Tests for Covariance Matrix with Fixed or Divergent Dimension

RONGMAO ZHANG*, LIANG PENG† AND RUODU WANG‡

Final version 30 May, 2013

Abstract

Testing covariance structure is of importance in many areas of statistical analysis, such as microarray analysis and signal processing. Conventional tests for finite dimensional covariance cannot be applied to high dimensional data in general, and tests for high dimensional covariance in the literature usually depend on some special structure of the matrix. In this paper, we propose some empirical likelihood ratio tests for testing whether a covariance matrix equals a given one or has a banded structure. The asymptotic distributions of the new tests are independent of the dimension.

Keywords: Covariance matrix, empirical likelihood tests, high-dimensional data, χ^2-distribution.

*Department of Mathematics, Zhejiang University, China. Email: `rmzhang@zju.edu.cn`

†School of Mathematics, Georgia Institute of Technology, USA. Email: `peng@math.gatech.edu`

‡Corresponding author. Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: `wang@uwaterloo.ca`

1 Introduction

Let $\boldsymbol{X}_i = (X_{i1}, \dots, X_{ip})^T$, $i = 1, 2, \dots, n$ be independent and identically distributed (i.i.d.) random vectors with mean $\mu = (\mu_1, \dots, \mu_p)^T$ and covariance matrix $\Sigma = (\sigma_{ij})_{1 \le i,j \le p}$. For a given covariance matrix Σ_0, it has been a long history for the study of testing

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_a : \Sigma \neq \Sigma_0. \tag{1.1}$$

Traditional methods for testing (1.1) with finite p include the likelihood ratio test (see [1]) and the scaled distance measure for positive definite Σ_0 defined as

$$V = \frac{1}{p} \mathrm{tr}\Big(S_n - I_p\Big)^2, \tag{1.2}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, I_p denotes the $p \times p$ identity matrix and S_n is the sample covariance matrix of $\Sigma_0^{-1/2} \boldsymbol{X}_i$ (see [12, 13] and [15]). When dealing with high dimensional data, the sample covariance in the likelihood ratio test is no longer invertible with probability one and tests based on a scaled distance may also fail as demonstrated in [14].

Since the above conventional tests cannot be employed for testing high dimensional covariance matrix, new methods are needed. When the high dimensional covariance matrix has a modest dimension p compared to the sample size n, i.e. $p/n \to c$ for some $c \in (0, \infty)$, [14] proposed a test by modifying the scaled distance measure V defined in (1.2) under the assumption that $\boldsymbol{X}_1$ has a normal distribution. When the dimension p is much larger than the sample size n, some special structure has to be imposed. [9] proposed a test which generalizes the result of [14] to the case of nonnormal distribution and large dimension by assuming that $\boldsymbol{X}_i = \Gamma \boldsymbol{Z}_i + \mu$ for some i.i.d. m dimensional random vectors $\{\boldsymbol{Z}_i\}$ with $\mathrm{E}\boldsymbol{Z}_1 = 0$, $\mathrm{var}(\boldsymbol{Z}_1) = I_m$, and Γ is a $p \times m$ constant matrix with $\Gamma\Gamma^T = \Sigma$.

Sparsity is another commonly employed special structure in analyzing high dimensional data such as variable selection and covariance matrix estimation. Estimating sparse covariance matrices has been actively studied in recent years. Some recent references include [3], [21], [9] and [4]. When the covariance matrix is assumed to be sparse and has a banded structure, it becomes important to test whether the covariance matrix possesses such a desired structure, i.e., to test

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau, \tag{1.3}$$

where $\tau < p$ is given and may depend on n. Recently, [5] proposed to use the maximum of the absolute values of sample covariances to test (1.3) when $\boldsymbol{X}_1$ has a multivariate normal distribution. However, it is known that the convergence rate of the normalized maximum to a Gumbel limit is very slow, which means such a test has a poor size in general. Although using maximum is very powerful in detecting the departure from the null hypothesis when at least one large departure exists, it is much less powerful than a test based on a Euclidian distance when many small departures from the null hypothesis happen.

To avoid assuming the sparse structure and normality condition in the testing problems (1.1) and (1.3), we propose to construct tests based on the equivalent testing problem $H_0 : ||\Sigma - \Sigma_0||_F^2 = 0$ against $H_a : ||\Sigma - \Sigma_0||_F^2 \neq 0$, where $||A||_F = \sqrt{\text{tr}(A^T A)}$ is the Frobenius norm of the matrix A.

Put $\boldsymbol{Y}_i = (\boldsymbol{X}_i - \mu)(\boldsymbol{X}_i - \mu)^T$ for $i = 1, \ldots, n$. Based on the fact that $\text{E}[\boldsymbol{Y}_i] = \Sigma$, one can test (1.1) by employing the well-known Hotelling one-sample T^2 statistic for a mean vector when p is finite, and its modified versions when p is divergent and some specific models are assumed; see for example [2] and [8].

Another popular test for a finite dimensional mean vector is the empirical likelihood ratio test proposed by [16, 17]. Recently, [11] and [7] extended it to the high dimensional case. It turns out that the asymptotic distribution of the empirical likelihood ratio test is a chi-square distribution for a fixed dimension and a normal distribution for a divergent dimension. That is, the limit depends on whether the dimension is fixed or divergent. Note that the methods in the above papers can also be used to construct an estimator for unknown parameters, which is called maximum empirical likelihood estimator.

Motivated by the empirical likelihood ratio test in [19] for testing a high dimensional mean vector, we propose to apply the empirical likelihood ratio test to two estimating equations, where one equation ensures the consistency of the proposed test and another one is used to improve the test power. It turns out that the proposed test puts no restriction on the sparse structure of the covariance matrix and normality of $\boldsymbol{X}_1$. When testing (1.3), a similar procedure can be employed; see Section 2 for more details.

The paper is organized as follows. In Section 2, we introduce the new methodologies and present the main results. Simulation results are given in Section 3. Section 4 proves the main results. Detailed proofs for lemmas used in Section 4 are put in the supplementary material ([22]).

2 Methodologies and main results

2.1 Testing a covariance matrix

Let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^T$, $i = 1, \ldots, n$ be independent and identically distributed observations with mean $\mu = (\mu_1, \ldots, \mu_p)^T$ and covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$.

When μ is known, for $i = 1, \ldots, n$, we define $\boldsymbol{Y}_i = (\boldsymbol{X}_i - \mu)(\boldsymbol{X}_i - \mu)^T$. Then $\mathrm{E}[\mathrm{tr}((\boldsymbol{Y}_1 - \Sigma_0)(\boldsymbol{Y}_2 - \Sigma_0))] = 0$ is equivalent to $||\Sigma - \Sigma_0||_F^2 = 0$, which is equivalent to $H_0 : \Sigma = \Sigma_0$. A direct application of the empirical likelihood ratio test to the above estimating equation may endure low power by noting that $\mathrm{E}[\mathrm{tr}((\boldsymbol{Y}_1 - \Sigma_0)(\boldsymbol{Y}_2 - \Sigma_0))] = ||\Sigma - \Sigma_0||_F^2 = O(\delta^2)$ rather than $O(\delta)$ if $||\Sigma - \Sigma_0||_F = O(\delta)$ and p is fixed. A brief simulation study and the power analysis in Section 2.3 confirm this fact. In order to improve the test power, we propose to add one more linear equation. Note that with prior information on the model or more specific alternative hypothesis, a more proper linear equation may be available. Without additional information, any linear equation that detects the change of order $||\Sigma - \Sigma_0||_F$ is a possible choice theoretically. Here we simply choose the following functional $\mathbf{1}_p^T(\boldsymbol{Y}_1+\boldsymbol{Y}_2-2\Sigma_0)\mathbf{1}_p$, where $\mathbf{1}_p = (1, \ldots, 1)^T \in \mathbb{R}^p$. More specifically, we propose to apply the empirical likelihood ratio test to the following two equations:

$$\mathrm{E}[\mathrm{tr}((\boldsymbol{Y}_1 - \Sigma_0)(\boldsymbol{Y}_2 - \Sigma_0))] = 0 \text{ and } \mathrm{E}[\mathbf{1}_p^T(\boldsymbol{Y}_1 + \boldsymbol{Y}_2 - 2\Sigma_0)\mathbf{1}_p] = 0. \tag{2.4}$$

Of course one can try other linear equations or add more equations to further improve the power. Theorems derived below can easily be extended to the case when $\mathbf{1}_p$ is replaced by any constant vector.

