Statistica Sinica **33** (2023), 2359-2380
doi:https://doi.org/10.5705/ss.202020.0486

USE OF RANDOM INTEGRATION TO TEST EQUALITY OF HIGH DIMENSIONAL COVARIANCE MATRICES

Yunlu Jiang[1], Canhong Wen[2], Yukang Jiang[3],
Xueqin Wang[2] and Heping Zhang[4]

[1]*Jinan University,* [2]*University of Science and Technology of China,*
[3]*Sun Yat-Sen University and* [4]*Yale University*

Abstract: Testing the equality of two covariance matrices is a fundamental problem in statistics, and especially challenging when the data are high dimensional. By means of a novel use of random integration, we test the equality of high-dimensional covariance matrices without assuming parametric distributions for the two underlying populations, even if the dimension is much larger than the sample size. The asymptotic properties of our test for an arbitrary number of covariates and sample size are studied in depth under a general multivariate model. The finite-sample performance of our test is evaluated using numerical studies. The empirical results demonstrate that the proposed test is highly competitive with existing tests in a wide range of settings, and particularly powerful when there exist a few large or many small diagonal disturbances between the two covariance matrices.

Key words and phrases: Covariance matrix, high-dimensional data, random integration.

1. Introduction

The need to test the equality of two covariance matrices arises in many important problems, including both classic experimental designs and analyses of high throughout omic data. For example, gene expression data are often used to classify disease types. Here, Igolkina et al. (2018) points out that variance in gene expression is an important characteristic of schizophrenia. Roberts, Catchpoole and Kennedy (2018) shows that many genes differ in the variances of their gene expressions between disease states. Comparing two covariance matrices is therefore essential when analyzing gene expression microarray data of two different groups. This comparison becomes difficult, because the number of samples is usually much smaller than the number of genes. Although, many methods have been proposed to test the equality of two covariance matrices, they tend to

Corresponding author: Heping Zhang, Department of Biostatistics, Yale University, New Haven, CT 06520- 8034, USA. E-mail: heping.zhang@yale.edu.

perform poorly in practice, especially when there are a few large or many small diagonal disturbances between the two covariance matrices.

Let $\mathbf{X}$ and $\mathbf{Y}$ be p-dimensional random variables with covariance matrices Σ_1 and Σ_2, respectively. Given independent samples $\mathscr{X}_m = \{\mathbf{X}_1, \ldots, \mathbf{X}_m\}$ from $\mathbf{X}$ and $\mathscr{Y}_n = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$ from $\mathbf{Y}$, we want to test

$$H_0 : \Sigma_1 = \Sigma_2 \quad \text{vs.} \quad H_1 : \Sigma_1 \neq \Sigma_2. \tag{1.1}$$

In the classic low-dimensional setting, Anderson (2003) proposes a likelihood ratio test (LRT) statistic, showing that it asymptotically follows a χ^2-distribution with degrees of freedom $p(p+1)/2$ under the multivariate normality assumption and H_0 when p is fixed.

In recent applications in the fields of gene expression (Pan et al. (2018)), neuroimaging (Le Bihan et al. (2001)), and risk management (Bollerslev, Meddahi and Nyawa (2019)), the dimension p can be much larger than the sample size n, that is, "large p small n" or "large p large n." In this setting, the sample covariance matrix does not converge to its population counterpart (Bai et al. (2009)), and the aforementioned classical methods for the low-dimensional case either are not applicable or perform poorly. As well summarized by Cai (2017), this problem is so important and difficult that it has attracted a great deal of attention. We briefly review some of the methods below.

Modified LRT-based methods have been considered by Bai et al. (2009), Jiang and Yang (2013), and Jiang and Qi (2015). Because $\Sigma_1 = \Sigma_2$ is equivalent to the Frobenius norm $tr(\Sigma_1 - \Sigma_2)^2 = 0$, Frobenius norm-based tests have also been proposed (Schott (2007); Srivastava and Yanagihara (2010)). However, both sets of methods assume multivariate normality. Li and Chen (2012) remove the normality assumption by using a linear combination of three one-sample U-statistics. The test proposed by Li and Chen (2012) can be applied without assuming parametric distributions for the two populations, and is very powerful when there are many small differences between two population covariance matrices. However, this test is found to lack power against sparse alternatives (Yang and Pan (2017)) or when the two covariance matrices differ slightly only in the diagonal (Wu and Li (2015)). Using random projection, Wu and Li (2015) construct a test statistic to improve the power when there are many small diagonal disturbances between the two covariance matrices, but again assumed normality. He et al. (2021) introduce adaptive testing to combine the finite-order U-statistics, which include the variants of Frobenius norm-based statistics.

Assuming sparsity for $\Sigma_1 - \Sigma_2$ under the alternative hypothesis, Cai, Liu and Xia (2013) introduce an extreme statistic that is robust with respect to the population distributions and is very powerful when there are only several large disturbances between the two population covariance matrices, and Chang et al. (2017) propose a computationally fast procedure. Zhu et al. (2017) construct a sparse-leading-eigenvalue-driven (sLED) test for when the signals are both sparse and weak, and prove that the sLED test achieves full power asymptotically when the sparse signal is sufficiently strong. However, these tests are either not powerful when there are many small disturbances between Σ_1 and Σ_2, or the theoretical properties require an explicit relationship between n and p or are too complicated for practical use.

The aforementioned tests often focus on specific situations. To accommodate more varied situations, several weighted combination tests have been proposed. For example, Yang and Pan (2017) propose a weighted test statistic based on random matrix theory. Zheng et al. (2020) introduce a power enhancement high-dimensional test. These tests can handle both sparse and nonsparse structures. However, the tests depend on a proper choice of weights, which is a challenging task. Furthermore, these procedures require an explicit relationship between n and p. Yu, Li and Xue, (2022) consider a scale-invariant power enhancement test based on Fisher's method, but again assumed normality.

In summary, although many methods exist, they have various limitations. By means of a novel use of random integration, we propose a method that tests the equality of two high-dimensional covariance matrices. Specifically, the proposed test possesses the following three merits:

1. It does not require a distributional assumption.
2. It works well for the paradigm of "large p", even when there are a few large or many small diagonal disturbances between the two covariance matrices.
3. The asymptotic theory is established under a general multivariate model with certain moment conditions, but without requiring an explicit relationship between p and n.

The rest of the paper is organized as follows. In Section 2, we introduce our test statistic using the random integration technique, and establish its asymptotic properties. In Section 3, we present simulation studies that evaluate the finite-sample performance of the proposed test. In Section 4, a real data set is analyzed to compare the proposed test with several existing methods. Section 5 concludes the paper. All technical proofs are presented in the Supplementary Material.

2. Methodology and Main Results

To introduce a proper statistic to test (1.1), especially when there are many small diagonal disturbances between the two covariance matrices, it would be helpful if we can strengthen the information of diagonal disturbances for $\Sigma_1 - \Sigma_2$. Unlike the random matrix projection method, which needs the normality assumption to strengthen the information on a line, we develop a random integration technique to strengthen this information using integrations. To this end, we first denote $\mathbf{X}^c = \mathbf{X} - \mu_1$ and $\mathbf{Y}^c = \mathbf{Y} - \mu_2$, where $\mu_1 = E\mathbf{X}$ and $\mu_2 = E\mathbf{Y}$. Then, $\Sigma_1 = E\mathbf{X}^c\mathbf{X}^{c\top}$ and $\Sigma_2 = E\mathbf{Y}^c\mathbf{Y}^{c\top}$. Note the following equivalences:

$$\begin{aligned}\Sigma_1 = \Sigma_2 &\Leftrightarrow E\mathbf{X}^c\mathbf{X}^{c\top} = E\mathbf{Y}^c\mathbf{Y}^{c\top} \\ &\Leftrightarrow \boldsymbol{\alpha}^\top E\mathbf{X}^c\mathbf{X}^{c\top}\boldsymbol{\alpha} = \boldsymbol{\alpha}^\top E\mathbf{Y}^c\mathbf{Y}^{c\top}\boldsymbol{\alpha}, \text{for any } \boldsymbol{\alpha} \in R^p \\ &\Leftrightarrow E\left\{\boldsymbol{\alpha}^\top\left(\mathbf{X}^c\mathbf{X}^{c\top} - \mathbf{Y}^c\mathbf{Y}^{c\top}\right)\boldsymbol{\alpha}\right\} = 0, \text{for any } \boldsymbol{\alpha} \in R^p.\end{aligned}$$

Thus, testing whether Σ_1 and Σ_2 amounts to testing whether

$$\mathrm{RI}_w(\mathbf{X},\mathbf{Y}) \triangleq \int E^2\left\{\boldsymbol{\alpha}^\top\left(\mathbf{X}^c\mathbf{X}^{c\top} - \mathbf{Y}^c\mathbf{Y}^{c\top}\right)\boldsymbol{\alpha}\right\} w(\boldsymbol{\alpha})d\boldsymbol{\alpha} = 0, \qquad (2.1)$$

where $w(\boldsymbol{\alpha})$ is a positive weight function. A critical observation is that $\mathrm{RI}_w(\mathbf{X},\mathbf{Y})$ can be evaluated easily for a certain properly chosen w. Theorem 1 enables us to derive an explicit form for (2.1) and obtain our difference measure of two covariances.

