The Annals of Statistics
2019, Vol. 47, No. 5, 2887–2921
https://doi.org/10.1214/18-AOS1768

TEST FOR HIGH-DIMENSIONAL CORRELATION MATRICES

BY SHURONG ZHENG*,[1], GUANGHUI CHENG†, JIANHUA GUO*,[2] AND HONGTU ZHU‡,[3]

Northeast Normal University, Guangzhou University† and University of North Carolina, Chapel Hill‡*

Testing correlation structures has attracted extensive attention in the literature due to both its importance in real applications and several major theoretical challenges. The aim of this paper is to develop a general framework of testing correlation structures for the one , two and multiple sample testing problems under a high-dimensional setting when both the sample size and data dimension go to infinity. Our test statistics are designed to deal with both the dense and sparse alternatives. We systematically investigate the asymptotic null distribution, power function and unbiasedness of each test statistic. Theoretically, we make great efforts to deal with the nonindependency of all random matrices of the sample correlation matrices. We use simulation studies and real data analysis to illustrate the versatility and practicability of our test statistics.

1. Introduction. Consider random samples obtained from K independent populations. Let $\mathbf{z}^{(\ell)}$ be a p-dimensional random vector for $\ell = 1, \dots, K$. We denote $\mathbf{z}_1^{(\ell)}, \dots, \mathbf{z}_{n_\ell}^{(\ell)}$ to be the n_ℓ independent samples of $\mathbf{z}^{(\ell)}$ for the ℓth population and $\bar{\mathbf{z}}^{(\ell)} = (\bar{z}_1^{(\ell)}, \dots, \bar{z}_p^{(\ell)})^{\mathrm{T}} = n_\ell^{-1} \sum_{i=1}^{n_\ell} \mathbf{z}_i^{(\ell)}$ as its sample mean. Then the sample covariance matrix and sample correlation matrix of $\{\mathbf{z}_i^{(\ell)} = (z_{1i}^{(\ell)}, \dots, z_{pi}^{(\ell)})^{\mathrm{T}} : i =$

Received May 2017; revised July 2018.

[1]Supported by National Natural Science Foundation of China (NNSFC) Grants 11522105 and 11690012.

[2]Supported by NNSFC Grants 11690012 and 11631003.

[3]Supported in part by NIH Grants MH086633 and MH116527, NSF Grants SES-1357666 and DMS-1407655, Cancer Prevention Research Institute of Texas Grant RR150054, and the endowed Bao-Shan Jing Professorship in Diagnostic Imaging.

[4]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

MSC2010 subject classifications. Primary 62H15; secondary 62H10.

Key words and phrases. Dense alternatives, global testing, sample correlation matrices, sparse alternatives.

$1, \dots, n_\ell\}$ are, respectively, given by

$$\mathbf{S}_\ell = (n_\ell - 1)^{-1} \sum_{i=1}^{n_\ell} (\mathbf{z}_i^{(\ell)} - \bar{\mathbf{z}}^{(\ell)})(\mathbf{z}_i^{(\ell)} - \bar{\mathbf{z}}^{(\ell)})^T \quad \text{and}$$

$$\widehat{\mathbf{R}}_\ell = [\operatorname{diag}(\mathbf{S}_\ell)]^{-1/2} \mathbf{S}_\ell [\operatorname{diag}(\mathbf{S}_\ell)]^{-1/2},$$

where $\operatorname{diag}(\mathbf{S}_\ell)$ is a diagonal matrix constructed from the diagonal elements of $\mathbf{S}_\ell$. There has been growing interest in the development of methods and theory for hypothesis testing on correlation structures $\{\mathbf{R}_\ell\}_{\ell=1}^K$ in different settings [Aitkin (1969), Browne (1978), Cole (1968), Gao et al. (2017), Jennrich (1970), Kullback (1967), Schott (1996), Schott (2005), Zhou, Han, Zhang and Liu (2015), Debashis and Alexander (2014)]. See, for example Anderson (2003) and Cai (2017) for overviews of statistical challenges associated with such developments.

1.1. *Existing literature.* Under the classical setting with fixed p as $\min_\ell\{n_\ell\} \to \infty$, there are three major testing problems corresponding to $K = 1$, $K = 2$ and $K > 2$, respectively. As $K = 1$, it is *one sample testing problem* that focuses on testing $H_{01} : \mathbf{R}_1 = \mathbf{R}_*$ against $H_{A1} : \mathbf{R}_1 \neq \mathbf{R}_*$, where $\mathbf{R}_1$ is the population correlation matrix and $\mathbf{R}_*$ is a specific correlation matrix. An interesting asymptotic result is that the test statistic

$$(n_1 - 1) \log(|\mathbf{R}_*|/|\widehat{\mathbf{R}}_1|) - p + \operatorname{tr}(\mathbf{R}_*^{-1}\widehat{\mathbf{R}}_1)$$

is asymptotically distributed as a linear form in $0.5p(p-1)$ independent χ_1^2 random variables, and not in general $\chi^2_{0.5p(p-1)}$ unless $\mathbf{R}_* = \mathbf{I}_p$ [Aitkin (1969), Bartlett and Rajalakshman (1953), Kullback (1967)], where $\mathbf{I}_p$ is the $p \times p$ identity matrix. This fact shows that testing the correlation matrix is a more difficult task than testing the covariance matrix. As $K = 2$, it is *two sample testing problem* that tests $H_{02} : \mathbf{R}_1 = \mathbf{R}_2$ against $H_{A2} : \mathbf{R}_1 \neq \mathbf{R}_2$, where $\mathbf{R}_1$ and $\mathbf{R}_2$ are two population correlation matrices. Several test statistics as distances between $\widehat{\mathbf{R}}_1$ and $\widehat{\mathbf{R}}_2$ and their asymptotic distributions have been studied in the literature [Aitkin (1969), Jennrich (1970), Larntz and Perlman (1985)]. As $K > 2$, it is *multiple sample testing problem* that tests $H_{0K} : \mathbf{R}_1 = \cdots = \mathbf{R}_K$ against $H_{AK} : \text{not} H_{0K}$. Many test statistics and their asymptotic distributions have been extended from the case $K = 2$ to $K > 2$ [Browne (1978), Cole (1968), Gupta, Johnson and Nagar (2013), Kullback (1967), Schott (1996)].

Recently, ultra-high dimensional data arise from a variety of applications, including neuroimaging and genetics; that is, both p and $\min_\ell\{n_\ell\}$ converge to infinity. Testing correlation structures $\{\mathbf{R}_\ell\}_{\ell=1}^K$ in this high-dimensional setting has attracted extensive attention in the past decade due to both its importance in real applications and two major theoretical challenges, including high dimensionality and dependency [Cai (2017), Debashis and Alexander (2014)]. In this case, the test statistics developed for the classical setting either do not perform well or are no longer applicable. Therefore, under the high-dimensional setting, a collection of new testing statistics have been developed in the last few years for both the one and two population testing problems [Bodwin, Zhang and Nobel (2018), Cai (2017), Cai and Zhang (2016), Gao et al. (2017), Schott (2005), Zhou et al. (2015)]. For the one sample case, the existing results focus on the test of short-range dependence, which includes independency as a special case, since the standard random matrix theory results are not directly applicable for a composite null. Moreover, the existing testing statistics are particularly powerful under either a "sparse" alternative or a dense alternative. For instance, Zhou et al. (2015) proposed several extreme value statistics to test the equality of two large U-statistic based correlation matrices, which include the rank-based correlation matrices as special cases.

1.2. *Our contributions*. The aim of this paper is to provide a general framework of testing correlation structures $\{\mathbf{R}_\ell\}_{\ell=1}^K$ for the one, two and multiple sample testing problems as $p \to \infty$. Compared with the existing literature discussed above, we make four major contributions as follows:

(I) For the first time, we develop a set of test statistics to test correlation structures $\{\mathbf{R}_\ell\}_{\ell=1}^K$ for the one, two and multiple sample testing problems under the high-dimensional setting. Our test statistics are designed to deal with both the dense and sparse alternatives. Specifically, they are the sum or the maximum of two terms, including a term for the dense alternative and the other for the sparse alternative.

(II) We propose the test statistics for testing $H_{01} : \mathbf{R}_1 = \mathbf{R}_*$ as $K = 1$ and then derive its asymptotic distribution and power function, even when $\mathbf{R}_*$ is an arbitrary correlation matrix. We make great efforts to deal with the nonindependent elements of population random vectors during the derivation. In contrast, the existing results based on the standard random matrix theory [Bai and Silverstein (2004)] are limited to the covariance matrix or independent correlation [Gao et al. (2017), Li and Xue (2015), Qiu and Chen (2012), Shao and Zhou (2014)].

(III) Similar to testing $H_{01} : \mathbf{R}_1 = \mathbf{R}_*$, we derive the asymptotic distribution of the test statistics and stride to deal with the nonindependency of the two random matrices of the sample correlation matrices for testing $H_{02} : \mathbf{R}_1 = \mathbf{R}_2$.

(IV) To the best of our knowledge, we propose the first test statistic for testing $H_{0K} : \mathbf{R}_1 = \cdots = \mathbf{R}_K$ under the high-dimensional setting and then establish its asymptotic distribution under both H_{0K} and H_{AK} without assuming the normality. We also stride to deal with the nonindependency of all random matrices of the sample correlation matrices.

The rest of this paper is organized as follows. Section 2 focuses on the one sample problem, whereas Section 3 focuses on two- and multiple-sample testing problems. In each section, we propose the test statistics and establish its asymptotic distribution, power function and unbiasedness. Section 5 will present simulation studies. We apply the test statistics to the ADHD data sets in Section 6. All proofs are collected in the Appendices.

2. Test statistics for one sample testing problem. In this section, we focus on the one sample problem of testing $H_{01}:\mathbf{R}_1=\mathbf{R}_*$ against $H_{A1}:\mathbf{R}_1\neq\mathbf{R}_*$. This section consists of three parts. In Section 2.1, we describe two proposed test statistics. We characterize its asymptotic null distributions in Section 2.2 and its power properties in Section 2.3.

2.1. *Test statistics.* We first introduce two terms as follows:

$$L_{n,1}=\operatorname{tr}\big[(\widehat{\mathbf{R}}_1-\mathbf{R}_*)^2\big]\quad\text{and}$$
$$T_{n,1}=\max_{1\le h<j\le p} n_1(\hat{r}_{1hj}-r_{*1hj})^2\big(\hat{\theta}_{1hj}^{-1}\delta_{\{\mathbf{R}_*\neq\mathbf{I}_p\}}+\delta_{\{\mathbf{R}_*=\mathbf{I}_p\}}\big),$$

where $\hat{\theta}_{1hj}=n_1^{-1}\sum_{i=1}^{n_1}\{(z_{hi}^{(1)}-\bar{z}_h^{(1)})(z_{ji}^{(1)}-\bar{z}_j^{(1)})/(s_{1hh}s_{1jj})^{1/2}-0.5\hat{r}_{1hj}[(z_{hi}^{(1)}-\bar{z}_h^{(1)})^2/s_{1hh}+(z_{ji}^{(1)}-\bar{z}_j^{(1)})^2/s_{1jj}]\}^2$ with $\widehat{\mathbf{R}}_1=(\widehat{r}_{1hj})_{h,j=1}^p$, $\mathbf{R}_*=(r_{*1hj})_{h,j=1}^p$, $\mathbf{S}_1=(s_{1hj})_{h,j=1}^p$ and $\delta_{\{\cdot\}}$ is an indicator function. The first term $L_{n,1}$ is designed for the dense alternative, whereas $T_{n,1}$ is for the sparse alternative.

Based on $L_{n,1}$ and $T_{n,1}$, we propose a weighted test statistic $M_{n,1}$ as follows:

$$M_{n,1}=L_{n,1}+C_0\delta_{\{T_{n,1}>s^*(n_1,p)\}},\tag{2.1}$$

where the second term of $M_{n,1}$ is a hard thresholding, C_0 is a large positive number and $s^*(n_1,p)$ is a scalar threshold depending on (n_1,p). The choices of C_0 and $s^*(n_1,p)$ will be given in the following Remark 2.1. For a given significance level α, we construct the acceptance region of $M_{n,1}$ to be

$$\big\{\big(\mathbf{z}_1^{(1)},\dots,\mathbf{z}_{n_1}^{(1)}\big):(M_{n,1}-\mu_{z0})/\big[2(n_1-1)^{-1}\operatorname{tr}\big(\mathbf{R}_*^2\big)\big]\le q_{1-\alpha}\big\},\tag{2.2}$$

where $q_{1-\alpha}$ is the $(1-\alpha)100\%$ quantile of $N(0,1)$ and μ_{z0} will be specified below.

We also propose a maximum test statistic $M'_{n,1}$ as follows:

$$\begin{aligned}M'_{n,1}=\max\big\{&(L_{n,1}-\mu_{z0})/\big[2(n_1-1)^{-1}\operatorname{tr}\big(\mathbf{R}_*^2\big)\big],\\ &C'_0(T_{n,1}-4\log p+\log\log p)\big\},\end{aligned}\tag{2.3}$$

where C'_0 is a positive constant and different C'_0 represents different contributions from $L_{n,1}$ and $T_{n,1}$ to $M'_{n,1}$. For a given significance level α, we construct the acceptance region of $M'_{n,1}$ to be

$$\big\{\big(\mathbf{z}_1^{(1)},\dots,\mathbf{z}_{n_1}^{(1)}\big):M'_{n,1}\le c_\alpha\big\},$$

where c_α is a critical value and the choices of c_α and C'_0 will be given in Remark 2.1.

2.2. *Null distribution.* Our first theoretical result is to characterize the limiting null distribution of $L_{n,1}$. We introduce two assumptions that will be used later.

Assumption (a) specifies the moment assumption of $\mathbf{z}_i^{(\ell)}$. Assumption (b) specifies the ratio of the dimension of $\mathbf{z}_i^{(\ell)}$ to the sample size n_ℓ. We then introduce Assumption (a) as follows.

ASSUMPTION (a). $\mathbf{z}_i^{(\ell)}$ has the independent component structure $\mathbf{z}_i^{(\ell)} = \boldsymbol{\mu}_z^{(\ell)} + \boldsymbol{\Sigma}_\ell^{1/2}\mathbf{w}_{zi}^{(\ell)}$, where $\mathbf{w}_{zi}^{(\ell)} = (w_{z1i}^{(\ell)}, \dots, w_{zpi}^{(\ell)})^T$ with independently and identically distributed (i.i.d.) elements $w_{zji}^{(\ell)}$s', $\mathrm{E}(w_{zji}^{(\ell)}) = 0$, $\mathrm{E}[(w_{zji}^{(\ell)})^2] = 1$ and the kurtosis of $w_{zji}^{(\ell)}$ is equal to $\beta_\ell = \mathrm{E}[(w_{zji}^{(\ell)})^4] - 3$. That is, $\{w_{zji}^{(\ell)}\}$ are standardized i.i.d. random variables only requiring that the fourth moment exists. The spectral norm of $\mathbf{R}_\ell$ is bounded.

Assumption (a) imposes the independent component structure on $\mathbf{z}_i^{(\ell)}$, which has been commonly used in random matrix theory [Bai and Silverstein (2004), Chen, Zhang and Zhong (2010)]. It only requires the existence of moments until the fourth order. The identically distributed assumption is not critical for most theoretical developments below.

We state Assumption (b) as follows.

ASSUMPTION (b). The ratio of the dimension p to the sample size n_ℓ tends to a constant, that is, $p/n_\ell \to y_\ell \in (0, \infty)$.

Assumption (b) gives the convergence regime of the data dimension and the sample sizes. It assumes that the data dimension increases proportionally with the sample size, even when the limit y_ℓ can be extremely small (or large). Therefore, the data dimension may be much smaller (or greater) than the sample size.

Our first theoretical result quantifies the limiting distribution of the statistic $\mathrm{tr}[(\widehat{\mathbf{R}}_1 - \mathbf{R}_*)^2]$. Let $\xrightarrow{L}$ denote the convergence in distribution.

