Supplement to "Two-Sample Covariance Matrix Testing And Support Recovery in High-Dimensional and Sparse Settings"

Tony Cai[1], Weidong Liu[1,2] and Yin Xia[1]

Abstract

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

1 Proof of Proposition 1

According to the proof of Theorem 1 and letting $d = 1$ in (29), we have

$$\begin{aligned}&\mathsf{P}_{H_0}(M_{ij} \geq q_\alpha + 4\log p - \log\log p)\\&= (1+o(1))\mathsf{P}(|N(0,1)| \geq q_\alpha + 4\log p - \log\log p).\end{aligned}$$

Then we can get, under (C2) (or (C2)*), for $0 < \alpha < 1$,

$$\begin{aligned}\text{Type I error} = \mathsf{P}_{H_0}(\Phi_\alpha = 1) &\leq \sum_{1\leq i\leq j\leq p} \mathsf{P}_{H_0}(M_{ij} \geq q_\alpha + 4\log p - \log\log p)\\&\leq -\log(1-\alpha) + o(1). \qquad \blacksquare\end{aligned}$$

[1]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, tcai@wharton.upenn.edu. The research was supported in part by NSF FRG Grant DMS-0854973.

[2]Department of Mathematics and Institute of Natural Sciences, Shanghai Jiao Tong University, China.

2 Proof of Proposition 2

Proposition 2 (i) is a directly result of Theorem 2. For (ii), by the analysis in Li and Chen (2012), pages 914 and 915, if

$$\mathrm{DIST}_{LC}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) := \frac{\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2}{\sqrt{\frac{1}{n_1^2}\|\mathbf{\Sigma}_1\|_F^4 + \frac{1}{n_2^2}\|\mathbf{\Sigma}_2\|_F^4}} = o(1), \tag{1}$$

then $\beta_{LC}(\alpha) \to \alpha$. By some elementary calculations, we have $\max_{i,j}\{\theta_{ij1} + \theta_{ij2}\} = O(1)$ and $\max_{i,j}|\sigma_{ij1} - \sigma_{ij2}| = O(c_{n,p}/\sqrt{n})$ for $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{S}(s_p, c_{n,p})$. Hence, $\mathrm{DIST}_{LC}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = O(s_p c_{n,p}^2/p)$. This shows that if $s_p = o(p/c_{n,p}^2)$, the asymptotic power $\lim_{(n,p)\to\infty} \beta_{LC}(\alpha) = \alpha$. For (iii), Lemma 2.3 in Srivastava and Yanagihara (2010) shows that, if

$$\mathrm{DIST}_{SY}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = n\Big|\frac{tr(\mathbf{\Sigma}_1^2)/p}{(tr(\mathbf{\Sigma}_1)/p)^2} - \frac{tr(\mathbf{\Sigma}_2^2)/p}{(tr(\mathbf{\Sigma}_2)/p)^2}\Big| = o(1), \tag{2}$$

then $\beta_{SY}(\alpha) \to \alpha$. It can be verified that, for $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{S}(s_p, c_{n,p})$, $\mathrm{DIST}_{SY}(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = O(\sqrt{n} s_p c_{n,p}/p)$. This shows that, if $s_p = o(p/(\sqrt{n} c_{n,p}))$, then the asymptotic power $\lim_{(n,p)\to\infty} \beta_{SY}(\alpha) = \alpha$. ∎

3 Proof of Proposition 3

By (34) in the proof of Theorem 5, we only need to show that

$$\mathsf{P}\Big(\max_{(i,j)\in A\setminus E_0} M_{ij} \geq 4\log p - \log\log p + q_\alpha\Big) \to \alpha,$$

where the set A is defined in the proof of Theorem 1 and

$$E_0 = \{(i,j) : 1 \leq i \leq j \leq p, \sigma_{ij1} \neq \sigma_{ij2}\} \cup \{(i,i) : 1 \leq i \leq p\}.$$

Since $s_0(p) = o(p)$, we have $\mathrm{Card}(E_0) = o(p^2)$. The rest proof follows exactly the same as that of Theorem 1. ∎

4 Proof of Lemmas 3-5

Proof of Lemma 3. We only prove (20) in Lemma 3 because the proof of (21) is similar. Without loss of generality, we assume that $\mathsf{E}\mathbf{X} = 0$ and $\mathsf{Var}(X_i) = 1$ for

$1 \leq i \leq p$. Let

$$\tilde{\theta}_{ij1} = \frac{1}{n_1} \sum_{k=1}^{n_1} \Big[X_{ki} X_{kj} - \widetilde{\sigma}_{ij1} \Big]^2 \quad \text{with } \widetilde{\sigma}_{ij1} = \frac{1}{n_1} \sum_{k=1}^{n_1} X_{ki} X_{kj}.$$