Theorem 1. *If $w(\boldsymbol{\alpha})$ is a p-dimensional standard normal density function, then*

$$RI(\mathbf{X},\mathbf{Y}) \triangleq RI_w(\mathbf{X},\mathbf{Y}) = \{tr(\Sigma_1 - \Sigma_2)\}^2 + 2tr\left(\Sigma_1 - \Sigma_2\right)^2, \qquad (2.2)$$

and $RI(\mathbf{X},\mathbf{Y}) \geq 0$, with the equality holding if and only if $\Sigma_1 = \Sigma_2$.

Remark 1. The p-dimensional standard normal random vector can be expressed as the independent product between a uniformly distributed random vector on the unit sphere S_1^{p-1} and the radial random variable. We choose the standard normal random density function to treat every unit vector equally in the integration. Moreover, the multivariate normal distribution is a special case of a multivariate stable distribution, and multivariate stable distributions are used as weight functions by Chen, Meintanis and Zhu (2019). As future work, it may be worthwhile considering other weight functions, such as the uniform distribution (Zhu et al. (2017); Kim, Balakrishnan and Wasserman (2020)) or the Bernoulli distribution (Qiu, Xu and Zhu (2021).

Note that the second term on the right-hand size of (2.2) is the Frobenius norm of the difference between the two covariance matrices. The test can be designed to be powerful when there are many small disturbances between Σ_1 and Σ_2. In contrast to Li and Chen (2012), the same test can be powerful when the two covariance matrices differ only by a small amount in the diagonal, owing to the first term on the right-hand side of (2.2), which represents the square of the difference between the diagonal elements of the two covariance matrices. If the nonzero signals of the difference between the diagonal elements of Σ_1 and Σ_2 are weakly dense with almost the same sign, $\mathrm{RI}(\mathbf{X},\mathbf{Y})$ should be more powerful than a test statistic in which $[tr(\Sigma_1-\Sigma_2)]^2$ is replaced with $\sum_{k=1}^{p}(\Sigma_{1,kk}-\Sigma_{2,kk})^2$. In addition, the estimation of $\mathrm{RI}(\mathbf{X},\mathbf{Y})$ does not need consistent estimates of the covariance matrices, because $\mathrm{RI}(\mathbf{X},\mathbf{Y})$ is based on the trace of the matrices. In the following, we obtain an unbiased estimator of $\mathrm{RI}(\mathbf{X},\mathbf{Y})$, the test statistic we need to test (1.1).

Denote

$$
\begin{aligned}
A_m^1 &= \frac{1}{m(m-1)}\sum_{i\neq j}(\mathbf{X}_i^\top\mathbf{X}_i)(\mathbf{X}_j^\top\mathbf{X}_j) - \frac{2}{m(m-1)(m-2)}\sum_{i,j,k}^{\star}\mathbf{X}_i^\top\mathbf{X}_j\mathbf{X}_k^\top\mathbf{X}_k \\
&\quad + \frac{1}{m(m-1)(m-2)(m-3)}\sum_{i,j,k,l}^{\star}\mathbf{X}_i^\top\mathbf{X}_j\mathbf{X}_k^\top\mathbf{X}_l, \\
B_n^1 &= \frac{1}{n(n-1)}\sum_{i\neq j}(\mathbf{Y}_i^\top\mathbf{Y}_i)(\mathbf{Y}_j^\top\mathbf{Y}_j) - \frac{2}{n(n-1)(n-2)}\sum_{i,j,k}^{\star}\mathbf{Y}_i^\top\mathbf{Y}_j\mathbf{Y}_k^\top\mathbf{Y}_k \\
&\quad + \frac{1}{n(n-1)(n-2)(n-3)}\sum_{i,j,k,l}^{\star}\mathbf{Y}_i^\top\mathbf{Y}_j\mathbf{Y}_k^\top\mathbf{Y}_l, \\
C_{m,n}^1 &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(\mathbf{X}_i^\top\mathbf{X}_i)(\mathbf{Y}_j^\top\mathbf{Y}_j) - \frac{1}{nm(m-1)}\sum_{i,k}^{\star}\sum_{j=1}^{n}\mathbf{Y}_j^\top\mathbf{Y}_j\mathbf{X}_i^\top\mathbf{X}_k \\
&\quad - \frac{1}{mn(n-1)}\sum_{i,k}^{\star}\sum_{j=1}^{m}\mathbf{X}_j^\top\mathbf{X}_j\mathbf{Y}_i^\top\mathbf{Y}_k + \frac{1}{m(m-1)n(n-1)}\sum_{i,k}^{\star}\sum_{j,l}^{\star}\mathbf{X}_i^\top\mathbf{X}_k\mathbf{Y}_j^\top\mathbf{Y}_l, \\
A_m^2 &= \frac{1}{m(m-1)}\sum_{i\neq j}(\mathbf{X}_i^\top\mathbf{X}_j)^2 - \frac{2}{m(m-1)(m-2)}\sum_{i,j,k}^{\star}\mathbf{X}_i^\top\mathbf{X}_j\mathbf{X}_j^\top\mathbf{X}_k \\
&\quad + \frac{1}{m(m-1)(m-2)(m-3)}\sum_{i,j,k,l}^{\star}\mathbf{X}_i^\top\mathbf{X}_j\mathbf{X}_k^\top\mathbf{X}_l, \\
B_n^2 &= \frac{1}{m(m-1)}\sum_{i\neq j}(\mathbf{Y}_i^\top\mathbf{Y}_j)^2 - \frac{2}{m(m-1)(m-2)}\sum_{i,j,k}^{\star}\mathbf{Y}_i^\top\mathbf{Y}_j\mathbf{Y}_j^\top\mathbf{Y}_k
\end{aligned}
$$

$$+\frac{1}{m(m-1)(m-2)(m-3)}\sum_{i,j,k,l}^{\star}\mathbf{Y}_i^\top\mathbf{Y}_j\mathbf{Y}_k^\top\mathbf{Y}_l,$$

$$C_{m,n}^2=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(\mathbf{X}_i^\top\mathbf{Y}_j)^2-\frac{1}{nm(m-1)}\sum_{i,k}^{\star}\sum_{j=1}^{n}\mathbf{X}_i^\top\mathbf{Y}_j\mathbf{Y}_j^\top\mathbf{X}_k$$

$$-\frac{1}{mn(n-1)}\sum_{i,k}^{\star}\sum_{j=1}^{m}\mathbf{Y}_i^\top\mathbf{X}_j\mathbf{X}_j^\top\mathbf{Y}_k+\frac{1}{m(m-1)n(n-1)}\sum_{i,k}^{\star}\sum_{j,l}^{\star}\mathbf{X}_i^\top\mathbf{Y}_j\mathbf{X}_k^\top\mathbf{Y}_l,$$

where $\sum^\star$ denotes a summation over mutually distinct indices. Then, the proposed sample test statistic is

$$\mathrm{RI}_{m,n}=A_m^1-2C_{m,n}^1+B_n^1+2(A_m^2-2C_{m,n}^2+B_n^2), \tag{2.3}$$

which is an unbiased estimator of $\mathrm{RI}(\mathbf{X},\mathbf{Y})$.

Remark 2. The computation cost is of the order of $\max\{pn^4, pm^4\}$ if we compute $\mathrm{RI}_{m,n}$ directly. The computational burden comes from the last two sums in A_m^1, B_n^1, A_m^2, and B_n^2 and the last three sums in $C_{m,n}^1$ and $C_{m,n}^2$. Because $\mathrm{RI}_{m,n}$ is invariant under the location shift, we can assume without loss of generality that $\boldsymbol{\mu}_1=\boldsymbol{\mu}_2=0$. Under this assumption, the last two sums in A_m^1, B_n^1, A_m^2, and B_n^2 and the last three sums in $C_{m,n}^1$ and $C_{m,n}^2$ are all of a smaller order than the first. Thus, we can first transform the data $\mathbf{X}_i$ to $\mathbf{X}_i-\bar{\mathbf{X}}$ and $\mathbf{Y}_j$ to $\mathbf{Y}_j-\bar{\mathbf{Y}}$, and then compute only the first term in $A_m^1, B_n^1, A_m^2, B_n^2, C_{m,n}^1$, and $C_{m,n}^2$. This reduces the computation cost to the order of $\max\{pn^2, pm^2\}$, without affecting the asymptotic properties of our proposed test. Using similar techniques to those of Zhong Li and Santo (2019), we can reduce the computation cost for the third terms in A_m^1 and B_n^1, the last two sums in A_m^2 and B_n^2, and the last three in $C_{m,n}^2$ to the order of $\max\{pn^2, pm^2\}$, and can maintain their unbiasedness. However, the second terms in A_m^1 and B_n^1 and the second and the last three in $C_{m,n}^1$ cannot reduce the order of $\max\{pn^2, pm^2\}$ to maintain their unbiasedness using this technique.

