

Article

# Generalized Nonparametric Composite Tests for High-Dimensional Data

Xiaoli Kong<sup>1</sup>, Alejandro Villasante-Tezanos<sup>2</sup>  and Solomon W. Harrar<sup>3,\*</sup> <sup>1</sup> Department of Mathematics, Wayne State University, Detroit, MI 48202, USA; kong@wayne.edu<sup>2</sup> Department of Biostatistics and Data Science, School of Public and Population Health, University of Texas Medical Branch, Galveston, TX 77555, USA; alvillas@utmb.edu<sup>3</sup> Dr. Bing Zhang Department of Statistics, University of Kentucky, Lexington, KY 40506, USA

\* Correspondence: solomon.harrar@uky.edu

**Abstract:** In this paper, composite high-dimensional nonparametric tests for two samples are proposed, by using component-wise Wilcoxon–Mann–Whitney-type statistics. No distributional assumption, moment condition, or parametric model is required for the development of the tests and the theoretical results. Two approaches are employed, for estimating the asymptotic variance of the composite statistic, leading to two tests. In both cases, banding of the covariance matrix to estimate variance of the test statistic is involved. An adaptive algorithm, for selecting the banding window width, is proposed. Numerical studies are provided, to show the favorable performance of the new tests in finite samples and under varying degrees of dependence.

**Keywords:** high dimension; two-sample test; Wilcoxon–Mann–Whitney; nonparametric;  $\alpha$ -mixing



**Citation:** Kong, X.; Villasante-Tezanos, A.; Harrar, S.W. Generalized Nonparametric Composite Tests for High-Dimensional Data. *Symmetry* **2022**, *14*, 1153. <https://doi.org/10.3390/sym14061153>

Academic Editor: Dalibor Štys

Received: 15 April 2022

Accepted: 29 May 2022

Published: 2 June 2022

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



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

## 1. Introduction

Recent advances in technology have allowed for the collection of data in high frequency and resolution. Sparked by these advances, high-dimensional data have been a subject of theoretical and applied investigations, in the last few decades. High-dimensional refers to the situation where the dimension is (much) larger than the sample size. Examples include data from genomic studies, biological studies, financial studies, satellite imaging, modern diagnostic and intervention modalities, etc. To analyze these data, in particular in the context of group or treatment comparison, the asymptotic theory requires both the sample sizes and dimensions to diverge. For an extensive account of the literature on this subject, we refer the reader to the review article by Harrar and Kong [1].

The classical parametric tests (e.g., [2]) are not applicable because the sample covariance matrices are singular. Even in the nonsingular situation, these tests suffer from low power [3]. The problem is worse, if the distribution of the data is heavy-tailed, or if the data is measured in a nonmetric scale, as the usual mean-based hypothesis and inference are not well defined anymore.

Nonparametric methods are well known for being more robust against nonnormality and other conditions, than their parametric counterparts. Classical nonparametric theory formulates tests, considering distribution functions rather than parameters. The challenges with these formulations are that (a) the alternative hypothesis is difficult to interpret and (b) the tests can not be pivoted to construct confidence intervals. To overcome these challenges, some characteristic of the distribution functions is, often, investigated to compare treatments. In our interest, to be as general as possible, we will employ a nonparametric test statistic, for quantifying group or treatment differences.

In the finite dimensional situation (small  $p$  and large  $n$ ), the problem of testing hypotheses formulated in terms of the nonparametric relative effects have been considered in several papers (e.g., [4–8]). A related topic to high-dimensional inference is the asymptotic setting, where the number of treatments goes to infinity but the sample size per

treatment could be fixed or large. These problems have been thoroughly investigated in nonparametric situations [9–13]. High-dimensional theory for nonparametric methods is underdeveloped. To the best of our knowledge, there are only two papers that have studied this problem [14,15]. Wang and Akritas [14] focuses on hypotheses formulated in terms of marginal distributions, while Kong and Harrar [15] considered the more general situation, in which the hypotheses of interest are formulated in terms of the nonparametric relative effects. The underlying idea in both works [14,15] is to make the stochastic comparison of each marginal distribution relative to the overall (average of all marginal) distributions. Therefore, the variables must be commensurate, for these methods to be appropriate. It is one aim of the present manuscript to overcome this challenge, by defining the relative effects, marginally.

Consider two mutually independent random samples  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i} \in \mathbb{R}^p$  for  $i \in \{1, 2\}$ . For each  $i$ , denote  $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijp})^\top$ , where the random variable  $X_{ijk}$  is the  $k^{\text{th}}$  variable of the  $j^{\text{th}}$  subject from the  $i^{\text{th}}$  sample group. Denote the total sample size by  $n = n_1 + n_2$ . To be as general as possible, we assume the nonparametric model

$$X_{ijk} \sim F_{ik}, \text{ for } j = 1, \dots, n_i,$$

where  $F_{ik}(\cdot)$  is an arbitrary non-degenerate distribution function. In order to accommodate binary, ordered categorical, discrete, and continuous data types in a unified manner, we will use the normalized version of the distribution function, defined as

$$F_{ik}(x) = \frac{1}{2} \{ F_{ik}^+(x) + F_{ik}^-(x) \} = P(X_{i1k} < x) + \frac{1}{2} P(X_{i1k} = x),$$

where  $F_{ik}^-(x) = P(X_{i1k} < x)$  and  $F_{ik}^+(x) = P(X_{i1k} \leq x)$  are the left and right continuous versions of the distribution function [16–18].

In our investigation to compare group effects, we study the so-called nonparametric relative group effects. The *relative effect* for the  $k^{\text{th}}$  variable is defined by

$$\omega_k = P(X_{11k} < X_{21k}) + \frac{1}{2} P(X_{11k} = X_{21k}) = \int F_{1k} dF_{2k}. \quad (1)$$

When  $\omega_k$  is greater than  $1/2$ , we interpret that observations on the  $k^{\text{th}}$  variable in the first sample tend to have smaller values, than observations on the  $k^{\text{th}}$  variable in the second sample and vice versa, if  $\omega_k$  is smaller than  $1/2$ . If  $\omega_k = 1/2$ , the two variables are tendentiously equal. It is obvious that if  $F_{1k} = F_{2k}$ , then  $\omega_k = 1/2$ . However,  $\omega_k = 1/2$  does not necessarily imply  $F_{1k} = F_{2k}$ . For example, assume  $X_{i1k}$  has normal distribution with mean  $\mu_{ik}$  and standard deviation  $\sigma_{ik}$ , for each  $i$ . It is easy to see that  $\omega_k = \frac{1}{2}$ , if and only if  $\mu_{1k} = \mu_{2k}$ , but  $F_{1k}$  and  $F_{2k}$  may have different standard deviations. Motivated by this, we consider a nonparametric hypothesis testing, about  $\omega_k$ 's in the high-dimensional framework. The multivariate relative effect of interest combines all marginal effects,  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_p)^\top$ . The global non-parametric hypothesis in terms of these relative effects is

$$H_0 : \boldsymbol{\omega} = \frac{1}{2} \mathbf{1}_p. \quad (2)$$

A parametric hypothesis could be a special case of (2), e.g., for the multivariate normal distribution.

In this paper, we investigate nonparametric tests, for testing the hypothesis  $H_0$  in high dimension. Our tests are applicable, even in situations where the variables are measured in binary, ordinal, and continuous scales, or a mixture of these. The main contributions are: (a) we construct a nonparametric test in high dimension. A composite of variable-by-variable univariate Wilcoxon–Mann–Whitney type tests of Brunner and Munzel [4] is proposed. We develop the asymptotic theory of the test statistics, without requiring any distributional assumption, moment condition, or parametric model. Motivated by Gregory et al. [19] and Zhang and Wang [20], two versions of the scaling parameter are

constructed. (b) We provide an adaptive algorithm for the asymptotic variance estimation. (c) We have shown, with a simulation study, that the proposed tests have superior performance compared to existing methods, especially for heavy-tailed data, where the parametric tests are not applicable.

The remainder of the paper is organized as follows. We propose the rank-based-test statistic in Section 2. The asymptotic results for the test statistic and the results are used to construct hypothesis tests are stated in Section 3. The performance of the tests is investigated via simulation study in Section 4, together with the adaptive algorithm for selecting the banding window width. Finally, we will end the paper with some conclusions in Section 5. All technical details and proofs of lemmas are shifted to the Appendix A.

## 2. Test Statistic

### 2.1. Preliminaries

For any  $k \in \{1, \dots, p\}$ , define the transformed variables  $Y_{1jk}$  and  $Y_{2jk}$  by

$$Y_{1jk} = F_{2k}(X_{1jk}) \text{ for } j \in \{1, \dots, n_1\}, \text{ and } Y_{2jk} = F_{1k}(X_{2jk}), \text{ for } j \in \{1, \dots, n_2\}.$$

Let  $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijp})^\top$  for  $i \in \{1, 2\}$  and  $j \in \{1, \dots, n_i\}$  be a vector of the transformed variables. Note that  $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in_i}$  are iid, and the two samples  $\{\mathbf{Y}_{1j} : j = 1, \dots, n_1\}$  and  $\{\mathbf{Y}_{2j} : j = 1, \dots, n_2\}$  are, still, independent. Further, let  $\bar{\mathbf{Y}}_i = (\bar{Y}_{i1}, \dots, \bar{Y}_{ip})^\top$  and  $\mathbf{S}_i = (S_{ikk'})_{k,k'=1}^p$  be the sample mean vector and sample covariance matrix for the  $i$ th sample, where  $\bar{Y}_{ik}$  and  $S_{ikk}$  (also denoted by  $S_{ik}^2$ ) are the sample mean and sample variance, respectively, of the  $k$ -th variable, and  $S_{ikk'}$  is the sample covariance between the  $k$ -th and  $k'$ -th variables.

Note that  $Y_{ijk} \in [0, 1]$  and is a non-degenerate random variable. Moments of any order of  $Y_{ijk}$  exist, and  $\sigma_{ik}^2 = \text{Var}(Y_{ijk}) > 0$ . In particular,

$$E(Y_{1jk}) = 1 - \omega_k \quad \text{and} \quad E(Y_{2jk}) = \omega_k.$$

The null hypothesis (2) is equivalent to the equality of mean vectors of  $\mathbf{Y}_{11}$  and  $\mathbf{Y}_{21}$ , i.e.,  $\boldsymbol{\omega} = \mathbf{1}_p - \boldsymbol{\omega}$ . The problem of testing equality of mean vectors has been, extensively, studied under the parametric and semiparametric contexts (see [1]). One of the test statistics that has received considerable attention for testing

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ against } H_0 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

is a composite statistic, formed by averaging the squared two-sample  $t$ -statistic over the  $p$  dimensions. More specifically,

$$T_n = p^{-1} \sum_{k=1}^p t_k^2, \quad \text{where } t_k = \frac{\bar{Y}_{2k} - \bar{Y}_{1k}}{\sqrt{S_{1k}^2/n_1 + S_{2k}^2/n_2}}. \quad (3)$$

Intuitively, this statistic quantifies the “total” separation between two groups, across all the dimensions.

To maintain an approximate balance between the two samples, we assume  $n_1$  and  $n_2$  have the same order of magnitude, i.e.,  $n_i/n \rightarrow \lambda_i \in (0, 1)$  as  $n_1, n_2 \rightarrow \infty$ . Further, to standardize  $T_n$  in high dimension, various estimators for variance of  $T_n$  have been proposed (e.g., [19–21]). Srivastava and Du [21] makes multivariate normality assumption, which was later, generalized, to a nonnormal situation by Srivastava et al. [22].

The test by Gregory et al. [19] assumed  $\{t_k^2 : k = 1, \dots, p\}$  to be an  $\alpha$ -mixing sequence and proposed the so-called Generalized Component Test (GCT) statistic, by normalizing  $T_n$  as

$$T_{\text{GCT}} = \frac{\sqrt{p}(T_n - \tilde{\zeta}_n)}{\tilde{\zeta}_n}. \quad (4)$$

Two estimators of the centering parameter  $\tilde{\xi}_n$  suitable for moderate- $p$  and large- $p$  were proposed. For moderate- $p$ , more precisely  $p = o(n^2)$ , the center estimator is  $\tilde{\xi}_n = 1$ . For large- $p$ , specifically  $p = o(n^6)$ , it was shown that  $E[T_n] = 1 + n^{-1}a_n + n^{-2}b_n + O(n^{-3})$  under some moment conditions and, thus, center estimator

$$\tilde{\xi}_n = 1 + \frac{1}{n} \tilde{a}_n + \frac{1}{n^2} \tilde{b}_n, \tag{5}$$

was proposed, where the rather lengthy formulas for  $\tilde{a}_n$  and  $\tilde{b}_n$  can be found in their paper. The finite sample corrections in the second and third terms on the right-hand side of (5) guard against finite sample bias, from the large-sample approximation of the center. This bias could accumulate considerably, as the dimension  $p$  gets large. The scaling quantity  $\tilde{\xi}_n$  was estimated by a smoothed version of the sample autocovariance function of  $\{t_k^2 : k = 1, \dots, p\}$  as

$$\tilde{\xi}_n = \sum_{|r| < L} w(r/L) \tilde{\gamma}(r), \text{ where } \tilde{\gamma}(r) = \frac{1}{p - |r|} \sum_{k=1}^{p-|r|} (t_k^2 - T_n)(t_{k+|r|}^2 - T_n), \tag{6}$$

where  $L$  is a user-selected window width, and  $w(\cdot)$  is an even-weight function. Such weight functions are, commonly, used in spectral-density estimation (e.g., [23]) and satisfy  $w(0) = 1, |w(x)| \leq 1$  for  $|x| \leq 1$  and  $w(x) = 0$  for  $|x| > 1$ . Two examples of smoothing weight functions that were used in Gregory et al. [19] are the Parzen window and the trapezoid window [24], defined by

$$w_p(r/L) = \begin{cases} 1 - 6|r/L|^2 + 6|r/L|^3, & \text{if } 0 \leq |r| < L/2; \\ 2(1 - |r/L|)^3, & \text{if } L/2 \leq |r| \leq L; \\ 0 & \text{if } |r| > L, \end{cases}$$

and

$$w_t(r/L) = \begin{cases} 1, & \text{if } 0 \leq |r| < L/2; \\ 1 - \frac{|r| - \lfloor L/2 \rfloor}{L - \lfloor L/2 \rfloor}, & \text{if } L/2 \leq |r| \leq L; \\ 0 & \text{if } |r| > L, \end{cases}$$

respectively, where  $\lfloor L/2 \rfloor$  denotes the largest integer not exceeding  $L/2$ .