By the proof of Lemma 2 in Cai and Liu (2011), we have for any $M > 0$, there exists a constant C such that

$$\mathsf{P}\Big(\max_{i,j} |\hat{\theta}_{ij1} - \tilde{\theta}_{ij1}| \geq C\sqrt{\log p/n} \Big) = O(p^{-M} + n^{-\epsilon/8}). \tag{3}$$

Write

$$\tilde{\theta}_{ij1} - \theta_{ij1} = \frac{1}{n_1} \sum_{k=1}^{n_1} \Big[(X_{ki}X_{kj})^2 - \mathsf{E}(X_{ki}X_{kj})^2 \Big] - \widetilde{\sigma}_{ij1}^2 + \sigma_{ij1}^2.$$

We can see that

$$\mathsf{P}\Big(\max_{i,j} |\widetilde{\sigma}_{ij1} - \sigma_{ij1}| \geq C\sqrt{\log p/n} \Big) = O(p^{-M} + n^{-\epsilon/8}). \tag{4}$$

We first assume that (C2) holds. It suffices to show that

$$\mathsf{P}\Big(\max_{i,j} \Big| \frac{1}{n_1} \sum_{k=1}^{n_1} \Big[(X_{ki}X_{kj})^2 - \mathsf{E}(X_{ki}X_{kj})^2 \Big] \Big| \geq C\sqrt{\frac{\log p}{n}} \Big) = O(p^{-M}). \tag{5}$$

Define

$$\hat{X}_{kj} = X_{kj} I\{|X_{kj}| \leq \tau\sqrt{\log(p+n)}\},$$

where τ is sufficiently large. We have

$$\begin{aligned} |\mathsf{E}(X_{ki}X_{kj})^2 - \mathsf{E}(X_{ki}\hat{X}_{kj})^2| &\leq C\Big(\mathsf{E}X_{kj}^4 I\{|X_{kj}| \geq \tau\sqrt{\log(p+n)}\} \Big)^{1/2} \\ &\leq C(n+p)^{-\tau^2\eta/2} \Big(\mathsf{E}X_{kj}^4 \exp\Big(2^{-1}\eta X_{kj}^2\Big) \Big)^{1/2} \\ &\leq C(n+p)^{-\tau^2\eta/2}, \end{aligned} \tag{6}$$

where C does not depend on n, p. Thus it follows that

$$\begin{aligned} &\mathsf{P}\Big(\max_{i,j} \Big| \frac{1}{n_1} \sum_{k=1}^{n_1} \Big[(X_{ki}X_{kj})^2 - \mathsf{E}(X_{ki}X_{kj})^2 \Big] \Big| \geq C\sqrt{\frac{\log p}{n}} \Big) \\ &\quad \leq \mathsf{P}\Big(\max_{i,j} \Big| \frac{1}{n_1} \sum_{k=1}^{n_1} \Big[(X_{ki}\hat{X}_{kj})^2 - \mathsf{E}(X_{ki}\hat{X}_{kj})^2 \Big] \Big| \geq 2^{-1}C\sqrt{\frac{\log p}{n}} \Big) \\ &\qquad + np\mathsf{P}\Big(|X_{11}| \geq \tau\sqrt{\log(p+n)} \Big). \end{aligned}$$

Note that

$$np\mathsf{P}\Big(|X_{11}| \geq \tau\sqrt{\log(p+n)}\Big) \leq np(n+p)^{-\tau^2\eta}\mathsf{E}\exp\Big(\eta X_{11}^2\Big) = O(p^{-M}).$$

Let $t = \eta(8\tau^2)^{-1}\sqrt{\log p/n_1}$ and $\hat{Z}_{kij} = (X_{ki}\hat{X}_{kj})^2 - \mathsf{E}(X_{ki}\hat{X}_{kj})^2$. Then we have

$$\begin{aligned}
&\mathsf{P}\Big(\frac{1}{n_1}\sum_{k=1}^{n_1}\Big[(X_{ki}\hat{X}_{kj})^2 - \mathsf{E}(X_{ki}\hat{X}_{kj})^2\Big] \geq C\sqrt{\frac{\log p}{n}}\Big)\\
&\leq \exp\Big(-Ct\sqrt{n_1\log p}\Big)\prod_{k=1}^{n_1}\mathsf{E}\exp\Big(t\hat{Z}_{kij}\Big)\\
&\leq \exp\Big(-Ct\sqrt{n_1\log p}\Big)\prod_{k=1}^{n_1}\Big(1+\mathsf{E}t^2\hat{Z}_{kij}^2\exp\Big(t|\hat{Z}_{kij}|\Big)\Big)\\
&\leq \exp\Big(-Ct\sqrt{n_1\log p}+\sum_{k=1}^{n_1}\mathsf{E}t^2\hat{Z}_{kij}^2\exp\Big(t|\hat{Z}_{kij}|\Big)\Big)\\
&\leq \exp\Big(-C\eta(8\tau^2)^{-1}\log p + c_{\tau,\eta}\log p\Big)\\
&\leq Cp^{-M},
\end{aligned}$$