2.1. Asymptotic properties

To establish the limiting distribution of $\mathrm{RI}_{m,n}$, we assume the following three conditions:

E1. There exist a $p\times m_1$ matrix Γ_1, a $p\times m_2$ matrix Γ_2, m_1-dimensional random vectors $\{\mathbf{Z}_{1j}\}_{j=1}^m$, and m_2-dimensional random vectors $\{\mathbf{Z}_{2j}\}_{j=1}^n$, such that $\mathbf{X}_j=\boldsymbol{\mu}_1+\Gamma_1\mathbf{Z}_{1j}$ for $j=1,\ldots,m$, and $\mathbf{Y}_j=\boldsymbol{\mu}_2+\Gamma_2\mathbf{Z}_{2j}$ for $j=1,\ldots,n$. In addition Γ_i for $i=1,2$, and $\mathbf{Z}_{ij}=(\mathbf{Z}_{ij1},\ldots,\mathbf{Z}_{ijm_i})^\top$ for $i=1, j=1,\ldots,m$

and $i = 2, j = 1, \ldots, n$ satisfy the following:

(a) $\Gamma_1 \Gamma_1^\top = \Sigma_1$ and $\Gamma_2 \Gamma_2^\top = \Sigma_2$, with $\min\{m_1, m_2\} \geq p$.

(b) $\{\mathbf{Z}_{1j}\}_{j=1}^m$ and $\{\mathbf{Z}_{2j}\}_{j=1}^n$ are independent and identically distributed (i.i.d.), with $E\mathbf{Z}_{1j} = \mathbf{0}$ and $\mathrm{Var}(\mathbf{Z}_{1j}) = \mathbf{I}_{m_1}$, and $E\mathbf{Z}_{2j} = \mathbf{0}$ and $\mathrm{Var}(\mathbf{Z}_{2j}) = \mathbf{I}_{m_2}$, where $\mathbf{I}_{m_i}$ is the $m_i \times m_i$ identity matrix.

(c) $\sup_{i,k} E|Z_{ijk}|^8 < \infty$ and $EZ_{ijk}^4 = 3 + \Delta_i$, for some constant Δ_i. Furthermore,

$$E\left(Z_{ijl_1}^{\varsigma_1} \cdots Z_{ijl_q}^{\varsigma_q}\right) = E\left(Z_{ijl_1}^{\varsigma_1}\right) \cdots E\left(Z_{ijl_q}^{\varsigma_q}\right) \tag{2.4}$$

for any positive integers q and ς_l such that $\sum_{l=1}^q \varsigma_l \leq 8$, and $l_1, l_2, \ldots, l_q$ are distinct indices.

E2. As $\min\{m, n\} \to \infty$, $p \to \infty$, and for any $i, j, k, l \in \{1, 2\}$, $tr(\Sigma_i \Sigma_j) \to \infty$ and

$$tr\{(\Sigma_i \Sigma_j)(\Sigma_k \Sigma_l)\} = o\{tr(\Sigma_i \Sigma_j) tr(\Sigma_k \Sigma_l)\}. \tag{2.5}$$

E3. As $\min\{m, n\} \to \infty$, $m/(m+n) \to \tau \in (0, 1)$.

Remark 3. Condition E1 yields a general multivariate model for high-dimensional data analysis that includes commonly used distributions such as the multivariate normal distribution (Chen, Zhang and Zhong (2010); Srivastava and Yanagihara (2010); Li and Chen (2012)). According to Chen and Qin (2010), $\min\{m_1, m_2\} \geq p$ means that the rank and eigenvalues of Σ_1 or Σ_2 are not affected by the transformation. According to Chen and Qin (2010) and He and Chen (2018), (2.4) can be viewed as a pseudo-independent condition of $\mathbf{Z}_{ij}$, that is a relaxed independence relation that allows some margin over probabilities (Kim and Lesser (2008)). Obviously, if $\mathbf{Z}_{ij}$ has independent components, then (2.4) is true.

In Condition E2, we do not require a direct relationship between p and n. We know of cases in which E2 does not imply a relation between p and n. For example, if all the eigenvalues of Σ_i are bounded away from zero and infinity, E2 holds. Some of the commonly encountered covariance structures satisfy Condition E2 (Chen, Zhang and Zhong (2010)).

Condition E3 is a standard regularity assumption in two-sample problems, and guarantees that m and n go to infinity proportionally.

Theorem 2. *Under Conditions* E1–E3, *as* $\min\{m, n\} \to \infty$, *we have*

$$\frac{RI_{m,n} - RI(\mathbf{X}, \mathbf{Y})}{\sigma_{m,n}} \xrightarrow{\mathscr{D}} \mathcal{N}(0, 1),$$

where $\sigma^2_{m,n}$ *is defined in* (A.2) *in the Supplementary Material.*

Under H_0, we can obtain $\mathrm{RI}(\mathbf{X},\mathbf{Y})=0$ and

$$\sigma^2_{0,m,n}=24\left(\frac{1}{m}+\frac{1}{n}\right)^2 tr^2(\Sigma^2).$$

Therefore, we obtain the following corollary.

Corollary 1. *Under Conditions* E1−E3 *and* $H_0:\Sigma_1=\Sigma_2=\Sigma$, *as* $\min\{m,n\}\to\infty$, *we have*

$$\frac{RI_{m,n}}{\sigma_{0,m,n}}\xrightarrow{\mathscr{D}}\mathcal{N}(0,1).$$

To formulate a test procedure, we need to estimate $\sigma_{0,m,n}$. Because $EA^2_m=tr(\Sigma_1^2)$ and $EB^2_n=tr(\Sigma_2^2)$, the following is a consistent estimate $\hat\sigma_{0,m,n}$ of $\sigma_{0,m,n}$ under H_0:

$$\hat\sigma_{0,m,n}=2\sqrt{6}\left(\frac{1}{m}A^2_m+\frac{1}{n}B^2_n\right).$$

Furthermore, the following theorem ensures that $\hat\sigma_{0,m,n}$ is ratio consistent to $\sigma_{0,m,n}$.

Theorem 3. *Under Conditions* E1−E3 *and* $H_0:\Sigma_1=\Sigma_2=\Sigma$, *as* $\min\{m,n\}\to\infty$, *we have*

$$\frac{RI_{m,n}}{\hat\sigma_{0,m,n}}\xrightarrow{\mathscr{D}}\mathcal{N}(0,1). \tag{2.6}$$

As shown in the Supplementary Material,

$$\frac{A^2_m}{tr(\Sigma_1^2)}\xrightarrow{P}1,\quad \frac{B^2_n}{tr(\Sigma_2^2)}\xrightarrow{P}1,\quad\text{and}\quad \frac{\hat\sigma_{0,m,n}}{\sigma_{0,m,n}}\xrightarrow{P}1.$$

Theorem 3 follows from Corollary 1 and Slutsky's theorem. Therefore, the proposed test with a nominal θ level of significance rejects H_0 if $\mathrm{RI}_{m,n}\geq\hat\sigma_{0,m,n}z_\theta$, where z_θ is the upper-θ quantile of $\mathcal{N}(0,1)$. The approximation results in the Supplementary Material indicate that the standard normal distribution is an adequate substitute for the null distribution of $\mathrm{RI}_{m,n}/\hat\sigma_{0,m,n}$.

Next, we study the power of our proposed test. Let $g_{m,n}(\Sigma_1,\Sigma_2;\theta)=P(\mathrm{RI}_{m,n}\geq\hat\sigma_{0,m,n}z_\theta|H_1)$ be the power of the proposed test under $H_1:\Sigma_1\neq\Sigma_2$. Let $\mathrm{SNR}_{m,n}(\Sigma_1,\Sigma_2)=\mathrm{RI}(\mathbf{X},\mathbf{Y})/\sigma_{m,n}$ and $\gamma_{m,n}=\{tr(\Sigma_1^2)/m+tr(\Sigma_2^2)/n\}/\mathrm{RI}(\mathbf{X},\mathbf{Y})$. Then, we obtain Theorem 4.

