*The Annals of Statistics*
2019, Vol. 47, No. 6, 3300–3334
https://doi.org/10.1214/18-AOS1779


# HYPOTHESIS TESTING ON LINEAR STRUCTURES OF HIGH-DIMENSIONAL COVARIANCE MATRIX

BY SHURONG ZHENG[*,1], ZHAO CHEN[†,2], HENGJIAN CUI[‡,3] AND RUNZE LI[§,4]

*Northeast Normal University*[*], *Fudan University*[†], *Pennsylvania State University*[‡] *and Capital Normal University*[§]

This paper is concerned with test of significance on high-dimensional covariance structures, and aims to develop a unified framework for testing commonly used linear covariance structures. We first construct a consistent estimator for parameters involved in the linear covariance structure, and then develop two tests for the linear covariance structures based on entropy loss and quadratic loss used for covariance matrix estimation. To study the asymptotic properties of the proposed tests, we study related high-dimensional random matrix theory, and establish several highly useful asymptotic results. With the aid of these asymptotic results, we derive the limiting distributions of these two tests under the null and alternative hypotheses. We further show that the quadratic loss based test is asymptotically unbiased. We conduct Monte Carlo simulation study to examine the finite sample performance of the two tests. Our simulation results show that the limiting null distributions approximate their null distributions quite well, and the corresponding asymptotic critical values keep Type I error rate very well. Our numerical comparison implies that the proposed tests outperform existing ones in terms of controlling Type I error rate and power. Our simulation indicates that the test based on quadratic loss seems to have better power than the test based on entropy loss.

**1. Introduction.** High-dimensional data analysis has become increasingly important in various research fields. Fan and Li (2006) gave a brief review of regularization methods to deal with several challenges in high-dimensional data analysis. Bai and Saranadasa (1996) demonstrated the impact of dimensionality for test of two-sample high-dimensional normal means. This paper aims to develop powerful tests for high-dimensional covariance structure without the normality assumption.

Received March 2018; revised August 2018.

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

[2]Chen is the corresponding author, and supported by NNSFC Grants 11690014 and 11690015, and partly supported by the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science, Ministry of Education.

[3]Supported by NNSFC Grants 11471223, 11071022 and 11231010.

[4]Supported by NSF Grants DMS-1512422 and DMS-1820702, and NIH Grants P50 DA039838 and U19AI089672.

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

*Key words and phrases.* Covariance matrix structure, random matrix theory, test of compound symmetric structure, test of banded structure, test of sphericity.

Test of covariance structure is of great importance in multivariate data analysis. Under normality assumption, various tests for covariance matrix have been developed in the classical multivariate analysis (see, e.g., Anderson (2003)). However, these tests become invalid when the dimension $p$ of data is large relative to the sample size $n$ (see Ledoit and Wolf (2002)). Alternatives to the classical tests of covariance structure have been developed in the literature (see Birke and Dette (2005), Srivastava (2005), Srivastava and Reid (2012)). Several authors have studied testing whether a covariance matrix equals the identity matrix. Johnstone (2001) derived the Tracy–Wisdom law of the largest eigenvalue of the sample covariance matrix for normal distribution with covariance matrix being the identity matrix and $p/n \to y \in (0, 1)$. Without normality assumption, Bai et al. (2009) proposed correcting the LRT for testing whether the covariance matrix equals a known one (or equivalently testing whether the covariance matrix equals the identity matrix), and derived the limiting null distribution when $p/n \to y \in (0, 1)$ by using results from modern random matrix theory (RMT) (see Bai and Silverstein (2004), Zheng (2012)). Wang et al. (2013) redefined the LRT when $y \in (0, 1)$. Wang (2014) further investigated the asymptotic power of the LRT. Jiang, Jiang and Yang (2012) studied a corrected LRT when $y \in (0, 1]$. They discussed the LRT for the case $y = 1$ and showed that the performance of the corrected LRT when $y = 1$ is quite different from that when $y \in (0, 1)$. Cai and Ma (2013) tested the covariance matrix being a given matrix from a minimax point of view and allowed $p/n \to \infty$.

Sphericity testing is another important problem under high-dimensional settings. When $p, n \to \infty$, Chen, Zhang and Zhong (2010) studied testing sphericity for high-dimensional covariance matrices. Wang and Yao (2013) also studied testing sphericity for large-dimensional data. Under the normality assumption, Jiang and Yang (2013) obtained the limiting null distributions of LRTs for test of sphericity, test of independence, the equality test of covariance matrices, and the identity test of covariance matrix using moment generating function technique, under the assumption that $p < n$ and $p/n \to y \in (0, 1]$. Jiang and Qi (2015) further obtained the limiting null distributions of test statistics studied in Jiang and Yang (2013) under the normality assumption and $p < n - c$ for some $0 \le c \le 4$. As an extension of test of sphericity, testing banded structure of covariance matrices has been considered. Cai and Jiang (2011) tested banded structure of covariance matrices by limiting law of coherence of random matrices. This test enjoys high power for sparse alternatives if $\log p = o(n^{1/3})$. Qiu and Chen (2012) studied testing banded structures based on U-statistics under the assumption $p/n \to y \in (0, \infty)$.

This paper intends to develop a unified framework for testing linear covariance structures when $p/n \to y \in (0, \infty)$ and without the normality assumption. Not only several commonly used structures such as test of sphericity, compound symmetric structure and banded structure are included, but also many more structures can be covered by selecting the proper basis matrices. To begin with, we propose estimating the parameters involved in the linear covariance structure by the squared loss; then develop two tests for these covariance structures based on the entropy loss and quadratic loss used for covariance matrix estimation in the classical multivariate analysis (see Muirhead (1982)). We demonstrate that many existing tests for specific covariance structures are special cases of the newly proposed tests. Furthermore, to establish the asymptotic theory of the proposed tests, we first study asymptotic properties of some useful functionals of high-dimensional sample covariance matrix. We further prove that these functionals converge in probability, and their joint distribution weakly converges to a bivariate normal distribution. These asymptotic results are of their own significance in spectral analysis of RMT. Finally, using these asymptotic results, we derive the limiting distributions of the two proposed tests under both null and alternative hypotheses, and the power functions of these two tests. We further show that the test based on quadratic loss is asymptotically unbiased in the sense that the power under the alternative hypothesis is always greater than the significance level.

We conduct Monte Carlo simulation study to examine the finite sample performance of the two tests. Our simulation results show that the limiting null distributions of the proposed tests approximate their null distributions quite well, and the corresponding asymptotic critical values keep Type I error rate very well. Our numerical comparison implies that the proposed tests outperform existing ones in terms of controlling Type I error rate and power. Our simulation indicates that the test based on quadratic loss seems to have higher power than the test based on entropy loss.

The rest of this paper is organized as follows. In Section 2, we propose an estimation procedure for parameters involved in the linear covariance matrix structure, and develop two tests for linear structure. We further derive the asymptotic distributions of these two tests under the null and alternative hypotheses. In Section 3, we conduct Monte Carlo simulation to compare the finite sample performance of the proposed tests with existing ones. Theoretical proofs and technical lemmas are given in Section 4.

## 2. Tests on linear structures of covariance matrices.

Suppose that $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is an independent and identically distributed random sample from a $p$-dimensional population $\mathbf{x}$ with mean $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$, and covariance matrix $\mathrm{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$. Following the commonly adopted assumptions in the literature of RMT (see Bai and Silverstein (2004)), we impose the following two assumptions.

ASSUMPTION A. Assume that the $p$-dimensional population $\mathbf{x}$ satisfies the independent component structure that can be represented as $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{w}$, where $\mathbf{w} = (w_1, \dots, w_p)^T$, and $w_1, \dots, w_p$ are independent and identically distributed and $E(w_j) = 0$, $E(w_j^2) = 1$ and $E(w_j^4) = \kappa < \infty$, for $1 \le j \le p$.

ASSUMPTION B. Denote by $y_{n-1} = p/(n-1)$. Assume that $y_{n-1} \to y \in (0, \infty)$.

Assumption A relaxes the normality assumption by imposing the moment conditions. This assumption is often used in random matrix theories. Regarding to the representation in Assumption A, it is natural to assume that $w_j$ is standardized so that $E(w_j) = 0$ and $E(w_j^2) = 1$. If $w_j$ is with finite kurtosis, then Assumption A is satisfied. Of course, multivariate normal distribution satisfies Assumption A. Many other distributions may also satisfy Assumption A.

Assumption B allows that $p$ diverges as $n$ grows to infinity. This assumption implies that we are interested in studying the asymptotic behaviors of test procedures under the statistical settings in which both the dimension $p$ and the sample size $n$ are allowed to tend to infinity. Assumption B allows that $p$ may be less than or greater than the sample size. Hereafter, we omit the subscript $n$ in $p_n$ for simplicity. Denote by $\bar{\mathbf{x}}$ and $\mathbf{S}_n$ the sample mean and sample covariance matrix, respectively. That is,

$$\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i, \qquad \mathbf{S}_n = (n-1)^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \tag{2.1}$$

2.1. *Estimation.* Linear structure for covariance matrix $\boldsymbol{\Sigma}$ means that $\boldsymbol{\Sigma}$ can be represented as a linear combination of prespecified symmetric $p \times p$ matrices $(\mathbf{A}_1, \dots, \mathbf{A}_K)$ with fixed and finite $K$. That is,

$$\boldsymbol{\Sigma} = \theta_1\mathbf{A}_1 + \theta_2\mathbf{A}_2 + \cdots + \theta_K\mathbf{A}_K, \tag{2.2}$$

where $\{\theta_j, j = 1, \dots, K\}$ are unknown parameters. Here, $\mathbf{A}_1, \dots, \mathbf{A}_K$ are a set of basis matrices, and they are assumed to be linearly independent. For example, Anderson (1973) provided various covariance matrices with different linear structures. In particular, the author showed that the covariance for $\mathbf{x} = \sum_{k=1}^{K}\mathbf{U}_k\boldsymbol{\zeta}_k + \mathbf{e}$ satisfies the linear covariance structure, where $\boldsymbol{\zeta}_k \sim N(\mathbf{0}, \theta_k\mathbf{I}_p)$, $\mathbf{e} \sim N(\mathbf{0}, \theta_0\mathbf{I}_p)$, $\zeta_1, \dots, \zeta_K, \mathbf{e}$ are independent and $\mathbf{I}_p$ is the identity matrix. Several other useful linear covariance structures are given in Section 2.4.

Under normality assumption, the parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)^T$ can be estimated by the maximum likelihood estimate. The theoretical property and related computational issue have been studied in Anderson (1973) and Zwiernik, Uhler and Richards (2017) when $p$ is fixed and finite. Without assuming a specific distribution on $\mathbf{W}$ such as the normality of $\mathbf{W}$, we propose estimating $\boldsymbol{\theta}$ by minimizing the following squared loss function:

$$\min_{\boldsymbol{\theta}} \operatorname{tr}(\mathbf{S}_n - \theta_1\mathbf{A}_1 - \cdots - \theta_K\mathbf{A}_K)^2. \tag{2.3}$$

Let $\mathbf{C}$ be a $K \times K$ matrix with $(i, j)$-element being $\operatorname{tr}\mathbf{A}_i\mathbf{A}_j$ and $\mathbf{a}$ be a $K \times 1$ vector with $j$th element being $\operatorname{tr}\mathbf{S}_n\mathbf{A}_j$. Further define $\mathbf{D} = \mathbf{C}^{-1}$. Minimizing (2.3) yields a least squares type estimator for $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = \mathbf{D}\mathbf{a}. \tag{2.4}$$

It can be shown that under Assumptions A and B, $\hat{\theta}_k = \theta_k + O_p(n^{-1})$, $k = 1, \dots, K$, by using (2.10) in Theorem 2.1 below.

2.2. *Tests*. In this section, we develop two tests for the linear structures of covariance matrices:

$$H_0 : \mathbf{\Sigma} = \theta_1\mathbf{A}_1 + \theta_2\mathbf{A}_2 + \cdots + \theta_K\mathbf{A}_K. \tag{2.5}$$

For simplicity, denote $\mathbf{\Sigma}_0 = \theta_1\mathbf{A}_1 + \theta_2\mathbf{A}_2 + \cdots + \theta_K\mathbf{A}_K$. A natural estimator of $\mathbf{\Sigma}$ is the sample covariance matrix $\mathbf{S}_n$. With the linear structure assumption, we may estimate $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ given in (2.4), and then under $H_0$, a natural estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}}_0 = \hat{\theta}_1\mathbf{A}_1 + \cdots + \hat{\theta}_K\mathbf{A}_K$. Let $L(\cdot,\cdot)$ be a loss function to measure the deviation between $\widehat{\mathbf{\Sigma}}_0$ and $\mathbf{S}_n$. Intuitively, we reject the null hypothesis if $L(\widehat{\mathbf{\Sigma}}_0, \mathbf{S}_n) > \delta_0$ for a given critical value $\delta_0$. Motivated by the entropy loss (EL) $L(\widehat{\mathbf{\Sigma}}_0, \mathbf{S}_n) = \operatorname{tr}\mathbf{S}_n\widehat{\mathbf{\Sigma}}_0^{-1} - \log(|\mathbf{S}_n\widehat{\mathbf{\Sigma}}_0^{-1}|) - p$ (James and Stein (1961), Muirhead (1982)), we propose our first test for $H_0$. For $p < n-1$,

$$T_{n1} = \operatorname{tr}\mathbf{S}_n\widehat{\mathbf{\Sigma}}_0^{-1} - \log\big(\big|\mathbf{S}_n\widehat{\mathbf{\Sigma}}_0^{-1}\big|\big) - p,$$

where $|\cdot|$ stands for the determinant of a matrix. Denote by $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$ the eigenvalues of $\mathbf{S}_n^{1/2}\widehat{\mathbf{\Sigma}}_0^{-1}\mathbf{S}_n^{1/2}$. Then we can write $T_{n1}$ as

$$T_{n1} = p\left(p^{-1}\sum_{j=1}^{p}\lambda_j - p^{-1}\sum_{j=1}^{p}\log\lambda_j\right) - p.$$

This motivates us to further extend the test to the situation that $p > n-1$ by defining

$$T_{n1} = (n-1)\left(p^{-1}\sum_{j=1}^{n-1}\lambda_j - (n-1)^{-1}\sum_{j=1}^{n-1}\log\lambda_j\right) - (n-1).$$

Define $q = \min\{p, n-1\}$. $T_{n1}$ can be written in a unified form for $p < n-1$ and $p \ge n-1$:

$$T_{n1} = q\left(p^{-1}\sum_{j=1}^{q}\lambda_j - q^{-1}\sum_{j=1}^{q}\log\lambda_j\right) - q. \tag{2.6}$$

Since this test is motivated by the entropy loss, we refer this test as *EL-test*. Motivated by the quadratic loss (QL), another popular loss function in covariance matrix estimation (see Haff (1980), Muirhead (1982), Olkin and Selliah (1977)), we propose our second test statistic

$$T_{n2} = \operatorname{tr}\big(\mathbf{S}_n\widehat{\mathbf{\Sigma}}_0^{-1} - \mathbf{I}_p\big)^2, \tag{2.7}$$

and refer the corresponding test as *QL-test*.

2.3. *New results on random matrix and limiting distributions of tests*. In order to derive the limiting distributions of $T_{n1}$ and $T_{n2}$, we develop new theory on large dimensional random matrix. In this section, we first present some useful theoretical results which are necessary to prove our main theorems.