where $c_{\tau,\eta}$ is a positive constant depending only on τ and η. Similarly, we can show that

$$\mathsf{P}\Big(\frac{1}{n_1}\sum_{k=1}^{n_1}\Big[(X_{ki}\hat{X}_{kj})^2 - \mathsf{E}(X_{ki}\hat{X}_{kj})^2\Big] \leq -C\sqrt{\frac{\log p}{n}}\Big) \leq Cp^{-M}.$$

Thus (5) is proved.

It remains to prove the lemma under (C2*). Define

$$Y_{ij,k} = (X_{ki}X_{kj})^2, \quad \hat{Y}_{ij,k} = Y_{ij,k}I\{|Y_{ij,k}| \leq n/(\log p)^8\}.$$

Then as in (6), we can show that $|\mathsf{E}Y_{ij,k} - \mathsf{E}\hat{Y}_{ij,k}| \leq Cn^{-\gamma_0/4}$. Note that $\varepsilon_n = (\log p)^{-1}$. It follows that

$$\begin{aligned}
&\mathsf{P}\Big(\max_{i,j}\Big|\sum_{k=1}^{n_1}(Y_{ij,k} - \mathsf{E}Y_{ij,k})\Big| \geq \frac{n\varepsilon_n}{\log p}\Big)\\
&\leq \mathsf{P}\Big(\max_{i,j}\Big|\sum_{k=1}^{n_1}(\hat{Y}_{ij,k} - \mathsf{E}\hat{Y}_{ij,k})\Big| \geq 2^{-1}\frac{n\varepsilon_n}{\log p}\Big) + \mathsf{P}\Big(\max_{i,j,k}|Y_{ij,k}| \geq \frac{n}{(\log p)^8}\Big)\\
&\leq Cp^2\exp(-C(\log p)^4) + Cn^{-\epsilon/8}, \qquad (7)
\end{aligned}$$

where the last inequality follows from Bernstein's inequality and (C2*). The lemma is proved. ∎

Proof of Lemma 4. Without loss of generality, we assume that $\boldsymbol{\mu}_1 = 0$ and $\boldsymbol{\mu}_2 = 0$. Set

$$\begin{aligned} Z_{ij,k} &= \frac{n_2}{n_1}(X_{ki}X_{kj} - \sigma_{ij1}), \quad 1 \le k \le n_1, \\ Z_{ij,k} &= -(Y_{ki}Y_{kj} - \sigma_{ij2}), \quad n_1 + 1 \le k \le n_1 + n_2. \end{aligned}$$

By (4), (5) and (7), we obtain for any $M > 0$,

$$\begin{aligned} &\mathsf{P}\Big(\max_{ij}\Big|\frac{1}{n_1}\sum_{k=1}^{n_1} Z_{ij,k}^2 - \frac{n_2^2}{n_1^2}\theta_{ij1}\Big| \ge C\frac{\varepsilon_n}{\log p}\Big) = O(p^{-M} + n^{-\epsilon/8}), \\ &\mathsf{P}\Big(\max_{ij}\Big|\frac{1}{n_2}\sum_{k=n_1+1}^{n_2} Z_{ij,k}^2 - \theta_{ij2}\Big| \ge C\frac{\varepsilon_n}{\log p}\Big) = O(p^{-M} + n^{-\epsilon/8}). \end{aligned} \tag{8}$$

We can write

$$\frac{(\tilde{\sigma}_{ij1} - \tilde{\sigma}_{ij2} - \sigma_{ij1} + \sigma_{ij2})^2}{\theta_{ij1}/n_1 + \theta_{ij2}/n_2} = \frac{\Big(\sum_{k=1}^{n_1+n_2} Z_{ij,k}\Big)^2}{\sum_{k=1}^{n_1+n_2} Z_{ij,k}^2} \times \frac{\sum_{k=1}^{n_1+n_2} Z_{ij,k}^2}{n_2^2\theta_{ij1}/n_1 + n_2\theta_{ij2}}.$$