More recently, Zhang and Wang [20] proposed a more efficient scaling quantity and, therefore, a potentially more powerful test (MPT) statistic can be defined, by making  $\alpha$ -mixing assumption on the original data. Denote  $\sigma_{ikk'} = Cov(Y_{i1k}, Y_{i1k'})$  and  $\sigma_{ik}^2 = Var(Y_{i1k})$  for  $i \in \{1, 2\}$ . Applying Lemma 2.1 of Zhang and Wang [20], it follows that  $Cov(t_k^2, t_{k'}^2) = \gamma_{kk'} + O(n^{-1/2})$  for any  $k, k' \in \{1, \dots, p\}$ , where

$$\gamma_{kk'} = \frac{2(\sigma_{1kk'} / \lambda_1 + \sigma_{2kk'} / \lambda_2)^2}{(\sigma_{1k}^2 / \lambda_1 + \sigma_{2k}^2 / \lambda_2)(\sigma_{1k'}^2 / \lambda_1 + \sigma_{2k'}^2 / \lambda_2)}. \tag{7}$$

Estimating  $\sigma_{ikk'}$  and  $\sigma_{ik}^2$ , by the sample covariance  $S_{ikk'}$  and sample variance  $S_{ik}^2$ , respectively, the quantity  $\gamma_{kk'}$  can be estimated by

$$\tilde{\gamma}_{kk'} = \frac{2(S_{1kk'} / \lambda_1 + S_{2kk'} / \lambda_2)^2}{(S_{1k}^2 / \lambda_1 + S_{2k}^2 / \lambda_2)(S_{1k'}^2 / \lambda_1 + S_{2k'}^2 / \lambda_2)}.$$

Note that the assumptions of the aforementioned Lemma 2.1, of Zhang and Wang [20], are automatically satisfied for  $Y_{ijk}$ . Zhang and Wang [20] proposed the estimator of  $\text{Var}(\sqrt{p}T_n)$  as

$$\tilde{\tau}_n^2 = \frac{1}{p} \sum_{|k-k'| \leq L} \tilde{\gamma}_{kk'}, \quad (8)$$

where  $L = \lceil p^\epsilon \rceil$  is a window width for some  $0 < \epsilon < 1$ . The banded partial sum is intended to capture the important correlations among neighboring observations and facilitates consistent estimation of the asymptotic variance. Zhang and Wang [20] suggested selecting  $L = \lceil p^{3/8} \rceil$  when  $p \geq 300$  for practical applications. Incidentally, they, also, noted from their simulation study that the type I error rate and power of GCT change, drastically, with varying window width. We will discuss more about window-width selection in Section 4. They, numerically, validated that the estimator  $\tilde{\tau}_n^2$  provides an improvement over  $\tilde{\zeta}_n^2$ , for scaling  $T_n$ . An intuitive explanation is that  $\tilde{\tau}_n^2$  makes use of the replications in the sample, directly, to reduce the variability.

The MPT statistic is, then, defined by normalizing  $T_n$  as

$$T_{\text{MPT}} = \frac{\sqrt{p}(T_n - \tilde{\zeta}_n)}{\tilde{\tau}_n}. \quad (9)$$

It was noted in Zhang and Wang [20] that both GCT and MPT yield liberal type I error rates, when using the large- $p$  version of the centering parameter  $\tilde{\zeta}_n$ , given in (5). This happens, especially, in small  $n$ , which may be due to the fact that the third and fourth moments involved in the estimator  $\tilde{\zeta}_n$ , often, require large sample size to achieve a reasonable accuracy. Both Gregory et al. [19] and Zhang and Wang [20] proposed to reject null hypothesis with two-tailed rejection regions.

However, the transformed variables  $\{Y_{ijk}\}$  are not observable and, thus, the test statistics  $T_{\text{GCT}}$  or  $T_{\text{MPT}}$  cannot be, directly, computed to test the nonparametric null hypothesis (2). Alternatively, the empirical distribution functions will be substituted in places of the distribution functions, to construct valid nonparametric tests. This strategy is investigated in detail in the next subsection.

## 2.2. Nonparametric Tests

In our approach, we seek rank-based estimates for the relative effects and use these estimates to construct a nonparametric test. To that end, define  $\hat{F}_{ik} = \frac{1}{2}(\hat{F}_{ik}^+ + \hat{F}_{ik}^-)$ , where  $\hat{F}_{ik}^+$  and  $\hat{F}_{ik}^-$  are the right- and left-continuous empirical distribution functions, respectively. More specifically,  $\hat{F}_{ik}(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} c(x - X_{ijk})$ , where  $c(u) = 0, 1/2$ , or  $1$  according to  $u < 0, u = 0$ , or  $u > 0$ . For  $k \in \{1, \dots, p\}$ , define  $\hat{Y}_{1jk} = \hat{F}_{2k}(X_{1jk})$  for  $j \in \{1, \dots, n_1\}$  and  $\hat{Y}_{2jk} = \hat{F}_{1k}(X_{2jk})$  for  $j \in \{1, \dots, n_2\}$ .

It is natural to estimate the relative effects  $\omega_k$ , by replacing the distribution functions  $F_{ik}$  with their empirical counterparts  $\hat{F}_{ik}$ . That is,

$$\hat{\omega}_k = \int \hat{F}_{1k} d\hat{F}_{2k} = \frac{1}{n_2} \sum_{j=1}^{n_2} \hat{F}_{1k}(X_{2jk}) = \frac{1}{n_1 n_2} \sum_{\ell=1}^{n_1} \sum_{j=1}^{n_2} c(X_{2jk} - X_{1\ell k}),$$

which is, actually, the sample mean of  $\hat{Y}_{2jk}$  for  $k \in \{1, \dots, p\}$  and can be computed in terms of component-wise ranks as [4],

$$\hat{\omega}_k = \frac{1}{n_1 n_2} \sum_{j=1}^{n_2} \left( R_{2jk} - R_{jk}^{(2)} \right) = \frac{1}{n_1} \left( \bar{R}_{2 \cdot k} - \frac{n_2 + 1}{2} \right).$$

To see this, observe that  $\widehat{Y}_{2jk} = \frac{1}{n-n_i}(R_{2jk} - R_{jk}^{(2)})$  and  $\sum_{j=1}^{n_i} R_{jk}^{(i)} = \frac{n_i(n_i-1)}{2}$ , where  $R_{ijk}$  refers to the (mid-)rank of  $X_{ijk}$  among all  $n$  observations of the  $k$ -th variable within the two samples,  $R_{jk}^{(i)}$  refers to the (mid-)rank of  $X_{ijk}$ , among the  $n_i$  observations of the  $k$ -th variable, within sample  $i$ , and  $\bar{R}_{i\cdot k} = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ijk}$ . More precisely,  $R_{ijk} = n_1 \widehat{F}_{1k}(X_{ijk}) + n_2 \widehat{F}_{2k}(X_{ijk}) + 1/2$  and  $R_{jk}^{(i)} = n_i \widehat{F}_{ik}(X_{ijk}) + 1/2$ .

Note that  $\widehat{\omega}_k$  is the Wilcoxon–Mann–Whitney-type  $U$ -statistic. It is well known that, even in the presence of ties, the estimator  $\widehat{\omega}_k$  is  $L_2$ -consistent for  $\omega_k$ , under the asymptotic framework that  $n_1$  and  $n_2$  diverge proportionally [4,17]. It is, also, well known that, under the same asymptotic framework,  $\sqrt{n}(\widehat{\omega}_k - \omega_k)/\sigma_k$  has, asymptotically, a standard normal distribution,  $\mathcal{N}(0, 1)$ , where  $\sigma_k^2 = n(\sigma_{1k}^2/n_1 + \sigma_{2k}^2/n_2)$  (see, also, Theorem 2 of [6]). An  $L_2$ -consistent rank-based estimator for  $\sigma_k^2$  can be constructed, as given in Brunner and Munzel [4]

$$\widehat{\sigma}_k^2 = n(\widehat{\sigma}_{1k}^2/n_1 + \widehat{\sigma}_{2k}^2/n_2),$$

where

$$\widehat{\sigma}_{ik}^2 = \frac{1}{(n - n_i)^2(n_i - 1)} \sum_{j=1}^{n_i} \left( R_{ijk} - R_{jk}^{(i)} - \bar{R}_{i\cdot k} + \frac{n_i + 1}{2} \right)^2 \tag{10}$$

is the sample variance of  $\widehat{Y}_{ijk}$  for  $j \in \{1, \dots, n_i\}$ . Thus, it follows, immediately, from Slutsky’s theorem that  $\sqrt{n}(\widehat{\omega}_k - \omega_k)/\widehat{\sigma}_k$  has, asymptotically, a standard normal distribution,  $\mathcal{N}(0, 1)$ . Some of these results are summarized in Lemmas A1–A7 of Appendix A.

Under the null hypothesis (2), denote

$$\widehat{t}_k = \sqrt{n}(\widehat{\omega}_k - 1/2)/\widehat{\sigma}_k.$$

Obviously,  $\widehat{t}_k$  and  $t_k$  have the same asymptotic distribution under the null hypothesis (2). In the following Lemma, the mean of  $\widehat{t}_k^2$  is given, up to the order of  $n^{-1}$ .

**Lemma 1.** Assume  $\lim_{n \rightarrow \infty} n_i/n \rightarrow \lambda_i \in (0, 1)$  for  $i = 1, 2$  and  $\min\{\sigma_{1k}^2, \sigma_{2k}^2\} > 0$ , for all  $k \in \{1, \dots, p\}$ . Then, under  $H_0$ ,  $E(\widehat{t}_k^2) = 1 + O(n^{-1})$ , for any  $k \in \{1, \dots, p\}$ . Moreover,  $\sup_k E[\widehat{t}_k^{2r}] < \infty$ , for any integer  $r > 1$ .

The covariance of  $\widehat{t}_k^2$  and  $\widehat{t}_{k'}^2$ , for any  $k, k' \in \{1, \dots, p\}$ , are given up to the order of  $n^{-1/2}$ , in Lemma 2. It is shown that the covariance  $\text{Cov}(\widehat{t}_k^2, \widehat{t}_{k'}^2)$  shares the same dominating term as  $\text{Cov}(t_k^2, t_{k'}^2)$ .

**Lemma 2.** Assume  $\lim_{n \rightarrow \infty} n_i/n \rightarrow \lambda_i \in (0, 1)$  for  $i = 1, 2$  and  $\min\{\sigma_{1k}^2, \sigma_{2k}^2\} > 0$  for all  $k \in \{1, \dots, p\}$ . Then, under  $H_0$ ,  $\text{Cov}(\widehat{t}_k^2, \widehat{t}_{k'}^2) = \gamma_{kk'} + O(n^{-1/2})$ , for any  $k, k' \in \{1, \dots, p\}$ , where  $\gamma_{kk'}$  is defined in (7).

The core component of our test is the composite statistic

$$\widehat{T}_n = \frac{1}{p} \sum_{k=1}^p \widehat{t}_k^2. \tag{11}$$

The motivation behind this construction is that in many high-dimensional applications, e.g., transcriptomics, the effect on any individual variable (gene expression) may be small, and composite statistics will have high power in detecting the cumulative effect. To develop

a test statistic for testing the null hypothesis (2), based on (11), there are two options for scaling  $\hat{T}_n$ . One is replacing  $t_k$  by  $\hat{t}_k$  in (6), i.e., define

$$\hat{\zeta}_n = \sum_{|r|<L} w(r/L) \hat{\gamma}(r), \quad \text{where } \hat{\gamma}(r) = \frac{1}{p-|r|} \sum_{k=1}^{p-|r|} (\hat{t}_k^2 - \hat{T}_n)(\hat{t}_{k+|r|}^2 - \hat{T}_n), \quad (12)$$

and employ similar results, as in Gregory et al. [19]. The other option is using a scale similar to (8) and define

$$\hat{\tau}_n^2 = \frac{1}{p} \sum_{|k-k'|\leq L} \hat{\gamma}_{kk'}, \quad \text{where } \hat{\gamma}_{kk'} = \frac{2(\hat{\sigma}_{1kk'}/\lambda_1 + \hat{\sigma}_{2kk'}/\lambda_2)^2}{(\hat{\sigma}_{1k}^2/\lambda_1 + \hat{\sigma}_{2k}^2/\lambda_2)(\hat{\sigma}_{1k'}^2/\lambda_1 + \hat{\sigma}_{2k'}^2/\lambda_2)}, \quad (13)$$

$\hat{\sigma}_{ik}^2$  is the  $L_2$ -consistent estimator for  $\sigma_{ik}^2$ , defined in (10), and  $\hat{\sigma}_{ikk'}$  is the  $L_2$ -consistent estimator of  $\sigma_{ikk'}$ , defined by (see Lemma A8),

$$\hat{\sigma}_{ikk'} = \frac{1}{(n-n_i)^2(n_i-1)} \sum_{j=1}^{n_i} \left( R_{ijk} - R_{jk}^{(i)} - \bar{R}_{i\cdot k} + \frac{n_i+1}{2} \right) \left( R_{ijk'} - R_{jk'}^{(i)} - \bar{R}_{i\cdot k'} + \frac{n_i+1}{2} \right), \quad (14)$$

which is the sample covariance of  $\hat{Y}_{ijk}$  and  $\hat{Y}_{ijk'}$  for  $j \in \{1, \dots, n_i\}$ . The comparison of the two scaling parameters is investigated, numerically, in Section 4. The main difference between these two scaling quantities given in (12) and (13) is that (13) uses the data more efficiently and, hence, is expected to enhance the power of the test. The window width can be, similarly, set as  $L = \lceil p^\epsilon \rceil$ , for some  $0 < \epsilon < 1$ , as for MPT. However, the choice of  $\epsilon$  should depend not only on the dimension  $p$  but also on the underlying distribution of the data. It can be seen from the simulation that there is no fixed number  $\epsilon$ , for which the window width  $L = \lceil p^\epsilon \rceil$  works for all models. Thus, we provide an algorithm in Section 4 for choosing  $L$ , for use in practice.

