Statistics and Its Interface Volume 13 (2020) 27–36

A test on linear hypothesis of k-sample means in high-dimensional data*

Mingxiang Cao[†,¶], Peng Sun, Daojiang He[‡], Rui Wang, and Xingzhong Xu[§]

In this paper, a new test procedure is proposed to test a linear hypothesis of k-sample mean vectors in high-dimensional normal models with heteroskedasticity. The motivation is on the basis of the generalized likelihood ratio method and the Bennett transformation. The asymptotic distributions of the new test are derived under null and local alternative hypotheses under mild conditions. Simulation results show that the new test can control the nominal level reasonably and has greater power than competing tests in some cases. Moreover, numerical studies illustrate that our proposed test can also be applied to non-normal data.

AMS 2000 subject classifications: Primary 62H15; secondary 62E20.

Keywords and phrases: High-dimensional data, Linear hypothesis, k-sample, Generalized likelihood ratio method, Bennett transformation.

1. INTRODUCTION

Due to the fast development of advanced technology, high-dimensional data are appearing in more and more fields, such as computational biology, medicine, meteorology, finance and so on. For example, DNA and protein test data, weather and environmental data, social survey data, economic data and financial data are all high-dimensional. High-dimensional data are characterized by high data dimensions and relatively small sample sizes. So in high-dimensional settings, the powers or properties of classical test statistics are not ideal, or classical test statistics are not well-defined. Therefore, it is vital to establish statistical theory for high-dimensional data.

Hypothesis testing for high-dimensional means has been a hot topic in recent years. [3] obtained a test concerning a single mean vector by modifying Hotelling's T^2 test statistic. [19] proposed test statistics for the hypothesis of two-sample mean vectors by using a diagonal matrix to replace the sample covariance matrix in Hotelling's T^2 test statistic. [9], [2] and [11] all used U-statistics to construct tests for two-sample Behrens-Fisher problem. [7], [12] and [13] proposed scale-invariant tests. [22] investigated an empirical likelihood ratio test for a mean vector. [25] gave a test statistic using the idea of likelihood ratio and union-intersection testing, which was named as generalized likelihood ratio test. Their simulation results showed that their test has good power performance, especially when the variables are correlated. However, the asymptotic distributions were not obtained in their paper. Recently, [24] proposed a least favorable direction test for multivariate analysis of variance via the generalized likelihood ratio method, and obtained asymptotic distributions under the spiked and non-spiked models.

Besides the aforementioned hypothesis testing on mean vectors, it is of interest to test hypotheses concerning linear combinations of k mean vectors. Assume $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}$ are independent and identically distributed random vectors with the p-dimensional multivariate normal $\mathcal{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ distribution, where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are unknown parameters with $\boldsymbol{\Sigma}_i$ being positive definite for $i \in \{1, \ldots, k\}$. Of interest is to test the hypothesis

$$(1)\quad H_0 : \sum_{i=1}^{k} \omega_i \boldsymbol{\mu}_i = 0 \quad vs. \quad H_1 : \sum_{i=1}^{k} \omega_i \boldsymbol{\mu}_i \neq 0,$$

where $\omega_1, \ldots, \omega_k$ are known scalars with $\omega_1^2 + \cdots + \omega_k^2 \neq 0$. For the hypothesis testing in (1), [16] proposed a Dempster trace test ([10]) by Bennett transformation ([4]). [8] used the Bennett transformation to obtain three different test statistics, which are respectively similar to the test statistics in [3], [20] and [17]. [14] proposed a test for multiple linear combinations of mean vectors with unequal covariance matrices. [26] gave a test for a general linear hypothesis testing problem based on U-statistics.

The main goal of this paper is to propose a new test statistic for the linear hypothesis problem (1). Our proposed test can also be applied to the two-sample Behrens-Fisher

*The authors thank the Editor-in-Chief, Professor Yuedong Wang and two anonymous referees for their constructive comments, suggestions and detailed advice that vastly improved this article.

†Cao's research is supported by the National Natural Science Foundation of China (Nos. 11601008, 11526070) and Doctor Startup Foundation of Anhui Normal University (No. 2016XJJ101).

‡He's research is supported by the National Natural Science Foundation of China (No. 11201005).

§Xu's research is supported by the National Natural Science Foundation of China (No. 11471035).

¶Corresponding author. ORCID: 0000-0002-2041-7044.

problem. It will be shown that the new test has better behavior than some existing tests such as [16] and [26] in both size and power. We will discuss such differences between the new test and existing tests through numerical comparisons.

The organization of this paper is as follows. Section 2 introduces the generalized likelihood ratio and Bennett transformation methods. In Section 3, we present our main results and give their proofs. Simulation studies are carried out in Section 4 to compare our proposed test with some existing tests. Section 5 contains some conclusions.

2. PRELIMINARIES

For testing the hypothesis in (1), the generalized likelihood ratio method can not be directly applied because the populations have unequal covariance matrices. Since the Bennett transformation can make k-sample statistical inference problems into one-sample problems, we use the Bennett transformation for the hypothesis in (1). We first introduce the method of Bennett transformation. Without loss of generality, let n_1 be the smallest of $n_1, \dots, n_k$. Define

$$\boldsymbol{U}_j = \omega_1 \boldsymbol{X}_{1j} + \sum_{i=2}^{k} \omega_i \sqrt{\frac{n_1}{n_i}} \Big(\boldsymbol{X}_{ij} - \frac{1}{n_1} \sum_{m=1}^{n_1} \boldsymbol{X}_{im} + \frac{1}{\sqrt{n_1 n_i}} \sum_{l=1}^{n_i} \boldsymbol{X}_{il} \Big), \tag{2}$$

$j \in \{1, \dots, n_1\}$. Simple calculations show that $\boldsymbol{U}_1, \dots, \boldsymbol{U}_{n_1}$ are independent and identically distributed with the $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\mu} = \sum_{i=1}^{k} \omega_i \boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma} = \sum_{i=1}^{k} \frac{n_1 \omega_i^2}{n_i} \boldsymbol{\Sigma}_i$. Therefore, the hypothesis in (1) is equivalent to

$$H_0 : \boldsymbol{\mu} = 0 \quad vs. \quad H_1 : \boldsymbol{\mu} \neq 0. \tag{3}$$

Now we can propose test procedures for (3) based on $\boldsymbol{U}_1, \dots, \boldsymbol{U}_{n_1}$.

Likelihood-based tests often posses desirable properties; however, in high-dimensional settings, the classical likelihood ratio method is not well-defined when the data dimension p is larger than sample size $n := n_1 + \cdots + n_k$. For this case, [25] and [24] considered a generalized likelihood ratio method and proposed new tests for one-sample mean vector and MANOVA problems, respectively. Moreover, their tests outperform some existing tests in many cases. Motivated by the merits of generalized likelihood ratio method, we here use this method to test the hypothesis in (3). Next, we introduce the generalized likelihood ratio method according to [25] and [24]. By the union-intersection method, the hypothesis in (3) is equivalent to

$$H_0 : \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\mu} = 0 \quad vs. \quad H_1 : \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\mu} \neq 0 \tag{4}$$

for all p-dimensional real value vectors $\boldsymbol{\alpha}$. The likelihood function of $\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{U}_1, \dots, \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{U}_{n_1}$ is

$$L(\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\mu}, \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\alpha}) = (2\pi \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\alpha})^{-\frac{n_1}{2}} exp \left\{ - \frac{\sum_{i=1}^{n_1} (\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{U}_i - \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\mu})^2}{2 \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\alpha}} \right\}.$$

Thus we have

$$\sup_{\substack{\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\mu} \\ \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\alpha}}} L(\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\mu}, \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\alpha}) = (2e\pi)^{-\frac{n_1}{2}} \left(\frac{n_1 - 1}{n_1} \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{S} \boldsymbol{\alpha} \right)^{-\frac{n_1}{2}} =: L(\boldsymbol{\alpha}) \tag{5}$$

and

$$\sup_{\substack{\boldsymbol{\mu}=0 \\ \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\alpha}}} L(\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\mu}, \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\alpha}) = (2e\pi)^{-\frac{n_1}{2}} \Big(\frac{n_1 - 1}{n_1} \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{S} \boldsymbol{\alpha} + \boldsymbol{\alpha}^{\mathrm{T}} \overline{\boldsymbol{U}}\,\overline{\boldsymbol{U}}^{\mathrm{T}} \boldsymbol{\alpha} \Big)^{-\frac{n_1}{2}} =: L_{H_0}(\boldsymbol{\alpha}), \tag{6}$$

where $\boldsymbol{S} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (\boldsymbol{U}_i - \overline{\boldsymbol{U}})(\boldsymbol{U}_i - \overline{\boldsymbol{U}})^{\mathrm{T}}$ is the sample covariance matrix and $\overline{\boldsymbol{U}} = \frac{1}{n_1} \sum_{i=1}^{n_1} \boldsymbol{U}_i$ is the sample mean.