Let $\{w_{ki}, k, i = 1, 2, \dots\}$ be a double array of independent and identically distributed random variables with mean 0 and variance 1. Let $\mathbf{w}_i = (w_{1i}, w_{2i}, \dots, w_{pi})^T$, and $\mathbf{w}_1, \dots, \mathbf{w}_n$ be independent and identically distributed random samples from a $p$-dimensional distribution with mean 0 and covariance matrix $\mathbf{I}_p$. To derive the limiting distributions of $T_{n1}$ and $T_{n2}$, we investigate the limiting distributions of the functionals of the eigenvalues of sample covariance matrix

$$\mathbf{F} = (n-1)^{-1} \sum_{i=1}^{n} (\mathbf{w}_i - \bar{\mathbf{w}})(\mathbf{w}_i - \bar{\mathbf{w}})^T, \tag{2.8}$$

where $\bar{\mathbf{w}} = n^{-1} \sum_{i=1}^{n} \mathbf{w}_i$. Throughout this paper, denote $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}^{1/2}$. Thus, it follows by Assumption A that

$$\mathbf{F} = \boldsymbol{\Gamma}^{-1} \mathbf{S}_n (\boldsymbol{\Gamma}^T)^{-1} \quad \text{and} \quad \mathbf{S}_n = \boldsymbol{\Gamma} \mathbf{F} \boldsymbol{\Gamma}^T. \tag{2.9}$$

To study the asymptotic behaviors of $T_{n1}$ and $T_{n2}$ under $H_0$ and $H_1$, we establish the asymptotic properties of $\mathbf{F}$. Theorems 2.1 and 2.2 will be repeatedly used in the proofs in (c) of Theorems 2.3 and 2.4. Suppose that $\lambda_j$'s, $j = 1, \dots, p$ are the real eigenvalues of $\mathbf{F}$. The empirical spectral distribution (ESD) of $\mathbf{F}$ is defined by $G_p(\lambda) = p^{-1} \sum_{j=1}^{p} I(\lambda_j \le \lambda)$, where $I(\cdot)$ is the indicator function. Note that the definition of ESD is suitable for both random and nonrandom matrices. Denote the spectral norm of a matrix $\mathbf{A}$ (the maximum eigenvalue) by $\|\mathbf{A}\|$ hereafter.

THEOREM 2.1. *Let $\mathbf{C}_k$, $k = 0, 1$, and 2, be $p \times p$ deterministic symmetric matrices. Under Assumptions A and B, the following statements are valid*:

(a) *If $\|\mathbf{C}_0\| = O(p)$, $\operatorname{tr} \mathbf{C}_0 = O(p)$ and $\operatorname{tr} \mathbf{C}_0^2 = O(p^2)$, then*

$$p^{-1} \operatorname{tr} \mathbf{F}\mathbf{C}_0 - p^{-1} \operatorname{tr} \mathbf{C}_0 = o_p(1). \tag{2.10}$$

(b) *If $\|\mathbf{C}_1\| = O(p)$, $\|\mathbf{C}_2\| = O(p)$, $\operatorname{tr}(\mathbf{C}_1^q) = O(p^q)$, $\operatorname{tr}(\mathbf{C}_2^q) = O(p^q)$, and $\operatorname{tr}(\mathbf{C}_1\mathbf{C}_2)^q = O(p^q)$ for $q = 1, 2$, then*

$$p^{-1} \operatorname{tr} \mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_2 - p^{-1} \operatorname{tr} \mathbf{C}_1\mathbf{C}_2 - y_{n-1}(p^{-1} \operatorname{tr} \mathbf{C}_1)(p^{-1} \operatorname{tr} \mathbf{C}_2) = o_p(1).$$

If we take $\mathbf{C}_1 = \mathbf{I}_p$, the identity matrix, in (b), then under the condition of (b), it follows that

$$p^{-1} \operatorname{tr} \mathbf{F}^2\mathbf{C}_2 - (1 + y_{n-1}) p^{-1} \operatorname{tr} \mathbf{C}_2 = o_p(1). \tag{2.11}$$

Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be $p \times p$ deterministic symmetric matrices. Define

$$u_1 = \begin{cases} 0 & \text{if } \|\mathbf{C}_1\| \text{ is bounded,} \\ 3/2 & \text{if } \|\mathbf{C}_1\| = O(p), \operatorname{tr}(\mathbf{C}_1^q) = O(p^q) \text{ for } q = 1, 2, 3, 4, \end{cases}$$

and

$$u_2 = \begin{cases} 0 & \text{if } \|\mathbf{C}_2\| \text{ is bounded,} \\ 1/2 & \text{if } \|\mathbf{C}_2\| = O(p), \operatorname{tr}(\mathbf{C}_2^q) = O(p^q) \text{ for } q = 1, 2. \end{cases}$$

Define $\boldsymbol{\mu}_n^{(1)} = (p^{-u_1}\mu_{n1}^{(1)}, p^{-u_2}\mu_{n2}^{(1)})^T$ with

$$\mu_{n1}^{(1)} = [\operatorname{tr}\mathbf{C}_1^2 + y_{n-1}p^{-1}(\operatorname{tr}\mathbf{C}_1)^2] + y_{n-1}p^{-1}\operatorname{tr}\mathbf{C}_1^2 + y_{n-1}(\kappa - 3)p^{-1}\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_1\mathbf{e}_i)^2,$$

where we call $\kappa = E(w_j^4)$, and $\mu_{n2}^{(1)} = \operatorname{tr}\mathbf{C}_2$ with $\mathbf{e}_i$ being the $i$th column of the $p \times p$ dimensional identity matrix. Further define $2 \times 2$ symmetric matrix $\boldsymbol{\Sigma}_n^{(1)}$ with $(i, j)$-th element being $\sigma_{nij}^{(1)}$ as follows:

$$\begin{aligned} \sigma_{n11}^{(1)} = p^{-2u_1}\Bigg\{& 8n^{-1}\operatorname{tr}\mathbf{C}_1^4 + 4(\kappa - 3)n^{-1}\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_1^2\mathbf{e}_i)^2 + 4(n^{-1}\operatorname{tr}\mathbf{C}_1^2)^2 \\ &+ 8(n^{-1}\operatorname{tr}\mathbf{C}_1)^2(n^{-1}\operatorname{tr}\mathbf{C}_1^2) + 4(\kappa - 3)(n^{-1}\operatorname{tr}\mathbf{C}_1)^2 n^{-1}\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_1\mathbf{e}_i)^2 \\ &+ 8(n^{-1}\operatorname{tr}\mathbf{C}_1)n^{-1}\Bigg[2\operatorname{tr}(\mathbf{C}_1^3) + (\kappa - 3)\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_1\mathbf{e}_i)(\mathbf{e}_i^T\mathbf{C}_1^2\mathbf{e}_i)\Bigg]\Bigg\}, \\ \sigma_{n22}^{(1)} = p^{-2u_2}\Bigg[& 2n^{-1}\operatorname{tr}\mathbf{C}_2^2 + (\kappa - 3)n^{-1}\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_2\mathbf{e}_i)^2\Bigg], \\ \sigma_{n12}^{(1)} = p^{-(u_1+u_2)}\Bigg\{& 4n^{-1}\operatorname{tr}(\mathbf{C}_1^2\mathbf{C}_2) + 2(\kappa - 3)n^{-1}\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_1^2\mathbf{e}_i)(\mathbf{e}_i^T\mathbf{C}_2\mathbf{e}_i) \\ &+ 2(n^{-1}\operatorname{tr}\mathbf{C}_1)n^{-1}\Bigg[2\operatorname{tr}(\mathbf{C}_1\mathbf{C}_2) + (\kappa - 3)\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_1\mathbf{e}_i)(\mathbf{e}_i^T\mathbf{C}_2\mathbf{e}_i)\Bigg]\Bigg\}. \end{aligned}$$

These expressions come from the proof of the following theorem.

THEOREM 2.2. *Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be $p \times p$ deterministic symmetric matrices and let $\boldsymbol{\Sigma}^{(1)} = \lim_{n\to\infty}\boldsymbol{\Sigma}_n^{(1)} > 0$ with either $\|\mathbf{C}_1\|$ being bounded or $\|\mathbf{C}_1\| = O(p)$, $\operatorname{tr}(\mathbf{C}_1^q) = O(p^q)$ for $q = 1, 2, 3, 4$, and either $\|\mathbf{C}_2\|$ being bounded or $\|\mathbf{C}_2\| = O(p)$, $\operatorname{tr}(\mathbf{C}_2^q) = O(p^q)$ for $q = 1, 2$. Under Assumptions A and B, it follows that*

$$\begin{pmatrix} p^{-u_1}\operatorname{tr}\mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_1 \\ p^{-u_2}\operatorname{tr}\mathbf{F}\mathbf{C}_2 \end{pmatrix} - \boldsymbol{\mu}_n^{(1)} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}^{(1)}).$$

*Hereafter, "$\xrightarrow{d}$" stands for convergence in distribution as $n \to \infty$.*

REMARK 2.1. Theorems 2.1 and 2.2 are established by using related techniques of RMT. The major goals in RMT are to investigate the asymptotic behaviors of the eigenvalues and the convergence of the sequence of ESDs. However, the limiting spectral distribution (LSD) is possibly defective. That is, total mass is less then one when some eigenvalues tend to $\infty$ (see Bai and Silverstein (2010)). Under the RMT framework, the existence of well-defined LSD is a common and necessary assumption. To apply the RMT, we need to impose the restrictive assumptions on the target matrices $\mathbf{C}_k$, $k = 0, 1$ and 2 in Theorems 2.1 and 2.2. The intuitive explanation is that the difference between the largest and the smallest eigenvalue is not too much and does not increase too fast as $p \to \infty$. Note that the $\mathbf{C}_k$, $k = 0, 1$ and 2 in these two theorems are not $\mathbf{\Sigma}$. Thus, Theorems 2.1 and 2.2 are applicable for a wide range of covariance structures.

We next study the asymptotic properties of the proposed test statistics. We first establish the limiting null distributions of EL-test and QL-test by RMT. Before presenting the main results, we provide an useful lemma about the spectral distributions of random matrices that will be used to establish the limiting distribution of $T_{n1}$. The technical details and proofs are given in the Supplementary Material (Zheng et al. (2019)). Define

$$\begin{aligned} \alpha_1(y) &= (1 - y^{-1}) \log(1 - y) - 1, \qquad y < 1, \\ \alpha_2(y) &= y^{-1}\alpha_1(y^{-1}) - y^{-1}\log(y^{-1}), \qquad y > 1; \end{aligned} \tag{2.12}$$

the mean functions

$$\begin{aligned} m_{12}(y) &= 0.5\log(1 - y) - 0.5(\kappa - 3)y, \qquad y < 1, \\ m_{22}(y) &= m_{12}(y^{-1}), \qquad y > 1; \end{aligned} \tag{2.13}$$

and the covariance functions

$$\begin{aligned} \nu_{11}(y) &= \nu_{12}(y) = \nu_{21}(y) = (\kappa - 1)y, \qquad y < 1, \\ \nu_{22}(y) &= -2\log(1 - y) + (\kappa - 3)y, \qquad y < 1; \\ \upsilon_{11}(y) &= (\kappa - 1)y, \qquad y > 1, \\ \upsilon_{12}(y) &= \upsilon_{21}(y) = \kappa - 1, \qquad y > 1, \\ \upsilon_{22}(y) &= -2\log(1 - y^{-1}) + (\kappa - 3)y^{-1}, \qquad y > 1. \end{aligned} \tag{2.14}$$

LEMMA 2.1. *Suppose that Assumptions A and B hold. Let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of $\mathbf{F}$ in (2.8) and $\mathbf{V}_{n1} = \{\nu_{ij}(y_{n-1})\}_{i,j}$ and $\mathbf{V}_{n2} = \{\upsilon_{ij}(y_{n-1})\}_{i,j}$ be $2 \times 2$ matrices whose entries are defined in (2.14), respectively. Then it follows the* *asymptotic normality*, (a) *if* $p < n-1$,

$$\mathbf{V}_{n1}^{-1/2}\left[\begin{pmatrix}\sum_{j=1}^{p}\lambda_j - p \\ \sum_{j=1}^{p}\log\lambda_j - p\alpha_1(y_{n-1})\end{pmatrix} - \begin{pmatrix}0\\ m_{12}(y_{n-1})\end{pmatrix}\right] \xrightarrow{d} N(\mathbf{0},\mathbf{I}_2); \tag{2.15}$$

*and* (b) *if* $p > n-1$,

$$\mathbf{V}_{n2}^{-1/2}\left[\begin{pmatrix}\sum_{j=1}^{n-1}\lambda_j - p \\ \sum_{j=1}^{n-1}\log\lambda_j - p\alpha_2(y_{n-1})\end{pmatrix} - \begin{pmatrix}0\\ m_{22}(y_{n-1})\end{pmatrix}\right] \xrightarrow{d} N(\mathbf{0},\mathbf{I}_2); \tag{2.16}$$

*and* (c) *if* $p = n-1$, *then* (2.15) *still holds by replacing* $\alpha_1(y_{n-1})$, $\nu_{ij}(y_{n-1})$ *and* $m_{12}(y_{n-1})$ *by* $-1$, $\nu_{ij}(y_n)$ *and* $m_{12}(y_n)$.

REMARK 2.2. Lemma 2.1 establishes the central limit theorem for the functional of eigenvalues of random matrix. It shows that the asymptotic behaviors of eigenvalues are quite different between the cases $p < n-1$ and $p > n-1$. When $y_{n-1} \to 0$ as $n \to \infty$, $\alpha_1(y_{n-1}) \to 0$, $m_{12}(y_{n-1}) \to 0$; When $y_{n-1} \to 1_-$, $\alpha_1(y_{n-1}) \to -1$, $m_{12}(y_{n-1}) = -\infty$; When $y_{n-1} \to 1_+$, $\alpha_2(y_{n-1}) \to -1$, $m_{22}(y_{n-1}) = -\infty$; When $y_{n-1} \to \infty$, $\alpha_2(y_{n-1}) \to 0$, $m_{22}(y_{n-1}) \to 0$.

We next present the limiting null distributions of the proposed tests.

THEOREM 2.3. *Suppose that Assumptions* A *and* B *hold. Denote by* $\sigma_{n1}^2(y) = -2y - 2\log(1-y)$, $y < 1$. *Using the same notation in Lemma* 2.1, *we have the following results under* $H_0$ *in* (2.5):

(a) *For* $p < n-1$,

$$\frac{T_{n1} + p\alpha_1(y_{n-1}) + m_{12}(y_{n-1})}{\sigma_{n1}(y_{n-1})} \xrightarrow{d} N(0,1). \tag{2.17}$$

*Moreover, for* $p = n-1$, $\sigma_{n1}^{-1}(y_n)\{T_{n1} - p + m_{12}(y_n)\} \xrightarrow{d} N(0,1)$.

(b) *For* $p > n-1$,

$$\frac{T_{n1} + p\alpha_2(y_{n-1}) + m_{22}(y_{n-1})}{\sigma_{n1}(y_{n-1}^{-1})} \xrightarrow{d} N(0,1). \tag{2.18}$$

(c) *Let* $\mathbf{B} = \sum_{k=1}^{K} d_k\mathbf{A}_k$, *where* $\mathbf{d} = (d_1,\ldots,d_K)^T = \mathbf{Dc}$ *and* $\mathbf{c}$ *be a* $K$-*dimensional vector with the kth entry being* $\operatorname{tr}\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1}$. *Assume that* $\sigma_{n2}^2 = y_{n-1}^2 -$

$(\kappa-1)y_{n-1}^3+2y_{n-1}^3p^{-1}\operatorname{tr}(\boldsymbol{\Sigma}_0\mathbf{B})^2+(\kappa-3)y_{n-1}^3p^{-1}\sum_{i=1}^p(\mathbf{e}_i^T\boldsymbol{\Gamma}^T\mathbf{B}\boldsymbol{\Gamma}\mathbf{e}_i)^2$ *has finite limit. It follows that*

$$\frac{T_{n2}-(p+\kappa-2)y_{n-1}}{2\sigma_{n2}}\xrightarrow{d}N(0,1). \tag{2.19}$$

REMARK 2.3. When the population is Gaussian, then $\kappa=3$ and the limiting distributions can be simplified as follows:

$$\frac{T_{n1}+p\alpha_1(y_{n-1})+m_{12}(y_{n-1})}{\sigma_{n1}(y_{n-1})}\xrightarrow{d}N(0,1),\qquad p<n-1,$$

$$\frac{T_{n1}-p+m_{12}(y_n)}{\sigma_{n1}(y_n)}\xrightarrow{d}N(0,1),\qquad p=n-1,$$

$$\frac{T_{n1}+p\alpha_2(y_{n-1})+m_{22}(y_{n-1})}{\sigma_{n1}(y_{n-1}^{-1})}\xrightarrow{d}N(0,1),\qquad p>n-1,$$

$$\frac{T_{n2}-(p+1)y_{n-1}}{2\sqrt{y_{n-1}^2-2y_{n-1}^3+2y_{n-1}^3p^{-1}\operatorname{tr}(\boldsymbol{\Sigma}_0\mathbf{B})^2}}\xrightarrow{d}N(0,1),$$

where

$$\alpha_1(y_{n-1})=[1-(y_{n-1})^{-1}]\log(1-y_{n-1})-1;$$

$$\alpha_2(y_{n-1})=y_{n-1}^{-1}[(1-y_{n-1})\log(1-y_{n-1}^{-1})-1]+y_{n-1}^{-1}\log y_{n-1};$$

$$m_{12}(y_{n-1})=0.5\log(1-y_{n-1});$$

$$m_{22}(y_{n-1})=0.5\log[1-(y_{n-1})^{-1}].$$

Moreover, especially for test of sphericity in Gaussian population, we have $[T_{n2}-(p+1)y_{n-1}]/(2y_{n-1})\xrightarrow{d}N(0,1)$. The statistic $T_{n1}$ is just from the original form of James and Stein's loss function.

