Ann Inst Stat Math (2017) 69:365–387
DOI 10.1007/s10463-015-0543-8

On testing the equality of high dimensional mean vectors with unequal covariance matrices

Jiang Hu[1] · **Zhidong Bai**[1] · **Chen Wang**[2] · **Wei Wang**[1]

Received: 17 September 2014 / Revised: 11 August 2015 / Published online: 8 October 2015

Abstract In this article, we focus on the problem of testing the equality of several high dimensional mean vectors with unequal covariance matrices. This is one of the most important problems in multivariate statistical analysis and there have been various tests proposed in the literature. Motivated by Bai and Saranadasa (Stat Sin 6:311–329, 1996) and Chen and Qin (Ann Stat 38:808–835, 2010), we introduce a test statistic and derive the asymptotic distributions under the null and the alternative hypothesis. In addition, it is compared with a test statistic recently proposed by Srivastava and Kubokawa (J Multivar Anal 115:204–216, 2013). It is shown that our test statistic performs better especially in the large dimensional case.

Keywords High-dimensional data · Hypothesis testing · MANOVA

✉ Jiang Hu
huj156@nenu.edu.cn

Zhidong Bai
baizd@nenu.edu.cn

Chen Wang
wangchen2351@gmail.com

Wei Wang
wangw044@nenu.edu.cn

1 KLASMOE and School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, Jilin, People's Republic of China

2 Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore

1 Introduction

In the last three decades, more and more large dimensional data sets appear in scientific research. When the dimension of data or the number of parameters becomes large, the classical methods could reduce statistical efficiency significantly. In order to analyze those large data sets, many new statistical techniques, such as large dimensional multivariate statistical analysis based on the random matrix theory, have been developed. In this article, we consider the problem of testing the equality of several high dimensional mean vectors with unequal covariance matrices, which is also called multivariate analysis of variance (MANOVA) problem. This problem is one of the most common multivariate statistical procedures in the social science, medical science, pharmaceutical science and genetics. For example, a kind of disease may have several treatments. In the past, doctors only concern which treatments can cure the disease, and the standard clinical cure is low dimension. However, nowadays researchers want to know whether the treatments alter some of the proteins or genes, thus then the high dimensional MANOVA is needed.

Suppose there are $k(k \geq 3)$ groups and $X_{i1}, \ldots, X_{in_i}$ are p-variate independent and identically distributed (i.i.d.) random samples vectors from the ith group, which have mean vector μ_i and covariance matrix Σ_i. We consider the problem of testing the hypothesis:

$$H_0 \colon \mu_1 = \cdots = \mu_k \quad \text{vs} \quad H_1 \colon \exists i \neq j, \quad \mu_i \neq \mu_j. \tag{1}$$

Notice that here we do not need normality assumption. The MANOVA problem has been discussed intensively in the literature about multivariate statistic analysis. For example, for normally distributed groups, when the total sample size $n = \sum_{i=1}^{k} n_i$ is considerably larger than the dimension p, statistics that have been commonly used are likelyhood ratio test statistic (Wilks 1932), generalized T^2 statistic (Lawley 1938; Hotelling 1947) and Pillai statistic (Pillai 1955). When p is larger than the sample size n, Dempster (1958, 1960) firstly considered this problem in the case of two sample problem. Since then, more high dimensional tests have been proposed by Bai and Saranadasa (1996), Fujikoshi et al. (2004), Srivastava and Fujikoshi (2006), Srivastava (2007), Schott (2007), Srivastava and Du (2008), Srivastava (2009), Srivastava and Yanagihara (2010), Chen and Qin (2010) and Srivastava et al. (2011, 2013). And recently, Cai and Xia (2014) proposed a statistic to test the equality of multiple high-dimensional mean vectors under common covariance matrix. Also, one can refer to the book (Fujikoshi et al. 2011) for more details.

The statistic of testing (1) we proposed in this article is motivated by Bai and Saranadasa (1996) and Chen and Qin (2010). Firstly, let us review the two test statistics briefly. For $k = 2$ and $\Sigma_1 = \Sigma_2 = \Sigma$, Bai and Saranadasa (1996) proposed the test statistic

$$T_{\text{bs}} = (\bar{X}_1 - \bar{X}_2)'(\bar{X}_1 - \bar{X}_2) - \frac{n_1 + n_2}{n_1 n_2} \text{tr} S_n, \tag{2}$$

and showed that under some conditions, as $\min\{p, n_1, n_2\} \to \infty$, $p/(n_1 + n_2) \to y > 0$ and $n_1/(n_1 + n_2) \to \kappa \in (0, 1)$

$$\frac{T_{\text{bs}} - \|\mu_1 - \mu_2\|^2}{\sqrt{\text{Var}(T_{\text{bs}})}} \xrightarrow{d} N(0, 1).$$

Here

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}, \quad S_n = \frac{1}{n_1 + n_2 - 2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)'(X_{ij} - \bar{X}_i)$$

and

$$\text{Var}(T_{\text{bs}}) = \frac{2(n_1 + n_2)^2(n_1 + n_2 - 1)}{n_1^2 n_2^2(n_1 + n_2 - 2)}\text{tr}\,\Sigma^2(1 + o(1)).$$

In addition, Bai and Saranadasa gave a ratio-consistent estimator of $\text{tr}\,\Sigma^2$ (in the sense that $\widehat{\text{tr}\,\Sigma^2}/\text{tr}\,\Sigma^2 \to 1$), that was

$$\widehat{\text{tr}\,\Sigma^2} = \frac{(n_1 + n_2 - 2)^2}{(n_1 + n_2)(n_1 + n_2 - 3)}\left(\text{tr}\,S_n^2 - \frac{1}{n_1 + n_2 - 2}(\text{tr}\,S_n)^2\right).$$

If $\Sigma_1 \neq \Sigma_2$, Chen and Qin (2010) gave a test statistic

$$T_{\text{cq}} = \frac{\sum_{i\neq j}^{n_1} X_{1i}' X_{1j}}{n_1(n_1 - 1)} + \frac{\sum_{i\neq j}^{n_2} X_{1i}' X_{2j}}{n_2(n_2 - 1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} X_{1i}' X_{2j}}{n_1 n_2},$$

which can be expressed as

$$T_{\text{cq}} = (\bar{X}_1 - \bar{X}_2)'(\bar{X}_1 - \bar{X}_2) - n_1^{-1}\text{tr}\,S_1 - n_2^{-1}\text{tr}\,S_2. \tag{3}$$

Here and throughout this paper, the sample covariance matrix of the ith group is denoted as

$$S_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)'(X_{ij} - \bar{X}_i).$$

Also they proved that under some conditions

$$\frac{T_{\text{cq}} - \|\mu_1 - \mu_2\|^2}{\sqrt{\text{Var}(T_{\text{cq}})}} \xrightarrow{d} N(0, 1)$$

where

$$\text{Var}(T_{\text{cq}}) = \left(\frac{2}{n_1(n_1 - 1)}\text{tr}(\Sigma_1^2) + \frac{2}{n_2(n_2 - 1)}\text{tr}(\Sigma_2^2) + \frac{4}{n_1 n_2}\text{tr}(\Sigma_1\Sigma_2)\right)(1 + o(1)).$$

And then Chen and Qin (2010) gave the ratio-consistent estimators of $\mathrm{tr}\Sigma_i^2$ and $\mathrm{tr}(\Sigma_1\Sigma_2)$, that were

$$\widehat{\mathrm{tr}(\Sigma_i^2)} = \frac{1}{n_i(n_i-1)}\mathrm{tr}\left(\sum_{j\neq k}^{n_i}(X_{ij}-\bar{X}_{i(j,k)})X'_{ij}(X_{ik}-\bar{X}_{i(j,k)})X'_{ik}\right) \tag{4}$$

and

$$\widehat{\mathrm{tr}(\Sigma_1\Sigma_2)} = \frac{1}{n_1 n_2}\mathrm{tr}\left(\sum_{l=1}^{n_1}\sum_{k=1}^{n_2}(X_{1l}-\bar{X}_{1(l)})X'_{1l}(X_{2k}-\bar{X}_{2(k)})X'_{2k}\right). \tag{5}$$

Here $\bar{X}_{i(j,k)}$ is the ith sample mean after excluding X_{ij} and X_{ik}, and $\bar{X}_{i(l)}$ is the ith sample mean without X_{il}.

When $\Sigma_1 = \Sigma_2$, it is apparent that the test statistic proposed by Chen and Qin (2010) reduces to the one obtained by Bai and Saranadasa (1996). Compared to Bai and Saranadasa (1996) and Chen and Qin (2010) generalized the test to the case when $\Sigma_1 \neq \Sigma_2$, and used different estimators of the variance. This is indeed a significant improvement to remove the assumption $\Sigma_1 = \Sigma_2$, because such an assumption is hard to verify for high-dimensional data. Thus based on these properties, we propose a statistic of testing the equality of more than two high dimensional mean vectors with unequal covariance matrices.

