Multi-sample comparison using spatial signs for infinite dimensional data

Joydeep Chowdhury[*][1]

[1]*King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*
e-mail: joydeepchowdhury01@gmail.com

Probal Chaudhuri[2]

[2]*Indian Statistical Institute, Kolkata, India*
e-mail: probal@isical.ac.in

Abstract: We consider an analysis of variance type problem, where the sample observations are random elements in an infinite dimensional space. This scenario covers the case, where the observations are random functions. For such a problem, we propose a test based on spatial signs. We develop an asymptotic implementation as well as a bootstrap implementation and a permutation implementation of this test and investigate their size and power properties. We compare the performance of our test with that of several mean based tests of analysis of variance for functional data studied in the literature. Interestingly, our test not only outperforms the mean based tests in several non-Gaussian models with heavy tails or skewed distributions, but in some Gaussian models also. Further, we also compare the performance of our test with the mean based tests in several models involving contaminated probability distributions. Finally, we demonstrate the performance of these tests in three real datasets: a Canadian weather dataset, a spectrometric dataset on chemical analysis of meat samples and a dataset on orthotic measurements on volunteers.

Contents

*Corresponding author.

1. Introduction

Analysis of variance based on ranks of real valued observations was studied by
[27]. Suppose that we have $n = \sum_{k=1}^{K} n_k$ independent real valued observations
$\{X_{ki} : i = 1, \ldots, n_k;\ k = 1, \ldots, K\}$ from K groups. We are interested in test-
ing whether the distributions of the groups are same or not. We rank all the
observations together, and let r_{ki} be the rank of the ith observation in the kth
group. Let $\bar{r}_{k\cdot} = n_k^{-1} \sum_{i=1}^{n_k} r_{ki}$ and $\bar{r}_{\cdot\cdot} = n^{-1} \sum_{k=1}^{K} n_k \bar{r}_{k\cdot} = 0.5(n+1)$. The test
statistic for the Kruskal-Wallis test is

$$W_n = (n-1) \frac{\sum_{k=1}^{K} n_k (\bar{r}_{k\cdot} - \bar{r}_{\cdot\cdot})^2}{\sum_{k=1}^{K} \sum_{i=1}^{n_k} (r_{ki} - \bar{r}_{\cdot\cdot})^2},$$

and a large value of W_n indicates that the distributions of the groups are not
same. When the groups have the same continuous distribution, the distribu-
tion of W_n is independent of the underlying distribution of the groups. The
usual mean based ANOVA test is developed assuming that the distributions of
the groups are Gaussian. It is well-known that the Kruskal-Wallis test exhibits
better performance than the mean based ANOVA test when the groups have
non-Gaussian distributions with heavy tails.

In [29], [30] and [31], several nonparametric tests based on multivariate spa-
tial signs and ranks were investigated. In [12, 13], spatial signs were employed to
generalize the Kruskal-Wallis test in a multivariate analysis of variance setup.
Those spatial signs and ranks are intrinsically related to spatial quantiles inves-
tigated in [11] and [25]. Consider independent multivariate observations $\{\mathbf{X}_{ki} :$
$i = 1, \ldots, n_k;\ k = 1, \ldots, K\}$ from K groups, and $n = \sum_{k=1}^{K} n_k$. Define $\mathbf{r}(\mathbf{x}) =$
$n^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathbf{s}(\mathbf{x} - \mathbf{X}_{ki})$ and $\bar{\mathbf{r}}_{k\cdot} = n_k^{-1} \sum_{i=1}^{n_k} \mathbf{r}(\mathbf{X}_{ki})$, where $\mathbf{s}(\mathbf{x} - \mathbf{y}) = (\mathbf{x} -$
$\mathbf{y})/\|\mathbf{x}-\mathbf{y}\|$ for $\mathbf{x} \neq \mathbf{y}$ and $\mathbf{s}(\mathbf{0}) = \mathbf{0}$. Note that when the observations are univari-
ate, $\mathbf{r}(\mathbf{X}_{ki}) = n^{-1}(2\,\text{rank}(\mathbf{X}_{ki}) - (n+1))$, where $\text{rank}(\mathbf{X}_{ki})$ is the rank of the uni-
variate observation $\mathbf{X}_{ki}$ within the pooled sample. Let $\mathbf{r}_k(\mathbf{x}) = n_k^{-1} \sum_{i=1}^{n_k} \mathbf{s}(\mathbf{x} -$
$\mathbf{X}_{ki})$ and $\hat{\mathbf{\Sigma}}_n = (n-K)^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathbf{r}_k(\mathbf{X}_{ki}) \mathbf{r}_k(\mathbf{X}_{ki})^t$. The test statistic for
the Choi-Marden one-way ANOVA test is

$$M_n = \sum_{k=1}^{K} n_k (\bar{\mathbf{r}}_{k\cdot})^t \hat{\mathbf{\Sigma}}_n^{-1} \bar{\mathbf{r}}_{k\cdot},$$

which is used to test the equality of the distributions of the groups. Unlike the
Kruskal-Wallis test for univariate data, the Choi-Marden test is not distribution-
free under the null hypothesis, and the authors implemented it using the asymp-
totic distribution of the test statistic. But, like the Kruskal-Wallis test, the

Choi-Marden test usually exhibits better powers and asymptotic relative efficiencies when the underlying distributions of the groups are non-Gaussian with heavy tails. When the observations $\mathbf{X}_{ki}$ are random functions, considered as random elements in an appropriate L_2 space, instead of finite dimensional random vectors, the Choi-Marden one-way ANOVA test cannot be directly applied for them. This is because in this case, the underlying space is an infinite dimensional Hilbert space, and the estimated covariance operator $\hat{\mathbf{\Sigma}}_n$ in an infinite dimensional space is not invertible and cannot be used to standardize the test statistic. Methodology for functional data, which are infinite dimensional in nature, and also high dimensional data requires non-involvement of the inverse of the sample covariance (see, e.g., [26, 22] for some methods in high dimensional data developed without using the inverse of the sample covariance).

Analysis of variance for functional data has been investigated by several authors. An ANOVA test for functional data based on the L_2 distance between the group means in the sample was proposed in [15]. In [33], [37] and [35], hypothesis testing in a linear model involving functional responses was investigated. In [8], a similar testing procedure was described in functional linear models and ANOVA, and demonstrated its utility in investigating brain electrical activity. In [39], a test of ANOVA was described, where the test statistic is obtained by integrating the pointwise F-statistic for the functional observations. In [38], another test of ANOVA was proposed, where the test statistic is the supremum of the pointwise F-statistic. In [23], a test of ANOVA was introduced, where the functional observations are projected on a number of principal components of the sample covariance operator, rendering the test as a test for finite dimensional observations. In [18], a permutation test of ANOVA was described for functional data. In [14], a test of ANOVA for functional data was described based on random projections. In [1], the procedures by [33] and [14] was adapted in an ANOVA problem for spatially correlated functional data. In [21], an L_2-norm based ANOVA procedure was used, which is similar to those proposed by [33] and [37], for weakly dependent functional time series. All these tests in the literature involving functional data are based on the mean of the response, as in the classical univariate and multivariate ANOVA problems. Recently, an ANOVA procedure was proposed in [34] based on the pairwise distances of the observations.

In this paper, we construct a test for one-way ANOVA for infinite dimensional data based on spatial signs, and describe implementations of this test based on the asymptotic distribution as well as a permutation procedure and a bootstrap procedure. We found that this test exhibits superior performance than the mean based tests in non-Gaussian heavy-tail processes, contaminated Gaussian and non-Gaussian processes and skewed processes. In the case of some Gaussian models also, our test outperforms the mean based tests. We demonstrate the usefulness of our testing procedure in simulated and real datasets.

2. A test based on spatial signs

In the context of functional data analysis, sample observations are considered as real valued random functions defined over some domain, which may be an interval, a rectangle, or some other set. The space of such functions can be considered as an L_2 space by equipping it with an appropriate L_2-norm. It is well known that such an L_2 space will be a separable Hilbert space when it is associated with a Borel σ-field and a σ-finite measure. In other words, one can model those random functions as random elements in a separable Hilbert space. The advantage of this approach is that the methodology is readily applicable to diverse types of data that can be viewed as elements in a separable Hilbert space, and we develop an ANOVA procedure for such data.

We consider a sample of n independent observations in a separable Hilbert space $\mathcal{H}$, which is divided into K groups with the kth group having n_k observations for $k = 1, \ldots, K$. The members of the kth group are denoted as $\mathbf{X}_{ki}$ for $i = 1, \ldots, n_k$. We assume that all the observations are independent, and for each of the groups, the observations within that group have identical distribution. Let the probability distribution of the kth group be denoted as P_k. We are interested in testing whether all the distributions P_k are the same or not. So, the null hypothesis is

$$\text{H}_0 : \ P_1 = P_2 = \cdots = P_K. \tag{2.1}$$

We assume that the distributions P_k are non-atomic probability distributions on the separable Hilbert space $\mathcal{H}$ equipped with the Borel σ-field. Also, the treatment effects are assumed to alter the underlying distributions of the groups. The literature on ANOVA for functional data described in section 1 is usually concerned with treatment effects which specifically alter the means of the groups. Our setup, which is inspired from the setups considered in [27] for univariate observations and in [12] for multivariate observations, covers the case where the treatment effects alter the means of the underlying distributions of the groups, as well as cases where the treatment effects alter the underlying the distributions of the groups but not necessarily the means of the groups.

Define $\mathbf{s}(\mathbf{x}) = \|\mathbf{x}\|^{-1}\mathbf{x}$ for $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{s}(\mathbf{0}) = \mathbf{0}$. Consider the quantity $E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})]$, which is the expected direction vector from the random element $\mathbf{X}_{l1}$ to the random element $\mathbf{X}_{k1}$. If the distributions P_k and P_l are same, then we have $E[\mathbf{s}(\mathbf{X}_{k1}-\mathbf{X}_{l1})] = E[\mathbf{s}(\mathbf{X}_{l1}-\mathbf{X}_{k1})] = -E[\mathbf{s}(\mathbf{X}_{k1}-\mathbf{X}_{l1})] = \mathbf{0}$. Define $\mathbf{R}(\mathbf{x}) = n^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathbf{s}(\mathbf{x}-\mathbf{X}_{ki})$. Note that $\mathbf{R}(\mathbf{x})$ is the average spatial sign of the point $\mathbf{x}$ with respect to the combined sample, which is the average of all direction vectors from the data points to $\mathbf{x}$. $\mathbf{R}(\mathbf{x})$ can also be viewed as the spatial rank of $\mathbf{x}$ in the combined sample of all the observations (cf. [12], [30]). Also define $\bar{\mathbf{R}}_k = n_k^{-1} \sum_{i=1}^{n_k} \mathbf{R}(\mathbf{X}_{ki})$. Note that $E[\bar{\mathbf{R}}_k] = n^{-1} \sum_{l=1}^{K} n_l E[\mathbf{s}(\mathbf{X}_{k1}-\mathbf{X}_{l1})] = \mathbf{0}$ for all k under H_0 in (2.1). So, if one is interested to test H_0, the test may be carried out based on the magnitudes of $\bar{\mathbf{R}}_k$, $k = 1, \ldots, K$. A high magnitude will indicate that H_0 in (2.1) is not true.

Define $\mathbf{U}_n = \left(\sqrt{n_1}\bar{\mathbf{R}}_1, \ldots, \sqrt{n_K}\bar{\mathbf{R}}_K \right)$. Note that $\mathbf{U}_n$ is a random element in the product Hilbert space $\mathcal{H} \times \mathcal{H} \times \cdots \times \mathcal{H} = \mathcal{H}^K$. Clearly, $E[\mathbf{U}_n] = \mathbf{0}$ under

the null hypothesis H_0 in (2.1). Based on this idea discussed above, we define the test statistic

$$SS_n = \|\mathbf{U}_n\|^2 = \sum_{k=1}^{K} n_k \left\| \bar{\mathbf{R}}_k \right\|^2.$$

We shall reject H_0 in (2.1) when SS_n is significantly large. We call the resulting test the SS test. The normalization using the $\sqrt{n_k}$ terms in the definition of $\mathbf{U}_n$ is done so that its asymptotic distribution is non-degenerate, as described in Theorem 2.1. The asymptotic distribution of SS_n under H_0 in (2.1) is obtained from this theorem in Corollary 2.2, which is used to derive the asymptotic implementation of the test.

$\bar{\mathbf{R}}_k$ can be expressed as a weighted sum of sample means of spatial signs. Define $\boldsymbol{\nu}_{kl} = E\left[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})\right]$ and $\widehat{\boldsymbol{\nu}}_{kl} = (n_k n_l)^{-1} \sum_{i_k=1}^{n_k} \sum_{i_l=1}^{n_l} (\mathbf{X}_{ki_k} - \mathbf{X}_{li_l})$. Let $\widehat{\boldsymbol{\nu}} = (\widehat{\boldsymbol{\nu}}_{11}, \ldots, \widehat{\boldsymbol{\nu}}_{KK})'$ be the vector containing all the sample means of spatial signs. Let $\mathbf{0}_K$ denote the vector of K null elements, and let $\mathbf{a}_k = (\mathbf{z}_1', \ldots, \mathbf{z}_k', \ldots, \mathbf{z}_K')'$, where $\mathbf{z}_i = \mathbf{0}_K$ for all $i \neq k$, and $\mathbf{z}_k = n^{-1}(n_1, \ldots, n_K)'$. So, for each k, $\mathbf{a}_k$ is a K^2-dimensional vector of weights whose sum is 1. Then, $\sqrt{n_k}\bar{\mathbf{R}}_k = (\sqrt{n_k/n})\mathbf{a}_k'\sqrt{n}\widehat{\boldsymbol{\nu}}$ for all k. Define the matrix of weights $\mathbf{A}_n = \left((\sqrt{n_1/n})\mathbf{a}_1, \ldots, (\sqrt{n_K/n})\mathbf{a}_K \right)'$. We have $\mathbf{U}_n = \mathbf{A}_n\sqrt{n}\widehat{\boldsymbol{\nu}}$. It can be verified that when $n^{-1}n_k \to \lambda_k \in (0,1)$ for all k as $n \to \infty$, $\mathbf{A}_n$ converges to a non-null matrix, say, $\mathbf{A}$, and the covariance operator of $\sqrt{n}\widehat{\boldsymbol{\nu}}$ also converges, say, to $\boldsymbol{\Gamma}$. Then, the covariance operator of $\mathbf{U}_n$ converges to $\mathbf{A}\boldsymbol{\Gamma}\mathbf{A}'$ as $n \to \infty$. Note that $\mathbf{A}\boldsymbol{\Gamma}\mathbf{A}'$ is a $K \times K$ matrix of covariance operators. The (k_1, k_2)th element of this matrix is $\boldsymbol{\sigma}_{k_1 k_2}$, which is described in Theorem 2.1. A similar approach for deriving the asymptotic distribution of a rank-based statistic for univariate data was developed in [6].

Theorem 2.1. *Assume that $n^{-1}n_k \to \lambda_k \in (0,1)$ for all k as $n \to \infty$. Let $\mathbf{X}_i$, $\mathbf{Y}_j$ and $\mathbf{Z}_k$ be independent random elements having the distributions of the ith, the jth and the kth group of the sample, respectively. Define $\mathbf{C}(i,j,k) = Cov(E\left[\mathbf{s}\left(\mathbf{X}_i - \mathbf{Z}_k\right) \mid \mathbf{Z}_k\right], E\left[\mathbf{s}\left(\mathbf{Y}_j - \mathbf{Z}_k\right) \mid \mathbf{Z}_k\right])$ and $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}_{k_1 k_2})_{K \times K}$, where*

$$\boldsymbol{\sigma}_{k_1 k_2} = \sqrt{\lambda_{k_1} \lambda_{k_2}} \sum_{l=1}^{K} \lambda_l \left[\mathbf{C}(k_1, k_2, l) - \mathbf{C}(l, k_2, k_1) - \mathbf{C}(k_1, l, k_2) \right]$$

$$+ \sum_{l_1=1}^{K} \sum_{l_2=1}^{K} \lambda_{l_1} \lambda_{l_2} \mathbf{C}(l_1, l_2, k_1) \mathbb{I}(k_1 = k_2).$$

Then, $[\mathbf{U}_n - E\left[\mathbf{U}_n\right]] \xrightarrow{w} G\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ as $n \to \infty$, where $G\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ denotes a Gaussian random element in $\mathcal{H}^K$ with mean $\mathbf{0}$ and covariance operator $\boldsymbol{\Sigma}$.

Corollary 2.2. *Assume that $n^{-1}n_k \to \lambda_k \in (0,1)$ for all k as $n \to \infty$. Under H_0 in (2.1), we have $E\left[\mathbf{U}_n\right] = \mathbf{0}$. Consequently, $\mathbf{U}_n \xrightarrow{w} G\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ and $SS_n = \|\mathbf{U}_n\|^2 \xrightarrow{w} \|\mathbf{W}\|^2$ as $n \to \infty$, where $\mathbf{W}$ is a random element having distribution $G\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$.*

The p-value of the SS test in the asymptotic implementation is given by $P[\|\mathbf{W}\|^2 \geq$ the observed value of $SS_n]$. The p-value can be numerically computed using a Monte Carlo procedure. However, since $\boldsymbol{\Sigma}$ is unknown, to implement the test, we need to estimate the covariance operator $\boldsymbol{\Sigma}$ from the sample. Let $\otimes$ be the outer product on $\mathcal{H}$ such that for every $\mathbf{x}, \mathbf{y} \in \mathcal{H}$, $\mathbf{x} \otimes \mathbf{y} : \mathcal{H} \to \mathcal{H}$ is a linear operator defined by $(\mathbf{x} \otimes \mathbf{y})(\mathbf{w}) = \langle \mathbf{w}, \mathbf{x} \rangle \mathbf{y}$. Note that for $i, j, k \in \{1, \ldots, K\}$,

$$\mathbf{C}(i,j,k) = E\left[E\left[\mathbf{s}(\mathbf{X}_i - \mathbf{Z}_k) \mid \mathbf{Z}_k\right] \otimes E\left[\mathbf{s}(\mathbf{Y}_j - \mathbf{Z}_k) \mid \mathbf{Z}_k\right]\right] \\ - E\left[\mathbf{s}(\mathbf{X}_i - \mathbf{Z}_k)\right] \otimes E\left[\mathbf{s}(\mathbf{Y}_j - \mathbf{Z}_k)\right],$$

where $\mathbf{X}_i$, $\mathbf{Y}_j$ and $\mathbf{Z}_k$ are independent random elements having the distributions of the ith, the jth and the kth group, respectively. Let

$$\mathbf{C}_n(i,j,k)$$
$$= \frac{1}{n_k-1} \sum_{l_k=1}^{n_k} \left[\left(\frac{1}{n_i}\sum_{l_i=1}^{n_i} \mathbf{s}\left(\mathbf{X}_{il_i} - \mathbf{X}_{kl_k}\right)\right) \otimes \left(\frac{1}{n_j}\sum_{l_j=1}^{n_j} \mathbf{s}\left(\mathbf{X}_{jl_j} - \mathbf{X}_{kl_k}\right)\right)\right]$$
$$- \left(\frac{1}{n_i n_k}\sum_{l_i=1}^{n_i}\sum_{l_k=1}^{n_k} \mathbf{s}\left(\mathbf{X}_{il_i} - \mathbf{X}_{kl_k}\right)\right) \otimes \left(\frac{1}{n_j n_k}\sum_{l_j=1}^{n_j}\sum_{l_k=1}^{n_k} \mathbf{s}\left(\mathbf{X}_{jl_j} - \mathbf{X}_{kl_k}\right)\right),$$

where $i, j, k \in \{1, \ldots, K\}$. The estimate of $\boldsymbol{\Sigma}$ is defined as

$$\widehat{\boldsymbol{\Sigma}}_n = \left(\boldsymbol{\sigma}_{k_1 k_2}^{(n)}\right)_{K \times K}, \text{ where}$$
$$\boldsymbol{\sigma}_{k_1 k_2}^{(n)} = \frac{\sqrt{n_{k_1} n_{k_2}}}{n} \sum_{l=1}^{K} \frac{n_l}{n} \left[\mathbf{C}_n(k_1, k_2, l) - \mathbf{C}_n(l, k_2, k_1) - \mathbf{C}_n(k_1, l, k_2)\right]$$
$$+ \sum_{l_1=1}^{K}\sum_{l_2=1}^{K} \frac{n_{l_1} n_{l_2}}{n^2} \mathbf{C}_n(l_1, l_2, k_1)\mathbb{I}(k_1 = k_2).$$

Using the estimated covariance operator $\widehat{\boldsymbol{\Sigma}}_n$, we generate independent observations from the distribution $G\left(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_n\right)$. Let us denote the generated observations as $\mathbf{W}_1, \ldots, \mathbf{W}_N$, where N is a suitably large number. Then the approximate p-value of the SS test is given by the proportion of $\|\mathbf{W}_i\|^2$ values greater than or equal to the observed value of SS_n. A similar estimation procedure for the p-value is described in [15] for their test of ANOVA for functional data.

For the asymptotic validity of this procedure, the estimated covariance operator needs to asymptotically consistent, which is established in the following theorem.

Theorem 2.3. *Assume that $n^{-1}n_k \to \lambda_k \in (0,1)$ for all k as $n \to \infty$. Then, $\widehat{\boldsymbol{\Sigma}}_n \longrightarrow \boldsymbol{\Sigma}$ almost surely as $n \to \infty$ in the operator norm.*

When the number of observations in any of the groups is small, the asymptotic null distribution of the test statistic described in Corollary 2.2 may not

be a good approximation of its finite sample distribution. For such cases, we can implement the test based on a bootstrap procedure (see, e.g., [17]) or a permutation procedure.

