

This article was downloaded by: [University of Pennsylvania]

On: 29 April 2013, At: 08:44

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Journal of the American Statistical Association

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/uasa20>

### Two-Sample Covariance Matrix Testing and Support Recovery in High-Dimensional and Sparse Settings

Tony Cai<sup>a</sup>, Weidong Liu<sup>b</sup> & Yin Xia<sup>a</sup>

<sup>a</sup> Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, 19104

<sup>b</sup> Department of Mathematics, Institute of Natural Sciences and MOE-LSC, Shanghai Jiao Tong University, Shanghai, China

Accepted author version posted online: 02 Jan 2013. Published online: 15 Mar 2013.

To cite this article: Tony Cai, Weidong Liu & Yin Xia (2013): Two-Sample Covariance Matrix Testing and Support Recovery in High-Dimensional and Sparse Settings, Journal of the American Statistical Association, 108:501, 265-277

To link to this article: <http://dx.doi.org/10.1080/01621459.2012.758041>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Two-Sample Covariance Matrix Testing and Support Recovery in High-Dimensional and Sparse Settings

Tony CAI, Weidong LIU, and Yin XIA

In the high-dimensional setting, this article considers three interrelated problems: (a) testing the equality of two covariance matrices  $\Sigma_1$  and  $\Sigma_2$ ; (b) recovering the support of  $\Sigma_1 - \Sigma_2$ ; and (c) testing the equality of  $\Sigma_1$  and  $\Sigma_2$  row by row. We propose a new test for testing the hypothesis  $H_0: \Sigma_1 = \Sigma_2$  and investigate its theoretical and numerical properties. The limiting null distribution of the test statistic is derived and the power of the test is studied. The test is shown to enjoy certain optimality and to be especially powerful against sparse alternatives. The simulation results show that the test significantly outperforms the existing methods both in terms of size and power. Analysis of a prostate cancer dataset is carried out to demonstrate the application of the testing procedures. When the null hypothesis of equal covariance matrices is rejected, it is often of significant interest to further investigate how they differ from each other. Motivated by applications in genomics, we also consider recovering the support of  $\Sigma_1 - \Sigma_2$  and testing the equality of the two covariance matrices row by row. New procedures are introduced and their properties are studied. Applications to gene selection are also discussed. Supplementary materials for this article are available online.

KEY WORDS: Extreme value Type I distribution; Gene selection; Hypothesis testing; Sparsity.

## 1. INTRODUCTION

Testing the equality of two covariance matrices  $\Sigma_1$  and  $\Sigma_2$  is an important problem in multivariate analysis. Many statistical procedures including the classical Fisher's linear discriminant analysis rely on the fundamental assumption of equal covariance matrices. This testing problem has been well studied in the conventional low-dimensional setting. See, for example, Sugiura and Nagao (1968), Gupta and Giri (1973), Perlman (1980), Gupta and Tang (1984), O'Brien (1992), and Anderson (2003). In particular, the likelihood ratio test (LRT) is commonly used and enjoys certain optimality under regularity conditions.

Driven by a wide range of contemporary scientific applications, analysis of high-dimensional data is of significant current interest. In the high-dimensional setting, where the dimension can be much larger than the sample size, the conventional testing procedures such as the LRT either perform poorly or are not even well defined. Several tests for the equality of two large covariance matrices have been proposed. For example, Schott (2007) introduced a test based on the Frobenius norm of the difference of the two covariance matrices. Srivastava and Yanagihara (2010) constructed a test that relied on a measure of distance by  $\text{tr}(\Sigma_1^2)/(\text{tr}(\Sigma_1))^2 - \text{tr}(\Sigma_2^2)/(\text{tr}(\Sigma_2))^2$ . Both of these two tests are designed for the multivariate normal populations. Li and Chen (2012) proposed a test using a linear combination of three one-sample  $U$  statistics, which was also motivated by an unbiased estimator of the Frobenius norm of  $\Sigma_1 - \Sigma_2$ .

In many applications, such as gene selection in genomics, the covariance matrices of the two populations can be either equal

or quite similar in the sense that they only possibly differ in a small number of entries. In such a setting, under the alternative the difference of the two covariance matrices  $\Sigma_1 - \Sigma_2$  is sparse. The above-mentioned tests, which are all based on the Frobenius norm, are not powerful against such sparse alternatives. See Section 3.3 for further details.

The first goal of this article is to develop a test that is powerful against sparse alternatives and robust with respect to the population distributions. Let  $X$  and  $Y$  be two  $p$  variate random vectors with covariance matrices  $\Sigma_1$  and  $\Sigma_2$ , respectively. Let  $\{X_1, \dots, X_{n_1}\}$  be independent and identically distributed (iid) random samples from  $X$  and let  $\{Y_1, \dots, Y_{n_2}\}$  be iid random samples from  $Y$  that are independent of  $\{X_1, \dots, X_{n_1}\}$ . We wish to test the hypotheses

$$H_0: \Sigma_1 = \Sigma_2 \quad \text{versus} \quad H_1: \Sigma_1 \neq \Sigma_2 \quad (1)$$

based on the two samples. We are particularly interested in the high-dimensional setting where  $p$  can be much larger than  $n = \max(n_1, n_2)$ . In many applications, if the null hypothesis  $H_0: \Sigma_1 = \Sigma_2$  is rejected, it is often of significant interest to further investigate in which way the two covariance matrices differ from each other. Motivated by applications to gene selection, in this article we also consider two related problems, recovering the support of  $\Sigma_1 - \Sigma_2$  and testing the equality of the two covariance matrices row by row.

We propose a test for the hypotheses in Equation (1) based on the maximum of the standardized differences between the entries of the two-sample covariance matrices and investigate its theoretical and numerical properties. The limiting null distribution of the test statistic is derived. It is shown that the distribution of the test statistic converges to a Type I extreme value distribution under the null  $H_0$ . This fact implies that the proposed test has the prespecified significance level asymptotically. The power of the test is investigated. The theoretical analysis shows that the proposed test enjoys certain optimality

Tony Cai is Dorothy Silberberg Professor of Statistics, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: [tcai@wharton.upenn.edu](mailto:tcai@wharton.upenn.edu)). Weidong Liu is Professor, Department of Mathematics, Institute of Natural Sciences and MOE-LSC, Shanghai Jiao Tong University, Shanghai, China (E-mail: [liuweidong99@gmail.com](mailto:liuweidong99@gmail.com)). Yin Xia is Ph.D. student, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: [xiayin@wharton.upenn.edu](mailto:xiayin@wharton.upenn.edu)). The research of Tony Cai and Yin Xia was supported in part by NSF FRG grant DMS-0854973. Weidong Liu's research was supported by NSFC, Grant No. 11201298, the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, Foundation for the Author of National Excellent Doctoral Dissertation of PR China and the startup fund from Shanghai Jiao Tong University (SJTU).

against a large class of sparse alternatives in terms of the power. We show that it only requires one of the entries of  $\Sigma_1 - \Sigma_2$  having a magnitude more than  $C\sqrt{\log p/n}$  in order for the test to correctly reject the null hypothesis  $H_0$ . It is also shown that this lower bound  $C\sqrt{\log p/n}$  is rate optimal.

In addition to the theoretical properties, we also consider the numerical performance of the proposed testing procedure using both simulated and real datasets. The numerical results show that the new test significantly outperforms the existing methods both in terms of size and power. A prostate cancer dataset is used to illustrate our testing procedures for gene selection.

In addition to the global test of equal covariance matrices, we also consider recovery of the support of  $\Sigma_1 - \Sigma_2$  as well as testing  $\Sigma_1$  and  $\Sigma_2$  row by row. Support recovery can also be viewed as simultaneous testing of the equality of individual entries between the two covariance matrices. We introduce a procedure for support recovery based on the thresholding of the standardized differences between the entries of the two covariance matrices. It is shown that under certain conditions, the procedure recovers the true support of  $\Sigma_1 - \Sigma_2$  exactly with probability tending to 1. The procedure is also shown to be minimax rate optimal.

The problem of testing the equality of two covariance matrices row by row is motivated by applications in genomics. A commonly used approach in microarray analysis is to select “interesting” genes by applying multiple testing procedures on the two-sample  $t$  statistics. This approach has been successful in finding genes with significant changes in the mean expression levels between diseased and nondiseased populations. It has been noted recently that these mean-based methods lack the ability to discover genes that change their relationships with other genes, and new methods that are based on the change in the gene’s dependence structure are thus needed to identify these genes. See, for example, Ho et al. (2008), Hu et al. (2009), and Hu, Qiu, and Glazko (2010). In this article, we propose a procedure that simultaneously tests the  $p$  null hypotheses that the corresponding rows of the two covariance matrices are equal to each other. Asymptotic null distribution of the test statistics is derived and properties of the test are studied. It is shown that the procedure controls the family-wise error rate (FWER) at a prespecified level. Applications to gene selection are also considered.

The rest of this article is organized as follows. Section 2 introduces the procedure for testing the equality of two covariance matrices. Theoretical properties of the test are investigated in Section 3. After Section 3.1 in which basic definitions and assumptions are given, Section 3.2 develops the asymptotic null distribution of the test statistic and presents the optimality results for testing against sparse alternatives. Comparisons with other tests are given in Section 3.3. Section 4 considers support recovery of  $\Sigma_1 - \Sigma_2$  and testing the two covariance matrices row by row. Section 5 investigates the numerical performance of the proposed test by simulations and by an analysis of a prostate cancer dataset. Section 6 discusses our results and other related work. The proofs of the main results are given in Section 7.

## 2. THE TESTING PROCEDURE

Given two iid random samples  $\{X_1, \dots, X_{n_1}\}$  from a  $p$ -variate distribution with the covariance matrix  $\Sigma_1 = (\sigma_{ij1})_{p \times p}$

and  $\{Y_1, \dots, Y_{n_2}\}$  from a distribution with the covariance matrix  $\Sigma_2 = (\sigma_{ij2})_{p \times p}$ , define the sample covariance matrices

$$(\hat{\sigma}_{ij1})_{p \times p} := \hat{\Sigma}_1 = \frac{1}{n_1} \sum_{k=1}^{n_1} (X_k - \bar{X})(X_k - \bar{X})',$$

$$(\hat{\sigma}_{ij2})_{p \times p} := \hat{\Sigma}_2 = \frac{1}{n_2} \sum_{k=1}^{n_2} (Y_k - \bar{Y})(Y_k - \bar{Y})',$$

where  $\bar{X} = \frac{1}{n_1} \sum_{k=1}^{n_1} X_k$  and  $\bar{Y} = \frac{1}{n_2} \sum_{k=1}^{n_2} Y_k$ . The null hypothesis  $H_0 : \Sigma_1 = \Sigma_2$  is equivalent to  $H_0 : \max_{1 \leq i \leq j \leq p} |\sigma_{ij1} - \sigma_{ij2}| = 0$ . A natural approach to testing this hypothesis is to compare the sample covariances  $\hat{\sigma}_{ij1}$  and  $\hat{\sigma}_{ij2}$  and to base the test on the maximum differences. It is important to note that the sample covariances  $\hat{\sigma}_{ij1}$ ’s and  $\hat{\sigma}_{ij2}$ ’s are in general heteroscedastic and can possibly have a wide range of variability. It is thus necessary to first standardize  $\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2}$  before making a comparison among different entries.

To be more specific, define the variances  $\theta_{ij1} = \text{var}((X_i - \mu_{1i})(X_j - \mu_{1j}))$  and  $\theta_{ij2} = \text{var}((Y_i - \mu_{2i})(Y_j - \mu_{2j}))$ . Given the two samples  $\{X_1, \dots, X_{n_1}\}$  and  $\{Y_1, \dots, Y_{n_2}\}$ ,  $\theta_{ij1}$  and  $\theta_{ij2}$  can be, respectively, estimated by

$$\hat{\theta}_{ij1} = \frac{1}{n_1} \sum_{k=1}^{n_1} [(X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) - \hat{\sigma}_{ij1}]^2,$$

$$\bar{X}_i = \frac{1}{n_1} \sum_{k=1}^{n_1} X_{ki},$$

and

$$\hat{\theta}_{ij2} = \frac{1}{n_2} \sum_{k=1}^{n_2} [(Y_{ki} - \bar{Y}_i)(Y_{kj} - \bar{Y}_j) - \hat{\sigma}_{ij2}]^2,$$

$$\bar{Y}_i = \frac{1}{n_2} \sum_{k=1}^{n_2} Y_{ki}.$$

Such an estimator of the variance has been used in Cai and Liu (2011) in the context of adaptive estimation of a sparse covariance matrix. Given  $\hat{\theta}_{ij1}$  and  $\hat{\theta}_{ij2}$ , the variance of  $\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2}$  can then be estimated by  $\hat{\theta}_{ij1}/n_1 + \hat{\theta}_{ij2}/n_2$ .