In order to obtain an independent paired data $(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$, we split the sample into two subsamples with size $N = [n/2]$. That is, for $i = 1, 2, \ldots, N$, we define $\boldsymbol{R}_i(\Sigma) = (e_i(\Sigma), v_i(\Sigma))^T$, where

$$e_i(\Sigma) = \mathrm{tr}((\boldsymbol{Y}_i - \Sigma)(\boldsymbol{Y}_{i+N} - \Sigma)) \text{ and } v_i(\Sigma) = \mathbf{1}_p^T(\boldsymbol{Y}_i + \boldsymbol{Y}_{i+N} - 2\Sigma)\mathbf{1}_p.$$

Based on $\{\boldsymbol{R}_i(\Sigma)\}_{i=1}^N$, we define the empirical likelihood ratio function for Σ as

$$L_1(\Sigma) = \sup\{\prod_{i=1}^N (Np_i) : \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \boldsymbol{R}_i(\Sigma) = 0, p_1 \geq 0, \ldots, p_N \geq 0\}.$$

When μ is unknown, instead of using $\{\boldsymbol{R}_i(\Sigma)\}_{i=1}^N$, we use $\{\boldsymbol{R}_i^*(\Sigma)\}_{i=1}^N$ where μ is replaced by the sample means. That is, put $\overline{\boldsymbol{X}^1} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{X}_i$, $\overline{\boldsymbol{X}^2} = \frac{1}{N}\sum_{i=N+1}^{2N} \boldsymbol{X}_i$, and define

$$\boldsymbol{Y}_i^* = (\boldsymbol{X}_i - \overline{\boldsymbol{X}^1})(\boldsymbol{X}_i - \overline{\boldsymbol{X}^1})^T \text{ and } \boldsymbol{Y}_{N+i}^* = (\boldsymbol{X}_{N+i} - \overline{\boldsymbol{X}^2})(\boldsymbol{X}_{N+i} - \overline{\boldsymbol{X}^2})^T$$

for $i = 1, \ldots, N$. Put $\boldsymbol{R}_i^*(\Sigma) = (e_i^*(\Sigma), v_i^*(\Sigma))^T$, where

$$e_i^*(\Sigma) = \mathrm{tr}((\boldsymbol{Y}_i^* - \frac{(N-1)\Sigma}{N})(\boldsymbol{Y}_{i+N}^* - \frac{(N-1)\Sigma}{N}))$$

and

$$v_i^*(\Sigma) = \mathbf{1}_p^T(\boldsymbol{Y}_i^* + \boldsymbol{Y}_{i+N}^* - \frac{2(N-1)\Sigma}{N})\mathbf{1}_p.$$

As before, we define the empirical likelihood ratio function for Σ as

$$L_2(\Sigma) = \sup\{\prod_{i=1}^N (Np_i) : \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \boldsymbol{R}_i^*(\Sigma) = 0, p_1 \geq 0, \ldots, p_N \geq 0\}.$$

First we show that Wilks' theorem holds for the above empirical likelihood ratio tests without imposing any special structure on $\boldsymbol{X}_1$.

Theorem 2.1. *Suppose that* $\mathrm{E}[v_1^2(\Sigma)] > 0$ *and for some* $\delta > 0$*,*

$$\max\left\{\mathrm{E}|e_1(\Sigma)|^{2+\delta}/(\mathrm{E}[e_1^2(\Sigma)])^{\frac{2+\delta}{2}}, \mathrm{E}|v_1(\Sigma)|^{2+\delta}/(\mathrm{E}[v_1^2(\Sigma)])^{\frac{2+\delta}{2}}\right\} = o(N^{\frac{\delta+\min\{2,\delta\}}{4}}). \tag{2.5}$$

Then under $H_0 : \Sigma = \Sigma_0$*,* $-2\log L_1(\Sigma_0)$ *converges in distribution to a chi-square distribution with two degrees of freedom as* $n \to \infty$*. In addition, if*

$$(\mathrm{tr}(\Sigma^2))^2 = o(N^2\mathrm{E}[e_1^2(\Sigma)]) \quad and \quad (\sum_{i=1}^p\sum_{j=1}^p \sigma_{ij})^2 = o(N\mathrm{E}[v_1^2(\Sigma)]), \tag{2.6}$$

then under $H_0 : \Sigma = \Sigma_0$*,* $-2\log L_2(\Sigma_0)$ *also converges in distribution to a chi-square distribution with two degrees of freedom as* $n \to \infty$*.*

Using Theorem 2.1, one can test $H_0 : \Sigma = \Sigma_0$ against $H_a : \Sigma \neq \Sigma_0$. That is, one rejects H_0 at level α when $-2\log L_1(\Sigma_0) > \xi_{1-\alpha}$ if μ is known, or when $-2\log L_2(\Sigma_0) > \xi_{1-\alpha}$ if μ is unknown, where $\xi_{1-\alpha}$ denotes the $(1-\alpha)$-th quantile of a chi-square distribution with two degrees of freedom.

Write the $p \times p$ matrix $\boldsymbol{Y}_1$ as a $q = p^2$ dimensional vector and denote the covariance matrix of such a vector by $\Theta = (\theta_{ij})_{q\times q}$. Conditions in Theorem 2.1 can be guaranteed by imposing some conditions on the moments and dimensionality of $\boldsymbol{X}_1$ such as the following assumptions:

A1: $\liminf_{n\to\infty} \frac{1}{q}\mathrm{tr}(\Theta^2) > 0$ and $\liminf_{n\to\infty} \frac{1}{q}\mathbf{1}_q^T\Theta\mathbf{1}_q > 0$;

A2: For some $\delta > 0$, $\frac{1}{p^2}\sum_{i=1}^p\sum_{j=1}^p \mathrm{E}|(X_{1,i}-\mu_i)(X_{1,j}-\mu_j)-\sigma_{ij}|^{2+\delta} = O(1)$;

A3: $p = o\Big(n^{\frac{\delta+\min(2,\delta)}{4(2+\delta)}}\Big)$.

Corollary 2.2. *Under conditions A1–A3 and $H_0 : \Sigma = \Sigma_0$, $-2\log L_1(\Sigma_0)$ converges in distribution to a chi-square distribution with two degrees of freedom as $n \to \infty$. Further, if*

$$\max_{1\le i\le p} \sigma_{ii} < C_0 \quad \textit{for some constant} \quad C_0 > 0, \tag{2.7}$$

then $-2\log L_2(\Sigma_0)$ also converges in distribution to a chi-square distribution with two degrees of freedom as $n \to \infty$.

Remark 2.1. Condition (2.5) requires that the second moment of (e_1, v_1) is not too small compared to a higher-order moment of (e_1, v_1), which ensures that Lyapunov central limit theorem holds for $\frac{1}{\sqrt{N}}\sum_{i=1}^N e_i(\Sigma_0)$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^N v_i(\Sigma_0)$. Condition (2.6) makes sure that the mean vector can be replaced by the sample mean. It is obvious that (2.6) and (2.7) hold when p is fixed.

Note that condition A1 is only related to the covariance matrix and condition A2 holds obviously if

$$\frac{1}{p^2}\sum_{i=1}^p\sum_{j=1}^p \mathrm{E}|X_{1,i}X_{1,j}|^{2+\delta} < \infty \quad \text{or} \quad \frac{1}{p}\sum_{i=1}^p \mathrm{E}|X_{1i}|^{4+2\delta} < \infty.$$

Condition A3 imposes some restriction on p, but it can be removed if $\boldsymbol{X}_i$ has some special dependence structure. For example, Theorem 2.1 can be applied to the following setting studied in [8], [2] and [9]:

(B) (Multivariate model). Assume that the sample has the following decomposition:

$$\boldsymbol{X}_i = \Gamma\boldsymbol{Z}_i + \mu, \tag{2.8}$$

where Γ is a $p \times m$ constant matrix with $\Gamma\Gamma^T = \Sigma$ and $\{\boldsymbol{Z}_i = (Z_{i1}, \cdots, Z_{im})^T\}$ is a sequence of m dimensional i.i.d random vectors with $\mathrm{E}\boldsymbol{Z}_i = 0$, $\mathrm{var}(\boldsymbol{Z}_i) = I_m$, $\mathrm{E}(Z_{11}^4) = \ldots = \mathrm{E}(Z_{1m}^4) = 3 + \Delta > 1$ and uniformly bounded 8th moment. Further assume that for any integers $l_v \geq 0$ and $h \geq 1$ with $\sum_{v=1}^h l_v = 8$,

$$\mathrm{E}(Z_{1i_1}^{l_1} Z_{1i_2}^{l_2} \ldots Z_{1i_h}^{l_h}) = \mathrm{E}(Z_{1i_1}^{l_1})\mathrm{E}(Z_{1i_2}^{l_2}) \ldots \mathrm{E}(Z_{1i_h}^{l_h}), \tag{2.9}$$

where $i_1, \ldots, i_h$ are distinct.

Note that if $\boldsymbol{X}_i$ has a multivariate normal distribution, then **(B)** holds.

Corollary 2.3. *Suppose* **(B)** *holds with* $\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} > 0$. *Then, under* $H_0 : \Sigma = \Sigma_0$, *both* $-2 \log L_1(\Sigma_0)$ *and* $-2 \log L_2(\Sigma_0)$ *converge in distribution to a chi-square distribution with two degrees of freedom as* $n \to \infty$.

Remark 2.2. Note that condition $\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} > 0$ for model **(B)** implies that $\mathrm{E}[v_1^2(\Sigma)] > 0$; see the proof of Lemma 4.4. For testing $H_0 : \Sigma = I_p$, [9] proposed a test based on the above model and required $p \to \infty$ as $n \to \infty$. In comparison, the proposed empirical likelihood ratio tests work for both fixed and divergent p.

Remark 2.3. When one is interested in testing $H_0 : \mu = \mu_0$ & $\Sigma = \Sigma_0$, it is straightforward to combine the proposed empirical likelihood ratio test with that in [19] for testing a high dimensional mean.

