JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2024, VOL. 119, NO. 545, 511–524: Theory and Methods
https://doi.org/10.1080/01621459.2022.2126781

Fisher's Combined Probability Test for High-Dimensional Covariance Matrices

Xiufan Yu[a], Danning Li[b], and Lingzhou Xue[c]

[a]Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN; [b]KLAS and School of Mathematics & Statistics, Northeast Normal University, Changchun, China; [c]Department of Statistics, Pennsylvania State University, University Park, PA

ABSTRACT

Testing large covariance matrices is of fundamental importance in statistical analysis with high-dimensional data. In the past decade, three types of test statistics have been studied in the literature: quadratic form statistics, maximum form statistics, and their weighted combination. It is known that quadratic form statistics would suffer from low power against sparse alternatives and maximum form statistics would suffer from low power against dense alternatives. The weighted combination methods were introduced to enhance the power of quadratic form statistics or maximum form statistics when the weights are appropriately chosen. In this article, we provide a new perspective to exploit the full potential of quadratic form statistics and maximum form statistics for testing high-dimensional covariance matrices. We propose a scale-invariant power-enhanced test based on Fisher's method to combine the p-values of quadratic form statistics and maximum form statistics. After carefully studying the asymptotic joint distribution of quadratic form statistics and maximum form statistics, we first prove that the proposed combination method retains the correct asymptotic size under the Gaussian assumption, and we also derive a new Lyapunov-type bound for the joint distribution and prove the correct asymptotic size of the proposed method without requiring the Gaussian assumption. Moreover, we show that the proposed method boosts the asymptotic power against more general alternatives. Finally, we demonstrate the finite-sample performance in simulation studies and a real application. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received June 2020
Accepted Setember 2022

KEYWORDS
Fisher's method; High-dimensional inference; Joint limiting law; Large covariance structure; Lyapunov-type bound; Power enhancement

1. Introduction

Hypothesis testing on large covariance matrices has received considerable attention in the past decade. The covariance matrices not only have the fundamental importance in multivariate statistics such as discriminant analysis, principal component analysis, and clustering (Anderson 2003), but also play a vital role in various research topics in biological science, finance, operations research including portfolio allocation (Goldfarb and Iyengar 2003), gene-set testing (Chen and Qin 2010), and gene-set clustering (Chang et al. 2017).

Let $\mathbf{X}$ and $\mathbf{Y}$ represent two independent p-dimensional random vectors with covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, respectively. We are interested in testing whether these two covariance matrices are equal, that is, $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, which has been extensively explored (Anderson 2003). The likelihood ratio test (LRT) enjoys the optimality in the classical setting where the dimension is fixed (Sugiura and Nagao 1968; Perlman 1980). Johnstone (2008) studied Roy's largest root test when the sample size grows in proportion to the dimension. Bai et al. (2009) and Jiang and Yang (2013) studied the modified LRTs and proved their asymptotic normality when the dimension grows at a slower rate than the sample size. However, the likelihood function is not well-defined due to the singular sample covariance matrix in the high-dimensional setting where the dimension grows at a faster rate than the sample size.

Over the past decade, statisticians have made a lot of efforts to tackle the challenges in the high-dimensional setting that the dimension can be much larger than the sample size. Three types of test statistics have been focused for testing large covariance matrices. First, quadratic form statistics were studied to test against the dense alternatives, which can be written in terms of the Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ with many small differences between two covariance matrices. When the dimension is on the same order of the sample size, Schott (2007) proposed a test statistic based on the sum of squared differences between two sample covariance matrices, and Srivastava and Yanagihara (2010) used a consistent estimator of $\mathrm{tr}(\mathbf{\Sigma}_1^2)/[\mathrm{tr}(\mathbf{\Sigma}_1)]^2 - \mathrm{tr}(\mathbf{\Sigma}_2^2)/[\mathrm{tr}(\mathbf{\Sigma}_2)]^2$ to construct a new test statistic. Li and Chen (2012) introduced an unbiased estimator of the Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ to allow for the ultra-high dimensionality that the dimension grows much faster than the sample size. Recently, He et al. (2021) proposed the adaptive testing to combine the finite-order U-statistics that includes the variants of quadratic form statistics. Second, maximum form statistics were explored to account for the sparse alternatives with only a few large differences between two covariance matrices, which can be

CONTACT Danning Li lidn040@nenu.edu.cn KLAS and School ofMathematics & Statistics, Northeast Normal University, Changchun, China; Lingzhou Xue lzxue@psu.edu Department of Statistics, Pennsylvania State University, University Park, PA.
Xiufan Yu and Danning Li are co-first authors who contributed equally to this article.
Danning Li and Lingzhou Xue are co-corresponding authors.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

written in terms of the entry-wise maximum norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$. Cai, Liu, and Xia (2013) studied the maximal standardized differences between two sample covariance matrices to test against the sparse alternative, and Chang et al. (2017) proposed a perturbed-based maximum test using a data-driven approach to determine the rejection region. Third, Li and Xue (2015), Yang and Pan (2017), and Li, Xue, and Zou (2018) used a weighted combination of quadratic form statistics and maximum form statistics to test against the dense or sparse alternatives, which shares the similar philosophy with the power enhancement method (Fan, Liao, and Yao 2015) for testing cross-sectional dependence.

Similar to these weighted combination tests, we are motivated by combining the strengths of quadratic form statistics and maximum form statistics to boost the power against the dense or sparse alternatives. It is important to combine the power of these two different statistics in real-world applications such as financial studies and genetic association studies. For instance, the anomalies in financial markets may come from the mispricing of a few assets or a systematic market mispricing (Fan, Liao, and Yao 2015; Yu, Yao, and Xue 2019), and the phenotype may be affected by a few causal variants or a large number of mutants (Liu et al. 2019).

It is worth pointing out that these weighted combination tests critically depend on the proper choice of weights to combine two different types of test statistics. There may exist a non-negligible discrepancy on the different magnitudes between quadratic form statistics and maximum form statistics in practice, which makes the choice of weights a very challenging task. As a promising alternative to Fan, Liao, and Yao (2015), Li and Xue (2015), Yang and Pan (2017), and Li, Xue, and Zou (2018), we provide a new perspective to exploit the full potential of quadratic form statistics and maximum form statistics for testing high-dimensional covariance matrices.

We propose a scale-invariant power-enhanced test based on Fisher's method (Fisher 1925) to combine the p-values of quadratic form statistics and maximum form statistics. To study the asymptotic property, we need to solve several nontrivial challenges in the theoretical analysis and then derive the asymptotic joint distribution of quadratic form statistics and maximum form statistics under the null hypothesis. We prove that the asymptotic null distribution of the proposed combination test statistic does not depend on the unknown parameters (see Theorem 2 under the Gaussian assumption and Theorems 3 and 4 without requiring the Gaussian assumption). More specifically, the proposed statistic follows a Chi-squared distribution with 4 degrees of freedom asymptotically under the null hypothesis. We also show the consistent asymptotic power against the union of dense alternatives and sparse alternatives, which is more general than the designated alternative in the weighted combination test. It is worth pointing out that Fisher's method achieves the asymptotic optimality with respect to Bahadur relative efficiency. Moreover, we demonstrate the numerical properties in simulation studies and a real application to gene-set testing.

In recent literature, Fan, Liao, and Yao (2015), Li and Xue (2015), Li, Xue, and Zou (2018), and He et al. (2021) studied the one-sample covariance test of a special covariance structure such as the identity or bandedness. Specifically, Fan, Liao, and Yao (2015), Li and Xue (2015), and He et al. (2021) considered the one-sample test for large covariance matrices that $H_0 : \mathbf{\Sigma} = \mathbf{I}$ under the restricted Gaussian or weak dependence assumption among entries of $\mathbf{X}$, and Li, Xue, and Zou (2018) studied the one-sample test that $H_0 : \mathbf{\Sigma}$ is a banded matrix under the Gaussian assumption. However, these methods and theories do not apply to the more challenging setting for two-sample test of the general covariance structure in large covariance matrices. It is significantly more challenging to deal with the complicated dependence in the two-sample tests for large covariance matrices even under the Gaussian assumption. To address this challenges, we use a decorrelation technique to address the complex nonlinear dependence in high dimensional covariances. Shi et al. (2019) used the decorrelation to study the linear hypothesis testing for high-dimensional generalized linear models. But the nonlinear dependence in the two-sample covariance testing is much more challenging than the linear hypothesis testing in Shi et al. (2019). To the best of our knowledge, our work presents the first proof of the asymptotic independence result of quadratic form statistics and maximum form statistics for testing two-sample covariance matrices, which provides the essential theoretical guarantee for Fisher's method to combine their p-values.

In the theoretical analysis, we make significant efforts to effectively relax the Gaussian or weak dependence assumption when deriving the asymptotic null distribution of the proposed test statistic. Existing works including Fan, Liao, and Yao (2015), Li and Xue (2015), Li, Xue, and Zou (2018), and He et al. (2021) require the restricted Gaussian or weak dependence assumption when deriving the asymptotic properties. To this end, we derive a new Lyapunov-type bound for the joint distribution of quadratic form statistics and maximum form statistics under the high-dimensional setting (see Theorems 4), which extends the known Lyapunov-type bound for quadratic form statistics (Bentkus 2005) or for maximum form statistics (Chernozhukov, Chetverikov, and Kato 2013, 2017). To the best of our knowledge, there does not exist such Lyapunov-type bound for their joint distribution in the current literature. Our theoretical result fills this gap and makes a separate contribution to the literature.

The rest of this article is organized as follows. Section 2 presents the preliminaries, and Section 3 introduces the proposed method for testing large covariance matrices. Section 4 studies the asymptotic properties. Section 5 examines the numerical properties in simulation studies, and Section 6 explores an empirical study on testing gene-sets. Section 7 includes the concluding remarks. Proofs and additional numerical results are presented in the supplementary materials.

2. Preliminaries

Let $\mathbf{X}$ and $\mathbf{Y}$ be two independent random vectors in $\mathbb{R}^p$ with covariance matrices $\mathbf{\Sigma}_1 = (\sigma_{ij1})_{p\times p}$ and $\mathbf{\Sigma}_2 = (\sigma_{ij2})_{p\times p}$, respectively. Without loss of generality, we assume they have zero means. Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}\}$ be iid samples of $\mathbf{X}$, and $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2}\}$ be iid samples of $\mathbf{Y}$. The problem of interest is to test whether two covariance matrices are equal,

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}. \tag{2.1}$$

We first revisit the quadratic form statistic (Li and Chen 2012) to test against the dense alternative and the maximum form statistic (Cai, Liu, and Xia 2013) to test against the sparse alternative. The dense alternative can be written in terms of the Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ and the sparse alternative can be written using the entry-wise maximum norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$.