When the null hypothesis is true, the distributions of all the groups are same, and the natural estimator of that common underlying distribution is the empirical distribution of the pooled sample of all the groups. So, in our bootstrap procedure, we draw a random sample of size n with replacement from the pooled sample and assign them to the groups so that given the original sample $\{\mathbf{X}_{ki} \mid k = 1, \ldots, K; i = 1, \ldots, n_k\}$, the observations in the bootstrap sample $\{\mathbf{X}_{ki}^* \mid k = 1, \ldots, K; i = 1, \ldots, n_k\}$ are independent and identically distributed, and for every k and i, we have $P[\mathbf{X}_{ki}^* = \mathbf{X}_{lj}] = n^{-1}$ for all l and j. From the bootstrap sample, we compute the bootstrap value of the statistic SS_n^*. By repeating this procedure M_b times independently, we get M_b values of SS_n^*. The p-value of this testing procedure is given by the proportion of the M_b values of SS_n^* higher than the actual value of SS_n. When an observation is repeated in the bootstrap sample, we use the convention of $\mathbf{s}(\mathbf{0}) = \mathbf{0}$ while computing the statistic SS_n^*.

Next, we describe the permutation implementation. Since the underlying distributions of all the groups in the sample are identical when the null hypothesis is true, we generate a random permutation of the pooled sample, then assign the first n_1 elements in that permuted sample to the first group, the next n_2 elements to the next group, and so on. From this permuted sample, we compute the value of the test statistic, which is denoted as $SS_n^\#$. We get M_p values of $SS_n^\#$ by repeating this procedure M_p times independently. The p-value of the permutation implementation of the test is given by the proportion of the M_p values of $SS_n^\#$ higher than the actual value of SS_n.

The advantage of the permutation implementation is that no observation is repeated in the generated sample unlike the bootstrap implementation. However, in our simulation analysis, we did not find any discernible difference between the performances of the two implementations. In principle, the permutation implementation can also be developed based on all possible permutations of the sample instead of a fixed number of random permutations. However, the number of all permutations of $(1, \ldots, n)$ increases very rapidly with n, which makes the permutation implementation computationally very expensive even for a moderate n. The test based on random permutations is prescribed to cover for such situations. We use the permutation implementation of the test for small sample sizes.

The validity of the implementations of our test is established in the following theorem.

Theorem 2.4. *Let $n^{-1}n_k \to \lambda_k \in (0,1)$ for all k as $n \to \infty$. Under H_0 in (2.1), for every $0 < \alpha < 1$, the size of a level α test based on the asymptotic procedure or the bootstrap procedure or the permutation procedure described above converges to α as $n \to \infty$ and $M_b, M_p \to \infty$.*

Note that the testing procedure is based on the magnitudes of the quantities $\bar{\mathbf{R}}_k$'s, and the distance between $\bar{\mathbf{R}}_k$ and $E[\bar{\mathbf{R}}_k]$ converges to 0 in proba-

bility for all k as $n \to \infty$ irrespective of H_0 in (2.1). Also, for all k, $E[\bar{\mathbf{R}}_k] = n^{-1} \sum_{l=1}^{K} n_l E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})] \longrightarrow \sum_{l=1}^{K} \lambda_l E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})]$ as $n \to \infty$. Recall that when the underlying distributions of the K groups are identical, i.e., when H_0 in (2.1) holds, $E[\bar{\mathbf{R}}_k] = \mathbf{0}$ for all k. When the distributions of the groups are not identical such that the asymptotic limit of $E[\bar{\mathbf{R}}_k]$ is nonzero for at least one k, the power of all the implementations of the test converges to 1 as the sample size increases, which we state in the next theorem.

Theorem 2.5. *Assume that $n^{-1} n_k \to \lambda_k \in (0,1)$ for all k as $n \to \infty$. When H_0 in (2.1) is not true such that $\sum_{l=1}^{K} \lambda_l E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})] \neq \mathbf{0}$ for any k, we have $\|E[\mathbf{U}_n]\| \to \infty$ as $n \to \infty$, and $SS_n \xrightarrow{P} \infty$ as $n \to \infty$. As a consequence, for every $0 < \alpha < 1$, the power of a level α test based on each of the asymptotic procedure, the bootstrap procedure and the permutation procedure converges to 1 as $n \to \infty$ and $M_b, M_p \to \infty$.*

The standard setup of ANOVA with location shift is a special case of our general setup. In the literature on ANOVA for functional data, most of the authors concentrated on this setup with certain additional assumptions required for their tests. In this setup, it is assumed that

$$\mathbf{X}_{ki} = \boldsymbol{\mu}_k + \boldsymbol{\epsilon}_{ki}, \tag{2.2}$$

where $\boldsymbol{\epsilon}_{ki}$'s are independent random elements in $\mathcal{H}$ having identical distribution P_0, and $\boldsymbol{\mu}_k$'s are fixed elements in $\mathcal{H}$. We are interested in testing whether the $\boldsymbol{\mu}_k$'s are identical for all k or not, i.e.,

$$H_0' : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_K. \tag{2.3}$$

When H_0' in (2.3) is true, all the underlying distributions of the K groups are same in the setup (2.2), and hence H_0 in (2.1) holds. Note that the existence of the mean or covariance or any other moment of the error distribution P_0 is not assumed in the setup (2.2). This relaxation enables the methodology developed here to be applicable in heavy-tailed processes which have no moments. In the ANOVA with location shift setup for functional data considered in [15], [33], [37], [35], [39], [38], [23] and [18], it is additionally assumed that the error distribution P_0 has mean $\mathbf{0}$ and a covariance operator. Thus our setup (2.2) covers the ANOVA with location shift setup considered by the other authors, but does not assume the existence of the mean or the covariance operator of $\boldsymbol{\epsilon}_{ki}$'s.

The validity and the asymptotic consistency of the different implementations of the SS test in the ANOVA with location shift setup follow from the result below.

Theorem 2.6. *Let $n^{-1} n_k \to \lambda_k \in (0,1)$ for all k as $n \to \infty$. Under H_0' in (2.3), for every $0 < \alpha < 1$, the sizes of a level α test based on the asymptotic procedure, the bootstrap procedure and the permutation procedure described above converge to α as $n \to \infty$ and $M_b, M_p \to \infty$. Further, when H_0' in (2.3) is not true and when the support of P_0 is not contained in a straight line in $\mathcal{H}$, we have $\sum_{l=1}^{K} \lambda_l E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})] \neq \mathbf{0}$ for every k, and hence for every $0 < \alpha < 1$, the*

power of a level α test based on each of the procedures converges to 1 *as* $n \to \infty$ *and* $M_b, M_p \to \infty$.

The condition that the support of P_0 is not contained in a straight line in $\mathcal{H}$ essentially means that P_0 is not degenerate in any direction in $\mathcal{H}$, and this condition holds for all the common non-degenerate distributions like the Gaussian processes, the t processes, etc.

We have framed the null hypothesis as the equality of the distributions underlying the groups. However, the test statistic SS_n is based on spatial signs, and can be used in a heteroscedastic situation also. The methodology developed here covers heteroscedastic situations, where the null hypothesis is

$$\mathrm{H}_0^* : \sum_{l=1}^{K} \lambda_l E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})] = \mathbf{0} \text{ for all } k,$$

where $\lambda_k = \lim_{n \to \infty} n^{-1} n_k \in (0, 1)$ for all k. If $\sqrt{n}(n^{-1} n_k - \lambda_k) \to 0$ as $n \to \infty$ for all k, then using arguments similar to those in the proof of Corollary 2.2, it can be verified that $SS_n \xrightarrow{w} \|\mathbf{W}\|^2$ as $n \to \infty$ under H_0^*, where $\mathbf{W}$ is as described in Corollary 2.2. This indicates that results analogous to Theorem 2.4 and Theorem 2.5 can be established under H_0^*, which guarantee the validity of the test procedures for testing H_0^*.

The null hypothesis H_0^* holds in many heteroscedastic setups. For example, consider any distribution P, which is symmetric with respect to the origin $\mathbf{0}$, i.e., for P being the distribution of the random element $\mathbf{X}$, it is also the distribution of $-\mathbf{X}$. Let $c_1, \ldots, c_K$ be positive constants, $\boldsymbol{\mu}$ be a fixed element, and P_k be the distribution of the random element $\boldsymbol{\mu} + c_k \mathbf{X}$ for $k = 1, \ldots, K$. Let $\mathbf{X}'$ be an independent copy of $\mathbf{X}$. Then,

$$\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1}) \stackrel{d}{=} \mathbf{s}(c_k \mathbf{X} - c_l \mathbf{X}') \stackrel{d}{=} \mathbf{s}(-c_k \mathbf{X} + c_l \mathbf{X}') \stackrel{d}{=} -\mathbf{s}(c_k \mathbf{X} - c_l \mathbf{X}')$$

for all k and l, which implies that $E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})] = \mathbf{0}$ for all k and l, and H_0^* is satisfied. Examples of such symmetric distributions P include all centered Gaussian processes and centered t processes.

3. Comparison of different tests

In this section, we shall compare the asymptotic and finite sample performance of the SS test with those of several other tests in the literature. Theorem 2.6 in the preceding section implies that our SS test is consistent under any fixed alternative. Similar consistency under fixed alternative holds for several of the mean based tests. Hence, in order to compare the asymptotic performance of these tests, we consider a class of shrinking alternatives:

$$\boldsymbol{\mu}_k = (1/\sqrt{n})\boldsymbol{\delta}_k \text{ for } k = 1, \ldots, K, \text{ where } \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K \in \mathcal{H} \text{ are fixed.} \tag{3.1}$$

Similar shrinking alternatives were considered in [39] and [38]. Note that the Frechét derivative of $\mathbf{s}(\mathbf{x}) = \|\mathbf{x}\|^{-1}\mathbf{x}$ exists at all $\mathbf{x} \neq \mathbf{0}$, and we denote it as

$\mathbf{s}^{(1)}(\mathbf{x})(\cdot)$. The expression for $\mathbf{s}^{(1)}(\mathbf{x})(\cdot)$ is given in the proof of Lemma B.1. From the following theorem, we derive the asymptotic power of the SS test under the class of shrinking alternatives in (3.1).

Theorem 3.1. *Let* $\mathbf{X}$ *and* $\mathbf{X}'$ *be independent random elements having identical distribution* P_0. *Let* $n^{-1}n_k \to \lambda_k \in (0,1)$ *for all* k *as* $n \to \infty$, *and* $\bar{\boldsymbol{\delta}} = \sum_{l=1}^{K} \lambda_l \boldsymbol{\delta}_l$. *Assume that* P_0 *is non-atomic and not contained in any straight line in* $\mathcal{H}$, *and* $E\left[\|\mathbf{X} - \mathbf{X}'\|^{-1}\right] < \infty$. *Then, under* (3.1), $E\left[\mathbf{U}_n\right] \to \mathbf{U}_0 = (\mathbf{u}_1, \ldots, \mathbf{u}_K)$ *as* $n \to \infty$, *where* $\mathbf{u}_k = E\left[\mathbf{s}^{(1)}\left(\mathbf{X} - \mathbf{X}'\right)\right]\left(\sqrt{\lambda_k}\left(\boldsymbol{\delta}_k - \bar{\boldsymbol{\delta}}\right)\right)$ *for all* $k = 1, \ldots, K$. *Consequently,* $\mathbf{U}_n \xrightarrow{w} G\left(\mathbf{U}_0, \boldsymbol{\Sigma}\right)$ *as* $n \to \infty$, *and* $SS_n = \|\mathbf{U}_n\|^2 \xrightarrow{w} \|\mathbf{U}_0 - \sum_{i=1}^{\infty}\langle \mathbf{U}_0, \boldsymbol{\beta}_i \rangle\|^2 + \sum_{i=1}^{\infty}\left(\langle \mathbf{U}_0, \boldsymbol{\beta}_i \rangle + \sqrt{\alpha_i}\mathbf{Z}_i\right)^2$, *where* $\alpha_1, \alpha_2, \ldots$ *is the decreasing sequence of eigenvalues of* $\boldsymbol{\Sigma}$, $\boldsymbol{\beta}_i$ *is the corresponding eigenvector of* α_i *for all* i, *and* $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ *are independent standard normal random variables. Further,* $\mathbf{U}_0 \neq \mathbf{0}$ *if* $\boldsymbol{\delta}_k \neq \boldsymbol{\delta}_l$ *for at least one pair of* k *and* l.

3.1. Description of other tests

We describe below the tests in the literature with which we compare the performance of the SS test in this paper. We compare the finite sample powers against fixed alternatives and the asymptotic powers against shrinking alternatives of these tests. Though our testing procedures are also applicable for heteroscedastic data, several of the testing procedures with which we compare the performance of our test were presented for the homoscedastic case (see, e.g., [33, 37, 35, 39]). For this reason, we restrict the comparison of performances in the simulation study to the homoscedastic setup. The real datasets we consider for the comparison of performances also do not exhibit heteroscedasticity.

The tests in the literature, which we consider here, are developed for the case where the observations are real-valued square-integrable random functions on an interval $[a, b]$, i.e., $\mathbf{X}_{ki}(\cdot) : [a, b] \to \mathbb{R}$ is a random function for all k and i. So, $\mathbf{X}_{ki}(\cdot)$ is an element of the $L_2[a, b]$ space, which is a separable Hilbert space, and $\|\mathbf{x}\| = \left(\int \|\mathbf{x}(t)\|^2 \mathrm{d}t\right)^{1/2}$ for $\mathbf{x} \in L_2[a, b]$. Define $\bar{\mathbf{X}}_{k\cdot}(t) = n_k^{-1}\sum_{i=1}^{n_k} \mathbf{X}_{ki}(t)$ and $\bar{\mathbf{X}}_{\cdot\cdot}(t) = n^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k} \mathbf{X}_{ki}(t)$ for all $t \in [a, b]$.

We call the test of ANOVA for functional data proposed by [15] as the CFF test. From the methodology developed in [37], an L_2 norm based test of ANOVA can be derived, which we call the ZC test. In [33] and [35], another test for functional linear models was proposed, and we call the corresponding test of ANOVA as the F-type test. The test of ANOVA described in [39] is denoted here as the GPF test, and the test proposed in [38] is called the F-max test following the authors. We call the permutation test given by [18] as the GS test. The random projection based test proposed in [14] is denoted here as the CAFB test. The test by [23] is denoted as the HR test. The detailed descriptions of the test statistics and the implementations of the tests are presented in Appendix A.

We shall compare the finite sample performance of the SS test with all the above tests. For the asymptotic power comparison against shrinking alternatives,

we consider all the tests except the F-type test, the CAFB test and the GS test. The F-type test is excluded from the asymptotic power comparison because the F-type test and the ZC test have identical asymptotic performance. The CAFB test and the GS test are also excluded from asymptotic power comparison because the authors of these tests did not study their asymptotic distributions, and we found it difficult to derive their asymptotic distributions.

For the comparison of the asymptotic power against shrinking alternatives, we need the mathematical expressions of the asymptotic powers of these tests in this setup. The asymptotic power of the F-max test under (3.1) is derived from Proposition 3 in [38], and that of the GPF test under (3.1) is obtained from Proposition 3 in [39]. From Theorem 4.17 and Remark 4.11 in [36], we get that the ZC test and the F-type test are asymptotically equivalent, and hence have the same asymptotic power. The expressions of the asymptotic powers of the CFF test, the ZC test and the HR test under the class of shrinking alternatives described in (3.1), where $\mathcal{H} = L_2[a, b]$, are derived in Appendix A.

3.2. Asymptotic power study of different tests

We compare the asymptotic powers of the tests under the class of shrinking alternatives described in (3.1) in probability models with 3 groups. Note that for a random element $\mathbf{V}$ following Gaussian process G, its corresponding t process with k degrees of freedom can be obtained by dividing $\mathbf{V}$ by $\sqrt{\chi/k}$, where χ is a chi-square random variable with k degrees of freedom independent of $\mathbf{V}$. We denote this t process as $t_{k,G}$. The derivation of the asymptotic powers of the tests except the SS test requires that $E\left[\|\mathbf{X}\|^2\right] < \infty$, and for $\mathbf{X}$ following a $t_{k,G}$ distribution with $k \geq 3$, $E\left[\|\mathbf{X}\|^2\right] < \infty$. We take $\mathcal{H} = L_2[a, b]$, and consider the standard Brownian motion (SBM) over $[a, b]$ along with its corresponding t_3 and t_4 processes as particular cases of P_0, where P_0 is as described below (2.2). The interval $[a, b]$ is taken as $[0.25, 0.75]$, which ensures that $\mathbf{X}_{ki}(t)$ is non-degenerate at every $t \in [a, b]$ for P_0 being any of the processes mentioned above.

We take the shifts in the groups under (3.1) as $\boldsymbol{\delta}_1 = \mathbf{0}$, $\boldsymbol{\delta}_2 = c_1\boldsymbol{\eta}_1$ and $\boldsymbol{\delta}_3 = c_2\boldsymbol{\eta}_2$. Here, c_1, c_2 are real valued constants and $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathcal{H}$ are defined as $\boldsymbol{\eta}_1(t) = t$ and $\boldsymbol{\eta}_2(t) = (t - 0.25)(0.75 - t)$ for $t \in [0.25, 0.75]$. Also, we take $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$. The nominal level is fixed at 5%. To generate asymptotic power curves, we fix c_1 and vary c_2 in an interval.

The asymptotic powers of the tests in the setup described above are presented in Figure 1. We observe that the performance of the SS test is better than all the other tests for both the Gaussian and the non-Gaussian processes.

3.3. Finite sample performance in simulated data

We compare the finite sample power of the SS test, the CFF test, the ZC test, the F-max test, the GPF test, the HR test, the CAFB test and the GS test in the setup (2.2) with 3 groups. We consider the same space $\mathcal{H} = L_2[a, b]$ with
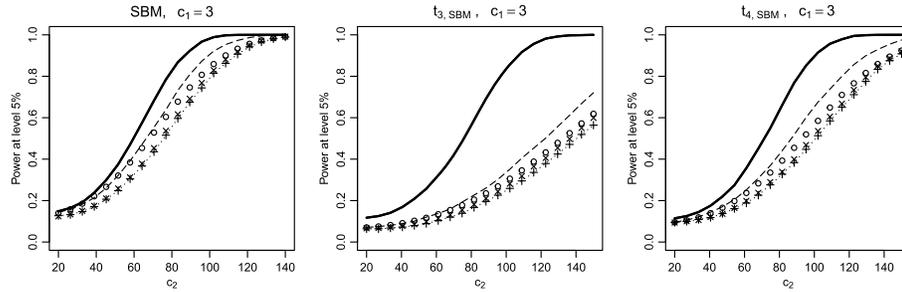
FIG 1. *Asymptotic power curves of the SS test (−), CFF test (++), ZC test (···), F-max test (− −), GPF test (××) and HR test (∘∘) under nominal level 5% in SBM and associated t processes.*

$[a, b] = [0.25, 0.75]$ as in subsection 3.2. For the cases of the distribution P_0, we consider the following three collections:

- *Gaussian and t processes:* The cases of the distribution P_0 are the SBM and its associated t_1 and t_3 processes.
- *Contaminated models:* Here, the distribution P_0 is replaced with a mixture distribution $(1-p)P_0 + pP^{(s)}$, where p is the proportion of contamination, s is a positive number and $P^{(s)}$ is the distribution of the random element $s\mathbf{X}$, where $\mathbf{X}$ follows P_0. We take $p = 0.25$ and $s = 5$. The cases of the distribution P_0, which are replaced with the respective contaminated models, are the SBM and its associated t_1 and t_3 processes as above.
- *Skewed distributions:* Three skewed distributions are considered:

    1. Geometric Brownian motion: P_0 is the distribution of the process $\mathbf{G}_1(t) = \exp[\mathbf{B}(t)]$, where $\mathbf{B}(t)$ is an SBM.

    2. Squared Brownian motion: P_0 is the distribution of the process $\mathbf{G}_2(t) = (\mathbf{B}(t))^2$, where $\mathbf{B}(t)$ is an SBM.

    3. Squared t process: P_0 is the distribution of the process $\mathbf{G}_3(t) = (\mathbf{T}(t))^2$, where $\mathbf{T}(t)$ is the $t_{3,\text{SBM}}$ process.

Note that the covariance operator of P_0 exists for all the cases above except the $t_{1,\text{SBM}}$ process and the contaminated $t_{1,\text{SBM}}$ process. Most of the other tests, which we consider here for comparison of powers, require the existence of the covariance operator for the validity of their theoretical properties. However, in this finite sample power study, we consider in addition the heavy-tailed $t_{1,\text{SBM}}$ and the contaminated $t_{1,\text{SBM}}$ processes to investigate the finite-sample performance of the tests under such non-Gaussian heavy-tailed distributions.

The observations from these distributions, which are random functions, are generated on an equispaced grid of length 100 on the interval $[a, b]$.

The shifts in the 3 groups are taken as $\boldsymbol{\mu}_1 = \mathbf{0}$, $\boldsymbol{\mu}_2 = c_1\boldsymbol{\eta}_1$ and $\boldsymbol{\mu}_3 = c_2\boldsymbol{\eta}_2$, where c_1, c_2 are constants and $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathcal{H}$ are defined by $\boldsymbol{\eta}_1(t) = t$ and $\boldsymbol{\eta}_2(t) = (t - 0.25)(0.75 - t)$ for $t \in [0.25, 0.75]$, which are the same as considered

Estimated sizes of the SS test, CFF test, ZC test, F-max test, GPF test, F-type test, HR test, CAFB test and GS test in the Gaussian and t processes (nominal level 5%).