We assume the following general multivariate model:

(a) $X_{ij} = \Gamma_i Z_{ij} + \mu_i$, for $i = 1, \ldots k$, $j = 1 \ldots, n_i$, where Γ_i is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_i\Gamma'_i = \Sigma_i$, and $\{Z_{ij}\}_{j=1}^{n_i}$ are m-variate i.i.d. random vectors satisfying $E(Z_{ij}) = 0$ and $\mathrm{Var}(Z_{ij}) = I_m$, the $m \times m$ identity matrix;
(b) $Z_{ij} = (z_{ij1}, \ldots, z_{ijm})'$, with $E(z_{ijl_1}^{\alpha_1} z_{ijl_2}^{\alpha_2} \cdots z_{ijl_q}^{\alpha_q}) = E(z_{ijl_1}^{\alpha_1})E(z_{ijl_2}^{\alpha_2})\ldots E(z_{ijl_q}^{\alpha_q})$ and $E(z_{ijk}^4) < \infty$, for a positive integer q such that $\sum_{l=1}^{q}\alpha_l \leq 8$ and $l_1 \neq l_2 \neq \cdots \neq l_q$;
(c) $\frac{n_i}{n} \to k_i \in (0,1) i = 1, \ldots k$, as $n \to \infty$. Here $n = \sum_{i=1}^{k} n_i$;
(d) $\mathrm{tr}(\Sigma_l\Sigma_d\Sigma_l\Sigma_h) = o[\mathrm{tr}(\Sigma_l\Sigma_d)\mathrm{tr}(\Sigma_l\Sigma_h)]$, $d, l, h \in \{1, 2, \ldots, k\}$;
(e) $(\mu_d - \mu_l)'\Sigma_d(\mu_d - \mu_h) = o[n^{-1}\mathrm{tr}\{(\sum_{i=1}^{k}\Sigma_i)^2\}]$, $d, l, h \in \{1, 2, \ldots, k\}$.

It should be noted that all random variables and parameters here and later depend on n. For simplicity we omit the subscript n from all random variables except those statistics defined later.

Now we construct our test. Consider the statistic

$$\begin{aligned} T_n^{(k)} &= \sum_{i<j}^{k}(\bar{X}_i - \bar{X}_j)'(\bar{X}_i - \bar{X}_j) - (k-1)\sum_{i=1}^{k} n_i^{-1}\mathrm{tr}S_i \\ &= (k-1)\sum_{i=1}^{k}\frac{1}{n_i(n_i-1)}\sum_{k_1\neq k_2} X'_{ik_1}X_{ik_2} - \sum_{i<j}^{k}\frac{2}{n_i n_j}\sum_{k_1,k_2} X'_{ik_1}X_{jk_2}. \end{aligned}$$

When $k=2$, apparently $T_n^{(2)}$ is the Chen–Qin test statistic. Next we will calculate the mean and variance of $T_n^{(k)}$. Unlike the method used in Chen and Qin (2010), we give a much simpler procedure. From $X_{ij}=\Gamma_i Z_{ij}+\mu_i$, we can rewrite $T_n^{(k)}-\sum_{i<j}\|\mu_i-\mu_j\|^2$ as $T_1^{(k)}+T_2^{(k)}$, where

$$T_1^{(k)}=(k-1)\sum_{i=1}^{k}\frac{1}{n_i(n_i-1)}\sum_{k_1\neq k_2}Z'_{ik_1}\Gamma'_i\Gamma_i Z_{ik_2}-\sum_{i<j}^{k}\frac{2}{n_in_j}\sum_{k_1,k_2}Z'_{ik_1}\Gamma'_i\Gamma_j Z_{jk_2}$$

$$T_2^{(k)}=\sum_{i=1}^{k}\frac{2}{n_i}(k\mu_i-\sum_{j=1}^{k}\mu_j)'\sum_{k}\Gamma_i Z_{ik_1}.$$

Thus we can show immediately that

$$E(T_n^{(k)})=\sum_{i<j}^{k}\|\mu_i-\mu_j\|^2$$

and

$$\mathrm{Var}(T_n^{(k)})=\sum_{i=1}^{k}\frac{2(k-1)^2}{n_i(n_i-1)}\mathrm{tr}(\Sigma_i^2)+\sum_{i<j}^{k}\frac{4}{n_in_j}\mathrm{tr}(\Sigma_i\Sigma_j)$$
$$+4\sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{k}\mu_j-k\mu_i\right)'\Sigma_i\left(\sum_{j=1}^{k}\mu_j-k\mu_i\right).$$

Then we have the following theorem:

Theorem 1 *Under the assumptions* (*a*)–(*e*)*, we obtain that as* $p\to\infty$ *and* $n\to\infty$*,*

$$\frac{T_n^{(k)}-\sum_{i<j}^{k}\|\mu_i-\mu_j\|^2}{\sqrt{\mathrm{Var}(T_n^{(k)})}}\xrightarrow{d}N(0,1). \qquad (6)$$

It is worth noting that under H_0, assumption (e) is trivially satisfied and $E(T_n^{(k)})=0$. What is more, under H_1 and assumptions (a)–(e), $\mathrm{Var}(T_n^{(k)})=(\sigma_n^{(k)})^2(1+o(1))$, where

$$(\sigma_n^{(k)})^2=\sum_{i=1}^{k}\frac{2(k-1)^2}{n_i(n_i-1)}\mathrm{tr}(\Sigma_i^2)+\sum_{i<j}^{k}\frac{4}{n_in_j}\mathrm{tr}(\Sigma_i\Sigma_j).$$

Then Theorem 1 is still true if the denominator of (6) is replaced by $\sigma_n^{(k)}$. Therefore, to complete the construction of our test statistic, we only need to find a ratio-consistent estimator of $(\sigma_n^{(k)})^2$ and substitute it into the denominator of (6). There are many estimators for $(\sigma_n^{(k)})^2$, and in this paper we choose two of them:

Lemma 2 [Uniformly minimum variance unbiased estimators (UMVUE)] *Under the assumptions* (*a*)–(*d*)*, we obtain that as* $p \to \infty$ *and* $n \to \infty$*,*

$$\frac{\widehat{\operatorname{tr}\Sigma_i^2}}{\operatorname{tr}\Sigma_i^2} \xrightarrow{p} 1 \quad \textit{and} \quad \frac{\operatorname{tr}(\widehat{\Sigma_i\Sigma_j})}{\operatorname{tr}(\Sigma_i\Sigma_j)} \xrightarrow{p} 1$$

where $i \neq j \in \{1, 2, \ldots, k\}$*,*

$$\widehat{\operatorname{tr}\Sigma_i^2} = \frac{(n_i-1)^2}{(n_i+1)(n_i-2)}\left(\operatorname{tr}S_i^2 - \frac{1}{n_i-1}\operatorname{tr}^2 S_i\right) \tag{7}$$

and

$$\operatorname{tr}(\widehat{\Sigma_i\Sigma_j}) = \operatorname{tr}S_iS_j. \tag{8}$$

Remark 3 Under the normality assumption (7) and (8) are uniformly minimum variance unbiased estimators. The proof of this lemma was given in Bai and Saranadasa (1996) and Srivastava (2009), and we omit it in this paper.

Lemma 4 [Unbiased nonparametric estimators (UNE)] *Under the assumptions* (*a*)–(*d*)*, we obtain that as* $p \to \infty$ *and* $n \to \infty$*,*

$$\frac{\widehat{\operatorname{tr}\Sigma_i^2}}{\operatorname{tr}\Sigma_i^2} \xrightarrow{p} 1 \quad \textit{and} \quad \frac{\operatorname{tr}(\widehat{\Sigma_i\Sigma_j})}{\operatorname{tr}(\Sigma_i\Sigma_j)} \xrightarrow{p} 1$$

where $i \neq j \in \{1, 2, \ldots, k\}$*,*

$$\begin{aligned}\widehat{\operatorname{tr}\Sigma_i^2} &= \frac{1}{(n_i)_6} \\ &\quad\times \sum_{\substack{k_1,\ldots,k_6\\ \text{distinct}}} (X_{ik_1}-X_{ik_2})'(X_{ik_3}-X_{ik_4})(X_{ik_3}-X_{ik_5})'(X_{ik_1}-X_{ik_6})\end{aligned} \tag{9}$$

and

$$\begin{aligned}\operatorname{tr}(\widehat{\Sigma_i\Sigma_j}) &= \frac{1}{(n_i)_3(n_j)_3} \\ &\quad\times \sum_{\substack{k_1,k_2,k_3 \text{ distict}\\ k_4k_5,k_6 \text{ distinct}}} (X_{ik_1}-X_{ik_2})'(X_{jk_4}-X_{jk_5})(X_{jk_4}-X_{jk_6})'(X_{ik_1}-X_{ik_3}).\end{aligned} \tag{10}$$

Here $(n)_l = n(n-1)\ldots(n-l+1)$.