Li and Chen (2012) proposed a quadratic-form test after reformulating the null hypothesis (2.1) into its equivalent form based on the squared Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$, that is,

$$H_0 : \|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2 = 0.$$

To construct the test statistic, given the simple fact that $\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2 = \mathrm{tr}\{(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2\} = \mathrm{tr}(\mathbf{\Sigma}_1^2) + \mathrm{tr}(\mathbf{\Sigma}_2^2) - 2\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$, Li and Chen (2012) proposed a test statistic T_{n_1,n_2} in the form of linear combination of unbiased estimators for each term, namely,

$$T_{n_1,n_2} = A_{n_1} + B_{n_2} - 2C_{n_1,n_2}, \tag{2.2}$$

where

$$\begin{aligned} A_{n_1} = {} & \frac{1}{n_1(n_1-1)} \sum_{u \neq v} \left(\mathbf{X}_u' \mathbf{X}_v\right)^2 \\ & - \frac{2}{n_1(n_1-1)(n_1-2)} \sum_{u,v,k}^{*} \mathbf{X}_u' \mathbf{X}_v \mathbf{X}_v' \mathbf{X}_k \\ & + \frac{1}{n_1(n_1-1)(n_1-2)(n_1-3)} \sum_{u,v,k,l}^{*} \mathbf{X}_u' \mathbf{X}_v \mathbf{X}_k' \mathbf{X}_l, \\ B_{n_2} = {} & \frac{1}{n_2(n_2-1)} \sum_{u \neq v} \left(\mathbf{Y}_u' \mathbf{Y}_v\right)^2 \\ & - \frac{2}{n_2(n_2-1)(n_2-2)} \sum_{u,v,k}^{*} \mathbf{Y}_u' \mathbf{Y}_v \mathbf{Y}_v' \mathbf{Y}_k \\ & + \frac{1}{n_2(n_2-1)(n_2-2)(n_2-3)} \sum_{u,v,k,l}^{*} \mathbf{Y}_u' \mathbf{Y}_v \mathbf{Y}_k' \mathbf{Y}_l, \end{aligned}$$

and

$$\begin{aligned} C_{n_1,n_2} = {} & \frac{1}{n_1 n_2} \sum_{u} \sum_{v} \left(\mathbf{X}_u' \mathbf{Y}_v\right)^2 \\ & - \frac{1}{n_1 n_2 (n_1-1)} \sum_{u,k}^{*} \sum_{v} \mathbf{X}_u' \mathbf{Y}_v \mathbf{Y}_v' \mathbf{X}_k \\ & - \frac{1}{n_1 n_2 (n_2-1)} \sum_{u,k}^{*} \sum_{v} \mathbf{Y}_u' \mathbf{X}_v \mathbf{X}_v' \mathbf{Y}_k \\ & + \frac{1}{n_1 n_2 (n_1-1)(n_2-1)} \sum_{u,k}^{*} \sum_{v,l}^{*} \mathbf{X}_u' \mathbf{Y}_v \mathbf{X}_k' \mathbf{Y}_l. \end{aligned}$$

Here, $\sum^*$ denotes summation over mutually distinct indices. A_{n_1}, B_{n_2}, and C_{n_1,n_2} are the unbiased estimators under H_0 for $\mathrm{tr}(\mathbf{\Sigma}_1^2)$, $\mathrm{tr}(\mathbf{\Sigma}_2^2)$ and $\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$, respectively. Then, the expected value of T_{n_1,n_2} is zero under the null hypothesis. Li and Chen (2012) proved that the asymptotic distribution of T_{n_1,n_2} is a normal distribution. Let z_α be the upper α quantile of the standard normal distribution, and $\widehat{\sigma}_{0,n_1,n_2} = \frac{2}{n_2} A_{n_1} + \frac{2}{n_2} B_{n_2}$ is a consistent estimator of the leading term $\sigma_{0,n_1,n_2} = (\frac{2}{n_1} + \frac{2}{n_2})\mathrm{tr}(\mathbf{\Sigma}^2)$ in the standard deviation of T_{n_1,n_2} under H_0. Please refer to Section 2 of Li and Chen (2012). The test rejects the null hypothesis at the significance level α if

$$\frac{T_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \geq z_\alpha. \tag{2.3}$$

As an alternative to the quadratic form statistic (Li and Chen 2012), Cai, Liu, and Xia (2013) studied the null hypothesis (2.1) in terms of the maximal absolute difference of two covariance matrices, that is,

$$H_0 : \max_{1 \leq i \leq j \leq p} |\sigma_{ij1} - \sigma_{ij2}| = 0.$$

Cai, Liu, and Xia (2013) proposed a maximum test statistic M_{n_1,n_2} based on the maximum of standardized differences between $\widehat{\sigma}_{ij1}$'s and $\widehat{\sigma}_{ij2}$'s. The maximum form statistic is written as

$$M_{n_1,n_2} = \max_{1 \leq i \leq j \leq p} \frac{\left(\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2}\right)^2}{\widehat{\theta}_{ij1}/n_1 + \widehat{\theta}_{ij2}/n_2}, \tag{2.4}$$

where the denominator $\widehat{\theta}_{ij1}/n_1 + \widehat{\theta}_{ij1}/n_2$ estimates the variance $\theta_{ij1}/n_1 + \theta_{ij1}/n_2$ of $\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2}$ to account for the heteroscedasticity of $\widehat{\sigma}_{ij1}$'s and $\widehat{\sigma}_{ij2}$'s among different entries. Here $\theta_{ij1} = \mathrm{var}(X_{u,i}X_{u,j})$ and $\theta_{ij2} = \mathrm{var}(Y_{v,i}Y_{v,j})$. The specific forms of $\widehat{\theta}_{ij1}$ and $\widehat{\theta}_{ij2}$ are

$$\begin{aligned} \widehat{\theta}_{ij1} &= \frac{1}{n_1} \sum_{u=1}^{n_1} \left[(X_{u,i} - \overline{X}_i)(X_{u,j} - \overline{X}_j) - \widehat{\sigma}_{ij1}\right]^2, \\ \overline{X}_i &= \frac{1}{n_1} \sum_{u=1}^{n_1} X_{ui}, \end{aligned}$$

and

$$\begin{aligned} \widehat{\theta}_{ij2} &= \frac{1}{n_2} \sum_{v=1}^{n_2} \left[(Y_{v,i} - \overline{Y}_i)(Y_{v,j} - \overline{Y}_j) - \widehat{\sigma}_{ij2}\right]^2, \\ \overline{Y}_i &= \frac{1}{n_2} \sum_{v=1}^{n_2} Y_{vi}. \end{aligned}$$

Cai, Liu, and Xia (2013) proved that the asymptotic null distribution of M_{n_1,n_2} is a Type I extreme value distribution (also known as the Gumbel distribution). Thus, the test rejects the null hypothesis at a significance level α if

$$M_{n_1,n_2} \geq q_\alpha + 4\log p - \log\log p, \tag{2.5}$$

where q_α is the upper α quantile of the Gumbel distribution.

3. Methodology

Li and Chen (2012) and Cai, Liu, and Xia (2013) have their respective power for testing high-dimensional covariance matrices. The quadratic form statistic T_{n_1,n_2} is powerful against the dense alternative, where the difference between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ under the squared Frobenius norm is no smaller than the order of $\mathrm{tr}(\mathbf{\Sigma}_1^2)/n_1 + \mathrm{tr}(\mathbf{\Sigma}_2^2)/n_2$. The maximum form statistic M_{n_1,n_2} is powerful against the sparse alternative, where at least one entry of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ has the magnitude larger than the order of $\sqrt{\log p/n}$. However, T_{n_1,n_2} performs poorly against the sparse alternative

and M_{n_1,n_2} performs poorly against the dense alternative, which will be further illustrated in Sections 4.4 and 5.

Fan, Liao, and Yao (2015), Li and Xue (2015), Yang and Pan (2017), Li, Xue, and Zou (2018), and Yu et al. (2022) studied the weighted combination $J = J_0 + J_1$ to achieve the power enhancement, where J_0 is built on the extreme value form statistic and J_1 is constructed from the asymptotically pivotal statistic. It is worth pointing out that, with the proper weighted combination, J enjoys the power enhancement properties (Fan, Liao, and Yao 2015):

(i) J is at least powerful as J_1,
(ii) the size distortion due to the addition of J_0 is asymptotically negligible, and
(iii) power is improved under the designated alternatives.

For testing large covariance matrices, Yang and Pan (2017) proposed $J_1 = (1 - (s_p + \xi_1)^{-1})M_n$ and $J_0 = n^{\frac{1}{s_p+\xi_1}+\frac{1}{\xi_2}} \cdot \max_{1\le i,j\le p}(\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2})^2$, where M_n is a so-called macro-statistic that performs well against the dense alternative, and s_p is the number of distinct entries in two covariance matrices. Note that the quantities ξ_1 and ξ_2 are carefully chosen such that $J_0 \to 0$ in probability under the null hypothesis H_0.

As a promising alternative, we propose a scale-invariant combination procedure based on Fisher's method (Fisher 1925) to combine both strengths of T_{n_1,n_2} and M_{n_1,n_2}. Let $\Phi(\cdot)$ be the cumulative distribution function of $N(0,1)$ and $G(y) = \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{y}{2}\right)\right)$ be the cumulative distribution function of the Gumbel distribution. Specifically, we combine the p-values of T_{n_1,n_2} and M_{n_1,n_2} after the negative natural logarithm transformation, that is,

$$F_{n_1,n_2} = -2\log p_T - 2\log p_M, \tag{3.1}$$

where $p_T = 1 - \Phi\left(T_{n_1,n_2}/\widehat{\sigma}_{0,n_1,n_2}\right)$ and $p_M = 1 - G(M_{n_1,n_2} - 4\log p + \log\log p)$ are the p-values associated with the test statistics T_{n_1,n_2} and M_{n_1,n_2}, respectively.

Let c_α denote the upper α quantile of a Chi-squared distribution with 4 degrees of freedom (i.e., χ^2_4). We reject the null hypothesis at the significance level α if

$$F_{n_1,n_2} \ge c_\alpha. \tag{3.2}$$

Unlike the weighted statistic $J = J_0 + J_1$, F_{n_1,n_2} does not need to estimate s_p or choose ξ_1 and ξ_2 to construct the proper weights, which may be nontrivial to deal with in practice. The inappropriate choice of s_p, ξ_1 and ξ_2 may lead to the size distortion or loss of power. In contrast, F_{n_1,n_2} is scale-invariant as the p-values always take values between 0 and 1, and the asymptotic null distribution of F_{n_1,n_2} (i.e., χ^2_4) does not depend on any hyper-parameters. As we will show in Sections 4.2 and 4.3, F_{n_1,n_2} achieves the desired nominal significance level asymptotically, even without requiring the Gaussian assumption. In Section 4.4, we will show that the proposed method boosts the power against either sparse or dense alternatives. Moreover, Fisher's method achieves the asymptotic optimality with respect to Bahadur relative efficiency (Littell and Folks 1971, 1973).

Remark 3.1. The idea of combining p-values has been widely used as an important technique for data fusion or meta analysis (Hedges and Olkin 2014). Recently, the Cauchy combination of p-values was used for testing high-dimensional mean vectors in (Liu and Xie 2020), and the minimum combination of p-values from the finite-order U-statistics was used for testing two-sample high-dimensional covariance matrices in (He et al. 2021). However, neither Liu and Xie (2020) nor He et al. (2021) studied the combination of p-values of T_{n_1,n_2} and M_{n_1,n_2}, and it is fundamentally challenging to study the asymptotic joint distribution of T_{n_1,n_2} and M_{n_1,n_2}. We will solve this open problem in Sections 4.2 and 4.3.

4. Asymptotic Properties

This section presents the asymptotic properties of our proposed Fisher's combined probability test F_{n_1,n_2}. Section 4.1 defines useful notations and presents the assumptions. Section 4.2 first studies the joint limiting distribution of two test statistics M_{n_1,n_2} and T_{n_1,n_2} under the null hypothesis and then prove the correct asymptotic size. Section 4.3 studies the asymptotic joint distribution of M_{n_1,n_2} and T_{n_1,n_2} and the asymptotic size without requiring the Gaussian assumption. Section 4.4 proves the consistent asymptotic power of our proposed method under general conditions under the alternative hypotheses.

4.1. Assumptions

For any matrix $\mathbf{A}$, let $\lambda_i(\mathbf{A})$ be the ith largest eigenvalue of $\mathbf{A}$. Throughout the rest of Section 4, we need the following assumptions to carry out the theoretical analysis.

Assumption 1. As $n_1, n_2, p \to \infty$, suppose that

(i) $n_1/(n_1+n_2) \to \gamma$, for some constant $\gamma \in (0,1)$.
(ii) For any $j \in \{1,2\}$ and there exists a constant $1 > \delta_0 > 0$, $\mathrm{tr}\left(\mathbf{\Sigma}_j^2\right) \to \infty$ and

$$[\log\{p \vee (n_1+n_2)\}]^{3(1+\delta_0)}\lambda_1^2(\mathbf{\Sigma}_j) \le C(\mathrm{tr}\left(\mathbf{\Sigma}_j^2\right)).$$

where $\lambda_1(\mathbf{\Sigma}_j)$ denotes the largest eigenvalue of $\mathbf{\Sigma}_j$ and $C > 0$ is a fixed constant that does not depend on n_1, n_2 and p.