Define the standardized statistics

$$M_{ij} =: \frac{(\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2})^2}{\hat{\theta}_{ij1}/n_1 + \hat{\theta}_{ij2}/n_2}, \quad 1 \leq i \leq j \leq p. \quad (2)$$

We consider the following test statistic for testing the hypothesis  $H_0 : \Sigma_1 = \Sigma_2$ ,

$$M_n =: \max_{1 \leq i \leq j \leq p} M_{ij} = \max_{1 \leq i \leq j \leq p} \frac{(\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2})^2}{\hat{\theta}_{ij1}/n_1 + \hat{\theta}_{ij2}/n_2}. \quad (3)$$

The asymptotic behavior of the test statistic  $M_n$  will be studied in detail in Section 3. Intuitively,  $M_{ij}$  are approximately square of standard normal variables under the null  $H_0$  and they are only “weakly dependent” under suitable conditions. The test statistic  $M_n$  is the maximum of  $p(p+1)/2$  such variables and so the value of  $M_n$  is close to  $2 \log p^2$  under  $H_0$ , based on the extreme values of normal random variables. More precisely, we show in Section 3 that, under certain regularity conditions,  $M_n - 4 \log p + \log \log p$  converges to a Type I extreme value distribution under the null hypothesis  $H_0$ . Based on this result, for a given significance level  $0 < \alpha < 1$ , we define the test  $\Phi_\alpha$

by

$$\Phi_\alpha = I(M_n \geq q_\alpha + 4 \log p - \log \log p), \quad (4)$$

where  $q_\alpha$  is the  $1 - \alpha$  quantile of the Type I extreme value distribution with the cumulative distribution function  $\exp(-\frac{1}{\sqrt{8\pi}} \exp(-\frac{x}{2}))$ , that is,

$$q_\alpha = -\log(8\pi) - 2 \log \log(1 - \alpha)^{-1}. \quad (5)$$

The hypothesis  $H_0 : \Sigma_1 = \Sigma_2$  is rejected whenever  $\Phi_\alpha = 1$ .

The test  $\Phi_\alpha$  is particularly well suited for testing against sparse alternatives. It is consistent if one of the entries of  $\Sigma_1 - \Sigma_2$  has a magnitude more than  $C\sqrt{\log p/n}$  for some constant  $C > 0$  and no other special structure of  $\Sigma_1 - \Sigma_2$  is required. It will be shown that the test  $\Phi_\alpha$  is an asymptotically  $\alpha$ -level test and enjoys certain optimality against sparse alternatives.

The standardized statistics  $M_{ij}$  are also useful for identifying the support of  $\Sigma_1 - \Sigma_2$ . That is, they can be used to estimate the positions at which the two covariance matrices differ from each other. This is of interest in many applications including gene selection. We show in Section 4.1 that under certain conditions, the support of  $\Sigma_1 - \Sigma_2$  can be correctly recovered by thresholding the  $M_{ij}$ .

### 3. THEORETICAL ANALYSIS OF SIZE AND POWER

We now turn to an analysis of the properties of the test  $\Phi_\alpha$  including the asymptotic size and power. The asymptotic size of the test is obtained by deriving the limiting distribution of the test statistic  $M_n$  under the null. We are particularly interested in the power of the test  $\Phi_\alpha$  under the sparse alternatives where  $\Sigma_1 - \Sigma_2$  is sparse in the sense that  $\Sigma_1 - \Sigma_2$  only contains a small number of nonzero entries.

#### 3.1 Definitions and Assumptions

We begin by introducing some basic definitions and technical assumptions. Throughout the article, denote  $\|\mathbf{a}\|_2 = \sqrt{\sum_{j=1}^p a_j^2}$  for the usual Euclidean norm of a vector  $\mathbf{a} = (a_1, \dots, a_p)^T \in \mathbb{R}^p$ . For a matrix  $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p \times q}$ , define the spectral norm  $\|\mathbf{A}\|_2 = \sup_{\|\mathbf{x}\|_2 \leq 1} \|\mathbf{A}\mathbf{x}\|_2$  and the Frobenius norm  $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$ . For two sequences of real numbers  $\{a_n\}$  and  $\{b_n\}$ , write  $a_n = O(b_n)$  if there exists a constant  $C$  such that  $|a_n| \leq C|b_n|$  holds for all sufficiently large  $n$ , write  $a_n = o(b_n)$  if  $\lim_{n \rightarrow \infty} a_n/b_n = 0$ , and write  $a_n \asymp b_n$  if there exist constants  $C > c > 0$  such that  $c|b_n| \leq |a_n| \leq C|b_n|$  for all sufficiently large  $n$ . The two-sample sizes are assumed to be comparable, that is,  $n_1 \asymp n_2$ . Let  $n = \max(n_1, n_2)$ .

Let  $\mathbf{R}_1 =: (\rho_{ij1})$  and  $\mathbf{R}_2 =: (\rho_{ij2})$  be the correlation matrices of  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively. For a fixed constant  $\alpha_0 > 0$ , define

$$s_j = s_j(\alpha_0) := \text{card}\{i : |\rho_{ij1}| \geq (\log p)^{-1-\alpha_0} \text{ or } |\rho_{ij2}| \geq (\log p)^{-1-\alpha_0}\}.$$

So  $s_j$  is the cardinality of the set of indices  $i$  such that either the  $i$ th variable of  $\mathbf{X}$  is highly correlated ( $> (\log p)^{-1-\alpha_0}$ ) with the  $j$ th variable of  $\mathbf{X}$  or the  $i$ th variable of  $\mathbf{Y}$  is highly correlated with the  $j$ th variable of  $\mathbf{Y}$ . For  $0 < r < 1$ , define the set

$$\Lambda(r) = \{1 \leq i \leq p : |\rho_{ij1}| > r \text{ or } |\rho_{ij2}| > r \text{ for some } j \neq i\}.$$

So  $\Lambda(r)$  is the set of indices  $i$  such that either the  $i$ th variable of  $\mathbf{X}$  is highly correlated ( $> r$ ) with some other variable of  $\mathbf{X}$  or the  $i$ th variable of  $\mathbf{Y}$  is highly correlated ( $> r$ ) with some other variable of  $\mathbf{Y}$ .

To obtain the asymptotic distribution of  $M_n$ , we assume that  $\max_{1 \leq j \leq p} s_j$  is not too large, and the set  $\Lambda(r)$  is also not too large for some  $r < 1$ .

**(C1).** Suppose that there exists a subset  $\Upsilon \subset \{1, 2, \dots, p\}$  with  $\text{Card}(\Upsilon) = o(p)$  and a constant  $\alpha_0 > 0$  such that for all  $\gamma > 0$ ,  $\max_{1 \leq j \leq p, j \notin \Upsilon} s_j(\alpha_0) = o(p^\gamma)$ . Moreover, there exist some constant  $r < 1$  and a sequence of numbers  $\Lambda_{p,r}$  such that  $\text{Card}(\Lambda(r)) \leq \Lambda_{p,r} = o(p)$ .

It is easy to check that if the eigenvalues of the correlation matrices  $\mathbf{R}_1$  and  $\mathbf{R}_2$  are bounded from above, then Condition (C1) is easily satisfied. In fact, if  $\lambda_{\max}(\mathbf{R}_1) \leq C_0$  and  $\lambda_{\max}(\mathbf{R}_2) \leq C_0$  for some constant  $C_0 > 0$ , then  $\max_j s_j \leq C(\log p)^{2+2\alpha_0}$ . The condition  $\max_{j, j \notin \Upsilon} s_j = o(p^\gamma)$  is quite weak. It allows the largest eigenvalues to be of order  $p/(\log p)^{2+2\alpha_0}$ . It is also easy to see that the condition  $\text{Card}(\Lambda(r)) = o(p)$  for some  $r < 1$  is trivially satisfied if all the correlations are bounded away from  $\pm 1$ , that is,

$$\max_{1 \leq i < j \leq p} |\rho_{ij1}| \leq r < 1 \text{ and } \max_{1 \leq i < j \leq p} |\rho_{ij2}| \leq r < 1. \quad (6)$$

We do not require the distributions of  $\mathbf{X}$  and  $\mathbf{Y}$  to be Gaussian. Instead we shall impose more general moment conditions.

**(C2).** *Sub-Gaussian-type tails:* Suppose that  $\log p = o(n^{1/5})$  and  $n_1 \asymp n_2$ . There exist some constants  $\eta > 0$  and  $K > 0$  such that

$$\begin{aligned} \mathbb{E} \exp(\eta(X_i - \mu_{i1})^2/\sigma_{ii1}) &\leq K, \\ \mathbb{E} \exp(\eta(Y_i - \mu_{i2})^2/\sigma_{ii2}) &\leq K \quad \text{for all } i. \end{aligned}$$

Furthermore, we assume that for some constants  $\tau_1 > 0$  and  $\tau_2 > 0$ ,

$$\min_{1 \leq i \leq j \leq p} \frac{\theta_{ij1}}{\sigma_{ii1}\sigma_{jj1}} \geq \tau_1 \quad \text{and} \quad \min_{1 \leq i \leq j \leq p} \frac{\theta_{ij2}}{\sigma_{ii2}\sigma_{jj2}} \geq \tau_2. \quad (7)$$

**(C2\*).** *Polynomial-type tails:* Suppose that for some  $\gamma_0, c_1 > 0$ ,  $p \leq c_1 n^{\gamma_0}$ ,  $n_1 \asymp n_2$  and for some  $\epsilon > 0$

$$\begin{aligned} \mathbb{E}|(X_i - \mu_{i1})/\sigma_{ii1}^{1/2}|^{4\gamma_0+4+\epsilon} &\leq K, \\ \mathbb{E}|(Y_i - \mu_{i2})/\sigma_{ii2}^{1/2}|^{4\gamma_0+4+\epsilon} &\leq K \quad \text{for all } i. \end{aligned}$$

Furthermore, we assume (7) holds.

In addition to the moment conditions, we also need the following technical condition for some of the results. It holds if  $\mathbf{X}$  and  $\mathbf{Y}$  have elliptically contoured distributions, see, for example, Anderson (2003, pp. 47–54).

**(C3).** Suppose that there exist  $\kappa_1, \kappa_2 \geq \frac{1}{3}$  such that for any  $i, j, k, l \in \{1, 2, \dots, p\}$ ,

$$\begin{aligned} \mathbb{E}(X_i - \mu_{i1})(X_j - \mu_{j1})(X_k - \mu_{k1})(X_l - \mu_{l1}) &= \kappa_1(\sigma_{ij1}\sigma_{kl1} + \sigma_{ik1}\sigma_{jl1} + \sigma_{il1}\sigma_{jk1}), \\ \mathbb{E}(Y_i - \mu_{i2})(Y_j - \mu_{j2})(Y_k - \mu_{k2})(Y_l - \mu_{l2}) &= \kappa_2(\sigma_{ij2}\sigma_{kl2} + \sigma_{ik2}\sigma_{jl2} + \sigma_{il2}\sigma_{jk2}). \end{aligned} \quad (8)$$

*Remark 1.* Conditions (C1) and (C3) are only needed for the limiting null distribution of the test statistic. (C3) holds for the elliptically contoured distributions (Anderson 2003) with  $\kappa_1 \equiv \frac{1}{3} \mathbb{E}(X_i - \mu_{i1})^4 / [\mathbb{E}(X_i - \mu_{i1})^2]^2$  and  $\kappa_2 \equiv \frac{1}{3} \mathbb{E}(Y_i - \mu_{i2})^4 / [\mathbb{E}(Y_i - \mu_{i2})^2]^2$ . Results on the power of the test  $\Phi_\alpha$  do not rely on either (C1) or (C3). Conditions (C2) and (C2\*) are moment conditions on  $X$  and  $Y$ . We only require either (C2) or (C2\*) to hold. This is much weaker than the Gaussian assumption required in the literature such as Schott (2007) and Srivastava and Yanagihara (2010). Condition (7) is satisfied with  $\tau_1 = \tau_2 = 1$ , if  $X \sim N(\mu_1, \Sigma_1)$  and  $Y \sim N(\mu_2, \Sigma_2)$ . In the non-Gaussian case, if (C3) and (6) hold, then (7) holds with  $\tau_1 = \kappa_1 - r^2/3$  and  $\tau_2 = \kappa_2 - r^2/3$ .

### 3.2 Limiting Null Distribution and Optimality

We are now ready to present the asymptotic null distribution of  $M_n$ . The following theorem shows that  $M_n - 4 \log p + \log \log p$  converges weakly under  $H_0$  to an extreme value distribution of Type I with the distribution function  $F(t) = \exp(-1/\sqrt{8\pi} e^{-t/2})$ .

*Theorem 1.* Suppose that (C1), (C2) (or (C2\*)), and (C3) hold. Then under  $H_0$ , for any  $t \in \mathbb{R}$ ,

$$\begin{aligned} &P(M_n - 4 \log p + \log \log p \leq t) \\ &\rightarrow \exp\left(-\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{t}{2}\right)\right), \end{aligned} \quad (9)$$

as  $n, p \rightarrow \infty$ . Furthermore, under  $H_0$ , the convergence in (9) is uniform for all  $X$  and  $Y$  satisfying (C1), (C2) (or (C2\*)), and (C3).

The limiting behavior of  $M_n$  is similar to that of the largest off-diagonal entry  $L_n$  of the sample correlation matrix in Jiang (2004), wherein the article derived the asymptotic distribution of  $L_n$  under the assumption that the components  $X_1, \dots, X_p$  are independent. Further extensions and improvements are given in Zhou (2007), Liu, Lin, and Shao (2008), and Cai and Jiang (2011, 2012). A key assumption in these articles is the independence between the components. The techniques in their proofs cannot be used to obtain (9), since dependent components are allowed in Theorem 1. The proof of (9) requires different techniques.

The limiting null distribution given in (9) shows that the test  $\Phi_\alpha$  defined in (4) is an asymptotically level  $\alpha$  test. That is,

$$P(\text{Type I error}) = P_{H_0}(\Phi_\alpha = 1) \rightarrow \alpha. \quad (10)$$

The limiting null distribution in Theorem 1 requires Conditions (C1) and (C3). However, without (C1) and (C3), the size of the test can still be effectively controlled.

*Proposition 1.* Under (C2) (or (C2\*)), for  $0 < \alpha < 1$ ,

$$P(\text{Type I error}) = P_{H_0}(\Phi_\alpha = 1) \leq -\log(1 - \alpha) + o(1). \quad (11)$$

Note that for commonly used significant level  $\alpha = 0.05$ ,  $-\log(1 - \alpha) = 0.05129$ , which is close to 0.05, and for  $\alpha = 0.01$ ,  $-\log(1 - \alpha) = 0.01005$ , which is even closer to the nominal level. It is easy to see that  $-\log(1 - \alpha) \approx \alpha$  for small  $\alpha$ . So, the test  $\Phi_\alpha$  effectively controls the probability of Type I error without (C1) and (C3).

