Weighted Statistic in Detecting Faint and Sparse Alternatives for High-Dimensional Covariance Matrices

Yang, Qing; Pan, Guangming

2017

Yang, Q., & Pan, G. (2017). Weighted Statistic in Detecting Faint and Sparse Alternatives for High-Dimensional Covariance Matrices. *Journal of the American Statistical Association*, 112(517), 188-200.

<https://hdl.handle.net/10356/85597>

<https://doi.org/10.1080/01621459.2015.1122602>

© 2017 American Statistical Association. This is the author created version of a work that has been peer reviewed and accepted for publication in *Journal of the American Statistical Association*, published by Taylor & Francis on behalf of American Statistical Association. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1080/01621459.2015.1122602>].

Downloaded on 26 Apr 2023 19:21:50 SGT

Weighted Statistic in Detecting Faint and Sparse Alternatives for High-dimensional Covariance Matrices

Qing Yang and Guangming Pan

Division of Mathematical Sciences, Nanyang Technological University

Singapore, 637371

email: qyang1@e.ntu.edu.sg; gmpan@ntu.edu.sg

Abstract

This paper considers testing equality of two population covariance matrices when the data dimension p diverges with the sample size n ($p/n \rightarrow c > 0$). We propose a weighted test statistic which is data-driven and powerful in both faint alternatives (many small disturbances) and sparse alternatives (several large disturbances). Its asymptotic null distribution is derived by large random matrix theory without assuming the existence of a limiting cumulative distribution function of the population covariance matrix. The simulation results confirm that our statistic is powerful against all alternatives, while other tests given in the literature fail in at least one situation.

Keywords: Linear Spectral Statistic; Empirical spectral distribution; Large random matrix theory; Stieltjes transform; Faint alternative; Sparse alternative.

1 Introduction

Testing whether two populations share the same covariance matrix is of enduring interest. This has been well studied in the low-dimensional setting. The most popular one is the classical likelihood ratio test (LRT), see Anderson (2003). Specifically speaking, let $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T, i = 1, \dots, n_1$ and $y_j = (y_{1j}, y_{2j}, \dots, y_{pj})^T, j = 1, \dots, n_2$ be independent

samples from two p -dimensional populations. The conventional LRT statistic is proposed as:

$$R = -2\log L, \quad L = \frac{|Q_1|^{\frac{n_1}{2}} |Q_2|^{\frac{n_2}{2}}}{|c_1 Q_1 + c_2 Q_2|^{\frac{n_1+n_2}{2}}},$$

where

$$Q_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \bar{x})(x_i - \bar{x})^T, \quad Q_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \bar{y})(y_i - \bar{y})^T,$$

$$c_1 = \frac{n_1}{n_1 + n_2}, \quad c_2 = \frac{n_2}{n_1 + n_2}, \quad \bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i, \quad \bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i.$$

When p is fixed, it is well known that under the multivariate normal populations assumption, as $n_1, n_2 \rightarrow \infty$, $R \xrightarrow{D} \chi_{1/2p(p+1)}^2$ under the null hypothesis $H_0 : \Sigma_1 = \Sigma_2$. Here, “ $\xrightarrow{D}$ ” means convergence in distribution. For other testing statistics in the low-dimensional setting one can refer to Wilks (1932), Bartlett (1937), Box (1949), Sugiura and Nagao (1968), Gupta and Giri (1973), Gupta and Tang (1984), Perlman (1980), Muirhead (1982), O’Brien (1992) and so on.

However, nowadays high-dimensional data with relatively small sample size have been attracting many researchers’ attention due to the rapid development of modern technology. Such data examples can be found in microarrays (measure gene expression), finance (time series data), climate data (spatial data) and neuroimaging (diffusion tensor imaging). In the high-dimensional setting, the aforementioned classical methods will fail or perform poorly. More specifically, for the LRT statistic R , when the data dimension p is larger than the sample size n_1 or n_2 , either the numerator of L is 0 and subsequently the test statistic $R \rightarrow \infty$, or the denominator of L is 0 and then R is not defined. Even when p is less than the sample size but p tends to ∞ together with the sample size, i.e. $p/n_i \rightarrow c_i \in (0, 1)$, Bai et al. (2009) illustrated that this classical LRT still encounters a power loss.

Regarding the test for the equality of two high-dimensional covariance matrices, there are already several methods in the literature. Broadly speaking, we can classify them into three categories. The first type is the corrected LRTs—these methods provide some corrections to the conventional LRT so that they could be used in the diverging dimension testing. For example, by applying large random matrix theory, Bai et al. (2009) proposed to study the rescaled statistic R based on the asymptotic properties for linear spectral statistics of random F -matrices. Rather than the traditional χ^2 approximation, they showed an asymptotic $N(0, 1)$ null distribution. For other recent works one can refer to Jiang et al. (2012) and

Jiang and Yang (2013), where the authors also derived the central limit theorem (CLT) under H_0 by analyzing the moments of their modified statistics. However, these corrected LRTs either need the multivariate normal distribution assumption or are not applicable if the dimension p is larger than the sample size n .

The second type is the Frobenius norm based tests—these methods are motivated by the fact that the null hypothesis $H_0 : \Sigma_1 = \Sigma_2$ is equivalent to the Frobenius norm $tr(\Sigma_1 - \Sigma_2)^2 = 0$. Various tests related to this Frobenius norm are provided. For example, Schott (2007) and Srivastava (2007b) considered two tests based on a consistent estimator of this Frobenius norm and a lower bound on this norm respectively. Srivastava and Yanagihara (2010) made some further modifications on the idea of the Frobenius norm. They also took into account the difference between $tr\Sigma_1$ and $tr\Sigma_2$ and developed their test according to a consistent estimator of $tr\Sigma_1^2/(tr\Sigma_1)^2 - tr\Sigma_2^2/(tr\Sigma_2)^2$. These tests could be used when p is larger than n , but they share a common assumption that they are also designed for the multivariate normal populations. Li and Chen (2012) removed this parametric distribution assumption by using a linear combination of three one-sample U statistics. But essentially, their statistic was also motivated by an unbiased estimator of the Frobenius norm $tr(\Sigma_1 - \Sigma_2)^2$. Unfortunately, as argued in Cai, Liu and Xia (2013), the tests based on the Frobenius norm are not powerful against sparse alternatives, i.e. the difference of two covariance matrices $(\Sigma_1 - \Sigma_2)$ is sparse under the alternative hypothesis H_1 .

Finally, to deal with sparse alternatives, Cai, Liu and Xia (2013) suggested a test based on the maximum of the standardized differences between the entries of two sample covariance matrices. They showed that it converges to a Type I extreme distribution under H_0 and that it is also robust with respect to the population distributions.

From our simulation results, we observe that when there are only several large disturbances between two population covariance matrices Σ_1 and Σ_2 (see Table 8 Model ES- $\Sigma^{(2)}$), Cai, Liu and Xia (2013)'s test is powerful but the powers of those Frobenius norm based tests decrease extremely fast when the dimension p increases. This supports the claim about the sparse alternative in Cai, Liu and Xia (2013). However, when there are many small disturbances between Σ_1 and Σ_2 (for ease of presentation, we will call it “faint alternative” later in our paper), see Table 8 Model EF- $\Sigma^{(1)}$, we see that a totally opposite phenomenon appears. The Frobenius norm based tests now show satisfactory powers but Cai, Liu and Xia (2013)'s test fails.

In view of the above facts, this paper aims at developing a test which keeps powerful for both sparse alternatives and faint alternatives in the high-dimensional setting. The resulting weighted test statistic is proposed from two aspects. On the one hand, from a macro perspective, we use the characteristic function of the empirical spectral distribution (ESD) to capture the significant difference between two population covariance matrices by noting the relationship between the population covariance matrix and the ESD. The macro-statistic proposed here is the main term and its limiting distribution enables us to conduct the hypothesis testing. To enhance the power of the statistic, on the other hand, from a micro perspective, we use the maximum of the squared differences between the entries of two sample covariance matrices as an adjustment term to capture sparse alternatives. The weighted test statistic is then proposed by putting these two perspectives in a flexible way, which is applicable when the data dimension is larger than the sample size and follows a limiting χ_2^2 distribution without specific population distribution assumptions. Simulation studies show that it enjoys satisfactory powers in both faint alternatives and sparse alternatives. What's more, we also design some alternatives between these two extreme cases, which are called "reasonable faint reasonable sparse" in the simulation and the results show that our weighted test statistic always outperforms others.

Another main contribution of this paper is to establish central limit theorems of the linear spectral statistics of two-sample by removing a widely used condition in the random matrix theory that F^Σ , the empirical distribution function of the population matrix, converges in distribution to a proper cumulative distribution function. Model 6 in the part of simulations is such an example where the corresponding empirical distribution function has no limit while the simulation results (see Table 7) confirm that central limit theorems still hold in this case.

The rest of the paper is organized as follows. Section 2 introduces the proposed statistic from the above two aspects. We establish the limiting distribution of the test statistic under the null hypothesis and discuss the power performance under the alternative. Simulations to test the approximation accuracy of our statistic's limiting distribution and several tests' performance comparison are outlined in Section 3. All the proof details are relegated to the Appendix.

2 Methodology and Theory

Suppose that $\mathbf{x}_1^0, \dots, \mathbf{x}_n^0$ are i.i.d p -dimensional random samples with mean zero and covariance matrix Σ_1 satisfying $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0) = \Sigma_1^{\frac{1}{2}}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \Sigma_1^{\frac{1}{2}}X$ and $\mathbf{y}_1^0, \dots, \mathbf{y}_n^0$ are i.i.d p -dimensional random samples with mean zero and covariance matrix Σ_2 satisfying $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0) = \Sigma_2^{\frac{1}{2}}(\mathbf{y}_1, \dots, \mathbf{y}_n) = \Sigma_2^{\frac{1}{2}}Y$ where $\Sigma_i^{\frac{1}{2}}$ denotes the Hermitian square root of the nonnegative definite Hermitian matrix Σ_i , $i = 1, 2$. The aim is to test

$$H_0 : \Sigma_1 = \Sigma_2 = \Sigma \quad \text{vs} \quad H_1 : \Sigma_1 \neq \Sigma_2. \quad (2.1)$$

Section 2.1 proposes a macro-statistic and we establish its asymptotic property by applying random matrix theory. Section 2.2 develops a micro-statistic by applying the maximum of the squared differences between the entries of two sample covariance matrices to capture sparse alternatives. The weighted test statistic that takes the foregoing two statistics into construction is reported in Section 2.3.

2.1 Macro-perspective

Throughout the paper, we impose the following three conditions.

Condition 1.

$$X = (\mathbf{x}_1, \dots, \mathbf{x}_n) = (x_{ij})_{p \times n}, \quad Y = (\mathbf{y}_1, \dots, \mathbf{y}_n) = (y_{ij})_{p \times n}, \quad 1 \leq i \leq p, \quad 1 \leq j \leq n,$$

where $\{x_{ij}\}$ are independent real random variables with mean zero, variance one and $Ex_{11}^4 = 3$ and $\{y_{ij}\}$ are also independent real random variables with mean zero, variance one and $Ey_{11}^4 = 3$. The matrices $\mathbf{X}$ and $\mathbf{Y}$ are independent.

Condition 2. $\Sigma_1 = (\sigma_{ij1})_{p \times p}$ and $\Sigma_2 = (\sigma_{ij2})_{p \times p}$ are two population covariance matrices with bounded spectral norms.

Condition 3. Let $p = p(n)$. Assume that $\frac{p}{n} \rightarrow c \in (0, \infty)$.

For two covariance matrices Σ_1 and Σ_2 , we define their respective sample covariance matrices by

$$B_{n1} = \frac{1}{n} \Sigma_1^{\frac{1}{2}} X X^T \Sigma_1^{\frac{1}{2}} = (\hat{\sigma}_{ij1})_{p \times p}, \quad B_{n2} = \frac{1}{n} \Sigma_2^{\frac{1}{2}} Y Y^T \Sigma_2^{\frac{1}{2}} = (\hat{\sigma}_{ij2})_{p \times p}. \quad (2.2)$$

Since we will use random matrix theory to construct a statistic from the macro-prospective view, we briefly introduce some relevant tools below. Given any matrix A with real eigenvalues, denote the empirical distribution function of the eigenvalues of A by F^A

$$F^A(x) = \frac{1}{p} \sum_{k=1}^p I(\lambda_k \leq x),$$

where $\lambda_k, k = 1, \dots, p$ are the eigenvalues of A . For any cumulative distribution function (c.d.f.) F , its Stieltjes transform is defined by

$$m_F(z) = \int \frac{1}{x - z} dF(x), \quad \Im z \neq 0.$$

By the inverse formula

$$F\{[x_1, x_2]\} = \frac{1}{\pi} \lim_{\epsilon \rightarrow 0+} \int_{x_1}^{x_2} \Im(m_F(x + i\epsilon)) dx,$$

where x_1 and x_2 are continuity points of F , one may recover $F(x)$ from its Stieltjes transform $m_F(z)$.