THEOREM 2.1. *If Assumptions* (a)–(b) *hold for* $\ell = 1$ *and under* H_{01}, *we conclude that*:

(I.1) $(L_{n,1} - \mu_{z0})/[2(n_1 - 1)^{-1}\,\mathrm{tr}(\mathbf{R}_*^2)] \xrightarrow{L} N(0, 1)$;

(I.2) $(M_{n,1} - \mu_{z0})/[2(n_1 - 1)^{-1}\,\mathrm{tr}(\mathbf{R}_*^2)] \xrightarrow{L} N(0, 1)$ *holds under some additional conditions, including* $s^*(n_1, p) - 4\log p \to +\infty$ *and* (C1), (C2) *and* (C3) *in Cai, Liu and Xia* (2013) *for* $\mathbf{R}_*$ *and* $\mathbf{z}_i^{(1)}$, *where* μ_{z0} *is defined as*

$$\frac{(n_1^2 - n_1 - 1)p^2}{n_1(n_1 - 1)^2} - \frac{2n_1^2 + n_1 + 1}{(n_1 - 1)^3}\,\mathrm{tr}(\mathbf{R}_*^2) + \frac{(n_1^2 - 3n_1)}{(n_1 - 1)^3}\sum_{j=1}^{p}\sum_{j'=1}^{p}(r_{*1jj'})^4$$

$$+ \frac{n_1 p \beta_1}{(n_1 - 1)^2} - \frac{2n_1\beta_1}{(n_1 - 1)^2}\sum_{j=1}^{p}\sum_{j'=1}^{p} r_{*\frac{3}{2}jj'}(r_{*\frac{1}{2}jj'})^3$$

$$
+ \frac{(n_1^2 - 5n_1)\beta_1}{2(n_1 - 1)^3} \sum_{j=1}^{p} \left[r_{*2jj} \sum_{j=1}^{p} (r_{*\frac{1}{2}jj'})^4 \right]
$$

$$
+ \frac{(n_1^2 - 3n_1)\beta_1}{2(n_1 - 1)^3} \sum_{j=1}^{p} \sum_{j'=1}^{p} \left[(r_{*1jj'})^2 \sum_{h=1}^{p} (r_{*\frac{1}{2}hj})^2 (r_{*\frac{1}{2}hj'})^2 \right]
$$

with r_{*khj} *being the* (h, j) *entry of* $\mathbf{R}_*^k$ *for* $k = 1/2, 1, 3/2$ *and* 2.

REMARK 2.1. When $\mathbf{R}_* = \mathbf{I}_p$, Cai and Jiang (2011) proved that $\max_{1 \le h < j \le p} n_1(\hat{r}_{1hj} - r_{*1hj})^2 - 4\log p + \log\log p$ converged to a type I extreme value distribution function $F(t) = \exp[-(8\pi)^{-1/2}\exp(-t/2)]$ under H_{01}. When $\mathbf{R}_* \neq \mathbf{I}_p$, similar to (22) of Cai and Zhang (2016), we conclude that $\max_{1 \le h < j \le p} n_1(\hat{r}_{1hj} - r_{*1hj})^2 \hat{\theta}_{1hj}^{-1} - 4\log p + \log\log p$ converges to the type I extreme value distribution function under H_{01} and (C1), (C2) and (C3) in Cai, Liu and Xia (2013) for $\mathbf{R}_*$ and $\mathbf{z}_i^{(1)}$. The choices of $s^*(n_1, p)$, C_0, C_0' and c_r are given as follows:

- *Choice of the threshold* $s^*(n_1, p)$: The test statistic $M_{n,1}$ mainly targets at $L_{n,1}$. For simplicity, the threshold is taken to be

$$
s^*(n_1, p) = [4 + (\log\log n_1 - 1)^2](\log p - 0.25\log\log p) + u_0,
$$

  where u_0 satisfies $\exp[-(8\pi)^{-1/2}\exp(-u_0/2)] = 0.99$. The threshold ensures that even if n_1 and p are small, the probability of the event $\{T_{n,1} > s^*(n_1, p)\}$ is bounded by 0.01 under H_{01}. The probability of the event $\{T_{n,1} > s^*(n_1, p)\}$ becomes negligible under H_{01} when either n_1 or p is relatively large.
- *Choice of the constant* C_0: The role of C_0 is to ensure that the second term of $M_{n,1}$ acts as the main term in $M_{n,1}$ when $T_{n,1} > s^*(n_1, p)$. It is enough to require that $C_0/[2(n_1 - 1)^{-1}\operatorname{tr}(\mathbf{R}_*^2)]$ is far away from $q_{1-\alpha}$. For simplicity, let C_0 be p^2 throughout this paper.
- *Choice of the constant* C_0' *and the critical value* c_α: Theorem 2.1 shows that $(L_{n,1} - \mu_{z0})/[2(n_1 - 1)^{-1}\operatorname{tr}(\mathbf{R}_*^2)]$ is asymptotically distributed as $N(0, 1)$ under H_{01}. To balance the contribution of $L_{n,1}$ and that of $T_{n,1}$, C_0' should be relatively small for extremely dense $\mathbf{R}_1 - \mathbf{R}_*$, whereas C_0' should be large for extremely sparse $\mathbf{R}_1 - \mathbf{R}_*$. However, it is unknown whether $\mathbf{R}_1 - \mathbf{R}_*$ is dense or sparse, so we choose C_0' such that $(L_{n,1} - \mu_{z0})/[2(n_1 - 1)^{-1}\operatorname{tr}(\mathbf{R}_*^2)]$ and $C_0'(T_{n,1} - 4\log p + \log\log p)$ have the same $(1-\alpha/2)100\%$ quantile, where α is the significance level. That is, we have $C_0' = q_{1-\alpha/2}/u_0'$ and $c_\alpha = q_{1-\alpha/2}$, where u_0' satisfies $\exp[-(8\pi)^{-1/2}\exp(-u_0'/2)] = 1 - \alpha/2$. Then we have $P(M_{n,1}' > q_{1-\alpha/2}) \le \alpha$ under H_{01}.

PROOF. We will give the skeletons of the proof of Theorem 2.1. The details of the proof are placed in Appendices. The proof proceeds in three steps.

Skeleton of Step 1. To obtain the expansion of $\mathrm{tr}[(\hat{\mathbf{R}}_1-\mathbf{R}_*)^2]$ as follows:

$$\begin{aligned}\mathrm{tr}\big[(\hat{\mathbf{R}}_1-\mathbf{R}_*)^2\big] &= \mathrm{tr}\big(\mathbf{S}_1^2\big)+\mathrm{tr}\big(\mathbf{R}_*^2\big)-2\,\mathrm{tr}(\mathbf{S}_1\mathbf{R}_*)\\ &\quad -2\,\mathrm{tr}\big\{\mathbf{S}_1^2\big[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p\big]\big\}+2\,\mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p\big]\mathbf{S}_1\mathbf{R}_*\big\}\\ &\quad +2\,\mathrm{tr}\big\{\mathbf{S}_1^2\big[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p\big]^2\big\}-1.5\,\mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p\big]^2\mathbf{S}_1\mathbf{R}_*\big\}\\ &\quad +\mathrm{tr}\big\{\mathbf{S}_1\big[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p\big]\mathbf{S}_1\big[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p\big]\big\}\\ &\quad -0.5\,\mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p\big]\mathbf{S}_1\big[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p\big]\mathbf{R}_*\big\}+o_p(1),\end{aligned}$$

we assume $\mathrm{Cov}(\mathbf{z}_1^{(1)})=\mathbf{R}_1$ without loss of generality. This step is mainly to use the Taylor expansions of $s_{1jj}^{-1}=1-(s_{1jj}-1)+(s_{1jj}-1)^2+o(p^{-1})$ and $s_{1jj}^{-1/2}=1-\frac{1}{2}(s_{1jj}-1)+\frac{3}{8}(s_{1jj}-1)^2+o(p^{-1})$ with $\mathbf{S}_1=(s_{1ij})_{i,j=1}^p$.

Skeleton of Step 2. We want to derive the limits of the following four terms: $\mathrm{tr}\{\mathbf{S}_1^2[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]^2\}$, $\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]^2\mathbf{S}_1\mathbf{R}_*\}$, $\mathrm{tr}\{\mathbf{S}_1[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]\mathbf{S}_1\times[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]\}$ and $\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]\mathbf{S}_1[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]\mathbf{R}_*\}$ in probability.

Skeleton of Step 3. We want to derive the limiting null distribution of $(\mathrm{tr}(\mathbf{S}_1^2)-\mathrm{E}\,\mathrm{tr}(\mathbf{S}_1^2),\ \mathrm{tr}(\mathbf{S}_1\mathbf{R}_*)-\mathrm{E}\,\mathrm{tr}(\mathbf{S}_1\mathbf{R}_*),\ \mathrm{tr}\{\mathbf{S}_1^2[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]\}-\mathrm{E}\,\mathrm{tr}\{\mathbf{S}_1^2[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]\},\ \mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]\mathbf{S}_1\mathbf{R}_*\}-\mathrm{E}\,\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p]\mathbf{S}_1\mathbf{R}_*)\})$. Thus by the delta method, we obtain the central limit theorem (*CLT*) of $\mathrm{tr}[(\widehat{\mathbf{R}}_1-\mathbf{R}_*)^2]$. Because these terms involve $\mathrm{diag}(\mathbf{S}_1)-\mathbf{I}_p$, we cannot directly use the random matrix theory on linear spectral statistics of $\mathbf{S}_1$ to obtain the limiting distribution of these terms. To solve the problem, we construct four martingale difference sequences to establish the CLT of these terms. Especially, the derivation of the CLT for the case $\mathbf{R}_*\neq\mathbf{I}_p$ is much more difficult than the derivation for the case $\mathbf{R}_*=\mathbf{I}_p$. □

COROLLARY 2.1. *Under the assumptions of Theorem 2.1, we have the following results*:

- *If $\mathbf{R}_*$ is the identity matrix $\mathbf{I}_p$, then μ_{z0} reduces to*

$$\mu_{z0}=\frac{(n_1^2-n_1-1)p^2}{n_1(n_1-1)^2}-\frac{(n_1^2+4n_1+1)p}{(n_1-1)^3}-\frac{3pn_1\beta_1}{(n_1-1)^3}.$$

- *If the population is Gaussian, then μ_{z0} reduces to*

$$\frac{(n_1^2-n_1-1)p^2}{n_1(n_1-1)^2}-\frac{2n_1^2+n_1+1}{(n_1-1)^3}\mathrm{tr}\big(\mathbf{R}_*^2\big)+\frac{(n_1^2-3n_1)}{(n_1-1)^3}\sum_{j=1}^p\sum_{j'=1}^p(r_{*1jj'})^4$$

where $r_{*1jj'}$ is the (j,j') entry of $\mathbf{R}_*$.

Theorem 2.1 provides a unified framework of testing $H_{01}:\mathbf{R}_1=\mathbf{R}_*$ for an arbitrary $\mathbf{R}_*$. Our test statistics account for both dense and sparse alternatives, and they work for $\mathbf{z}_i^{(1)}$ satisfying the independent component structure specified in Assumption (a) and a ratio of $p/n_1 = y_1$ in Assumption (b). Technically, to prove Theorem 2.1, we develop a set of novel tools to deal with the dependence between $\mathbf{S}_1$ and $\mathrm{diag}(\mathbf{S}_1)$, which is technically nontrivial and is of independent interest for handling the sample correlation in more general settings. In contrast, Gao et al. (2017) only established the CLT of the sample correlation matrices of a high-dimensional vector whose elements have an identity correlated structure $\mathbf{R}_* = \mathbf{I}_p$. Moreover, their theoretical result involves some two-dimensional contour integrals, which can be difficult to compute.

2.3. *Power properties and optimality.* We examine the power properties of $M_{n,1}$ and $M'_{n,1}$. We first establish the asymptotic distribution of the statistic $\mathrm{tr}[(\widehat{\mathbf{R}}_1 - \mathbf{R}_*)^2]$ under the alternative hypothesis H_{A1}.

THEOREM 2.2. *Assuming that Assumptions* (a) *and* (b) *hold for* $\ell = 1$, *we have*

$$\frac{\mathrm{tr}[(\widehat{\mathbf{R}}_1 - \mathbf{R}_*)^2] - \mu_{zA}}{\sigma_{zA}} \xrightarrow{L} N(0, 1),$$

where μ_{zA} *and* σ_{zA} *depend on the alternative population correlation matrix* $\mathbf{R}_1$ *and will be given in the Appendix.*

Given the result in Theorem 2.2, we can characterize the properties of the power functions, which is given by

$$g_1(\mathbf{R}_1, \alpha) = P\big((M_{n,1} - \mu_{z0})/\big[2(n_1 - 1)^{-1}\,\mathrm{tr}\big(\mathbf{R}_*^2\big)\big] > q_{1-\alpha}\big),$$
$$g'_1(\mathbf{R}_1, \alpha) = P\big(M'_{n,1} > c_\alpha\big).$$

In the following, we will study the properties of the power functions $g_1(\mathbf{R}_1, \alpha)$ and $g'_1(\mathbf{R}_1, \alpha)$.

COROLLARY 2.2. *Assuming that Assumptions* (a) *and* (b) *hold for* $\ell = 1$, *we have the following results*:

- *If* $\mathrm{tr}[(\mathbf{R}_1 - \mathbf{R}_*)^2] > c_0 > 0$, *then* $g_1(\mathbf{R}_1, \alpha) > \alpha$ *when the sample size* n_1 *is large enough and* c_0 *is any given small constant.*
- *If* $\mathrm{tr}[(\mathbf{R}_1 - \mathbf{R}_*)^2]$ *tends to infinity, then* $g_1(\mathbf{R}_1, \alpha)$ *and* $g'_1(\mathbf{R}_1, \alpha)$ *are close to one as* $n_1 \to \infty$.
- *If the absolute value of at least one entry of* $\mathbf{R}_1 - \mathbf{R}_*$ *is greater than* $n_1^{-1/2}\sqrt{\log(p)\log(n_1)}$ *and the conditions* (C1), (C2) *and* (C3) *in Cai, Liu and Xia* (2013) *hold for* $\mathbf{R}_1$ *and* $\mathbf{z}_i^{(1)}$, *then* $g_1(\mathbf{R}_1, \alpha)$ *and* $g'_1(\mathbf{R}_1, \alpha)$ *are close to one as* $n_1 \to \infty$.

Corollary 2.2 shows that the proposed test $M_{n,1}$ is asymptotically unbiased. In the Appendix, we will prove that (i) for the dense alternative $\mathrm{tr}[(\mathbf{R}_1-\mathbf{R}_*)^2]\to\infty$, the power functions tend to one; (ii) for the sparse alternative, if the absolute value of at least one entry of $\mathbf{R}_1-\mathbf{R}_*$ is greater than $n_1^{-1/2}\sqrt{\log(p)\log(n_1)}$, then the power functions will be close to one.

Similar to Cai and Ma (2013), we define

$$\Theta_1^*(b_1,b_{10})=\{\mathbf{R}_1:\|\mathbf{R}_1-\mathbf{R}_*\|_F>b_1\sqrt{p/n_1}\text{ or }\|\mathbf{R}_1-\mathbf{R}_*\|_\infty>b_{10}\sqrt{\log p/n_1}\},$$

where b_1, b_{10} are positive constants, $\|\mathbf{R}_1-\mathbf{R}_*\|_F=\{\mathrm{tr}[(\mathbf{R}_1-\mathbf{R}_*)^2]\}^{1/2}$ and $\|\mathbf{R}_1-\mathbf{R}_*\|_\infty=\max_{1\le i\le j\le p}|\mathbf{e}_i^T(\mathbf{R}_1-\mathbf{R}_*)\mathbf{e}_j|$ with $\mathbf{e}_i$ and $\mathbf{e}_j$ being the ith column and jth column of the $p\times p$ identity matrix, respectively.

THEOREM 2.3. *Let $0<\alpha<\beta<1$. Suppose that as $p/n_1\to y_1>0$ as $n_1\to+\infty$. Then there exist two constants $0<b_1,b_{10}<1$ such that for any test ϕ with the significance level α for testing $H_{01}:\mathbf{R}_1=\mathbf{R}_*$, we have*

$$\limsup_{n_1\to\infty}\inf_{\mathbf{R}_1\in\Theta_1^*(b_1,b_{10})}\mathrm{E}_{\mathbf{R}_1}\phi<\beta,$$

where $\mathrm{E}_{\mathbf{R}_1}$ is the expectation under the population correlation matrix being $\mathbf{R}_1$.

Theorem 2.3 shows that no level α test can distinguish the null hypothesis from the alternative hypothesis with the power tending to one as $p/n_1\to y_1>0$, $\|\mathbf{R}_1-\mathbf{R}_*\|_F=O(\sqrt{p/n_1})$ or $\|\mathbf{R}_1-\mathbf{R}_*\|_\infty>b_{10}\sqrt{\log p/n_1}$. Then Theorem 2.3 gives the lower bound for the optimality of our proposed procedure.

3. Extensions to two and multiple sample testing problems. This section consists of two parts. In Section 3.1, we focus on the two-sample problem of testing $H_{02}:\mathbf{R}_1=\mathbf{R}_2$ against $H_{A2}:\mathbf{R}_1\neq\mathbf{R}_2$. In Section 3.2, we consider the multiple sample testing problem.