Theorem 4. *Under Conditions* E1−E3 *and* $H_1:\Sigma_1\neq\Sigma_2$, *we have*

$$\lim_{m,n\to\infty}g_{m,n}(\Sigma_1,\Sigma_2;\theta)\geq\lim_{m,n\to\infty}\Phi\left\{-\sqrt{2}z_\theta+SNR_{m,n}(\Sigma_1,\Sigma_2)\right\},$$

where $\Phi(\cdot)$ *is the cumulative standard normal distribution function.*

Theorem 4 indicates that the power of our proposed test is bounded from below. The power is determined mainly by $\mathrm{SNR}_{m,n}(\Sigma_1, \Sigma_2)$. From (A.2) in the Supplementary Material, we have

$$\sigma_{m,n}^2 \leq 24\left\{\frac{tr(\Sigma_1^2)}{m} + \frac{tr(\Sigma_2^2)}{n}\right\}^2 \\ +20\max\{2+\Delta_1, 2+\Delta_2\}\left\{\frac{tr(\Sigma_1^2)}{m} + \frac{tr(\Sigma_2^2)}{n}\right\}\mathrm{RI}(\mathbf{X}, \mathbf{Y}),$$

that is,

$$\mathrm{SNR}_{m,n}(\Sigma_1, \Sigma_2) \geq \left(24\gamma_{m,n}^2 + 20\max\{2+\Delta_1, 2+\Delta_2\}\gamma_{m,n}\right)^{-1/2}.$$

Therefore, when $\gamma_{m,n} \to 0$ as $\min\{m, n\} \to \infty$, we have $\mathrm{SNR}_{m,n}(\Sigma_1, \Sigma_2) \to \infty$. Thus, we have

$$\lim_{m,n\to\infty} g_{m,n}(\Sigma_1, \Sigma_2; \theta) = 1.$$

We consider the following three cases.

Case I: Let $\Sigma_1 = r\mathbf{I}_p + AR(0.1)$ and $\Sigma_2 = AR(0.1)$, where $AR(\rho) = (a_{ij})_{p\times p}$ is a covariance matrix with $a_{ij} = \rho^{|i-j|}$, for $i, j = 1, \ldots, p$. Then, we have Corollary 2.

Corollary 2. *Under Conditions* E1−E3 *and* $\lim_{m,n,p\to\infty}\sqrt{p}(m+n)r^2 = c$, *where* $0 < c < \infty$, *we have*

$$\lim_{m,n\to\infty} g_{m,n}(\Sigma_1, \Sigma_2; \theta) = 1.$$

Corollary 2 shows that our proposed test is very powerful when there are many small diagonal disturbances between the two covariance matrices. In addition, under the conditions in Corollary 2, the signal-to-noise ratio of the method proposed by Li and Chen (2012) diminishes to zero. Therefore, the power of their test has a low bounded from below.

Case II: Let $\Sigma_1 = \mathbf{I}_p$ and $\Sigma_2 = \mathbf{I}_p + H(\varpi_0, \varpi_1, p_0)$, where $H(\varpi_0, \varpi_1, p_0) = (h_{ij})_{p\times p}$ with $h_{ij} = 0$, except $h_{ii} = \varpi_0$ for $i = 1, \ldots, p_0$, and $h_{i,i+1} = h_{i+1,i} = \varpi_1$ for $i = 1, \ldots, p_0 - 1$. Then, we have Corollary 3.

Corollary 3. *Under Conditions* E1−E3, $p/\{(m+n)p_0^2\varpi_0^2\} = o(1)$ *and* $np_0\varpi_0 \to \infty$, *we have*

$$\lim_{m,n\to\infty} g_{m,n}(\Sigma_1, \Sigma_2; \theta) = 1.$$

Following Cai, Liu and Xia (2013), we take $\varpi_0 = O(\sqrt{p/(m+n)})$ and $p_0 = p^{1/4}$. Corollary 3 shows that the proposed test is powerful. The method proposed by Cai, Liu and Xia (2013) is also powerful under this case.

Case III: Let $\Sigma_1 = \mathbf{I}_p$ and $\Sigma_2 = \mathbf{I}_p + M$, where M is a $p \times p$ matrix with $M_{ii} = 0$ and $M_{ij} = \omega_1$, for $i \neq j$. Then, we have Corollary 4.

Corollary 4. *Under Conditions* E1–E3*,* $(m+n)p\omega_1^2 \to \infty$ *as* $m, n, p \to \infty$*. Then, we have*

$$\lim_{m,n\to\infty} g_{m,n}(\Sigma_1, \Sigma_2; \theta) = 1.$$

Corollary 4 indicates that our proposed test is also very powerful under some conditions when the diagonals are the same and there are many small non-diagonal disturbances between the two covariance matrices.

3. Simulation Studies

We carry out numerical simulations to investigate the finite-sample performance of our proposed method. We also consider the methods proposed by Li and Chen (2012) (LC), Cai, Liu and Xia (2013) (Cai), Zhu et al. (2017) (sLED), and Chang et al. (2017) ($\Psi_{B,\alpha}$), the T2 method proposed by Zheng et al. (2020), and a scale-invariant power enhancement test based on Fisher's method (Yu, Li and Xue, (2022)) ($F_{m,n}$). We set the nominal level of significance at 0.05. We choose the sample sizes $m = n = 60, 100$ and $m = 200, n = 60$, and the dimension is $p = 300, 500, 800, 1000, 1200, 1500$. All empirical sizes and powers are calculated from 1,000 replications. For Σ_1 and Σ_2, we consider the following six scenarios:

Scenario 1: $\Sigma_1^{(1)} = \mathbf{I}_p$, the identity matrix, $\Sigma_2^{(1)} = \mathbf{I}_p + H(0.04, 0.2, \lfloor 0.3p \rfloor)$, where $H(\varpi_0, \varpi_1, k) = (h_{ij})_{p\times p}$ with $h_{ij} = 0$, except $h_{ii} = \varpi_0$ for $i = 1, \ldots, k$ and $h_{i,i+1} = h_{i+1,i} = \varpi_1$ for $i = 1, \ldots, k-1$;

Scenario 2: $\Sigma_1^{(2)} = \mathbf{I}_p$, $\Sigma_2^{(2)} = \mathbf{I}_p + H(0.04, 0.2, p)$;

Scenario 3: $\Sigma_1^{(3)} = \mathbf{I}_p$, $\Sigma_2^{(3)} = \mathbf{I}_p + \Sigma_*^{(3)}$, where $\Sigma_*^{(3)} = \left(\sigma_{ij,*}^{(3)}\right)$ is a $p \times p$ matrix with $\sigma_{ii,*}^{(3)} = 0$ for $i = 1, \ldots, p$, and $\sigma_{ij,*}^{(3)} = 1/\sqrt{p}$ for $i \neq j$.

Scenario 4: $\Sigma_1^{(4)} = \mathbf{I}_p$, $\Sigma_2^{(4)} = \mathbf{I}_p + H(4, 0.05, \lfloor 0.02p \rfloor)$;

Scenario 5: $\Sigma_1^{(5)} = \Sigma_*^{(5)} + \delta_0 \mathbf{I}_p$ and $\Sigma_2^{(5)} = \Sigma_*^{(5)} + \delta_0 \mathbf{I}_p + U$, where $\Sigma_*^{(5)} = \left(\sigma_{ij,*}^{(5)}\right) = D^{1/2} C D^{1/2}$, $D = \mathrm{diag}(d_1, \ldots, d_p)$ and $d_1, \ldots, d_p \overset{i.i.d.}{\sim} \mathrm{Unif}(0.5, 2.5)$, $C = (c_{ij})$, with $c_{ii} = 1$ and $c_{ij} = 0.5$ for $5(k-1)+1 \leq i \neq j \leq 5k$, where

$k = 1, \ldots, [p/5]$, and $c_{ij} = 0$ otherwise. U is a $p \times p$ symmetric matrix with four nonzero entries from $\mathrm{Unif}(0, 4) \times \max_{1 \le j \le p} \sigma^{(5)}_{jj,*}$ randomly located in the upper triangle, and another four located in the lower triangle by symmetry. Furthermore, $\delta_0 = |\min\{\lambda_{min}(\Sigma^{(5)}_* + U), \lambda_{min}(\Sigma^{(5)}_*)\}| + 0.05$, where $\lambda_{min}(A)$ denotes the minimum eigenvalue of a symmetric matrix A.

Scenario 6: $\Sigma^{(6)}_1 = 0.2 \times \mathbf{I}_p + AR(0.1)$, $\Sigma^{(6)}_2 = AR(0.1)$, where $AR(\rho) = (a_{ij})_{p \times p}$ is a covariance matrix with $a_{ii} = 1$ and $a_{ij} = \rho^{|i-j|}$, for $i \ne j$;

Scenario 7: $\Sigma^{(7)}_1 = 0.1 \times \mathbf{I}_p + AR(0.2)$, $\Sigma^{(7)}_2 = AR(0.24)$.