Suppose for the moment we have the following condition

$$F^{\Sigma_i} \xrightarrow{D} H_i, \tag{2.3}$$

where H_i is a proper c.d.f., then with probability 1 (see Bai and Silverstein (2004)), $F^{B_{ni}} \xrightarrow{D} F^{c, H_i}$, whose Stieltjes transform $m_{F^{c, H_i}} \equiv m_i$ is the unique solution to

$$m_i = \int \frac{1}{\lambda(1 - c - czm_i) - z} dH_i(\lambda), \quad i = 1, 2 \tag{2.4}$$

in the set $\{m \in \mathbb{C} : -\frac{1-c}{z} + cm \in \mathbb{C}^+\}$. In view of (2.3) and (2.4), one can see that if $\Sigma_1 = \Sigma_2$, then $H_1 = H_2$, which leads to $m_{F^{c, H_1}} = m_{F^{c, H_2}}$. Since the Stieltjes transform and its c.d.f. have a one-to-one correspondence, this implies that the limiting distributions of $F^{B_{n1}}$ and $F^{B_{n2}}$ are the same. This motivates us to use the difference $(F^{B_{n1}} - F^{B_{n2}})$ to capture the equality between Σ_1 and Σ_2 . However, there is no central limit theorem for $(F^{B_{n1}} - F^{B_{n2}})$, as argued by Bai and Silverstein (2004). Therefore we instead consider the corresponding characteristic functions of the ESDs by taking into account the bijection property between distribution functions and characteristic functions. Here we would like to clarify that we do not need condition (2.3) later. Writing down equality (2.4) under condition (2.3) just serves as a motivation for proposing our statistic.

The characteristic function of $F^{B_{nk}}$ is $\int e^{itx} dF^{B_{nk}}(x)$, $t \in \mathbb{R}$, $k = 1, 2$. Then the difference turns out to be

$$\int e^{itx} d(F^{B_{n1}}(x) - F^{B_{n2}}(x)). \quad (2.5)$$

Before formulating our macro-statistic, we first establish central limit theorems (CLT) of general linear spectral statistics (LSS) of the form

$$\int f(x) dp[F^{B_{n1}}(x) - F^{B_{n2}}(x)] \quad (2.6)$$

in the following Theorem 1.

Theorem 1. *Let f_1 and f_2 be functions on $\mathbb{R}$ analytic on an open interval containing*

$$[x_l, x_r] = [\liminf_p \lambda_{\min}^\Sigma I_{(0,1)}(c)(1 - \sqrt{c})^2, \limsup_p \lambda_{\max}^\Sigma (1 + \sqrt{c})^2]. \quad (2.7)$$

Suppose that f_1 and f_2 are not linearly dependent. Set

$$\varphi_n(f_k) = \int f_k(x) dp[F^{B_{n1}}(x) - F^{B_{n2}}(x)], \quad k = 1, 2, \quad (2.8)$$

$$\Sigma_n = \begin{pmatrix} \sigma_n^2(f_1) & \text{cov}_n(f_1, f_2) \\ \text{cov}_n(f_1, f_2) & \sigma_n^2(f_2) \end{pmatrix}. \quad (2.9)$$

Then under Conditions 1-3, we have under the null hypothesis H_0 ,

$$\Sigma_n^{-\frac{1}{2}} \begin{pmatrix} \varphi_n(f_1) \\ \varphi_n(f_2) \end{pmatrix} \xrightarrow{D} \mathbf{N}(\mathbf{0}, I_2), \quad (2.10)$$

where

$$\begin{aligned} \sigma_n^2(f_k) &= \sigma_{1n}^2(f_k) + \sigma_{2n}^2(f_k), \quad \text{cov}_n(f_1, f_2) = \text{cov}_{1n}(f_1, f_2) + \text{cov}_{2n}(f_1, f_2), \\ \underline{B}_{n1} &= \frac{1}{n} X^T \Sigma X, \quad \underline{B}_{n2} = \frac{1}{n} Y^T \Sigma Y, \quad \underline{m}_{1n}(z) = m_{F^{\underline{B}_{n1}}}(z), \quad \underline{m}_{2n}(z) = m_{F^{\underline{B}_{n2}}}(z), \\ \sigma_{qn}^2(f_k) &= -\frac{1}{2\pi^2} \iint \frac{f_k(z_1) f_k(z_2)}{(\underline{E} \underline{m}_{qn}(z_1) - \underline{E} \underline{m}_{qn}(z_2))^2} \underline{E} \underline{m}'_{qn}(z_1) \underline{E} \underline{m}'_{qn}(z_2) dz_1 dz_2, \quad q = 1, 2, \\ \text{cov}_{qn}(f_1, f_2) &= -\frac{1}{2\pi^2} \iint \frac{f_1(z_1) f_2(z_2)}{(\underline{E} \underline{m}_{qn}(z_1) - \underline{E} \underline{m}_{qn}(z_2))^2} \underline{E} \underline{m}'_{qn}(z_1) \underline{E} \underline{m}'_{qn}(z_2) dz_1 dz_2, \quad q = 1, 2. \end{aligned} \quad (2.11)$$

The contours in (2.11) are closed and are taken in the positive direction in the complex plane, each enclosing the interval $[x_l, x_r]$ defined in (2.7).

Remark 1. This theorem can be viewed as a generalization of CLT of Bai and Silverstein (2004) to the two-sample case. More importantly, we do not assume that the ESD of Σ converges weakly to some distribution function while such an assumption was imposed in Bai and Silverstein (2004).

Now we are in a position to propose the macro-statistic. Consider the characteristic function (2.5) and choose $f_1 = \cos(tx)$, $f_2 = \sin(tx)$ in Theorem 1. For simplicity, set

$$Re_n(t) = \int \cos(tx) dp[F^{B_{n1}}(x) - F^{B_{n2}}(x)], \quad Im_n(t) = \int \sin(tx) dp[F^{B_{n1}}(x) - F^{B_{n2}}(x)],$$

$$\Sigma_n(t) = \begin{pmatrix} \sigma_n^2(\cos(tx)) & cov_n(\cos(tx), \sin(tx)) \\ cov_n(\cos(tx), \sin(tx)) & \sigma_n^2(\sin(tx)) \end{pmatrix}, \quad (2.12)$$

where $\Sigma_n(t)$ is calculated according to equations (2.9) and (2.11). In order to use $\Sigma_n(t)$ in the following macro-statistic, we first need to verify its invertibility. It can be proved that $(E\mathbf{m}_{1n}(z) - E\mathbf{m}_{2n}(z))$ converges to zero due to $\Sigma_1 = \Sigma_2 = \Sigma$ and hence we write $E\mathbf{m}_{qn}(z)$, $q = 1, 2$ as a common $E\mathbf{m}_n(z)$. It follows that

$$\sigma_n^2(f_k) = -\frac{1}{\pi^2} \iint \frac{f_k(z_1)f_k(z_2)}{(E\mathbf{m}_n(z_1) - E\mathbf{m}_n(z_2))^2} E\mathbf{m}'_n(z_1)E\mathbf{m}'_n(z_2) dz_1 dz_2, \quad k = 1, 2,$$

$$cov_n(f_1, f_2) = -\frac{1}{\pi^2} \iint \frac{f_1(z_1)f_2(z_2)}{(E\mathbf{m}_n(z_1) - E\mathbf{m}_n(z_2))^2} E\mathbf{m}'_n(z_1)E\mathbf{m}'_n(z_2) dz_1 dz_2.$$

For any two functions f and g we define a map

$$\langle f, g \rangle = -\frac{1}{\pi^2} \iint \frac{f(z_1)g(z_2)}{(E\mathbf{m}_n(z_1) - E\mathbf{m}_n(z_2))^2} E\mathbf{m}'_n(z_1)E\mathbf{m}'_n(z_2) dz_1 dz_2, \quad (2.13)$$

where the contours here are defined the same as those in (2.11). It can be checked that this map is an inner product. Then the Cauchy-Schwarz inequality implies that

$$|\langle f, g \rangle|^2 \leq \langle f, f \rangle \cdot \langle g, g \rangle.$$

Equality holds if and only if f and g are linearly dependent. Since now we choose $f_1 = \cos(tx)$ and $f_2 = \sin(tx)$ in $\Sigma_n(t)$

$$|\langle f_1, f_2 \rangle|^2 < \langle f_1, f_1 \rangle \cdot \langle f_2, f_2 \rangle,$$

which implies that $\Sigma_n(t)$ is always invertible.

Thus we can define

$$\begin{pmatrix} \widetilde{Re}_n(t) \\ \widetilde{Im}_n(t) \end{pmatrix} = \Sigma_n^{-\frac{1}{2}}(t) \begin{pmatrix} Re_n(t) \\ Im_n(t) \end{pmatrix}. \quad (2.14)$$

Here, if a square matrix A has the eigen-decomposition $A = UDU^T$, $A^{\frac{1}{2}}$ is defined to be $A^{\frac{1}{2}} = UD^{\frac{1}{2}}U^T$. Our macro-statistic is then proposed as

$$M_n = \int_{T_1}^{T_2} (|\widetilde{Re}_n(t)|^2 + |\widetilde{Im}_n(t)|^2) dU(t), \quad (2.15)$$

where $U(t)$ is a weight function with its support on a compact interval $I = [T_1, T_2]$. The asymptotic distribution of M_n under the null hypothesis is given in the following Theorem 2.

Theorem 2. *Under the conditions given in Theorem 1, we have under the null hypothesis H_0 ,*

$$M_n \xrightarrow{D} \chi_2^2,$$

where M_n is defined in equation (2.15) and χ_2^2 means a chi-squared distribution with 2 degrees of freedom.

From our simulation results (Table 8), we can see the macro-statistic M_n performs well and even better than the tests in the literature when there are enough disturbances. However, it is not feasible for testing extreme sparse alternative and the empirical powers decrease fast when p increases, just like those Frobenius norm based tests. To make sure that the resulting weighted test statistic is also effective in the sparse alternative case, we introduce an adjustment term from another perspective.

2.2 Micro-perspective

Our macro statistic and those Frobenius norm based tests in the literature fail in the sparse alternative because essentially their efficacy depends on the sum of the difference of all entries. To capture the sparse difference, instead of using the sum, we consider the maximum entry, i.e. $\max_{1 \leq i, j \leq p} |(B_{n1} - B_{n2})_{ij}|$, where B_{n1} and B_{n2} are sample covariance matrices defined in (2.2). The key point here is that although as a whole, the sample covariance matrix is not a satisfactory estimator in the high-dimensional setting, each of its individual elements is still a consistent estimator. In view of this, we can propose the following micro-statistic

$$S_n = \max_{1 \leq i, j \leq p} (\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2})^2, \quad (2.16)$$

where $B_{n1} = (\hat{\sigma}_{ij1})_{p \times p}$, $B_{n2} = (\hat{\sigma}_{ij2})_{p \times p}$. We study the asymptotic property of S_n in Theorem 5.

Theorem 3. Under Conditions 1-3 and $Ex_{11}^8, Ey_{11}^8 < \infty$, we have under the null hypothesis H_0 ,

$$S_n = \max_{1 \leq i, j \leq p} (\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2})^2 = O_p(n^{-\frac{1}{2} + \epsilon}), \quad \forall \quad 0 < \epsilon < \frac{1}{2}. \quad (2.17)$$

Remark 2. Here ϵ can be arbitrarily small. Thus under the null hypothesis $n^\iota \cdot S_n \xrightarrow{i.p.} 0$ as long as $\iota < \frac{1}{2}$, which is used in the construction of the weighted test statistic.

We would like to point out that S_n will not necessarily converge to zero in the alternative hypothesis. To see this write

$$\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2} = (\hat{\sigma}_{ij1} - \sigma_{ij1}) - (\hat{\sigma}_{ij2} - \sigma_{ij2}) + (\sigma_{ij1} - \sigma_{ij2}).$$

This implies that

$$\max_{ij} |(\sigma_{ij1} - \sigma_{ij2})| \leq \max_{ij} |\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2}| + \max_{ij} |(\hat{\sigma}_{ij1} - \sigma_{ij1}) - (\hat{\sigma}_{ij2} - \sigma_{ij2})|.$$

By Theorem 3 $\max_{ij} |(\hat{\sigma}_{ij1} - \sigma_{ij1}) - (\hat{\sigma}_{ij2} - \sigma_{ij2})| \xrightarrow{i.p.} 0$. We then conclude that $\max_{ij} |\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2}|$ is strictly great than a positive constant as long as $\max_{ij} |(\sigma_{ij1} - \sigma_{ij2})| > 0$.

