightarrow \tau_\alpha \in (0, 1), \alpha = 1, \dots, k$.
- C3. $K(\mathbf{y}, \mathbf{y}')$ is a reproduced kernel such that $E_y[\tilde{K}^2(\mathbf{y}, \mathbf{y})] \leq \infty$.

Condition C1 assumes that the null hypothesis is satisfied and the common distribution function is F . Condition C2 is a regularity condition for k -sample problems and it requires that the group sample sizes tend to ∞ proportionally. Condition C3 is required such that $\tilde{K}(\mathbf{y}, \mathbf{y}')$ is square-integrable and expression (19) is valid.

In fact, under Condition C3, using (19) and the Cauchy–Schwarz inequality, we obtain the following results:

$$\begin{aligned} E[\tilde{K}(\mathbf{y}, \mathbf{y})] &= \sum_{r=1}^{\infty} \lambda_r < \sqrt{E[\tilde{K}^2(\mathbf{y}, \mathbf{y})]} < \infty, \\ E_{\mathbf{y}, \mathbf{y}'}[\tilde{K}(\mathbf{y}, \mathbf{y}')^2] &= \sum_{r=1}^{\infty} \lambda_r^2 < E_{\mathbf{y}, \mathbf{y}'}[\tilde{K}(\mathbf{y}, \mathbf{y})\tilde{K}(\mathbf{y}', \mathbf{y}')] = E_y^2[\tilde{K}(\mathbf{y}, \mathbf{y})] < \infty, \end{aligned} \quad (21)$$

where $\mathbf{y}, \mathbf{y}' \stackrel{\text{i.i.d.}}{\sim} F$. These inequalities hold due to the square-integrability assumption and the properties of Mercer's expansion. Now, let us state the following theorem that establishes the asymptotic distribution of $\tilde{T}_n$.

Theorem 1. Under Conditions C1–C3, as $n \rightarrow \infty$, we have $\tilde{T}_n \xrightarrow{\mathcal{L}} \tilde{T}$, where

$$\tilde{T} \stackrel{d}{=} \sum_{r=1}^{\infty} \lambda_r A_r, \quad A_r \stackrel{\text{i.i.d.}}{\sim} \chi_{k-1}^2. \quad (22)$$

It is worth highlighting that the limit null distribution of the proposed test statistic differs from the one derived in [18] (Theorem 1) and it offers a more straightforward alternative compared to the limit null distribution obtained by [17] (Theorem 1). This explains why the limit null distribution presented by [17] is not employed to approximate the null distribution of their test statistic. However, as demonstrated in Section 3, it is indeed feasible to utilize this distribution if desired.

2.4. Mean and Variance of $\tilde{T}_n$

Based on (13), (14) and (18), through some simple calculation, under Condition C1, we have

$$\begin{aligned} E[\tilde{V}_{\alpha\alpha}] &= \frac{1}{n_\alpha} E[\tilde{K}(\mathbf{y}, \mathbf{y})], E[\tilde{V}_{\alpha\beta}] = 0, \\ \text{Var}[\tilde{V}_{\alpha\alpha}] &= \frac{1}{n_\alpha^3} \text{Var}[\tilde{K}(\mathbf{y}, \mathbf{y})] + \frac{2(n_\alpha - 1)}{n_\alpha^3} E[\tilde{K}^2(\mathbf{y}, \mathbf{y}')], \\ \text{Var}[\tilde{V}_{\alpha\beta}] &= \frac{1}{n_\alpha n_\beta} E[\tilde{K}^2(\mathbf{y}, \mathbf{y}')], \text{ and } \text{Cov}(\tilde{V}_{\alpha\alpha}, \tilde{V}_{\alpha\beta}) = 0, \alpha \neq \beta, \end{aligned} \quad (23)$$

where $\mathbf{y}, \mathbf{y}' \stackrel{\text{i.i.d.}}{\sim} F$.

Theorem 2. Under Condition C1, we have $E(\tilde{T}_n) = (k - 1) E[\tilde{K}(\mathbf{y}, \mathbf{y})]$, and

$$\text{Var}(\tilde{T}_n) = \left[\sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \right] \text{Var}[\tilde{K}(\mathbf{y}, \mathbf{y})] + 2 \left[(k - 1) - \sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \right] E[\tilde{K}^2(\mathbf{y}, \mathbf{y}')],$$

where $\mathbf{y}, \mathbf{y}' \stackrel{i.i.d.}{\sim} F$.

Note that under Condition C2, we have $\sum_{\alpha=1}^k (n - n_{\alpha})^2 / (n^2 n_{\alpha}) \leq \sum_{\alpha=1}^k n_{\alpha}^{-1} \rightarrow 0$ as $n \rightarrow \infty$. Then as $n \rightarrow \infty$, we have $\text{Var}(\tilde{T}_n) = 2(k-1) \mathbb{E}[\tilde{K}^2(\mathbf{y}, \mathbf{y}')] [1 + o(1)]$.

2.5. Asymptotic Power

In this subsection, we examine the asymptotic power of the proposed test under the following local alternative hypothesis:

$$H_{1n} : \boldsymbol{\mu}_{\alpha} = \boldsymbol{\mu} + n^{-(1/2-\Delta)} \mathbf{h}_{\alpha}, \alpha = 1, \dots, k, \quad (24)$$

where $0 < \Delta \leq 1/2$ and $\mathbf{h}_{\alpha}, \alpha = 1, \dots, k$ are constant elements in the RKHS $\mathcal{H}$ such that

$$0 < \sum_{\alpha=1}^k \frac{n_{\alpha}}{n} \|\mathbf{h}_{\alpha} - \bar{\mathbf{h}}\|^2 \leq B_{\bar{\mathbf{h}}} < \infty \text{ with } \bar{\mathbf{h}} = \frac{1}{n} \sum_{\alpha=1}^k n_{\alpha} \mathbf{h}_{\alpha},$$

and

$$\sum_{\alpha=1}^k \sigma_{\alpha}^2 > 0, \text{ where } \sigma_{\alpha}^2 = \mathbb{E}[\langle \mathbf{x}_{\alpha 1} - \boldsymbol{\mu}_{\alpha}, \mathbf{h}_{\alpha} - \bar{\mathbf{h}} \rangle]^2, \alpha = 1, \dots, k.$$

It is important to note that when $\Delta = 1/2$, the local hypothesis (24) simplifies into a fixed alternative hypothesis $H_1 : \boldsymbol{\mu}_{\alpha} = \boldsymbol{\mu} + \mathbf{h}_{\alpha}$ for $\alpha = 1, \dots, k$. However, when $0 < \Delta < 1/2$, Equation (24) represents a strict local alternative hypothesis. In this case, as $n \rightarrow \infty$, the strict local hypothesis tends to converge toward the null hypothesis. Detecting a strict local alternative hypothesis becomes exceedingly challenging in such scenarios. A test is typically considered root- n -consistent if it can detect a strict local alternative hypothesis with a probability approaching 1 as $n \rightarrow \infty$. A root- n -consistent test is considered effective because it achieves the best possible detection rate for a local alternative hypothesis as n grows.

Through (9) and (24), T_n can be written as

$$T_n = \tilde{T}_n + 2S_n + n^{2\Delta} \sum_{\alpha=1}^k \frac{n_{\alpha}}{n} \|\mathbf{h}_{\alpha} - \bar{\mathbf{h}}\|^2, \quad (25)$$

where

$$S_n = n^{-(1/2-\Delta)} \sum_{\alpha=1}^k n_{\alpha} \langle \bar{\mathbf{x}}_{\alpha} - \boldsymbol{\mu}_{\alpha}, \mathbf{h}_{\alpha} - \bar{\mathbf{h}} \rangle.$$

Theorem 3. Assume that $|K(\mathbf{y}, \mathbf{y}')| \leq B_K$ for all $\mathbf{y}, \mathbf{y}' \in \mathcal{R}^p$ for some $B_K < \infty$. Then, we have $\text{Var}(\tilde{T}_n) \leq 32(k-1)B_K^2$ and $\text{Var}(S_n) = n^{2\Delta} \sum_{\alpha=1}^k n_{\alpha} \sigma_{\alpha}^2 / n$, where $\sum_{\alpha=1}^k n_{\alpha} \sigma_{\alpha}^2 / n \leq 4B_K B_{\bar{\mathbf{h}}} < \infty$.

Theorem 4. Assume that $|K(\mathbf{y}, \mathbf{y}')| \leq B_K$ for all $\mathbf{y}, \mathbf{y}' \in \mathcal{R}^p$ for some $B_K < \infty$. Then, under Condition C2 and the local alternative hypothesis (24), as $n \rightarrow \infty$, we have (a) $\tilde{T}_n / \sqrt{\text{Var}(S_n)} \xrightarrow{P} 0$; (b) $S_n / \sqrt{\text{Var}(S_n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$; and (c)

$$\Pr(T_n \geq C_{\epsilon}) = \Phi \left[\frac{n^{\Delta} \sum_{\alpha=1}^k \tau_{\alpha} \|\mathbf{h}_{\alpha} - \mathbf{h}^*\|^2}{2 \sqrt{\sum_{\alpha=1}^k \tau_{\alpha} \sigma_{\alpha}^{*2}}} \right] [1 + o(1)] \rightarrow 1,$$

where C_{ϵ} denotes the upper 100ϵ percentile of $\tilde{T}_n$ with ϵ being the given significance level, $\tau_{\alpha}; \alpha = 1, \dots, k$ are defined in Condition C2; $\mathbf{h}^* = \sum_{\alpha=1}^k \tau_{\alpha} \mathbf{h}_{\alpha}; \sigma_{\alpha}^{*2} = \mathbb{E}[\langle \mathbf{x}_{\alpha 1} - \boldsymbol{\mu}_{\alpha}, \mathbf{h}_{\alpha} - \mathbf{h}^* \rangle]^2$; and $\Phi(\cdot)$ denotes the cumulative distribution of $\mathcal{N}(0, 1)$.