2.2 Testing bandedness

Suppose $\{\boldsymbol{X}_i\}$ is a sequence of i.i.d. normal random vectors with covariance matrix $\Sigma = (\sigma_{ij})_{1 \leq i,j \leq p}$. [5] proposed to use the maximum of the absolute values of the sample correlations to test a banded structure

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau, \tag{2.10}$$

where $\tau < p$. It is known that the rate of convergence of the above maximum to a Gumbel distribution is very slow in general, which results in a poor size (see also in Section 3, simulation results). Using the maximum as a test statistic is powerful when at least a large deviation from the null hypothesis exists. However, when many small deviations from the null hypothesis exist, a test based on the maximum is much less efficient than a test based on a Euclidian distance such as the test in [20]. Here we modify the empirical likelihood ratio tests in Section 2.1 to test the above banded structure as follows.

For a matrix M, define the matrix $M^{(\tau)}$ as $(M^{(\tau)})_{ij} = (M)_{ij} I(|i-j| \geq \tau)$, where $I(\cdot)$ denotes the indicator function. Put

$$e'_i(\Sigma) = \mathrm{tr}((\boldsymbol{Y}_i^{(\tau)} - \Sigma^{(\tau)})(\boldsymbol{Y}_{N+i}^{(\tau)} - \Sigma^{(\tau)})),$$

$$v'_i(\Sigma) = \mathbf{1}_p^T(\boldsymbol{Y}_i^{(\tau)} + \boldsymbol{Y}_{N+i}^{(\tau)} - 2\Sigma^{(\tau)})\mathbf{1}_p,$$

$$e_i^{*\prime}(\Sigma) = \mathrm{tr}((\boldsymbol{Y}_i^{*(\tau)} - \frac{N-1}{N}\Sigma^{(\tau)})(\boldsymbol{Y}_{N+i}^{*(\tau)} - \frac{N-1}{N}\Sigma^{(\tau)}))$$

and

$$v_i^{*\prime}(\Sigma) = \mathbf{1}_p^T(\boldsymbol{Y}_i^{*(\tau)} + \boldsymbol{Y}_{N+i}^{*(\tau)} - \frac{2(N-1)}{N}\Sigma^{(\tau)})\mathbf{1}_p.$$

Then $\Sigma^{(\tau)}$ is zero under H_0 in (2.10). Based on $\boldsymbol{R}'_i(\Sigma) = (e'_i(\Sigma), v'_i(\Sigma))^T$ and $\boldsymbol{R}_i^{*\prime}(\Sigma) = (e_i^{*\prime}(\Sigma), v_i^{*\prime}(\Sigma))^T$, we define the empirical likelihood ratio functions for Σ as

$$L_3(\Sigma) = \sup\{\prod_{i=1}^N (Np_i) : \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \boldsymbol{R}'_i(\Sigma) = 0, p_i \geq 0, i = 1, \ldots, N\}$$

for the case of a known mean and

$$L_4(\Sigma) = \sup\{\prod_{i=1}^N (Np_i) : \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \boldsymbol{R}_i^{*\prime}(\Sigma) = 0, p_i \geq 0, i = 1, \ldots, N\}$$

for the case of an unknown mean. Similar to the proof of Theorem 2.1, we can show that $-2\log L_3(\Sigma_0)$ and $-2\log L_4(\Sigma_0)$ converge in distribution to a chi-square distribution with two degrees of freedom as $n \to \infty$ under some moment conditions.

Theorem 2.4. *Suppose that* $\mathrm{E}[v_1'^2(\Sigma)] > 0$ *and for some* $\delta > 0$,

$$\max\left\{\mathrm{E}|e'_1(\Sigma)|^{2+\delta}/(\mathrm{E}[e_1'^2(\Sigma)])^{\frac{2+\delta}{2}}, \mathrm{E}|v'_1(\Sigma)|^{2+\delta}/(\mathrm{E}[v_1'^2(\Sigma)])^{\frac{2+\delta}{2}}\right\} = o(N^{\frac{\delta+\min\{2,\delta\}}{4}}). \tag{2.11}$$

Then under H_0 in (2.10), $-2\log L_3(\Sigma_0)$ converges in distribution to a chi-square distribution with two degrees of freedom as $n \to \infty$, where Σ_0 is any matrix such that $\Sigma_0^{(\tau)} = 0$. In addition, if

$$\mathrm{E}\{\sum_{i=1}^{N}(e_i^{*\prime}(\Sigma) - e_i'(\Sigma))^2 + [\sum_{i=1}^{N}(e_i^{*\prime}(\Sigma) - e_i'(\Sigma))]^2\} = o(N\mathrm{E}[e_1'^2(\Sigma)])$$

and

$$\mathrm{E}\{\sum_{i=1}^{N}(v_i^{*\prime}(\Sigma) - v_i'(\Sigma))^2 + [\sum_{i=1}^{N}(v_i^{*\prime}(\Sigma) - v_i'(\Sigma))]^2\} = o(N\mathrm{E}[v_1'^2(\Sigma)]),$$

then under H_0 in (2.10), $-2\log L_4(\Sigma_0)$ also converges in distribution to a chi-square distribution with two degrees of freedom as $n \to \infty$.

In order to compare with [5], we use a different linear functional so as to easily verify conditions when $\boldsymbol{X}_i$ has a multivariate normal distribution. More specifically, for a $p \times p$ matrix M we define the matrix $M^{[\tau]}$ as

$$(M^{[\tau]})_{ij} = (M)_{ij}\{I(i \le (p-\tau)/2, j > (p+\tau)/2) + I(j \le (p-\tau)/2, i > (p+\tau)/2)\}.$$

Put $\tilde{v}_i'(\Sigma) = \mathbf{1}_p^T(\boldsymbol{Y}_i^{[\tau]} + \boldsymbol{Y}_{N+i}^{[\tau]} - 2\Sigma^{[\tau]})\mathbf{1}_p$ and

$$\tilde{v}_i^{*\prime}(\Sigma) = \mathbf{1}_p^T(\boldsymbol{Y}_i^{*[\tau]} + \boldsymbol{Y}_{N+i}^{*[\tau]} - \frac{2(N-1)}{N}\Sigma^{[\tau]})\mathbf{1}_p.$$

Based on $\tilde{\boldsymbol{R}}_i^{*\prime}(\Sigma) = (e_i^{*\prime}(\Sigma), \tilde{v}_i^{*\prime}(\Sigma))^T$, we define the empirical likelihood ratio function for Σ as

$$L_5(\Sigma) = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i\tilde{\boldsymbol{R}}_i^{*\prime}(\Sigma) = 0, p_i \ge 0, i = 1, \ldots, N\}.$$

Theorem 2.5. *Assume $\boldsymbol{X}_i \sim N(\mu, \Sigma)$,*

$$C_1 \le \liminf_{n\to\infty} \min_{1\le i\le p} \sigma_{ii} \le \limsup_{n\to\infty} \max_{1\le i\le p} \sigma_{ii} \le C_2$$

for some constants $0 < C_1 \le C_2 < \infty$ and $\tau = o((\sum_{1\le i,j\le p}\sigma_{ij}^2)^{1/2})$. Then under H_0 in (2.10), $-2\log L_5(\Sigma_0)$ converges in distribution to a chi-square distribution with two degrees of freedom as $n \to \infty$, where Σ_0 is any matrix such that $\Sigma_0^{(\tau)} = 0$.

Remark 2.4. Condition (2.11) is similar to (2.5) to ensure that central limit theorem can be employed. The other two conditions in Theorem 2.4 are similar to (2.6) and they make sure that the mean vector can be replaced by the sample mean. The test in [5] requires that $\tau = o(p^s)$ for all $s > 0$ and $\log p = o(n^{1/3})$. However, the new test in Theorem 2.5 only imposes conditions between τ and p. Also note that $\tau = o(p^{1/2})$ is sufficient for $\tau = o((\sum_{1\le i,j\le p} \sigma_{ij}^2)^{1/2})$.

2.3 Power analysis

In this subsection we analyze the powers of our new tests. Denote $\pi_{11} = \mathrm{E}(e_1^2(\Sigma))$, $\pi_{22} = \mathrm{E}(v_1^2(\Sigma))$, $\zeta_{n1} = \mathrm{tr}((\Sigma-\Sigma_0)^2)/\sqrt{\pi_{11}}$, $\zeta_{n2} = 2\mathbf{1}_p^T(\Sigma-\Sigma_0)\mathbf{1}_p/\sqrt{\pi_{22}}$ and $\nu = N(\zeta_{n1}^2+\zeta_{n2}^2)$. Let ξ_β denote the β-quantile of a chi-square distribution with two degrees of freedom and let $\chi^2_{2,\nu}$ denote a noncentral chi-square distribution with two degrees of freedom and noncentrality parameter ν.

Theorem 2.6. *Under conditions of Corollary 2.3 and $H_a : \Sigma \neq \Sigma_0$, we have as $n \to \infty$*

$$P\{-2\log L_j(\Sigma_0) > \xi_{1-\alpha}\} = P\{\chi^2_{2,\nu} > \xi_{1-\alpha}\} + o(1) \tag{2.12}$$

for $j = 1, 2$.