REMARK 2.4. The EL-test $T_{n1}$ is just equivalent to the corrected LRT and the QL test $T_{n2}$ is just equivalent to the corrected John's test for test of sphericity (see Theorems 2.1 and 2.2 in Wang and Yao (2013)). Moreover, as we mentioned in the Introduction, the performance of the EL-test between $p<n-1$ and $p>n-1$ are quite different. This is the reason why the limiting null distribution of EL-test statistic is presented in part (a) and (b) of Theorem 2.3 separately.

The limiting null distributions can be used to construct the rejection regions of $T_{n1}$ and $T_{n2}$. We next establish the asymptotic power functions.

Suppose that under $H_1:\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_1$, where $\boldsymbol{\Sigma}_1$ cannot be represented as a linear combination of the selected matrices. Using the estimation procedure proposed in Section 2.3, there still exists the linear approximation for $\boldsymbol{\Sigma}_1$. Denoted by $\mathbf{a}_1^*=(\operatorname{tr}\boldsymbol{\Sigma}_1\mathbf{A}_1,\dots,\operatorname{tr}\boldsymbol{\Sigma}_1\mathbf{A}_K)^T$, $\boldsymbol{\theta}_1^*=(\theta_{11}^*,\dots,\theta_{K1}^*)^T=\mathbf{D}\mathbf{a}_1^*$, and $\boldsymbol{\Sigma}_1^*=$

$\theta_{11}^*\mathbf{A}_1 + \cdots + \theta_{K1}^*\mathbf{A}_K$ can be viewed as the best linear approximation to $\mathbf{\Sigma}_1$ under $H_1$. Recall that $\hat{\boldsymbol{\theta}}_1$ defined in (2.4) is the estimator of $\boldsymbol{\theta}^*$ based on observations and $\hat{\theta}_{k1} = \theta_{k1}^* + O_p(1/n), k = 1, \ldots, K$. Naturally, $\mathbf{I}_p - \mathbf{\Sigma}_1\mathbf{\Sigma}_1^{*-1}$ measures the approximation error between the alternative $\mathbf{\Sigma}_1$ and the null hypothesis. $\mathbf{\Sigma}_1\mathbf{\Sigma}_1^{*-1}$ is the core of the entropy loss and the quadratic loss. After symmetrization, we obtain the error denoted by $\mathbf{E} = \mathbf{\Sigma}_1^{*-1}(\mathbf{I}_p - \mathbf{\Sigma}_1\mathbf{\Sigma}_1^{*-1})$. Notice that $\mathbf{E} = 0$ under $H_0$.

Let $(h_1, \ldots, h_K)^T = \mathbf{D}(\operatorname{tr}\mathbf{E}\mathbf{A}_1, \ldots, \operatorname{tr}\mathbf{E}\mathbf{A}_K)^T$ and $G_p(t)$ be the ESD of $\mathbf{\Gamma}^T(\mathbf{\Sigma}_1^{*-1} + \sum_{k=1}^K h_k\mathbf{A}_k)\mathbf{\Gamma}$, where recall $\mathbf{\Gamma} = \mathbf{\Sigma}_1^{1/2}$ under $H_1$. Let $b_0 = y_{n-1} \times p^{-1}\operatorname{tr}\mathbf{\Sigma}_1\mathbf{\Sigma}_1^{*-1}$, $\mathbf{E}_0 = -(b_0\mathbf{I}_p - \mathbf{\Sigma}_1\mathbf{\Sigma}_1^{*-1})\mathbf{E} + b_0\mathbf{\Sigma}_1^{*-1}$ and $(h_1^*, \ldots, h_K^*)^T = \mathbf{D}(\operatorname{tr}\mathbf{E}_0\mathbf{A}_1, \ldots, \operatorname{tr}\mathbf{E}_0\mathbf{A}_K)^T$. Let $\mathbf{B}_* = \mathbf{\Sigma}_1^{*-1} + \sum_{k=1}^K h_k\mathbf{A}_k$ and $\mathbf{B}_{1*} = \mathbf{\Sigma}_1^{*-1} + \sum_{k=1}^K h_k^*\mathbf{A}_k$. Assume that $\|\mathbf{\Gamma}^T\mathbf{\Sigma}_1^{*-1}\mathbf{\Gamma}\| = O(p)$, $\operatorname{tr}[\mathbf{\Gamma}^T\mathbf{\Sigma}_1^{*-1}\mathbf{\Gamma}]^q = O(p^q)$ for $q = 1, 2, 3, 4$, $\operatorname{tr}[\mathbf{\Gamma}^T\mathbf{B}_*\mathbf{\Gamma}]^q = O(p)$ for $q = 1, 2$ and $\operatorname{tr}[\mathbf{\Gamma}^T\mathbf{B}_{1*}\mathbf{\Gamma}]^q = O(p^q)$ for $q = 1, 2$ and $G_p(t)$ has the nondegenerated LSD $G(t)$. Under such alternative hypothesis, we obtain the following limiting distributions in Theorem 2.4.

THEOREM 2.4. *Suppose that Assumptions* A *and* B *hold, and the limits of* $\sigma_{nj}^{(1)}$, $j = 1, 2, 3$, *exist. Then under* $H_1 : \mathbf{\Sigma} = \mathbf{\Sigma}_1$ *that cannot be represented as the linear combination of given matrices, it follows that*:

(a) *for* $p < n - 1$, *satisfying* $G_p(t) \to G(t)$,

$$\frac{T_{n1} - pF_1^{y_{n-1},G} - \mu_1^{(1)}}{\sigma_{n1}^{(1)}} \xrightarrow{d} N(0, 1),$$

(b) *for* $p > n - 1$, *satisfying* $G_p(t) \to G(t)$,

$$\frac{T_{n1} - pF_2^{y_{n-1},G} - \mu_2^{(1)}}{\sigma_{n2}^{(1)}} \xrightarrow{d} N(0, 1),$$

(c)

$$\frac{T_{n2} - \mu_3^{(1)}}{\sigma_{n3}^{(1)}} \xrightarrow{d} N(0, 1),$$

*where* $\mu_j^{(1)}$, $j = 1, 2, 3$, $F_j^{y_{n-1},G}$, $j = 1, 2$ *and* $\sigma_{nj}^{(1)}$, $j = 1, 2, 3$, *are given in the proof of Theorem* 2.4.

For fixed significance level $\alpha$, the corresponding power of the test based on the statistic $T_{n2}$ is $\beta_{T_{n2}}(\mathbf{\Sigma}_1) = 1 - \Phi((\mu_0 - \mu_3^{(1)})/\sigma_{n3}^{(1)} - 2q_{\alpha/2}\sigma/\sigma_{n3}^{(1)}) + \Phi((\mu_0 - \mu_3^{(1)})/\sigma_{n3}^{(1)} + 2q_{\alpha/2}\sigma/\sigma_{n3}^{(1)})$, where $q_{\alpha/2}$ is the $\alpha/2$ quantile of $N(0, 1)$, $\mu_0 = (p + \kappa - 2)y_{n-1}$ and $\sigma = \sigma_{n2}$ defined in Theorem 2.3. The following theorem shows that QL-test is asymptotically unbiased in the sense that $\beta_{T_{n2}}(\mathbf{\Sigma}_1) \geq \alpha \geq \beta_{T_{n2}}(\mathbf{\Sigma}_0)$, for any $\mathbf{\Sigma}_1$ belongs to certain alternative. Let $I_{\{\cdot\}}$ be an indicator function.

THEOREM 2.5. *Suppose that Assumptions* A *and* B *are satisfied and the limit of* $\sigma_{n3}^{(1)}$ *exists. Under* $H_1 : \mathbf{\Sigma} = \mathbf{\Sigma}_1$, *and* $\mathbf{\Sigma}_1$ *satisfies that the empirical spectral distribution* $p^{-1}\sum_{j=1}^{p} I_{\{\tilde{\lambda}_j \le t\}}$ *weakly converges to some distribution function with* $\tilde{\lambda}_j$*'s being the eigenvalues of* $\mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1}\mathbf{\Gamma} = \mathbf{I}_p + \mathbf{A}$, $\mathbf{A} \ge 0$ *and* $\operatorname{tr}\mathbf{A}^2 > \delta > 0$, *then for the prefixed significance level* $\alpha$,

$$\beta_{T_{n2}}(\mathbf{\Sigma}_1) > \alpha,$$

*when* $n$ *is sufficiently large and* $\delta$ *is any given small constant. Furthermore, if* $p^{-1}\operatorname{tr}\mathbf{A} \to c_1 \neq 0$, *then* $\beta_{T_{n2}} \to 1$ *as* $n \to \infty$.

REMARK 2.5. Note that $\mu_0 = (p+\kappa-2)y_{n-1}$. The the proof of Theorem 2.5 reveals a nice property that for large $p$, $\mu_3^{(1)} - \mu_0 \ge p\{(1+y)c_1^2 + 2yc_1\}$, which tends to $\infty$ at rate $p$. This implies that the power of QL-test increases to one quickly. This is consistent with our numerical studies in Section 3.

2.4. *Examples*. In this section, we demonstrate how the proposed tests of linear structures of covariance matrices provide a unified framework for many existing tests on covariance matrix by several examples, some of which are new to literature.

EXAMPLE 2.1. Test of sphericity has been well studied since the sphericity structure is the simplest linear structure of covariance matrix. Let $\mathbf{A}_1 = \mathbf{I}_p$. The test of sphericity is to test the null hypothesis

$$H_{10} : \mathbf{\Sigma} = \theta_1 \mathbf{A}_1 \tag{2.20}$$

for an unknown positive constant $\theta_1$ versus $H_{11} : \mathbf{\Sigma} \neq \theta_1\mathbf{A}_1$ for any positive constant $\theta_1$.

Under $H_{10}$, $\theta_1$ can be estimated by $\hat{\theta}_1 = p^{-1}\operatorname{tr}\mathbf{S}_n = \theta_1 + O_p(n^{-1})$ under Assumptions A and B. When $\kappa = 3$ (e.g., under normality assumption), by Theorem 2.3, we have the following limiting null distribution of $T_{n1}$. For $p < n-1$,

$$\frac{T_{n1} + (p-n+1.5)\log(1-y_{n-1}) - p}{\sqrt{-2y_{n-1} - 2\log(1-y_{n-1})}} \xrightarrow{d} N(0,1), \tag{2.21}$$

and for $p > n-1$,

$$\frac{T_{n1} + (n-0.5-p)\log(1-y_{n-1}^{-1}) - (n-1) + (n-1)\log(y_{n-1})}{\sqrt{-2y_{n-1}^{-1} - 2\log(1-y_{n-1}^{-1})}} \xrightarrow{d} N(0,1).$$

It can be easily verified that $T_{n1}$ equals $(2/n)$ times the logarithm of the LRT under normality assumption when $p < n-1$. The LRT has been well studied for fixed and finite dimension $p$ under normality assumption (Section 10.7 of Anderson (2003)). Recently Jiang and Yang (2013) derived the asymptotic distribution of the

LRT with $y \in (0, 1]$ for normal data. Chen, Zhang and Zhong (2010) demonstrated that the classical LRT may become invalid for high-dimensional data and proposed a test based on U-statistics with $p, n \to \infty$. Wang and Yao (2013) proposed the corrected LRT with $p/n \to (0, 1)$. Both Jiang and Yang (2013) and Wang and Yao (2013) derived the limiting null distribution, which is the same as that in (2.21), but Jiang and Yang (2013) imposes normality assumption.

For test of sphericity, the $T_{n2}$ becomes

$$T_{n2} = \operatorname{tr}\big[\mathbf{S}_n/(p^{-1}\operatorname{tr}\mathbf{S}_n) - \mathbf{I}_p\big]^2. \tag{2.22}$$

Under normality assumption, it follows by Theorem 2.3 that

$$\frac{T_{n2} - (p+1)y_{n-1}}{2y_{n-1}} \xrightarrow{d} N(0, 1).$$

The test statistic in (2.22) coincides with the corrected John's test proposed by Wang and Yao (2013) with $y \in (0, \infty)$. Wang and Yao (2013) further showed that the power of their proposed corrected John's test is similar to that of Chen, Zhang and Zhong (2010), and the corrected LRT had greater power than the corrected John's test and Chen, Zhang and Zhong's (2010) test, when the dimension $p$ is not large relative to the sample size $n$. But when $p$ is large relative to $n$ ($p < n$), the corrected LRT had smaller power than the corrected John's test and the test proposed by Chen, Zhang and Zhong (2010).

EXAMPLE 2.2. The compound symmetric structure of high-dimensional covariance matrix is another commonly used linear structure of covariance matrix. Let $\mathbf{A}_1 = \mathbf{I}_p$ and $\mathbf{A}_2 = \mathbf{1}_p\mathbf{1}_p^T$, where $\mathbf{1}_p$ stands for a $p$-dimensional column vector with all elements being 1. Testing compound symmetric structure is to test

$$H_{20} : \boldsymbol{\Sigma} = \theta_1\mathbf{A}_1 + \theta_2\mathbf{A}_2,$$

where $\theta_1 > 0$ and $-1/(p-1) < \theta_2/(\theta_1 + \theta_2) < 1$ versus $H_{21} : \boldsymbol{\Sigma} \neq \theta_1\mathbf{A}_1 + \theta_2\mathbf{A}_2$. Under normality assumption, Kato, Yamada and Fujikoshi (2010) studied the asymptotic behavior of the corresponding LRT when $p < n$, and Srivastava and Reid (2012) proposed a new test statistic for $H_{20}$ even if $p \geq n$. Without normality assumption, the EL and QL tests can be used to test the compound symmetric structure. By (2.4), $\theta_1$ and $\theta_2$ can be estimated by

$$\hat{\theta}_1 = p^{-1}(p-1)^{-1}\big(p\operatorname{tr}\mathbf{S}_n - \mathbf{1}_p^T\mathbf{S}_n\mathbf{1}_p\big),$$
$$\hat{\theta}_2 = p^{-1}(p-1)^{-1}\big(\mathbf{1}_p^T\mathbf{S}_n\mathbf{1}_p - \operatorname{tr}\mathbf{S}_n\big),$$

respectively. Thus, both $\hat{\theta}_1$ and $\hat{\theta}_2$ are $n$-consistent under Assumptions A and B. By Theorem 2.3, it follows that

$$\frac{T_{n2} - (p+\kappa-2)y_{n-1}}{2y_{n-1}} \xrightarrow{d} N(0, 1),$$

since $\sigma_{n2} = y_{n-1}$ in this example. While it seems that the limiting distributions of $T_{n1}$ given in (2.17) and (2.18) cannot be further simplified.

