Testing the equality of two covariance matrices for non-normal high-dimensional data

Jieqiong Shen
School of Computer and Data Engineering
NingboTech University
Ningbo 315100
China
and
College of Mathematics
Sichuan University
Chengdu 610064
China
jieqiongshen123@163.com

Abstract. This paper focuses on testing the equality of two high-dimensional covariance matrices without the normality assumption. Two new test statistics are proposed by taking the ratio between the unbiased and consistent estimators of the trace of the covariance matrices. Under some mild assumptions, the proposed test statistics are proved to be asymptotically normal. Furthermore, numerical simulations demonstrate that the proposed tests have good size and power with varying dimensions and sample sizes.

Keywords: high-dimensional covariance matrices, hypothesis test, equality, non-normality.

1. Introduction

Let $X_{ij} = (x_{ij1}, \ldots, x_{ijp})'$, $j = 1, \ldots, N_i$, be independent and identically distributed vectors coming from a p-dimensional population $\mathfrak{F}_i$ with the mean vector μ_i and covariance matrix Σ_i, where N_i denotes the sample size of the i-th population, $i = 1, 2$. We devote to checking on the equality of two covariance matrices, that is, testing the following hypotheses:

$$H_0 : \Sigma_1 = \Sigma_2 = \Sigma \quad \text{vs.} \quad H_1 : \Sigma_1 > \Sigma_2 \text{ or } \Sigma_2 > \Sigma_1, \tag{1}$$

where $A > B$ means that $A - B$ is positive definite. Here we just pay attention to the case $p > N_i$, i.e., so-called high-dimensional situation, because data with the dimension larger than the sample sizes are increasing encountered in recent statistical studies, such as gene data analyses, see [1, 4] for a broader introduction.

As noted in [7], if we know that $\mathfrak{F}_1$ has a larger covariance matrix than $\mathfrak{F}_2$, one may advise us to take larger sample from $\mathfrak{F}_1$ to offset the largeness to some

degree; therefore, it is interesting to consider the testing problem (1), in which H_1 is so-called one-side alternative. By using a lower bound on a measure of distance between the null and alternative hypotheses, Srivastava and Yanagihara [7] constructed their test procedure based on the unbiased estimators of $\mathrm{tr}\Sigma_i$ and $\mathrm{tr}\Sigma_i^2$.

It is worth pointing out that H_1 is a special case of $H_1' : \Sigma_1 \neq \Sigma_2$. Thus, the test procedures, designed to test the hypotheses $H_0 : \Sigma_1 = \Sigma_2$ vs. $H_1' : \Sigma_1 \neq \Sigma_2$, still work for testing the hypotheses (1). We also mention that there exist many strategies to test the hypotheses $H_0 : \Sigma_1 = \Sigma_2$ vs. $H_1' : \Sigma_1 \neq \Sigma_2$. For instance, based on the Frobenius norm of the difference of two covariances, Schott [5] proposed a test statistic which is an unbiased estimator of $\mathrm{tr}(\Sigma_1-\Sigma_2)^2$.

Note that the test procedures, both in [5] and [7] mentioned above, heavily relied on the normality assumption. They used an estimator of $\mathrm{tr}\Sigma_i^2$ proposed by [2, 6], which is given by

$$\widehat{\mathrm{tr}\Sigma_i^2}_{(1)} = \frac{(N_i-1)^2}{(N_i-2)(N_i+1)}\left(\mathrm{tr}S_i^2 - \frac{1}{N_i-1}(\mathrm{tr}S_i)^2\right), \tag{2}$$

where $S_i = \frac{1}{N_i-1}\sum_{j=1}^{N_i} Y_{ij}Y_{ij}'$ with $Y_{ij} = X_{ij} - \bar{X}_i$ and $\bar{X}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}$, $i = 1, 2$. The estimator $\widehat{\mathrm{tr}\Sigma_i^2}_{(1)}$ given in (2) is unbiased under the normality assumption, but is biased under the non-normality assumption. Therefore, Srivastava et al. [8] defined an unbiased estimator of $\mathrm{tr}\Sigma_i^2$ dispensing with the normality condition, which is given by

$$\widehat{\mathrm{tr}\Sigma_i^2}_{(2)} = \frac{1}{f}\Big((N_i-2)(N_i-1)^3\mathrm{tr}S_i^2 - N_i(N_i-1)\mathrm{tr}D_i^2 + (N_i-1)^2(\mathrm{tr}S_i)^2\Big), \tag{3}$$

where $f = N_i(N_i-1)(N_i-2)(N_i-3)$ and $D_i = \mathrm{diag}(Y_{i1}'Y_{i1}, \ldots, Y_{iN_i}'Y_{iN_i})$, $i = 1, 2$. Moreover, Srivastava et al. [8] proposed a test procedure by employing $\widehat{\mathrm{tr}\Sigma_i^2}_{(2)}$ to modify the test statistic proposed in [5].