Remark 2.5. Note that under model **(B)**, $\pi_{11} = O(\mathrm{tr}(\Sigma^2)^2)$ and $\pi_{22} = O(\mathbf{1}_p^T\Sigma\mathbf{1}_p)^2$ (see the proof of Lemma 4.4). Therefore, $\zeta_{n1} = O(\mathrm{tr}((\Sigma-\Sigma_0)^2)/\mathrm{tr}(\Sigma^2))$ and $\zeta_{n2} = O(\mathbf{1}_p^T(\Sigma-\Sigma_0)\mathbf{1}_p/(\mathbf{1}_p^T\Sigma\mathbf{1}_p))$ are both natural measures of distance between the null hypothesis and the real model.

Remark 2.6. For a test only using the first estimating equation in (2.4), one needs $\sqrt{n}\zeta_{n1} \to \infty$ to ensure the probability of rejecting H_0 goes to one. Thus it is less powerful than the test using both estimating equations in (2.4) when $\sqrt{n}\zeta_{n2} \to \infty$ and $\sqrt{n}\zeta_{n1}$ is bounded from infinity.

From the above theorem, we conclude that the new test rejects H_0 with probability tending to one when either $\sqrt{n}\zeta_{n1}$ or $\sqrt{n}|\zeta_{n2}|$ goes to infinity. To compare with the power of the test given in [9], we consider the testing problem $H_0 : \Sigma = I_p$ against $H_a : \Sigma \neq I_p$, where $\Sigma = I_p + (dI(|i-j| \leq \tau))_{1\leq i,j\leq p}$ for some positive $d = d(n) \to 0$ as $n \to \infty$. Note that the term $\sqrt{n^2\rho_{2,n}^2 + n\rho_{2,n}}$ in (3.6) of [9] is a typo and it should be $\sqrt{\rho_{2,n}^2 + \rho_{2,n}}$. It is easy to verify that the power of the test in [9] tends to one when $nd^2\tau \to \infty$ for the above example. On the other hand, similar to Theorem 4 in [9], $\sqrt{n}|\zeta_{n2}| \to \infty$ is equivalent to $\sqrt{n}|2\mathbf{1}_p^T(\Sigma - \Sigma_0)\mathbf{1}_p|/p \to \infty$, thus the proposed empirical likelihood ratio test only needs $nd^2\tau^2 \to \infty$ to ensure that the power tends to one. Hence, when $\Sigma = I_p + (d1(|i-j| \leq \tau))_{1\leq i,j\leq p}$ and $\tau = \tau(n) \to \infty$, the proposed empirical likelihood ratio test has a larger local power than the test in [9]. For some other settings, the test in [9] may be more powerful.

For testing the banded structure in Theorems 2.4 and 2.5, we have similar power results. Here we focus on Theorem 2.5. Let $\kappa_{11} = \mathrm{E}(e_1'^2(\Sigma))$, and $\kappa_{22} = \mathrm{E}(\tilde{v}_1'^2(\Sigma))$. Define $\zeta_{n1}' = \mathrm{tr}((\Sigma^{(\tau)})^2)/\sqrt{\kappa_{11}}$, $\zeta_{n2}' = 2\mathbf{1}_p^T\Sigma^{[\tau]}\mathbf{1}_p/\sqrt{\kappa_{22}}$ and $\nu' = N(\zeta_{n1}'^2 + \zeta_{n2}'^2)$.

Theorem 2.7. *Under conditions of Theorem 2.5, when H_0 in (2.10) is false, we have as $n \to \infty$*

$$P\{-2\log L_5(\Sigma_0) > \xi_{1-\alpha}\} = P\{\chi^2_{2,\nu'} > \xi_{1-\alpha}\} + o(1), \tag{2.13}$$

where Σ_0 is any matrix such that $\Sigma_0^{(\tau)} = 0$.

Remark 2.7. As we argue in the introduction, the size of the test in [5] is poor for testing a banded structure. Since the power analysis for the test in [5] is not available, theoretical comparison is impossible. Instead, a simulation comparison is given in the next section, which clearly shows that the proposed test is much more powerful than the test in [5] when many small deviations from the null hypothesis exist. On the other hand, the test in [5] is more powerful when only a large deviation exists. In that case, one can add more equations or replace the second equation by a more relevant one in the proposed empirical likelihood ratio test so as to catch this sparsity effectively.

3 Simulation

In this section we investigate the finite sample behavior of the proposed empirical likelihood ratio tests in terms of both size and power, and compare them with the test in [9] for testing $H_0 : \Sigma = I_p$ and the test in [5] for testing a banded structure.

First we consider testing $H_0 : \Sigma = I_p$ against $H_a : \Sigma \neq I_p$. Draw $1,000$ random samples with sample size $n = 50$ or 200 from the random variable $W_1 + (\delta/n^{1/4})W_2$, where $W_1 \sim N(0, I_p)$, $W_2 \sim N(0, (\sigma_{ij})_{1\le i,j\le p})$ with $\sigma_{ij} = 0.5^{|i-j|}I(|i-j| < \tau)$, and W_1 is independent of W_2. When the sample size is small, it turns out that the size of the proposed empirical likelihood ratio test is a bit large and some calibration is necessary. Here we propose the following bootstrap calibration for the empirical likelihood ratio function $L_2(I_p)$ in Theorem 2.1.

For a given sample $\{\boldsymbol{R}_i^*(I_p)\}_{i=1}^N$, we draw 300 bootstrap samples with size N, say $\{\tilde{\boldsymbol{R}}_i^{*(b)}(I_p)\}_{i=1}^N$ with $b = 1, \cdots, 300$. Based on each bootstrap sample $\{\tilde{\boldsymbol{R}}_i^{*(b)}(I_p)\}_{i=1}^N$, we compute the bootstrapped empirical likelihood ratio function

$$\begin{aligned} L_2^{(b)}(I_p) &= \sup\{\prod_{i=1}^N (Np_i) : p_1 \ge 0, \cdots, p_N \ge 0, \sum_{i=1}^N p_i = 1, \\ &\quad \sum_{i=1}^N p_i \tilde{\boldsymbol{R}}_i^{*(b)}(I_p) = \frac{1}{N}\sum_{j=1}^N \boldsymbol{R}_j^*(I_p)\}. \end{aligned}$$

Then the bootstrap calibrated empirical likelihood ratio test with level γ will reject the null hypothesis $H_0 : \Sigma = I_p$ whenever $-2\log L_2(I_p)$ is larger than the $[300(1-\gamma)]$-th largest value of $\{-2\log L_2^{(b)}(I_p)\}_{b=1}^{300}$. More details on calibration for empirical likelihood ratio test can be found in [18]. We denote the empirical likelihood ratio test based on $-2\log L_2(I_p)$, its bootstrap calibrated version, and the test in [9] by $EL(\gamma)$, $BCEL(\gamma)$

and $CZZ(\gamma)$ respectively, where γ is the significance level.

Table 1 reports the sizes ($\delta = 0$) and powers ($\delta = 1$) of these three tests with level 0.05 by considering $\tau = 10$ and $p = 25, 50, 100, 200, 400, 800$. As we can see, i) the empirical likelihood ratio test has a large size for the small sample size $n = 50$, but the bootstrap calibrated version has an accurate size, which is comparable to the size of the test in [9]; ii) the test in [9] is more powerful for $n = 50$, but less powerful when $n = 200$; iii) for a large sample size, the empirical likelihood ratio test has no need to calibrate.

Next we consider testing $H_0 : \sigma_{ij} = 0$ for $|i - j| \geq \tau$ by drawing $1,000$ random samples from $\tilde{W} + (\delta/n^{1/4})\bar{W}$, where $\tilde{W} \sim N(0, (0.5^{|i-j|}I(|i - j| < \tau))_{1 \leq i,j \leq p})$, $\bar{W} = (\sum_{i=1}^{k} W_i/\sqrt{k}, \cdots, \sum_{i=p}^{p+k} W_i/\sqrt{k})^T$, $W_1, \cdots, W_{p+k}$ are iid with $N(0, 1)$ and independent of $\tilde{W}$. We consider the proposed empirical likelihood ratio test based on Theorem 2.5 ($EL(\gamma)$) and a similar bootstrap calibrated version as in testing $H_0 : \Sigma = I_p$ ($BCEL(\gamma)$), and compare them with the test based on maximum in [5] ($CJ(\gamma)$).

Table 2 reports the sizes ($\delta = 0$) and powers ($\delta = 1$) of these three tests with level 0.05 by considering $\tau = 5$, $k = \tau + 10$ and $p = 25, 50, 100, 200, 400, 800$. From Table 2, we observe that i) the empirical likelihood ratio test has a large size for the small sample size $n = 50$, but the bootstrap calibrated version has an accurate size, which is more accurate than the size of the test in [5]; ii) the test in [5] has little power for all considered cases, and is much less powerful than the proposed empirical likelihood ratio test; iii) for a large sample size, the empirical likelihood ratio test has no need to calibrate.

It is expected that the test based on the maximum statistic in [5] should be more powerful than a test based on a Euclidian distance when a large departure, instead of many small departures, from the null hypothesis happens. To examine this, we test $H_0 : \sigma_{ij} = 0$ for $|i - j| \geq \tau$ by drawing 1,000 random samples with size $n = 200$ from $\tilde{W} + \delta\bar{W}$, where $\tilde{W} \sim N(0, (0.5^{|i-j|}I(|i - j| < \tau))_{1 \leq i,j \leq p})$, $\bar{W} = (\bar{W}_1, \cdots, \bar{W}_p)^T$ with $\bar{W}_1 = \bar{W}_{\tau+1} \sim N(0, 1)$ and $\bar{W}_j = 0$ for $j \neq 1, \tau + 1$. Again, $\tilde{W}$ and $\bar{W}_1$ are independent. We take $\tau = 5$, level 0.05 and $\delta = 0.6, 0.7, 0.8$. This is the sparse case in

Table 1: Sizes and powers are reported for the proposed empirical likelihood method ($EL(\gamma)$), its bootstrap calibrated version ($BCEL(\gamma)$) and the test in [9] ($CZZ(\gamma)$) with significance level $\gamma = 0.05$ for tesing $H_0 : \Sigma = I_p$. We choose $\tau = 10$.