Theorem 4 shows that the proposed test is indeed a root- n -consistent test.

3. Methods for Implementing the Proposed Test

In this section, we will outline three different approaches for approximating the null distribution of T_n (8) in order to conduct the proposed test. These methods include a parametric bootstrap approach, a random permutation technique, and a χ^2 -approximation method. We will evaluate and compare their performance in the next section.

3.1. Parametric Bootstrap Method

Theorem 1 reveals that the asymptotic null distribution of T_n takes the form of a chi-squared-type mixture denoted as $\tilde{T}$ (22). The coefficients of this mixture are determined by the unknown eigenvalues $\lambda_r, r = 1, 2, \dots$ of $\tilde{K}(\mathbf{y}, \mathbf{y}')$, where $\mathbf{y}, \mathbf{y}'$ are independently and identically distributed according to the common distribution function F representing the k distributions when the null hypothesis is valid. Consequently, in order to estimate the asymptotic null distribution of T_n , it is essential to consistently estimate $\lambda_r, r = 1, 2, \dots$. This consistency can be achieved by utilizing the empirical eigenvalues of the centered Gram matrix, as suggested by [20], to construct a reliable estimator for $\tilde{T}$ (22).

Let us recall that $n = \sum_{\alpha=1}^k n_\alpha$ represents the total sample size. We pool the k samples (1) and denote it as

$$\mathbf{y}_1, \dots, \mathbf{y}_n. \quad (26)$$

Under the null hypothesis, $\mathbf{y}_1, \dots, \mathbf{y}_n$ are independently and identically distributed from the common distribution F . Let K represent the $n \times n$ Gram matrix, where the (i, j) th entry is defined as $K(\mathbf{y}_i, \mathbf{y}_j)$ for $i, j = 1, \dots, n$. Additionally, let $\mathbf{1}_n$ denote a vector of ones with dimensions $n \times 1$, and I_n denote the identity matrix of size $n \times n$. Then, the matrix $P_n = I_n - \mathbf{1}_n \mathbf{1}_n^\top / n$ is a projection matrix of size $n \times n$ with rank $n - 1$.

Now, define $\tilde{K}_n = P_n K P_n$, commonly referred to as the centered Gram matrix. Its (i, j) th entry is given by

$$\tilde{K}_n(\mathbf{y}_i, \mathbf{y}_j) = K(\mathbf{y}_i, \mathbf{y}_j) - n^{-1} \sum_{v=1}^n K(\mathbf{y}_i, \mathbf{y}_v) - n^{-1} \sum_{u=1}^n K(\mathbf{y}_u, \mathbf{y}_j) + n^{-2} \sum_{u=1}^n \sum_{v=1}^n K(\mathbf{y}_u, \mathbf{y}_v), \quad (27)$$

where $i, j = 1, \dots, n$. As n approaches infinity, for any fixed i and j , we can observe that, by the law of large numbers:

$$\tilde{K}_n(\mathbf{y}_i, \mathbf{y}_j) \xrightarrow{\mathcal{L}} \tilde{K}(\mathbf{y}_i, \mathbf{y}_j) = K(\mathbf{y}_i, \mathbf{y}_j) - E_{\mathbf{y}'}[K(\mathbf{y}_i, \mathbf{y}')] - E_{\mathbf{y}}[K(\mathbf{y}, \mathbf{y}_j)] + E_{\mathbf{y}, \mathbf{y}'}[K(\mathbf{y}, \mathbf{y}')].$$

Let $\omega_1, \dots, \omega_q$ be all the non-zero eigenvalues of $\tilde{K}_n$ that can be obtained via an eigen-decomposition of $\tilde{K}_n$. Set $\hat{\lambda}_r = \omega_r / n, r = 1, \dots, q$. Then, following [20], we can show that under Condition C3, the distribution of $\tilde{T}$ can be consistently estimated by

$$\hat{\tilde{T}} = \sum_{r=1}^q \hat{\lambda}_r A_r, \quad A_r \stackrel{\text{i.i.d.}}{\sim} \chi_{k-1}^2. \quad (28)$$

The parametric bootstrap method can be described as follows. Let us choose a large value for N , for example, $N = 1000$. Using expression (28), we can obtain a sample $\hat{\tilde{T}}^{(1)}, \dots, \hat{\tilde{T}}^{(N)}$ of $\hat{\tilde{T}}$ by independently generating $A_r, r = 1, \dots, q$ a total of N times. Now, let T_{obs} represent the observed test statistic calculated using (8) based on the k samples (1). Using the parametric bootstrap method, we can conduct the proposed test by calculating the approximate p -value, which is given by $N^{-1} \sum_{i=1}^N I\{\hat{\tilde{T}}^{(i)} \geq T_{obs}\}$, where $I\{S\}$ is an indicator function that takes 1 when S is a true event and 0 otherwise.

3.2. Random Permutation Method

We can also approximate the null distribution of T_n using a random permutation method. Let $\ell_1, \dots, \ell_n$ represent a random permutation of the indices $1, \dots, n$ from the pooled sample (26). Consequently, the sequence

$$\mathbf{y}_{\ell_1}, \dots, \mathbf{y}_{\ell_n}, \quad (29)$$

forms a permutation of the pooled sample (26). To create permuted samples, we utilize the first n_1 observations in the permuted pooled sample (29) as the first permuted sample, the next n_2 observations as the second permuted sample, and so on, until we obtain k permuted samples. These permuted samples are denoted as

$$\mathbf{y}_{\alpha i}^*, i = 1, \dots, n_\alpha; \alpha = 1, \dots, k. \quad (30)$$

The permuted test statistic, denoted as T_n^* , is calculated using (8) but with the k samples (1) replaced by the k permuted samples (30).

The random permutation method proceeds as follows. Let N be a sufficiently large number, for instance, $N = 1000$. Suppose we repeat the permutation process described above N times, resulting in N permuted test statistics denoted as $T_n^{*(i)}, i = 1, \dots, N$. Then, we can use the empirical distribution of $T_n^{*(i)}, i = 1, \dots, N$ to approximate the null distribution of T_n . Recall that T_{obs} represents the test statistic computed using (8) based on the k original samples (1). Following the random permutation method, the proposed test can be conducted by calculating the approximated p -value, given by $N^{-1} \sum_{i=1}^N I\{T_n^{*(i)} \geq T_{obs}\}$.

3.3. Welch–Satterthwaite χ^2 -Approximation Method

The parametric bootstrap method and the random permutation method are effective for controlling size but can be computationally intensive, particularly with large total sample sizes. To address this issue, we can utilize the well-known Welch–Satterthwaite (W–S) χ^2 -approximation method [21,22]. This method is known to be reliable for approximating the distribution of a chi-squared-type mixture. Theorem 1 demonstrates that the asymptotic null distribution of T_n is a chi-squared-type mixture $\tilde{T}$ (22).

The core concept of the W–S χ^2 -approximation method is to approximate the null distribution of T_n using that of a random variable of the form: $W \stackrel{d}{=} \beta \chi_d^2$, where β and d are unknown parameters. These parameters can be determined by matching the means and variances of $\tilde{T}_n$ and W , where $\tilde{T}_n$, defined in (10), has the same distribution as T_n under the null hypothesis. Specifically, the mean and variance of W are βd and $2\beta^2 d$, respectively, while the mean and variance of $\tilde{T}_n$ are given in Theorem 2. Equating the means and variances of $\tilde{T}_n$ and W , we obtain

$$\beta = \frac{\text{Var}(\tilde{T}_n)}{2\text{E}(\tilde{T}_n)} \quad \text{and} \quad d = \frac{2\text{E}^2(\tilde{T}_n)}{\text{Var}(\tilde{T}_n)}. \quad (31)$$

To implement the W–S χ^2 -approximation method, we need to consistently estimate $\text{E}(\tilde{T}_n)$ and $\text{Var}(\tilde{T}_n)$ based on the pooled sample (26) from the k samples (1). According to Theorem 2, these estimates can be obtained as follows:

$$\begin{aligned} \hat{\text{E}}(\tilde{T}_n) &= (k-1)\hat{\text{E}}[\tilde{K}(\mathbf{y}, \mathbf{y})], \text{ and} \\ \widehat{\text{Var}}(\tilde{T}_n) &= \left[\sum_{\alpha=1}^k \frac{(n-n_\alpha)^2}{n^2 n_\alpha} \right] \widehat{\text{Var}}[\tilde{K}(\mathbf{y}, \mathbf{y})] + 2 \left[(k-1) - \sum_{\alpha=1}^k \frac{(n-n_\alpha)^2}{n^2 n_\alpha} \right] \hat{\text{E}}[\tilde{K}^2(\mathbf{y}, \mathbf{y}')], \end{aligned}$$

where

$$\begin{aligned}\widehat{\mathbb{E}}[\tilde{K}(\mathbf{y}, \mathbf{y})] &= n^{-1} \sum_{i=1}^n \tilde{K}_n(\mathbf{y}_i, \mathbf{y}_i), \\ \widehat{\text{Var}}[\tilde{K}(\mathbf{y}, \mathbf{y})] &= (n-1)^{-1} \sum_{i=1}^n \left\{ \tilde{K}_n(\mathbf{y}_i, \mathbf{y}_i) - n^{-1} \sum_{j=1}^n \tilde{K}_n(\mathbf{y}_j, \mathbf{y}_j) \right\}^2, \text{ and} \\ \widehat{\mathbb{E}}[\tilde{K}^2(\mathbf{y}, \mathbf{y}')] &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} [\tilde{K}_n(\mathbf{y}_i, \mathbf{y}_j)]^2,\end{aligned}$$

with $\tilde{K}_n(\mathbf{y}_i, \mathbf{y}_j)$ being defined in (27). Substituting these estimators into (31), we obtain

$$\hat{\beta} = \frac{\widehat{\text{Var}}(\tilde{T}_n)}{2\widehat{\mathbb{E}}(\tilde{T}_n)} \quad \text{and} \quad \hat{d} = \frac{2(\widehat{\mathbb{E}}(\tilde{T}_n))^2}{\widehat{\text{Var}}(\tilde{T}_n)}.$$

Let $\chi_d^2(\alpha^*)$ denote the upper $100\alpha^*$ percentile of χ_d^2 , where α^* is the given significance level, and let T_{obs} denote the observed test statistic computed using (8) based on the k samples (1). Then, through the W-S χ^2 -approximation method, the proposed test can be conducted via rejecting the null hypothesis when $T_{obs} > \hat{\beta}\chi_d^2(\alpha^*)$ or when the approximated p -value $\Pr(\chi_d^2 > T_{obs}/\hat{\beta})$ is less than α^* .