Remark 5 By assumption (a), the unbiasedness of estimators $\widehat{\mathrm{tr}\Sigma_i^2}$ and $\widehat{\mathrm{tr}\Sigma_i\Sigma_j}$ can be easily proved and their ratio-consistency can be found in Li and Chen (2012).

Remark 6 Li and Chen (2012) mentioned that the computation of the estimators in Lemma 4 would be extremely heavy if the sample sizes are very large. Thus to increase the computation speed, we simplify the estimator (9) further to:

$$\widehat{\mathrm{tr}\Sigma_i^2} = \frac{1}{n_i(n_i-3)}\|\Theta_i\|_2^2 - \frac{2}{n_i(n_i-2)(n_i-3)}\|\Theta_i\|_{1,2}^2 + \frac{1}{n_i(n_i-1)(n_i-2)(n_i-3)}(\|\Theta_i\|_1)^2$$

where $\Theta_i = X_i'X_i - \mathrm{Diag}[X_i'X_i]$, $X_i = (X_{i1}, \dots X_{in_i})_{p\times n_i}$ and $\mathrm{Diag}[X_i'X_i]$ is a diagonal matrix consisting of the diagonal elements of $X_i'X_i$. Notice that for any matrix $A = (a_{ij})_{m\times n}$, the norm $\|\cdot\|_q$ is entrywise norm, i.e., $\|A\|_q = (\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^q)^{1/q}$ and the norm $\|\cdot\|_{p,q}$ is $L_{p,q}$ norm, i.e.,

$$\|A\|_{p,q} = \left(\sum_{i=1}^m \left(\sum_{j=1}^n |a_{ij}|^p\right)^{q/p}\right)^{1/q}.$$

What is more, from a direct calculation we can show that the estimator (10) is exactly equal to the estimator (8) in Lemma 2. That is because,

$$\begin{aligned}(10) = & \frac{1}{(n_i-1)(n_j-1)}\sum_{k_1,k_4}(X_{ik_1}'X_{jk_4})^2 - \frac{1}{(n_i-1)n_j(n_j-1)}\sum_{k_1}^{n_i}\left(\sum_{k_4}^{n_j}X_{ik_1}'X_{jk_4}\right)^2 \\ & -\frac{1}{n_i(n_i-1)(n_j-1)}\sum_{k_4}^{n_j}\left(\sum_{k_1}^{n_i}X_{ik_1}'X_{jk_4}\right)^2 \\ & +\frac{1}{n_i(n_i-1)n_j(n_j-1)}\left(\sum_{k_1,k_4}X_{ik_1}{}'X_{jk_4}\right)^2 \\ = & \frac{1}{(n_i-1)(n_j-1)}\mathrm{tr}X_iX_i'X_jX_j' - \frac{n_j}{(n_i-1)(n_j-1)}\mathrm{tr}\bar{X}_i\bar{X}_i'X_jX_j' \\ & -\frac{n_i}{(n_i-1)(n_j-1)} + \frac{n_in_j}{(n_i-1)(n_j-1)}\mathrm{tr}\bar{X}_i\bar{X}_i'\bar{X}_j\bar{X}_j' = \mathrm{tr}S_iS_j.\end{aligned}$$

Apparently, using the simplified formulas instead of the original ones can make the computation much faster.

Now, by combining Theorem 1 and Lemma 2 (or Lemma 4), we obtain our test statistic under H_0 and have the following theorem:

Theorem 7 *Under H_0 and the assumptions* (a)–(d), *we obtain that as $p \to \infty$ and $n \to \infty$,*

$$T_{\text{our}} = T_n^{(k)}/\hat{\sigma}_n^{(k)} \xrightarrow{d} N(0,1),$$

where $(\hat{\sigma}_n^{(k)})^2 = \sum_{i=1}^k \frac{8}{n_i(n_i-1)}\widehat{\mathrm{tr}(\Sigma_i^2)} + \sum_{i<j}^k \frac{4}{n_i n_j}\widehat{\mathrm{tr}(\Sigma_i\Sigma_j)}$ *with* $\widehat{\mathrm{tr}(\Sigma_i^2)}$ *and* $\widehat{\mathrm{tr}(\Sigma_i\Sigma_j)}$ *given in Lemma* 2 *or Lemma* 4.

Remark 8 When the number of groups k is small, the hypothesis H_0 can be considered as a multiple hypothesis of testing each two sample. And the test for each sub-hypothesis can be tested by Chen and Qin (2010). However, for each sub-hypothesis, there is a test statistic of Chen and Qin (2010). The problem is how do we set up the critical value for the simultaneous test of the compound hypothesis H_0. In the literature, there is a famous Bonferroni correction method can be used. But it is well known that Bonfferoni correction is much conservative. Form this theorem, we can see that using our test, one may set up an asymptotically exact test.

Due to Theorem 7, the test with an α level of significance rejects H_0 if $T_{\text{our}} > \xi_\alpha$ where ξ_α is the upper α quantile of $N(0,1)$. Next we will discuss the power properties of the proposed test. Denote $\|\mu\| = \sum_{i<j}^k \|\mu_i - \mu_j\|^2$. From the above conclusions, we can easily obtain that

$$T_{\text{our}} - \frac{\|\mu\|}{\sqrt{\mathrm{Var}(T_n^{(k)})}} \xrightarrow{d} N(0,1).$$

This implies

$$\beta_{nT}(\|\mu\|) = P_{H_1}(T_{\text{our}} > \xi_\alpha) = \Phi\left(-\xi_\alpha + \frac{\|\mu\|}{\sigma_n^{(k)}}\right) + o(1),$$

where Φ is the standard normal distribution function.

2 Other tests and simulations

Due to the fact that the commonly used likelihood ratio test performs badly when dimension is large has been considered in a lot of literature such as Bai and Saranadasa (1996), Bai et al. (2009), Jiang et al. (2012) and Jiang and Yang (2013), the discussion of the likelihood ratio test is left out in this paper. Recently, Srivastava and Kubokawa (2013) proposed a test statistic of testing the equality of mean vectors of several groups with a common unknown non-singular covariance matrix. Denote $\mathbf{1}_r = (1,\ldots,1)'$ as an r-vector with all the elements equal to one and define $Y = (X_{11},\ldots,X_{1n_1},\ldots,X_{k1},\ldots,X_{kn_k})$, $L = (I_{k-1}, -\mathbf{1}_{k-1})_{(k-1)\times k}$ and

$$E = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_{n_k} \end{pmatrix}_{n\times k}.$$

Then it is proposed that

$$T_{\mathrm{sk}} = \frac{\mathrm{tr}(BD_S^{-1} - (n-k)p(k-1)(n-k-2)^{-1})}{\sqrt{2c_{p,n}(k-1)(\mathrm{tr}R^2 - (n-k)^{-1}p^2)}},$$

where $B = Y'E(E'E)^{-1}L'[L(E'E)^{-1}L']^{-1}L(E'E)^{-1}E'Y$, $D_S = \mathrm{Diag}[(n-k)^{-1}Y(I_n - E(E'E)^{-1}E')Y]$, $R = D_S^{-1/2}Y(I_n - E(E'E)^{-1}E')YD_S^{-1/2}$ and $c_{p,n} = 1 + \mathrm{tr}(R^2)/p^{3/2}$. Notice that Diag$[A]$ denotes a diagonal matrix with the same diagonal elements as the diagonal elements of matrix A. Under the null hypothesis and the condition $n = O(p^\delta)$ with $\delta > 1/2$, T_{sk} is asymptotically distributed as $N(0, 1)$. That is as $n, p \to \infty$,

$$P_{H_0}(T_{\mathrm{sk}} > \xi_\alpha) \to \Phi(-\xi_\alpha).$$

In this section we compare the performance of the proposed statistics T_{our} and T_{sk} in finite samples by simulation. Notice that the data is generated from the model

$$X_{ij} = \Gamma_i Z_{ij} + \mu_i, \quad i = 1, \ldots, k, \; j = 1, \ldots, n_i$$

where Γ_i is a $p \times p$ such that $\Gamma_i^2 = \Sigma_i$. Here $Z_{ij} = (z_{ij1}, \ldots, z_{ijp})'$ and z_{ijk}'s are independent random variables which are distributed as one of the following three distributions:

$$\text{(i) } N(0, 1), \quad \text{(ii) } (\chi_2^2 - 2)/2, \quad \text{(iii) } (\chi_8^2 - 8)/4.$$

For the covariance matrix Σ_i, $i = 1, 2, 3$, we consider the following two cases:

Case 1 $\Sigma_i = \Gamma_i = I_p$;
Case 2 $\Sigma_i = \Gamma_i^2 = W_i \Psi_i W_i$, $W_i = \mathrm{Diag}[w_{i1}, \ldots, w_{ip}]$, $w_{ij} = 2*i + (p - j + 1)/p$, $\Psi_i = (\phi_{jk}^{(i)})$, $\phi_{jj}^{(i)} = 1$, $\phi_{jk}^{(i)} = (-1)^{j+k}(0.2 \times i)^{|j-k|^{0.1}}$, $j \neq k$.