When $p > n_1$, the sample covariance matrix $\boldsymbol{S}$ is singular with probability one. Hence $\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{S} \boldsymbol{\alpha} = 0$ for some $\boldsymbol{\alpha}$, namely $L(\boldsymbol{\alpha}) = \infty$ for $\boldsymbol{\alpha} \in \boldsymbol{\mathcal{G}} = \{\boldsymbol{\alpha} : \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{S} \boldsymbol{\alpha} = 0\}$. If there exists a vector $\boldsymbol{\alpha}^* \in \boldsymbol{\mathcal{G}}$ to make $L_{H_0}(\boldsymbol{\alpha}^*)$ be the smallest, then based on ${\boldsymbol{\alpha}^*}^{\mathrm{T}} \boldsymbol{U}_1, \dots, {\boldsymbol{\alpha}^*}^{\mathrm{T}} \boldsymbol{U}_{n_1}$, we can construct a test statistic to achieve the largest discrepancy of distributions between the null hypothesis and the alternative hypothesis. This $\boldsymbol{\alpha}^*$ is called the least favorable direction in [24]. So the test statistic for testing the hypothesis in (3) is defined as

$$R(\boldsymbol{U}) = \min_{\substack{L(\boldsymbol{\alpha}) = \infty \\ \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha} = 1}} L_{H_0}(\boldsymbol{\alpha}),$$

where $\boldsymbol{U} = (\boldsymbol{U}_1, \dots, \boldsymbol{U}_{n_1})$. Moreover, when $R(\boldsymbol{U})$ is small enough, we reject the null hypothesis H_0.

By (5) and (6),

$$R(\boldsymbol{U}) = \min_{\substack{\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{S} \boldsymbol{\alpha} = 0 \\ \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha} = 1}} (2\pi)^{-n_1/2} \left(\boldsymbol{\alpha}^{\mathrm{T}} \overline{\boldsymbol{U}}\,\overline{\boldsymbol{U}}^{\mathrm{T}} \boldsymbol{\alpha} \right)^{-n_1/2} e^{-n_1/2},$$

which is equivalent to the following statistic

$$T(\boldsymbol{U}) = \max_{\substack{\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{S} \boldsymbol{\alpha} = 0 \\ \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha} = 1}} \boldsymbol{\alpha}^{\mathrm{T}} \overline{\boldsymbol{U}}\,\overline{\boldsymbol{U}}^{\mathrm{T}} \boldsymbol{\alpha}, \tag{7}$$

which is the test statistic we propose.

3. MAIN RESULTS AND PROOFS

Following the same steps as in [25], we get Lemma 3.1, which provides an explicit expression of our proposed test statistic $T(\boldsymbol{U})$ in (7). We omit the proof.

Lemma 3.1. *The generalized likelihood ratio test statistic for the hypothesis (1) is*

$$T(\boldsymbol{U}) = \frac{1}{\mathbf{1}_{n_1}^T(\boldsymbol{U}^T\boldsymbol{U})^{-1}\mathbf{1}_{n_1}} = \overline{\boldsymbol{U}}^T(\boldsymbol{I}_p - \boldsymbol{P_S})\overline{\boldsymbol{U}}, \tag{8}$$

where $\mathbf{1}_{n_1}$ denotes an n_1-dimensional vector whose components are all equal to 1, $\boldsymbol{I}_p$ is a $p\times p$ identity matrix and $\boldsymbol{P_S}$ is the orthogonal projection matrix onto the column space of $\boldsymbol{S}$.

Although [25] gave the explicit expression of test statistic, the asymptotic distributions were not obtained. [24] proposed a least favorable direction test and derived asymptotic distributions for the MANOVA problem. However, their theoretical results require that the number of populations is at least two, so they are not directly applicable to the one-sample case. Motivated by [24], we here give the theoretical results of the generalized likelihood ratio test statistic for one-sample testing.

It is difficult to obtain the asymptotic distribution of $T(\boldsymbol{U})$ via (8). In the following, we get an equivalent expression of $T(\boldsymbol{U})$ which is different from that in (8). Let $\boldsymbol{P}_{\mathbf{1}_{n_1}} = \boldsymbol{I}_{n_1} - \frac{1}{n_1}\mathbf{1}_{n_1}\mathbf{1}_{n_1}^{\mathrm{T}}$. Denote the spectral decomposition of $\boldsymbol{P}_{\mathbf{1}_{n_1}}$ as $\boldsymbol{Q}\begin{pmatrix}\boldsymbol{I}_{n_1-1} & 0\\ 0 & 0\end{pmatrix}\boldsymbol{Q}^{\mathrm{T}}$, where $\boldsymbol{Q} = (\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_{n_1})$ is an orthogonal matrix with $\boldsymbol{Q}_{n_1} = \frac{1}{\sqrt{n_1}}\mathbf{1}_{n_1}$. Therefore,

$$\begin{aligned}(n_1-1)\boldsymbol{S} &= \sum_{i=1}^{n_1}(\boldsymbol{U}_i - \overline{\boldsymbol{U}})(\boldsymbol{U}_i - \overline{\boldsymbol{U}})^{\mathrm{T}} = \boldsymbol{U}\boldsymbol{P}_{\mathbf{1}_{n_1}}\boldsymbol{U}^{\mathrm{T}}\\ &= \boldsymbol{U}\boldsymbol{Q}\begin{pmatrix}\boldsymbol{I}_{n_1-1} & 0\\ 0 & 0\end{pmatrix}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{Y}\begin{pmatrix}\boldsymbol{I}_{n_1-1} & 0\\ 0 & 0\end{pmatrix}\boldsymbol{Y}^{\mathrm{T}}\\ &= \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \sum_{i=1}^{n_1-1}\boldsymbol{Y}_i\boldsymbol{Y}_i^{\mathrm{T}},\end{aligned} \tag{9}$$

where $\boldsymbol{V} = (\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{n_1-1})$ and $\boldsymbol{Y}^{\mathrm{T}} = (\boldsymbol{V},\boldsymbol{Y}_{n_1})^{\mathrm{T}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}$ has the matrix normal distribution $\mathcal{N}_{n_1\times p}(\boldsymbol{Q}^{\mathrm{T}}\mathbf{1}_{n_1}\boldsymbol{\mu}^{\mathrm{T}}, \boldsymbol{I}_{n_1}\otimes\boldsymbol{\Sigma})$ where $\otimes$ denotes the Kronecker product of matrices. Hence, $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{n_1}$ are independent, $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{n_1-1}$ has the $\mathcal{N}_p(0,\boldsymbol{\Sigma})$ distribution and $\boldsymbol{Y}_{n_1}$ is distributed as $\mathcal{N}_p(\sqrt{n_1}\boldsymbol{\mu},\boldsymbol{\Sigma})$. Therefore, $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{\alpha} = 0$ is equivalent to $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V} = 0$. Furthermore, we have

$$\begin{aligned}T(\boldsymbol{U}) &= \max_{\substack{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{\alpha}=0\\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}=1}}\boldsymbol{\alpha}^{\mathrm{T}}\overline{\boldsymbol{U}}\overline{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{\alpha} = \max_{\substack{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}=0\\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}=1}}\frac{1}{n_1^2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{U}\mathbf{1}_{n_1}\mathbf{1}_{n_1}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\alpha}\\ &= \max_{\substack{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}=0\\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}=1}}\frac{1}{n_1^2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{Q}^{\mathrm{T}}\mathbf{1}_{n_1}\mathbf{1}_{n_1}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}\\ &= \max_{\substack{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}=0\\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}=1}}\frac{1}{n_1}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Y}\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}\\ &= \frac{1}{n_1}\max_{\substack{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}=0\\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}=1}}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Y}_{n_1}\boldsymbol{Y}_{n_1}^{\mathrm{T}}\boldsymbol{\alpha}.\end{aligned}$$

Lemma 3.2. *Assume $p \ge n_1 - 1$. Let $\boldsymbol{P_V}$ be the orthogonal projection matrix on the column space of $\boldsymbol{V}$. Then*

$$T(\boldsymbol{U}) = \frac{1}{n_1}\boldsymbol{Y}_{n_1}^T(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{Y}_{n_1}. \tag{10}$$