Under normality assumption, when $p < n - 1$, $T_{n1}$ equals $(2/n)$ times the logarithm of the LRT in Kato, Yamada and Fujikoshi (2010). Srivastava and Reid's (2012) method is different from Kato, Yamada and Fujikoshi (2010) and our proposed tests since Srivastava and Reid (2012) tested the compound symmetric structure of covariance matrix by testing the independence of random variables. The details are as follows. Let $\mathbf{G}$ be the orthogonal matrix with the first column being the $p^{-1/2}\mathbf{1}_p$ and the $i$th column being $i^{-1/2}(i-1)^{-1/2}(1, \dots, 1, -i+1, 0, \dots, 0)^T$. Thus, $\mathbf{G}^T\mathbf{\Sigma}\mathbf{G}$ is a diagonal matrix with the first diagonal element being $\theta_1[1 + (p-1)\theta_2]$ and the remaining diagonal elements being $\theta_1(1-\theta_2)$. Thus, Srivastava and Reid (2012) cast the testing problem $H_{20}$ as testing the independence of the first random variable and the remaining $p - 1$ random variables. Kato, Yamada and Fujikoshi's (2010) and Srivastava and Reid's (2012) tests are both proposed for the normal case. However, Kato, Yamada and Fujikoshi's (2010) test is only valid when $p < n$, and Srivastava and Reid's (2012) test still works when $p \geq n$.

EXAMPLE 2.3. Denote by $\sigma_{ij}$ the $(i, j)$-entry of $\mathbf{\Sigma}$. Here, a $(K-1)$-banded covariance matrix means that $\sigma_{i,j} = \sigma_{j,i} = \theta_k$ if $|i - j| = k - 1$, $k = 1, \dots, K$, and $\sigma_{ij} = 0$ if $|i - j| \geq K$. Let $\mathbf{A}_1 = \mathbf{I}_p$ and $\mathbf{A}_k$, $2 \leq k \leq K$, be a $p \times p$ matrix with $(i, j)$-element being 1 if $|i - j| = k - 1$ and 0 otherwise. Testing the $(K-1)$-banded covariance matrix is equivalent to test

$$H_{30} : \mathbf{\Sigma} = \theta_1\mathbf{A}_1 + \cdots + \theta_K\mathbf{A}_K,$$

where $\theta_k$'s are unknown parameters. By (2.4), we have

$$\hat{\theta}_1 = p^{-1}\operatorname{tr}\mathbf{S}_n,$$
$$\hat{\theta}_k = \frac{1}{2}(p-k+1)^{-1}\operatorname{tr}\mathbf{S}_n\mathbf{A}_k \qquad \text{for } 2 \leq k \leq K.$$

When $K$ is a finite positive integer, it can be shown that $\hat{\theta}_k = \theta_k + O_p(n^{-1})$ if $p/n \to y \in (0, \infty)$.

Qiu and Chen (2012) proposed a test for banded covariance matrix based on U-statistic. Their test is different from our proposed testing methods. For general $K$, the limiting null distributions in Theorem 2.3 cannot be further simplified. For $K = 2$, we may obtain a closed form for $d_k$, $k = 1, 2$ in Theorem 2.3(c). Specifically, let $\alpha = \theta_1(2\theta_2)^{-1}$, $\beta = -\alpha + \operatorname{sgn}(\alpha)\sqrt{\alpha^2 - 1}$. Then it follows by some calculations that

$$d_1 = \frac{\beta}{(1-\beta^2)[1-\beta^{2(p+1)}]\theta_2} \times \left[\frac{2(\beta^2 - \beta^{2(p+1)})}{p(1-\beta^2)} - 1 - \beta^{2(p+1)}\right],$$

$$d_2 = \frac{\beta}{(1-\beta^2)[1-\beta^{2(p+1)}]\theta_2} \times \left[\frac{4(\beta^3-\beta^{2p+1})}{p(1-\beta^2)} - 2(1-p^{-1})\beta - 2(1-p^{-1})\beta^{2p+1}\right].$$

Then $\mathbf{B} = d_1\mathbf{I}_p + d_2\mathbf{A}_2$. Thus, $\sigma_{n2}^2$ can be obtained. Then testing $H_{30}$ can be carried out by using the proposed EL and QL tests.

EXAMPLE 2.4. The factor model assumes that $\mathbf{X}$ can be represented as $\mathbf{X} = v_1\mathbf{U}_1 + \cdots + v_{K-1}\mathbf{U}_{K-1} + \boldsymbol{\varepsilon}$, where $v_1, \dots, v_{K-1}$ are random variables and $\mathbf{U}_k$, $k = 1, \dots, K-1$ are random vectors. Suppose that $v_1, \dots, v_{K-1}$, $\mathbf{U}_1, \dots, \mathbf{U}_{K-1}$ and $\boldsymbol{\varepsilon}$ are mutually independent and $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \theta_1\mathbf{I}_p$. Conditioning on $\mathbf{U}_k$, $k = 1, \dots, K-1$, the covariance matrix of factor model has the structure $\boldsymbol{\Sigma} = \theta_1\mathbf{I}_p + \theta_2\mathbf{U}_1\mathbf{U}_1^T + \cdots + \theta_K\mathbf{U}_{K-1}\mathbf{U}_{K-1}^T$, where $\theta_{k+1} = \mathrm{Var}(v_k)$ for $k = 1, \dots, K-1$. Let $\mathbf{A}_1 = \mathbf{I}_p$, and $\mathbf{A}_{k+1} = \mathbf{U}_k\mathbf{U}_k^T$ for $k = 1, \dots, K-1$. Thus, it is of interest to test

$$H_{40} : \boldsymbol{\Sigma} = \theta_1\mathbf{A}_1 + \cdots + \theta_K\mathbf{A}_K,$$

where $\theta_k$'s are unknown parameters. Generally, $\mathbf{U}_k$ are orthogonal such that $\mathbf{U}_s^T\mathbf{U}_t = p$ for $s = t$ and 0 for $s \neq t$. The parameters can be estimated by

$$\hat{\theta}_1 = (p-K+1)^{-1}\left(\mathrm{tr}\,\mathbf{S}_n - p^{-1}\sum_{k=1}^{K-1}\mathbf{U}_k^T\mathbf{S}_n\mathbf{U}_k\right),$$

$$\hat{\theta}_{k+1} = p^{-2}(\mathbf{U}_k^T\mathbf{S}_n\mathbf{U}_k - p\hat{\theta}_1)$$

for $k = 1, \dots, K-1$. Thus, when $K$ is finite and $p/n$ has a finite positive limit, $\hat{\theta}_k = \theta_k + O_p(n^{-1})$ under $H_{40}$ and Assumptions A and B. Then testing $H_{40}$ can be carried out by using the proposed EL and QL tests.

EXAMPLE 2.5. In this example, we consider testing the particular pattern of covariance matrix. For even $p$ which is fixed and finite, McDonald (1974) considered

$$H_{50} : \boldsymbol{\Sigma} = \begin{pmatrix} \theta_1\mathbf{I}_{p/2} + \theta_2\mathbf{1}_{p/2}\mathbf{1}_{p/2}^T & \theta_3\mathbf{I}_{p/2} \\ \theta_3\mathbf{I}_{p/2} & \theta_1\mathbf{I}_{p/2} + \theta_2\mathbf{1}_{p/2}\mathbf{1}_{p/2}^T \end{pmatrix}.$$

Let $\mathbf{A}_1 = \mathbf{I}_p$,

$$\mathbf{A}_2 = \begin{pmatrix} \mathbf{1}_{p/2}\mathbf{1}_{p/2}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{p/2}\mathbf{1}_{p/2}^T \end{pmatrix} \quad \text{and} \quad \mathbf{A}_3 = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{p/2} \\ \mathbf{I}_{p/2} & \mathbf{0} \end{pmatrix}.$$

Then $H_{50}$ can be written as $H_{50} : \boldsymbol{\Sigma} = \theta_1\mathbf{A}_1 + \theta_2\mathbf{A}_2 + \theta_3\mathbf{A}_3$. Thus, the proposed EL and QL tests can be used to $H_{50}$ with high-dimensional data.

## 3. Simulation studies and application.

3.1. *Practical implementation issues.* The limiting distributions derived in Theorem 2.3 involve the unknown parameter $\kappa$. Thus, we need to estimate $\kappa$ in practice. By Assumption A and some direct calculations, it follows that

$$\operatorname{Var}\{(\mathbf{x}-\boldsymbol{\mu})^T(\mathbf{x}-\boldsymbol{\mu})\} = 2\operatorname{tr}(\boldsymbol{\Sigma}^2) + (\kappa-3)\sum_{j=1}^{p}\sigma_{jj}^2, \tag{3.1}$$

where $\sigma_{jj}$ is the $j$th diagonal element of $\boldsymbol{\Sigma}$. This enables us to construct a moment estimator for $\kappa$. Specifically, we estimate $\sigma_{jj}$ by $\hat{\sigma}_{jj} = s_{jj}$, where $s_{jj}$ is the $j$th diagonal element of $\mathbf{S}_n$. A natural estimator for $\operatorname{Var}\{(\mathbf{x}-\boldsymbol{\mu})^T(\mathbf{x}-\boldsymbol{\mu})\}$ is

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

Under $H_0$, a natural estimator of $\boldsymbol{\Sigma}$ is $\hat{\boldsymbol{\Sigma}}_0 = \sum_{k=1}^{K}\hat{\theta}_k\mathbf{A}_k$ defined in Section 2.2. As a result, we may estimate $\operatorname{tr}\boldsymbol{\Sigma}^2$ by using $\operatorname{tr}\hat{\boldsymbol{\Sigma}}_0^2$. Thus, we may estimate $\kappa$ by

$$\hat{\kappa}_0 = 3 + \frac{\widehat{V} - 2\operatorname{tr}(\hat{\boldsymbol{\Sigma}}_0^2)}{\sum_{j=1}^{p}s_{jj}^2}. \tag{3.2}$$

It can be shown that $\hat{\kappa}_0$ is a consistent estimator of $\kappa$ under $H_0$. The corresponding test statistics control Type I error rate very well in our simulation study.

As shown in Theorem 2.2, in general, $p^{-1}\operatorname{tr}\mathbf{S}_n^2 - p^{-1}\operatorname{tr}\boldsymbol{\Sigma}^2$ does not tend to zero. This implies that $\operatorname{tr}\mathbf{S}_n^2$ may not serve as an estimator of $\operatorname{tr}\boldsymbol{\Sigma}^2$. According to Theorem 2.2 and ignoring the higher order term, a natural estimator for $\operatorname{tr}\boldsymbol{\Sigma}^2$ is

$$\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)} = (n-1)\{\operatorname{tr}(\mathbf{S}_n^2) - (n-1)^{-1}[\operatorname{tr}(\mathbf{S}_n)]^2\}/n.$$

This estimator is calibrated by the ratio $p/(n-1) = y_{n-1}$. This leads to another estimator of $\kappa$ given by

$$\hat{\kappa}_1 = 3 + \frac{n\widehat{V} - 2\{(n-1)\operatorname{tr}(\mathbf{S}_n^2) - [\operatorname{tr}(\mathbf{S}_n)]^2\}}{n\sum_{j=1}^{p}s_{jj}^2}. \tag{3.3}$$

Chen and Qin (2010) also studied the issue of estimation of $\operatorname{tr}(\boldsymbol{\Sigma}^2)$ and proposed the following estimator:

$$\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)} = \frac{1}{n(n-1)}\operatorname{tr}\left[\sum_{j\neq k}^{n}(\mathbf{x}_j - \bar{\mathbf{x}}_{(j,k)})\mathbf{x}_j^T(\mathbf{x}_k - \bar{\mathbf{x}}_{(j,k)})\mathbf{x}_k^T\right], \tag{3.4}$$

where $\bar{\mathbf{x}}_{(j,k)}$ is the sample mean after excluding $\mathbf{x}_j$ and $\mathbf{x}_k$. This leads to another estimator of $\kappa$:

$$\hat{\kappa}_2 = 3 + \frac{\widehat{V} - 2n^{-1}(n-1)^{-1}\operatorname{tr}[\sum_{j\neq k}^{n}(\mathbf{x}_j - \bar{\mathbf{x}}_{(j,k)})\mathbf{x}_j^T(\mathbf{x}_k - \bar{\mathbf{x}}_{(j,k)})\mathbf{x}_k^T]}{\sum_{j=1}^{p}s_{jj}^2}. \tag{3.5}$$

We compare the performance of $\hat{\kappa}_0$, $\hat{\kappa}_1$ and $\hat{\kappa}_2$ by Monte Carlo simulation study. Simulation results are reported in the Supplementary Material (Zheng et al.

(2019)). From our numerical comparison, $\hat{\kappa}_0$ performs very well across all scenarios of all simulation examples in Section 3.2 and its sample standard deviation is much less than those of $\hat{\kappa}_1$ and $\hat{\kappa}_2$. This implies that $\hat{\kappa}_0$ is more stable than $\hat{\kappa}_1$ and $\hat{\kappa}_2$. Thus, we will use $\hat{\kappa}_0$ throughout our numerical examples in Section 3.2.

3.2. *Numerical studies.* We illustrate the proposed testing procedure by a real data example in the Supplemental Material (Zheng et al. (2019)). In this section, we focus on assessing the finite sample performance of the proposed tests including their Type I error rates and powers. All simulations are conducted by using R code. We generate $n$ random samples from a population $\mathbf{x} = \mathbf{\Sigma}^{1/2}\mathbf{w}$, where $\mathbf{\Sigma}$ will be set according to the hypothesis to be tested, and $\mathbf{w}$ is defined in the previous section. In order to examine the performance of the proposed tests under different distributions, we consider the elements of $\mathbf{w}$ being independent and identically distributed as (a) N(0, 1) or (b) Gamma(4, 2)$-2$. Both distributions have means 0 and variances 1. For each setting, we conduct 1000 Monte Carlo simulations. The Monte Carlo simulation error rate is $1.96\sqrt{0.05 \times 0.95/1000} \approx 0.0135$ at level 0.05. In the numerical studies, we consider four different covariance matrix structures, which have been studied in the literature.