We now turn to an analysis of the power of the test  $\Phi_\alpha$  given in (4). We shall define the following class of matrices:

$$\mathcal{U}(c) = \left\{ (\Sigma_1, \Sigma_2) : \max_{1 \leq i \leq j \leq p} \frac{|\sigma_{ij1} - \sigma_{ij2}|}{\sqrt{\theta_{ij1}/n_1 + \theta_{ij2}/n_2}} \geq c\sqrt{\log p} \right\}. \quad (12)$$

The next theorem shows that the null parameter set in which  $\Sigma_1 = \Sigma_2$  is asymptotically distinguishable from  $\mathcal{U}(4)$  by the test  $\Phi_\alpha$ . That is,  $H_0$  is rejected by  $\Phi_\alpha$  with overwhelming probability if  $(\Sigma_1, \Sigma_2) \in \mathcal{U}(4)$ .

*Theorem 2.* Suppose that (C2) or (C2\*) holds. Then as  $n, p \rightarrow \infty$ ,

$$\inf_{(\Sigma_1, \Sigma_2) \in \mathcal{U}(4)} P(\Phi_\alpha = 1) \rightarrow 1. \quad (13)$$

It can be seen from Theorem 2 that it only requires one of the entries of  $\Sigma_1 - \Sigma_2$  having a magnitude more than  $C\sqrt{\log p/n}$  in order for the test  $\Phi_\alpha$  to correctly reject  $H_0$ . This lower bound is rate optimal. Denote by  $\mathcal{P}$  the collection of distributions satisfying (C2) or (C2\*). Let  $\mathcal{T}_\alpha$  be the set of  $\alpha$ -level tests over  $\mathcal{P}$ , that is,  $P(T_\alpha = 1) \leq \alpha$  under  $H_0$  over all distributions in  $\mathcal{P}$  for any  $T_\alpha \in \mathcal{T}_\alpha$ .

*Theorem 3.* Suppose that (C2) or (C2\*) holds. Let  $\alpha, \beta > 0$  and  $\alpha + \beta < 1$ . Then there exists a constant  $c_0 > 0$  such that for all large  $n$  and  $p$ ,

$$\inf_{(\Sigma_1, \Sigma_2) \in \mathcal{U}(c_0)} \sup_{T_\alpha \in \mathcal{T}_\alpha} P(T_\alpha = 1) \leq 1 - \beta. \quad (14)$$

The difference between  $\Sigma_1$  and  $\Sigma_2$  is measured by  $\max_{1 \leq i \leq j \leq p} |\sigma_{ij1} - \sigma_{ij2}|$  in Theorem 3. Another measurement is based on the Frobenius norm  $\|\Sigma_1 - \Sigma_2\|_F$ . Suppose that  $\Sigma_1 - \Sigma_2$  is a sparse matrix with  $c_0(p)$  nonzero entries. That is,

$$c_0(p) = \sum_{i=1}^p \sum_{j=1}^p I\{\sigma_{ij1} - \sigma_{ij2} \neq 0\}.$$

We introduce the following class of matrices for  $\Sigma_1 - \Sigma_2$ :

$$\begin{aligned} \mathcal{V}(c) = \left\{ (\Sigma_1, \Sigma_2) : \max_i |\sigma_{i11}| \leq K, \max_i |\sigma_{i22}| \leq K, \right. \\ \left. \|\Sigma_1 - \Sigma_2\|_F^2 \geq cc_0(p) \frac{\log p}{n} \right\}. \end{aligned}$$

Note that on  $\mathcal{V}(c)$ , we have  $\max_{1 \leq i \leq j \leq p} |\sigma_{ij1} - \sigma_{ij2}| \geq \sqrt{c \log p/n}$ . Thus, for sufficiently large constant  $c$ ,

$$\inf_{(\Sigma_1, \Sigma_2) \in \mathcal{V}(c)} P(\Phi_\alpha = 1) \rightarrow 1$$

as  $n, p \rightarrow \infty$ . The following theorem shows that the lower bound for  $\|\Sigma_1 - \Sigma_2\|_F^2$  in  $\mathcal{V}(c)$  is rate optimal. That is, no  $\alpha$ -level test can reject  $H_0$  with overwhelming probability uniformly over  $\mathcal{V}(c_0)$  for some  $c_0 > 0$ .

*Theorem 4.* Suppose that (C2) or (C2\*) holds. Assume that  $c_0(p) \leq p^r$  for some  $0 < r < 1/2$ . Let  $\alpha, \beta > 0$  and  $\alpha + \beta < 1$ . There exists a constant  $c_0 > 0$  such that for all large  $n$  and  $p$ ,

$$\inf_{(\Sigma_1, \Sigma_2) \in \mathcal{V}(c_0)} \sup_{T_\alpha \in \mathcal{T}_\alpha} P(T_\alpha = 1) \leq 1 - \beta.$$

Note that for every  $c > 0$ , there exists some constant  $K(c) > 0$  such that for any  $0 < c_0 < K(c)$ ,  $\mathcal{V}(c) \subset \mathcal{U}(c_0)$ . Thus, Theorem 3 follows from Theorem 4 directly.

### 3.3 Comparison with Other Tests

We now compare the power of the test  $\Phi_\alpha$  with those of the tests in Li and Chen (2012), Srivastava and Yanagihara (2010), and Schott (2007) under the sparse alternatives. Define the following class of matrices:

$$\mathcal{S}(s_p, c_{n,p}) = \left\{ (\Sigma_1, \Sigma_2) : \sum_{j=1}^p \sum_{i=1}^p I\{\sigma_{ij1} - \sigma_{ij2} \neq 0\} \leq s_p, \right. \\ \left. K^{-1} \leq \min_{1 \leq i \leq p} |\sigma_{iil}| \leq \max_{1 \leq i \leq p} |\sigma_{iil}| \leq K \text{ for } l = 1, 2, \right. \\ \left. 4\sqrt{\log p} \leq \max_{1 \leq i \leq j \leq p} \frac{|\sigma_{ij1} - \sigma_{ij2}|}{\sqrt{\theta_{ij1}/n_1 + \theta_{ij2}/n_2}} \leq c_{n,p} \right\},$$

where we assume  $s_p = o(\min\{p/c_{n,p}^2, p/(\sqrt{nc_{n,p}})\})$ . Note that  $\mathcal{S}(s_p, c_{n,p}) \subset \mathcal{U}(4)$ . Hence, the power of the test  $\Phi_\alpha$  converges to one by Theorem 2.

We now turn to an analysis of the powers of the tests given in Li and Chen (2012), Schott (2007), and Srivastava and Yanagihara (2010). We first recall the conditions used in these three articles.

**(D1).** For any  $i, j, k, l \in \{1, 2\}$ ,

$$\text{tr}(\Sigma_k \Sigma_l) \rightarrow \infty \text{ and } \text{tr}\{(\Sigma_i \Sigma_j)(\Sigma_k \Sigma_l)\} = o\{\text{tr}(\Sigma_i \Sigma_j)\text{tr}(\Sigma_k \Sigma_l)\}.$$

**(D2).**  $\lim_{p \rightarrow \infty} \frac{1}{p} \text{tr}(\Sigma_1^i) = \gamma_{i1} \in (0, \infty)$  and  $\lim_{p \rightarrow \infty} \frac{1}{p} \text{tr}(\Sigma_2^i) = \gamma_{i2} \in (0, \infty)$ .

**(D3).**  $p/n_1 \rightarrow b_1$  and  $p/n_2 \rightarrow b_2$ , where  $b_1, b_2 \in (0, \infty)$ .

Let  $\beta_{LC}(\alpha)$ ,  $\beta_{SY}(\alpha)$ , and  $\beta_{Sc}(\alpha)$  be, respectively, the powers of the tests given in Li and Chen (2012), Srivastava and Yanagihara (2010), and Schott (2007) with the significance level being controlled at level  $\alpha$ . We shall make the comparisons under the normality condition since two of these tests require this assumption.

*Proposition 2.* Suppose that  $X \sim N(\mu_1, \Sigma_1)$  and  $Y \sim N(\mu_2, \Sigma_2)$ . Assume that  $(\Sigma_1, \Sigma_2) \in \mathcal{S}(s_p, c_{n,p})$ .

(i) For the proposed test  $\Phi_\alpha$  defined in (4),

$$\lim_{n,p \rightarrow \infty} \mathbb{P}(\Phi_\alpha = 1) = 1.$$

(ii) Suppose (D1) holds, then

$$\lim_{n,p \rightarrow \infty} \beta_{LC}(\alpha) = \alpha.$$

(iii) Suppose (D2) holds for  $i = 1, 2, 3$ , and 4, then

$$\lim_{n,p \rightarrow \infty} \beta_{SY}(\alpha) = \alpha.$$

Proposition 2 shows that under the class of sparse alternatives in  $\mathcal{S}(s_p, c_{n,p})$ , the tests given in Li and Chen (2012) and Srivastava and Yanagihara (2010) would suffer from trivial power while the proposed test  $\Phi_\alpha$  enjoys the full power. This fact is

also illustrated in some of the simulation results given in Section 5.

For the test given in Schott (2007), there is no result on the limiting distribution under the alternative. But from the theoretical analysis given in Schott (2007), under Condition (D2) for  $i = 1, \dots, 8$  and Condition (D3), one can see that to ensure the power tending to one, it is required that

$$\text{DIST}_{Sc}(\Sigma_1, \Sigma_2) = \|\Sigma_1 - \Sigma_2\|_F^2 \rightarrow \infty. \quad (15)$$

Note that, under (D2) for  $i = 1, \dots, 8$  and (D3), since  $s_p = o(p/c_{n,p}^2)$ , we have  $\text{DIST}_{Sc}(\Sigma_1, \Sigma_2) = o(1)$  for  $(\Sigma_1, \Sigma_2) \in \mathcal{S}(s_p, c_{n,p})$  and (15) is thus not satisfied.

Condition (C1) is generally much weaker than (D2) for  $i = 1, \dots, 4$ . For simplicity, let us assume the diagonal entries  $\sigma_{i1} = \sigma_{i2} = 1$ . Put  $\Upsilon$  in (C1) as

$$\Upsilon = \left\{ 1 \leq i \leq p : \sum_{j=1}^p \sigma_{ij1}^2 \geq \log p \text{ or } \sum_{j=1}^p \sigma_{ij2}^2 \geq \log p \right\}.$$

Since  $\text{tr}(\Sigma_i^2) = \|\Sigma_i\|_F^2 = O(p)$  for  $i = 1$  and 2, we have  $\text{Card}(\Upsilon) = O(p/\log p)$ . So  $\sum_{j=1}^p (\sigma_{ij1}^2 + \sigma_{ij2}^2) \leq 2 \log p$  uniformly for  $i \notin \Upsilon$ . This implies that, uniformly for  $i \notin \Upsilon$ ,  $\sigma_{i(k)1}^2 \leq k^{-1} \sum_{j=1}^p \sigma_{ij1}^2 = O(k^{-1} \log p)$  for  $1 \leq k \leq p$ , where  $\sigma_{i(1)1}^2 \geq \dots \geq \sigma_{i(p)1}^2$ . Similar inequality holds for  $\sigma_{ij2}$ . Hence, (D2) implies  $\max_{j \notin \Upsilon} s_j(\alpha) = O(p^\gamma)$  for any  $\alpha > 0$  and  $\gamma > 0$ , which is stronger than the condition on  $\max_{j \notin \Upsilon} s_j(\alpha_0)$  in (C1). Additionally, Condition (C1) allows for  $\text{tr}(\Sigma_i^2) \asymp (p^2/(\log p)^{2+2\alpha_0})$ .

## 4. SUPPORT RECOVERY OF $\Sigma_1 - \Sigma_2$ AND APPLICATION TO GENE SELECTION

We have focused on testing the equality of two covariance matrices  $\Sigma_1$  and  $\Sigma_2$  in Sections 2 and 3. As mentioned in the introduction, if the null hypothesis  $H_0 : \Sigma_1 = \Sigma_2$  is rejected, further exploring in which ways they differ is also of significant interest in practice. Motivated by applications in gene selection, we consider in this section two related problems, one is recovering the support of  $\Sigma_1 - \Sigma_2$  and another is identifying the rows on which the two covariance matrices differ from each other.

### 4.1 Support Recovery of $\Sigma_1 - \Sigma_2$

The goal of support recovery is to find the positions at which the two covariance matrices differ from each other. The problem can also be viewed as simultaneous testing of equality of individual entries between the two covariance matrices. Denote the support of  $\Sigma_1 - \Sigma_2$  by

$$\Psi = \Psi(\Sigma_1, \Sigma_2) = \{(i, j) : \sigma_{ij1} \neq \sigma_{ij2}\}. \quad (16)$$

In certain applications, the variances along the diagonal play a more important role than the covariances. For example, in a differential variability analysis of gene expression, Ho et al. (2008) proposed to select genes based on the differences in the variances. In this section, we shall treat the variances along the diagonal differently from the off-diagonal covariances for support recovery. Since there are  $p$  diagonal elements,  $M_{ii}$  along the diagonal are thresholded at  $2 \log p$  based on the extreme values of normal variables. The off-diagonal entries  $M_{ij}$  with

$i \neq j$  are thresholded at a different level. More specifically, set  $\hat{\Psi}(\tau) = \{(i, i) : M_{ii} \geq 2 \log p\} \cup \{(i, j) : M_{ij} \geq \tau \log p, i \neq j\}$ , (17)

where  $M_{ij}$  are defined in (2) and  $\tau$  is the threshold constant for the off-diagonal entries.