P_0	(n_1, n_2, n_3)	SS	CFF	ZC	F-max	GPF	F-type	HR	CAFB	GS
SBM	(20, 20, 20)	0.054	0.054	0.060	0.064	0.061	0.051	0.081	0.032	0.048
SBM	(4, 4, 4)	0.056	0.110	0.125	0.044	0.102	0.013	0.305	0.031	0.052
$t_{1,\text{SBM}}$	(20, 20, 20)	0.053	0.014	0.016	0.017	0.015	0.014	0.025	0.025	0.055
$t_{1,\text{SBM}}$	(4, 4, 4)	0.058	0.041	0.080	0.021	0.045	0.001	0.183	0.039	0.057
$t_{3,\text{SBM}}$	(20, 20, 20)	0.049	0.053	0.054	0.053	0.051	0.048	0.081	0.030	0.052
$t_{3,\text{SBM}}$	(4, 4, 4)	0.055	0.089	0.095	0.029	0.090	0.002	0.270	0.031	0.054

in subsection 3.2. Note that when $c_1 = c_2 = 0$, H_0' in (2.3) is satisfied. So, taking $c_1 = c_2 = 0$, we estimate the sizes of the tests at nominal level 5%. To generate power curves, we fix c_1 and vary c_2 in an interval, and estimate the powers of the tests corresponding to those values of c_1 and c_2. Specifically, we choose 20 equispaced values of c_2 in an interval which spans most of the power range of the SS test. The specific values of the constants vary among the distributions.

To investigate the performances of the asymptotic and the permutation implementations of the SS test, we consider two cases of group sizes. We take $n_1 = n_2 = n_3 = 20$ in the first case, where the asymptotic implementation of the SS test is used, and $n_1 = n_2 = n_3 = 4$ in the second case, where the permutation implementation is used. We estimate the sizes and the powers of the tests under nominal level 5% based on 1000 independent replications.

To get observations from the chosen distributions of $\mathbf{X}_{ki}$ in (2.2), we first generate observations $\boldsymbol{\epsilon}_{ki}$ from P_0, then add the appropriate $\boldsymbol{\mu}_k$ to get the $\mathbf{X}_{ki}$ observations. For each of the cases of P_0 in the power study, we generate 1000 independent sets $\{\boldsymbol{\epsilon}_{ki} \mid \boldsymbol{\epsilon}_{ki} \sim P_0 \text{ are independent}, i = 1, \ldots, n_k; k = 1, 2, 3\}$, then obtain the corresponding 1000 independent sets $\{\mathbf{X}_{ki} = \boldsymbol{\mu}_k + \boldsymbol{\epsilon}_{ki} \mid \boldsymbol{\mu}_1 = \mathbf{0}, \boldsymbol{\mu}_2 = c_1\boldsymbol{\eta}_1, \boldsymbol{\mu}_3 = c_2\boldsymbol{\eta}_2, i = 1, \ldots, n_k; k = 1, 2, 3\}$. For different values of c_2, the sets of observations $\{\boldsymbol{\epsilon}_{ki}\}$ are kept same. This eliminates the random fluctuation in the estimated power curves arising from the random sampling process, and the estimated power curves are relatively smooth. The 1000 independent sets of the $\boldsymbol{\epsilon}_{ki}$ observations used to estimated the sizes of the tests for different underlying distributions P_0 are generated independently from those in the power study.

Results for the Gaussian and t processes

The estimated sizes of the tests in the the Gaussian and t processes are presented in Table 1, and the estimated power curves of the tests are plotted in Figure 2. We note that only the SS test and the GS test have sizes close to the nominal level irrespective of the group sizes or underlying distributions. The F-max test, the F-type test and the CAFB test tend to have lower sizes for small group sizes or non-Gaussian heavy tail processes (i.e., $t_{1,\text{SBM}}$ and $t_{3,\text{SBM}}$), while the sizes of the CFF test, the ZC test and the GPF test are sometimes significantly higher
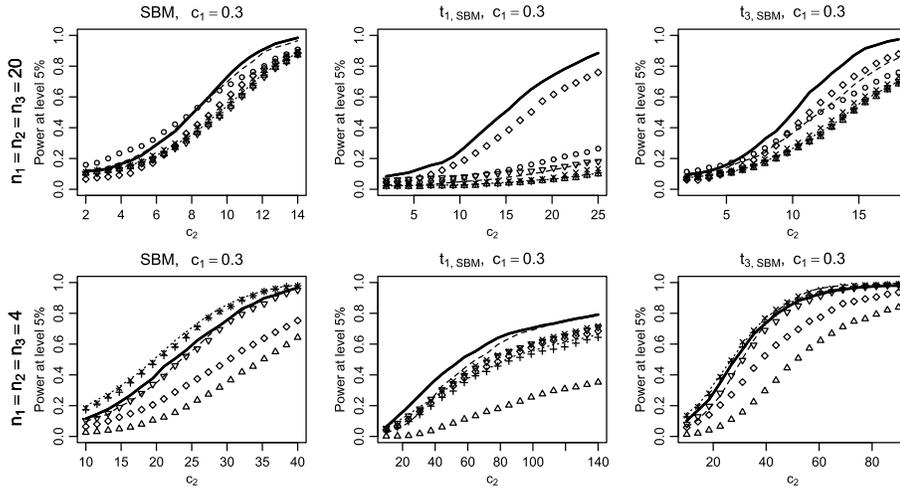
FIG 2. *SS test (−), CFF test (++), ZC test (···), F-max test (− −), GPF test (××), F-type test (△), HR test (○), CAFB test (◇) and GS test (▽) under nominal level 5% in SBM and associated t processes:* 1st *row:* $n_1 = n_2 = n_3 = 20$, 2nd *row:* $n_1 = n_2 = n_3 = 4$.

than the nominal level and sometimes somewhat lower than the nominal level. The size of the HR test is always significantly higher than the nominal level for small group sizes, and so it is omitted from the power comparison for the case of $n_1 = n_2 = n_3 = 4$.

For $n_1 = n_2 = n_3 = 20$, we note in Figure 2 that the performance of the SS test in case of SBM is sometimes narrowly beaten by only the HR test. However, the HR test also exhibits an estimated size over 8.5% in that case, considerably higher than the nominal level, indicating it is biased towards rejecting the null hypothesis. For the non-Gaussian heavy tail processes (i.e., $t_{1,SBM}$ and $t_{3,\text{SBM}}$), the performance of the SS test is significantly better than all other tests.

When the group sizes are small, i.e., $n_1 = n_2 = n_3 = 4$, and P_0 is the SBM, we notice that the power of the SS test is narrowly exceeded by that of the CFF test, the ZC test and the GPF test. However, in this case, the CFF test, the ZC test and the GPF test have sizes over 10% indicating their bias towards rejecting the null hypothesis. In the case of P_0 being $t_{1,\text{SBM}}$ here, the SS test exhibits better performance than all other tests. For P_0 being $t_{3,\text{SBM}}$, the power curve of the SS test is very close to that of the best performing test.

Results for the contaminated models

Since the SS test exhibits superior performance in distributions having heavy tails, we were motivated to investigate its performance in contaminated models. The contaminated models introduce impurities in the covariance operator of the groups but do not affect the means. We consider the same choices of $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$, n_1, n_2 and n_3 as in the uncontaminated models. Under the nominal level 5%, we

TABLE 2
Estimated sizes of the SS test, CFF test, ZC test, F-max test, GPF test, F-type test, HR test, CAFB test and GS test in the Gaussian and t processes under covariance impurity (nominal level 5%).

P_0	(n_1, n_2, n_3)	SS	CFF	ZC	F-max	GPF	F-type	HR	CAFB	GS
SBM (impurity)	$(20, 20, 20)$	0.046	0.047	0.046	0.045	0.046	0.043	0.049	0.027	0.045
SBM (impurity)	$(4, 4, 4)$	0.058	0.034	0.056	0.012	0.026	0.000	0.198	0.032	0.037
$t_{1,\text{SBM}}$ (impurity)	$(20, 20, 20)$	0.053	0.019	0.018	0.017	0.019	0.016	0.023	0.039	0.057
$t_{1,\text{SBM}}$ (impurity)	$(4, 4, 4)$	0.054	0.024	0.047	0.008	0.030	0.000	0.172	0.030	0.045
$t_{3,\text{SBM}}$ (impurity)	$(20, 20, 20)$	0.054	0.039	0.039	0.032	0.041	0.038	0.052	0.028	0.048
$t_{3,\text{SBM}}$ (impurity)	$(4, 4, 4)$	0.043	0.031	0.042	0.014	0.038	0.000	0.196	0.027	0.038

FIG 3. *SS test $(-)$, CFF test $(++)$, ZC test $(\cdots)$, F-max test $(--)$, GPF test $(\times\times)$, F-type test $(\triangle)$, HR test $(\circ)$, CAFB test $(\diamond)$ and GS test $(\triangledown)$ under nominal level 5% in SBM and associated t processes with impurities: 1st row: $n_1 = n_2 = n_3 = 20$, 2nd row: $n_1 = n_2 = n_3 = 4$.*

estimate the sizes and the powers of the tests and present them in Table 2 and Figure 3, respectively.

In Table 2, we note that the estimated sizes of the SS test is always close to the nominal level 5%. The estimated sizes of the GS test is close to the nominal level most of the times. The sizes for the HR test are significantly high for small group sizes, i.e., $n_1 = n_2 = n_3 = 4$, and hence it is omitted in the power comparison for $n_1 = n_2 = n_3 = 4$ in Figure 3. The sizes of the other tests are often considerably lower than the nominal level. In Figure 3, we notice that the power of the SS test is nearly always higher than the powers of all other tests. In comparison to the cases presented in Figure 2, the difference between the performances of the SS test and other tests increases for SBM and $t_{3,\text{SBM}}$ and $n_1 = n_2 = n_3 = 20$.
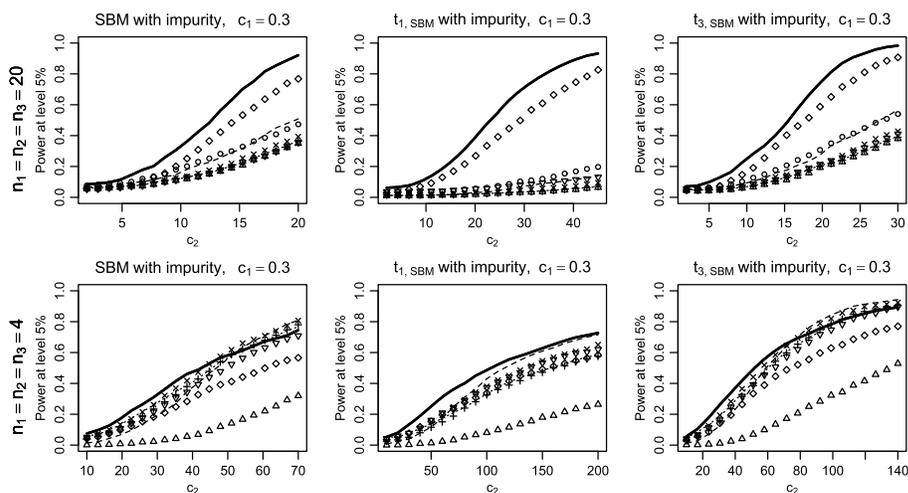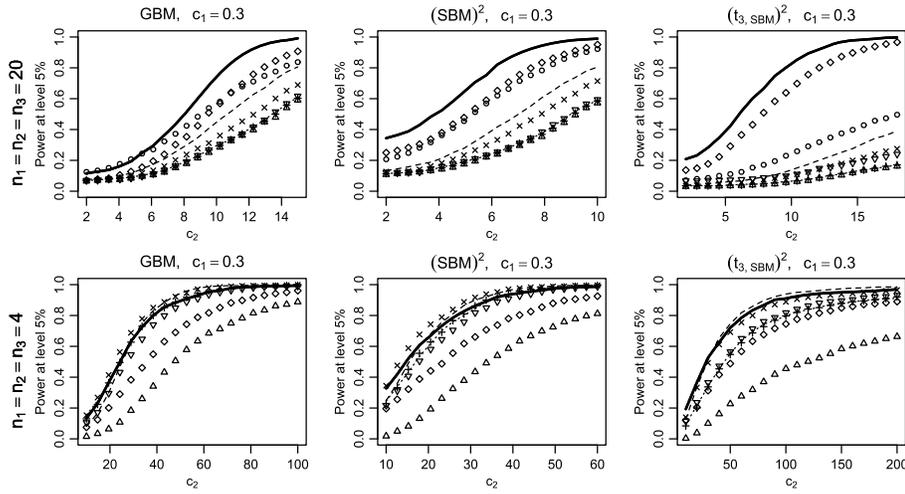
FIG 4. *SS test (−), CFF test (++), ZC test (···), F-max test (− −), GPF test (××), F-type test (△), HR test (○), CAFB test (◇) and GS test (▽) under nominal level 5% in the skewed models:* 1st *row:* $n_1 = n_2 = n_3 = 20$, 2nd *row:* $n_1 = n_2 = n_3 = 4$.

Results for the skewed distributions

We consider three skewed distributions as described before. The third case, which is the squared t process given by $\mathbf{G}_3(t) = (\mathbf{T}(t))^2$, $\mathbf{T}(t)$ being the $t_{3,\text{SBM}}$ process, can also be considered as a skewed and heavy-tailed process. However, the geometric Brownian motion also has heavy tails. We denote the case of P_0 being the geometric Brownian motion as GBM, the case of the squared Brownian motion as $(\text{SBM})^2$ and the case of the squared $t_{3,\text{SBM}}$ process as $(t_{3,\text{SBM}})^2$. The estimated sizes under the nominal level 5% are presented in Table 3. The estimated powers under the nominal level 5% are depicted in Figure 4.

TABLE 3
Estimated sizes of the SS test, CFF test, ZC test, F-max test, GPF test, F-type test, HR test, CAFB test and GS test in the skewed distributions (nominal level 5%).

P_0	(n_1, n_2, n_3)	SS	CFF	ZC	F-max	GPF	F-type	HR	CAFB	GS
GBM	(20, 20, 20)	0.051	0.043	0.044	0.051	0.048	0.038	0.093	0.029	0.046
GBM	(4, 4, 4)	0.047	0.076	0.092	0.048	0.082	0.007	0.281	0.035	0.054
$(\text{SBM})^2$	(20, 20, 20)	0.055	0.035	0.034	0.041	0.037	0.031	0.106	0.034	0.035
$(\text{SBM})^2$	(4, 4, 4)	0.047	0.061	0.070	0.049	0.068	0.002	0.361	0.029	0.044
$(t_{3,SBM})^2$	(20, 20, 20)	0.051	0.020	0.022	0.031	0.030	0.022	0.049	0.030	0.049
$(t_{3,SBM})^2$	(4, 4, 4)	0.044	0.036	0.041	0.037	0.041	0.000	0.285	0.026	0.043

In Table 3, the estimated sizes of the SS test is again found to be always very close to the nominal level 5%. The estimated sizes of the GS test is also close to the nominal level. The estimated sizes for the HR test are significantly high for small group sizes, i.e., $n_1 = n_2 = n_3 = 4$, and so we omit it in the power comparison for $n_1 = n_2 = n_3 = 4$ in Figure 4. In Figure 4, it can be seen that

the power of the SS test is almost uniformly higher than the powers of all other tests for higher sample sizes, i.e., $n_1 = n_2 = n_3 = 20$. But for the lower sample sizes $n_1 = n_2 = n_3 = 4$ also, the power of the SS test is either the highest or very close to the highest among the tests.

Comparison of the asymptotic and permutation implementations

We compare the sizes of the asymptotic and the permutation implementations of the SS test for 3 groups with $n_1 = n_2 = n_3 = 5, 10, 20$ and 40, and compare their powers with $n_1 = n_2 = n_3 = 10$ for the distributions considered previously in this subsection. The number of random permutations taken is 1000, and the sizes and the powers are estimated based on 1000 independent replications with nominal level 5%. All the results pertaining to this investigation are presented in Appendix C. We notice there that the powers of the two implementations are very close to each other. However, when $n_1 = n_2 = n_3 = 5$, the size of the asymptotic implementation is sometimes somewhat smaller than the nominal level. The estimated sizes for the permutation implementation are always close to the nominal level. These findings indicate that the permutation implementation has power similar to the asymptotic implementation, while the asymptotic implementation may be slightly more conservative when the group sizes are small. This points to the utility of the permutation implementation for small group sizes. We suggest using the permutation implementation when the minimum group size is less than or equal to 5 and the total sample size is less than or equal to 20.

The permutation implementation is found to be computationally expensive compared to the asymptotic implementation. We found that when $n_1 = n_2 = n_3 = 5$, the computation of a single p-value using the permutation implementation takes on average nearly twice the time taken by the asymptotic implementation. When $n_1 = n_2 = n_3 = 10, 20$ and 40, the factors are on average around 7, 22 and 60, respectively.

The code for the SS test is written in R and the numerically heavy portions are implemented using the Rcpp package.

For univariate, multivariate and high-dimensional data, several authors investigated and demonstrated the utility of Studentizing the bootstrap or the permutation test statistic (see, e.g., [24], [16] and [32]). In [16], it was demonstrated that Studentization leads to better performance of the test particularly when the underlying distribution is non-Gaussian or heavy-tailed. However, in the setup considered in this paper, Studentization of the permutation or the bootstrap procedures would significantly increase the computational complexity of the testing method due to the involvement of the covariance operator, the calculation of which in our case is found to be quite expensive computationally. Further, the performance of our testing procedures is found to be considerably less affected by the non-Gaussian tails compared to other tests in the literature. We investigated the performance of the Studentized test statistic in all the probability models described in this subsection, and found that despite taking

considerably more time to yield output, the Studentized test shows no significant gain in estimated power over the usual non-Studentized test. In fact, we found that the estimated powers and sizes of the non-Studentized test are very close to the those of the Studentized test. Based on these findings, we do not recommend using Studentization with our procedures.

4. Analysis of real data

We compare the performance of the tests in three real datasets: the Canadian Temperature data, the Tecator data and the Orthosis data.

The Canadian Temperature dataset, analyzed in [36], contains the daily temperature records over 365 days at 35 Canadian weather stations situated in three regions: 15 in Eastern Canada, 15 in Western Canada, and 5 in Northern Canada. One may be interested to know whether there is a statistically significant effect of location on the temperature curves over theses three regions. This leads to an ANOVA problem with three groups with a total of 35 observations, where each observation is a daily temperature curve, and the three groups are the three regions. Since the sizes of the groups are small, we apply the permutation implementation of the SS test, and the other tests are implemented based on the advices of their authors.

The Tecator dataset, available in the 'caret' package in R, contains the percentage values of moisture, fat and protein contents of 215 meat samples along with their absorbance spectra in the wavelength range 850–1050 nm. Moisture, fat and protein contents were measured by analytical chemistry, while a Tecator Infratec Food and Feed Analyzer was used to record the absorbance spectrum. We divide the observations in three groups by the value of the protein content such that the 33% of the observations with the lowest protein content are in the first group, the 33% with the highest protein content are in the third group and the rest in the second group. The group sizes are found to be 72, 69 and 74. We are interested to know whether their is a significant effect of the protein content over the absorption spectra curves, and we get an ANOVA problem with three groups.

The Orthosis dataset, analyzed in [39], contains orthotic measurements on a volunteer under four experimental conditions. For each condition, 10 observations were recorded, and each observation depicts the moments of force over a fixed time duration at the knee of the volunteer. Each observation is recorded at a fixed grid of length 256, and we consider them as functional observations. We are interested in whether the experimental conditions have same effect, and consequently get an ANOVA problem with 4 groups with 10 observations in each group. On visual inspection, we found that the observations corresponding to one of the conditions are far away from the other observations, and so we discard that group from our analysis. Hence, we get an ANOVA problem with 3 groups with 10 observations in each group.

The p-values of all the tests in each of the 3 datasets are presented in Table 4. We notice that all the tests exhibit very small p-values in every dataset.

TABLE 4
P-values of the tests in the real datasets.

Data	SS	CFF	ZC	F-max	GPF	F-type	HR	CAFB	GS
Canada	0	0	0	0	0	0	0	0	0
Tecator	0	0.001	0.002	0.001	0.002	0.002	0.002	0	0.002
Orthosis	0	0	0	0	0	0	0	0	0

TABLE 5
Estimated sizes and powers of the tests in the size and power study in real datasets under nominal level 5%. In Canadian Temperature data, $n_1 = n_2 = n_3 = 4$, in Tecator data, $n_1 = n_2 = n_3 = 23$, and in Orthosis data, $n_1 = n_2 = 3$ and $n_3 = 4$.

Data	Size/Power	SS	CFF	ZC	F-max	GPF	F-type	HR	CAFB	GS
Canada	Size	0.055	0.069	0.071	0.038	0.073	0.009	0.303	0.032	0.051
Canada	Power	0.997	0.993	0.993	0.994	0.983	0.741	1	0.921	0.984
Tecator	Size	0.053	0.047	0.049	0.049	0.048	0.042	0.067	0.033	0.041
Tecator	Power	0.501	0.339	0.347	0.498	0.326	0.319	0.359	0.476	0.322
Orthosis	Size	0.041	0.028	0.094	0.020	0.104	0	0.138	0.034	0.050
Orthosis	Power	0.535	0.448	0.613	0.076	0.678	0.006	0.573	0.200	0.539

Since the p-values of all the tests in all the datasets are found to be statistically significant, we carry out a size and power study of the tests based on subsampling in each of the datasets. To estimate the sizes of the tests, we divide each of the K groups randomly in K subgroups, carry out the tests on those subgroups and then take the mean of the rejection rates within the groups as an estimate of the size in this subsampling process. If a group has too few observations to draw K distinct random subgroups from it, we omit it from the size study. To estimate the powers of the tests in a dataset, we draw the K random subgroups from the K groups separately, carry out the tests. The nominal level is 5% everywhere. The sizes and the powers of the tests are estimated based on 1000 independent replications of the above procedure. Similar procedures of size and power study in real datasets were described in [10] and [20]. When the number of observations in the subsamples are small, we implement the SS test based on the permutation procedure with 1000 random permutations.

In the Canadian Temperature data, since the third group has only 5 observations, it is omitted from the size study in subsampling but included in the power study. The size of each of the 3 subgroups in this dataset is fixed at 4. Due to the small size of the subgroups, the SS test is implemented based on the permutation procedures. In Table 5, we present the estimated sizes and the estimated powers of the tests in the size and power study. The nominal level is 5% everywhere. The SS test is implemented based on the permutation procedure in those two datasets. In the Tecator dataset, the tests are implemented using their usual asymptotic procedure.