We first compare the convergence rates of the estimators (7) and (9) based on the above models, see Figs. 1 and 2. Here the dimension $p = 100$ and the sample sizes n are from 10 to 1000. The results are based on 1000 replications. From these two figures we can easily find that in both cases, the UNE (9) and UMVUE (7) are almost the same if the data sets come from standard normal distribution. But UNE is much better than UMVUE if the data sets come from χ^2 distribution, especially when n is small.

Next let us see the performance of the estimator $\widehat{\mathrm{tr}\Sigma_i \Sigma_j} = \mathrm{tr}S_i S_j$ in Case 1 and Case 2 (see Figs. 3, 4). Also the dimension $p = 100$ and the sample sizes $n_1 = n_2$ are from 10 to 1000. The results are based on 1000 replications. both cases, the estimator $\widehat{\mathrm{tr}\Sigma_i \Sigma_j} = \mathrm{tr}S_i S_j$ performs very well at all the three distributions. Thus when the sample size n is large, we can safely use these estimators in the applications.

Now we examine the attained significance level (ASL) of the test statistics T_{our} and T_{sk} compared to the nominal value $\alpha = 0.05$, and then examine their attained power.

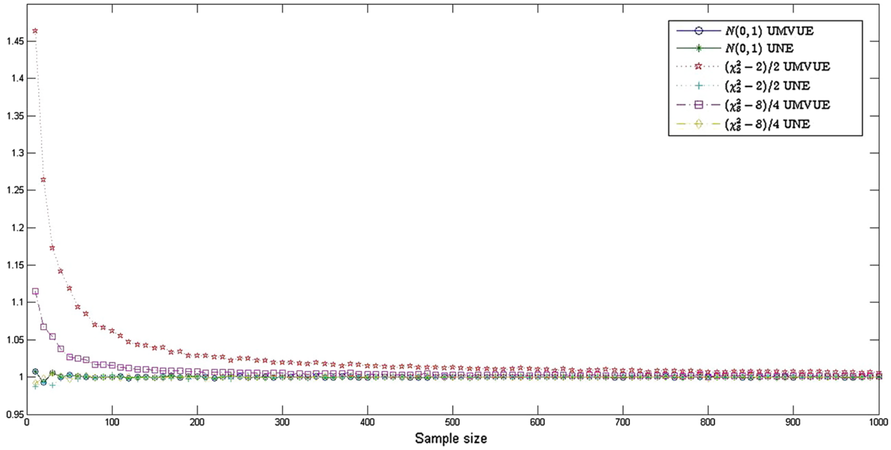

Fig. 1 Graph of the ratio between the estimators $\frac{1}{p}\widehat{\mathrm{tr}\Sigma_1^2}$ and the true value $\frac{1}{p}\mathrm{tr}\Sigma_1^2$ in Case 1, i.e., $\Sigma_1 = I_p$

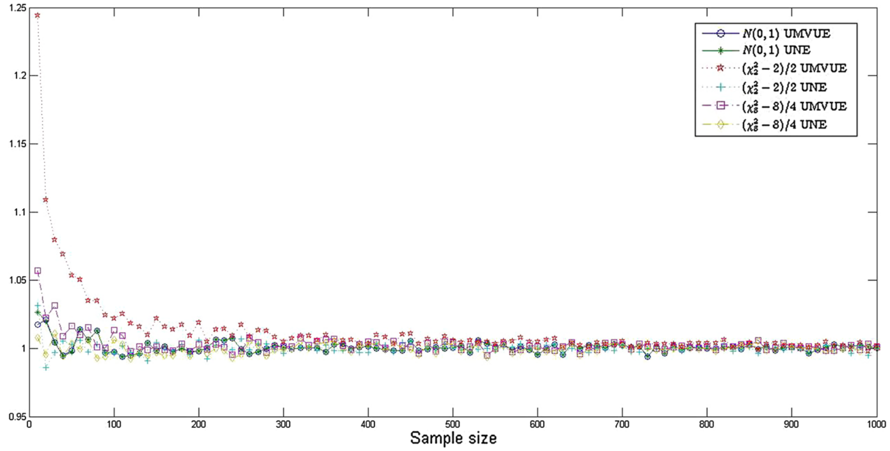

Fig. 2 Graph of the ratio between the estimators $\frac{1}{p}\widehat{\mathrm{tr}\Sigma_1^2}$ and the true value $\frac{1}{p}\mathrm{tr}\Sigma_1^2$ in Case 2, i.e., $\Sigma_1 = W_1\Psi_1W_1$

The ASL is computed as $\hat{\alpha} = \#(T > \xi_{1-\alpha})/r$ where T are values of the test statistic T_{our} or T_{sk} obtained from data simulated under H_0, r is the number of replications and $\xi_{1-\alpha}$ is the $100(1-\alpha)$ % quantile of the standard normal distribution. The attained power of the test T_{our} and T_{sk} is also computed as $\hat{\beta} = \#(T > \xi_{1-\alpha})/r$, where T are values of the test statistic T_{our} or T_{sk} computed from data simulated under the alternative.

For simulation, we consider the problem of testing the equality of three mean vectors, that is, $k = 3$. Choose $p \in \{20, 50, 100, 500, 800\}$, $n_1 = 0.5 \times n^*$, $n_2 = n^*$, $n_3 = 1.5 \times n^*$, where $n^* \in \{20, 50, 100, 200\}$. For the null hypothesis, without loss of generality we choose $\mu_1 = \mu_2 = \mu_3 = \mathbf{0}$. For the alternative hypothesis, we choose $\mu_1 = \mathbf{0}$, $\mu_2 = (u_1, \ldots, u_p)'$ and $\mu_3 = -\mu_2$, where $u_i = (-1)^i v_i$ with v_i are i.i.d.

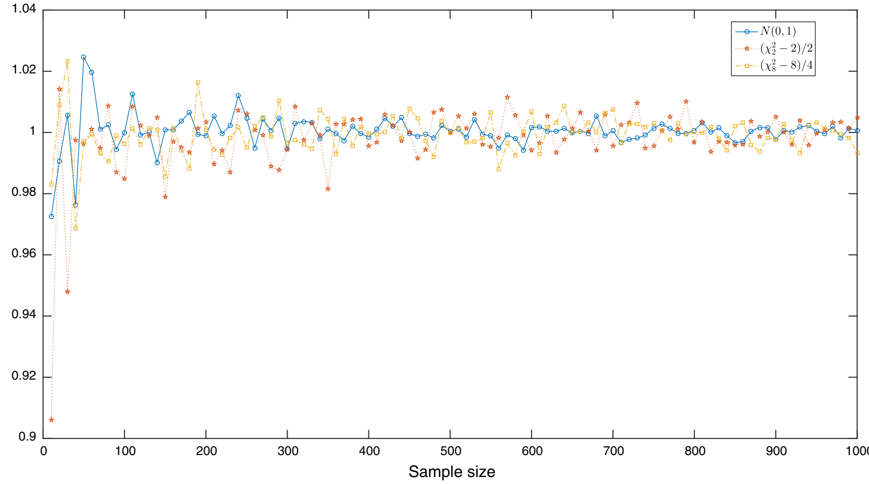

Fig. 3 Graph of the ratio between the estimators $\frac{1}{p}\mathrm{tr}\widehat{\Sigma_1\Sigma_2}$ and the true value $\frac{1}{p}\mathrm{tr}\Sigma_1\Sigma_2$ in Case 1, i.e., $\Sigma_1 = \Sigma_2 = I_p$

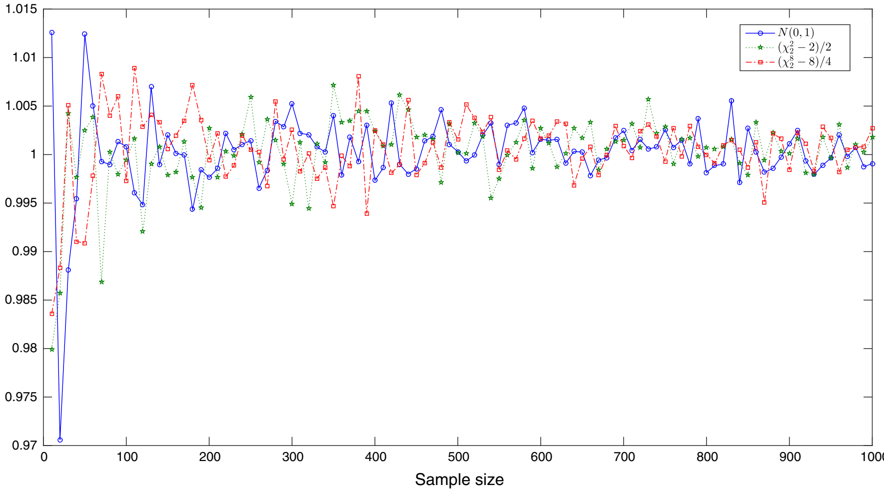

Fig. 4 Graph of the ratio between the estimators $\frac{1}{p}\mathrm{tr}\widehat{\Sigma_1\Sigma_2}$ and the true value $\frac{1}{p}\mathrm{tr}\Sigma_1\Sigma_2$ in Case 2, i.e., $\Sigma_1 = W_1\Psi_1 W_1$ and $\Sigma_2 = W_2\Psi_2 W_2$

$U(0, a)$ which denotes uniform distribution with the support $(0, a)$. Here in Case 1 we choose $a = 0.1$ and in Case 2 we choose $a = 0.2$, respectively.