The following theorem shows that with  $\tau = 4$  the estimator  $\hat{\Psi}(4)$  recovers the support  $\Psi$  exactly with probability tending to 1 when the magnitudes of nonzero entries are above certain thresholds. Define

$$\mathcal{W}_0(c) = \left\{ (\Sigma_1, \Sigma_2) : \min_{(i,j) \in \Psi} \frac{|\sigma_{ij1} - \sigma_{ij2}|}{\sqrt{\theta_{ij1}/n_1 + \theta_{ij2}/n_2}} \geq c\sqrt{\log p} \right\}.$$

We have the following result.

*Theorem 5.* Suppose that (C2) or (C2\*) holds. Then as  $n, p \rightarrow \infty$ ,

$$\inf_{(\Sigma_1, \Sigma_2) \in \mathcal{W}_0(4)} P(\hat{\Psi}(4) = \Psi) \rightarrow 1.$$

The choice of the thresholding constant  $\tau = 4$  is optimal for the exact recovery of the support  $\Psi$ . Consider the class of  $s_0(p)$ -sparse matrices,

$$\mathcal{S}_0 = \left\{ A = (a_{ij})_{p \times p} : \max_i \sum_{j=1}^p I\{a_{ij} \neq 0\} \leq s_0(p) \right\}.$$

We show in the following theorem that, for any threshold constant  $\tau < 4$ , the probability of recovering the support exactly tends to 0 for all  $\Sigma_1 - \Sigma_2 \in \mathcal{S}_0$ . This is mainly because the threshold level  $\tau \log p$  is too small to ensure that the zero entries of  $\Sigma_1 - \Sigma_2$  are estimated by zero. We assume that

(C1\*). There exists a subset  $\Theta \subset \{1, 2, \dots, p\}$  with  $\text{Card}(\Theta) \geq p^\gamma$  for all  $0 < \gamma < 1$  such that  $|\rho_{ij1}| \leq C(\log p)^{-1-\alpha_0}$  and  $|\rho_{ij2}| \leq C(\log p)^{-1-\alpha_0}$  for all  $i \neq j \in \Theta$  and some  $\alpha_0 > 0$ .

It is easy to see that (C1\*) is much weaker than (C1) because it allows  $p - p^\gamma$  variables to be arbitrarily correlated.

*Theorem 6.* Suppose that (C1\*), (C2) (or (C2\*)), and (C3) hold. Let  $0 < \tau < 4$ . If  $s_0(p) = O(p^{1-\tau_1})$  with some  $\tau/4 < \tau_1 < 1$ , then as  $n, p \rightarrow \infty$ ,

$$\sup_{\Sigma_1 - \Sigma_2 \in \mathcal{S}_0} P(\hat{\Psi}(\tau) = \Psi) \rightarrow 0.$$

In addition, the minimum separation of the magnitudes of the nonzero entries must be at least of order  $\sqrt{\log p/n}$  to enable the exact recovery of the support. Consider, for example,  $s_0(p) = 1$  in Theorem 4. Then Theorem 4 indicates that no test can distinguish  $H_0$  and  $H_1$  uniformly on the space  $\mathcal{V}(c_0)$  with probability tending to one. It is easy to see that  $\mathcal{V}(c_0) \subseteq \mathcal{W}_0(c_1)$  for some  $c_1 > 0$ . Hence, there is no estimate that can recover the support  $\Psi$  exactly uniformly on the space  $\mathcal{W}_0(c_1)$  with probability tending to one.

We have so far focused on the exact recovery of the support of  $\Sigma_1 - \Sigma_2$  with probability tending to 1. It is sometimes desirable to recover the support under a less stringent criterion such as the FWER, defined by  $\text{FWER} = P(V \geq 1)$ , where  $V$  is the number of false discoveries. The goal of using the threshold level  $4 \log p$  is to ensure  $\text{FWER} \rightarrow 0$ . To control the FWER at a prespecified

level  $\alpha$  for some  $0 < \alpha < 1$ , a different threshold level is needed. For this purpose, we shall set the threshold level at  $4 \log p - \log \log p + q_\alpha$  where  $q_\alpha$  is given in (5). Define

$$\hat{\Psi}^* = \{(i, i) : M_{ii} \geq 2 \log p\} \cup \{(i, j) : M_{ij} \geq 4 \log p - \log \log p + q_\alpha, i \neq j\}.$$

We have the following result.

*Proposition 3.* Suppose that (C1), (C2) (or (C2\*)), and (C3) hold. Under  $\Sigma_1 - \Sigma_2 \in \mathcal{S}_0 \cap \mathcal{W}_0(4)$  with  $s_0(p) = o(p)$ , we have as  $n, p \rightarrow \infty$ ,

$$P(\hat{\Psi}^* \neq \Psi) \rightarrow \alpha.$$

Proposition 3 assumes Conditions (C1) and (C3). As in Proposition 1, it can be shown that  $P(\hat{\Psi}^* \neq \Psi) \leq -\log(1 - \alpha) + o(1)$  under Condition (C2) (or (C2\*)) only, without assuming (C1) or (C3).

## 4.2 Testing Rows of Two Covariance Matrices

As discussed in the introduction, the standard methods for gene selection are based on the comparisons of the means and thus lack the ability to select genes that change their relationships with other genes. It is of significant practical interest to develop methods for gene selection, which capture the changes in the gene's dependence structure.

Several methods have been proposed in the literature. It was noted in Ho et al. (2008) that the changes of variances are biologically interesting and are associated with changes in coexpression patterns in different biological states. Ho et al. (2008) proposed to test  $H_{0i} : \sigma_{ii1} = \sigma_{ii2}$  and select the  $i$ th gene if  $H_{0i}$  is rejected. Hu et al. (2009) and Hu, Qiu, and Glazko (2010) introduced methods that are based on simultaneous testing of the equality of the joint distributions of each row between two-sample correlation matrices/covariance distance matrices.

The covariance provides a natural measure of the association between two genes and it can also reflect the changes of variances. Motivated by these applications, in this section we consider testing the equality of two covariance matrices row by row. Let  $\sigma_{i,1}$  and  $\sigma_{i,2}$  be the  $i$ th row of  $\Sigma_1$  and  $\Sigma_2$ , respectively. We consider testing simultaneously the hypotheses

$$H_{0i} : \sigma_{i,1} = \sigma_{i,2}, \quad 1 \leq i \leq p.$$

We shall use the FWER to measure the Type I errors for the  $p$  tests. The support estimate  $\hat{\Psi}(4)$  defined in (17) can be used to test the hypotheses  $H_{0i}$  by rejecting  $H_{0i}$  if the  $i$ th row of  $\hat{\Psi}(4)$  is nonzero. Suppose that (C1) and (C2) (or (C2\*)) hold. Then it can be shown easily that for this test, the  $\text{FWER} \rightarrow 0$ .

A different test is needed to simultaneously test the hypotheses  $H_{0i}$ ,  $1 \leq i \leq p$ , at a prespecified FWER level  $\alpha$  for some  $0 < \alpha < 1$ . Define

$$M_i = \max_{1 \leq j \leq p, j \neq i} M_{ij}.$$

The null distribution of  $M_i$  can be derived similarly under the same conditions as in Theorem 1.

*Theorem 7.* Suppose the null hypothesis  $H_{0i}$  holds. Then under the conditions in Theorem 1, for any given  $x \in \mathbb{R}$ ,

as  $n, p \rightarrow \infty$ ,

$$P(M_i - 2 \log p + \log \log p \leq x) \rightarrow \exp\left(-\frac{1}{\sqrt{\pi}} \exp\left(-\frac{x}{2}\right)\right). \quad (18)$$

Based on the limiting null distribution of  $M_i$  given in (18), we introduce the following test for testing a single hypothesis  $H_{0i}$ ,

$$\Phi_{i,\alpha} = I(M_i \geq \alpha_p \text{ or } M_{ii} \geq 2 \log p) \quad (19)$$

with  $\alpha_p = 4 \log p - \log \log p + q_\alpha$ , where  $q_\alpha$  is given in (5).

$H_{0i}$  is rejected and the  $i$ th gene is selected if  $\Phi_{i,\alpha} = 1$ . It can be shown that the tests  $\Phi_{i,\alpha}$ ,  $1 \leq i \leq p$  together controls the overall FWER at level  $\alpha$  asymptotically.

*Theorem 8.* Let  $F = \{i : \sigma_{i,1} \neq \sigma_{i,2}, 1 \leq i \leq p\}$  and suppose  $\text{Card}(F) = o(p)$ . Then under the conditions in Theorem 1,

$$\text{FWER} = P\left(\max_{i \in F^c} \Phi_{i,\alpha} = 1\right) \rightarrow \alpha.$$

Theorem 8 requires Conditions (C1) and (C3). However, similar to Proposition 1, it can be shown that the FWER can still be effectively controlled without (C1) and (C3) in the sense that  $\text{FWER} \leq -\log(1 - \alpha) + o(1)$  under Condition (C2) (or (C2\*)) only.

## 5. NUMERICAL RESULTS

In this section, we turn to the numerical performance of the proposed methods. The goal is to first investigate the numerical performance of the test  $\Phi_\alpha$  and the support recovery procedure through simulation studies and then illustrate our methods by applying them to the analysis of a prostate cancer dataset. The test  $\Phi_\alpha$  is compared with four other tests, the LRT, the tests given in Schott (2007) and Li and Chen (2012), which are both based on an unbiased estimator of  $\|\Sigma_1 - \Sigma_2\|_F^2$ , and the test proposed in Srivastava and Yanagihara (2010), which is based on a measure of distance by  $\text{tr}(\Sigma_1^2)/(\text{tr}(\Sigma_1))^2 - \text{tr}(\Sigma_2^2)/(\text{tr}(\Sigma_2))^2$ . The LRT is not well defined when  $p > n$ , so the comparison with the LRT is made only for the case  $p \leq n$ .

We first introduce the matrix models used in the simulations. Let  $\mathbf{D} = (d_{ij})$  be a diagonal matrix with diagonal elements  $d_{ii} = \text{Unif}(0.5, 2.5)$  for  $i = 1, \dots, p$ . Denote by  $\lambda_{\min}(\mathbf{A})$  the minimum eigenvalue of a symmetric matrix  $\mathbf{A}$ . The following four models under the null,  $\Sigma_1 = \Sigma_2 = \Sigma^{(i)}$ ,  $i = 1, 2, 3$  and 4, are used to study the size of the tests.

- Model 1:  $\Sigma^{*(1)} = (\sigma_{ij}^{*(1)})$ , where  $\sigma_{ii}^{*(1)} = 1$ ,  $\sigma_{ij}^{*(1)} = 0.5$  for  $5(k-1) + 1 \leq i \neq j \leq 5k$ , where  $k = 1, \dots, \lfloor p/5 \rfloor$  and  $\sigma_{ij}^{*(1)} = 0$  otherwise.  $\Sigma^{(1)} = \mathbf{D}^{1/2} \Sigma^{*(1)} \mathbf{D}^{1/2}$ .
- Model 2:  $\Sigma^{*(2)} = (\sigma_{ij}^{*(2)})$ , where  $\sigma_{ij}^{*(2)} = 0.5^{|i-j|}$  for  $1 \leq i, j \leq p$ .  $\Sigma^{(2)} = \mathbf{D}^{1/2} \Sigma^{*(2)} \mathbf{D}^{1/2}$ .
- Model 3:  $\Sigma^{*(3)} = (\sigma_{ij}^{*(3)})$ , where  $\sigma_{ii}^{*(3)} = 1$ ,  $\sigma_{ij}^{*(3)} = 0.5 * \text{Bernoulli}(1, 0.05)$  for  $i < j$  and  $\sigma_{ji}^{*(3)} = \sigma_{ij}^{*(3)}$ .  $\Sigma^{(3)} = \mathbf{D}^{1/2} (\Sigma^{*(3)} + \delta \mathbf{I}) / (1 + \delta) \mathbf{D}^{1/2}$  with  $\delta = |\lambda_{\min}(\Sigma^{*(3)})| + 0.05$ .
- Model 4:  $\Sigma^{(4)} = \mathbf{O} \Delta \mathbf{O}$ , where  $\mathbf{O} = \text{diag}(\omega_1, \dots, \omega_p)$  and  $\omega_1, \dots, \omega_p \stackrel{\text{iid}}{\sim} \text{Unif}(1, 5)$  and  $\Delta = (a_{ij})$  and  $a_{ij} =$

$(-1)^{i+j} 0.4^{|i-j|^{1/10}}$ . This model was used in Srivastava and Yanagihara (2010).

To evaluate the power of the tests, let  $\mathbf{U} = (u_{kl})$  be a matrix with eight random nonzero entries. The locations of four nonzero entries are selected randomly from the upper triangle of  $\mathbf{U}$ , each with a magnitude generated from  $\text{Unif}(0, 4) * \max_{1 \leq j \leq p} \sigma_{jj}$ . The other four nonzero entries in the lower triangle are determined by symmetry. We use the following four pairs of covariance matrices  $(\Sigma_1^{(i)}, \Sigma_2^{(i)})$ ,  $i = 1, 2, 3$ , and 4, to compare the power of the tests, where  $\Sigma_1^{(i)} = \Sigma^{(i)} + \delta \mathbf{I}$  and  $\Sigma_2^{(i)} = \Sigma^{(i)} + \mathbf{U} + \delta \mathbf{I}$ , with  $\delta = |\min\{\lambda_{\min}(\Sigma^{(i)} + \mathbf{U}), \lambda_{\min}(\Sigma^{(i)})\}| + 0.05$ .