By the self-normalized large deviation theorem for independent random variables (Theorem 1, Jing, Shao and Wang (2003)), we can get

$$\max_{1\le i\le j\le p} \mathsf{P}\Big(\frac{\Big(\sum_{k=1}^{n_1+n_2} Z_{ij,k}\Big)^2}{\sum_{k=1}^{n_1+n_2} Z_{ij,k}^2} \ge x^2\Big) \le C(1 - \Phi(x))$$

uniformly for $0 \le x \le (8\log p)^{1/2}$. This, together with (8), proves the lemma. ∎

Proof of Lemma 5. When $d = 1$, we have by the tail probabilities of normal distribution,

$$\mathsf{P}\Big(|\boldsymbol{N}_1|_{\min} \ge y_p^{1/2} \pm \epsilon_n(\log p)^{-1/2}\Big) = (1 + o(1))\frac{2p^{-2}}{\sqrt{8\pi}}\exp(-x/2).$$

This implies (31) in Lemma 5. It remains to prove the lemma when $d \ge 2$. Note that for any $(i, j) \in A \setminus A_0$ and $(k, l) \in A \setminus A_0$, we have

$$\mathsf{Cov}(X_iX_j, X_kX_l) = \mathsf{E}X_iX_jX_kX_l + O((\log p)^{-2-2\alpha_0}).$$

Define graph $G_{abcd} = (V_{abcd}, E_{abcd})$, where $V_{abcd} = \{a, b, c, d\}$ is the set of vertices and E_{abcd} is the set of edges. There is an edge between $i \ne j \in \{a, b, c, d\}$ if and only if $|\sigma_{ij}| \ge (\log p)^{-1-\alpha_0}$. We say G_{abcd} is a three vertices graph (3-G) if the number of different vertices in V_{abcd} is 3. Similarly, G_{abcd} is a four vertices graph (4-G) if the number of different vertices in V_{abcd} is 4. A vertex in G_{abcd} is said to be *isolated* if there is no edge connected to it. Note that for any $1 \le m_1 \ne m_2 \le q$, $G_{i_{m_1}j_{m_1}i_{m_2}j_{m_2}}$ is 3-G or 4-G. We say a graph $\mathcal{G} := G_{i_{m_1}j_{m_1}i_{m_2}j_{m_2}}$ satisfy $(\star)$ if

$$\begin{aligned}(\star): \quad & \text{If } \mathcal{G} \text{ is 4-G, then there is at least one isolated vertex in } \mathcal{G};\\ & \quad \text{otherwise } \mathcal{G} \text{ is 3-G and } E_{i_{m_1}j_{m_1}i_{m_2}j_{m_2}} = \emptyset.\end{aligned}$$

For any $G_{i_{m_1}j_{m_1}i_{m_2}j_{m_2}}$ satisfying $(\star)$, by (C3),

$$|\mathsf{E}X_{i_{m_1}}X_{j_{m_1}}X_{i_{m_2}}X_{j_{m_2}}| = O((\log p)^{-1-\alpha_0}), \tag{9}$$

where $O(1)$ is uniformly for $i_{m_1}, j_{m_1}, i_{m_2}, j_{m_2}$. We now define the following set

$$\begin{aligned}
\mathcal{I} &= \{1 \le k_1 < \cdots < k_d \le q\},\\
\mathcal{I}_0 &= \{1 \le k_1 < \cdots < k_d \le q : \text{for some } m_1, m_2 \in \{k_1, \cdots, k_d\} \text{ with } m_1 \ne m_2\\
&\qquad \mathcal{G} := G_{i_{m_1}j_{m_1}i_{m_2}j_{m_2}} \text{ does not satisfy } (\star)\},\\
\mathcal{I}_0^c &= \{1 \le k_1 < \cdots < k_d \le q : \text{for any } m_1, m_2 \in \{k_1, \cdots, k_d\} \text{ and } m_1 \ne m_2,\\
&\qquad \mathcal{G} \text{ satisfies } (\star)\}.
\end{aligned}$$

It is easy to see that $\mathcal{I} = \mathcal{I}_0 \cup \mathcal{I}_0^c$. For any subset S of $\{k_1, \ldots, k_d\}$, we say that S satisfies $(\star\star)$ if

$$(\star\star) \qquad \text{for any } m_1 \ne m_2 \in S,\ G_{i_{m_1}j_{m_1}i_{m_2}j_{m_2}} \text{ satisfies } (\star).$$

For $2 \le l \le d$, let

$$\begin{aligned}
\mathcal{I}_{0l} &= \{1 \le k_1 < \cdots < k_d \le q : \text{the largest cardinality of } S \text{ is } l, \text{ where}\\
&\qquad\qquad S \text{ is any subset of } \{k_1 < \ldots, k_d\} \text{ satisfies } (\star\star).\},\\
\mathcal{I}_{01} &= \{1 \le k_1 < \cdots < k_d \le q : \text{for any } m_1, m_2 \in \{k_1, \cdots, k_d\} \text{ with } m_1 \ne m_2\\
&\qquad \mathcal{G} := G_{i_{m_1}j_{m_1}i_{m_2}j_{m_2}} \text{ does not satisfy } (\star)\}.
\end{aligned}$$