(n,p)	$EL(0.05)$	$BCEL(0.05)$	$CZZ(0.05)$	$EL(0.05)$	$BCEL(0.05)$	$CZZ(0.05)$
	$\delta = 0$	$\delta = 0$	$\delta = 0$	$\delta = 1$	$\delta = 1$	$\delta = 1$
$(50, 25)$	0.127	0.054	0.053	0.296	0.118	0.219
$(50, 50)$	0.148	0.065	0.067	0.324	0.136	0.216
$(50, 100)$	0.138	0.068	0.038	0.317	0.125	0.212
$(50, 200)$	0.168	0.081	0.041	0.310	0.113	0.221
$(50, 400)$	0.151	0.071	0.045	0.342	0.145	0.242
$(50, 800)$	0.154	0.064	0.041	0.337	0.137	0.219
$(200, 25)$	0.065	0.048	0.052	0.348	0.305	0.179
$(200, 50)$	0.058	0.052	0.041	0.336	0.298	0.162
$(200, 100)$	0.068	0.054	0.059	0.353	0.319	0.179
$(200, 200)$	0.056	0.051	0.058	0.358	0.322	0.155
$(200, 400)$	0.069	0.064	0.051	0.374	0.343	0.180
$(200, 800)$	0.058	0.047	0.050	0.366	0.338	0.182

Table 2: Sizes and powers are reported for the proposed empirical likelihood method ($EL(\gamma)$), its bootstrap calibrated version ($BCEL(\gamma)$) and the test in [5] ($CJ(\gamma)$) with significance level $\gamma = 0.05$ for testing $H_0 : \sigma_{ij} = 0$ for all $|i - j| \geq \tau$. We choose $\tau = 5$, $k = \tau + 10$.

(n, p)	$EL(0.05)$	$BCEL(0.05)$	$CZZ(0.05)$	$EL(0.05)$	$BCEL(0.05)$	$CZZ(0.05)$
	$\delta = 0$	$\delta = 0$	$\delta = 0$	$\delta = 1$	$\delta = 1$	$\delta = 1$
$(50, 25)$	0.118	0.036	0.015	0.272	0.093	0.017
$(50, 50)$	0.124	0.049	0.010	0.266	0.097	0.018
$(50, 100)$	0.126	0.057	0.005	0.268	0.099	0.004
$(50, 200)$	0.128	0.058	0.003	0.268	0.100	0.001
$(50, 400)$	0.113	0.053	0.002	0.282	0.121	0.001
$(50, 800)$	0.128	0.062	0.001	0.281	0.109	0.000
$(200, 25)$	0.078	0.062	0.019	0.288	0.253	0.034
$(200, 50)$	0.074	0.059	0.033	0.323	0.286	0.020
$(200, 100)$	0.057	0.053	0.019	0.332	0.304	0.044
$(200, 200)$	0.066	0.046	0.024	0.293	0.263	0.032
$(200, 400)$	0.061	0.052	0.020	0.336	0.304	0.016
$(200, 800)$	0.053	0.046	0.026	0.317	0.297	0.025

which we expect the CJ test to be favored. The powers of $CJ(0.05)$ are 0.074, 0.268 and 0.642 for $\delta = 0.6, 0.7, 0.8$, respectively, while the powers of $EL(0.05)$ are 0.066 for all $\delta = 0.6, 0.7, 0.8$. This confirms the advantage of using maximum when a large departure occurs. However, as we argue in the introduction, the proposed empirical likelihood ratio test is quite flexible in taking information into account. Since only one large departure exists, the second equation in the proposed empirical likelihood ratio test should be replaced by an estimating equation related with this sparsity. Here, we use the first 40% data to get the sample variance $\hat{\sigma}_{ij}$ and find the positions of the largest four values of $|\hat{\sigma}_{ij}|$ for $i - j \geq \tau$. Next we use the remaining 60% data to formulate the empirical likelihood ratio test through replacing $\tilde{v}^{*\prime}$ in the second estimating equation of $L_5(\Sigma)$ by the sum of values at the identified four positions of the covariances $(\mathbf{Y}_i + \mathbf{Y}_{N+i})$. For this modified empirical likelihood ratio test we find that the empirical size is 0.061 and powers are 0.106, 0.255 and 0.542 for $\delta = 0.6, 0.7, 0.8$ respectively. As we can see, the empirical likelihood ratio test with the new second equation improves the power significantly and becomes comparable with the CJ test based on the maximum statistic. In conclusion, the proposed empirical likelihood ratio test is powerful and flexible.

4 Proofs

Without loss of generality, we assume $\mu_0 = 0$ throughout. For simplicity, we use $||\cdot||$ to denote the L_2 norm of a vector or matrix and write $e_i(\Sigma_0) = e_i$, $v_i(\Sigma_0) = v_i$, $e_i^*(\Sigma_0) = e_i^*$, $v_i^*(\Sigma_0) = v_i^*$, $e_i'(\Sigma_0) = e_i'$, $\tilde{v}_i'(\Sigma_0) = \tilde{v}_i'$, $e_i^{*\prime}(\Sigma_0) = e_i^{*\prime}$, $\tilde{v}_i^{*\prime}(\Sigma_0) = \tilde{v}_i^{*\prime}$, $\pi_{11} = \mathrm{E}(e_1^2(\Sigma_0))$ and $\pi_{22} = \mathrm{E}(v_1^2(\Sigma_0))$. We first collect some lemmas and leave the proofs in a supplementary file.

Lemma 4.1. *Under condition* (2.5) *in Theorem 2.1, we have*

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big(\frac{e_i}{\sqrt{\pi_{11}}}, \frac{v_i}{\sqrt{\pi_{22}}} \Big)^T \xrightarrow{d} N(0, I_2). \tag{4.14}$$

Further,

$$\frac{\sum_{i=1}^N e_i^2}{N\pi_{11}} - 1 \xrightarrow{p} 0, \quad \frac{\sum_{i=1}^N v_i^2}{N\pi_{22}} - 1 \xrightarrow{p} 0, \quad \frac{\sum_{i=1}^N e_i v_i}{N\sqrt{\pi_{11}\pi_{22}}} \xrightarrow{p} 0, \tag{4.15}$$

$$\max_{1\le i\le N} |e_i/\sqrt{\pi_{11}}| = o_p(N^{1/2}), \quad \max_{1\le i\le N} |v_i/\sqrt{\pi_{22}}| = o_p(N^{1/2}). \tag{4.16}$$

Lemma 4.2. *Under conditions* (2.5) *and* (2.6) *in Theorem 2.1, we have*

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \Big(\frac{e_i^*}{\sqrt{\pi_{11}}}, \frac{v_i^*}{\sqrt{\pi_{22}}} \Big)^T \xrightarrow{d} N(0, I_2). \tag{4.17}$$

Further,

$$\frac{\sum_{i=1}^N e_i^{*2}}{N\pi_{11}} - 1 \xrightarrow{p} 0, \quad \frac{\sum_{i=1}^N v_i^{*2}}{N\pi_{22}} - 1 \xrightarrow{p} 0, \quad \frac{\sum_{i=1}^N e_i^* v_i^*}{N\sqrt{\pi_{11}\pi_{22}}} \xrightarrow{p} 0, \tag{4.18}$$

$$\max_{1\le i\le N} |e_i^*/\sqrt{\pi_{11}}| = o_p(N^{1/2}), \quad \max_{1\le i\le N} |v_i^*/\sqrt{\pi_{22}}| = o_p(N^{1/2}). \tag{4.19}$$

Lemma 4.3. *Under conditions of Corollary 2.2, for any $\delta > 0$, we have*

$$\mathrm{E}|e_1|^{2+\delta} \le q^\delta \Big(\sum_{i=1}^p \sum_{j=1}^p \mathrm{E}|X_{1i}X_{1j} - \sigma_{ij}|^{2+\delta} \Big)^2$$

and

$$\mathrm{E}|v_1|^{2+\delta} \le 2^{4+\delta} q^{1+\delta} \sum_{i=1}^p \sum_{j=1}^p \mathrm{E}|X_{1i}X_{1j} - \sigma_{ij}|^{2+\delta}.$$

Lemma 4.4. *Under conditions of Corollary 2.3, we have*

$$\mathrm{E}e_1^4/(\mathrm{E}e_1^2)^2 = O(1) \ \ \textit{and} \ \ \mathrm{E}v_1^4/(\mathrm{E}v_1^2)^2 = O(1).$$