2.3 The weighted test statistic

We are now in a position to propose the weighted test statistic. Using the macro-statistic M_n as the major term and treating the micro-statistic S_n as the adjustment term, we suggest the following test statistic

$$T_n = \left(1 - \frac{1}{s_p + \zeta_1}\right) M_n + n^{\frac{1}{s_p + \zeta_1}} \cdot n^{\frac{1}{\zeta_2}} S_n, \quad (2.18)$$

where M_n and S_n are defined in (2.15) and (2.16) respectively. The parameters ζ_1, ζ_2 are two positive constants such that $\frac{1}{\zeta_1} + \frac{1}{\zeta_2} < \frac{1}{2}$. Note that under H_0 , $n^{\frac{1}{s_p + \zeta_1}} \cdot n^{\frac{1}{\zeta_2}} S_n \xrightarrow{i.p.} 0$ according to Remark 2. The variable s_p is used to measure the level of difference between Σ_1 and Σ_2 , defined as

$$s_p = \sum_{i=1}^p \sum_{j=1}^p I\{\sigma_{ij1} - \sigma_{ij2} \neq 0\}, \quad (2.19)$$

where I is the indicator function. Obviously, under H_0 , $s_p = 0$. The idea behind using such a weighted statistic is to give M_n more weight in the case of faint alternatives and give S_n more weight in the case of sparse alternatives. We would like to emphasize that we put weight $\frac{1}{s_p + \zeta_1}$ into the power of n instead of being a coefficient before S_n . In this way, power

of T_n can be significantly amplified under the alternative hypothesis. We put an additional term $n^{\frac{1}{\zeta_2}}$ here to gain more power when s_p tends to infinity in the case of sparse alternatives, in which the coefficient $n^{\frac{1}{s_p + \zeta_1}}$ would not give any weight to S_n . Here we would like to point out that in this case ($s_p \rightarrow \infty$), M_n alone already works well. After adding the micro-term to it, T_n works even better. One can refer to Section 2.4 for both theoretical support and numerical results (Table 3). The asymptotic distribution of T_n under the null hypothesis is an immediate conclusion of M_n 's and S_n 's asymptotic properties.

Theorem 4. *Let ζ_1, ζ_2 be two positive constants such that $\frac{1}{\zeta_1} + \frac{1}{\zeta_2} < \frac{1}{2}$. Given the conditions in Theorem 3, under the null hypothesis H_0 , we have*

$$T_n \xrightarrow{D} (1 - \frac{1}{\zeta_1})\chi_2^2.$$

For the practical implementation, we need to estimate unknown quantities in T_n . When calculating M_n , we estimate $E\bar{m}_{kn}(z)$ by its sample version $\bar{m}_{kn}(z) = m_{F\bar{E}_{nk}}(z)$, ($k = 1, 2$) in (2.11) and get M_n^* . Moreover, $\bar{m}_{kn}(z)$ is a uniformly consistent estimator of $E\bar{m}_{kn}(z)$ on the contours defined in Theorem 1 (one may refer to our proof of the random part in the Appendix and Bai and Silverstein (2004)). By Slutsky's theorem,

$$M_n^* \xrightarrow{D} \chi_2^2.$$

The choice of the constants ζ_1, ζ_2 will be discussed in the simulation.

In order to estimate s_p in (2.19), we use the following nice estimator from Cai, Liu and Xia (2013)

$$\hat{s}_p = \sum_{i=1}^p I\{M_{ii} \geq 2 \log p\} + \sum_{i \neq j}^p \sum_{j=1}^p I\{M_{ij} \geq 4 \log p\}, \quad (2.20)$$

where M_{ij} is defined in equation (2) in Cai, Liu and Xia (2013). Using notations in our paper, M_{ij} can be written as $M_{ij} = \frac{n(\hat{\sigma}_{ij1} - \hat{\sigma}_{ij2})^2}{\hat{\theta}_{ij1} + \hat{\theta}_{ij2}}$, where

$$\begin{aligned} \hat{\theta}_{ij1} &= \frac{1}{n} \sum_{k=1}^n [(x_{ki}^0 - \bar{x}_i^0)(x_{kj}^0 - \bar{x}_j^0) - \hat{\sigma}_{ij1}]^2, & \bar{x}_i^0 &= \frac{1}{n} \sum_{k=1}^n x_{ki}^0, \\ \hat{\theta}_{ij2} &= \frac{1}{n} \sum_{k=1}^n [(y_{ki}^0 - \bar{y}_i^0)(y_{kj}^0 - \bar{y}_j^0) - \hat{\sigma}_{ij2}]^2, & \bar{y}_i^0 &= \frac{1}{n} \sum_{k=1}^n y_{ki}^0. \end{aligned}$$

Theorem 5 in Cai, Liu and Xia (2013) tells us that $\hat{s}_p$ is consistent under H_0 and extreme sparse alternatives in our paper. We record this estimator's values in our simulations (see

Table 1). Various models considered in the simulation can be found in Section 3. Under H_0 , one can see that for different models in this paper, the estimates $\hat{s}_p$ are always close to the true one $s_p = 0$. This accuracy guarantees the theoretical null distribution in Theorem 4. On the other hand, we report the values of $\hat{s}_p$ under the extreme sparse model $\Sigma^{(2)}$, where the true value $s_p = 4$. The reason why we also pay attention to $\hat{s}_p$ under this alternative is that the micro-term is the leading term in this case. The inaccuracy of $\hat{s}_p$ (say $\hat{s}_p$ is very large) will result in a small micro-term and subsequently a poor power. The closeness between $\hat{s}_p$ and s_p ensures that the coefficient term $n^{\frac{1}{\hat{s}_p + \zeta_1}}$ before S_n is near $n^{\frac{1}{s_p + \zeta_1}}$ and thus helps to amplify the role of S_n under the extreme sparse alternative.

Table 1: Average values of $\hat{s}_p$ under H_0 for different models $\Sigma_1 = \Sigma_2 = \Sigma^{(i)}, i = 0, \dots, 5$ and under sparse alternative Model 2 $\Sigma_1 = \Sigma^{(0)}, \Sigma_2 = \Sigma^{(2)}, \alpha = 0.05, 500$ replications.

Standard errors are given in parentheses.

p	$H_0 : \Sigma_1 = \Sigma_2 = \Sigma^{(i)}, i = 0, \dots, 5$							$H_1 : \Sigma_1 = \Sigma^{(0)}, \Sigma_2 = \Sigma^{(2)}$	
	s_p	$\Sigma^{(0)}$	$\Sigma^{(1)}$	$\Sigma^{(2)}$	$\Sigma^{(3)}$	$\Sigma^{(4)}$	$\Sigma^{(5)}$	s_p	$\hat{s}_p$
40	0	0.4500 (0.8252)	0.4850 (0.7956)	0.4850 (0.8143)	0.3950 (0.6938)	0.4350 (0.7474)	0.4350 (0.8302)	4.0000	3.9050 (1.1717)
80	0	0.3450 (0.6308)	0.4300 (0.7734)	0.4400 (0.8603)	0.3800 (0.7058)	0.3550 (0.6938)	0.4850 (0.8266)	4.0000	3.6600 (1.1537)
100	0	0.3650 (0.6587)	0.4900 (0.8205)	0.3700 (0.6748)	0.5300 (0.8501)	0.4750 (0.8202)	0.3900 (0.7284)	4.0000	3.4950 (1.1819)
150	0	0.3450 (0.6843)	0.3200 (0.6321)	0.4250 (0.8293)	0.3900 (0.7003)	0.4550 (0.7881)	0.4750 (0.7431)	4.0000	3.4450 (1.0968)
200	0	0.3450 (0.7475)	0.3600 (0.6948)	0.4750 (0.7761)	0.4450 (0.8064)	0.3850 (0.7207)	0.4100 (0.7647)	4.0000	3.2350 (1.1207)
250	0	0.3350 (0.6283)	0.4500 (0.8492)	0.3700 (0.8039)	0.3200 (0.6928)	0.3300 (0.6954)	0.4200 (0.7853)	4.0000	3.2000 (1.1028)
280	0	0.3800 (0.7606)	0.4600 (0.8196)	0.3300 (0.6582)	0.3700 (0.6443)	0.4000 (0.7700)	0.4700 (0.8501)	4.0000	3.2000 (1.0751)
300	0	0.3900 (0.9011)	0.3850 (0.7413)	0.3050 (0.6356)	0.3150 (0.7673)	0.3650 (0.7379)	0.4700 (0.7760)	4.0000	3.3300 (1.2076)

Now denote

$$T_n^* = \left(1 - \frac{1}{\hat{s}_p + \zeta_1}\right) M_n^* + n^{\frac{1}{\hat{s}_p + \zeta_1}} \cdot n^{\frac{1}{\zeta_2}} S_n.$$

We can easily get Corollary 1.

Corollary 1. *Given the conditions in Theorem 4, under the null hypothesis H_0 , we have*

$$T_n^* \xrightarrow{D} \left(1 - \frac{1}{\zeta_1}\right) \chi_2^2.$$

Remark 3. M_n^* and T_n^* are the test statistics used in practice. With a bit abuse of notation, we still use M_n and T_n to indicate the macro-statistic and the weighted test statistic in our simulations.

Remark 4. One may wonder why not directly combine those Frobenius norm based tests in the literature, which also perform not bad in the faint alternative, with an adjustment term. We use the macro-statistic M_n in (2.15) due to two considerations: First M_n outperforms these Frobenius norm based tests (see Table 8). Intuitively speaking, according to the Taylor expansion of trigonometric functions $\cos(x)$ and $\sin(x)$, the macro-statistic M_n makes use of all moments while the existing Frobenius norm based tests only use the first two moments. Then it is reasonable to expect that M_n enjoys a better power performance. Secondly these existing tests require much stronger conditions than M_n does such as the multivariate normal distributions or tail assumptions.

2.4 Power discussion

Theorem 5. *In addition to the assumptions in Theorem 4, if the following condition hold under H_1 ,*

$$\limsup_{n,p \rightarrow \infty} \left| p \int e^{itx} d(EF^{B_{n1}}(x) - EF^{B_{n2}}(x)) \right| = \infty, \quad (2.21)$$

then

$$P \left\{ M_n > \chi_2^2[\alpha] \middle| H_1 \right\} \rightarrow 1 \quad \text{as } p, n \rightarrow \infty.$$

Or if the following condition hold under H_1 ,

$$\max_{ij} |\sigma_{ij1} - \sigma_{ij2}| > \varepsilon > 0, \quad (2.22)$$

then

$$P \left\{ n^{\frac{1}{s_p + \zeta_1}} \cdot n^{\frac{1}{\zeta_2}} S_n > \left(1 - \frac{1}{\zeta_1}\right) \cdot \chi_2^2[\alpha] \middle| H_1 \right\} \rightarrow 1 \quad \text{as } p, n \rightarrow \infty.$$

Subsequently, let the condition (2.21) or (2.22) hold under H_1 , then

$$P \left\{ T_n > \left(1 - \frac{1}{\zeta_1}\right) \cdot \chi_2^2[\alpha] \middle| H_1 \right\} \rightarrow 1 \quad \text{as } p, n \rightarrow \infty,$$

where the value $\chi_2^2[\alpha]$ denotes a quantile of a χ_2^2 distribution and the probability to the right of the quantile is α . i.e. if $X \sim \chi_2^2$, then $P\{X > \chi_2^2[\alpha]\} = \alpha$.

Condition (2.22) is clear and the conclusion following it is obvious according to Remark 2.

Given the condition (2.21), the proof of Theorem 5 is also straightforward since according to the proof of random parts in A.2, $R_n^{(1)}$ and $R_n^{(3)}$ still follow the central limit theorem and are bounded in probability. But the nonrandom part in A.1 now tends to infinity under H_1 .