The sample sizes are taken to be  $n_1 = n_2 = n$  with  $n = 60$  and 100, while the dimension  $p$  varies over the values 50, 100, 200, 400, and 800. For each model, data are generated from multivariate normal distributions with mean zero and covariance matrices  $\Sigma_1$  and  $\Sigma_2$ . The nominal significant level for all the tests is set at  $\alpha = 0.05$ . The actual sizes and powers in percents for the four models, reported in Table 1, are estimated from 5000 replications.

It can be seen from Table 1 that the estimated sizes of our test  $\Phi_\alpha$  are close to the nominal level 0.05 in all the cases. This reflects the fact that the null distribution of the test statistic  $M_n$  is well approximated by its asymptotic distribution. For Models 1–3, the estimated sizes of the tests in Schott (2007) and Li and Chen (2012) are also close to 0.05. But both tests suffer from the size distortion for Model 4, the actual sizes are around 0.10 for both tests. The LRT has serious size distortion (all is equal to 1). Srivastava and Yanagihara (2010)'s test has actual sizes close to the nominal significance level in all the cases.

The power results in Table 1 show that the proposed test has much higher power than the other tests in all settings. The number of nonzero off-diagonal entries of  $\Sigma_1 - \Sigma_2$  does not change when  $p$  grows, so the estimated powers of all tests tend to decrease when the dimension  $p$  increases. It can be seen in Table 1 that the powers of Schott (2007), Li and Chen (2012), and Srivastava and Yanagihara (2010)'s tests decrease extremely fast as  $p$  grows. However, the power of the proposed test  $\Phi_\alpha$  remains high even when  $p = 800$ , especially in the case of  $n = 100$ . Overall, for the sparse models, the new test significantly outperforms all the other three tests.

We also carried out simulations for non-Gaussian distributions including Gamma distribution,  $t$ -distribution, and normal distribution contaminated with exponential distribution. Similar phenomena as those in the Gaussian case are observed. For reasons of space, these simulation results are given in the supplementary material (Cai, Liu, and Xia 2012).

### 5.1 Support Recovery

We now consider the simulation results on recovering the support of  $\Sigma_1 - \Sigma_2$  in the first three models with  $\mathbf{D} = \mathbf{I}$  and the fourth model with  $\mathbf{O} = \mathbf{I}$  under the normal distribution. For  $i = 1, 2, 3$ , and 4, let  $\mathbf{U}^{(i)}$  be a matrix with 50 random nonzero entries, each with a magnitude of 2 and let  $\Sigma_1^{(i)} = (\Sigma^{(i)} + \delta \mathbf{I}) / (1 + \delta)$  and  $\Sigma_2^{(i)} = (\Sigma^{(i)} + \mathbf{U}^{(i)} + \delta \mathbf{I}) / (1 + \delta)$  with  $\delta = |\min\{\lambda_{\min}(\Sigma^{(i)} + \mathbf{U}^{(i)}), \lambda_{\min}(\Sigma^{(i)})\}| + 0.05$ . After normalization, the nonzero elements of  $\Sigma_2^{(i)} - \Sigma_1^{(i)}$  have



Table 1.  $N(0, 1)$  random variables. Models 1–4 empirical sizes and powers in percents.  $\alpha = 0.05$ .  $n = 60$  and  $100$ . 5000 replications

$n$	$p$	Model 1					Model 2					Model 3					Model 4				
		50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
Empirical size																					
60	$\Phi_\alpha$	5.0	4.6	5.0	6.1	6.1	5.5	5.4	5.0	5.3	6.0	5.5	5.3	5.6	5.9	5.9	4.5	4.5	4.6	4.6	5.2
	Likelihood	100.0	100.0	NA	NA	NA	100.0	100.0	NA	NA	NA	100.0	100.0	NA	NA	NA	100.0	100.0	NA	NA	NA
	Schott	7.1	5.7	5.3	5.6	5.6	6.6	6.1	5.4	4.8	4.6	6.4	5.5	5.3	5.4	5.2	10.2	10.8	10.3	9.0	10.3
	Li-Chen	7.1	5.8	5.3	5.5	5.6	6.6	6.0	5.2	4.8	4.9	6.3	5.2	5.3	5.4	5.1	10.0	10.9	10.2	9.1	10.4
	Srivastava	5.6	5.4	4.7	4.8	3.8	5.2	4.7	5.1	4.9	3.7	5.2	5.0	4.7	4.1	4.0	5.1	5.9	5.2	4.4	5.3
100	$\Phi_\alpha$	4.8	4.1	4.4	4.9	4.8	4.5	4.2	4.3	5.0	4.9	4.6	5.1	4.5	4.5	4.9	4.2	3.9	3.7	4.0	4.2
	Likelihood	100.0	100.0	100.0	NA	NA	100.0	100.0	100.0	NA	NA	100.0	100.0	100.0	NA	NA	100.0	100.0	100.0	NA	NA
	Schott	6.4	5.3	5.0	5.2	4.8	5.5	5.0	4.9	5.0	5.2	5.0	5.3	5.2	5.2	4.8	9.2	9.2	9.2	9.9	9.7
	Li-Chen	6.5	5.3	5.1	5.1	4.7	5.7	5.0	5.1	5.2	5.2	4.9	5.1	5.0	5.1	4.9	9.1	9.4	9.2	9.9	9.6
	Srivastava	5.2	5.5	5.0	5.0	4.1	4.9	5.3	5.0	5.4	4.5	4.7	4.9	5.0	4.7	4.7	5.6	5.7	4.8	5.0	4.7
Empirical power																					
60	$\Phi_\alpha$	87.9	74.1	47.6	40.4	38.8	86.7	56.2	60.9	46.3	25.8	90.3	77.6	49.7	42.5	40.1	81.8	72.1	48.2	40.5	32.7
	Schott	31.8	14.2	8.3	6.7	6.2	34.5	9.2	8.9	6.6	5.2	35.0	15.5	8.3	6.7	6.4	28.4	13.5	8.2	7.1	7.2
	Li-Chen	31.2	13.9	8.1	6.4	5.9	33.7	8.9	8.7	6.7	5.4	34.0	15.4	8.3	6.7	6.4	27.8	13.1	7.9	6.9	7.0
	Srivastava	28.7	11.3	5.9	4.6	3.9	36.7	7.8	6.7	5.1	3.7	37.6	14.7	6.2	4.6	4.1	15.5	6.2	5.4	4.3	4.7
100	$\Phi_\alpha$	99.9	99.4	94.1	91.6	93.0	99.9	94.6	98.1	95.9	79.3	100.0	99.7	95.0	92.9	93.7	99.6	99.2	93.9	92.5	89.4
	Schott	57.4	24.7	10.0	7.9	7.0	63.0	14.1	12.9	9.2	5.8	63.6	27.3	10.3	7.9	7.0	51.3	21.9	9.4	7.9	6.5
	Li-Chen	57.1	24.3	9.8	7.6	7.1	62.5	14.1	12.8	8.9	5.4	63.2	27.1	10.2	8.0	6.9	50.4	21.4	9.4	7.9	6.2
	Srivastava	58.1	28.3	8.9	5.9	5.4	73.8	12.0	10.9	6.7	4.6	75.0	32.0	9.1	6.2	5.2	24.9	8.1	5.5	4.7	5.0

magnitude between 0.74 and 0.86 for  $i = 1, 2, 3$ , and 4 in our generated models.

The accuracy of the support recovery is evaluated by the similarity measure  $s(\hat{\Psi}, \Psi)$  defined by

$$s(\hat{\Psi}, \Psi) = \frac{|\hat{\Psi} \cap \Psi|}{\sqrt{|\hat{\Psi}| |\Psi|}}$$

where  $\hat{\Psi} = \hat{\Psi}(4)$  and  $\Psi$  is the support of  $\Sigma_1^{(i)} - \Sigma_2^{(i)}$  and  $|\cdot|$  denotes the cardinality. Note that  $0 \leq s(\hat{\Psi}, \Psi) \leq 1$  with  $s(\hat{\Psi}, \Psi) = 0$  indicating disjointness and  $s(\hat{\Psi}, \Psi) = 1$  indicating exact recovery. We summarize the average (standard deviation) values of  $s(\hat{\Psi}, \Psi)$  in percents for all models with  $n = 100$  in Table 2 based on 100 replications. The values are close to one, and hence the supports are well recovered by our procedure.

To better illustrate the elementwise recovery performance, heat maps of the nonzeros identified out of 100 replications for  $p = 50$  and  $100$  are shown in Figures 1 and 2. These heat maps suggest that, in all models, the estimated support by  $\hat{\Psi}(4)$  is close to the true support.

### 5.2 Real Data Analysis

We now illustrate our methods by applying them to the analysis of a prostate cancer dataset (Singh et al. 2002), which is avail-

able at <http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi>. The dataset consists of two classes of gene expression data that came from 52 prostate tumor patients and 50 prostate normal patients. This dataset has been analyzed in several articles on classification in which the two covariance matrices are assumed to be equal; see, for example, Xu, Brock, and Parrish (2009). The equality of the two covariance matrices is an important assumption for the validity of these classification methods. It is thus interesting to test whether such an assumption is valid.

There are a total of 12,600 genes. To control the computational costs, only the 5000 genes with the largest absolute values of the  $t$  statistics are used. Let  $\Sigma_1$  and  $\Sigma_2$  be, respectively, the covariance matrices of these 5000 genes in tumor and normal samples. We apply the test  $\Phi_\alpha$  defined in (4) to test the hypotheses  $H_0 : \Sigma_1 = \Sigma_2$  versus  $H_1 : \Sigma_1 \neq \Sigma_2$ . Based on the asymptotic distribution of the test statistic  $M_n$ , the  $p$ -value is calculated to be 0.0058 and the null hypothesis  $H_0 : \Sigma_1 = \Sigma_2$  is thus rejected at commonly used significant levels such as  $\alpha = 0.05$  or  $\alpha = 0.01$ . Based on this test result, it is therefore not reasonable to assume  $\Sigma_1 = \Sigma_2$  in applying a classifier to this dataset.

We then apply three methods to select genes with changes in variances/covariances between the two classes. The first is the differential variability analysis (Ho et al. 2008), which chooses the genes with different variances between two classes. In our procedure, the  $i$ th gene is selected if  $M_{ii} \geq 2 \log p$ . As a result, 21 genes are selected. The second and third methods are based on the differential covariance analysis, which is similar to the differential covariance distance vector analysis in Hu, Qiu, and Glazko (2010), but replacing the covariance distance matrix in their article by the covariance matrix. The second method selects the  $i$ th gene, if the  $i$ th row of  $\hat{\Psi}(4)$  is nonzero. This leads to the selection of 43 genes. The third method, which tests the covariance matrices row by row and is defined in (19), controls the FWER at  $\alpha = 0.1$ , and is able to find 52 genes. As

Table 2. Average (standard deviation) of  $s(\hat{\Psi}, \Psi)$  in percents

	Model 1	Model 2	Model 3	Model 4
$p = 50$	97.4 (2.5)	89.8 (4.3)	91.3 (4.1)	93.0 (3.8)
$p = 100$	95.3 (3.1)	86.8 (5.1)	86.2 (5.3)	89.7 (4.4)
$p = 200$	78.8 (5.6)	80.1 (5.9)	82.4 (5.8)	84.0 (6.4)
$p = 400$	75.6 (6.5)	73.2 (7.7)	79.4 (6.1)	77.8 (7.0)
$p = 800$	64.0 (7.5)	61.8 (8.8)	72.9 (6.1)	68.1 (7.6)

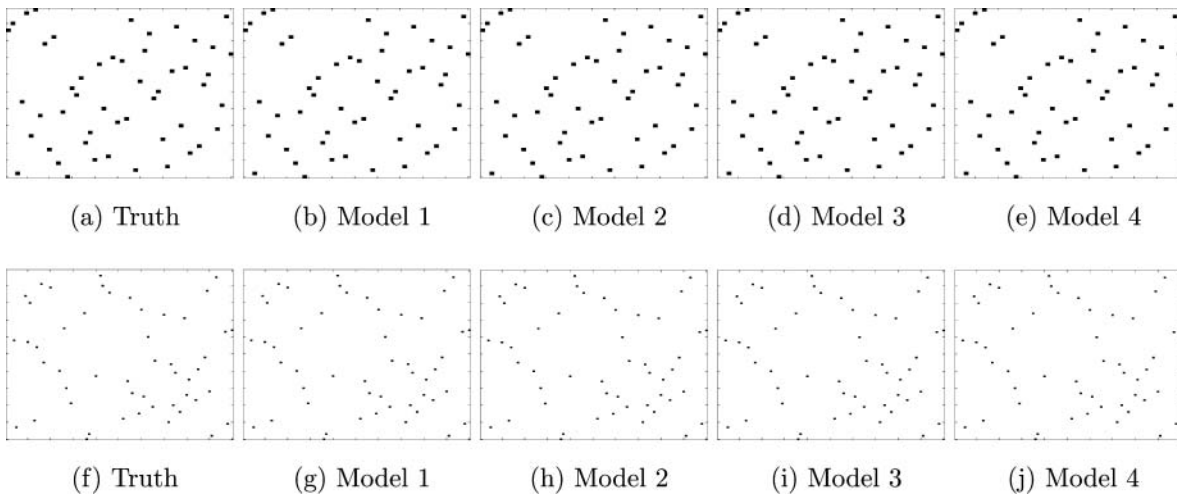


Figure 1. Heat maps of the frequency of the 0's identified for each entry of  $\Sigma_1 - \Sigma_2$  ( $n = 100$ ,  $p = 50$  for the top row and  $p = 100$  for the second row) out of 100 replications. White indicates 1000's identified out of 100 runs; black, 0/100.

expected, the gene selection based on the covariances could be more powerful than the one that is based only on the variances. The tests provide valuable information as these identified genes can be selected for a follow-up study.