EXAMPLE 3.1. This example is designed to compare the performance of proposed testing procedures and the test proposed in Srivastava and Reid (2012) for hypothesis $H_{20}$ in Example 2.2. We set the covariance matrix structure as $\mathbf{\Sigma} = \theta_1 \mathbf{I}_p + \theta_2 \mathbf{1}_p \mathbf{1}_p^T + \theta_3 \mathbf{u}_p \mathbf{u}_p^T$, where $\mathbf{u}_p$ is a $p$-dimensional random vector following uniform distribution over $[-1, 1]$. The third term is to examine the empirical power when $\theta_3 \neq 0$. In our simulation, we set $(\theta_1, \theta_2) = (6, 1)$ and $\theta_3 = 0.0, 0.5, 1.0$, respectively. We set $\theta_3 = 0$ to examine Type I error rates and $\theta_3 = 0.5, 1.0$ to study the powers of the proposed tests. The sample size is set as $n = 100, 200$ and the dimension is taken to be $p = 50, 100, 500, 1000$. The percentages of rejecting $H_{20}$ at level 0.05 over 1000 simulations are summarized in Table 1, where the labels QL, EL and SR stand for the QL-test, the EL-test and the test proposed by Srivastava and Reid (2012), respectively. The top panel with $\theta_3 = 0$ in Table 1 is Type I error rates for different testing methods. Table 1 indicates that both QL and EL tests retain Type I error rates reasonably well across different sample sizes and dimensions. As Srivastava and Reid (2012) mentioned, SR test can control Type I error rate under normal assumption. But when the population distribution departures from the normality, SR test fails to control Type I error rate, even the sample size increases from 100 to 200. This is expected since the SR test is derived based on multivariate normality assumption. As the sample size increases from $n = 100$ to $n = 200$, both QL-test and EL-test control Type I error rate better. The empirical powers are listed in the panel with $\theta_3 = 0.5$ or 1.0 in Table 1, from which we can see that QL-test has higher power than EL-test for all cases in this example, and SR test for most cases in this example. For normal samples with $p = 50$ and 100,

TABLE 1
*Simulation results for $H_{20}$ (in percentage of rejecting $H_{20}$ over 1000 replications)*

| | | | $W_j \sim N(0,1)$ | | | | $W_j \sim \text{Gamma}(4,2)-2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_3$ | $n$ | Test | $p = 50$ | 100 | 500 | 1000 | 50 | 100 | 500 | 1000 |
| 0 | 100 | QL | 5.23 | 5.40 | 5.12 | 5.19 | 6.48 | 5.99 | 5.64 | 5.41 |
| | | EL | 5.32 | 6.51 | 5.12 | 5.18 | 5.77 | 6.35 | 5.46 | 5.54 |
| | | SR | 4.90 | 5.01 | 4.91 | 4.98 | 9.60 | 8.96 | 8.15 | 8.04 |
| 0.5 | 100 | QL | 40.25 | 80.58 | 100.0 | 100.0 | 41.01 | 80.70 | 100.0 | 100.0 |
| | | EL | 13.42 | 11.38 | 99.78 | 100.0 | 13.74 | 11.42 | 99.74 | 100.0 |
| | | SR | 24.46 | 59.04 | 99.99 | 100.0 | 41.22 | 73.86 | 100.0 | 100.0 |
| 1 | 100 | QL | 95.88 | 99.97 | 100.0 | 100.0 | 95.98 | 99.97 | 100.0 | 100.0 |
| | | EL | 53.53 | 29.97 | 100.0 | 100.0 | 53.71 | 30.18 | 100.0 | 100.0 |
| | | SR | 87.90 | 99.67 | 100.0 | 100.0 | 93.61 | 99.87 | 100.0 | 100.0 |
| 0 | 200 | QL | 5.22 | 5.14 | 5.12 | 5.19 | 6.32 | 5.78 | 5.31 | 5.34 |
| | | EL | 5.18 | 5.12 | 5.05 | 5.13 | 5.94 | 5.42 | 5.23 | 5.31 |
| | | SR | 4.98 | 4.93 | 4.93 | 5.03 | 9.95 | 9.23 | 8.44 | 8.43 |
| 0.5 | 200 | QL | 79.86 | 99.32 | 100.0 | 100.0 | 78.56 | 99.28 | 100.0 | 100.0 |
| | | EL | 42.00 | 58.62 | 100.0 | 100.0 | 41.22 | 58.62 | 100.0 | 100.0 |
| | | SR | 61.74 | 95.79 | 100.0 | 100.0 | 75.78 | 98.24 | 100.0 | 100.0 |
| 1 | 200 | QL | 99.98 | 100.0 | 100.0 | 100.0 | 99.96 | 100.0 | 100.0 | 100.0 |
| | | EL | 97.23 | 99.53 | 100.0 | 100.0 | 96.86 | 99.55 | 100.0 | 100.0 |
| | | SR | 99.81 | 100.0 | 100.0 | 100.0 | 99.91 | 100.0 | 100.0 | 100.0 |

SR test is more powerful than EL-test. We notice that the empirical power of our proposed methods increases as the dimension increases. It is consistent with our theoretical results. The power of all three tests increases significantly when the value of $\theta_3$ increases from 0.5 to 1 or the sample size $n$ increases from 100 to 200.

In summary, QL-test performs the best in terms of retaining Type I error rate and power. The SR test cannot control Type I error rate for nonnormal samples. EL-test can control Type I error rate well, but is less powerful than QL-test. Under normality assumption, EL-test is equivalent to the LRT test, which is the most powerful test in the traditional setting. For a high-dimensional setting, the EL-test corresponds to the corrected LRT, whose power can be improved by the QL-test for $H_{k0}$, $k = 2, 3, 4$ and 5 from this example and simulation examples below. Additional numerical comparison with a test proposed by Zhong et al. (2017) is given in Section S.5.2 in Zheng et al. (2019). The proposed QL- and EL-test both outperform the test proposed by Zhong et al. (2017).

EXAMPLE 3.2. To test covariance matrix structure in $H_{30}$, the banded covariance structure, we construct a banded matrix defined in Example 2.3 with width of band $K = 3$. Therefore, the null hypothesis $H_{30}$ has the linear decomposition $\mathbf{\Sigma} = \theta_1 \mathbf{I}_p + \theta_2 \mathbf{A}_2 + \theta_3 \mathbf{A}_3 + \theta_4 \mathbf{u}_p \mathbf{u}_p{}^T$, where $\mathbf{A}_2$ and $\mathbf{A}_3$ are defined in

TABLE 2

*Simulation results for $H_{30}$ (in percentage of rejecting $H_{30}$ over 1000 replications)*

| | | | $W_j \sim N(0,1)$ | | | | $W_j \sim$ Gamma(4, 2)−2 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_4$ | $n$ | Test | $p = 50$ | 100 | 500 | 1000 | 50 | 100 | 500 | 1000 |
| 0 | 100 | QL | 5.30 | 5.19 | 5.29 | 5.34 | 6.61 | 6.19 | 5.83 | 5.89 |
| | | EL | 5.31 | 6.34 | 5.20 | 5.36 | 5.83 | 6.32 | 5.53 | 5.81 |
| | | QC | 5.00 | 5.50 | 5.50 | 5.90 | 5.00 | 5.80 | 6.10 | 5.20 |
| 0.5 | 100 | QL | 46.01 | 84.91 | 100.0 | 100.0 | 45.13 | 83.62 | 100.0 | 100.0 |
| | | EL | 15.20 | 12.23 | 99.90 | 100.0 | 15.35 | 12.13 | 99.90 | 100.0 |
| | | QC | 17.00 | 70.00 | 100.0 | 100.0 | 24.00 | 85.30 | 100.0 | 100.0 |
| 1 | 100 | QL | 97.49 | 99.98 | 100.0 | 100.0 | 96.96 | 99.98 | 100.0 | 100.0 |
| | | EL | 60.45 | 33.68 | 100.0 | 100.0 | 59.80 | 33.71 | 100.0 | 100.0 |
| | | QC | 83.00 | 100.0 | 100.0 | 100.0 | 78.00 | 99.90 | 100.0 | 100.0 |
| 0 | 200 | QL | 5.25 | 5.19 | 5.16 | 5.10 | 6.44 | 5.84 | 5.56 | 5.55 |
| | | EL | 5.24 | 5.15 | 5.04 | 4.94 | 6.01 | 5.46 | 5.25 | 5.26 |
| | | QC | 6.00 | 5.40 | 5.50 | 4.50 | 4.00 | 4.50 | 5.80 | 5.10 |
| 0.5 | 200 | QL | 85.30 | 99.66 | 100.0 | 100.0 | 84.09 | 99.64 | 100.0 | 100.0 |
| | | EL | 48.46 | 65.39 | 100.0 | 100.0 | 47.62 | 65.20 | 100.0 | 100.0 |
| | | QC | 49.00 | 100.0 | 100.0 | 100.0 | 49.00 | 99.70 | 100.0 | 100.0 |
| 1 | 200 | QL | 99.99 | 100.0 | 100.0 | 100.0 | 99.98 | 100.0 | 100.0 | 100.0 |
| | | EL | 98.27 | 99.80 | 100.0 | 100.0 | 98.00 | 99.76 | 100.0 | 100.0 |
| | | QC | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

Example 2.3 and $\mathbf{u}_p$ is generated by the same way in Example 3.1. We take $(\theta_1, \theta_2, \theta_3) = (6.0, 1.0, 0.5)$ and $\theta_4 = 0$ to examine Type I error rates and take $\theta_4 = 0.5, 1$ to examine the powers. In the simulation studies, we still set sample size $n = 100, 200$ and dimension $p = 50, 100, 500, 1000$. The percentages of rejecting $H_{30}$ at level 0.05 over 1000 simulations are listed in Table 2. In this example, we compare the test proposed by Qiu and Chen (2012) for the banded covariance matrix with our proposed tests, and refer their test as "QC" test hereinafter. Table 2 indicates that QL, EL and QC tests control Type I error rates well. QC test is supposed to control Type I error rates well and has high power since it is particularly proposed for testing banded matrix. Table 2 indicates that QL-test has higher power than QC test in our simulation settings, in particular, for $p = 50$. From our simulation experience, we find that QC test requires more computing time than QL and EL tests since QC test is a $U$-statistic method.

EXAMPLE 3.3. In this example, we examine Type I error rates and empirical powers of the proposed tests for $H_{40}$ and $H_{50}$ defined in Examples 2.4 and 2.5, respectively. We first investigate the performance of the QL and EL tests for $H_{40}$. We generate several mutually orthogonal factors. Suppose that $\mathbf{u}_k^*$, $k = 1, \dots, K$ are independent and identically distributed random vectors following $N(\mathbf{0}, \mathbf{I}_p)$. Let

TABLE 3
*Simulation results for Example 3.3 (in percentage of rejecting null hypothesis over 1000 replications)*

| | | | $W_j \sim N(0,1)$ | | | | $W_j \sim \text{Gamma}(4,2)-2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_4$ | $n$ | Test | $p=50$ | 100 | 500 | 1000 | 50 | 100 | 500 | 1000 |
| | | | Results for $H_{40}$ | | | | | | | |
| 0 | 100 | QL | 5.46 | 5.46 | 6.02 | 6.27 | 6.89 | 6.53 | 6.58 | 6.94 |
| | | EL | 5.40 | 6.40 | 5.79 | 6.20 | 6.03 | 6.42 | 6.30 | 6.52 |
| 0.5 | 100 | QL | 99.99 | 100.0 | 100.0 | 100.0 | 99.96 | 100.0 | 100.0 | 100.0 |
| | | EL | 97.58 | 86.91 | 100.0 | 100.0 | 97.38 | 87.25 | 100.0 | 100.0 |
| 1 | 100 | QL | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | EL | 100.0 | 99.93 | 100.0 | 100.0 | 99.99 | 99.93 | 100.0 | 100.0 |
| 0 | 200 | QL | 5.32 | 5.28 | 5.42 | 5.52 | 6.57 | 6.05 | 5.89 | 6.05 |
| | | EL | 5.30 | 5.22 | 5.33 | 5.65 | 6.15 | 5.61 | 5.60 | 5.65 |
| 0.5 | 200 | QL | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | EL | 99.99 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 1 | 200 | QL | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | EL | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | Results for $H_{50}$ | | | | | | | |
| 0 | 100 | QL | 5.28 | 5.19 | 5.23 | 5.41 | 6.59 | 6.15 | 5.84 | 6.16 |
| | | EL | 5.27 | 6.33 | 5.17 | 5.48 | 5.85 | 6.35 | 5.61 | 5.73 |
| 0.5 | 100 | QL | 99.64 | 100.0 | 100.0 | 100.0 | 99.61 | 100.0 | 100.0 | 100.0 |
| | | EL | 85.90 | 57.86 | 100.0 | 100.0 | 85.89 | 59.17 | 100.0 | 100.0 |
| 1 | 100 | QL | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | EL | 99.87 | 97.67 | 100.0 | 100.0 | 99.81 | 97.86 | 100.0 | 100.0 |
| 0 | 200 | QL | 5.25 | 5.11 | 5.08 | 5.18 | 6.40 | 5.84 | 5.58 | 5.62 |
| | | EL | 5.25 | 5.15 | 5.06 | 5.05 | 6.01 | 5.44 | 5.29 | 5.30 |
| 0.5 | 200 | QL | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | EL | 99.75 | 99.99 | 100.0 | 100.0 | 99.72 | 99.99 | 100.0 | 100.0 |
| 1 | 200 | QL | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | EL | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

$\mathbf{u}_1 = \mathbf{u}_1^*$ and $\mathbf{u}_k = (\mathbf{I}_p - \mathbf{P}_k)\mathbf{u}_k^*$, where $\mathbf{P}_k$ is the projection matrix on $\mathbf{u}_1, \dots, \mathbf{u}_{k-1}$ for $k = 2, \dots, K$. Providing the vectors $\mathbf{u}_k$, we have the covariance matrix structure

$$\boldsymbol{\Sigma} = \theta_0 \mathbf{I}_p + \sum_{k=1}^{K} \theta_k \mathbf{u}_k \mathbf{u}_k^T$$

for the factor model defined in Example 2.4. In this simulation, we set $K = 4$ and the coefficient vector $(\theta_0, \theta_1, \theta_2, \theta_3)^T = (4, 3, 2, 1)^T$. Similarly, $\theta_4 = 0$ is for Type I error rates and $\theta_4 = 0.5, 1.0$ is for powers. We summarize simulation results in the top panel of Table 3. Both QL and EL tests control Type I error rates well and have high powers at $\theta_4 = 0.5$ and 1.0. For some cases such as $p = 100$ and $\theta_4 = 0.5$, the QL-test has slightly higher power than the EL-test.

We next investigate the performance of QL and EL tests for the covariance matrix with a special pattern $H_{50}$. We represent the covariance matrix as a linear combination

$$\boldsymbol{\Sigma} = \sum_{k=1}^{4} \theta_k \mathbf{A}_k,$$

where $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ are defined in Example 2.5 and $\mathbf{A}_4 = \mathbf{u}_p \mathbf{u}_p^T$ with $\mathbf{u}_p \sim N_p(\mathbf{0}, \mathbf{I}_p)$. We set the first three coefficients $(\theta_1, \theta_2, \theta_3) = (6.0, 0.5, 0.1)$ and $\theta_4 = 0.0, 0.5$ and 1.0 for examining Type I error rates and powers, respectively. We summarize the simulation results in the bottom panel of Table 3, which shows that both QL and EL tests can control Type I error rates, and have high power as well, although QL-test has higher power than EL-test for $(n, p) = (100, 100)$ and $\theta_5 = 0.5$.