3.1. *Extension to two sample testing problem.*

3.1.1. *Test statistics and their null distributions for two-sample testing problem.* Let $\widehat{\mathbf{R}}_\ell=(\hat r_{\ell hj})_{h,j=1}^p$ and $\mathbf{S}_\ell=(s_{\ell hj})_{h,j=1}^p$ for $\ell=1,2$. We introduce two terms as follows:

$$L_{n,2}=\mathrm{tr}\big[(\widehat{\mathbf{R}}_1-\widehat{\mathbf{R}}_2)^2\big]\quad\text{and}$$
$$T_{n,2}=\max_{1\le h<j\le p}\big(n_1^{-1}\hat\theta_{1hj}+n_2^{-1}\hat\theta_{2hj}\big)^{-1}(\hat r_{1hj}-\hat r_{2hj})^2,$$

where $\hat{\theta}_{\ell hj}$ is defined as

$$\hat{\theta}_{\ell hj} = n_\ell^{-1} \sum_{i=1}^{n_\ell} \biggl\{ \frac{(z_{hi}^{(\ell)} - \bar{z}_h^{(\ell)})(z_{ji}^{(\ell)} - \bar{z}_j^{(\ell)})}{(s_{\ell hh} s_{\ell jj})^{1/2}} - 0.5 \hat{r}_{\ell hj} \biggl[\frac{(z_{hi}^{(\ell)} - \bar{z}_h^{(\ell)})^2}{s_{\ell hh}} + \frac{(z_{ji}^{(\ell)} - \bar{z}_j^{(\ell)})^2}{s_{\ell jj}} \biggr] \biggr\}^2.$$

The first term $L_{n,2}$ is introduced to deal with the dense alternative, whereas the second term $T_{n,2}$ is for the sparse alternative.

We propose a weighted test statistic $M_{n,2}$ as follows:

$$M_{n,2} = L_{n,2} + C_{0,2} \delta_{\{T_{n,2} > s(n_1, n_2, p)\}}, \tag{3.1}$$

where $C_{0,2}$ and the threshold $s(n_1, n_2, p)$ will be given in Remark 3.1. For a given significance level α, we construct an acceptance region of $M_{n,2}$ to be

$$\bigl\{ \{\mathbf{z}_i^{(\ell)} : i = 1, \dots, n_\ell\}_{\ell=1}^2 : (M_{n,2} - \hat{\mu}_{z12}) / \bigl\{ 2\bigl[(n_1 - 1)^{-1} + (n_2 - 1)^{-1}\bigr] \hat{a} \bigr\} \le q_{1-\alpha} \bigr\},$$

where $\hat{\mu}_{z12}$ and $\hat{a}$ will be defined below.

We also propose a maximum test statistic $M'_{n,2}$ as follows:

$$M'_{n,2} = \max\bigl\{ (L_{n,2} - \hat{\mu}_{z12}) / \bigl\{ 2\bigl[(n_1 - 1)^{-1} + (n_2 - 1)^{-1}\bigr] \hat{a} \bigr\}, C'_{0,2}(T_{n,2} - 4 \log p + \log\log p) \bigr\}, \tag{3.2}$$

where $C'_{0,2}$ is a positive constant. For a given significance level α, we construct an acceptance region of $M'_{n,2}$ to be

$$\bigl\{ \{\mathbf{z}_i^{(\ell)} : i = 1, \dots, n_\ell\}_{\ell=1}^2 : M'_{n,2} \le c_{\alpha,2} \bigr\},$$

where the positive constant $C'_{0,2}$ and the critical value $c_{\alpha,2}$ will be given in Remark 3.1.

We establish the asymptotic null distribution of $L_{n,2}$ as follows.

THEOREM 3.1. *Let* $\mathbf{R}$ *be the common correlation matrix* $\mathbf{R} = \mathbf{R}_1 = \mathbf{R}_2$. *Assuming that Assumptions* (a) *and* (b) *hold for* $\ell = 1$ *and* 2 *and under* H_{02}, *we conclude that*:

(II.1) $(L_{n,2} - \mu_{z12}) / \{2[(n_1 - 1)^{-1} + (n_2 - 1)^{-1}] \operatorname{tr}(\mathbf{R}^2)\} \xrightarrow{L} N(0, 1)$;

(II.2) $(M_{n,2} - \mu_{z12}) / \{2[(n_1 - 1)^{-1} + (n_2 - 1)^{-1}] \operatorname{tr}(\mathbf{R}^2)\} \xrightarrow{L} N(0, 1)$ *holds under some additional conditions, including* $s(n_1, n_2, p) - 4 \log p \to +\infty$ *and*

(C1), (C2) *and* (C3) *of Cai, Liu and Xia* (2013) *for* $\mathbf{R}$ *and* $\mathbf{z}_i^{(\ell)}$, *where* μ_{z12} *is given by*

$$\mu_{z12} = \sum_{\ell=1}^{2} \frac{(n_\ell^2 - n_\ell - 1)p^2}{n_\ell(n_\ell - 1)^2} - \sum_{\ell=1}^{2} \frac{2n_\ell^2 + n_\ell + 1}{(n_\ell - 1)^3} \operatorname{tr}(\mathbf{R}^2) + \sum_{\ell=1}^{2} \frac{\beta_\ell n_\ell p}{(n_\ell - 1)^2}$$
$$- \sum_{\ell=1}^{2} \frac{\beta_\ell 2pn_\ell}{(n_\ell - 1)^2} b_0 + \sum_{\ell=1}^{2} \frac{\beta_\ell p(n_\ell^2 - 5n_\ell)}{2(n_\ell - 1)^3} c_0 + \sum_{\ell=1}^{2} \frac{p(n_\ell - 3)n_\ell}{2(n_\ell - 1)^3} d_\ell,$$

with r_{0khj} *being the* (h, j) *entry of* $\mathbf{R}^k$ *for* $k = 1/2, 1, 2/3$ *and* 2 *and*

$$b_0 = p^{-1} \sum_{j=1}^{p} \sum_{j'=1}^{p} r_{0\frac{3}{2}jj'} (r_{0\frac{1}{2}jj'})^3,$$

$$c_0 = p^{-1} \sum_{j=1}^{p} r_{02jj} \sum_{j'=1}^{p} (r_{0\frac{1}{2}jj'})^4,$$

$$d_\ell = p^{-1} \sum_{j=1}^{p} \sum_{j'=1}^{p} \left[2(r_{01jj'})^4 + \beta_\ell (r_{01jj'})^2 \sum_{h=1}^{p} (r_{0\frac{1}{2}hj})^2 (r_{0\frac{1}{2}hj'})^2 \right].$$

REMARK 3.1. Similar to (22) of Cai and Zhang (2016), we conclude that

$$\max_{1 \le h < j \le p} \left(n_1^{-1}\hat{\theta}_{1hj} + n_2^{-1}\hat{\theta}_{2hj}\right)^{-1} (\hat{r}_{1hj} - \hat{r}_{2hj})^2 - 4\log p + \log\log(p)$$

converges to the type I extreme value distribution function under H_{02} and (C1), (C2) and (C3) in Cai, Liu and Xia (2013) for $\mathbf{R}$ and $\mathbf{z}_i^{(\ell)}$. The choices of $s(n_1, n_2, p)$, $C_{0,2}$, $C'_{0,2}$ and $c_{\alpha,2}$ are given as follows:

- *Choice of the threshold* $s(n_1, n_2, p)$*:* The test statistic $M_{n,2}$ mainly targets at $L_{n,2}$. For simplicity, we set $s(n_1, n_2, p)$ as

  $$s(n_1, n_2, p) = \left[4 + \left(\log\log(n_1 + n_2) - 1\right)^2\right](\log p - 0.25 \log\log p) + u'_0,$$

  where u'_0 satisfies $\exp[-(8\pi)^{-1/2} \exp(-u'_0/2)] = 0.99$. The threshold ensures that even for small n_1, n_2 and p, the probability of the event $\{T_{n,2} > s(n_1, n_2, p)\}$ is bounded by 0.01 under H_{02}. The probability of the event $\{T_{n,2} > s(n_1, n_2, p)\}$ becomes negligible under H_{02} when either n_1, n_2 or p is moderately large.
- *Choice of* $C_{0,2}$, $C'_{0,2}$ *and* $c_{\alpha,2}$*:* The constants $C_{0,2}$, $C'_{0,2}$ and $c_{\alpha,2}$ are the same as C_0, C'_0 and c_α in Remark 2.1. Moreover, $P(M'_{n,2} > q_{1-\alpha/2}) \le \alpha$ under H_{02}.

PROOF OF THEOREM 3.1. We will give the skeletons of the proof of Theorem 3.1. The details of the proof are placed in the Appendices. The proof proceeds in three steps.

Skeleton of Step 1. It is assumed that $\mathrm{Cov}(\mathbf{z}_i^{(\ell)}) = \mathbf{R}_\ell$ holds for $\ell = 1, 2$ without loss of generality. We obtain the expansion of $\mathrm{tr}[(\widehat{\mathbf{R}}_1 - \widehat{\mathbf{R}}_2)^2]$ as follows:

$$\begin{aligned}
&\mathrm{tr}\big[(\widehat{\mathbf{R}}_1 - \widehat{\mathbf{R}}_2)^2\big] \\
&\quad = \mathrm{tr}\big(\mathbf{S}_1^2\big) + \mathrm{tr}\big(\mathbf{S}_2^2\big) - 2\,\mathrm{tr}(\mathbf{S}_1\mathbf{S}_2) - 2\,\mathrm{tr}\big[\mathbf{S}_1^2\big(\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p\big)\big] \\
&\qquad - 2\,\mathrm{tr}\big[\mathbf{S}_2^2\big(\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p\big)\big] \\
&\qquad + 2\,\mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p\big]\mathbf{S}_1\mathbf{S}_2\big\} + 2\,\mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p\big]\mathbf{S}_1\mathbf{S}_2\big\} \\
&\qquad + 2\sum_{\ell=1}^{2} \mathrm{tr}\big[\mathbf{S}_\ell^2\big(\mathrm{diag}(\mathbf{S}_\ell) - \mathbf{I}_p\big)^2\big] \\
&\qquad + \sum_{\ell=1}^{2} \mathrm{tr}\big[\mathbf{S}_\ell\big(\mathrm{diag}(\mathbf{S}_\ell) - \mathbf{I}_p\big)\mathbf{S}_\ell\big(\mathrm{diag}(\mathbf{S}_\ell) - \mathbf{I}_p\big)\big] \\
&\qquad - 1.5\,\mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p\big]^2\mathbf{S}_1\mathbf{S}_2\big\} - 1.5\,\mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p\big]^2\mathbf{S}_1\mathbf{S}_2\big\} \\
&\qquad - 0.5\,\mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p\big]\mathbf{S}_1\big[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p\big]\mathbf{S}_2\big\} \\
&\qquad - 0.5\,\mathrm{tr}\big\{\mathbf{S}_1[\mathrm{diag}\,\mathbf{S}_2 - \mathbf{I}_p]\mathbf{S}_2[\mathrm{diag}\,\mathbf{S}_2 - \mathbf{I}_p]\big\} \\
&\qquad - \mathrm{tr}\big\{\big[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p\big]\big[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p\big]\mathbf{S}_1\mathbf{S}_2\big\} \\
&\qquad - \mathrm{tr}\big\{\mathbf{S}_1\big[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p\big]\mathbf{S}_2\big[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p\big]\big\} + o_p(1).
\end{aligned}$$

This step is mainly to use the Taylor expansions of $s_{\ell jj}^{-1} = 1 - (s_{\ell jj} - 1) + (s_{\ell jj} - 1)^2 + o(p^{-1})$ and $s_{\ell jj}^{-1/2} = 1 - \frac{1}{2}(s_{\ell jj} - 1) + \frac{3}{8}(s_{\ell jj} - 1)^2 + o(p^{-1})$ with $\mathbf{S}_\ell = (s_{\ell ij})_{i,j=1}^{p}$ for $\ell = 1, 2$.

Skeleton of Step 2. We want to derive the limits of the following ten terms in probability: $\mathrm{tr}[\mathbf{S}_1^2(\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p)^2]$, $\mathrm{tr}[\mathbf{S}_2^2(\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p)^2]$, $\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p]^2\mathbf{S}_1\mathbf{S}_2\}$, $\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p]^2\mathbf{S}_1\mathbf{S}_2\}$, $\mathrm{tr}[\mathbf{S}_1(\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p)\mathbf{S}_1(\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p)]$, $\mathrm{tr}[\mathbf{S}_2(\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p)\mathbf{S}_2(\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p)]$, $\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p]\mathbf{S}_1[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p]\mathbf{S}_2\}$, $\mathrm{tr}\{\mathbf{S}_1[\mathrm{diag}\,\mathbf{S}_2 - \mathbf{I}_p]\mathbf{S}_2[\mathrm{diag}\,\mathbf{S}_2 - \mathbf{I}_p]\}$, $\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p][\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p]\mathbf{S}_1\mathbf{S}_2\}$, $\mathrm{tr}\{\mathbf{S}_1[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p]\mathbf{S}_2[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p]\}$.

Skeleton of Step 3. We want to derive the limiting null distribution of $(\mathrm{tr}(\mathbf{S}_1^2) - \mathrm{E}\,\mathrm{tr}(\mathbf{S}_1^2), \mathrm{tr}(\mathbf{S}_2^2) - \mathrm{E}\,\mathrm{tr}(\mathbf{S}_2^2), \mathrm{tr}(\mathbf{S}_1\mathbf{S}_2) - \mathrm{E}\,\mathrm{tr}(\mathbf{S}_1\mathbf{S}_2), \mathrm{tr}[\mathbf{S}_1^2(\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p)] - \mathrm{E}\,\mathrm{tr}[\mathbf{S}_1^2(\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p)], \mathrm{tr}[\mathbf{S}_2^2(\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p)] - \mathrm{E}\,\mathrm{tr}[\mathbf{S}_2^2(\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p)], \mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p]\mathbf{S}_1\mathbf{S}_2\} - \mathrm{E}\,\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p]\mathbf{S}_1\mathbf{S}_2\}, \mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p]\mathbf{S}_1\mathbf{S}_2\} - \mathrm{E}\,\mathrm{tr}\{[\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p]\mathbf{S}_1\mathbf{S}_2)\}$. Thus by the delta method, we obtain the CLT of $\mathrm{tr}[(\widehat{\mathbf{R}}_1 - \widehat{\mathbf{R}}_2)^2]$. Because these terms involve the product of any two or three terms among $\mathbf{S}_1$, $\mathrm{diag}(\mathbf{S}_1) - \mathbf{I}_p$, $\mathbf{S}_2$ and $\mathrm{diag}(\mathbf{S}_2) - \mathbf{I}_p$, the CLT for Theorem 2.1 is not directly applicable. Thus, in order to derive the CLT of these terms, eight new martingale difference sequences are constructed. Especially, the derivation of the

CLT for the two population case $\mathbf{R}_1 = \mathbf{R}_2$ is very different from and more difficult than the derivation for the one population case $\mathbf{R}_1 = \mathbf{R}_*$. $\square$