Inspired by [8], we aim to propose some new and more powerful test procedures for testing the hypotheses (1) without the normal condition. The test statistics are constructed by taking the ratio between the unbiased and consistent estimators of the trace of the covariance matrices. The rest paper is organized as follows. Section 2 reviews the estimators of $\mathrm{tr}\Sigma_i^l/p$, $l = 1, 2, 3, 4$, and the properties of these estimators for general populations under some mild assumptions. In Section 3, two new test statistics are proposed, whose asymptotic distributions are derived to be normal. Simulation results are reported in Section 4. A conclusion is made in Section 5.

2. Notations and preliminaries

In this section, we collect some important results needed in the later sections. We begin with the description of the general multivariate model, including

the multivariate normal distribution as a special case. Assume that the data $\{X_{ij}\}_{j=1}^{N_i}$ satisfy

$$X_{ij} = \Lambda_i Z_{ij} + \mu_i,\ i = 1, 2, j = 1, \ldots, N_i. \tag{4}$$

Here, Λ_i is a $p \times p$ matrix such that $\Lambda_i \Lambda_i' = \Sigma_i = (\sigma_{ist})_{p\times p}$, and $Z_{ij} = (z_{ij1}, \ldots, z_{ijp})'$ has independent and identically distributed elements, yielding that $\mathbb{E}Z_{ij} = \mathbf{0}$, $\mathrm{Cov}(Z_{ij}) = I_p$, where I_p denotes a $p \times p$ identity matrix. Note that the model (4) is widely used in the high dimensional multivariate analysis, such as [2, 8, 3].

Throughout the sequel we abbreviate $a_l = \frac{1}{p}\mathrm{tr}\Sigma^l$, $a_{li} = \frac{1}{p}\mathrm{tr}\Sigma_i^l$, $l = 1, 2, 3, 4$, $i = 1, 2$. Then the estimators of a_{li} can be stated as follows:

$$\widehat{a}_{1i} = \frac{1}{p}\mathrm{tr}S_i, \tag{5}$$

$$\widehat{a}_{2i} = \frac{1}{p}\widehat{\mathrm{tr}\Sigma_{i\,(2)}^2}, \tag{6}$$

$$\widehat{a}_{3i} = \frac{\tau_1}{p}\left(\mathrm{tr}S_i^3 - \frac{3}{n_i}\mathrm{tr}S_i^2\mathrm{tr}S_i + \frac{2}{n_i^2}(\mathrm{tr}S_i)^3\right), \tag{7}$$

$$\begin{aligned}\widehat{a}_{4i} = \frac{\tau_2}{p}\Bigg(&\mathrm{tr}S_i^4 - \frac{4}{n_i}\mathrm{tr}S_i^3\mathrm{tr}S_i - \frac{2n_i^2 + 3n_i - 6}{n_i^3 + n_i^2 + 2n_i}(\mathrm{tr}S_i^2)^2 \\ &+ \frac{10n_i + 12}{n_i^3 + n_i^2 + 2n_i}\mathrm{tr}S_i^2(\mathrm{tr}S_i)^2 - \frac{5n_i + 6}{n_i^4 + n_i^3 + 2n_i^2}(\mathrm{tr}S_i)^4\Bigg),\end{aligned} \tag{8}$$

where $n_i = N_i - 1$, and

$$\tau_1 = \frac{n_i^4}{(n_i - 1)(n_i - 2)(n_i + 2)(n_i + 4)},$$
$$\tau_2 = \frac{n_i^7 + n_i^6 + 2n_i^5}{(n_i + 1)(n_i + 2)(n_i + 4)(n_i + 6)(n_i - 1)(n_i - 2)(n_i - 3)}.$$

One can see [8, 9] and references therein for details. To consider the unbiasedness, consistence and asymptotic normality of these estimators, the following assumptions are needed:

Assumption 1. *Both N_i and p tend to infinity with $N_i = O(p^\varepsilon)$, $1/2 < \varepsilon < 1$.*

Assumption 2. *For $N_1 \le N_2$, $0 < N_1/N_2 \le 1$.*

Assumption 3. *$0 < a_{2i} < \infty$, $a_{4i}/p = o(1)$, and $\sum_{s,t=1}^p \sigma_{ist}^4/p^2 = o(1)$.*

Assumption 4. *$\mathbb{E}(z_{ijk}^4) = \Delta_i + 3$ with $\Delta_i < \infty$, where z_{ijk} is the k-th component of the vector $Z_{ij} = (z_{ij1}, \ldots, z_{ijp})'$, and $\mathbb{E}\left(\prod_{k=1}^p z_{ijk}^{\gamma_k}\right) = \prod_{k=1}^p \mathbb{E}(z_{ijk}^{\gamma_k})$ for integers $\gamma_k \ge 0$ with $\sum_{k=1}^p \gamma_k \le 8$.*

Assumption 5. *$\mathrm{tr}(\Sigma_i^s \odot \Sigma_i^t) = o(\mathrm{tr}\Sigma_i^{s+t})$ for integers s, t with $1 \le s, t \le 2$, where $\odot$ denotes the Hadamard product of two matrices.*

Assumptions $1 \sim 4$ are used frequently, see for instance [8]. Indeed, Assumption 4 indicates that z_{ijk} has the finite fourth moment and the existence of the moments of z_{ijk} are up to the order eight. Assumption 5 is also practical and holds for many common covariances. For example, let Σ_i be compound symmetric structure, i.e., $\Sigma_i = (1-\rho)I_p + \rho J_p$, where $0 < \rho < 1$ and J_p is the $p \times p$ matrix with all elements being one. It is easy to show that $\mathrm{tr}(\Sigma_i^s \odot \Sigma_i^t) = o(p)$ and $\mathrm{tr}\Sigma_i^{s+t} = O(p^{s+t})$ for integers $1 \le s, t \le 2$, which indicates that Assumption 5 is reasonable.