Next let's have a look at the condition (2.21) more closely. To this end, we write

$$p \int e^{itz} d[EF^{B_{n1}}(x) - EF^{B_{n2}}(x)] = -\frac{1}{2\pi i} \oint e^{itz} p[Em_{F^{B_{n1}}}(z) - Em_{F^{B_{n2}}}(z)]dz$$

and denote

$$\begin{aligned} M_n^{(2)}(z) &= p[Em_{F^{B_{n1}}}(z) - Em_{F^{B_{n2}}}(z)] \\ &= p[Em_{F^{B_{n1}}}(z) - m_1^0(z)] + p[m_1^0(z) - m_2^0(z)] - p[Em_{F^{B_{n2}}}(z) - m_2^0(z)] \\ &= p[Em_{F^{B_{n1}}}(z) - m_1^0(z)] + n[\underline{m}_1^0(z) - \underline{m}_2^0(z)] - p[Em_{F^{B_{n2}}}(z) - m_2^0(z)]. \end{aligned} \quad (2.23)$$

Here $m_i^0(z), i = 1, 2$ is the unique solution to equation (2.4) by replacing H_i by F^{Σ_i} . i.e.

$$m_i^0(z) = \int \frac{1}{\lambda(1 - c - czm_i^0(z)) - z} dF^{\Sigma_i}(\lambda), \quad i = 1, 2. \quad (2.24)$$

And

$$\underline{m}_i^0(z) = -\frac{1 - c}{z} + cm_i^0(z). \quad (2.25)$$

We abbreviate $\underline{m}_1^0(z), \underline{m}_2^0(z)$ to $\underline{m}_1^0, \underline{m}_2^0$ for the sake of simplicity in this section. The equalities (2.24) and (2.25) imply that $\underline{m}_1^0, \underline{m}_2^0$ satisfy the equations

$$\frac{1}{\underline{m}_1^0} = -z + c \int \frac{\lambda}{1 + \lambda \underline{m}_1^0} dF^{\Sigma_1}(\lambda), \quad \frac{1}{\underline{m}_2^0} = -z + c \int \frac{\lambda}{1 + \lambda \underline{m}_2^0} dF^{\Sigma_2}(\lambda),$$

which result in

$$n(\underline{m}_1^0 - \underline{m}_2^0) \left[c \int \frac{\lambda^2 dF^{\Sigma_2}(\lambda)}{(1 + \lambda \underline{m}_1^0)(1 + \lambda \underline{m}_2^0)} - \frac{1}{\underline{m}_1^0 \underline{m}_2^0} \right] = -\frac{p}{\underline{m}_1^0} \int \frac{1}{1 + \lambda \underline{m}_1^0} d(F^{\Sigma_1}(\lambda) - F^{\Sigma_2}(\lambda)).$$

Obviously, the term $\left| c \int \frac{\lambda^2 dF^{\Sigma_2}(\lambda)}{(1 + \lambda \underline{m}_1^0)(1 + \lambda \underline{m}_2^0)} - \frac{1}{\underline{m}_1^0 \underline{m}_2^0} \right| < \infty$. Denote the right hand side as RHS , i.e.

$$RHS = -\frac{p}{\underline{m}_1^0} \int \frac{1}{1 + \lambda \underline{m}_1^0} d(F^{\Sigma_1}(\lambda) - F^{\Sigma_2}(\lambda)). \quad (2.26)$$

Then as long as $|RHS| \rightarrow \infty$, the term $|n(\underline{m}_1^0 - \underline{m}_2^0)| \rightarrow \infty$.

For easier and more concrete understanding, we now give some examples to show different cases when RHS can tend to infinity under $H_1 : \Sigma_1 \neq \Sigma_2$, and subsequently (2.21) holds.

Let $A_1(s)$ be the $s \times s$ covariance matrix of MA(1) model with the parameter $\theta_1 = 0.4$. Let $A_2(s)$ be the $s \times s$ covariance matrix of MA(2) model with the parameters $\theta_1 = \theta_2 = 0.7$. And $B = 0.7I_{p-s}$ is a $(p-s) \times (p-s)$ matrix. Define the following two sets of matrices:

$$Q_1(s) = \text{BlkDiag}(A_1(s), B), \quad Q_2(s) = \text{BlkDiag}(A_2(s), B), \quad 0 < s \leq p.$$

Here “BlkDiag” indicates a block diagonal matrix. The condition $s > 0$ corresponds to the alternative hypothesis $H_1 : Q_1(s) \neq Q_2(s)$. It is well known that the ESDs of the covariance matrices of MA models have limits. If their limiting spectral distributions are different, i.e. $F^{Q_1(s)}(\lambda)$ and $F^{Q_2(s)}(\lambda)$ have different limits, obviously $|RHS| \rightarrow \infty$ from (2.26). However, it is possible that $F^{Q_1(s)}(\lambda)$ and $F^{Q_2(s)}(\lambda)$ have the same limit. For example, $s = p^r$, $0 < r < 1$. But from (2.26), we can see as long as $s \rightarrow \infty$, $|RHS| \rightarrow \infty$. Then $|n(\underline{m}_1^0 - \underline{m}_2^0)| \rightarrow \infty$. Together with the results that $p[Em_{F^{B_{nk}}}(z) - m_k(z)]$, $k = 1, 2$ are bounded in this example (see Bai and Silverstein (2004)), we can conclude that $|M_n^{(2)}(z)| \rightarrow \infty$ and thus condition (2.21) holds.

We also conduct a series of simulations to investigate our test's performance under these two different cases.

Case 1: If $F^{Q_1(s)}(\lambda)$ and $F^{Q_2(s)}(\lambda)$ have the same limit. We choose $s = p^{\frac{1}{4}}$, $p^{\frac{1}{4}}$ or $p^{\frac{1}{3}}$. Other parameters' settings are same as in the Section 3. The empirical sizes are recorded in Table 2 and the empirical powers are in Table 3, from which we can see that T_n always outperforms others and M_n alone already can catch enough power.

Case 2: If $F^{Q_1(s)}(\lambda)$ and $F^{Q_2(s)}(\lambda)$ have different limits. This can be realized when we choose $s = O(p)$. Our simulation results show that in this case all the powers of T_n are close to 1. We do not attach our numerical results here because these powers can be obviously expected from Table 3, where the powers are almost ones when s is only $O(\sqrt{p})$.

In summary, whenever Σ_1 and Σ_2 have different limiting spectral distributions, the condition (2.21) holds. Moreover, even if their limits are the same, (2.21) is still true as long as their eigenvalues differ in infinite positions. More specifically, in the above example, $s \rightarrow \infty$ is sufficient. As long as (2.21) holds, theoretically both M_n and T_n would enjoy powers tending to one.

Table 2: Empirical sizes when $\Sigma_1 = \Sigma_2 = Q_k(s)$, $k = 1, 2$, $s = p^{\frac{1}{4}}, p^{\frac{1}{3}}, p^{\frac{1}{2}}$, $\alpha = 0.05$, 500 replications. Standard errors are given in parentheses.

$\Sigma_1 = \Sigma_2 = Q_1(s)$							$\Sigma_1 = \Sigma_2 = Q_2(s)$					
p	s	T_n	M_n	Chen	Tony	Schott	s	T_n	M_n	Chen	Tony	Schott
40	$p^{\frac{1}{4}}$	0.0740 (1.9921)	0.0660 (1.9736)	0.0560 (1.0375)	0.0500 (2.2676)	0.0560 (1.0174)	$p^{\frac{1}{4}}$	0.0820 (2.2288)	0.0560 (2.0626)	0.0620 (1.0860)	0.0560 (2.3955)	0.0660 (1.0742)
80		0.0460 (1.7473)	0.0440 (1.7316)	0.0680 (1.0478)	0.0500 (2.3634)	0.0580 (1.0295)		0.0860 (2.1496)	0.0640 (1.9892)	0.0640 (1.0798)	0.0580 (2.5275)	0.0700 (1.0715)
100		0.0580 (1.8257)	0.0500 (1.8064)	0.0560 (0.9957)	0.0380 (2.3289)	0.0560 (1.0023)		0.0660 (1.9200)	0.0460 (1.7428)	0.0580 (1.0499)	0.0480 (2.4674)	0.0520 (1.0302)
150		0.0460 (1.7794)	0.0380 (1.7705)	0.0520 (0.9995)	0.0540 (2.3559)	0.0380 (0.9874)		0.0800 (1.8785)	0.0460 (1.7052)	0.0580 (1.0240)	0.0480 (2.4103)	0.0540 (0.9963)
200		0.0400 (1.7993)	0.0300 (1.7889)	0.0600 (0.9921)	0.0520 (2.4404)	0.0580 (0.9798)		0.0760 (1.9252)	0.0460 (1.8200)	0.0460 (1.0403)	0.0540 (2.3408)	0.0400 (1.0070)
250		0.0560 (1.9074)	0.0400 (1.8954)	0.0600 (1.0572)	0.0500 (2.3498)	0.0600 (1.0463)		0.0640 (1.9025)	0.0420 (1.7788)	0.0300 (0.9981)	0.0560 (2.4232)	0.0360 (0.9791)
280		0.0780 (2.0797)	0.0620 (2.0705)	0.0380 (0.9590)	0.0420 (2.4267)	0.0460 (0.9545)		0.0720 (1.9558)	0.0540 (1.8909)	0.0400 (0.9568)	0.0460 (2.4388)	0.0420 (0.9492)
300		0.0520 (1.8117)	0.0440 (1.8038)	0.0480 (0.9651)	0.0440 (2.3406)	0.0480 (0.9544)		0.0760 (2.0560)	0.0520 (1.9987)	0.0420 (0.9912)	0.0360 (2.2621)	0.0420 (0.9648)
40	$p^{\frac{1}{3}}$	0.0600 (1.9143)	0.0480 (1.8677)	0.0560 (1.0404)	0.0440 (2.2879)	0.0520 (1.0184)	$p^{\frac{1}{3}}$	0.0860 (2.3550)	0.0620 (2.1324)	0.0580 (1.0266)	0.0440 (2.2240)	0.0480 (1.0209)
80		0.0560 (1.9684)	0.0460 (1.9530)	0.0440 (0.9628)	0.0560 (2.4642)	0.0460 (0.9429)		0.0720 (2.2071)	0.0540 (2.0538)	0.0520 (1.0303)	0.0480 (2.3726)	0.0680 (1.0355)
100		0.0540 (1.9845)	0.0440 (1.9770)	0.0580 (0.9502)	0.0340 (2.2280)	0.0520 (0.9429)		0.0680 (1.8020)	0.0460 (1.6519)	0.0660 (1.0396)	0.0580 (2.5928)	0.0620 (1.0332)
150		0.0660 (1.8626)	0.0580 (1.8510)	0.0440 (1.0158)	0.0360 (2.3595)	0.0460 (0.9931)		0.0580 (1.8655)	0.0360 (1.7136)	0.0640 (1.0290)	0.0740 (2.6230)	0.0660 (1.0027)
200		0.0760 (1.9958)	0.0600 (1.9876)	0.0520 (1.0192)	0.0620 (2.5566)	0.0540 (1.0008)		0.0640 (1.7699)	0.0360 (1.6760)	0.0540 (1.0380)	0.0560 (2.5912)	0.0540 (1.0155)
250		0.0760 (1.8593)	0.0580 (1.8442)	0.0600 (0.9996)	0.0540 (2.4344)	0.0600 (0.9823)		0.0840 (1.9034)	0.0380 (1.7885)	0.0600 (0.9873)	0.0500 (2.3359)	0.0520 (0.9677)
280		0.0640 (1.9096)	0.0500 (1.9035)	0.0460 (0.9799)	0.0640 (2.5914)	0.0440 (0.9742)		0.0720 (1.8905)	0.0420 (1.8177)	0.0520 (1.0175)	0.0500 (2.4945)	0.0580 (1.0114)
300		0.0500 (1.7659)	0.0400 (1.7591)	0.0440 (0.9948)	0.0420 (2.4273)	0.0440 (0.9763)		0.0860 (1.9128)	0.0540 (1.8395)	0.0440 (1.0022)	0.0340 (2.3297)	0.0380 (0.9837)
40	$p^{\frac{1}{2}}$	0.0760 (1.9798)	0.0680 (1.9623)	0.0580 (1.0558)	0.0520 (2.4720)	0.0640 (1.0344)	$p^{\frac{1}{2}}$	0.0720 (2.1146)	0.0360 (1.8842)	0.0700 (1.1152)	0.0440 (2.2868)	0.0760 (1.1015)
80		0.0840 (2.1441)	0.0700 (2.1270)	0.0440 (0.9843)	0.0260 (2.1801)	0.0420 (0.9648)		0.0700 (1.7272)	0.0700 (1.5474)	0.0880 (1.1558)	0.0480 (2.4129)	0.0800 (1.1474)
100		0.0820 (2.1086)	0.0600 (2.0855)	0.0600 (1.0133)	0.0700 (2.5551)	0.0580 (1.0121)		0.0880 (2.0458)	0.0420 (1.8822)	0.0640 (1.0783)	0.0380 (2.3287)	0.0780 (1.0864)
150		0.0600 (1.8447)	0.0480 (1.8379)	0.0660 (0.9948)	0.0500 (2.5142)	0.0600 (0.9734)		0.0960 (2.0365)	0.0500 (1.9121)	0.0580 (1.0538)	0.0480 (2.4081)	0.0520 (1.0327)
200		0.0580 (1.8716)	0.0500 (1.8633)	0.0580 (1.0663)	0.0420 (2.3318)	0.0680 (1.0532)		0.0800 (1.7629)	0.0280 (1.6964)	0.0420 (0.9712)	0.0380 (2.2894)	0.0400 (0.9418)
250		0.0580 (1.9167)	0.0460 (1.9107)	0.0500 (1.0383)	0.0340 (2.1651)	0.0540 (1.0261)		0.0840 (1.8951)	0.0500 (1.8066)	0.0460 (1.0059)	0.0500 (2.3693)	0.0400 (0.9825)
280		0.0560 (1.8180)	0.0500 (1.8122)	0.0520 (1.0169)	0.0440 (2.3093)	0.0560 (0.9978)		0.0860 (1.6479)	0.0320 (1.5469)	0.0520 (1.0029)	0.0500 (2.3804)	0.0600 (0.9863)
300		0.0640 (1.8168)	0.0540 (1.8158)	0.0640 (1.0457)	0.0520 (2.6512)	0.0620 (1.0285)		0.0680 (1.6228)	0.0280 (1.5217)	0.0360 (0.9290)	0.0460 (2.3164)	0.0380 (0.9325)

