Annals of the Institute of Statistical Mathematics (2024) 76:579–615
https://doi.org/10.1007/s10463-024-00896-8

Multi-sample hypothesis testing of high-dimensional mean vectors under covariance heterogeneity

Lixiu Wu[1] · **Jiang Hu**[1]

Received: 19 December 2022 / Revised: 18 December 2023 / Accepted: 4 January 2024 /
Published online: 22 March 2024

Abstract
In this paper, we focus on the hypothesis testing problem of the mean vectors of high-dimensional data in the multi-sample case. We propose two maximum-type statistics and apply a parametric bootstrap technique to compute the critical values. Unlike previous hypothesis testing methods that heavily depend on the structural assumptions of the unknown covariance matrix, the proposed methods accommodate a general covariance structure. Additionally, we introduce screening-based testing procedures to enhance the power of our tests. These test procedures do not require the use of approximate limiting distributions for the test statistics. Finally, we obtain and verify the theoretical properties through simulation studies.

Keywords Multi-sample hypothesis · High dimension · Parametric bootstrap · Maximum-type statistics

1 Introduction

High-dimensional data are essential to the integration of the new generation of information technology as society develops. More and more sectors, like genomics and biomedical research, will require the processing of high-dimensional data. Runcie and Mukherjee (2013) estimated the covariance matrix of high-dimensional genetic data by Bayesian analysis, Pan and Shen (2011) generalized the variable thresholds test and applied these methods to the field of genetic data analysis. In medical research, Zou et al. (2008) proposed a fractional ALFF approach to analyze brain data. In the area of high-dimensional big data analysis,

✉ Jiang Hu
huj156@nenu.edu.cn

Lixiu Wu
wulx838@nenu.edu.cn

[1] KLASMOE and School of Mathematics and Statistics, Northeast Normal University, 5268 Renmin Street, Changchun, Jilin, China

both the dimension and sample size are frequently rather large, with the dimension frequently being significantly bigger than the sample size. The following is an introduction to high-dimensional multivariate analysis of variance.

Let $\mathcal{X}_1 = \{\boldsymbol{X}_{11}, \dots, \boldsymbol{X}_{1n_1}\}, \dots, \mathcal{X}_K = \{\boldsymbol{X}_{K1}, \dots, \boldsymbol{X}_{Kn_K}\}$ represent K groups of random vectors, where $\boldsymbol{X}_{ij} = (X_{ij1}, \dots, X_{ijp})', i = 1, \dots, K, j = 1, \dots, n_i$. Assume that K is known and each group to be composed of independent and identically distributed (i.i.d.) observations, each endowed with mean vectors $\boldsymbol{\mu}_1 = (\mu_{11}, \dots, \mu_{1p})', \dots, \boldsymbol{\mu}_K = (\mu_{K1}, \dots, \mu_{Kp})'$ and associated covariance matrices $\boldsymbol{\Sigma}_1 = (\sigma_{1,kl})_{1 \le k,l \le p}, \dots, \boldsymbol{\Sigma}_K = (\sigma_{K,kl})_{1 \le k,l \le p}$, respectively. In this paper, the "covariance heterogeneity" means that the covariance matrices $\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_K$ are allowed to be not identical. This article aims to test the hypothesis

$$\boldsymbol{H}_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_K \quad v.s. \quad \boldsymbol{H}_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j \quad \textit{for some } i \neq j. \tag{1}$$

Numerous studies have addressed the two-sample high-dimensional hypothesis testing problem (i.e., $K = 2$), such as those by Bai and Saranadasa (1996), Cai et al. (2014), Huang et al. (2021), Srivastava (2009) and Chen and Qin (2010). These results indicate that the sum of squares-type statistics are robustly analyzed and powerful. In situations where the covariance matrices are unknown and unequal, Chang et al. (2017) computed the critical value using bootstrap techniques based on maximum type statistics. Based on Bai and Saranadasa (1996); Schott (2007) presented a one-way MANOVA test for $K > 2$ with common covariance matrices. Under certain mild assumptions, Aoshima and Yata (2015) examined the asymptotic normality of the multi-sample test statistic. In cases where the covariance matrices differ, Hu et al. (2017) proposed test statistics for the high-dimensional multi-sample mean vectors hypothesis testing problem, drawing upon Chen and Qin (2010)'s two-sample U-statistic. Yamada and Himeno (2015) were concerned with the problem of multi-sample mean vectors hypothesis testing without assuming common covariance matrices. Zhang and Xu (2009) extended Scheffe's transformation method to get an approximate solution of the test statistic. The critical value in the traditional hypothesis testing procedure was often calculated by determining the limiting null distribution of the test statistics and in the high-dimensional cases, there are numerous assumptions regarding the structure of the covariance matrix (see Cai and Xia 2014; Zhong et al. 2013; Kong and Harrar 2021; Srivastava 2009 for example). However, some real data may not satisfy these structure assumptions. For example, in Wolen and Miles (2012), we know that in the same path, the gene expression level is highly correlated, which will lead to the covariance matrices being more complex, thus the existing methods of hypotheses testing based on certain covariance matrices may not be applicable.

Inspired by Chang et al. (2017) and Chernozhukov et al. (2013), we propose in this paper several test procedures that do not rely on the covariance matrix structural assumptions. Next, we introduce the main idea in Chang et al. (2017). Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_{n_1} \in \mathbb{R}^p$ and $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{n_2} \in \mathbb{R}^p$ be two groups i.i.d. random vectors with mean vectors $\mathbb{E}(\boldsymbol{X}_i) = \boldsymbol{\mu}_1$ and $\mathbb{E}(\boldsymbol{Y}_i) = \boldsymbol{\mu}_2$, respectively. Considering the following hypothesis testing problem

$$\boldsymbol{H}_{00}:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2 \quad v.s. \quad \boldsymbol{H}_{10}:\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2. \tag{2}$$

For the two-sample hypothesis testing problem (2), Chang et al. (2017) proposed two maximum-type statistics

$$T_1=\max_{1\le l\le p}\frac{\sqrt{n_1n_2}|\bar{X}_l-\bar{Y}_l|}{\sqrt{n_1+n_2}} \quad \text{and} \quad T_2=\max_{1\le l\le p}\frac{\sqrt{n_1n_2}|\bar{X}_l-\bar{Y}_l|}{\sqrt{n_2\hat{\sigma}_{1l}^2+n_1\hat{\sigma}_{2l}^2}},$$

where T_1 and T_2 are referred to as non-studentized statistic and studentized statistic, respectively. Here $\bar{X}_l=n_1^{-1}\sum_{i=1}^{n_1}X_{il}$ (resp. $\bar{Y}_l=n_2^{-1}\sum_{i=1}^{n_2}Y_{il}$) and $\hat{\sigma}_{1l}^2=n_1^{-1}\sum_{i=1}^{n_1}\left(X_{il}-\bar{X}_l\right)^2$ (resp. $\hat{\sigma}_{2l}^2=n_2^{-1}\sum_{i=1}^{n_2}\left(Y_{il}-\bar{Y}_l\right)^2$). Instead of requiring the limiting null distribution of the test statistics as in earlier methods, the critical values were determined using parametric bootstrap techniques. Therefore, the multi-sample mean vectors hypothesis testing approaches are suggested consequently. Let $\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_K$ be the covariance matrices and $\tilde{\boldsymbol{\Sigma}}_1,\ldots,\tilde{\boldsymbol{\Sigma}}_K$ be the estimators of $\boldsymbol{\Sigma}_1$, $\ldots,\boldsymbol{\Sigma}_K$, respectively. For $i=1,\ldots,K$ and $j=1,\ldots,K$, define

$$\boldsymbol{D}_{ij}=diag\left(n_i\boldsymbol{\Sigma}_j+n_j\boldsymbol{\Sigma}_i\right) \text{ and } \tilde{\boldsymbol{D}}_{ij}=diag\left(n_i\tilde{\boldsymbol{\Sigma}}_j+n_j\tilde{\boldsymbol{\Sigma}}_i\right).$$

Now we introduce the test statistics for problem (1) as follows:

$$\boldsymbol{T}_{ns}=\begin{bmatrix}\frac{\sqrt{n_1n_2}(\bar{X}_1-\bar{X}_2)}{\sqrt{n_1+n_2}}\\ \frac{\sqrt{n_1n_3}(\bar{X}_1-\bar{X}_3)}{\sqrt{n_1+n_3}}\\ \vdots\\ \frac{\sqrt{n_{K-1}n_K}(\bar{X}_{K-1}-\bar{X}_K)}{\sqrt{n_{K-1}+n_K}}\end{bmatrix}$$

and

$$\boldsymbol{T}_s=\begin{bmatrix}\sqrt{n_1n_2}\tilde{\boldsymbol{D}}_{12}^{-1/2}\left(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2\right)\\ \sqrt{n_1n_3}\tilde{\boldsymbol{D}}_{13}^{-1/2}\left(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_3\right)\\ \vdots\\ \sqrt{n_{K-1}n_K}\tilde{\boldsymbol{D}}_{K-1K}^{-1/2}\left(\bar{\boldsymbol{X}}_{K-1}-\bar{\boldsymbol{X}}_K\right)\end{bmatrix}.$$

In the sequel, we call $\boldsymbol{T}_{ns}$ and $\boldsymbol{T}_s$ non-studentized vector and studentized vector, respectively. If the infinite norm of $\boldsymbol{T}_{ns}$ or $\boldsymbol{T}_s$ is above a certain threshold, we reject $\boldsymbol{H}_0$. The critical values of the proposed test are computed using parametric bootstrap techniques, and the details are postponed in the next section.

The rest of the article is organized as follows. In Sect. 2, we present the proposed test methods in detail. In Sect. 3, we establish the theoretical properties of the proposed tests. Simulation studies are presented in Sect. 4. In Sect. 5, we present an example of real data analysis. In Sect. 6, we have added a discussion section to summarize the contributions and limitations of the proposed approach. All theoretical proofs are postponed in Appendix.

Notation. Before presenting our methods, we provide some notations for convenient use later. In the sequel, we denote $|\boldsymbol{B}|_\infty = \max_{1\le l_1\le k, 1\le l_2\le p} |b_{l_1 l_2}|$ for any $k\times p$ matrix $\boldsymbol{B} = (b_{l_1 l_2})$. Let $\mathbf{1}(A)$ be an indicator function, i.e., if the event A happens $\mathbf{1}(A) = 1$, and if the event A does not happen $\mathbf{1}(A) = 0$. For $\mathcal{X}_i = \{\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i}\}$, $i = 1, \dots, K$ and $\boldsymbol{X}_{ij} = \left(X_{ij1}, \dots, X_{ijp}\right)'$, define $\bar{X}_{il} = n_i^{-1}\sum_{j=1}^{n_i} X_{ijl}$ and $\hat{\sigma}_{il}^2 = n_i^{-1}\sum_{j=1}^{n_i}\left(X_{ijl} - \bar{X}_{il}\right)^2$, $i = 1, \dots, K, l = 1, \dots, p$.

2 Methodology

In this section, we introduce testing methods: one based on the non-studentized statistic and the other on the studentized statistic. Additionally, we outline screening-based testing procedures for each method.

2.1 Non-studentized statistic-based test procedure

We first present the testing method based on the non-studentized statistic. Subsequently, we employ a screening process to filter out irrelevant variables and conduct the test in a reduced-dimensional space.

2.1.1 Non-studentized statistic-based test without screening

In this section, we detail the procedure of the non-studentized statistic based test. For the test problem (1), we propose the non-studentized test statistic T_{nsmax} as

$$T_{nsmax} = \max_{1\le l\le p, 1\le i<j\le K}\left\{\frac{\sqrt{n_i n_j}|\bar{X}_{il} - \bar{X}_{jl}|}{\left(n_i + n_j\right)^{1/2}}\right\},$$

which is actually the infinite norm of $\boldsymbol{T}_{ns}$. The estimator of covariance matrix of $\boldsymbol{T}_{ns}$ can be denoted by $\tilde{\boldsymbol{R}}_{ns}$, and can be written in the following form:

$$\tilde{\boldsymbol{R}}_{ns} = \begin{bmatrix} \frac{\left(n_2\tilde{\boldsymbol{\Sigma}}_1 + n_1\tilde{\boldsymbol{\Sigma}}_2\right)}{N_{12}} & \frac{\sqrt{n_2 n_3}\tilde{\boldsymbol{\Sigma}}_1}{N_{12}^{1/2}N_{13}^{1/2}} & \frac{\sqrt{n_2 n_4}\tilde{\boldsymbol{\Sigma}}_1}{N_{12}^{1/2}N_{14}^{1/2}} & \\ \frac{\sqrt{n_2 n_3}\tilde{\boldsymbol{\Sigma}}_1}{N_{12}^{1/2}N_{13}^{1/2}} & \frac{\left(n_3\tilde{\boldsymbol{\Sigma}}_1 + n_1\tilde{\boldsymbol{\Sigma}}_3\right)}{N_{13}} & \frac{\sqrt{n_3 n_4}\tilde{\boldsymbol{\Sigma}}_1}{N_{13}^{1/2}N_{14}^{1/2}} & \cdots \\ \frac{\sqrt{n_2 n_4}\tilde{\boldsymbol{\Sigma}}_1}{N_{12}^{1/2}N_{14}^{1/2}} & \frac{\sqrt{n_3 n_4}\tilde{\boldsymbol{\Sigma}}_1}{N_{13}^{1/2}N_{14}^{1/2}} & \frac{\left(n_4\tilde{\boldsymbol{\Sigma}}_1 + n_1\tilde{\boldsymbol{\Sigma}}_4\right)}{N_{14}} & \\ & \vdots & & \ddots \end{bmatrix}.$$

Here $N_{ij} = n_i + n_j$ for $1 \le i,j \le K$. The critical value can be computed by

$$c_{ns,\alpha} = \inf\left\{t \in \mathbb{R} : P\left(\left|\boldsymbol{W}_{ns}\right|_\infty \ge t | \mathcal{X}_1, \dots, \mathcal{X}_K\right) \le \alpha\right\},$$

where $\boldsymbol{W}_{ns} \sim N(\mathbf{0}, \tilde{\boldsymbol{R}}_{ns})$ be a $K(K-1)p/2$-dimensional Gaussian random vector. Next, we aim to estimate the critical value. We generate random vectors $\left\{\boldsymbol{W}_{ns,l}\right\}_{l=1}^{N_{ns}} \sim N\left(\mathbf{0}, \tilde{\boldsymbol{R}}_{ns}\right)$ and use $\hat{c}_{ns,\alpha}$ to estimate the critical value, denoted as follows:

$$\hat{c}_{ns,\alpha} = \inf\left\{ x \in \mathbb{R} : N_{ns}^{-1} \sum_{l=1}^{N_{ns}} \mathbf{1}\left\{\left|\boldsymbol{W}_{ns,l}\right| \propto \leq x\right\} \geq 1-\alpha \right\}.$$

Thus, we reject the null hypothesis $\boldsymbol{H}_0$ whenever $T_{nsmax} \geq \hat{c}_{ns,\alpha}$.

2.1.2 Non-studentized statistic-based test with screening

To simplify the testing procedure and improve the power of the test, we propose a screening-based testing procedure. Define the set $\boldsymbol{S} = \{1 \leq l \leq p : \mu_{1l} = \mu_{2l} = \cdots = \mu_{Kl}\}$. We first eliminate irrelevant features indexed by $\boldsymbol{S}$. Thus, we need to estimate the set $\boldsymbol{S}$. Let $K_s = K(K-1)p/2$, we define the estimator of $\boldsymbol{S}$ as follows:

$$\hat{\boldsymbol{S}} = \left\{ 1 \leq l \leq p : T_{sl}^{sc} \leq \left(1 + \left\{2\log\left(K_s\right)\right\}^{-1}\right)\sqrt{2\log\left(K_s\right)} + \sqrt{0.0002\log(1/\alpha)} \right\},$$

where

$$T_{sl}^{sc} = \max\left\{ \frac{\sqrt{n_1 n_2}|\bar{X}_{1l} - \bar{X}_{2l}|}{\left(n_2\hat{\sigma}_{1l}^2 + n_1\hat{\sigma}_{2l}^2\right)^{1/2}}, \frac{\sqrt{n_1 n_3}|\bar{X}_{1l} - \bar{X}_{3l}|}{\left(n_3\hat{\sigma}_{1l}^2 + n_1\hat{\sigma}_{3l}^2\right)^{1/2}}, \ldots, \frac{\sqrt{n_{K-1}n_K}\left|\bar{X}_{(K-1)l} - \bar{X}_{Kl}\right|}{\left(n_K\hat{\sigma}_{K-1l}^2 + n_{K-1}\hat{\sigma}_{Kl}^2\right)^{1/2}} \right\}.$$

The estimator of set $\boldsymbol{S}$ is inspired by the following test. We recharacterize the original global test with the following marginal hypothesis test: for $1 \leq l \leq p$,

$$\boldsymbol{H}_{0l} : \mu_{1l} = \cdots = \mu_{Kl} \quad v.s. \quad \boldsymbol{H}_{1l} : \mu_{il} \neq \mu_{jl} \text{ for some } i \neq j. \tag{3}$$

For the test problem 3, we define the test statistic T_{sl}^{sc}, and thus obtain $\hat{\boldsymbol{S}}$. We next focus only on the following hypothesis

$$\boldsymbol{H}_{0S} : \boldsymbol{\mu}_1^s = \cdots = \boldsymbol{\mu}_K^s \quad v.s. \quad \boldsymbol{H}_{1S} : \boldsymbol{\mu}_i^s \neq \boldsymbol{\mu}_j^s, \text{ for some } i \neq j, \tag{4}$$

where $\boldsymbol{\mu}_1^s, \ldots, \boldsymbol{\mu}_K^s$ are the sub-vectors of $\boldsymbol{\mu}_1 \in \mathbb{R}^p, \ldots, \boldsymbol{\mu}_K \in \mathbb{R}^p$ with the subscripts not in the index set $\hat{\boldsymbol{S}}$, respectively. We give the test statistic

$$T_{nsmax}^s = \max_{l \notin \hat{S}}\left\{ \frac{\sqrt{n_1 n_2}|\bar{X}_{1l} - \bar{X}_{2l}|}{\left(n_2 + n_1\right)^{1/2}}, \frac{\sqrt{n_1 n_3}|\bar{X}_{1l} - \bar{X}_{3l}|}{\left(n_3 + n_1\right)^{1/2}}, \ldots, \frac{\sqrt{n_{K-1}n_K}\left|\bar{X}_{(K-1)l} - \bar{X}_{Kl}\right|}{\left(n_{K-1} + n_K\right)^{1/2}} \right\}.$$

The critical value can be denoted as follows

$$c^s_{ns,\alpha} = \inf\left\{x \in \mathbb{R} : P\left(\max_{l \notin \hat{\boldsymbol{S}}_{all}} |W_{nsl}| \geq x | \mathcal{X}_1, \ldots, \mathcal{X}_K\right) \leq \alpha\right\},$$

where $\boldsymbol{W}_{ns}$ has been defined previously. We reject the null hypothesis $\boldsymbol{H}_{0S}$ whenever $T^s_{nsmax} \geq c^s_{ns,\alpha}$. Furthermore, $\hat{\boldsymbol{S}}_{all} = \{l + ip : i = 0, 1, \ldots, K(K-1)/2, l \in \hat{\boldsymbol{S}}\}$. If $Card(\hat{\boldsymbol{S}}_{all}) = K(K-1)p/2$, we don't reject $\boldsymbol{H}_{0S}$.

2.2 Studentized statistic-based test procedure

We present testing approaches based on the studentized statistic, considering both screening and non-screening conditions. Subsequently, we will provide a detailed exposition.