For the moderate- $p$  setting, specifically  $p = o(n^2)$ , we propose the test statistic analogous to (4) and (9) as

$$\hat{T}_{\text{GCT}} := \frac{\sqrt{p}(\hat{T}_n - 1)}{\hat{\zeta}_n} \quad \text{and} \quad \hat{T}_{\text{MPT}} := \frac{\sqrt{p}(\hat{T}_n - 1)}{\hat{\tau}_n}.$$

In the large- $p$  setting, the center (5) computed in terms of  $\hat{Y}_{ijk}$  may be used. However, as stated in Zhang and Wang [20], using the large- $p$  center estimator tends to give liberal type I error rates.

### 3. Main Results

For a sequence of random variables  $\{Z_k : k = 1, 2, \dots\}$ , define the mixing coefficient  $\alpha(r)$  as

$$\alpha(r) = \sup_{k \geq 1} \left\{ \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| : A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+r}^\infty \right\},$$

for  $r = 1, 2, \dots$ , where  $\mathcal{F}_a^b = \sigma\{Z_k : a \leq k \leq b\}$  denotes the  $\sigma$ -field generated by the random variables  $Z_a, \dots, Z_b$ . Here,  $\alpha(r)$  measures the dependency among the components that are at least  $r$  indices apart. A sequence of random variables is  $\alpha$ -mixing (strong-mixing), if the mixing coefficient  $\alpha(r)$  goes to 0 as  $r$  goes to infinity. The  $\alpha$ -mixing condition assumed on the sequence  $\{Z_k : k = 1, \dots, p\}$ , basically, requires the dependence between two observations to decay, as the separation between the observations ( $r$ ) increases. It prescribes weak dependence among the components of the random vector, which is commonly assumed in time-series analysis, repeated measures, or some other types of data. Two-sample tests on  $\alpha$ -mixing sequence random variables were considered by many authors ([14,15,19,20,25–27], etc.).

To derive the asymptotic distribution of  $\widehat{T}_n$  and construct a test statistic, we require the following assumptions that are analogous to Zhang and Wang [20].

- C1: For any  $i \in \{1, 2\}$  and  $j \in \{1, \dots, n_i\}$ , the sequence  $\{X_{ijk} : k = 1, \dots, p\}$  has a mixing coefficient  $\alpha(r)$ , which satisfies  $\sum_{r=1}^{\infty} r[\alpha(r)]^{\nu/(2+\nu)} < \infty$ , for some  $\nu > 0$ .
- C2: For any  $i \in \{1, 2\}$ ,  $j \in \{1, \dots, n_i\}$  and  $k \in \{1, \dots, p\}$ ,  $X_{ijk} \sim F_{ik}$  is non-degenerate.
- C3:  $\lim_{n \rightarrow \infty} n_i/n \rightarrow \lambda_i \in (0, 1)$  for  $i = 1, 2$  where  $n = n_1 + n_2$ .

**Remark 1.** Similar to Zhang and Wang [20], a sufficient condition for  $\sum_{r=1}^{\infty} r[\alpha(r)]^{\nu/(2+\nu)} < \infty$  in C1 is  $\alpha(r) = O(r^{-c})$ , for some  $c > 2 + 4/\nu$ . It provides a condition on the mixing coefficient  $\alpha(r)$  on  $r$ , for a given  $\nu > 0$ ; if satisfied, the window width  $L$  is not required in the proof of the consistency of the estimator for the scaling parameter, see Theorem 2.

Applying the Theorem 5.2 of Bradley [28] to the marginal distribution function  $F_{ik}$ , which is a Borel function with a single argument, the  $\alpha$ -mixing assumption on the original sequence  $\{X_{ijk} : k = 1, \dots, p\}$  will be inherited by the transformed sequence  $\{Y_{ijk}, k = 1, \dots, p\}$ , with the same mixing coefficient. Under condition C1, the sequence of squared  $t$ -statistics derived from  $\{X_{ijk}\}$  is an  $\alpha$ -mixing process, with the same mixing coefficient  $\alpha(r)$ , by Lemma III.3 in the Supplementary Material of Zhang and Wang [20], so the sequence derived from  $\{Y_{ijk}\}$ . Strictly speaking, the mixing coefficients of the two samples are not required to be equal. Condition C1 can be stated in terms of the maximum of the two coefficients from the two samples. Note that squared Wilcoxon–Mann–Whitney-type statistic  $\widehat{t}_k^2$ , derived from  $\{\widehat{Y}_{ijk}\}$ , is actually defined through basic arithmetic operations on the Borel functions  $c(X_{2jk} - X_{1\ell k})$  for  $\ell = 1, \dots, n_1$ , and  $j = 1, \dots, n_2$ . Thus,  $\widehat{t}_k^2$  can be viewed as a Borel function  $h_k$  of the  $n$  observations on the  $k$ -th variable, i.e.,

$$\widehat{t}_k^2 = h_k(X_{11k}, \dots, X_{1n_1k}, X_{21k}, \dots, X_{2n_2k}).$$

Therefore, along the same arguments as in Zhang and Wang [20], the sequence  $\{\widehat{t}_k^2 : k = 1, \dots, p\}$  is  $\alpha$ -mixing with the same coefficient  $\alpha(r)$ , under condition C1. The asymptotic distribution of  $\widehat{T}_n$  is, then, established in Theorem 1.

**Theorem 1.** Suppose  $\{X_{ijk} : k = 1, \dots, p\}$  for any  $i \in \{1, 2\}$  and  $j \in \{1, \dots, n_i\}$  are sequences of random variables satisfying conditions C1, C2, and C3. Then, under  $H_0$ ,

$$\frac{\sqrt{p}(\widehat{T}_n - E[\widehat{T}_n])}{\sqrt{\text{Var}(\sqrt{p}\widehat{T}_n)}} \xrightarrow{d} \mathcal{N}(0, 1), \text{ as } n, p \rightarrow \infty.$$

**Proof.** Under conditions C1, C2, and C3, the sequence  $\{\widehat{t}_{ik}^2, k = 1, \dots, p\}$  has the same  $\alpha$ -mixing property as  $\{t_{ik}^2, k = 1, \dots, p\}$ . Moment conditions that are analogous to those in Theorem 2.2 of Zhang and Wang [20] are satisfied by Lemmas 1 and 2. Therefore, the remainder of the proof follows, along the same lines.  $\square$

**Remark 2.** As pointed out earlier, asymptotic normality of  $T_n$  was argued in Gregory et al. [19], without proof under different conditions. Similar arguments could be made here, to have asymptotic normality of  $\widehat{T}_n$ . Besides the  $\alpha$ -mixing condition on the  $\{\widehat{t}_k^2 : k = 1, \dots, p\}$  and conditions C2 and C3, it is, additionally, required that the autocovariance function of the sequence  $\{\widehat{t}_k^2 : k = 1, \dots, p\}$  satisfies  $\lim_{n \rightarrow \infty} \frac{1}{p-r} \sum_{k=1}^{p-r} \text{Cov}(\widehat{t}_k^2, \widehat{t}_{k+r}^2) = \gamma(r)$ , for each lag  $r > 0$ . Other conditions assumed in Gregory et al. [19] are, automatically, satisfied in our setup.

The central limit theorem stated above involves centering  $E(\widehat{T}_n)$  and scaling  $\sqrt{\text{Var}(\sqrt{p}\widehat{T}_n)}$  parameters. In both statistics  $\widehat{T}_{GCT}$  and  $\widehat{T}_{MPT}$ , the centering parameter  $\tilde{\zeta}_n$  is set to 1, whereas the scaling parameter  $\sqrt{\text{Var}(\sqrt{p}\widehat{T}_n)}$  is estimated by  $\widehat{\zeta}_n$  and  $\widehat{\tau}_n$ , respectively. If the estimators for

the scaling parameter are consistent, then the asymptotic normality of  $\hat{T}_{GCT}$  and  $\hat{T}_{MPT}$  follow from Slutsky's theorem. In Lemma 3, the consistency of  $\hat{\gamma}_{kk'}$  is proven and, subsequently, in Theorem 2, the consistency of  $\hat{\tau}_n$  is established.

**Lemma 3.** Suppose  $\{X_{ijk} : k = 1, \dots, p\}$  for any  $i \in \{1, 2\}$  and  $j \in \{1, \dots, n_i\}$  are sequences of random variables satisfying conditions C1, C2, and C3. Then, under  $H_0$ ,  $\hat{\gamma}_{kk'} = \gamma_{kk'} + O_p(n^{-1/2})$ , for any  $k, k' \in \{1, \dots, p\}$ , where  $\gamma_{kk'}$  and  $\hat{\gamma}_{kk'}$  are as defined in (7) and (13), respectively.

**Theorem 2.** Suppose  $\{X_{ijk} : k = 1, \dots, p\}$  for any  $i \in \{1, 2\}$  and  $j \in \{1, \dots, n_i\}$  are sequences of random variables satisfying conditions C1, C2, and C3. Assume  $L = \lceil p^\epsilon \rceil$ , for some  $0 < \epsilon < 1$ . Then, under  $H_0$  and for  $\hat{\tau}_n$ , defined in (13),

$$\hat{\tau}_n^2 - \text{Var}(\sqrt{p}\hat{T}_n) = O_p(n^{-1/2}) + O(L^{-1}), \text{ as } n, p \rightarrow \infty.$$

Further, if  $\alpha(r) = O(r^{-c})$ , for some  $c > 1$ , then

$$\hat{\tau}_n^2 - \text{Var}(\sqrt{p}\hat{T}_n) = O_p(n^{-1/2}) + O(p^{1-c\nu/(2+\nu)}),$$

for some  $\nu > 2/(c-1)$  as  $n, p \rightarrow \infty$ .

**Proof.** Similar to the proof of Theorem 2.5 of Zhang and Wang [20],  $\square$

The asymptotic distribution of the test statistic  $\hat{T}_{MPT}$  is stated in Corollary 1.

**Corollary 1.** Suppose  $\{X_{ijk} : k = 1, \dots, p\}$  for any  $i \in \{1, 2\}$  and  $j \in \{1, \dots, n_i\}$  are sequences of random variables satisfying conditions C1, C2, and C3. Then, under  $H_0$  and  $p = o(n^2)$ ,

$$\hat{T}_{MPT} = \frac{\sqrt{p}(\hat{T}_n - 1)}{\hat{\tau}_n} \xrightarrow{d} \mathcal{N}(0, 1), \text{ as } n, p \rightarrow \infty,$$

where  $\hat{\tau}_n$  is as defined in (13) and  $L = \lceil p^\epsilon \rceil$  for some  $0 < \epsilon < 1$ .

**Proof.** The consistency of  $\hat{\tau}_n$  is stated in Theorem 2. By Lemma 1,  $E(\hat{t}_k^2) = 1 + O(n^{-1})$ . Therefore,

$$\sqrt{p}(E[\hat{T}_n] - 1) = \frac{\sqrt{p}}{p} \sum_{i=1}^p (E[\hat{t}_k^2] - 1) = o(1), \text{ when } p = o(n^2).$$

The desired result follows from Theorem 1, by applying Slutsky's theorem.  $\square$

Under the assumptions discussed in Remark 2, the sample autocovariance  $\hat{\gamma}(r)$  is consistent for  $\gamma(r)$ . Therefore, the consistency of  $\hat{\zeta}_n$  follows, if a suitable window width  $L$  is chosen. Consequently, one can establish the asymptotic distribution of the statistic  $\hat{T}_{GCT}$  as

$$\hat{T}_{GCT} = \frac{\sqrt{p}(\hat{T}_n - 1)}{\hat{\zeta}_n} \xrightarrow{d} \mathcal{N}(0, 1), \text{ as } n, p \rightarrow \infty,$$

where  $\hat{\zeta}_n$  is defined in (12) for either of the Parzen or trapezoid-window-weight function  $w(\cdot)$ . To be consistent, one may choose the same window width  $L$  as in Corollary 1. It was noted in Zhang and Wang [20] that GCT's estimate of scaling parameter does not, effectively, use the sample replications, as they rely on a single term  $(\hat{t}_k^2 - \hat{T}_n)(\hat{t}_{k'}^2 - \hat{T}_n)$ . This might be a negative number, while  $\text{Cov}(\hat{t}_k^2, \hat{t}_{k'}^2)$  has been proven to be positive. Instead, MPT's estimate of the scaling parameter benefits from both increasing the number of replications and the dimension. Thus, MPT's estimate is likely to produce more stable and consistent results over different samples, while the GCT's estimate could vary drastically. From the

simulation study in Section 4, we can see features from the two proposed ranked-version tests, referred to as RGCT and RMPT, respectively.