In the power study of the tests in the Canadian Temperature data, we found that the estimated powers of all the tests except the F-type test and the CAFB test are very close to 1. In the corresponding size study, we found that the estimated size of the HR test is very far from the nominal level. On the other hand, the estimated size of the F-type test is rather low compared to the nominal

level. In the size study in the Tecator data, we found that all the tests have their estimated sizes close to the nominal level. In the corresponding power study, the SS test demonstrates the highest power, though the F-max test and the CAFB test have powers close to that of the SS test. In the power study in the Orthosis data, we see that the ZC test, the GPF test, the HR test and the GS test exhibit higher estimated powers than the SS test. However, the estimated sizes of the ZC test, the GPF test and the HR test in the corresponding size study are found to be significantly higher than the nominal level, which indicates that their powers may be overestimated. The GS test has estimated size very close to the nominal level and slightly higher power than the SS test in this dataset.

In the case of the Tecator data, the second derivatives of the curves may also be considered as observations and a similar size and power study may be carried out. However, when we carried out this study, we found that all the p-values of the tests are 0, and the all the powers to be 1.

5. Concluding remarks

In this paper, we have developed a test for multisample comparison using spatial signs. The sample is assumed to consist of Hilbert space valued random elements. We described three implementations for this test: one is based on the asymptotic distribution of the test statistic, one is based on a bootstrap procedure and the other is based on a permutation procedure. The asymptotic implementation is faster, and it is recommended for moderate to large sample sizes. But for small sample sizes, its size tends to be lower than the nominal level. We observed the bootstrap and the permutation procedures to behave almost identically, and so choose the permutation procedure for the implementation in suitable scenarios. The permutation procedure, though computationally costly than the asymptotic procedure, exhibits satisfactory performance under the null hypothesis irrespective of the sample size unlike the asymptotic procedure. Further, its power is very similar to the asymptotic implementation. So, the permutation procedure is recommended for small sample sizes.

In our simulation study, we found that the test exhibits substantially superior performance than other mean based ANOVA for functional data in the literature when the underlying distribution is non-Gaussian with heavy tails or skewed. In the Gaussian models considered by us, we found that our test beats most of the other mean based tests, if not all, and the performance of our test is almost always very close to the best among the mean based tests. Our test also exhibits better performance in contaminated models than the other tests.

The above discussion implies the advantage of using our test in a general situation, where it is unknown whether the data are truly Gaussian and without any contamination.

Appendix A: Description of the other tests and asymptotic powers

The test statistic of the CFF test ([15]) is $CFF_n = \sum_{k<l} n_k \left\| \bar{\mathbf{X}}_{k\cdot} - \bar{\mathbf{X}}_{l\cdot} \right\|^2$. We have used the implementation of the CFF test suggested in [15] for homoscedastic samples because all the simulation models we consider are homoscedastic and the real data we consider do not exhibit heteroscedasticity, and we take 1000 bootstrap samples for its implementation.

The test statistic of the ZC test ([37]) is $ZC_n = \sum_{k=1}^{K} n_k \left\| \bar{\mathbf{X}}_{k\cdot} - \bar{\mathbf{X}}_{\cdot\cdot} \right\|^2$, and the test statistic of the F-type test ([33, 35]) is

$$F_n = \frac{(K-1)^{-1} \sum_{k=1}^{K} n_k \left\| \bar{\mathbf{X}}_{k\cdot} - \bar{\mathbf{X}}_{\cdot\cdot} \right\|^2}{(n-K)^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left\| \mathbf{X}_{ki} - \bar{\mathbf{X}}_{k\cdot} \right\|^2}.$$

The parameter estimation for implementing the ZC test and the F-type test for moderate or large sample sizes was done following the so-called 'naive method' (see subsection 2.1 in [18]) instead of the 'bias-reduced method', because we did not see any significant difference in performances for those two implementations, and the 'naive method' involves less computation. For small samples sizes, these two tests are implemented based on 1000 bootstrap samples using the R package 'fdANOVA' (see [19]).

The test statistic of the GPF test ([39]) is

$$FG_n = \int_{\mathcal{T}} \frac{(K-1)^{-1} \sum_{k=1}^{K} n_k \left(\bar{\mathbf{X}}_{k\cdot}(t) - \bar{\mathbf{X}}_{\cdot\cdot}(t) \right)^2}{(n-K)^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(\mathbf{X}_{ki}(t) - \bar{\mathbf{X}}_{k\cdot}(t) \right)^2} \mathrm{d}t,$$

and the test statistic of the F-max test ([38]) is

$$FM_n = \sup_{t \in [a,b]} \frac{(K-1)^{-1} \sum_{k=1}^{K} n_k \left(\bar{\mathbf{X}}_{k\cdot}(t) - \bar{\mathbf{X}}_{\cdot\cdot}(t) \right)^2}{(n-K)^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(\mathbf{X}_{ki}(t) - \bar{\mathbf{X}}_{k\cdot}(t) \right)^2}.$$

The GPF test and the F-max test are implemented according to the advice of the respective authors, and the number of bootstrap samples used for the implementation of the F-max test is 1000.

The GS test ([18]) is a permutation test of ANOVA based on the basis function representation of the functional observations, where the test statistic is similar to that of the F-type test. The GS test is implemented based on the function in the R package 'fdANOVA' using the Fourier basis and the BIC criterion and 1000 permutations.

The CAFB test ([14]) is based on random projections of the sample. To implement the test, we generate 30 random directions following standard Brownian motion on which the sample observations are projected. Then, we carry out the univariate ANOVA procedure described in section 3 in [5] on the projected observations as suggested in [14]. The test of ANOVA for the functional observations is then decided based on the p-values of the univariate ANOVA. The reader is referred to [14] for further details on the implementation of this test, including the p-value computation.

We need to define a few quantities to describe the test statistic of the HR test ([23]), which is based on d principal component scores of the sample in the following way. Define $\mathbf{\Omega}_{nk} : [a,b] \times [a,b] \to \mathbb{R}$ as

$$\mathbf{\Omega}_{nk}(s,t) = n_k^{-1} \sum_{i=1}^{n_k} \left(\mathbf{X}_{ki}(s) - \bar{\mathbf{X}}_{k\cdot}(s)\right)\left(\mathbf{X}_{ki}(t) - \bar{\mathbf{X}}_{k\cdot}(t)\right)$$

for $k = 1, \ldots, K$. Let $\mathbf{\Omega}_n = n^{-1}\sum_{k=1}^{K} n_k \mathbf{\Omega}_{nk}$ and $\phi_{n1}, \ldots, \phi_{nd}$ be the eigenfunctions corresponding to the d largest eigenvalues of $\mathbf{\Omega}_n$. Define

$$\bar{\boldsymbol{\xi}}_{ki} = \left(\langle \mathbf{X}_{ki}, \phi_{n1}\rangle, \ldots, \langle \mathbf{X}_{ki}, \phi_{nd}\rangle\right)^t, \quad \bar{\boldsymbol{\xi}}_{k\cdot} = n_k^{-1}\sum_{i=1}^{n_k} \bar{\boldsymbol{\xi}}_{ki},$$

$$\boldsymbol{\Psi}_{nk} = n_k^{-1}\sum_{i=1}^{n_k}\left(\bar{\boldsymbol{\xi}}_{ki} - \bar{\boldsymbol{\xi}}_{k\cdot}\right)\left(\bar{\boldsymbol{\xi}}_{ki} - \bar{\boldsymbol{\xi}}_{k\cdot}\right)^t, \quad \bar{\boldsymbol{\xi}}_{\cdot\cdot} = \left(\sum_{k=1}^{K} n_k \boldsymbol{\Psi}_{nk}^{-1}\right)^{-1}\sum_{k=1}^{K} n_k \boldsymbol{\Psi}_{nk}^{-1}\bar{\boldsymbol{\xi}}_{k\cdot}.$$

The test statistic of the HR test is defined as

$$HR_n = \sum_{k=1}^{K} n_k \left(\bar{\boldsymbol{\xi}}_{k\cdot} - \bar{\boldsymbol{\xi}}_{\cdot\cdot}\right)^t \boldsymbol{\Psi}_{nk}^{-1}\left(\bar{\boldsymbol{\xi}}_{k\cdot} - \bar{\boldsymbol{\xi}}_{\cdot\cdot}\right).$$

In [23], there was no clear direction on the number of principal components to be used for implementing the HR test. So, we have taken the number so that the chosen principal components explain at least 90% of the variance present in the sample. In other words, we choose the number d such that the sum of the d largest eigenvalues of the sample covariance operator $\mathbf{\Omega}_n$ is at least 90% of the sum of all its eigenvalues.

We derive here the asymptotic powers of the CFF test, the ZC test and the HR test under the class of shrinking alternatives described in (3.1), where $\mathcal{H} = L_2[a,b]$.

For $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)^t$, $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_p)^t$, where $\mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathcal{H}$, define $\mathbf{u}^t\mathbf{v}$ as $\mathbf{u}^t\mathbf{v} = \sum_{i=1}^{p}\langle \mathbf{u}_i, \mathbf{v}_i\rangle$. Matrix operations between elements in $\mathcal{H}$ or scalars are defined analogously. For the result on the HR test, we need to further define the following quantities. Define $\mathbf{\Omega}_k, \mathbf{\Omega} : [a,b] \times [a,b] \to \mathbb{R}$ as $\mathbf{\Omega}_k(s,t) = E\left[(\mathbf{X}_{k1}(s) - E\left[\mathbf{X}_{k1}(s)\right])(\mathbf{X}_{k1}(t) - E\left[\mathbf{X}_{k1}(t)\right])\right]$ and $\mathbf{\Omega} = \sum_{k=1}^{K} \lambda_k \mathbf{\Omega}_k$ for $k = 1, \ldots, K$. Let $\phi_1, \ldots, \phi_d$ be the eigenfunctions corresponding to the d largest eigenvalues of $\mathbf{\Omega}$. Define $\mathbf{d}_k = (\langle \boldsymbol{\delta}_k, \phi_1\rangle, \ldots, \langle \boldsymbol{\delta}_k, \phi_d\rangle)^t$, $\boldsymbol{\xi}_{ki} = (\langle \mathbf{X}_{ki}, \phi_1\rangle, \ldots, \langle \mathbf{X}_{ki}, \phi_d\rangle)^t$, $\boldsymbol{\Psi}_k = E\left[(\boldsymbol{\xi}_{k1} - E\left[\boldsymbol{\xi}_{k1}\right])(\boldsymbol{\xi}_{k1} - E\left[\boldsymbol{\xi}_{k1}\right])^t\right]$ and

$$\mathbf{A}_{Kd \times Kd}$$

$$= \mathbb{I}_{Kd} - \begin{bmatrix} \sqrt{\lambda_1}\boldsymbol{\Psi}_1^{-1/2} \\ \vdots \\ \sqrt{\lambda_K}\boldsymbol{\Psi}_K^{-1/2} \end{bmatrix} \left(\sum_{k=1}^{K} \lambda_k \boldsymbol{\Psi}_k^{-1}\right)^{-1} \left[\sqrt{\lambda_1}\boldsymbol{\Psi}_1^{-1/2}, \ldots, \sqrt{\lambda_K}\boldsymbol{\Psi}_K^{-1/2}\right].$$

One can verify that $\mathbf{A}_{Kd\times Kd}$ is an idempotent matrix of degree $d(K-1)$. Let $\mathbf{V}_{Kd\times Kd}$ be such that

$$\mathbf{A}_{Kd\times Kd} = \mathbf{V}_{Kd\times Kd} \begin{bmatrix} \mathbb{I}_{d(K-1)} & \mathbf{0}_{d(K-1)\times d} \\ \mathbf{0}_{d\times d(K-1)} & \mathbf{0}_{d\times d} \end{bmatrix} \mathbf{V}^t_{Kd\times Kd}.$$

Define $\mathbf{M}_{Kd\times 1} = \mathbf{V}^t_{Kd} \left[\sqrt{\lambda_1}\boldsymbol{\Psi}_1^{-1/2}\mathbf{d}_1^t, \ldots, \sqrt{\lambda_K}\boldsymbol{\Psi}_K^{-1/2}\mathbf{d}_K^t \right]^t$, and let $\widetilde{\mathbf{M}}_{d(K-1)\times 1}$ be the truncated column vector obtained by dropping the bottom d elements from $\mathbf{M}_{Kd\times 1}$. We have the following theorem on the asymptotic powers of the three tests.

Theorem A.1. *Consider the class of shrinking alternatives described in* (3.1), *and assume* $E[\|\mathbf{X}\|^2] < \infty$, *where* $\mathbf{X}$ *has distribution* P_0, *and that* $n^{-1}n_k \to \lambda_k \in (0,1)$ *for all* k *as* $n \to \infty$. *Let* $\mathbf{Y}_1, \ldots, \mathbf{Y}_K$ *be independent zero-mean Gaussian random elements with the covariance operator of* P_0 *denoted as* $\boldsymbol{\Gamma}$. *For* $\alpha \in (0,1)$, *let* $CFF(\alpha)$, $ZC(\alpha)$ *and* $HR(\alpha)$ *denote the asymptotic powers at level* α *of the CFF test, the ZC test, and the HR test, respectively, under* (3.1) *with* $n \to \infty$. *Then, we have the following.*

(a) *$CFF(\alpha) = P[\sum_{k<l}\lambda_k\|\lambda_k\boldsymbol{\delta}_k + \mathbf{Y}_k - \sqrt{\lambda_k/\lambda_l}(\lambda_l\boldsymbol{\delta}_l + \mathbf{Y}_l)\|^2 \geq Q_1(1-\alpha)]$, where $Q_1(1-\alpha)$ is the $(1-\alpha)$-quantile of the distribution of the random variable $\sum_{k<l}\lambda_k\|\mathbf{Y}_k - \sqrt{\lambda_k/\lambda_l}\mathbf{Y}_l\|^2$.*

(b) *Let $\gamma_1, \gamma_2, \ldots$ be the sequence of eigenvalues of $\boldsymbol{\Gamma}$, and $\boldsymbol{\chi}_1^2, \boldsymbol{\chi}_2^2, \ldots$ be an infinite sequence of independent χ^2_{K-1} random variables. Further, let $\mathbf{I}_K$ denote a $K \times K$ matrix of identity operators, $\mathbf{p}_0 = \left(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K}\right)^t$, and $\mathbf{Z} = \left(\sqrt{\lambda_1}\boldsymbol{\delta}_1 + \mathbf{Y}_1, \ldots, \sqrt{\lambda_K}\boldsymbol{\delta}_K + \mathbf{Y}_K\right)^t$. Then, $ZC(\alpha) = P[\mathbf{Z}^t(\mathbf{I}_K - \mathbf{p}_0\mathbf{p}_0^t)\mathbf{Z} \geq Q_2(1-\alpha)]$, where $Q_2(1-\alpha)$ is the $(1-\alpha)$-quantile of the distribution of the random variable $\sum_{i=1}^{\infty}\gamma_i\boldsymbol{\chi}_i^2$.*

(c) *Let $\boldsymbol{\chi}$ be a random variable following a non-central chi-square distribution with $((K-1)d)$ degrees of freedom and $\left\|\widetilde{\mathbf{M}}_{d(K-1)\times 1}\right\|^2$ being its noncentrality parameter. Then, $HR(\alpha) = P[\boldsymbol{\chi} \geq Q_3(1-\alpha)]$, where $Q_3(1-\alpha)$ is the $(1-\alpha)$-quantile of the central chi-square distribution with $((K-1)d)$ degrees of freedom.*

Appendix B: Proof of mathematical results

Proof of Theorem 2.1. Define $\tilde{\mathbf{s}}(\mathbf{X}, \mathbf{Y}) = \mathbf{s}(\mathbf{X} - \mathbf{Y}) - E[\mathbf{s}(\mathbf{X} - \mathbf{Y})]$, where $\mathbf{X}$ and $\mathbf{Y}$ are independent random elements. Define

$$\mathbf{S}_k = \frac{1}{n}\sum_{l=1}^{K}n_l \left[\frac{1}{n_k}\sum_{i_k=1}^{n_k}E\left[\tilde{\mathbf{s}}(\mathbf{X}_{ki_k}, \mathbf{X}_{l1}) \mid \mathbf{X}_{ki_k}\right] + \frac{1}{n_l}\sum_{i_l=1}^{n_l}E\left[\tilde{\mathbf{s}}(\mathbf{X}_{k1}, \mathbf{X}_{li_l}) \mid \mathbf{X}_{li_l}\right] \right],$$

$$\mathbf{V}_{kl}$$
$$= \frac{1}{n_k n_l}\sum_{i_k=1}^{n_k}\sum_{i_l=1}^{n_l}\left[\tilde{\mathbf{s}}(\mathbf{X}_{ki_k}, \mathbf{X}_{li_l}) - E[\tilde{\mathbf{s}}(\mathbf{X}_{ki_k}, \mathbf{X}_{l1}) \mid \mathbf{X}_{ki_k}] - E[\tilde{\mathbf{s}}(\mathbf{X}_{k1}, \mathbf{X}_{li_l}) \mid \mathbf{X}_{li_l}]\right],$$

where $i_k = 1, \ldots, n_k$, $i_l = 1, \ldots, n_l$ and $k, l = 1, \ldots, K$. We have $\bar{\mathbf{R}}_k - E\left[\bar{\mathbf{R}}_k\right] = \mathbf{S}_k + n^{-1} \sum_{l=1}^{K} n_l \mathbf{V}_{kl}$. Also, define $\mathbf{W}_n = \left(\sqrt{n_1}\mathbf{S}_1, \ldots, \sqrt{n_K}\mathbf{S}_K\right)$, and $\bar{\mathbf{s}}(\mathbf{X}, \mathbf{Y}) = \tilde{\mathbf{s}}(\mathbf{X}, \mathbf{Y}) - E\left[\tilde{\mathbf{s}}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X}\right] - E\left[\tilde{\mathbf{s}}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y}\right]$. Note that

$$E\left[\bar{\mathbf{s}}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X}\right] = E\left[\bar{\mathbf{s}}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y}\right] = E\left[\bar{\mathbf{s}}(\mathbf{X}, \mathbf{Y})\right] = \mathbf{0} \qquad (\text{B.1})$$

$$\text{and } \|\bar{\mathbf{s}}(\mathbf{X}, \mathbf{Y})\| \leq 6. \qquad (\text{B.2})$$

From (B.1), we get that for every case other than $(k, i_k, l_1, i_{l_1}) = (k, j_k, l_2, j_{l_2})$ and $(k, i_k, l_1, i_{l_1}) = (l_2, j_{l_2}, k, j_k)$,

$$E\left[\left\langle \bar{\mathbf{s}}\left(\mathbf{X}_{ki_k}, \mathbf{X}_{l_1 i_{l_1}}\right), \bar{\mathbf{s}}\left(\mathbf{X}_{kj_k}, \mathbf{X}_{l_2 j_{l_2}}\right)\right\rangle\right] = 0. \qquad (\text{B.3})$$

Since $\mathbf{V}_{kl} = (n_k n_l)^{-1} \sum_{i_k=1}^{n_k} \sum_{i_l=1}^{n_l} \bar{\mathbf{s}}(\mathbf{X}_{ki_k}, \mathbf{X}_{li_l})$, from (B.2) and (B.3) we get that for all k,

$$E\left[\left\|\sqrt{n_k}\frac{1}{n}\sum_{l=1}^{K} n_l \mathbf{V}_{kl}\right\|^2\right] \longrightarrow 0 \text{ as } n \to \infty,$$

$$\text{which implies } \|[\mathbf{U}_n - E[\mathbf{U}_n]] - \mathbf{W}_n\| = \sqrt{\sum_{k=1}^{K}\left\|\sqrt{n_k}\frac{1}{n}\sum_{l=1}^{K} n_l \mathbf{V}_{kl}\right\|^2} \xrightarrow{P} 0 \qquad (\text{B.4})$$

as $n \to \infty$. Next, we consider $\mathbf{S}_k$. It can be verified that

$$\text{Cov}\left(\sqrt{n_{k_1}}\mathbf{S}_{k_1}, \sqrt{n_{k_2}}\mathbf{S}_{k_2}\right)$$

$$= \frac{\sqrt{n_{k_1} n_{k_2}}}{n} \sum_{l=1}^{K} \frac{n_l}{n}\left[\mathbf{C}(k_1, k_2, l) - \mathbf{C}(k_1, l, k_2) - \mathbf{C}(l, k_2, k_1)\right]$$

$$+ \sum_{l_1=1}^{K}\sum_{l_2=1}^{K} \frac{n_{l_1} n_{l_2}}{n^2}\mathbf{C}(l_1, l_2, k_1)\mathbb{I}(k_1 = k_2). \qquad (\text{B.5})$$

Since $n^{-1}n_k \to \lambda_k \in (0, 1)$ for all k as $n \to \infty$, from (B.5), we have for all k_1 and k_2,

$$\text{Cov}\left(\sqrt{n_{k_1}}\mathbf{S}_{k_1}, \sqrt{n_{k_2}}\mathbf{S}_{k_2}\right) \longrightarrow \boldsymbol{\sigma}_{k_1 k_2} \quad \text{as } n \to \infty. \qquad (\text{B.6})$$

Since $E[\mathbf{W}_n] = \mathbf{0}$, from (B.6) and an application of Theorem 1.1 in [28], we get $\mathbf{W}_n \xrightarrow{w} G(\mathbf{0}, \boldsymbol{\Sigma})$ as $n \to \infty$, which, along with (B.4), implies that $[\mathbf{U}_n - E[\mathbf{U}_n]] \xrightarrow{w} G(\mathbf{0}, \boldsymbol{\Sigma})$ as $n \to \infty$. $\qquad \square$

Proof of Corollary 2.2. When H_0 in (2.1) is true, $\mathbf{X}_{k1}$ is an independent copy of $\mathbf{X}_{l1}$ for $k \neq l$, which, along with the fact that $\mathbf{s}(\mathbf{0}) = \mathbf{0}$ by definition, gives $E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})] = \mathbf{0}$ for all k, l. So,

$$E\left[\sqrt{n_k}\bar{\mathbf{R}}_k\right] = \sqrt{n_k}\sum_{l=1}^{K} \frac{n_l}{n}E\left[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})\right] = \mathbf{0}$$

for all k, and consequently, $E[\mathbf{U}_n] = \mathbf{0}$. Therefore from Theorem 2.1, we have $\mathbf{U}_n \xrightarrow{w} G(\mathbf{0}, \boldsymbol{\Sigma})$ as $n \to \infty$. Hence, from an application of the mapping theorem [3, Theorem 2.7, p. 21], we get $\|\mathbf{U}_n\|^2 \xrightarrow{w} \|\mathbf{W}\|^2$ as $n \to \infty$, where $\mathbf{W}$ is a random element having distribution $G(\mathbf{0}, \boldsymbol{\Sigma})$. $\square$

Proof of Theorem 2.3. Note that the class of compact operators on a separable Hilbert space is closed [2, p. 164] and separable [2, p. 168] with respect to the operator norm. So, the class of compact operators on a separable Hilbert space is a separable Banach space with respect to the operator norm, and we denote it as $\mathcal{C}$. We have

$$\mathbf{C}_n(k_1, k_2, l) = \frac{n_l}{n_l - 1} \mathbf{C}_n^{(1)}(k_1, k_2, l) - \mathbf{C}_n^{(2)}(k_1, l) \otimes \mathbf{C}_n^{(2)}(k_2, l), \tag{B.7}$$

where $\mathbf{C}_n^{(1)}(j, k, l) = \dfrac{1}{n_l n_j n_k} \sum\limits_{i_l=1}^{n_l} \sum\limits_{i_j=1}^{n_j} \sum\limits_{i_k=1}^{n_k} \mathbf{s}\left(\mathbf{X}_{ji_j} - \mathbf{X}_{li_l}\right) \otimes \mathbf{s}\left(\mathbf{X}_{ki_k} - \mathbf{X}_{li_l}\right)$

and $\mathbf{C}_n^{(2)}(k, l) = \dfrac{1}{n_k n_l} \sum\limits_{i_k=1}^{n_k} \sum\limits_{i_l=1}^{n_l} \mathbf{s}\left(\mathbf{X}_{ki_k} - \mathbf{X}_{li_l}\right)$ for $j, k, l \in \{1, \ldots, K\}$.