2.2.1 Studentized statistic-based test without screening

For the hypothesis testing problem denoted as (1), we define the studentized test statistic as

$$T_{smax} = \max_{1 \leq l \leq p, 1 \leq i < j \leq K} \left\{ \frac{\sqrt{n_i n_j}\left|\bar{X}_{il} - \bar{X}_{jl}\right|}{\left(n_i \hat{\sigma}^2_{jl} + n_j \hat{\sigma}^2_{il}\right)^{1/2}} \right\},$$

which is the infinite norm of $\boldsymbol{T}_s$. The estimator of the covariance matrix of $\boldsymbol{T}_s$ can be denoted by $\tilde{\boldsymbol{R}}_s$, and can be written in the following form:

$$\tilde{\boldsymbol{R}}_s = \begin{bmatrix} \tilde{\boldsymbol{D}}_{12}^{-\frac{1}{2}}(n_2\tilde{\boldsymbol{\Sigma}}_1 + n_1\tilde{\boldsymbol{\Sigma}}_2)\tilde{\boldsymbol{D}}_{12}^{-\frac{1}{2}} & \tilde{\boldsymbol{D}}_{12}^{-\frac{1}{2}}\sqrt{n_2n_3}\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{D}}_{13}^{-\frac{1}{2}} & \tilde{\boldsymbol{D}}_{12}^{-\frac{1}{2}}\sqrt{n_2n_4}\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{D}}_{14}^{-\frac{1}{2}} & \\ \tilde{\boldsymbol{D}}_{13}^{-\frac{1}{2}}\sqrt{n_2n_3}\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{D}}_{12}^{-\frac{1}{2}} & \tilde{\boldsymbol{D}}_{13}^{-\frac{1}{2}}(n_3\tilde{\boldsymbol{\Sigma}}_1 + n_1\tilde{\boldsymbol{\Sigma}}_3)\tilde{\boldsymbol{D}}_{13}^{-\frac{1}{2}} & \tilde{\boldsymbol{D}}_{13}^{-\frac{1}{2}}\sqrt{n_3n_4}\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{D}}_{14}^{-\frac{1}{2}} & \cdots \\ \tilde{\boldsymbol{D}}_{14}^{-\frac{1}{2}}\sqrt{n_2n_4}\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{D}}_{12}^{-\frac{1}{2}} & \tilde{\boldsymbol{D}}_{14}^{-\frac{1}{2}}\sqrt{n_3n_4}\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{D}}_{13}^{-\frac{1}{2}} & \tilde{\boldsymbol{D}}_{14}^{-\frac{1}{2}}(n_1\tilde{\boldsymbol{\Sigma}}_4 + n_4\tilde{\boldsymbol{\Sigma}}_1)\tilde{\boldsymbol{D}}_{14}^{-\frac{1}{2}} & \\ & \vdots & & \ddots \end{bmatrix}.$$

Analogous to non-studentized statistics, let $\boldsymbol{W}_s \sim N\left(\boldsymbol{0}, \tilde{\boldsymbol{R}}_s\right)$, and the critical value can be denoted as

$$c_{s,\alpha} = \inf\left\{x \in \mathbb{R} : P\left(\left|\boldsymbol{W}_s\right|_\infty \geq x | \mathcal{X}_1, \ldots, \mathcal{X}_K\right) \leq \alpha\right\}.$$

The test procedure is that we first generate $\left\{\boldsymbol{W}_{s,l}\right\}_{l=1}^{N_s} \sim N\left(\boldsymbol{0}, \tilde{\boldsymbol{R}}_s\right)$, and then, the critical value can be estimated as

$$\hat{c}_{s,\alpha} = \inf\left\{x \in \mathbb{R} : N_s^{-1}\sum_{l=1}^{N_s} \boldsymbol{1}\left\{\left|\boldsymbol{W}_{s,l}\right|_\infty \leq x\right\} \geq 1 - \alpha\right\}.$$

Thus, we reject the null hypothesis $\boldsymbol{H}_0$ whenever $T_{smax} \geq \hat{c}_{s,\alpha}$.

2.2.2 Studentized statistic-based test with screening

Next we describe the screening-based testing procedure under studentized statistic. Based on the screening procedure in Sect. 2.1.2, we obtain the hypothesis testing problem 4, and next we perform the hypothesis testing with the new statistic. We give the test statistic

$$T^s_{smax} = \max_{l \notin \hat{S}} \left\{ \frac{\sqrt{n_1 n_2}|\bar{X}_{1l} - \bar{X}_{2l}|}{\left(n_2 \hat{\sigma}^2_{1l} + n_1 \hat{\sigma}^2_{2l}\right)^{1/2}}, \frac{\sqrt{n_1 n_3}|\bar{X}_{1l} - \bar{X}_{3l}|}{\left(n_3 \hat{\sigma}^2_{1l} + n_1 \hat{\sigma}^2_{3l}\right)^{1/2}}, \ldots, \frac{\sqrt{n_{K-1} n_k}|\bar{X}_{(K-1)l} - \bar{X}_{Kl}|}{\left(n_{K-1} \hat{\sigma}^2_{Kl} + n_K \hat{\sigma}^2_{(K-1)l}\right)^{1/2}} \right\}.$$

The critical value can be denoted as follows:

$$c^s_{s,\alpha} = \inf \left\{ x \in \mathbb{R} : P\left(\max_{l \notin \hat{S}_{all}} |W_{sl}| \geq x | \mathcal{X}_1, \ldots, \mathcal{X}_K \right) \leq \alpha \right\},$$

where $\boldsymbol{S}$, $\hat{\boldsymbol{S}}_{all}$ and $\boldsymbol{W}_s$ have been defined previously. We reject the null hypothesis $\boldsymbol{H}_{0S}$ whenever $T^s_{smax} \geq c^s_{s,\alpha}$. If $Card(\hat{\boldsymbol{S}}_{all}) = p$, we accept $\boldsymbol{H}_{0S}$.

In Algorithm 1, we let $\hat{\boldsymbol{\Phi}}_{ns,\alpha} = \mathbf{1}\left\{T_{nsmax} \geq \hat{c}_{ns,\alpha}\right\}$ and $\hat{\boldsymbol{\Phi}}_{s,\alpha} = \mathbf{1}\{T_{smax} \geq \hat{c}_{s,\alpha}\}$. In non-studentized test, we reject the null hypothesis $\boldsymbol{H}_0$ whenever $\hat{\boldsymbol{\Phi}}_{ns,\alpha} = 1$, and in studentized test, we reject the null hypothesis $\boldsymbol{H}_0$ whenever $\hat{\boldsymbol{\Phi}}_{s,\alpha} = 1$. The detailed algorithms for the bootstrap procedures for $\boldsymbol{\Phi}_{ns,\alpha}$ and $\boldsymbol{\Phi}_{s,\alpha}$ are stated in Algorithm 1.

Algorithm 1 Bootstrap procedures for $\boldsymbol{\Phi}_{ns,\alpha}$ and $\boldsymbol{\Phi}_{s,\alpha}$

Input: $\{\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}\}, \{\boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2n_2}\}, \ldots, \{\boldsymbol{X}_{K1}, \ldots, \boldsymbol{X}_{Kn_K}\}$.
Output: $\hat{\boldsymbol{\Phi}}_{ns,\alpha}$ and $\hat{\boldsymbol{\Phi}}_{s,\alpha}$.
1: ***Procedures***
2: Compute $\tilde{\boldsymbol{R}}_{ns}$ or $\tilde{\boldsymbol{R}}_s$.
3: $T_{nsmax} = \max_{1 \leq l \leq p, 1 \leq i < j \leq K} \left\{ \frac{\sqrt{n_i n_j}|\bar{X}_{il} - \bar{X}_{jl}|}{(n_i + n_j)^{1/2}} \right\}$ or
$T_{smax} = \max_{1 \leq l \leq p, 1 \leq i < j \leq K} \left\{ \frac{\sqrt{n_i n_j}|\bar{X}_{il} - \bar{X}_{jl}|}{\left(n_i \hat{\sigma}^2_{jl} + n_j \hat{\sigma}^2_{il}\right)^{1/2}} \right\}$.
4: $\left\{\boldsymbol{W}_{ns,1}, \ldots, \boldsymbol{W}_{ns,N_{ns}}\right\}$ are independent random vectors from $N\left(\mathbf{0}, \tilde{\boldsymbol{R}}_{ns}\right)$ and $\left\{\boldsymbol{W}_{s,1}, \ldots, \boldsymbol{W}_{s,N_s}\right\}$ are independent random vectors from $N\left(\mathbf{0}, \tilde{\boldsymbol{R}}_s\right)$.
5: $\hat{c}_{ns,\alpha} = \inf\{x \in \mathbb{R} : N_{ns}{}^{-1} \sum_{l=1}^{N_{ns}} \mathbf{1}\left\{|\boldsymbol{W}_{ns,l}|_\infty \leq x\right\} \geq 1 - \alpha\}$ and $\hat{c}_{s,\alpha} = \inf\{x \in \mathbb{R} : N_s{}^{-1} \sum_{l=1}^{N_s} \mathbf{1}\left\{|\boldsymbol{W}_{s,l}|_\infty \leq x\right\} \geq 1 - \alpha\}$.
6: $\hat{\boldsymbol{\Phi}}_{ns,\alpha} = \mathbf{1}\left\{T_{nsmax} \geq \hat{c}_{ns,\alpha}\right\}$ or $\hat{\boldsymbol{\Phi}}_{s,\alpha} = \mathbf{1}\{T_{smax} \geq \hat{c}_{s,\alpha}\}$.

In Algorithm 2, we let $\hat{\boldsymbol{\Phi}}^s_{ns,\alpha} = \mathbf{1}\left\{T^s_{nsmax} \geq \hat{c}^s_{ns,\alpha}\right\}$ and $\hat{\boldsymbol{\Phi}}^s_{s,\alpha} = \mathbf{1}\{T^s_{smax} \geq \hat{c}^s_{s,\alpha}\}$. In non-studentized with screening test, we reject the null hypothesis $\boldsymbol{H}_0$ whenever $\hat{\boldsymbol{\Phi}}^s_{ns,\alpha} = 1$, and in studentized with screening test, we reject the null hypothesis $\boldsymbol{H}_0$

whenever $\hat{\boldsymbol{\Phi}}^{s}_{s,\alpha} = 1$. The detailed algorithms for the bootstrap procedures for $\boldsymbol{\Phi}^{s}_{ns,\alpha}$ and $\boldsymbol{\Phi}^{s}_{s,\alpha}$ are stated in Algorithm 2.

Algorithm 2 Bootstrap procedures for $\boldsymbol{\Phi}^{s}_{ns,\alpha}$ and $\boldsymbol{\Phi}^{s}_{s,\alpha}$

Input: $\{\boldsymbol{X}_{11},\ldots,\boldsymbol{X}_{1n_1}\},\{\boldsymbol{X}_{21},\ldots,\boldsymbol{X}_{2n_2}\},\ldots,\{\boldsymbol{X}_{K1},\ldots,\boldsymbol{X}_{Kn_K}\}$.
Output: $\hat{\boldsymbol{\Phi}}^{s}_{ns,\alpha}$ and $\hat{\boldsymbol{\Phi}}^{s}_{s,\alpha}$.
1: ***Procedures***
2: Use $T^{sc}_{sl} = \max\left\{\frac{\sqrt{n_1n_2}|\bar{X}_{1l}-\bar{X}_{2l}|}{(n_2\hat{\sigma}^2_{1l}+n_1\hat{\sigma}^2_{2l})^{1/2}}, \frac{\sqrt{n_1n_3}|\bar{X}_{1l}-\bar{X}_{3l}|}{(n_3\hat{\sigma}^2_{1l}+n_1\hat{\sigma}^2_{3l})^{1/2}}, \ldots, \frac{\sqrt{n_{K-1}n_K}|\bar{X}_{(K-1)l}-\bar{X}_{Kl}|}{\left(n_K\hat{\sigma}^2_{K-1l}+n_{K-1}\hat{\sigma}^2_{Kl}\right)^{1/2}}\right\}$.
to get $\hat{\boldsymbol{S}}$ and $\boldsymbol{\mu}^s_1,\ldots,\boldsymbol{\mu}^s_K$.
3: Compute $\tilde{\boldsymbol{R}}_{ns}$ or $\tilde{\boldsymbol{R}}_s$.
4: $T^s_{nsmax} = \max_{1\le l\le p, l\notin\hat{\boldsymbol{S}}, 1\le i<j\le K}\left\{\frac{\sqrt{n_in_j}|\bar{X}_{il}-\bar{X}_{jl}|}{(n_i+n_j)^{1/2}}\right\}$,
$T^s_{smax} = \max_{1\le l\le p, l\notin\hat{\boldsymbol{S}}, 1\le i<j\le K}\left\{\frac{\sqrt{n_in_j}|\bar{X}_{il}-\bar{X}_{jl}|}{\left(n_i\hat{\sigma}^2_{jl}+n_j\hat{\sigma}^2_{il}\right)^{1/2}}\right\}$.
5: $\{\boldsymbol{W}_{ns,1},\ldots,\boldsymbol{W}_{ns,N_{ns}}\}$ are independent random vectors from $N\left(\boldsymbol{0},\tilde{\boldsymbol{R}}_{ns}\right)$ and $\{\boldsymbol{W}_{s,1},\ldots,\boldsymbol{W}_{s,N_s}\}$ are independent random vectors from $N\left(\boldsymbol{0},\tilde{\boldsymbol{R}}_s\right)$.
6: $\hat{c}^s_{ns,\alpha} = \inf\{x\in\mathbb{R} : P(N^{-1}_{ns}\sum_{l=1}^{N_{ns}}\max_{l'\notin\hat{\boldsymbol{S}}_{all}}|W_{ns,ll'}|\ge x|\mathcal{X}_1,\ldots,\mathcal{X}_K)\le\alpha\}$ and $\hat{c}^s_{s,\alpha} = \inf\{x\in\mathbb{R} : P(N^{-1}_{s}\sum_{l=1}^{N_s}\max_{l'\notin\hat{\boldsymbol{S}}_{all}}|W_{s,ll'}|\ge x|\mathcal{X}_1,\ldots,\mathcal{X}_K)\le\alpha\}$.
7: $\hat{\boldsymbol{\Phi}}^s_{ns,\alpha} = \boldsymbol{1}\left\{T^s_{nsmax}\ge\hat{c}^s_{ns,\alpha}\right\}$ or $\hat{\boldsymbol{\Phi}}^s_{s,\alpha} = \boldsymbol{1}\left\{T^s_{smax}\ge\hat{c}^s_{s,\alpha}\right\}$.

3 Theoretical properties

Compared with many previous methods, the proposed test procedures do not need to make any structural assumptions on the covariance matrix and correlation matrix and do not require the data structure and the limit distributions of statistics. Consequently, the proposed test methods are broadly applicable. As covariance matrices and correlation matrices are unknown, their estimators are required. It is only necessary to ensure the validity of the following equations: $|\boldsymbol{R}_{ns}-\tilde{\boldsymbol{R}}_{ns}|_\infty = o_p(1)$ and $|\boldsymbol{R}_s-\tilde{\boldsymbol{R}}_s|_\infty = o_p(1)$ hold, where

$$\boldsymbol{R}_{ns} = \begin{bmatrix} \frac{(n_2\boldsymbol{\Sigma}_1+n_1\boldsymbol{\Sigma}_2)}{N_{12}} & \frac{\sqrt{n_2n_3}\boldsymbol{\Sigma}_1}{N_{12}^{1/2}N_{13}^{1/2}} & \frac{\sqrt{n_2n_4}\boldsymbol{\Sigma}_1}{N_{12}^{1/2}N_{14}^{1/2}} & \\ \frac{\sqrt{n_2n_3}\boldsymbol{\Sigma}_1}{N_{12}^{1/2}N_{13}^{1/2}} & \frac{(n_3\boldsymbol{\Sigma}_1+n_1\boldsymbol{\Sigma}_3)}{N_{13}} & \frac{\sqrt{n_3n_4}\boldsymbol{\Sigma}_1}{N_{13}^{1/2}N_{14}^{1/2}} & \cdots \\ \frac{\sqrt{n_2n_4}\boldsymbol{\Sigma}_1}{N_{12}^{1/2}N_{14}^{1/2}} & \frac{\sqrt{n_3n_4}\boldsymbol{\Sigma}_1}{N_{13}^{1/2}N_{14}^{1/2}} & \frac{(n_4\boldsymbol{\Sigma}_1+n_1\boldsymbol{\Sigma}_4)}{N_{14}} & \\ & \vdots & & \ddots \end{bmatrix},$$

and

$$\boldsymbol{R}_s = \begin{bmatrix} \boldsymbol{D}_{12}^{-\frac{1}{2}}(n_2\boldsymbol{\Sigma}_1 + n_1\boldsymbol{\Sigma}_2)\boldsymbol{D}_{12}^{-\frac{1}{2}} & \boldsymbol{D}_{12}^{-\frac{1}{2}}\sqrt{n_2n_3}\boldsymbol{\Sigma}_1\boldsymbol{D}_{13}^{-\frac{1}{2}} & \boldsymbol{D}_{12}^{-\frac{1}{2}}\sqrt{n_2n_4}\boldsymbol{\Sigma}_1\boldsymbol{D}_{14}^{-\frac{1}{2}} & \cdots \\ \boldsymbol{D}_{13}^{-\frac{1}{2}}\sqrt{n_2n_3}\boldsymbol{\Sigma}_1\boldsymbol{D}_{12}^{-\frac{1}{2}} & \boldsymbol{D}_{13}^{-\frac{1}{2}}(n_3\boldsymbol{\Sigma}_1 + n_1\boldsymbol{\Sigma}_3)\boldsymbol{D}_{13}^{-\frac{1}{2}} & \boldsymbol{D}_{13}^{-\frac{1}{2}}\sqrt{n_3n_4}\boldsymbol{\Sigma}_1\boldsymbol{D}_{14}^{-\frac{1}{2}} & \cdots \\ \boldsymbol{D}_{14}^{-\frac{1}{2}}\sqrt{n_2n_4}\boldsymbol{\Sigma}_1\boldsymbol{D}_{12}^{-\frac{1}{2}} & \boldsymbol{D}_{14}^{-\frac{1}{2}}\sqrt{n_3n_4}\boldsymbol{\Sigma}_1\boldsymbol{D}_{13}^{-\frac{1}{2}} & \boldsymbol{D}_{14}^{-\frac{1}{2}}(n_1\boldsymbol{\Sigma}_4 + n_4\boldsymbol{\Sigma}_1)\boldsymbol{D}_{14}^{-\frac{1}{2}} & \\ & \vdots & & \ddots \end{bmatrix}.$$

In this section, we give the non-asymptotic bounds on the distributional errors between the test statistics T_{smax} and T_{nsmax} and their Gaussian analog under the null hypothesis, which can ensure the test procedures introduced in the previous section work. Let

$$\hat{\boldsymbol{\Sigma}}_i = n_i^{-1} \sum_{j=1}^{n_i} \left(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i\right)\left(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i\right)', i = 1, \dots, K.$$

From Lemma 2, the following equations $|\boldsymbol{R}_{ns} - \hat{\boldsymbol{R}}_{ns}|_\infty = o_p(1)$ and $|\boldsymbol{R}_s - \hat{\boldsymbol{R}}_s|_\infty = o_p(1)$ hold. As a result, we may conduct the test directly using sample covariance matrices. The theoretical results of the article are built on the basis that the following two conditions hold. $\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i} (i = 1, \dots, K)$ are independent and identically distributed p-dimensional random vectors and $E(\boldsymbol{X}_{ij}) = \boldsymbol{\mu}_i$, $cov(\boldsymbol{X}_{ij}) = \boldsymbol{\Sigma}_i$, let $\boldsymbol{D}_i = diag(\boldsymbol{\Sigma}_i)$, define the following p-dimensional random vectors $\boldsymbol{U}_{i1}$

$$\boldsymbol{U}_{i1} = \boldsymbol{U}_{i1}\left(\boldsymbol{X}_{i1}\right) = \left(U_{i11}, \dots, U_{i1p}\right)' = \boldsymbol{D}_i^{-\frac{1}{2}}\boldsymbol{X}_{i1}, i = 1, \dots, K.$$

Furthermore, $\boldsymbol{R}_i$ is the correlation matrix of $\boldsymbol{X}_{i1}$ and $\hat{\boldsymbol{R}}_i$ is the sample correlation matrix of $\boldsymbol{X}_{i1}$. The conditions (A.1) and (A.2) will be used in the following results.