4. Simulation Studies

In this section, we delve into intensive simulation studies aimed at assessing the performance of the test we propose when compared to the energy test introduced by [12], which we denote as T_{SR} . Our proposed test employs the parametric bootstrap, the random permutation, and the W-S χ^2 -approximation methods as described in Section 3. For simplicity, we refer to the resulting tests as T_{PB} , T_{RP} , and T_{WS} , respectively.

For simplicity, we opt for the Gaussian Radial Basis Function (RBF) kernel, denoted as $K(\cdot, \cdot)$, which is defined as follows:

$$K(\mathbf{y}, \mathbf{y}') = \exp\left(-\frac{\|\mathbf{y} - \mathbf{y}'\|^2}{2\sigma^2}\right).$$

Here, σ^2 is referred to as the kernel width. It is worth noting that the Gaussian RBF kernel described above is bounded by 1, ensuring that Condition C3 is always met. Following the approach outlined in [20], we set σ to be equal to the median distance between observed vectors in the pooled sample (26).

We also employ the Average Relative Error (ARE) introduced by [23] to evaluate the overall effectiveness of a test in maintaining its nominal size. The ARE is calculated as follows: $ARE = 100M^{-1} \sum_{i=1}^M |\hat{\alpha}_i^* - \alpha^*|/\alpha^*$, where $\hat{\alpha}_i^*, i = 1, \dots, M$ represents the empirical sizes observed across M different simulation settings. A smaller ARE value indicates the better performance of a test in terms of size control. In this simulation study, we set the nominal size to $\alpha^* = 5\%$.

4.1. Simulation 1

We set $k = 3$ for simplicity. We generate the $k = 3$ samples (1) as follows. We set

$$\begin{aligned}\mathbf{y}_{1i} &= \boldsymbol{\mu} + \boldsymbol{\Gamma}(\mathbf{u}_{1i} + \delta_1 \mathbf{v}_{1i}), i = 1, \dots, n_1, \\ \mathbf{y}_{2i} &= \boldsymbol{\mu} + \boldsymbol{\Gamma}(\mathbf{u}_{2i} + \delta_2 \mathbf{v}_{2i}), i = 1, \dots, n_2, \\ \mathbf{y}_{3i} &= \boldsymbol{\mu} + \boldsymbol{\Gamma} \mathbf{u}_{3i}, i = 1, \dots, n_3,\end{aligned}$$

where $\mathbf{u}_{\alpha i}, i = 1, \dots, n_\alpha; \alpha = 1, 2, 3 \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, while $\mathbf{v}_{\alpha i} = (v_{\alpha i1}, \dots, v_{\alpha ip})^\top, i = 1, \dots, n_\alpha; \alpha = 1, 2$ are generated using the following three models:

Model 1. $v_{\alpha ir}, i = 1, \dots, n_\alpha; \alpha = 1, 2; r = 1, \dots, p \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

Model 2. $v_{\alpha,ir} = u_{\alpha,ir}/\sqrt{2}, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p$ with $u_{\alpha,ir}, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p \stackrel{\text{i.i.d.}}{\sim} t_4$.

Model 3. $v_{\alpha,ir} = (u_{\alpha,ir} - 1)/\sqrt{2}, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p$ with $u_{\alpha,ir}, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p \stackrel{\text{i.i.d.}}{\sim} \chi_1^2$.

It is important to note that δ_1 and δ_2 play pivotal roles as tuning parameters that govern the similarity of the distributions among the three generated samples. Specifically, when both δ_1 and δ_2 are set to zero ($\delta_1 = \delta_2 = 0$), the three samples generated from the three models exhibit identical distributions. If at least one of δ_i , where i can be 1 or 2, is non-zero, the samples still share the same mean vector but differ in their covariance matrices. Furthermore, it is worth mentioning that the test's power increases as both δ_1 and δ_2 increase.

We set μ as $h(1, \dots, p)^{\top} / \sqrt{\sum_{i=1}^p i^2}$ and Γ as $a[(1 - \rho)I_p + \rho J_p]$, where J_p denotes a matrix of ones with dimensions $p \times p$. To assess the performance of the considered tests across a range of dimensionality settings, we examine three cases: $p = 10$, $p = 100$, and $p = 500$. For each of these cases, we consider three sets of sample sizes (n_1, n_2, n_3) : $(20, 30, 40)$, $(80, 120, 160)$, and $(160, 240, 320)$. Additionally, we investigate three levels of correlation: $\rho = 0.1$, $\rho = 0.5$, and $\rho = 0.9$. These three values of ρ correspond to samples with varying degrees of correlation, ranging from nearly uncorrelated to moderately correlated and highly correlated. Notably, correlation increases as ρ grows. For simplicity, we set $h = 2$ and $a = 1.5$ across all three models.

In the case of the parametric bootstrap and random permutation methods, as well as the energy test, we use a total of $N = 1000$ replicates for computing the p -values at each simulation run, as described in Sections 3.1 and 3.2. It is worth noting that the W-S χ^2 -approximation method, which does not require generating replicates, is the least time-consuming among the methods considered. The empirical sizes and powers are computed based on 1000 simulation runs.

Table 1 provides an overview of the empirical sizes of T_{PB} , T_{RP} , T_{WS} , and T_{SR} , with the last row displaying the associated ARE values for the three different ρ values. Several observations can be made based on this table: Firstly, for nearly uncorrelated samples ($\rho = 0.1$), T_{WS} exhibits a slight tendency to be liberal, with an ARE value of 16.22 that is marginally higher than the ARE values of the other three tests. Secondly, when the generated samples are moderately correlated ($\rho = 0.5$) or highly correlated ($\rho = 0.9$), all four tests demonstrate fairly similar empirical sizes and ARE values, making them comparable in terms of size control. Finally, it is seen that the influence of sample sizes on size control is relatively minor, even though, in theory, a larger total sample size should result in better size control.

Figure 1 displays the empirical powers of all four tests in scenarios where all three generated samples have the same mean vectors, but they differ from each other in covariance matrices. Several conclusions can be drawn regarding these power values: Firstly, for $\rho = 0.1, 0.5$, and 0.9 , T_{PB} , T_{RP} , and T_{WS} exhibit similar empirical powers. This suggests that these three tests perform comparably, regardless of whether the data are nearly uncorrelated, moderately correlated, or highly correlated. Secondly, it is seen that under similar settings, as expected, the empirical powers of the tests generally increase with larger sample sizes. Finally, the empirical powers of T_{SR} consistently rank the lowest among all four tests. This indicates that T_{SR} is less powerful compared to the other three tests in these scenarios.

Table 1. Empirical sizes (in %) of Simulation 1.

			$\rho = 0.1$				$\rho = 0.5$				$\rho = 0.9$			
Model	p	(n_1, n_2, n_3)	T_{PB}	T_{RP}	T_{WS}	T_{SR}	T_{PB}	T_{RP}	T_{WS}	T_{SR}	T_{PB}	T_{RP}	T_{WS}	T_{SR}
1	10	(20, 30, 40)	4.0	4.5	5.2	4.9	4.4	4.0	4.7	4.1	5.7	5.5	6.1	5.9
		(80, 120, 160)	4.9	5.0	5.6	5.3	5.8	5.7	6.0	5.2	4.2	4.1	4.3	4.8
		(160, 240, 320)	5.1	5.1	5.2	4.9	4.0	4.2	4.4	3.9	4.6	5.4	4.9	4.3
	100	(20, 30, 40)	4.8	5.2	6.5	4.8	4.2	4.6	4.3	4.7	5.9	5.6	6.1	5.9
		(80, 120, 160)	4.6	4.4	4.9	3.9	4.8	5.1	5.3	5.3	5.1	5.3	5.3	5.5
		(160, 240, 320)	5.8	5.6	6.9	5.5	4.7	4.5	4.5	4.2	4.6	4.3	5.0	4.5
	500	(20, 30, 40)	5.0	5.4	5.8	4.9	4.3	4.4	4.7	4.8	3.8	4.4	4.5	4.5
		(80, 120, 160)	4.9	5.3	5.8	4.3	5.0	5.2	5.3	5.4	5.0	5.3	5.5	5.3
		(160, 240, 320)	6.2	6.2	7.0	5.7	6.4	6.2	6.4	6.4	4.8	4.4	5.0	5.2
2	10	(20, 30, 40)	4.9	5.5	6.6	5.7	4.4	4.5	5.0	4.9	5.6	5.0	5.7	5.6
		(80, 120, 160)	5.6	5.6	6.4	5.8	3.6	3.5	4.0	4.5	5.4	5.3	5.3	4.9
		(160, 240, 320)	5.7	5.8	6.8	6.3	4.9	5.5	5.8	5.2	5.3	5.5	5.7	5.1
	100	(20, 30, 40)	4.9	5.1	5.9	5.0	5.4	5.5	5.5	5.3	3.9	4.5	4.7	4.3
		(80, 120, 160)	5.5	5.6	6.6	5.5	4.5	4.3	4.9	3.8	5.8	5.6	5.7	5.6
		(160, 240, 320)	5.1	5.3	5.6	4.9	6.4	6.5	6.8	7.0	5.0	4.8	4.8	4.1
	500	(20, 30, 40)	5.0	5.1	5.4	5.5	5.9	5.8	6.0	6.6	5.0	5.2	5.3	5.5
		(80, 120, 160)	4.8	4.3	5.4	5.2	5.7	5.6	5.7	5.1	4.1	3.9	4.0	3.9
		(160, 240, 320)	4.2	4.3	5.0	4.9	4.6	4.4	4.8	4.4	5.4	5.7	5.4	5.9
3	10	(20, 30, 40)	4.2	5.0	5.6	4.8	5.4	6.0	6.5	5.7	4.9	4.9	5.5	4.8
		(80, 120, 160)	4.9	5.0	5.5	4.9	4.9	4.4	5.0	5.0	5.2	5.1	5.4	5.4
		(160, 240, 320)	4.5	4.6	5.5	4.9	5.0	5.0	5.4	5.0	6.0	5.9	5.8	5.8
	100	(20, 30, 40)	4.1	4.4	5.1	4.7	5.0	5.5	5.8	4.5	5.0	5.0	5.4	4.4
		(80, 120, 160)	3.1	3.5	4.0	4.5	4.8	5.2	5.0	4.4	5.0	5.0	5.0	5.0
		(160, 240, 320)	4.7	4.5	5.4	5.0	4.8	4.8	5.1	4.5	4.6	4.6	4.8	3.9
	500	(20, 30, 40)	4.9	4.7	5.4	6.0	5.0	4.6	5.4	4.5	4.1	4.1	4.5	3.9
		(80, 120, 160)	5.6	5.8	6.1	5.5	4.3	4.2	4.5	4.7	4.8	5.0	5.1	5.2
		(160, 240, 320)	5.0	4.8	5.5	6.0	4.4	4.5	4.6	4.5	5.5	5.4	5.6	6.0
ARE			9.04	9.33	16.22	8.67	10.67	12.52	11.56	11.70	9.26	8.74	9.19	11.56