Furthermore, by employing the results of [8, 9], the properties of $\widehat{a}_{li}$, $l = 1, 2, 3, 4$, can be summarized immediately by the following lemmas.

Lemma 2.1. *For $i = 1, 2$, under Assumptions $1 \sim 5$,*

(i) when the underlying distribution is normal, $\widehat{a}_{li}$ given above is unbiased and consistent estimator of a_{li}, $l = 1, 2, 3, 4$;

(ii) when the underlying distribution is not normal, $\widehat{a}_{1i}$ and $\widehat{a}_{2i}$ are still unbiased and consistent estimators of a_{1i} and a_{2i}, respectively, while $\widehat{a}_{3i}$ and $\widehat{a}_{4i}$ are consistent but biased estimators of a_{3i} and a_{4i}, respectively.

Lemma 2.2. *For $i = 1, 2$, under Assumptions $1 \sim 5$, one has*

$$U^{-\frac{1}{2}} (\widehat{a}_{1i} - a_{1i}, \widehat{a}_{2i} - a_{2i})' \xrightarrow{D} N_2(\mathbf{0}, I_2),$$

where

$$U = \begin{pmatrix} 2a_{2i}/(N_i p) & , & 4a_{3i}/(N_i p) \\ 4a_{3i}/(N_i p) & , & (4pa_{2i}^2 + 8N_i a_{4i} - 8a_{4i})/(N_i^2 p) \end{pmatrix}.$$

3. Two test procedures

Based on the consistent estimators of two distance functions β_1 and β_2 given in (9) and (14) respectively, we accordingly propose two test statistics for the testing problem (1) in this section.

3.1 The first test procedure

We firstly notice that a measure of distance between H_0 and H_1 can be given by

$$\beta_1 = \frac{\mathrm{tr}\Sigma_1^2}{\mathrm{tr}\Sigma_2^2} - 1, \tag{9}$$

where β_1 equals zero if H_0 is true, meanwhile, β_1 does not equal zero if H_1 holds. Then, the corresponding test statistic, based on the consistent estimator of β_1, can be constructed by

$$\widehat{\beta}_1 = \frac{\widehat{a}_{21}}{\widehat{a}_{22}} - 1, \tag{10}$$

where $\widehat{a}_{2i}$ is given in (6), $i = 1, 2$. A straightforward calculation, embedded in the proof of Theorem 3.1, shows that the asymptotic variance of $\widehat{\beta}_1$ is

$$\delta_1 = \frac{4pa_{21}^2 + 8(N_1 - 1)a_{41}}{N_1^2 pa_{22}^2} + \frac{4pa_{21}^2 a_{22}^2 + 8(N_2 - 1)a_{21}^2 a_{42}}{N_2^2 pa_{22}^4}. \tag{11}$$

Obviously, under $H_0 : \Sigma_1 = \Sigma_2 = \Sigma$, we have $a_{21} = a_{22} = a_2$ and $a_{41} = a_{42} = a_4$. Thus δ_1 reduces to

$$\delta_{10} = \frac{4}{N_1^2} + \frac{4}{N_2^2} + \left(\frac{8N_1 - 8}{N_1^2 p} + \frac{8N_2 - 8}{N_2^2 p}\right)\frac{a_4}{a_2^2}. \tag{12}$$

The following theorem establishes the asymptotic normality of $\widehat{\beta}_1$.

Theorem 3.1. *Under Assumptions* $1 \sim 5$*, we have*

$$(\delta_1)^{-\frac{1}{2}}(\widehat{\beta}_1 - \beta_1) \xrightarrow{D} N(0, 1).$$

In particular, under H_0*, we have*

$$(\delta_{10})^{-\frac{1}{2}}\widehat{\beta}_1 \xrightarrow{D} N(0, 1),$$

where δ_1 *and* δ_{10} *are given in* (11) *and* (12)*, respectively.*

Proof. Since two samples are drawn from two independent populations, by employing Lemma 2.2, we have the distribution of $(\widehat{a}_{21}, \widehat{a}_{22})'$ as follows:

$$V^{-\frac{1}{2}}(\widehat{a}_{21} - a_{21}, \widehat{a}_{22} - a_{22})' \xrightarrow{D} N_2(\mathbf{0}, I_2),$$

where $V = \mathrm{diag}(\xi_1, \xi_2)$ with $\xi_i = (4pa_{2i}^2 + 8N_i a_{4i} - 8a_{4i})/(N_i^2 p)$, $i = 1, 2$. Note that $\widehat{\beta}_1 = \widehat{a}_{21}/\widehat{a}_{22} - 1$ is a function of $\widehat{a}_{21}$ and $\widehat{a}_{22}$. Hence the partial derivatives of $\widehat{\beta}_1$ with respect to $\widehat{a}_{21}$ and $\widehat{a}_{22}$ are

$$\frac{\partial\widehat{\beta}_1}{\partial\widehat{a}_{21}} = \frac{1}{\widehat{a}_{22}}, \quad \frac{\partial\widehat{\beta}_1}{\partial\widehat{a}_{22}} = -\frac{\widehat{a}_{21}}{\widehat{a}_{22}^2}.$$