The ASL and the powers are obtained based on 10,000 replications and the 95 % quantile of the standard normal distribution is 1.64485. The four tables report the ASL and the power in the null hypothesis and the alternative hypothesis of the two tests. For illustration, in the tables we respectively use the estimators proposed in Lemmas 2 and 4 to obtain two different test statistics, $T_{\mathrm{our}}^{\mathrm{umvue}}$ and $T_{\mathrm{our}}^{\mathrm{une}}$. It is shown in Tables 1 and 3 that the ASL of the proposed tests $T_{\mathrm{our}}^{\mathrm{umvue}}$ and $T_{\mathrm{our}}^{\mathrm{une}}$ approximate $\alpha = 0.05$ well in both cases, and $T_{\text{our}}^{\text{une}}$ is even better at nonnormal distributions. Because it is shown that UNE is better than UMVUE if the data sets come from χ^2 distribution, especially when n is small. But the ASL of test T_{sk} in case 2 performs substantially worse. In addition, in Case 1 the test T_{sk} seems worse when dimension p is much larger than the sample size n^*. This is probably because T_{sk} is under common covariance matrix assumption and needs condition $n = O(p^\delta)$ with $\delta > 1/2$ to obtain the asymptotic distribution. As reported in Tables 2 and 4, the powers of the test $T_{\text{our}}^{\text{une}}$ perform better than T_{sk} in Case 2 and worse in Case 1. But actually in Case 1, when the dimension p and sample size n^* are large, the powers of the test $T_{\text{our}}^{\text{une}}$ are also good enough. Thus when the dimension is much larger than the sample size, or the dimension and the sample size are both large, our test statistic is recommended, as it is more stable.

3 Technical details

In this section we give the proof of Theorem 1. We restricted our attention to the case in which $k = 3$ for simplicity and the proof for the case of $k > 3$ is the same. Here we use the same method as in Chen and Qin (2010), hence some of the derivations are omitted. The main difference is that we need to verify the asymptotic normality of $T_n^{(3)}$. Because it does not follow by any means that the random variable $\alpha_n + \beta_n$ will converge in distribution to $\alpha + \beta$, if $\alpha_n \xrightarrow{d} \alpha$ and $\beta_n \xrightarrow{d} \beta$.

Denote $T_n^{(3)} = T_{n1}^{(3)} + T_{n2}^{(3)}$, where

$$T_{n1}^{(3)} = 2\sum_{k=1}^{3}\sum_{i\neq j}^{n_k} \frac{(X_{ki}-\mu_k)'(X_{kj}-\mu_k)}{n_k(n_k-1)} - 2\sum_{k<l}^{3}\sum_{i=1}^{n_k}\sum_{j=1}^{n_l} \frac{(X_{ki}-\mu_k)'(X_{lj}-\mu_l)}{n_k n_l}$$

and

$$T_{n2}^{(3)} = 2\sum_{k,l=1}^{3}\sum_{i=1}^{n_k} \frac{(X_{ki}-\mu_k)'(\mu_k-\mu_l)}{n_k} + \sum_{k<l}^{3} \| \mu_k - \mu_l \|^2 .$$

We can verify that $E(T_{n1}^{(3)}) = 0$, $E(T_{n2}^{(3)}) = \sum_{k<l}^{3} \| \mu_k - \mu_l \|^2$ and

$$\begin{aligned}\operatorname{Var}(T_{n2}^{(3)}) &= 4\sum_{k<l}^{3}(\mu_k-\mu_l)'(n_l^{-1}\Sigma_l + n_k^{-1}\Sigma_k)(\mu_k-\mu_l)\\ &\quad + 4\sum_{i\neq j\neq k}^{3}(\mu_i-\mu_j)'n_i^{-1}\Sigma_i(\mu_i-\mu_k).\end{aligned}$$

From condition (e), that is,

$$\operatorname{Var}\left(\frac{T_{n2}^{(3)} - \sum_{k<l}^{3} \| \mu_k - \mu_l \|^2}{\sigma_n^{(3)}}\right) = o(1),$$

Table 1 ASL of T_{our} and T_{sk} in Case 1

p	n^*	$z_{ijk} \stackrel{iid}{\sim} N(0,1)$			$z_{ijk} \stackrel{iid}{\sim} (\chi_2^2 - 2)/2$			$z_{ijk} \stackrel{iid}{\sim} (\chi_8^2 - 8)/4$		
		$T_{\mathrm{our}}^{\mathrm{umvue}}$	$T_{\mathrm{our}}^{\mathrm{une}}$	T_{sk}	$T_{\mathrm{our}}^{\mathrm{umvue}}$	$T_{\mathrm{our}}^{\mathrm{une}}$	T_{sk}	$T_{\mathrm{our}}^{\mathrm{umvue}}$	$T_{\mathrm{our}}^{\mathrm{une}}$	T_{sk}
20	20	0.0578	0.0672	0.0505	0.0427	0.0643	0.0497	0.0549	0.0655	0.0531
	50	0.0626	0.0648	0.0480	0.0534	0.0605	0.0472	0.0578	0.0594	0.0489
	100	0.0630	0.0626	0.0458	0.0543	0.0614	0.0466	0.0598	0.0604	0.0476
	200	0.0606	0.0612	0.0434	0.0547	0.0635	0.0438	0.0587	0.0586	0.0441
50	20	0.0561	0.0613	0.0509	0.0383	0.0604	0.0471	0.0457	0.0632	0.0503
	50	0.0551	0.0573	0.0454	0.0444	0.0597	0.0450	0.0538	0.0572	0.0443
	100	0.0596	0.0608	0.0465	0.0528	0.0545	0.0490	0.0561	0.0578	0.0458
	200	0.0548	0.0563	0.0453	0.0528	0.0580	0.0453	0.0570	0.0578	0.0472
100	20	0.0565	0.0571	0.0524	0.0346	0.0572	0.0428	0.0477	0.0556	0.0478
	50	0.0551	0.0573	0.0468	0.0471	0.0599	0.0476	0.0506	0.0578	0.0477
	100	0.0548	0.0580	0.0468	0.0499	0.0570	0.0471	0.0524	0.0556	0.0461
	200	0.0549	0.0535	0.0456	0.0541	0.0599	0.0455	0.0521	0.0588	0.0437
500	20	0.0514	0.0549	0.0365	0.0315	0.0543	0.0351	0.0473	0.0528	0.0391
	50	0.0535	0.0553	0.0427	0.0422	0.0515	0.0434	0.0498	0.0517	0.0428
	100	0.0562	0.0511	0.0513	0.0457	0.0510	0.0423	0.0500	0.0532	0.0454
	200	0.0546	0.0518	0.0509	0.0497	0.0477	0.0473	0.0530	0.0584	0.0506
800	20	0.0516	0.0536	0.0332	0.0323	0.0535	0.0327	0.0436	0.0575	0.0325
	50	0.0480	0.0532	0.0393	0.0359	0.0543	0.0387	0.0452	0.0537	0.0392
	100	0.0462	0.0518	0.0403	0.0442	0.0528	0.0421	0.0445	0.0518	0.0421
	200	0.0531	0.0509	0.0465	0.0507	0.0524	0.0473	0.0515	0.0531	0.0490