Both Gregory et al. [19] and Zhang and Wang [20] proposed to use a two-tailed rejection region. However, we believe that the hypothesis  $H_0$  should be rejected, when  $T_n$  or  $\hat{T}_n$  is large. That means a one-tailed rejection region should be used. The numerical results are shown in Figures 1–5.

## 4. Simulation

### 4.1. Simulation Design

The aim of this section is to compare the proposed rank-based tests RGCT and RMPT (for testing the hypothesis (2)), with their parametric versions GCT and MPT (for testing the hypothesis of equality of two mean vectors), through Monte Carlo simulation. As we explained before, we use the moderate- $p$  version statistics for all four tests. We employ one-tailed rejection regions and Parzen window weight for GCT and RGCT, and two-tailed rejection regions for MPT and RMPT. In order to allow correct interpretation of the powers vis-à-vis the achieved sizes, the empirical type I error rates as well as powers of these tests are presented in Figures 2–4. Throughout, the sizes and powers are calculated through 2000 replications, and the actual level of significance is set at  $\alpha = 0.05$ . The proportion of rejections, out of the 2000 runs, is recorded. We will investigate the effects of window width and the influence of dimensionality under various dependence structures and innovation distributions, while some of them may violate the model assumption C1 or moment assumptions stated in Gregory et al. [19] or Zhang and Wang [20].

The two groups are always independent, and the dependency only exists within the same subject. More precisely, for any  $i \in \{1, 2\}$  and  $j \in \{1, \dots, n_i\}$ , we generate data as

$$X_{ijk} = \mu_{ik} + \varepsilon_{ijk},$$

for  $k \in \{1, \dots, p\}$ , where  $\mu_{ik}$  is a constant and  $\{\varepsilon_{ijk} : k = 1, \dots, p\}$  is the error process. Let  $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ip})^\top$ . Under the null hypothesis, we set  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}_p$ . Under the alternative hypothesis, let  $\boldsymbol{\mu}_1 = \mathbf{0}_p$  and  $0 < \beta \leq 1$  be the proportion of nonzero elements in  $\boldsymbol{\mu}_2$ . We set the  $p\beta/2$  elements of  $\boldsymbol{\mu}_2$  to  $\delta$ , the next  $p\beta/2$  elements to  $-\delta$ , and the remaining  $p - p\beta$  elements to 0, i.e.,

$$\boldsymbol{\mu}_2 = (\underbrace{\delta, \dots, \delta}_{\beta p/2}, \underbrace{-\delta, \dots, -\delta}_{\beta p/2}, \underbrace{0, \dots, 0}_{1-\beta p}).$$

Four distributions for the innovation  $\varepsilon_{ijk}$  are considered,

Normal: standard normal distribution Normal (0, 1);

T:  $t$  distribution with degrees of freedom 3;

Gamma: centered Gamma distribution with shape parameter 4 and scale parameter 2;

Cauchy: Cauchy (0, 0.1) distribution.

Here, standard normal distribution serves as a benchmark; the  $t$  distribution is used to assess how the tests perform, when the higher-order moment conditions are violated of GCT and MPT; and the gamma and Cauchy distributions allow to evaluate robustness to skewed and heavy-tailed data, respectively. We will consider independent and three dependent error processes, but, otherwise, the errors in the two groups are generated by the same process.

IND (independent):  $\varepsilon_{ijk}$  is independently drawn from the innovation distribution for  $k = 1, \dots, p$ .

WD (weakly dependent):  $\varepsilon_{ijk}$  is generated, according to ARMA(2, 2), with autoregressive parameter  $\boldsymbol{\phi} = (0.4, -0.1)$  and moving-average parameters  $\boldsymbol{\theta} = (0.2, 0.3)$ .

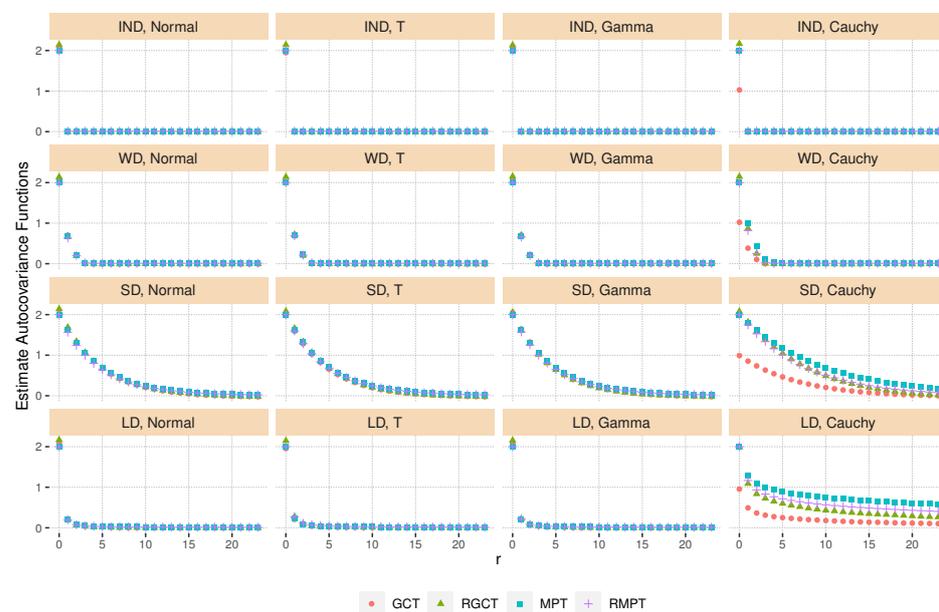
SD (strongly dependent):  $\varepsilon_{ijk}$  is generated, according to AR(1), with autoregressive parameters  $\phi = 0.9$ .

LD (long-range dependent): Let  $A = (a_{st})_{s,t=1}^p$ , where  $a_{st} = 0.5[(r + 1)^{2H} + (r - 1)^{2H} - 2r^{2H}]$ , for constant self-similarity parameters  $H = 0.7$  and  $r = |s - t|$ . Decompose  $A$  by Cholesky factorization, to get the matrix  $U$ , such that  $A = U^T U$ . Independently, draw  $\eta_{ijk}$  from the innovation distribution for  $k = 1, \dots, p$ . Let  $\boldsymbol{\eta}_{ij} = (\eta_{ij1}, \dots, \eta_{ijp})^T$  and set  $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \dots, \varepsilon_{ijp})^T = U^T \boldsymbol{\eta}_{ij}$ .

The long-range dependent process is generated, following the approach proposed by Hall et al. [29]. It was mentioned in Zhang and Wang [20] that the correlation for each of the four setups tends to 0, when the lag separation between the two components goes to infinity. However, the decaying speed of LD is that the correlation converges to 0 at the rate  $O(r^{-2(1-H)})$  (see [30]), which is much slower than the ARMA model. The condition for mixing coefficients C1 holds for the IND, WD, and SD models but not for the LD model. Thus, including the LD error process in the simulation helps us to evaluate the robustness of the tests, against the assumed dependence regularity.

#### 4.2. Adaptive Selection of Window Width $L$

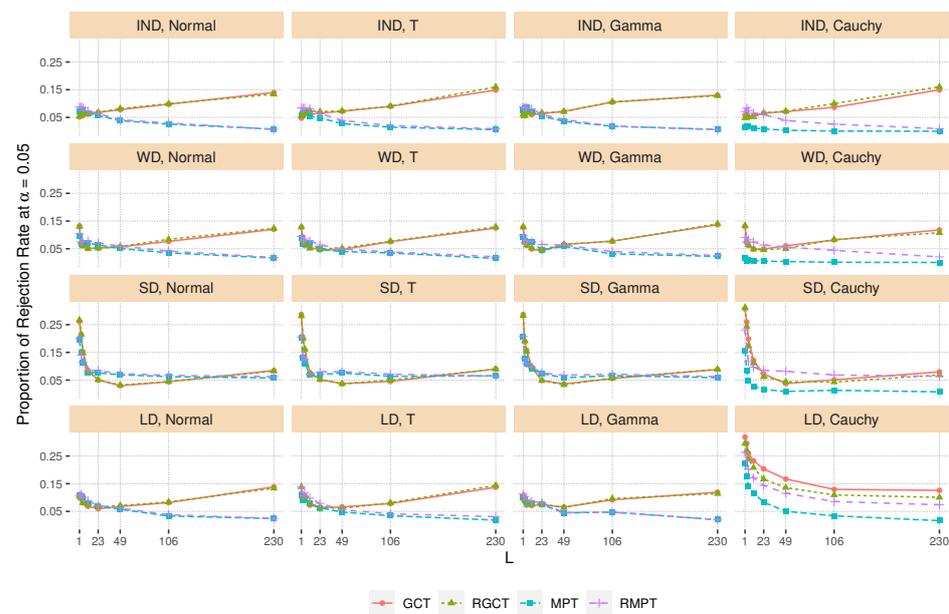
To compare the two estimators of the scaling parameter, estimated autocovariance of the sequence  $\{\hat{t}_k^2 : k = 1, \dots, p\}$  at lags  $r = 0, 1, \dots, L$ , i.e.,  $\hat{\gamma}(r)$  for RGCT and  $\frac{1}{p} \sum_{k-k'=r} \hat{\gamma}_{kk'}$  for RMPT are displayed in Figure 1, together with the parametric versions for GCT and MPT, on the original data. Estimates are averages over 2000 runs at each lag  $r$ , with dimension  $p = 500$ , sample size  $n_1 = 80, n_2 = 100$ , and for  $L = 500^{1/2} \approx 23$ . The four estimates are very close, except for LD and SD Cauchy. The similarity of the estimates verifies that the mixing structures of the sequence of squared  $t$ -statistics, derived from the original data and the rank-transformed data, are about the same. It, also, shows that the decaying speeds are not only related to the error processes but also to the innovation distributions. Thus, there does not appear to exist a fixed number  $\epsilon$ , such that the window-width choice  $L = \lceil p^\epsilon \rceil$  works, reasonably, for all data.



**Figure 1.** The estimates of the autocovariance function at varying lags  $r = 0, 1, \dots, L$ , from 2000 simulation runs. We set  $n_1 = 80, n_2 = 100$ , and  $p = 500$ .

To illustrate the sensitivity to the choice of the window width, Figure 2 reports a type I error rate, by setting the dimension  $p = 500$  and sample sizes  $n_1 = 80$  and  $n_2 = 100$ . A sequence of values,  $L = \lceil p^\epsilon \rceil$ , for  $\epsilon = i/8$ , and  $i = 0, 1, \dots, 7$ , i.e.,  $L = 1, 3, 5, 11, 23, 49, 106, 230$ , are experimented. It can be, clearly, seen from the plots that all four tests are sensitive to the window width. MPT and RMPT exhibit a slightly decreasing trend of the type I error rate,

while the trends of GCT and RGCT are not monotonic if the data are dependent. When the window gets wider, the type I errors of MPT and RMPT tend to be conservative, while GCT and RGCT have more rejections. Especially, if the data are strongly or long-range dependent, a wider window width is needed, to capture more correlations, but if the data are less dependent, a wider window width may include more noise, such that the scaling parameters of GCT and RGCT will be underestimated, while those for MPT and RMPT are overestimated.



**Figure 2.** The empirical type I errors at level  $\alpha = 0.05$  from 2000 simulation runs, for varying window widths  $L = (1, 3, 5, 11, 23, 49, 106, 230)$ . We set  $n_1 = 80, n_2 = 100$ , and  $p = 500$ .

Thus, instead of using a window width  $L = \lceil p^\epsilon \rceil$ , for a fixed  $\epsilon$ , we provide the following algorithm to select  $L$  for RMPT. According to Lemma 2,  $\text{Cov}(\hat{t}_k^2, \hat{t}_{k'}^2) = \gamma_{kk'} + O(n^{-1/2})$ . Since it was assumed that  $p = o(n^2)$  and  $p > n$ , usually, holds in real data application, we choose the window width  $L$  to be the smallest  $r$ , such that  $p^{-1} \sum_{|k-k'|=r+1} \hat{\gamma}_{kk'}$  is less than  $p^{-1/2}$ . In the case of  $n < p$ , the bound can be replaced with  $n^{-1/2}$ . From Figure 1, the decaying speed of RGCT and RMPT are about the same. So, we may choose  $L + 1$  to be the window width for RGCT. The extra one is, actually, not used, since the weight  $w_p(L/L) = 0$ . Note that we do not use  $\hat{\gamma}(r)$  as the selection criterion, since it relies only on a single unstable term and does not take advantage of the sample replications. Applying this algorithm to the original data (removing step 1 in Algorithm 1), we can obtain a window width  $L$  for MPT and, then,  $L + 1$  could be used as a window width for GCT. For all the remaining simulations and applications, we will use this algorithm to determine the window width.