Scenario 8: $\Sigma^{(8)}_1 = \Sigma^{(8)}_* + \lambda_0 \mathbf{I}_p$, $\Sigma^{(8)}_2 = \Sigma^{(8)}_* + Q + \lambda_0 \mathbf{I}_p$, where $\Sigma^{(8)}_* = \left(\sigma^{(8)}_{ij,*}\right)_{1 \le i,j \le p}$ with i.i.d. $\sigma^{(8)}_{ii,*} \sim \mathrm{Unif}(1, 2)$, and $\sigma^{(8)}_{ij,*} = \{(|i-j|+1)^{2H} + (|i-j|-1)^{2H} - 2(|i-j|)^{2H}\}/2$ with $H = 0.85$ for $i \ne j$. A perturbation matrix Q has $\lfloor 0.05p \rfloor$ random nonzero elements in the diagonal and nondiagonal. Here, $\lfloor 0.05p \rfloor/2$ nonzero elements are randomly allocated in the upper triangle of Q, and the others are in its lower triangle, by symmetry. The magnitudes of the nonzero elements are randomly generated from $\mathrm{Unif}(\tau/2, 3\tau/2)$, with $\tau = 8\max\{\max_{1 \le i \le p} \sigma^{(8)}_{ii,*}, (\log p)^{1/2}\}$ and $\lambda_0 = |\min\{\lambda_{min}(\Sigma^{(8)}_* + Q), \lambda_{min}(\Sigma^{(8)}_*)\}| + 0.05$.

Finally, the data are generated using $\mathbf{X}_i = \Sigma_1^{1/2} \mathbf{Z}_i$ for $i = 1, \ldots, m$ and $\mathbf{Y}_l = \Sigma_2^{1/2} \mathbf{Z}_{m+l}$ for $l = 1, \ldots, n$, where $\{\mathbf{Z}_i : i = 1, \ldots, m+n\}$ are independent p-dimensional random variables with i.i.d. coordinates Z_{ij}, for $j = 1, \ldots, p$. We consider the following four distributions for Z_{ij}:

1. The standard normal distribution $\mathcal{N}(0, 1)$;

2. A t-distribution with degrees of freedom 15, that is, $t(15)$;

3. A centralized gamma distribution with $a = 16, b = 0.25$, that is, $\Gamma(16, 0.25) - 4$;

4. A discrete distribution that has five possible values $-2, -1, 0, 3/2, 4$, with probabilities 1/12, 4/25, 13/24, 16/75, 1/600, ,respectively; that is,

$$Z_{ij} \sim \frac{1}{12}\delta_{-2} + \frac{4}{25}\delta_{-1} + \frac{13}{24}\delta_0 + \frac{16}{75}\delta_{3/2} + \frac{1}{600}\delta_4.$$

This distribution is used in Yang and Pan (2017), who show that the first four moments of Z_{ij} are the same as those of $\mathcal{N}(0, 1)$.

There are reasonably small disturbances between Σ_1 and Σ_2, and these two covariance matrices are reasonably sparse in Scenario 1. This is similar to the case considered by Yang and Pan (2017). In Scenario 2, there are many small disturbances between Σ_1 and Σ_2. This case was also considered by Yang and Pan (2017) and Li and Chen (2012). In Scenario 3, the two matrices differ only in their off-diagonal entries and have many weakly dense signals. Scenario 4 focuses on the sparse case, with $\lfloor 0.02p \rfloor$ features exerting larger signals. Scenario 5 examines the extremely sparse case, with eight nonzero signals in the nondiagonal entries. Scenario 5 was considered by Cai, Liu and Xia (2013). In Scenario 6, the two covariance matrices differ only in the diagonal. The two covariance matrices differ by a larger amount in Scenario 7. Scenarios 6 and 7 were also considered by Wu and Li (2015). Scenario 8 was studied by Chang et al. (2017), except that a perturbation matrix Q adds some nonzero elements to the diagonal.

In all scenarios, we first calculate the empirical p-values when $\Sigma_1 = \Sigma_2 = \Sigma^{(i)}$, for $i = 1, \ldots, 8$. The corresponding results are given in Tables 1−15 of the Supplementary Material. These results show that the estimated p-values of the proposed RI method and the other six methods are controlled fairly well around 0.05 for all cases except the method proposed by Cai, Liu and Xia (2013) for the discrete distribution.

Because the empirical power for $m = n = 100$ and $m = 200, n = 60$ are very similar to those for $m = n = 60$, we present only the empirical power with $\Sigma_1 = \Sigma_1^{(i)}$ and $\Sigma_2 = \Sigma_2^{(i)}$, $i = 1, \ldots, 8$, for $m = n = 60$ in Figures 1−4. The empirical power for $m = n = 100$ and $m = 200, n = 60$ are included in the Supplementary Material.

For Scenarios 1−2, we have the following findings:

1. Our proposed RI test is considerably more powerful than the other methods. LC is the second most powerful, suggesting its ability to detect covariance differences with many small disturbances, as also observed by Yang and Pan (2017). T2 is the third most powerful because it is a weighted statistic. The Cai, sLED, and $\Psi_{B,\alpha}$ methods have poor power, below 0.20 in almost all cases.

2. The power values of the RI, LC, T2, and $F_{m,n}$ methods increase from Scenario 1 to Scenario 2, which is expected because we increase the small disturbances in the differences between the two covariance matrices. Therefore, RI, LC, T2, and $F_{m,n}$ gain power as the number of deviations increases, even if these deviations are not large. This is expected because they are defined from the Frobenius norm of the difference in the two covariance ma-

trices. However, the other three methods are little affected as we increase the differences.

3. It is not surprising, but reassuring, that the power of our proposed method increases as the sample sizes n, m increase.

In Scenario 3, the two covariance matrices have many weakly dense signals in the nondiagonal entries. The power values of RI, LC, sLED, T2, and $F_{m,n}$ are close to one, but those of the Cai test and $\Psi_{B,\alpha}$ are lower. The results are verified by Corollary 4.

Scenario 4 contains strong sparse signals in the main diagonal of the difference between the two covariance matrices. All seven methods perform quite well, especially when the sample size m or n is large. The power of our proposed RI method is close to one.

Scenario 5 contains extremely sparse strong signals in the nondiagonal differences between the two covariance matrices. The power values of the Cai test, T2, $\Psi_{B,\alpha}$, and $F_{m,n}$ are high, but those of RI, LC, and sLED are lower.

When the two covariance matrices differ in the main diagonal with sparsely weak signals in Scenario 6, our proposed RI method still exhibits perfect power in all cases. The power values of the other six methods are poor. The results are verified by Corollary 2.

In Scenario 7, the two covariance matrices differ to a large degree. The power of the RI method is close to one, whereas the other six methods are much less powerful and not competitive with RI.

For Scenario 8, the two covariance matrices have long-range dependence, according to Chang et al. (2017). The power of the RI method is higher than that of the other six methods for large p.

In conclusion, the RI method has considerably higher and more stable power than the six existing methods do in a wide range of settings. Not only can it deal with cases with many small deviations in the difference of the covariance matrices, but it can also handle cases with sparsely strong or sparsely weak signals. Thus, it is applicable in a broader range of applications when testing the difference between covariance matrices.

4. Real-Data Analysis

In this section, we apply the proposed method to analyze a gene expression data set on breast cancer from a study reported by Schemidt et al. (2008). We downloaded the data set from `http://bioconductor.org/packages/release/data/experiment/html/breastCancerMAINZ.html`.