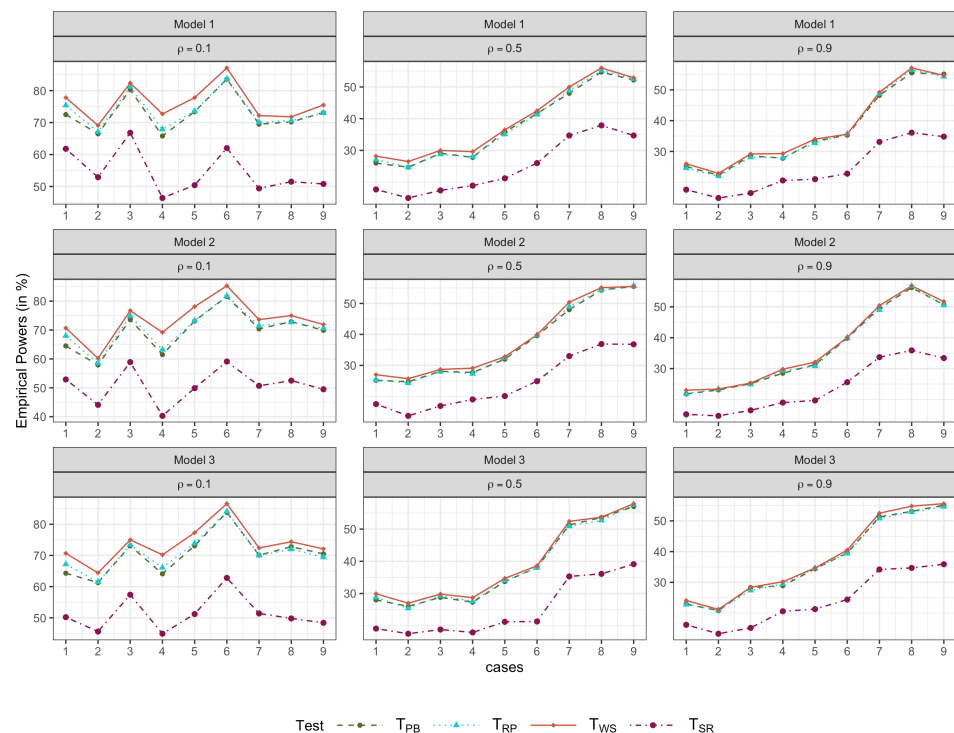


Figure 1. Simulation 1. The empirical powers (in %) of T_{PB} , T_{RP} , T_{WS} , and T_{SR} under different cases of $(p, n_1, n_2, n_3, \delta_1, \delta_2)$: 1. (10, 20, 30, 40, 1.2, 0.6), 2. (10, 80, 120, 160, 0.75, 0.375), 3. (10, 160, 240, 320, 0.65, 0.325), 4. (100, 20, 30, 40, 1.3, 0.65), 5. (100, 80, 120, 160, 0.85, 0.425), 6. (100, 160, 240, 320, 0.72, 0.36), 7. (500, 20, 30, 40, 1.65, 0.825), 8. (500, 80, 120, 160, 1, 0.5), 9. (500, 160, 240, 320, 0.8, 0.4).

4.2. Simulation 2

Certainly, it should be acknowledged that the MMD-based tests T_{PB} , T_{RP} , and T_{WS} may not consistently demonstrate superior performance when compared to the energy-distance-based test T_{SR} as in Simulation 1. To illustrate this point, in the context of this simulation study, we keep the same experimental framework as described in Simulation 1. However, we now introduce a new collection of three models, which are defined as follows:

Model 4. $v_{\alpha,ir}, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0.5, 1)$.

Model 5. $v_{\alpha,ir} = u_{\alpha,ir}/\sqrt{2} + 0.5, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p$ with $u_{\alpha,ir}, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p \stackrel{\text{i.i.d.}}{\sim} t_4$.

Model 6. $v_{\alpha,ir} = (u_{\alpha,ir} - 1)/\sqrt{2} + 0.5, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p$ with $u_{\alpha,ir}, i = 1, \dots, n_{\alpha}; \alpha = 1, 2; r = 1, \dots, p \stackrel{\text{i.i.d.}}{\sim} \chi_1^2$.

Please be aware that $v_{\alpha,ir}$, where i ranges from 1 to n_{α} , α takes values 1 and 2, and r ranges from 1 to p , is adjusted to ensure that $E(v_{\alpha,ir}) = 0.5$ and $\text{Var}(v_{\alpha,ir}) = 1$ across all three models. When both δ_1 and δ_2 are set to 0, the three generated samples follow the same distributions as in Simulation 1. Consequently, this implies that the empirical sizes of all four tests in Simulation 2 will be similar to those observed in Simulation 1. However, when at least one of δ_i (where i takes values 1 and 2) is non-zero, the three samples exhibit distinct mean vectors and covariance matrices, differing from those observed in Simulation 1. As a result, our focus should be on calculating the empirical powers of all four tests based on Models 4–6.

Figure 2 presents the empirical powers of all four tests based on Models 4–6, offering several noteworthy insights. First, it is evident that T_{PB} , T_{RP} , and T_{WS} demonstrate similar empirical powers. This implies that these three tests exhibit comparable performance regardless of whether the generated data follow a normal or non-normal distribution. Second, it is also observed that under similar settings, the empirical powers of the tests generally increase with larger sample sizes. Lastly, when ρ takes on values of 0.5 and 0.9, the empirical powers of T_{SR} surpass those of the other three tests. However, when $\rho = 0.1$, the empirical powers of all four tests are generally comparable. This indicates that, in the scenarios under consideration, T_{SR} demonstrates greater effectiveness compared to the other three tests when the correlation coefficient ρ is relatively high.

From these two simulation studies, it can be inferred that the proposed MMD-based tests T_{PB} , T_{RP} , and T_{WS} may outperform the energy-distance-based test T_{SR} when the differences in distributions are primarily in covariance matrices, while the reverse could be true when the differences in distribution involve both mean vectors and covariance matrices. Notably, the MMD-based test T_{WS} generally requires less computational effort compared to the bootstrap or permutation-based tests T_{PB} , T_{RP} , and T_{SR} .