By employing the delta method, the asymptotic distribution of $\widehat{\beta}_1$ can be derived to be normal with mean $\beta_1 = a_{21}/a_{22} - 1$ and variance

$$\begin{aligned}
\delta_1 &= \left(\frac{1}{a_{22}}, -\frac{a_{21}}{a_{22}^2}\right) V \left(\frac{1}{a_{22}}, -\frac{a_{21}}{a_{22}^2}\right)' \\
&= \frac{1}{a_{22}^2}\xi_1 + \frac{a_{21}^2}{a_{22}^4}\xi_2 \\
&= \frac{4pa_{21}^2 + 8(N_1 - 1)a_{41}}{N_1^2 pa_{22}^2} + \frac{4pa_{21}^2 a_{22}^2 + 8(N_2 - 1)a_{21}^2 a_{42}}{N_2^2 pa_{22}^4},
\end{aligned}$$

which completes the proof of Theorem 3.1. □

To use $(\delta_{10})^{-\frac{1}{2}}\widehat{\beta}_1$ in practice, we should estimate δ_{10}, which involves unknown terms a_2 and a_4. In fact, by Lemma 2.1, we can obtain that a consistent estimator of δ_{10} is

$$\widehat{\delta}_{10} = \frac{4}{N_1^2} + \frac{4}{N_2^2} + \left(\frac{8N_1 - 8}{N_1^2 p} + \frac{8N_2 - 8}{N_2^2 p}\right)\frac{\widehat{a}_4}{\widehat{a}_2^2},$$

where $\widehat{a}_4 = (N_1\widehat{a}_{41} + N_2\widehat{a}_{42})/(N_1 + N_2)$ and $\widehat{a}_2 = (N_1\widehat{a}_{21} + N_2\widehat{a}_{22})/(N_1 + N_2)$, with $\widehat{a}_{4i}$ and $\widehat{a}_{2i}$ given in (8) and (6), respectively, $i = 1, 2$. By employing Slutzky's theorem, under $H_0 : \Sigma_1 = \Sigma_2$ and Assumptions $1 \sim 5$, we have

$$T_1 := (\widehat{\delta}_{10})^{-\frac{1}{2}}\widehat{\beta}_1 \xrightarrow{D} N(0, 1). \tag{13}$$

Hence, we can test the hypotheses (1) at significance level α by

$$\text{rejecting } H_0 \Leftrightarrow |T_1| > Z_{\frac{\alpha}{2}},$$

where $Z_{\frac{\alpha}{2}}$ is upper $\frac{\alpha}{2}$ quantile of the standard normal distribution.

3.2 The second test procedure

We have proposed a test statistic based merely on $\mathrm{tr}\Sigma_1^2$ and $\mathrm{tr}\Sigma_2^2$ in the subsection above. However, we note that $\Sigma_1 = \Sigma_2$ meaning that $\mathrm{tr}\Sigma_1^2 = \mathrm{tr}\Sigma_2^2$ and $\mathrm{tr}\Sigma_1 = \mathrm{tr}\Sigma_2$. Realizing the difference between $\mathrm{tr}\Sigma_1$ and $\mathrm{tr}\Sigma_2$ may also have influence on the difference between Σ_1 and Σ_2. Therefore, $\mathrm{tr}\Sigma_i^2$ in conjuction with $\mathrm{tr}\Sigma_i$ may provide us another idea to construct the measure of distance between the null and alternative hypotheses. In fact, we can define another distance function

$$\beta_2 = \frac{\mathrm{tr}\Sigma_1^2/(\mathrm{tr}\Sigma_1)^2}{\mathrm{tr}\Sigma_2^2/(\mathrm{tr}\Sigma_2)^2} - 1. \tag{14}$$