REMARK 3.2. Under the null hypothesis H_{02}, we do not know the true $\mathbf{R}$, so we have to estimate the terms related to $\mathbf{R}$ in the asymptotic mean and variance. Let

$$\omega_1 = n_1/(n_1+n_2), \qquad \omega_2 = 1-\omega_1 \quad \text{and}$$

$$\mathbf{M}_{\ell i} = \mathbf{x}_i^{(\ell)}(\mathbf{x}_i^{(\ell)})^T \qquad \text{for } \ell = 1, 2,$$

$$\hat{a}_\ell = \frac{[\operatorname{tr}(\widehat{\mathbf{R}}_\ell^2) - (n_\ell^2 - n_\ell - 1)p^2 n_\ell^{-1}(n_\ell-1)^{-2} - \beta_\ell n_\ell p (n_\ell - 1)^{-2}](n_\ell-1)^2}{p(n_\ell^2 - n_\ell + 2)},$$
$$\ell = 1, 2,$$

where $\mathbf{x}_i^{(\ell)} = (x_{1i}^{(\ell)}, \dots, x_{pi}^{(\ell)})^T = [\operatorname{diag}(\mathbf{S}_\ell)]^{-1/2}(\mathbf{z}_i^{(\ell)} - \bar{\mathbf{z}}^{(\ell)})$ with $i = 1, \dots, n_\ell$. Then, we estimate $a_0 = p^{-1}\operatorname{tr}(\mathbf{R}^2)$, b_0, c_0, and d_ℓ as follows:

$$\hat{a}_0 = \omega_1\hat{a}_1 + \omega_2\hat{a}_2,$$

$$\hat{b}_0 = \sum_{\ell=1}^{2}\omega_\ell\beta_\ell^{-1}p^{-1}\left(n_\ell^{-1}\sum_{i=1}^{n_\ell}\operatorname{tr}\{\widehat{\mathbf{R}}_{\{1,2\}/\{\ell\}}\mathbf{M}_{\ell i}[\operatorname{diag}(\mathbf{M}_{\ell i}) - \mathbf{I}_p]\} - 2p\hat{a}_0\right),$$

$$\hat{c}_0 = \sum_{\ell=1}^{2}\omega_\ell\beta_\ell^{-1}p^{-1}\left(n_\ell^{-1}\sum_{i=1}^{n_\ell}\operatorname{tr}\{\widehat{\mathbf{R}}_{\{1,2\}/\{\ell\}}[\widehat{\mathbf{R}}_\ell - (n_\ell-1)^{-1}\mathbf{M}_{\ell i}]\right.$$
$$\left.\times[\operatorname{diag}(\mathbf{M}_{\ell i}) - \mathbf{I}_p]^2\} - 2p\hat{a}_0\right),$$

$$\hat{d}_\ell = p^{-1}\{n_\ell^{-1}\sum_{i=1}^{n_\ell}\operatorname{tr}\{[\widehat{\mathbf{R}}_\ell - \mathbf{M}_{\ell i}(n_\ell-1)^{-1}][\operatorname{diag}(\mathbf{M}_{\ell i}) - \mathbf{I}_p]$$
$$\times\widehat{\mathbf{R}}_{\{1,2\}/\{\ell\}}[\operatorname{diag}(\mathbf{M}_{\ell i}) - \mathbf{I}_p]\}\},$$

with letting $\beta_\ell^{-1} = 0$ if $\beta_\ell = 0$, $\widehat{\mathbf{R}}_{\{1,2\}/\{1\}} = \widehat{\mathbf{R}}_2$ and $\widehat{\mathbf{R}}_{\{1,2\}/\{2\}} = \widehat{\mathbf{R}}_1$. Finally, we can obtain an estimate of μ_{z12} as follows:

$$\hat{\mu}_{z12} = \sum_{\ell=1}^{2}\frac{(n_\ell^2 - n_\ell - 1)p^2}{n_\ell(n_\ell-1)^2} - \sum_{\ell=1}^{2}\frac{p(2n_\ell^2 + n_\ell + 1)}{(n_\ell-1)^3}\hat{a}_0 + \sum_{\ell=1}^{2}\frac{\beta_\ell n_\ell p}{(n_\ell-1)^2}$$
$$- \sum_{\ell=1}^{2}\frac{\beta_\ell 2pn_\ell}{(n_\ell-1)^2}\hat{b}_0 + \sum_{\ell=1}^{2}\frac{\beta_\ell p(n_\ell^2 - 5n_\ell)}{2(n_\ell-1)^3}\hat{c}_0 + \sum_{\ell=1}^{2}\frac{p(n_\ell-3)n_\ell}{2(n_\ell-1)^3}\hat{d}_\ell.$$

COROLLARY 3.1. *Under the same assumptions of Theorem* 3.1, *we concluded that*:

- $(L_{n,2} - \hat{\mu}_{z12})/\{2[(n_1-1)^{-1} + (n_2-1)^{-1}]p\hat{a}_0\} \xrightarrow{L} N(0,1)$;
- $(M_{n,2} - \hat{\mu}_{z12})/\{2[(n_1-1)^{-1} + (n_2-1)^{-1}]p\hat{a}_0\} \xrightarrow{L} N(0,1)$ *holds under some additional conditions, including* $s(n_1, n_2, p) - 4\log p \to +\infty$ *and* (C1), (C2) *and* (C3) *in Cai, Liu and Xia* (2013) *for* $\mathbf{R}$ *and* $\mathbf{z}_i^{(\ell)}$.

3.1.2. *Power properties and optimality.* In the following, we will study the power properties of the statistics $M_{n,2}$ and $M'_{n,2}$.

THEOREM 3.2. *Under Assumptions* (a) *and* (b), *we have*

$$\sigma_{A12}^{-1}\{\mathrm{tr}[(\widehat{\mathbf{R}}_1 - \widehat{\mathbf{R}}_2)^2] - \mu_{A12}\} \xrightarrow{L} N(0,1),$$

where μ_{A12} *and* σ_{A12} *depend on the alternative population correlation matrices* $\mathbf{R}_1$ *and* $\mathbf{R}_2$ *and will be given in the Appendix.*

Theorem 3.2 gives the asymptotic distribution of the statistic $\mathrm{tr}[(\widehat{\mathbf{R}}_1 - \widehat{\mathbf{R}}_2)^2]$ under the alternative hypothesis. The power function is given by

$$g_2(\mathbf{R}_1, \mathbf{R}_2, \alpha) = P\big((M_{n,2} - \hat{\mu}_{z12})/\big\{2\big[(n_1-1)^{-1} + (n_2-1)^{-1}\big]p\hat{a}_0\big\} > q_{1-\alpha}\big),$$

and $g'_2(\mathbf{R}_1, \mathbf{R}_2, \alpha) = P(M'_{n,2} > c_{\alpha,2})$. Then $g_2(\mathbf{R}_1, \mathbf{R}_2, \alpha)$ and $g'_2(\mathbf{R}_1, \mathbf{R}_2, \alpha)$ have the following properties.

COROLLARY 3.2. *Under the same assumptions of Theorem* 3.2, *we have the following results*:

- *If* $\mathrm{tr}[(\mathbf{R}_2 - \mathbf{R}_1)^2] > c_0 > 0$, *where* c_0 *is a positive scalar, then* $g_2(\mathbf{R}_1, \mathbf{R}_2, \alpha) > \alpha$ *when the sample size is large enough.*
- *If* $\mathrm{tr}[(\mathbf{R}_1 - \mathbf{R}_2)^2] \to \infty$, *then* $g_2(\mathbf{R}_1, \mathbf{R}_2, \alpha)$ *and* $g'_2(\mathbf{R}_1, \mathbf{R}_2, \alpha)$ *are close to one as* $n_1, n_2 \to \infty$.
- *If the absolute value of at least one entry of* $\mathbf{R}_1 - \mathbf{R}_2$ *is greater than* $[\log(p)\log(n_1+n_2)]^{1/2}/\sqrt{\min\{n_1, n_2\}}$ *and the conditions* (C1), (C2) *and* (C3) *in Cai, Liu and Xia* (2013) *hold for* $\mathbf{R}_\ell$ *and* $\mathbf{z}_i^{(\ell)}$, $\ell = 1, 2$, *then* $g_2(\mathbf{R}_1, \mathbf{R}_2, \alpha)$ *and* $g'_2(\mathbf{R}_1, \mathbf{R}_2, \alpha)$ *are close to one as* $n_1, n_2 \to \infty$.

Corollary 3.2 shows that the proposed test $M_{n,2}$ is asymptotically unbiased. In the Appendix, we will prove that for the dense alternative $\mathrm{tr}[(\mathbf{R}_1 - \mathbf{R}_2)^2] \to \infty$, the power functions tend to one. For the sparse alternative, if the absolute value of at least one entry of $\mathbf{R}_1 - \mathbf{R}_2$ is greater than $[\log(p)\log(n_1+n_2)]^{1/2}/\sqrt{\min\{n_1, n_2\}}$, then the power functions are close to one.

Similar to Cai and Ma (2013), we define

$$\Theta_2^*(b_2, b_{20}) = \{\mathbf{R}_1, \mathbf{R}_2 : \|\mathbf{R}_1 - \mathbf{R}_2\|_F > b_2 \min\{\sqrt{p/n_1}, \sqrt{p/n_2}\}$$
$$\text{or } \|\mathbf{R}_1 - \mathbf{R}_2\|_\infty > b_{20} \min\{\sqrt{\log p/n_1}, \sqrt{\log p/n_2}\},$$

where b_2, b_{20} are positive constants, $\|\mathbf{R}_1 - \mathbf{R}_2\|_F = \{\mathrm{tr}[(\mathbf{R}_1 - \mathbf{R}_2)^2]\}^{1/2}$ and $\|\mathbf{R}_1 - \mathbf{R}_2\|_\infty = \max_{1\le i\le j\le p} |\mathbf{e}_i^T(\mathbf{R}_1 - \mathbf{R}_2)\mathbf{e}_j|$ with $\mathbf{e}_i$ and $\mathbf{e}_j$ being the ith column and jth column of the $p \times p$ identity matrix, respectively.

THEOREM 3.3. *Let $0 < \alpha < \beta < 1$. Suppose that $p/n_i \to y_i > 0$ as $n_i \to \infty$ for $i = 1, 2$. Then there exist two constants $0 < b_2, b_{20} < 1$ such that for any test ϕ with the significance level α for testing $H_{02} : \mathbf{R}_1 = \mathbf{R}_2$, we have*

$$\limsup_{n_1,n_2\to\infty} \inf_{\mathbf{R}_1,\mathbf{R}_2\in\Theta_2^*(b_2,b_{20})} \mathrm{E}_{\mathbf{R}_1,\mathbf{R}_2}\phi < \beta,$$

where $\mathrm{E}_{\mathbf{R}_1,\mathbf{R}_2}$ is the expectation under the two population correlation matrix being $\mathbf{R}_1$ and $\mathbf{R}_2$.

Theorem 3.3 shows that no level α test can distinguish between the null hypothesis and all alternative hypotheses with the power tending to one as $p/n_i \to y_i > 0$ for $i = 1, 2$ and $\|\mathbf{R}_1 - \mathbf{R}_2\|_F = O(\min\{\sqrt{p/n_1}, \sqrt{p/n_2}\})$ or $\|\mathbf{R}_1 - \mathbf{R}_2\|_\infty > b_{20}\min\{\sqrt{\log p/n_1}, \sqrt{\log p/n_2}\}$. Then Theorem 3.3 gives the lower bound for the optimality of our proposed procedure.

3.2. *Test statistic for multiple sample testing problem.* We extend the test statistic from two samples to K samples. The one weighted test statistic is constructed as

$$M_{n,K} = \sum_{1\le\ell_1<\ell_2\le K} \omega_{\ell_1,\ell_2} \mathrm{tr}\big[(\widehat{\mathbf{R}}_{\ell_1} - \widehat{\mathbf{R}}_{\ell_2})^2\big] + T_{n,K},$$

where

$$T_{n,K} = C_0 \sum_{1\le\ell_1<\ell_2\le K} \omega_{\ell_1,\ell_2}\delta_{\{\max_{1\le h<j\le p}(n_{\ell_1}^{-1}\hat{\theta}_{\ell_1 hj}+n_{\ell_2}^{-1}\hat{\theta}_{\ell_2 hj})^{-1}(\hat{r}_{\ell_1 hj}-\hat{r}_{\ell_2 hj})^2>s(n_{\ell_1},n_{\ell_2},p)\}},$$

with $\{\omega_{\ell_1,\ell_2}, 1 \le \ell_1 < \ell_2 \le K\}$ being a vector of weights.

For simplicity, we focus on the asymptotic distribution of $M_{n,K}$.

We first present a key lemma as follows.

LEMMA 3.1. *Assume that Assumptions (a) and (b) hold for $\ell = 1, \dots, K$. Then, $\{\mathrm{tr}[(\widehat{\mathbf{R}}_{\ell_1} - \widehat{\mathbf{R}}_{\ell_2})^2] - \mu_{A\ell_1\ell_2}, 1 \le \ell_1 < \ell_2 \le K\}$ are asymptotically distributed as a multivariate normal distribution. Moreover, we have*

$$\big\{\mathrm{tr}\big[(\widehat{\mathbf{R}}_{\ell_1} - \widehat{\mathbf{R}}_{\ell_2})^2\big] - \mu_{A\ell_1\ell_2}, \mathrm{tr}\big[(\widehat{\mathbf{R}}_{\ell_3} - \widehat{\mathbf{R}}_{\ell_4})^2\big] - \mu_{A\ell_3\ell_4}\big\} \xrightarrow{L} N(\mathbf{0}_2, \mathbf{\Gamma}),$$

where $\mathbf{\Gamma} = \{\gamma_{Auu'}\}_{u,u'=1}^2$ with $\gamma_{A11} = \sigma_{A\ell_1\ell_2}$, $\gamma_{A22} = \sigma_{A\ell_3\ell_4}$ and

$$\gamma_{A12} = \begin{cases} \sigma_{A\ell_1\ell_2\ell_3\ell_4}, & \ell_1 < \ell_2 = \ell_3 < \ell_4; \\ 0, & \ell_1, \ell_2, \ell_3, \ell_4 \text{ are all unequal.} \end{cases}$$

Moreover, $\mu_{A\ell_1\ell_2}$, $\mu_{A\ell_3\ell_4}$ and $\gamma_{Auu'}$ have closed forms and will be given in the Appendix.

Based on Lemma 3.1, we can establish the asymptotic null distribution of $M_{n,K}$ as follows.

THEOREM 3.4. *Under the same assumptions of Lemma* 3.1 *and* H_{0K}, *we conclude that*:

(III.1) $\nu_K^{-1/2}\sum_{1\le\ell_1<\ell_2\le K}\omega_{\ell_1,\ell_2}\{\mathrm{tr}[(\widehat{\mathbf{R}}_{\ell_1}-\widehat{\mathbf{R}}_{\ell_2})^2]-\mu_{z\ell_1\ell_2}\}\xrightarrow{L}N(0,1)$.

(III.2) $\nu_K^{-1/2}[M_{n,K}-\sum_{1\le\ell_1<\ell_2\le K}\omega_{\ell_1,\ell_2}\mu_{z\ell_1\ell_2}]\xrightarrow{L}N(0,1)$ *under some additional conditions*, *including* $s(n_{\ell_1},n_{\ell_2},p)-4\log p\to+\infty$ *and* (C1), (C2) *and* (C3) *of Cai*, *Liu and Xia* (2013) *for* $\mathbf{R}$ *and* $\mathbf{z}_i^{(\ell)}$, *where* $\nu_K=4[\mathrm{tr}(\mathbf{R}^2)]^2u_K$ *is given by*

$$u_K=\sum_{1\le\ell_1<\ell_2\le K}\omega_{\ell_1\ell_2}^2[(n_{\ell_1}-1)^{-1}+(n_{\ell_2}-1)^{-1}]^2$$
$$+2\sum_{1\le\ell_1<\ell_2=\ell_3<\ell_4\le K}\omega_{\ell_1\ell_2}\omega_{\ell_3\ell_4}(n_{\ell_2}-1)^{-2}$$
$$+2\sum_{1\le\ell_1<\ell_3<\ell_2=\ell_4\le K}\omega_{\ell_1\ell_2}\omega_{\ell_3\ell_4}(n_{\ell_2}-1)^{-2}$$
$$+2\sum_{1\le\ell_1=\ell_3<\ell_2<\ell_4\le K}\omega_{\ell_1\ell_2}\omega_{\ell_3\ell_4}(n_{\ell_1}-1)^{-2},$$

and $\mu_{z\ell_1\ell_2}$ *can be similarly defined as* μ_{z12}.

REMARK 3.3. There are two important issues associated with $M_{n,K}$. The first one is to determine the weights ω_{ℓ_1,ℓ_2}. Since the asymptotic variance of $\mathrm{tr}[(\widehat{\mathbf{R}}_{\ell_1}-\widehat{\mathbf{R}}_{\ell_2})^2]-\mu_{z\ell_1\ell_2}$ is equal to $4[(n_{\ell_1}-1)^{-1}+(n_{\ell_2}-1)^{-1}]^2[\mathrm{tr}(\mathbf{R}^2)]^2$, a reasonable choice of ω_{ℓ_1,ℓ_2} is $\omega_{\ell_1,\ell_2}=[(n_{\ell_1}-1)^{-1}+(n_{\ell_2}-1)^{-1}]^{-1}$ for $1\le\ell_1<\ell_2\le K$. The second one is to estimate the asymptotic mean and variance under H_{0K}, since we do not know what the true $\mathbf{R}$ is. A good estimate of $p^{-1}\,\mathrm{tr}(\mathbf{R}^2)$ is $\sum_{\ell=1}^K(n_\ell-1)(n_1+\cdots+n_K-K)^{-1}\hat{a}_\ell$, where $\hat{a}_\ell$ was defined in Remark 3.2. Then the estimate of ν_K,

$$\hat{\nu}_K=\sum_{\ell=1}^K(n_\ell-1)(n_1+\cdots+n_K-K)^{-1}p\hat{a}_\ell u_K.$$

Furthermore, the estimate $\hat{\mu}_{z\ell_1\ell_2}$ can be obtained by replacing 1 and 2 by ℓ_1 and ℓ_2 in $\hat{\mu}_{z12}$ in Remark 3.2. The C_0 is the same as $C_{0,2}$ in Remark 3.1. The threshold $s(n_{\ell_1},n_{\ell_2},p)$ is obtained by replacing 1 and 2 by ℓ_1 and ℓ_2 in $s(n_1,n_2,p)$ in Remark 3.2.

4. Estimation of the kurtosis $\boldsymbol{\beta_1}$. To estimate β_1 in Theorems 2.1 and 3.1, we consider two cases as follows.