(A.1) $\max_{1\le i\le K} \max_{1\le l\le p} \left\{E(\left|U_{i1l}\right|^r)\right\}^{\frac{1}{r}} \le C_1$ for some $r \ge 4$ and $C_1 > 0$.

(A.2) $\max_{1\le i\le K} \max_{1\le l\le p} E(\exp(C_2\left|U_{i1l}\right|^\gamma)) \le C_3$ for some $C_2 > 0,\ C_3 > 1$ and $0 < \gamma \le 2$.

Proposition 1 *Assume that $\tilde{\boldsymbol{\Sigma}}_1, \dots, \tilde{\boldsymbol{\Sigma}}_K$ are any covariance matrix estimators such that $diag\left(\tilde{\boldsymbol{\Sigma}}_1\right) = diag\left(\hat{\boldsymbol{\Sigma}}_1\right), \dots, diag\left(\tilde{\boldsymbol{\Sigma}}_K\right) = diag\left(\hat{\boldsymbol{\Sigma}}_K\right)$. (i) If (A.1) holds, then under $\boldsymbol{H}_0$,*

$$\begin{aligned} &\sup_{x\ge 0} |P\left(|\boldsymbol{T}_v|_\infty > x\right) - P\left(|\boldsymbol{W}_v|_\infty > x|\mathcal{X}_1, \dots, \mathcal{X}_K\right)| \\ &\le CN^{-\frac{1}{8}}\{\log(K(K-1)pN/2)\}^{\frac{7}{8}} + C\theta_{N,p}^{\frac{1}{(r+1)}}\{\log(K(K-1)pN/2)\}^{\frac{3}{2}} \\ &+ \Theta_v^{\frac{1}{3}}\{\log(K(K-1)pN/2)\}^{\frac{3}{2}} + C\left(pn^{1-\frac{r}{2}}\right)^{\frac{2}{(r+2)}}\{\log(K(K-1)pN/2)\}^{\frac{3}{2}}; \end{aligned}$$

(ii) if (A.2) holds, then under $\boldsymbol{H}_0$,

$$\sup_{x\geq 0} |P(|\boldsymbol{T}_\nu|_\infty > x) - P(|\boldsymbol{W}_\nu|_\infty > x|\mathcal{X}_1,\dots,\mathcal{X}_K)|$$
$$\leq Cn^{-\frac{1}{8}}\{\log(K(K-1)pN/2)\}^{\frac{7}{8}} + Cn^{-\frac{1}{2}}\{\log(K(K-1)pN/2)\}^{\frac{3}{2}+\frac{1}{\gamma}}$$
$$+ C\Theta_\nu^{\frac{1}{3}}\{\log(K(K-1)pN/2)\}^{\frac{3}{2}} + n^{-1}\{\log(K(K-1)pN/2)\}^{\frac{2}{\gamma}+1};$$

where $N = n_1 + \cdots + n_K, \theta_{N,p} = pN^{1-\frac{r}{2}}, n = \min\left\{n_1,\dots,n_K\right\}, \Theta_\nu = |\tilde{\boldsymbol{R}}_\nu - \boldsymbol{R}_\nu|_\infty, \nu \in \{s, ns\}$.

Remark 1 It is obvious that $\left(\tilde{\boldsymbol{\Sigma}}_1,\dots,\tilde{\boldsymbol{\Sigma}}_K\right) = \left(\hat{\boldsymbol{\Sigma}}_1,\dots,\hat{\boldsymbol{\Sigma}}_K\right)$ satisfies the condition for the covariance matrices in the proposition.

Next, we state the asymptotic results for the proposed four test statistics.

Theorem 1 *Let* $\tilde{\boldsymbol{\Sigma}}_1 = \hat{\boldsymbol{\Sigma}}_1,\dots,\tilde{\boldsymbol{\Sigma}}_K = \hat{\boldsymbol{\Sigma}}_K$ *represent the sample covariance matrices. Assume that*: (i) (A.1) *holds, and* $p = O\left(n^{\frac{r}{2}-1}N^{-\delta}\right)$ *for some* $\delta > 0$; (ii) (A.2) *holds for some* $\gamma \geq \frac{1}{2}$, *and* $\log(pN) = o\left(n^{\frac{1}{7}}\right)$; *then, under either* (i) *or* (ii), *we have*

$$\lim_{n_1,\dots,n_K,p\to\infty} P_{\boldsymbol{H}_0}\left(T_{\nu max} > c_{\nu,\alpha}\right) = \alpha,$$

where $\nu \in \{ns, s\}$, $N = n_1 + \cdots + n_K$, *and* $n = \min\{n_1,\dots,n_K\}$.

Theorem 2 *Let* $\tilde{\boldsymbol{\Sigma}}_1 = \hat{\boldsymbol{\Sigma}}_1,\dots,\tilde{\boldsymbol{\Sigma}}_K = \hat{\boldsymbol{\Sigma}}_K$ *denote the sample covariance matrices, and consider the sequence* $\xi_n, n \geq 1, \xi_n > 0$. *As* $n \to \infty$, $\xi_n \to 0$, *and* $\xi_n\sqrt{\log(p)} \to \infty$. *Assume that either* (A.1) *holds with* $p = O\left(n^{\frac{r}{2}-1}N^{-\delta}\right)$ *or* (A.2) *holds with* $\log(pN) = o\left(n^{\frac{\gamma}{3\gamma+2}}\right)$. (i) *If* $\frac{\max_{1\leq l\leq p}|\mu_{il}-\mu_{jl}|}{\max_{1\leq l\leq p}\left(\sigma_{il}^2/n_i+\sigma_{jl}^2/n_j\right)^{1/2}} \geq \left(1+\xi_n\right)\eta(p,\alpha)$ *for* $i,j \in (1,\dots,K)$, *then*

$$\lim_{n_1,\dots,n_K,p\to\infty} P_{\boldsymbol{H}_1}\left(T_{nsmax} > c_{ns,\alpha}\right) = 1.$$

And (ii) *if* $\max_{1\leq l\leq p} \frac{|\mu_{il}-\mu_{jl}|}{\left(\sigma_{il}^2/n_i+\sigma_{jl}^2/n_j\right)^{\frac{1}{2}}} \geq \left(1+\xi_n\right)\eta(p,\alpha)$ *for* $i,j \in (1,\dots,K)$, *then*

$$\lim_{n_1,\dots,n_K,p\to\infty} P_{\boldsymbol{H}_1}\left(T_{smax} > c_{s,\alpha}\right) = 1.$$

Where $N = n_1 + \cdots + n_K, n = \min\{n_1,\dots,n_K\}$, $\eta(p,\alpha) = \sqrt{2\log(3p)} + \sqrt{2\log(1/\alpha)}$ *for* $0 \leq \alpha \leq 1$.

Theorem 3 *Consider* $\tilde{\boldsymbol{\Sigma}}_1 = \hat{\boldsymbol{\Sigma}}_1,\dots,\tilde{\boldsymbol{\Sigma}}_K = \hat{\boldsymbol{\Sigma}}_K$ *as the sample covariance matrices. Assume that*: (i) (A.1) *holds with* $p = O\left(n^{\frac{r}{2}-1}N^{-\delta}\right)$; (ii) (A.2) *holds with* $\gamma \geq \frac{1}{2}$ *and* $\log(pN) = o\left(n^{\frac{1}{7}}\right)$; *then, under either* (i) *or* (ii), *we have*

$$\limsup_{n_1,\dots,n_K,p\to\infty} P_{\boldsymbol{H}_0}\left(T^s_{\nu\max} > c^s_{\nu,\alpha}\right) \le \alpha + o_p(1),$$

where $\nu \in \{ns, s\}$, $N = n_1 + \cdots + n_K$, *and* $n = \min\{n_1, \dots, n_K\}$.

Theorem 4 *Let* $\tilde{\boldsymbol{\Sigma}}_1 = \hat{\boldsymbol{\Sigma}}_1, \dots, \tilde{\boldsymbol{\Sigma}}_K = \hat{\boldsymbol{\Sigma}}_K$ *represent the sample covariance matrices, and consider the sequence* $\xi_n, n \ge 1$, $\xi_n > 0$. *As* $n \to \infty$, $\xi_n \to 0$, *and* $\xi_n\sqrt{\log(p)} \to \infty$. *Assume that* (A.1) *holds with* $p = O(n^{\frac{r}{2}-1}N^{-\delta}), \delta > 0$ *or* (A.2) *holds with* $\log(pN) = o(n^{\frac{\gamma}{3\gamma+2}})$. *If* $\frac{\max_{1\le l\le p}|\mu_{il}-\mu_{jl}|}{\max_{1\le l\le p}(\sigma^2_{il}/n_i+\sigma^2_{jl}/n_j)^{\frac{1}{2}}} \ge (1+\xi_n)\eta(p,\alpha)$ *for all* $i,j \in (1,\dots,K)$, *then we have*

$$\lim_{n_1,\dots,n_K,p\to\infty} P_{\boldsymbol{H}_1}\left(T^s_{nsmax} > c^s_{ns,\alpha}\right) = 1;$$

if $\max_{1\le l\le p} \frac{|\mu_{il}-\mu_{jl}|}{\left(\sigma^2_{il}/n_i+\sigma^2_{jl}/n_j\right)^{\frac{1}{2}}} \ge \left(1+\xi_n\right)\eta(p,\alpha)$ *for all* $i,j \in (1,\dots,K)$, *then we have*

$$\lim_{n_1,\dots,n_K,p\to\infty} P_{\boldsymbol{H}_1}\left(T^s_{smax} > c^s_{s,\alpha}\right) = 1.$$

Where $N = n_1 + \cdots + n_K, n = \min\left\{n_1, \dots, n_K\right\}$, $\eta(p,\alpha) = \sqrt{2\log(3p)} + \sqrt{2\log(1/\alpha)}$ *for* $0 \le \alpha \le 1$.

4 Simulation studies

In this section, we present the findings derived from the multi-sample hypothesis testing simulations. For the sake of simplicity, we only present the results of the three-sample case. The simulation results include the following four methods: non-studentized test without screening $\boldsymbol{\Phi}_{ns,\alpha}$, non-studentized test with screening $\boldsymbol{\Phi}^s_{ns,\alpha}$, studentized test without screening $\boldsymbol{\Phi}_{s,\alpha}$, studentized test with screening $\boldsymbol{\Phi}^s_{s,\alpha}$. To show the performance of the proposed tests, we compare them with the method proposed by Hu et al. (2017), denoted by $\boldsymbol{T}^{umvue}$. We generate samples based on different sample sizes and dimensions. The empirical sizes and powers are based on 1, 000 replications. To produce samples and compute the critical values for the suggested test, we use the sample covariance matrix and the Monte Carlo approach based on $M = 1{,}000$. The four models are of the following forms.

Model I: Let $\boldsymbol{X}_{ij} = \boldsymbol{\Sigma}_i\boldsymbol{Z}_{ij} + \boldsymbol{\mu}_i$, $i \in \{1,2,3\}, j = 1,\dots,n_i$, where $\boldsymbol{\Sigma}_i = \boldsymbol{I}_p$, and $\boldsymbol{Z}_{ij} = (Z_{ij1,\dots,Z_{ijp}})'$, $Z_{ijk}{'s}$ are independent random variables which are distributed as distribution: $(\chi^2_{50} - 50)/10$.

Model II: Let $\boldsymbol{X}_{ij} \overset{i.i.d.}{\sim} \boldsymbol{t}_{\omega_i}\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)$, $i \in \{1,2,3\}$, $j = 1,\dots,n_i$, $\boldsymbol{\Sigma}_i = \boldsymbol{D}_i^{1/2}\boldsymbol{H}_i\boldsymbol{D}_i^{1/2}$, where $\boldsymbol{D}_i = Diag[D_{i1}, \dots, D_{ip}]$, for $1 \le l \le p$, $D_{1\,l} \sim U(0.75, 0.85)$, $D_{2\,l} \sim U(0.75, 0.8)$, $D_{3\,l} \sim U(0.65, 0.75)$, $\boldsymbol{H}_i = (h_{i,kl})_{1\le k,l\le p}$, where $\omega_1 = 4$, $\omega_2 = 7$, $\omega_3 = 8$, and $h_{1,kl} = 0.8^{|k-l|}, h_{2,kl} = 0.75^{|k-l|}, h_{3,kl} = 0.02^{|k-l|}$.

Table 1 Empirical sizes of $\boldsymbol{\Phi}_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^s_{ns,\alpha}$, $\boldsymbol{\Phi}^s_{s,\alpha}$, and $\boldsymbol{T}^{umvue}$ at 5% nominal significance are considered when $n_1 = n_2 = n_3 = 40, p = 100, 200, 500$; $n_1 = n_2 = n_3 = 80, p = 100, 200, 500$; $n_1 = n_2 = n_3 = 120, p = 100, 200, 500$; $n_1 = n_2 = n_3 = 240, p = 100, 200, 500$

$n_1 = n_2 = n_3$	p	Model I					Model II				
		$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^s_{ns,\alpha}$	$\boldsymbol{\Phi}^s_{s,\alpha}$	$\boldsymbol{T}^{umvue}$	$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^s_{ns,\alpha}$	$\boldsymbol{\Phi}^s_{s,\alpha}$	$\boldsymbol{T}^{umvue}$
40	100	0.039	0.092	0.106	0.099	0.080	0.030	0.072	0.090	0.075	0.070
	200	0.039	0.093	0.113	0.097	0.070	0.035	0.066	0.077	0.076	0.050
	500	0.031	0.089	0.096	0.106	0.020	0.059	0.092	0.138	0.099	0.029
80	100	0.041	0.064	0.079	0.068	0.070	0.038	0.060	0.078	0.060	0.040
	200	0.041	0.058	0.078	0.076	0.050	0.052	0.053	0.084	0.053	0.015
	500	0.036	0.077	0.080	0.077	0.060	0.055	0.056	0.096	0.061	0.020
120	100	0.043	0.055	0.069	0.067	0.050	0.042	0.051	0.075	0.058	0.040
	200	0.046	0.060	0.069	0.062	0.040	0.042	0.056	0.076	0.063	0.030
	500	0.041	0.068	0.078	0.070	0.070	0.034	0.048	0.076	0.054	0.033
240	100	0.04	0.051	0.06	0.056	0.033	0.042	0.050	0.062	0.052	0.040
	200	0.045	0.065	0.066	0.065	0.063	0.053	0.042	0.072	0.049	0.047
	500	0.047	0.059	0.061	0.059	0.043	0.043	0.046	0.063	0.046	0.027

Table 2 Empirical sizes of $\boldsymbol{\Phi}_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^s_{ns,\alpha}$, $\boldsymbol{\Phi}^s_{s,\alpha}$, and $\boldsymbol{T}^{umvue}$ at 5% nominal significance are considered when $n_1 = n_2 = n_3 = 40, p = 100, 200, 500$; $n_1 = n_2 = n_3 = 80, p = 100, 200, 500$; $n_1 = n_2 = n_3 = 120, p = 100, 200, 500$; $n_1 = n_2 = n_3 = 240, p = 100, 200, 500$

$n_1 = n_2 = n_3$	p	Model III					Model IV				
		$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^s_{ns,\alpha}$	$\boldsymbol{\Phi}^s_{s,\alpha}$	$\boldsymbol{T}^{umvue}$	$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^s_{ns,\alpha}$	$\boldsymbol{\Phi}^s_{s,\alpha}$	$\boldsymbol{T}^{umvue}$
40	100	0.035	0.067	0.091	0.074	0.040	0.043	0.090	0.093	0.090	0.060
	200	0.046	0.055	0.106	0.067	0.090	0.050	0.098	0.099	0.110	0.030
	500	0.052	0.084	0.124	0.088	0.060	0.034	0.098	0.100	0.104	0.110
80	100	0.040	0.053	0.079	0.053	0.100	0.035	0.055	0.073	0.070	0.050
	200	0.040	0.059	0.086	0.067	0.060	0.040	0.073	0.080	0.079	0.040
	500	0.047	0.059	0.083	0.059	0.063	0.038	0.062	0.072	0.070	0.040
120	100	0.052	0.043	0.084	0.053	0.065	0.048	0.055	0.070	0.064	0.080
	200	0.043	0.071	0.073	0.071	0.075	0.043	0.063	0.072	0.069	0.070
	500	0.041	0.052	0.073	0.054	0.075	0.046	0.054	0.072	0.069	0.072
240	100	0.044	0.051	0.076	0.052	0.090	0.048	0.047	0.060	0.056	0.040
	200	0.046	0.055	0.069	0.055	0.080	0.045	0.043	0.060	0.053	0.053
	500	0.041	0.054	0.070	0.055	0.060	0.046	0.056	0.068	0.066	0.067

Model III: Let $\boldsymbol{X}_{ij} \overset{i.i.d.}{\sim} \boldsymbol{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i \in \{1, 2, 3\}, j = 1, \dots, n_i$, where $\boldsymbol{\Sigma}_i = \boldsymbol{W}_i \boldsymbol{\Psi}_i \boldsymbol{W}_i$, $\boldsymbol{W}_i = Diag[W_{i1}, \dots, W_{ip}]$, $W_{il} = 0.02 \times i + (p - l + i)/p$, $\boldsymbol{\Psi}_i = (\phi^i_{kl})$, $\phi^i_{kk} = 1, \phi^i_{kl} = (-1)^{k+l}(0.2 \times i)^{|k-l|^{0.1}}, k \neq l$.