Similar to β_1, β_2 also takes the value zero if H_0 is true, nonzero if H_1 is true, since $\mathrm{tr}\Sigma_i^2/(\mathrm{tr}\Sigma_i)^2$ is a monotone increasing function of the ordered eigenvalues of Σ_i. Thus, we can construct the second test statistic for testing the hypotheses (1) by

$$\widehat{\beta}_2 = \frac{\widehat{a}_{21}/(\widehat{a}_{11})^2}{\widehat{a}_{22}/(\widehat{a}_{12})^2} - 1, \tag{15}$$

where $\widehat{a}_{2i}$ and $\widehat{a}_{1i}$ are given by (6) and (5), respectively. Mention that $\widehat{\beta}_2$ is a consistent estimator of β_2. A simple calculation in the proof of Theorem 3.2 indicates that the asymptotic variance of $\widehat{\beta}_2$ is

$$\delta_2 = \frac{a_{12}^4}{a_{22}^2}\eta_1 + \frac{a_{21}^2 a_{12}^8}{a_{11}^4 a_{22}^4}\eta_2, \tag{16}$$

where

$$\eta_i = \frac{8a_{2i}^3}{N_i p a_{1i}^6} - \frac{16a_{2i}a_{3i}}{N_i p a_{1i}^5} + \frac{4pa_{2i}^2 + 8(N_i-1)a_{4i}}{N_i^2 p a_{1i}^4}, \quad i = 1, 2.$$

Obviously, under $H_0 : \Sigma_1 = \Sigma_2 = \Sigma$, we have $a_{l1} = a_{l2} = a_l$, $l = 1, 2, 3, 4$. Thus δ_2 reduces to

$$\delta_{20} = \frac{a_1^4 (\eta_{10} + \eta_{20})}{a_2^2}, \tag{17}$$

with $\eta_{i0} = \frac{8a_2^3}{N_i p a_1^6} - \frac{16a_2a_3}{N_i p a_1^5} + \frac{4pa_2^2+8(N_i-1)a_4}{N_i^2 p a_1^4}$. Furthermore, the asymptotic normality of $\widehat{\beta}_2$ can obtained by the following theorem.

Theorem 3.2. *Under Assumptions* $1 \sim 5$*, we have*

$$(\delta_2)^{-\frac{1}{2}}(\widehat{\beta}_2 - \beta_2) \xrightarrow{D} N(0, 1).$$

In particular, under H_0*, we have*

$$(\delta_{20})^{-\frac{1}{2}}\widehat{\beta}_2 \xrightarrow{D} N(0, 1),$$

where δ_2 *and* δ_{20} *are given in* (16) *and* (17)*, respectively.*

Proof. To simplify the presentation, we set $\omega_i = \widehat{a}_{2i}/\widehat{a}_{1i}^2$, $i = 1, 2$. First, we derive the distribution of (ω_1, ω_2). Note that ω_i is a function of $\widehat{a}_{1i}$ and $\widehat{a}_{2i}$, hence the partial derivatives of ω_i with respect to $\widehat{a}_{1i}$ and $\widehat{a}_{2i}$ are

$$\frac{\partial \omega_i}{\partial \widehat{a}_{1i}} = -\frac{2\widehat{a}_{2i}}{\widehat{a}_{1i}^3}, \quad \frac{\partial \omega_i}{\partial \widehat{a}_{2i}} = \frac{1}{\widehat{a}_{1i}^2}.$$

By employing the delta method and Lemma 2.2, the asymptotic distribution of ω_i can be derived to be normal with mean a_{2i}/a_{1i}^2 and variance

$$\begin{aligned}\eta_i &= \left(-\frac{2a_{2i}}{a_{1i}^3}, \frac{1}{a_{1i}^2}\right) U \left(-\frac{2a_{2i}}{a_{1i}^3}, \frac{1}{a_{1i}^2}\right)' \\ &= \frac{8a_{2i}^3}{N_i p a_{1i}^6} - \frac{16a_{2i}a_{3i}}{N_i p a_{1i}^5} + \frac{4pa_{2i}^2 + 8(N_i-1)a_{4i}}{N_i^2 p a_{1i}^4}.\end{aligned}$$

Thus, the distribution of (ω_1, ω_2) is obtained as follows:

$$W^{-\frac{1}{2}} \left(\omega_1 - a_{21}/a_{11}^2, \omega_2 - a_{22}/a_{12}^2\right)' \xrightarrow{D} N_2(\mathbf{0}, I_2),$$

where $W = \mathrm{diag}(\eta_1, \eta_2)$.

Next, we can proceed along the lines of the proof of Theorem 3.1 and then one gives us the distribution of $\widehat{\beta}_2 = \omega_1/\omega_2 - 1$. This finishes the proof of Theorem 3.2. □

To use $(\delta_{20})^{-\frac{1}{2}}\widehat{\beta}_2$ in practice, we consider a consistent estimator of δ_{20} given by

$$\widehat{\delta}_{20} = \frac{\widehat{a}_1^4\,(\widehat{\eta}_{10}+\widehat{\eta}_{20})}{\widehat{a}_2^2},$$

where

$$\widehat{\eta}_{i0} = \frac{8\widehat{a}_2^3}{N_i p\widehat{a}_1^6} - \frac{16\widehat{a}_2\widehat{a}_3}{N_i p\widehat{a}_1^5} + \frac{4p\widehat{a}_2^2 + 8(N_i-1)\widehat{a}_4}{N_i^2 p\widehat{a}_1^4}, \quad i=1,2,$$

with $\widehat{a}_l = (N_1\widehat{a}_{l1} + N_2\widehat{a}_{l2})/(N_1+N_2)$, $l=1,2,3,4$. We have

$$T_2 := (\widehat{\delta}_{20})^{-\frac{1}{2}}\widehat{\beta}_2 \xrightarrow{D} N(0,1). \tag{18}$$

Hence, we can test the hypotheses (1) at significance level α by

$$\text{rejecting } H_0 \Leftrightarrow |T_2| > Z_{\frac{\alpha}{2}},$$

where $Z_{\frac{\alpha}{2}}$ is upper $\frac{\alpha}{2}$ quantile of the standard normal distribution.