Clearly $\mathcal{I}_0^c = \mathcal{I}_{0d}$ and $\mathcal{I}_0 = \cup_{l=1}^{d-1}\mathcal{I}_{0l}$. We can prove that $\mathrm{Card}(\mathcal{I}_{0l}) \le C_d q^{l+2\gamma(d-l)}$ and $\mathrm{Card}(\mathcal{I}_0^c) = (1+o(1))C_q^d$. We claim that

$$\sum_{\mathcal{I}_0^c} \mathsf{P}\Big(|\boldsymbol{N}_d|_{\min} \ge y_p^{1/2} \pm \varepsilon_n(\log p)^{-1/2}\Big) = (1+o(1))\frac{1}{d!}\Big(\frac{1}{\sqrt{8\pi}}\exp(-\frac{x}{2})\Big)^d \tag{10}$$

and

$$\sum_{\mathcal{I}_0} \mathsf{P}\Big(|\boldsymbol{N}_d|_{\min} \ge y_p^{1/2} \pm \varepsilon_n(\log p)^{-1/2}\Big) = o(1). \tag{11}$$

By (10) and (11), Lemma 5 is proved.

We first prove (11). For $1 \le a \ne b \le q$, define the indicator function

$$
\begin{aligned}
d((i_a, j_a), (i_b, j_b)) &= 1 \quad \text{if } G_{i_a j_a i_b j_b} \text{ does not satisfy } (\star); \\
&= 0 \quad \text{otherwise }.
\end{aligned}
$$

We further divide $\mathcal{I}_{0l}$ as follow. Let $(k_1, \ldots, k_d) \in \mathcal{I}_{0l}$ and let $S_\star \subset (k_1, \ldots, k_d)$ be the largest cardinality subset satisfying $(\star\star)$. (If there are more than two subsets that attain the the largest cardinality, then we can choose any one of them.) Define

$$
\begin{aligned}
\mathcal{I}_{0l1} &= \{(k_1, \ldots, k_d) \in \mathcal{I}_{0l} : \text{there exists an } a \notin S_\star \text{ such that for some } b_1, b_2 \in S_\star \\
&\qquad \text{with } b_1 \ne b_2,\ d((i_a, j_a), (i_{b_1}, j_{b_1})) = 1,\ d((i_a, j_a), (i_{b_2}, j_{b_2})) = 1\}, \\
\mathcal{I}_{0l2} &= \mathcal{I}_{0l} \setminus \mathcal{I}_{0l1}.
\end{aligned}
$$

Note that $\mathcal{I}_{011} = \emptyset$ and $\mathcal{I}_{012} = \mathcal{I}_{01}$. Recall that d is fixed and $l \le d-1$. We can prove that $\mathrm{Card}(\mathcal{I}_{0l1}) \le Cq^{l-1+2\gamma(d-l+1)}$ and $\mathrm{Card}(\mathcal{I}_{0l2}) \le C_d q^{l+2\gamma(d-l)}$. Write $S_\star = (b_1, \ldots, b_l)$ and $x_p = y_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}$. For any $(k_1, \ldots, k_d) \in \mathcal{I}_{0l}$,

$$
\begin{aligned}
\mathsf{P}\Big(|\boldsymbol{N}_d|_{\min} \ge y_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\Big) &\le \mathsf{P}\Big(|N_{b_1}| \ge x_p, \ldots, |N_{b_l}| \ge x_p\Big) \\
&= \frac{1}{(2\pi)^{l/2} |\boldsymbol{U}_l|^{1/2}} \int_{|\boldsymbol{y}|_{\min} \ge x_p} \exp\Big(-\frac{1}{2} \boldsymbol{y}^T \boldsymbol{U}_l^{-1} \boldsymbol{y}\Big) d\boldsymbol{y},
\end{aligned}
$$

where $\boldsymbol{U}_l$ is the covariance matrix of $(N_{b_1}, \ldots, N_{b_l})$. By (9), we have $\|\boldsymbol{U}_l - \boldsymbol{I}_l\|_2 = O((\log p)^{-1-\alpha_0})$. Let $|\boldsymbol{y}|_{\max} = \max_{1 \le i \le l} |y_i|$ for $\boldsymbol{y} = (y_1, \ldots, y_l)$. Then