Lemma 4.5. *Under conditions of Theorem 2.5, we have*

$$\mathrm{E}e_1'^4/(\mathrm{E}e_1'^2)^2 = O(1), \quad \mathrm{E}\tilde{v}_1'^4/(\mathrm{E}\tilde{v}_1'^2)^2 = O(1), \tag{4.20}$$

$$\mathrm{E}\{\sum_{i=1}^N (e_i^{*\prime} - e_i')^2 + [\sum_{i=1}^N (e_i^{*\prime} - e_i')]^2\} = o(N\mathrm{E}[e_1'^2]), \tag{4.21}$$

$$\mathrm{E}\{\sum_{i=1}^N (\tilde{v}_i^{*\prime} - \tilde{v}_i')^2 + [\sum_{i=1}^N (\tilde{v}_i^{*\prime} - \tilde{v}_i')]^2\} = o(N\mathrm{E}[\tilde{v}_1'^2]). \tag{4.22}$$

Proof of Theorem 2.1. Put $\hat{e}_i = e_i/\sqrt{\pi_{11}}$, $\hat{v}_i = v_i/\sqrt{\pi_{22}}$ and $\hat{\boldsymbol{R}}_i = (\hat{e}_i, \hat{v}_i)^T$ for $i = 1, \cdots, N$. Then it is easy to see that $-2\log L_1(\Sigma_0) = 2\sum_{i=1}^N \log\{1 + \rho^T \hat{\boldsymbol{R}}_i\}$, where $\rho = (\rho_1, \rho_2)^T$ satisfies

$$\frac{1}{N}\sum_{i=1}^N \frac{\hat{\boldsymbol{R}}_i}{1 + \rho^T \hat{\boldsymbol{R}}_i} = 0. \tag{4.23}$$

Using Lemma 4.1 and similar arguments in the proof of (2.14) in [17], we can show that

$$||\rho|| = O_p(N^{-1/2}). \tag{4.24}$$

Then it follows from (4.16) and (4.24) that

$$\max_{1\le i\le N} |\frac{\rho^T \hat{\boldsymbol{R}}_i}{1 + \rho^T \hat{\boldsymbol{R}}_i}| = o_p(1). \tag{4.25}$$

By (4.23), we have

$$\begin{aligned} 0 &= \tfrac{1}{N}\textstyle\sum_{i=1}^N \frac{\rho^T \hat{\boldsymbol{R}}_i}{1+\rho^T \hat{\boldsymbol{R}}_i} \\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^N \rho^T \hat{\boldsymbol{R}}_i \{1 - \rho^T \hat{\boldsymbol{R}}_i + \frac{(\rho^T \hat{\boldsymbol{R}}_i)^2}{1+\rho^T \hat{\boldsymbol{R}}_i}\} \\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^N \rho^T \hat{\boldsymbol{R}}_i - \tfrac{1}{N}\sum_{i=1}^N (\rho^T \hat{\boldsymbol{R}}_i)^2 + \tfrac{1}{N}\sum_{i=1}^N \frac{(\rho^T \hat{\boldsymbol{R}}_i)^3}{1+\rho^T \hat{\boldsymbol{R}}_i} \\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^N \rho^T \hat{\boldsymbol{R}}_i - \tfrac{1+o_p(1)}{N}\sum_{i=1}^N (\rho^T \hat{\boldsymbol{R}}_i)^2, \end{aligned}$$

which implies

$$\frac{1}{N}\sum_{i=1}^N \rho^T \hat{\boldsymbol{R}}_i = \frac{1 + o_p(1)}{N}\sum_{i=1}^N (\rho^T \hat{\boldsymbol{R}}_i)^2. \tag{4.26}$$

Using (4.23)–(4.25) and Lemma 4.1, we have

$$
\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\boldsymbol{R}}_i}{1+\rho^T\hat{\boldsymbol{R}}_i} \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\{1-\rho^T\hat{\boldsymbol{R}}_i+\frac{(\rho^T\hat{\boldsymbol{R}}_i)^2}{1+\rho^T\hat{\boldsymbol{R}}_i}\} \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i-\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^T\rho+\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\boldsymbol{R}}_i(\rho^T\hat{\boldsymbol{R}}_i)^2}{1+\rho^T\hat{\boldsymbol{R}}_i} \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i-\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^T\rho+O_p(\max_{1\le i\le N}||\frac{\hat{\boldsymbol{R}}_i}{1+\rho^T\hat{\boldsymbol{R}}_i}||\frac{1}{N}\sum_{i=1}^{N}(\rho^T\hat{\boldsymbol{R}}_i)^2) \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i-\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^T\rho+o_p(N^{1/2}\rho^T\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^T\rho) \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i-\frac{1}{N}\sum_{i=1}^{n}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^T\rho+o_p(N^{1/2}),
\end{aligned}
$$

which implies that

$$
\rho=\{\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^T\}^{-1}\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i+o_p(N^{-1/2}). \tag{4.27}
$$

Hence, using Taylor expansion, (4.26), (4.27) and Lemma 4.1, we have

$$
\begin{aligned}
& -2\log L_1(\Sigma_0) \\
= \quad & 2\textstyle\sum_{i=1}^{N}\rho^T\hat{\boldsymbol{R}}_i-(1+o_p(1))\sum_{i=1}^{N}(\rho^T\hat{\boldsymbol{R}}_i)^2 \\
= \quad & (1+o_p(1))\rho^T\textstyle\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^T\rho \qquad (4.28)\\
= \quad & (1+o_p(1))(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i)^T(\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i\hat{\boldsymbol{R}}_i^T)^{-1}(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\boldsymbol{R}}_i)+o_p(1) \\
\xrightarrow{d} \quad & \chi_2^2 \quad \text{as } n\to\infty.
\end{aligned}
$$

Similarly we can show that $-2\log L_2(\Sigma_0)\xrightarrow{d}\chi_2^2$ by using Lemma 4.2. □

Proof of Corollary 2.2. First we prove the case of known μ. Lemma 4.3 implies that under condition A2,

$$\mathrm{E}|e_1|^{2+\delta} = O(q^{2+\delta}) \quad \text{and} \quad \mathrm{E}|v_1|^{2+\delta} = O(q^{2+\delta}).$$

Further, under condition A1, we have for a constant $C > 0$, $\pi_{11} = \mathrm{tr}(\Theta^2) \geq qC$ and $\pi_{22} = \mathbf{1}_q^T \Theta \mathbf{1}_q \geq qC$. Thus,

$$\mathrm{E}|e_1|^{2+\delta}/\pi_{11}^{(2+\delta)/2} = O(q^{(2+\delta)/2}) = O(p^{2+\delta})$$

and

$$\mathrm{E}|v_1|^{2+\delta}/\pi_{22}^{(2+\delta)/2} = O(q^{(2+\delta)/2}) = O(p^{2+\delta}).$$

Therefore, (2.5) in Theorem 2.1 follows from condition A3, i.e., Corollary 2.2 holds for the case of known μ.

Next we prove the case of unknown μ. Since (2.5) is satisfied, by Theorem 2.1, it is enough to show that condition (2.6) holds. Under condition $\max_{1\leq i\leq p} \sigma_{ii} < C_0$, we have

$$(\mathrm{tr}(\Sigma^2))^2 = \left(\sum_{1\leq i,j\leq p} \sigma_{ij}^2 \right)^2 \leq q^2 (\max_{1\leq i\leq p} \sigma_{ii}^2) \leq C_0^2 q^2 \tag{4.29}$$

and

$$\left(\mathbf{1}_p^T \Sigma \mathbf{1}_p\right)^2 \leq q^2 (\max_{1\leq i\leq p} \sigma_{ii}^2) \leq C_0^2 q^2. \tag{4.30}$$

On the other hand, under condition A1, there exists a constant $C > 0$ such that

$$\pi_{11} = \mathrm{tr}(\Theta^2) \geq qC \quad \text{and} \quad \pi_{22} = \mathbf{1}_q^T \Theta \mathbf{1}_q \geq qC. \tag{4.31}$$

Note that condition A3 implies that $p = o(n^{1/4})$ and $q = o(n^{1/2})$. Thus, by (4.29), (4.30) and (4.31), we have

$$N\mathrm{E}e_1^2 = N\pi_{11} \geq CNq \geq (\mathrm{tr}(\Sigma^2))^2$$

and

$$\sqrt{N}\mathrm{E}v_1^2 = \sqrt{N}\pi_{22} \geq \sqrt{N}qC > \left(\mathbf{1}_p^T \Sigma \mathbf{1}_p\right)^2.$$

Hence, (2.6) holds and the proof of Corollary 2.2 is complete. □

Proof of Corollary 2.3. It follows from Lemma 4.4 that (2.5) in Theorem 2.1 holds with $\delta = 2$. Hence Corollary 2.3 follows from Theorem 2.1 when μ is known.

When μ is unknown, it follows from Lemma 4.4 that (2.5) holds. Further, through the proof of Lemma 4.4, we have

$$\mathrm{E}[e_1^2] \geq C^2(\mathrm{tr}(\Sigma^2))^2 \text{ and } \mathrm{E}[v_1^2] \geq C(\mathbf{1}_p^T \Sigma \mathbf{1}_p)^2,$$

i.e., condition (2.6) holds. Thus, by Theorem 2.1, Corollary 2.3 holds for unknown μ. □

Proof of Theorem 2.4. Since the required moment conditions are satisfied, it follows from the same arguments as in the proof of Theorem 2.1. □

Proof of Theorem 2.5. Using Lemma 4.5, the proof of Theorem 2.5 follows from the same arguments as in the proof of Theorem 2.1. □

Proof of Theorem 2.6. We only show the case of known μ since the case of unknown μ can be proved similarly.