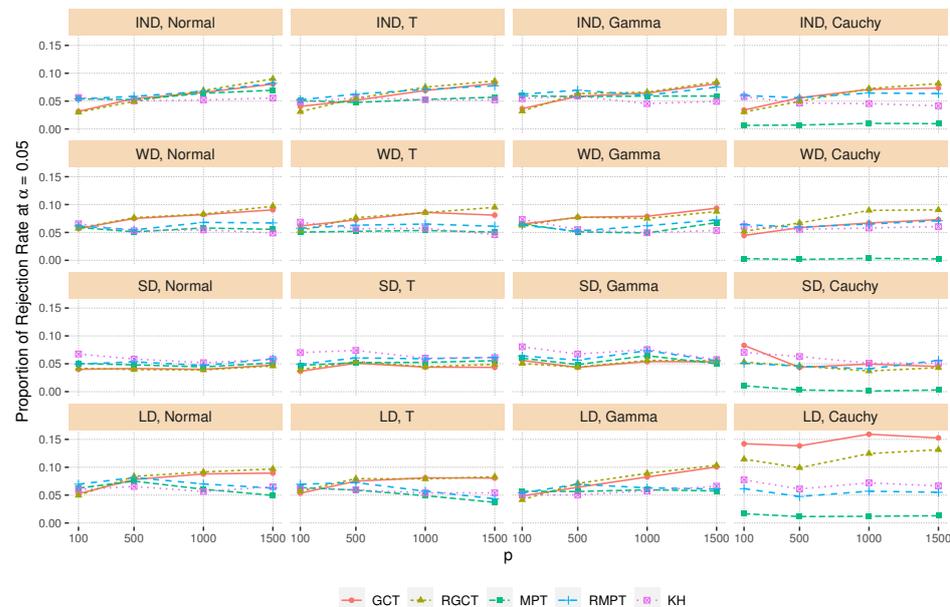
**Algorithm 1:** Data adaptive window width selection

**Data:** Two samples of data  $X_{ij}$  with  $p$ -dimension, for  $i \in \{1, 2\}$  and  $j \in \{1, \dots, n_i\}$ .  
**Result:** Return window width  $L$  and scaling estimate  $\hat{\tau}_n$ .

- 1 Applying rank transformation to the original data, to get  $\hat{Y}_{ij}$ ;
- 2 Set  $r = 0$  and  $\hat{\tau}_n = p^{-1} \sum_{|k-k'|=0} \hat{\gamma}_{kk'}$ ;
- 3 **while** ( $r < p$ ) **do**
- 4     Set  $\hat{\tau}_{nr} = p^{-1} \sum_{|k-k'|=r+1} \hat{\gamma}_{kk'}$ ;
- 5     **if**  $\{\hat{\tau}_{nr} > \min(p^{-1/2}, n^{-1/2})\}$  **then**
- 6          $\hat{\tau}_n = \hat{\tau}_n + \hat{\tau}_{nr}$ ;
- 7          $r = r + 1$ ;
- 8     **else**
- 9         **break**;
- 10    **end**
- 11 **end**
- 12 Output  $L = r$  and  $\hat{\tau}_n$ .

4.3. Type I Error Rate

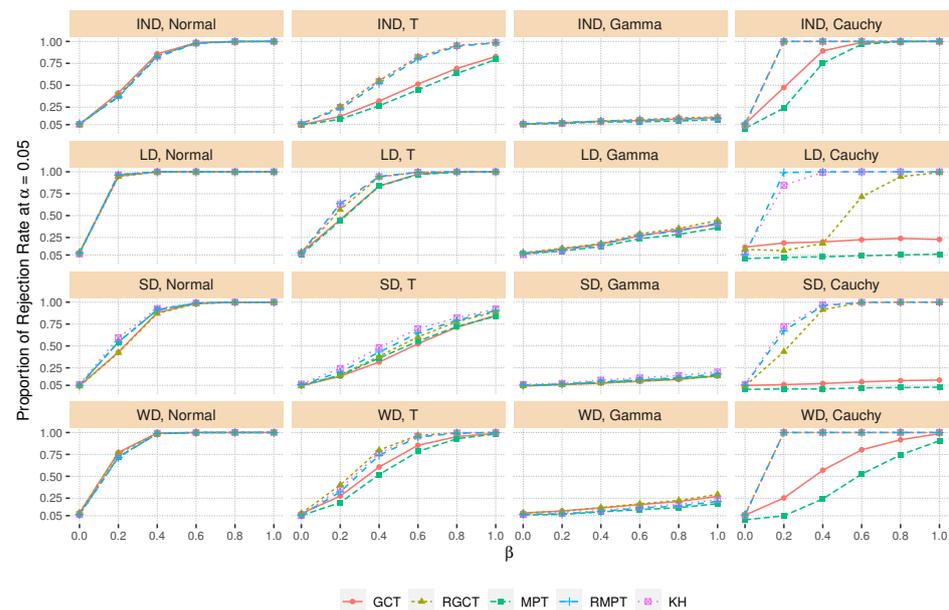
The effect of dimensionality is investigated in Figure 3, by considering  $p = 100, 500, 1000,$  and  $1500$ , for fixed sample sizes  $n_1 = 80$  and  $n_2 = 100$ . In addition, a two-sample rank-based test, given in Kong and Harrar [15] (referred to as KH), is included, always with the other four tests, although it tests the overall relative effects. From Figure 3, MPT tends to be conservative for Cauchy innovations. That is reasonable, since the moments of Cauchy distribution do not exist. When the dimension gets larger, some inflated type I errors for GCT and RGCT are observed, especially for LD dependence that violates condition C1. Test KH has a few inflated type I errors, under SD dependence. Speaking overall, RMPT always has a type I error rate close to the nominal level, which also confirms that the window width was selected well.



**Figure 3.** The empirical type I errors at level  $\alpha = 0.05$  from 2000 simulation runs, for varying dimensions  $p = 100, 500, 1000,$  and  $1500$ , with  $n_1 = 80$  and  $n_2 = 100$ . Window widths are calculated from Algorithm 1.

#### 4.4. Power Comparison

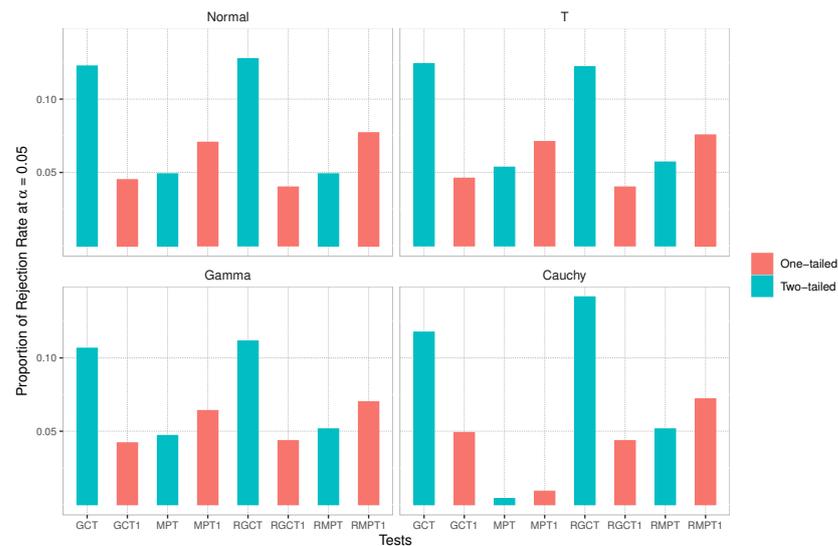
For power comparison, we fix dimension  $p = 500$  and sample sizes  $n_1 = 80$  and  $n_2 = 100$ . We plot the powers of the tests against  $\beta \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ . The signal magnitude  $\delta$  is chosen as follows:  $\delta = 0.1$  for IND, 0.2 for LD and WD, and 0.5 for SD. As shown in Figure 4, there is a clear advantage of the rank tests RGCT and RMPT, for heavily tailed distributions, such as Cauchy, while GCT and MPT tend to be conservative, under correlated error processes. This is expected, because the model violates most of the moment assumptions needed for the theoretical results of GCT and MPT. All five tests are comparable in terms of power, for the other innovation distributions. Notably, they all have low power under centered Gamma innovation, increasing signal magnitude  $\delta$ , which is needed for obtaining higher power.



**Figure 4.** The empirical rejection frequencies at nominal level  $\alpha = 0.05$  from 2000 simulation runs for varying  $\beta = (0, 0.2, 0.4, 0.6, 0.8, 1)$ . Fix  $p = 500$ ,  $n_1 = 80$ , and  $n_2 = 100$ , and window widths are calculated from Algorithm 1. The signal magnitude  $\delta = 0.1$  for IND, 0.2 for LD and WD, and 0.5 for SD.

#### 4.5. One-Tailed versus Two-Tailed Tests

We close the simulation study, by investigating type I error rates for one-tailed and two-tailed tests. Figure 5 provides achieved type I error rates, in the case of the SD-dependence model. For convenience, two-tailed tests are labeled as GCT, MPT, RGCT, and RMPT, while the corresponding one-tailed tests are labeled as GCT1, MPT1, RGCT1, and RMPT1. The plot shows that two-tailed GCT and RGCT tend to have liberal type I error rates. Although there is not much difference, the type I errors of one-tailed MPT and RMPT are closer to the nominal level ( $\alpha = 0.05$ ). Thus, we recommend to use the one-tailed rejection region for GCT and RGCT and keep using the two-tailed rejection region for MPT and RMPT for comparison. It can, also, be seen from the power plot in Figure 4, that the one-tailed GCT and RGCT are comparable to the two-tailed MPT and RMPT.



**Figure 5.** The empirical type I errors at level  $\alpha = 0.05$  from 2000 simulation runs for one-tailed and two-tailed tests, under SD dependence. Set  $p = 500$ ,  $n_1 = 80$ , and  $n_2 = 100$ , and window widths are calculated using Algorithm 1.

## 5. Conclusions

In this work, we developed nonparametric tests for high-dimensional data, in two samples. The hypothesis is formulated in terms of the nonparametric marginal relative effects. These effects are meaningful and well defined, even for data measured on an ordinal scale or for heavily tailed data. The only stipulation is that the marginal distributions of the data are non-degenerate.

Related research by Gregory et al. [19] or, more recently, Zhang and Wang [20] developed a test for equality of mean vectors, where a composite of the squares of marginal  $t$  statistics are employed. Our test statistics are in the same vein, but we use the nonparametric Wilcoxon–Mann–Whitney-type statistic of Brunner and Munzel [4], instead of the  $t$  statistic. In many high-dimensional applications, such as transcriptomics, the effect on any individual variable (gene expression) may be small, and composite statistics will have high power in detecting the cumulative effect.

We proposed two tests that differ in the way the asymptotic variance of the composite statistic is estimated. In both cases, the test statistics are shown to, asymptotically, follow a standard normal distribution, under  $\alpha$ -mixing (strong-mixing) dependence. The estimation of the asymptotic variances involves banding the covariance matrix of the marginal squared Wilcoxon–Mann–Whitney-type statistics, to guard against overestimation or underestimation. We demonstrated that the length of the banding window is related to the distribution of data as well as the strength of dependence, among the variables. We provided an algorithm for, adaptively, selecting the window width from the data. The algorithm was shown to improve the performance of the asymptotic variance estimator and, also, plays a crucial role in controlling the type I error rate.

The finite sample performance of the proposed tests was studied via simulation and compared with their parametric counterparts as well as another rank-based test. Generally, the parametric tests were found to be liberal in type I error control, when the dimension is large or for heavily tailed distributions, whereas the proposed tests have satisfactory performance overall.

Nonparametric tests, for relative effects in high dimensions, are not well studied [14,15]. For available methods to be applicable, the variables must be commensurate, so that the relative effects, which are measured to the average of all marginal distributions, would be appropriate. The present manuscript overcomes this challenge, by defining the relative effects, marginally. Comparing our proposed tests, overall RGCT is more reliable, when sample sizes are small. However, efficiency may be gained, by using the replications in

estimating the asymptotic variance, as in RMPT, if the sample size per group is relatively large. All these nonparametric methods are, currently, available only for two groups and, by extension, to multiple treatment groups, which is an important problem for future research.

**Author Contributions:** Conceptualization, S.W.H. and X.K.; methodology, S.W.H., X.K. and A.V.-T.; validation, S.W.H. and X.K.; formal analysis, S.W.H. and X.K.; investigation, S.W.H. and X.K.; writing—original draft preparation, S.W.H., X.K. and A.V.-T., writing—review and editing, S.W.H., X.K. and A.V.-T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors are grateful to the four anonymous referees, for critically reading the original version of the manuscript and making valuable suggestions that led to great improvements. The authors are, also, thankful to the editor for the orderly handling of the manuscript.

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

## Appendix A

We present some key Lemmas, needed for the proof of the main results, below. The proofs of Lemmas A1, A2, A5, and A7 can be found, for example, in Brunner and Munzel [4]. For the sake of completeness, Lemmas A3, A4, and A6 are stated for proving Lemmas A8, 1, and 2, which are somewhat new.

**Lemma A1.** For any  $1 \leq k \leq p$  and  $1 \leq j \leq n_2$ ,  $E[\widehat{F}_{1k}(X_{2jk})] = E[F_{1k}(X_{2jk})] = \omega_k$ .

**Lemma A2.** For any  $1 \leq k \leq p$  and  $1 \leq j \leq n_2$ ,  $E\left[\left\{\widehat{F}_{1k}(X_{2jk}) - F_{1k}(X_{2jk})\right\}^2\right] \leq 1/n_1$ .