Remark 4.1. Assumption 1 is analogous to (A1) and (A2) in Li and Chen (2012), where the first condition is standard for two-sample asymptotic analysis, and the second one describes the extent of high dimensionality and the dependence which can be accommodated by the proposed tests. Sharing the spirit, Assumption 1 does not impose explicit requirements on relationships between p and n_1, n_2, but rather requires a mild condition (ii) regarding the covariance matrices, which can be satisfied if eigenvalues of two covariance matrices are bounded. The condition (ii) assumes a specific rate to prove Lemma 2 that is useful for exploring the asymptotic joint distribution, whereas (A2) in Li and Chen (2012) does not need to assume any specific rate for studying the quadratic form statistics.

Before proceeding, we define some useful notations. For any set $\mathcal{A}$, $\mathrm{card}(\mathcal{A})$ denotes the cardinality of $\mathcal{A}$. For $0 < r < 1$, let $\mathcal{V}_i(r) = \left\{1 \le j \le p : \frac{|\sigma_{ij1}|}{\sqrt{\sigma_{ii1}\sigma_{jj1}}} \ge r \text{ or } \frac{|\sigma_{ij2}|}{\sqrt{\sigma_{ii2}\sigma_{jj2}}} \ge r\right\}$ be the set of indices j such that X_j (or Y_j) is highly correlated (whose correlation $> r$) with X_i (or Y_i) for a given $i \in \{1,\ldots,p\}$. And

for any $\alpha > 0$, let $s_i(\alpha) = \text{card}(\mathcal{V}_i((\log p)^{-1-\alpha}))$, $i = 1, \ldots, p$ be the number of indices j in the set $\mathcal{V}_i((\log p)^{-1-\alpha})$. Moreover, define $\mathcal{W}(r) = \{1 \le i \le p : \mathcal{V}_i(r) \neq \varnothing\}$ such that, $\forall i \in \mathcal{W}(r)$, X_i (or Y_i) is highly correlated with some other variable of $\mathbf{X}$ (or $\mathbf{Y}$).

Assumption 2. There exists a subset $\Upsilon \subset \{1, 2, \ldots, p\}$ with $\text{card}(\Upsilon) = o(p)$ and some constant $\alpha_0 > 0$, such that for all $\kappa > 0$, $\max_{1 \le i \le p, i \notin \Upsilon} s_i(\alpha_0) = o(p^\kappa)$. In addition, there exists a constant $0 < r_0 < 1$, such that $\text{card}(\mathcal{W}(r_0)) = o(p)$. Furthermore, we assume that for some constants $0 < \tau_1 < \tau_2$, $\min_{1 \le i \le p} \min(\sigma_{ii1}, \sigma_{ii2}) \ge \tau_1$ and $\max_{1 \le i \le p} \max(\sigma_{ii1}, \sigma_{ii2}) \le \tau_2$

Remark 4.2. Assumption 2 was introduced by Cai, Liu, and Xia (2013) such that $\max_{1 \le i \le p, i \notin \Upsilon} s_i(\alpha_0)$ and $\mathcal{W}(r_0)$ are moderate for $\alpha_0 > 0$ and $0 < r_0 < 1$. It is satisfied if the eigenvalues of covariance matrices are bounded from above and correlations are bounded away from ± 1.

In Section 4.2, we assume that both $\mathbf{X}$ and $\mathbf{Y}$ are Gaussian random vectors. The Gaussian assumption facilitates the use of a new decorrelation technique to address the complex nonlinear dependence in high dimensional covariance matrices in the theoretical analysis of the proposed scale-invariant combination test. Li and Xue (2015), Li, Xue, and Zou (2018), and He et al. (2021) studied the asymptotic joint distribution of the maximum test statistic and the quadratic test statistic for one-sample covariance test of a special covariance structure such as the identity or bandedness under the Gaussian or weak dependence assumption. Please see the first paragraph of Section 2 in Li and Xue (2015), the first paragraph of Section 2 in Li, Xue, and Zou (2018), and Condition 2.3 in He et al. (2021) for more details. However, we need to deal with a general covariance structure in two-sample covariance test, and the nonlinear dependence is fundamentally more challenging than the dependence in the one-sample covariance test even under the Gaussian assumption. In Section 4.2, we use a decorrelation technique to address the complex nonlinear dependence in high dimensional covariance. Further, it will require new ideas to relax the Gaussian or weak dependence assumption in two-sample covariance test. In Section 4.3, without requiring the restricted Gaussian or weak dependence assumption, we derive a new Lyapunov-type bound for the joint distribution of quadratic form statistics and maximum form statistics and prove the correct asymptotic size of the proposed method.

4.2. Asymptotic Size under the Gaussian Assumption

Now, we present the joint limiting law for M_{n_1,n_2} and T_{n_1,n_2} under the null hypothesis.

Theorem 1. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are Gaussian random vectors, Assumptions 1–2 hold, and $\log p = o((n_1 + n_2)^{\frac{1}{5}})$. Under the null hypothesis H_0, for any $x, y \in \mathbb{R}$, we have

$$P\left(\frac{T_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \le x,\ M_{n_1,n_2} - 4\log p + \log\log p \le y\right) \to \Phi(x) \cdot G(y) \tag{4.1}$$

as $n_1, n_2, p \to \infty$, where $G(y) = \exp\left(-\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{y}{2}\right)\right)$ (or $\Phi(x)$) is the cumulative distribution function of the Gumbel distribution (or the standard normal distribution).

Given Theorems 1–2 from Li and Chen (2012) and Theorem 1 from Cai, Liu, and Xia (2013), Theorem 1 implies that M_{n_1,n_2} and T_{n_1,n_2} are asymptotically independent. To the best of our knowledge, we present the first proof of the asymptotic independence between quadratic form statistics and maximum form statistics for testing two-sample covariance matrices.

In the sequel, we provide a high-level intuition to prove the asymptotic independence result (4.1). First of all, it is worth mentioning that under Assumption 1, all the third-moment and fourth-moment terms in A_{n_1}, B_{n_2}, and C_{n_1,n_2} are of small order than the leading second-moment terms, which may be neglected when deriving the asymptotic normality. Hence, in theoretical analysis, we may consider the simplified statistic of T_{n_1,n_2} defined by

$$\widetilde{T}_{n_1,n_2} = \frac{1}{n_1(n_1-1)} \sum_{u \neq v} \left(\mathbf{X}_u' \mathbf{X}_v\right)^2 + \frac{1}{n_2(n_2-1)} \sum_{u \neq v} \left(\mathbf{Y}_u' \mathbf{Y}_v\right)^2 - \frac{2}{n_1 n_2} \sum_u \sum_v \left(\mathbf{X}_u' \mathbf{Y}_v\right)^2. \tag{4.2}$$

As pointed out by Li and Chen (2012), $\widetilde{T}_{n_1,n_2}$ and T_{n_1,n_2} shares the same asymptotic behavior.

Compared with the one-sample covariance tests in Li and Xue (2015), Li, Xue, and Zou (2018), and He et al. (2021), it is significantly more difficult to analyze the asymptotic joint distribution given the complicated dependence in the two-sample tests for large covariance matrices. To address this challenge, we use a decorrelation technique to address the complex nonlinear dependence in high dimensional covariance matrices. Specifically, we introduce a decorrelated statistic $T^*_{n_1,n_2}$. Under H_0 : $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, we may partition $\mathbf{X}$ and $\mathbf{Y}$ as

$$\mathbf{X}_{p\times 1} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} \quad \text{and}$$

$$\mathbf{Y}_{p\times 1} = \begin{pmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mathbf{0}_{p-q} \\ \mathbf{0}_q \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}\right),$$

where $\mathbf{X}^{(1)}, \mathbf{Y}^{(1)} \in \mathbb{R}^{p-q}$, $\mathbf{X}^{(2)}, \mathbf{Y}^{(2)} \in \mathbb{R}^q$ for integer q satisfying $q = O(\log p)$. Let $\mathbf{Z}_1 = \mathbf{X}^{(1)} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}^{(2)}$, $\mathbf{Z}_2 = \mathbf{X}^{(2)}$, $\mathbf{W}_1 = \mathbf{Y}^{(1)} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{Y}^{(2)}$, $\mathbf{W}_2 = \mathbf{Y}^{(2)}$. It's easy to see that $\mathbf{Z}_1$ is independent of $\mathbf{Z}_2$, and the same results hold for $\mathbf{W}_1$ and $\mathbf{W}_2$. Back to the sample level, we have that $\{\mathbf{Z}_{1u}\}_{u=1}^{n_1}$ and $\{\mathbf{W}_{1v}\}_{v=1}^{n_2}$ iid follow $N_{p-q}(\mathbf{0}, \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})$. Following the pattern of $\widetilde{T}_{n_1,n_2}$ in (4.2), we define

$$T^*_{n_1,n_2} = \frac{1}{n_1(n_1-1)} \sum_{u \neq v} \left(\mathbf{Z}_{1u}' \mathbf{Z}_{1v}\right)^2 + \frac{1}{n_2(n_2-1)} \sum_{u \neq v} \left(\mathbf{W}_{1u}' \mathbf{W}_{1v}\right)^2 - \frac{2}{n_1 n_2} \sum_u \sum_v \left(\mathbf{Z}_{1u}' \mathbf{W}_{1v}\right)^2. \tag{4.3}$$

$\{\mathbf{Z}_{1u}\}_{u=1}^{n_1}$ and $\{\mathbf{W}_{1v}\}_{v=1}^{n_2}$ are regarded as a decorrelated version of $\{\mathbf{X}_u\}_{u=1}^{n_1}$ and $\{\mathbf{Y}_v\}_{v=1}^{n_2}$, respectively. $T^*_{n_1,n_2}$ is regarded as the

$\widetilde{T}_{n_1,n_2}$ statistic derived from the decorrelated samples. The above decorrelation shares a similar philosophy with Shi et al. (2019). We should point out that Shi et al. (2019) used the decorrelation to study the linear hypothesis testing for high-dimensional generalized linear models, but the nonlinear dependence in the two-sample covariance testing is much more challenging than the linear hypothesis testing.

Next, we study the joint distribution of M_{n_1,n_2} and $\widetilde{T}_{n_1,n_2}$. Let A be the event associated with the maximum statistic M_{n_1,n_2} and B be the event associated with the quadratic statistic $\widetilde{T}_{n_1,n_2}$. We use the simple fact that $A = \cup_i A_i$. Then, we may rewrite the joint probability $P(A \cap B)$ into the probability for a union of events, that is, $P(A \cap B) = P((\cup_i A_i) \cap B)$. Then, we give the proof sketch to derive the upper bound $P(A \cap B) - P(A)P(B) \le o(1)$. We begin with a union bound to obtain that $P(\cup_i (A_i \cap B)) \le \sum_i P(A_i \cap B)$. In order to deal with the joint probability of $A_i \cap B$, we further decompose the quadratic statistic into two parts: $T^*_{n_1,n_2}$ is independent of A_i, and the remaining term $\widetilde{T}_{n_1,n_2} - T^*_{n_1,n_2}$ is associated with A_i. Consequently, B can be written as $B = B_i^c \cup B_i$, in which B_i^c represents to the event corresponding to $T^*_{n_1,n_2}$. Therefore, $\sum_i P(A_i \cap B) \le \sum_i P(A_i \cap B_i^c) + \sum_i P(A_i \cap B_i) \le \sum_i P(A_i)P(B_i^c) + \sum_i P(B_i)$. Lemma 2 suggests $T^*_{n_1,n_2}$ is sufficiently close to $\widetilde{T}_{n_1,n_2}$ so that we have $P(B_i^c) \approx P(B)$, $\sum_i P(A_i) \to P(A)$ and $\sum_i P(B_i) = o(1)$. The lower bound $o(1) \le P(A \cap B) - P(A)P(B)$ can be similarly derived from the Bonferroni inequality. Therefore, we can prove the asymptotic independence given that $|P(A \cap B) - P(A)P(B)| = o(1)$.

In what follows, we present three useful lemmas to prove (4.1) in Theorem 1.