$$
\begin{aligned}
&\frac{1}{(2\pi)^{l/2} |\boldsymbol{U}_l|^{1/2}} \int_{|\boldsymbol{y}|_{\min} \ge x_p} \exp\Big(-\frac{1}{2} \boldsymbol{y}^T \boldsymbol{U}_l^{-1} \boldsymbol{y}\Big) d\boldsymbol{y} \\
&= \frac{1}{(2\pi)^{l/2} |\boldsymbol{U}_l|^{1/2}} \int_{|\boldsymbol{y}|_{\min} \ge x_p, |\boldsymbol{y}|_{\max} \le (\log p)^{1/2+\alpha_0/4}} \exp\Big(-\frac{1}{2} \boldsymbol{y}^T \boldsymbol{U}_l^{-1} \boldsymbol{y}\Big) d\boldsymbol{y} \\
&\quad + O\Big(\exp\Big(-(\log p)^{1+\alpha_0/2}/4\Big)\Big) \\
&= \frac{1 + O((\log p)^{-\alpha_0/2})}{(2\pi)^{l/2}} \int_{|\boldsymbol{y}|_{\min} \ge x_p, |\boldsymbol{y}|_{\max} \le (\log p)^{1/2+\alpha_0/4}} \exp\Big(-\frac{1}{2} \boldsymbol{y}^T \boldsymbol{y}\Big) d\boldsymbol{y} \\
&\quad + O\Big(\exp\Big(-(\log p)^{1+\alpha_0/2}/4\Big)\Big) \\
&= \frac{1 + O((\log p)^{-\alpha_0/2})}{(2\pi)^{l/2}} \int_{|\boldsymbol{y}|_{\min} \ge x_p} \exp\Big(-\frac{1}{2} \boldsymbol{y}^T \boldsymbol{y}\Big) d\boldsymbol{y} \\
&\quad + O\Big(\exp\Big(-(\log p)^{1+\alpha_0/2}/4\Big)\Big) \\
&= O\Big(p^{-2l}\Big). \qquad (12)
\end{aligned}
$$

This implies that

$$
\sum_{\mathcal{I}_{0l1}} \mathsf{P}\Big(|\boldsymbol{N}_d|_{\min} \ge y_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\Big) \le Cp^{-2+4\gamma(d-l+1)} = o(1). \qquad (13)
$$

For $(k_1,\ldots,k_d)\in\mathcal{I}_{0l2}$, let $\bar a=\min\{a: a\in(k_1,\ldots,k_d), a\notin S_\star\}$. Without loss of generality and for notation briefness, we can assume that $d((i_{\bar a},j_{\bar a}),(i_{b_1},j_{b_1}))=1$. Then we have

$$\begin{aligned}&\sum_{\mathcal{I}_{0l2}}\mathsf{P}\Big(|\boldsymbol{N}_d|_{\min}\ge y_p^{1/2}\pm\epsilon_n(\log p)^{-1/2}\Big)\\&\quad\le\sum_{\mathcal{I}_{0l2}}\mathsf{P}\Big(|N_{\bar a}|\ge x_p,|N_{b_1}|\ge x_p,\ldots,|N_{b_l}|\ge x_p\Big),\end{aligned}$$

Because $(k_1,\ldots,k_d)\in\mathcal{I}_{0l2}$, by (C3), we can show that $\mathsf{Cov}(N_{\bar a},N_{b_j})=O((\log p)^{-1-\alpha_0})$ for $2\le j\le l$. Recall that $S_\star=(b_1,\ldots,b_l)$. We have $\mathsf{Cov}(N_{b_i},N_{b_j})=O((\log p)^{-1-\alpha_0})$ for $1\le i\ne j\le l$. Let $\boldsymbol{V}_l$ be the covariance matrix of $(N_{\bar a},N_{b_1},\ldots,N_{b_l})$. It follows that $\|\boldsymbol{V}_l-\bar{\boldsymbol{V}}_l\|_2=O((\log p)^{-1-\alpha_0})$, where $\bar{\boldsymbol{V}}_l=$diag$(\boldsymbol{D},\boldsymbol{I}_{l-1})$ and $\boldsymbol{D}$ is the covariance matrix of $(N_{\bar a},N_{b_1})$.