**Lemma A3.** For any  $1 \leq k, k' \leq p$ ,  $1 \leq \ell_1 \neq \ell_2 \leq n_1$  and  $1 \leq j \leq n_2$ ,

$$E\left[\left\{c(X_{2jk} - X_{1\ell_1 k}) - \omega_k\right\}\left\{c(X_{2jk'} - X_{1\ell_2 k'}) - \omega_{k'}\right\}\right] = \sigma_{2kk'}.$$

**Lemma A4.** For any  $1 \leq k, k' \leq p$ ,  $1 \leq \ell \leq n_1$  and  $1 < j_1 \neq j_2 \leq n_2$ ,

$$E\left[\left\{c(X_{2j_1 k} - X_{1\ell k}) - \omega_k\right\}\left\{c(X_{2j_2 k'} - X_{1\ell k'}) - \omega_{k'}\right\}\right] = \sigma_{1kk'}.$$

**Lemma A5.** As  $\min\{n_1, n_2\} \rightarrow \infty$  and  $n_i/n \rightarrow \lambda_i \in (0, 1)$ , two statistics,  $\sqrt{n}(\widehat{\omega}_k - \omega_k)$  and  $\sqrt{n}(\bar{Y}_{2k} - \bar{Y}_{1k} + 1 - 2\omega_k)$ , asymptotically, have the same distribution.

**Lemma A6.** As  $\min\{n_1, n_2\} \rightarrow \infty$  and  $n_i/n \rightarrow \lambda_i \in (0, 1)$ ,  $\text{Var}(\widehat{\omega}_k) = \sigma_{1k}^2/n_1 + \sigma_{2k}^2/n_2 + O(n^{-2})$ .

**Lemma A7.** As  $\min\{n_1, n_2\} \rightarrow \infty$  and  $n_i/n \rightarrow \lambda_i \in (0, 1)$ ,  $\widehat{\sigma}_{ik}^2$  defined in (10) is an  $L_2$  consistency estimator of  $\sigma_{ik}^2$ .

**Lemma A8.** As  $\min\{n_1, n_2\} \rightarrow \infty$  and  $n_i/n \rightarrow \lambda_i \in (0, 1)$ ,  $\widehat{\sigma}_{ikk'}$  defined in (14) is an  $L_2$  consistency estimator of  $\sigma_{ikk'}$  for any  $k \neq k'$ .

**Proof.** The proof of the two  $i$ 's are similar, so, without loss of generality, we only prove it for  $i = 2$ . First,

$$\sigma_{2kk'} = \text{Cov}(Y_{21k}, Y_{21k'}) = E[Y_{21k}Y_{21k'}] - \omega_k\omega_{k'},$$

and

$$\begin{aligned} \hat{\sigma}_{2kk'} &= \frac{1}{n_1^2(n_2 - 1)} \sum_{j=1}^{n_2} \left( R_{2jk} - R_{jk}^{(2)} - \bar{R}_{2\cdot k} + \frac{n_2 + 1}{2} \right) \left( R_{2jk'} - R_{jk'}^{(2)} - \bar{R}_{2\cdot k'} + \frac{n_2 + 1}{2} \right) \\ &= \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} \left\{ \hat{F}_{1k}(X_{2jk}) - \hat{\omega}_k \right\} \left\{ \hat{F}_{1k'}(X_{2jk'}) - \hat{\omega}_{k'} \right\} \\ &= \frac{n_2}{n_2 - 1} \left( \frac{1}{n_2} \sum_{j=1}^{n_2} \hat{F}_{1k}(X_{2jk}) \hat{F}_{1k'}(X_{2jk'}) - \hat{\omega}_k \hat{\omega}_{k'} \right). \end{aligned}$$

Thus, it suffices to show that, as  $n \rightarrow \infty$ ,

$$\frac{1}{n_2} \sum_{j=1}^{n_2} \hat{F}_{1k}(X_{2jk}) \hat{F}_{1k'}(X_{2jk'}) \xrightarrow{L_2} E[Y_{21k} Y_{21k'}], \text{ and } \hat{\omega}_k \hat{\omega}_{k'} \xrightarrow{L_2} \omega_k \omega_{k'}. \tag{A1}$$

It is known from Lemma A6 that  $\hat{\omega}_k$  and  $\hat{\omega}_{k'}$  are  $L_2$ -consistent to  $\omega_k$  and  $\omega_{k'}$ , respectively. Thus, applying  $c_r$ -inequality for  $r = 2$ , it follows that

$$\begin{aligned} E[\{\hat{\omega}_k \hat{\omega}_{k'} - \omega_k \omega_{k'}\}^2] &= E[\{\hat{\omega}_k(\hat{\omega}_{k'} - \omega_{k'}) + (\hat{\omega}_k - \omega_k)\omega_{k'}\}^2] \\ &\leq 2E[\hat{\omega}_k^2(\hat{\omega}_{k'} - \omega_{k'})^2] + 2E[(\hat{\omega}_k - \omega_k)^2\omega_{k'}^2] \leq 2E[(\hat{\omega}_{k'} - \omega_{k'})^2] + 2E[(\hat{\omega}_k - \omega_k)^2] \rightarrow 0, \end{aligned}$$

since  $|\hat{\omega}_k| \leq 1$  and  $|\omega_{k'}| \leq 1$ . It proves the second  $L_2$  consistency in (A1).

On the other hand,  $\frac{1}{n_2} \sum_{j=1}^{n_2} F_{1k}(X_{2jk}) F_{1k'}(X_{2jk'}) \xrightarrow{L_2} E[Y_{21k} Y_{21k'}]$ , since

$$E\left[\left\{\frac{1}{n_2} \sum_{j=1}^{n_2} F_{1k}(X_{2jk}) F_{1k'}(X_{2jk'}) - E[Y_{21k} Y_{21k'}]\right\}^2\right] = \frac{1}{n_2^2} \sum_{j=1}^{n_2} E\left[\{Y_{2jk} Y_{2jk'} - E[Y_{2jk} Y_{2jk'}]\}^2\right] \leq \frac{1}{n_2} \rightarrow 0,$$

where the last inequality is because  $|Y_{2jk} Y_{2jk'} - E[Y_{2jk} Y_{2jk'}]| \leq 1$ . Thus, to prove the first  $L_2$  consistency in (A1), it suffices to prove

$$E\left[\left\{\frac{1}{n_2} \sum_{j=1}^{n_2} \left(\hat{F}_{1k}(X_{2jk}) \hat{F}_{1k'}(X_{2jk'}) - F_{1k}(X_{2jk}) F_{1k'}(X_{2jk'})\right)\right\}^2\right] \rightarrow 0.$$

Applying  $c_r$ -inequality for  $r = 2$ , it follows that

$$\begin{aligned} &\left\{\sum_{j=1}^{n_2} \left(\hat{F}_{1k}(X_{2jk}) \hat{F}_{1k'}(X_{2jk'}) - F_{1k}(X_{2jk}) F_{1k'}(X_{2jk'})\right)\right\}^2 \\ &= \left\{\sum_{j=1}^{n_2} \hat{F}_{1k}(X_{2jk}) \left(\hat{F}_{1k'}(X_{2jk'}) - F_{1k'}(X_{2jk'})\right) + \left(\hat{F}_{1k}(X_{2jk}) - F_{1k}(X_{2jk})\right) F_{1k'}(X_{2jk'})\right\}^2 \\ &\leq 2\left\{\sum_{j=1}^{n_2} \hat{F}_{1k}(X_{2jk}) \left(\hat{F}_{1k'}(X_{2jk'}) - F_{1k'}(X_{2jk'})\right)\right\}^2 + 2\left\{\sum_{j=1}^{n_2} \left(\hat{F}_{1k}(X_{2jk}) - F_{1k}(X_{2jk})\right) F_{1k'}(X_{2jk'})\right\}^2. \end{aligned}$$

Taking the expectation on the first term, it follows that

$$\begin{aligned} &\frac{2}{n_2^2} E\left[\left\{\sum_{j=1}^{n_2} \hat{F}_{1k}(X_{2jk}) \left(\hat{F}_{1k'}(X_{2jk'}) - F_{1k'}(X_{2jk'})\right)\right\}^2\right] \\ &\leq \frac{2}{n_2^2} \sum_{j=1}^{n_2} E\left[\hat{F}_{1k}^2(X_{2jk}) \left(\hat{F}_{1k'}(X_{2jk'}) - F_{1k'}(X_{2jk'})\right)^2\right] \\ &\quad + \frac{2}{n_2^2} \sum_{j_1 \neq j_2}^{n_2} E\left[\hat{F}_{1k}(X_{2j_1k}) \left(\hat{F}_{1k'}(X_{2j_1k'}) - F_{1k'}(X_{2j_1k'})\right) \hat{F}_{1k}(X_{2j_2k}) \left(\hat{F}_{1k'}(X_{2j_2k'}) - F_{1k'}(X_{2j_2k'})\right)\right]. \end{aligned}$$

Since  $\widehat{F}_{1k}^2(X_{2jk}) \leq 1$ , and by Lemma A2,

$$\begin{aligned} & \frac{2}{n_2^2} \sum_{j=1}^{n_2} \mathbb{E} \left[ \widehat{F}_{1k}^2(X_{2jk}) \left( \widehat{F}_{1k'}(X_{2jk'}) - F_{1k'}(X_{2jk'}) \right)^2 \right] \\ & \leq \frac{2}{n_2^2} \sum_{j=1}^{n_2} \mathbb{E} \left[ \left( \widehat{F}_{1k'}(X_{2jk'}) - F_{1k'}(X_{2jk'}) \right)^2 \right] \leq \frac{2}{n_1 n_2} \rightarrow 0, \end{aligned}$$

and

$$\begin{aligned} & \frac{2}{n_2^2} \sum_{j_1 \neq j_2}^{n_2} \mathbb{E} \left[ \widehat{F}_{1k}(X_{2j_1k}) \left( \widehat{F}_{1k'}(X_{2j_1k'}) - F_{1k'}(X_{2j_1k'}) \right) \widehat{F}_{1k}(X_{2j_2k}) \left( \widehat{F}_{1k'}(X_{2j_2k'}) - F_{1k'}(X_{2j_2k'}) \right) \right] \\ & = \frac{2}{n_1^4 n_2^2} \sum_{j_1 \neq j_2}^{n_2} \sum_{\ell_1, \ell_2, \ell_3, \ell_4=1}^{n_1} \mathbb{E} \left[ c(X_{2j_1k} - X_{1\ell_1k}) \left\{ c(X_{2j_1k'} - X_{1\ell_2k'}) - F_{1k}(X_{2j_1k'}) \right\} \right. \\ & \qquad \qquad \qquad \left. c(X_{2j_2k} - X_{1\ell_3k}) \left\{ c(X_{2j_2k'} - X_{1\ell_4k'}) - F_{1k}(X_{2j_2k'}) \right\} \right] \\ & \leq \frac{2n_2(n_2 - 1)(n_1^3 + 2n_1(n_1 - 1))}{n_1^4 n_2^2} \rightarrow 0, \text{ as } n_1, n_2 \rightarrow \infty. \end{aligned}$$

The last inequality is because the expectations of terms in the summation are zero, if  $\{\ell_1, \ell_2, \ell_3, \ell_4\}$  are all different,  $\ell_2$  is not the same of any  $\{\ell_1, \ell_3, \ell_4\}$ , or  $\ell_4$  is not the same of any  $\{\ell_1, \ell_2, \ell_3\}$  by Lemma A1. We divide it by two cases, when the expectations may not be 0,  $\ell_2 = \ell_4$ , or  $\ell_2 \neq \ell_4$ . If  $\ell_2 = \ell_4$ , the above summation is less than  $n_2(n_2 - 1)n_1^3$ . If  $\ell_2 \neq \ell_4$ , then it should have  $\{\ell_1, \ell_3\} = \{\ell_2, \ell_4\}$ , and the above summation is less than  $2n_2(n_2 - 1)n_1(n_1 - 1)$ .