Table 2 Power of T_{our} and T_{sk} in Case 1

p	n^*	$z_{ijk} \overset{iid}{\sim} N(0,1)$			$z_{ijk} \overset{iid}{\sim} (\chi_2^2-2)/2$			$z_{ijk} \overset{iid}{\sim} (\chi_8^2-8)/4$		
		$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}	$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}	$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}
20	20	0.0941	0.0985	0.1079	0.0615	0.0931	0.1035	0.0828	0.0972	0.1031
	50	0.1673	0.1655	0.2188	0.1380	0.1630	0.2295	0.1538	0.1579	0.2191
	100	0.3111	0.3205	0.4649	0.3009	0.3180	0.4775	0.3139	0.3206	0.4678
	200	0.6710	0.9888	0.8453	0.6571	0.9902	0.8497	0.6664	0.9888	0.8529
50	20	0.1039	0.1158	0.1380	0.0748	0.1130	0.1494	0.0913	0.1147	0.1434
	50	0.2380	0.2431	0.3708	0.2102	0.2353	0.3957	0.2244	0.2372	0.3808
	100	0.5390	0.5459	0.7787	0.5218	0.5419	0.7879	0.5324	0.5446	0.7797
	200	0.9342	1.0000	0.9918	0.9344	1.0000	0.9932	0.9388	1.0000	0.9931
100	20	0.1280	0.1404	0.1843	0.0954	0.1357	0.2076	0.1154	0.1379	0.1918
	50	0.3493	0.3536	0.5712	0.3183	0.3577	0.6043	0.3433	0.3535	0.5871
	100	0.7787	0.7707	0.9574	0.7621	0.7799	0.9609	0.7774	0.7781	0.9581
	200	0.9966	1.0000	0.9999	0.9959	1.0000	1.0000	0.9965	1.0000	0.9999
500	20	0.2908	0.3120	0.4691	0.2206	0.2993	0.5203	0.2723	0.2989	0.4775
	50	0.8576	0.8690	0.9884	0.8413	0.8765	0.9937	0.8643	0.8683	0.9917
	100	1.0000	0.9998	1.0000	0.9997	0.9998	1.0000	1.0000	0.9999	1.0000
	200	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
800	20	0.3921	0.4103	0.6118	0.3028	0.4119	0.6763	0.3706	0.4001	0.6312
	50	0.9670	0.9662	0.9997	0.9547	0.9668	0.9997	0.9641	0.9681	0.9999
	100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	200	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 3 ASL of T_{our} and T_{sk} in Case 2

p	n^*	$z_{ijk} \overset{iid}{\sim} N(0,1)$			$z_{ijk} \overset{iid}{\sim} (\chi_2^2 - 2)/2$			$z_{ijk} \overset{iid}{\sim} (\chi_8^2 - 8)/4$		
		$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}	$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}	$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}
20	20	0.0723	0.0707	0.0153	0.0679	0.0788	0.0103	0.0739	0.0705	0.0120
	50	0.0688	0.0673	0.0087	0.0686	0.0657	0.0105	0.0654	0.0697	0.0084
	100	0.0708	0.0695	0.0100	0.0705	0.0730	0.0085	0.0664	0.0670	0.0096
	200	0.0695	0.0656	0.0083	0.0622	0.0710	0.0061	0.0640	0.0664	0.0104
50	20	0.0755	0.0747	0.0098	0.0815	0.0736	0.0116	0.0646	0.0755	0.0092
	50	0.0711	0.0722	0.0061	0.0676	0.0725	0.0069	0.0685	0.0735	0.0066
	100	0.0683	0.0694	0.0074	0.0623	0.0707	0.0052	0.0658	0.0695	0.0070
	200	0.0687	0.0688	0.0064	0.0673	0.0645	0.0066	0.0645	0.0709	0.0059
100	20	0.0744	0.0747	0.0086	0.0710	0.0705	0.0067	0.0745	0.0722	0.0074
	50	0.0687	0.0720	0.0067	0.0674	0.0665	0.0056	0.0744	0.0724	0.0072
	100	0.0679	0.0719	0.0056	0.0649	0.0641	0.0046	0.0752	0.0690	0.0067
	200	0.0699	0.0705	0.0057	0.0695	0.0684	0.0052	0.0673	0.0674	0.0056
500	20	0.0736	0.0710	0.0044	0.0755	0.0722	0.0048	0.0690	0.0754	0.0043
	50	0.0708	0.0714	0.0043	0.0703	0.0700	0.0043	0.0656	0.0681	0.0032
	100	0.0703	0.0705	0.0035	0.0713	0.0613	0.0019	0.0660	0.0728	0.0023
	200	0.0694	0.0692	0.0032	0.0686	0.0680	0.0030	0.0694	0.0709	0.0019
800	20	0.0730	0.0753	0.0041	0.0797	0.0797	0.0037	0.0736	0.0767	0.0032
	50	0.0680	0.0697	0.0025	0.0686	0.0738	0.0026	0.0754	0.0733	0.0026
	100	0.0686	0.0667	0.0022	0.0667	0.0651	0.0017	0.0680	0.0683	0.0017
	200	0.0705	0.0686	0.0015	0.0688	0.0678	0.0028	0.0691	0.0620	0.0016

Table 4 Power of T_{our} and T_{sk} in Case 2

p	n^*	$z_{ijk} \overset{iid}{\sim} N(0,1)$			$z_{ijk} \overset{iid}{\sim} (\chi_2^2-2)/2$			$z_{ijk} \overset{iid}{\sim} (\chi_8^2-8)/4$		
		$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}	$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}	$T_{\text{our}}^{\text{umvue}}$	$T_{\text{our}}^{\text{une}}$	T_{sk}
20	20	0.1952	0.0998	0.0573	0.1855	0.1998	0.0529	0.1859	0.1962	0.0557
	50	0.3860	0.1420	0.1377	0.3814	0.3973	0.1345	0.3927	0.3881	0.1406
	100	0.6769	0.2225	0.3454	0.6783	0.6782	0.3486	0.6864	0.6799	0.3457
	200	0.9464	0.7934	0.7498	0.9491	0.9999	0.7450	0.9516	0.9998	0.7475
50	20	0.2096	0.1111	0.0486	0.2031	0.2215	0.0500	0.2016	0.2087	0.0502
	50	0.4201	0.1583	0.1327	0.4246	0.4257	0.1326	0.4350	0.4329	0.1385
	100	0.7369	0.2442	0.3462	0.7425	0.7423	0.3588	0.7506	0.7471	0.3583
	200	0.9779	0.8457	0.7835	0.9796	1.0000	0.7763	0.9816	1.0000	0.7899
100	20	0.2148	0.1050	0.0471	0.2112	0.2230	0.0471	0.2163	0.2188	0.0458
	50	0.4575	0.1529	0.1286	0.4564	0.4583	0.1305	0.4631	0.4571	0.1292
	100	0.7860	0.2482	0.3601	0.7838	0.7839	0.3549	0.7797	0.7863	0.3482
	200	0.9899	0.8846	0.7939	0.9897	1.0000	0.7994	0.9912	1.0000	0.7949
500	20	0.2465	0.1134	0.0342	0.2430	0.2478	0.0331	0.2450	0.2460	0.0350
	50	0.5226	0.1787	0.1127	0.5220	0.5162	0.1119	0.5284	0.5198	0.1161
	100	0.8594	0.2807	0.3373	0.8624	0.8637	0.3371	0.8595	0.8619	0.3381
	200	0.9990	0.9400	0.8238	0.9995	1.0000	0.8171	0.9993	1.0000	0.8212
800	20	0.2474	0.1237	0.0290	0.2586	0.2572	0.0320	0.2579	0.2597	0.0317
	50	0.5397	0.1867	0.1063	0.5550	0.5482	0.1061	0.5466	0.5352	0.1033
	100	0.8250	0.2984	0.3359	0.8850	0.8893	0.3459	0.8855	0.8865	0.3370
	200	0.9998	0.9548	0.8152	0.9997	1.0000	0.8252	0.9996	1.0000	0.8247

we get

$$\frac{T_n^{(3)} - \sum_{k<l}^{3} \parallel \mu_k - \mu_l \parallel^2}{\sqrt{\mathrm{Var}(T_n^{(3)})}} = \frac{T_{n1}^{(3)}}{\sigma_n^{(3)}} + o_p(1).$$

Next we will prove the asymptotic normality of $T_{n1}^{(3)}$. Without loss of generality we assume that $\mu_1 = \mu_2 = \mu_3 = 0$. Let $Y_i = X_{1i}(i = 1, \ldots, n_1)$, $Y_{j+n_1} = X_{2j}(j = 1, \ldots, n_2)$, $Y_{j+n_1+n_2} = X_{3j}(j = 1, \ldots, n_3)$. For $i \neq j$,

$$\phi_{ij} = \begin{cases} 2n_1^{-1}(n_1-1)^{-1}Y_i'Y_j, & \text{if } i, j \in \{1, 2, \ldots, n_1\}; \\ -n_1^{-1}n_2^{-1}Y_i'Y_j, & \text{if } i \in \{1, 2, \ldots, n_1\} \text{and} j \in \{n_1+1, \ldots, n_1+n_2\}; \\ 2n_2^{-1}(n_2-1)^{-1}Y_i'Y_j, & \text{if } i, j \in \{n_1+1, \ldots, n_1+n_2\}; \\ -n_2^{-1}n_3^{-1}Y_i'Y_j, & \text{if } i \in \{n_1+1, \ldots, n_1+n_2\} \\ & \text{and } j \in \{n_1+n_2+1, \ldots, n_1+n_2+n_3\}; \\ 2n_3^{-1}(n_3-1)^{-1}Y_i'Y_j, & \text{if } i, j \in \{n_1+n_2+1, \ldots, n_1+n_2+n_3\}; \\ -n_3^{-1}n_1^{-1}Y_i'Y_j, & \text{if } i \in \{1, 2, \ldots, n_1\} \\ & \text{and } j \in \{n_1+n_2+1, \ldots, n_1+n_2+n_3\}. \end{cases}$$