Finally, we apply the support recovery procedure  $\hat{\Psi}(4)$  to  $\Sigma_1 - \Sigma_2$ . For a visual comparison between  $\Sigma_1$  and  $\Sigma_2$ , Figure 2 plots the heat map of  $\Sigma_1 - \Sigma_2$  of the 200 largest absolute values of two-sample  $t$  statistics. It can be seen from Figure 2 that the estimated support of  $\Sigma_1 - \Sigma_2$  is quite sparse.

### 6. DISCUSSION

We introduced in this article a test for the equality of two covariance matrices, which is proved to have the prespecified significance level asymptotically and to enjoy certain optimality in terms of its power. In particular, it was shown theoretically and numerically that the test is especially powerful against sparse alternatives. Support recovery and testing two covariance matrices row by row with applications to gene selection are also considered.

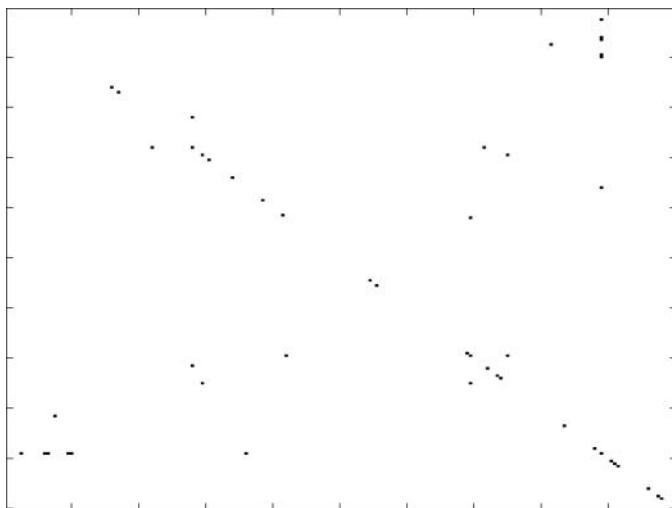


Figure 2. Heat map of the selected genes by exact recovery.

There are several possible extensions of our method. For example, an interesting direction is to generalize the procedure to testing the hypothesis of homogeneity of several covariance matrices,  $H_0 : \Sigma_1 = \dots = \Sigma_K$ , where  $K \geq 2$ . A test statistic similar to  $M_n$  can be constructed to test this hypothesis and analogous theoretical results can be developed. We shall report the details of the results elsewhere in the future as a significant amount of additional work is still needed.

Much recent attention has focused on the estimation of large covariance and precision matrices. The current work here is related to the estimation of covariance matrices. An adaptive thresholding estimator of sparse covariance matrices was introduced in Cai and Liu (2011). The procedure is based on the standardized statistics  $\hat{\sigma}_{ij}/\hat{\theta}_{ij}^{1/2}$ , which is closely related to  $M_{ij}$ . In this article, the asymptotic distribution of  $M_n = \max_{1 \leq i \leq j \leq p} M_{ij}$  is obtained. It gives a justification on the family-wise error of simultaneous tests  $H_{0ij} : \sigma_{ij1} = \sigma_{ij2}$  for  $1 \leq i \leq j \leq p$ . For example, by thresholding  $M_{ij}$  at level  $4 \log p - \log \log p + q_\alpha$ , the family-wise error is controlled asymptotically at level  $\alpha$ , that is,

$$\text{FWER} = \mathbb{P}\left(\max_{(i,j) \in G} M_{ij} \geq 4 \log p - \log \log p + q_\alpha\right) \rightarrow \alpha,$$

where  $G = \{(i, j) : \sigma_{ij1} = \sigma_{ij2}\}$  and  $\text{Card}(G^c) = o(p^2)$ . These tests are useful in the studies of differential coexpression in genetics; see de la Fuente (2010).

The test introduced in the article is based on the asymptotic results. When the sample size is small, say  $n \leq 30$ , the critical value derived from the asymptotic distribution is not sufficiently accurate and modification is thus needed. The following “normal cut-off” method can be used instead. Let  $(X_k^*, 1 \leq k \leq n_1)$  and  $(Y_k^*, 1 \leq k \leq n_2)$  be generated from iid  $N(0, I_{p \times p})$ . Let  $M_n^*$  be the maximum test statistic constructed from  $(X_k^*, 1 \leq k \leq n_1)$  and  $(Y_k^*, 1 \leq k \leq n_2)$ . It follows from Theorem 1 that under (C1)–(C3), we have

$$\sup_{y \in R} |\mathbb{P}(M_n \geq y) - \mathbb{P}(M_n^* \geq y)| \rightarrow 0$$

under the null, where  $M_n$  is the maximum test statistic computed from the original data. Let  $y_n(\alpha)$  be the critical value such that

$P(M_n^* \geq y_n(\alpha)) = \alpha$ . Then we have  $P(M_n \geq y_n(\alpha)) \rightarrow \alpha$  under  $H_0$ . So  $y_n(\alpha)$  can be used as the critical value when the sample size is small and it can be obtained easily by simulation. Some additional simulation results illustrating the case with a small sample size ( $n = 30$ ) are given in the supplementary material (Cai, Liu, and Xia 2012).

### 7. PROOFS

We prove the main results in this section. Throughout this section, we denote by  $C, C_1, C_2, \dots$ , constants that do not depend on  $n, p$  and may vary from place to place. We begin by collecting some technical lemmas that will be used in the proofs of the main results. These technical lemmas together with Proposition 1–3 are proved in the supplementary material (Cai, Liu, and Xia 2012).

#### 7.1 Technical Lemmas

The first lemma is the classical Bonferroni’s inequality.

*Lemma 1* (Bonferroni inequality). Let  $B = \cup_{t=1}^p B_t$ . For any  $k < \lfloor p/2 \rfloor$ , we have

$$\sum_{t=1}^{2k} (-1)^{t-1} E_t \leq P(B) \leq \sum_{t=1}^{2k-1} (-1)^{t-1} E_t,$$

where  $E_t = \sum_{1 \leq i_1 < \dots < i_t \leq p} P(B_{i_1} \cap \dots \cap B_{i_t})$ .

The second lemma comes from Berman (1962).

*Lemma 2.* [Berman (1962)] If  $X$  and  $Y$  have a bivariate normal distribution with expectation zero, unit variance and correlation coefficient  $\rho$ , then

$$\lim_{c \rightarrow \infty} \frac{P(X > c, Y > c)}{[2\pi(1 - \rho)^{1/2}c^2]^{-1} \exp(-\frac{c^2}{1+\rho})(1 + \rho)^{1/2}} = 1,$$

uniformly for all  $\rho$  such that  $|\rho| \leq \delta$ , for any  $\delta, 0 < \delta < 1$ .

The next lemma is on the large deviations for  $\hat{\theta}_{ij1}$  and  $\hat{\theta}_{ij2}$ .

*Lemma 3.* Under the conditions of (C2) or (C2\*), there exists some constant  $C > 0$  such that

$$P\left(\max_{i,j} |\hat{\theta}_{ij1} - \theta_{ij1}| / \sigma_{ii1} \sigma_{jj1} \geq C \frac{\varepsilon_n}{\log p}\right) = O(p^{-1} + n^{-\epsilon/8}), \tag{20}$$

and

$$P\left(\max_{i,j} |\hat{\theta}_{ij2} - \theta_{ij2}| / \sigma_{ii2} \sigma_{jj2} \geq C \frac{\varepsilon_n}{\log p}\right) = O(p^{-1} + n^{-\epsilon/8}), \tag{21}$$

where  $\varepsilon_n = \max((\log p)^{1/6}/n^{1/2}, (\log p)^{-1}) \rightarrow 0$  as  $n, p \rightarrow \infty$ .

Define

$$\begin{aligned} \tilde{\Sigma}_1 &= (\tilde{\sigma}_{ij1})_{p \times p} = \frac{1}{n_1} \sum_{k=1}^{n_1} (X - \mu_1)(X - \mu_1)^T, \\ \tilde{\Sigma}_2 &= (\tilde{\sigma}_{ij2})_{p \times p} = \frac{1}{n_2} \sum_{k=1}^{n_2} (Y - \mu_2)(Y - \mu_2)^T. \end{aligned}$$

Let  $\Lambda$  be any subset of  $\{(i, j) : 1 \leq i \leq j \leq p\}$  and  $|\Lambda| = \text{Card}(\Lambda)$ .

*Lemma 4.* Under the conditions of (C2) or (C2\*), we have for some constant  $C > 0$  that

$$\begin{aligned} P\left(\max_{(i,j) \in \Lambda} \frac{(\tilde{\sigma}_{ij1} - \tilde{\sigma}_{ij2} - \sigma_{ij1} + \sigma_{ij2})^2}{\theta_{ij1}/n_1 + \theta_{ij2}/n_2} \geq x^2\right) \\ \leq C|\Lambda|(1 - \Phi(x)) + O(p^{-1} + n^{-\epsilon/8}) \end{aligned}$$

uniformly for  $0 \leq x \leq (8 \log p)^{1/2}$  and  $\Lambda \subseteq \{(i, j) : 1 \leq i \leq j \leq p\}$ .

The proofs of Lemmas 3 and 4 are given in the supplementary material (Cai, Liu, and Xia 2012).

#### 7.2 Proof of Theorem 1

Without loss of generality, we assume that  $\mu_1 = 0, \mu_2 = 0, \sigma_{ii1} = \sigma_{ii2} = 1$  for  $1 \leq i \leq p$ . Let

$$\hat{M}_n = \max_{1 \leq i \leq j \leq p} \frac{(\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2})^2}{\theta_{ij1}/n_1 + \theta_{ij2}/n_2},$$

and

$$\tilde{M}_n = \max_{1 \leq i \leq j \leq p} \frac{(\tilde{\sigma}_{ij1} - \tilde{\sigma}_{ij2})^2}{\theta_{ij1}/n_1 + \theta_{ij2}/n_2}.$$

Note that under the event  $\{|\hat{\theta}_{ij1}/\theta_{ij1} - 1| \leq C\varepsilon_n/\log p, |\hat{\theta}_{ij2}/\theta_{ij2} - 1| \leq C\varepsilon_n/\log p\}$ ,

$$|M_n - \hat{M}_n| \leq C\hat{M}_n \frac{\varepsilon_n}{\log p},$$

$$\begin{aligned} |\hat{M}_n - \tilde{M}_n| &\leq C(n_1 + n_2) \left( \max_{1 \leq i \leq p} \bar{X}_i^4 + \max_{1 \leq i \leq p} \bar{Y}_i^4 \right) \\ &\quad + C(n_1 + n_2)^{1/2} \tilde{M}_n^{1/2} \left( \max_{1 \leq i \leq p} \bar{X}_i^2 + \max_{1 \leq i \leq p} \bar{Y}_i^2 \right). \end{aligned}$$

By the exponential inequality,  $\max_{1 \leq i \leq p} |\bar{X}_i| + \max_{1 \leq i \leq p} |\bar{Y}_i| = O_P(\sqrt{\log p/n})$ . Thus, by Lemma 3, it suffices to show that for any  $x \in \mathbb{R}$ ,

$$P(\tilde{M}_n - 4 \log p + \log \log p \leq x) \rightarrow \exp\left(-\frac{1}{\sqrt{8\pi}} \exp\left(\frac{x}{2}\right)\right) \tag{22}$$

as  $n, p \rightarrow \infty$ . Let  $\rho_{ij} = \rho_{ij1} = \rho_{ij2}$  under  $H_0$ . Define

$$\begin{aligned} A_j &= \{i : |\rho_{ij}| \geq (\log p)^{-1-\alpha_0}\}, \\ A &= \{(i, j) : 1 \leq i \leq j \leq p\}, \\ A_0 &= \{(i, j) : 1 \leq i \leq p, i \leq j, j \in A_i\}, \\ B_0 &= \{(i, j) : i \in \Lambda(r) \cup \Upsilon, i < j \leq p\} \cup \{(i, j) : j \in \Lambda(r) \cup \Upsilon, 1 \leq i < j\}, \\ D_0 &= A_0 \cup B_0. \end{aligned}$$

By the definition of  $D_0$ , for any  $(i_1, j_1) \in A \setminus D_0$  and  $(i_2, j_2) \in A \setminus D_0$ , we have  $|\rho_{kl}| \leq r$  for any  $k \neq l \in \{i_1, j_1, i_2, j_2\}$ . Set  $y_p = x + 4 \log p - \log \log p$ ,

$$\tilde{M}_{A \setminus D_0} = \max_{(i,j) \in A \setminus D_0} \frac{(\tilde{\sigma}_{ij1} - \tilde{\sigma}_{ij2})^2}{\theta_{ij1}/n_1 + \theta_{ij2}/n_2},$$

and

$$\tilde{M}_{D_0} = \max_{(i,j) \in D_0} \frac{(\tilde{\sigma}_{ij1} - \tilde{\sigma}_{ij2})^2}{\theta_{ij1}/n_1 + \theta_{ij2}/n_2}.$$

Then  $|P(\tilde{M}_n \geq y_p) - P(\tilde{M}_{A \setminus D_0} \geq y_p)| \leq P(\tilde{M}_{D_0} \geq y_p)$ . Note that  $\text{Card}(\Lambda(r)) + \text{Card}(\Upsilon) = o(p)$  and for all  $\gamma > 0$ ,

$\max_{j \notin \gamma} s_j = o(p^\gamma)$ . This implies that  $\text{Card}(D_0) \leq C_0 p^{1+\gamma} + o(p^2)$  for any  $\gamma > 0$ . By Lemma 4, we obtain that for any fixed  $x \in \mathbb{R}$ ,

$$\mathbb{P}(\tilde{M}_{D_0} \geq y_p) \leq \text{Card}(D_0) \times Cp^{-2} + o(1) = o(1).$$