Table 3: Empirical powers when comparing $\Sigma_1 = Q_1(s)$ with $\Sigma_2 = Q_2(s)$, $s = p^{\frac{1}{4}}, p^{\frac{1}{3}}, p^{\frac{1}{2}}$, $\alpha = 0.05$, 500 replications. Standard errors are given in parentheses.

s	p	40	80	100	150	200	250	280	300
$p^{\frac{1}{4}}$	T_n	0.9880 (13.2382)	0.8420 (4.9571)	0.8180 (5.5140)	0.6320 (4.3557)	0.7020 (4.8550)	0.5600 (4.0807)	0.5500 (3.6313)	0.4900 (3.7897)
	M_n	0.9860 (12.7033)	0.7260 (4.4418)	0.7320 (5.1292)	0.4580 (4.0201)	0.5340 (4.5637)	0.3420 (3.8341)	0.3180 (3.3561)	0.2760 (3.5595)
	Chen	0.9120 (2.1537)	0.7560 (1.6429)	0.6500 (1.4665)	0.4600 (1.2646)	0.5560 (1.3041)	0.4220 (1.1741)	0.4200 (1.1939)	0.3740 (1.1571)
	Tony	0.3260 (3.9766)	0.2220 (3.2866)	0.1900 (3.3982)	0.1020 (2.9992)	0.1420 (3.2213)	0.1640 (3.4399)	0.1120 (2.9929)	0.1140 (3.1687)
	Schott	0.9040 (2.3555)	0.7620 (1.7435)	0.6700 (1.5145)	0.4660 (1.2945)	0.5640 (1.3424)	0.4460 (1.1941)	0.4220 (1.2237)	0.3700 (1.1835)
$p^{\frac{1}{3}}$	T_n	0.9960 (14.8951)	0.9840 (8.1806)	0.9880 (7.1700)	0.9560 (7.2183)	0.9360 (6.1956)	0.8100 (4.9530)	0.8620 (4.8164)	0.8400 (5.0241)
	M_n	0.9940 (14.3148)	0.9640 (7.7303)	0.9780 (6.8586)	0.9080 (6.9079)	0.8080 (5.9828)	0.6340 (4.7515)	0.6460 (4.6848)	0.6140 (4.9292)
	Chen	0.8980 (2.2609)	0.8900 (1.9456)	0.9200 (1.8734)	0.8060 (1.4651)	0.8300 (1.3543)	0.7020 (1.3144)	0.7300 (1.2622)	0.7480 (1.2361)
	Tony	0.3060 (4.2498)	0.3000 (3.8509)	0.3180 (4.1103)	0.2020 (3.6571)	0.1960 (3.6245)	0.1720 (3.4808)	0.1720 (3.5013)	0.1720 (3.3484)
	Schott	0.9040 (2.4496)	0.8900 (2.1191)	0.9200 (2.0271)	0.8040 (1.5608)	0.8320 (1.4244)	0.7320 (1.3465)	0.7400 (1.3212)	0.7580 (1.2793)
$p^{\frac{1}{2}}$	T_n	1.0000 (9.9134)	1.0000 (10.8705)	1.0000 (9.3918)	1.0000 (8.2616)	1.0000 (9.9587)	1.0000 (10.4633)	1.0000 (10.0185)	1.0000 (8.2403)
	M_n	1.0000 (9.6260)	1.0000 (10.6514)	1.0000 (9.2207)	1.0000 (8.1677)	0.9980 (9.8567)	1.0000 (10.4233)	0.9980 (9.9443)	0.9940 (8.1611)
	Chen	1.0000 (2.5955)	0.9960 (2.0551)	1.0000 (1.8750)	1.0000 (1.6762)	0.9960 (1.5927)	0.9980 (1.3816)	0.9980 (1.3782)	0.9960 (1.4239)
	Tony	0.6520 (4.6651)	0.6140 (4.4371)	0.5620 (4.2565)	0.5020 (3.6810)	0.4460 (4.1157)	0.4260 (3.7099)	0.4000 (3.8993)	0.3560 (3.7899)
	Schott	1.0000 (3.1050)	0.9960 (2.3718)	1.0000 (2.1717)	1.0000 (1.8932)	0.9960 (1.7942)	0.9980 (1.5445)	1.0000 (1.5181)	0.9960 (1.5565)

3 Simulations

Simulations are carried out in this section to investigate the numerical performance of our test statistic. In subsection 3.1, we report the empirical sizes to measure the suggested χ_2^2 approximation accuracy for the macro-statistic M_n and the weighted test statistic T_n in Theorems 2 and 4 respectively. In subsection 3.2, we compare the numerical performance of our proposed test statistic with three other literature tests, which were proposed in Schott (2007), Li and Chen (2012) and Cai, Liu and Xia (2013). In the tables shown below, for ease of presentation, they are abbreviated to Schott, Chen and Tony, respectively. In fact, we also recorded the performance of the tests in Bai et al. (2009), Jiang and Yang (2013) and Srivastava and Yanagihara (2010), which are mentioned in the introduction as well. We find that the first two tests suffer from size distortion if p is close to or larger than n . Even when $p < n$, they fail to be superior to others. And the test in Srivastava and Yanagihara (2010) always performs worse than Chen. So in order to save space, we do not present them here in this paper.

Throughout the simulation, the nominal significant level is set at $\alpha = 0.05$. All empirical sizes, powers and standard errors are calculated from 500 replications. We choose the sample size $n = 100$ and change the dimension p from 40 to 300. We simply let $\zeta_1 = \zeta_2 = \zeta$ in (2.18). Theoretically speaking, as long as $\zeta > 4$, when $n, p \rightarrow \infty$, the sizes should cluster round 0.05 and the smaller ζ is, the more powerful T_n can be. However, in practice, we cannot allow the sample size and the dimension to be infinity, which forces us to select larger ζ to stabilize empirical sizes. One can see that, under H_0 , the micro term is roughly $n^{\frac{1}{\zeta_1}} \cdot n^{\frac{1}{\zeta_2}} \cdot O_p(n^{-\frac{1}{2}+\epsilon}) = O_p(n^{\frac{2}{\zeta}-\frac{1}{2}}n^\epsilon)$. Suppose $\zeta = 10$, even after ignoring the ϵ term, one can calculate that $n^{\frac{2}{\zeta}-\frac{1}{2}} = n^{-0.3} \approx 0.25$, which is far from the theoretical value zero. Keeping in mind that we should first stabilize the empirical size and then try to amplify the power, the more noticeably the better, we tried $\zeta = 12, 14, 16, 18, 20$ and find that $\zeta_1 = \zeta_2 = \zeta = 20$ is the best choice.

We would like to point out that actually we also run simulations when the nominal significant level is 0.01 and the results show that $\zeta_1 = \zeta_2 = \zeta = 10$ is enough to stabilize empirical sizes. But it is burdensome to list all simulated results as for $\alpha = 0.05$ again here. Since the magnitude of ζ (or the micro-term) has the most significant influence on the extreme faint model (see following Model 2), we show the numerical results for $\Sigma^{(2)}$ as a representative in Table 9. We can see that, as expected, when ζ becomes smaller, the powers

are larger but the significance levels are worse.

Let $H(\tau_0, \tau_1, r) = (h_{ij})_{p \times p}$, where $h_{ij} = 0$ except $h_{ii} = \tau_0, i = 1, \dots, r$ and $h_{i,i+1} = h_{i,i-1} = \tau_1, i = 1, \dots, r-1$. Here τ_0, τ_1 are used to measure the level of faint alternatives and r is used to measure the level of sparse alternative. We first introduce seven covariance matrix models that are used in the simulations.

- **Model 0:** $\Sigma^{(0)} = I_p$, the identity matrix.

- **Model 1:** (Extreme faint (EF), $\tau_0 = 0.04, \tau_1 = 0.2, r = p$).

$\Sigma^{(1)} = I_p + H(0.04, 0.2, p)$. This matrix can also be considered as the covariance matrix of MA(1) model with the parameter $\theta_1 = 0.2$, which is also used in Li and Chen (2012). It has many faint disturbances compared with $\Sigma^{(0)}$ and thus we call it extreme faint (EF) case in the power comparison Table 8.

- **Model 2:** (Extreme sparse (ES), $\tau_0 = 1, \tau_1 = 1.9, r = 2$).

$\Sigma^{(2)} = I_p + H(1, 1.9, 2)$. This model only has four large disturbances compared with $\Sigma^{(0)}$, which is related to the extreme sparse (ES) alternative.

- **Model 3:** (Reasonable faint, reasonable sparse (RFRS1), $\tau_0 = 0.5, \tau_1 = 0.5, r = p/5$).

$\Sigma^{(3)} = I_p + H(0.5, 0.5, p/5)$. We use models 3, 4 and 5 to consider the cases between the extreme faint alternative in Model 1 and the extreme sparse alternative in Model 2. So we select $\tau_0 = \tau_1 = 0.5$, which lies in the interval of the values in Models 1 and 2. The choice of r is in a similar way.

- **Model 4:** (Reasonable faint, reasonable sparse (RFRS2), $\tau_0 = 0.5, \tau_1 = 0.5, r = p/10$).

$\Sigma^{(4)} = I_p + H(0.5, 0.5, p/10)$. As in Model 3, the value of r here is still between 2 (in Model 2) and p (in Model 1), but it decreases a lot compared with Model 3.

- **Model 5:** (Reasonable faint, reasonable sparse (RFRS3), $\tau_0 = 0.5, \tau_1 = 0.5, r = 10$).

$\Sigma^{(5)} = I_p + H(0.5, 0.5, 10)$. Unlike previous two models where r changes with p , we fix r to a constant value in $\Sigma^{(5)}$.

- **Model 6:** (No limiting distribution).

$\Sigma^{(6)} = (1 + (-1)^p * 0.5) * I_p$. This model has different diagonal values for odd and even dimensions so that $F^{\Sigma^{(6)}}$ does not have a limiting cumulative distribution, which

can be used to test whether our test statistic is still applicable without assuming the existence of the limiting empirical spectral distribution.

We consider two sets of data distributions that satisfy the conditions in this paper.

- Data 1: The entries of X and Y are i.i.d variables from $N(0, 1)$.
- Data 2: The entries of X and Y are i.i.d variables from the following discrete distribution

$$\xi = \frac{1}{12}\delta_{-2} + \frac{4}{25}\delta_{-1} + \frac{13}{24}\delta_0 + \frac{16}{75}\delta_{\frac{3}{2}} + \frac{1}{600}\delta_4.$$

The discrete distribution ξ only has five possible values $-2, -1, 0, \frac{3}{2}$ and 4 , with probabilities $\frac{1}{12}, \frac{4}{25}, \frac{13}{24}, \frac{16}{75}$ and $\frac{1}{600}$ respectively. It is easy to check that ξ 's first four moments are the same as those of $N(0, 1)$ and hence ξ satisfies Condition 1.

3.1 Approximation Accuracy

Empirical sizes are used and tabulated to check the χ_2^2 approximation accuracy in Theorem 2 and Theorem 4. We choose the weight function $U(t)$ in (2.15) to be a uniform distribution and other weight functions behave similarly.

First, when data are generated from Data 1 and the common population matrices ($\Sigma_1 = \Sigma_2 = \Sigma^{(i)}, i = 0, 1, \dots, 5$) are from Models 0–5, the empirical sizes are shown in Table 4, from which we can see that the empirical sizes of M_n are around 0.05, which indicates that the χ_2^2 approximation in Theorem 2 is accurate. Some of the empirical sizes of T_n are slightly larger. This is reasonable since, theoretically speaking, the micro-term in the construction of T_n converges to zero in probability as n, p tend to infinity. While in our simulation, $n = 100$ and thus we can anticipate a nonnegligible micro-term's value, which leads to a larger empirical size. In fact, we also tried simulations when n becomes larger and the results show satisfactory empirical sizes that are all close to 0.05. In order to save space, we do not attach them any more.

Although our theory assumes large pairs of (n, p) , one may be interest in getting some idea about how small values of (n, p) need to get for the approximation to perform poorly. To this end, we consider Model 0 and Model 1 and run simulations in Table 5, where we start n from 30 and p from 20 to investigate the performance of M_n and T_n . We can see that in general, the performance gets better as (n, p) becomes larger. i.e. the empirical

sizes are closer to 0.05 and the empirical powers become larger. Other models show similar phenomenon and we no longer report them here.

Moreover, theoretically our macro-statistic M_n 's null asymptotic behavior does not depend on the parametric population distribution or the limiting empirical spectral distribution assumption. To support this numerically, we apply Data 2 with Models 0–5 to see the influence of data distributions on empirical sizes. These results are shown in Table 6. And then we apply Model 6 with Data 1 and Data 2 to check the case when $F^{\Sigma_1} = F^{\Sigma_2}$ has no limiting distribution in Table 7. We consider both odd and even values for p such that $\Sigma^{(6)}$ has a different expression. From these two tables, we can see that no matter $p/n < 1$, $p/n = 1$ or $p/n > 1$, the empirical sizes are always close to 0.05, which agrees with our theoretical conclusion.