First we consider the case of $\nu = o(N)$. Note that under the alternative hypothesis H_a, $\mathrm{E}\boldsymbol{Y}_1 = \Sigma$ and write for $1 \leq i \leq N$,

$$e_i(\Sigma_0) = e_i(\Sigma) + \mathrm{tr}((\Sigma - \Sigma_0)^2) + \mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_i + \boldsymbol{Y}_{N+i} - 2\Sigma))$$

and $v_i(\Sigma_0) = v_i(\Sigma) + 2\mathbf{1}_p^T(\Sigma - \Sigma_0)\mathbf{1}_p$, where $q = p^2$. As a result, we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big(\frac{e_i(\Sigma_0)}{\sqrt{\pi_{11}}}, \frac{v_i(\Sigma_0)}{\sqrt{\pi_{22}}} \Big)^T \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big(\frac{e_i(\Sigma)}{\sqrt{\pi_{11}}}, \frac{v_i(\Sigma)}{\sqrt{\pi_{22}}} \Big)^T + \sqrt{N}\Big(\zeta_{n1}, \zeta_{n2}\Big)^T + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big(\eta_i(\Sigma), 0\Big)^T, \qquad (4.32)
\end{aligned}$$

where $\eta_i(\Sigma) = \mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_i + \boldsymbol{Y}_{N+i} - 2\Sigma))/\sqrt{\pi_{11}}$. Since $\mathrm{E}[\eta_i(\Sigma)] = 0$ and

$$\begin{aligned}
\mathrm{E}[\eta_i(\Sigma)]^2 &= 4\mathrm{E}(\mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_1 - \Sigma))^2)/\pi_{11} \\
&\leq 4\mathrm{E}(\mathrm{tr}((\Sigma - \Sigma_0)^2)\mathrm{tr}((\boldsymbol{Y}_1 - \Sigma)^2))/\pi_{11} \qquad (4.33) \\
&= O[\mathrm{tr}((\Sigma - \Sigma_0)^2)/\sqrt{\pi_{11}}] = o(1),
\end{aligned}$$

we have

$$\frac{1}{N}\sum_{i=1}^{N}\eta_i^2(\Sigma) = o_p(1) \quad \text{and} \quad \frac{\max_{1\le i\le N}|\eta_i(\Sigma)|}{\sqrt{N}} \le \sqrt{\frac{\sum_{i=1}^{N}\eta_i^2(\Sigma)}{N}} \xrightarrow{p} 0. \tag{4.34}$$

Hence it follows from Lemma 4.1 that

$$V_N \xrightarrow{d} N(0, I_2), \tag{4.35}$$

where

$$V_N = \begin{pmatrix} V_{N1} \\ V_{N2} \end{pmatrix} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\begin{pmatrix} \frac{e_i(\Sigma_0)}{\sqrt{\pi_{11}}} \\ \frac{v_i(\Sigma_0)}{\sqrt{\pi_{22}}} \end{pmatrix} - \begin{pmatrix} \zeta_{n1} \\ \zeta_{n2} \end{pmatrix}\right\}.$$

Put $W_i = (\frac{e_i(\Sigma_0)}{\sqrt{\pi_{11}}}, \frac{v_i(\Sigma_0)}{\sqrt{\pi_{22}}})^T$. Then it follows from the proof of Theorem 2.1 that

$$\begin{aligned}
&-2\log L_1(\Sigma_0) \\
=\ & (1+o_p(1))(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N} W_i)^T(\tfrac{1}{N}\sum_{i=1}^{N} W_i W_i^T)^{-1}\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N} W_i + o_p(1) \\
=\ & (1+\zeta_{n1}^2+\zeta_{n2}^2)^{-1}[(1+\zeta_{n2}^2)(V_{N1}+\sqrt{N}\zeta_{n1})^2 - 2\zeta_{n1}\zeta_{n2}(V_{N1}+\sqrt{N}\zeta_{n1}) \qquad (4.36)\\
& \times(V_{n2}+\sqrt{N}\zeta_{n2}) + (1+\zeta_{n1}^2)(V_{n2}+\sqrt{N}\zeta_{n2})^2] + o_p(1) \\
=\ & (V_{N1}+\sqrt{N}\zeta_{n1})^2(1+o_p(1)) + (V_{n2}+\sqrt{N}\zeta_{n2})^2(1+o_p(1)) + o_p(1).
\end{aligned}$$

If the limit of $\nu = N(\zeta_{n1}^2+\zeta_{n2}^2)$, say ν_0, is finite, then it follows from (4.35) and (4.36) that $-2\log L_1(\Sigma_0)$ converges in distribution to a noncentral chi-square distribution with two degrees of freedom and noncentrality parameter ν_0. If ν goes to infinite, the limit of the right hand side of (2.12) is 1. By (4.36), we have

$$\begin{aligned}
&-2\log L_1(\Sigma_0) \\
\ge\ & (\tfrac{N\zeta_{n1}^2}{2} - V_{N1}^2)(1+o_p(1)) + (\tfrac{N\zeta_{n2}^2}{2} - V_{N2}^2)(1+o_p(1)) + o_p(1) \qquad (4.37)\\
=\ & \tfrac{\nu}{2}(1+o_p(1)) - (V_{N1}^2+V_{N2}^2)(1+o_p(1)) + o_p(1) \xrightarrow{p} \infty,
\end{aligned}$$

which implies that the limit of the left-hand side of (2.12) is also 1. Thus (2.12) holds when $\nu = o(N)$.

For the case of $\liminf \nu/N > 0$, we first consider the case of $\liminf \zeta_{n2}^2 > 0$. Since $\sum_{i=1}^N p_i \boldsymbol{R}_i(\Sigma_0) = 0$ implies that $\sum_{i=1}^N p_i \nu_i(\Sigma_0) = 0$, we have

$$L_1(\Sigma_0) \le \sup\{\prod_{i=1}^N (Np_i) : p_1 \ge 0, \cdots, p_N \ge 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i v_i(\Sigma_0) = 0\} \tag{4.38}$$
$$= \sup\{\prod_{i=1}^N (Np_i) : p_1 \ge 0, \cdots, p_N \ge 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \frac{v_i(\Sigma_0)}{\sqrt{\pi_{22}}} = 0\}.$$

Define

$$L^*(\theta) = \sup\{\prod_{i=1}^N (Np_i) : p_1 \ge 0, \cdots, p_N \ge 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i (\frac{v_i(\Sigma_0)}{\sqrt{\pi_{22}}} - \zeta_{n2}) = \theta\}.$$

Put $\theta^* = \frac{1}{N}\sum_{i=1}^N (\frac{v_i(\Sigma_0)}{\sqrt{\pi_{22}}} - \zeta_{n2})$. Then

$$\log L^*(\theta^*) = 0. \tag{4.39}$$

Since $\mathrm{E}\{v_i(\Sigma_0)/\sqrt{\pi_{22}} - \zeta_{n2}\} = \mathrm{E}\{v_i(\Sigma)/\sqrt{\pi_{22}}\} = 0$ and $\mathrm{E}\{v_i(\Sigma_0)/\sqrt{\pi_{22}} - \zeta_{n2}\}^2 = 1$ under $H_a : \Sigma \ne \Sigma_0$, we have by using Chebyshev's inequality that

$$P(|\theta^*| > N^{-2/5}) \to 0. \tag{4.40}$$

Using $\mathrm{E}\{v_i(\Sigma_0)/\sqrt{\pi_{22}} - \zeta_{n2}\}^2 = 1$, similar to the proof of (4.37), we can show that

$$-2\log L^*(\theta_1^*) \xrightarrow{p} \infty \text{ and } -2\log L^*(\theta_2^*) \xrightarrow{p} \infty,$$

where $\theta_1^* = N^{-1/4}$ and $\theta_2^* = -N^{-1/4}$, which satisfy $N(\theta_1^*)^2 = o(N)$ and $N(\theta_2^*)^2 = o(N)$. It follows from [10] that the set $\{\theta : -2\log L^*(\theta) \le c\} =: I_c$ is convex for any c. Take $c = \min\{-2\log L^*(\theta_1^*), -2\log L^*(\theta_2^*)\}/2$. By (4.39), we have that $\theta^* \in I_c$. Thus, if $-\zeta_{n2} \in I_c$, then $-a\zeta_{n2} + (1-a)\theta^* \in I_c$ for any $a \in [0,1]$, which implies that one of θ_1^* and θ_2^* must belong to I_c. As a result, we have

$$P(|\theta^*| \le N^{-2/5}, -\zeta_{n2} \in I_c) \le P(\theta_1^* \in I_c \text{ or } \theta_2^* \in I_c)$$
$$= P(\min\{-2\log L^*(\theta_1^*), -2\log L^*(\theta_2^*)\} = 0) \to 0$$

which, together with (4.40), implies

$$
\begin{aligned}
&P(-2\log L^*(-\zeta_{n2}) > c)\\
=&P(-\zeta_{n2} \notin I_c) \geq 1 - P(|\theta^*| \leq N^{-2/5}, -\zeta_{n2} \in I_c) - P(|\theta^*| > N^{-2/5}) \to 1,
\end{aligned}
$$

and therefore

$$
-2\log L^*(-\zeta_{n2}) \xrightarrow{p} \infty \tag{4.41}
$$

since $c \xrightarrow{p} \infty$. Hence, combining with (4.38), we have

$$
P(-2\log L_1(\Sigma_0) > \xi_{1-\alpha}) \geq P(-2\log L^*(-\zeta_{n2}) > \xi_{1-\alpha}) \to 1
$$

when $\liminf \zeta_{n2}^2 > 0$.