First, we shall consider the term $\mathbf{C}_n^{(1)}(k_1, k_2, l)$ and the following cases of the triplet (k_1, k_2, l) separately: (a) k_1, k_2, l are all distinct, (b) $k_1 = k_2 \neq l$, (c) $k_1 = l \neq k_2$, (d) $k_2 = l \neq k_1$ and (e) $k_1 = k_2 = l$. These five cases cover all possible cases of (k_1, k_2, l).

Assume that k_1, k_2, l are all distinct. Define $\boldsymbol{\Phi}_1(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{s}(\mathbf{x} - \mathbf{z}) \otimes \mathbf{s}(\mathbf{y} - \mathbf{z})$ for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{H}$. Since $\boldsymbol{\Phi}_1(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is a finite-rank operator, it belongs to the separable Banach space $\mathcal{C}$. Therefore,

$$\mathbf{C}_n^{(1)}(k_1, k_2, l) = \frac{1}{n_l n_{k_1} n_{k_2}} \sum_{i_l=1}^{n_l} \sum_{i_{k_1}=1}^{n_{k_1}} \sum_{i_{k_2}=1}^{n_{k_2}} \boldsymbol{\Phi}_1\left(\mathbf{X}_{k_1 i_{k_1}}, \mathbf{X}_{k_2 i_{k_2}}, \mathbf{X}_{li_l}\right)$$

is a 3-sample UB-statistic [4, p. 15]. Since $\|\boldsymbol{\Phi}_1\| \leq 1$ and $n^{-1} n_k \to \lambda_k \in (0, 1)$ for all k as $n \to \infty$, from Theorem 3.2.1 in [4, p. 79], we get

$$\mathbf{C}_n^{(1)}(k_1, k_2, l) \xrightarrow{a.s.} E\left[\mathbf{s}\left(\mathbf{X}_{k_1 1} - \mathbf{X}_{l1}\right) \otimes \mathbf{s}\left(\mathbf{X}_{k_2 1} - \mathbf{X}_{l1}\right)\right] \tag{B.8}$$

as $n \to \infty$ in the operator norm. Next, assume $k_1 = k_2 \neq l$. Define $\boldsymbol{\Phi}_2(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = 2^{-1}[\mathbf{s}(\mathbf{x}_1 - \mathbf{y}) \otimes \mathbf{s}(\mathbf{x}_2 - \mathbf{y}) + \mathbf{s}(\mathbf{x}_2 - \mathbf{y}) \otimes \mathbf{s}(\mathbf{x}_1 - \mathbf{y})]$ and $\boldsymbol{\Phi}_3(\mathbf{x}, \mathbf{y}) = \mathbf{s}(\mathbf{x} - \mathbf{y}) \otimes \mathbf{s}(\mathbf{x} - \mathbf{y})$. Clearly, $\boldsymbol{\Phi}_2(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}), \boldsymbol{\Phi}_3(\mathbf{x}, \mathbf{y}) \in \mathcal{C}$, and since $\|\boldsymbol{\Phi}_2\| \leq 1$ and $\|\boldsymbol{\Phi}_3\| \leq 1$, from Theorem 3.2.1 in [4, p. 79], we get

$$\mathbf{C}_n^{(1)}(k_1, k_2, l) = \frac{n_{k_1} - 1}{n_{k_1}} \frac{1}{n_l \binom{n_{k_1}}{2}} \sum_{i_l=1}^{n_l} \sum_{1 \leq i_{k_1} < j_{k_1} \leq n_{k_1}} \boldsymbol{\Phi}_2\left(\mathbf{X}_{k_1 i_{k_1}}, \mathbf{X}_{k_1 j_{k_1}}; \mathbf{X}_{li_l}\right)$$

$$+ \frac{1}{n_{k_1}} \frac{1}{n_l n_{k_1}} \sum_{i_l=1}^{n_l} \sum_{i_{k_1}=1}^{n_{k_1}} \boldsymbol{\Phi}_3\left(\mathbf{X}_{k_1 i_{k_1}}, \mathbf{X}_{li_l}\right)$$

$$\xrightarrow{a.s.} E\left[\mathbf{s}\left(\mathbf{X}_{k_1 1} - \mathbf{X}_{l1}\right) \otimes \mathbf{s}\left(\mathbf{X}_{k_1 2} - \mathbf{X}_{l1}\right)\right] \tag{B.9}$$

as $n \to \infty$ in the operator norm. Next, assume $k_1 = l \neq k_2$. Define $\boldsymbol{\Phi}_4\left(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}\right)$ $= 2^{-1}[\mathbf{s}\left(\mathbf{x}_1 - \mathbf{x}_2\right) \otimes \mathbf{s}\left(\mathbf{y} - \mathbf{x}_2\right) + \mathbf{s}\left(\mathbf{x}_2 - \mathbf{x}_1\right) \otimes \mathbf{s}\left(\mathbf{y} - \mathbf{x}_1\right)]$. Since $\|\boldsymbol{\Phi}_4\| \leq 1$, we get

$$\mathbf{C}_n^{(1)}(k_1, k_2, l) = \frac{n_{k_1} - 1}{n_{k_1}} \frac{1}{\binom{n_{k_1}}{2} n_{k_2}} \sum_{1 \leq i_{k_1} < j_{k_1} \leq n_{k_1}} \sum_{i_{k_2}=1}^{n_{k_2}} \boldsymbol{\Phi}_4\left(\mathbf{X}_{k_1 i_{k_1}}, \mathbf{X}_{k_1 j_{k_1}}; \mathbf{X}_{k_2 i_{k_2}}\right)$$

$$\xrightarrow{a.s.} E\left[\mathbf{s}\left(\mathbf{X}_{k_1 1} - \mathbf{X}_{k_1 2}\right) \otimes \mathbf{s}\left(\mathbf{X}_{k_2 1} - \mathbf{X}_{k_1 2}\right)\right] \tag{B.10}$$

as $n \to \infty$ in the operator norm. Next, assume $k_2 = l \neq k_1$. Arguing similarly as in the case $k_1 = l \neq k_2$, we get

$$\mathbf{C}_n^{(1)}(k_1, k_2, l) \xrightarrow{a.s.} E\left[\mathbf{s}\left(\mathbf{X}_{k_1 1} - \mathbf{X}_{k_2 2}\right) \otimes \mathbf{s}\left(\mathbf{X}_{k_2 1} - \mathbf{X}_{k_2 2}\right)\right] \tag{B.11}$$

as $n \to \infty$ in the operator norm. Finally, assume $k_1 = k_2 = l$. Define

$$\boldsymbol{\Phi}_5\left(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\right) = (3!)^{-1} \sum_{\sigma} \mathbf{s}\left(\mathbf{x}_{\sigma(1)} - \mathbf{x}_{\sigma(3)}\right) \otimes \mathbf{s}\left(\mathbf{x}_{\sigma(2)} - \mathbf{x}_{\sigma(3)}\right),$$

where the sum is carried out over all 6 permutations of $(1, 2, 3)$, and

$$\boldsymbol{\Phi}_6\left(\mathbf{x}_1, \mathbf{x}_2\right) = 2^{-1}\left[\mathbf{s}\left(\mathbf{x}_1 - \mathbf{x}_2\right) \otimes \mathbf{s}\left(\mathbf{x}_1 - \mathbf{x}_2\right) + \mathbf{s}\left(\mathbf{x}_2 - \mathbf{x}_1\right) \otimes \mathbf{s}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right].$$

Since $\|\boldsymbol{\Phi}_5\| \leq 1$ and $\|\boldsymbol{\Phi}_6\| \leq 1$, from Theorem 3.1.1 in [4, p. 73], we get

$$\mathbf{C}_n^{(1)}(k_1, k_2, l)$$
$$= \frac{1}{n_{k_1}^3} \sum_{h_{k_1}=1}^{n_{k_1}} \sum_{i_{k_1}=1}^{n_{k_1}} \sum_{j_{k_1}=1}^{n_{k_1}} \mathbf{s}\left(\mathbf{X}_{k_1 i_{k_1}} - \mathbf{X}_{k_1 h_{k_1}}\right) \otimes \mathbf{s}\left(\mathbf{X}_{k_1 j_{k_1}} - \mathbf{X}_{k_1 h_{k_1}}\right)$$
$$= \frac{(n_{k_1} - 1)(n_{k_1} - 2)}{n_{k_1}^2} \frac{1}{\binom{n_{k_1}}{3}} \sum_{1 \leq h_{k_1} < i_{k_1} < j_{k_1} \leq n_{k_1}} \boldsymbol{\Phi}_5\left(\mathbf{X}_{k_1 i_{k_1}}, \mathbf{X}_{k_1 j_{k_1}}, \mathbf{X}_{k_1 h_{k_1}}\right)$$
$$+ \frac{n_{k_1} - 1}{n_{k_1}^2} \frac{1}{\binom{n_{k_1}}{2}} \sum_{1 \leq i_{k_1} < j_{k_1} \leq n_{k_1}} \boldsymbol{\Phi}_3\left(\mathbf{X}_{k_1 i_{k_1}}, \mathbf{X}_{k_1 j_{k_1}}\right)$$
$$\xrightarrow{a.s.} E\left[\mathbf{s}\left(\mathbf{X}_{k_1 1} - \mathbf{X}_{k_1 3}\right) \otimes \mathbf{s}\left(\mathbf{X}_{k_1 2} - \mathbf{X}_{k_1 3}\right)\right] \tag{B.12}$$

as $n \to \infty$ in the operator norm. Therefore, from (B.8), (B.9), (B.10), (B.11) and (B.12), we get that for all $k_1, k_2, l \in \{1, \ldots, K\}$,

$$\mathbf{C}_n^{(1)}(k_1, k_2, l) \xrightarrow{a.s.} E\left[\mathbf{s}\left(\mathbf{X}_{k_1 1} - \mathbf{X}_{l1}\right) \otimes \mathbf{s}\left(\mathbf{X}_{k_2 1} - \mathbf{X}_{l1}\right)\right]$$
$$= E\left[E\left[\mathbf{s}(\mathbf{X}_{k_1 1} - \mathbf{X}_{l1}) \mid \mathbf{X}_{l1}\right] \otimes E\left[\mathbf{s}(\mathbf{X}_{k_2 1} - \mathbf{X}_{l1}) \mid \mathbf{X}_{l1}\right]\right] \tag{B.13}$$

as $n \to \infty$ in the operator norm, where, if $k_1 = l$ or $k_2 = l$, $\mathbf{X}_{k_1 1}$ or $\mathbf{X}_{k_2 1}$ are considered to be independent copies of $\mathbf{X}_{l1}$, respectively. Next, we consider the

term $\mathbf{C}_n^{(2)}(k,l)$. Note that if $k = l$, $\mathbf{C}_n^{(2)}(k,l) = \mathbf{0}$ and $E\left[\mathbf{s}(\mathbf{X}_{k_11} - \mathbf{X}_{l1})\right] = \mathbf{0}$. When $k \neq l$, $\mathbf{C}_n^{(2)}(k,l)$ is a 2-sample UB statistic, and we have

$$\mathbf{C}_n^{(2)}(k,l) \xrightarrow{a.s.} E\left[\mathbf{s}(\mathbf{X}_{k_11} - \mathbf{X}_{l1})\right] \tag{B.14}$$

as $n \to \infty$ in the operator norm. Therefore, from (B.7), (B.13) and (B.14), we have

$$\mathbf{C}_n(k_1, k_2, l) \xrightarrow{a.s.} \mathbf{C}(k_1, k_2, l). \tag{B.15}$$

as $n \to \infty$ in the operator norm. Since $n^{-1}n_k \to \lambda_k$ for all k as $n \to \infty$, from (B.15) we have $\boldsymbol{\sigma}_{k_1 k_2}^{(n)} \xrightarrow{a.s.} \boldsymbol{\sigma}_{k_1 k_2}$ as $n \to \infty$ in the operator norm for all k_1, k_2, and hence $\widehat{\boldsymbol{\Sigma}}_n \xrightarrow{a.s.} \boldsymbol{\Sigma}$ as $n \to \infty$ in the operator norm. $\qquad\square$

Proof of Theorem 2.4. The proof for the test based on the asymptotic procedure follows from Corollary 2.2.

We now prove the theorem for the bootstrap implementation. Let $\mathcal{S}_{(n_1,\ldots,n_K)}$ denote the original sample $\{\mathbf{X}_{ki} \mid k = 1,\ldots,K; i = 1,\ldots,n_k\}$ and define $\mathcal{S} = \cup\{\mathcal{S}_{(n_1,\ldots,n_K)} \mid n_1 \geq 1, \ldots, n_K \geq 1\}$. Let $\{\mathbf{X}_{ki}^* \mid k = 1,\ldots,K; i = 1,\ldots,n_k\}$ denote a bootstrap sample. Define $\mathbf{R}^*(\mathbf{x}) = n^{-1}\sum_{k=1}^K \sum_{i=1}^{n_k} \mathbf{s}(\mathbf{x} - \mathbf{X}_{ki}^*)$, $\bar{\mathbf{R}}_k^* = n_k^{-1}\sum_{i_k=1}^{n_k} \mathbf{R}^*\left(\mathbf{X}_{ki_k}^*\right)$, $\mathbf{U}_n^* = \left(\sqrt{n_1}\bar{\mathbf{R}}_1^*, \ldots, \sqrt{n_K}\bar{\mathbf{R}}_K^*\right)$ and $SS_n^* = \sum_{k=1}^K n_k \left\|\bar{\mathbf{R}}_k^*\right\|^2 = \|\mathbf{U}_n^*\|^2$. Define $P^*[\cdot] = P[\cdot \mid \mathcal{S}]$, $E^*[\cdot] = E[\cdot \mid \mathcal{S}]$, $\mathrm{Cov}^*[\cdot,\cdot] = \mathrm{Cov}[\cdot,\cdot \mid \mathcal{S}]$ and $E^*[\cdot \mid \mathbf{Z}] = E[\cdot \mid \mathbf{Z}, \mathcal{S}]$ for any random element $\mathbf{Z}$. Also, let "$\xrightarrow{P^*}$" imply convergence in probability given $\mathcal{S}$, and "$\xrightarrow{w^*}$" imply convergence in distribution given $\mathcal{S}$. Define $\tilde{\mathbf{s}}^*\left(\mathbf{X}_{ki_k}^*, \mathbf{X}_{li_l}^*\right) = \mathbf{s}\left(\mathbf{X}_{ki_k}^* - \mathbf{X}_{li_l}^*\right) - E^*\left[\mathbf{s}\left(\mathbf{X}_{ki_k}^* - \mathbf{X}_{li_l}^*\right)\right]$. Note that

$$E^*\left[\tilde{\mathbf{s}}^*\left(\mathbf{X}_{ki_k}^*, \mathbf{X}_{li_l}^*\right) \mid \mathbf{X}_{ki_k}^*\right] = E^*\left[\tilde{\mathbf{s}}^*\left(\mathbf{X}_{ki_k}^*, \mathbf{X}_{l1}^*\right) \mid \mathbf{X}_{ki_k}^*\right]$$
$$\text{and } E\left[\tilde{\mathbf{s}}^*\left(\mathbf{X}_{ki_k}^*, \mathbf{X}_{li_l}^*\right) \mid \mathbf{X}_{li_l}^*\right] = E\left[\tilde{\mathbf{s}}^*\left(\mathbf{X}_{k1}^*, \mathbf{X}_{li_l}^*\right) \mid \mathbf{X}_{li_l}^*\right].$$

Also, define

$$\mathbf{S}_k^* = \frac{1}{n}\sum_{l=1}^K n_l \left[\frac{1}{n_k}\sum_{i_k=1}^{n_k} E^*\left[\tilde{\mathbf{s}}^*(\mathbf{X}_{ki_k}^*, \mathbf{X}_{l1}^*)|\mathbf{X}_{ki_k}^*\right] + \frac{1}{n_l}\sum_{i_l=1}^{n_l} E\left[\tilde{\mathbf{s}}^*(\mathbf{X}_{k1}^*, \mathbf{X}_{li_l}^*)|\mathbf{X}_{li_l}^*\right]\right]$$

and

$$\mathbf{V}_{kl}^* = \frac{1}{n_k n_l}\sum_{i_k=1}^{n_k}\sum_{i_l=1}^{n_l}\left[\tilde{\mathbf{s}}^*(\mathbf{X}_{ki_k}^*, \mathbf{X}_{li_l}^*) - E^*\left[\tilde{\mathbf{s}}^*\left(\mathbf{X}_{ki_k}^*, \mathbf{X}_{l1}^*\right)|\mathbf{X}_{ki_k}^*\right]\right.$$
$$\left. - E^*\left[\tilde{\mathbf{s}}^*\left(\mathbf{X}_{k1}^*, \mathbf{X}_{li_l}^*\right)|\mathbf{X}_{li_l}^*\right]\right].$$

We have $\bar{\mathbf{R}}_k^* - E^*\left[\bar{\mathbf{R}}_k^*\right] = \mathbf{S}_k^* + n^{-1}\sum_{l=1}^K n_l \mathbf{V}_{kl}^*$. Let $\mathbf{W}_n^* = \left(\sqrt{n_1}\mathbf{S}_1^*, \ldots, \sqrt{n_K}\mathbf{S}_K^*\right)$.