3.2 Tests Comparison

In this section we use Data 1 to generate random samples and consider the comparison between Model 0 ($\Sigma^{(0)}$) and five different Models 1–5 ($\Sigma^{(i)}, i = 1, \dots, 5$). Empirical sizes are reported in Table 4 and empirical powers are reported in Table 8. We compare our test statistic T_n with three other tests that are abbreviated to Schott, Chen and Tony according to our notations.

One may observe that regarding to powers, when p is small, the standard errors of T_n and M_n are relatively large, however, they decrease fast as p becomes larger. Realizing the fact that the test Tony has standard errors all round 10 in the extreme sparse Model (see Table 8 ES ($\Sigma^{(2)}$)), which is the alternative that this test performs the best, we believe our tests' standard errors are acceptable.

From Table 8, we can observe that (1) for the extreme faint alternative $\Sigma^{(1)}$, both T_n and M_n show significant power superiority over all other tests. Tony's test does not work well in this case; (2) for the extreme sparse alternative $\Sigma^{(2)}$, Tony reaches the largest powers, whereas powers of Chen and Schott decrease fast when p increases. Same phenomenon appears for M_n , however, after adding the micro-statistic, T_n shows significant increase in powers and is only slightly worse than those of Tony; (3) for the reasonable faint and reasonable sparse alternatives $\Sigma^{(3)}$, $\Sigma^{(4)}$ and $\Sigma^{(5)}$, our statistic T_n always outperforms the remaining statistics. (4) as r (which is used to measure the level of sparse alternative) decreases from $\Sigma^{(3)} \rightarrow \Sigma^{(4)} \rightarrow \Sigma^{(5)} \rightarrow \Sigma^{(2)}$, which means more and more sparse, the powers'

increase from M_n to T_n becomes more and more significant. This coincides with the purpose of adding our micro-statistic.

Acknowledgement The authors would like to thank the editor, the associate editor and the referees for their constructive comments which have led to significant improvement of the paper. G. M. Pan (corresponding author) was partially supported by a MOE Tier 2 grant 2014-T2-2-060 and by a MOE Tier 1 Grant RG25/14 at the Nanyang Technological University, Singapore.

Table 4: Empirical sizes when $\Sigma_1 = \Sigma_2 = \Sigma^{(i)}, i = 0, \dots, 5, \alpha = 0.05, 500$ replications.

Standard errors are given in parentheses.

p	$\Sigma^{(i)}$	T_n	M_n	Chen	Tony	Schott	$\Sigma^{(i)}$	T_n	M_n	Chen	Tony	Schott
40	$\Sigma^{(0)}$	0.0540 (1.8228)	0.0440 (1.8117)	0.0640 (1.0761)	0.0360 (2.2671)	0.0600 (1.0522)	$\Sigma^{(1)}$	0.0620 (1.7912)	0.0480 (1.7692)	0.0480 (1.0186)	0.0540 (2.5592)	0.0540 (1.0202)
80		0.0540 (1.7280)	0.0320 (1.7199)	0.0380 (0.9688)	0.0220 (2.1221)	0.0480 (0.9532)		0.0460 (1.8718)	0.0380 (1.8621)	0.0480 (1.0094)	0.0400 (2.3160)	0.0540 (0.9918)
100		0.0400 (1.5884)	0.0240 (1.5812)	0.0580 (1.0243)	0.0560 (2.3572)	0.0560 (1.0060)		0.0580 (1.9996)	0.0480 (1.9813)	0.0680 (1.0326)	0.0540 (2.4936)	0.0640 (1.0233)
150		0.0520 (1.6327)	0.0400 (1.6250)	0.0420 (1.0142)	0.0400 (2.3956)	0.0420 (0.9908)		0.0640 (1.7951)	0.0460 (1.7841)	0.0360 (0.9955)	0.0400 (2.3668)	0.0440 (0.9911)
200		0.0640 (1.8131)	0.0440 (1.8139)	0.0500 (0.9900)	0.0380 (2.2049)	0.0520 (0.9775)		0.0540 (1.7297)	0.0360 (1.7094)	0.0400 (1.0059)	0.0600 (2.5035)	0.0360 (0.9831)
250		0.0540 (1.7626)	0.0320 (1.7565)	0.0420 (1.0379)	0.0400 (2.4045)	0.0480 (1.0265)		0.0760 (1.8559)	0.0520 (1.8518)	0.0460 (0.9938)	0.0560 (2.5257)	0.0460 (0.9844)
280		0.0720 (1.8352)	0.0460 (1.8303)	0.0480 (1.0139)	0.0420 (2.3822)	0.0560 (0.9920)		0.0620 (1.7694)	0.0400 (1.7676)	0.0760 (1.0675)	0.0500 (2.6247)	0.0540 (1.0546)
300		0.0620 (1.8751)	0.0460 (1.8737)	0.0480 (0.9991)	0.0560 (2.4162)	0.0420 (0.9851)		0.0640 (1.7436)	0.0340 (1.7384)	0.0460 (0.9869)	0.0520 (2.5236)	0.0520 (0.9696)
40	$\Sigma^{(2)}$	0.0860 (2.0903)	0.0480 (1.9203)	0.0600 (1.0648)	0.0460 (2.4811)	0.0600 (1.0422)	$\Sigma^{(3)}$	0.0640 (1.8032)	0.0400 (1.7563)	0.0440 (1.0572)	0.0340 (2.1601)	0.0480 (1.0355)
80		0.0660 (1.9924)	0.0480 (1.8806)	0.0660 (1.0464)	0.0480 (2.4776)	0.0680 (1.0297)		0.0680 (1.9343)	0.0500 (1.8915)	0.0520 (1.0731)	0.0400 (2.3191)	0.0560 (1.0492)
100		0.0660 (1.9012)	0.0440 (1.8567)	0.0660 (1.0599)	0.0720 (2.6672)	0.0600 (1.0336)		0.0720 (1.8714)	0.0420 (1.8590)	0.0580 (1.0275)	0.0580 (2.3717)	0.0640 (1.0193)
150		0.0660 (1.7445)	0.0400 (1.7133)	0.0580 (1.0170)	0.0500 (2.3720)	0.0540 (0.9992)		0.0700 (1.9436)	0.0360 (1.9226)	0.0380 (1.0091)	0.0620 (2.6725)	0.0380 (0.9815)
200		0.0720 (1.8975)	0.0340 (1.8156)	0.0440 (1.0107)	0.0400 (2.3168)	0.0500 (1.0028)		0.0780 (1.9285)	0.0400 (1.8864)	0.0400 (0.9963)	0.0400 (2.3665)	0.0440 (0.9733)
250		0.0640 (1.8927)	0.0380 (1.8617)	0.0480 (0.9978)	0.0660 (2.5341)	0.0560 (0.9690)		0.0760 (1.7064)	0.0440 (1.6680)	0.0600 (1.0678)	0.0460 (2.4396)	0.0620 (1.0535)
280		0.0600 (1.6772)	0.0340 (1.6466)	0.0580 (0.9895)	0.0540 (2.7010)	0.0580 (0.9782)		0.0920 (1.8741)	0.0500 (1.8456)	0.0500 (1.0243)	0.0420 (2.3635)	0.0520 (1.0122)
300		0.0760 (1.9907)	0.0520 (1.9667)	0.0440 (0.9826)	0.0660 (2.4821)	0.0480 (0.9646)		0.0840 (1.7980)	0.0340 (1.7596)	0.0400 (1.0266)	0.0480 (2.5644)	0.0400 (1.0019)
40	$\Sigma^{(4)}$	0.0400 (1.7215)	0.0240 (1.6737)	0.0560 (1.0726)	0.0460 (2.3261)	0.0580 (1.0495)	$\Sigma^{(5)}$	0.0760 (1.9729)	0.0580 (1.9251)	0.0540 (1.0583)	0.0580 (2.5052)	0.0540 (1.0464)
80		0.0580 (1.8313)	0.0440 (1.8074)	0.0700 (1.0336)	0.0440 (2.4189)	0.0660 (1.0345)		0.0920 (1.9698)	0.0620 (1.9619)	0.0540 (1.0388)	0.0520 (2.3286)	0.0520 (1.0183)
100		0.0660 (1.7841)	0.0460 (1.7606)	0.0600 (1.0512)	0.0300 (2.2004)	0.0540 (1.0444)		0.0820 (2.0252)	0.0640 (2.0062)	0.0500 (1.0081)	0.0480 (2.5364)	0.0480 (0.9910)
150		0.0760 (1.8769)	0.0460 (1.8635)	0.0580 (1.0954)	0.0560 (2.5915)	0.0600 (1.0747)		0.0600 (1.7137)	0.0400 (1.7046)	0.0580 (0.9823)	0.0460 (2.4583)	0.0520 (0.9692)
200		0.0860 (1.9828)	0.0540 (1.9741)	0.0480 (0.9995)	0.0280 (2.2912)	0.0380 (0.9951)		0.0700 (1.8981)	0.0560 (1.8785)	0.0520 (1.0043)	0.0500 (2.3703)	0.0480 (0.9890)
250		0.0760 (1.9288)	0.0600 (1.9021)	0.0520 (1.0485)	0.0380 (2.4374)	0.0500 (1.0103)		0.0900 (1.8637)	0.0560 (1.8504)	0.0420 (0.9804)	0.0520 (2.4034)	0.0540 (0.9688)
280		0.0760 (1.8835)	0.0500 (1.8638)	0.0480 (1.0113)	0.0380 (2.3085)	0.0360 (0.9903)		0.0680 (1.8568)	0.0580 (1.8365)	0.0440 (1.0207)	0.0460 (2.2897)	0.0400 (1.0076)
300		0.0700 (1.7099)	0.0340 (1.7004)	0.0440 (1.0196)	0.0340 (2.2338)	0.0440 (1.0026)		0.0860 (1.7823)	0.0480 (1.7561)	0.0520 (0.9637)	0.0400 (2.3842)	0.0560 (0.9409)

Table 5: Performance for small values of (n, p) , $\alpha = 0.05$, 500 replications.

Standard errors are given in parentheses.