## 4. Technical proofs.

4.1. *Proofs of Theorems* 2.1 *and* 2.2. Recall the definitions of $\bar{\mathbf{x}}$ and $\mathbf{S}_n$ in (2.1) and $\mathbf{F}$ in (2.8), it follows that under Assumption A,

$$\mathbf{F} = \boldsymbol{\Gamma}^{-1} \mathbf{S}_n (\boldsymbol{\Gamma}^T)^{-1} \quad \text{and} \quad \mathbf{S}_n = \boldsymbol{\Gamma} \mathbf{F} \boldsymbol{\Gamma}^T. \tag{4.1}$$

PROOF OF THEOREM 2.1. Using Chebyshev's inequality, the proof of (2.10) and (2.11) is completed by showing that

$$\mathrm{E} p^{-1} \operatorname{tr} \mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_2 - p^{-1} \operatorname{tr} \mathbf{C}_1\mathbf{C}_2 - y_{n-1}(p^{-1} \operatorname{tr} \mathbf{C}_1)(p^{-1} \operatorname{tr} \mathbf{C}_2) = o(1)$$

and

$$\begin{aligned} \mathrm{E} p^{-1} \operatorname{tr} \mathbf{F}\mathbf{C}_0 - p^{-1} \operatorname{tr} \mathbf{C}_0 &= o(1), \\ \operatorname{Var}(p^{-1} \operatorname{tr} \mathbf{F}\mathbf{C}_0) &= o(1), \\ \operatorname{Var}\{p^{-1} \operatorname{tr}(\mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_2)\} &= o(1). \end{aligned}$$

We have

$$p^{-1} \operatorname{tr} \mathbf{F}\mathbf{C}_0 = p^{-1} \sum_{i=1}^{n} \boldsymbol{\gamma}_i^T \mathbf{C}_0 \boldsymbol{\gamma}_i - p^{-1} n \bar{\boldsymbol{\gamma}}^T \mathbf{C}_0 \bar{\boldsymbol{\gamma}},$$

where $\boldsymbol{\gamma}_i = (n-1)^{-1/2} \mathbf{w}_i$ and $\bar{\boldsymbol{\gamma}} = n^{-1} \sum_{i=1}^n \boldsymbol{\gamma}_i$. Then $p^{-1} \mathrm{E} \sum_{i=1}^n \boldsymbol{\gamma}_i^T \mathbf{C}_0 \boldsymbol{\gamma}_i = n(n-1)^{-1} p^{-1} \operatorname{tr} \mathbf{C}_0$ and $p^{-1} n \mathrm{E} \bar{\boldsymbol{\gamma}}^T \mathbf{C}_0 \bar{\boldsymbol{\gamma}} = (n-1)^{-1} p^{-1} \operatorname{tr} \mathbf{C}_0 \to 0$. Thus,

$$\mathrm{E} p^{-1} \operatorname{tr} \mathbf{F}\mathbf{C}_0 - p^{-1} \operatorname{tr} \mathbf{C}_0 \to 0. \tag{4.2}$$

Moreover,

$$
\begin{aligned}
&\mathrm{E}\left(p^{-1}\sum_{i=1}^{n}\boldsymbol{\gamma}_i^T\mathbf{C}_0\boldsymbol{\gamma}_i - p^{-1}(n-1)^{-1}n\operatorname{tr}\mathbf{C}_0\right)^2 \\
&\quad = np^{-2}\mathrm{E}\big(\boldsymbol{\gamma}_1^T\mathbf{C}_0\boldsymbol{\gamma}_1 - (n-1)^{-1}\operatorname{tr}\mathbf{C}_0\big)^2 \to 0. \qquad (4.3)
\end{aligned}
$$

By (4.2) and (4.3), we have

$$p^{-1}\sum_{i=1}^{n}\boldsymbol{\gamma}_i^T\mathbf{C}_0\boldsymbol{\gamma}_i - p^{-1}\operatorname{tr}\mathbf{C}_0 = o_p(1).$$

Similarly, we have

$$p^{-1}n\bar{\boldsymbol{\gamma}}^T\mathbf{C}_0\bar{\boldsymbol{\gamma}} = o_p(1).$$

Thus, we have

$$p^{-1}\operatorname{tr}\mathbf{F}\mathbf{C}_0 - p^{-1}\operatorname{tr}\mathbf{C}_0 = o_p(1).$$

Moreover, we have

$$
\begin{aligned}
p^{-1}\operatorname{tr}\mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_2 &= p^{-1}\sum_{i\neq j}\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\mathbf{C}_2\boldsymbol{\gamma}_i + p^{-1}\sum_{i}\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\mathbf{C}_2\boldsymbol{\gamma}_i \\
&\quad - 2p^{-1}n\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\mathbf{F}\mathbf{C}_2\bar{\boldsymbol{\gamma}} + p^{-1}n^2\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}^T\mathbf{C}_2\bar{\boldsymbol{\gamma}},
\end{aligned}
$$

where $p^{-1}\mathrm{E}\sum_{i\neq j}\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\mathbf{C}_2\boldsymbol{\gamma}_i = [n/(n-1)]p^{-1}\operatorname{tr}\mathbf{C}_1\mathbf{C}_2$,

$$
\begin{aligned}
&p^{-1}\sum_{i}\mathrm{E}\{\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\mathbf{C}_2\boldsymbol{\gamma}_i\} \\
&\quad = [n/(n-1)^2]p^{-1}\operatorname{tr}\mathbf{C}_2\operatorname{tr}\mathbf{C}_1 + 2n/[p(n-1)^2]\operatorname{tr}\mathbf{C}_1\mathbf{C}_2 + o(1),
\end{aligned}
$$

and $\mathrm{E}p^{-1}n\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\mathbf{F}\mathbf{C}_2\bar{\boldsymbol{\gamma}} = n/[p(n-1)^2]\operatorname{tr}\mathbf{C}_1\mathbf{C}_2 + o(1)$ and $\mathrm{E}p^{-1}n^2\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}^T \times \mathbf{C}_2\bar{\boldsymbol{\gamma}} = o(1)$. Then

$$p^{-1}\mathrm{E}\operatorname{tr}\mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_2 - p^{-1}\operatorname{tr}\mathbf{C}_1\mathbf{C}_2 - y_{n-1}(p^{-1}\operatorname{tr}\mathbf{C}_1)(p^{-1}\operatorname{tr}\mathbf{C}_2) \to 0.$$

Similarly, we can prove

$$
\begin{aligned}
\operatorname{Var}\left[p^{-1}\sum_{i\neq j}\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\mathbf{C}_2\boldsymbol{\gamma}_i\right] &\to 0, \\
\operatorname{Var}\left[p^{-1}\sum_{i}\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\mathbf{C}_2\boldsymbol{\gamma}_i\right] = p^{-2}\operatorname{Var}[\boldsymbol{\gamma}_1^T\mathbf{C}_1\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^T\mathbf{C}_2\boldsymbol{\gamma}_1] &\to 0, \\
\operatorname{Var}[p^{-1}n\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\mathbf{F}\mathbf{C}_2\bar{\boldsymbol{\gamma}}] \to 0, \qquad \operatorname{Var}[p^{-1}n^2\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}^T\mathbf{C}_2\bar{\boldsymbol{\gamma}}] &\to 0.
\end{aligned}
$$

Thus, we have

$$\Big[p^{-1}\sum_{i\neq j}\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\mathbf{C}_2\boldsymbol{\gamma}_i - p^{-1}\operatorname{tr}\mathbf{C}_1\mathbf{C}_2\Big] = o_p(1),$$

$$\Big[p^{-1}\sum_{i}\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\mathbf{C}_2\boldsymbol{\gamma}_i - y_{n-1}(p^{-1}\operatorname{tr}\mathbf{C}_1)(p^{-1}\operatorname{tr}\mathbf{C}_2)$$
$$- 2n/[p(n-1)^2]\operatorname{tr}\mathbf{C}_1\mathbf{C}_2\Big] = o_p(1)$$

and

$$\{p^{-1}n\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\mathbf{F}\mathbf{C}_2\bar{\boldsymbol{\gamma}} - n/[p(n-1)^2]\operatorname{tr}\mathbf{C}_1\mathbf{C}_2\} = o_p(1),$$
$$\{p^{-1}n^2\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}^T\mathbf{C}_2\bar{\boldsymbol{\gamma}}\} = o_p(1).$$

Then we have

$$[p^{-1}\operatorname{tr}\mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_2 - p^{-1}\operatorname{tr}\mathbf{C}_1\mathbf{C}_2 - y_{n-1}(p^{-1}\operatorname{tr}\mathbf{C}_1)(p^{-1}\operatorname{tr}\mathbf{C}_2)] = o_p(1).$$

This completes the proof of Theorem 2.1. □

PROOF OF THEOREM 2.2. Recall $\boldsymbol{\gamma}_i = (n-1)^{-1/2}\mathbf{w}_i$, $i = 1, \dots, n$, and $\bar{\boldsymbol{\gamma}} = n^{-1}\sum_{i=1}^n \boldsymbol{\gamma}_i$. To derive its limiting distribution, we first calculate $\mathrm{E}(p^{-u_1}\operatorname{tr}\mathbf{F}\mathbf{C}_1 \times \mathbf{F}\mathbf{C}_1, p^{-u_2}\operatorname{tr}\mathbf{F}\mathbf{C}_2)^T$. Under Assumptions A and B, $\mathrm{E}[\operatorname{tr}(\mathbf{F}\mathbf{C}_2)] = \operatorname{tr}\mathbf{C}_2$ and it follows by some calculations that

$$\begin{aligned}
&p^{-u_1}\mathrm{E}\operatorname{tr}\mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_1\\
&\quad= p^{-u_1}\Big(\mathrm{E}\sum_{i\neq j}\operatorname{tr}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\mathbf{C}_1 + \mathrm{E}\sum_{i=1}^n\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\mathbf{C}_1\boldsymbol{\gamma}_i\\
&\qquad - 2n\mathrm{E}\sum_j\operatorname{tr}\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\mathbf{C}_1 + \mathrm{E}n^2\operatorname{tr}\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}^T\mathbf{C}_1\Big)\\
&\quad= p^{-u_1}\Big\{[\operatorname{tr}\mathbf{C}_1^2 + y_{n-1}p^{-1}(\operatorname{tr}\mathbf{C}_1)^2] + y_{n-1}p^{-1}\operatorname{tr}\mathbf{C}_1^2\\
&\qquad + y_{n-1}(\kappa-3)p^{-1}\sum_{i=1}^p(\mathbf{e}_i^T\mathbf{C}_1\mathbf{e}_i)^2\Big\} + o(1).
\end{aligned}$$

To establish the asymptotic normality of $(p^{-u_1}\operatorname{tr}\mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_1, p^{-u_2}\operatorname{tr}\mathbf{F}\mathbf{C}_2)^T$, it suffices to establish the asymptotic normality of $p^{-u_1}\operatorname{tr}\mathbf{F}\mathbf{C}_1\mathbf{F}\mathbf{C}_1 + bp^{-u_2}\operatorname{tr}\mathbf{F}\mathbf{C}_2$ for any constant $b$. Define

$$A_{n1} = p^{-u_1}\left(\sum_{i\neq j} \operatorname{tr} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T \mathbf{C}_1 \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T \mathbf{C}_1 + \sum_{i=1}^n \boldsymbol{\gamma}_i^T \mathbf{C}_1 \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T \mathbf{C}_1 \boldsymbol{\gamma}_i\right) + p^{-u_2} b \sum_{i=1}^n \boldsymbol{\gamma}_i^T \mathbf{C}_2 \boldsymbol{\gamma}_i$$

and

$$A_{n2} = p^{-u_1}\left(-2n \sum_j \operatorname{tr} \bar{\boldsymbol{\gamma}} \bar{\boldsymbol{\gamma}}^T \mathbf{C}_1 \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T \mathbf{C}_1 + n^2 \operatorname{tr} \bar{\boldsymbol{\gamma}} \bar{\boldsymbol{\gamma}}^T \mathbf{C}_1 \bar{\boldsymbol{\gamma}} \bar{\boldsymbol{\gamma}}^T \mathbf{C}_1\right) - p^{-u_2} b n \bar{\boldsymbol{\gamma}}^T \mathbf{C}_2 \bar{\boldsymbol{\gamma}}.$$

It follows that

$$p^{-u_1} \operatorname{tr} \mathbf{F}\mathbf{C}_1 \mathbf{F}\mathbf{C}_1 + p^{-u_2} b \operatorname{tr} \mathbf{F}\mathbf{C}_2 = A_{n1} + A_{n2}.$$

Because $\operatorname{Var}(A_{n2}) = o(1)$, it is sufficient to deal with $A_{n1}$. We will use Lindeberg CLT on martingale difference sequence to establish the asymptotic normality of $A_{n1}$. Let $\mathrm{E}_\ell(Z)$ be the conditional expectation of $Z$ given $\{\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_\ell\}$. Then it can be verified that $\{(\mathrm{E}_\ell - \mathrm{E}_{\ell-1}) A_{n1}, \ell = 1, \dots, n\}$ is a martingale difference sequence. Define

$$\delta_{1\ell} = (\mathrm{E}_\ell - \mathrm{E}_{\ell-1}) \sum_{i\neq j} \operatorname{tr} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T \mathbf{C}_1 \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T \mathbf{C}_1,$$

$$\delta_{2\ell} = \boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell \boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell - \mathrm{E} \boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell \boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell,$$

$$\delta_{3\ell} = b\left[\boldsymbol{\gamma}_\ell^T \mathbf{C}_2 \boldsymbol{\gamma}_\ell - (n-1)^{-1} \operatorname{tr} \mathbf{C}_2\right].$$

Then $(\mathrm{E}_\ell - \mathrm{E}_{\ell-1}) A_{n1} = p^{-u_1} \delta_{1\ell} + p^{-u_1} \delta_{2\ell} + p^{-u_2} \delta_{3\ell}$. We may simplify $\delta_{1\ell}$ as follows:

$$\begin{aligned}
\delta_{1\ell} &= \mathrm{E}_\ell \sum_{i\neq j} \operatorname{tr} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T \mathbf{C}_1 \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T \mathbf{C}_1 - \mathrm{E}_{\ell-1} \sum_{i\neq j} \operatorname{tr} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T \mathbf{C}_1 \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T \mathbf{C}_1 \\
&= \frac{2(n-\ell)}{(n-1)}\left[\boldsymbol{\gamma}_\ell^T \mathbf{C}_1^2 \boldsymbol{\gamma}_\ell - (n-1)^{-1} \operatorname{tr} \mathbf{C}_1^2\right] \\
&\quad + 2 \sum_{j\leq \ell-1}\left[\boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell - (n-1)^{-1} \boldsymbol{\gamma}_j^T \mathbf{C}_1^2 \boldsymbol{\gamma}_j\right].
\end{aligned}$$

Rewrite

$$\begin{aligned}
\delta_{2\ell} &= \boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell \boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell - \mathrm{E} \boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell \boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell \\
&= \left[\boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell - (n-1)^{-1} \operatorname{tr} \mathbf{C}_1\right]^2 - \mathrm{E}\left[\boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell - (n-1)^{-1} \operatorname{tr} \mathbf{C}_1\right]^2 \\
&\quad + 2(n-1)^{-1} \operatorname{tr} \mathbf{C}_1\left[\boldsymbol{\gamma}_\ell^T \mathbf{C}_1 \boldsymbol{\gamma}_\ell - (n-1)^{-1} \operatorname{tr} \mathbf{C}_1\right].
\end{aligned}$$

Because the Lindeberg condition of the martingale difference sequence $(p^{-u_1}\delta_{1\ell} + p^{-u_1}\delta_{2\ell} + p^{-u_2}\delta_{3\ell})$ may be easily verified, it is sufficient to derive the limit of $\sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}[p^{-u_1}\delta_{1\ell} + p^{-u_1}\delta_{2\ell} + p^{-u_2}\delta_{3\ell}]^2$. Bai and Silverstein (2010) pointed out that the CLT of the linear spectral statistics of the sample covariance matrix $S_n$ from $\{w_{ij} I_{(|w_{ij}| \le \eta_n \sqrt{n})}, i = 1, \dots, p, j = 1, \dots, n\}$ is the same as the sample covariance matrix $S_n$ from $\{w_{ij}, i = 1, \dots, p, j = 1, \dots, n\}$ where $\eta_n \to 0$ and $\eta_n \sqrt{n} \to \infty$. Then it follows by using Lemma 9.1 in Bai and Silverstein (2010) that

$$p^{-2u_1} \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\delta_{1\ell}\delta_{2\ell}$$
$$= p^{-2u_1}\left\{4(n-1)^{-2} \operatorname{tr} \mathbf{C}_1 \left[2 \operatorname{tr} \mathbf{C}_1^3 + (\kappa - 3) \sum_{i=1}^{p} (\mathbf{e}_i^T \mathbf{C}_1 \mathbf{e}_i)(\mathbf{e}_i^T \mathbf{C}_1^2 \mathbf{e}_i)\right]\right\}$$
$$+ o_p(1).$$

By some calculations, we have

$$p^{-(u_1+u_2)} \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\delta_{1\ell}\delta_{3\ell}$$
$$= 2bp^{-(u_1+u_2)}(n-1)^{-1}\left[2 \operatorname{tr}(\mathbf{C}_1^2 \mathbf{C}_2) + (\kappa - 3) \sum_{i=1}^{p} \mathbf{e}_i^T \mathbf{C}_1^2 \mathbf{e}_i \mathbf{e}_i^T \mathbf{C}_2 \mathbf{e}_i\right]$$
$$+ o_p(1)$$

and

$$p^{-(u_1+u_2)} \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\delta_{2\ell}\delta_{3\ell}$$
$$= 2bp^{-(u_1+u_2)}(\kappa - 3)(n-1)^{-2} \operatorname{tr} \mathbf{C}_1 \sum_{i=1}^{p} \mathbf{e}_i^T \mathbf{C}_1 \mathbf{e}_i \mathbf{e}_i^T \mathbf{C}_2 \mathbf{e}_i$$
$$+ 4bp^{-(u_1+u_2)}(n-1)^{-2} \operatorname{tr} \mathbf{C}_1 \operatorname{tr} \mathbf{C}_1 \mathbf{C}_2 + o_p(1)$$

by using Lemma 9.1 in Bai and Silverstein (2010) again. We next deal with the squared terms. We can show that

$$p^{-2u_1} \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\delta_{1\ell}^2$$
$$= 4p^{-2u_1}(n-1)^{-1}\left[2 \operatorname{tr} \mathbf{C}_1^4 + (\kappa - 3) \sum_{i=1}^{p} (\mathbf{e}_i^T \mathbf{C}_1^2 \mathbf{e}_i)^2\right]$$
$$+ 4p^{-2u_1}[(n-1)^{-1} \operatorname{tr} \mathbf{C}_1^2]^2 + o_p(1),$$

$$p^{-2u_1}\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}\delta_{2\ell}^2 = 4p^{-2u_1}(n-1)^{-2}(\operatorname{tr}\mathbf{C}_1)^2\left[2\operatorname{tr}\mathbf{C}_1^2+(\kappa-3)\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_1\mathbf{e}_i)^2\right] + o_p(1),$$

$$p^{-2u_2}\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}\delta_{3\ell}^2 = b^2p^{-2u_2}(n-1)^{-1}\left[2\operatorname{tr}\mathbf{C}_2^2+(\kappa-3)\sum_{i=1}^{p}(\mathbf{e}_i^T\mathbf{C}_2\mathbf{e}_i)^2\right]+o_p(1).$$

Thus, we can further derive $\sum_{\ell=1}^n \mathrm{E}_{\ell-1}(p^{-u_1}\delta_{1\ell}+p^{-u_1}\delta_{2\ell}+p^{-u_2}\delta_{3\ell})^2$.