Set

$$\tilde{M}_{A \setminus D_0} = \max_{(i,j) \in A \setminus D_0} \frac{(\tilde{\sigma}_{ij1} - \tilde{\sigma}_{ij2})^2}{\theta_{ij1}/n_1 + \theta_{ij2}/n_2} =: \max_{(i,j) \in A \setminus D_0} U_{ij}^2,$$

where  $U_{ij} = \frac{\tilde{\sigma}_{ij1} - \tilde{\sigma}_{ij2}}{\sqrt{\theta_{ij1}/n_1 + \theta_{ij2}/n_2}}$ . It suffices to show that for any  $x \in \mathbb{R}$ ,

$$\begin{aligned} \mathbb{P}(\tilde{M}_{A \setminus D_0} - 4 \log p + \log \log p \leq x) \\ \rightarrow \exp\left(-\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right)\right) \end{aligned}$$

as  $n, p \rightarrow \infty$ . We arrange the two-dimensional indices  $\{(i, j) : (i, j) \in A \setminus D_0\}$  in any ordering and set them as  $\{(i_m, j_m) : 1 \leq m \leq q\}$  with  $q = \text{Card}(A \setminus D_0)$ . Let  $n_2/n_1 \leq K_1$  with  $K_1 \geq 1$ ,  $\theta_{k1} = \theta_{i_k j_{k1}}, \theta_{k2} = \theta_{i_k j_{k2}}$  and define

$$\begin{aligned} Z_{lk} &= \frac{n_2}{n_1} (X_{li_k} X_{lj_{k1}} - \sigma_{i_k j_{k1}}), 1 \leq l \leq n_1, \\ Z_{lk} &= -(Y_{li_k} Y_{lj_{k2}} - \sigma_{i_k j_{k2}}), n_1 + 1 \leq l \leq n_1 + n_2, \\ V_k &= \frac{1}{\sqrt{n_2^2 \theta_{k1}/n_1 + n_2 \theta_{k2}}} \sum_{l=1}^{n_1+n_2} Z_{lk}, \\ \hat{V}_k &= \frac{1}{\sqrt{n_2^2 \theta_{k1}/n_1 + n_2 \theta_{k2}}} \sum_{l=1}^{n_1+n_2} \hat{Z}_{lk}, \end{aligned}$$

where  $\hat{Z}_{lk} = Z_{lk} I\{|Z_{lk}| \leq \tau_n\} - \mathbb{E}Z_{lk} I\{|Z_{lk}| \leq \tau_n\}$ , and  $\tau_n = \eta^{-1} 8K_1 \log(p+n)$  if (C2) holds,  $\tau_n = \sqrt{n}/(\log p)^8$  if (C2\*) holds. Note that  $\max_{(i,j) \in A \setminus D_0} U_{ij}^2 = \max_{1 \leq k \leq q} V_k^2$ . We have, if (C2) holds, then

$$\begin{aligned} \max_{1 \leq k \leq q} \frac{1}{\sqrt{n}} \sum_{l=1}^{n_1+n_2} \mathbb{E}|Z_{lk}| I\{|Z_{lk}| \geq \eta^{-1} 8K_1 \log(p+n)\} \\ \leq C\sqrt{n} \max_{1 \leq l \leq n_1+n_2} \max_{1 \leq k \leq q} \mathbb{E}|Z_{lk}| I\{|Z_{lk}| \geq \eta^{-1} 8K_1 \log(p+n)\} \\ \leq C\sqrt{n}(p+n)^{-4} \max_{1 \leq l \leq n_1+n_2} \max_{1 \leq k \leq q} \mathbb{E}|Z_{lk}| \exp(\eta|Z_{lk}|/(2K_1)) \\ \leq C\sqrt{n}(p+n)^{-4}, \end{aligned}$$

and if (C2\*) holds, then

$$\begin{aligned} \max_{1 \leq k \leq q} \frac{1}{\sqrt{n}} \sum_{l=1}^{n_1+n_2} \mathbb{E}|Z_{lk}| I\{|Z_{lk}| \geq \sqrt{n}/(\log p)^8\} \\ \leq C\sqrt{n} \max_{1 \leq l \leq n_1+n_2} \max_{1 \leq k \leq q} \mathbb{E}|Z_{lk}| I\{|Z_{lk}| \geq \sqrt{n}/(\log p)^8\} \\ \leq Cn^{-\gamma_0 - \epsilon/8}. \end{aligned}$$

Hence, we have

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq k \leq q} |V_k - \hat{V}_k| \geq (\log p)^{-1}\right) \\ \leq \mathbb{P}\left(\max_{1 \leq k \leq q} \max_{1 \leq l \leq n_1+n_2} |Z_{lk}| \geq \tau_n\right) \\ \leq np \max_{1 \leq j \leq p} [\mathbb{P}(X_j^2 \geq \tau_n/2) + \mathbb{P}(Y_j^2 \geq \tau_n/2)] \\ = O(p^{-1} + n^{-\epsilon/8}). \end{aligned} \tag{23}$$

Note that

$$\begin{aligned} \left| \max_{1 \leq k \leq q} V_k^2 - \max_{1 \leq k \leq q} \hat{V}_k^2 \right| \leq 2 \max_{1 \leq k \leq q} |\hat{V}_k| \max_{1 \leq k \leq q} |V_k - \hat{V}_k| \\ + \max_{1 \leq k \leq q} |V_k - \hat{V}_k|^2. \end{aligned} \tag{24}$$

By (23) and (24), it suffices to prove that for any  $x \in \mathbb{R}$ , as  $n, p \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq k \leq q} \hat{V}_k^2 - 4 \log p + \log \log p \leq x\right) \\ \rightarrow \exp\left(-\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right)\right). \end{aligned} \tag{25}$$

By Bonferroni inequality (see Lemma 1), for any integer  $m$  with  $0 < m < q/2$ ,

$$\begin{aligned} \sum_{d=1}^{2m} (-1)^{d-1} \sum_{1 \leq k_1 < \dots < k_d \leq q} \mathbb{P}\left(\bigcap_{j=1}^d E_{k_j}\right) \leq \mathbb{P}\left(\max_{1 \leq k \leq q} \hat{V}_k^2 \geq y_p\right) \\ \leq \sum_{d=1}^{2m-1} (-1)^{d-1} \sum_{1 \leq k_1 < \dots < k_d \leq q} \mathbb{P}\left(\bigcap_{j=1}^d E_{k_j}\right), \end{aligned} \tag{26}$$

where  $E_{k_j} = \{\hat{V}_{k_j}^2 \geq y_p\}$ . Let  $\tilde{Z}_{lk} = \hat{Z}_{lk}/(n_2 \theta_{k1}/n_1 + \theta_{k2})^{1/2}$  for  $1 \leq k \leq q$  and  $\mathbf{W}_l = (\tilde{Z}_{lk_1}, \dots, \tilde{Z}_{lk_d})$ , for  $1 \leq l \leq n_1 + n_2$ . Define  $|\mathbf{a}|_{\min} = \min_{1 \leq i \leq d} |a_i|$  for any vector  $\mathbf{a} \in \mathbb{R}^d$ . Then we have

$$\mathbb{P}\left(\bigcap_{j=1}^d E_{k_j}\right) = \mathbb{P}\left(\left|n_2^{-1/2} \sum_{l=1}^{n_1+n_2} \mathbf{W}_l\right|_{\min} \geq y_n^{1/2}\right).$$

By Theorem 1 in Zaitsev (1987), we have

$$\begin{aligned} \mathbb{P}\left(\left|n_2^{-1/2} \sum_{l=1}^{n_1+n_2} \mathbf{W}_l\right|_{\min} \geq y_n^{1/2}\right) \\ \leq \mathbb{P}(|N_d|_{\min} \geq y_p^{1/2} - \epsilon_n (\log p)^{-1/2}) \\ + c_1 d^{5/2} \exp\left(-\frac{n^{1/2} \epsilon_n}{c_2 d^3 \tau_n (\log p)^{1/2}}\right), \end{aligned} \tag{27}$$

where  $c_1 > 0$  and  $c_2 > 0$  are absolutely constants,  $\epsilon_n \rightarrow 0$ , which will be specified later and  $\mathbf{N}_d =: (N_{k_1}, \dots, N_{k_d})$  is a normal vector with  $\mathbb{E}\mathbf{N}_d = 0$  and  $\text{cov}(\mathbf{N}_d) = \frac{n_1}{n_2} \text{cov}(\mathbf{W}_1) + \text{cov}(\mathbf{W}_{n_1+1})$ . Recall that  $d$  is a fixed integer not depending on  $n, p$ . Because  $\log p = o(n^{1/5})$ , we can let  $\epsilon_n \rightarrow 0$  sufficiently slow such that

$$c_1 d^{5/2} \exp\left(-\frac{n^{1/2} \epsilon_n}{c_2 d^3 \tau_n (\log p)^{1/2}}\right) = O(p^{-M}) \tag{28}$$

for any large  $M > 0$ . It follows from (26), (27), and (28) that

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq k \leq q} \hat{V}_k^2 \geq y_p\right) \leq \sum_{d=1}^{2m-1} (-1)^{d-1} \sum_{1 \leq k_1 < \dots < k_d \leq q} \\ \times \mathbb{P}(|N_d|_{\min} \geq y_p^{1/2} - \epsilon_n (\log p)^{-1/2}) + o(1). \end{aligned} \tag{29}$$

Similarly, using Theorem 1 in Zaitsev (1987) again, we can get

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq k \leq q} \hat{V}_k^2 \geq y_p\right) \geq \sum_{d=1}^{2m} (-1)^{d-1} \sum_{1 \leq k_1 < \dots < k_d \leq q} \\ \times \mathbb{P}(|N_d|_{\min} \geq y_p^{1/2} + \epsilon_n (\log p)^{-1/2}) - o(1). \end{aligned} \tag{30}$$

We need the following technical lemma whose proof is involved, and is given in the supplementary material (Cai, Liu, and Xia 2012).

*Lemma 5.* For any fixed integer  $d \geq 1$  and real number  $x \in \mathbb{R}$ ,

$$\begin{aligned} & \sum_{1 \leq k_1 < \dots < k_d \leq q} \mathbf{P}(|N_d|_{\min} \geq y_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}) \\ &= \frac{1}{d!} \left( \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right) \right)^d (1 + o(1)). \end{aligned} \tag{31}$$

Now submitting (31) into (29) and (30), we get

$$\begin{aligned} & \limsup_{n, p \rightarrow \infty} \mathbf{P}\left(\max_{1 \leq k \leq q} \hat{V}_k^2 \geq y_p\right) \\ & \leq \sum_{d=1}^{2m} (-1)^{d-1} \frac{1}{d!} \left( \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right) \right)^d \end{aligned}$$

and

$$\begin{aligned} & \liminf_{n, p \rightarrow \infty} \mathbf{P}\left(\max_{1 \leq k \leq q} \hat{V}_k^2 \geq y_p\right) \\ & \geq \sum_{d=1}^{2m-1} (-1)^{d-1} \frac{1}{d!} \left( \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right) \right)^d \end{aligned}$$

for any positive integer  $m$ . Letting  $m \rightarrow \infty$ , we obtain (25).  $\square$

### 7.3 Proofs of the Other Theorems

*Proof of Theorem 2.* Recall that

$$M_n^1 = \max_{1 \leq i \leq j \leq p} \frac{(\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2} - \sigma_{ij1} + \sigma_{ij2})^2}{\hat{\theta}_{ij1}/n_1 + \hat{\theta}_{ij2}/n_2}.$$

By Lemmas 3 and 4,

$$\mathbf{P}\left(M_n^1 \leq 4 \log p - \frac{1}{2} \log \log p\right) \rightarrow 1 \tag{32}$$

as  $n, p \rightarrow \infty$ . By Lemma 3, the inequalities

$$\max_{1 \leq i \leq j \leq p} \frac{(\sigma_{ij1} - \sigma_{ij2})^2}{\hat{\theta}_{ij1}/n_1 + \hat{\theta}_{ij2}/n_2} \leq 2M_n^1 + 2M_n \tag{33}$$

and

$$\max_{1 \leq i \leq j \leq p} \frac{(\sigma_{ij1} - \sigma_{ij2})^2}{\theta_{ij1}/n_1 + \theta_{ij2}/n_2} \geq 16 \log p,$$

which yields that  $\mathbf{P}(M_n \geq q_\alpha + 4 \log p - \log \log p) \rightarrow 1$  as  $n, p \rightarrow \infty$ .  $\square$

*Proof of Theorem 4.* It suffices to take  $\mathcal{T}_\alpha$  to be the set of  $\alpha$ -level tests over the normal distributions, since it contains all the  $\alpha$ -level tests over the collection of distributions satisfying (C2) or (C2\*). Let  $\mathcal{M}$  denote the set of all subsets of  $\{1, \dots, p\}$  with cardinality  $c_0(p)$ . Let  $\hat{m}$  be a random subset of  $\{1, \dots, p\}$ , which is uniformly distributed on  $\mathcal{M}$ . We construct a class of  $\Sigma_1, \mathcal{N} = \{\Sigma_{\hat{m}}, \hat{m} \in \mathcal{M}\}$ , such that

$$\sigma_{ij} = 0 \text{ for } i \neq j \text{ and } \sigma_{ii} - 1 = \rho \mathbf{1}_{i \in \hat{m}},$$

for  $i, j = 1, \dots, p$  and  $\rho = c\sqrt{\log p/n}$ , where  $c > 0$  will be specified later. Let  $\Sigma_2 = \mathbf{I}$  and  $\Sigma_1$  be uniformly distributed on