Table 3 Empirical powers of model I and model II about $\boldsymbol{\Phi}_{ns,\alpha}$, $\boldsymbol{\Phi}^{s}_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^{s}_{s,\alpha}$, $\boldsymbol{T}^{umvue}$ when $n_1 = n_2 = n_3 = 40, p = 500$

β	Q	Model I					Model II				
		$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^{s}_{ns,\alpha}$	$\boldsymbol{\Phi}^{s}_{s,\alpha}$	$\boldsymbol{T}^{umvue}$	$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^{s}_{ns,\alpha}$	$\boldsymbol{\Phi}^{s}_{s,\alpha}$	$\boldsymbol{T}^{umvue}$
0.1	2	0.180	0.413	0.470	0.430	0.176	0.2	0.264	0.480	0.510	0.040
	4	0.203	0.557	0.593	0.581	0.133	0.412	0.602	0.634	0.624	0.060
	6	0.327	0.820	0.853	0.847	0.643	0.596	0.752	0.770	0.624	0.070
	8	0.353	0.843	0.850	0.860	0.847	0.704	0.816	0.864	0.850	0.110
	10	0.440	0.903	0.937	0.918	0.903	0.822	0.916	0.928	0.922	0.170
	12	0.637	0.977	0.997	0.997	0.987	0.832	0.918	0.934	0.925	0.250
	16	0.577	0.952	0.983	0.980	0.947	0.864	0.960	0.972	0.970	0.460
	20	0.787	1	0.997	1	0.994	0.956	0.986	0.996	0.990	0.540
	40	0.863	1	0.997	1	1	1	1	1	1	0.940
	80	0.997	1	1	1	1	0.996	1	1	1	1
0.15	2	0.307	0.856	0.843	0.870	0.310	0.562	0.742	0.772	0.750	0.042
	4	0.500	0.876	0.903	0.896	0.590	0.834	0.928	0.942	0.936	0.060
	6	0.667	0.997	0.997	0.997	0.890	0.876	0.978	0.98	0.982	0.150
	8	0.643	0.989	0.990	0.992	0.900	0.950	0.996	0.996	0.998	0.240
	10	0.860	1	1	1	1	0.972	0.998	0.996	0.996	0.350
	12	0.853	1	1	1	1	0.984	0.998	1	1	0.560
	16	0.906	1	1	1	1	1	1	1	1	0.780
	20	1	1	1	1	1	1	1	1	1	0.800
	40	1	1	1	1	1	1	1	1	1	0.990
	80	1	1	1	1	1	1	1	1	1	1
0.2	2	0.487	0.863	0.893	0.894	0.360	0.804	0.916	0.918	0.918	0.052
	4	0.687	1	1	1	0.910	0.996	0.992	0.998	0.994	0.070
	6	0.833	1	1	1	0.980	0.984	0.996	0.999	0.998	0.310
	8	0.843	0.956	1	0.967	1	1	1	1	1	0.510
	10	0.903	1	1	1	1	1	1	1	1	0.680
	12	0.953	1	1	1	1	1	1	1	1	0.710
	16	0.993	1	1	1	1	1	1	1	1	0.903
	20	0.997	1	1	1	1	1	1	1	1	0.910
	40	1	1	1	1	1	1	1	1	1	1
	80	1	1	1	1	1	1	1	1	1	1

Model IV: Let $\boldsymbol{X}_{ij} \overset{i.i.d.}{\sim} N\big(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\big)$, $i \in \{1, 2, 3\}$, $j = 1, \dots, n_i$ and $\boldsymbol{\Sigma}_i = \boldsymbol{D}_i^{1/2}\boldsymbol{H}_i\boldsymbol{D}_i^{1/2}$, $\boldsymbol{D}_i = Diag[D_{i1}, \dots, D_{ip}]$, for $1 \le l \le p$, $D_{il} \sim U(0.95, 1)$, $\boldsymbol{H}_i = (h_{i,kl})_{1\le k,l\le p}$, $h_{1,kl} = 0.5^{|k-l|}, h_{2,kl} = 0.6^{|k-l|}, h_{3,kl} = 0.6^{|k-l|}$.

Under the null hypothesis, we assume that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{0}$. Under the alternative hypothesis, we assume that the nonzero components of $\mu_i, i = 1, 2, 3$ are $\mu_{1\,l} = \left[\beta\hat{\sigma}_{1,ll}\log(p)\left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}\right)\right]^{\frac{1}{2}}$, $\mu_{2\,l} = \left[4\beta\hat{\sigma}_{2,ll}\log(p)\left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}\right)\right]^{\frac{1}{2}}$,

Table 4 Empirical powers of model III and model IV about $\boldsymbol{\Phi}_{ns,\alpha}$, $\boldsymbol{\Phi}^{s}_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^{s}_{s,\alpha}$, $\boldsymbol{T}^{umvue}$ when $n_1 = n_2 = n_3 = 40, p = 500$

β	Q	Model III					Model IV				
		$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^{s}_{ns,\alpha}$	$\boldsymbol{\Phi}^{s}_{s,\alpha}$	$\boldsymbol{T}^{umvue}$	$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^{s}_{ns,\alpha}$	$\boldsymbol{\Phi}^{s}_{s,\alpha}$	$\boldsymbol{T}^{umvue}$
0.1	2	0.230	0.432	0.447	0.440	0.093	0.227	0.312	0.410	0.353	0.107
	4	0.413	0.683	0.703	0.690	0.110	0.32	0.600	0.660	0.603	0.253
	6	0.530	0.767	0.820	0.810	0.160	0.397	0.612	0.633	0.630	0.437
	8	0.653	0.860	0.867	0.860	0.190	0.700	0.850	0.880	0.870	0.677
	10	0.783	0.883	0.930	0.920	0.213	0.733	0.910	0.923	0.917	0.807
	12	0.820	0.936	0.960	0.947	0.260	0.547	0.917	0.937	0.937	0.883
	16	0.870	0.977	0.977	0.977	0.402	0.867	0.927	0.970	0.967	0.677
	20	0.927	0.990	0.993	0.993	0.510	0.810	0.962	0.987	0.980	0.997
	40	0.993	1	1	1	0.970	0.983	1	1	1	1
	80	1	1	1	1	1	1	1	1	1	1
0.15	2	0.547	0.742	0.767	0.760	0.110	0.480	0.707	0.717	0.720	0.203
	4	0.710	0.907	0.923	0.923	0.123	0.650	0.887	0.867	0.896	0.457
	6	0.883	0.967	0.980	0.977	0.180	0.848	0.873	0.923	0.892	0.763
	8	0.927	0.990	0.990	0.990	0.260	0.953	0.966	0.983	0.980	0.907
	10	0.980	0.997	0.997	0.997	0.370	0.923	1	1	1	0.980
	12	0.987	1	1	1	0.460	0.986	0.992	0.997	0.997	0.997
	16	1	1	1	1	0.683	0.997	0.993	1	0.993	1
	20	1	1	1	1	0.890	0.993	1	1	1	1
	40	1	1	1	1	1	1	1	1	1	1
	80	1	1	1	1	1	1	1	1	1	1
0.2	2	0.836	0.908	0.923	0.920	0.110	0.470	0.873	0.893	0.883	0.267
	4	0.953	0.973	0.997	0.993	0.207	0.836	0.965	0.973	0.97	0.630
	6	0.987	0.990	0.997	0.997	0.317	0.86	1	0.997	1	0.923
	8	0.993	1	1	1	0.407	0.933	1	1	1	0.987
	10	0.997	1	1	1	0.520	0.993	1	1	1	0.997
	12	1	1	1	1	0.680	1	1	1	1	0.997
	16	1	1	1	1	0.900	0.997	1	1	1	1
	20	1	1	1	1	0.963	1	1	1	1	1
	40	1	1	1	1	1	1	1	1	1	1
	80	1	1	1	1	1	1	1	1	1	1

$\mu_{3l} = \left[8\beta\hat{\sigma}_{3,ll}\log(p)\left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}\right)\right]^{\frac{1}{2}}$. Where $\hat{\sigma}_{1,ll}$, $\hat{\sigma}_{2,ll}$,$\hat{\sigma}_{3,ll}$ are the l-th diagonal elements of $\hat{\boldsymbol{\Sigma}}_1$, $\hat{\boldsymbol{\Sigma}}_2$ and $\hat{\boldsymbol{\Sigma}}_3$, respectively. Both $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ have Q nonzero terms that are sampled from $\{1, \ldots, p\}$, respectively. Simulation results on the test $\boldsymbol{\Phi}_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^{s}_{ns,\alpha}$, $\boldsymbol{\Phi}^{s}_{s,\alpha}$, $\boldsymbol{T}^{umvue}$ are summarized in Tables 1, 2, 3, 4, 5 and 6.

The empirical sizes of the four models are summarized in Tables 1 and 2. The empirical sizes are reasonably close to the nominal level of 0.05 for all the tests. Especially for $\boldsymbol{\Phi}_{ns,\alpha}$, even when the sample size is small, the empirical sizes are

Table 5 Empirical powers of model I and model II about $\boldsymbol{\Phi}_{ns,\alpha}$, $\boldsymbol{\Phi}^{s}_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^{s}_{s,\alpha}$, $\boldsymbol{T}^{umvue}$ when $n_1 = n_2 = n_3 = 80, p = 500$

β	Q	Model I					Model II				
		$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^{s}_{ns,\alpha}$	$\boldsymbol{\Phi}^{s}_{s,\alpha}$	$\boldsymbol{T}^{umvue}$	$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^{s}_{ns,\alpha}$	$\boldsymbol{\Phi}^{s}_{s,\alpha}$	$\boldsymbol{T}^{umvue}$
0.1	2	0.283	0.406	0.417	0.410	0.150	0.170	0.363	0.413	0.383	0.020
	4	0.450	0.573	0.587	0.577	0.360	0.183	0.753	0.733	0.753	0.050
	6	0.613	0.706	0.763	0.747	0.680	0.463	0.763	0.780	0.772	0.140
	8	0.620	0.803	0.820	0.807	0.801	0.383	0.710	0.760	0.733	0.150
	10	0.770	0.866	0.887	0.883	0.860	0.727	0.903	0.937	0.920	0.240
	12	0.777	0.883	0.923	0.918	0.904	0.603	0.897	0.906	0.902	0.350
	16	0.916	0.976	0.980	0.980	0.962	0.727	0.963	0.980	0.973	0.540
	20	0.926	0.970	0.978	0.972	0.963	0.63	0.956	0.970	0.966	0.600
	40	1	1	1	1	1	0.977	1	1	1	0.990
	80	1	1	1	1	1	0.997	1	1	1	0.990
0.15	2	0.530	0.643	0.690	0.680	0.230	0.457	0.743	0.767	0.756	0.050
	4	0.853	0.863	0.913	0.897	0.580	0.578	0.940	0.950	0.950	0.120
	6	0.887	0.943	0.958	0.953	0.890	0.830	0.977	0.980	0.986	0.190
	8	0.967	0.990	0.990	0.990	0.970	0.850	0.983	0.987	0.986	0.290
	10	0.983	1	1	1	1	0.880	0.997	1	1	0.610
	12	0.990	1	1	1	1	0.980	0.993	0.996	0.993	0.630
	16	1	1	1	1	1	0.963	1	1	1	0.880
	20	1	1	1	1	1	0.990	1	1	1	0.920
	40	1	1	1	1	1	1	1	1	1	0.980
	80	1	1	1	1	1	1	1	1	1	1
0.2	2	0.87	0.883	0.910	0.890	0.420	0.803	0.893	0.920	0.896	0.060
	4	0.970	0.980	0.987	0.983	0.820	0.753	0.993	0.993	0.996	0.230
	6	0.990	1	1	1	1	0.960	0.996	1	1	0.370
	8	1	0.997	1	1	1	1	1	1	1	0.600
	10	1	1	1	1	1	0.997	1	1	1	0.740
	12	1	1	1	1	1	1	1	1	1	0.760
	16	1	1	1	1	1	0.993	1	1	1	0.970
	20	1	1	1	1	1	1	1	1	1	1
	40	1	1	1	1	1	1	1	1	1	1
	80	1	1	1	1	1	1	1	1	1	1

still close to the nominal level of 0.05. When the sample size is too small (e.g., $n_1 = n_2 = n_3 = 40$), the tests $\boldsymbol{\Phi}^{s}_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^{s}_{s,\alpha}$ and $\boldsymbol{T}^{umvue}$ show a slightly inflated result. However, as the sample size increases, the proposed test method demonstrates improvement. It is worth noting that the screening-based tests $\boldsymbol{\Phi}_{s,\alpha}$ and $\boldsymbol{\Phi}^{s}_{s,\alpha}$ have higher empirical sizes. From these results, we can also find that the empirical sizes of test $\boldsymbol{T}^{umvue}$ are more affected by the model settings, which makes sense since the test $\boldsymbol{T}^{umvue}$ heavily depends on the structure of the covariance matrix.

Table 6 Empirical powers of model III and model IV about $\boldsymbol{\Phi}_{ns,\alpha}$, $\boldsymbol{\Phi}^s_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^s_{s,\alpha}$, $\boldsymbol{T}^{umvue}$ when $n_1 = n_2 = n_3 = 80, p = 500$

β	Q	Model III					Model IV				
		$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^s_{ns,\alpha}$	$\boldsymbol{\Phi}^s_{s,\alpha}$	$\boldsymbol{T}^{umvue}$	$\boldsymbol{\Phi}_{ns,\alpha}$	$\boldsymbol{\Phi}_{s,\alpha}$	$\boldsymbol{\Phi}^s_{ns,\alpha}$	$\boldsymbol{\Phi}^s_{s,\alpha}$	$\boldsymbol{T}^{umvue}$
0.1	2	0.180	0.420	0.450	0.440	0.080	0.256	0.350	0.370	0.357	0.150
	4	0.270	0.620	0.630	0.630	0.110	0.446	0.546	0.607	0.587	0.290
	6	0.450	0.760	0.810	0.773	0.120	0.613	0.740	0.763	0.753	0.400
	8	0.510	0.876	0.920	0.900	0.170	0.756	0.833	0.863	0.847	0.640
	10	0.500	0.827	0.873	0.830	0.160	0.800	0.892	0.903	0.916	0.700
	12	0.560	0.923	0.940	0.936	0.180	0.843	0.957	0.943	0.968	0.900
	16	0.730	0.996	0.987	0.987	0.330	0.906	0.957	0.973	0.963	0.920
	20	0.717	0.976	0.990	0.987	0.510	0.940	0.980	0.980	0.980	0.930
	40	0.936	1	1	1	0.970	0.993	0.997	0.997	0.997	0.997
	80	0.993	1	1	1	1	1	1	1	1	1
0.15	2	0.277	0.662	0.693	0.687	0.110	0.627	0.690	0.720	0.693	0.200
	4	0.433	0.897	0.903	0.910	0.140	0.853	0.862	0.900	0.874	0.410
	6	0.602	0.963	0.983	0.977	0.180	0.917	0.954	0.967	0.963	0.720
	8	0.750	0.927	0.993	0.997	0.310	0.956	0.970	0.970	0.970	0.820
	10	0.810	0.937	1	0.999	0.420	1	1	1	1	0.900
	12	0.860	0.993	0.993	0.993	0.460	0.997	1	1	1	0.940
	16	0.892	1	1	1	0.610	1	1	1	1	0.980
	20	0.970	1	1	1	0.900	1	1	1	1	1
	40	0.997	1	1	1	1	1	1	1	1	1
	80	1	1	1	1	1	1	1	1	1	1
0.2	2	0.463	0.930	0.930	0.930	0.100	0.883	0.887	0.916	0.888	0.260
	4	0.708	0.962	0.983	0.997	0.270	0.973	1	1	1	0.580
	6	0.580	1	0.983	1	0.290	0.994	1	1	1	0.82
	8	0.743	0.992	0.993	0.998	0.480	1	1	1	1	0.920
	10	0.936	1	1	1	0.620	1	1	1	1	1
	12	0.91	1	1	1	0.640	1	1	1	1	1
	16	0.994	1	1	1	0.890	1	1	1	1	1
	20	0.997	1	1	1	0.980	1	1	1	1	1
	40	1	1	1	1	0.990	1	1	1	1	1
	80	1	1	1	1	1	1	1	1	1	1

We give simulation results for power with sample sizes of 40 and 80. The empirical powers are summarized in Tables 3–6. In model I, the powers of $\boldsymbol{T}^{umvue}$ are significantly lower than $\boldsymbol{\Phi}^s_{ns,\alpha}$, $\boldsymbol{\Phi}_{s,\alpha}$, $\boldsymbol{\Phi}^s_{s,\alpha}$ with sparse signals and small values of $\mu_{il}, i = 1, 2, 3$. The powers of $\boldsymbol{\Phi}_{ns,\alpha}$ are lower than $\boldsymbol{T}^{umvue}$ when $n_1 = n_2 = n_3 = 40$, but the powers of $\boldsymbol{\Phi}_{ns,\alpha}$ are significantly higher than $\boldsymbol{T}^{umvue}$ when $n_1 = n_2 = n_3 = 80$. Thus, the proposed tests significantly outperform $\boldsymbol{T}^{umvue}$ with sparse signals and small values of nonzero items. As the values of the nonzero term increase and the alternative hypothesis becomes less sparse, the powers of all tests tend to 1. Both

Table 7 The values of the proposed test and the estimate of the critical values we compute as following

Tests	The values of the test statistic	Critical value estimators
$\boldsymbol{\Phi}_{ns,\alpha}$	44.0500	0.5900
$\boldsymbol{\Phi}_{s,\alpha}$	711.5348	3.7300
$\boldsymbol{\Phi}^s_{ns,\alpha}$	24.8578	0.5855
$\boldsymbol{\Phi}^s_{s,\alpha}$	711.5348	3.6765

$\boldsymbol{\Phi}^s_{ns,\alpha}$ and $\boldsymbol{\Phi}^s_{s,\alpha}$ are more powerful than $\boldsymbol{\Phi}_{ns,\alpha}$ and $\boldsymbol{\Phi}_{s,\alpha}$, respectively. Moreover, the difference between $\boldsymbol{\Phi}^s_{ns,\alpha}$ and $\boldsymbol{\Phi}_{ns,\alpha}$ is even greater. $\boldsymbol{\Phi}^s_{ns,\alpha}$ performs the best on all of the proposed tests. In model II, when the alternative hypothesis is sparse and the values of $\mu_{il}, i = 1, 2, 3$ are small, the proposed tests significantly outperform $\boldsymbol{T}^{umvue}$. As the values of the nonzero term increase and the alternative hypothesis less sparse, the powers of all tests tend to 1. Both $\boldsymbol{\Phi}^s_{ns,\alpha}$ and $\boldsymbol{\Phi}^s_{s,\alpha}$ demonstrate higher power than $\boldsymbol{\Phi}_{ns,\alpha}$ and $\boldsymbol{\Phi}_{s,\alpha}$, respectively. Moreover, the difference between $\boldsymbol{\Phi}^s_{ns,\alpha}$ and $\boldsymbol{\Phi}_{ns,\alpha}$ is more significant. In model III, the proposed tests significantly outperform $\boldsymbol{T}^{umvue}$ in sparse cases under small values of $\mu_{il}, i = 1, 2, 3$ setting. Both $\boldsymbol{\Phi}^s_{ns,\alpha}$ and $\boldsymbol{\Phi}^s_{s,\alpha}$ exhibit higher power compared to $\boldsymbol{\Phi}_{ns,\alpha}$ and $\boldsymbol{\Phi}_{s,\alpha}$, respectively. Additionally, the difference between $\boldsymbol{\Phi}^s_{ns,\alpha}$ and $\boldsymbol{\Phi}_{ns,\alpha}$ is more pronounced. The powers of all tests incline toward 1 as signal strength rises. In model IV, it is also $\boldsymbol{\Phi}^s_{ns,\alpha}$ and $\boldsymbol{\Phi}^s_{s,\alpha}$ that performed best. In sparse cases under small values of $\mu_{il}, i = 1, 2, 3$ setting, the powers of the proposed test are significantly better than that of $\boldsymbol{T}^{umvue}$. The powers of all tests converge to 1 when the values of the nonzero term increase and the alternative hypothesis is less sparse. In summary, the simulation results indicate that the proposed tests exhibit significantly higher power than $\boldsymbol{T}^{umvue}$ in sparse cases under weak signal settings. Additionally, the screening-based testing procedures uniformly enhance the power performance of the tests and simplify the testing procedures.

5 Real data analysis

In the real data analysis section, we conducted an analysis of the rice data, which were collected from https://www.muratkoklu.com/datasets/. The original dataset consists of 75, 000 samples from five regions, namely Arborio, Basmati, Ipsala, Jasmine, and Karacadag. Each sample is characterized by 106 features, including Area, Perimeter, Eccentricity, Extent, Perimeter, etc. Within each of these five data groups, there are 15,000 samples. After completing the data preprocessing, each group retained 14,999 samples. We denote each set of sample sizes as $n_1 = n_2 = n_3 = n_4 = n_5 = 14999$, respectively. Therefore, the total sample size is $N = 74995$. After normalization of the data we performed hypothesis testing analysis. The question we want to investigate was to test whether there are significant differences in the rice data from the five regions, which is equivalent to the hypothesis testing problem $\boldsymbol{H}_{0sj} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = \boldsymbol{\mu}_5$ versus $\boldsymbol{H}_{1sj} : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for some $i \neq j$. We use 1, 000 replications when using the bootstrap method to find the critical values of the tests. The values of the test statistics and the approximate critical values are shown in Table 7.

From the results presented in Table 7, we find that the values of the test statistics of the four proposed tests are greater than the critical values. Based on these findings, we can conclude that there are significant differences among the rice features of the five regions.