Case 1: When $\mathbf{R}_1$ is unknown, the covariance matrix $\mathbf{\Sigma}_1$ is unknown. We may use an estimator of β_1 by Zheng et al. (2018) as follows:

$$\tilde{\beta}_1 = \frac{n_1 \widehat{V} - 2\{(n_1 - 1)\operatorname{tr}(\mathbf{S}_1^2) - [\operatorname{tr}(\mathbf{S}_1)]^2\}}{n_1 \sum_{j=1}^p s_{1jj}^2},$$

where $\mathbf{S}_1 = (s_{1\ell j})_{\ell,j=1}^p$ and

$$\widehat{V} = (n_1 - 1)^{-1} \sum_{i=1}^{n_1} \left\{ (\mathbf{z}_i^{(1)} - \bar{\mathbf{z}}^{(1)})^T (\mathbf{z}_i^{(1)} - \bar{\mathbf{z}}^{(1)}) - n_1^{-1} \sum_{i=1}^{n_1} [(\mathbf{z}_i^{(1)} - \bar{\mathbf{z}}^{(1)})^T (\mathbf{z}_i^{(1)} - \bar{\mathbf{z}}^{(1)})] \right\}^2.$$

Case 2: For $\mathbf{R}_1 = \mathbf{R}_*$ for a prespecified correlation matrix $\mathbf{R}_*$, we may estimate β_1 as follows:

$$\hat{\beta}_1 = p^{-1} \sum_{\ell=1}^p \left[\frac{(n_1 - 1)^{-1} \sum_{i=1}^{n_1} (\xi_{\ell i} - \bar{\xi}_\ell)^2 - 2\bar{\xi}_\ell^2}{\bar{\xi}_\ell^2 \sum_{j=1}^p (r_{*\frac{1}{2}\ell j})^4} \right],$$

where $\xi_{\ell i} = (x_{\ell i}^{(1)} - \bar{x}_\ell^{(1)})^2$, $\bar{\xi}_\ell = n_1^{-1} \sum_{i=1}^{n_1} \xi_{\ell i}$, and $\mathbf{R}_*^{1/2} = (r_{*\frac{1}{2}\ell j})_{\ell,j=1}^p$.

The following lemma gives the consistency of the estimator $\hat{\beta}_1$ under the null hypothesis $H_{01} : \mathbf{R}_1 = \mathbf{R}_*$.

LEMMA 4.1. *Suppose that* $\mathrm{E}[(z_{\ell 1}^{(1)})^8] \le c$ *holds for all* $\ell = 1, \dots, p$, *where* c *is a positive constant. Under the null hypothesis* H_{01} *and Assumptions* (a)–(b), *we have*

$$\hat{\beta}_1 = \beta_1 + o_p(1).$$

PROOF. Without loss of generality, assume $\mathrm{E} z_{\ell 1}^{(1)} = 0$ and $\mathrm{Var}(z_{\ell 1}^{(1)}) = 1$ for all ℓ. Let $\mathbf{e}_\ell$ be the ℓth column of the $p \times p$ identity matrix. We can show the following results:

$$n_1^{-1} \sum_{i=1}^{n_1} z_{\ell i}^{(1)} = o_p(1), \tag{4.1}$$

$$n_1^{-1} \sum_{i=1}^{n_1} (z_{\ell i}^{(1)})^2 = \mathrm{E}(z_{\ell 1}^{(1)})^2 + o_p(1), \tag{4.2}$$

$$n_1^{-1} \sum_{i=1}^{n_1} (z_{\ell i}^{(1)})^4 = \mathrm{E}(z_{\ell 1}^{(1)})^4 + o_p(1), \tag{4.3}$$

where $o_p(1)$ is uniform for $\ell = 1, \dots, p$. For instance, to prove (4.1), we have

$$\mathrm{E}\left[\left(n_1^{-1}\sum_{i=1}^{n_1} z_{\ell i}^{(1)} - \mathrm{E}z_{\ell 1}^{(1)}\right)^2\right] = n_1^{-1}\mathrm{E}\big(z_{\ell 1}^{(1)} - \mathrm{E}z_{\ell 1}^{(1)}\big)^2$$
$$= n_1^{-1}\mathbf{e}_\ell^T \boldsymbol{\Sigma}_* \mathbf{e}_\ell \le n_1^{-1} c = o(1),$$

where $o(1)$ is uniform for all $\ell = 1, \dots, p$. It follows that

$$\bar{\xi}_\ell = n_1^{-1}\sum_{i=1}^{n_1}\big(z_{\ell 1}^{(1)}\big)^2 - \left[n_1^{-1}\sum_{i=1}^{n_1} z_{\ell i}^{(1)}\right]^2 = \mathbf{e}_\ell^T \boldsymbol{\Sigma}_* \mathbf{e}_\ell + o_p(1),$$
$$n_1^{-1}\sum_{i=1}^{n_1}\xi_{\ell i}^2 = \mathrm{E}\big(z_{\ell 1}^{(1)}\big)^4 + o_p(1).$$

Therefore, we have

$$(n_1 - 1)^{-1}\sum_{i=1}^{n_1}(\xi_{\ell i} - \bar{\xi}_\ell)^2/\bar{\xi}_\ell^2 = \frac{\mathrm{E}(z_{\ell 1}^{(1)})^4 - (\mathbf{e}_\ell^T \boldsymbol{\Sigma}_* \mathbf{e}_\ell)^2}{(\mathbf{e}_\ell^T \boldsymbol{\Sigma}_* \mathbf{e}_\ell)^2} + o_p(1)$$
$$= 2 + \beta_1 \sum_{j=1}^{p}\big[\mathbf{e}_j^T \mathbf{R}_*^{1/2}\mathbf{e}_\ell\big]^4 + o_p(1),$$

which yields $\hat{\beta}_1 = \beta_1 + o_p(1)$. This completes the proof of Lemma 4.1. □

5. Simulation studies. In this section, we carried out simulation studies to evaluate the finite-sample performance of the proposed test statistics in terms of the empirical test size and power. We consider both one sample testing problem and two sample testing problem. For the one sample testing problem, we set the dimension p to be $p = 50, 100, 200, 500$ and 1000 and the sample size n_1 to be $n_1 = 100, 120, 200$ and 300. The data were generated according to $\mathbf{z}_i^{(1)} = \mathbf{R}^{1/2}\mathbf{w}_{zi}^{(1)}$ for $i = 1, \dots, n_1$, where the elements of $\mathbf{w}_{zi}^{(1)}$ were independently and identically generated from Gaussian population $N(0, 1)$ or $\mathrm{Gamma}(4, 2) - 2$. For the two sample testing problem, we set $p = 50, 100, 200, 500$ and 1000 and $(n_1, n_2) = (100, 100)$, $(150, 150)$, $(200, 200)$. The data were generated according to $\mathbf{z}_i^{(\ell)} = \mathbf{R}_\ell^{1/2}\mathbf{w}_{zi}^{(\ell)}$ for $i = 1, \dots, n_\ell$ and $\ell = 1, 2$, where the elements of $\mathbf{w}_{zi}^{(\ell)}$ were independently and identically generated from Gaussian population $N(0, 1)$ or $\mathrm{Gamma}(4, 2) - 2$. For the three sample testing problem, we set $p = 50, 100, 200, 500$ and 1000 and $(n_1, n_2, n_3) = (100, 100, 100)$, $(100, 100, 100)$, $(100, 100, 200)$ and $(100, 200, 200)$. The data were generated according to $\mathbf{z}_i^{(\ell)} = \mathbf{R}_\ell^{1/2}\mathbf{w}_{zi}^{(\ell)}$ for $i = 1, \dots, n_\ell$ and $\ell = 1, 2, 3$, where the elements of $\mathbf{w}_{zi}^{(\ell)}$ were independently and identically generated from Gaussian population $N(0, 1)$ or

Gamma(4, 2) − 2. We set the nominal size to be 5%, run 2000 replications for empirical sizes and 1000 replications for empirical powers for each setting.

We consider nine different sets of population correlation matrices for $\mathbf{R}_\ell$. For the two sample testing problem, we compare our tests denoted as "FDS" for $M_{n,2}$ and "MAX" for $M'_{n,2}$ with the extreme statistic test, denoted as "CZ" in Cai and Zhang (2016). However, for the one sample testing problem, we cannot find any competing method when $\mathbf{R}_*$ is not an identity matrix, so we do not include any alternative method. When $\mathbf{R}_*$ is an identity matrix, we compare our test "FDS" for $M_{n,1}$ and "MAX" for $M'_{n,1}$ with "GHPY" in Gao et al. (2017) and "LX" in Li and Xue (2015). For the three-sample testing problem, since we cannot find any competing method, we do not include any alternative method. For the sake of space, we selectively present some key results in Tables 1–3 and include additional results in the Supplementary Material [Zheng et al. (2018)]. The first three models are designed for the one sample testing problem, whereas the middle four ones are for the two sample testing problem and the last three ones are for the three-sample testing problem. The ten different models of population correlation matrices are summarized as follows.

- Model 1.1: The population correlation matrix is set as $\mathbf{R}_1 = \mathbf{R}_* = (\rho^{|j'-j|})_{j',j=1}^p$, where ρ is taken as 0.0 and 0.5.
- Model 1.2: The population correlation matrix is set as $\mathbf{R}_1 = \mathbf{R}_* + \epsilon 0.02(\mathbf{1}_p\mathbf{1}_p^T - \mathbf{I}_p) + (1-\epsilon)2.5\sqrt{(\log p)/n}(\mathbf{e}_2\mathbf{e}_1^T + \mathbf{e}_1\mathbf{e}_2^T)$, where $\mathbf{R}_* = \mathbf{I}_p$, $\mathbf{1}_p$ is a $p \times 1$ vector of ones and $\mathbf{e}_k$ is the kth column of the $p \times p$ identity matrix. When $\epsilon = 1$, the signal pattern of $\mathbf{R}_1 - \mathbf{R}_*$ is dense. When $\epsilon = 0$, the signal pattern of $\mathbf{R}_1 - \mathbf{R}_*$ is sparse.
- Model 1.3: The population correlation matrix is set as $\mathbf{R}_1 = \mathbf{R}_* + \epsilon\sqrt{\log p \log n} \times (\mathbf{e}_2\mathbf{e}_1^T + \mathbf{e}_1\mathbf{e}_2^T)$, where $\mathbf{R}_* = (0.25^{|i-j|})_{i,j=1}^p$ and $\epsilon = 0.09$ and 0.12. In this case, the signal pattern of $\mathbf{R} - \mathbf{R}_*$ is sparse.
- Model 2.1: The population correlation matrices are set as $\mathbf{R}_1 = \mathbf{R}_2 = \{\rho^{|j'-j|}\}_{j',j=1}^p$ with $\rho = 0.00, 0.25, 0.50$ and 0.75. The simulation results for $\rho = 0.25$ and 0.75 are included in the Supplementary Material [Zheng et al. (2018)].
- Model 2.2: The population correlation matrices are set as $\mathbf{R}_1 = (0.5^{|j'-j|})_{j',j=1}^p$ and $\mathbf{R}_2 = \mathbf{R}_1 + \epsilon(\mathbf{1}_p\mathbf{1}_p^T - \mathbf{I}_p)$ with $\epsilon = 0.05$ and 0.08. In this case, the signal pattern of $\mathbf{R}_2 - \mathbf{R}_1$ is dense.
- Model 2.3: The population correlation matrices are set as $\mathbf{R}_1 = (\rho^{|j'-j|})_{j',j=1}^p$ and $\mathbf{R}_2 = \mathbf{R}_1 + \epsilon_p(\mathbf{e}_2\mathbf{e}_1^T + \mathbf{e}_1\mathbf{e}_2^T)$ with $\rho = 0.05, 0.08, 0.10$, and 0.12 and $\epsilon_p = \exp(0.008p)/[1+\exp(0.008p)]$. In this case, the signal pattern of $\mathbf{R}_2 - \mathbf{R}_1$ is sparse. The simulation results for $\rho = 0.05, 0.10$ and 0.12 are included in the Supplementary Material [Zheng et al. (2018)].
- Model 2.4: The population correlation matrices are set as $\mathbf{R}_1 = \mathbf{I}_p$ and $\mathbf{R}_2 = (\rho^{|j'-j|})_{j',j=1}^p$ with $\epsilon = 0.2, 0.225, 0.25$ and 0.275. In this case, the signal pat-

TABLE 1
Empirical sizes in Model 1.1 *and empirical powers in Models* 1.2–1.3 *for* H_{01} (*in percentage*)

			$w_{ij} \sim N(0,1)$					$w_{ij} \sim \text{Gamma}(4,2) - 2$				
ϵ	n	Methods	$p = 50$	100	200	500	1000	50	100	200	500	1000
						Empirical sizes in Model 1.1						
0.0	100	FDS	6.20	4.75	5.90	5.75	6.25	5.45	6.25	6.60	6.55	5.90
		MAX	4.25	3.25	3.50	3.10	3.30	4.60	5.00	4.30	4.60	4.75
		GHPY	4.85	4.15	4.70	4.10	5.10	4.85	4.50	4.70	5.75	4.90
		LX	3.15	1.90	1.70	0.90	0.90	4.35	4.55	3.55	3.10	3.10
	200	FDS	6.15	5.15	5.90	6.05	5.95	5.70	5.55	6.35	5.55	5.85
		MAX	4.40	3.40	5.05	4.10	3.80	4.85	4.45	5.10	4.75	4.65
		GHPY	5.80	4.80	4.90	4.85	4.95	5.20	4.35	5.35	4.80	5.10
		LX	5.35	3.40	3.55	2.15	2.00	5.20	4.90	5.40	4.90	5.00
	300	FDS	5.15	4.60	5.15	5.65	5.05	6.25	6.05	6.50	5.85	6.40
		MAX	3.90	3.80	3.80	4.45	3.50	4.55	4.40	5.10	4.15	4.75
		GHPY	4.45	5.20	4.65	5.10	5.25	4.95	5.50	5.65	4.80	5.85
		LX	5.25	4.25	3.90	2.80	2.75	5.15	5.20	6.05	5.15	5.65
0.5	100	FDS	5.50	5.25	5.90	5.95	6.35	5.75	6.00	5.70	5.80	5.95
		MAX	5.20	4.80	5.50	5.65	5.15	5.10	4.45	4.55	4.75	4.10
	200	FDS	5.25	6.00	5.30	5.05	6.25	5.60	5.40	5.75	5.45	5.10
		MAX	4.60	4.65	4.45	4.45	4.90	4.65	4.55	4.20	3.85	4.35
	300	FDS	5.85	6.00	5.00	5.75	5.05	4.50	5.75	6.35	5.80	6.05
		MAX	5.05	4.25	4.60	4.20	4.00	3.70	4.65	5.10	4.30	4.30
						Empirical powers in Model 1.2						
1.0	100	FDS	28.9	59.4	90.8	99.9	100.0	28.1	58.1	92.1	100.0	100.0
		MAX	21.0	49.8	86.8	99.8	100.0	20.4	49.6	88.3	99.9	100.0
		GHPY	19.6	49.5	86.9	100.0	100.0	20.4	46.9	87.4	99.3	100.0
		LX	7.10	7.6	8.9	15.7	23.3	10.7	13.4	16.1	24.4	34.3
	200	FDS	58.9	93.3	100.0	100.0	100.0	55.8	92.9	99.9	100.0	100.0
		MAX	48.9	90.8	100.0	100.0	100.0	48.8	90.2	99.8	100.0	100.0
		GHPY	51.6	90.4	99.9	100.0	100.0	50.9	89.8	100.0	100.0	100.0
		LX	17.8	28.3	49.7	86.90	97.55	22.9	34.0	54.0	85.9	96.8
	300	FDS	83.4	99.9	100.0	100.0	100.0	81.9	99.7	100.0	100.0	100.0
		MAX	77.4	99.5	100.0	100.0	100.0	75.6	99.5	100.0	100.0	100.0
		GHPY	78.8	99.6	100.0	100.0	100.0	77.3	99.0	100.0	100.0	100.0
		LX	33.6	59.0	90.6	99.8	100.0	37.2	62.8	89.6	99.4	100.0
0.0	100	FDS	75.0	79.0	85.2	90.0	94.6	74.5	81.6	85.5	92.1	95.3
		MAX	81.6	84.4	90.1	93.3	96.8	82.2	87.0	90.1	94.9	97.2
		GHPY	7.6	6.2	5.5	4.6	5.5	9.0	5.9	5.1	5.7	5.1
		LX	85.4	88.1	92.2	94.7	97.4	87.3	90.0	92.8	95.8	97.9
	200	FDS	71.4	78.7	79.9	84.6	88.2	72.0	76.4	80.3	84.5	88.4
		MAX	77.9	84.8	84.3	88.4	91.7	78.9	82.3	86.1	89.3	91.6
		GHPY	10.4	6.2	5.5	5.3	4.9	10.1	5.6	6.4	5.1	5.5
		LX	83.3	88.3	88.1	90.6	93.7	83.7	86.3	88.7	91.8	93.0

TABLE 1
(*Continued*)

			$w_{ij} \sim N(0,1)$					$w_{ij} \sim \text{Gamma}(4,2)-2$				
ϵ	n	Methods	$p=50$	100	200	500	1000	50	100	200	500	1000
	300	FDS	71.2	76.6	80.8	84.3	87.5	71.9	76.8	80.4	85.7	86.4
		MAX	77.2	81.6	86.0	88.8	90.5	77.5	82.7	85.4	88.9	90.4
		GHPY	8.4	7.0	5.2	5.3	5.1	9.2	6.7	6.8	4.7	5.1
		LX	83.2	85.0	88.2	90.7	92.7	82.6	86.6	88.4	91.6	92.5
			Empirical powers in Model 1.3									
0.09	100	FDS	44.8	44.1	47.0	46.8	47.2	38.4	39.7	38.5	38.0	38.6
		MAX	53.8	53.4	55.0	54.2	54.5	46.3	47.3	46.1	44.3	44.4
	200	FDS	99.2	99.8	99.8	99.9	99.9	97.0	97.3	97.4	97.5	98.1
		MAX	99.4	99.6	99.9	99.9	100.0	98.3	98.5	98.3	98.2	98.7
	300	FDS	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0
		MAX	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9	100.0	100.0
0.12	100	FDS	93.6	94.6	96.0	96.0	96.3	84.9	84.5	82.6	78.8	78.5
		MAX	96.2	96.7	97.5	97.5	97.3	89.4	88.5	84.4	82.5	82.3
	200	FDS	100.0	100.0	100.0	100.0	100.0	99.8	99.8	99.7	99.3	99.3
		MAX	100.0	100.0	100.0	100.0	100.0	99.9	99.8	99.9	99.6	99.6
	300	FDS	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		MAX	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

tern of $\mathbf{R}_2 - \mathbf{R}_1$ is between dense case and sparse case. The simulation results are included in the Supplementary Material [Zheng et al. (2018)].