4. Numerical simulations

In this section, the performance of the proposed tests T_1 and T_2 given in (13) and (18) respectively, are evaluated by Monte Carlo simulation. For the purpose of comparison, we also consider two existing methods: one was discussed in [5], referred as T_{SC}, and the other was proposed by [8], referred as T_{SR}. Throughout the simulation, we set $\mu_1 = \mu_2 = 0$ and $N_1 = N_2$ without loss of generality.

We generate the samples from two scenarios of distributions:

$$\text{(I)}\ \ z_{ija} \sim N(0,1), \quad \text{(II)}\ \ z_{ija} \sim \Gamma(4,0.5) - 2.$$

In the scenario (I), z_{ija} is a standard normal distributed random variable, and in the scenario (II), z_{ija} is actually a centralized Gamma distributed random variable.

To investigate the empirical sizes of these tests, we set the null hypothesis as follows:

$$H_0 : \Sigma_1 = \Sigma_2 = \Sigma = (\sigma_{st})_{p\times p}, \tag{19}$$

where $\sigma_{st} = \mathrm{I}(s=t) + \exp(-|s-t|/2)$, $s,t = 1,\ldots,p$, and $\mathrm{I}(\cdot)$ denotes the indicator function.

To discuss the empirical powers of T_1, T_2, T_{SC} and T_{SR}, we design the following two alternative hypotheses:

$$H_1^a : \Sigma_1 = \Sigma, \quad \Sigma_2 = \Sigma + \text{banded}(5, 0.4)$$

and

$$H_1^b : \Sigma_1 = \Sigma, \quad \Sigma_2 = \Sigma + 0.8I_p,$$

where Σ is given in (19) and banded$(5, 0.4)$ denotes a banded matrix with the bandwidth 5, i.e., we can generate $\widetilde{X}_{ij} = (\widetilde{x}_{ij1}, \ldots, \widetilde{x}_{ijp})'$ from a moving average model of order 5: $\widetilde{x}_{ija} = \widetilde{z}_{ija} + \sum_{l=1}^{5} 0.4\widetilde{z}_{ija-l}$, with $\widetilde{z}_{ija}$ generated from the scenario (I) or (II). The nominal significance level is set at $\alpha = 0.05$ and the number of replications is 1000.

Begin by sampling $N_1 = N_2 = 120$ observations generated from the scenario (II) with $p = 240$. We first show the QQ plots of the proposed tests T_1 and T_2, under H_0, H_1^a and H_1^b in Figures 1, 2 and 3, respectively. These figures indicate that the normality result appears to be satisfied, which validates the theoretical results of Theorem 3.1 and Theorem 3.2. We also obtain some QQ plots when the samples come from the scenario (I). However, there are not substantially different from Figures $1 \sim 3$ and thus we do not list those here.

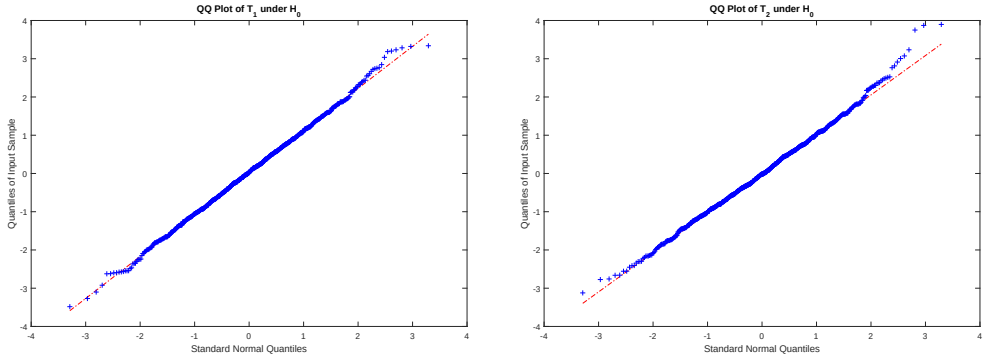

Figure 1: QQ Plots of T_1 (the left panel) and T_2 (the right panel) under H_0, with the sample generated from the scenario (II) of Gamma distribution and $p = 240$, $N_1 = N_2 = 120$.

Table 1 reports the sizes of the tests T_1, T_2, T_{SC} and T_{SR}. The results show that under the scenario (I) of normal distribution, the sizes of these tests are similar and reasonable, which are close to the nominal level 0.05. However, under the scenario (II) of Gamma distribution, T_{SC} encounters serious size distortion, while the proposed two tests T_1 and T_2 as well as T_{SR} still have reasonable sizes. This is understandable because T_{SC} is designed based on the normality assumption, while others are not.