Using virtually identical arguments used while proving (B.4), one can show that

$$\left\|\left[\mathbf{U}_n^* - E^*\left[\mathbf{U}_n^*\right]\right] - \mathbf{W}_n^*\right\| = \sqrt{\sum_{k=1}^K \left\|\sqrt{n_k}\frac{1}{n}\sum_{l=1}^K n_l \mathbf{V}_{kl}^*\right\|^2} \xrightarrow{P^*} 0 \quad \text{as } n \to \infty.$$

(B.16)

Note that for any k, l, i_k, i_l,

$$E^*\left[\mathbf{s}\left(\mathbf{X}_{ki_k}^* - \mathbf{X}_{li_l}^*\right)\right] = n^{-2}\sum_{q=1}^K\sum_{r=1}^K\sum_{i_q=1}^{n_q}\sum_{i_r=1}^{n_r}\mathbf{s}\left(\mathbf{X}_{qi_q} - \mathbf{X}_{ri_r}\right) = \mathbf{0},$$

which implies $E^*\left[\bar{\mathbf{R}}_k^*\right] = \mathbf{0}$ for all k, and hence $E^*\left[\mathbf{U}_n^*\right] = \mathbf{0}$. Hence, from (B.16), we get

$$\left\|\mathbf{U}_n^* - \mathbf{W}_n^*\right\| = \sqrt{\sum_{k=1}^K\left\|\sqrt{n_k}\frac{1}{n}\sum_{l=1}^K n_l\mathbf{V}_{kl}^*\right\|^2} \xrightarrow{P^*} 0 \quad \text{as } n \to \infty. \qquad (\text{B.17})$$

Next, define $F_k^*\left(\mathbf{X}_{li_l}^*\right) = E^*\left[\tilde{\mathbf{s}}^*\left(\mathbf{X}_{k1}^*, \mathbf{X}_{li_l}^*\right) \mid \mathbf{X}_{li_l}^*\right]$ for all k, l and i_l,

$$\mathbf{E}_n = \frac{1}{n^3}\sum_{m=1}^K\sum_{k=1}^K\sum_{l=1}^K\sum_{i_m=1}^{n_m}\sum_{i_k=1}^{n_k}\sum_{i_l=1}^{n_l}\mathbf{s}\left(\mathbf{X}_{ki_k} - \mathbf{X}_{mi_m}\right) \otimes \mathbf{s}\left(\mathbf{X}_{li_l} - \mathbf{X}_{mi_m}\right)$$

$$- \left\{\frac{1}{n^2}\sum_{k=1}^K\sum_{l=1}^K\sum_{i_k=1}^{n_k}\sum_{i_l=1}^{n_l}\mathbf{s}(\mathbf{X}_{ki_k} - \mathbf{X}_{li_l})\right\} \otimes \left\{\frac{1}{n^2}\sum_{k=1}^K\sum_{l=1}^K\sum_{i_k=1}^{n_k}\sum_{i_l=1}^{n_l}\mathbf{s}(\mathbf{X}_{ki_k} - \mathbf{X}_{li_l})\right\}$$

and

$$\mathbf{E} = \sum_{m=1}^K\sum_{k=1}^K\sum_{l=1}^K\lambda_k\lambda_l\lambda_m E\left[\mathbf{s}\left(\mathbf{X}_{k1} - \mathbf{X}_{m1}\right) \otimes \mathbf{s}\left(\mathbf{X}_{l1} - \mathbf{X}_{m1}\right)\right]$$

$$- \left\{\sum_{k=1}^K\sum_{l=1}^K\lambda_k\lambda_l E\left[\mathbf{s}\left(\mathbf{X}_{k1} - \mathbf{X}_{l1}\right)\right]\right\} \otimes \left\{\sum_{k=1}^K\sum_{l=1}^K\lambda_k\lambda_l E\left[\mathbf{s}\left(\mathbf{X}_{k1} - \mathbf{X}_{l1}\right)\right]\right\}.$$

Note that for all k and l, $E^*\left[F_k^*\left(\mathbf{X}_{li_l}^*\right)\right] = \mathbf{0}$, and one can verify that for all k, l_1, l_2,

$$E^*\left[F_{l_1}^*\left(\mathbf{X}_{ki_k}^*\right) \otimes F_{l_2}^*\left(\mathbf{X}_{ki_k}^*\right)\right] = \mathbf{E}_n. \qquad (\text{B.18})$$

Using the fact that $\tilde{\mathbf{s}}^*\left(\mathbf{X}_{ki_k}^*, \mathbf{X}_{li_l}^*\right) = -\tilde{\mathbf{s}}^*\left(\mathbf{X}_{li_l}^*, \mathbf{X}_{ki_k}^*\right)$, we derive

$$\mathbf{S}_k^* = -\frac{1}{n}\sum_{\substack{l=1\\l\neq k}}^K\frac{n_l}{n_k}\sum_{i_k=1}^{n_k}F_l^*\left(\mathbf{X}_{ki_k}^*\right) + \frac{1}{n}\sum_{\substack{l=1\\l\neq k}}^K\sum_{i_l=1}^{n_l}F_k^*\left(\mathbf{X}_{li_l}^*\right). \qquad (\text{B.19})$$

Since $\text{Cov}^* \left(F_{l_1}^* \left(\mathbf{X}_{ki_k}^* \right), F_{l_2}^* \left(\mathbf{X}_{li_l}^* \right) \right) = \mathbf{0}$ for $(k, i_k) \neq (l, i_l)$ and for all l_1, l_2, from (B.18) and (B.19), we get

$$\text{Cov}^* \left(\sqrt{n_{k_1}} \mathbf{S}_{k_1}^*, \sqrt{n_{k_2}} \mathbf{S}_{k_2}^* \right) = \left(\mathbb{I} \left(k_1 = k_2 \right) - \frac{\sqrt{n_{k_1} n_{k_2}}}{n} \right) \mathbf{E}_n. \qquad \text{(B.20)}$$

Using arguments similar to those in the proof of Theorem 2.3, we get that $\mathbf{E}_n \xrightarrow{a.s.} \mathbf{E}$ as $n \to \infty$ in the operator norm, and hence, from (B.20), we get

$$\text{Cov}^* \left(\sqrt{n_{k_1}} \mathbf{S}_{k_1}^*, \sqrt{n_{k_2}} \mathbf{S}_{k_2}^* \right) \xrightarrow{a.s.} \left(\mathbb{I} \left(k_1 = k_2 \right) - \sqrt{\lambda_{k_1} \lambda_{k_2}} \right) \mathbf{E} \quad \text{as } n \to \infty. \qquad \text{(B.21)}$$

Define $\mathbf{\Sigma}^* = \left(\boldsymbol{\sigma}_{k_1 k_2}^* \right)_{K \times K}$, where $\boldsymbol{\sigma}_{k_1 k_2}^* = \left(\mathbb{I} \left(k_1 = k_2 \right) - \sqrt{\lambda_{k_1} \lambda_{k_2}} \right) \mathbf{E}$ for $k_1, k_2 = 1, \ldots, K$. Since $E^* \left[\mathbf{S}_k^* \right] = \mathbf{0}$ for all k, we have $E \left[\mathbf{W}_n^* \right] = \mathbf{0}$, which, along with (B.21) and an application of Theorem 1.1 in [28], implies that $\mathbf{W}_n^* \xrightarrow{w^*} G \left(\mathbf{0}, \mathbf{\Sigma}^* \right)$ as $n \to \infty$. Therefore, from (B.17), we get $\mathbf{U}_n^* \xrightarrow{w^*} G \left(\mathbf{0}, \mathbf{\Sigma}^* \right)$ as $n \to \infty$, and since $\mathbf{\Sigma}^*$ is independent of $\mathcal{S}$, we have the unconditional convergence

$$\mathbf{U}_n^* \xrightarrow{w} G \left(\mathbf{0}, \mathbf{\Sigma}^* \right) \quad \text{as } n \to \infty. \qquad \text{(B.22)}$$

It can be verified that when the null hypothesis H_0 in (2.1) is true, we have $\mathbf{\Sigma}^* = \mathbf{\Sigma}$, where $\mathbf{\Sigma}$ is as defined in Theorem 2.1. From Corollary 2.2 and using arguments similar to those used in its proof, we get that under H_0 in (2.1), the asymptotic distribution of SS_n^* is identical to that of SS_n. Hence under H_0 in (2.1), for every $0 < \alpha < 1$, the size of a level α test based on the bootstrap procedure converges to α as $n \to \infty$ and $M_b \to \infty$.

Next, we prove the theorem for the permutation method. Let $\boldsymbol{\sigma}_n$ denote a random permutation of $(1, \ldots, n)$, and $\{ \mathbf{X}_{ki}^{\#} \mid k = 1, \ldots, K; i = 1, \ldots, n_k \}$ denote the permuted sample obtained by applying $\boldsymbol{\sigma}_n$ to the original sample. Define $\mathbf{R}^{\#}(\mathbf{x}) = n^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathbf{s}(\mathbf{x} - \mathbf{X}_{ki}^{\#})$ and $\bar{\mathbf{R}}_k^{\#} = n_k^{-1} \sum_{i_k=1}^{n_k} \mathbf{R}^{\#} \left(\mathbf{X}_{ki_k}^{\#} \right)$. So, $SS_n^{\#} = \sum_{k=1}^{K} n_k \left\| \bar{\mathbf{R}}_k^{\#} \right\|^2$. Given the original sample, all possible realizations of the permuted sample $\{ \mathbf{X}_{ki}^{\#} \mid k = 1, \ldots, K; i = 1, \ldots, n_k \}$ are equally likely. Also, under the null hypothesis H_0 in (2.1), the underlying distributions of all the classes are identical. So, the joint distribution of $\{ \mathbf{X}_{ki}^{\#} \mid k = 1, \ldots, K; i = 1, \ldots, n_k \}$ is the same as that of the original sample. Hence, under H_0 in (2.1), the distribution of $SS_n^{\#}$ is the same as the distribution of the original test statistic SS_n for all n. Therefore, under H_0 in (2.1), the asymptotic distribution of $SS_n^{\#}$ is the same as that of SS_n, which is given in Corollary 2.2. So, as $n \to \infty$ and $M \to \infty$, the empirical distribution of $SS_n^{\#}$ converges to the asymptotic null distribution of SS_n. Hence under H_0 in (2.1), for every $0 < \alpha < 1$, the size of a level α test based on the permutation procedure converges to α as $n \to \infty$ and $M \to \infty$. $\square$

Proof of Theorem 2.5. Since for all k, $E[\bar{\mathbf{R}}_k] \to \sum_{l=1}^{K} \lambda_l E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})]$ as $n \to \infty$, if $\sum_{l=1}^{K} \lambda_l E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})] \neq \mathbf{0}$ for any k, we have $\left\| E \left[\sqrt{n_k} \bar{\mathbf{R}}_k \right] \right\| \to \infty$

as $n \to \infty$ for that k, and this implies $\|E[\mathbf{U}_n]\| \to \infty$ as $n \to \infty$. Consequently, from Theorem 2.1, we have $\|\mathbf{U}_n\| \geq \|E[\mathbf{U}_n]\| - \|\mathbf{U}_n - E[\mathbf{U}_n]\| \xrightarrow{P} \infty$ as $n \to \infty$. Therefore,

$$SS_n = \|\mathbf{U}_n\|^2 \xrightarrow{P} \infty \quad \text{as } n \to \infty. \tag{B.23}$$

From (B.23) and Corollary 2.2, we get that for every $0 < \alpha < 1$, the power of a level α test based on the asymptotic procedure converges to 1 as $n \to \infty$.

Next, from (B.22), we get that SS_n^* is stochastically bounded whether the null or the alternative is true. Hence, from (B.23), we get that for every $0 < \alpha < 1$, the power of a level α test based on the bootstrap procedure converges to 1 as $n \to \infty$ and $M_b \to \infty$.

Next we prove the theorem for the permutation procedure. Recall $\bar{\mathbf{R}}_k^{\#}$ defined in the proof of Theorem 2.4. We shall show that $n_k\|\bar{\mathbf{R}}_k^{\#}\|^2$ is stochastically bounded for all k, and this would imply that $SS_n^{\#}$ is stochastically bounded under the null as well as the alternative hypotheses. Then the proof would follow from (B.23). It can be verified that

$$n_k\left\|\bar{\mathbf{R}}_k^{\#}\right\|^2$$

$$= n_k \sum_{\substack{l_1=1 \\ l_1 \neq k}}^{K} \sum_{\substack{l_2=1 \\ l_2 \neq k}}^{K} \frac{1}{n^2 n_k^2} \sum_{i_1=1}^{n_k} \sum_{\substack{i_2=1 \\ i_1 \neq i_2}}^{n_k} \sum_{j_1=1}^{n_{l_1}} \sum_{\substack{j_2=1 \\ (l_1,j_1) \neq (l_2,j_2)}}^{n_{l_2}} \left\langle \mathbf{s}\left(\mathbf{X}_{ki_1}^{\#} - \mathbf{X}_{l_1 j_1}^{\#}\right), \mathbf{s}\left(\mathbf{X}_{ki_2}^{\#} - \mathbf{X}_{l_2 j_2}^{\#}\right)\right\rangle$$

$$+ n_k \sum_{\substack{l=1 \\ l \neq k}}^{K} \frac{n_l^2}{n^2} \frac{1}{n_k^2 n_l^2} \sum_{i_1=1}^{n_k} \sum_{\substack{i_2=1 \\ i_1 \neq i_2}}^{n_k} \sum_{j=1}^{n_l} \left\langle \mathbf{s}\left(\mathbf{X}_{ki_1}^{\#} - \mathbf{X}_{lj}^{\#}\right), \mathbf{s}\left(\mathbf{X}_{ki_2}^{\#} - \mathbf{X}_{lj}^{\#}\right)\right\rangle$$

$$+ n_k \sum_{\substack{l_1=1 \\ l_1 \neq k}}^{K} \sum_{\substack{l_2=1 \\ l_2 \neq k}}^{K} \frac{1}{n^2 n_k^2} \sum_{i=1}^{n_k} \sum_{j_1=1}^{n_{l_1}} \sum_{\substack{j_2=1 \\ (l_1,j_1) \neq (l_2,j_2)}}^{n_{l_2}} \left\langle \mathbf{s}\left(\mathbf{X}_{ki}^{\#} - \mathbf{X}_{l_1 j_1}^{\#}\right), \mathbf{s}\left(\mathbf{X}_{ki}^{\#} - \mathbf{X}_{l_2 j_2}^{\#}\right)\right\rangle$$

$$+ n_k \sum_{\substack{l=1 \\ l \neq k}}^{K} \frac{n_l^2}{n^2} \frac{1}{n_k^2 n_l^2} \sum_{i=1}^{n_k} \sum_{j=1}^{n_l} \left\langle \mathbf{s}\left(\mathbf{X}_{ki}^{\#} - \mathbf{X}_{lj}^{\#}\right), \mathbf{s}\left(\mathbf{X}_{ki}^{\#} - \mathbf{X}_{lj}^{\#}\right)\right\rangle.$$

Let $E^{\#}[\cdot]$ denote the conditional expectation given the original sample. We get that for (k, i_1), (k, i_2), (l_1, j_1) and (l_2, j_2) all distinct,

$$E^{\#}\left[\left\langle \mathbf{s}\left(\mathbf{X}_{ki_1}^{\#} - \mathbf{X}_{l_1 j_1}^{\#}\right), \mathbf{s}\left(\mathbf{X}_{ki_2}^{\#} - \mathbf{X}_{l_2 j_2}^{\#}\right)\right\rangle\right] = \mathbf{0},$$

for (k, i_1), (k, i_2) and (l, j) all distinct,

$$E^{\#}\left[\left\langle \mathbf{s}\left(\mathbf{X}_{ki_1}^{\#} - \mathbf{X}_{lj}^{\#}\right), \mathbf{s}\left(\mathbf{X}_{ki_2}^{\#} - \mathbf{X}_{lj}^{\#}\right)\right\rangle\right]$$
$$= \frac{(n-3)!}{n!} \sum_{(i,j,k) \in \mathcal{P}_3^n} \left\langle \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_k\right), \mathbf{s}\left(\mathbf{Y}_j - \mathbf{Y}_k\right)\right\rangle,$$

for (k, i), (l_1, j_1) and (l_2, j_2) all distinct,

$$
\begin{aligned}
&E^{\#}\left[\left\langle \mathbf{s}\left(\mathbf{X}_{ki}^{\#} - \mathbf{X}_{l_1 j_1}^{\#}\right),\ \mathbf{s}\left(\mathbf{X}_{ki}^{\#} - \mathbf{X}_{l_2 j_2}^{\#}\right)\right\rangle\right] \\
&= \frac{(n-3)!}{n!} \sum_{(i,j,k)\in\mathcal{P}_3^n} \left\langle \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_k\right),\ \mathbf{s}\left(\mathbf{Y}_j - \mathbf{Y}_k\right)\right\rangle,
\end{aligned}
$$

and for $(k, i) \neq (l, j)$,

$$
\begin{aligned}
&E^{\#}\left[\left\langle \mathbf{s}\left(\mathbf{X}_{ki}^{\#} - \mathbf{X}_{lj}^{\#}\right),\ \mathbf{s}\left(\mathbf{X}_{ki}^{\#} - \mathbf{X}_{lj}^{\#}\right)\right\rangle\right] \\
&= \frac{(n-2)!}{n!} \sum_{(i,j)\in\mathcal{P}_2^n} \left\langle \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_j\right),\ \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_j\right)\right\rangle,
\end{aligned}
$$

where $\mathcal{P}_k^n$ is the collection of all permutations of size k drawn from $(1, \ldots, n)$, and $\mathbf{Y}_i$, $\mathbf{Y}_j$ and $\mathbf{Y}_k$ are the ith, j^{th} and the kth elements of the sample, respectively. Define

$$
R_{n,1} = \frac{(n-3)!}{n!} \sum_{(i,j,k)\in\mathcal{P}_3^n} \left\langle \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_k\right),\ \mathbf{s}\left(\mathbf{Y}_j - \mathbf{Y}_k\right)\right\rangle,
$$

$$
\text{and } R_{n,2} = \frac{(n-2)!}{n!} \sum_{(i,j)\in\mathcal{P}_2^n} \left\langle \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_j\right),\ \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_j\right)\right\rangle.
$$

From the above discussion, we get

$$
\begin{aligned}
&E\left[n_k \left\|\bar{\mathbf{R}}_k^{\#}\right\|^2\right] \\
&= E\left[R_{n,1}\right]\left[\sum_{l\neq k} \frac{(n_k + n_l - 2)n_l}{n^2} + \sum_{l_1 \neq l_2 \neq k} \frac{n_{l_1} n_{l_2}}{n^2}\right] + E\left[R_{n,2}\right]\frac{1}{n}\sum_{l\neq k}\frac{n_l}{n},
\end{aligned}
$$
(B.24)

where $R_{n,1} = \dfrac{(n-3)!}{n!} \displaystyle\sum_{(i,j,k)\in\mathcal{P}_3^n} \left\langle \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_k\right),\ \mathbf{s}\left(\mathbf{Y}_j - \mathbf{Y}_k\right)\right\rangle$

and $R_{n,2} = \dfrac{(n-2)!}{n!} \displaystyle\sum_{(i,j)\in\mathcal{P}_2^n} \left\langle \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_j\right),\ \mathbf{s}\left(\mathbf{Y}_i - \mathbf{Y}_j\right)\right\rangle.$

Now, from an application of the Cauchy-Schwarz inequality, we get that $|R_{n,1}| \leq 1$ and $|R_{n,2}| \leq 1$. Therefore, from (B.24), we have $E[n_k\|\bar{\mathbf{R}}_k^{\#}\|^2] \leq 2K$, and consequently $n_k\|\bar{\mathbf{R}}_k^{\#}\|^2$ is stochastically bounded for all k. Hence, $SS_n^{\#}$ is stochastically bounded.

Hence, from (B.23), we get that for every $0 < \alpha < 1$, the power of a level α test based on the permutation procedure converges to 1 as $n \to \infty$ and $M \to \infty$. $\quad\square$

Proof of Theorem 2.6. Under H_0' in (2.3), H_0 in (2.1) holds, and from Theorem 2.4, we get that for every $0 < \alpha < 1$, the sizes of a level α test based on the asymptotic and the permutation procedures converge to α as $n \to \infty$ and $M \to \infty$.

Next, let $\mathbf{X}$ and $\mathbf{X}'$ be independent random elements having distribution P_0. Consider a random element $\mathbf{Z}$ such that $\mathbf{Z} = (\mathbf{X} + \boldsymbol{\mu}_k) - \mathbf{X}'$ with probability λ_k for $k = 1, \ldots, K$. The spatial distribution of $\mathbf{Z}$ at a point $\mathbf{z}$ is defined as $\mathbf{S}(\mathbf{z}) = E[\mathbf{s}(\mathbf{z} - \mathbf{Z})]$ [9, p. 1205]. Since the support of P_0 is not contained in a straight line in $\mathcal{H}$, it follows from the proof of Theorem 3.1 in [9] that $\mathbf{S}(\mathbf{z}_1) \neq \mathbf{S}(\mathbf{z}_2)$ for $\mathbf{z}_1 \neq \mathbf{z}_2$. So, $\mathbf{S}(\boldsymbol{\mu}_k) = \mathbf{0}$ for all k implies $\boldsymbol{\mu}_k = \boldsymbol{\mu}_l$ for all $k \neq l$. Note that $\mathbf{S}(\boldsymbol{\mu}_k) = \sum_{l=1}^{K} \lambda_l E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})]$, and $E[\bar{\mathbf{R}}_k] \to \mathbf{S}(\boldsymbol{\mu}_k)$ as $n \to \infty$ since $n^{-1} n_k \to \lambda_k$ as $n \to \infty$ for all k. When H_0' is false, $\boldsymbol{\mu}_k \neq \boldsymbol{\mu}_l$ for some $k \neq l$, which implies $\mathbf{S}(\boldsymbol{\mu}_k) \neq \mathbf{0}$ for some k, and the proof follows from Theorem 2.5. $\square$

We need the following lemma for the proof of Theorem 3.1. The arguments used in its proof are similar to those in the proof of Proposition 2.1 in [7].

Lemma B.1. *Define* $\mathbf{f} : \mathcal{H} \to \mathcal{H}$ *by* $\mathbf{f}(\mathbf{a}) = E[\mathbf{s}(\mathbf{X} - \mathbf{X}' + \mathbf{a})]$. *Assume that the probability measure* P_0 *is non-atomic and not contained in any straight line in* $\mathcal{H}$, *and* $E\left[\|\mathbf{X} - \mathbf{X}'\|^{-1}\right] < \infty$, *where* $\mathbf{X}$ *and* $\mathbf{X}'$ *are independent random elements having identical distribution* P_0. *Then, the Frechét derivative of* $\mathbf{f}(\cdot)$ *at* $\mathbf{0}$ *is* $E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right]$. *Further, for any* $\mathbf{h} \neq \mathbf{0}$, $E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right](\mathbf{h}) \neq \mathbf{0}$.

Proof. The Frechét derivative of $\mathbf{s}(\mathbf{x}) = \|\mathbf{x}\|^{-1} \mathbf{x}$ at $\mathbf{x}$, denoted as $\mathbf{s}^{(1)}(\mathbf{x})$, exists for all $\mathbf{x} \neq \mathbf{0}$, and it is a bounded linear operator on $\mathcal{H}$ defined by

$$\mathbf{s}^{(1)}(\mathbf{x})(\mathbf{h}) = \frac{1}{\|\mathbf{x}\|}\left(\mathbf{h} - \left\langle \mathbf{h}, \frac{\mathbf{x}}{\|\mathbf{x}\|}\right\rangle \frac{\mathbf{x}}{\|\mathbf{x}\|}\right), \text{ where } \mathbf{h} \in \mathcal{H}.$$

Note that $\left\|\mathbf{s}^{(1)}(\mathbf{x})\right\| \leq \|\mathbf{x}\|^{-1}$. Since by assumption $E\left[\|\mathbf{X} - \mathbf{X}'\|^{-1}\right] < \infty$, which implies $P[\mathbf{X} - \mathbf{X}' = \mathbf{0}] = 0$, we get that $E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right]$ exists and is the Frechét derivative of $\mathbf{f}(\mathbf{a}) = E[\mathbf{s}(\mathbf{X} - \mathbf{X}' + \mathbf{a})]$ at $\mathbf{a} = \mathbf{0}$.