For $j = 2, 3, \ldots, n_1 + n_2 + n_3$, denote $V_{nj} = \sum_{i=1}^{j-1} \phi_{ij}$, $S_{nm} = \sum_{j=2}^{m} V_{nj}$ and $\mathcal{F}_{nm} = \sigma\{Y_1, Y_2, \ldots, Y_m\}$ which is the σ algebra generated by $\{Y_1, Y_2, \ldots, Y_m\}$. Then we have

$$T_n^{(3)} = 2 \sum_{j=2}^{n_1+n_2+n_3} V_{nj}.$$

It is easy to verify that $\{S_{nm}, \mathcal{F}_{nm}\}_{m=1}^{n}$ forms a sequence of zero mean and square integrable martingale. Then the asymptotic normality of $T_n^{(3)}$ can be proved by employing Corollary 3.1 in Hall and Heyde (1980) with routine verification of the following:

$$\frac{\sum_{j=2}^{n_1+n_2+n_3} E[V_{nj}^2 | \mathcal{F}_{n,j-1}]}{(\sigma_n^{(3)})^2} \xrightarrow{P} \frac{1}{4}. \tag{11}$$

and

$$\sum_{j=2}^{n_1+n_2+n_3} (\sigma_n^{(3)})^{-2} E[V_{nj}^2 I(|V_{nj}| > \epsilon\sigma_n^{(3)}) | \mathcal{F}_{n,j-1}] \xrightarrow{P} 0. \tag{12}$$

Thus next we prove (11) and (12) respectively.

3.1 Proof of (11)

Verify that

$$E(V_{nj}^2|\mathcal{F}_{n,j-1}) = E\left[\left(\sum_{i=1}^{j-1}\phi_{ij}\right)^2\middle|\mathcal{F}_{n,j-1}\right] = E\left(\sum_{i_1,i_2=1}^{j-1}\phi_{i_1j}\phi_{i_2j}\middle|\mathcal{F}_{n,j-1}\right)$$
$$= \sum_{i_1,i_2=1}^{j-1} c_{i_1j}c_{i_2j}Y_{i_1}'E(Y_jY_j'|\mathcal{F}_{n,j-1})Y_{i_2} = \sum_{i_1,i_2=1}^{j-1} c_{i_1j}c_{i_2j}Y_{i_1}'E(Y_jY_j')Y_{i_2}$$
$$= \sum_{i_1,i_2=1}^{j-1} c_{i_1j}c_{i_2j}Y_{i_1}'\tilde{\Sigma}_jY_{i_2},$$

where c_{ij} is the coefficient of ϕ_{ij} and if $j \in [1, n_1]$, $\tilde{\Sigma}_j = \Sigma_1$; if $j \in [n_1+1, n_1+n_2]$, $\tilde{\Sigma}_j = \Sigma_2$; if $j \in [n_1+n_2+1, n_1+n_2+n_3]$, $\tilde{\Sigma}_j = \Sigma_3$.

Denote

$$\eta_n = \sum_{j=2}^{n_1+n_2+n_3} E(V_{nj}^2|\mathcal{F}_{n,j-1}).$$

Then we have

$$E(\eta_n) = \frac{2\mathrm{tr}(\Sigma_1^2)}{n_1(n_1-1)} + \frac{2\mathrm{tr}(\Sigma_2^2)}{n_2(n_2-1)} + \frac{2\mathrm{tr}(\Sigma_3^2)}{n_3(n_3-1)}$$
$$+ \frac{\mathrm{tr}(\Sigma_1\Sigma_2)}{n_1n_2} + \frac{\mathrm{tr}(\Sigma_1\Sigma_3)}{n_1n_3} + \frac{\mathrm{tr}(\Sigma_3\Sigma_2)}{n_3n_2} = \frac{1}{4}(\sigma_n^{(3)})^2.$$

Now consider

$$E(\eta_n^2) = E\left[\sum_{j=2}^{n_1+n_2+n_3}\sum_{i_1,i_2=1}^{j-1} c_{i_1j}c_{i_2j}Y_{i_1}'\tilde{\Sigma}_jY_{i_2}\right]^2 = 2E(A) + E(B), \tag{13}$$

where

$$A = \sum_{2\le j_1<j_2}^{n_1+n_2+n_3}\sum_{i_1,i_2=1}^{j_1-1}\sum_{i_3,i_4=1}^{j_2-1} c_{i_1j_1}c_{i_2j_1}c_{i_3j_2}c_{i_4j_2}Y_{i_1}'\tilde{\Sigma}_{j_1}Y_{i_2}Y_{i_3}'\tilde{\Sigma}_{j_2}Y_{i_4}$$

and

$$B = \sum_{j=2}^{n_1+n_2+n_3}\sum_{i_1,i_2=1}^{j-1}\sum_{i_3,i_4=1}^{j-1} c_{i_1j}c_{i_2j}c_{i_3j}c_{i_4j}Y_{i_1}'\tilde{\Sigma}_jY_{i_2}Y_{i_3}'\tilde{\Sigma}_jY_{i_4}.$$

The term B can be further partitioned as $B = B_1 + B_2 + B_3$, where

$$E(B_1) = E\left(\sum_{j=2}^{n_1} \sum_{i_1,i_2=1}^{j-1} \sum_{i_3,i_4=1}^{j-1} c_{i_1 j} c_{i_2 j} c_{i_3 j} c_{i_4 j} Y_{i_1}' \Sigma_1 Y_{i_2} Y_{i_3}' \Sigma_1 Y_{i_4}\right)$$

$$E(B_2) = E\left(\sum_{j=n_1+1}^{n_1+n_2} \sum_{i_1,i_2=1}^{j-1} \sum_{i_3,i_4=1}^{j-1} c_{i_1 j} c_{i_2 j} c_{i_3 j} c_{i_4 j} Y_{i_1}' \Sigma_2 Y_{i_2} Y_{i_3}' \Sigma_2 Y_{i_4}\right)$$

$$E(B_3) = E\left(\sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{i_1,i_2=1}^{j-1} \sum_{i_3,i_4=1}^{j-1} c_{i_1 j} c_{i_2 j} c_{i_3 j} c_{i_4 j} Y_{i_1}' \Sigma_3 Y_{i_2} Y_{i_3}' \Sigma_3 Y_{i_4}\right).$$

We only compute $E(B_3)$ here as $E(B_1)$ and $E(B_2)$ can be computed following the same procedure. As $\mu_1 = \mu_2 = \mu_3 = 0$, we only need to consider i_1, i_2, i_3 and i_4 in these three cases: (a) $(i_1 = i_2) \neq (i_3 = i_4)$; (b) $(i_1 = i_3) \neq (i_2 = i_4)$ or $(i_1 = i_4) \neq (i_2 = i_3)$; (c) $i_1 = i_2 = i_3 = i_4$. Thus we obtain that

$$E(B_3) = E(B_{31}) + E(B_{32}) + E(B_{33}),$$

where

$$\begin{aligned} E(B_{31}) &= O(n^{-8}) E\left(\sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{i_1 \neq i_2}^{j-1} Y_{i_1}' \Sigma_3 Y_{i_1} Y_{i_2}' \Sigma_3 Y_{i_2}\right) \\ &= O(n^{-5}) \sum_{i,j=1}^{3} \mathrm{tr}\,\Sigma_3 \Sigma_i \mathrm{tr}\,\Sigma_3 \Sigma_j \\ E(B_{32}) &= O(n^{-8}) E\left(\sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{i_1 \neq i_2}^{j-1} Y_{i_1}' \Sigma_3 Y_{i_2} Y_{i_2}' \Sigma_3 Y_{i_1}\right) \\ &= O(n^{-5}) \sum_{i,j=1}^{3} \mathrm{tr}\,\Sigma_i \Sigma_3 \Sigma_j \Sigma_3 \end{aligned}$$

and

$$\begin{aligned} E(B_{33}) &= O(n^{-8}) E\left(\sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{i=1}^{j-1} Y_i' \Sigma_3 Y_i Y_i' \Sigma_3 Y_i\right) \\ &= O(n^{-6}) \left(\sum_{i=1}^{3} \left(E(Z_{i1}' \Gamma_i' \Sigma_3 \Gamma_i Z_{i1} - \mathrm{tr}\,\Sigma_3 \Sigma_i)^2 + \mathrm{tr}^2 \Sigma_3 \Sigma_i\right)\right) \\ &= O(n^{-6}) \sum_{i=1}^{3} (\mathrm{tr}(\Sigma_3 \Sigma_i)^2 + \mathrm{tr}^2 \Sigma_3 \Sigma_i). \end{aligned}$$