- Model 3.1: The three population correlation matrices are taken as $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{R}_3 = \mathbf{I}_p$. The model is used for evaluating the empirical performance on Type I errors of the proposed test $M_{n,3}$.
- Model 3.2: The three population correlation matrices are taken as $\mathbf{R}_1 = \mathbf{R}_2 = I_p$ and $\mathbf{R}_3 = \mathbf{I}_p + 0.03(\mathbf{1}\mathbf{1}^T - \mathbf{I}_p)_{i,j=1}^p$. Here, $\mathbf{R}_3 - \mathbf{R}_1$ or $\mathbf{R}_3 - \mathbf{R}_2$ is dense.
- Model 3.3: The three population correlation matrices are taken as $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{I}_p$ and $\mathbf{R}_3 = (\delta_{\{i=j\}} + 0.2\delta_{\{|i-j|=1\}})_{i,j=1}^p$. Here, $\mathbf{R}_3 - \mathbf{R}_1$ or $\mathbf{R}_3 - \mathbf{R}_2$ is a little sparse.

Overall, the Type I error rates for our tests "FDS" and "MAX" are relatively accurate for all sample sizes, for all dimensions, for all correlation matrices, and for the two different distributions of error terms. For the one sample testing problem, "FDS" and "MAX" can deal with an arbitrary correlation matrix $\mathbf{R}_*$, whereas other test statistics "GHPY" and "LX" cannot. It seems that both ρ and p have some minor impact on its Type I error rates. The proposed tests 'FDS" and "MAX" perform very well for both sparse and dense alternatives. Consistent with our expectations, the statistical powers for rejecting the null hypothesis increase as ϵ, n and p increase. It seems that "MAX" has a little better performance than "FDS."

TABLE 2
Empirical sizes in Model 2.1 *and empirical powers in Models* 2.2–2.3 *for* H_{02} (*in percentage*)

ϵ	(n_1, n_2)	Methods	$w_{ij} \sim N(0,1)$				$w_{ij} \sim \text{Gamma}(4,2) - 2$			
			$p = 50$	100	200	500	50	100	200	500
			Empirical sizes in Model 2.1							
0.0	(100, 100)	FDS	5.85	6.00	6.20	6.85	4.85	4.65	5.15	5.50
		MAX	4.65	5.15	4.90	5.05	3.50	3.45	3.85	4.10
		CZ	5.00	5.10	5.85	5.25	3.25	3.70	3.55	3.10
	(150, 150)	FDS	5.85	6.45	5.50	5.50	4.85	5.55	4.80	4.65
		MAX	4.70	5.55	5.10	5.15	4.00	4.35	4.25	3.55
		CZ	4.40	4.45	4.80	5.05	3.35	3.95	4.25	2.80
	(200, 200)	FDS	5.30	5.20	5.25	5.80	4.85	4.95	5.40	5.25
		MAX	4.25	4.55	4.45	4.75	3.60	3.70	4.00	4.20
		CZ	4.20	5.45	4.50	4.70	3.60	4.00	3.00	3.35
0.5	(100, 100)	FDS	6.15	6.75	5.60	6.25	5.30	4.20	5.50	5.40
		MAX	5.30	5.75	5.00	5.60	3.45	4.15	4.60	4.55
		CZ	5.55	5.45	5.50	5.55	4.00	4.40	4.15	3.95
	(150, 150)	FDS	6.10	5.95	6.70	6.25	5.40	5.75	5.70	5.05
		MAX	5.30	4.75	5.85	4.85	4.05	4.04	4.20	4.45
		CZ	5.60	5.55	4.75	5.00	3.75	3.60	3.85	4.00
	(200, 200)	FDS	5.20	6.30	6.30	6.90	5.45	4.90	6.10	5.15
		MAX	4.20	5.30	5.05	5.95	4.50	3.65	5.65	4.30
		CZ	4.20	4.55	4.95	5.30	4.15	3.70	4.30	4.05
			Empirical powers in Model 2.2							
0.05	(100, 100)	FDS	49.6	83.6	99.2	100.0	48.6	82.8	99.9	100.0
		MAX	42.5	78.8	98.7	100.0	41.1	77.6	99.9	100.0
		CZ	9.8	10.4	12.6	14.3	9.7	9.8	10.3	10.7
	(150, 150)	FDS	71.8	96.7	100.0	100.0	70.8	97.3	100.0	100.0
		MAX	64.7	94.9	99.9	100.0	62.2	95.5	100.0	100.0
		CZ	14.4	15.0	16.8	16.6	11.5	11.8	14.1	14.3
	(200, 200)	FDS	84.9	99.6	100.0	100.0	82.4	99.8	100.0	100.0
		MAX	79.5	99.2	100.0	100.0	77.3	99.5	100.0	100.0
		CZ	14.9	17.5	19.5	23.6	14.6	15.5	17.9	20.6
0.08	(100, 100)	FDS	87.7	99.8	100.0	100.0	90.1	99.4	100.0	100.0
		MAX	84.2	99.7	100.0	100.0	85.8	99.3	99.9	100.0
		CZ	18.4	21.8	24.1	30.1	19.3	20.5	21.6	26.1
	(150, 150)	FDS	98.5	100.0	100.0	100.0	97.9	100.0	100.0	100.0
		MAX	97.5	100.0	100.0	100.0	97.2	100.0	100.0	100.0
		CZ	31.4	32.8	40.2	43.4	26.0	31.0	36.6	40.1
	(200, 200)	FDS	99.7	100.0	100.0	100.0	99.8	100.0	100.0	100.0
		MAX	99.5	100.0	100.0	100.0	99.6	100.0	100.0	100.0
		CZ	37.6	45.4	53.8	62.3	37.3	42.8	49.6	58.7

TABLE 2
(*Continued*)

			$w_{ij} \sim N(0,1)$				$w_{ij} \sim$ Gamma(4, 2) − 2			
ϵ	(n_1, n_2)	Methods	$p = 50$	100	200	500	50	100	200	500
						Empirical powers in Model 2.3				
0.08	(100, 100)	FDS	65.5	86.8	99.6	99.6	61.8	78.5	97.6	96.5
		MAX2	72.8	90.6	99.8	99.8	68.6	83.3	98.8	97.5
		CZ	78.7	93.2	99.9	99.9	76.0	87.7	99.5	98.3
	(150, 150)	FDS	94.1	99.4	100.0	100.0	92.6	98.5	99.9	100.0
		MAX	96.5	99.6	100.0	100.0	95.2	99.3	99.9	100.0
		CZ	97.7	99.9	100.0	100.0	97.0	99.5	99.9	100.0
	(200, 200)	FDS	99.5	100.0	100.0	100.0	99.1	99.9	100.0	100.0
		MAX	99.9	100.0	100.0	100.0	99.5	100.0	100.0	100.0
		CZ	99.9	100.0	100.0	100.0	99.7	100.0	100.0	100.0

For the two- and three-sample testing problems, "FDS" and "MAX" also can deal with arbitrary correlation matrices. It seems that ρ, p, and the error distribution have little impact on its Type I error rates. The proposed tests "FDS" and "MAX" perform reasonably well for sparse alternatives, dense alternatives, and between sparse and dense alternatives. It seems that "MAX" is slightly better than "FDS."

TABLE 3
Empirical test sizes in Model 3.1 *and empirical powers in Models* 3.2 *and* 3.3 *for* H_{03} (*in percentage*)

		$w_{ij} \sim N(0,1)$					$w_{ij} \sim$ Gamma(4, 2) − 2				
(n_1, n_2, n_3)		$p = 50$	100	200	500	1000	50	100	200	500	1000
						Empirical test sizes in Model 3.1					
(50, 100, 100)	FDS	6.10	6.60	5.95	6.20	6.60	5.45	5.50	6.00	5.85	5.30
(100, 100, 100)	FDS	6.30	5.90	5.20	5.85	5.35	5.90	4.85	5.15	5.90	5.60
(100, 100, 200)	FDS	4.70	5.60	5.65	5.75	5.25	5.00	5.35	4.80	5.35	5.15
(100, 200, 200)	FDS	5.85	5.45	5.60	5.70	4.55	5.75	4.85	5.40	4.90	5.25
						Empirical powers in Model 3.2					
(50, 100, 100)	FDS	18.5	44.9	86.4	100.0	100.0	14.4	36.3	81.6	100.0	100.0
(100, 100, 100)	FDS	22.1	52.5	91.8	100.0	100.0	18.6	50.1	91.0	100.0	100.0
(100, 100, 200)	FDS	28.8	74.1	99.6	100.0	100.0	26.8	71.2	98.9	100.0	100.0
(100, 200, 200)	FDS	38.9	88.2	99.9	100.0	100.0	39.2	85.8	99.8	100.0	100.0
						Empirical powers in Model 3.3					
(50, 100, 100)	FDS	39.7	37.4	37.1	40.3	53.1	30.1	28.8	32.1	33.3	35.8
(100, 100, 100)	FDS	47.2	45.9	47.7	50.4	50.1	40.6	43.2	41.5	45.1	46.8
(100, 100, 200)	FDS	68.1	69.8	70.3	73.6	73.1	64.2	65.2	65.6	68.8	67.8
(100, 200, 200)	FDS	84.5	87.5	87.2	89.3	88.8	83.2	85.0	85.1	86.3	86.1

6. Real data analysis.

6.1. *Alzheimer's Disease Neuroimaging Initiative* (*ADNI*) *data*. "Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator, Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org."[5]

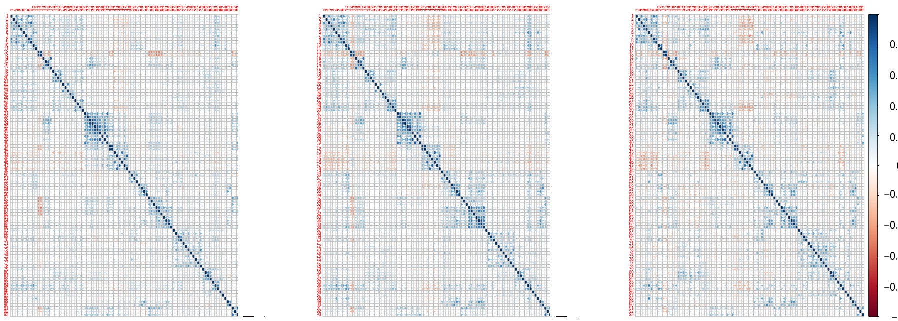
We consider 749 T1 weighted images collected at the baseline of ADNI1, consisting of 206 normal subjects, 364 mild cognitive impairment (MCI) subjects and 179 Alzheimer's disease (AD) subjects. These scans were performed on a 1.5T MRI scanners using a sagittal MPRAGE sequence and the typical protocol includes the following parameters: repetition time (TR) = 2400 ms, inversion time (TI) = 1000 ms, flip angle = 8°, and field of view (FOV) = 24 cm with a $256 \times 256 \times 170$ mm^3 acquisition matrix in the x, y and z dimensions, which yields a voxel size of $1.25 \times 1.26 \times 1.2$ mm^3.

The T1-weighted images were processed using the Hierarchical Attribute Matching Mechanism for Elastic Registration (HAMMER) pipeline. The processing steps include anterior commissure and posterior commissure correction, skull-stripping, cerebellum removal, intensity inhomogeneity correction and segmentation. We performed automatic regional labeling by labeling the template and by transferring the labels following the deformable registration of subject images. Finally, we labeled 93 regions of interest (ROIs) and computed their volumes for each subject.

6.2. *Group comparisons*. We are interested in characterizing differences among the three correlation matrices of ROI volumes for normal subjects, MCI subjects and AD subjects, which are denoted as R_{NC}, R_{MCI} and R_{AD}, respectively. Statistically, we test three two sample testing problems, including $R_{\text{NC}} = R_{\text{MCI}}$, $R_{\text{NC}} = R_{\text{AD}}$ and $R_{\text{MCI}} = R_{\text{AD}}$, and one three sample testing problem, that is, $R_{\text{NC}} = R_{\text{MCI}} = R_{\text{AD}}$.

We applied the test statistics $M_{n,2}$ and $M_{n,3}$ to carry out these tests as follows. First, for each ROI, we fitted a linear regression model with its ROI volume as response and age, gender and whole brain volume as covariates by using data obtained from all subjects. Second, for each group, we calculated its correlated matrix based on the residuals of all ROIs obtained from the first step. Figure 1 presents the

[5]ADNI manuscript citation guidelines. https://adni.loni.usc.edu/wp-content/uploads//how_to_apply/ADNI_DSP_Policy.pdf



(a) Normal (b) MCI (c) AD

FIG. 1. *Graphical display of correlation matrices of normal subjects, MCI subjects and AD subjects.*

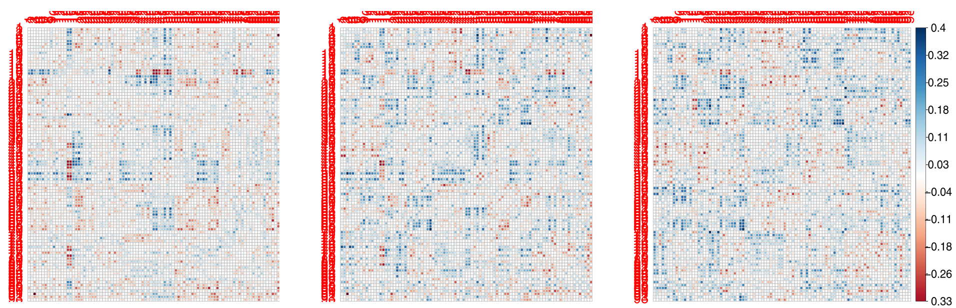
sample correlation matrices corresponding to the three groups. Figure 2 presents the difference of the sample correlation matrices among the three groups. Then we clustered the 93 ROIs according to the correlation matrix of the normal control group. For example, Cluster 1 includes the large area of prefrontal cortex, and its functions span over the frontoparietal control network (orbitofrontal cortex, middle frontal gyrus), default node network and ventral attention network. This region has been implicated in decision making, complex cognitive behavior, processing of higher information, decision making, personal expression, social behavior moderating, attention, memory, recognizing faces, characters, etc. Third, we calculated the p-value of testing $R_{\mathrm{NC}} = R_{\mathrm{MCI}}$, that of $R_{\mathrm{NC}} = R_{\mathrm{AD}}$, and that of $R_{\mathrm{MCI}} = R_{\mathrm{AD}}$ as 1.23×10^{-9}, 0 and 1.78×10^{-11}, respectively. Fourth, we calculated the p-value of testing $R_{\mathrm{NC}} = R_{\mathrm{MCI}} = R_{\mathrm{AD}}$ as 0.