Proof. Suppose $\boldsymbol{Y}_{n_1}$ has a decomposition $\boldsymbol{Y}_{n_1} = \boldsymbol{V}\boldsymbol{a}+\boldsymbol{b}$ for a p-dimensional vector $\boldsymbol{a}$ and an (n_1-1)-dimensional vector $\boldsymbol{b}$ which is orthogonal to the columns of $\boldsymbol{V}$. Then $(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{Y}_{n_1} = (\boldsymbol{I}_p - \boldsymbol{P_V})(\boldsymbol{V}\boldsymbol{a}+\boldsymbol{b}) = \boldsymbol{b}$. Under the conditions $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V} = 0$ and $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha} = 1$,

$$\begin{aligned}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Y}_{n_1}\boldsymbol{Y}_{n_1}^{\mathrm{T}}\boldsymbol{\alpha} &= \boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{V}\boldsymbol{a}+\boldsymbol{b})(\boldsymbol{V}\boldsymbol{a}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{\alpha}\\ &= \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\alpha} \le \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}\\ &= \boldsymbol{Y}_{n_1}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{Y}_{n_1}.\end{aligned} \tag{11}$$

Moreover, the equality in (11) holds if and only if $\boldsymbol{\alpha} = \pm\frac{(\boldsymbol{I}_p-\boldsymbol{P_V})\boldsymbol{Y}_{n_1}}{\sqrt{\boldsymbol{Y}_{n_1}^{\mathrm{T}}(\boldsymbol{I}_p-\boldsymbol{P_V})\boldsymbol{Y}_{n_1}}}$. Hence, the proof of theorem is completed. □

Although there are different expressions for $T(\boldsymbol{U})$ in (8) and (10), the following lemma states that they are equivalent.

Lemma 3.3. *Assume $p \ge n_1 - 1$. Then it gets*

$$T(\boldsymbol{U}) = \frac{1}{n_1}\boldsymbol{Y}_{n_1}^T(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{Y}_{n_1} = \overline{\boldsymbol{U}}^T(\boldsymbol{I}_p - \boldsymbol{P_S})\overline{\boldsymbol{U}}.$$

Proof. Note that

$$\begin{aligned}\frac{1}{n_1}\boldsymbol{Y}_{n_1}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{Y}_{n_1} &= \frac{1}{n_1}\boldsymbol{Q}_{n_1}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{U}\boldsymbol{Q}_{n_1}\\ &= \frac{1}{n_1^2}\mathbf{1}_{n_1}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{U}\mathbf{1}_{n_1}\\ &= \overline{\boldsymbol{U}}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\overline{\boldsymbol{U}} = \overline{\boldsymbol{U}}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_S})\overline{\boldsymbol{U}}.\end{aligned}$$

The last equality holds owing to

$$\begin{aligned}\boldsymbol{P_S} &= \boldsymbol{S}\boldsymbol{S}^{+} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}})^{+}\\ &= \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{P_V},\end{aligned}$$

where $\boldsymbol{S}^{+}$ denotes the Moore-Penrose inverse of $\boldsymbol{S}$. □

In order to get the asymptotic distribution of the new test statistic, we first state five conditions, where $\lambda_1 \ge \cdots \ge \lambda_p$ are the eigenvalues of $\boldsymbol{\Sigma}$:

(A1) $\lambda_1 = o(n_1^{-1}\mathrm{tr}(\boldsymbol{\Sigma}))$.
(A2) $\lambda_1 - \lambda_p = O(n_1^{-1}\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)})$.

(A3) $n_i/n_1 \to \rho_i \geq 1$ for $i \in \{2, \ldots, k\}$ where k is fixed.
(A4) $\mathrm{tr}(\mathbf{\Sigma}^4) = o(\mathrm{tr}^2(\mathbf{\Sigma}^2))$.
(A5) $\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = O(n_1^{-1}\sqrt{\mathrm{tr}(\mathbf{\Sigma}^2)})$.

(A1) gives $n_1 = o(p)$ which means the data dimension p can be much larger than the data size n_1. (A2) indicates that the range of the eigenvalues of $\mathbf{\Sigma}$ is not too large. If all the eigenvalues of $\mathbf{\Sigma}$ are uniformly bounded, (A2) holds. (A3) implies that all sample sizes have the same growth rate except constant terms, which is a standard assumption for multi-sample asymptotic analysis. (A4) is often used in high-dimensional mean hypothesis. (A5) is called as the local alternative.

Now we consider the following example about k population covariance matrices $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_k$ which satisfy (A4). Suppose the eigenvalues $\lambda_{i1}(\mathbf{\Sigma}_i), \ldots, \lambda_{ip}(\mathbf{\Sigma}_i)$ of $\mathbf{\Sigma}_i$ obey $\lambda_{ij}(\mathbf{\Sigma}_i) = a_{ij}p^{\delta_{ij}}$ and $\lambda_{il}(\mathbf{\Sigma}_i) = c_{il}$ for $j = 1, \ldots, m_i$, $l = m_i + 1, \ldots, p$ and $i = 1, \ldots, k$, where m_i is a fixed constant, and a_{ij} and c_{il} are all positive, unknown and uniformly bounded constants as $n, p \to \infty$, such that the condition $\lambda_{i1}(\mathbf{\Sigma}_i) \geq \cdots \geq \lambda_{ip}(\mathbf{\Sigma}_i)$ holds. Here we require $\frac{1}{2} > \delta_{i1} \geq \cdots \geq \delta_{im_i} > 0$ for $i = 1, \ldots, k$. This covariance structure is called a spiked covariance model, which has been considered in literature such as [16], [6], [5], [15], [18], [23] and the references therein. It is noted that δ_{i1} in [16] was less than $\frac{1}{4}$ or $\frac{1}{8}$ in order to satisfy condition (A.2) or (A.3) of their paper, which is stronger than our condition (A4).

In order to prove our main results, the following lemmas are needed, which are respectively from Lemmas 5 and 6 in [24].

Lemma 3.4. *([24]) Let Υ^T have the matrix normal distribution $\mathcal{N}_{m\times n}(\mathbf{0}, \boldsymbol{I}_{mn})$. Then for any random variables $\xi_1, \ldots, \xi_n$ independent of Υ, we have*

$$\|\Upsilon^T \mathbf{\Omega} \Upsilon - \mathrm{tr}(\mathbf{\Omega})\boldsymbol{I}_m\| = O_P(\sqrt{m\mathrm{tr}(\mathbf{\Omega}^2)} + m \max_{1\leq i\leq n} |\xi_i|),$$

where $\|\cdot\|$ is the spectral norm of a matrix and $\mathbf{\Omega} = \mathrm{diag}(\xi_1, \cdots, \xi_n)$.

Lemma 3.5. *([24]) If $\lambda_1 = o(n_1^{-1}\mathrm{tr}(\mathbf{\Sigma}))$, then*

$$\mathrm{tr}((\boldsymbol{I}_p - \boldsymbol{P_V})\mathbf{\Sigma}) = \mathrm{tr}(\mathbf{\Sigma}) - \frac{(n_1-1)\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})} + o_P(\sqrt{\mathrm{tr}(\mathbf{\Sigma}^2)})$$

and

$$\mathrm{tr}((\boldsymbol{I}_p - \boldsymbol{P_V})\mathbf{\Sigma})^2 = \mathrm{tr}(\mathbf{\Sigma}^2)(1 + o_P(1)).$$

Theorem 3.1. *Suppose the conditions (A1)-(A4) hold. Then under the local alternative hypothesis (A5),*

$$\frac{T(\boldsymbol{U}) - n_1^{-1}(\mathrm{tr}(\mathbf{\Sigma}) - (n_1-1)\mathrm{tr}(\mathbf{\Sigma}^2)/\mathrm{tr}(\mathbf{\Sigma}))}{n_1^{-1}\sqrt{2\mathrm{tr}(\mathbf{\Sigma}^2)}} \xrightarrow{d} \mathcal{N}\left(\frac{n_1\boldsymbol{\mu}^T\boldsymbol{\mu}}{\sqrt{2\mathrm{tr}(\mathbf{\Sigma}^2)}}, 1\right), \tag{12}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Proof. Let $\boldsymbol{Q}_*\mathbf{\Lambda}\boldsymbol{Q}_*^{\mathrm{T}}$ be the spectral decomposition of $\mathbf{\Sigma}$. Define $\boldsymbol{Y}_{n_1} = \sqrt{n_1}\boldsymbol{\mu} + \boldsymbol{Q}_*\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{Z}$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ and $\boldsymbol{Z} = (z_1, \ldots, z_p)^{\mathrm{T}}$ has the $\mathcal{N}_p(0, \boldsymbol{I}_p)$ distribution. Then

$$\begin{aligned} T(\boldsymbol{U}) &= \frac{1}{n_1}\boldsymbol{Y}_{n_1}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{Y}_{n_1} \\ &= \frac{1}{n_1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}_*^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{Q}_*\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{Z} + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{\mu} \\ &\quad + \frac{2}{\sqrt{n_1}}\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{Q}_*\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{Z} =: \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned}$$