6 Conclusion and discussion

This article proposes four multi-sample mean vector hypothesis testing methods. For the test problem $\boldsymbol{H}_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_K$ v.s. $\boldsymbol{H}_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for some $i \neq j$, we propose the test statistics T_{nsmax} and T_{smax}. To simplify the test procedure and enhance the powers of the test, we propose the screening-based hypothesis method. The corresponding test statistics are T^s_{nsmax} and T^s_{smax}. For the test procedure, instead of directly seeking the limiting distribution of the test statistic under the null hypothesis, we utilize normal distributions $N(\mathbf{0}, \tilde{\boldsymbol{R}}_{ns})$ and $N(\mathbf{0}, \tilde{\boldsymbol{R}}_s)$ to approximate them. We then employ the Monte Carlo method to generate an infinite number of normals with vectors to obtain the critical values of the tests. We give the bound of $T_{nsmax}(T_{smax})$ and $|\boldsymbol{W}_{ns}|_\infty(|\boldsymbol{W}_s|_\infty)$ as Proposition 1 when conditions $|\boldsymbol{R}_{ns} - \tilde{\boldsymbol{R}}_{ns}|_\propto = o_p(1)$ and $|\boldsymbol{R}_s - \tilde{\boldsymbol{R}}_s|_\propto = o_p(1)$ are satisfied. The first type of error of the proposed method can be controlled within 0.05, and their powers tend to 1. The validity of the proposed tests is verified in the simulation section, and a real data analysis section is included to illustrate the application of the proposed tests. The four methods are built upon two maximum-type statistics, distinguishing them from previous hypothesis testing methods that rely on the limiting null distributions of the statistics. As a result, we only need to utilize a parametric bootstrap technique to compute the critical value of the test. The suggested techniques allow for the general covariance structure and do not rely much on the structural assumptions of the unknown covariance matrix. The proposed tests are widely used because they do not make assumptions about the data structure.

In general, the screening-based testing procedures simplify the testing steps and improve the powers of the test, but the improvement of the power of the test $\boldsymbol{\Phi}^s_s$ is not very impressive in the high-dimensional case, we will try to solve this problem in our subsequent work, and we will push the mean vector testing method to a less restrictive situation so that it can be more widely applied.

Acknowledgements The authors would like to thank the Editor, the Associate Editor and two referees for their constructive comments. Jiang Hu was supported by National Natural Science Foundation of China (Grant Nos. 12171078, 12292980, 12292982) and Fundamental Research Funds for the Central Universities (Grant No. 2412023YQ003).

A. Appendix

In this section, we present the technical proofs of the results stated in Sect. 3. Before that, we give some auxiliary results for convenient use later.

Lemma 1 (Lemma 3.1 in Chernozhukov et al.(2013) *$\boldsymbol{X}$ and $\boldsymbol{Y}$ are p-dimensional random vectors, we assume that $\boldsymbol{X} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_1\right), \boldsymbol{Y} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_2\right)$ and there exist constants $C_1 > c_1 > 0$ such that for every $1 \le l \le p$, $c_1 \le \sigma_{1,ll}$, $\sigma_{2,ll} \le C_1$, where $\sigma_{1,ll} = diag\left(\boldsymbol{\Sigma}_1\right)$ and $\sigma_{2,ll} = diag\left(\boldsymbol{\Sigma}_2\right)$. Then, there exists a constant C' depending only on C_1 and c_1 such that*

$$\sup_{x \in R} \left| P\left(\max_{1 \le l \le p} X_l \le x \right) - P\left(\max_{1 \le l \le p} Y_l \le x \right) \right| \le C' \Lambda^{\frac{1}{3}} \{1 \vee \log(p/\Lambda)\}^{\frac{2}{3}},$$

where $\Lambda = \left|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\right|_\infty$.

Lemma 2 (*Lemma* 3 *in Chang et al.* 2017) *Suppose that n_i, $p > 2$ and $\log(p) \le n_i$, $\theta_{n_i,p} = pn_i^{1-\frac{r}{2}}, v_{ir}(\boldsymbol{X}) = \max_{1 \le l \le p} (E|D_i^{-1/2} X_{i1l}|^r)^{1/r}, R_i = (r_{i,kl}), \hat{R}_i = (\hat{r}_{i,kl}), \hat{r}_{i,kl} = \sum_{j=1}^{n_i} (X_{ijk} - \bar{X}_{ik})(X_{ijl} - \bar{X}_{il}) / \sqrt{\sum_{j=1}^{n_i} (X_{ijk} - \bar{X}_{ik})^2 \sum_{j=1}^{n_i} (X_{ijl} - \bar{X}_{il})^2}$, for all $i = 1, \ldots, K$ established.*

(*i*) *Assume that condition* (*A*.1) *holds. Then, there exist constants $C_2, C_3 > 0$ independent of n_i and p such that, with a probability of at least $1 - C_2\left\{ n_i^{-1} + \theta_{n_i,p}^{\frac{2}{(2+r)}} \right\}$,*

$$\max \left(\left| \boldsymbol{D}_i^{-\frac{1}{2}} \hat{\boldsymbol{\Sigma}}_i \boldsymbol{D}_i^{-\frac{1}{2}} - \boldsymbol{R}_i \right|_\infty, \left| \hat{\boldsymbol{R}}_i - \boldsymbol{R}_i \right|_\infty \right) \le C_3 \left[v_{i4}^2 n_i^{-\frac{1}{2}} \left\{ \log(pn_i) \right\}^{\frac{1}{2}} + v_{ir}^2 \theta_{n_i,p}^{\frac{2}{(r+2)}} + v_{ir}^2 \theta_{n_i,p}^{\frac{2}{r}} \log(p) \right].$$

(*ii*) *Assume that condition* (*A*.2) *holds. Then, there exist constants $C_4, C_5 > 0$ independent of n_i and p such that, with a probability of at least $1 - C_4 n_i^{-1}$,*

$$\max \left(\left| \boldsymbol{D}_i^{-\frac{1}{2}} \hat{\boldsymbol{\Sigma}}_i \boldsymbol{D}_i^{-\frac{1}{2}} - \boldsymbol{R}_i \right|_\infty, \left| \hat{\boldsymbol{R}}_i - \boldsymbol{R}_i \right|_\infty \right) \le C_5 \left[n_i^{-\frac{1}{2}} \left\{ \log(pn_i) \right\}^{\frac{1}{2}} + n_i^{-1} \log(pn_i)^{\frac{2}{\gamma}} \right].$$

(*iii*) *Assume that $v_{i4}, i = 1, \ldots, K$ are uniformly bounded, then for $0 < t \le \sqrt{n_i}$,*

$$P\left\{ \sqrt{n_i} \left| \hat{\boldsymbol{D}}_i^{-\frac{1}{2}} \left(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i \right) \right|_\infty \ge t \right\} \le Cp \exp(-ct^2) + n_i^{-1},$$

where $\hat{\boldsymbol{D}}_i = diag\left(\hat{\sigma}_{i1}^2, \ldots, \hat{\sigma}_{ip}^2\right)$, $\hat{\boldsymbol{\mu}}_i = \left(\bar{X}_{i1}, \ldots, \bar{X}_{ip}\right)'$ and $C, c > 0$ are constants independent of p.

The following states the auxiliary theory of the procedures for testing the equality of means, which is based on the idea of Gaussian approximation. $\boldsymbol{H}_1, \boldsymbol{H}_2, \dots, \boldsymbol{H}_N$ are independent random vectors in $\mathbb{R}^{Kp}$, suppose each $\boldsymbol{H}_i$ is centered and has a finite covariance matrix $E[\boldsymbol{H}_i\boldsymbol{H}_i']$. Which $\boldsymbol{H}_i = (H_{i1}, \dots, H_{iKp})'$, $\boldsymbol{B}_i = diag(E[\boldsymbol{H}_i\boldsymbol{H}_i'])$, $\sigma_{il}^2 = var(H_{il}) = \sigma_{i,ll}$ and define

$$T_0 := \max_{1\le l\le Kp} N^{-\frac{1}{2}} \sum_{i=1}^{N} \frac{H_{il}}{\sigma_{il}}.$$

Let $\boldsymbol{G}_1, \dots, \boldsymbol{G}_N$ are a sequence of independent centered Gaussian random vectors in $\mathbb{R}^{Kp}$ such that each $\boldsymbol{G}_i$ has the same covariance matrix as $\boldsymbol{H}_i$ and $\boldsymbol{G}_i \sim N(\boldsymbol{0}, E[\boldsymbol{H}_i\boldsymbol{H}_i'])$. The following Z_0 is the Gaussian analogue of T_0 can be defined as

$$Z_0 := \max_{1\le l\le Kp} N^{-\frac{1}{2}} \sum_{i=1}^{N} \frac{G_{il}}{\sigma_{il}}.$$

For $l = 1, \dots, Kp$ and $r \ge 1$, define the moments

$$M_r(\boldsymbol{H}) := \max_{1\le l\le Kp} \left(N^{-1} \sum_{i=1}^{N} E|V_{il}|^r \right)^{1/r},$$

where $\boldsymbol{V}_i = \boldsymbol{V}_i(\boldsymbol{H}) = (V_{i1}, \dots, V_{iKp})' = \boldsymbol{B}_i^{-\frac{1}{2}}\boldsymbol{H}_i$, $i = 1, \dots, N$. For $0 \le t \le 1$, set $u(t) = \max\left(u_X(t), u_G(t)\right)$, which is the maximum $(1-t)$-quantiles of $\max_{1\le i\le N} |\boldsymbol{V}_i|_\infty$ and $\max_{1\le i\le N} \left|\boldsymbol{B}_i^{-\frac{1}{2}}\boldsymbol{G}_i\right|_\infty$, respectively.

Lemma 3 (Theorem 2.2 in Chernozhukov et al. 2013) *Assume that $\sigma_{i,ll}$ is bounded away from 0 and ∞. Then, for any $0 \le t \le 1$,*

$$\begin{aligned}
&\sup_{x\in R} \left| P\left(T_0 \le x\right) - P\left(Z_0 \le x\right) \right| \\
&\le C\left[\left(M_3^{\frac{3}{4}} \vee M_4^{\frac{1}{2}} \right) N^{-\frac{1}{8}} \{\log(KpN/t)\}^{\frac{7}{8}} + u(t)N^{-\frac{1}{2}} \{\log(KpN/t)\}^{\frac{3}{2}} + t \right],
\end{aligned} \tag{5}$$

where $C > 0$ is a constant independent of N, p and t.

Remark 2 The bound of (5) can be easily reduced under some conditions, respectively. According to Markov's inequality, for $u > 0$,

$$P\left(\max_{1\le i\le N} \left|\boldsymbol{B}_i^{-\frac{1}{2}}\boldsymbol{G}_i\right|_\infty > u \right) \le 2KNp(1 - \Phi(u)) \le u^{-1} \exp\left\{ \log(KpN) - u^2/2 \right\}.$$

Since

$$u_G(t) = \inf\left\{ u \geq 0 : P\left(\max_{1\leq i\leq N} \left| \boldsymbol{B}_i^{-\frac{1}{2}} \boldsymbol{G}_i \right|_\propto > u \right) \leq t \right\},$$

for $0 \leq t \leq 1$, $u_G(t) \leq \sqrt{2\log(KpN/t)}$. If $\max_{1\leq i\leq N} \max_{1\leq l\leq Kp} \left\{ E(\left|V_{il}\right|^r) \right\}^{\frac{1}{r}} \leq C_4'$ for $r \geq 4$ and $C_4' > 0$ holds, according to Markov's inequality, for all $u > 0$, let

$$\mu_{N,r} := \left\{ E(\max_{1\leq i\leq N} |\boldsymbol{V}_i|_\propto^r) \right\}^{\frac{1}{r}},$$

thus $P\left(\max_{1\leq i\leq N} |\boldsymbol{V}_i|_\propto > u \right) \leq u^{-r} E\left(\max_{1\leq i\leq N} |\boldsymbol{V}_i|_\propto^r \right) = u^{-r}\mu_{N,r}^r$. And since

$$u_X(t) := \inf\left\{ u \geq 0 : P\left(\max_{1\leq i\leq N} |\boldsymbol{V}_i|_\propto \right) > u) \leq t \right\},$$

hence $u_X(t) \leq t^{-\frac{1}{r}}\mu_{N,r}$. By the inequality that

$$\mu_{N,r} = \left\{ E\left(\max_{1\leq i\leq N} |\boldsymbol{V}_i|_\propto^r \right) \right\}^{\frac{1}{r}} \leq (KpN)^{\frac{1}{r}} \max_{1\leq l\leq Kp, 1\leq i\leq N} \left(E\left(\left|\boldsymbol{V}_{il}\right|^r\right) \right)^{\frac{1}{r}}.$$

And taking $t = \min\left\{ 1, \left[\mu_{N,r} N^{-\frac{1}{2}} \{\log(KpN)\}^{\frac{3}{2}} \right]^{\frac{r}{(r+1)}} \right\}$, the bound of (5) is

$$N^{-\frac{1}{8}} \{\log(KpN)\}^{\frac{7}{8}} + \theta_{N,p}^{\frac{1}{(r+1)}} \{\log(KpN)\}^{\frac{3}{2}}. \tag{6}$$

Moreover, if $\max_{1\leq i\leq N} \max_{1\leq l\leq Kp} E(\exp(C_5'\left|V_{il}\right|^\gamma)) \leq C_6'$ for some $C_5' > 0$, $C_6' > 1$ and $0 < \gamma \leq 2$ holds, then $u_X(t) \leq \{\log\{KpN/t\}\}^{\frac{1}{\gamma}}$. Let $t = N^{-\frac{1}{2}}$, thus the bound of (5) is

$$N^{-\frac{1}{8}} \{\log(KpN)\}^{\frac{7}{8}} + N^{-\frac{1}{2}} \{\log(KpN)\}^{\frac{3}{2}+\frac{1}{\gamma}}, \tag{7}$$

where $\theta_{N,p} = KpN^{1-\frac{r}{2}}$.

In the sequel, we give the proofs of Proposition 1 and Theorems 1–4. For simplicity, we only consider the three-sample case, i.e., $K = 3$, and the proofs are analogous when $K > 3$.

A. 1 Proof of Proposition 1

Because the non-studentized statistic T_{nsmax} can be handled similarly, we only provide the proof of the studentized statistic T_{smax} for the sake of clarity. To begin with, observe that for every $x \in \mathbb{R}$, $|x| = \max(-x, x)$. For $t > 0$ and the definition of T_{smax}, the distribution of test statistic can be written as follows:

$$
\begin{aligned}
&P\left(T_{smax} > t\right) \\
=&P\left(\max_{1\le l\le p}\left(\frac{|\bar{X}_{1l}-\bar{X}_{2l}|}{\sqrt{\hat{\sigma}_{1l}^2/n_1+\hat{\sigma}_{2l}^2/n_2}},\frac{|\bar{X}_{1l}-\bar{X}_{3l}|}{\sqrt{\hat{\sigma}_{1l}^2/n_1+\hat{\sigma}_{3l}^2/n_3}},\frac{|\bar{X}_{2l}-\bar{X}_{3l}|}{\sqrt{\hat{\sigma}_{2l}^2/n_2+\hat{\sigma}_{3l}^2/n_3}}\right)\ge t\right) \\
=&P\left(\max_{1\le l\le p}\left(\max\left(\frac{\bar{X}_{1l}-\bar{X}_{2l}}{\sqrt{\hat{\sigma}_{1l}^2/n_1+\hat{\sigma}_{2l}^2/n_2}},-\frac{\bar{X}_{1l}-\bar{X}_{2l}}{\sqrt{\hat{\sigma}_{1k}^2/n_1+\hat{\sigma}_{2l}^2/n_2}}\right),\right.\right. \\
&\quad\max\left(\frac{\bar{X}_{1l}-\bar{X}_{3l}}{\sqrt{\hat{\sigma}_{1l}^2/n_1+\hat{\sigma}_{3l}^2/n_3}},-\frac{\bar{X}_{1l}-\bar{X}_{3l}}{\sqrt{\hat{\sigma}_{2k}^2/n_2+\hat{\sigma}_{3l}^2/n_3}}\right), \\
&\quad\left.\left.\max\left(\frac{\bar{X}_{2l}-\bar{X}_{3l}}{\sqrt{\hat{\sigma}_{2l}^2/n_2+\hat{\sigma}_{3l}^2/n_3}},-\frac{\bar{X}_{2l}-\bar{X}_{3l}}{\sqrt{\hat{\sigma}_{2l}^2/n_2+\hat{\sigma}_{3l}^2/n_3}}\right)\right)\ge t\right).
\end{aligned}
$$

Establish a new set of dilated random vectors $\boldsymbol{X}_{11}^e,\ldots,\boldsymbol{X}_{1n_1}^e$, $\boldsymbol{X}_{21}^e,\ldots,\boldsymbol{X}_{2n_2}^e$, $\boldsymbol{X}_{31}^e,\ldots,\boldsymbol{X}_{3n_3}^e$ taking values in $\mathbb{R}^{2p}$, given by

$$
\begin{aligned}
\boldsymbol{X}_{1i}^e &= \left(X_{1i1}^e,\ldots,X_{1i2p}^e\right)' = \left(\boldsymbol{X}_{1i}',-\boldsymbol{X}_{1i}'\right)', \\
\boldsymbol{X}_{2i}^e &= \left(X_{2i1}^e,\ldots,X_{2i2p}^e\right)' = \left(\boldsymbol{X}_{2i}',-\boldsymbol{X}_{2i}'\right)', \\
\boldsymbol{X}_{3i}^e &= \left(X_{3i1}^e,\ldots,X_{3i2p}^e\right)' = \left(\boldsymbol{X}_{3i}',-\boldsymbol{X}_{3i}'\right)'.
\end{aligned}
$$

From this point of view, we have $P\left(T_{smax} > t\right) = P\left(T_{smax}^e > t\right)$, where

$$
\begin{aligned}
\bar{X}_{1l}^e &= n_1^{-1}\sum_{i=1}^{n_1} X_{1il}^e, \quad \left(\hat{\sigma}_{1l}^e\right)^2 = n_1^{-1}\sum_{i=1}^{n_1}(X_{1il}^e-\bar{X}_{1l}^e)^2, \\
T_{s1}^e &= \max_{1\le l\le p}\frac{\bar{X}_{1l}^e-\bar{X}_{2l}^e}{\sqrt{(\hat{\sigma}_{1l}^e)^2/n_1+(\hat{\sigma}_{2l}^e)^2/n_2}}; \\
\bar{X}_{2l}^e &= n_2^{-1}\sum_{i=1}^{n_2} X_{2il}^e, \quad \left(\hat{\sigma}_{2l}^e\right)^2 = n_2^{-1}\sum_{i=1}^{n_2}\left(X_{2il}^e-\bar{X}_{2l}^e\right)^2, \\
T_{s2}^e &= \max_{1\le l\le p}\frac{\bar{X}_{1l}^e-\bar{X}_{3l}^e}{\sqrt{\left(\hat{\sigma}_{1l}^e\right)^2/n_1+\left(\hat{\sigma}_{3l}^e\right)^2/n_3}}; \\
\bar{X}_{3l}^e &= n_3^{-1}\sum_{i=1}^{n_3} X_{3il}^e, \quad \left(\hat{\sigma}_{3l}^e\right)^2 = n_3^{-1}\sum_{i=1}^{n_3}(X_{3il}^e-\bar{X}_{3l}^e)^2, \\
T_{s3}^e &= \max_{1\le l\le p}\frac{\bar{X}_{2l}^e-\bar{X}_{3l}^e}{\sqrt{\left(\hat{\sigma}_{2l}^e\right)^2/n_2+(\hat{\sigma}_{3l}^e)^2/n_3}}; \\
T_{smax}^e &= \max(T_{s1}^e,T_{s2}^e,T_{s3}^e).
\end{aligned}
$$