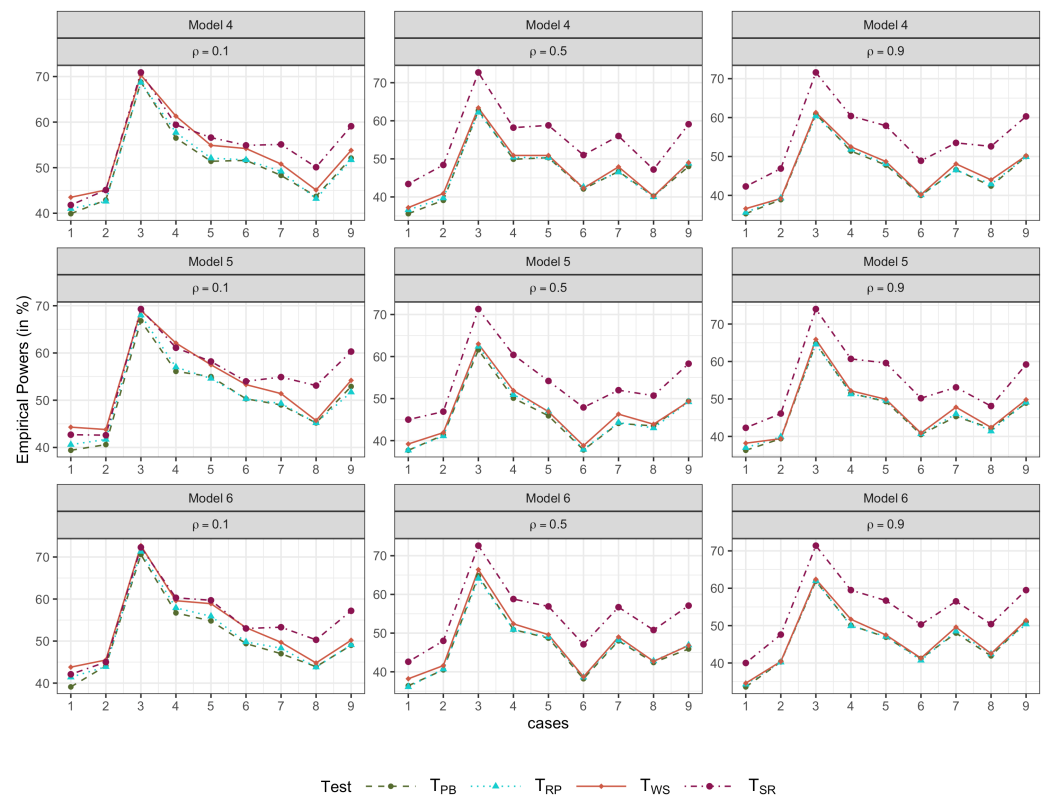


Figure 2. Simulation 2. The empirical powers (in %) of T_{PB} , T_{RP} , T_{WS} , and T_{SR} under different cases of $(p, n_1, n_2, n_3, \delta_1, \delta_2)$: 1. (10, 20, 30, 40, 0.37, 0.185), 2. (10, 80, 120, 160, 0.195, 0.097), 3. (10, 160, 240, 320, 0.18, 0.09), 4. (100, 20, 30, 40, 0.142, 0.071), 5. (100, 80, 120, 160, 0.068, 0.034), 6. (100, 160, 240, 320, 0.044, 0.022), 7. (500, 20, 30, 40, 0.06, 0.03), 8. (500, 80, 120, 160, 0.028, 0.014), 9. (500, 160, 240, 320, 0.022, 0.011).

5. Application to the Corneal Surface Data

The corneal surface data are briefly mentioned in Section 1. They were acquired during a keratoconus study, a collaborative project involving Ms. Nancy Tripoli and Dr. Kenneth L. Cohen from the Department of Ophthalmology at the University of North Carolina, Chapel Hill. This dataset comprises 150 observations with each corneal surface having more than 6000 height measurements. It can be categorized into four distinct groups: a group of 43 healthy corneas (referred to as the normal cornea group), a group of 14 corneas with unilateral suspect characteristics, a group of 21 corneas with suspect map features, and a group of 72 corneas clinically diagnosed with keratoconus. It is important to note that the corneal surfaces within the normal, unilateral suspect, and suspect map groups exhibit similar shapes, but they significantly differ from the corneal surfaces observed in the clinical keratoconus group (refer to Figure 1 in [24] for visualization). In the process of reconstructing a corneal surface, ref. [24] utilized the Zernike regression model to fit the height measurements associated with the corneal surface. The height of the corneal surface at a specific radius r and angle θ is denoted as $h(r, \theta)$, while $\hat{h}(r, \theta)$ represents the height estimated through the fitted model within the predefined region of interest. This region of interest spans from $r = 0$ to $r = r_0$ and from $\theta = 0$ to $\theta = 2\pi$, with r_0 being a predetermined positive constant. To naturally represent each corneal surface, a feature vector is constructed, consisting of values $\hat{h}(r_i, \theta_j)$, where i ranges from 1 to K and j ranges from 1 to L . These values are obtained by evaluating the fitted corneal surface $\hat{h}(r, \theta)$ at a grid of points defined as $r_i = r_0 i / K$ and $\theta_j = 2\pi j / L$ for $i = 1, \dots, K$ and $j = 1, \dots, L$. For simplicity, we choose to set $K = 20$ and $L = 100$, resulting in a feature vector with dimensions of 2000 for each corneal surface.

For simplicity, we put the fitted feature vectors for the complete corneal surface dataset collectively into a feature matrix with dimensions of 150×2000 . In this matrix, each row corresponds to a feature vector representing a corneal surface. Specifically, the initial 43 rows of the feature matrix correspond to observations from the normal group, sequentially followed by 14 rows from the unilateral suspect group, 21 rows from the suspect map group, and lastly, 72 rows from the clinical keratoconus group.

Our objective is to examine whether there are significant differences in the distributions among various corneal surface groups, referred to as multi-sample problems for the equality of distributions for high-dimensional data, given that the high-dimensional feature vectors represent the observations of the corneal surface data. In this application, we employ T_{PB} , T_{RP} , T_{WS} , and T_{SR} to address these problems. For both the parametric bootstrap and random permutation methods, as well as the energy test, we perform a total of $N = 10,000$ replicates to compute the associated p -values. For simplicity, we denote the normal, unilateral suspect, suspect map, and clinical keratoconus groups as NOR, UNI, SUS, and CLI, respectively.

Table 2 displays the results obtained from the application of four statistical tests, namely T_{PB} , T_{RP} , T_{WS} , and T_{SR} , to assess the equality of distributions among different corneal surface groups. A careful examination of these results yields several noteworthy insights. To begin with, when considering the comparison among the corneal surface groups labeled “NOR vs. UNI vs. SUS vs. CLI”, it is evident that all four tests reject the null hypothesis. This rejection signifies that there exists at least one significant difference among the distributions of these four corneal surface groups. Consequently, further investigation is warranted to identify which specific group or groups differ from the others. Secondly, it is important to highlight the outcome for the comparison involving “NOR vs. UNI vs. SUS”. In this case, none of the four tests reject the null hypothesis at the 5% significance level. This outcome indicates that the normal, unilateral suspect, and suspect map groups share a similar distribution pattern. Thirdly, across the remaining three comparisons, all four tests consistently reject the null hypothesis. These results align with the observations depicted in Figure 1 of [24], which illustrates the distinctiveness of corneal surfaces within the clinical keratoconus group when compared to the other three groups. Fourthly, when focusing on the comparison “NOR vs. UNI vs. SUS”, it is worth noting that the p -values obtained from T_{PB} , T_{RP} , and T_{WS} are quite similar, suggesting their comparable performance. This consistency in p -values is also reflected in the empirical sizes presented in Table 1. Lastly, when analyzing the cases involving CLI, it becomes evident that the p -values generated by T_{SR} consistently exhibit larger values compared to those produced by the other three tests. This discrepancy implies that T_{SR} may have a lower sensitivity in detecting distribution differences when compared to the other tests, indicating the potentially reduced statistical power in this real data example.

Table 2. p -values (in %) for testing the distribution equality of corneal surface groups.

Case	Method			
	T_{PB}	T_{RP}	T_{WS}	T_{SR}
NOR vs. UNI vs. SUS vs. CLI	0	0	0.00004	0.09
NOR vs. UNI vs. SUS	31.2	32.3	33.3	38.9
NOR vs. UNI vs. CLI	0.02	0	0.0012	0.25
NOR vs. SUS vs. CLI	0	0	0.00001	0.01
UNI vs. SUS vs. CLI	0.01	0	0.0012	0.24

Notice that the test T_{PB} is bootstrap-based and the tests T_{RP} and T_{SR} are permutation-based. This means that their p -values are obtained via bootstrapping or permutating numerous random samples to compute the associated p -values as described in Sections 3.1 and 3.2. Thus, the p -values of these tests are random, i.e., they are different at different instances. However, the p -value of T_{WS} remains fixed. In order to investigate this clearly, we performed 500 iterations of T_{PB} , T_{RP} , T_{WS} , and T_{SR} on the case “NOR vs. UNI vs. SUS”.

The boxplots of the corresponding p -values of the four tests are presented in Figure 3. It is evident that the p -values obtained from T_{WS} remain fixed, whereas those derived from T_{PB} , T_{RP} , and T_{SR} exhibit variability. This contrast underscores the fact that p -values resulting from bootstrap-based or permutation-based tests indeed differ across various instances.

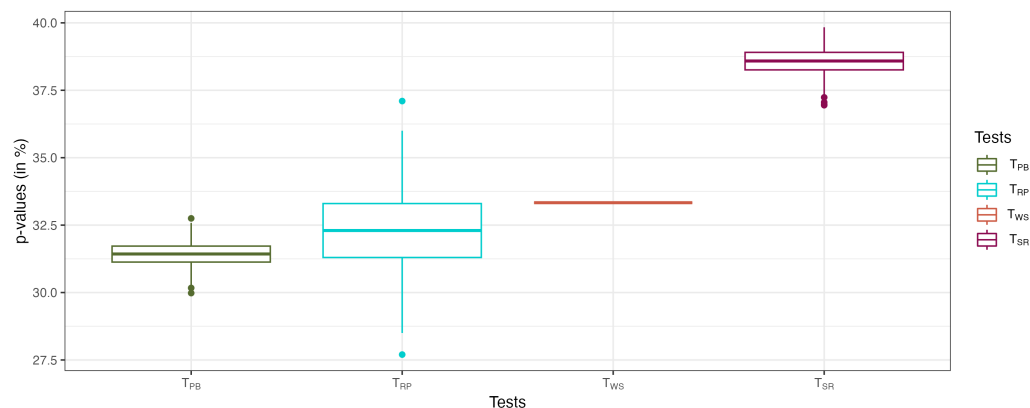


Figure 3. Boxplots of p -values of T_{PB} , T_{RP} , T_{WS} , and T_{SR} when applied to the case “NOR vs. UNI vs. SUS” of the corneal surface data 500 times.

6. Concluding Remarks

Testing whether multiple high-dimensional samples adhere to the same distribution is a common area of research interest. This paper introduces and investigates a novel MMD-based test to address this question. The null distribution of this test is approximated using three methods: parametric bootstrap, random permutation, and the W-S χ^2 -approximation approach. Results from two simulation studies and a real data application demonstrate that the proposed test exhibits effective size control and superior statistical power compared to the energy test introduced by [12] when the differences among sample distributions are primarily related to covariance matrices rather than mean vectors. Thus, the proposed test is generally well suited for conducting multi-sample equal distribution testing on high-dimensional data. We particularly recommend its use in scenarios where distribution differences are associated with covariance matrices. Conversely, when distribution differences predominantly pertain to means, the energy test is a more powerful choice. However, in practice, determining whether distribution differences are related to means or covariance matrices can be challenging. Therefore, we suggest considering both the new test and the energy test as viable options. Nevertheless, it is important to note that implementing the proposed test comes with certain challenges. Both the parametric bootstrap and random permutation methods can be computationally intensive, leading to variable p -values across different applications. In contrast, the W-S χ^2 -approximation method offers computational efficiency and produces fixed p -values. However, its accuracy is limited as it solely relies on matching two cumulants of the test statistic under the null hypothesis.