Next, we shall show that the operator $E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right]$ is bounded below. The arguments in this part of the proof is similar to those in the proof of Proposition 2.1 in [7]. Let $\mathbf{P_h}(\cdot)$ be the orthogonal projection operator on the orthogonal complement of $\mathbf{h}$. Note that

$$\left\langle \mathbf{h}, E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right](\mathbf{h})\right\rangle = E\left[\frac{1}{\|\mathbf{X} - \mathbf{X}'\|}\left\|\mathbf{P_h}\left(\frac{\mathbf{X} - \mathbf{X}'}{\|\mathbf{X} - \mathbf{X}'\|}\right)\right\|^2\right].$$

Consider the set of all subspaces $\mathcal{K} \subseteq \mathcal{H}$ such that $\text{Var}(\langle \mathbf{v}, \|\mathbf{X} - \mathbf{X}'\|^{-1}(\mathbf{X} - \mathbf{X}')\rangle) = 0$ for all $\mathbf{v} \in \mathcal{K}$. Suppose this set is non-empty. Then, by Zorn's lemma, it has a maximal element, which we denote as $\mathcal{K}_0$. The maximality of $\mathcal{K}_0$ implies that $\text{Var}\left(\left\langle \mathbf{v}, \|\mathbf{X} - \mathbf{X}'\|^{-1}(\mathbf{X} - \mathbf{X}')\right\rangle\right) > 0$ for all $\mathbf{v} \in \mathcal{K}_0^\perp$, the orthogonal complement of $\mathcal{K}_0$. Since $\mathbf{X}$ and $\mathbf{X}'$ are independent random elements having

distribution P_0, which is not contained in a straight line, $\mathcal{K}_0^\perp$ must have dimension at least 2. Let $\mathbf{v}_1, \mathbf{v}_2$ be two orthogonal unit vectors in $\mathcal{K}_0^\perp$. Define $g(t) = \mathrm{Var}\left(\left\langle \cos(t)\mathbf{v}_1 + \sin(t)\mathbf{v}_2, \|\mathbf{X} - \mathbf{X}'\|^{-1}(\mathbf{X} - \mathbf{X}')\right\rangle\right)$ for $0 \le t \le 2\pi$. Note that $g(t) > 0$ for all t, because $\cos(t)\mathbf{v}_1 + \sin(t)\mathbf{v}_2 \in \mathcal{K}_0^\perp$. Since $g(\cdot)$ is continuous on a compact set, we get that there is $l_0 > 0$ such that for all unit vectors in span $\{\mathbf{v}_1, \mathbf{v}_2\}$, $\mathrm{Var}\left(\left\langle \mathbf{v}, \|\mathbf{X} - \mathbf{X}'\|^{-1}(\mathbf{X} - \mathbf{X}')\right\rangle\right) \ge l_0$. The orthogonal complement of $\mathbf{h}$ and span $\{\mathbf{v}_1, \mathbf{v}_2\}$ must have nonempty intersection, which implies that there is a unit vector $\mathbf{v}_0 \in \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ such that $\langle \mathbf{v}_0, \mathbf{h}\rangle = 0$. Hence, for all $\mathbf{y} \in \mathcal{H}$, $\|\mathbf{P_h}(\mathbf{y})\|^2 \ge \langle \mathbf{v}_0, \mathbf{y}\rangle^2$. We take M large enough such that $\mathrm{Var}\left[\left\langle \mathbf{v}_0, \|\mathbf{X} - \mathbf{X}'\|^{-1}(\mathbf{X} - \mathbf{X}')\,\mathbb{I}(\|\mathbf{X} - \mathbf{X}'\| \le M)\right\rangle\right] > (l_0/2)$. Therefore,

$$
\begin{aligned}
&\left\langle \mathbf{h}, E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right](\mathbf{h})\right\rangle \\
&\ge E\left[\frac{1}{\|\mathbf{X} - \mathbf{X}'\|}\left\|\mathbf{P_h}\left(\frac{\mathbf{X} - \mathbf{X}'}{\|\mathbf{X} - \mathbf{X}'\|}\right)\right\|^2 \mathbb{I}(\|\mathbf{X} - \mathbf{X}'\| \le M)\right] \\
&\ge \frac{1}{M}E\left[\left\langle \mathbf{v}_0, \|\mathbf{X} - \mathbf{X}'\|^{-1}(\mathbf{X} - \mathbf{X}')\,\mathbb{I}(\|\mathbf{X} - \mathbf{X}'\| \le M)\right\rangle^2\right] \\
&\ge \frac{1}{M}\mathrm{Var}\left[\left\langle \mathbf{v}_0, \|\mathbf{X} - \mathbf{X}'\|^{-1}(\mathbf{X} - \mathbf{X}')\,\mathbb{I}(\|\mathbf{X} - \mathbf{X}'\| \le M)\right\rangle\right] \ge \frac{l_0}{2M} > 0.
\end{aligned}
$$
(B.25)

From (B.25), we get that for every $\mathbf{h} \in \mathcal{H}$,

$$
0 < \frac{l_0}{2M}\|\mathbf{h}\|^2 \le \left\langle \mathbf{h}, E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right](\mathbf{h})\right\rangle \le \left\|E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right](\mathbf{h})\right\|\|\mathbf{h}\|,
$$

which implies $\left\|E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right](\mathbf{h})\right\| \ge \frac{l_0}{2M}\|\mathbf{h}\| > 0.$ (B.26)

Hence, from (B.26), we get that for any $\mathbf{h} \ne \mathbf{0}$, $E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right](\mathbf{h}) \ne \mathbf{0}$.   □

Proof of Theorem 3.1. Define $\mathbf{f} : \mathcal{H} \to \mathcal{H}$ by $\mathbf{f}(\mathbf{a}) = E[\mathbf{s}(\mathbf{X} - \mathbf{X}' + \mathbf{a})]$. From Lemma B.1, we get that the Frechét derivative of $\mathbf{f}(\cdot)$ at $\mathbf{0}$ is $E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right]$. Define $\mathbf{X}_k = \mathbf{X}_{k1} - \boldsymbol{\mu}_k$ for $k = 1, \ldots, K$. Then, $\mathbf{X}_1, \ldots, \mathbf{X}_K$ are independent and identically distributed random elements with common distribution P_0. Since $\mathbf{s}(\mathbf{x} - \mathbf{y}) = -\mathbf{s}(\mathbf{y} - \mathbf{x})$, we get $E[\mathbf{s}(\mathbf{X}_k - \mathbf{X}_l)] = -E[\mathbf{s}(\mathbf{X}_k - \mathbf{X}_l)] = \mathbf{0}$ for all k, l. Using this fact and a Taylor expansion of $\mathbf{f}(\cdot)$ at $\mathbf{0}$, we have

$$
\begin{aligned}
&E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})] \\
&= E\left[\mathbf{s}^{(1)}(\mathbf{X} - \mathbf{X}')\right]\left(\frac{1}{\sqrt{n}}(\boldsymbol{\delta}_k - \boldsymbol{\delta}_l)\right) + o\left(\frac{1}{\sqrt{n}}\|\boldsymbol{\delta}_k - \boldsymbol{\delta}_l\|\right).
\end{aligned}
$$
(B.27)

Since $n^{-1}n_k \to \lambda_k \in (0, 1)$ for all k as $n \to \infty$, from (B.27) we get

$$
E\left[\sqrt{n_k}\bar{\mathbf{R}}_k\right] = \sqrt{n_k}\sum_{l=1}^{K}\frac{n_l}{n}E[\mathbf{s}(\mathbf{X}_{k1} - \mathbf{X}_{l1})]
$$

$$\to E\left[\mathbf{s}^{(1)}\left(\mathbf{X}-\mathbf{X}'\right)\right]\left(\sqrt{\lambda_k}\left(\boldsymbol{\delta}_k-\bar{\boldsymbol{\delta}}\right)\right) \tag{B.28}$$

as $n \to \infty$. From (B.28), we have $E\left[\mathbf{U}_n\right] \to \mathbf{U}_0 = (\mathbf{u}_1, \ldots, \mathbf{u}_K)$ as $n \to \infty$, where

$$\mathbf{u}_k = E\left[\mathbf{s}^{(1)}\left(\mathbf{X}-\mathbf{X}'\right)\right]\left(\sqrt{\lambda_k}\left(\boldsymbol{\delta}_k-\bar{\boldsymbol{\delta}}\right)\right)$$

for all $k = 1, \ldots, K$. From Slutsky's theorem and Theorem 2.1, we get $\mathbf{U}_n \xrightarrow{w} G(\mathbf{U}_0, \boldsymbol{\Sigma})$ as $n \to \infty$. Consequently, from an application of the Karhunen-Loeve expansion and the mapping theorem [3, Theorem 2.7, p. 21], we get $\|\mathbf{U}_n\|^2 \xrightarrow{w} \|\mathbf{U}_0 - \sum_{i=1}^{\infty}\langle\mathbf{U}_0, \boldsymbol{\beta}_i\rangle\|^2 + \sum_{i=1}^{\infty}\left(\langle\mathbf{U}_0, \boldsymbol{\beta}_i\rangle + \sqrt{\alpha_i}\mathbf{Z}_i\right)^2$ as $n \to \infty$.

Let $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K)^t$, $\mathbf{D} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_K)$, $\mathbf{L} = (\lambda_1, \ldots, \lambda_K)^t$, $\mathbf{I}_K$ be the $K \times K$ identity matrix and $\mathbf{1}_K$ be the column vector of length K with all elements equal to 1. Then, $\mathbf{U}_0 = E\left[\mathbf{s}^{(1)}\left(\mathbf{X}-\mathbf{X}'\right)\right]\left(\sqrt{\mathbf{D}}\left(\mathbf{I}_K - \mathbf{1}_K\mathbf{L}^t\right)\boldsymbol{\delta}\right)$. Note that $\left(\mathbf{I}_K - \mathbf{1}_K\mathbf{L}^t\right)$ is a $K \times K$ idempotent matrix with trace $(K-1)$ and its null space contains $\mathbf{1}_K$. So, for any $\boldsymbol{\delta}$ such that $\boldsymbol{\delta}_k \neq \boldsymbol{\delta}_l$ for some k and l, $\left(\mathbf{I}_K - \mathbf{1}_K\mathbf{L}^t\right)\boldsymbol{\delta} \neq \mathbf{0}$, which implies

$$\mathbf{U}_0 = E\left[\mathbf{s}^{(1)}\left(\mathbf{X}-\mathbf{X}'\right)\right]\left(\sqrt{\mathbf{D}}\left(\mathbf{I}_K - \mathbf{1}_K\mathbf{L}^t\right)\boldsymbol{\delta}\right) \neq \mathbf{0}$$

since $\sqrt{\mathbf{D}}$ is non-singular, and $E\left[\mathbf{s}^{(1)}\left(\mathbf{X}-\mathbf{X}'\right)\right](\mathbf{h}) \neq \mathbf{0}$ for any $\mathbf{h} \neq \mathbf{0}$ from Lemma B.1. $\qquad\square$

Proof of Theorem A.1. Proof of (a): Since $E[\|\mathbf{X}\|^2] < \infty$ and $\sqrt{n_k}n^{-\frac{1}{2}} \to \sqrt{\lambda_k}$ as $n \to \infty$ for all k, from Theorem 1.1 in [28] and Slutsky's Theorem, we get

$$\left(\sqrt{n_1}\left(\bar{\mathbf{X}}_{1\cdot} - E[\mathbf{X}]\right), \ldots, \sqrt{n_K}\left(\bar{\mathbf{X}}_{K\cdot} - E[\mathbf{X}]\right)\right)$$
$$\xrightarrow{w}\left(\sqrt{\lambda_1}\boldsymbol{\delta}_1 + \mathbf{Y}_1, \ldots, \sqrt{\lambda_K}\boldsymbol{\delta}_K + \mathbf{Y}_K\right) \tag{B.29}$$

as $n \to \infty$. Define $\mathbf{g}(\cdot), \mathbf{g}_n(\cdot) : \mathcal{H}^K \to \mathbb{R}$ as

$$\mathbf{g}(\mathbf{x}_1, \ldots, \mathbf{x}_K) = \sum_{k<l}\left\|\mathbf{x}_k - \lambda_l^{-1/2}\lambda_k^{1/2}\mathbf{x}_l\right\|^2$$

$$\text{and } \mathbf{g}_n(\mathbf{x}_1, \ldots, \mathbf{x}_K) = \sum_{k<l}\left\|\mathbf{x}_k - n_l^{-1/2}n_k^{1/2}\mathbf{x}_l\right\|^2.$$

Since $\mathbf{g}(\cdot)$ is a continuous function and for every $\mathbf{x}_1, \ldots, \mathbf{x}_K$, $|\mathbf{g}_n(\mathbf{x}_1, \ldots, \mathbf{x}_K) - \mathbf{g}(\mathbf{x}_1, \ldots, \mathbf{x}_K)| \longrightarrow 0$ as $n \to \infty$, from (B.29), Theorem 2.7 in [3, p. 21] and an application of Slutsky's Theorem, we get

$$CFF_n = \mathbf{g}_n\left(\sqrt{n_1}\left(\bar{\mathbf{X}}_{1\cdot} - E[\mathbf{X}]\right), \ldots, \sqrt{n_K}\left(\bar{\mathbf{X}}_{K\cdot} - E[\mathbf{X}]\right)\right)$$
$$\xrightarrow{w} \mathbf{g}\left(\sqrt{\lambda_1}\boldsymbol{\delta}_1 + \mathbf{Y}_1, \ldots, \sqrt{\lambda_K}\boldsymbol{\delta}_K + \mathbf{Y}_K\right)$$
$$= \sum_{k<l}\left\|\sqrt{\lambda_k}\boldsymbol{\delta}_k + \mathbf{Y}_k - \lambda_l^{-1/2}\lambda_k^{1/2}\left(\sqrt{\lambda_l}\boldsymbol{\delta}_l + \mathbf{Y}_l\right)\right\|^2$$

as $n \to \infty$. Since under H_0 described in (2.1) in the main paper, $CFF_n \xrightarrow{w} \sum_{k<l} \left\| \mathbf{Y}_k - \lambda_l^{-1/2} \lambda_k^{1/2} \mathbf{Y}_l \right\|^2$ as $n \to \infty$, the proof of part *(a)* of the theorem follows.

Proof of (b): From the arguments in [36, p. 151], we get that under the null hypothesis H_0 in (2.1) in the main paper,

$$ZC_n \xrightarrow{w} \sum_{i=1}^{\infty} \gamma_i \boldsymbol{\chi}_i^2 \quad \text{as } n \to \infty. \tag{B.30}$$

Let $\mathbf{Z}_n = \left(\sqrt{n_1} \left(\bar{\mathbf{X}}_{1.} - E[\mathbf{X}] \right), \dots, \sqrt{n_K} \left(\bar{\mathbf{X}}_{K.} - E[\mathbf{X}] \right) \right)^t$. Since $E[\|\mathbf{X}\|^2] < \infty$ and $\sqrt{n_k} n^{-\frac{1}{2}} \to \sqrt{\lambda_k}$ for all k as $n \to \infty$, from Theorem 1.1 in [28] and Slutsky's Theorem, we get that under the shrinking alternatives described in (3.1) in the main paper, $\mathbf{Z}_n \xrightarrow{w} \mathbf{Z}$ as $n \to \infty$. Let $\mathbf{p}_n = \left(\sqrt{n_1}, \dots, \sqrt{n_K} \right)^t$. We have $(\mathbf{I}_K - (\mathbf{p}_n \mathbf{p}_n^t)/n) \to (\mathbf{I}_K - \mathbf{p}_0 \mathbf{p}_0^t)$ as $n \to \infty$. Since $ZC_n = \sum_{k=1}^{K} n_k \left\| \bar{\mathbf{X}}_{k.} - \bar{\mathbf{X}}_{..} \right\|^2 = \mathbf{Z}_n^t \left(\mathbf{I}_K - (\mathbf{p}_n \mathbf{p}_n^t)/n \right) \mathbf{Z}_n$, from Theorem 2.7 in [3, p. 21] and an application of Slutsky's theorem, we get that under the shrinking alternatives described in (3.1),

$$ZC_n \xrightarrow{w} \mathbf{Z}^t \left(\mathbf{I}_K - \mathbf{p}_0 \mathbf{p}_0^t \right) \mathbf{Z} \quad \text{as } n \to \infty. \tag{B.31}$$

The proof follows from (B.30) and (B.31).

Proof of (c): Under the condition of the theorem, from [28] we get that there exist independent Gaussian processes $\mathbf{W}_k$ with mean $\mathbf{0}$ and covariance operator $\boldsymbol{\Omega}_k$, $k = 1, \dots, K$, such that

$$\max_{1 \le k \le K} \left\| n_k^{-1/2} \sum_{i=1}^{n_k} [\mathbf{X}_{ki} - \boldsymbol{\mu}_k] - \mathbf{W}_k \right\| \xrightarrow{P} 0$$

as $n \to \infty$. Let $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_d$ be the eigenfunctions corresponding to the d largest eigenvalues of $\boldsymbol{\Omega}$. Define $\mathbf{Z}_k = \mathbf{W}_k - \sqrt{\lambda_k} \boldsymbol{\delta}_k$ for $k = 1, \dots, K$. Define

$$\mathbf{z}_k = (\langle \mathbf{Z}_k, \boldsymbol{\phi}_1 \rangle, \dots, \langle \mathbf{Z}_k, \boldsymbol{\phi}_K \rangle)^t, \quad \mathbf{z}_{..} = \left(\sum_{l=1}^{K} \lambda_l \boldsymbol{\Psi}_l^{-1} \right)^{-1} \sum_{l=1}^{K} \sqrt{\lambda_l} \boldsymbol{\Psi}_l^{-1} \mathbf{z}_l$$

and $\quad \mathbf{y}_k = \boldsymbol{\Psi}_k^{-1/2} \mathbf{z}_k$.

Note that $\mathbf{z}_k$'s are independent d-variate normal random vectors with dispersion matrix $\boldsymbol{\Psi}_k$ and mean vector $\sqrt{\lambda_k} \mathbf{d}_k$. Consequently, $\mathbf{y}_k$'s are independent d-variate normal random vectors with the identity matrix $\mathbf{I}_d$ as dispersion and mean vector $\sqrt{\lambda_k} \boldsymbol{\Psi}_k^{-1/2} \mathbf{d}_k$. Using arguments similar to those used in the proof of Theorem 2.1 in [23], one can show that

$$HR_n = \sum_{k=1}^{K} \left(\mathbf{z}_k - \sqrt{\lambda_k} \mathbf{z}_{..} \right)^t \boldsymbol{\Psi}_k^{-1} \left(\mathbf{z}_k - \sqrt{\lambda_k} \mathbf{z}_{..} \right) + o_P(1) \quad \text{as } n \to \infty. \tag{B.32}$$

Now,

$$\sum_{k=1}^{K} \left(\mathbf{z}_k - \sqrt{\lambda_k} \mathbf{z}_{..} \right)^t \boldsymbol{\Psi}_k^{-1} \left(\mathbf{z}_k - \sqrt{\lambda_k} \mathbf{z}_{..} \right)$$

$$= \sum_{k=1}^{K} \mathbf{y}_k^t \mathbf{y}_k - \left(\sum_{k=1}^{K} \sqrt{\lambda_k} \boldsymbol{\Psi}_k^{-1/2} \mathbf{y}_k \right)^t \left(\sum_{k=1}^{K} \lambda_k \boldsymbol{\Psi}_k^{-1} \right)^{-1} \left(\sum_{k=1}^{K} \sqrt{\lambda_k} \boldsymbol{\Psi}_k^{-1/2} \mathbf{y}_k \right)$$

$$= \left[\mathbf{y}_1^t, \ldots, \mathbf{y}_K^t \right] \left[\mathbf{y}_1^t, \ldots, \mathbf{y}_K^t \right]^t$$

$$- \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{bmatrix}^t \begin{bmatrix} \sqrt{\lambda_1} \boldsymbol{\Psi}_1^{-1/2} \\ \vdots \\ \sqrt{\lambda_K} \boldsymbol{\Psi}_K^{-1/2} \end{bmatrix} \left(\sum_{k=1}^{K} \lambda_k \boldsymbol{\Psi}_k^{-1} \right)^{-1} \left[\sqrt{\lambda_1} \boldsymbol{\Psi}_1^{-1/2}, \ldots, \sqrt{\lambda_K} \boldsymbol{\Psi}_K^{-1/2} \right] \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{bmatrix}$$

$$= \left[\mathbf{y}_1^t, \ldots, \mathbf{y}_K^t \right] \mathbf{A}_{Kd \times Kd} \left[\mathbf{y}_1^t, \ldots, \mathbf{y}_K^t \right]^t. \tag{B.33}$$

Since $\mathbf{A}_{Kd \times Kd}$ is an idempotent matrix of rank $d(K-1)$, from (B.33), we get that the random variable $\sum_{k=1}^{K} \left(\mathbf{z}_k - \sqrt{\lambda_k} \mathbf{z}_{..} \right)^t \boldsymbol{\Psi}_k^{-1} \left(\mathbf{z}_k - \sqrt{\lambda_k} \mathbf{z}_{..} \right)$ follows a noncentral chi square distribution with $d(K-1)$ degrees of freedom and non-centrality parameter $\left\| \widetilde{\mathbf{M}}_{d(K-1) \times 1} \right\|^2$. The proof is complete from (B.32). $\square$

Recall that in section 2, we have estimated the covariance operator $\boldsymbol{\Sigma}$ in Theorem 2.1 and Corollary 2.2 by $\widehat{\boldsymbol{\Sigma}}_n$. To generate observations from $G\left(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_n \right)$, the estimate $\widehat{\boldsymbol{\Sigma}}_n$ needs to be a non-negative definite operator. The non-negative definiteness of $\widehat{\boldsymbol{\Sigma}}_n$ is established in the following theorem under the condition that each of the K groups has at least two distinct observations. Clearly, this condition is satisfied almost surely when the underlying distributions of the groups are non-atomic.