The term I can be rewritten as $\sum_{i=1}^{p} \frac{1}{n_1}\lambda_i((\boldsymbol{I}_p - \boldsymbol{P_V})\mathbf{\Sigma})z_i^2$. Conditional on $\boldsymbol{V}$, $\mathrm{Var(I)} = \frac{2}{n_1^2}\mathrm{tr}((\boldsymbol{I}_p - \boldsymbol{P_V})\mathbf{\Sigma})^2$. So the Lindeberg condition

$$\max_{1\leq i\leq p} \frac{\frac{1}{n_1^2}\lambda_i^2((\boldsymbol{I}_p - \boldsymbol{P_V})\mathbf{\Sigma})}{\mathrm{Var(I)}} \leq \frac{\lambda_1^2}{2\mathrm{tr}(\mathbf{\Sigma}^2)(1 + o_P(1))} \xrightarrow{pr} 0$$

holds by Lemma 3.5. The limit is obtained by the condition (A4) and $\xrightarrow{pr}$ denotes convergence in probability. Then

$$\frac{\mathrm{I} - \frac{1}{n_1}\mathrm{tr}((\boldsymbol{I}_p - \boldsymbol{P_V})\mathbf{\Sigma})}{\sqrt{\frac{2}{n_1^2}\mathrm{tr}(\mathbf{\Sigma}^2)}} \xrightarrow{d} \mathcal{N}(0, 1),$$

which, combined with Lemma 3.5, gives

$$\frac{\mathrm{I} - n_1^{-1}(\mathrm{tr}(\mathbf{\Sigma}) - (n_1-1)\mathrm{tr}(\mathbf{\Sigma}^2)/\mathrm{tr}(\mathbf{\Sigma}))}{n_1^{-1}\sqrt{2\mathrm{tr}(\mathbf{\Sigma}^2)}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{13}$$

For III, conditional on $\boldsymbol{V}$, $\mathrm{E(III)} = 0$ and $\mathrm{Var(III)} = \frac{4}{n_1}\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I}_p - \boldsymbol{P_V})\mathbf{\Sigma}(\boldsymbol{I}_p - \boldsymbol{P_V})\boldsymbol{\mu} \leq \frac{4}{n_1}\lambda_1\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = o(n_1^{-2}\mathrm{tr}(\mathbf{\Sigma}^2))$. Therefore

$$\frac{\mathrm{III}}{\sqrt{n_1^{-2}\mathrm{tr}(\mathbf{\Sigma}^2)}} \xrightarrow{pr} 0. \tag{14}$$

Next we consider II. By Lemma 3.4, we have

$$\lambda_{n_1-1}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}) = (1 + o_P(1))\mathrm{tr}(\mathbf{\Sigma}). \tag{15}$$

It follows from the distribution of $\boldsymbol{V}$ that $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{\mu}$ has the $\mathcal{N}_{n_1-1}(\mathbf{0}, \boldsymbol{\mu}^{\mathrm{T}}\mathbf{\Sigma}\boldsymbol{\mu}\boldsymbol{I}_{n_1-1})$ distribution, which, combined with the law of large numbers, results in

$$\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{\mu} = O_P(n_1^{-1}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{\Sigma}\boldsymbol{\mu}) = O_P(n_1^{-1}\lambda_1\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}). \tag{16}$$

By (15) and (16), we get

$$\begin{aligned} \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{P_V}\boldsymbol{\mu} &= \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{V}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{\mu} \\ &\leq \lambda_1((\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1})\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{\mu} \\ &= \lambda_{n_1-1}^{-1}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{\mu} \end{aligned} \tag{17}$$

$$= (1 + o_P(1))\mathrm{tr}^{-1}(\boldsymbol{\Sigma})O_P(n_1^{-1}\lambda_1\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu})$$
$$= o_P(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}),$$

where the last equality holds by the condition (A1). So according to (17) and the condition (A5),

$$\text{(18)} \qquad \mathrm{II} = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} - \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{P_V}\boldsymbol{\mu} = (1 + o_P(1))\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}$$
$$= O_P(n_1^{-1}\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}).$$

Lastly, the conclusion of theorem is obtained from (13), (14) and (18). □

In order to formulate a test procedure, we need to give asymptotically ratio-consistent estimators of $\mathrm{tr}(\boldsymbol{\Sigma})$ and $\mathrm{tr}(\boldsymbol{\Sigma}^2)$. Let $n_{11} := [n_1/2] + 1$ and $n_{12} := n_1 - n_{11}$, where $[x]$ is the integer part of x for $x \geq 0$, then $\overline{\boldsymbol{U}}_{n_{11}}$, $\boldsymbol{S}_{n_{11}}$ and $\overline{\boldsymbol{U}}_{n_{12}}$, $\boldsymbol{S}_{n_{12}}$ stand for the sample mean vectors and covariance matrices of the first n_{11} samples and the remaining n_{12} samples, respectively.

Lemma 3.6. *Suppose the conditions (A1), (A3) and (A4) hold. Then* $\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}})$ *and* $\mathrm{tr}(\boldsymbol{S})$ *are the asymptotically ratio-consistent estimators of* $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ *and* $\mathrm{tr}(\boldsymbol{\Sigma})$*, respectively.*

Proof. First, from $\mathrm{E}\{\mathrm{tr}(\boldsymbol{S})\} = \mathrm{tr}(\boldsymbol{\Sigma})$ and $\mathrm{Var}\{\mathrm{tr}(\boldsymbol{S})\} = \frac{2}{n_1-1}\mathrm{tr}(\boldsymbol{\Sigma}^2)$, we have $\mathrm{tr}(\boldsymbol{S}) = \mathrm{tr}(\boldsymbol{\Sigma}) + O_P(\sqrt{n_1^{-1}\mathrm{tr}(\boldsymbol{\Sigma}^2)})$.

Note that

$$\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}}) = \frac{\sum_{i=1}^{n_{11}}(\boldsymbol{U}_i - \overline{\boldsymbol{U}}_{n_{11}})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{U}_i - \overline{\boldsymbol{U}}_{n_{l1}})}{(n_{11}-1)(n_{12}-1)}$$

where $\boldsymbol{W} := \sum_{j=n_{11}+1}^{n_1}(\boldsymbol{U}_j - \overline{\boldsymbol{U}}_{n_{12}})(\boldsymbol{U}_j - \overline{\boldsymbol{U}}_{n_{12}})^{\mathrm{T}}$. It is easy to get $\mathrm{E}\{\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}})\} = \mathrm{tr}(\boldsymbol{\Sigma}^2)$ and

$$\begin{aligned}\mathrm{Var}\Big\{\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}})\Big\} &= \mathrm{Var}\Big\{\mathrm{E}[\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}})|\boldsymbol{W}]\Big\} \\ &\quad + \mathrm{E}\Big\{\mathrm{Var}[\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}})|\boldsymbol{W}]\Big\} \\ &= \frac{2\{n_1\mathrm{tr}(\boldsymbol{\Sigma}^4) + \mathrm{tr}^2(\boldsymbol{\Sigma}^2)\}}{(n_{11}-1)(n_{12}-1)}.\end{aligned}$$

Thus we have

$$\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}}) = (1 + o_P(n_1^{-\frac{1}{2}}))\mathrm{tr}(\boldsymbol{\Sigma}^2).$$

This completes the proof of Lemma 3.6. □

Remark 3.1. *The estimator* $\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}})$ *of* $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ *is based on [1], which is given in Lemma 3.1. It should be noted that the requirements for obtaining asymptotically ratio-consistent estimator of* $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ *in [1] are different from (A1), (A3) and (A4). Our assumption on* $\boldsymbol{\Sigma}$ *in (A4) is weaker than those assumptions (A-iv and A-v) in [1].*