We say the graph $G_{i_{\bar a}j_{\bar a}i_{b_1}j_{b_1}}$ is aG-bE if $G_{i_{\bar a}j_{\bar a}i_{b_1}j_{b_1}}$ is a-G and there are b edges in $E_{i_{\bar a}j_{\bar a}i_{b_1}j_{b_1}}$ for $a=3,4$ and $b=0,1,2,3,4$. We further divide $\mathcal{I}_{0l2}$ into two parts:

$$\begin{aligned}\mathcal{I}_{0l2,1}&=\Big\{(k_1,\ldots,k_d)\in\mathcal{I}_{0l2}:\ G_{i_{\bar a}j_{\bar a}i_{b_1}j_{b_1}}\text{ is 3G-1E or 4G-2E}\Big\},\\\mathcal{I}_{0l2,2}&=\Big\{(k_1,\ldots,k_d)\in\mathcal{I}_{0l2}:\ G_{i_{\bar a}j_{\bar a}i_{b_1}j_{b_1}}\text{ is 4G-3E or 4G-4E}\Big\}.\end{aligned}$$

Note that $\mathcal{I}_{0l2}=\mathcal{I}_{0l2,1}\cup\mathcal{I}_{0l2,2}$. We can prove that Card$(\mathcal{I}_{0l2,2})\le Cp^{2l-2+2(d-l-1)\gamma}\times p^{1+3\gamma}$, where C is a constant depending only on the fixed number d. This, together with (12), implies that

$$\begin{aligned}&\sum_{\mathcal{I}_{0l2,2}}\mathsf{P}\Big(|N_{\bar a}|\ge x_p,|N_{b_1}|\ge x_p,\ldots,|N_{b_l}|\ge x_p\Big)\\&\quad\le\sum_{\mathcal{I}_{0l2,2}}\mathsf{P}\Big(|N_{b_1}|\ge x_p,\ldots,|N_{b_l}|\ge x_p\Big)\\&\quad\le Cp^{2l-1+(2d-2l+1)\gamma}\times p^{-2l}=o(1).\end{aligned}\tag{14}$$

For $(k_1,\ldots,k_d)\in\mathcal{I}_{0l2,1}$, by (C3), we have for large p,

$$\begin{aligned}\frac{1}{\theta^{1/2}_{i_{\bar a}j_{\bar a}1}\theta^{1/2}_{i_{b_1}j_{b_1}1}}|\mathsf{E}X_{i_{\bar a}}X_{j_{\bar a}}X_{i_{b_1}}X_{j_{b_1}}|&\le\max\{|\rho_{i_{\bar a}i_{b_1}}\rho_{j_{\bar a}j_{b_1}}|,|\rho_{i_{\bar a}j_{b_1}}\rho_{j_{\bar a}i_{b_1}}|\}+O((\log p)^{-1-\alpha_0})\\&\le r+O((\log p)^{-1-\alpha_0})<(r+1)/2\end{aligned}$$

uniformly for all $\bar a$ and b_1. Recall the covariance matrix $\boldsymbol{V}_l$ of $(N_{\bar a},N_{b_1},\ldots,N_{b_l})$ satisfying $\|\boldsymbol{V}_l-$diag$(\boldsymbol{D},\boldsymbol{I}_{l-1})\ \|_2=O((\log p)^{-1-\alpha_0})$. Using the similar argument as that in (12), we can obtain that

$$\sum_{\mathcal{I}_{0l2,1}}\mathsf{P}\Big(|N_{\bar a}|\ge x_p,|N_{b_1}|\ge x_p,\ldots,|N_{b_l}|\ge x_p\Big)$$

$$\leq C \sum_{\mathcal{I}_{0l2,1}} \Big[\mathsf{P}\Big(|N_{\bar{a}}| \geq x_p, |N_{b_1}| \geq x_p \Big) \times p^{-2l+2} + \exp\Big(-(\log p)^{1+\alpha_0/2}/4 \Big) \Big]$$

$$\leq C \sum_{\mathcal{I}_{0l2,1}} \Big[p^{-2-(2-2r)/(3+r)} \times p^{-2l+2} + \exp\Big(-(\log p)^{1+\alpha_0/2}/4 \Big) \Big], \tag{15}$$

where the last inequality follows from Lemma 2. We can show that $\mathrm{Card}(\mathcal{I}_{0l2,1}) \leq Cp^{2l-2+2(d-l-1)\gamma} \times p^{2+2\gamma}$. Hence it follows from (15) that

$$\sum_{\mathcal{I}_{0l2,1}} \mathsf{P}\Big(|N_{\bar{a}}| \geq x_p, |N_{b_1}| \geq x_p, \ldots, |N_{b_l}| \geq x_p \Big) \leq Cp^{-(2-2r)/(3+r)+2(d-l)\gamma} = o(1). \tag{16}$$

Combining (13), (14) and (16) yields (11).