Applying Lindeberg CLT on martingale difference sequence on $p^{-u_1}\delta_{1\ell}+p^{-u_1}\delta_{2\ell}+p^{-u_2}\delta_{3\ell}$, it follows that for any $b$, $p^{-u_1}(\operatorname{tr}\mathbf{FC}_1\mathbf{FC}_1-\mathrm{E}\operatorname{tr}\mathbf{FC}_1\mathbf{FC}_1)+p^{-u_2}b(\operatorname{tr}\mathbf{FC}_2-\mathrm{E}\operatorname{tr}\mathbf{FC}_2)$ converges to a normal distribution with mean 0 and variance $\lim_{n\to\infty}\sum_{\ell=1}^n \mathrm{E}_{\ell-1}(p^{-u_1}\delta_{1\ell}+p^{-u_1}\delta_{2\ell}+p^{-u_2}\delta_{3\ell})^2$. The proof of Theorem 2.2 is completed by calculating the corresponding mean vector and covariance matrix, which equal $\boldsymbol{\mu}_n^{(1)}$ and $\boldsymbol{\Sigma}_n^{(1)}$ given in Section 2.4. □

4.2. *Proofs of Theorems 2.3, 2.4 and 2.5.* Recall $\boldsymbol{\Sigma}_0=\theta_1\mathbf{A}_1+\cdots+\theta_K\mathbf{A}_K$ and $\widehat{\boldsymbol{\Sigma}}_0=\hat\theta_1\mathbf{A}_1+\cdots+\hat\theta_K\mathbf{A}_K$. Using the identity $\hat\Sigma_0^{-1}-\boldsymbol{\Sigma}_0^{-1}=-\boldsymbol{\Sigma}_0^{-1}(\hat\Sigma_0-\boldsymbol{\Sigma}_0)\hat\Sigma_0^{-1}$, it follows that

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_0^{-1} &= \boldsymbol{\Sigma}_0^{-1}-\sum_{k=1}^{K}(\hat\theta_k-\theta_k)\boldsymbol{\Sigma}_0^{-1}\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1}\\ &\quad+\sum_{i,j}[(\hat\theta_i-\theta_i)(\hat\theta_j-\theta_j)\widehat{\boldsymbol{\Sigma}}_0^{-1}\mathbf{A}_i\boldsymbol{\Sigma}_0^{-1}\mathbf{A}_j\boldsymbol{\Sigma}_0^{-1}]\\ &= \boldsymbol{\Sigma}_0^{-1}\left[\mathbf{I}_p-\sum_{k=1}^{K}(\hat\theta_k-\theta_k)\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1}\right]\\ &\quad+\sum_{i,j}(\hat\theta_i-\theta_i)(\hat\theta_j-\theta_j)\widehat{\boldsymbol{\Sigma}}_0^{-1}\mathbf{A}_i\boldsymbol{\Sigma}_0^{-1}\mathbf{A}_j\boldsymbol{\Sigma}_0^{-1}.\end{aligned}$$

Then we have $\mathbf{S}_n\widehat{\boldsymbol{\Sigma}}_0^{-1}=\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1}[\mathbf{I}_p-\sum_{k=1}^K(\hat\theta_k-\theta_k)\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1}]+\sum_{i,j}(\hat\theta_i-\theta_i)(\hat\theta_j-\theta_j)\mathbf{S}_n\widehat{\boldsymbol{\Sigma}}_0^{-1}\mathbf{A}_i\boldsymbol{\Sigma}_0^{-1}\mathbf{A}_j\boldsymbol{\Sigma}_0^{-1}$. Under $H_0$, by $\hat\theta_k=\theta_k+O_p(1/n)$, then the trace of the second term is of order $o_p(1)$.

PROOF OF THEOREM 2.3(a) AND (b). When $p < n-1$, it follows by the Taylor expansion of $\hat{\theta}_k$ at $\theta_k$, $k = 1, \ldots, K$ that under $H_0$,

$$
\begin{aligned}
&\operatorname{tr}\mathbf{S}_n\widehat{\boldsymbol{\Sigma}}_0^{-1} - \log|\mathbf{S}_n\widehat{\boldsymbol{\Sigma}}_0^{-1}| \\
&\quad = \operatorname{tr}\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1} - \sum_{k=1}^{K} p(\hat{\theta}_k - \theta_k)p^{-1}\operatorname{tr}\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1}\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1} \\
&\qquad - \log|\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1}| + \sum_{k=1}^{K} p(\hat{\theta}_k - \theta_k)p^{-1}\operatorname{tr}\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1} + o_p(1) \\
&\quad = \operatorname{tr}\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1} - \log|\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1}| + o_p(1) \\
&\quad = \operatorname{tr}\mathbf{F} - \log|\mathbf{F}| + o_p(1) \qquad (4.4)
\end{aligned}
$$

since $p^{-1}\operatorname{tr}\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1}\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1} - p^{-1}\operatorname{tr}\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1} = o_p(1)$ under $H_0$ by (2.10) with setting $\mathbf{C}_0 = \boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_0^{-1}\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Gamma}$. By Lemma 2.1 and when $p < n-1$, we have

$$
\frac{T_{n1} + p\alpha_1(y_{n-1}) + m_{12}(y_{n-1})}{\sigma_{n1}(y_{n-1})} \xrightarrow{d} N(0, 1),
$$

where $\sigma_{n1}^2(y_{n-1}) = -2y_{n-1} - 2\log(1 - y_{n-1})$.

This completes the proof of Theorem 2.3(a) with $p < n-1$. Similarly, Theorem 2.3(a) with $p = n-1$ can be obtained.

Similarly, for $p > n-1$, a direct application of Lemma 2.1 to (4.4) leads to

$$
\frac{T_{n1} + p\alpha_2(y_{n-1}) + m_{22}(y_{n-1})}{\sqrt{y_{n-1}^{-2}v_{11}(y_{n-1}) + v_{22}(y_{n-1}) - 2y_{n-1}^{-1}v_{12}(y_{n-1})}} \xrightarrow{d} N(0, 1).
$$

That is,

$$
\frac{T_{n1} + p\alpha_2(y_{n-1}) + m_{22}(y_{n-1})}{\sigma_{n1}(y_{n-1}^{-1})} \xrightarrow{d} N(0, 1).
$$

This completes the proof of Theorem 2.3(b). □

PROOF OF THEOREM 2.3(c). Under $H_0$, $\hat{\theta}_k = \theta_k + O_p(1/n)$, it follows that

$$
\begin{aligned}
&\operatorname{tr}(\mathbf{S}_n\hat{\boldsymbol{\Sigma}}_0^{-1} - \mathbf{I}_p)^2 \\
&\quad = \operatorname{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1} - \mathbf{I}_p)^2 + p\sum_{k=1}^{K}(\hat{\theta}_k - \theta_k)2p^{-1}\operatorname{tr}\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1}\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1} \\
&\qquad - p\sum_{k=1}^{K}(\hat{\theta}_k - \theta_k)2p^{-1}\operatorname{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_0^{-1})^2\mathbf{A}_k\boldsymbol{\Sigma}_0^{-1} + o_p(1). \qquad (4.5)
\end{aligned}
$$

Taking $\mathbf{C}_0 = \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_0^{-1} \mathbf{A}_k \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Gamma}$ for $k = 1, \dots, K$, it follows by Theorem 2.1 that

$$\begin{aligned}
&\operatorname{tr}(\mathbf{S}_n \hat{\boldsymbol{\Sigma}}_0^{-1} - \mathbf{I}_p)^2 \\
&\quad = \operatorname{tr}(\mathbf{S}_n \boldsymbol{\Sigma}_0^{-1} - \mathbf{I}_p)^2 - 2y_{n-1} \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k) \operatorname{tr} \mathbf{A}_k \boldsymbol{\Sigma}_0^{-1} + o_p(1) \\
&\quad = \operatorname{tr}(\mathbf{S}_n \boldsymbol{\Sigma}_0^{-1} - \mathbf{I}_p)^2 - 2y_{n-1}(\operatorname{tr} \widehat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_0^{-1} - p) + o_p(1).
\end{aligned}$$

By the definition of $\hat{\boldsymbol{\theta}}$ in Section 2.1, it follows that

$$\operatorname{tr} \widehat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_0^{-1} = \operatorname{tr} \mathbf{S}_n \mathbf{B} = \operatorname{tr} \mathbf{F} \boldsymbol{\Gamma}^T \mathbf{B} \boldsymbol{\Gamma},$$

where $\mathbf{B}$ is defined in Theorem 2.3(c) and $\operatorname{tr} \boldsymbol{\Sigma}_0 \mathbf{B} = \operatorname{tr} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_0^{-1} = p$. As a result, we have

$$\begin{aligned}
&\operatorname{tr}(\mathbf{S}_n \widehat{\boldsymbol{\Sigma}}_0^{-1} - \mathbf{I}_p)^2 \\
&\quad = \operatorname{tr}(\mathbf{F} - \mathbf{I}_p)^2 - 2y_{n-1}(\operatorname{tr} \widehat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_0^{-1} - p) + o_p(1) \\
&\quad = -2 \operatorname{tr} \mathbf{F} - 2y_{n-1} \operatorname{tr} \widehat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_0^{-1} + \operatorname{tr} \mathbf{F}^2 + 2y_{n-1} p + p + o_p(1) \\
&\quad = -2 \operatorname{tr} \mathbf{F} - 2y_{n-1} \operatorname{tr} \mathbf{F} \boldsymbol{\Gamma}^T \mathbf{B} \boldsymbol{\Gamma} + \operatorname{tr} \mathbf{F}^2 + 2y_{n-1} p + p + o_p(1) \\
&\quad = \operatorname{tr} \mathbf{F}^2 + \operatorname{tr} \mathbf{F}(-2\mathbf{I}_p - 2y_{n-1} \boldsymbol{\Gamma}^T \mathbf{B} \boldsymbol{\Gamma}) + 2y_{n-1} p + p + o_p(1).
\end{aligned}$$

By Theorem 2.2, we have

$$\frac{\operatorname{tr}(\mathbf{S}_n \widehat{\boldsymbol{\Sigma}}_0^{-1} - \mathbf{I}_p)^2 - p y_{n-1} - (\kappa - 2)y}{2\sigma} \to N(0, 1),$$

where

$$\begin{aligned}
\sigma^2 &= 4^{-1} \Bigg[ 4(\kappa - 1)(y + 2y^2) + 8y^3 p^{-1} \operatorname{tr}(\boldsymbol{\Sigma}_0 \mathbf{B})^2 \\
&\quad + 4(\kappa - 3) y^3 p^{-1} \sum_{i=1}^{p} (\mathbf{e}_i^T \boldsymbol{\Gamma}^T \mathbf{B} \boldsymbol{\Gamma} \mathbf{e}_i)^2 \\
&\quad + 4y^2 + 4(\kappa - 1)(y + 2y^2 + y^3) - 8(\kappa - 1) y(1 + y)^2 \Bigg] \\
&= 4^{-1} \Bigg[ 4y^2 + 8y^3 p^{-1} \operatorname{tr}(\boldsymbol{\Sigma}_0 \mathbf{B})^2 \\
&\quad + 4(\kappa - 3) y^3 p^{-1} \sum_{i=1}^{p} (\mathbf{e}_i^T \boldsymbol{\Gamma}^T \mathbf{B} \boldsymbol{\Gamma} \mathbf{e}_i)^2 - 4(\kappa - 1) y^3 \Bigg]
\end{aligned}$$

$$= y^2 + 2y^3 p^{-1} \operatorname{tr}(\boldsymbol{\Sigma}_0 \mathbf{B})^2 + (\kappa - 3) y^3 p^{-1} \sum_{i=1}^{p} (\mathbf{e}_i^T \boldsymbol{\Gamma}^T \mathbf{B} \boldsymbol{\Gamma} \mathbf{e}_i)^2 - (\kappa - 1) y^3. \qquad \square$$

PROOF OF THEOREM 2.4(c). Due to the space limit, the proof of Theorem 2.4(a) and (b) are given in the Supplementary Material (Zheng et al. (2019)).