Corollary 3.1. *Let*

$$\widehat{T} = \frac{T(\boldsymbol{U}) - n_1^{-1}(\mathrm{tr}(\boldsymbol{S}) - (n_1-1)\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}})/\mathrm{tr}(\boldsymbol{S}))}{n_1^{-1}\sqrt{2\mathrm{tr}(\boldsymbol{S}_{n_{11}}\boldsymbol{S}_{n_{12}})}}$$

be our new test. It follows from Theorem 3.1 and Lemma 3.6 that $\widehat{T} \xrightarrow{d} \mathcal{N}\left(\frac{n_1\boldsymbol{\mu}^T\boldsymbol{\mu}}{\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}}, 1\right)$ *under the local alternative (A5). It is evident that* $\widehat{T} \xrightarrow{d} \mathcal{N}(0,1)$ *under the null hypothesis. Thus, the rejection region is* $\left\{(\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}, \ldots, \boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{kn_k}) : \widehat{T} > \xi_\alpha\right\}$ *where* ξ_α *represents the upper* α *quantile of* $\mathcal{N}(0,1)$*. Furthermore, the power of* $\widehat{T}$ *is* $\beta(\widehat{T}) = \Phi\left(-\xi_\alpha + \frac{n_1\boldsymbol{\mu}^T\boldsymbol{\mu}}{\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}}\right)$ *where* $\Phi(x)$ *means the cumulative distribution function of* $\mathcal{N}(0,1)$.

Remark 3.2. *If* $k = 1$, $\omega_1 = 1$ *and* $k = 2$, $\omega_1 = -\omega_2 = 1$, *we get respectively the generalized likelihood ratio tests for one-sample problem and two-sample Behrens-Fisher problem from* $\widehat{T}$*. Moreover, it is noted that the asymptotic power* $\beta(\widehat{T}) = \Phi\left(-\xi_\alpha + \frac{n\gamma(1-\gamma)\boldsymbol{\mu}^T\boldsymbol{\mu}}{\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}_*^2)}}\right)$ *is the same as that of the test in [9], where* $\gamma = \frac{1}{1+\rho_2}$ *and* $\boldsymbol{\Sigma}_* = \gamma\boldsymbol{\Sigma}_2 + (1-\gamma)\boldsymbol{\Sigma}_1$.

Remark 3.3. *The asymptotic power of the test in [16] is the same as ours. However, in order to get the asymptotic power, the assumption (A.2) used in [16] is stronger than our assumption (A4).*

From the viewpoint of asymptotic behavior, our new test does not suffer loss of power from Remarks 3.2 and 3.3 when using the Bennett transformation. And the simulation results in Section 4 also show that our test outperforms some competing tests in some cases.

4. SIMULATION STUDIES

In this section we compare our proposed test with some existing tests by simulation. Although our new test is obtained under a normal model, we also consider the non-normal model in simulation to illustrate the robustness of our new test. We generate $\boldsymbol{X}_{ki}$, $i \in \{1, \ldots, n_k\}$ from the following three models:

- Model 1: $\boldsymbol{X}_{ki}$ is distributed as a p-variate $\mathcal{N}_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ random vector.
- Model 2: $\boldsymbol{X}_{ki}$ is distributed as a p-variate scaled mixture of normal distributions $0.7\mathcal{N}_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) + 0.3\mathcal{N}_p(\boldsymbol{\mu}_k, 17/3\boldsymbol{\Sigma}_k)$.
- Model 3: $\boldsymbol{X}_{ki}$ has the p-variate t distribution with mean $\boldsymbol{\mu}_k$, covariance matrix $\boldsymbol{\Sigma}_k$ and 35 degrees of freedom.

Let $\boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{\Lambda}_k\boldsymbol{Q}_k$ be the spectral decomposition of $\boldsymbol{\Sigma}_k$, we generate $\boldsymbol{Q}_k$ from the Harr distribution. Let $p = 50, 100, 150, 200$ and 300. Empirical sizes and powers are computed under the nominal level $\alpha = 0.05$ with 5000 replications.

All of simulation results are presented by some figures. In addition, we put our numerical simulation results in the supplement material "http://intlpress.com/site/pub/files/_supp/sii/2020/0013/0001/SII-2020-0013-0001-s003.pdf".

[16] proposed a Dempster trace test for the linear hypothesis of k-sample means with unequal covariance matrices by Bennett transformation in high-dimensional data, namely

$$T_N = \widetilde{\sigma}^{-1}\sqrt{p}\left\{\frac{n_1\overline{\boldsymbol{U}}^{\mathrm{T}}\overline{\boldsymbol{U}}}{\mathrm{tr}\boldsymbol{S}} - 1\right\},$$

where $\widetilde{\sigma}^2 = \frac{\sqrt{2\widehat{a}_2}}{\widehat{a}_1}$, $\widehat{a}_1 = \frac{1}{p}\mathrm{tr}(\boldsymbol{S})$ and $\widehat{a}_2 = \frac{(n_1-1)^2}{p(n_1-2)(n_1+1)} \times \{\mathrm{tr}(\boldsymbol{S}^2) - \frac{1}{n_1-1}\mathrm{tr}^2(\boldsymbol{S})\}$.

[26] considered the general linear hypothesis testing problem in high-dimensional data with heteroscedasticity. For hypothesis (1), their test is given by

$$T_Z = \widetilde{\sigma}_*^{-1}\Big\{\sum_{l=1}^{k}\frac{\omega_l^2}{n_l(n_l-1)}\sum_{i\neq j}^{n_l}\boldsymbol{X}_{li}^{\mathrm{T}}\boldsymbol{X}_{lj} + \sum_{l\neq s}^{k}\frac{\omega_l\omega_s}{n_l n_s}\sum_{i,j}\boldsymbol{X}_{li}^{\mathrm{T}}\boldsymbol{X}_{sj}\Big\},$$

where

$$\widetilde{\sigma}_*^2 = 2\left\{\sum_{l=1}^{k}\frac{\omega_l^2\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_l^2)}}{n_l(n_l-1)} + \sum_{l\neq s}^{k}\frac{\omega_l\omega_s\mathrm{tr}(\boldsymbol{F}_l\boldsymbol{F}_s)}{n_l n_s}\right\},$$

$$\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_l^2)} = \frac{(n_l-1)}{n_l(n_l-2)(n_l-3)}\Big\{(n_l-1)(n_l-2)\mathrm{tr}(\boldsymbol{F}_l^2) + \mathrm{tr}^2(\boldsymbol{F}_l) - \frac{n_l}{(n_l-1)}\sum_{i=1}^{n_l}\|\boldsymbol{X}_{li}-\overline{\boldsymbol{X}}\|^4\Big\},$$

and $\boldsymbol{F}_l$ is the sample covariance matrix of lth population for $l \in \{1,\ldots,k\}$.

In the simulation, we set $k=3$ and $\omega_1=\omega_2=\omega_3=1$. For $\boldsymbol{\Lambda}_i = \mathrm{diag}(\lambda_{i1},\cdots,\lambda_{ip})$, we consider the following four cases.

- Case 1: $\lambda_{1i} = p^{2/(8+i)}$ for $i \in \{1,\ldots,9\}$ and $\lambda_{1i} = 0.025$ for $i \in \{10,\ldots,p\}$. $\lambda_{2i} = p^{1/(2+i)}$ for $i \in \{1,\ldots,7\}$ and $\lambda_{2i} = 0.05$ for $i \in \{8,\ldots,p\}$. $\lambda_{3i} = p^{3/(7+i)}$ for $i \in \{1,\ldots,5\}$ and $\lambda_{3i} = 0.075$ for $i \in \{6,\ldots,p\}$.
- Case 2: $\lambda_{1i} = p^{2/(8+i)}$ for $i \in \{1,\ldots,9\}$ and $\lambda_{1i} = 0.05$ for $i \in \{10,\ldots,p\}$. $\lambda_{2i} = p^{1/(2+i)}$ for $i \in \{1,\ldots,7\}$ and $\lambda_{2i} = 0.07$ for $i \in \{8,\ldots,p\}$. $\lambda_{3i} = p^{4/(8+i)}$ for $i \in \{1,2,3\}$ and $\lambda_{3i} = 0.09$ for $i \in \{4,\ldots,p\}$.
- Case 3: $\lambda_{1i} = p^{4/(12+i)}$ for $i \in \{1,\ldots,9\}$ and $\lambda_{1i} = 0.07$ for $i \in \{10,\ldots,p\}$. $\lambda_{2i} = p^{1/(2+i)}$ for $i \in \{1,\ldots,7\}$ and $\lambda_{2i} = 0.05$ for $i \in \{8,\ldots,p\}$. $\lambda_{3i} = p^{3/(7+i)}$ for $i \in \{1,\ldots,5\}$ and $\lambda_{3i} = 0.075$ for $i \in \{6,\ldots,p\}$.
- Case 4: $\lambda_{1i} = p^{4/(12+i)}$ for $i \in \{1,\ldots,9\}$ and $\lambda_{1i} = 0.075$ for $i \in \{10,\ldots,p\}$. $\lambda_{2i} = p^{2/(8+i)}$ for $i \in \{1,\ldots,7\}$ and $\lambda_{2i} = 0.025$ for $i \in \{8,\ldots,p\}$. $\lambda_{3i} = p^{3/(7+i)}$ for $i \in \{1,\ldots,5\}$ and $\lambda_{3i} = 0.05$ for $i \in \{6,\ldots,p\}$.