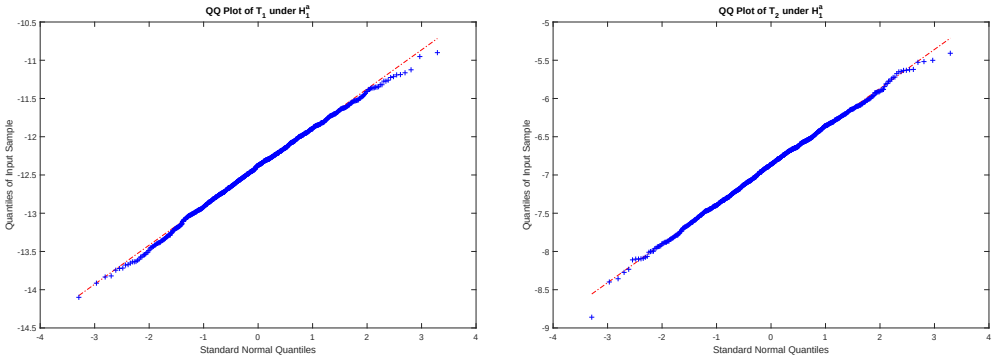

Figure 2: QQ Plots of T_1 (the left panel) and T_2 (the right panel) under H_1^a, with the sample generated from the scenario (II) of Gamma distribution and $p = 240$, $N_1 = N_2 = 120$.

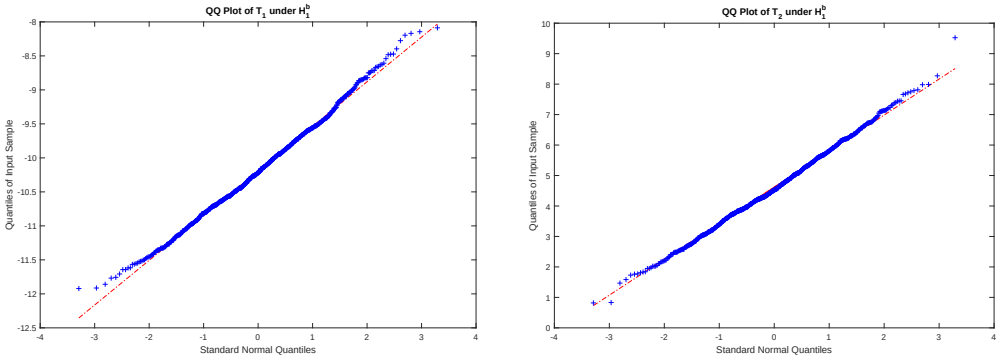

Figure 3: QQ Plots of T_1 (the left panel) and T_2 (the right panel) under H_1^b, with the sample generated from the scenario (II) of Gamma distribution and $p = 240$, $N_1 = N_2 = 120$.

Table 1: Empirical sizes of the tests T_1, T_2, T_{SC} and T_{SR}

p	$N_1 = N_2$	Scenario (I)				Scenario (II)			
		T_{SC}	T_{SR}	T_1	T_2	T_{SC}	T_{SR}	T_1	T_2
60	20	0.040	0.042	0.062	0.079	0.126	0.066	0.072	0.079
	40	0.046	0.046	0.049	0.067	0.128	0.046	0.066	0.061
	80	0.044	0.048	0.057	0.065	0.149	0.064	0.056	0.052
	120	0.054	0.054	0.048	0.056	0.150	0.054	0.072	0.057
120	20	0.032	0.036	0.064	0.058	0.107	0.044	0.074	0.076
	40	0.056	0.058	0.055	0.064	0.108	0.044	0.061	0.054
	80	0.049	0.049	0.052	0.050	0.130	0.050	0.057	0.047
	120	0.058	0.059	0.043	0.045	0.142	0.043	0.065	0.055
180	20	0.042	0.044	0.069	0.069	0.110	0.052	0.078	0.075
	40	0.046	0.052	0.060	0.052	0.126	0.046	0.054	0.051
	80	0.046	0.050	0.056	0.065	0.130	0.047	0.056	0.047
	120	0.045	0.045	0.059	0.059	0.140	0.061	0.070	0.062
240	20	0.034	0.044	0.075	0.061	0.105	0.045	0.075	0.081
	40	0.046	0.049	0.068	0.068	0.126	0.049	0.057	0.051
	80	0.051	0.055	0.059	0.045	0.130	0.048	0.067	0.053
	120	0.055	0.058	0.040	0.058	0.157	0.060	0.060	0.061