	M_n						T_n					
p	20	40	60	80	100	120	20	40	60	80	100	120
n	Empirical Sizes: $\Sigma_1 = \Sigma_2 = \Sigma^{(0)}$											
30	0.1280 (3.0897)	0.0260 (1.5759)	0.0480 (1.7474)	0.0620 (1.9709)	0.0520 (1.9309)	0.0700 (2.1283)	0.1740 (3.1905)	0.0760 (1.6595)	0.1020 (1.7821)	0.1260 (2.0503)	0.1360 (1.9576)	0.1580 (2.1569)
50	0.0500 (1.9497)	0.0860 (2.3627)	0.0440 (1.7843)	0.0820 (2.2099)	0.0420 (1.9266)	0.0600 (1.9603)	0.0720 (2.0181)	0.1160 (2.4077)	0.0840 (1.8217)	0.1280 (2.2444)	0.0820 (1.9412)	0.0980 (1.9982)
80	0.0560 (1.9138)	0.0740 (2.2007)	0.0260 (1.6161)	0.0320 (1.5795)	0.0260 (1.6951)	0.0360 (1.6241)	0.0680 (1.9517)	0.0880 (2.2274)	0.0360 (1.6284)	0.0540 (1.5837)	0.0460 (1.6897)	0.0560 (1.6354)
100	0.0320 (1.7187)	0.0620 (1.9690)	0.0460 (1.8209)	0.0380 (1.8053)	0.0500 (1.9100)	0.0440 (1.7900)	0.0380 (1.7429)	0.0760 (1.9853)	0.0680 (1.8424)	0.0600 (1.8213)	0.0580 (1.9123)	0.0640 (1.7946)
120	0.0400 (1.7296)	0.0700 (2.1244)	0.0400 (1.8291)	0.0320 (1.8485)	0.0560 (1.9607)	0.0520 (1.9094)	0.0460 (1.7466)	0.0760 (2.1328)	0.0500 (1.8477)	0.0560 (1.8508)	0.0680 (1.9700)	0.0640 (1.9101)
150	0.0300 (1.6222)	0.0460 (1.8431)	0.0520 (1.9462)	0.0580 (2.0854)	0.0520 (1.9652)	0.0620 (2.0723)	0.0360 (1.6332)	0.0500 (1.8525)	0.0700 (1.9488)	0.0700 (2.0870)	0.0560 (1.9725)	0.0700 (2.0713)
180	0.0420 (1.6734)	0.0500 (1.8079)	0.0380 (1.8867)	0.0720 (2.1003)	0.0400 (1.9395)	0.0540 (1.9114)	0.0420 (1.6874)	0.0580 (1.8182)	0.0460 (1.8950)	0.0940 (2.1081)	0.0480 (1.9479)	0.0660 (1.9133)
n	Empirical Sizes: $\Sigma_1 = \Sigma_2 = \Sigma^{(1)}$											
30	0.1140 (2.6801)	0.0260 (1.5153)	0.0620 (2.0603)	0.0560 (2.0849)	0.0780 (2.1307)	0.0800 (2.3844)	0.1660 (2.7977)	0.0720 (1.6027)	0.1520 (2.1237)	0.1500 (2.1485)	0.1640 (2.1397)	0.2100 (2.4009)
50	0.0380 (1.8798)	0.1140 (2.7742)	0.0640 (1.8320)	0.0660 (1.9991)	0.0420 (1.8905)	0.0460 (1.8301)	0.0740 (1.9431)	0.1580 (2.8278)	0.0900 (1.8574)	0.1180 (2.0484)	0.1000 (1.9133)	0.1000 (1.8639)
80	0.0540 (1.9860)	0.0740 (2.1364)	0.0320 (1.6171)	0.0380 (1.6256)	0.0360 (1.8823)	0.0260 (1.5927)	0.0720 (2.0309)	0.0940 (2.1586)	0.0540 (1.6365)	0.0560 (1.6316)	0.0660 (1.8825)	0.0320 (1.6118)
100	0.0840 (2.2023)	0.0720 (2.0167)	0.0460 (1.7821)	0.0460 (1.9133)	0.0460 (1.7991)	0.0420 (1.7802)	0.0980 (2.2380)	0.0800 (2.0220)	0.0500 (1.8000)	0.0640 (1.9260)	0.0520 (1.8083)	0.0580 (1.7859)
120	0.0380 (1.8143)	0.0420 (1.8603)	0.0500 (1.8212)	0.0380 (1.7850)	0.0540 (1.9116)	0.0520 (2.0351)	0.0460 (1.8272)	0.0540 (1.8741)	0.0600 (1.8424)	0.0560 (1.7869)	0.0720 (1.9226)	0.0740 (2.0355)
150	0.0300 (1.6969)	0.0380 (2.0070)	0.0580 (2.0062)	0.0560 (1.9289)	0.0560 (2.0360)	0.0580 (2.0087)	0.0320 (1.7056)	0.0480 (2.0160)	0.0660 (2.0068)	0.0640 (1.9311)	0.0760 (2.0485)	0.0700 (2.0070)
180	0.0580 (1.9439)	0.0440 (1.7852)	0.0660 (1.9817)	0.0660 (2.0402)	0.0420 (1.9122)	0.0480 (1.8880)	0.0680 (1.9613)	0.0480 (1.7937)	0.0740 (1.9839)	0.0780 (2.0518)	0.0500 (1.9205)	0.0560 (1.8914)
n	Empirical Powers: $\Sigma_1 = \Sigma^{(0)}$, $\Sigma_2 = \Sigma^{(1)}$											
30	0.2520 (5.2216)	0.0760 (2.1546)	0.1700 (2.9669)	0.1700 (2.8711)	0.1720 (3.0608)	0.1660 (2.8238)	0.3400 (5.3377)	0.1640 (2.2483)	0.2920 (3.0189)	0.3140 (2.9609)	0.3060 (3.0571)	0.3000 (2.8469)
50	0.1520 (2.6961)	0.3420 (4.9754)	0.2780 (3.4870)	0.3120 (4.2122)	0.3260 (4.0013)	0.3200 (3.4075)	0.1940 (2.7679)	0.4040 (5.0386)	0.3740 (3.5275)	0.4000 (4.2430)	0.4400 (4.0187)	0.4040 (3.4204)
80	0.2540 (3.5369)	0.5020 (5.0646)	0.4740 (4.1775)	0.4560 (3.7933)	0.6360 (5.3523)	0.6440 (4.9251)	0.2940 (3.5950)	0.5740 (5.0800)	0.5420 (4.1867)	0.5320 (3.8130)	0.6960 (5.3576)	0.6980 (4.9436)
100	0.4720 (5.0880)	0.6480 (6.0930)	0.6580 (5.3184)	0.7480 (6.0987)	0.7660 (5.5324)	0.8300 (5.7472)	0.5200 (5.1262)	0.6720 (6.1135)	0.7080 (5.3414)	0.7960 (6.1142)	0.7920 (5.5429)	0.8660 (5.7503)
120	0.5460 (4.8153)	0.7520 (6.1405)	0.8060 (6.5990)	0.8640 (6.8303)	0.8620 (6.7894)	0.9380 (7.9930)	0.6000 (4.8300)	0.7860 (6.1519)	0.8240 (6.6187)	0.8740 (6.8510)	0.8960 (6.8016)	0.9520 (8.0041)
150	0.7480 (5.8086)	0.8720 (6.4742)	0.9580 (7.8308)	0.9720 (8.8625)	0.9880 (8.7937)	0.9840 (8.5211)	0.7740 (5.8282)	0.8860 (6.4872)	0.9660 (7.8474)	0.9760 (8.8792)	0.9900 (8.8109)	0.9880 (8.5316)
180	0.8220 (6.2073)	0.9640 (8.4041)	0.9880 (8.9920)	0.9900 (9.5451)	0.9980 (9.8397)	0.9920 (9.7358)	0.8480 (6.2329)	0.9680 (8.4217)	0.9900 (9.0001)	0.9940 (9.5590)	0.9980 (9.8475)	0.9940 (9.7431)

Table 6: Empirical sizes when Data 2 is used to generate random variables, $\Sigma_1 = \Sigma_2 = \Sigma^{(i)}, i = 0, \dots, 5$, $\alpha = 0.05$, 500 replications. Standard errors are given in parentheses.

	Model					
p	$\Sigma^{(0)}$	$\Sigma^{(1)}$	$\Sigma^{(2)}$	$\Sigma^{(3)}$	$\Sigma^{(4)}$	$\Sigma^{(5)}$
40	0.0560 (2.0442)	0.0580 (2.0546)	0.0500 (1.8382)	0.0500 (1.9110)	0.0740 (2.0462)	0.0460 (1.8797)
80	0.0460 (1.7888)	0.0440 (1.8186)	0.0520 (1.7330)	0.0500 (1.7652)	0.0380 (1.7170)	0.0700 (1.9920)
100	0.0460 (1.8526)	0.0620 (1.9120)	0.0500 (1.8135)	0.0460 (1.8047)	0.0560 (1.8108)	0.0360 (1.9166)
150	0.0780 (2.0965)	0.0640 (2.0412)	0.0580 (1.8511)	0.0660 (2.0291)	0.0780 (2.0869)	0.0380 (1.7471)
200	0.0600 (2.0453)	0.0580 (1.9698)	0.0500 (1.8871)	0.0740 (1.9305)	0.0540 (1.8898)	0.0720 (2.0941)
250	0.0400 (1.8568)	0.0460 (1.7868)	0.0500 (1.9156)	0.0320 (1.6646)	0.0460 (1.7398)	0.0700 (2.0409)
280	0.0680 (2.0142)	0.0680 (2.0216)	0.0680 (1.9591)	0.0520 (2.0004)	0.0560 (1.9606)	0.0540 (1.9024)
300	0.0520 (1.9802)	0.0420 (1.8099)	0.0520 (1.8700)	0.0620 (1.8985)	0.0620 (1.8987)	0.0400 (1.6792)

Table 7: Empirical sizes when Model 6 is used as the common covariance matrix, $\Sigma_1 = \Sigma_2 = \Sigma^{(6)}$, $\alpha = 0.05$, 500 replications. Standard errors are given in parentheses.

	p						
Data	40	75	100	155	180	205	250
Data1	0.0480 (1.9439)	0.0640 (2.0356)	0.0500 (1.8951)	0.0620 (1.9237)	0.0580 (1.8647)	0.0380 (1.7433)	0.0500 (1.7123)
Data2	0.0660 (1.8910)	0.0620 (2.0795)	0.0500 (1.8159)	0.0400 (1.7202)	0.0300 (1.7373)	0.0440 (1.8141)	0.0420 (1.7035)

Table 8: Empirical powers when comparing $\Sigma^{(0)}$ with $\Sigma^{(i)}$, $i = 1, \dots, 5$, $\alpha = 0.05$, 500 replications. Standard errors are given in parentheses.

Model	p	40	80	100	150	200	250	280	300
(EF) ($\Sigma^{(1)}$)	T_n	0.6300 (5.0621)	0.7140 (4.8443)	0.8280 (5.5565)	0.8960 (6.4576)	0.9140 (6.3336)	0.9360 (6.4515)	0.9340 (6.5932)	0.9240 (6.9729)
	M_n	0.5740 (5.0234)	0.6480 (4.8253)	0.7620 (5.5428)	0.8660 (6.4507)	0.8860 (6.3150)	0.9140 (6.4483)	0.9020 (6.5870)	0.8860 (6.9643)
	Chen	0.5720 (1.0603)	0.5220 (0.9925)	0.6060 (1.0397)	0.6100 (1.0193)	0.6320 (0.9304)	0.6040 (0.9079)	0.5980 (0.9694)	0.6000 (0.9843)
	Tony	0.1840 (2.9730)	0.1340 (2.8639)	0.1360 (2.8862)	0.1040 (2.6634)	0.0900 (2.5681)	0.0860 (2.6900)	0.0880 (2.6384)	0.0920 (2.5624)
	Schott	0.5660 (1.0864)	0.5220 (1.0170)	0.6120 (1.0655)	0.6160 (1.0437)	0.6280 (0.9468)	0.6200 (0.9364)	0.6020 (0.9856)	0.6000 (1.0237)
(ES) ($\Sigma^{(2)}$)	T_n	0.9960 (16.0215)	0.9540 (6.9591)	0.9060 (4.5987)	0.8320 (3.6467)	0.7800 (3.0218)	0.7700 (3.0937)	0.8020 (2.7045)	0.7880 (2.7966)
	M_n	0.9800 (14.7974)	0.6540 (6.0119)	0.4060 (3.7960)	0.1620 (2.8471)	0.0880 (2.1522)	0.0960 (2.2876)	0.0920 (2.1451)	0.0780 (2.1065)
	Chen	0.9720 (1.8304)	0.7800 (1.3885)	0.6760 (1.2696)	0.3960 (1.1144)	0.2820 (1.0529)	0.2380 (1.0476)	0.1840 (1.0499)	0.1820 (1.0661)
	Tony	1.0000 (10.2594)	1.0000 (9.7961)	0.9980 (9.4946)	1.0000 (9.5904)	0.9940 (10.3418)	0.9920 (10.7164)	0.9940 (9.5167)	0.9800 (10.7966)
	Schott	0.9740 (2.0370)	0.7920 (1.4549)	0.6700 (1.3177)	0.4140 (1.1306)	0.3040 (1.0731)	0.2400 (1.0591)	0.1860 (1.0472)	0.1880 (1.0670)
(RFRS1) ($\Sigma^{(3)}$)	T_n	0.9480 (7.3511)	0.9960 (9.5364)	1.0000 (9.9126)	1.0000 (11.6220)	1.0000 (11.3727)	1.0000 (10.7691)	1.0000 (11.2679)	1.0000 (10.3879)
	M_n	0.9100 (7.1481)	0.9920 (9.4112)	1.0000 (9.7924)	1.0000 (11.5154)	1.0000 (11.3153)	1.0000 (10.7022)	1.0000 (11.2237)	1.0000 (10.3555)
	Chen	0.8440 (1.2814)	0.9000 (1.1159)	0.9060 (1.0337)	0.9180 (1.0590)	0.9400 (1.0218)	0.9400 (1.0490)	0.9160 (1.0077)	0.9440 (0.9462)
	Tony	0.4980 (3.8250)	0.4500 (3.5565)	0.4860 (3.9784)	0.4220 (3.7552)	0.4760 (3.6260)	0.3780 (3.6692)	0.3800 (3.5257)	0.3840 (3.7313)
	Schott	0.8580 (1.3475)	0.9100 (1.1653)	0.9200 (1.0857)	0.9280 (1.1036)	0.9380 (1.0876)	0.9440 (1.0907)	0.9280 (1.0519)	0.9500 (1.0009)
(RFRS2) ($\Sigma^{(4)}$)	T_n	0.4800 (4.2135)	0.7360 (5.2423)	0.8020 (5.9625)	0.9080 (6.3158)	0.9000 (6.5572)	0.9040 (6.5113)	0.9240 (5.9868)	0.9160 (6.2987)
	M_n	0.3740 (4.0246)	0.6340 (5.1323)	0.7080 (5.8633)	0.8400 (6.2302)	0.8300 (6.4907)	0.8440 (6.4300)	0.8400 (5.9476)	0.8260 (6.2474)
	Chen	0.3780 (1.1409)	0.4700 (1.1480)	0.4980 (1.0169)	0.4780 (1.0202)	0.4960 (1.0608)	0.4920 (1.0373)	0.5140 (0.9981)	0.5220 (1.0047)
	Tony	0.2900 (3.9076)	0.2900 (3.6082)	0.2860 (3.6126)	0.2940 (3.4061)	0.2480 (3.3221)	0.2640 (3.6536)	0.2580 (3.5314)	0.2160 (3.2606)
	Schott	0.3820 (1.1592)	0.4840 (1.1753)	0.5100 (1.0363)	0.4840 (1.0660)	0.5060 (1.0788)	0.4940 (1.0599)	0.5120 (1.0093)	0.5320 (1.0289)
(RFRS3) ($\Sigma^{(5)}$)	T_n	0.9960 (9.6016)	0.8780 (5.5091)	0.8060 (5.8478)	0.5600 (4.0342)	0.4660 (3.5745)	0.3460 (3.1817)	0.2580 (2.6235)	0.2300 (2.6359)
	M_n	0.9820 (9.4544)	0.8120 (5.3536)	0.7240 (5.7447)	0.3960 (3.9242)	0.3120 (3.5267)	0.2260 (3.1081)	0.1320 (2.5450)	0.1320 (2.5314)
	Chen	0.9260 (1.2702)	0.5660 (1.0586)	0.5040 (1.0399)	0.3060 (1.0779)	0.2220 (1.0519)	0.2040 (1.0513)	0.1360 (1.0009)	0.1140 (0.9598)
	Tony	0.5420 (3.9761)	0.3600 (3.6851)	0.2760 (3.5482)	0.2000 (3.4899)	0.1480 (3.2834)	0.1420 (3.2534)	0.1380 (3.1370)	0.1200 (3.0449)
	Schott	0.9340 (1.3467)	0.5780 (1.0855)	0.5100 (1.0634)	0.3080 (1.0725)	0.2260 (1.0447)	0.2080 (1.0616)	0.1500 (1.0046)	0.1160 (0.9635)

Table 9: Numerical results for Model 2 ($\Sigma^{(2)}$) under different $\zeta_1 = \zeta_2 = \zeta$ when $\alpha = 0.01$, 500 replications. Standard errors are given in parentheses.