Lemma 1 (Asymptotic Normality). Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are Gaussian random vectors and Assumption 1 holds, and $q = O(\log p)$. Under the null hypothesis H_0, as $n_1, n_2, p \to \infty$, we have

$$\frac{T^*_{n_1,n_2}}{2\left(n_1^{-1} + n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)} \xrightarrow{d} N(0,1). \tag{4.4}$$

Lemma 2 (Exponential Decay). Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are Gaussian random vectors and Assumption 1 holds. Under the null hypothesis H_0, for any $0 < \epsilon < 1$, there exist positive constants c, C that do not depend on p, n_1, n_2, such that

$$P\left(\frac{\left|\widetilde{T}_{n_1,n_2} - T^*_{n_1,n_2}\right|}{2\left(n_1^{-1} + n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)} \ge \epsilon\right) \le C\exp\left[-c\epsilon[\log\{p \vee (n_1 + n_2)\}]^{1+\delta_0}\right]. \tag{4.5}$$

where $1 > \delta_0 > 0$ is defined in Assumption 1(ii).

Remark 4.3. Lemma 2 presents the concentration inequality for the two-sample degenerate U-statistic $\widetilde{T}_{n_1,n_2} - T^*_{n_1,n_2}$. The classical concentration inequality for degenerate U-statistics $U_m(f)$ in Proposition 2.3(c) of Arcones and Gine (1993) requires that $\|f\|_\infty \le c$, where f is a real-valued function of m variables. However, $\|f\|_\infty \le c$ does not hold for the U-statistic $\widetilde{T}_{n_1,n_2} - T^*_{n_1,n_2}$. Thus, we use the truncation technique and the concentration inequality of Arcones and Gine (1993) to prove Lemma 2.

As a final step, Lemma 3 derives the joint limiting distribution of the test statistic M_{n_1,n_2} and the simplified statistic $\widetilde{T}_{n_1,n_2}$, which directly implies Theorem 1.

Lemma 3. Under the assumptions of Theorem 1, $\forall x, y \in \mathbb{R}$, as $n_1, n_2, p \to \infty$, we have

$$P\left(\frac{\widetilde{T}_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \le x,\ M_{n_1,n_2} - 4\log p + \log\log p \le y\right) \to \Phi(x) \cdot G(y). \tag{4.6}$$

Lemma 1 shows that such decorrelation procedure does not affect the asymptotic behavior of the quadratic test statistic. Lemma 2 depicts the tail behavior of the difference between $\widetilde{T}_{n_1,n_2}$ and $T^*_{n_1,n_2}$ with explicit decaying rate. Lemmas 1 and 2 lay the foundation of replacing $\widetilde{T}_{n_1,n_2}$ with $T^*_{n_1,n_2}$ in the theoretical analysis.

Given the explicit joint distribution of M_{n_1,n_2} and T_{n_1,n_2}, we proceed to present the asymptotic properties of our proposed Fisher's test F_{n_1,n_2}. Recall that c_α is the upper α-quantile of χ^2_4 distribution and $F_{n_1,n_2} = -2\log(p_M) - 2\log(p_T)$ rejects H_0 if F_{n_1,n_2} is as extreme as c_α. On top of the asymptotic independence established in Section 4.2 and by simple probability transformation, it's easy to obtain the null distribution of F_{n_1,n_2}, and therefore, the asymptotic size of the test. The results are formally presented in Theorem 2.

Theorem 2 (Asymptotic Size). Under the same assumptions as in Theorem 1, the Fisher's test achieves accurate asymptotic size, that is, under the null hypothesis,

$$P\left(F_{n_1,n_2} \ge c_\alpha\right) \to \alpha \quad \text{as } n_1, n_2, p \to \infty.$$

Remark 4.4. Besides Fisher's method, the asymptotic independence result makes it feasible to combine p-values using other approaches such as Tippett's method (Tippett 1931), Stouffer's method (Stouffer et al. 1949), and Cauchy combination (Liu and Xie 2020).

4.3. Asymptotic Size without Requiring the Gaussian Assumption

The Gaussian assumption is essential to prove the asymptotic independence result in Theorem 1 and the correct asymptotic size property in Theorem 2. However, the Gaussian assumption can be violated in real applications. In the sequel, we study the asymptotic joint distribution of quadratic form statistics and maximum form statistics and prove the accurate asymptotic size of the proposed method without requiring the Gaussian assumption.

To study the asymptotic size without the normality, we assume the following condition:

Assumption 3. Let $\mathbf{X}_u = \mathbf{\Gamma}_1 \boldsymbol{\zeta}_{1u}$ and $\mathbf{Y}_v = \mathbf{\Gamma}_2 \boldsymbol{\zeta}_{2v}$ for $u = 1, \ldots, n_1, v = 1, \ldots, n_2$, where $\mathbf{\Gamma}_1 \mathbf{\Gamma}_1^T = \mathbf{\Sigma}_1$, $\mathbf{\Gamma}_2 \mathbf{\Gamma}_2^T = \mathbf{\Sigma}_2$. For any $i = 1$ or 2, $\{\boldsymbol{\zeta}_{ij}\}_{j=1}^{n_i}$ are iidrandom vectors with $\mathrm{E}\boldsymbol{\zeta}_{ij} = 0$ and $\operatorname{cov}(\boldsymbol{\zeta}_{ij}) = \mathbf{I}_{m_i}$ with $m_i \ge p$. Furthermore, for any $1 \le k \le p$, $\mathrm{E}\zeta_{ijk}^4 = 3$ and $\mathrm{E}\zeta_{ijk}^8 < \infty$. Further for all integers s and $\rho_v \ge 0$ such that $\sum_{v=1}^s \rho_v \le 8$ and $\mathrm{E}\zeta_{ijk_1}^{\rho_1} \ldots \zeta_{ijk_s}^{\rho_s} = \mathrm{E}(\zeta_{ijk_1}^{\rho_1}) \ldots \mathrm{E}(\zeta_{ijk_s}^{\rho_s})$ for $k_1 \ne k_2 \ne \cdots \ne k_s$.

Remark 4.5. Assumption 3 is analogous to the assumption (A3) in Li and Chen (2012) or (C3) in Cai, Liu, and Xia (2013).

Assumption 3 includes the Gaussian assumption as the special example, and it also holds for the elliptically contoured distributions (Anderson 2003).

Theorem 3 presents the asymptotic independence result and the accurate asymptotic size property when the dimension p grows at a slower rate than the sample size n_1 or n_2.

Theorem 3. Suppose that Assumptions 1–3 hold. Under the null hypothesis H_0, for any $x, y \in \mathbb{R}$, when $p = o((n_1+n_2)^{1/7})$ and $\mathbf{\Sigma}$ is invertible, we have

$$P\left(\frac{T_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \le x,\ M_{n_1,n_2} - 4\log p + \log\log p \le y\right) \to \Phi(x)\cdot G(y) \tag{4.7}$$

as $n_1, n_2, p \to \infty$, where $G(y) = \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{y}{2}\right)\right)$ (or $\Phi(x)$) is the cumulative distribution function of the Gumbel distribution (or the standard normal distribution). Moreover,

$$P\left(F_{n_1,n_2} \ge c_\alpha\right) \to \alpha \quad \text{as } n_1, n_2, p \to \infty.$$

The Lyapunov-type bound from Bentkus (2005) provides the essential tool to prove Theorem 3. The conditions, $p = o((n_1+n_2)^{1/7})$ and $\mathbf{\Sigma}$ is invertible, are necessary conditions to apply the Lyapunov-type bound from Bentkus (2005). Let $\mathbf{Z}_k = \text{Vec}(Z_{k1}, \ldots, Z_{kd})$ be a d-dimensional vector of interest in $\mathbb{R}^d$ with mean $\mathbf{0}$ and covariance $\mathbf{\Omega} = (\omega_{ij})_{d\times d}$ for $1 \le k \le m$. Let $\eta_1, \ldots, \eta_m$ be independent Gaussian random vectors in $\mathbb{R}^d$ with mean $\mathbf{0}$ and covariance $\text{cov}(\eta_k) = \mathbf{\Omega}$. By definition, η_k's have the same covariance matrix as $\mathbf{Z}_k$'s. If $\mathcal{A}$ is the set of all measurable convex sets in $\mathbb{R}^d$, Bentkus (2005) proved that

$$\sup_{A\in\mathcal{A}}\left|P\left(\frac{\sum_{i=1}^m \mathbf{Z}_i}{\sqrt{m}} \in A\right) - P\left(\frac{\sum_{i=1}^m \eta_i}{\sqrt{m}} \in A\right)\right| \le Cd^{1/4}m^{-3/2}\sum_{i=1}^m \mathrm{E}\|\mathbf{\Omega}^{-1/2}\mathbf{Z}_i\|_2^3,$$

where C does not depend on d and m. After carefully constructing the random vector $\mathbf{Z}_k$ and the convex set A, we can combine the above Lyapunov-type bound and simplified proofs of Theorem 1 under the Gaussian assumption to obtain the asymptotic independence result and the accurate asymptotic size property without requiring the Gaussian assumption.

In the sequel, we study the asymptotic size of the proposed method without requiring the Gaussian assumption when p is larger than n_1 or n_2. We assume the following condition:

Assumption 4. Suppose that

(i) there is a fixed constant $t_0 > 0$ such that $E[e^{t_0 X_{ij}^2}] < \infty$ and $E[e^{t_0 Y_{kj}^2}] < \infty$ for all $1 \le i \le n_1, 1 \le k \le n_2, 1 \le j \le p$.
(ii) $(\log p)^{13} = o((n_1+n_2))$.

Theorem 4. Given Assumptions 1–4, under H_0, we have

$$P\left(\frac{T_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \le x,\ M_{n_1,n_2} - 4\log p + \log\log p \le y\right) \to \Phi(x)\cdot G(y) \quad \text{as } n_1, n_2, p \to \infty,$$

where $G(y) = \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{y}{2}\right)\right)$ (or $\Phi(x)$) is the cumulative distribution function of the Gumbel distribution (or the standard normal distribution). Moreover, we have

$$P\left(F_{n_1,n_2} \ge c_\alpha\right) \to \alpha \quad \text{as } n_1, n_2, p \to \infty.$$

Theorem 4 presents the asymptotic independence result and the accurate asymptotic size property when p is larger than n_1 or n_2 as $n_1, n_2, p \to \infty$. Note that the Lyapunov-type bound from Bentkus (2005) no longer works under such a high-dimensional setting. To prove Theorem 4, we follow Chernozhukov, Chetverikov, and Kato (2013) to prove a Lyapunov-type bound for the joint distribution of quadratic form statistics and maximum form statistics. Given this bound, we can combine it and simplified proofs of Theorem 1 to obtain the desired result without requiring the Gaussian assumption.

In the current literature, Bentkus (2005) derived the Lyapunov-type bound for quadratic form statistics, and Chernozhukov, Chetverikov, and Kato (2013, 2017) derived the Lyapunov-type bound for maximum form statistics. There is no existing Lyapunov-type bound for the joint distribution of quadratic form statistics and maximum form statistics under the high-dimensional setting. Thus, our result fills this gap and makes a separate contribution to the literature.

4.4. Asymptotic Power

Two classes of alternative hypotheses (i.e., dense alternatives and sparse alternatives) have received a lot of attention when testing the high-dimensional covariance matrices. For dense alternatives, as shown in Li and Chen (2012), the parameter space of interest is defined using the squared Frobenius norm $\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2 = \text{tr}\{(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2\}$, that is,

$$\mathcal{G}_d = \left\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \frac{1}{n_1}\text{tr}(\mathbf{\Sigma}_1^2) + \frac{1}{n_2}\text{tr}(\mathbf{\Sigma}_2^2) = o\left(\text{tr}\{(\mathbf{\Sigma}_1-\mathbf{\Sigma}_2)^2\}\right)\right\}. \tag{4.8}$$

The distributions under $H_1 : (\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{G}_d$ may include many small nonzero entries of about the same size in $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$, which is why we called $\mathcal{G}_d$ the "dense alternatives." For sparse alternatives, as shown in Cai, Liu, and Xia (2013), the parameter space of interest is defined using the entry-wise maximum norm $\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_{\max} = \max_{1\le i\le j\le p}|\sigma_{1,ij} - \sigma_{2,ij}|$, that is,

$$\mathcal{G}_s = \left\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \max_{1\le i\le j\le p}\frac{|\sigma_{ij1} - \sigma_{ij2}|}{\sqrt{\theta_{ij1}/n_1 + \theta_{ij2}/n_2}} \ge 4\sqrt{\log p}\right\}. \tag{4.9}$$

The distributions under $H_1 : (\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{G}_s$ may include only a few large nonzero entries in $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$, which is why we called $\mathcal{G}_s$ the "sparse alternatives."