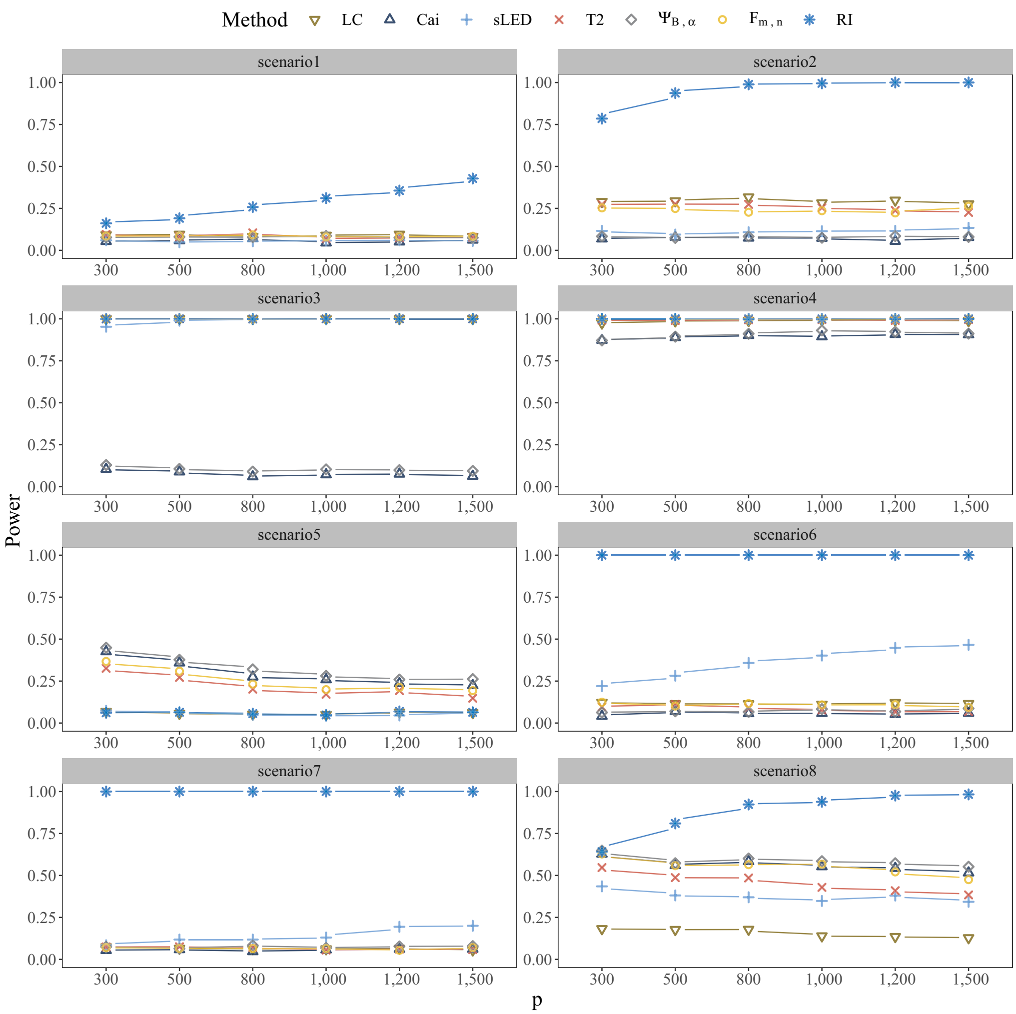


Figure 1. Empirical power for Scenarios 1−8, with Z_{ij} following $\mathcal{N}(0, 1)$ and $m = n = 60$.

The data set contains gene expression patterns from 200 tumors of patients who were not treated by systemic therapy after surgery, and consists of 22,283 features. There are three groups of tumor grades: 29 well differentiated tumors (group 1), 136 moderately differentiated tumors (group 2), and 35 poor/undifferentiated tumors (group 3). This data set has been analyzed in the literature under the assumption that the two covariance matrices are equal; for example, see Teschendorff and Caldas (2008) and Haibe-Kains et al. (2012). The equality of the two covariance matrices is a very important assumption for the validity of the reported findings. We test whether this assumption is valid.

Following Gentleman et al. (2016), for quality control and computational

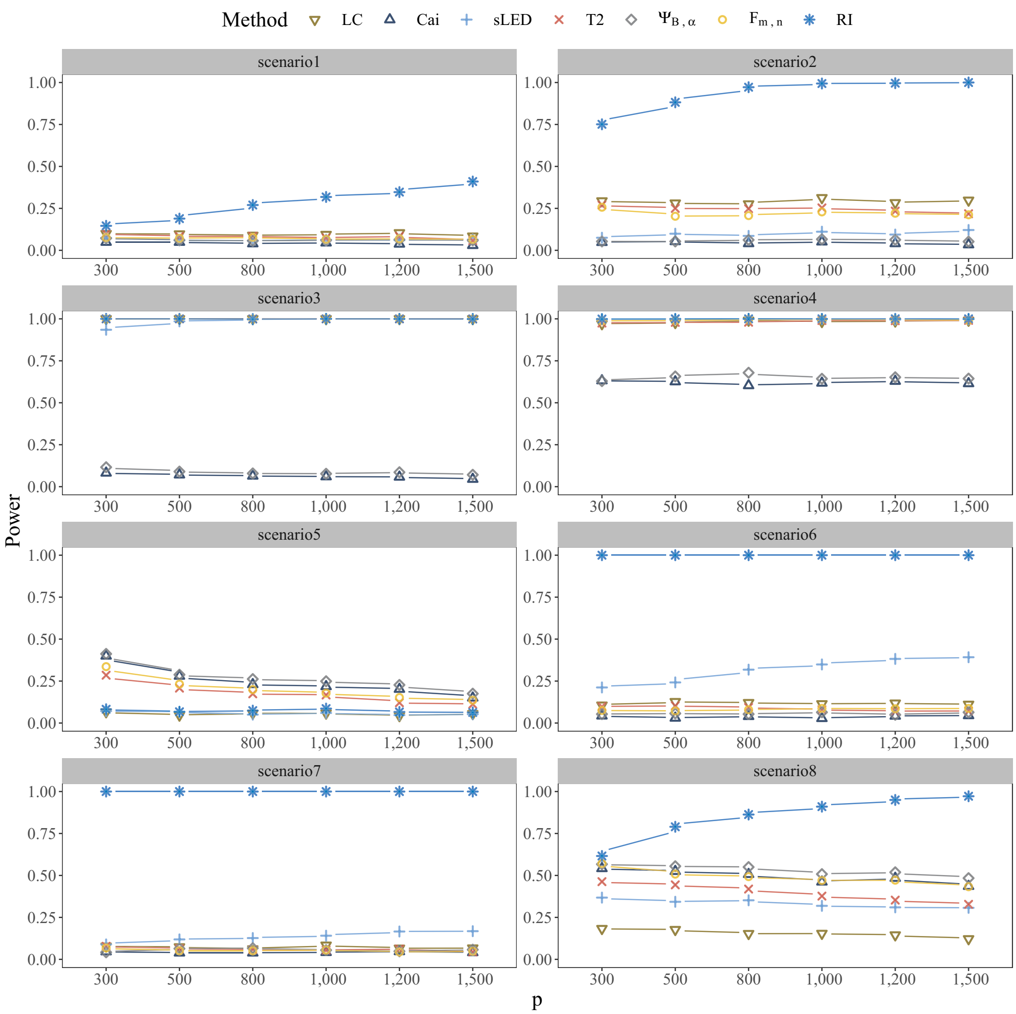


Figure 2. Empirical power for Scenarios 1−8, with Z_{ij} following $t(15)$ and $m = n = 60$.

burden considerations, we first select the features for which more than 50% of their intensities are greater than five and their coefficients of variation (CV) fall within the range (0.22, 1.0). The intensities are gene expressions, as measured by Affymetrix hgu133a technology, and the coefficient of variation is the standard deviation divided by the absolute value of the mean. We also screen the features using predetermined cutoffs that remove low-quality features, while retaining high-quality features, as previously done for this data set (Sherafatian (2018); Chong et al. (2018); Schiffman et al. (2008)). After these selection steps, 1,193 features remain in our analysis. Let Σ_1, Σ_2, and Σ_3 be the covariance matrices of these 1,193 features with groups 1, 2 and 3, respectively. We standardized each

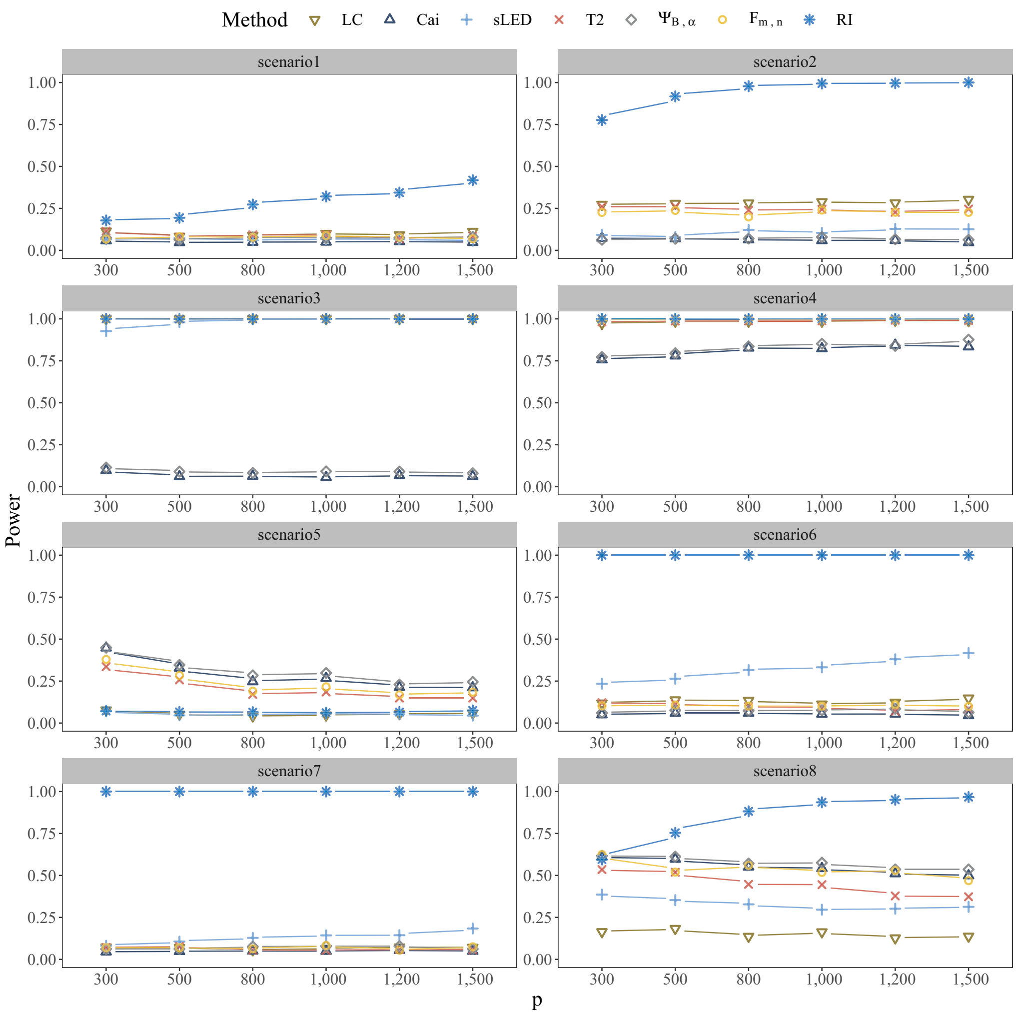


Figure 3. Empirical power for Scenarios 1–8, with Z_{ij} following $\Gamma(16, 0.25) - 4$ and $m = n = 60$.