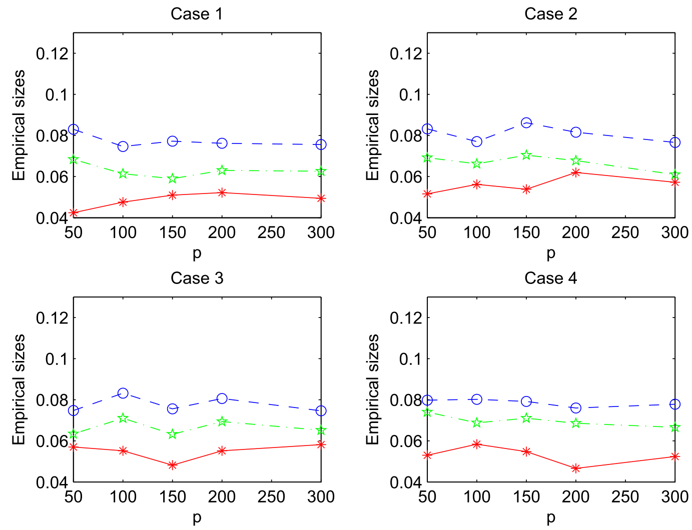

Figure 1. The Empirical sizes of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 1 when $n_1 = 15$, $n_2 = 20$.

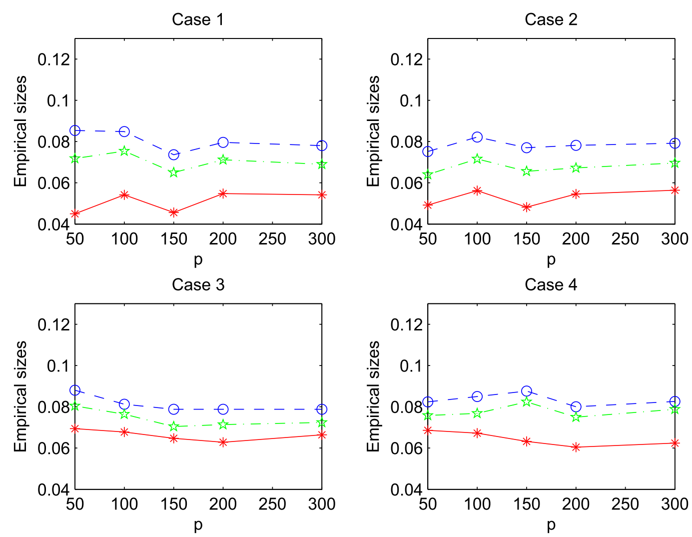

Figure 2. The Empirical sizes of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 1 when $n_1 = 15$, $n_2 = 40$.

For data sizes, we set $n_1 = 15$, $n_2 = 20$, $n_3 = 25$ and $n_1 = 15$, $n_2 = 30$, $n_3 = 55$, respectively. In the power simulation, we set $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{0}$ and $\boldsymbol{\mu}_3 = (u_1,\ldots,u_p)^{\mathrm{T}}$ where $u_{2j-1} = 0$ and $u_{2j}'s$ are i.i.d. $U(-a,a)$. a is taken as 0 and 0.5 for empirical size and power, respectively. Because there have the similar patterns with those under Case 1 and Model 1, we here only consider the case of H_1 for different config-

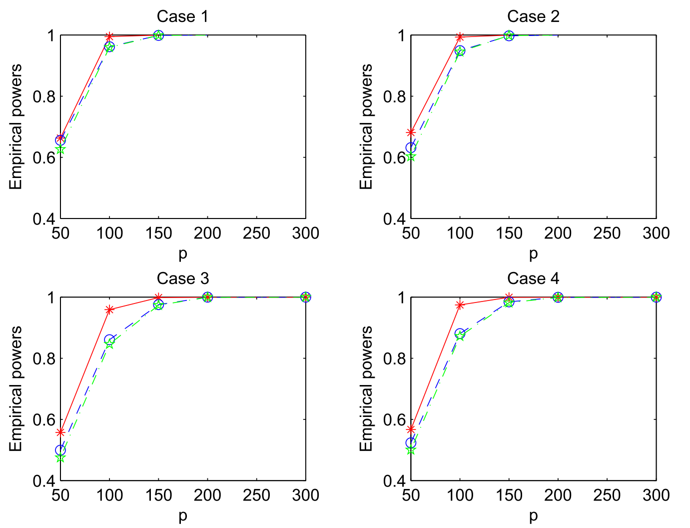

Figure 3. The Empirical powers of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 1 when $n_1 = 15$, $n_2 = 20$.

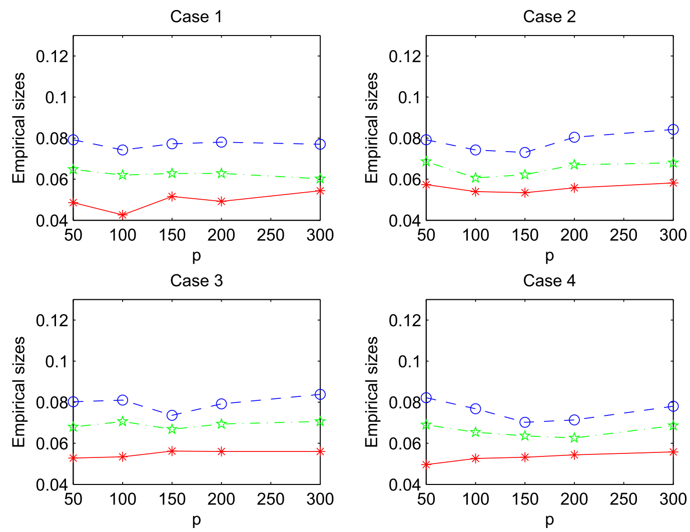

Figure 5. The Empirical sizes of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 2 when $n_1 = 15$, $n_2 = 20$.

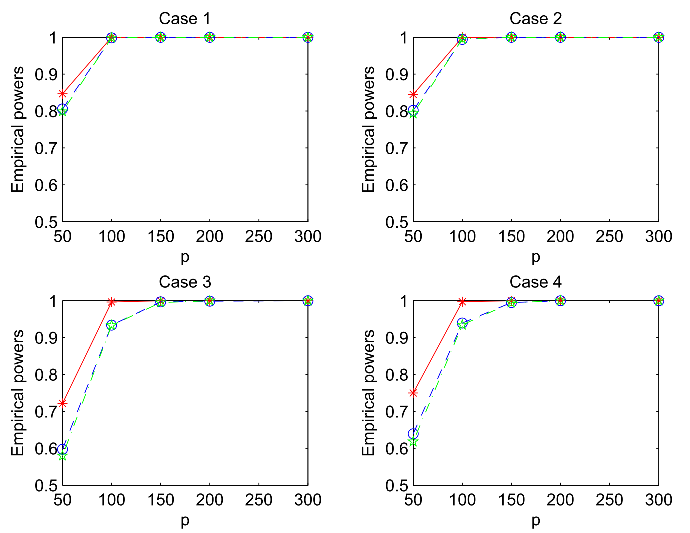

Figure 4. The Empirical powers of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 1 when $n_1 = 15$, $n_2 = 40$.