Theorem B.2. *The operator $\widehat{\boldsymbol{\Sigma}}_n$ is non-negative definite whenever each of the K groups has at least two distinct observations.*

Proof. The non-negative definiteness of $\widehat{\boldsymbol{\Sigma}}_n$ is established subject to the condition that each of the K groups, $\{ \mathbf{X}_{ki} \mid i = 1, \ldots, n_k \}$, where $k = 1, \ldots, K$, has at least two distinct observations. In that case, for every group of observations $\{ \mathbf{X}_{ki} | i = 1, \ldots, n_k \}$, we can find a real-valued function $F_k(\cdot)$, such that

$$\frac{1}{n_k} \sum_{i=1}^{n_k} \left\{ F_k\left(\mathbf{X}_{ki} \right) - \frac{1}{n_k} \sum_{j=1}^{n_k} F_k\left(\mathbf{X}_{kj} \right) \right\}^2 = 1. \tag{B.34}$$

For example, suppose $\mathbf{X}_{ki_1}$ and $\mathbf{X}_{ki_2}$ be distinct observations in the kth group. Then the sample variance of the projections on $(\mathbf{X}_{ki_1} - \mathbf{X}_{ki_2})$ is positive, i.e.,

$$\frac{1}{n_k} \sum_{i=1}^{n_k} \left\{ \langle \mathbf{X}_{ki_1} - \mathbf{X}_{ki_2}, \mathbf{X}_{ki} \rangle - \frac{1}{n_k} \sum_{j=1}^{n_k} \langle \mathbf{X}_{ki_1} - \mathbf{X}_{ki_2}, \mathbf{X}_{kj} \rangle \right\}^2 > 0,$$

and one can define

$$F_k(\mathbf{x})$$

$$= \left[\frac{1}{n_k} \sum_{i=1}^{n_k} \left\{ \langle \mathbf{X}_{ki_1} - \mathbf{X}_{ki_2}, \mathbf{X}_{ki} \rangle - \frac{1}{n_k} \sum_{j=1}^{n_k} \langle \mathbf{X}_{ki_1} - \mathbf{X}_{ki_2}, \mathbf{X}_{kj} \rangle \right\}^2 \right]^{-\frac{1}{2}} \times$$

$$\langle \mathbf{X}_{ki_1} - \mathbf{X}_{ki_2}, \mathbf{x} \rangle .$$

We use these functions and the fact that every covariance operator is non-negative definite so establish the non-negative definiteness of $\widehat{\boldsymbol{\Sigma}}_n$.

Define

$$\mathbf{D}_n^{(1)}(i,j,k)$$

$$= \frac{1}{n_k} \sum_{l_k=1}^{n_k} \left[\left(\frac{1}{n_i} \sum_{l_i=1}^{n_i} \sqrt{\frac{n_k}{n_k-1}} \mathbf{s} \left(\mathbf{X}_{il_i} - \mathbf{X}_{kl_k} \right) \right) \otimes \right.$$

$$\left. \left(\frac{1}{n_j} \sum_{l_j=1}^{n_j} \sqrt{\frac{n_k}{n_k-1}} \mathbf{s} \left(\mathbf{X}_{jl_j} - \mathbf{X}_{kl_k} \right) \right) \right]$$

$$- \left(\frac{1}{n_i n_k} \sum_{l_i=1}^{n_i} \sum_{l_k=1}^{n_k} \sqrt{\frac{n_k}{n_k-1}} \mathbf{s} \left(\mathbf{X}_{il_i} - \mathbf{X}_{kl_k} \right) \right) \otimes$$

$$\left(\frac{1}{n_j n_k} \sum_{l_j=1}^{n_j} \sum_{l_k=1}^{n_k} \sqrt{\frac{n_k}{n_k-1}} \mathbf{s} \left(\mathbf{X}_{jl_j} - \mathbf{X}_{kl_k} \right) \right)$$

and

$$\mathbf{D}_n^{(2)}(i,j,k)$$

$$= \left(\frac{1}{n_i n_k} \sum_{l_i=1}^{n_i} \sum_{l_k=1}^{n_k} \frac{\mathbf{s} \left(\mathbf{X}_{il_i} - \mathbf{X}_{kl_k} \right)}{\sqrt{n_k-1}} \right) \otimes \left(\frac{1}{n_j n_k} \sum_{l_j=1}^{n_j} \sum_{l_k=1}^{n_k} \frac{\mathbf{s} \left(\mathbf{X}_{jl_j} - \mathbf{X}_{kl_k} \right)}{\sqrt{n_k-1}} \right),$$

where $i, j, k \in \{1, \ldots, K\}$. It can be verified that $\mathbf{C}_n(i,j,k) = \mathbf{D}_n^{(1)}(i,j,k) + \mathbf{D}_n^{(2)}(i,j,k)$ for all $i, j, k \in \{1, \ldots, K\}$. Define the two operators $\widehat{\boldsymbol{\Sigma}}_n^{(1)}$ and $\widehat{\boldsymbol{\Sigma}}_n^{(2)}$ as

$$\widehat{\boldsymbol{\Sigma}}_n^{(1)} = \left(\boldsymbol{\sigma}_{k_1 k_2}^{(1,n)} \right)_{K \times K} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_n^{(2)} = \left(\boldsymbol{\sigma}_{k_1 k_2}^{(2,n)} \right)_{K \times K},$$

where

$$\boldsymbol{\sigma}_{k_1 k_2}^{(1,n)} = \frac{\sqrt{n_{k_1} n_{k_2}}}{n} \sum_{l=1}^{K} \frac{n_l}{n} \left[\mathbf{D}_n^{(1)}(k_1, k_2, l) - \mathbf{D}_n^{(1)}(l, k_2, k_1) - \mathbf{D}_n^{(1)}(k_1, l, k_2) \right]$$

$$+ \sum_{l_1=1}^{K} \sum_{l_2=1}^{K} \frac{n_{l_1} n_{l_2}}{n^2} \mathbf{D}_n^{(1)}(l_1, l_2, k_1) \mathbb{I}(k_1 = k_2)$$

and

$$\boldsymbol{\sigma}_{k_1 k_2}^{(2,n)} = \frac{\sqrt{n_{k_1} n_{k_2}}}{n} \sum_{l=1}^{K} \frac{n_l}{n} \left[\mathbf{D}_n^{(2)}(k_1, k_2, l) - \mathbf{D}_n^{(2)}(l, k_2, k_1) - \mathbf{D}_n^{(2)}(k_1, l, k_2) \right]$$
$$+ \sum_{l_1=1}^{K} \sum_{l_2=1}^{K} \frac{n_{l_1} n_{l_2}}{n^2} \mathbf{D}_n^{(2)}(l_1, l_2, k_1) \mathbb{I}(k_1 = k_2).$$

Then, we have $\widehat{\boldsymbol{\Sigma}}_n = \widehat{\boldsymbol{\Sigma}}_n^{(1)} + \widehat{\boldsymbol{\Sigma}}_n^{(2)}$. We shall show that both $\widehat{\boldsymbol{\Sigma}}_n^{(1)}$ and $\widehat{\boldsymbol{\Sigma}}_n^{(2)}$ are non-negative definite, which would imply that $\widehat{\boldsymbol{\Sigma}}_n$ is also non-negative definite.

For $k = 1, \ldots, K$, define the K independent random elements $\tilde{\mathbf{X}}_k$ by

$$\tilde{\mathbf{X}}_k = \mathbf{X}_{ki} \text{ with probability } \frac{1}{n_k}, \quad i = 1, \ldots, n_k.$$

Define

$$\tilde{\mathbf{S}}_k^{(1)} = \sum_{l=1}^{K} \frac{n_l}{n} \left[\sqrt{\frac{1}{n_k - 1}} \frac{1}{n_l} \sum_{i_l=1}^{n_l} \mathbf{s} \left(\tilde{\mathbf{X}}_k - \mathbf{X}_{l i_l} \right) + \sqrt{\frac{1}{n_l - 1}} \frac{1}{n_k} \sum_{i_k=1}^{n_k} \mathbf{s} \left(\mathbf{X}_{k i_k} - \tilde{\mathbf{X}}_l \right) \right]$$

and

$$\tilde{\mathbf{S}}_k^{(2)} = \sum_{l=1}^{K} \frac{n_l}{n} \left\{ \frac{1}{n_k n_l} \sum_{i_k=1}^{n_k} \sum_{i_l=1}^{n_l} \mathbf{s} \left(\mathbf{X}_{k i_k} - \mathbf{X}_{l i_l} \right) \right\} \left[\frac{F_l \left(\tilde{\mathbf{X}}_l \right)}{\sqrt{n_l (n_l - 1)}} + \frac{F_k \left(\tilde{\mathbf{X}}_k \right)}{\sqrt{n_k (n_k - 1)}} \right].$$

Also define

$$\tilde{\mathbf{W}}_n^{(1)} = \left(\sqrt{n_1} \tilde{\mathbf{S}}_1^{(1)}, \ldots, \sqrt{n_K} \tilde{\mathbf{S}}_K^{(1)} \right) \quad \text{and} \quad \tilde{\mathbf{W}}_n^{(2)} = \left(\sqrt{n_1} \tilde{\mathbf{S}}_1^{(2)}, \ldots, \sqrt{n_K} \tilde{\mathbf{S}}_K^{(2)} \right).$$

Let $\widetilde{\mathrm{Cov}}(\cdot, \cdot)$ denote the covariance between functions of $\tilde{\mathbf{X}}_k$ with respect to the distributions of the K random elements $\tilde{\mathbf{X}}_k$, $k = 1, \ldots, K$, conditional on all the observations $\mathbf{X}_{ki}$. Then, using (B.34), it can be verified that

$$\widetilde{\mathrm{Cov}} \left(\sqrt{n_{k_1}} \tilde{\mathbf{S}}_{k_1}^{(1)}, \sqrt{n_{k_2}} \tilde{\mathbf{S}}_{k_2}^{(1)} \right) = \boldsymbol{\sigma}_{k_1 k_2}^{(1,n)} \quad \text{and} \quad \widetilde{\mathrm{Cov}} \left(\sqrt{n_{k_1}} \tilde{\mathbf{S}}_{k_1}^{(2)}, \sqrt{n_{k_2}} \tilde{\mathbf{S}}_{k_2}^{(2)} \right) = \boldsymbol{\sigma}_{k_1 k_2}^{(2,n)}$$

for all k_1, k_2. This implies that

$$\widetilde{\mathrm{Cov}} \left(\tilde{\mathbf{W}}_n^{(1)}, \tilde{\mathbf{W}}_n^{(1)} \right) = \widehat{\boldsymbol{\Sigma}}_n^{(1)} \quad \text{and} \quad \widetilde{\mathrm{Cov}} \left(\tilde{\mathbf{W}}_n^{(2)}, \tilde{\mathbf{W}}_n^{(2)} \right) = \widehat{\boldsymbol{\Sigma}}_n^{(2)},$$

which in turn imply that $\widehat{\boldsymbol{\Sigma}}_n^{(1)}$ and $\widehat{\boldsymbol{\Sigma}}_n^{(2)}$ are covariance operators, and hence non-negative definite. Therefore, $\widehat{\boldsymbol{\Sigma}}_n$ is also non-negative definite whenever each of the groups has at least two distinct observations. $\qquad \square$

TABLE 6
Estimated sizes of the asymptotic and the permutation implementations of the SS test under nominal level 5%; $n_1 = n_2 = n_3 = 5, 10, 20$ and 40.

P_0	$n_1 = n_2 = n_3 = 5$		$n_1 = n_2 = n_3 = 10$	
	Asymptotic	Permutation	Asymptotic	Permutation
SBM	0.050	0.055	0.054	0.051
$t_{1,\text{SBM}}$	0.047	0.047	0.052	0.045
$t_{3,\text{SBM}}$	0.058	0.057	0.050	0.050
SBM with impurity	0.046	0.047	0.053	0.047
$t_{1,\text{SBM}}$ with impurity	0.046	0.047	0.048	0.047
$t_{3,\text{SBM}}$ with impurity	0.058	0.057	0.055	0.050
GBM	0.037	0.041	0.055	0.049
$(\text{SBM})^2$	0.036	0.047	0.053	0.057
$(t_{3,\text{SBM}})^2$	0.034	0.053	0.042	0.047

P_0	$n_1 = n_2 = n_3 = 20$		$n_1 = n_2 = n_3 = 40$	
	Asymptotic	Permutation	Asymptotic	Permutation
SBM	0.057	0.043	0.051	0.053
$t_{1,\text{SBM}}$	0.055	0.053	0.057	0.052
$t_{3,\text{SBM}}$	0.046	0.043	0.054	0.046
SBM with impurity	0.055	0.047	0.049	0.044
$t_{1,\text{SBM}}$ with impurity	0.055	0.042	0.059	0.047
$t_{3,\text{SBM}}$ with impurity	0.054	0.050	0.045	0.041
GBM	0.052	0.044	0.050	0.046
$(\text{SBM})^2$	0.056	0.051	0.052	0.047
$(t_{3,\text{SBM}})^2$	0.047	0.049	0.053	0.049

Appendix C: Comparison of asymptotic and permutation implementations of the SS test

Here, we present the level and the power comparisons of the asymptotic and the permutation implementations of the SS test. We present the estimated sizes and the power curves of the two implementations at nominal level 5% in the models considered in subsection 3.3. For the size study, we consider 3 groups and the underlying distribution P_0 of the groups are as in subsection 3.3 with 4 cases of the group sizes: $n_1 = n_2 = n_3 = 5, 10, 20$ and 40. For the power study, we consider the sizes of the 3 groups as 10, 10 and 10, and the underlying distribution P_0 of the groups are as in subsection 3.3. The shifts of the 3 groups, $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$, are kept the same as in subsection 3.3. The data generation process is also the same. The number of random permutations taken is 1000, and the sizes and the powers of the test procedures are estimated using 1000 independent replications.

The estimated sizes of the two implementations are presented in Table 6. The estimated power curves are presented in Figure 5. We note that the powers of the asymptotic and the permutation implementations are virtually identical. But the size of the asymptotic implementation is for a few cases of the underlying distribution somewhat lower than the nominal level of 5% when the group sizes are small, i.e., 5. The estimated sizes of the permutation implementation are satisfactory for both small and larger sample sizes.
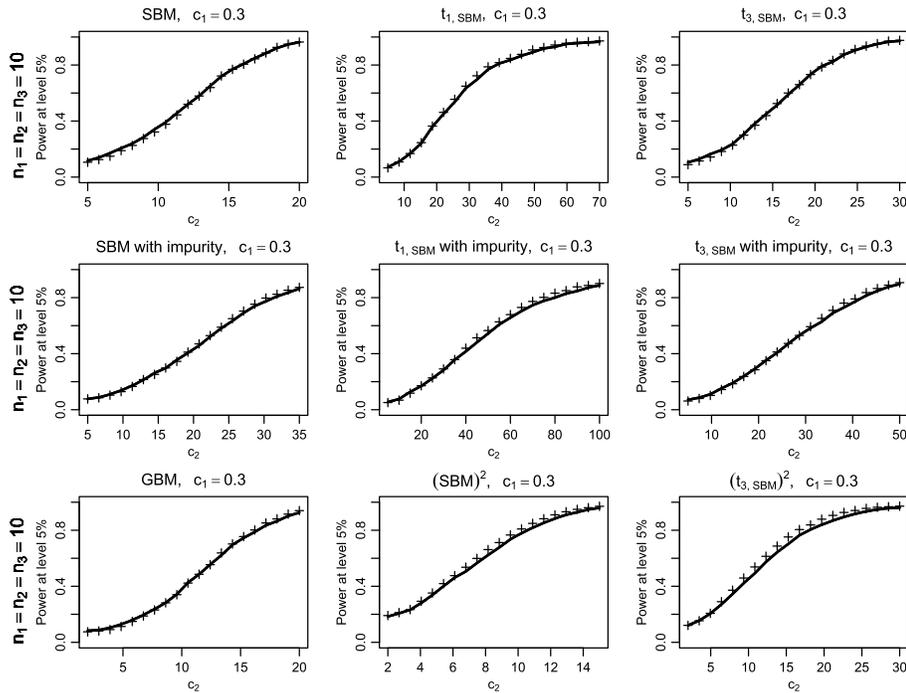
Fig 5. *Estimated power curves of the asymptotic implementation (−) and the permutation implementation (++) of the SS test under nominal level 5%.*

R Codes

The R functions to compute the SS test are available in https://github.com/joydeepchowdhury/spatialsign_functional_anova.

References

[1] ARISTIZABAL, J.-P., GIRALDO, R. and MATEU, J. (2019). Analysis of variance for spatially correlated functional data: application to brain data. *Spatial Statistics* **32** 100381. MR3980907

[2] BHATIA, R. (2009). *Notes on Functional Analysis* **50**. Hindustan Book Agency. MR2477477

[3] BILLINGSLEY, P. (2013). *Convergence of Probability Measures*. John Wiley & Sons. MR0233396

[4] BOROVSKIKH, I. I. V. (1996). *U-statistics in Banach Spaces*. VSP. MR1419498

[5] BRUNNER, E., DETTE, H. and MUNK, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association* **92** 1494–1502. MR1615259

[6] BRUNNER, E., KONIETSCHKE, F., PAULY, M. and PURI, M. L. (2017). Rank-based procedures in factorial designs: Hypotheses about nonparametric treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 1463–1485. MR3731671

[7] CARDOT, H., CÉNAC, P. and ZITT, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli* **19** 18–43. MR3019484

[8] CAUSEUR, D., SHEU, C.-F., PERTHAME, E. and RUFINI, F. (2019). A functional generalized F-test for signal detection with applications to event-related potentials significance analysis. *Biometrics*. MR4098559

[9] CHAKRABORTY, A. and CHAUDHURI, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The Annals of Statistics* **42** 1203–1231. MR3224286

[10] CHAKRABORTY, A. and CHAUDHURI, P. (2015). A Wilcoxon–Mann–Whitney-type test for infinite-dimensional data. *Biometrika* **102** 239–246. MR3335109

[11] CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association* **91** 862–872. MR1395753

[12] CHOI, K. and MARDEN, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *Journal of the American Statistical Association* **92** 1581–1590. MR1615267

[13] CHOI, K. and MARDEN, J. (2002). Multivariate analysis of variance using spatial ranks. *Sociological Methods & Research* **30** 341–366. MR1873709

[14] CUESTA-ALBERTOS, J. and FEBRERO-BANDE, M. (2010). A simple multiway ANOVA for functional data. *Test* **19** 537–557. MR2746001

[15] CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2004). An anova test for functional data. *Computational Statistics & Data Analysis* **47** 111–122. MR2087932

[16] DELAIGLE, A., HALL, P. and JIN, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73** 283–301. MR2815777

[17] EFRON, B. and TIBSHIRANI, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press. MR1270903

[18] GÓRECKI, T. and SMAGA, Ł. (2015). A comparison of tests for the oneway ANOVA problem for functional data. *Computational Statistics* **30** 987–1010. MR3433439

[19] GÓRECKI, T. and SMAGA, Ł. (2019). fdANOVA: an R software package for analysis of variance for univariate and multivariate functional data. *Computational Statistics* **34** 571–597. MR3953673

[20] GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research* **13** 723–773. MR2913716

[21] Guo, J. and Chen, Y. (2019). An L^2-norm based ANOVA test for the equality of weakly dependent functional time series. *Statistics and Its Interface* **12** 167–180. MR3876418

[22] Harrar, S. W. and Kong, X. (2022). Recent developments in high-dimensional inference for multivariate data: Parametric, semiparametric and nonparametric approaches. *Journal of Multivariate Analysis* **188** 104855. MR4353864

[23] Horváth, L. and Rice, G. (2015). An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Revista Matemática Complutense* **28** 505–548. MR3379038

[24] Janssen, A. and Pauls, T. (2003). How do bootstrap and permutation tests work? *The Annals of Statistics* **31** 768–806. MR1994730

[25] Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics* **25** 435–477. MR1439309

[26] Kong, X. and Harrar, S. W. (2020). High-dimensional rank-based inference. *Journal of Nonparametric Statistics* **32** 294–322. MR4101381

[27] Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* **47** 583–621.

[28] Kundu, S., Majumdar, S. and Mukherjee, K. (2000). Central limit theorems revisited. *Statistics & Probability Letters* **47** 265–275. MR1747487

[29] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics* **5** 201–213. MR1346895

[30] Möttönen, J., Oja, H. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics* **25** 542–552. MR1439313

[31] Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Springer Science & Business Media. MR2598854

[32] Pauly, M., Brunner, E. and Konietschke, F. (2015). Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77** 461–473. MR3310535

[33] Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statistica Sinica* **14** 1239–1257. MR2126351

[34] Shinohara, R. T., Shou, H., Carone, M., Schultz, R., Tunc, B., Parker, D., Martin, M. L. and Verma, R. (2019). Distance-based analysis of variance for brain connectivity. *Biometrics*. MR4098560

[35] Zhang, J.-T. (2011). Statistical inferences for linear models with functional responses. *Statistica Sinica* **21** 1431–1451. MR2827530

[36] Zhang, J.-T. (2013). *Analysis of Variance for Functional Data. Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. CRC Press.

[37] Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics* **35** 1052–1079. MR2341698

[38] ZHANG, J.-T., CHENG, M.-Y., WU, H.-T. and ZHOU, B. (2019). A new test for functional one-way ANOVA with applications to ischemic heart screening. *Computational Statistics & Data Analysis* **132** 3–17. MR3913131

[39] ZHANG, J.-T. and LIANG, X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics* **41** 51–71. MR3181132