of the 1,193 features to have a mean of zero. Then, we apply the LC, Cai, sLED, T2, $\Psi_{B,\alpha}$, $F_{m,n}$, and RI methods to separately test (a) $H_0^{(1,2)} : \Sigma_1 = \Sigma_2$ and (b) $H_0^{(2,3)} : \Sigma_2 = \Sigma_3$. The p-values for these methods are reported in Table 1. From Table 1, the RI method rejects both $H_0^{(1,2)}$ and $H_0^{(2,3)}$ at the significance level of 0.05, but the LC, sLED, T2, and $F_{m,n}$ methods reject $H_0^{(2,3)}$ only.

To visualize the comparison of the different methods, we plot heat maps of $(\hat{\Sigma}_1 - \hat{\Sigma}_2)$ and $(\hat{\Sigma}_2 - \hat{\Sigma}_3)$ for the top 100 features with the largest absolute values of the two-sample t-statistics, where $\hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3$ are the sample covariance matrices based on the selected 100 features. The corresponding results appear in

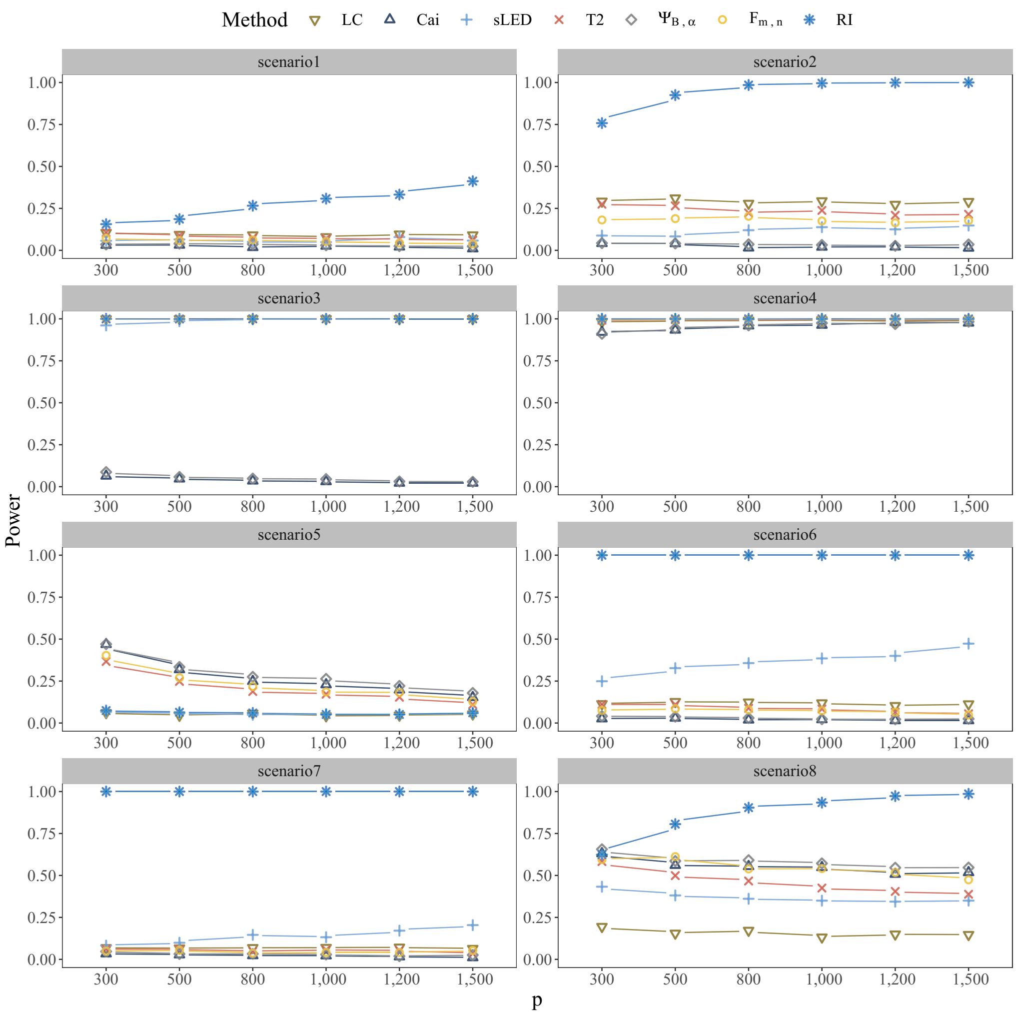


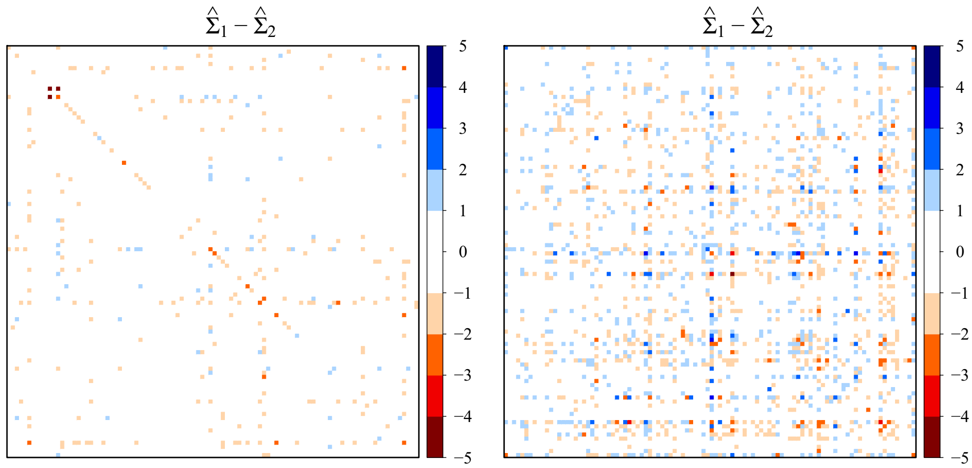
Figure 4. Empirical power for Scenarios 1−8, with Z_{ij} following the discrete distribution $(1/12)\delta_{-2} + (4/25)\delta_{-1} + (13/24)\delta_0 + (16/75)\delta_{3/2} + (1/600)\delta_4$ and $m = n = 60$.

Figure 5, which shows that $(\hat{\Sigma}_2 - \hat{\Sigma}_3)$ has stronger signals than $(\hat{\Sigma}_1 - \hat{\Sigma}_2)$. Many moderate disturbances are present in $(\hat{\Sigma}_2 - \hat{\Sigma}_3)$. The maximum absolute value in the elements of the estimator of $(\hat{\Sigma}_2 - \hat{\Sigma}_3)$ for the 1193 features is 8.972, which is larger than the maximum absolute value in $(\hat{\Sigma}_1 - \hat{\Sigma}_2)$, that is, 4.648. However, the diagonal in $(\hat{\Sigma}_1 - \hat{\Sigma}_2)$ has much stronger signals than those in $(\hat{\Sigma}_2 - \hat{\Sigma}_3)$.

Both Equation (2.2) and the simulation results of Scenarios 4−5 reveal that that the proposed method is more powerful than the other three methods when there are more signals in the diagonal for the difference of two covariance matrices. This explains why our proposed method rejects $H_0^{(1,2)}$, whereas the others did

Table 1. The p-values of the LC, Cai, sLED, T2, $\Psi_{B,\alpha}$, $F_{m,n}$, and RI methods for a gene expression data set.