	Empirical sizes					Empirical powers				
	T_n	M_n	Chen	Tony	Schott	T_n	M_n	Chen	Tony	Schott
p	$\zeta_1 = \zeta_2 = 20$									
40	0.0120 (1.8849)	0.0060 (1.7527)	0.0140 (1.0446)	0.0020 (2.0790)	0.0160 (1.0373)	0.9300 (7.7640)	0.7760 (6.6542)	0.9380 (1.6944)	1.0000 (10.0712)	0.9400 (1.8618)
80	0.0060 (1.9159)	0.0020 (1.8183)	0.0240 (1.0419)	0.0060 (2.4714)	0.0240 (1.0408)	0.7840 (6.5251)	0.4380 (5.7155)	0.5860 (1.3603)	0.9980 (10.2031)	0.6140 (1.4110)
100	0.0120 (1.8903)	0.0060 (1.8037)	0.0140 (1.0664)	0.0100 (2.4511)	0.0140 (1.0520)	0.7060 (4.9852)	0.2100 (4.3072)	0.3940 (1.1736)	0.9960 (10.0151)	0.3960 (1.1978)
150	0.0120 (1.8537)	0.0100 (1.8211)	0.0180 (1.0074)	0.0120 (2.3763)	0.0200 (0.9926)	0.4780 (3.3850)	0.0460 (2.7090)	0.2240 (1.1261)	0.9920 (10.3298)	0.2460 (1.1287)
200	0.0140 (1.9352)	0.0080 (1.9246)	0.0120 (0.9545)	0 (2.2711)	0.0160 (0.9553)	0.4120 (3.3140)	0.0340 (2.6462)	0.1200 (1.0494)	0.9820 (10.5755)	0.1420 (1.0572)
250	0.0100 (1.9213)	0.0060 (1.8949)	0.0060 (1.0028)	0.0200 (2.7660)	0.0060 (0.9810)	0.3320 (2.8216)	0.0240 (2.2184)	0.0920 (1.0257)	0.9840 (10.2784)	0.0880 (1.0241)
280	0.0160 (1.9355)	0.0160 (1.9258)	0.0100 (0.9981)	0.0160 (2.4823)	0.0120 (0.9849)	0.3460 (2.9970)	0.0220 (2.3677)	0.0600 (1.0110)	0.9840 (9.8260)	0.0620 (1.0036)
300	0.0080 (1.7917)	0.0020 (1.7799)	0.0060 (0.9934)	0.0080 (2.5247)	0.0100 (0.9778)	0.3200 (2.8490)	0.0140 (2.1381)	0.0440 (0.9898)	0.9620 (10.5591)	0.0540 (0.9861)
p	$\zeta_1 = \zeta_2 = 12$									
40	0.2000 (5.0337)	0.1660 (4.9074)	0.0260 (1.0543)	0.0200 (2.7160)	0.0200 (1.0265)	0.9780 (7.8072)	0.7880 (6.4078)	0.9440 (1.8070)	1.0000 (10.7521)	0.9540 (2.0012)
80	0.0080 (1.6424)	0.0060 (1.5383)	0.0120 (1.0143)	0.0060 (2.4299)	0.0160 (0.9937)	0.8600 (6.2253)	0.3480 (5.0070)	0.5980 (1.2631)	0.9940 (10.0040)	0.6260 (1.3096)
100	0.0240 (1.9309)	0.0080 (1.8422)	0.0100 (0.9584)	0.0140 (2.6772)	0.0100 (0.9543)	0.8020 (5.0384)	0.1800 (3.9209)	0.4240 (1.2361)	0.9940 (9.7737)	0.4480 (1.2768)
150	0.0180 (1.8230)	0.0140 (1.7612)	0.0080 (0.9725)	0.0120 (2.4907)	0.0040 (0.9562)	0.7040 (3.9636)	0.0740 (2.9166)	0.2040 (1.0786)	0.9840 (10.0350)	0.2280 (1.1016)
200	0.0200 (1.8413)	0.0080 (1.7893)	0.0080 (1.0255)	0.0100 (2.3281)	0.0080 (1.0040)	0.6640 (3.6006)	0.0440 (2.5606)	0.1380 (1.0846)	0.9900 (9.2716)	0.1400 (1.0941)
250	0.0040 (1.7145)	0.0020 (1.6909)	0.0080 (1.0133)	0.0140 (2.4638)	0.0060 (0.9977)	0.5820 (3.6608)	0.0360 (2.4664)	0.0740 (1.0970)	0.9800 (10.3587)	0.0780 (1.1030)
280	0.0160 (1.7408)	0.0040 (1.6741)	0.0100 (1.0313)	0.0060 (2.3884)	0.0120 (1.0104)	0.5980 (3.3007)	0.0160 (2.1783)	0.0500 (0.9962)	0.9780 (9.6994)	0.0580 (0.9960)
300	0.0200 (1.9044)	0.0120 (1.8395)	0.0140 (1.0005)	0.0060 (2.3052)	0.0140 (0.9862)	0.6160 (3.2729)	0.0220 (2.0108)	0.0920 (1.0912)	0.9720 (10.6678)	0.0980 (1.0976)
p	$\zeta_1 = \zeta_2 = 10$									
40	0.0220 (2.0402)	0.0120 (1.8598)	0.0140 (1.0033)	0.0080 (2.4141)	0.0140 (0.9975)	0.9660 (7.8646)	0.7880 (5.9344)	0.9280 (1.8703)	1.0000 (10.7234)	0.9460 (2.0616)
80	0.0180 (1.6828)	0.0040 (1.5182)	0.0160 (1.0783)	0.0040 (2.2794)	0.0180 (1.0525)	0.9140 (6.6272)	0.3940 (5.0173)	0.5680 (1.3879)	0.9920 (10.3642)	0.5840 (1.4569)
100	0.0160 (1.7597)	0.0040 (1.5947)	0.0020 (0.9683)	0.0100 (2.3336)	0.0020 (0.9569)	0.8620 (5.3713)	0.2400 (4.1762)	0.4500 (1.2244)	0.9900 (10.2167)	0.4480 (1.2716)
150	0.0220 (1.8568)	0.0060 (1.7288)	0.0160 (1.0464)	0.0120 (2.3758)	0.0160 (1.0418)	0.7960 (3.8717)	0.0680 (2.8037)	0.1980 (1.0824)	0.9900 (10.4795)	0.2200 (1.0982)
200	0.0160 (1.7284)	0.0020 (1.6433)	0.0080 (1.0129)	0.0060 (2.2976)	0.0120 (1.0017)	0.7300 (4.1102)	0.0420 (2.6573)	0.1240 (1.1057)	0.9720 (10.2634)	0.1360 (1.1218)
250	0.0160 (1.8451)	0.0080 (1.8232)	0.0060 (0.9953)	0.0060 (2.5552)	0.0080 (0.9808)	0.7520 (3.4901)	0.0300 (2.3560)	0.0860 (1.1002)	0.9760 (9.9409)	0.0920 (1.1091)
280	0.0120 (1.6638)	0 (1.6216)	0.0080 (0.9821)	0.0100 (2.3906)	0.0080 (0.9639)	0.7140 (3.7477)	0.0380 (2.3015)	0.0700 (1.0232)	0.9820 (10.3535)	0.0760 (1.0325)
300	0.0200 (1.8469)	0.0080 (1.7694)	0.0080 (1.0107)	0.0140 (2.4371)	0.0120 (0.9938)	0.7320 (3.2906)	0.0200 (2.0287)	0.0540 (0.9671)	0.9780 (9.4514)	0.0600 (0.9775)

References

- T.W. Anderson (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed.), New York: Wiley-Interscience.
- Z.D. Bai, J.W. Silverstein (2004). CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices. *The Annals of Probability*, 32, 553-605.
- Z.D. Bai, D.D. Jiang, J.F. Yao, S.R. Zheng (2009). Corrections to LRT on Large-Dimensional Covariance Matrix by RMT. *The Annals of Statistics*, 37, 3822-3840.
- M.S. Bartlett (1937). Properties and sufficiency and statistical tests. *Proc. R. Soc. Lond. A*, 160, 268-282.
- P.J. Bickel, E. Levina (2008). Covariance Regularization by Thresholding. *The Annals of Statistics*, 36, 2577-2604.
- P. Billingsley (1968). *Convergence of probability measures*, John Wiley, New York 228, 229.
- G.E.P. Box (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36, 317-346.
- T. Cai, W.D. Liu, Y. Xia (2013). Two-Sample Covariance Matrix Testing and Support Recovery in High-Dimensional and Sparse Settings. *Journal of the American Statistical Association*, 108, 265-277.
- C.H. Guo, N.J. Higham (2006). A Schur-Newton method for the matrix p -th root and its inverse, *SIAM Journal on Matrix Analysis and Applications*, 28(3).
- A.K. Gupta, J. Tang (1984). Distribution of Likelihood Ratio Statistic for Testing Equality of Covariance Matrices of Multivariate Gaussian Models. *Biometrika*, 71, 555-559.
- D.S. Gupta, N. Giri (1973). Properties of Tests Concerning Covariance Matrices of Normal Distributions. *The Annals of Statistics*, 6, 1222-1224.
- D. Jiang, T. Jiang, F. Yang (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Stat. Plann. Inference*, 142, 2241-2256.

- T.F. Jiang, F. Yang (2013). Central Limit Theorems for Classical Likelihood Ratio Tests for High-Dimensional Normal Distributions. *The Annals of Statistics*, 41, 2029-2074.
- J. Li, S.X. Chen (2012). Two Sample Tests for High-dimensional Covariance Matrices. *The Annals of Statistics*, 40, 908-940.
- A. Lytova, L. Pastur (2009). Central Limit Theorem for Linear Eigenvalue Statistics of Random Matrices with Independent Entries. *The Annals of Probability*, 37, 1778-1840.
- R.J. Muirhead (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- P.C. O'Brien (1992). Robust Procedures for Testing Equality of Covariance Matrices. *Biometrics*, 48, 819-827.
- O. Ledoit, M. Wolf (2004). A well conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88, 365-411.
- M.D. Perlman (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.
- A.J. Rothman, E. Levina, J. Zhu (2009). Generalized Thresholding of Large Covariance Matrices. *Journal of the American Statistical Association*, 104, 177-186.
- J.R. Schott (2007). A test for the equality of covariance matrices when the dimension is large relative to the same sizes. *Computational Statistics and Data Analysis*, 51, 6535-6542.
- J.W. Silverstein (1995). Strong Convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55, 331-339.
- M.S. Srivastava (2007a). Multivariate theory for analyzing high-dimensional data. *J. Japan Statist. Soc.*, 37, 53-86.
- M.S. Srivastava (2007b). Testing the equality of two covariance matrices and independence of two sub-vectors with fewer observations than the dimension, International Conference on Advances in Interdisciplinary Statistics and Combinatorics, Oct 12-14, 2007, University of North Carolina at Greensboro, 2007, NC, USA.

- M.S. Srivastava, H. Yannagihara (2010). Testing the Equality of Several Covariance Matrices with Fewer Observations than the Dimension. *Journal of Multivariate Analysis*, 101, 1319-1329.
- N. Sugiura, H. Nagao (1968). Unbiasedness of Some Test Criteria for the Equality of One or Two Covariance Matrices. *The Annals of Mathematical Statistics*, 39, 1682-1692.
- S.S. Wilks (1932). Certain generalizations in the analysis of variance. *Biometrika*, 24, 471-494.