Similarly, the second term,

$$\begin{aligned} & \frac{2}{n_2^2} \mathbb{E} \left[ \left\{ \sum_{j=1}^{n_2} \left( \widehat{F}_{1k}(X_{2jk}) - F_{1k}(X_{2jk}) \right) F_{1k'}(X_{2jk'}) \right\}^2 \right] \\ & = \frac{2}{n_2^2} \sum_{j=1}^{n_2} \mathbb{E} \left[ \left( \widehat{F}_{1k}(X_{2jk}) - F_{1k}(X_{2jk}) \right)^2 F_{1k'}^2(X_{2jk'}) \right] \\ & \quad + \frac{2}{n_2^2} \sum_{j_1 \neq j_2}^{n_2} \mathbb{E} \left[ \left( \widehat{F}_{1k}(X_{2j_1k}) - F_{1k}(X_{2j_1k}) \right) F_{1k'}(X_{2j_1k'}) \left( \widehat{F}_{1k}(X_{2j_2k}) - F_{1k}(X_{2j_2k}) \right) F_{1k'}(X_{2j_2k'}) \right]. \end{aligned}$$

Since  $F_{1k'}^2(X_{2jk'}) \leq 1$ , and by Lemma A2,

$$\frac{2}{n_2^2} \sum_{j=1}^{n_2} \mathbb{E} \left\{ \left( \widehat{F}_{1k}(X_{2jk}) - F_{1k}(X_{2jk}) \right)^2 F_{1k'}^2(X_{2jk'}) \right\} \leq \frac{2}{n_2^2} \sum_{j=1}^{n_2} \mathbb{E} \left\{ \left( \widehat{F}_{1k}(X_{2jk}) - F_{1k}(X_{2jk}) \right)^2 \right\} \leq \frac{2}{n_1 n_2} \rightarrow 0,$$

and

$$\begin{aligned} & \frac{2}{n_2^2} \sum_{j_1 \neq j_2}^{n_2} \mathbb{E} \left[ \left( \widehat{F}_{1k}(X_{2j_1k}) - F_{1k}(X_{2j_1k}) \right) F_{1k'}(X_{2j_1k'}) \left( \widehat{F}_{1k}(X_{2j_2k}) - F_{1k}(X_{2j_2k}) \right) F_{1k'}(X_{2j_2k'}) \right] \\ & = \frac{2}{n_1^2 n_2^2} \sum_{j_1 \neq j_2}^{n_2} \sum_{\ell_1, \ell_2=1}^{n_1} \mathbb{E} \left[ \left\{ c(X_{2j_1k} - X_{1\ell_1k}) - F_{1k}(X_{2j_1k}) \right\} \left\{ c(X_{2j_2k} - X_{1\ell_2k}) - F_{1k}(X_{2j_2k}) \right\} F_{1k'}(X_{2j_1k'}) F_{1k'}(X_{2j_2k'}) \right] \\ & = \frac{2}{n_1^2 n_2^2} \sum_{j_1 \neq j_2}^{n_2} \sum_{\ell_1=1}^{n_1} \mathbb{E} \left[ \left\{ c(X_{2j_1k} - X_{1\ell_1k}) - F_{1k}(X_{2j_1k}) \right\} \left\{ c(X_{2j_2k} - X_{1\ell_1k}) - F_{1k}(X_{2j_2k}) \right\} F_{1k'}(X_{2j_1k'}) F_{1k'}(X_{2j_2k'}) \right] \\ & \leq \frac{2(n_2 - 1)}{n_1 n_2} \rightarrow 0. \quad \square \end{aligned}$$

**Proof of Lemma 1.** For any  $k \in \{1, \dots, p\}$ , denote  $\Delta_n = \widehat{\sigma}_{1k}^2 / \lambda_1 - \sigma_{1k}^2 / \lambda_1 + \widehat{\sigma}_{2k}^2 / \lambda_2 - \sigma_{2k}^2 / \lambda_2$ . It can be seen from Lemma A7 that  $\widehat{\sigma}_{1k}^2 = \sigma_{1k}^2 + O_p(n^{-1/2})$  and  $\widehat{\sigma}_{2k}^2 = \sigma_{2k}^2 +$

$O_p(n^{-1/2})$ . Thus  $\Delta_n = O_p(n^{-1/2})$ . From Lemma A6,  $n(\hat{\omega}_k - 1/2)^2 = O_p(1)$  under  $H_0$ . Recall that  $\sigma_k^2 = \sigma_{1k}^2/\lambda_1 + \sigma_{2k}^2/\lambda_2$ . Therefore,

$$\hat{t}_k^2 = \frac{(\hat{\omega}_k - 1/2)^2}{\hat{\sigma}_{1k}^2/n_1 + \hat{\sigma}_{2k}^2/n_2} = \frac{n(\hat{\omega}_k - 1/2)^2}{\Delta_n + \sigma_k^2} = n\sigma_k^{-2}(\hat{\omega}_k - 1/2)^2(1 + \sigma_k^{-2}\Delta_n) + O_p(n^{-1}),$$

where the last equality follows from Taylor’s expansion. Taking the expectation, by Lemma A6, it follows that

$$n\sigma_k^{-2}E[(\hat{\omega}_k - 1/2)^2] = 1 + O(n^{-1}).$$

For the second term,

$$(\hat{\omega}_k - 1/2)^2\Delta_n = (\hat{\omega}_k - 1/2)^2(\hat{\sigma}_{1k}^2/\lambda_1 - \sigma_{1k}^2/\lambda_1) + (\hat{\omega}_k - 1/2)^2(\hat{\sigma}_{2k}^2/\lambda_2 - \sigma_{2k}^2/\lambda_2).$$

It suffices to show that  $E[(\hat{\omega}_k - 1/2)^2(\hat{\sigma}_{2k}^2 - \sigma_{2k}^2)] = O(n^{-2})$ , while the proof for the other term is similar. Note that, by Lemma A7, it follows that

$$\begin{aligned} & (\hat{\omega}_k - 1/2)^2(\hat{\sigma}_{2k}^2 - \sigma_{2k}^2) \\ &= \frac{n_2}{n_2 - 1}(\hat{\omega}_k - 1/2)^2 \left\{ \int \hat{F}_{1k}^2 d\hat{F}_{2k} - \int F_{1k}^2 dF_{2k} - \hat{\omega}_k^2 + 1/4 \right\} + O_p(n^{-2}) \\ &= \frac{n_2}{n_2 - 1}(\hat{\omega}_k - 1/2)^2 \left\{ \int (\hat{F}_{1k}^2 - F_{1k}^2) d\hat{F}_{2k} - \int F_{1k}^2 d(\hat{F}_{2k} - F_{2k}) - (\hat{\omega}_k^2 - 1/4) \right\} + O_p(n^{-2}). \end{aligned}$$

Taking the expectation, the first term becomes

$$\begin{aligned} & E\left[(\hat{\omega}_k - 1/2)^2 \int (\hat{F}_{1k}^2 - F_{1k}^2) d\hat{F}_{2k}\right] = \frac{1}{n_2} E\left[(\hat{\omega}_k - 1/2)^2 \sum_{j=1}^{n_2} \left\{ \hat{F}_{1k}^2(X_{2jk}) - F_{1k}^2(X_{2jk}) \right\}\right] \\ &= \frac{1}{n_2^3} \sum_{j_1, j_2=1}^{n_2} E\left[\left\{ \hat{F}_{1k}(X_{2j_1k}) - 1/2 \right\} \left\{ \hat{F}_{1k}(X_{2j_2k}) - 1/2 \right\} \left\{ \hat{F}_{1k}(X_{2jk}) - F_{1k}(X_{2jk}) \right\} \left\{ \hat{F}_{1k}(X_{2jk}) + F_{1k}(X_{2jk}) \right\}\right] \\ &= \frac{1}{n_1^4 n_2^3} \sum_{\ell_1, \ell_2, \ell_3, \ell_4=1}^{n_1} \sum_{j_1, j_2=1}^{n_2} E\left[\left\{ c(X_{2j_1k} - X_{1\ell_1k}) - 1/2 \right\} \left\{ c(X_{2j_2k} - X_{1\ell_2k}) - 1/2 \right\} \cdot \right. \\ & \qquad \qquad \qquad \left. \left\{ c(X_{2jk} - X_{1\ell_3k}) - F_{1k}(X_{2jk}) \right\} \left\{ c(X_{2jk} - X_{1\ell_4k}) + F_{1k}(X_{2jk}) \right\}\right] \\ &= \frac{1}{n_1^4 n_2^3} \left\{ \sum_{\ell_1=\ell_2 \neq \ell_3 \neq \ell_4} \sum_{j_1 \neq j_2}^{n_2} + \sum_{\ell_1 \neq \ell_2 \neq \ell_3 \neq \ell_4} \sum_{j_1=j_2}^{n_2} \right\} E[\dots] + O(n^{-2}) = O(n^{-2}), \end{aligned}$$

since  $E[c(X_{2j_1k} - X_{1\ell_1k}) - 1/2] = 0$  for any  $\ell_1 \in \{1, \dots, n_1\}$  and  $E\left[\left\{ c(X_{2jk} - X_{1\ell_3k}) - F_{1k}(X_{2jk}) \right\} \left\{ c(X_{2jk} - X_{1\ell_4k}) + F_{1k}(X_{2jk}) \right\}\right] = 0$  for any  $\ell_3 \neq \ell_4$ . The expectation of the second term is

$$\begin{aligned} & E\left[(\hat{\omega}_k - 1/2)^2 \int F_{1k}^2 d(\hat{F}_{2k} - F_{2k})\right] = \frac{1}{n_2} E\left[(\hat{\omega}_k - 1/2)^2 \sum_{j=1}^{n_2} \left( F_{1k}^2(X_{2jk}) - E[F_{1k}^2(X_{2jk})] \right)\right] \\ &= \frac{1}{n_1^2 n_2^3} \sum_{\ell_1, \ell_2=1}^{n_1} \sum_{j_1, j_2=1}^{n_2} E\left[\left\{ c(X_{2j_1k} - X_{1\ell_1k}) - 1/2 \right\} \left\{ c(X_{2j_2k} - X_{1\ell_2k}) - 1/2 \right\} \left\{ F_{1k}^2(X_{2jk}) - E[F_{1k}^2(X_{2jk})] \right\}\right] \\ &= \frac{1}{n_1^2 n_2^3} \left\{ \sum_{\ell_1=\ell_2} \sum_{j_1 \neq j_2}^{n_2} + 2 \sum_{\ell_1 \neq \ell_2} \sum_{j_1 \neq j_2}^{n_2} + \sum_{\ell_1 \neq \ell_2} \sum_{j_1=j_2}^{n_2} \right\} E[\dots] + O(n^{-2}) = O(n^{-2}), \end{aligned}$$

The following results, on expectation of the third term, can be proven, similarly,

$$E\left[(\hat{\omega}_k - 1/2)^2(\hat{\omega}_k^2 - 1/4)\right] = E\left[(\hat{\omega}_k - 1/2)^3(\hat{\omega}_k + 1/2)\right] = O(n^{-2}).$$

Using the same arguments as above, it can be shown that  $E[(\hat{\omega}_k - 1/2)^{2r}] = O(n^{-r})$ . Finally, we show  $\sup_k E[\hat{t}_k^{2r}] < \infty$ , for any  $r > 1$ . Since  $1 + \sigma_k^{-2}\Delta_n = \hat{\sigma}_k^2/\sigma_k^{-2} \geq 0$  and  $(\hat{\omega}_k - 1/2)^{2r} \geq 0$ , applying Hölder inequality,

$$E[\hat{t}_k^{2r}] = \sigma_k^{-2r} E[n^r (\hat{\omega}_k - 1/2)^{2r} (1 + \sigma_k^{-2}\Delta_n)^{-r}] \leq \sigma_k^{-2r} \left\{ E[n^{r+1} (\hat{\omega}_k - 1/2)^{2r+2}] \right\}^{\frac{2r}{2r+2}} \left\{ E[(1 + \sigma_k^{-2}\Delta_n)^{-2r(r+1)}] \right\}^{\frac{1}{2r+2}}.$$

First  $E[n^{r+1} (\hat{\omega}_k - 1/2)^{2r+2}] = O(1)$ . It was seen in Lemma A7 that  $\Delta_n$  converges to 0 in  $L_2$ . Thus, applying Taylor’s expansion

$$E[(1 + \sigma_k^{-2}\Delta_n)^{-2r(r+1)}] = E[1 - 2r(r+1)\sigma_k^{-2}\Delta_n + O_p(\Delta_n^2)] \leq 1 + \sigma_k^{-2}E[|\Delta_n|] + O(n^{-1}) \leq \infty.$$

This proves that  $E[\hat{t}_k^{2r}] < \infty$  for any  $k$ . Thus the desired result follows. □

**Proof of Lemma 2.** Under  $H_0$ ,

$$\hat{t}_k^2 = \frac{(\hat{\omega}_k - 1/2)^2}{\hat{\sigma}_{1k}^2/n_1 + \hat{\sigma}_{2k}^2/n_2} = \frac{n(\hat{\omega}_k - 1/2)^2}{\sigma_{1k}^2/\lambda_1 + \sigma_{2k}^2/\lambda_2} \cdot \frac{\sigma_{1k}^2/\lambda_1 + \sigma_{2k}^2/\lambda_2}{\hat{\sigma}_{1k}^2/\lambda_1 + \hat{\sigma}_{2k}^2/\lambda_2} = \frac{n(\hat{\omega}_k - 1/2)^2}{\sigma_{1k}^2/\lambda_1 + \sigma_{2k}^2/\lambda_2} + O_p(n^{-1/2}),$$

where the last equality follows from Taylor’s expansion. Next, consider

$$\begin{aligned} & \text{Cov}\left(n(\hat{\omega}_k - 1/2)^2, n(\hat{\omega}_{k'} - 1/2)^2\right) \\ &= n^2 E\left[(\hat{\omega}_k - 1/2)^2 (\hat{\omega}_{k'} - 1/2)^2\right] - n^2 E[(\hat{\omega}_k - 1/2)^2] E[(\hat{\omega}_{k'} - 1/2)^2]. \end{aligned} \tag{A2}$$

By Lemma A6, the second term in (A2) becomes

$$\begin{aligned} & n^2 \left\{ \sigma_{1k}^2/n_1 + \sigma_{2k}^2/n_2 + O(n^{-2}) \right\} \left\{ \sigma_{1k'}^2/n_1 + \sigma_{2k'}^2/n_2 + O(n^{-2}) \right\} \\ &= (\sigma_{1k}^2/\lambda_1 + \sigma_{2k}^2/\lambda_2) (\sigma_{1k'}^2/\lambda_1 + \sigma_{2k'}^2/\lambda_2) + O(n^{-1}). \end{aligned}$$