Next we consider the case of $\liminf \zeta_{n1} > 0$. Define

$$
\pi_{33} = \mathrm{E}\{\mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_i + \boldsymbol{Y}_{N+i} - 2\Sigma))\}^2 \text{ and } \zeta_{n3} = \frac{\mathrm{tr}((\Sigma - \Sigma_0)^2)}{\sqrt{\pi_{11} + \pi_{33}}}.
$$

As before, we have

$$
\begin{aligned}
&L_1(\Sigma_0)\\
&\leq \sup\{\textstyle\prod_{i=1}^N (Np_i) : p_1 \geq 0, \cdots, p_N \geq 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i e_i(\Sigma_0) = 0\} \qquad (4.42)\\
&= \sup\{\textstyle\prod_{i=1}^N (Np_i) : p_1 \geq 0, \cdots, p_N \geq 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \frac{e_i(\Sigma_0)}{\sqrt{\pi_{11}+\pi_{33}}} = 0\}.
\end{aligned}
$$

Define

$$
\begin{aligned}
L^{**}(\theta) = \sup\{&\prod_{i=1}^N (Np_i) : p_1 \geq 0, \cdots, p_N \geq 0, \sum_{i=1}^N p_i = 1,\\
&\sum_{i=1}^N p_i\Big(\frac{e_i(\Sigma_0)}{\sqrt{\pi_{11} + \pi_{33}}} - \zeta_{n3}\Big) = \theta\}.
\end{aligned}
$$

Since $e_1(\Sigma)$ and $\mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_1 + \boldsymbol{Y}_{N+1} - 2\Sigma))$ are two uncorrelated variables with zero means, we have

$$
\mathrm{Var}(e_1(\Sigma) + \mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_1 + \boldsymbol{Y}_{N+1} - 2\Sigma))) = \pi_{11} + \pi_{33}.
$$

As we have shown in the proof of Lemma 4.4, $\mathrm{E}|e_1(\Sigma)|^4 = o(N\pi_{11}^2)$. Following the same lines for estimating $\mathrm{E}(v_1^4)$ in the end of the proof of Lemma 4.4, we have

$$\mathrm{E}\{\mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_1 + \boldsymbol{Y}_{N+1} - 2\Sigma))\}^4 = O(\pi_{33}^2).$$

Then it follows that

$$\begin{aligned} & \mathrm{E}\{e_1(\Sigma) + \mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_1 + \boldsymbol{Y}_{N+1} - 2\Sigma))\}^4 \\ \leq\ & 8(\mathrm{E}|e_1(\Sigma)|^4 + \mathrm{E}\{\mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_1 + \boldsymbol{Y}_{N+1} - 2\Sigma))\}^4) \\ =\ & o(N(\pi_{11} + \pi_{33})^2). \end{aligned}$$

Write

$$\frac{e_i(\Sigma_0)}{\sqrt{\pi_{11} + \pi_{33}}} - \zeta_{n3} = \frac{e_i(\Sigma) + \mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_i + \boldsymbol{Y}_{N+i} - 2\Sigma))}{\sqrt{\pi_{11} + \pi_{33}}}.$$

Then we have

$$\mathrm{E}(\frac{e_i(\Sigma_0)}{\sqrt{\pi_{11} + \pi_{33}}} - \zeta_{n3})^4 = \frac{\mathrm{E}(e_i(\Sigma) + \mathrm{tr}((\Sigma - \Sigma_0)(\boldsymbol{Y}_i + \boldsymbol{Y}_{N+i} - 2\Sigma)))^4}{(\pi_{11} + \pi_{33})^2} = o(N).$$

This ensures the validity of Wilks' theorem for $-2\log L^{**}(0)$, that is, $-2\log L^{**}(0)$ converges in distribution to a chi-square distribution with one degree of freedom. Similar to the proof of (4.37), we can show that

$$-2\log L^{**}(\theta_1^*) \xrightarrow{p} \infty \text{ and } -2\log L^{**}(\theta_2^*) \xrightarrow{p} \infty,$$

where $\theta_1^* = N^{-1/4}$ and $\theta_2^* = -N^{-1/4}$, which satisfy $N(\theta_1^*)^2 = o(N)$ and $N(\theta_2^*)^2 = o(N)$.

Put $\theta^{**} = \frac{1}{N}\sum_{i=1}^N(\frac{e_i(\Sigma_0)}{\sqrt{\pi_{11}+\pi_{33}}} - \zeta_{n3})$. Then

$$\log L^{**}(\theta^{**}) = 0. \tag{4.43}$$

Since

$$\begin{aligned} & \mathrm{E}\{e_i(\Sigma_0)/\sqrt{\pi_{11} + \pi_{33}} - \zeta_{n3}\} \\ =\ & \mathrm{E}\{\tfrac{e_i(\Sigma)+\mathrm{tr}((\Sigma-\Sigma_0)(\boldsymbol{Y}_i+\boldsymbol{Y}_{N+i}-2\Sigma))}{\sqrt{\pi_{11}+\pi_{33}}}\} = 0 \end{aligned}$$

and

$$\mathrm{E}\{\frac{e_i(\Sigma_0)}{\sqrt{\pi_{11}+\pi_{33}}}-\zeta_{n3}\}^2=\mathrm{E}\{\frac{e_i(\Sigma)+\mathrm{tr}((\Sigma-\Sigma_0)(\boldsymbol{Y}_i+\boldsymbol{Y}_{N+i}-2\Sigma))}{\sqrt{\pi_{11}+\pi_{33}}}\}^2=1$$

under $H_a:\Sigma\neq\Sigma_0$, we have from Chebyshev's inequality that

$$P(|\theta^{**}|>N^{-2/5})\to 0. \tag{4.44}$$

By (4.33), we have $\pi_{33}/\pi_{11}=O(\zeta_{n1})$, which implies that there exists a constant $M>0$ such that

$$\zeta_{n3}/N^{-1/4}=N^{1/4}\zeta_{n1}\frac{\sqrt{\pi_{11}}}{\sqrt{\pi_{11}+\pi_{33}}}\geq N^{1/4}\zeta_{n1}\{1+M\zeta_{n1}\}^{-1/2}\to\infty$$

since $\liminf\zeta_{n1}>0$.

Using (4.43), (4.44), and the same arguments in proving (4.41), we have $-2\log L^{**}(-\zeta_{n3})\xrightarrow{p}\infty$. Hence, combining with (4.42), we have

$$P(-2\log L_1(\Sigma_0)>\xi_{1-\alpha})\geq P(-2\log L^{**}(-\zeta_{n3})>\xi_{1-\alpha})\to 1$$

when $\liminf\zeta_{n1}^2>0$. Therefore (2.12) holds when $\liminf\zeta_{n1}>0$. This completes the proof of Theorem 2.6. □

Proof of Theorem 2.7. The proof is similar to that of Theorem 2.6. □

Acknowledgement

We thank the editor Professor Runze Li, an associate editor and two reviewers for their constructive comments. Zhang's research was supported by NSFC grants 11171074 and 10801118, and the Fundamental Research Funds for the Central Universities. Peng's research was supported by NSF grant DMS-1005336. Wang's research was partly supported by the Bob Price Fellowship at the Georgia Institute of Technology.

Supplement

The supplementary file [22] contains detailed proofs of Lemmas 4.1–4.5.

References

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* Third edition. Wiley, New York.

[2] Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica* **6**, 311–329.

[3] Bickel, P. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36**, 199–227.

[4] Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106**, 672–684.

[5] Cai, T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39**, 1496–1525.

[6] Cai, T., Zhang, C. H. and Zhou, H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38**, 2118–2144.

[7] Chen, S., Peng, L. and Qin, Y. (2009). Empirical likelihood methods for high dimension. *Biometrika* **96**, 711–722.

[8] Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38**, 808–835.

[9] Chen, S., Zhang, L. and Zhong, P. (2010). Tests for High-Dimensional Covariance Matrices. *J. Amer. Statist. Assoc.* **105**, 810-819.

[10] Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58**, 109–127.

[11] Hjort, N. L., McKeague, I. W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37**, 1079–1111.

[12] John, S. (1971). Some optimal multivariate tests. *Biometrika* **58**, 123–127.

[13] John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59**, 169–173.

[14] Ledoit, O., and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compare to the sample size. *Ann. Statist.* **30**, 1081–1102.

[15] Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1**, 700–709.

[16] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

[17] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.

[18] Owen, A. (2001). *Empirical Likelihood*. Chapman & Hall/CRC.

[19] Peng, L., Qi, Y. and Wang, F. (2013). Empirical likelihood method for high dimensional means. *Statistical Science*. To appear.

[20] Qiu, Y. and Chen, S. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.* **40**, 1285-1314.

[21] Rothman, A. J., Levina, E., and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* **104**, 177–186.

[22] Zhang, R. M., Peng, L. and Wang, R. (2013). Supplement to "Tests for covariance matrix with fixed or divergent dimension". DOI: 10.1214/13-AOS1136SUPP.