An intriguing question arises naturally: Can we enhance the accuracy of the proposed test by matching three cumulants of the test statistic? Recent work by [18] suggests that this is indeed possible. However, deriving the third cumulant of the test statistic presents a current challenge and requires further investigation. Another aspect to consider is the choice of kernel width. While the paper opts for simplicity by utilizing the median distance between observed vectors in the pooled sample, it is worth exploring the kernel width choice recommended by [18] to potentially enhance the test’s statistical power. These avenues for future research promise exciting developments and warrant further exploration.

Author Contributions: Conceptualization, J.-T.Z. and A.A.C.; methodology, J.-T.Z., T.Z. and Z.P.O.; software, Z.P.O. and T.Z.; validation, J.-T.Z. and T.Z.; formal analysis, Z.P.O.; investigation, J.-T.Z. and Z.P.O.; resources, Z.P.O.; data curation, Z.P.O.; writing—original draft preparation, J.-T.Z. and Z.P.O.; writing—review and editing, J.-T.Z. and T.Z.; visualization, A.A.C. and T.Z.; supervision,

J.-T.Z.; project administration, J.-T.Z.; funding acquisition, J.-T.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Zhang's research was partially funded by the National University of Singapore academic research grant (22-5699-A0001).

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://tandf.figshare.com/articles/dataset/Linear_hypothesis_testing_with_functional_data/6063026/1?file=10914914 (accessed on 20 October 2023).

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

Abbreviations

The following abbreviations are used in this manuscript:

MMD	Maximum Mean Discrepancy
EDF	Empirical Distribution Function
RKHS	Reproducing Kernel Hilbert Spaces
RBF	Radial Basis Function
ARE	Average Relative Error

Appendix A. Technical Proofs

Proof of (6). We have

$$\begin{aligned}
 T_n &= \sum_{\alpha=1}^k n_{\alpha} \|\bar{\mathbf{x}}_{\alpha} - \bar{\mathbf{x}}\|^2 \\
 &= \sum_{\alpha=1}^k n_{\alpha} (\|\bar{\mathbf{x}}_{\alpha}\|^2 - 2\langle \bar{\mathbf{x}}_{\alpha}, \bar{\mathbf{x}} \rangle + \|\bar{\mathbf{x}}\|^2) \\
 &= \sum_{\alpha=1}^k n_{\alpha} \|\bar{\mathbf{x}}_{\alpha}\|^2 - n \|\bar{\mathbf{x}}\|^2 \\
 &= \sum_{\alpha=1}^k n_{\alpha} \|\bar{\mathbf{x}}_{\alpha}\|^2 - \sum_{\alpha=1}^k \sum_{\beta=1}^k \frac{n_{\alpha} n_{\beta}}{n} \langle \bar{\mathbf{x}}_{\alpha}, \bar{\mathbf{x}}_{\beta} \rangle \\
 &= \sum_{\alpha=1}^k \frac{n_{\alpha}(n-n_{\alpha})}{n} \|\bar{\mathbf{x}}_{\alpha}\|^2 - 2 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_{\alpha} n_{\beta}}{n} \langle \bar{\mathbf{x}}_{\alpha}, \bar{\mathbf{x}}_{\beta} \rangle \\
 &= \sum_{\alpha \neq \beta}^k \frac{n_{\alpha} n_{\beta}}{n} \|\bar{\mathbf{x}}_{\alpha}\|^2 - 2 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_{\alpha} n_{\beta}}{n} \langle \bar{\mathbf{x}}_{\alpha}, \bar{\mathbf{x}}_{\beta} \rangle \\
 &= \sum_{1 \leq \alpha < \beta \leq k} \frac{n_{\alpha} n_{\beta}}{n} (\|\bar{\mathbf{x}}_{\alpha}\|^2 + \|\bar{\mathbf{x}}_{\beta}\|^2 - 2\langle \bar{\mathbf{x}}_{\alpha}, \bar{\mathbf{x}}_{\beta} \rangle) \\
 &= \sum_{1 \leq \alpha < \beta \leq k} \frac{n_{\alpha} n_{\beta}}{n} \|\bar{\mathbf{x}}_{\alpha} - \bar{\mathbf{x}}_{\beta}\|^2.
 \end{aligned}$$

□

Proof of Theorem 1. Under Condition C1, we have $F_1 = \dots = F_k = F$. Let $\mathbf{y}, \mathbf{y}' \stackrel{\text{i.i.d.}}{\sim} F$. Under Condition C3, we have Mercer's expansion (19). Through (14) and (20), we have

$$\lambda_r \mathbb{E}_{\mathbf{y}}[\psi_r(\mathbf{y})] = \int_{\mathcal{X}} \mathbb{E}_{\mathbf{y}} \tilde{K}(\mathbf{y}, \mathbf{y}') \psi_r(\mathbf{y}') dF(\mathbf{y}') = 0.$$

This, together with (20), implies that

$$\mathbb{E}[\psi_r(\mathbf{y})] = 0 \text{ whenever } \lambda_r \neq 0 \text{ and } \text{Var}[\psi_r(\mathbf{y})] = \int_{\mathcal{X}} \psi_r^2(\mathbf{y}) dF(\mathbf{y}) = 1. \quad (\text{A1})$$

Set $z_{r,\alpha i} = \psi_r(\mathbf{y}_{\alpha i})$, $i = 1, \dots, n_{\alpha}$; $\alpha = 1, \dots, k$. Under Condition C1, we have $\mathbf{y}_{\alpha i} \stackrel{\text{i.i.d.}}{\sim} F$.

It follows that for a fixed $r = 1, 2, \dots$, $z_{r,\alpha i}$, $i = 1, \dots, n_{\alpha}$; $\alpha = 1, \dots, k$ are i.i.d. with mean 0 and variance 1. For different r , $z_{r,\alpha i}$, $i = 1, \dots, n_{\alpha}$; $\alpha = 1, \dots, k$ are uncorrelated. Then, through (18) and (19), we have

$$\begin{aligned}
 \tilde{V}_{\alpha\alpha} &= \frac{1}{n_{\alpha}^2} \sum_{i=1}^{n_{\alpha}} \sum_{j=1}^{n_{\alpha}} \left(\sum_{r=1}^{\infty} \lambda_r z_{r,\alpha i} z_{r,\alpha j} \right) = \sum_{r=1}^{\infty} \lambda_r \left(\frac{1}{n_{\alpha}^2} \sum_{i=1}^{n_{\alpha}} \sum_{j=1}^{n_{\alpha}} z_{r,\alpha i} z_{r,\alpha j} \right) \\
 &= \sum_{r=1}^{\infty} \lambda_r \bar{z}_{r,\alpha}^2, \text{ and} \\
 \tilde{V}_{\alpha\beta} &= \frac{1}{n_{\alpha} n_{\beta}} \sum_{i=1}^{n_{\alpha}} \sum_{j=1}^{n_{\beta}} \left(\sum_{r=1}^{\infty} \lambda_r z_{r,\alpha i} z_{r,\beta j} \right) = \sum_{r=1}^{\infty} \lambda_r \bar{z}_{r,\alpha} \bar{z}_{r,\beta},
 \end{aligned}$$

where $\bar{z}_{r,\alpha} = n_\alpha^{-1} \sum_{i=1}^{n_\alpha} z_{r,\alpha i}$, $\alpha = 1, \dots, k$; $r = 1, 2, \dots$. Through (17), we have

$$\begin{aligned}\tilde{T}_n &= \sum_{r=1}^{\infty} \lambda_r \left(\sum_{\alpha=1}^k \frac{n_\alpha(n-n_\alpha)}{n} \bar{z}_{r,\alpha}^2 - 2 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_\alpha n_\beta}{n} \bar{z}_{r,\alpha} \bar{z}_{r,\beta} \right) \\ &= \sum_{r=1}^{\infty} \lambda_r \left[\sum_{\alpha=1}^k n_\alpha (\bar{z}_{r,\alpha} - \bar{z}_r)^2 \right] \\ &= \sum_{r=1}^{\infty} \lambda_r A_{n,r},\end{aligned}$$

where $A_{n,r} = \sum_{\alpha=1}^k n_\alpha (\bar{z}_{r,\alpha} - \bar{z}_r)^2$, $r = 1, 2, \dots$ and $\bar{z}_r = n^{-1} \sum_{\alpha=1}^k n_\alpha \bar{z}_{r,\alpha}$.