Li and Chen (2012) and Cai, Liu, and Xia (2013) provided power analysis of tests T_{n_1,n_2} and M_{n_1,n_2} over the dense alternative $\mathcal{G}_d$ and the sparse alternative $\mathcal{G}_s$, respectively. However, T_{n_1,n_2} performs poorly under the sparse alternative $\mathcal{G}_s$, and M_{n_1,n_2} performs poorly under the dense alternative $\mathcal{G}_d$. In addition to the numerical demonstration in simulation studies (Section 5), we provide the following theoretical examples to demonstrate this fact. On the one hand, we assume that $\sigma_{ij1} -$

$\sigma_{ij2} = \frac{1}{\sqrt{n_1 p}}$ for $1 \le i < j \le p$ for $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$. When $\lambda_1(\mathbf{\Sigma}_1)$ and $\lambda_1(\mathbf{\Sigma}_2)$ are finite, it is easy to show that $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{G}_d$ and also that $\frac{|\sigma_{ij1}-\sigma_{ij2}|}{\sqrt{\theta_{ij1}/n_1+\theta_{ij2}/n_2}} < C\sqrt{\frac{1}{p}}$ and $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \notin \mathcal{G}_s$. By Lemma 4 from Cai, Liu, and Xia (2013), as $n_1, n_2, p \to \infty$, we have

$$
\begin{aligned}
&P\left(\max_{1\le i\le j\le p} \frac{\left(\widehat{\sigma}_{ij1}-\widehat{\sigma}_{ij2}\right)^2}{\theta_{ij1}/n_1+\theta_{ij2}/n_2} - 4\log p + \log\log p \ge q_\alpha\right)\\
&\le P\left(\max_{1\le i\le j\le p} \frac{\left(\widehat{\sigma}_{ij1}-\widehat{\sigma}_{ij2}-(\sigma_{ij1}-\sigma_{ij2})\right)^2}{\theta_{ij1}/n_1+\theta_{ij2}/n_2}\right.\\
&\qquad\left. -4\log p + \log\log p + \frac{C^2}{p} + \frac{C\log p}{\sqrt{p}} \ge q_\alpha\right)\\
&+ P\left(\max_{1\le i\le j\le p} \frac{\left(\widehat{\sigma}_{ij1}-\widehat{\sigma}_{ij2}-(\sigma_{ij1}-\sigma_{ij2})\right)^2}{\theta_{ij1}/n_1+\theta_{ij2}/n_2}\right.\\
&\qquad\left. > 4\log p - \frac{1}{2}\log\log p\right) \to \alpha < 1.
\end{aligned}
$$

Thus, M_{n_1,n_2} has the low asymptotic power under this dense alternative in $\mathcal{G}_d$.

On the other hand, we assume that $\sigma_{121} - \sigma_{122} = 5\sqrt{\frac{\log p}{n}}$ for $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$. When $\lambda_1(\mathbf{\Sigma}_1)$ and $\lambda_1(\mathbf{\Sigma}_2)$ are finite, it is easy to show that $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{G}_s$ and $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \notin \mathcal{G}_d$. Under Assumption 5(iii), by Theorem 1 from Li and Chen (2012), as $n_1, n_2, p \to \infty$, we have

$$
\begin{aligned}
&P\left(\frac{T_{n_1,n_2}}{\sigma_{0,n_1,n_2}} \ge z_\alpha\right)\\
&= P\left(\frac{T_{n_1,n_2} - \mathrm{tr}(\mathbf{\Sigma}_1-\mathbf{\Sigma}_2)^2}{\sigma_{n_1,n_2}} \ge \frac{\sigma_{0,n_1,n_2}}{\sigma_{n_1,n_2}} z_\alpha - \frac{\mathrm{tr}(\mathbf{\Sigma}_1-\mathbf{\Sigma}_2)^2}{\sigma_{n_1,n_2}}\right)\\
&\quad \to \alpha < 1,
\end{aligned}
$$

where

$$
\begin{aligned}
\sigma^2_{n_1,n_2} = \sum_{i=1}^{2}\Bigg[&\frac{4\mathrm{tr}^2(\mathbf{\Sigma}_i^2)}{n_i^2} + \frac{8\mathrm{tr}\{(\mathbf{\Sigma}_i^2 - \mathbf{\Sigma}_1\mathbf{\Sigma}_2)^2\}}{n_i}\\
&+ \frac{4\Delta_i \mathrm{tr}(\mathbf{\Gamma}_i^T(\mathbf{\Sigma}_1-\mathbf{\Sigma}_2)\mathbf{\Gamma}_i) \circ \mathbf{\Gamma}_i^T(\mathbf{\Sigma}_1-\mathbf{\Sigma}_2)\mathbf{\Gamma}_i)}{n_i}\Bigg]\\
&+ \frac{8}{n_1 n_2}\mathrm{tr}^2(\mathbf{\Sigma}_1\mathbf{\Sigma}_2).
\end{aligned}
$$

The above inequality holds by using the fact that $\frac{\mathrm{tr}(\mathbf{\Sigma}_1-\mathbf{\Sigma}_2)^2}{\sigma_{0,n_1,n_2}} \to 0$ and $\frac{\sigma_{0,n_1,n_2}}{\sigma_{n_1,n_2}} \to 1$ as $n_1, n_2 \to \infty$. Thus, T_{n_1,n_2} has the low asymptotic power under this sparse alternative in $\mathcal{G}_s$.

As an effective combination test, our proposed method F_{n_1,n_2} exploits the full potential of M_{n_1,n_2} and T_{n_1,n_2} to boost their respective power against more general alternatives. To establish the power enhancement, we consider the following conditions in Assumption 5:

Assumption 5. As $n_1, n_2, p \to \infty$, suppose that

(i) $n_1/(n_1+n_2) \to \gamma$, for some constant $\gamma \in (0,1)$.
(ii) For any $i,j,k,l \in \{1,2\}$ $\mathrm{tr}(\mathbf{\Sigma}_k\mathbf{\Sigma}_l) \to \infty$ and $\mathrm{tr}\{\mathbf{\Sigma}_i\mathbf{\Sigma}_j\mathbf{\Sigma}_k\mathbf{\Sigma}_l\} = o\{\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j)\,\mathrm{tr}(\mathbf{\Sigma}_k\mathbf{\Sigma}_l)\}$ as $n_1, n_2, p \to \infty$.
(iii) Let $\mathbf{X}_u = \mathbf{\Gamma}_1\zeta_{1u}$ and $\mathbf{Y}_v = \mathbf{\Gamma}_2\zeta_{2v}$ for $u = 1,\dots,n_1$, $v = 1,\dots,n_2$, where $\mathbf{\Gamma}_1\mathbf{\Gamma}_1^T = \mathbf{\Sigma}_1$, $\mathbf{\Gamma}_2\mathbf{\Gamma}_2^T = \mathbf{\Sigma}_2$. For any $i = 1$ or 2, $\{\zeta_{ij}\}_{j=1}^{n_i}$ are iid random vectors with $\mathrm{E}\zeta_{ij} = 0$ and $\mathrm{cov}(\zeta_{ij}) = \mathbf{I}_{m_i}$ with $m_i \ge p$. For any $1 \le k \le p$, $\mathrm{E}\zeta_{ijk}^4 = 3+\Delta_i$ and $\mathrm{E}\zeta_{ijk}^8 < \infty$, where Δ_i is a constant that describes the difference between the fourth moment of ζ_{ijk} and that of the standard normal distribution. Further for all integers s and $\rho_v \ge 0$ such that $\sum_{v=1}^{s}\rho_v \le 8$ and $\mathrm{E}\zeta_{ijk_1}^{\rho_1}\dots\zeta_{ijk_s}^{\rho_s} = E(\zeta_{ijk_1}^{\rho_1})\dots\mathrm{E}(\zeta_{ijk_s}^{\rho_s})$ for $k_1 \ne k_2 \ne \dots \ne k_s$.
(iv) Suppose that X_{kj} and Y_{lj} ($1 \le k \le n_1, 1 \le l \le n_2$ and $1 \le j \le p$) satisfy sub-Gaussian tail conditions with $\log p = o(n^{1/5})$ or polynomial-type tail conditions with $p = o(n^{\gamma_0})$ for some constant $\gamma_0 \ge 1$.

Remark 4.6. Assumption 5 is similar to (A1), (A2), and (A3) in Li and Chen (2012) and (C2) or (C2*) in Cai, Liu, and Xia (2013). The condition (i) is standard in the literature. The conditions (ii) and (iii) assume a general multivariate model as in Li and Chen (2012), The condition (iv) is similar to the moment condition (C2) or (C2*) in Cai, Liu, and Xia (2013), and it is weaker than the Gaussian assumption.

Theorem 5 presents the enhanced power of the proposed method uniformly over $\mathcal{G}_d \cup \mathcal{G}_s$.

Theorem 5 (Asymptotic Power). Suppose that Assumption 5 holds. The proposed Fisher's method achieves the consistent asymptotic power, that is, under the alternative hypothesis,

$$\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_d\cup\mathcal{G}_s} P\left(F_{n_1,n_2} \ge c_\alpha\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty.$$

Remark 4.7. (Bahadur Efficiency) Littell and Folks (1971, 1973) studied the asymptotic optimality of using Fisher's method to combine independent tests. They proved for the two asymptotically independent tests $p_T = 1 - \Phi\left(T_{n_1,n_2}/\widehat{\sigma}_{0,n_1,n_2}\right)$ and $p_M = 1 - G(M_{n_1,n_2} - 4\log p + \log\log p)$, the Fisher's method delivers the largest exact Bahadur slope among all reasonable combination approaches, indicating the fastest decay rate for the p-values under the alternatives. Fisher's combined test is asymptotically optimal in terms of Bahadur relative efficiency.

5. Simulation Studies

This section examines the finite-sample performance of our Fisher's combined probability test, compared to the tests proposed by Cai, Liu, and Xia (2013) (referred to as the CLX test in the following context) and Li and Chen (2012) (referred to as the LC test). To demonstrate the advantage of using Fisher's combination, we also compare with an equally weighted combination test (referred to as the Weighted test in later discussion). More specifically, the test statistic is defined as $J = \widehat{\sigma}^{-1}_{n_1,n_2}T_{n_1,n_2} + (M_{n_1,n_2} - 4\log p + \log\log p)$. The distribution of J is calculated by the convolution of $\Phi'(\cdot)$ and $G'(\cdot)$.

In the simulation studies, we generated two random samples $\{\mathbf{X}_1,\dots,\mathbf{X}_{n_1}\}$ iid from $N_p(\mathbf{0}, \mathbf{\Sigma}_1)$ and $\{\mathbf{Y}_1,\dots,\mathbf{Y}_{n_2}\}$ iid from $N_p(\mathbf{0}, \mathbf{\Sigma}_2)$. The sample sizes are taken to be $n_1 = n_2 = N$ with $N = 100$ and 200, while the dimension p varies over the values 100, 200, 500, 800, and 1000. The significance level is set to be

0.05 for all the tests. For each simulation setting, the average number of rejections is reported over 1000 replications.

We also conducted simulation studies to demonstrate the numerical performance with respect to non-Gaussian data. We consider several non-Gaussian distributions in the data generating process, including the gamma distribution, t distribution, uniform distribution, contaminated normal distribution, and so on. For space consideration, we present these results in Section S.4 of the supplementary materials.

Under the null hypothesis H_0, we set $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}^{*(i)}, i = 1, \ldots, 7$, and consider the following seven models to evaluate the testing size.

(i) $\mathbf{\Sigma}^{*(1)} = \mathbf{I}_p$.

(ii) $\mathbf{\Sigma}^{*(2)} = (\mathbf{\Omega}^{*(2)})^{-1}$, where $\omega_{ij}^{*(2)} = 0.5^{|i-j|}$.