Table 2: Empirical powers of the tests T_1, T_2, T_{SC} and T_{SR} under H_1^a

p	$N_1 = N_2$	Scenario (I)				Scenario (II)			
		T_{SC}	T_{SR}	T_1	T_2	T_{SC}	T_{SR}	T_1	T_2
60	20	0.817	0.808	0.951	0.239	0.864	0.784	0.929	0.267
	40	0.999	0.999	1.000	0.673	1.000	0.999	1.000	0.640
	80	1.000	1.000	1.000	0.996	1.000	1.000	1.000	0.994
	120	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
120	20	0.866	0.852	0.997	0.362	0.920	0.843	0.991	0.352
	40	1.000	1.000	1.000	0.883	1.000	1.000	1.000	0.881
	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	120	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
180	20	0.877	0.869	1.000	0.399	0.927	0.857	1.000	0.423
	40	1.000	1.000	1.000	0.946	1.000	1.000	1.000	0.946
	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	120	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
240	20	0.881	0.867	1.000	0.427	0.920	0.857	1.000	0.428
	40	1.000	1.000	1.000	0.964	1.000	1.000	1.000	0.962
	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	120	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tables 2 and 3 provide the powers of these tests under the alternative hypotheses H_1^a and H_1^b, respectively. We observe from Table 2 that under H_1^a, many entries of the powers of these tests approach to 1 in both scenarios (I) and (II), which is especially obvious for T_1. Despite the fact that the power of T_2 is lower than other tests when N_i is small, it increases rapidly and approaches to 1 as N_i and p increase. From Table 3, we find that under the alternative hypothesis H_1^b, the proposed tests T_1 and T_2 are more powerful than T_{SC} and T_{SR}. For instance, under the scenario (I) of normal distribution, when $p = 240$ and $N_1 = N_2 = 80$, the powers of T_1 and T_2 are 1 and 0.848, respectively, while the powers of T_{SC} and T_{SR} are 0.589 and 0.586, respectively.

Table 3: Empirical powers of the tests T_1, T_2, T_{SC} and T_{SR} under H_1^b

p	$N_1 = N_2$	Scenario (I)				Scenario (II)			
		T_{SC}	T_{SR}	T_1	T_2	T_{SC}	T_{SR}	T_1	T_2
60	20	0.088	0.093	0.572	0.226	0.210	0.096	0.490	0.219
	40	0.232	0.223	0.960	0.366	0.424	0.230	0.931	0.326
	80	0.523	0.531	1.000	0.688	0.719	0.521	1.000	0.680
	120	0.839	0.839	1.000	0.886	0.923	0.836	1.000	0.878
120	20	0.099	0.103	0.708	0.211	0.213	0.100	0.675	0.236
	40	0.230	0.225	0.998	0.393	0.390	0.222	0.998	0.416
	80	0.558	0.567	1.000	0.785	0.744	0.538	1.000	0.776
	120	0.832	0.834	1.000	0.967	0.938	0.833	1.000	0.965
180	20	0.092	0.105	0.795	0.217	0.207	0.094	0.752	0.214
	40	0.224	0.229	1.000	0.405	0.380	0.214	0.999	0.389
	80	0.554	0.556	1.000	0.856	0.769	0.561	1.000	0.801
	120	0.842	0.845	1.000	0.984	0.935	0.848	1.000	0.973
240	20	0.093	0.093	0.839	0.215	0.189	0.101	0.802	0.199
	40	0.200	0.208	1.000	0.428	0.389	0.215	1.000	0.423
	80	0.589	0.586	1.000	0.848	0.773	0.575	1.000	0.852
	120	0.839	0.838	1.000	0.976	0.939	0.817	1.000	0.991

5. Conclusion

In this concluding section, we provide a short summary of our main results on the hypothesis test about the homogeneity of covariance matrices for non-normal high-dimensional data. Firstly, two new test statistics are proposed using the ratio between the unbiased and consistent estimators of the trace of covariance matrices. Secondly, the asymptotic normality of the proposed test statistics are proved. Finally, numerical simulations demonstrate that the proposed tests, which are flexible to the underlying distribution, perform well and better in several cases than some existing tests and are thus recommended.

Acknowledgements

The author is deeply indebted to the referees for their many useful comments which greatly improved the manuscript.

References

[1] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, A.J. Levine, *Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays*, Proceedings of the National Academy of Sciences of the United States of America, 96 (1999), 6745-6750.

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

[3] S.X. Chen, L.X. Zhang, P.S. Zhong, *Tests for high-dimensional covariance matrices*, Journal of the American Statistical Association, 105 (2010), 810-819.

[4] M. Dettling, P. Bühlmann, *Boosting for tumor classification with gene expression data,* Bioinformatics, 19 (2003), 1061-1069.

[5] J.R. Schott, *A test for the equality of covariance matrices when the dimension is large relative to the sample sizes*, Computational Statistics and Data Analysis, 51 (2007), 6535-6542.

[6] M.S. Srivastava, *Some tests concerning the covariance matrix in high dimensional data*, Journal of the Japan Statistical Society, 35 (2005), 251-272.

[7] M.S. Srivastava, H. Yanagihara, *Testing the equality of several covariance matrices with fewer observations than the dimension,* Journal of Multivariate Analysis, 101 (2010), 1319-1329.

[8] M.S. Srivastava, H. Yanagihara, T. Kubokawa, *Tests for covariance matrices in high dimension with less sample size,* Journal of Multivariate Analysis, 130 (2014), 89-309.

[9] X. Tian, Y. Lu, W. Li, *A robust test for sphericity of high-dimensional covariance matrices*, Journal of Multivariate Analysis, 141 (2015), 217-227.

Accepted: 22.06.2019