Let $\varphi_X(t) = E(e^{itX})$ denote the characteristic function of a random variable X . Set $\tilde{T}_n^{(q)} = \sum_{r=1}^q \lambda_r A_{n,r}$. Then, we have $|\varphi_{\tilde{T}_n}(t) - \varphi_{\tilde{T}_n^{(q)}}(t)| \leq |t| [E(\tilde{T}_n - \tilde{T}_n^{(q)})^2]^{1/2}$. For any given $r = 1, 2, \dots$, through the central limit theorem, under Conditions C2 and C3, as $n \rightarrow \infty$, we have $\sqrt{n_\alpha} \bar{z}_{r,\alpha} \xrightarrow{\mathcal{L}} N(0, 1)$, $\alpha = 1, \dots, k$. Set $w_{n,r} = [\sqrt{n_1} \bar{z}_{r,1}, \dots, \sqrt{n_k} \bar{z}_{r,k}]^\top$. Then, as $n \rightarrow \infty$, we have $w_{n,r} \xrightarrow{\mathcal{L}} w_r \sim N_k(0, I_k)$ and w_r , $r = 1, 2, \dots$ are independently identically distributed. Set $\delta_{n,k} = [\sqrt{n_1/n}, \dots, \sqrt{n_k/n}]^\top$. Under Condition C2, we have $\delta_{n,k} \rightarrow \delta_k = [\sqrt{\tau_1}, \dots, \sqrt{\tau_k}]^\top$. Thus, $G_n = I_k - \delta_{n,k} \delta_{n,k}^\top \rightarrow G = I_k - \delta_k \delta_k^\top$. Both G_n and G are idempotent matrices with rank $k - 1$. It follows that

$$A_{n,r} = w_{n,r}^\top G_n w_{n,r} \xrightarrow{\mathcal{L}} A_r = w_r^\top G w_r \sim \chi_{k-1}^2, \quad (\text{A2})$$

which is a chi-squared distribution with $k - 1$ degrees of freedom and A_r , $r = 1, 2, \dots$ are independent. It follows that as $n \rightarrow \infty$, we have $\text{Var}(A_{n,r}) = 2(k - 1)[1 + o(1)]$ and $E(A_{n,r}) = (k - 1)[1 + o(1)]$. Therefore, as $n \rightarrow \infty$, we have

$$\begin{aligned}E(\tilde{T}_n - \tilde{T}_n^{(q)})^2 &= E\left(\sum_{r=q+1}^{\infty} \lambda_r A_{n,r}\right)^2 \\ &= \text{Var}\left(\sum_{r=q+1}^{\infty} \lambda_r A_{n,r}\right) + E^2\left(\sum_{r=q+1}^{\infty} \lambda_r A_{n,r}\right) \\ &\leq \left[\sum_{r=q+1}^{\infty} \sqrt{\text{Var}(\lambda_r A_{n,r})}\right]^2 + \left[\sum_{r=q+1}^{\infty} E(\lambda_r A_{n,r})\right]^2 \\ &= (k^2 - 1) \left(\sum_{r=q+1}^{\infty} \lambda_r\right)^2 [1 + o(1)].\end{aligned}$$

It follows that

$$|\varphi_{\tilde{T}_n}(t) - \varphi_{\tilde{T}_n^{(q)}}(t)| \leq |t|(k^2 - 1)^{1/2} \left(\sum_{r=q+1}^{\infty} \lambda_r\right) [1 + o(1)]. \quad (\text{A3})$$

Let t be fixed. Under Condition C3 and (21), as $q \rightarrow \infty$, we have $\sum_{r=q+1}^{\infty} \lambda_r \rightarrow 0$. Thus, for any given $\epsilon > 0$, there exist N_1 and Q_1 , depending on $|t|$ and ϵ , such that as $n > N_1$ and $q > Q_1$, we have

$$|\varphi_{\tilde{T}_n}(t) - \varphi_{\tilde{T}_n^{(q)}}(t)| \leq \epsilon. \quad (\text{A4})$$

For any fixed $q > Q_1$, through (A2), as $n \rightarrow \infty$, we have $\tilde{T}_n^{(q)} \xrightarrow{\mathcal{L}} \tilde{T}^{(q)} \stackrel{d}{=} \sum_{r=1}^q \lambda_r A_r$, $A_r \stackrel{\text{i.i.d.}}{\sim} \chi_{k-1}^2$. Thus, there exists N_2 , depending on q and ϵ , such that as $n > N_2$, we have

$$|\varphi_{\tilde{T}_n^{(q)}}(t) - \varphi_{\tilde{T}^{(q)}}(t)| \leq \epsilon. \quad (\text{A5})$$

Recall that $\tilde{T} = \sum_{r=1}^{\infty} \lambda_r A_r$, $A_r \stackrel{\text{i.i.d.}}{\sim} \chi_{k-1}^2$. Along the same lines as those for proving (A4), we can show that there exists Q_2 , depending on $|t|$ and ϵ , such that as $q > Q_2$, we have

$$|\varphi_{\tilde{T}^{(q)}}(t) - \varphi_{\tilde{T}}(t)| \leq \epsilon. \quad (\text{A6})$$

It follows from (A4)–(A6) that for any $n \geq \max(N_1, N_2)$ and $q \geq \max(Q_1, Q_2)$, we have

$$|\varphi_{\tilde{T}_n}(t) - \varphi_{\tilde{T}}(t)| \leq |\varphi_{\tilde{T}_n}(t) - \varphi_{\tilde{T}_n^{(q)}}(t)| + |\varphi_{\tilde{T}_n^{(q)}}(t) - \varphi_{\tilde{T}^{(q)}}(t)| + |\varphi_{\tilde{T}^{(q)}}(t) - \varphi_{\tilde{T}}(t)| \leq 3\epsilon.$$

The convergence in distribution of $\tilde{T}_n$ to $\tilde{T}$ follows as we can let $\epsilon \rightarrow 0$. $\square$

Proof of Theorem 2. Under Condition C1, let $\mathbf{y}, \mathbf{y}' \stackrel{\text{i.i.d.}}{\sim} F$. Then, through (23), we have $E(\tilde{V}_{\alpha\alpha}) = n_\alpha^{-1} E[\tilde{K}(\mathbf{y}, \mathbf{y})]$ and $E(\tilde{V}_{\alpha\beta}) = 0$. Thus, we have

$$E(\tilde{T}_n) = \sum_{\alpha=1}^k \frac{n_\alpha(n - n_\alpha)}{n} \left\{ n_\alpha^{-1} E[\tilde{K}(\mathbf{y}, \mathbf{y})] \right\} = (k-1) E[\tilde{K}(\mathbf{y}, \mathbf{y})].$$

Through (23) again, we have

$$\begin{aligned} \text{Var}(\tilde{T}_n) &= \sum_{\alpha=1}^k \left[\frac{n_\alpha(n - n_\alpha)}{n} \right]^2 \text{Var}(\tilde{V}_{\alpha\alpha}) + 4 \sum_{1 \leq \alpha < \beta \leq k} \left(\frac{n_\alpha n_\beta}{n} \right)^2 \text{Var}(\tilde{V}_{\alpha\beta}) \\ &= \sum_{\alpha=1}^k \frac{n_\alpha^2(n - n_\alpha)^2}{n^2} \left\{ \frac{1}{n_\alpha^3} \text{Var}[\tilde{K}(\mathbf{y}, \mathbf{y})] + \frac{2(n_\alpha - 1)}{n_\alpha^3} E[\tilde{K}^2(\mathbf{y}, \mathbf{y}')] \right\} \\ &+ 4 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_\alpha^2 n_\beta^2}{n^2} \left\{ \frac{1}{n_\alpha n_\beta} E[\tilde{K}^2(\mathbf{y}, \mathbf{y}')] \right\} \\ &= \sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \text{Var}[\tilde{K}(\mathbf{y}, \mathbf{y})] + 2 \left[\sum_{\alpha=1}^k \frac{(n - n_\alpha)^2(n_\alpha - 1)}{n^2 n_\alpha} + 2 \sum_{1 \leq \alpha < \beta \leq k} \frac{n_\alpha n_\beta}{n^2} \right] E[\tilde{K}^2(\mathbf{y}, \mathbf{y}')] \\ &= \sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \text{Var}[\tilde{K}(\mathbf{y}, \mathbf{y})] + 2 \left[(k-1) - \sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \right] E[\tilde{K}^2(\mathbf{y}, \mathbf{y}')] \end{aligned}$$

$\square$

Proof of Theorem 3. First, through (12), we have $|\tilde{K}(\mathbf{y}, \mathbf{y})|, |\tilde{K}(\mathbf{y}, \mathbf{y}')| \leq 4B_K$ for all $\mathbf{y}, \mathbf{y}' \in \mathcal{R}^p$. Thus, we have

$$\text{Var}[\tilde{K}(\mathbf{y}, \mathbf{y})] \leq E[\tilde{K}^2(\mathbf{y}, \mathbf{y})] \leq 16B_K^2, \text{ and } E[\tilde{K}^2(\mathbf{y}, \mathbf{y}')] \leq 16B_K^2.$$

Then, through Theorem 2, we have

$$\begin{aligned} \text{Var}(\tilde{T}_n) &= \left[\sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \right] \text{Var}[\tilde{K}(\mathbf{y}_{\alpha 1}, \mathbf{y}_{\alpha 1})] + 2 \left[(k-1) - \sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \right] E[\tilde{K}^2(\mathbf{y}_{\alpha 1}, \mathbf{y}_{\beta 1})] \\ &\leq 16B_K^2 \left[\sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \right] + 2 \cdot 16B_K^2 \left[(k-1) - \sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \right] \\ &= 16B_K^2 \left[2(k-1) - \sum_{\alpha=1}^k \frac{(n - n_\alpha)^2}{n^2 n_\alpha} \right] \\ &\leq 32B_K^2(k-1). \end{aligned}$$

Finally, since $\bar{x}_\alpha, \alpha = 1, \dots, k$ are independent, we have

$$\text{Var}(S_n) = n^{-(1-2\Delta)} \sum_{\alpha=1}^k n_\alpha^2 \cdot \frac{\sigma_\alpha^2}{n_\alpha} = n^{2\Delta} \sum_{\alpha=1}^k \frac{n_\alpha}{n} \sigma_\alpha^2.$$