Without loss of generality, we only need to focus on

$$T_{smax}^{+} = \max_{1\le l\le p}\left\{ \frac{\bar{X}_{1l}-\bar{X}_{2l}}{\left(\hat{\sigma}_{1l}^2/n_1+\hat{\sigma}_{2l}^2/n_2\right)^{1/2}}, \frac{\bar{X}_{1l}-\bar{X}_{3l}}{\left(\hat{\sigma}_{1l}^2/n_1+\hat{\sigma}_{3l}^2/n_3\right)^{1/2}}, \frac{\bar{X}_{2l}-\bar{X}_{3l}}{\left(\hat{\sigma}_{2l}^2/n_2+\hat{\sigma}_{3l}^2/n_3\right)^{1/2}} \right\}.$$

Let $N = n_1 + n_2 + n_3$, denote by

$$s_{1l}^2 = \frac{n_1}{N}\left(\sigma_{1l}^2 + \frac{n_1}{n_2}\sigma_{2l}^2\right),\ \hat{s}_{1l}^2 = \frac{n_1}{N}\left(\hat{\sigma}_{1l}^2 + \frac{n_1}{n_2}\hat{\sigma}_{2l}^2\right),\ l = 1,\dots,p,$$

$$\boldsymbol{D}_1 = \begin{bmatrix} s_{11}^2 & & \\ & \ddots & \\ & & s_{1p}^2 \end{bmatrix}, \hat{\boldsymbol{D}}_1 = \begin{bmatrix} \hat{s}_{11}^2 & & \\ & \ddots & \\ & & \hat{s}_{1p}^2 \end{bmatrix}.$$

$$s_{2l}^2 = \frac{n_1}{N}\left(\sigma_{1l}^2 + \frac{n_1}{n_3}\sigma_{3l}^2\right),\ \hat{s}_{2l}^2 = \frac{n_1}{N}\left(\hat{\sigma}_{1l}^2 + \frac{n_1}{n_3}\hat{\sigma}_{3l}^2\right),\ l = 1,\dots,p,$$

$$\boldsymbol{D}_2 = \begin{bmatrix} s_{21}^2 & & \\ & \ddots & \\ & & s_{2p}^2 \end{bmatrix}, \hat{\boldsymbol{D}}_2 = \begin{bmatrix} \hat{s}_{21}^2 & & \\ & \ddots & \\ & & \hat{s}_{2p}^2 \end{bmatrix}.$$

$$s_{3l}^2 = \frac{n_2}{N}\left(\sigma_{2l}^2 + \frac{n_2}{n_3}\sigma_{3l}^2\right),\ \hat{s}_{3l}^2 = \frac{n_2}{N}\left(\hat{\sigma}_{2l}^2 + \frac{n_2}{n_3}\hat{\sigma}_{3l}^2\right),\ l = 1,\dots,p,$$

$$\boldsymbol{D}_3 = \begin{bmatrix} s_{31}^2 & & \\ & \ddots & \\ & & s_{3p}^2 \end{bmatrix}, \hat{\boldsymbol{D}}_3 = \begin{bmatrix} \hat{s}_{31}^2 & & \\ & \ddots & \\ & & \hat{s}_{3p}^2 \end{bmatrix}.$$

And

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_1 & & \\ & \boldsymbol{D}_2 & \\ & & \boldsymbol{D}_3 \end{bmatrix}, \hat{\boldsymbol{D}} = \begin{bmatrix} \hat{\boldsymbol{D}}_1 & & \\ & \hat{\boldsymbol{D}}_2 & \\ & & \hat{\boldsymbol{D}}_3 \end{bmatrix}.$$

First, we define a pooled sequence of random vectors $\left\{\boldsymbol{\alpha}_i = \left(\alpha_{i1},\dots,\alpha_{i3p}\right)'\right\}_{i=1}^{N}$ as follows:

$$\boldsymbol{\alpha}_i = \begin{cases} \begin{bmatrix} \boldsymbol{I}_{p\times p} \\ \boldsymbol{I}_{p\times p} \\ \boldsymbol{0}_{p\times p} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{1i} - \boldsymbol{\mu}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_{1i} - \boldsymbol{\mu}_1 \\ \boldsymbol{X}_{1i} - \boldsymbol{\mu}_1 \\ \boldsymbol{0}_p \end{bmatrix}, & 1 \le i \le n_1; \\ \begin{bmatrix} -\frac{n_1}{n_2}\boldsymbol{I}_{p\times p} \\ \boldsymbol{0}_{p\times p} \\ \boldsymbol{I}_{p\times p} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{2(i-n_1)} - \boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} -\frac{n_1}{n_2}\left(\boldsymbol{X}_{2(i-n_1)} - \boldsymbol{\mu}_2\right) \\ \boldsymbol{0}_p \\ \boldsymbol{X}_{2(i-n_1)} - \boldsymbol{\mu}_2 \end{bmatrix}, & n_1 + 1 \le i \le n_1 + n_2; \\ \begin{bmatrix} \boldsymbol{0}_{p\times p} \\ -\frac{n_1}{n_2}\boldsymbol{I}_{p\times p} \\ -\frac{n_2}{n_3}\boldsymbol{I}_{p\times p} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{3(i-n_1-n_2)} - \boldsymbol{\mu}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0}_p \\ -\frac{n_1}{n_3}\left(\boldsymbol{X}_{3(i-n_1-n_2)} - \boldsymbol{\mu}_3\right) \\ -\frac{n_2}{n_3}\left(\boldsymbol{X}_{3(i-n_1-n_2)} - \boldsymbol{\mu}_3\right) \end{bmatrix}, & n_1 + n_2 + 1 \le i \le n_1 + n_2 + n_3; \end{cases}$$

$$\tilde{\boldsymbol{\alpha}}_i = \boldsymbol{D}^{-1/2}\boldsymbol{\alpha}_i,\ \hat{\boldsymbol{\alpha}}_i = \hat{\boldsymbol{D}}^{-1/2}\boldsymbol{\alpha}_i.$$

By the above notation and under the null hypothesis $\boldsymbol{H}_0$, we have that

$$\begin{aligned} T^*_{smax} &= N^{-1/2} \max_{1\le l\le 3p} \sum\nolimits_{i=1}^{N} \tilde{\alpha}_{il} \\ &= \max_{1\le l\le p} \left\{ \frac{\bar{X}_{1l} - \bar{X}_{2l}}{\left(\sigma^2_{1l}/n_1 + \sigma^2_{2l}/n_2\right)^{1/2}}, \frac{\bar{X}_{1l} - \bar{X}_{3l}}{\left(\sigma^2_{1l}/n_1 + \sigma^2_{3l}/n_3\right)^{1/2}}, \frac{\bar{X}_{2l} - \bar{X}_{3l}}{\left(\sigma^2_{2l}/n_2 + \sigma^2_{3l}/n_3\right)^{1/2}} \right\} \end{aligned}$$

and

$$\begin{aligned} T^+_{smax} &= N^{-1/2} \max_{1\le l\le 3p} \sum\nolimits_{i=1}^{N} \hat{\alpha}_{il} \\ &= \max_{1\le l\le p} \left\{ \frac{\bar{X}_{1l} - \bar{X}_{2l}}{\left(\hat{\sigma}^2_{1l}/n_1 + \hat{\sigma}^2_{2l}/n_2\right)^{1/2}}, \frac{\bar{X}_{1l} - \bar{X}_{3l}}{\left(\hat{\sigma}^2_{1l}/n_1 + \hat{\sigma}^2_{3l}/n_3\right)^{1/2}}, \frac{\bar{X}_{2l} - \bar{X}_{3l}}{\left(\hat{\sigma}^2_{2l}/n_2 + \hat{\sigma}^2_{3l}/n_3\right)^{1/2}} \right\}. \end{aligned}$$

Conditional on $\{\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3\}$, the 3p-dimensional random vector $\boldsymbol{W}_s = \left(W_{s1}, \dots, W_{s3p}\right)'$ is a centered Gaussian random vector with covariance matrix $\tilde{\boldsymbol{R}}_s$. Let

$$Z^+_{smax} = \max_{1\le l\le 3p} W_{sl}.$$

We will demonstrate that the distribution of T^+_{smax} is sufficiently close to the conditional distribution of Z^+_{smax} given $\{\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3\}$ with high probability. We define the following random vectors. When $1 \le i \le n_1$, let $\boldsymbol{g}_i \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_1\right)$,

$$\boldsymbol{G}_i = \begin{bmatrix} \boldsymbol{I}_{p\times p} \\ \boldsymbol{I}_{p\times p} \\ \boldsymbol{0}_{p\times p} \end{bmatrix} \left(\boldsymbol{g}_i\right) \text{ and } \tilde{\boldsymbol{G}}_i = \boldsymbol{D}^{-1/2}\boldsymbol{G}_i.$$

When $n_1 + 1 \le i \le n_1 + n_2$, let $\boldsymbol{g}_i \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_2\right)$,

$$\boldsymbol{G}_i = \begin{bmatrix} -\frac{n_1}{n_2}\boldsymbol{I}_{p\times p} \\ \boldsymbol{0}_{p\times p} \\ \boldsymbol{I}_{p\times p} \end{bmatrix} \boldsymbol{g}_i \text{ and } \tilde{\boldsymbol{G}}_i = \boldsymbol{D}^{-1/2}\boldsymbol{G}_i.$$

When $n_1 + n_2 + 1 \le i \le n_1 + n_2 + n_3$, let $\boldsymbol{g}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_3)$,

$$\boldsymbol{G}_i = \begin{bmatrix} \boldsymbol{0}_{p\times p} \\ \frac{n_1}{n_3}\boldsymbol{I}_{p\times p} \\ -\frac{n_2}{n_3}\boldsymbol{I}_{p\times p} \end{bmatrix} \boldsymbol{g}_i \text{ and } \tilde{\boldsymbol{G}}_i = \boldsymbol{D}^{-1/2}\boldsymbol{G}_i.$$

By the notation of $\boldsymbol{G}_i$ and $\tilde{\boldsymbol{R}}_1$, we define

$$Z^*_{smax} = \max_{1\le l\le 3p} N^{-\frac{1}{2}} \sum\nolimits_{i=1}^{N} \tilde{G}_{il}.$$

Note that Z^*_{smax} and T^*_{smax} are random variables. By Lemma 3, we have the following Berry–Esseen type bound

$$d = \sup_{x\in R} \left| P\left(T^*_{smax} \le x\right) - P\left(Z^*_{smax} \le x\right) \right|, \tag{8}$$

where the order of d depends on the moment conditions imposed on $\boldsymbol{\alpha}_i$ as described in (6) and (7). For the Gaussian maximum Z^*_{smax} and Z^+_{smax} given above, obtained from Lemma 1 to bound:

$$\hat{d} = \sup_{x\in R} |P\left(Z^*_{smax} \le x\right) - P\left(Z^+_{smax} \le x|\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3\right)| \le C\Theta_s^{\frac{1}{3}} \left\{1 + \log(3p/\Theta_s)\right\}^{\frac{2}{3}}, \tag{9}$$

where $\Theta_s = |\tilde{R}_s - R_s|_\infty$. For $0 < t_{12} \le \frac{1}{2}$, define the event

$$\boldsymbol{\varepsilon}_{n_1,n_2}\left(t_{12}\right) = \left\{ \max_{1\le l\le p} \left| \frac{\hat{\sigma}^2_{vl}}{\sigma^2_{vl}} - 1 \right| \le t_{12}, v = 1, 2 \right\}.$$

On $\boldsymbol{\varepsilon}_{n_1,n_2}\left(t_{12}\right)$, we have $1 - t_{12} \le \left(\hat{s}_{1l}/s_{1l}\right)^2 \le 1 + t_{12}$ for all $l = 1, \dots, p$. The inequality $1 + \frac{1}{2}\mu - \frac{1}{2}\mu^2 \le (1+\mu)^{1/2} \le 1 + \frac{1}{2}\mu$ holds for $\mu \ge -1$, and thus, we can apply the inequality for $0 < t_{12} \le 1/2$ to obtain that

$$\boldsymbol{\varepsilon}_{n_1,n_2}(t_{12}) \subseteq \boldsymbol{A}_{n_1,n_2}(t_{12}) = \left\{ \max_{1\le l\le p} \left| \frac{\hat{s}_{1l}}{s_{1l}} - 1 \right| \le \frac{t_{12}\left(1 + t_{12}\right)}{2} \right\}.$$

Similarly, we define the event

$$\boldsymbol{\varepsilon}_{n_1,n_3}(t_{13}) = \left\{ \max_{1\le l\le p} \left| \frac{\hat{\sigma}^2_{vl}}{\sigma^2_{vl}} - 1 \right| \le t_{13},\ v = 1, 3 \right\},$$

and

$$\boldsymbol{\varepsilon}_{n_2,n_3}\left(t_{23}\right) = \left\{ \max_{1\le l\le p} \left| \frac{\hat{\sigma}_{vl}^2}{\sigma_{vl}^2} - 1 \right| \le t_{23},\ v = 2, 3 \right\}.$$

We have that

$$\boldsymbol{\varepsilon}_{n_1,n_3}\left(t_{13}\right) \subseteq \boldsymbol{A}_{n_1,n_3}\left(t_{13}\right) = \left\{ \max_{1\le l\le p} \left|\hat{s}_{2l}/s_{2l} - 1\right| \le \frac{t_{13}(1+t_{13})}{2} \right\}$$

and

$$\boldsymbol{\varepsilon}_{n_2,n_3}(t_{23}) \subseteq \boldsymbol{A}_{n_2,n_3}\left(t_{23}\right) = \left\{ \max_{1\le l\le p} \left|\hat{s}_{3l}/s_{3l} - 1\right| \le \frac{t_{23}(1+t_{23})}{2} \right\}.$$

Apply Theorem 3 in Chernozhukov et al. (2015),

$$P\left(x-\varepsilon < Z_{smax}^{*} \le x\right) \le 4\varepsilon\left(1 + (2\log(3p))^{1/2}\right) := \Delta_1. \tag{10}$$

Consequently, combination of inequalities (8) (9) (10) gives, for arbitrary $x \in \mathbb{R}$ and $\varepsilon > 0$,

$$\begin{aligned} P\left(T_{smax}^{+} > x\right) &\le P\left(T_{smax}^{*} > x-\varepsilon\right) + P\left(\left(T_{smax}^{+} - T_{smax}^{*}\right) > \varepsilon\right) \\ &\le \sup_{x\in R}\left|P\left(T_{smax}^{*} > x-\varepsilon\right) - P\left(Z_{smax}^{*} > x-\varepsilon\right)\right| \\ &\quad + P\left(Z_{smax}^{*} > x-\varepsilon\right) + P\left(T_{smax}^{+} - T_{smax}^{*} > \varepsilon\right) \\ &\le d + \Delta_1 + P(Z_{smax}^{*} > x) + P((T_{smax}^{+} - T_{smax}^{*}) > \varepsilon) \\ &\le d + \Delta_1 + \hat{d} + P(Z_{smax}^{+} > x|\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3) + P((T_{smax}^{+} - T_{smax}^{*}) > \varepsilon). \end{aligned} \tag{11}$$

From the above inequalities, we can get the following conclusion

$$\begin{aligned} &P\left(T_{smax}^{+} > x\right) - P\left(Z_{smax}^{+} > x|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right) \\ &\le d + \hat{d} + \Delta_1 + P\left(\left(T_{smax}^{+} - T_{smax}^{*}\right) > \varepsilon\right). \end{aligned}$$

Similarly, we can get the inverse inequality as follows

$$P\left(Z_{smax}^{+} > x\right) - P\left(T_{smax}^{+} > x|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right) \le d + \hat{d} + \Delta_1 + P\left(\left(T_{smax}^{*} - T_{smax}^{+}\right) > \varepsilon\right).$$

Thus, we can obtain the following result. For every $x \in \mathbb{R}$, we have that

$$\begin{aligned} &\sup_{x\in R}\left|P\left(T_{smax}^{+} > x\right) - P\left(Z_{smax}^{+} > x|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right)\right| \\ &\le \Delta_1 + d + \hat{d} + \max\left\{P\left(\left(T_{smax}^{+} - T_{smax}^{*}\right) > \varepsilon\right), P\left(\left(T_{smax}^{*} - T_{smax}^{+}\right) > \varepsilon\right)\right\}. \end{aligned} \tag{12}$$

The next major task is to calculate the probability of

$$P\left((T_{smax}^{+} - T_{smax}^{*}) > \varepsilon\right) \quad \text{and } P\left((T_{smax}^{*} - T_{smax}^{+}) > \varepsilon\right).$$

Under the null hypothesis, we know that

$$\left|\max_{1\le l\le p}\frac{\sqrt{n_1n_2}\left(\bar{X}_{1l}-\bar{X}_{2l}\right)}{\sqrt{n_2\hat{\sigma}_{1l}^2+n_1\hat{\sigma}_{2l}^2}}-\max_{1\le l\le p}\frac{\sqrt{n_1n_2}\left(\bar{X}_{1l}-\bar{X}_{2l}\right)}{\sqrt{n_2\sigma_{1l}^2+n_1\sigma_{2l}^2}}\right| \tag{13}$$
$$\le \max_{1\le l\le p}\left|\hat{s}_{1l}/s_{1l}-1\right|\sqrt{n_1}\left|\hat{\boldsymbol{D}}_{1,2}\left(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_2\right)\right|_{\propto},$$

$$\left|\max_{1\le l\le p}\frac{\sqrt{n_1n_3}(\bar{X}_{1l}-\bar{X}_{3l})}{\sqrt{n_3\hat{\sigma}_{1l}^2+n_1\hat{\sigma}_{3l}^2}}-\max_{1\le l\le p}\frac{\sqrt{n_1n_3}(\bar{X}_{1l}-\bar{X}_{3l})}{\sqrt{n_3\sigma_{1l}^2+n_1\sigma_{3l}^2}}\right| \tag{14}$$
$$\le \max_{1\le l\le p}\left|\hat{s}_{2l}/s_{2l}-1\right|\sqrt{n_1}\left|\hat{\boldsymbol{D}}_{1,3}(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_3)\right|_{\propto},$$

$$\left|\max_{1\le l\le p}\frac{\sqrt{n_2n_3}\left(\bar{X}_{2l}-\bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}_{2l}^2+n_2\hat{\sigma}_{3l}^2}}-\max_{1\le l\le p}\frac{\sqrt{n_2n_3}\left(\bar{X}_{2l}-\bar{X}_{3l}\right)}{\sqrt{n_3\sigma_{2l}^2+n_2\sigma_{3l}^2}}\right| \tag{15}$$
$$\le \max_{1\le l\le p}\left|\hat{s}_{3l}/s_{3l}-1\right|\sqrt{n_2}\left|\hat{\boldsymbol{D}}_{2,3}\left(\hat{\boldsymbol{\mu}}_2-\hat{\boldsymbol{\mu}}_3\right)\right|_{\propto},$$

where

$$\hat{\boldsymbol{D}}_{1,2}=diag\left(\hat{\boldsymbol{\Sigma}}_1+\frac{n_1}{n_2}\hat{\boldsymbol{\Sigma}}_2\right),\ \hat{\boldsymbol{D}}_{2,3}=diag\left(\hat{\boldsymbol{\Sigma}}_2+\frac{n_2}{n_3}\hat{\boldsymbol{\Sigma}}_3\right),\ \hat{\boldsymbol{D}}_{1,3}=diag\left(\hat{\boldsymbol{\Sigma}}_1+\frac{n_1}{n_3}\hat{\boldsymbol{\Sigma}}_3\right).$$