$\mathcal{N}$ . Let  $\mu_\rho$  be the distribution of  $\Sigma_1 - \mathbf{I}$ . Note that  $\mu_\rho$  is a probability measure on  $\{\Delta \in \mathcal{S}(c_0(p)) : \|\Delta\|_F^2 = c_0(p)\rho^2\}$ , where  $\mathcal{S}(c_0(p))$  is the class of matrices with  $c_0(p)$  nonzero entries. Let  $d\mathbf{P}_1(\{X_n, Y_n\})$  be the likelihood function given  $\Sigma_1$  being uniformly distributed on  $\mathcal{N}$  and

$$L_{\mu_\rho} := L_{\mu_\rho}(\{X_n, Y_n\}) = \mathbf{E}_{\mu_\rho} \left( \frac{d\mathbf{P}_1(\{X_n, Y_n\})}{d\mathbf{P}_0(\{X_n, Y_n\})} \right),$$

where  $\mathbf{E}_{\mu_\rho}(\cdot)$  is the expectation on  $\Sigma_1$ . By the arguments in Baraud (2002, p. 595), it suffices to show that  $\mathbf{E}L_{\mu_\rho}^2 \leq 1 + o(1)$ . It is easy to see that

$$L_{\mu_\rho} = \mathbf{E}_{\hat{m}} \left( \prod_{i=1}^n \frac{1}{|\Sigma_{\hat{m}}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{Z}_i^T (\Sigma_{\hat{m}} - \mathbf{I}) \mathbf{Z}_i\right) \right),$$

where  $\Sigma_{\hat{m}} = \Sigma_{\hat{m}}^{-1}$  and  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  are iid multivariate normal vectors with mean vector  $\mathbf{0}$  and covariance matrix  $\mathbf{I}$ . Thus,

$$\begin{aligned} \mathbf{E}L_{\mu_\rho}^2 &= \mathbf{E} \left( \frac{1}{\binom{p}{k_p}} \sum_{m \in \mathcal{M}} \left( \prod_{i=1}^n \frac{1}{|\Sigma_m|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{Z}_i^T (\Sigma_m - \mathbf{I}) \mathbf{Z}_i\right) \right) \right)^2 \\ &= \frac{1}{\binom{p}{k_p}^2} \sum_{m, m' \in \mathcal{M}} \mathbf{E} \left( \prod_{i=1}^n \frac{1}{|\Sigma_m|^{1/2}} \frac{1}{|\Sigma_{m'}|^{1/2}} \right. \\ & \quad \left. \times \exp\left(-\frac{1}{2} \mathbf{Z}_i^T (\Sigma_m + \Sigma_{m'} - 2\mathbf{I}) \mathbf{Z}_i\right) \right). \end{aligned}$$

Set  $\Sigma_m + \Sigma_{m'} - 2\mathbf{I} = (a_{ij})$ . Then  $a_{ij} = 0$  for  $i \neq j$ ,  $a_{jj} = 0$  if  $j \in (m \cup m')^c$ ,  $a_{jj} = 2(\frac{1}{1+\rho} - 1)$  if  $j \in m \cap m'$  and  $a_{jj} = \frac{1}{1+\rho} - 1$  if  $j \in m \setminus m'$  and  $m' \setminus m$ . Let  $t = |m \cap m'|$ . Then we have

$$\begin{aligned} \mathbf{E}L_{\mu_\rho}^2 &= \frac{1}{\binom{p}{k_p}} \sum_{t=0}^{k_p} \binom{k_p}{t} \binom{p-k_p}{k_p-t} \\ & \quad \times \frac{1}{(1+\rho)^{k_p n}} (1+\rho)^{(k_p-t)n} \left( \frac{1+\rho}{1-\rho} \right)^{tn/2} \\ & \leq p^{k_p} (p-k_p)! / p! \sum_{t=0}^{k_p} \binom{k_p}{t} \left( \frac{k_p}{p} \right)^t \left( \frac{1}{1-\rho^2} \right)^{tn/2} \\ & = (1+o(1)) \left( 1 + \frac{k_p}{p(1-\rho^2)^{n/2}} \right)^{k_p}, \end{aligned}$$

for  $r < 1/2$ . So

$$\begin{aligned} \mathbf{E}L_{\mu_\rho}^2 &\leq \exp\left(k_p \log(1+k_p p^{c^2-1})\right) (1+o(1)) \\ &\leq \exp\left(k_p^2 p^{c^2-1}\right) (1+o(1)) = 1+o(1) \end{aligned}$$

by letting  $c$  be sufficiently small. Theorem 4 is proved.  $\square$

*Proof of Theorem 5.* The proof of Theorem 5 is similar to that of Theorem 2. In fact, by (32) and a similar inequality as (33), we can get

$$\mathbf{P}\left(\min_{(i,j) \in \Psi} M_{ij} \geq 4 \log p\right) \rightarrow 1 \tag{34}$$

uniformly for  $(\Sigma_1, \Sigma_2) \in \mathcal{W}_0(4)$ .  $\square$

*Proof of Theorem 6.* Without loss of generality, we assume  $\Theta = \{1, 2, \dots, p_0\}$  with  $p_0 = \text{Card}(\theta) \geq p^\gamma$  for all

$0 < \gamma < 1$ . Let  $A_1$  be the largest subset of  $\Theta \setminus \{1\}$  such that  $\sigma_{1k1} = \sigma_{1k2}$  for all  $k \in A_1$ . Let  $i_1 = \min\{j : j \in A_1, j > 1\}$ . Then we have  $|i_1 - 1| \leq s_0(p)$ . Also,  $\text{Card}(A_1) \geq p_0 - s_0(p)$ . Similarly, let  $A_l$  be the largest subset of  $A_{l-1} \setminus \{i_{l-1}\}$  such that  $\sigma_{i_{l-1}k1} = \sigma_{i_{l-1}k2}$  for all  $k \in A_l$  and  $i_l = \min\{j : j \in A_l, j > i_{l-1}\}$ . We can see that  $i_l - i_{l-1} \leq s_0(p)$  for  $l < p_0/s_0(p)$  and  $\text{Card}(A_l) \geq \text{Card}(A_{l-1}) - s_0(p) \geq p_0 - (s_0(p) + 1)l$ . Let  $l = \lceil p^{\tau_2} \rceil$  with  $\tau/4 < \tau_2 < \tau_1$ . Let  $\Sigma_{1l}$  and  $\Sigma_{2l}$  be the covariance matrices of  $(X_{i_0}, \dots, X_{i_l})$  and  $(Y_{i_0}, \dots, Y_{i_l})$ . Then the entries of  $\Sigma_{1l}$  and  $\Sigma_{2l}$  are the same except for the diagonal. Hence, it follows from the proof of Theorem 1 that

$$\begin{aligned} & \mathbb{P} \left( \max_{0 \leq j < k \leq l} M_{i_j i_k} - 4 \log l + \log \log l \leq x \right) \\ & \rightarrow \exp \left( -\frac{1}{\sqrt{8\pi}} \exp \left( -\frac{x}{2} \right) \right) \end{aligned}$$

uniformly for all  $\Sigma_1 - \Sigma_2 \in \mathcal{S}_0$ . This implies that

$$\inf_{\Sigma_1 - \Sigma_2 \in \mathcal{S}_0} \mathbb{P} \left( \max_{0 \leq j < k \leq l} M_{i_j i_k} \geq c \log p \right) \rightarrow 1$$

for all  $\tau < c < 4\tau_2$ . By the definition of  $\hat{\Psi}(\tau)$  and the fact  $\sigma_{i_j i_k 1} = \sigma_{i_j i_k 2}$  for all  $0 \leq j < k \leq l$ , Theorem 6 is proved.  $\square$

*Proof of Theorems 7 and 8.* The proofs of Theorems 7 and 8 are similar to that of Theorem 1. Note that by Lemmas 3 and 4,

$$\begin{aligned} & \left| \mathbb{P} \left( \max_{i \in F^c} \Phi_{i,\alpha} = 1 \right) - \mathbb{P} \left( \max_{(i,j) \in A \setminus E_1} M_{ij} \geq \alpha_p \right) \right| \\ & \leq C \text{Card}(F) p \times p^{-2} + o(1) = o(1), \end{aligned}$$

where  $E_1$  is defined by

$$E_1 = \{(i, j) : i \in F, i \leq j \leq p\} \cup \{(i, i) : 1 \leq i \leq p\}.$$

By replacing  $A$ ,  $A_0$ , and  $B_0$  in the proof of Theorem 1 with  $A \setminus E_1$ ,  $A_0 \setminus E_1$ , and  $B_0 \setminus E_1$ , respectively, the rest proof of Theorem 8 follows exactly the same as that of Theorem 1. The proof of Theorem 7 is much simpler than that of Theorem 1. Due to the close similarity, the proof is omitted here.  $\square$

## SUPPLEMENTARY MATERIALS

In the supplement we prove Proposition 1–3 and the technical results, Lemmas 3, 4, and 5, which are used in the proofs of the main results. We also present more extensive simulation results comparing the numerical performance of the proposed test with that of other tests.

[Received October 2012. Revised November 2012.]

## REFERENCES

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis* (3rd ed.), New York: Wiley-Interscience. [265,267]  
 Baraud, Y. (2002), “Non-Asymptotic Minimax Rates of Testing in Signal Detection,” *Bernoulli*, 8, 577–606. [276]

Berman, S. M. (1962), “A Law of Large Numbers for the Maximum of a Stationary Gaussian Sequence,” *The Annals of Mathematical Statistics*, 33, 894–908. [274]  
 Cai, T., and Jiang, T. (2011), “Limiting Laws of Coherence of Random Matrices With Applications to Testing Covariance Structure and Construction of Compressed Sensing Matrices,” *The Annals of Statistics*, 39, 1496–1525. [268]  
 ——— (2012), “Phase Transition in Limiting Distributions of Coherence of High-Dimensional Random Matrices,” *Journal of Multivariate Analysis*, 107, 24–39. [268]  
 Cai, T., and Liu, W. D. (2011), “Adaptive Thresholding for Sparse Covariance Matrix Estimation,” *Journal of the American Statistical Association*, 106, 672–684. [266,273]  
 Cai, T., Liu, W. D., and Xia, Y. (2012), Supplement to “Two-Sample Covariance Matrix Testing and Support Recovery in High-Dimensional and Sparse Settings,” Technical Report. [271,274,276]  
 de la Fuente, A. (2010), “From ‘Differential Expression’ to ‘Differential Networking’—Identification of Dysfunctional Regulatory Networks in Diseases,” *Trends in Genetics*, 26, 326–333. [273]  
 Gupta, A. K., and Tang, J. (1984), “Distribution of Likelihood Ratio Statistic for Testing Equality of Covariance Matrices of Multivariate Gaussian Models,” *Biometrika*, 71, 555–559. [265]  
 Gupta, D. S., and Giri, N. (1973), “Properties of Tests Concerning Covariance Matrices of Normal Distributions,” *The Annals of Statistics*, 6, 1222–1224. [265]  
 Ho, J. W., Stefani, M., dos Remedios, C. G., and Charleston, M. A. (2008), “Differential Variability Analysis of Gene Expression and its Application to Human Diseases,” *Bioinformatics*, 24, 390–398. [266,269,270,272]  
 Hu, R., Qiu, X., and Glazko, G. (2010), “A New Gene Selection Procedure Based on the Covariance Distance,” *Bioinformatics*, 26, 348–354. [266,270,272]  
 Hu, R., Qiu, X., Glazko, G., Klevanov, L., and Yakovlev, A. (2009), “Detecting Intergene Correlation Changes in Microarray Analysis: A New Approach to Gene Selection,” *BMC Bioinformatics*, 10, 20. [266,270]  
 Jiang, T. (2004), “The Asymptotic Distributions of the Largest Entries of Sample Correlation Matrices,” *The Annals of Applied Probability*, 14, 865–880. [268]  
 Li, J., and Chen, S. (2012), “Two Sample Tests for High Dimensional Covariance Matrices,” *The Annals of Statistics*, 40, 908–940. [265,269,271]  
 Liu, W., Lin, Z. Y., and Shao, Q. M. (2008), “The Asymptotic Distribution and Berry-Esseen Bound of a New Test for Independence in High Dimension With an Application to Stochastic Optimization,” *The Annals of Applied Probability*, 18, 2337–2366. [268]  
 O’Brien, P. C. (1992), “Robust Procedures for Testing Equality of Covariance Matrices,” *Biometrics*, 48, 819–827. [265]  
 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. [265]  
 Schott, J. R. (2007), “A Test for the Equality of Covariance Matrices When the Dimension is Large Relative to the Sample Sizes,” *Computational Statistics and Data Analysis*, 51, 6535–6542. [265,268,269,271]  
 Singh, D., Febbo, P., Ross, K., Jackson, D., Manola, J., Ladd, C., Tamayo, P., Renshaw, A., D’Amico, A., and Richie, J. (2002), “Gene Expression Correlates of Clinical Prostate Cancer Behavior,” *Cancer Cell*, 1, 203–209. [272]  
 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. [265,268,269,271]  
 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, 1682–1692. [265]  
 Xu, P., Brock, G. N., and Parrish, R. S. (2009), “Modified Linear Discriminant Analysis Approaches for Classification of High-Dimensional Microarray Data,” *Computational Statistics and Data Analysis*, 53, 1674–1687. [272]  
 Zaitsev, A. Yu (1987), “On the Gaussian Approximation of Convolutions Under Multidimensional Analogues of S.N. Bernstein’s Inequality Conditions,” *Probability Theory and Related Fields*, 74, 535–566. [275]  
 Zhou, W. (2007), “Asymptotic Distribution of the Largest Off-Diagonal Entry of Correlation Matrices,” *Transactions of the American Mathematical Society*, 359, 5345–5364. [268]