The first term in (A2) is

$$\begin{aligned} & n^2 E\left[(\hat{\omega}_k - 1/2)^2 (\hat{\omega}_{k'} - 1/2)^2\right] \\ &= \frac{n^2}{n_1^4 n_2^4} \sum_{\ell_1, \ell_2, \ell_3, \ell_4=1}^{n_1} \sum_{j_1, j_2, j_3, j_4=1}^{n_2} E\left[\left\{c(X_{2j_1k} - X_{1\ell_1k}) - \omega_k\right\} \left\{c(X_{2j_2k} - X_{1\ell_2k}) - \omega_k\right\} \right. \\ & \qquad \qquad \qquad \left. \left\{c(X_{2j_3k'} - X_{1\ell_3k'}) - \omega_{k'}\right\} \left\{c(X_{2j_4k'} - X_{1\ell_4k'}) - \omega_{k'}\right\}\right] \\ &= \frac{n^2}{n_1^4 n_2^4} \sum_{\ell_1 \neq \ell_2 \neq \ell_3}^{n_1} \sum_{j_1 \neq j_2 \neq j_3}^{n_2} \left\{ E\left[\left\{c(X_{2j_1k} - X_{1\ell_2k}) - \omega_k\right\} \left\{c(X_{2j_1k} - X_{1\ell_3k}) - \omega_k\right\}\right] \right. \\ & \qquad \qquad \qquad E\left[\left\{c(X_{2j_2k'} - X_{1\ell_1k'}) - \omega_{k'}\right\} \left\{c(X_{2j_3k'} - X_{1\ell_1k'}) - \omega_{k'}\right\}\right] \\ & \qquad \qquad \qquad + E\left[\left\{c(X_{2j_2k} - X_{1\ell_1k}) - \omega_k\right\} \left\{c(X_{2j_3k} - X_{1\ell_1k}) - \omega_k\right\}\right] \\ & \qquad \qquad \qquad E\left[\left\{c(X_{2j_1k'} - X_{1\ell_2k'}) - \omega_{k'}\right\} \left\{c(X_{2j_1k'} - X_{1\ell_3k'}) - \omega_{k'}\right\}\right] \\ & \qquad \qquad \qquad + 4E\left[\left\{c(X_{2j_1k} - X_{1\ell_2k}) - \omega_k\right\} \left\{c(X_{2j_1k'} - X_{1\ell_3k'}) - \omega_{k'}\right\}\right] \\ & \qquad \qquad \qquad \left. E\left[\left\{c(X_{2j_2k} - X_{1\ell_1k}) - \omega_k\right\} \left\{c(X_{2j_3k'} - X_{1\ell_1k'}) - \omega_{k'}\right\}\right] \right\} \\ & + \frac{n^2}{n_1^4 n_2^4} \sum_{\ell_1 \neq \ell_2}^{n_1} \sum_{j_1 \neq j_2 \neq j_3 \neq j_4}^{n_2} \left\{ E\left[\left\{c(X_{2j_1k} - X_{1\ell_1k}) - \omega_k\right\} \left\{c(X_{2j_2k} - X_{1\ell_1k}) - \omega_k\right\}\right] \right. \\ & \qquad \qquad \qquad \left. E\left[\left\{c(X_{2j_3k'} - X_{1\ell_2k'}) - \omega_{k'}\right\} \left\{c(X_{2j_4k'} - X_{1\ell_2k'}) - \omega_{k'}\right\}\right] \right\} \end{aligned}$$

$$\begin{aligned}
 &+ 2E\left[\left\{c(X_{2j_1k} - X_{1\ell_1k}) - \omega_k\right\}\left\{c(X_{2j_2k'} - X_{1\ell_1k'}) - \omega_{k'}\right\}\right] \\
 &E\left[\left\{c(X_{2j_3k} - X_{1\ell_2k}) - \omega_k\right\}\left\{c(X_{2j_4k'} - X_{1\ell_2k'}) - \omega_{k'}\right\}\right] \\
 &+ \frac{n^2}{n_1^4 n_2^4} \sum_{\ell_1 \neq \ell_2 \neq \ell_3 \neq \ell_4} \sum_{j_1 \neq j_2}^{n_1} \sum_{j_3 \neq j_4}^{n_2} \left\{E\left[\left\{c(X_{2j_1k} - X_{1\ell_1k}) - \omega_k\right\}\left\{c(X_{2j_1k} - X_{1\ell_2k}) - \omega_k\right\}\right]\right. \\
 &E\left[\left\{c(X_{2j_2k'} - X_{1\ell_3k'}) - \omega_{k'}\right\}\left\{c(X_{2j_2k'} - X_{1\ell_4k'}) - \omega_{k'}\right\}\right] \\
 &+ 2E\left[\left\{c(X_{2j_1k} - X_{1\ell_1k}) - \omega_k\right\}\left\{c(X_{2j_1k'} - X_{1\ell_2k'}) - \omega_{k'}\right\}\right] \\
 &E\left[\left\{c(X_{2j_2k} - X_{1\ell_3k}) - \omega_k\right\}\left\{c(X_{2j_2k'} - X_{1\ell_4k'}) - \omega_{k'}\right\}\right] \left. + O(n^{-1})\right\} \\
 &= \frac{n^2}{n_1^4 n_2^4} \left\{ \sum_{\ell_1 \neq \ell_2 \neq \ell_3} \sum_{j_1 \neq j_2 \neq j_3}^{n_1} (\sigma_{1k'}^2 \sigma_{2k}^2 + \sigma_{1k}^2 \sigma_{2k'}^2 + 4\sigma_{1kk'} \sigma_{2kk'}) \right. \\
 &\quad \left. + \sum_{\ell_1 \neq \ell_2} \sum_{j_1 \neq j_2 \neq j_3 \neq j_4}^{n_1} (\sigma_{1k}^2 \sigma_{1k'}^2 + 2\sigma_{1kk'}^2) + \sum_{\ell_1 \neq \ell_2 \neq \ell_3 \neq \ell_4} \sum_{j_1 \neq j_2}^{n_2} (\sigma_{2k}^2 \sigma_{2k'}^2 + 2\sigma_{2kk'}^2) \right\} + O(n^{-1}) \\
 &= 2(\sigma_{1kk'} / \lambda_1 - \sigma_{2kk'} / \lambda_2)^2 + (\sigma_{1k}^2 / \lambda_1 + \sigma_{2k}^2 / \lambda_2)(\sigma_{1k'}^2 / \lambda_1 + \sigma_{2k'}^2 / \lambda_2) + O(n^{-1}),
 \end{aligned}$$

by Lemmas A3 and A4. Putting the simplified forms back in two terms of (A2), it follows that

$$\text{Cov}\left(n(\hat{\omega}_k - 1/2)^2, n(\hat{\omega}_{k'} - 1/2)^2\right) = 2(\sigma_{1kk'} / \lambda_1 - \sigma_{2kk'} / \lambda_2)^2 + O(n^{-1}).$$

Therefore,

$$\text{Cov}\left(\hat{t}_k^2, \hat{t}_{k'}^2\right) = \frac{2(\sigma_{1kk'} / \lambda_1 - \sigma_{2kk'} / \lambda_2)^2}{\sigma_{1k}^2 / \lambda_1 + \sigma_{2k}^2 / \lambda_2} + O(n^{-1/2}). \quad \square$$

**Proof of Lemma 3.** By Lemma 1,  $\sup_k E[\hat{t}_k^2 \hat{t}_{k'}^2] < \infty$  and, hence, by Cauchy–Schwarz inequality  $\gamma_{kk'} < \infty$ . By Lemmas A7 and A8, it follows that  $\hat{\sigma}_{1k}^2 = \sigma_{1k}^2 + O_p(n^{-1/2})$ ,  $\hat{\sigma}_{2k}^2 = \sigma_{2k}^2 + O_p(n^{-1/2})$ ,  $\hat{\sigma}_{1kk'} = \sigma_{1kk'} + O_p(n^{-1/2})$ , and  $\hat{\sigma}_{2kk'} = \sigma_{2kk'} + O_p(n^{-1/2})$  for any  $k, k' \in \{1, \dots, p\}$ . Thus, following similar arguments, as in the proof of Lemma 2.4 of Zhang and Wang [20], we can get the desired result.  $\square$

**References**

1. Harrar, S.W.; Kong, X. Recent developments in high-dimensional inference for multivariate data: parametric, semiparametric and nonparametric approaches. *J. Multivar. Anal.* **2022**, *188*, 104855. [\[CrossRef\]](#)
2. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, 3rd ed.; Wiley Series in Probability and Statistics; Wiley-Interscience: Hoboken, NJ, USA, 2003.
3. Bai, Z.; Saranadasa, H. Effect of high dimension: By an example of a two sample problem. *Stat. Sin.* **1996**, *6*, 311–329.
4. Brunner, E.; Munzel, U. The nonparametric Behrens-Fisher problem: asymptotic theory and a small sample approximation. *Biom. J.* **2000**, *42*, 17–25. [\[CrossRef\]](#)
5. Brunner, E.; Munzel, U.; Puri, M.L. The multivariate nonparametric Behrens-Fisher problem. *J. Stat. Plan. Inference* **2002**, *108*, 37–53. [\[CrossRef\]](#)
6. Brunner, E.; Konietzschke, F.; Pauly, M.; Puri, M.L. Rank-based procedures in factorial designs: hypotheses about non-parametric treatment effects. *J. R. Stat. Soc. Ser. B* **2017**, *79*, 1463–1485. [\[CrossRef\]](#)
7. Konietzschke, F.; Aguayo, R.R.; Staab, W. Simultaneous inference for factorial multireader diagnostic trials. *Stat. Med.* **2018**, *37*, 28–47. [\[CrossRef\]](#)
8. Dobler, D.; Friedrich, S.; Pauly, M. Nonparametric MANOVA in meaningful effects. *Ann. Inst. Stat. Math.* **2020**, *72*, 997–1022. [\[CrossRef\]](#)
9. Bathke, A.C.; Harrar, S.W. Nonparametric methods in multivariate factorial designs for large number of factor levels. *J. Stat. Plan. Inference* **2008**, *138*, 588–610. [\[CrossRef\]](#)
10. Harrar, S.W.; Bathke, A.C. Nonparametric methods for unbalanced multivariate data and many factor levels. *J. Multivar. Anal.* **2008**, *99*, 1635–1664. [\[CrossRef\]](#)
11. Bathke, A.C.; Harrar, S.W.; Madden, L.V. How to compare small multivariate samples using nonparametric tests. *Comput. Stat. Data Anal.* **2008**, *52*, 4951–4965. [\[CrossRef\]](#)
12. Burchett, W.W.; Ellis, A.R.; Harrar, S.W.; Bathke, A.C. Nonparametric inference for multivariate data: the R package nrmv. *J. Stat. Softw.* **2017**, *76*, 1–18. [\[CrossRef\]](#)

13. Bathke, A.C.; Harrar, S.W. Rank-based inference for multivariate data in factorial designs. In *Robust Rank-BASED and Nonparametric Methods*; Springer: Berlin, Germany, 2016; 121–139.
14. Wang, H.; Akritas, M.G. Rank test for heteroscedastic functional data. *J. Multivar. Anal.* **2010**, *101*, 1791–1805. [[CrossRef](#)]
15. Kong, X.; Harrar, S.W. High-dimensional rank-based inference. *J. Nonparametr. Stat.* **2020**, *32*, 294–322. [[CrossRef](#)]
16. Ruymgaart, F.H. *Statistique non Paramétrique Asymptotique: A Unified Approach to the Asymptotic Distribution Theory of Certain Midrank Statistics*; Springer: Berlin/Heidelberg, Germany, 1980; pp. 1–18.
17. Akritas, M.G.; Brunner, E. A unified approach to rank tests for mixed models. *J. Stat. Plan. Inference* **1997**, *61*, 249–277. [[CrossRef](#)]
18. Brunner, E.; Munzel, U.; Puri, M.L. Rank-score tests in factorial designs with repeated measures. *J. Multivar. Anal.* **1999**, *70*, 286–317. [[CrossRef](#)]
19. Gregory, K.B.; Carroll, R.J.; Baladandayuthapani, V.; Lahiri, S.N. A two-sample test for equality of means in high dimension. *J. Am. Stat. Assoc.* **2015**, *110*, 837–849. [[CrossRef](#)]
20. Zhang, H.; Wang, H. A more powerful test of equality of high-dimensional two-sample means. *Comput. Stat. Data Anal.* **2021**, *164*, 107318. [[CrossRef](#)]
21. Srivastava, M.S.; Du, M. A test for the mean vector with fewer observations than the dimension. *J. Multivar. Anal.* **2008**, *99*, 386–402. [[CrossRef](#)]
22. Srivastava, M.S.; Katayama, S.; Kano, Y. A two sample test in high dimensional data. *J. Multivar. Anal.* **2013**, *114*, 349–358. [[CrossRef](#)]
23. Brockwell, P.J.; Davis, R.A. *Time Series: Theory and Methods*; Springer Series in Statistics; Springer: New York, NY, USA, 2013.
24. Politis, D.N.; Romano, J.P. Bias-corrected nonparametric spectral estimation. *J. Time Ser. Anal.* **1995**, *16*, 67–103. [[CrossRef](#)]
25. Xu, G.; Lin, L.; Wei, P.; Pan, W. An adaptive two-sample test for high-dimensional means. *Biometrika* **2016**, *103*, 609–624. [[CrossRef](#)] [[PubMed](#)]
26. Chen, S.; Li, J.; Zhong, P. Two-sample and ANOVA tests for high dimensional means. *Ann. Stat.* **2019**, *47*, 1443–1474. [[CrossRef](#)]
27. Kong, X.; Harrar, S.W. High-dimensional MANOVA under weak conditions. *Stat. A J. Theor. Appl. Stat.* **2021**, *55*, 321–349. [[CrossRef](#)]
28. Bradley, R.C. Basic properties of strong mixing conditions. A Survey and some open questions. *Probab. Surv.* **2005**, *2*, 107–144. [[CrossRef](#)]
29. Hall, P.; Jing, B.; Lahiri, S. On the sampling window method for long-range dependent data. *Stat. Sin.* **1998**, *8*, 1189–1204.
30. Samorodnitsky, G. *Long Range Dependence*; Now Publishers Inc.: Hanover, MA, USA, 2007.