(iii) $\mathbf{\Sigma}^{*(3)}$ is a block diagonal matrix given by each block being $0.5\mathbf{I}_5 + 0.5\mathbb{1}_5\mathbb{1}_5'$.

(iv) $\mathbf{\Sigma}^{*(4)} = (\mathbf{\Sigma}^{(4)} + \delta\mathbf{I}_p)/(1+\delta)$, where $\sigma_{ii}^{(4)} = 1$, $\sigma_{ij}^{(4)} = 0.5 \times \text{Bernoulli}(1, 0.05)$ for $i < j$ and $\sigma_{ij}^{(4)} = \sigma_{ji}^{(4)}$, $\delta = |\lambda_{\min}(\mathbf{\Sigma}^{(4)})| + 0.05$.

(v) $\mathbf{\Sigma}^{*(5)} = \{\sigma_{ij}^{*(5)}\}_{p\times p}$, $\sigma_{ij}^{*(5)} = (-1)^{i+j}0.4^{|i-j|^{1/10}}$.

(vi) $\mathbf{\Sigma}^{*(6)} = \{\sigma_{ij}^{*(6)}\}_{p\times p}$, $\sigma_{ij}^{*(6)} = 0.9^{|i-j|}$.

(vii) $\mathbf{\Sigma}^{*(7)} = 0.5\mathbf{I}_p + 0.5\mathbf{1}_p\mathbf{1}_p'$.

Model (i) is the most commonly used multivariate standard normal distribution. Model (ii) and Model (iii) are the cases when the true covariance matrices have certain banded-type and block-type sparsity. Model (iv) is also a sparse matrix yet without any specific sparsity pattern. Model (v) was first proposed by Srivastava and Yanagihara (2010) and further studied in Cai, Liu, and Xia (2013). Model (vi) is the autoregressive matrix with strong positive autocorrelation, and Model (vii) is the compound symmetry matrix.

To evaluate the power of the tests, we consider the scenarios when the differences of the two covariance matrices satisfy certain structure. There are two types of alternatives we desire to look into: the sparse alternative H_s and the dense alternative H_d.

Generally speaking, the sparse alternative shares commonality among different models. Let $\mathbf{U}$ denote the difference between $\mathbf{\Sigma}_2$ and $\mathbf{\Sigma}_1$, that is, $\mathbf{U} = \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$. As in Cai, Liu, and Xia (2013), we consider the situation when $\mathbf{U}$ is a symmetric sparse matrix with eight random nonzero entries. The locations of four nonzero entries are randomly selected from the upper triangle of $\mathbf{U}$, each with a magnitude of Unif(0,4)$\times \max_{1\le j\le p}\sigma_{jj}^*$. The other four are determined by symmetry. Then we generate samples from these covariance pairs $\left(\mathbf{\Sigma}_1^{(i)}, \mathbf{\Sigma}_2^{(i)}\right)$, $i = 1, \ldots, 7$, in order to evaluate the power of the tests against sparse alternative, where $\mathbf{\Sigma}_1^{(i)} = \mathbf{\Sigma}^{*(i)} + \tau\mathbf{I}$ and $\mathbf{\Sigma}_2^{(i)} = \mathbf{\Sigma}^{*(i)} + \tau\mathbf{I} + \mathbf{U}$, with $\tau = |\min\{\lambda_{\min}(\mathbf{\Sigma}^{*(i)} + \mathbf{U}), \lambda_{\min}(\mathbf{\Sigma}^{*(i)})\}| + 0.05$.

In terms of the dense alternative setting, since the seven models differ a lot from each other, we shall discuss their corresponding alternative settings separately afterwards. To begin with, we shall take a look at the simplest case in Model (i). We consider its dense alternative to be the AR(1) model with parameter $\rho = 0.1$ and 0.2, denoted by $\mathbf{\Sigma}_\rho^{\text{AR}}$. In another word, we generate the copies of $\mathbf{X}$ from the p-dimensional standard normal while copies of $\mathbf{Y}$ from $N_p\left(\mathbf{0}, \mathbf{\Sigma}_\rho^{\text{AR}}\right)$. For Models (ii), (iii), and (iv), we set the covariance pairs to be $\mathbf{\Sigma}_1^{(i)} = \mathbf{\Sigma}^{*(i)}$ and $\mathbf{\Sigma}_2^{(i)} = \mathbf{\Sigma}^{*(i)} + \eta(\mathbf{I}_p - \mathbf{\Sigma}^{*(i)})$, where η is used to weaken the signals in an effort to avoid trivial power in all tests. We follow the same alternative hypothesis as in Srivastava and Yanagihara (2010) for Model (v) by letting $\mathbf{\Sigma}_1^{(5)} = \mathbf{\Sigma}^{*(5)}$ and $\mathbf{\Sigma}_2^{(5)} = \{\sigma_{ij}^{(5)}\}_{p\times p}$ with $\sigma_{ij}^{(5)} = (-1)^{i+j}0.6^{|i-j|^{1/10}}$. We use the AR(1) matrix with $\rho = 0.85$ as the dense alternative for Model (vi) and the compound symmetry matrix with correlation 0.4 for Model (vii).

For each covariance model, we generate samples independently from $N_p(\mathbf{0}, \mathbf{\Sigma}^{*(i)})$ to evaluate the size, and use different covariance pairs described above to examine the power against sparse and dense alternatives. The empirical size and power are calculated based on 1000 replications at significance level of 5% and the results are reported in Tables 1–3.

We have the following findings from the size and power comparisons in Tables 1–3:

1. Under H_0, the sizes of all four tests are well retained close to the nominal level 0.05 for Models (i)–(iv). The LC test suffers from the size distortion in Models (v) and (vii) while the CLX test is overly conservative in Models (vi) and (vii), because of the violations of the test assumptions on covariance matrices. This leads to the unsatisfactory size approximations using our proposed test and the Weighted test in Models (v)–(vii) as both of them rely on the performance of the LC and CLX tests.
2. The CLX test is demonstrated to be powerful under the sparse alternative H_s, however, its performance is not satisfactory under the dense alternative. In the meantime, the LC test remains a high power under the dense alternative H_d, whereas performs poorly against the sparse alternative with a tendency of decaying as dimension p grows large.
3. The Weighted test is able to maintain a relatively high power under either sparse or dense alternatives. However, we observe that the Weighted test behaves more in tandem with the CLX test compared to the LC test. This is because the CLX test statistic spreads more widely than the LC test statistic, see Figure 1 for a graphical illustration. By taking sum of the two statistics with equal weights, the Weighted test statistic is mostly driven by the CLX statistic. As a result, the Weighted test is more prone to boosting the power under the sparse alternative H_s but less effective in enhancing test power under the dense alternative H_d.
4. In comparison, our proposed Fisher's combined test exhibits competent results. Our proposed test performs as good as the CLX test under the sparse alternative and remains competitive power with the LC test when against the dense alternative. What's more, our proposed test outperforms the Weighted test under the dense alternative, together with comparable performance when against the sparse alternative.

In a summary, based on these simulation results, we find that the proposed method boosts the power against more general alternatives while retaining the desired significance level, which is consistent with our theoretical results in Section 4.

Table 1. Comparison of empirical size and power (%) for model (i) with Gaussian data.

N	p	100	200	500	800	1000	100	200	500	800	1000
		Size					Power under sparse alternative				
100	Proposed	5.7	5.2	4.6	4.7	5.0	99.4	95.0	91.7	85.9	72.6
	CLX	4.1	4.4	3.9	4.4	4.6	99.9	96.6	95.2	89.8	79.5
	LC	5.3	4.9	5.3	4.5	5.2	30.1	13.4	7.9	5.8	5.5
	Weighted	5.0	5.2	4.7	4.0	4.7	99.8	97.2	94.9	89.6	78.1
200	Proposed	6.0	4.9	4.8	4.7	4.2	100.0	100.0	99.9	99.9	99.9
	CLX	5.1	4.6	4.3	4.4	4.2	100.0	100.0	99.9	99.9	100.0
	LC	5.6	5.0	5.2	4.6	4.3	70.7	38.1	8.3	7.3	5.6
	Weighted	5.6	4.8	4.8	4.4	4.1	100.0	100.0	99.9	99.9	100.0
		Power under dense alternative									
		$\rho = 0.1$					$\rho = 0.2$				
100	Proposed	12.7	9.9	11.7	11.4	11.6	56.3	57.0	54.1	53.2	53.9
	CLX	5.1	4.2	5.1	5.0	5.4	12.2	9.8	7.7	6.9	7.1
	LC	14.2	11.5	13.0	14.5	12.3	63.9	62.9	63.4	64.3	64.1
	Weighted	6.6	7.0	6.7	6.9	7.4	27.9	22.3	20.5	17.1	17.0
200	Proposed	22.6	22.4	18.8	19.6	18.0	99.0	99.8	99.3	99.3	99.1
	CLX	5.4	6.5	4.4	4.7	4.2	45.0	42.7	33.0	26.4	27.4
	LC	24.6	24.9	26.8	24.9	25.5	98.2	99.0	99.3	99.5	99.3
	Weighted	10.7	11.1	7.4	7.1	7.0	86.7	85.9	79.4	75.4	75.7

NOTE: This table reports the frequencies of rejection by each method under the null and alternative hypotheses based on 1000 independent replications at the significance level 5%.

Table 2. Comparison of empirical size and power (%) for models (ii), (iii), and (iv) with Gaussian data.

		Model (ii)					Model (iii)					Model (iv)				
N	p	100	200	500	800	1000	100	200	500	800	1000	100	200	500	800	1000
		Size														
100	Proposed	4.7	4.6	6.0	5.5	5.4	6.6	6.1	4.6	4.3	4.9	5.4	5.5	5.6	4.4	6.2
	CLX	4.1	4.5	5.7	5.5	5.3	4.7	5.6	4.2	3.6	4.5	4.2	5.0	5.5	4.3	5.3
	LC	5.5	4.0	5.4	5.2	4.8	5.6	5.8	4.9	4.4	4.8	6.1	5.7	5.6	5.0	5.7
	Weighted	4.3	4.7	5.1	5.6	4.8	5.4	5.9	4.7	3.7	4.9	4.7	5.3	6.2	5.1	5.2
200	Proposed	6.0	5.1	4.3	4.4	4.9	6.4	6.0	5.9	5.1	4.2	5.6	5.7	6.4	4.1	5.0
	CLX	4.0	5.0	4.0	3.5	4.1	4.9	4.9	4.6	3.4	3.3	4.5	4.7	4.4	3.5	4.0
	LC	5.8	5.6	4.2	4.7	4.6	5.8	5.1	5.2	5.7	4.5	5.1	5.3	6.4	6.3	4.5
	Weighted	5.0	5.2	4.4	4.2	4.4	5.3	5.6	5.7	4.2	3.1	5.1	4.4	5.3	4.3	4.7
		Power under sparse alternative														
100	Proposed	99.1	80.4	79.6	71.6	83.6	94.3	87.1	78.2	73.2	79.1	94.0	94.4	92.2	74.4	90.8
	CLX	99.7	83.6	85.9	77.4	90.2	96.7	90.6	82.7	79.5	86.6	95.2	96.5	95.4	81.5	93.8
	LC	32.7	9.5	6.7	6.7	5.6	18.6	7.7	5.3	4.8	6.5	17.7	11.7	6.7	5.8	5.9
	Weighted	99.3	83.3	84.2	77.2	89.4	96.9	89.9	82.2	78.4	85.0	95.8	96.8	94.8	81.9	93.2
200	Proposed	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
	CLX	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	LC	42.2	19.6	7.9	6.4	8.0	59.7	13.4	12.0	8.7	6.0	30.2	22.8	9.7	7.1	7.4
	Weighted	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		Power under dense alternative														
100	Proposed	84.6	84.3	85.3	83.1	84.3	97.0	98.2	98.2	97.9	97.9	75.4	83.3	82.3	83.5	84.4
	CLX	14.3	11.9	10.0	7.1	6.2	36.7	28.4	20.1	17.7	16.6	10.4	7.7	5.8	6.4	6.7
	LC	87.3	89.3	90.5	90.3	91.1	97.0	97.8	98.9	98.5	98.6	81.3	88.7	88.8	90.6	90.5
	Weighted	44.8	33.7	30.2	27.4	23.1	76.5	72.6	64.3	59.6	55.8	33.1	29.1	22.4	26.8	25.6
200	Proposed	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0
	CLX	57.1	49.4	37.9	32.0	28.2	90.6	90.8	89.0	84.4	85.0	45.4	35.6	22.2	21.9	20.1
	LC	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0
	Weighted	98.0	98.2	97.1	92.9	92.1	100.0	100.0	100.0	100.0	100.0	95.0	92.0	88.4	87.5	87.8

NOTE: This table reports the frequencies of rejection by each method under the null and alternative hypotheses based on 1000 independent replications at the significance level 5%.