By the above definitions, we obtain that

$$
\begin{aligned}
&P((T^{+}_{smax} - T^{*}_{smax}) > \varepsilon)\\
&\leq P\left(\max_{1\leq l\leq p} \frac{\sqrt{n_1 n_2}\left(\bar{X}_{1l} - \bar{X}_{2l}\right)}{\sqrt{n_2\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{2l}}} - T^{*}_{smax} > \varepsilon, \max_{1\leq l\leq p} \frac{\sqrt{n_1 n_3}\left(\bar{X}_{1l} - \bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{3l}}} - T^{*}_{smax} > \varepsilon,\right.\\
&\qquad \left.\max_{1\leq l\leq p} \frac{\sqrt{n_2 n_3}\left(\bar{X}_{2l} - \bar{X}_{3l}\right)}{\sqrt{n_2\hat{\sigma}^2_{3l} + n_3\hat{\sigma}^2_{2l}}} - T^{*}_{smax} > \varepsilon\right)\\
&\leq P\left(\max_{1\leq l\leq p} \frac{\sqrt{n_1 n_2}\left(\bar{X}_{1l} - \bar{X}_{2l}\right)}{\sqrt{n_2\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{2l}}} - T^{*}_{smax} > \varepsilon\right) + P\left(\max_{1\leq l\leq p} \frac{\sqrt{n_1 n_3}\left(\bar{X}_{1l} - \bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{3l}}} - T^{*}_{smax} > \varepsilon\right)\\
&\quad + P\left(\max_{1\leq l\leq p} \frac{\sqrt{n_2 n_3}\left(\bar{X}_{2l} - \bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}^2_{2l} + n_2\hat{\sigma}^2_{3l}}} - T^{*}_{smax} > \varepsilon\right)\\
&\leq P\left(\left|\max_{1\leq l\leq p} \frac{\sqrt{n_1 n_2}\left(\bar{X}_{1l} - \bar{X}_{2l}\right)}{\sqrt{n_2\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{2l}}} - \max_{1\leq l\leq p} \frac{\sqrt{n_1 n_2}\left(\bar{X}_{1l} - \bar{X}_{2l}\right)}{\sqrt{n_2\sigma^2_{1l} + n_1\sigma^2_{2l}}}\right| > \varepsilon\right)\\
&\quad + P\left(\left|\max_{1\leq l\leq p} \frac{\sqrt{n_1 n_3}\left(\bar{X}_{1l} - \bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{3l}}} - \max_{1\leq l\leq p} \frac{\sqrt{n_1 n_3}\left(\bar{X}_{1l} - \bar{X}_{3l}\right)}{\sqrt{n_3\sigma^2_{1l} + n_1\sigma^2_{3l}}}\right| > \varepsilon\right)\\
&\quad + P\left(\left|\max_{1\leq l\leq p} \frac{\sqrt{n_2 n_3}\left(\bar{X}_{2l} - \bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}^2_{2l} + n_2\hat{\sigma}^2_{3l}}} - \max_{1\leq l\leq p} \frac{\sqrt{n_2 n_3}\left(\bar{X}_{2l} - \bar{X}_{3l}\right)}{\sqrt{n_3\sigma^2_{2l} + n_2\sigma^2_{3l}}}\right| > \varepsilon\right).
\end{aligned}
$$

Similarly, the following inequalities

$$
\begin{aligned}
P((T^{*}_{smax} - T^{+}_{smax}) > \varepsilon) &\leq P\left(\left|\max_{1\leq l\leq p} \frac{\sqrt{n_1 n_2}\left(\bar{X}_{1l} - \bar{X}_{2l}\right)}{\sqrt{n_2\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{2l}}} - \max_{1\leq l\leq p} \frac{\sqrt{n_1 n_2}\left(\bar{X}_{1l} - \bar{X}_{2l}\right)}{\sqrt{n_2\sigma^2_{1l} + n_1\sigma^2_{2l}}}\right| > \varepsilon\right)\\
&\quad + P\left(\left|\max_{1\leq l\leq p} \frac{\sqrt{n_1 n_3}\left(\bar{X}_{1l} - \bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{3l}}} - \max_{1\leq l\leq p} \frac{\sqrt{n_1 n_3}\left(\bar{X}_{1l} - \bar{X}_{3l}\right)}{\sqrt{n_3\sigma^2_{1l} + n_1\sigma^2_{3l}}}\right| > \varepsilon\right)\\
&\quad + P\left(\left|\max_{1\leq l\leq p} \frac{\sqrt{n_2 n_3}\left(\bar{X}_{2l} - \bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}^2_{2l} + n_2\hat{\sigma}^2_{3l}}} - \max_{1\leq l\leq p} \frac{\sqrt{n_2 n_3}\left(\bar{X}_{2l} - \bar{X}_{3l}\right)}{\sqrt{n_3\sigma^2_{2l} + n_2\sigma^2_{3l}}}\right| > \varepsilon\right)
\end{aligned}
$$

can be obtained. Thus, $P((T^{+}_{smax} - T^{*}_{smax}) > \varepsilon)$ and $P((T^{*}_{smax} - T^{+}_{smax}) > \varepsilon)$ can be controlled by the same boundary. First, we consider the upper bounds of

$$P\left(\left|\max_{1\le l\le p}\frac{\sqrt{n_1n_2}\left(\bar{X}_{1l}-\bar{X}_{2l}\right)}{\sqrt{n_2\hat{\sigma}_{1l}^2+n_1\hat{\sigma}_{2l}^2}}-\max_{1\le l\le p}\frac{\sqrt{n_1n_2}\left(\bar{X}_{1l}-\bar{X}_{2l}\right)}{\sqrt{n_2\sigma_{1l}^2+n_1\sigma_{2l}^2}}\right|>\varepsilon\right).$$

It follows that

$$\begin{aligned}
&P\left(\left|\max_{1\le l\le p}\frac{\sqrt{n_1n_2}\left(\bar{X}_{1l}-\bar{X}_{2l}\right)}{\sqrt{n_2\hat{\sigma}_{1l}^2+n_1\hat{\sigma}_{2l}^2}}-\max_{1\le l\le p}\frac{\sqrt{n_1n_2}\left(\bar{X}_{1l}-\bar{X}_{2l}\right)}{\sqrt{n_2\sigma_{1l}^2+n_1\sigma_{2l}^2}}\right|\ge\varepsilon\right)\\
&\le P\left(\left\{\max_{1\le l\le p}\left|\frac{\hat{s}_{1l}}{s_{1l}}-1\right|\sqrt{n_1}\left|\hat{\boldsymbol{D}}_{1,2}^{-1/2}(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_2)\right|_\infty\ge\varepsilon\right\}\bigcap\{\boldsymbol{A}_{n_1,n_2}(t_{12})\}\right)+P\left(\{\boldsymbol{A}^c_{n_1,n_2}(t_{12})\}\right)\\
&\le P\left(\max_{1\le l\le p}\sqrt{n_1}\left|\hat{\boldsymbol{D}}_{1,2}^{-1/2}(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_2)\right|_\infty\frac{t_{12}(1+t_{12})}{2}\ge\varepsilon\right)+P\left(\boldsymbol{A}^c_{n_1,n_2}\left(t_{12}\right)\right)\\
&\le P\left(\max_{1\le l\le p}\sqrt{n_1}\left|\hat{\boldsymbol{D}}_{1,2}^{-1/2}(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_2)\right|_\infty\ge\frac{2\varepsilon}{t_{12}(1+t_{12})}\right)+P\left(\boldsymbol{\varepsilon}^c_{n_1,n_2}\left(t_{12}\right)\right).
\end{aligned}\tag{16}$$

By Lemma 2

$$P\left(\max_{1\le l\le p}\sqrt{n_1}\left|\hat{\boldsymbol{D}}_{1,2}\left(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_2\right)\right|_\infty\ge\frac{2\varepsilon}{t_{12}(1+t_{12})}\right)\le cp\exp\left(-c\frac{4\varepsilon^2}{t_{12}^2(1+t_{12})^2}\right)+n_1^{-1};\tag{17}$$

and $P\left(\boldsymbol{\varepsilon}^c_{n_1,n_2}\left(t_{12}\right)\right)\le P\left(\max_{1\le l\le p}\left|\frac{\hat{\sigma}^2_{1l}}{\sigma^2_{1l}}-1\right|>t_{12}\right)+P\left(\max_{1\le l\le p}\left|\frac{\hat{\sigma}^2_{2l}}{\sigma^2_{2l}}-1\right|>t_{12}\right)$. Let

$$\begin{aligned}
t_{12}\asymp\max\Big\{&v_{14}^2n_1^{-\frac{1}{2}}\{\log(pn_1)\}^{\frac{1}{2}}+v_{1r}^2\theta_{n_1,p}^{\frac{2}{r+2}}+v_{1r}^2\theta_{n_1,p}^{\frac{2}{r}}\log(p),\\
&v_{24}^2n_2^{-\frac{1}{2}}\{\log(pn_2)\}^{\frac{1}{2}}+v_{2r}^2\theta_{n_2,p}^{\frac{2}{r+2}}+v_{2r}^2\theta_{n_2,p}^{\frac{2}{r}}\log(p)\Big\},
\end{aligned}$$

we can obtain that

$$P\left(\boldsymbol{\varepsilon}^c_{n_1,n_2}\left(t_{12}\right)\right)\le c\{n_1^{-1}+\theta_{n_1,p}^{\frac{2}{r+2}}\}+c\{n_2^{-1}+\theta_{n_2,p}^{\frac{2}{r+2}}\},\tag{18}$$

where $\theta_{n_1,p}=pn_1^{1-\frac{r}{2}}$ and $\theta_{n_2,p}=pn_2^{1-\frac{r}{2}}$. From (13), (16), (17) and (18), we have that

$$\begin{aligned}
&P\left(\left|\max_{1\le l\le p}\frac{\sqrt{n_1n_2}\left(\bar{X}_{1l}-\bar{X}_{2l}\right)}{\sqrt{n_2\hat{\sigma}_{1l}^2+n_1\hat{\sigma}_{2l}^2}}-\max_{1\le l\le p}\frac{\sqrt{n_1n_2}\left(\bar{X}_{1l}-\bar{X}_{2l}\right)}{\sqrt{n_2\sigma_{1l}^2+n_1\sigma_{2l}^2}}\right|\ge\varepsilon\right)\\
&\le cp\exp\left(-c\frac{4\varepsilon^2}{t_{12}^2(1+t_{12})^2}\right)+n_1^{-1}+c\left\{n_1^{-1}+\theta_{n_1,p}^{\frac{2}{r+2}}\right\}+c\left\{n_2^{-1}+\theta_{n_2,p}^{\frac{2}{r+2}}\right\}.
\end{aligned}\tag{19}$$

The bounds of terms

$$P\left(\left|\max_{1\le l\le p}\frac{\sqrt{n_1n_3}\left(\bar{X}_{1l}-\bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}_{1l}^2+n_1\hat{\sigma}_{3l}^2}}-\max_{1\le l\le p}\frac{\sqrt{n_1n_3}\left(\bar{X}_{1l}-\bar{X}_{3l}\right)}{\sqrt{n_3\sigma_{1l}^2+n_1\sigma_{3l}^2}}\right|\ge\varepsilon\right) \tag{20}$$

and

$$P\left(\left|\max_{1\le l\le p}\frac{\sqrt{n_2n_3}\left(\bar{X}_{2l}-\bar{X}_{3l}\right)}{\sqrt{n_3\hat{\sigma}_{2l}^2+n_2\hat{\sigma}_{3l}^2}}-\max_{1\le l\le p}\frac{\sqrt{n_2n_3}\left(\bar{X}_{2l}-\bar{X}_{3l}\right)}{\sqrt{n_3\sigma_{2l}^2+n_2\sigma_{3l}^2}}\right|\ge\varepsilon\right) \tag{21}$$

are analogous. Combining the results of the three parts, we obtain that

$$\begin{aligned}
&\max\{P((T^+_{smax}-T^*_{smax})>\varepsilon),P((T^*_{smax}-T^+_{smax})>\varepsilon)\}\\
&\quad\le Cp\exp\left(-c\frac{4\varepsilon^2}{t_{12}^2(1+t_{12})^2}\right)+n_1^{-1}+c\left\{n_1^{-1}+\theta_{n_1,p}^{\frac{2}{r+2}}\right\}+c\left\{n_2^{-1}+\theta_{n_2,p}^{\frac{2}{r+2}}\right\}\\
&\quad+Cp\exp\left(-c\frac{4\varepsilon^2}{t_{13}^2(1+t_{13})^2}\right)+n_1^{-1}+c\left\{n_1^{-1}+\theta_{n_1,p}^{\frac{2}{r+2}}\right\}+c\left\{n_3^{-1}+\theta_{n_3,p}^{\frac{2}{r+2}}\right\}\\
&\quad+Cp\exp\left(-c\frac{4\varepsilon^2}{t_{23}^2(1+t_{23})^2}\right)+n_2^{-1}+c\left\{n_2^{-1}+\theta_{n_2,p}^{\frac{2}{r+2}}\right\}+c\left\{n_3^{-1}+\theta_{n_3,p}^{\frac{2}{r+2}}\right\}.
\end{aligned}$$

Let $t=\max\left(t_{12},t_{13},t_{23}\right)$, we then have that

$$\begin{aligned}
&\max\{P((T^+_{smax}-T^*_{smax})>\varepsilon),P((T^*_{smax}-T^+_{smax})>\varepsilon)\}\\
&\quad\le Cp\exp\left(-c\frac{4\varepsilon^2}{t^2(1+t)^2}\right)+n_1^{-1}+c\left\{n_1^{-1}+\theta_{n_1,p}^{\frac{2}{r+2}}\right\}+n_1^{-1}\\
&\quad+c\left\{n_2^{-1}+\theta_{n_2,p}^{\frac{2}{(r+2)}}\right\}+n_2^{-1}+c\left\{n_3^{-1}+\theta_{n_3,p}^{\frac{2}{r+2}}\right\}.
\end{aligned} \tag{22}$$

Let $\varepsilon=ct(1+t)\{\log(pN)\}^{\frac{1}{2}}$, thus

$$\Delta_1\le 4c(1+t)\log(3pN).$$

By Lemma 3, we can solve for the bound on d:

$$d\le C\left(N^{-\frac{1}{8}}\{\log(3pN)\}^{\frac{7}{8}}+\theta_{N,p}^{\frac{1}{(r+1)}}\{\log(3pN)\}^{\frac{3}{2}}\right),$$

where $N = n_1 + n_2 + n_3$, $\theta_{N,p} = pN^{1-\frac{r}{2}}$. Substituting the above result into (12) gives the result of Proposition 1 (i). Under condition (A.2), the proof is similar, and thus, it is omitted. □

A.2 Proof of Theorem 1

The proof of the case of non-studentized test is similar to studentized, and thus, we only give proof of the test based on studentized statistic. Recall that

$$c_{s,\alpha} = \inf\left\{t \in \mathbb{R} : P\left(\left|\boldsymbol{W}_s\right|_\infty \geq t | \mathcal{X}_1, \dots, \mathcal{X}_K\right) \leq \alpha\right\},$$

then it follows from Proposition 1 and Lemma 2 that, under (A.1) or (A.2)

$$|P_{\boldsymbol{H}_0}\left(T_{smax} > c_{s,\alpha}\right) - P\left\{|\boldsymbol{W}_s|_\infty > c_{s,\alpha}|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right\}| \xrightarrow{p} 0,$$

as $N \to \infty$ According to Theorem 3 in Chernozhukov et al. (2015), we know the following inequalities hold

$$\alpha \geq P\left(|\boldsymbol{W}_s|_\infty > c_{s,\alpha}|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right)$$

and

$$\begin{aligned} & P\left(|\boldsymbol{W}_s|_\infty > c_{s,\alpha}|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right) \\ & \geq P\left(|\boldsymbol{W}_s|_\infty > c_{s,\alpha} - N^{-1}|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right) - P\left(c_{s,\alpha} - N^{-1} \leq |\boldsymbol{W}_s| \leq c_{s,\alpha}|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right) \\ & \geq \alpha - CN^{-1}\{\log(3p)\}^{\frac{1}{2}}. \end{aligned}$$

Thus, we have that

$$\lim_{n_1,n_2,n_3,p\to\infty} P_{\boldsymbol{H}_0}\left(T_{smax} > c_{s,\alpha}\right) = \alpha,$$

which completes the proof of Theorem 1. □

A.3 Proof of Theorem 2

For the statistic T_{nsmax}, the proof is similar, thus, we only give the proof of the test based on T_{smax}. Recall that for $1 \leq i < j \leq 3$,

$$T_{smax} = \begin{bmatrix} \hat{\boldsymbol{D}}_{12}\sqrt{n_1 n_2}\left(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2\right) \\ \hat{\boldsymbol{D}}_{13}\sqrt{n_1 n_3}\left(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_3\right) \\ \hat{\boldsymbol{D}}_{23}\sqrt{n_2 n_3}\left(\bar{\boldsymbol{X}}_2 - \bar{\boldsymbol{X}}_3\right) \end{bmatrix}_\infty,\quad \boldsymbol{\varepsilon}_{n_i,n_j}\left(t_{ij}\right) = \left\{\max_{1\leq l\leq p} |\frac{\hat{\sigma}^2_{vl}}{\sigma^2_{vl}} - 1| \leq t_{ij},\ v = i,j\right\}.$$

Let $\boldsymbol{W}_s|\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 \sim N\left(\boldsymbol{0}, \tilde{\boldsymbol{R}}_s\right)$, Borell (1975) showed that for every $u > 0$,

$$P\left\{|\boldsymbol{W}_s|_\infty \geq E\left(|\boldsymbol{W}_s|_\infty|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right) + u|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right\} \leq \exp\left(-\frac{u^2}{2}\right). \tag{23}$$

From the properties of the distribution of $|\boldsymbol{W}_s|_\infty$, the following inequalities hold

$$E\left(|\boldsymbol{W}_s|_\infty|\mathcal{X}_1,\ \mathcal{X}_2,\ \mathcal{X}_3\right) \leq (1 + \{2\log(3p)\}^{-1})\{2\log(3p)\}^{\frac{1}{2}},$$

which implies

$$c_{s,\alpha} \leq [1 + \{2\log(3p)\}^{-1}]\sqrt{2\log(3p)} + \sqrt{2\log(1/\alpha)}. \tag{24}$$

Thus the following inequalities hold:

$$\begin{aligned}
&P(T_{smax} \geq c_{s,\alpha}) \geq P(T^+_{smax} > c_{s,\alpha}) \\
&\geq P\Big(T^+_{smax} > \Big(1 + \{2\log(3p)\}^{-\frac{1}{4}} + \{2\log(3p)\}^{-\frac{1}{2}}\Big)\eta(p,\alpha), \\
&\qquad\qquad\qquad\qquad \boldsymbol{\varepsilon}_{n_1,n_2}(t_{12}), \boldsymbol{\varepsilon}_{n_1,n_3}(t_{13}), \boldsymbol{\varepsilon}_{n_2,n_3}(t_{23})\Big) \\
&\geq 1 - P\Big(T^+_{smax} \leq \Big(1 + \{2\log(3p)\}^{-\frac{1}{4}} + \{2\log(3p)\}^{-\frac{1}{2}}\Big)\eta(p,\alpha)\Big) - P\Big(\boldsymbol{\varepsilon}^c_{n_1,n_2}(t_{12})\Big) \\
&\quad - P\Big(\boldsymbol{\varepsilon}^c_{n_1,n_3}(t_{13})\Big) - P\Big(\boldsymbol{\varepsilon}^c_{n_2,n_3}(t_{23})\Big) \\
&\geq 1 - P\left(\max_{1\leq l\leq p} \frac{\sqrt{n_1n_2}|\bar{X}_{1l} - \bar{X}_{2l}|}{n_2\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{2l}} \leq \Big(1 + \{2\log(3p)\}^{-\frac{1}{4}} + \{2\log(3p)\}^{-\frac{1}{2}}\Big)\eta(p,\alpha)\right) \\
&\quad - P\left(\max_{1\leq l\leq p} \frac{\sqrt{n_1n_3}|\bar{X}_{1l} - \bar{X}_{3l}|}{n_3\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{3l}} \leq \Big(1 + \{2\log(3p)\}^{-\frac{1}{4}} + \{2\log(3p)\}^{-\frac{1}{2}}\Big)\eta(p,\alpha)\right) \\
&\quad - P\left(\max_{1\leq l\leq p} \frac{\sqrt{n_2n_3}|\bar{X}_{2l} - \bar{X}_{3l}|}{n_3\hat{\sigma}^2_{2l} + n_2\hat{\sigma}^2_{3l}} \leq (1 + \{2\log(3p)\}^{-\frac{1}{4}} + \{2\log(3p)\}^{-\frac{1}{2}})\eta(p,\alpha)\right) \\
&\quad - P\Big(\boldsymbol{\varepsilon}^c_{n_1,n_2}(t_{12})\Big) - P\Big(\boldsymbol{\varepsilon}^c_{n_1,n_3}(t_{13})\Big) - P\Big(\boldsymbol{\varepsilon}^c_{n_2,n_3}(t_{23})\Big).
\end{aligned} \tag{25}$$