Method	LC	Cai	sLED	T2	$\Psi_{B,\alpha}$	$F_{m,n}$	RI
$H_0^{(1,2)}$	0.105	0.104	0.330	0.555	0.056	0.061	1.110×10^{-16}
$H_0^{(2,3)}$	3.050×10^{-6}	0.093	0.000	8.157×10^{-9}	0.052	4.577×10^{-6}	3.030×10^{-9}



(a) (b)

Figure 5. (a) heat map of $(\hat{\Sigma}_1 - \hat{\Sigma}_2)$ for the selected 100 features; (b) heat map of $(\hat{\Sigma}_2 - \hat{\Sigma}_3)$ for the selected 100 features.

not. Consequently, our method is the only one able to detect a more subtle, but very important difference in this commonly analyzed data set.

5. Discussion

Conducting inferences using for high-dimensional covariance matrices is highly challenging. Here, we use a random integration technique to develop a two-covariance matrix test statistic. This test can be performed without estimating the covariance matrices, which is known to be extremely difficult in high-dimensional data. We investigate both the theoretical properties and the numerical performance of our method. Our results show that it is not only competitive, but also often much more powerful than existing methods in both simulation studies and a real-data analysis when there are many small diagonal disturbances between the two covariance matrices.

There are several issues that warrant further investigation. First, a general multivariate model is applied to obtain asymptotic results. Although it is a common assumption in the literature, it would be useful to investigate the asymptotic properties of our proposed method under weaker conditions, for example, assumption 2.1 in Han and Wu (2020). Second, we consider a two-sample test for high-dimensional covariance matrices. As in Zheng et al. (2020), it would be interesting to extend our method to test the equality of more than two covariance matrices. Third, we will extend the proposed method to test for high-dimensional correlation matrices, which is a more difficult task than testing the covariance matrices (Zheng et al. (2019)). Finally, we use the standard multivariate normal density function as the weight function to construct our test statistic. This is a common practice, but it is worth investigating other choices that may perform better under various settings.

Supplementary Material

The online Supplementary Material includes detailed proofs of the theoretical results and additional simulation results.

Acknowledgments

Wang and Zhang contributed equally to this article. Jiang's research was partially supported by NSFC(12171203) and the Natural Science Foundation of Guangdong (2019A1515011830, 2022A1515010045). Wen's research was partially supported by NSFC(12171449, 11801540) and the Natural Science Foundation of Anhui Province (BJ2040170017). Wang's research was partially supported by NSFC(12231017, 72171216, 71921001, 71991474), the International Science & Technology cooperation program of Guangdong, China (2016B050502007), the Key Research and Development Program of Guangdong, China (2019B020228001), and the Science and Technology Program of Guangzhou, China (202002030129). Zhang's research was partially supported by the U.S. National Institutes of Health (R01HG010171, R01MH116527) and National Science Foundation (DMS2112711).

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* 3rd Edition. Wiley, New York.

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

Bollerslev, T., Meddahi, N. and Nyawa, S. (2019). High-dimensional multivariate realized volatility estimation. *Journal of Econometrics* **212**, 116–136.

Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association* **108**, 265–277.

Cai, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annuals Review of Statistics and Its Application* **4**, 423– 446.

Chang, J., Zhou, W., Zhou, W.-X. and Wang, L. (2017). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics* **73**, 31–4.

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

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

Chen, F., Meintanis, S. G. and Zhu, L. (2019). On some characterizations and multidimensional criteria for testing homogeneity, symmetry and independence. *Journal of Multivariate Analysis* **173**, 125–144.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G. et al. (2018). Metaboanalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research* **46**, 486–494.

Gentleman, R., Carey, V., Huber, W. and Hahne, F. (2016). Genefilter: Methods for filtering genes from high-throughput experiments. R Package Version 1.54.2. European Bioinformatics Institute, Cambridge.

Haibe-Kains, B., Desmedt, C., Loi, S., Culhane, A. C., Bontempi, G., Quackenbush, J. et al. (2012). A three-gene model to robustly identify breast cancer molecular subtypes. *Journal of the National Cancer Institute* **104**, 311–325.

He, J. and Chen, S. X. (2018). High-dimensional two-sample covariance matrix testing via super-diagonals. *Statistica Sinica* **28**, 2671–2696.

Han, Y. and Wu, W. B. (2020). Test for high dimensional covariance matrices. *The Annals of Statistics* **48**, 3565–3588.

He, Y., Xu, G., Wu, C. and Pan, W. (2021). Asymptotically independent U-statistics in high-dimensional testing. *The Annals of Statistics* **49**, 154–181.

Igolkina, A., Aromokus, C., Newman, J., Evgrafov, O., Mcintyre, L., Nuzhdin, S. et al. (2018). Analysis of gene expression variance in schizophrenia using structural equation modeling. *Frontiers in Molecular Neuroscience* **11**, 192–192.

Jiang, T. and Qi, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scandinavian Journal of Statistics* **42**, 988–1009.

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

Kim, Y. and Lesser, V. (2008). Finding minimum data requirements using pseudo-independence. In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* **2**, 57–64.

Kim, I., Balakrishnan, S. and Wasserman, L. (2020). Robust multivariate nonparametric tests via projection averaging. *The Annals of Statistics* **48**, 3417–3441.

Le Bihan, D., Mangin, J.-F., Poupon, C., Clark, C. A., Pappata, S., Molko, N. et al. (2001), Diffusion tensor imaging: Concepts and applications. *Journal of Magentic Resonance Imaging* **13**, 534–546.

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

Pan, W., Tian, Y., Wang, X. and Zhang, H. (2018). Ball divergence: Nonparametric two sample test. *The Annals of Statistics* **46**, 1109–1137.

Qiu, T., Xu, W. and Zhu, L. (2021). Two-sample test in high dimensions through random selection. *Computational Statistics & Data Analysis* **106**, 107218.

Roberts, A. G., Catchpoole, D. R. and Kennedy, P. J. (2018). Variance-based feature selection for classification of cancer subtypes using gene expression data. In 2018 *International Joint Conference on Neural Networks(IJCNN)*. IEEE.

Schiffman, C., Petrick, L., Perttula, K., Yano, Y., Carlsson, H., Whitehead, T. et al. (2008). Filtering procedures for untargeted LC-MS metabolomics data. *BMC Bioinformatics* **20**, 334–334.

Schemidt, M., Böhm , D., von Törne, C., Steiner, E., Puhl, A., Pilch, H. et al. (2008). The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Research* **68**, 5405–5413.

Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics & Data Analysis* **51**, 6335–6542.

Sherafatian, M. (2018). Tree-based machine learning algorithms identified minimal set of miRNA biomarkers for breast cancer diagnosis and molecular subtyping. *Gene* **677**, 111–118.

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

Teschendorff, A. E. and Caldas, C. (2008). A robust classifier of high predictive value to identify good prognosis patients in ER-negative breast cancer. *Breast Cancer Research* **10**, R73.

Wu, T.-L. and Li, P. (2015). Tests for high-dimensional covariance matrices using random matrix projection. *arXiv:1511.01611.*

Yang, Q. and Pan, G. (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *Journal of the American Statistical Association* **112**, 188–200.

Yu, X., Li, D. and Xue, L. (2022). Fisher's combined probability test for high-dimensional covariance matrices. *Journal of the American Statistical Association.* DOI: 10.1080/01621459.2022.2126781.

Zheng, S., Lin, R., Guo, J. and Yin, G. (2020). Testing homogeneity of high-dimensional covariance matrices. *Statistica Sinica* **30**, 35–53.

Zhu, L., Lei, J., Devlin, B. and Roeder, K. (2017). Testing high-dimensional covariance matrices, with application to detecting schizophrenia risk genes. *The Annals of Applied Statistics* **11**, 1810–1831.

Zhu, L., Xu, K., Li, R. and Zhong, W. (2017). Projection correlation between two random vectors. *Biometrika* **104**, 829–843.

Zheng, S., Cheng, G., Guo, J. and Zhu, H. (2019). Test for high dimensional correlation matrices. *The Annals of Statistics* **47**, 2887–2921.

Zhong, P.-S., Li, R. and Santo, S. (2019). Homogeneity tests of covariance matrices with high-dimensional longitudinal data. *Biometrika* **106**, 619–634.

Yunlu Jiang

Department of Statistics & Data Science, College of Economics, Jinan University, Guangzhou, GD 510632, China.

E-mail: tjiangyl@jnu.edu.cn

Canhong Wen

International Institute of Finance, School of Management, University of Science and Technology of China, Hefei, AH 230026, China.

E-mail: wench@ustc.edu.cn

Yukang Jiang

Department of Statistical Science, School of Mathematics, Sun Yat-Sen University, Guangzhou, GD 510275, China.

E-mail: jiangyk3@mail2.sysu.edu.cn

Xueqin Wang

International Institute of Finance, School of Management, University of Science and Technology of China, Hefei, AH 230026, China.

E-mail: wangxq20@ustc.edu.cn

Heping Zhang

Department of Biostatistics, School of Public Health, Yale University, New Haven, CT 06520-8034, USA.

E-mail: heping.zhang@yale.edu

(Received November 2020; accepted January 2022)