6. Application to Gene-Set Testing

We demonstrate the power of our proposed test by identifying those sets of genes which potentially have significant differences in covariance matrices across different types of tumors. In biology, each gene does not work individually, but rather tends to function as groups to achieve complex biological tasks. Sets of genes are interpreted by Gene Ontology (GO) terms making use of the Gene Ontology system, in which genes are assigned to a set of predefined bins depending on their functional characteristics. The Gene Ontology covers three domains: biological process (BP), cellular component (CC) and molecular function (MF).

Table 3. Comparison of empirical size and power (%) for models (v), (vi), and (vii) with Gaussian data.

		Model (v)					Model (vi)					Model (vii)				
N	*p*	100	200	500	800	1000	100	200	500	800	1000	100	200	500	800	1000
									Size							
100	Proposed	9.6	8.9	10.8	9.9	9.4	6.9	5.1	4.2	4.9	3.6	8.6	8.9	6.8	8.9	6.2
	CLX	4.1	2.9	4.4	3.7	3.3	1.7	2.0	2.7	2.6	2.8	2.3	1.4	1.2	1.8	0.9
	LC	9.6	8.7	11.7	10.4	9.6	8.4	7.7	5.8	5.4	5.1	10.5	10.9	9.1	10.8	8.7
	Weighted	6.7	6.5	8.2	6.7	5.8	3.6	2.3	3.2	3.4	2.8	1.7	3.4	7.7	8.0	9.0
200	Proposed	9.3	10.7	8.8	9.0	11.3	6.8	5.9	4.2	3.4	4.7	10.3	8.6	6.7	7.5	7.3
	CLX	4.4	4.3	3.7	3.5	5.1	1.7	2.7	1.7	2.3	3.5	2.3	1.8	1.6	0.7	0.7
	LC	9.7	10.1	8.9	8.6	11.3	8.5	6.5	5.8	4.5	5.3	12.0	11.0	8.3	10.6	10.3
	Weighted	6.4	7.9	6.4	6.9	7.6	4.0	3.7	2.4	2.5	3.9	2.0	1.6	2.0	2.2	5.3
									Power under sparse alternative							
100	Proposed	93.7	97.2	83.5	97.5	79.8	93.7	97.8	96.8	86.4	74.2	98.2	95.1	78.9	75.3	70.8
	CLX	95.5	97.7	87.3	99.0	84.3	95.8	99.0	98.8	89.9	80.0	98.9	96.7	84.1	80.5	76.9
	LC	13.1	9.7	6.7	8.6	6.1	13.5	14.1	10.1	5.8	6.2	20.3	12.7	6.9	7.3	5.9
	Weighted	95.5	98.0	86.3	98.7	84.8	95.5	98.7	98.2	88.1	78.2	98.9	96.2	84.5	79.3	76.8
200	Proposed	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	CLX	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	LC	93.7	24.6	9.5	8.8	7.8	86.3	23.1	11.1	7.7	9.2	27.0	21.5	11.3	7.7	6.5
	Weighted	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
									Power under dense alternative							
100	Proposed	85.5	88.0	94.7	94.4	96.8	50.6	57.7	56.1	60.3	58.4	27.2	24.3	27.0	27.4	25.4
	CLX	57.5	61.5	72.2	72.5	78.1	10.2	12.0	7.3	6.9	6.9	10.1	9.1	8.6	9.3	9.4
	LC	85.5	87.8	94.8	94.8	97.0	55.2	63.8	67.2	71.9	71.8	29.6	26.9	30.3	29.4	26.8
	Weighted	77.8	82.5	90.4	91.2	93.6	26.9	27.3	18.9	18.8	18.5	19.2	17.0	18.5	18.7	17.6
200	Proposed	98.6	99.5	99.7	100.0	99.9	85.3	92.7	96.9	98.7	97.7	44.1	43.8	44.4	43.1	45.6
	CLX	87.4	92.2	96.3	97.9	97.7	26.9	28.3	21.5	20.4	18.5	18.7	16.7	17.1	18.0	18.0
	LC	98.5	99.4	99.8	99.9	100.0	89.5	95.0	98.2	98.8	98.8	45.1	47.3	48.4	46.3	49.3
	Weighted	96.9	98.7	99.5	100.0	99.9	58.6	65.4	64.7	64.3	61.7	35.8	32.5	33.5	31.7	33.5

NOTE: This table reports the frequencies of rejection by each method under the null and alternative hypotheses based on 1000 independent replications at the significance level 5%.

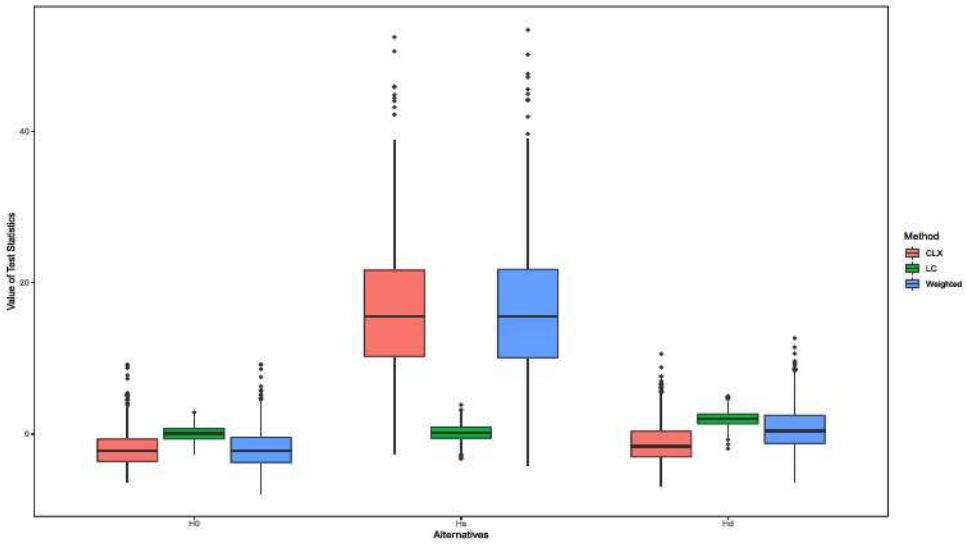

Figure 1. Boxplots of the LC, CLX, and Weighted test statistics. The statistics come from the simulation studies for Model (i) with $N = 100$ and $p = 500$ based on 1000 replications.

We consider the Acute Lymphoblastic Leukemia (ALL) data from the Ritz Laboratory at the Dana-Farber Cancer Institute (DFCI). The latest data is accessible at the ALL package (version 1.24.0) on the Bioconductor website, including the original version published by Chiaretti et al. (2004). The ALL dataset consists of microarrays expression measures of 12,625 probes on Affymetrix chip series HG-U95Av2 for 128 different individuals with acute lymphoblastic leukemia, which is a type of blood cancer in that bone marrow affects white blood cells. Based on the type of lymphocyte that the leukemia cells come from, the disease is classified into subgroups of T-cell ALL and B-cell ALL. In our study, we focus on a subset of the ALL data of 79 patients with the B-cell ALL. We are interested in two types of B-cell tumors: BCR/ABL and NEG, with sample sizes being 37 and 42, respectively.

Table 4. Summary of the dimension of gene-sets for three GO categories.

GO category	Total number	Min	First-Quantile	Median	Third-Quantile	Max
BP	1849	10	15	27	62	2153
CC	306	10	17	32	85	2181
MF	324	10	14	26	68	2148

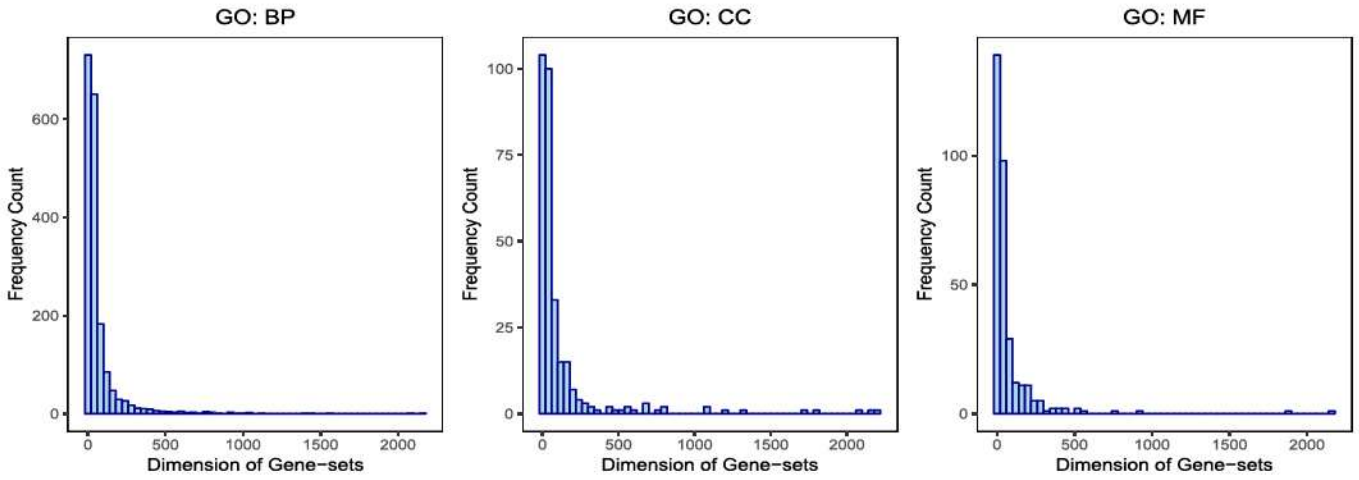

Figure 2. Histograms of the dimension of gene-sets for three GO categories.

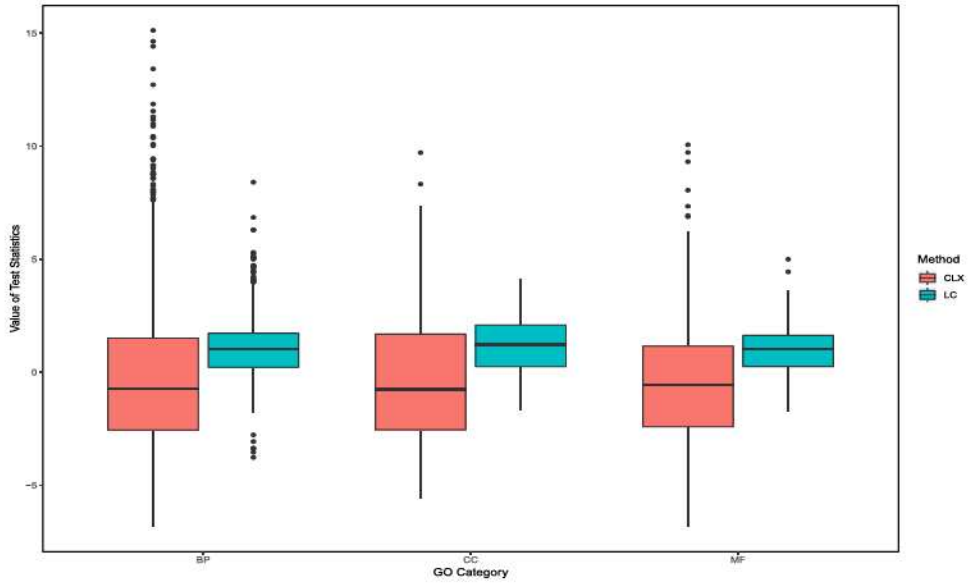

Figure 3. Boxplots of the LC and CLX test statistics for three GO categories.