Define the following equations

$$L_{ij} = \arg\max_{1\leq l\leq p} \sigma^{-1}_{lij}|\mu_{il} - \mu_{jl}|,\ \sigma^2_{lij} = \frac{\sigma^2_{il}}{n_i} + \frac{\sigma^2_{jl}}{n_j},\ \hat{\sigma}^2_{lij} = \frac{\hat{\sigma}^2_{il}}{n_i} + \frac{\hat{\sigma}^2_{jl}}{n_j}, 1 \leq i < j \leq 3.$$

For $1 \leq i < j \leq 3$, $0 < t_{ij} \leq \frac{1}{2}$, assume without loss of generality that $\mu_{iL_{ij}} - \mu_{jL_{ij}} > 0$, then on the event $\boldsymbol{\varepsilon}_{n_i,n_j}(t_{ij})$,

$$\max_{1\le l\le p}\frac{\sqrt{n_in_j}|\bar{X}_{il}-\bar{X}_{jl}|}{\left(n_j\hat{\sigma}_{il}^2+n_i\hat{\sigma}_{jl}^2\right)^{\frac{1}{2}}}\ge\frac{\left(\bar{X}_{iL_{ij}}-\mu_{iL_{ij}}\right)-\left(\bar{X}_{jL_{ij}}-\mu_{jL_{ij}}\right)}{\hat{\sigma}_{L_{ij}ij}}+\frac{\mu_{iL_{ij}}-\mu_{jL_{ij}}}{\hat{\sigma}_{L_{ij}ij}}$$
$$\ge\frac{\left(\bar{X}_{iL_{ij}}-\mu_{iL_{ij}}\right)-\left(\bar{X}_{jL_{ij}}-\mu_{jL_{ij}}\right)}{\hat{\sigma}_{L_{ij}ij}}+\left(1+\frac{1}{2}t_{ij}\right)^{-1}\frac{\mu_{iL_{ij}}-\mu_{jL_{ij}}}{\sigma_{L_{ij}ij}}.$$

Combining the above inequalities, let $(1+\{2\log(3p)\}^{-\frac{1}{4}}+\{2\log(3p)\}^{-\frac{1}{2}})\eta(p,\alpha))=C_{p,\eta}$, we have that

$$\begin{aligned}(25)\ge 1-P(&\frac{(\bar{X}_{1L_{12}}-\mu_{1L_{12}})-(\bar{X}_{2L_{12}}-\mu_{2L_{12}})}{\hat{\sigma}_{L_{12}12}}+\left(1+t_{12}/2\right)^{-1}\frac{\mu_{1L_{12}}-\mu_{2L_{12}}}{\sigma_{L_{12}12}}\le C_{p,\eta})\\
-P(&\frac{(\bar{X}_{1L_{13}}-\mu_{1L_{13}})-(\bar{X}_{3L_{13}}-\mu_{3L_{13}})}{\hat{\sigma}_{L_{13}13}}+\left(1+t_{13}/2\right)^{-1}\frac{\mu_{1L_{13}}-\mu_{3L_{13}}}{\sigma_{L_{13}13}}\le C_{p,\eta})\\
-P(&\frac{(\bar{X}_{2L_{23}}-\mu_{2L_{23}})-(\bar{X}_{3L_{23}}-\mu_{3L_{23}})}{\hat{\sigma}_{L_{23}23}}+\left(1+t_{23}/2\right)^{-1}\frac{\mu_{2L_{23}}-\mu_{3L_{23}}}{\sigma_{L_{23}23}}\le C_{p,\eta})\\
-P&\left(\varepsilon^c_{n_1,n_2}\left(t_{12}\right)\right)-P\left(\varepsilon^c_{n_1,n_3}\left(t_{13}\right)\right)-P\left(\varepsilon^c_{n_2,n_3}\left(t_{23}\right)\right).\end{aligned}\tag{26}$$

Furthermore, we choose $u_{12}=u_{12,n,p},\ u_{13}=u_{13,n,p},\ u_{23}=u_{23,n,p}$ such that

$$1+\xi_n=\left(1+t_{ij}/2\right)\left(1+\{2\log(3p)\}^{-\frac{1}{4}}+\{2\log(3p)\}^{-\frac{1}{2}}+u_{ij}\right),1\le i<j\le 3.$$

For $\xi_n>0$ such that $\xi_n\to 0$ and $\xi_n\sqrt{\log(3p)}\to\infty, 1\le i<j\le 3$, we have that

$$\frac{\max_{1\le l\le p}|\mu_{il}-\mu_{jl}|}{\left(\sigma_{il}^2/n_i+\sigma_{jl}^2/n_j\right)^{\frac{1}{2}}}\ge\left(1+t_{ij}/2\right)\left(1+\{2\log(3p)\}^{-\frac{1}{2}}+\{2\log(3p)\}^{-\frac{1}{4}}+u_{ij}\right)\eta(p,\alpha).$$

By Lemma 2, the following inequalities are obtained to deduce (26)

$$\begin{aligned}&P\left(\varepsilon^c_{n_1,n_2}\right)\left(t_{12}\right)+P\left(\varepsilon^c_{n_1,n_3}\right)\left(t_{13}\right)+P\left(\varepsilon^c_{n_2,n_3}\right)\left(t_{23}\right)\\
&\le c\left\{n_1^{-1}+\theta_{n_1,p}^{\frac{2}{r+2}}\right\}+c\left\{n_2^{-1}+\theta_{n_2,p}^{\frac{2}{r+2}}\right\}+c\left\{n_3^{-1}+\theta_{n_3,p}^{\frac{2}{r+2}}\right\},\end{aligned}$$

where

$$\begin{aligned}t_{ij}\asymp\max&\left\{v_{i4}^2n_i^{-\frac{1}{2}}\left\{\log(pn_i)\right\}^{\frac{1}{2}}+v_{ir}^2\theta_{n_i,p}^{\frac{2}{r+2}}+v_{ir}^2\theta_{n_i,p}^{\frac{2}{r}}\log(p),\right.\\
&\left.v_{j4}^2n_j^{-\frac{1}{2}}\left\{\log(pn_j)\right\}^{\frac{1}{2}}+v_{jr}^2,\theta_{n_j,p}^{\frac{2}{r+2}}+v_{jr}^2\theta_{n_j,p}^{\frac{2}{r}}\log(p)\right\},1\le i<j\le 3;\end{aligned}$$

and under condition (A.1) with $p = O\left(n^{\frac{r}{2}-1}N^{-\delta}\right)$, thus $p = O\left(n_i^{\frac{r}{2}-1-\delta}\right), i = 1,2,3,$ we can get $\theta_{n_1,p} = O\left(n_1^{-\delta}\right),\ \theta_{n_2,p} = O\left(n_2^{-\delta}\right),\ \theta_{n_3,p} = O\left(n_3^{-\delta}\right)$. Under these conditions, for $1 \le i < j \le 3$, let $t_{ij} \asymp \left\{ n_i^{-\frac{2\delta}{(r+2)}}, n_j^{-\frac{2\delta}{(r+2)}} \right\}$, we have that

$$P\left(\boldsymbol{\varepsilon}^c_{n_i,n_j}\left(t_{ij}\right)\right) \le C\left\{ n_i^{-1} + n_i^{-\frac{2\delta}{(r+2)}} \right\} + C\left\{ n_j^{-1} + n_j^{-\frac{2\delta}{(r+2)}} \right\}.$$

Summing up the above equations, we have the following inequalities

$$\begin{aligned}
(26) \ge 1 &- P\left(\frac{\left(\bar{X}_{1L_{12}} - \mu_{1L_{12}}\right) - \left(\bar{X}_{2L_{12}} - \mu_{2L_{12}}\right)}{\hat{\sigma}_{L_{12}12}} \le -u_{12}\eta(p,\alpha) \right) \\
&- P\left(\frac{\left(\bar{X}_{1L_{13}} - \mu_{1L_{13}}\right) - \left(\bar{X}_{3L_{13}} - \mu_{3L_{13}}\right)}{\hat{\sigma}_{L_{13}13}} \le -u_{13}\eta(p,\alpha) \right) \\
&- P\left(\frac{\left(\bar{X}_{2L_{23}} - \mu_{2L_{23}}\right) - \left(\bar{X}_{3L_{23}} - \mu_{3L_{23}}\right)}{\hat{\sigma}_{L_{23}23}} \le -u_{23}\eta(p,\alpha) \right) \\
&- P\left(\boldsymbol{\varepsilon}^c_{n_1,n_2}\left(t_{12}\right)\right) - P(\boldsymbol{\varepsilon}^c_{n_1,n_3}\left(t_{13}\right)) - P(\boldsymbol{\varepsilon}^c_{n_2,n_3}\left(t_{23}\right)\big) \\
\ge 1 &- C_{21}e^{-cu_{12}^2 log(p)} - C_{22}e^{-cu_{13}^2 log(p)} - C_{23}e^{-cu_{23}^2 log(p)} - n_1^{-1}n_2^{-1} - n_2^{-1}n_3^{-1} \\
&- n_1^{-1}n_3^{-1} - P\left(\boldsymbol{\varepsilon}^c_{n_1,n_2}\right)\left(t_{12}\right) - P\left(\boldsymbol{\varepsilon}^c_{n_1,n_3}\right)\left(t_{13}\right) - P\left(\boldsymbol{\varepsilon}^c_{n_2,n_3}\right)\left(t_{23}\right) \to 1.
\end{aligned}$$

Under $(A.2)$, for $1 \le i < j \le 3$, let

$$t_{ij} \asymp \max\{ n_i^{-\frac{1}{2}}\left(\log\left(pn_i\right)\right)^{\frac{1}{2}} + n_i^{-1}\left(\log\left(pn_i\right)\right)^{\frac{2}{\gamma}}, n_j^{-\frac{1}{2}}\left(\log\left(pn_j\right)\right)^{\frac{1}{2}} + n_j^{-1}\left(\log\left(pn_j\right)\right)^{\frac{2}{\gamma}} \},$$

thus

$$P\left(\boldsymbol{\varepsilon}^c_{n_1,n_2}\left(t_{12}\right)\right) + P\left(\boldsymbol{\varepsilon}^c_{n_1,n_3}\left(t_{13}\right)\right) + P\left(\boldsymbol{\varepsilon}^c_{n_2,n_3}\left(t_{23}\right)\right) \le cn_1^{-1} + cn_2^{-1} + cn_3^{-1}.$$

Following from (25) and (26), we can get the conclusion. □

A.4 Proof of Theorem 3

Define the event $\mathcal{B}_{1n} = \left\{ \hat{\boldsymbol{S}} = \{1, \dots, p\} \right\}$, $\mathcal{B}_{2n} = \{T_{smax} > (1 + \{2\log(3p)\}^{-1}) \sqrt{2\log(3p)} + \sqrt{0.0002\log(1/\alpha)}\}$ and $\boldsymbol{A} = \left\{T_{smax} > \left(1 + \{2\log(3p)\}^{-1}\right) \sqrt{2\log(3p)} + \sqrt{2\log(1/\alpha)}\right\}$. Under $\boldsymbol{H}_0$, we have that $P_{\boldsymbol{H}_0}\left(\mathcal{B}^c_{1n}\right) \le P_{\boldsymbol{H}_0}\left(\mathcal{B}_{2n}\right)$. By the fact that

$$P_{\boldsymbol{H}_0}\left(\mathcal{B}_{2n}\right) = P_{\boldsymbol{H}_0}(\mathcal{B}_{2n} \cap \boldsymbol{A}) + P_{\boldsymbol{H}_0}(\mathcal{B}_{2n} \cap \boldsymbol{A}^c),$$

and

$$P_{\boldsymbol{H}_0}\left\{\left\{T^s_{smax} > c^s_{s,\alpha}\right\}, \mathcal{B}_{1n}\right\} = 0,$$

we can obtain

$$P_{\boldsymbol{H}_0}\left\{T^s_{smax} > c^s_{s,\alpha}\right\} \le P_{\boldsymbol{H}_0}\left\{\left\{T^s_{smax} > c^s_{s,\alpha}\right\}, \mathcal{B}_{1n}\right\} + P_{\boldsymbol{H}_0}\left\{\mathcal{B}^c_{1n}\right\},$$
$$\le P_{\boldsymbol{H}_0}\left\{\left\{T^s_{smax} > c^s_{s,\alpha}\right\}, \mathcal{B}_{1n}\right\} + P_{\boldsymbol{H}_0}\left\{\mathcal{B}_{2n}\right\} \le P_{\boldsymbol{H}_0}(\mathcal{B}_{2n} \cap \boldsymbol{A}) + P_{\boldsymbol{H}_0}(\mathcal{B}_{2n} \cap \boldsymbol{A}^c).$$

Thus, we have that

$$\limsup_{n_1,n_2,n_3,p\to\infty} P_{\boldsymbol{H}_0}\left(T^s_{smax} > c^s_{s,\alpha}\right) \le \alpha + o_p(1).$$

The proof of the non-studentized analog $\boldsymbol{T}^s_{nsmax}$ can be constructed similarly, and thus, we complete the proof of Theorem 3. □

A.5 Proof of Theorem 4

Define the set $\mathcal{B}_{3n} = \left\{T_{smax} = T^s_{smax}\right\}$, and we know that

$$T^s_{smax} = \max_{l \notin \hat{S}} \left\{ \frac{\sqrt{n_1 n_2}|\bar{X}_{1l} - \bar{X}_{2l}|}{\left(n_2\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{2l}\right)}, \frac{\sqrt{n_2 n_3}|\bar{X}_{2l} - \bar{X}_{3l}|}{\left(n_3\hat{\sigma}^2_{2l} + n_2\hat{\sigma}^2_{3l}\right)}, \frac{\sqrt{n_1 n_3}|\bar{X}_{1l} - \bar{X}_{3l}|}{\left(n_3\hat{\sigma}^2_{1l} + n_1\hat{\sigma}^2_{3l}\right)} \right\}.$$

Because

$$c_{s,\alpha} \le \left(1 + \{2\log(3p)\}^{-1}\right)\sqrt{2\log(3p)} + \sqrt{2\log(1/\alpha)}$$

and $\lim_{n_1,n_2,n_3,p\to\infty} P_{\boldsymbol{H}_1}\left(T_{smax} > c_{s,\alpha}\right) = 1$, we have that

$$\lim_{n_1,n_2,n_3,p\to\infty} P_{\boldsymbol{H}_1}\left(T^s_{smax} > \left(1 + \{2\log(3p)\}^{-1}\right)\sqrt{2\log(3p)} + \sqrt{2\log(1/\alpha)}\right) = 1,$$

which implies $\lim_{n_1,n_2,n_3,p\to\infty} P_{\boldsymbol{H}_1}\left(\mathcal{B}^c_{3n}\right) = 0$. Because $\hat{\boldsymbol{S}}_{all} \subseteq \{1, \dots, 3p\}$ and the following three equations

$$P(\max_{l \notin \hat{\boldsymbol{S}}_{all}} \left\{|W_{sl}| > c_{s,\alpha}\right\}) \le P(|\boldsymbol{W}_s|_\infty > c_{s,\alpha}), \quad c_{s,\alpha} = \inf\{t \in R : P(|\boldsymbol{W}_s|_\infty \ge t) \le \alpha\},$$

$$c^s_{s,\alpha} = \inf\left\{t \in R : P(\max_{l \notin \hat{\boldsymbol{S}}_{all}} |W_{sl}| \ge t) \le \alpha\right\},$$

we have that $c^s_{s,\alpha} \leq c_{s,\alpha}$. Then, we know that

$$P_{\boldsymbol{H}_1}\left(T^s_{smax} > c^s_{s,\alpha}\right) \geq P_{\boldsymbol{H}_1}\left\{\left(T^s_{smax} > c^s_{s,\alpha}\right),\ \mathcal{B}_{3n}\right\}$$
$$\geq P_{\boldsymbol{H}_1}\left\{T_{smax} > c_{s,\alpha},\ \mathcal{B}_{3n}\right\} \geq P_{\boldsymbol{H}_1}\left\{T_{smax} > c_{s,\alpha}\right\} - P_{\boldsymbol{H}_1}\left\{\mathcal{B}^c_{3n}\right\} \to 1,$$

as $n_1, n_2, n_3, p \to \infty$. Similarly, the proof for T^s_{nsmax} is similar, and then, we complete the proof of this theorem. □

References

Aoshima, M., Yata, K. (2015). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. *Methodology and Computing in Applied Probability, 17*(2), 419–439.

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

Borell, C. (1975). The Brunn-Minkowski inequality in gauss space. *Inventiones Mathematicae, 30*(2), 207–216.

Cai, T. T., Xia, Y. (2014). High-dimensional sparse MANOVA. *Journal of Multivariate Analysis, 131*, 174–196.

Chang, J. Y., Zheng, C., Zhou, W. X., Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics, 73*(4), 1300–1310.

Chen, S. X., Qin, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics, 38*(2), 808–835.

Chernozhukov, V., Chetverikov, D., Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics, 41*(6), 2786–2189.

Chernozhukov, V., Chetverikov, D., Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields, 162*(1), 47–70.

Hu, J., Bai, Z. D., Wang, C., Wang, W. (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics, 69*(2), 365–387.

Huang, Y., Li, C. C., Li, R. Z., Yang, S. S. (2021). An overview of tests on high-dimensional means. *Journal of Multivariate Analysis, 188*, 104813.

Kong, X. L., Harrar, S. W. (2021). High-dimensional MANOVA under weak conditions. *Statistics, 55*(2), 321–349.

Pan, W., Shen, X. T. (2011). Adaptive tests for association analysis of rare variants. *Genetic Epidemiology, 35*(5), 381–388.

Runcie, D. E., Mukherjee, S. (2013). Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices. *Genetics, 194*(3), 753–767.

Schott, J. R. (2007). Some high-dimensional tests for a one-way MANOVA. *Journal of Multivariate Analysis, 98*(9), 1825–1839.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis, 100*(3), 518–532.

Tony Cai, T., Liu, W., Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(2), 349–372.

Wolen, A. R., Miles, M. F. (2012). Identifying gene networks underlying the neurobiology of ethanol and alcoholism. *Erlbaum Alcohol Research: Current Reviews, 34*(3), 306–317.

Yamada, T., Himeno, T. (2015). Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. *Journal of Multivariate Analysis, 139*, 7–27.

Zhang, J. T., Xu, J. F. (2009). On the k-sample Behrens-Fisher problem for high-dimensional data. *Science in China Series A: Mathematics, 52*(6), 1285–1304.

Zhong, P. S., Chen, S. X., Xu, M. Y. (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *The Annals of Statistics, 41*(6), 2820–2851.

Zou, Q. H., Zhu, C. Z., Yang, Y. H., Zuo, X. N., Long, X. Y., Cao, Q. J., Wang, Y. F., Zang, Y. F. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. *Journal of Neuroscience Methods, 172*(1), 137–141.