APPENDIX A: SOME EXPRESSIONS

Let r_{*khj} be the (h, j) entry of $\mathbf{R}_*^k$ and $r_{\ell khj}$ be the (h, j) entry of $\mathbf{R}_\ell^k$ for $k = 1/2, 1, 3/2, 2, 3$. Let $\mathbf{e}_j$ be the jth column of the $p \times p$ identity matrix.

A.1. Expressions of μ_{zA}, μ_{z0} and σ_{zA}^2 for one population in Theorem 2.1 and 2.2. *Expression of μ_{zA} and μ_{z0}*:

$$\begin{aligned}\mu_{zA} = {} & [n_1(n_1-1)+2](n_1-1)^{-2}\operatorname{tr}(\mathbf{R}_1^2) + (n_1^2 - n_1 - 1)p^2 n_1^{-1}(n_1-1)^{-2} \\ & + \beta_1 n_1 p(n_1-1)^{-2} - 2\operatorname{tr}(\mathbf{R}_1\mathbf{R}_*) + \operatorname{tr}(\mathbf{R}_*^2) \\ & - 4n_1(n_1-1)^{-2}\left[2\operatorname{tr}(\mathbf{R}_1^2) + \beta_1 \sum_{h=1}^p \sum_{j=1}^p r_{1\frac{3}{2}hj}(r_{1\frac{1}{2}hj})^3\right] \\ & + 2n_1(n_1-1)^{-2}\left[2\operatorname{tr}(\mathbf{R}_1\mathbf{R}_*) + \beta_1 \sum_{h=1}^p \sum_{j=1}^p \mathbf{e}_j^T \mathbf{R}_1^{1/2}\mathbf{R}_*\mathbf{e}_h (r_{1\frac{1}{2}hj})^3\right]\end{aligned}$$



(a) Normal-MCI (b) Normal-AD (c) MCI-AD

FIG. 2. *Graphical display of difference between correlation matrices of normal subjects, MCI subjects and AD subjects.*

$$
\begin{aligned}
&+ 2n_1(n_1-2)(n_1-1)^{-3}\Bigg[2\operatorname{tr}(\mathbf{R}_1^2) + \beta_1 \sum_{h=1}^{p} r_{12hh} \sum_{j=1}^{p} (r_{1\frac{1}{2}hj})^4\Bigg] \\
&- 1.5n_1(n_1-1)^{-2}\Bigg[2\operatorname{tr}(\mathbf{R}_1\mathbf{R}_*) + \beta_1 \sum_{j=1}^{p} \mathbf{e}_j^T \mathbf{R}_1\mathbf{R}_*\mathbf{e}_j \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^4\Bigg] \\
&+ n_1(n_1-2)(n_1-1)^{-3} \sum_{j=1}^{p}\sum_{j'=1}^{p} (r_{11jj'})^2 \\
&\times \Bigg[2(r_{11jj'})^2 + \beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 (r_{1\frac{1}{2}hj'})^2\Bigg] \\
&- 0.5n_1(n_1-1)^{-2} \sum_{j=1}^{p}\sum_{j'=1}^{p} (r_{11jj'})(r_{*1jj'}) \\
&\times \Bigg[2(r_{11jj'})^2 + \beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 (r_{1\frac{1}{2}hj'})^2\Bigg].
\end{aligned}
$$

When $\mathbf{R}_1 = \mathbf{R}_*$, we have

$$
\begin{aligned}
\mu_{z0} &= \mu_{zA} \\
&= (n_1^2 - n_1 - 1)p^2 n_1^{-1}(n_1-1)^{-2} - [2n_1^2 + n_1 + 1](n_1-1)^{-3}\operatorname{tr}(\mathbf{R}_*^2) \\
&\quad + \frac{(n_1^2 - 3n_1)}{(n_1-1)^3} \sum_{j'=1}^{p}\sum_{j=1}^{p} (r_{*1jj'})^4 + \beta_1\Bigg[n_1 p (n_1-1)^{-2} \\
&\quad - 2n_1(n_1-1)^{-2} \sum_{h=1}^{p}\sum_{j=1}^{p} (r_{*\frac{3}{2}hj})(r_{*\frac{1}{2}hj})^3 \\
&\quad + 0.5(n_1^2 - 5n_1)(n_1-1)^{-3} \sum_{j=1}^{p} (r_{*2jj}) \sum_{h=1}^{p} (r_{*\frac{1}{2}hj})^4 \\
&\quad + 0.5(n_1^2 - 3n_1)(n_1-1)^{-3} \sum_{j=1}^{p}\sum_{j'=1}^{p} (r_{*1jj'})^2 \sum_{h=1}^{p} (r_{*\frac{1}{2}hj})^2 (r_{*\frac{1}{2}hj'})^2\Bigg].
\end{aligned}
$$

Expression of σ_{zA}^2:

$$
\begin{aligned}
\sigma_{zA}^2 &= 8n_1^{-1}\operatorname{tr}[(\mathbf{R}_1\mathbf{R}_*)^2] + 4\beta_1 n_1^{-1} \sum_{j=1}^{p} (\mathbf{e}_j^T \mathbf{R}_1^{1/2}\mathbf{R}_*\mathbf{R}_1^{1/2}\mathbf{e}_j)^2 \\
&\quad + 4n_1^{-1} \sum_{j=1}^{p}\sum_{j'=1}^{p} \mathbf{e}_j^T \mathbf{R}_1\mathbf{R}_*\mathbf{e}_j \mathbf{e}_{j'}^T \mathbf{R}_1\mathbf{R}_*\mathbf{e}_{j'}
\end{aligned}
$$

$$\times \left[2(r_{11jj'})^2 + 4\beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 (r_{1\frac{1}{2}hj'})^2 \right]$$

$$+ 4n_1^{-1} \left[2\operatorname{tr}(\mathbf{R}_1^4) + \beta_1 \sum_{j=1}^{p} (r_{12jj})^2 \right] + 4[n_1^{-1}\operatorname{tr}(\mathbf{R}_1^2)]^2$$

$$+ 4n_1^{-1} \sum_{j=1}^{p} \sum_{j'=1}^{p} (r_{12jj})(r_{12j'j'}) \left[2(r_{11jj'})^2 + \beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 (r_{1\frac{1}{2}hj'})^2 \right]$$

$$- 8n_1^{-1} \sum_{j=1}^{p} \mathbf{e}_j^T \mathbf{R}_1 \mathbf{R}_* \mathbf{e}_j \left[2\mathbf{e}_j^T \mathbf{R}_1 \mathbf{R}_* \mathbf{R}_1 \mathbf{e}_j + \beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 \mathbf{e}_h^T \mathbf{R}_1^{1/2} \mathbf{R}_* \mathbf{R}_1^{1/2} \mathbf{e}_h \right]$$

$$- 8n_1^{-1} \left[2\operatorname{tr}(\mathbf{R}_1^3 \mathbf{R}_*) + \beta_1 \sum_{j=1}^{p} \mathbf{e}_j^T \mathbf{R}_1^{1/2} \mathbf{R}_* \mathbf{R}_1^{1/2} \mathbf{e}_j (r_{12jj}) \right]$$

$$+ 8n_1^{-1} \sum_{j=1}^{p} (r_{12jj}) \left[2\mathbf{e}_j^T \mathbf{R}_1 \mathbf{R}_* \mathbf{R}_1 \mathbf{e}_j + \beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 \mathbf{e}_h^T \mathbf{R}_1^{1/2} \mathbf{R}_* \mathbf{R}_1^{1/2} \mathbf{e}_h \right]$$

$$+ 8n_1^{-1} \sum_{j=1}^{p} \mathbf{e}_j^T \mathbf{R}_1 \mathbf{R}_* \mathbf{e}_j \left[2(r_{13jj}) + \beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 (r_{12hh}) \right]$$

$$- 8n_1^{-1} \sum_{j=1}^{p} \sum_{j'=1}^{p} \mathbf{e}_j^T \mathbf{R}_1 \mathbf{R}_* \mathbf{e}_j (r_{12j'j'}) \left[2(r_{11jj'})^2 + \beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 (r_{1\frac{1}{2}hj'})^2 \right]$$

$$- 8n_1^{-1} \sum_{j=1}^{p} (r_{12jj}) \left[2r_{13jj} + \beta_1 \sum_{h=1}^{p} (r_{1\frac{1}{2}hj})^2 (r_{12hh}) \right].$$

When $\mathbf{R}_1 = \mathbf{R}_*$, we have $\sigma_{zA}^2 = 4[n_1^{-1}\operatorname{tr}(\mathbf{R}_*^2)]^2$.

A.2. Expressions of $\mu_{A\ell_1\ell_2}$, $\mu_{z\ell_1\ell_2}$, $\sigma^2_{A\ell_1\ell_2}$ and $\sigma_{A\ell_1\ell_2\ell_2\ell_3}$ in Theorems 3.1–3.4.

$$\mu_{A\ell_1\ell_2}$$

$$= -2\operatorname{tr}(\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}) + \sum_{i=1}^{2} a_{\ell_i} \Big\{ (n_{\ell_i}^2 - n_{\ell_i} + 2) n_{\ell_i}^{-1} \operatorname{tr}(\mathbf{R}_{\ell_i}^2)$$

$$+ (n_{\ell_i}^2 - n_{\ell_i} - 1) p^2 n_{\ell_i}^{-2} + \beta_{\ell_i} p$$

$$- 4\left[2\operatorname{tr}(\mathbf{R}_{\ell_i}^2) + \beta_{\ell_i} \sum_{h=1}^{p} \sum_{j=1}^{p} (r_{\ell_i \frac{3}{2} hj})(r_{\ell_i \frac{1}{2} hj})^3 \right]$$

$$+ 2\left[2\operatorname{tr}(\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}) + \beta_{\ell_i}\sum_{h=1}^{p}\sum_{j=1}^{p}\mathbf{e}_h^T\mathbf{R}_{\ell_i}^{1/2}\mathbf{R}_{\{\ell_1,\ell_2\}\setminus\{\ell_i\}}\mathbf{e}_j(r_{\ell_i\frac{1}{2}hj})^3\right]$$

$$+ 2(n_{\ell_i}-2)(n_{\ell_i}-1)^{-1}\left[2\operatorname{tr}(\mathbf{R}_{\ell_i}^2) + \beta_{\ell_i}\sum_{j=1}^{p}(r_{\ell_i 2jj})\sum_{h=1}^{p}(r_{\ell_i\frac{1}{2}hj})^4\right]$$

$$- 1.5\sum_{h=1}^{p}\mathbf{e}_h^T\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}\mathbf{e}_h\left[2 + \beta_{\ell_i}\sum_{j=1}^{p}(r_{\ell_i\frac{1}{2}hj})^4\right]$$

$$+ (n_{\ell_i}-2)(n_{\ell_i}-1)^{-1}$$

$$\times \sum_{h=1}^{p}\sum_{j=1}^{p}(r_{\ell_i 1hj})^2\left[2(r_{\ell_i 1hj})^2 + \beta_{\ell_i}\sum_{j'=1}^{p}(r_{\ell_i\frac{1}{2}hj'})^2(r_{\ell_i\frac{1}{2}jj'})^2\right]$$

$$- 0.5\sum_{h=1}^{p}\sum_{j=1}^{p}(r_{\ell_1 1hj})(r_{\ell_2 1hj})\left[2(r_{\ell_i 1hj})^2 + \beta_{\ell_i}\sum_{j'=1}^{p}(r_{\ell_i\frac{1}{2}hj'})^2(r_{\ell_i\frac{1}{2}jj'})^2\right]\Bigg\},$$

where $a_{\ell_i} = n_{\ell_i}/[(n_{\ell_i}-1)^2]$, $r_{\ell_i khj}$ is the (h, j) entry of $(\mathbf{R}_{\ell_i})^k$ for $k = 1/2, 1, 3/2, 2, 3$ and $\mathbf{R}_{\{\ell_1,\ell_2\}\setminus\{\ell_1\}} = \mathbf{R}_{\ell_2}$, $\mathbf{R}_{\{\ell_1,\ell_2\}\setminus\{\ell_2\}} = \mathbf{R}_{\ell_1}$.

When $\mathbf{R}_{\ell_1} = \mathbf{R}_{\ell_2} = \mathbf{R}$, we have

$$\begin{aligned}
\mu_{zk_1\ell_2} &= \mu_{Az_1z_2} \\
&= \sum_{i=1}^{2}\frac{n_{\ell_i}}{(n_{\ell_i}-1)^3}\left[\frac{(n_{\ell_i}^2 - n_{\ell_i} - 1)p^2(n_{\ell_i}-1)}{n_{\ell_i}^2} - \frac{(2n_{\ell_i}^2 + n_{\ell_i} + 1)}{n_{\ell_i}}\operatorname{tr}(\mathbf{R}^2)\right. \\
&\quad + (n_{\ell_i}-3)\sum_{h=1}^{p}\sum_{j=1}^{p}(r_{01hj})^4 + \beta_{\ell_i}p(n_{\ell_i}-1) - 2\beta_{\ell_i}(n_{\ell_i}-1) \\
&\quad \times \sum_{h=1}^{p}\sum_{j=1}^{p}(r_{0\frac{3}{2}hj})(r_{0\frac{1}{2}hj})^3 + 0.5(n_{\ell_i}-5)\beta_{\ell_i}\sum_{h=1}^{p}(r_{02hh})\sum_{j=1}^{p}(r_{0\frac{1}{2}hj})^4 \\
&\quad \left. + 0.5(n_{\ell_i}-3)\beta_{\ell_i}\sum_{h=1}^{p}\sum_{j=1}^{p}(r_{01hj})^2\sum_{j'=1}^{p}(r_{0\frac{1}{2}hj'})^2(r_{0\frac{1}{2}jj'})^2\right],
\end{aligned}$$

where r_{0khj} is the (h, j) entry of $\mathbf{R}_k$ for $k = 1/2, 1, 3/2, 2, 3$. We have $\sigma_{A\ell_1\ell_2}^2 = a_{A\ell_1\ell_2} + b_{A\ell_1\ell_2}$ where

$$\begin{aligned}
a_{A\ell_1\ell_2} &= 8n_{\ell_1}^{-1}\operatorname{tr}[(\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2})^2] + 4\beta_{\ell_1}n_{\ell_1}^{-1}\sum_{h=1}^{p}(\mathbf{e}_h^T\mathbf{R}_{\ell_1}^{1/2}\mathbf{R}_{\ell_2}\mathbf{R}_{\ell_1}^{1/2}\mathbf{e}_h)^2 \\
&\quad + 4n_{\ell_1}^{-1}\sum_{h=1}^{p}\sum_{j'=1}^{p}\mathbf{e}_h^T\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}\mathbf{e}_h\mathbf{e}_{j'}^T\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}\mathbf{e}_{j'}
\end{aligned}$$

$$
\begin{aligned}
&\times \Biggl[2(r_{\ell_1 1 h j'})^2 + 4\beta_{\ell_1} \sum_{j=1}^{p} (r_{\ell_1 \frac{1}{2} h j})^2 (r_{\ell_1 \frac{1}{2} j j'})^2\Biggr] \\
&+ 4n_{\ell_1}^{-1}\Biggl[2\operatorname{tr}(\mathbf{R}_{\ell_1}^4) + \beta_{\ell_1} \sum_{h=1}^{p} (r_{\ell_1 2 h h})^2\Biggr] + 4[n_{\ell_1}^{-1} \operatorname{tr}(\mathbf{R}_{\ell_1}^2)]^2 \\
&+ 4n_{\ell_1}^{-1} \sum_{h=1}^{p} \sum_{j'=1}^{p} (r_{\ell_1 2 h h})(r_{\ell_1 2 j' j'}) \\
&\times \Biggl[2(r_{\ell_1 1 h j'})^2 + \beta_{\ell_1} \sum_{j=1}^{p} (r_{\ell_1 \frac{1}{2} h j})^2 (r_{\ell_1 \frac{1}{2} j j'})^2\Biggr] \\
&- 8n_{\ell_1}^{-1} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2} \mathbf{e}_h \Biggl[2\mathbf{e}_h^T \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1} \mathbf{e}_h \\
&+ \beta_{\ell_1} \sum_{j=1}^{p} (r_{\ell_1 \frac{1}{2} h j})^2 \mathbf{e}_j^T \mathbf{R}_{\ell_1}^{1/2} \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1}^{1/2} \mathbf{e}_j\Biggr] \\
&- 8n_{\ell_1}^{-1}\Biggl[2\operatorname{tr}(\mathbf{R}_{\ell_1}^3 \mathbf{R}_{\ell_2}) + \beta_{\ell_1} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_1}^{1/2} \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1}^{1/2} \mathbf{e}_h (r_{\ell_1 2 h h})\Biggr] \\
&+ 8n_{\ell_1}^{-1} \sum_{h=1}^{p} (r_{\ell_1 2 h h}) \Biggl[2\mathbf{e}_h^T \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1} \mathbf{e}_h \\
&+ \beta_{\ell_1} \sum_{j=1}^{p} (r_{\ell_1 \frac{1}{2} h j})^2 \mathbf{e}_j^T \mathbf{R}_{\ell_1}^{1/2} \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1}^{1/2} \mathbf{e}_j\Biggr] \\
&+ 8n_{\ell_1}^{-1} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2} \mathbf{e}_h \Biggl[2(r_{\ell_1 3 h h}) + \beta_{\ell_1} \sum_{j=1}^{p} (r_{\ell_1 \frac{1}{2} h j})^2 (r_{\ell_1 2 j j})\Biggr] \\
&- 8n_{\ell_1}^{-1} \sum_{h=1}^{p} \sum_{j'=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2} \mathbf{e}_h (r_{\ell_1 2 j' j'}) \\
&\times \Biggl[2(r_{\ell_1 1 h j'})^2 + \beta_{\ell_1} \sum_{j=1}^{p} (r_{\ell_1 \frac{1}{2} h j})^2 (r_{\ell_1 \frac{1}{2} j j'})^2\Biggr] \\
&- 8n_{\ell_1}^{-1} \sum_{h=1}^{p} (r_{\ell_1 2 h h}) \Biggl[2(r_{\ell_1 3 h h}) + \beta_{\ell_1} \sum_{j=1}^{p} (r_{\ell_1 \frac{1}{2} h j})^2 (r_{\ell_1 2 j j})\Biggr]
\end{aligned}
$$