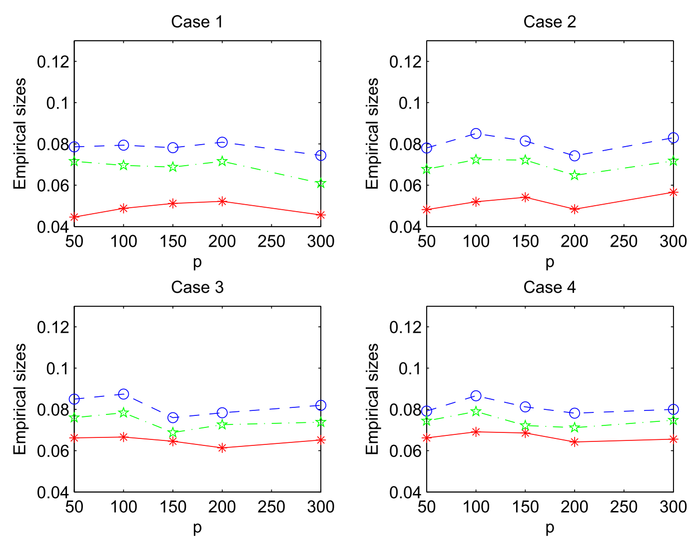

Figure 6. The Empirical sizes of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 2 when $n_1 = 15$, $n_2 = 40$.

urations of a under Case 1 and Model 1. All the simulation results are reported in Figures 1-13.

For testing hypotheses about a linear combination of means, Figures 1, 2, 5, 6, 9 and 10 show that the empirical sizes of the new test $\widehat{T}$ are around 5% and at most 6.94%. T_N and T_Z have empirical sizes around 7% and have the largest empirical sizes 8.8% and 8.24%, respectively. So $\widehat{T}$ can control the nominal size $\alpha = 0.05$ very well. And T_Z is superior to T_N in terms of controlling the nominal size. Regarding power, Figures 3, 4, 7, 8, 11 and 12 illustrate that T_N and T_Z have similar empirical powers, which are less than that of $\widehat{T}$ in all of our simulation studies. Moreover, when data sizes become larger, the empirical powers of the three tests also become larger. Figure 13 indicates that the proposed test $\widehat{T}$ has greater power than T_N and T_Z as a is about more than 0.4. However, T_N and T_Z are superior to our test as a is about less than 0.3.

In summary, our proposed test $\widehat{T}$ controls a given size reasonably and has greater power than competing tests for the linear hypothesis problem whenever samples are from the normal model (Model 1) or non-normal models (Models 2 and 3) as a tends to take a large value. Throughout the simulations, the results under Models 2 and 3 have a pattern similar to those under Model 1, which illustrate a degree of robustness of our test.

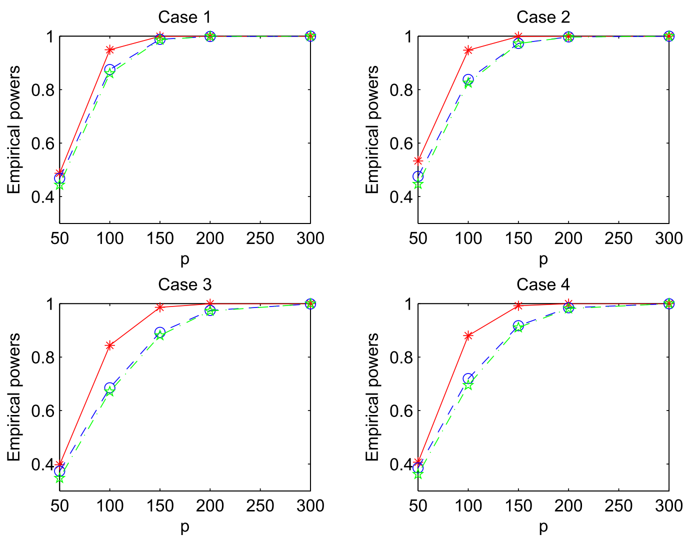

Figure 7. The Empirical powers of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 2 when $n_1 = 15$, $n_2 = 20$.

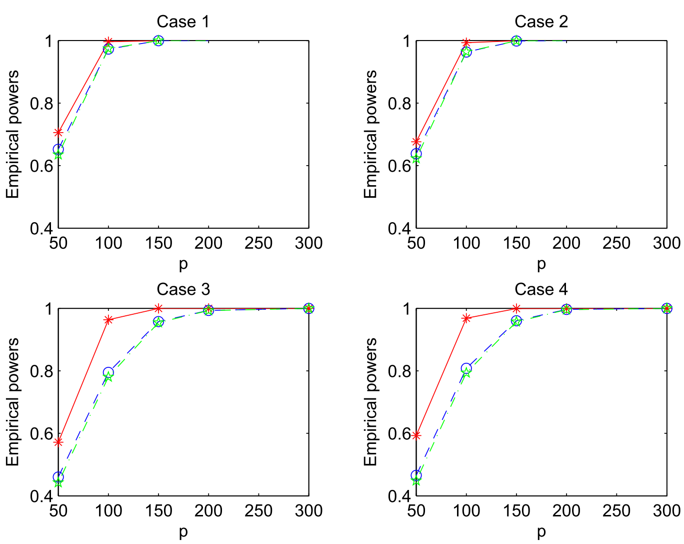

Figure 8. The Empirical powers of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 2 when $n_1 = 15$, $n_2 = 40$.

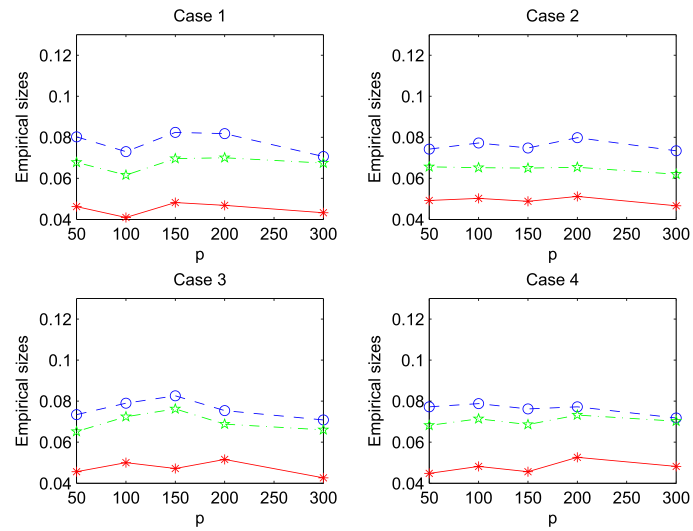

Figure 9. The Empirical sizes of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 3 when $n_1 = 15$, $n_2 = 20$.

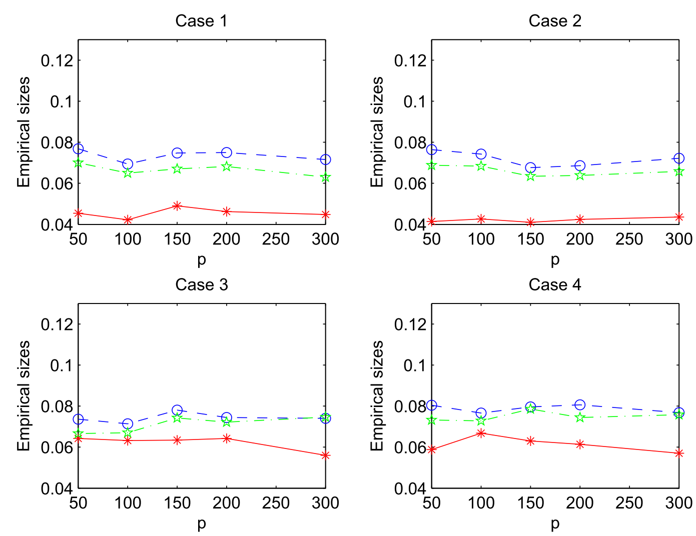

Figure 10. The Empirical sizes of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 3 when $n_1 = 15$, $n_2 = 40$.

5. CONCLUDING REMARKS

In this paper, we propose a new test for the hypotheses about a linear combination of k high-dimensional means and provide numerical studies. The new test procedure is based on the generalized likelihood ratio method and the Bennett transformation. The asymptotic distributions of the new test are obtained under the null and alternative hypotheses. Our proposed test can also be applied to the two-sample Behrens-Fisher problem. The numerical studies in this paper show that our proposed test can control the nominal size reasonably and has closer-to-nominal size and greater power than competing tests T_N and T_Z for the linear hypothesis problem in all of our simulations.

As pointed out by a reviewer, it is an important issue to investigate the asymptotic behavior of the proposed test under some special structures of covariance matrices such as a low dimensional factor model in [15] when our assumptions imposed on covariance matrix do not hold. The existed proof methods in literature may not be used, so it needs to find out some new proof methods. This is a very important problem in both theory and practice, we will leave this problem as a future study.