We next derive the power function of $T_{n2}$. First, we consider $\boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Gamma}$ being bounded spectral norm. When $\boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Gamma}$ has unbounded spectral norm, we only add a factor $p^{-3/2}$ to $T_{n2}$. Then the same results are obtained. Under $H_1$, it follows that

$$\begin{aligned} &\operatorname{tr}(\mathbf{S}_n \hat{\boldsymbol{\Sigma}}_0^{-1} - \mathbf{I}_p)^2 \\ &\quad = \operatorname{tr}(\mathbf{S}_n \boldsymbol{\Sigma}_1^{*-1} - \mathbf{I}_p)^2 + p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) 2p^{-1} \operatorname{tr} \mathbf{S}_n \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \\ &\qquad - p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) 2p^{-1} \operatorname{tr}(\mathbf{S}_n \boldsymbol{\Sigma}_1^{*-1})^2 \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} + o_p(1). \end{aligned}$$

Recall $\mathbf{F} = \sum_{i=1}^{n} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T - n \bar{\boldsymbol{\gamma}} \bar{\boldsymbol{\gamma}}^T$. Thus, we have

$$\begin{aligned} &p^{-1} \operatorname{tr} \mathbf{S}_n \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \\ &\quad = p^{-1} \sum_{i=1}^{n} \boldsymbol{\gamma}_i^T \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Gamma} \boldsymbol{\gamma}_i - (n/p) \bar{\boldsymbol{\gamma}}^T \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Gamma} \bar{\boldsymbol{\gamma}}. \end{aligned}$$

Because $p^{-1} \sum_{i=1}^{n} \mathrm{E} \boldsymbol{\gamma}_i^T \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Gamma} \boldsymbol{\gamma}_i = p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} + o(1)$ and

$$(n/p) \mathrm{E} \bar{\boldsymbol{\gamma}}^T \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Gamma} \bar{\boldsymbol{\gamma}} = [(n-1)p]^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1},$$

then we have

$$p^{-1} \mathrm{E} \operatorname{tr} \mathbf{S}_n \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} - p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \to 0.$$

Furthermore, we can show that

$$\operatorname{Var}\left[ p^{-1} \sum_{i=1}^{n} \boldsymbol{\gamma}_i^T \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Gamma} \boldsymbol{\gamma}_i \right] \to 0 \quad \text{and}$$

$$\operatorname{Var}[\bar{\boldsymbol{\gamma}}^T \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Gamma} \bar{\boldsymbol{\gamma}}] \to 0.$$

Then we have $p^{-1}\operatorname{tr}\mathbf{S}_n\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1} - p^{-1}\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1} = o_p(1)$. Moreover, we have

$$\begin{aligned}
&p^{-1}\mathrm{E}\operatorname{tr}[\mathbf{S}_n\boldsymbol{\Sigma}_1^{*-1}]^2\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}\\
&\quad= p^{-1}\mathrm{E}\sum_{i=1}^{n}\operatorname{tr}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\\
&\qquad+ p^{-1}\mathrm{E}\sum_{i\neq j}\operatorname{tr}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\\
&\quad= np^{-1}(n-1)^{-2}\Biggl[2\operatorname{tr}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})^2\\
&\qquad+(\kappa-3)\sum_{\ell=1}^{p}(\mathbf{e}_\ell^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\mathbf{e}_\ell)(\mathbf{e}_\ell^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\mathbf{e}_\ell)\Biggr]\\
&\qquad+np^{-1}(n-1)^{-2}(\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})(\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1})\\
&\qquad+np^{-1}(n-1)^{-1}\operatorname{tr}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})^2.
\end{aligned}$$

Thus, it follows that $p^{-1}\mathrm{E}\operatorname{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_1^{*-1})^2\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1} - y_{n-1}(p^{-1}\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})(p^{-1}\times\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}) - p^{-1}\operatorname{tr}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})^2 \to 0$. Similarly, it can be shown that

$$\operatorname{Var}\Biggl[p^{-1}\sum_{i=1}^{n}\operatorname{tr}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\Biggr]\to 0$$

and

$$\operatorname{Var}\Biggl[p^{-1}\mathrm{E}\sum_{i\neq j}\operatorname{tr}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\Biggr]\to 0.$$

Therefore, we obtain $p^{-1}\operatorname{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_1^{*-1})^2\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1} - y_{n-1}(p^{-1}\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})(p^{-1}\times\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}) - p^{-1}\operatorname{tr}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})^2 = o_p(1)$. As a result, it follows that

$$\begin{aligned}
&\operatorname{tr}(\mathbf{S}_n\hat{\boldsymbol{\Sigma}}_0^{-1}-\mathbf{I}_p)^2\\
&\quad=\operatorname{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_1^{*-1}-\mathbf{I}_p)^2+p\sum_{k=1}^{K}(\hat{\theta}_k-\theta_k^*)2p^{-1}\operatorname{tr}\mathbf{S}_n\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}\\
&\qquad-p\sum_{k=1}^{K}(\hat{\theta}_k-\theta_k^*)2p^{-1}\operatorname{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_1^{*-1})^2\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}+o_p(1)\\
&\quad=\operatorname{tr}[\mathbf{S}_n\boldsymbol{\Sigma}_1^{*-1}-\mathbf{I}_p]^2+2p\sum_{k=1}^{K}(\hat{\theta}_k-\theta_k^*)p^{-1}\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1}\mathbf{A}_k\boldsymbol{\Sigma}_1^{*-1}
\end{aligned}$$

$$
\begin{aligned}
&- 2p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) \big[ y_{n-1} \big( p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big) \big( p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \big) \\
&+ p^{-1} \operatorname{tr} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \big( \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big)^2 \big] \\
&+ o_p(1).
\end{aligned}
$$

Note that

$$
2p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} = 2 \operatorname{tr} \hat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} - 2 \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1}
$$

and

$$
\begin{aligned}
&2p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) \big[ y_{n-1} \big( p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big) \big( p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \big) \\
&\quad + p^{-1} \operatorname{tr} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \big( \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big)^2 \big] \\
&\qquad = 2 y_{n-1} p^{-1} \big( \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big) \big( \operatorname{tr} \hat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big) + 2 \operatorname{tr} \hat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_1^{*-1} \big( \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big)^2 \\
&\qquad\quad - 2 y_{n-1} p^{-1} \big[ \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big]^2 - 2 \operatorname{tr} \big[ \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big]^2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&2p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \\
&\quad - 2p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) \big[ y_{n-1} p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \\
&\quad + p^{-1} \operatorname{tr} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1} \big( \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big)^2 \big] \\
&\qquad = 2 \big( 1 - y_{n-1} p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big) \operatorname{tr} \hat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_1^{*-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} - 2 \operatorname{tr} \hat{\boldsymbol{\Sigma}}_0 \boldsymbol{\Sigma}_1^{*-1} \big( \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big)^2 \\
&\qquad\quad - 2 \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} + 2 y_{n-1} p^{-1} \big[ \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big]^2 + 2 \operatorname{tr} \big[ \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big]^2 \\
&\qquad = \operatorname{tr} \mathbf{S}_n \mathbf{B}_1 - 2 \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} + 2 y_{n-1} p^{-1} \big[ \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big]^2 + 2 \operatorname{tr} \big[ \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \big]^2,
\end{aligned}
$$

where $\mathbf{B}_1 = \sum_{k=1}^{K} \mathbf{e}_k^T \mathbf{D} \mathbf{g} \cdot \mathbf{A}_k$ with $\mathbf{g} = (\operatorname{tr} \mathbf{G} \mathbf{A}_1, \dots, \operatorname{tr} \mathbf{G} \mathbf{A}_K)^T$ and $\mathbf{G} = \boldsymbol{\Sigma}_1^{*-1} \times \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} 2(1 - y_{n-1} p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1}) - 2 \boldsymbol{\Sigma}_1^{*-1} (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1})^2$. Then we have

$$
\begin{aligned}
&\operatorname{tr} \big( \mathbf{S}_n \hat{\boldsymbol{\Sigma}}_0^{-1} - \mathbf{I}_p \big)^2 \\
&\qquad = \operatorname{tr} \big( \mathbf{S}_n \boldsymbol{\Sigma}_1^{*-1} - \mathbf{I}_p \big)^2 + 2p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) p^{-1} \operatorname{tr} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_1^{*-1} \mathbf{A}_k \boldsymbol{\Sigma}_1^{*-1}
\end{aligned}
$$

$$
\begin{aligned}
&- 2p \sum_{k=1}^{K} (\hat{\theta}_k - \theta_k^*) \big[ y_{n-1} p^{-1} \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} p^{-1} \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \mathbf{A}_k \mathbf{\Sigma}_1^{*-1} \\
&+ p^{-1} \operatorname{tr} \mathbf{A}_k \mathbf{\Sigma}_1^{*-1} \big( \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^2 \big] + o_p(1) \\
= {} & \operatorname{tr} \big( \mathbf{S}_n \mathbf{\Sigma}_1^{*-1} - \mathbf{I}_p \big)^2 + \operatorname{tr} \mathbf{S}_n \mathbf{B}_1 - 2 \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} + 2 y_{n-1} p^{-1} \big( \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} )^2 \big)^2 \\
&+ 2 \operatorname{tr} \big( \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} + o_p(1) \\
= {} & \operatorname{tr} \big( \mathbf{F} \mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1} \mathbf{\Gamma} \big)^2 + \operatorname{tr} \mathbf{F} \big( -2 \mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1} \mathbf{\Gamma} + \mathbf{\Gamma}^T \mathbf{B}_1 \mathbf{\Gamma} \big) \\
&+ p - 2 \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} + 2 y_{n-1} p^{-1} \big( \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^2 + 2 \operatorname{tr} \big( \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^2 + o_p(1).
\end{aligned}
$$

It follows by Theorem 2.2 that under $H_1$,

$$
\frac{\operatorname{tr}(\mathbf{S}_n \widehat{\mathbf{\Sigma}}_0^{-1} - \mathbf{I}_p)^2 - \mu_3^{(1)}}{\sigma_{n3}^{(1)}} \to N(0,1),
$$

where

$$
\begin{aligned}
&(\sigma_{n3}^{(1)})^2 \\
&\quad = 8 \big[ n^{-1} \operatorname{tr} \big( \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^4 \big] + 4 \big[ n^{-1} \operatorname{tr} \big( \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^2 \big]^2 \\
&\qquad + 8 \big( n^{-1} \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^2 \big[ n^{-1} \operatorname{tr} \big( \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^2 \big] \\
&\qquad + 16 \big( n^{-1} \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big) \big[ n^{-1} \operatorname{tr} \big( \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^3 \big] \\
&\qquad + 2 n^{-1} \operatorname{tr} \big( -2 \mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1} \mathbf{\Gamma} + \mathbf{\Gamma}^T \mathbf{B}_1 \mathbf{\Gamma} \big)^2 \\
&\qquad + 8 n^{-1} \operatorname{tr} \big[ \big( \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^2 \mathbf{\Sigma} \big( -2 \mathbf{\Sigma}_1^{*-1} + \mathbf{B}_1 \big) \big] \\
&\qquad + 8 \big( n^{-1} \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big) \big[ n^{-1} \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \mathbf{\Sigma} \big( -2 \mathbf{\Sigma}_1^{*-1} + \mathbf{B}_1 \big) \big] \\
&\qquad + 4(\kappa - 3) n^{-1} \sum_{i=1}^{p} (\mathbf{e}_i^T \mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \mathbf{\Gamma} \mathbf{e}_i)^2 \\
&\qquad + 4(\kappa - 3) \big( n^{-1} \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big)^2 \left[ n^{-1} \sum_{i=1}^{p} (\mathbf{e}_i^T \mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1} \mathbf{\Gamma} \mathbf{e}_i)^2 \right] \\
&\qquad + 8(\kappa - 3) \big( n^{-1} \operatorname{tr} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \big) \\
&\qquad \times \left[ n^{-1} \sum_{i=1}^{p} (\mathbf{e}_i^T \mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1} \mathbf{\Gamma} \mathbf{e}_i)(\mathbf{e}_i^T \mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1} \mathbf{\Sigma} \mathbf{\Sigma}_1^{*-1} \mathbf{\Gamma} \mathbf{e}_i) \right] \\
&\qquad + (\kappa - 3) n^{-1} \sum_{i=1}^{p} \big[ \mathbf{e}_i \big( -2 \mathbf{\Gamma}^T \mathbf{\Sigma}_1^{*-1} \mathbf{\Gamma} + \mathbf{\Gamma}^T \mathbf{B}_1 \mathbf{\Gamma} \big) \mathbf{e}_i \big]^2
\end{aligned}
$$

$$+4(\kappa-3)n^{-1}\sum_{i=1}^{p}\mathbf{e}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\mathbf{e}_i\mathbf{e}_i^T\boldsymbol{\Gamma}^T(-2\boldsymbol{\Sigma}_1^{*-1}+\mathbf{B}_1)\boldsymbol{\Gamma}\mathbf{e}_i$$

$$+4(\kappa-3)(n^{-1}\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})$$

$$\times\left[n^{-1}\sum_{i=1}^{p}\mathbf{e}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\mathbf{e}_i\mathbf{e}_i^T\boldsymbol{\Gamma}^T(-2\boldsymbol{\Sigma}_1^{*-1}+\mathbf{B}_1)\boldsymbol{\Gamma}\mathbf{e}_i\right]$$

and

$$\mu_3^{(1)}=\frac{(\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})^2}{n-1}+\frac{n-2}{n-1}\operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})^2-2\operatorname{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1}+p$$
$$+\frac{1}{n-1}\left[2\operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{*-1})^2+(\kappa-3)\sum_{i=1}^{p}(\mathbf{e}_i^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_1^{*-1}\boldsymbol{\Gamma}\mathbf{e}_i)^2\right].\qquad\square$$

**5. Discussion.** We have studied hypothesis testing on linear structure of high-dimensional covariance matrix, and developed two tests for the linear structure. Under the null hypothesis, the covariance matrix can be represented as a linear combination of a finite number of prespecified matrix bases. This implies that we may estimate the covariance matrix well. If the null hypothesis gets rejected, one may have to consider more general structure or unstructured covariance matrix, and conduct further study.

**Acknowledgments.** The authors wish to thank the Associate Editor and anonymous referees for their constructive comments, which lead to significant improvement of this work. The authors thank Professor Ping-Shou Zhong for sharing his code of Zhong et al. (2017), and the code was used for our numerical comparison in Section S.5.2 in Zheng et al. (2019).

## SUPPLEMENTARY MATERIAL

**Supplement to "Hypothesis testing on linear structures of high-dimensional covariance matrix"** (DOI: 10.1214/18-AOS1779SUPP; .pdf). This supplementary material consists of the technical proofs and additional numerical results.

## REFERENCES

ANDERSON, T. W.T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141. MR0331612

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

BAI, ZHIDONGZ. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

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

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

BIRKE, M. and DETTE, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. MR2189467

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

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

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

FAN, J. and LI, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. In *International Congress of Mathematicians*, *Vol. III* 595–622. Eur. Math. Soc., Zürich. MR2275698

HAFF, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8** 586–597. MR0568722

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, *Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

JIANG, D., JIANG, T. and YANG, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Statist. Plann. Inference* **142** 2241–2256. MR2911842

JIANG, T. and QI, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.* **42** 988–1009. MR3426306

JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

KATO, N., YAMADA, T. and FUJIKOSHI, Y. (2010). High-dimensional asymptotic expansion of LR statistic for testing intraclass correlation structure and its error bound. *J. Multivariate Anal.* **101** 101–112. MR2557621

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

MCDONALD, R. P. (1974). The measurement of factor indeterminacy. *Psychometrika* **39** 203–222. MR0350973

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0652932

OLKIN, I. and SELLIAH, J. B. (1977). Estimating covariances in a multivariate normal distribution. In *Statistical Decision Theory and Related Topics*, *II* (S. S. Gupta and D. S. Moore, eds.) 313–326. Academic Press, New York. MR0436407

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

SRIVASTAVA, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.* **35** 251–272. MR2328427

SRIVASTAVA, M. S. and REID, N. (2012). Testing the structure of the covariance matrix with fewer observations than the dimension. *J. Multivariate Anal.* **112** 156–171. MR2957293

WANG, C. (2014). Asymptotic power of likelihood ratio tests for high dimensional data. *Statist. Probab. Lett.* **88** 184–189. MR3178349

WANG, Q. and YAO, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.* **7** 2164–2192. MR3104916

WANG, C., YANG, J., MIAO, B. and CAO, L. (2013). Identity tests for high dimensional data using RMT. *J. Multivariate Anal.* **118** 128–137. MR3054095

ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional $F$-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263

ZHENG, S., CHEN, Z., CUI, H. and LI, R. (2019). Supplement to "Hypothesis testing on linear structures of high-dimensional covariance matrix." DOI:10.1214/18-AOS1779SUPP.

ZHONG, P.-S., LAN, W., SONG, P. X. K. and TSAI, C.-L. (2017). Tests for covariance structures with high-dimensional repeated measurements. *Ann. Statist.* **45** 1185–1213. MR3662452

ZWIERNIK, P., UHLER, C. and RICHARDS, D. (2017). Maximum likelihood estimation for linear Gaussian covariance models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1269–1292. MR3689318

S. ZHENG
KLAS AND SCHOOL OF MATHEMATICS AND STATISTICS
NORTHEAST NORMAL UNIVERSITY
CHANGCHUN, JILIN 130024
P.R. CHINA
E-MAIL: zhengsr@nenu.edu.cn

Z. CHEN
SCHOOL OF DATA SCIENCE
FUDAN UNIVERSITY
SHANGHAI 200433
P.R. CHINA
E-MAIL: zuc4@psu.edu

H. CUI
DEPARTMENT OF STATISTICS
CAPITAL NORMAL UNIVERSITY
BEIJING 100048
P.R. CHINA
E-MAIL: hjcui@bnu.edu.cn

R. LI
DEPARTMENT OF STATISTICS AND THE METHODOLOGY CENTER
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802-2111
USA
E-MAIL: rzli@psu.edu