Thus we obtain that

$$E(B_3) = o((\sigma_n^{(3)})^4).$$

As we can similarly get $E(B_1) = o((\sigma_n^{(3)})^4)$ and $E(B_2) = o((\sigma_n^{(3)})^4)$, we conclude that

$$E(B) = o((\sigma_n^{(3)})^4). \tag{14}$$

Using the same method of deriving (14), we have

$$2E(A) = \frac{1}{16}(\sigma_n^{(3)})^4(1 + o(1)),$$

which together with (13) and (14) implies

$$E(\eta_n^2) = \frac{1}{16}(\sigma_n^{(3)})^4 + o((\sigma_n^{(3)})^4).$$

Then we have

$$\mathrm{Var}(\eta_n) = E(\eta_n^2) - E^2(\eta_n) = o((\sigma_n^{(3)})^4).$$

Therefore we obtain

$$(\sigma_n^{(3)})^{-2} E\left\{\sum_{j=1}^{n_1+n_2+n_3} E(V_{nj}^2|\mathcal{F}_{n,j-1})\right\} = (\sigma_n^{(3)})^{-2} E(\eta_n) = \frac{1}{4}$$

and

$$(\sigma_n^{(3)})^{-4} \mathrm{Var}\left\{\sum_{j=1}^{n_1+n_2+n_3} E(V_{nj}^2|\mathcal{F}_{n,j-1})\right\} = (\sigma_n^{(3)})^{-4}\mathrm{Var}(\eta_n) = o(1),$$

which complete the proof of (11). □

3.2 Proof of (12)

As

$$\sum_{j=2}^{n_1+n_2+n_3} (\sigma_n^{(3)})^{-2} E\{V_{nj}^2 I(|V_{nj}| > \epsilon\sigma_n^{(3)})|\mathcal{F}_{n,j-1}\}$$
$$\le (\sigma_n^{(3)})^{-4}\epsilon^{-2} \sum_{j=1}^{n_1+n_2+n_3} E(V_{nj}^4|\mathcal{F}_{n,j-1}),$$

we just need to show that

$$E\left(\sum_{j=2}^{n_1+n_2+n_3} E(V_{nj}^4|\mathcal{F}_{n,j-1})\right) = o((\sigma_n^{(3)})^4).$$

Note that

$$\begin{aligned}
&E\left\{\sum_{j=2}^{n_1+n_2+n_3} E(V_{nj}^4|\mathcal{F}_{n,j-1})\right\}\\
&= \sum_{j=2}^{n_1+n_2+n_3} E(V_{nj}^4) = O(n^{-8}) \sum_{j=2}^{n_1+n_2+n_3} E\left(\sum_{i=1}^{j-1} Y_i'Y_j\right)^4\\
&= O(n^{-8}) \sum_{j=2}^{n_1+n_2+n_3} \sum_{s\neq t}^{j-1} E((Y_j'Y_s)^2(Y_t'Y_j)^2) + O(n^{-8}) \sum_{j=2}^{n_1+n_2+n_3} \sum_{s=1}^{j-1} E(Y_s'Y_j)^4\\
&= O(n^{-8}) \sum_{j=2}^{n_1+n_2+n_3} \sum_{s\neq t}^{j-1} E(Y_j'\tilde{\Sigma}_s Y_j Y_j'\tilde{\Sigma}_t Y_j) + O(n^{-8}) \sum_{j=2}^{n_1+n_2+n_3} \sum_{s=1}^{j-1} E(Y_s'Y_j)^4.
\end{aligned}$$

The first term of last equation has the order $o((\sigma_n^{(3)})^4)$ which can be proved by the same procedure in last subsection. It remains to consider the second term. As proved in Chen and Qin (2010), we have

$$\sum_{j=2}^{n_1+n_2} \sum_{s=1}^{j-1} E(Y_s'Y_j)^4 = O(n^2)\left(\sum_{i=1}^{2}(\mathrm{tr}^2(\Sigma_i^2) + \mathrm{tr}(\Sigma_i^4)) + \mathrm{tr}^2(\Sigma_1\Sigma_2) + \mathrm{tr}(\Sigma_1\Sigma_2)^2\right),$$

and

$$\begin{aligned}
\sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{s=1}^{j-1} E(Y_s'Y_j)^4 &= \sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{s=1}^{n_1} E(Y_s'Y_j)^4 + \sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{s=n_1+1}^{n_1+n_2} E(Y_s'Y_j)^4\\
&\quad + \sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{s=n_1+n_2+1}^{j-1} E(Y_s'Y_j)^4\\
&= O(n^2)\left(\mathrm{tr}^2(\Sigma_3^2) + \mathrm{tr}(\Sigma_3^4) + \sum_{i=1}^{2} \mathrm{tr}^2(\Sigma_i\Sigma_3)\right.\\
&\quad \left. + \sum_{i=1}^{2} \mathrm{tr}(\Sigma_i\Sigma_3)^2\right).
\end{aligned}$$

Thus we conclude that

$$
\begin{aligned}
&O(n^{-8}) \sum_{j=2}^{n_1+n_2+n_3} \sum_{s=1}^{j-1} E(Y_s'Y_j)^4 \\
&= O(n^{-8}) \sum_{j=2}^{n_1+n_2} \sum_{s=1}^{j-1} E(Y_s'Y_j)^4 + O(n^{-8}) \sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} \sum_{s=1}^{j-1} E(Y_s'Y_j)^4 \\
&= O(n^{-6}) \left(\sum_{i=1}^{3} (\mathrm{tr}^2(\Sigma_i^2) + \mathrm{tr}(\Sigma_i^4)) + \sum_{i<j}^{3} \mathrm{tr}^2(\Sigma_i \Sigma_j) + \sum_{i<j}^{3} \mathrm{tr}(\Sigma_i \Sigma_j)^2 \right) \\
&= o((\sigma_n^{(3)})^4).
\end{aligned}
$$

Then the proof of (12) is complete. $\square$

Acknowledgements The authors would like to thank the referees for their constructive comments that led to a substantial improvement of the paper. J. Hu was partially supported by CNSF 11301063. Z. D. Bai was partially supported by CNSF 11171057 and 11571067.

References

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

Bai, Z. D., Jiang, D. D., Yao, J. F., Zheng, S. R. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *The Annals of Statistics*, *37*(6B), 3822–3840.

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

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.

Dempster, A. P. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, *29*(1), 995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, *16*(1), 41.

Fujikoshi, Y., Himeno, T., Wakaki, H. (2004). Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size. *Journal of the Japan Statistical Society*, *34*(1), 19–26.

Fujikoshi, Y., Ulyanov, V. V., Shimizu, R. (2010). *Multivariate statistics: High-dimensional and large-sample approximations*. Hoboken, New Jersey: Wiley.

Hall, P., Heyde, C. C. (1980). *Martingale limit theory and its application*. New York: Academic press.

Hotelling, H. (1947). Multivariate quality control, illustrated by the air testing of sample bombsights. In C. Eisenhart, M.W. Hastay, E.A. Wallis (Eds.) *Selected Techniques of Statistical Analysis* (pp. 111–184). New York: McGraw-Hill.

Jiang, D., Jiang, T., Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *Journal of Statistical Planning and Inference*, *142*(8), 2241–2256.

Jiang, T., Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *The Annals of Statistics*, *41*(4), 2029–2074.

Lawley, D. N. (1938). A generalization of Fisher's z test. *Biometrika*, *30*(1/2), 180–187.

Li, J., Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 908–940.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *The Annals of Mathematical Statistics*, *26*(1), 117–121.

Schott, J. R. (2007). Some high-dimensional tests for a one-way MANOVA. *Journal of Multivariate Analysis*, *98*(9), 1825–1839.
Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, *37*(1), 53–86.
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.
Srivastava, M. S., Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, *99*(3), 386–402.
Srivastava, M. S., Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis*, *97*(9), 1927–1940.
Srivastava, M. S., Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, *115*, 204–216.
Srivastava, M. S., Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis*, *101*(6), 1319–1329.
Srivastava, M. S., Kollo, T., von Rosen, D. (2011). Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, *102*(6), 1090–1103.
Srivastava, M. S., Katayama, S., Kano, Y. (2013). A two sample test in high dimensional data. *Journal of Multivariate Analysis*, *114*, 349–358.
Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, *24*(3), 471–494.