Received 25 November 2018

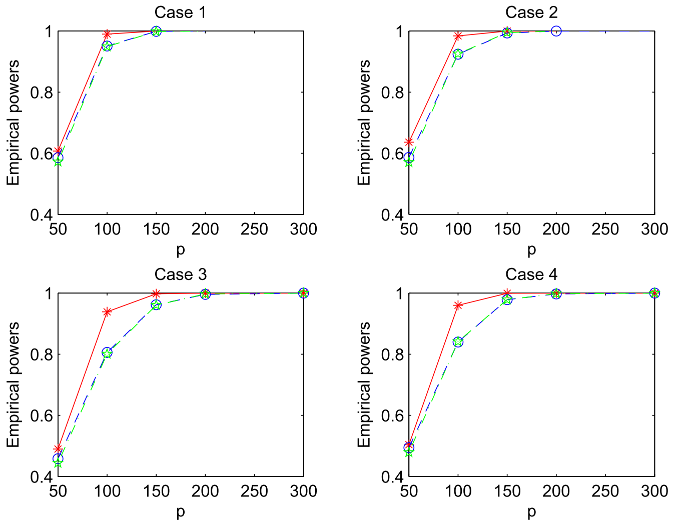

Figure 11. The Empirical powers of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 3 when $n_1 = 15$, $n_2 = 20$.

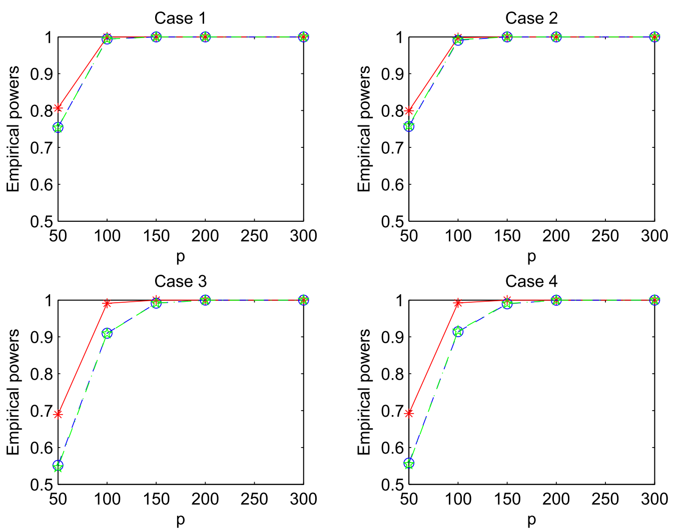

Figure 12. The Empirical powers of $\widehat{T}$ (red), T_N (blue) and T_Z (green) under Model 3 when $n_1 = 15$, $n_2 = 40$.

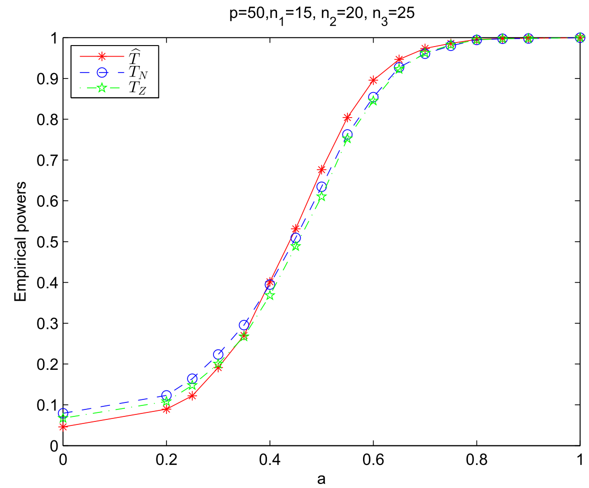

Figure 13. The Empirical powers of $\widehat{T}$, T_N and T_Z for different values of a under Case 1 and Model 1.

REFERENCES

[1] Aoshima, M. and Yata, K. (2011). Two-stage procedures for high-dimensional data. *Sequential Analysis* **30** 356–399. MR2855952

[2] Ahmad, M.R. (2014). A U-statistic approach for a high-dimensional two-sample mean testing problem under non-normality and Behrens-Fisher setting. *Annals of the Institute of Statistical Mathematics* **66** 33–61. MR3147544

[3] Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica* **6** 311–329. MR1399305

[4] Bennett, B.M. (1950). Note on a solution of the generalized Behrens-Fisher problem. *Annals of the Institute of Statistical Mathematics* **2** 87–90. MR0041399

[5] Birnbaum, A, Johnstone B., Nadler, B. and Paul, D. (2013). Minimax bounds for sparse PCA noisy high dimensional data. *The Annals of Statistics* **41** 1055–1084. MR3113803

[6] Cai, T., Ma, Z. and Wu, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *The Annals of Statistics* **41** 3074–3110. MR3161458

[7] Cai, T., Liu, W., Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society Series B* **76** 349–372. MR3164870

[8] Cao, M. (2014). Tests for mean vectors in high dimensional k-sample with unequal covariance matrices[D]. Beijing: Beijng Institute of Technology (in Chinese)

[9] Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics* **38** 808–835. MR2604697

[10] Dempster, A.P. (1958). A high dimensional two sample significant test. *Annals of Mathematical Statistics* **29** 995–1010. MR0112207

[11] Feng, L., Zou, C., et al. (2015). Two sample Behrens-Fisher problem for high-dimensional data. *Statistica Sinica* **25** 1297–1312. MR3409068

[12] Gregory, K.B., Carroll, R.J., et al. (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association* **110** 837–849. MR3367268

[13] Knavoot, J. and Samruam, C. (2015). A new test for the mean vector in high-dimensional data. *Songklanakarin Journal of Science and Technology* **37** 477–484.

[14] Li, H., Hu, J., et al. (2017). Test on the linear combinations of mean vectors in high-dimensional data. *Test* **26** 188–208. MR3613611

[15] Ma, Y., Lan, W., Wang, H. (2015). A high dimensional two-sample test under a low dimensional factor structure. *Journal of Multivariate Analysis* **140** 162–170. MR3372560

[16] Nishiyama, T., Hyodo. M., et al. (2013). Testing linear hypotheses of mean vectors for high dimension data with unequal covariance matrices. *Journal of Statistical Planning and Inference* **143** 1898–1911. MR3095080

[17] Park, J. and Ayyala, D. (2013). A test for the mean vector in large dimension and small samples. *Journal of Statistical Planning and Inference* **143** 929–943. MR3011304

[18] Passemier, D., Li, Z. and Yao, J. (2017). On estimation of the noise variance in high dimensional probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B* **79** 51–67. MR3597964

[19] Srivastava, M. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis* **99** 386–402. MR2396970
[20] Srivastava, M. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis* **100** 518–532. MR2483435
[21] Srivastava, M., Katayama, S., Kano, Y. (2013). A two sample test in high dimensional data. *Journal of Multivariate Analysis* **114** 349–358. MR2993891
[22] Wang, R., Peng, L., Qi, Y. (2013). Jackknife empirical likelihood test for equality of two high dimensional means. *Statistica Sinica* **23** 667–690. MR3086651
[23] Wang, R. and Xu, X. (2018a). On two sample mean tests under spiked covariances. *Journal of Multivariate Analysis* **167** 225–249. MR3830644
[24] Wang, R. and Xu, X. (2018b). The least favorable direction test for multivariate analysis of variance in high dimension. *Statistica Sinica*, under review.
[25] Zhao, J. and Xu, X. (2016). A generalized likelihood ratio test for normal mean when p is greater than n. *Computational Statistics and Data Analysis* **99** 91–104. MR3473084
[26] Zhou, B., Guo, J. and Zhang, J. (2017). High-dimensional general linear hypothesis testing under heteroscedasticity. *Journal of Statistical Planning and Inference* **188** 36–54. MR3648316

Mingxiang Cao
Department of Statistics
Anhui Normal University
China
E-mail address: caomingx@163.com

Peng Sun
Department of Statistics
Anhui Normal University
China
KLATASDS-MOE
School of Statistics
East China Normal University
China
E-mail address: 285794609@qq.com

Daojiang He
Department of Statistics
Anhui Normal University
China
E-mail address: djheahnu@163.com

Rui Wang
Department of Statistics
Beijing Institute of Technology
China
E-mail address: 446100240@qq.com

Xingzhong Xu
Department of Statistics
Beijing Institute of Technology
China
E-mail address: xuxz@bit.edu.cn