Furthermore, through the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \sum_{\alpha=1}^k \frac{n_\alpha}{n} \sigma_\alpha^2 &= \sum_{\alpha=1}^k \frac{n_\alpha}{n} E[(\mathbf{x}_{\alpha 1} - \boldsymbol{\mu}_\alpha, \mathbf{h}_\alpha - \bar{\mathbf{h}})]^2 \\ &\leq \sum_{\alpha=1}^k \frac{n_\alpha}{n} E(\|\mathbf{x}_{\alpha 1} - \boldsymbol{\mu}_\alpha\|^2) \|\mathbf{h}_\alpha - \bar{\mathbf{h}}\|^2 \\ &\leq \sum_{\alpha=1}^k \frac{n_\alpha}{n} \|\mathbf{h}_\alpha - \bar{\mathbf{h}}\|^2 E[\tilde{K}(\mathbf{y}_{\alpha 1}, \mathbf{y}_{\alpha 1})] \\ &\leq 4B_K \sum_{\alpha=1}^k \frac{n_\alpha}{n} \|\mathbf{h}_\alpha - \bar{\mathbf{h}}\|^2 \\ &\leq 4B_K B_h < \infty. \end{aligned}$$

$\square$

Proof of Theorem 4. Under the given conditions, through Theorems 2 and 3, we have

$$E(\tilde{T}_n) = (k-1) E[\tilde{K}(\mathbf{y}_{\alpha 1}, \mathbf{y}_{\alpha 1})], \text{ and } \text{Var}(\tilde{T}_n) \leq 32B_K^2(k-1),$$

and as $n \rightarrow \infty$,

$$\text{Var}(S_n) = n^{2\Delta} \sum_{\alpha=1}^k \tau_\alpha \sigma_\alpha^2 [1 + o(1)], \quad (\text{A7})$$

where τ_α , $\alpha = 1, \dots, k$ are given in Condition C2, It follows that as $n \rightarrow \infty$, we have $E(\tilde{T}_n)/\sqrt{\text{Var}(S_n)} \rightarrow 0$ and $\text{Var}(\tilde{T}_n)/\text{Var}(S_n) \rightarrow 0$. Thus, through the Markov inequality, as $n \rightarrow \infty$, we have

$$\Pr\left(\left|\frac{\tilde{T}_n}{\sqrt{\text{Var}(S_n)}}\right| > \varepsilon\right) = \Pr\left[\left(\frac{\tilde{T}_n}{\sqrt{\text{Var}(S_n)}}\right)^2 > \varepsilon^2\right] \leq \frac{1}{\varepsilon^2} \cdot \frac{E(\tilde{T}_n^2)}{\text{Var}(S_n)} \rightarrow 0,$$

for all $\varepsilon > 0$. Therefore, we have $\tilde{T}_n/\sqrt{\text{Var}(S_n)} \xrightarrow{P} 0$ and (a) is proved. To prove (b), notice that through the central limit theorem, as $n_\alpha \rightarrow \infty$, we have

$$u_\alpha = \sqrt{n_\alpha} \langle \bar{x}_\alpha - \mu_\alpha, \mathbf{h}_\alpha - \bar{\mathbf{h}} \rangle / \sigma_\alpha \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \alpha = 1, \dots, k.$$

Since $\bar{x}_\alpha, \alpha = 1, \dots, k$ are independent and $S_n = n^\Delta \sum_{\alpha=1}^k \sqrt{n_\alpha/n} \sigma_\alpha u_\alpha$, we have $S_n/\sqrt{\text{Var}(S_n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$. To prove (c), notice that as $n \rightarrow \infty$, through (25), (a) and (b), we have

$$\frac{T_n - n^{2\Delta} \sum_{\alpha=1}^k \frac{n_\alpha}{n} \|\mathbf{h}_\alpha - \bar{\mathbf{h}}\|^2}{2\sqrt{\text{Var}(S_n)}} = \frac{\tilde{T}_n}{2\sqrt{\text{Var}(S_n)}} + \frac{S_n}{\sqrt{\text{Var}(S_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Thus, as $n \rightarrow \infty$, and through (A7), we have

$$\begin{aligned} \Pr(T_n \geq C_\varepsilon) &= \Pr\left[\frac{T_n - n^{2\Delta} \sum_{\alpha=1}^k \frac{n_\alpha}{n} \|\mathbf{h}_\alpha - \bar{\mathbf{h}}\|^2}{2\sqrt{\text{Var}(S_n)}} \geq \frac{C_\varepsilon - n^{2\Delta} \sum_{\alpha=1}^k \frac{n_\alpha}{n} \|\mathbf{h}_\alpha - \bar{\mathbf{h}}\|^2}{2\sqrt{\text{Var}(S_n)}}\right] \\ &= 1 - \Phi\left(\frac{C_\varepsilon}{2n^\Delta \sqrt{\sum_{\alpha=1}^k \frac{n_\alpha}{n} \sigma_\alpha^2}} - \frac{n^{2\Delta} \sum_{\alpha=1}^k \frac{n_\alpha}{n} \|\mathbf{h}_\alpha - \bar{\mathbf{h}}\|^2}{2n^\Delta \sqrt{\sum_{\alpha=1}^k \frac{n_\alpha}{n} \sigma_\alpha^2}}\right) [1 + o(1)] \\ &= 1 - \Phi\left(-\frac{n^\Delta \sum_{\alpha=1}^k \tau_\alpha \|\mathbf{h}_\alpha - \mathbf{h}^*\|^2}{2\sqrt{\sum_{\alpha=1}^k \tau_\alpha \sigma_\alpha^{*2}}}\right) [1 + o(1)] \\ &= \Phi\left(\frac{n^\Delta \sum_{\alpha=1}^k \tau_\alpha \|\mathbf{h}_\alpha - \mathbf{h}^*\|^2}{2\sqrt{\sum_{\alpha=1}^k \tau_\alpha \sigma_\alpha^{*2}}}\right) [1 + o(1)] \\ &\rightarrow 1, \end{aligned}$$

where $\mathbf{h}^* = \sum_{\alpha=1}^k \tau_\alpha \mathbf{h}_\alpha$ and $\sigma_\alpha^{*2} = E[\langle \mathbf{x}_{\alpha 1} - \mu_\alpha, \mathbf{h}_\alpha - \mathbf{h}^* \rangle]^2$. Thus, the theorem is proved. $\square$

References

- Lehmann, E.L. *Nonparametrics: Statistical Methods Based on Ranks*; Springer: New York, NY, USA, 2006.
- Friedman, J.H.; Rafsky, L.C. Multivariate Generalizations of the Wald–Wolfowitz and Smirnov Two-Sample Tests. *Ann. Stat.* **1979**, *7*, 697–717.
- Schilling, M.F. Multivariate Two-Sample Tests Based on Nearest Neighbors. *J. Am. Stat. Assoc.* **1986**, *81*, 799–806.
- Baringhaus, L.; Franz, C. On a new multivariate two-sample test. *J. Multivar. Anal.* **2004**, *88*, 190–206.
- Rosenbaum, P.R. An exact distribution-free test comparing two multivariate distributions based on adjacency. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 515–530.
- Biswas, M.; Mukhopadhyay, M.; Ghosh, A.K. A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika* **2014**, *101*, 913–926.
- Li, J. Asymptotic normality of interpoint distances for high-dimensional data with applications to the two-sample problem. *Biometrika* **2018**, *105*, 529–546.
- Chen, H.; Friedman, J.H. A New Graph-Based Two-Sample Test for Multivariate and Object Data. *J. Am. Stat. Assoc.* **2017**, *112*, 397–409.
- Hall, P.; Tajvidi, N. Permutation Tests for Equality of Distributions in High-Dimensional Settings. *Biometrika* **2002**, *89*, 359–374.
- Wei, S.; Lee, C.; Wichers, L.; Marron, J.S. Direction-Projection-Permutation for High-Dimensional Hypothesis Tests. *J. Comput. Graph. Stat.* **2016**, *25*, 549–569.
- Ghosh, A.K.; Biswas, M. Distribution-free high-dimensional two-sample tests based on discriminating hyperplanes. *Test* **2016**, *25*, 525–547.
- Székel, G.J.; Rizzo, M.L. Testing for equal distributions in high dimension. *InterStat* **2004**, *5*, 1249–1272.

13. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A Kernel Two-Sample Test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
14. Sejdinovic, D.; Sriperumbudur, B.; Gretton, A.; Fukumizu, K. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Stat.* **2013**, *41*, 2263–2291.
15. Zhang, J.T.; Smaga, Ł. Two-sample test for equal distributions in separable metric space: new maximum mean discrepancy based approaches. *Electron. J. Stat.* **2022**, *16*, 4090–4132.
16. Zhou, B.; Ong, Z.P.; Zhang, J.T. A new MMD-based two-sample test for equal distributions in separable metric spaces. *Manuscript* **2023**, *in press*.
17. Balogoun, A.S.K.; Nkiet, G.M.; Ogouyandjou, C. k -Sample problem based on generalized maximum mean discrepancy. *arXiv* **2018**, arXiv:1811.09103.
18. Zhang, J.T.; Guo, J.; Zhou, B. Testing equality of several distributions in separable metric spaces: A maximum mean discrepancy based approach. *J. Econom.* **2022**, *in press*. <https://doi.org/10.1016/j.jeconom.2022.03.007>.
19. Zhang, J.T.; Guo, J.; Zhou, B. Linear hypothesis testing in high-dimensional one-way MANOVA. *J. Multivar. Anal.* **2017**, *155*, 200–216.
20. Gretton, A.; Fukumizu, K.; Harchaoui, Z.; Sriperumbudur, B.K. A Fast, Consistent Kernel Two-Sample Test. In *Advances in Neural Information Processing Systems 22*; Curran Associates, Inc.: New York, NY, USA, 2009; pp. 673–681.
21. Welch, B.L. The generalization of ‘student’s’ problem when several different population variances are involved. *Biometrika* **1947**, *34*, 28–35.
22. Satterthwaite, F.E. An Approximate Distribution of Estimates of Variance Components. *Biom. Bull.* **1946**, *2*, 110–114.
23. Zhang, J.T. Two-Way MANOVA with Unequal Cell Sizes and Unequal Cell Covariance Matrices. *Technometrics* **2011**, *53*, 426–439.
24. Smaga, Ł.; Zhang, J.T. Linear Hypothesis Testing with Functional Data. *Technometrics* **2019**, *61*, 99–110.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.