and

$$
\begin{aligned}
b_{A\ell_1\ell_2} = {} & 8n_{\ell_2}^{-1}\operatorname{tr}\big[(\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2})^2\big] + \frac{8}{n_{\ell_2}n_{\ell_1}}\big[\operatorname{tr}(\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2})\big]^2 \\
& + 4\beta_{\ell_2}n_{\ell_2}^{-1}\sum_{h=1}^{p}\big(\mathbf{e}_h^T\mathbf{R}_{\ell_2}^{1/2}\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}^{1/2}\mathbf{e}_h\big)^2 \\
& + 4n_{\ell_2}^{-1}\sum_{h=1}^{p}\sum_{j'=1}^{p}\mathbf{e}_h^T\mathbf{R}_{\ell_2}\mathbf{R}_{\ell_1}\mathbf{e}_h\mathbf{e}_{j'}^T\mathbf{R}_{\ell_2}\mathbf{R}_{\ell_1}\mathbf{e}_{j'} \\
& \times\left[2(r_{\ell_2 1hj'})^2 + \beta_{\ell_2}\sum_{j=1}^{p}(r_{\ell_2\frac{1}{2}hj})^2(r_{\ell_2\frac{1}{2}jj'})^2\right] \\
& + 4n_{\ell_2}^{-1}\left[2\operatorname{tr}\big(\mathbf{R}_{\ell_2}^4\big) + \beta_{\ell_2}\sum_{h=1}^{p}(r_{\ell_2 2hh})^2\right] + 4\big(n_{\ell_2}^{-1}\operatorname{tr}\mathbf{R}_{\ell_2}^2\big)^2 \\
& + 4n_{\ell_2}^{-1}\sum_{h=1}^{p}\sum_{j'=1}^{p}(r_{\ell_2 2hh})(r_{\ell_2 2j'j'}) \\
& \times\left[2(r_{\ell_2 1hj'})^2 + \beta_{\ell_2}\sum_{j=1}^{p}(r_{\ell_2\frac{1}{2}hj})^2(r_{\ell_2\frac{1}{2}jj'})^2\right] \\
& - 8n_{\ell_2}^{-1}\sum_{h=1}^{p}\mathbf{e}_h^T\mathbf{R}_{\ell_2}\mathbf{R}_{\ell_1}\mathbf{e}_h\Bigg[2\mathbf{e}_h^T\mathbf{R}_{\ell_2}\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}\mathbf{e}_h \\
& + \beta_{\ell_2}\sum_{j=1}^{p}(r_{\ell_2\frac{1}{2}hj})^2\mathbf{e}_j^T\mathbf{R}_{\ell_2}^{1/2}\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}^{1/2}\mathbf{e}_j\Bigg] \\
& - 8n_{\ell_2}^{-1}\left[2\operatorname{tr}\big(\mathbf{R}_{\ell_2}^3\mathbf{R}_{\ell_1}\big) + \beta_{\ell_2}\sum_{h=1}^{p}\mathbf{e}_h^T\mathbf{R}_{\ell_2}^{1/2}\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}^{1/2}\mathbf{e}_h(r_{\ell_2 2hh})\right] \\
& + 8n_{\ell_2}^{-1}\sum_{h=1}^{p}(r_{\ell_2 2hh})\Bigg[2\mathbf{e}_h^T\mathbf{R}_{\ell_2}\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}\mathbf{e}_h \\
& + \beta_{\ell_2}\sum_{j=1}^{p}(r_{\ell_2\frac{1}{2}hj})^2\mathbf{e}_j^T\mathbf{R}_{\ell_2}^{1/2}\mathbf{R}_{\ell_1}\mathbf{R}_{\ell_2}^{1/2}\mathbf{e}_j\Bigg] \\
& + 8n_{\ell_2}^{-1}\sum_{h=1}^{p}\mathbf{e}_h^T\mathbf{R}_{\ell_2}\mathbf{R}_{\ell_1}\mathbf{e}_h\left[2(r_{\ell_2 3hh}) + \beta_{\ell_2}\sum_{j=1}^{p}(r_{\ell_2\frac{1}{2}hj})^2(r_{\ell_2 2jj})\right] \\
& - 8n_{\ell_2}^{-1}\sum_{h=1}^{p}\sum_{j'=1}^{p}\mathbf{e}_h^T\mathbf{R}_{\ell_2}\mathbf{R}_{\ell_1}\mathbf{e}_h(r_{\ell_2 2j'j'})
\end{aligned}
$$

$$\times \left[2(r_{\ell_2 1hj'})^2 + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 \frac{1}{2} jj'})^2 \right]$$

$$- 8n_{\ell_2}^{-1} \sum_{h=1}^{p} (r_{\ell_2 2hh}) \left[2(r_{\ell_2 3hh}) + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 2jj}) \right].$$

When $\mathbf{R}_{\ell_1} = \mathbf{R}_{\ell_2} = \mathbf{R}$, we have $a_{A\ell_1\ell_2} = 4[n_{\ell_1}^{-1} \operatorname{tr}(\mathbf{R}^2)]^2$, $b_{A\ell_1\ell_2} = 8n_{\ell_1}^{-1} n_{\ell_2}^{-1} \times [\operatorname{tr}(\mathbf{R}^2)]^2 + 4[n_{\ell_2}^{-1} \operatorname{tr}(\mathbf{R}^2)]^2$ and $\sigma_{A\ell_1\ell_2}^2 = 4(n_{\ell_1}^{-1} + n_{\ell_2}^{-1})^2 [\operatorname{tr}(\mathbf{R}^2)]^2$.

For $\ell_1 \neq \ell_2 \neq k_3$, we have

$$\begin{aligned}
&\sigma_{A\ell_1\ell_2\ell_2\ell_3} \\
&\quad = 4n_{\ell_2}^{-1} \Bigg[2 \operatorname{tr}(\mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_3} \mathbf{R}_{\ell_2}) \\
&\qquad + \beta_{\ell_2} \sum_{j=1}^{p} \mathbf{e}_j^T \mathbf{R}_{\ell_2}^{1/2} \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2}^{1/2} \mathbf{e}_j \mathbf{e}_j^T \mathbf{R}_{\ell_2}^{1/2} \mathbf{R}_{\ell_3} \mathbf{R}_{\ell_2}^{1/2} \mathbf{e}_j \Bigg] \\
&\qquad + 4n_{\ell_2}^{-1} \sum_{h=1}^{p} \sum_{j'=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1} \mathbf{e}_h \mathbf{e}_{j'}^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_3} \mathbf{e}_{j'} \Bigg[2(r_{\ell_2 1hj'})^2 \\
&\qquad + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 \frac{1}{2} jj'})^2 \Bigg] \\
&\qquad + 4n_{\ell_2}^{-1} \left[2 \operatorname{tr}(\mathbf{R}_{\ell_2}^4) + \beta_{\ell_2} \sum_{h=1}^{p} (r_{\ell_2 2hh})^2 \right] + 4[n_{\ell_2}^{-1} \operatorname{tr}(\mathbf{R}_{\ell_2}^2)]^2 \\
&\qquad + 4n_{\ell_2}^{-1} \sum_{h=1}^{p} \sum_{j'=1}^{p} (r_{\ell_2 2hh})(r_{\ell_2 2j'j'}) \\
&\qquad \times \left[2(r_{\ell_2 1hj'})^2 + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 \frac{1}{2} jj'})^2 \right] \\
&\qquad - 2n_{\ell_2}^{-1} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_3} \mathbf{e}_h \Bigg[2\mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2} \mathbf{e}_h \\
&\qquad + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 \mathbf{e}_j^T \mathbf{R}_{\ell_2}^{1/2} \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2}^{1/2} \mathbf{e}_j \Bigg] \\
&\qquad - 2n_{\ell_2}^{-1} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1} \mathbf{e}_h \Bigg[2\mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_3} \mathbf{R}_{\ell_2} \mathbf{e}_h
\end{aligned}$$

$$+ \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 \mathbf{e}_j^T \mathbf{R}_{\ell_2}^{1/2} \mathbf{R}_{\ell_3} \mathbf{R}_{\ell_2}^{1/2} \mathbf{e}_j \Bigg]$$

$$- 4n_{\ell_2}^{-1} \Bigg[2\,\mathrm{tr}\big(\mathbf{R}_{\ell_2}^3 \mathbf{R}_{\ell_1}\big) + \beta_{\ell_2} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2}^{1/2} \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2}^{1/2} \mathbf{e}_h (r_{\ell_2 2hh}) \Bigg]$$

$$- 4n_{\ell_2}^{-1} \Bigg[2\,\mathrm{tr}\big(\mathbf{R}_{\ell_2}^3 \mathbf{R}_{\ell_3}\big) + \beta_{\ell_2} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2}^{1/2} \mathbf{R}_{\ell_3} \mathbf{R}_{\ell_2}^{1/2} \mathbf{e}_h (r_{\ell_2 2hh}) \Bigg]$$

$$+ 4n_{\ell_2}^{-1} \sum_{h=1}^{p} (r_{\ell_2 2hh}) \Bigg[2\mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2} \mathbf{e}_h$$

$$+ \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 \mathbf{e}_j^T \mathbf{R}_{\ell_2}^{1/2} \mathbf{R}_{\ell_1} \mathbf{R}_{\ell_2}^{1/2} \mathbf{e}_j \Bigg]$$

$$+ 4n_{\ell_2}^{-1} \sum_{h=1}^{p} (r_{\ell_2 2hh}) \Bigg[2\mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_3} \mathbf{R}_{\ell_2} \mathbf{e}_h$$

$$+ \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 \mathbf{e}_j^T \mathbf{R}_{\ell_2}^{1/2} \mathbf{R}_{\ell_3} \mathbf{R}_{\ell_2}^{1/2} \mathbf{e}_j \Bigg]$$

$$+ 4n_{\ell_2}^{-1} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1} \mathbf{e}_h \Bigg[2(r_{\ell_2 3hh}) + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 2jj}) \Bigg]$$

$$+ 4n_{\ell_2}^{-1} \sum_{h=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_3} \mathbf{e}_h \Bigg[2(r_{\ell_2 3hh}) + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 2jj}) \Bigg]$$

$$- 4n_{\ell_2}^{-1} \sum_{h=1}^{p} \sum_{j'=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_1} \mathbf{e}_h (r_{\ell_2 2j'j'})$$

$$\times \Bigg[2(r_{\ell_2 1hj'})^2 + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 \frac{1}{2} jj'})^2 \Bigg]$$

$$- 4n_{\ell_2}^{-1} \sum_{h=1}^{p} \sum_{j'=1}^{p} \mathbf{e}_h^T \mathbf{R}_{\ell_2} \mathbf{R}_{\ell_3} \mathbf{e}_h (r_{\ell_2 2j'j'})$$

$$\times \Bigg[2(r_{\ell_2 1hj'})^2 + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 \frac{1}{2} jj'})^2 \Bigg]$$

$$- 4n_{\ell_2}^{-1} \sum_{h=1}^{p} (r_{\ell_2 2hh}) \Bigg[2(r_{\ell_2 3hh}) + \beta_{\ell_2} \sum_{j=1}^{p} (r_{\ell_2 \frac{1}{2} hj})^2 (r_{\ell_2 2jj}) \Bigg].$$

When $\mathbf{R}_{\ell_1} = \mathbf{R}_{\ell_2} = \mathbf{R}_{\ell_3} = \mathbf{R}$, we have $\sigma_{A\ell_1\ell_2\ell_2\ell_3} = 4[n_{\ell_2}^{-1} \operatorname{tr}(\mathbf{R}^2)]^2$.

Acknowledgment. The authors thank the Editor, the Associate Editor and the referees for their constructive comments and suggestions that led to a significant improvement of this article.

SUPPLEMENTARY MATERIAL

Supplement to "Test for high-dimensional correlation matrices" (DOI: 10.1214/18-AOS1768SUPP; .pdf). This supplementary material consists of the technical proofs and additional numerical results.

REFERENCES

AITKIN, M. A. (1969). Some tests for correlation matrices. *Biometrika* **56** 443–446.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792

BARTLETT, M. S. and RAJALAKSHMAN, D. V. (1953). Goodness of fit tests for simultaneous autoregressive series. *J. Roy. Statist. Soc. Ser. B* **15** 107–124. MR0058181

BODWIN, K., ZHANG, K. and NOBEL, A. (2018). A testing based approach to the discovery of differentially correlated variable sets. *Ann. Appl. Stat.* **12** 1180–1203. MR3834299

BROWNE, M. W. (1978). The likelihood ratio test for the equality of correlation matrices. *Br. J. Math. Stat. Psychol.* **31** 209–217. MR0553468

CAI, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Ann. Rev. Stat. Appl.* **4** 423–446.

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618

CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557

CAI, T. T. and ZHANG, A. (2016). Inference for high-dimensional differential correlation matrices. *J. Multivariate Anal.* **143** 107–126. MR3431422

CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

COLE, N. (1968). The likelihood ratio test of the equality of correlation matrices. 1968-65: The L. L. Thurstone Psychometric Laboratory. Univ. North Carolina, Chapel Hill, NC.

DEBASHIS, P. and ALEXANDER, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718

GAO, J., HAN, X., PAN, G. and YANG, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402

GUPTA, A. K., JOHNSON, B. E. and NAGAR, D. K. (2013). Testing equality of several correlation matrices. *Rev. Colombiana Estadíst.* **36** 237–258. MR3242200

JENNRICH, R. I. (1970). An asymptotic χ^2 test for the equality of two correlation matrices. *J. Amer. Statist. Assoc.* **65** 904–912. MR0269023

KULLBACK, S. (1967). On testing correlation matrices. *J. Roy. Statist. Soc. Ser. C* **16** 80–85.

LARNTZ, K. and PERLMAN, M. D. (1985). A simple test for the equality of correlation matrices Technical Report 63.

LI, D. N. and XUE, L. Z. (2015). Joint limiting laws for high-dimensional independence tests. Available at arXiv:1512.08819V1.

QIU, Y. and CHEN, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.* **40** 1285–1314. MR3015026

SCHOTT, J. R. (1996). Testing for the equality of several correlation matrices. *Statist. Probab. Lett.* **27** 85–89. MR1394180

SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. MR2234197

SHAO, Q.-M. and ZHOU, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab.* **42** 623–648. MR3178469

ZHENG, S., CHENG, G., GUO, J. and ZHU, H. (2019). Supplement to "Test for high-dimensional correlation matrices." DOI:10.1214/18-AOS1768SUPP.

ZHENG, S., CHEN, Z., CUI, H. and LI, R. (2018). Hypothesis testing on linear structures of high dimensional covariance matrix. *Ann. Statist.* To appear.

ZHOU, C., HAN, F., ZHANG, X. S. and LIU, H. (2015). An Extreme-value approach for testing the equality of large U-statistic based correlation matrices. Available at arXiv:1502.03211.

SHURONG ZHENG
JIANHUA GUO
KLAS AND SCHOOL OF MATHEMATICS
AND STATISTICS
NORTHEAST NORMAL UNIVERSITY
CHANGCHUN, JILIN
CHINA
E-MAIL: zhengsr@nenu.edu.cn
jhguo@nenu.edu.cn

GUANGHUI CHENG
SCHOOL OF ECONOMICS AND STATISTICS
GUANGZHOU UNIVERSITY
GUANGZHOU
CHINA
E-MAIL: chenggh845@nenu.edu.cn

HONGTU ZHU
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA,
CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA
USA
E-MAIL: htzhu@email.unc.edu