Let us consider K gene sets $S_1, \ldots, S_K$, and $\mathbf{\Sigma}_{1S_k}$ and $\mathbf{\Sigma}_{2S_k}$ be the covariance matrices of two types of tumors, respectively. The null hypotheses we are interested are

$$H_{0,\text{category}} : \mathbf{\Sigma}_{1S_k} = \mathbf{\Sigma}_{2S_k}, \quad k = 1, \ldots, K$$

where category $\in \{BP, CC, MF\}$ because we classify the gene sets into three different GO categories and shall test each GO category separately.

To control the computational costs, we first perform a pre-screening procedure following the same criteria as in Dudoit, Keles, and van der Laan (2008) by choosing those probes that satisfy (i) the fluorescence intensities greater than 100 (absolute scale) for at least 25% of the 79 cell samples; (ii) the interquartile range (IQR) of the fluorescence intensities for the 79 cell samples greater than 0.5 (log base 2 scale). The preliminary gene-filtering retains 2391 probes. After that we then identify those GO terms annotating at least 10 of the 2391 filtered probes, which gives us 1849 unique GO terms in BP category, 306 in CC and 324 in MF for further analysis. Table 4 and Figure 2 summarize the dimension of gene-sets contained in each category.

We first take a look at the performance of the CLX test and the LC test in the boxplots of Figure 3. It can be observed that test statistics have quite different magnitudes, indicating difficulty

Table 5. Gene-set testing results at the nominal level $\alpha = 0.05$.

GO category	Total number of gene-sets	Number of significant gene-sets				
		CLX	LC	Bonferroni	Weighted	Proposed
BP	1849	297	505	451	485	615
CC	306	52	111	96	80	116
MF	324	38	78	61	78	96

Table 6. Gene-set testing results with the FDR control at $\alpha = 0.05$.

GO category	Total number of gene-sets	Number of significant gene-sets				
		CLX	LC	Bonferroni	Weighted	Proposed
BP	1849	0	126	81	7	254
CC	306	0	55	24	0	68
MF	324	0	20	4	0	26

in the approach of weighted summation combination of the two statistics.

We then apply our proposed Fisher's method to test the hypothesis, together with comparisons to the CLX and LC tests. We also compare our test with the natural Bonferroni combination. The test outcomes are reported in Table 5, with nominal level $\alpha = 0.05$ for each test. Furthermore, in order to control the false discovery rate (FDR), we apply the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg 1995) to each GO category, and the results are listed in Table 6, with nominal level $\alpha = 0.05$ for every category.

As shown in Table 6, our proposed test identifies much more significant gene-sets than the other methods. The LC identifies a few while the Bonferroni test identifies fewer significant gene-sets than the LC test does. This illustrates that the Bonferroni test is relatively conservative, which is consistent with what we expect. Unfortunately, the CLX test fails to declare any significance after we control the FDR using BH procedure. This is possibly because the signals in the differences are not strong enough for the CLX test to detect.

Further, we investigate the potential importance of those gene-sets that are not declared significant by the CLX and LC tests but are identified by our proposed Fisher test. Taking the GO term "GO:0005905" as an example, it refers to the clathrin-coated pit which functions in the cellular component (CC) gene ontology category. Protein evidence by Ezkurdia et al. (2014) confirms that the clathrin-coated pit works with several protein-coding genes, such as CLTCL1, PICALM, etc., that are closely related to human cancers. We also take a deep look at "GO:0035259," the glucocorticoid receptor binding, in the molecular function (MF) gene ontology category. Many genes contribute to this gene-set, among them, we pay special attention to STAT3, a protein-coding gene which plays an important role in the immune system by transmitting signals for the maturation of immune system cells, especially T-cells and B-cells. Researchers have observed that STAT3 gene mutations are highly correlated with cancers, especially blood cancers (Hodge, Hurt, and Farrar 2005; Jerez et al. 2012; Haapaniemi et al. 2015; Milner et al. 2015). In a short summary, our proposed test incorporates the information from the CLX statistic, which successfully enhances the power over the LC test, even though the LC test itself may not declare any significance.

7. Conclusion

This article studies the fundamental problem of testing high-dimensional covariance matrices. Unlike the existing quadratic form statistics, maximum form statistics, and their weighted combination, we provide a new perspective to exploit the full potential of quadratic form statistics and maximum form statistics. We propose a scale-invariant and computationally efficient power-enhanced test based on Fisher's method to combine their respective p-values. Theoretically, after deriving their joint limiting null distribution, we prove that the proposed combination method retains the correct asymptotic size and boosts the power against more general alternatives. In particular, we derive a new Lyapunov-type bound for the joint distribution and prove the correct asymptotic size of the proposed method without requiring the Gaussian assumption. Numerically, we demonstrate the finite-sample properties in simulation studies and the practical relevance in an application to gene-set testing.

Supplementary Materials

The supplementary materials provide technical lemmas, the complete proofs of lemmas and theorems, and additional numerical results for non-Gaussian data.

Acknowledgments

We would like to thank the Coeditor, Associate Editor, and referees for their helpful comments and suggestions.

Funding

Xiufan Yu and Lingzhou Xue were partially supported by the National Science Foundation grants DMS-1811552 and DMS-1953189. Danning Li was partially supported by the National Natural Science Foundation of China grant 12101116.

ORCID

Lingzhou Xue http://orcid.org/0000-0002-8252-0637

References

Anderson, T. (2003), *An Introduction to Multivariate Statistical Analysis* (Vol. 3), New York: Wiley. [511,517]

Arcones, M. A., and Gine, E. (1993), "Limit Theorems for U-Processes," *The Annals of Probability*, 21, 1494–1542. [516]

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. [511]

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Series B, 57, 289–300. [523]

Bentkus, V. (2005), "A Lyapunov Type Bound in R^d," *Theory of Probability & Its Applications*, 49, 311–323. [512,517]

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. [512,513,515,516,517,518,519]

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, 31–41. [511,512]

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. [511]

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013), "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *The Annals of Statistics*, 41, 2786–2819. [512,517]

——— (2017), "Central Limit Theorems and Bootstrap in High Dimensions," *The Annals of Probability*, 45, 2309–2352. [512,517]

Chiaretti, S., Li, X., Gentleman, R., Vitale, A., Vignetti, M., Mandelli, F., Ritz, J., and Foa, R. (2004), "Gene Expression Profile of Adult T-Cell Acute Lymphocytic Leukemia Identifies Distinct Subsets of Patients with Different Response to Therapy and Survival," *Blood*, 103, 2771–2778. [521]

Dudoit, S., Keles, S., and van der Laan, M. J. (2008), "Multiple Tests of Association with Biological Annotation Metadata," *Probability and Statistics: Essays in Honor of David A. Freedman*, 2, 153–218. [522]

Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., and Tress, M. L. (2014), "Multiple Evidence Strands Suggest that there may be as few as 19,000 Human Protein-Coding Genes," *Human Molecular Genetics*, 23, 5866–5878. [523]

Fan, J., Liao, Y., and Yao, J. (2015), "Power Enhancement in High-Dimensional Cross-Sectional Tests," *Econometrica*, 83, 1497–1541. [512,514]

Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd. [512,514]

Goldfarb, D., and Iyengar, G. (2003), "Robust Portfolio Selection Problems," *Mathematics of Operations Research*, 28, 1–38. [511]

Haapaniemi, E. M., Kaustio, M., Rajala, H. L., van Adrichem, A. J., Kainulainen, L., Glumoff, V., Doffinger, R., Kuusanmäki, H., Heiskanen-Kosma, T., and Trotta, L. (2015), "Autoimmunity, Hypogammaglobulinemia, Lymphoproliferation, and Mycobacterial Disease in Patients with Activating Mutations in STAT3," *Blood*, 125, 639–648. [523]

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. [511,512,514,515]

Hedges, L. V., and Olkin, I. (2014), *Statistical Methods for Meta-Analysis*, New York: Academic Press. [514]

Hodge, D. R., Hurt, E. M., and Farrar, W. L. (2005), "The Role of IL-6 and STAT3 in Inflammation and Cancer," *European Journal of Cancer*, 41, 2502–2512. [523]

Jerez, A., Clemente, M. J., Makishima, H., Koskela, H., LeBlanc, F., Ng, K. P., Olson, T., Przychodzen, B., Afable, M., and Gomez-Segui, I. (2012), "STAT3 Mutations Unify the Pathogenesis of Chronic Lymphoproliferative Disorders of NK Cells and T-Cell Large Granular Lymphocyte Leukemia," *Blood*, 120, 3048–3057. [523]

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. [511]

Johnstone, I. M. (2008), "Multivariate Analysis and Jacobi Ensembles: Largest Eigenvalue, Tracy-Widom Limits and Rates of Convergence," *The Annals of Statistics*, 36, 2638–2716. [511]

Li, D., and Xue, L. (2015), "Joint Limiting Laws for High-Dimensional Independence Tests," arXiv preprint arXiv:1512.08819. [512,514,515]

Li, D., Xue, L., and Zou, H. (2018), "Applications of Peter Hall's Martingale Limit Theory to Estimating and Testing High Dimensional Covariance Matrices," *Statistica Sinica*, 28, 2657–2670. [512,514,515]

Li, J., and Chen, S. X. (2012), "Two Sample Tests for High-Dimensional Covariance Matrices," *The Annals of Statistics*, 40, 908–940. [511,513,514,515,516,517,518]

Littell, R. C., and Folks, J. L. (1971), "Asymptotic Optimality of Fisher's Method of Combining Independent Tests," *Journal of the American Statistical Association*, 66, 802–806. [514,518]

——— (1973), "Asymptotic Optimality of Fisher's Method of Combining Independent Tests II," *Journal of the American Statistical Association*, 68, 193–194. [514,518]

Liu, Y., Chen, S., Li, Z., Morrison, A. C., Boerwinkle, E., and Lin, X. (2019), "ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies," *The American Journal of Human Genetics*, 104, 410–421. [512]

Liu, Y., and Xie, J. (2020), "Cauchy Combination Test: A Powerful Test with Analytic *p*-value Calculation under Arbitrary Dependency Structures," *Journal of the American Statistical Association*, 115, 393–402. [514,516]

Milner, J. D., Vogel, T. P., Forbes, L., Ma, C. A., Stray-Pedersen, A., Niemela, J. E., Lyons, J. J., Engelhardt, K. R., Zhang, Y., and Topcagic, N. (2015), "Early-onset Lymphoproliferation and Autoimmunity Caused by Germline STAT3 Gain-of-Function Mutations," *Blood*, 125, 591–599. [523]

Perlman, M. D. (1980), "Unbiasedness of the Likelihood Ratio Tests for Equality of Several Covariance Matrices and Equality of Several Multivariate Normal Populations," *The Annals of Statistics*, 8, 247–263. [511]

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, 6535–6542. [511]

Shi, C., Song, R., Chen, Z., and Li, R. (2019), "Linear Hypothesis Testing for High Dimensional Generalized Linear Models," *The Annals of Statistics*, 47, 2671–2703. [512,516]

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. [511,519]

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., and Williams Jr, R. M. (1949), *The American Soldier: Adjustment During Army Life*, Princeton: Princeton University Press. [516]

Sugiura, N., and Nagao, H. (1968), "Unbiasedness of Some Test Criteria for the Equality of One or Two Covariance Matrices," *The Annals of Mathematical Statistics*, 39, 1686–1692. [511]

Tippett, L. H. C. (1931), *The Methods of Statistics: An Introduction Mainly for Experimentalists*, London: Williams & Norgate Ltd. [516]

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. [512,514]

Yu, X., Li, D., Xue, L., and Li, R. (2022), "Power-Enhanced Simultaneous Test of High-Dimensional Mean Vectors and Covariance Matrices with Application to Gene-Set Testing," *Journal of the American Statistical Association* . [514]

Yu, X., Yao, J., and Xue, L. (2019), "Innovated Power Enhancement for Testing Multi-Factor Asset Pricing Models," Available at SSRN 3809369. [512]