It remains for us to show that (10). By (9), we have $\|\mathsf{Cov}(\boldsymbol{N}_d) - \boldsymbol{I}_d\|_2 = O((\log p)^{-1-\alpha_0})$ uniformly for $(k_1, \ldots, k_d) \in \mathcal{I}_0^c$. Then by exactly the same proof as that in (12), we can get

$$\mathsf{P}\Big(|\boldsymbol{N}_d|_{\min} \geq y_p^{1/2} \pm \varepsilon_n (\log p)^{-1/2} \Big) = (1+o(1)) \Big(\frac{2}{\sqrt{8\pi}} \exp\Big(-\frac{x}{2} \Big) \Big)^d p^{-2d}$$

uniformly for $(k_1, \ldots, k_d) \in \mathcal{I}_0^c$. This, together with the fact $\mathrm{Card}(\mathcal{I}_0^c) = (1+o(1)) C_q^d$, proves (10). ∎

5 Additional simulation results

In this section we present additional simulation results comparing the numerical performance of the proposed test with that of other tests, particularly in the non-Gaussian setting and the small sample size setting.

5.1 Simulation results for non-Gaussian distributions

For each model, we generate two independent random samples $\{\boldsymbol{X}_k\}_{k=1}^{n_1}$ and $\{\boldsymbol{Y}_l\}_{l=1}^{n_2}$ from multivariate models $\boldsymbol{X}_k = \Gamma_1 \boldsymbol{Z}_k^{(1)}$ and $\boldsymbol{Y}_k = \Gamma_2 \boldsymbol{Z}_k^{(2)}$, with $\Gamma_1 \Gamma_1^{'} = \boldsymbol{\Sigma}_1$, $\Gamma_2 \Gamma_2^{'} = \boldsymbol{\Sigma}_2$. We consider the following four types of distribution of $\boldsymbol{Z}_k^{(i)}$, $i = 1, 2$.

(I) The components of $\boldsymbol{Z}_k^{(i)} = (Z_{k1}, ..., Z_{kp})'$ are i.i.d. Gamma(10,1) random variables.

(II) $\boldsymbol{Z}_k^{(i)} = (Z_{k1}, ..., Z_{kp})'$ are generated by letting each component be an standard normal ($N(0,1)$) variable with probability 0.9 and an standard exponential random variable (exp(1)) with probability 0.1, where the components are independent.

(III) The components of $\boldsymbol{Z}_k^{(i)} = (Z_{k1}, ..., Z_{kp})'$ are i.i.d. t distributed random variables with 12 degrees of freedom.

The simulation results are summarized in Tables 1-3.

n	p	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2					Model 3					Model 4				
		Empirical size																			
60	Φ_α	0.048	0.048	0.049	0.049	0.045	0.049	0.047	0.049	0.051	0.043	0.050	0.046	0.046	0.043	0.048	0.039	0.037	0.044	0.036	0.041
	likelihood	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA
	Schott	0.086	0.073	0.072	0.065	0.063	0.090	0.085	0.076	0.067	0.068	0.095	0.089	0.082	0.079	0.081	0.103	0.097	0.106	0.099	0.093
	Chen	0.070	0.056	0.054	0.049	0.049	0.067	0.058	0.055	0.047	0.048	0.064	0.058	0.055	0.051	0.052	0.100	0.097	0.100	0.098	0.093
	Srivastava	0.062	0.051	0.053	0.046	0.044	0.057	0.049	0.046	0.047	0.043	0.051	0.054	0.048	0.047	0.042	0.058	0.048	0.052	0.048	0.050
100	Φ_α	0.042	0.043	0.043	0.040	0.035	0.043	0.042	0.043	0.039	0.038	0.044	0.044	0.042	0.039	0.034	0.042	0.035	0.034	0.028	0.030
	likelihood	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA
	Schott	0.082	0.073	0.077	0.073	0.074	0.084	0.079	0.076	0.076	0.075	0.092	0.087	0.086	0.083	0.083	0.108	0.105	0.094	0.097	0.094
	Chen	0.067	0.056	0.058	0.055	0.053	0.059	0.055	0.054	0.053	0.053	0.059	0.059	0.058	0.052	0.050	0.101	0.101	0.092	0.095	0.093
	Srivastava	0.056	0.057	0.056	0.048	0.048	0.060	0.053	0.058	0.050	0.047	0.058	0.059	0.056	0.050	0.050	0.053	0.055	0.046	0.050	0.050
		Empirical power																			
60	Φ_α	0.836	0.687	0.428	0.353	0.335	0.831	0.522	0.558	0.433	0.223	0.867	0.720	0.445	0.368	0.346	0.774	0.668	0.432	0.361	0.284
	Schott	0.389	0.214	0.145	0.111	0.101	0.437	0.149	0.151	0.113	0.091	0.442	0.233	0.145	0.117	0.102	0.354	0.184	0.129	0.105	0.094
	Chen	0.300	0.144	0.091	0.064	0.060	0.346	0.090	0.093	0.068	0.055	0.345	0.160	0.089	0.064	0.057	0.277	0.134	0.085	0.067	0.069
	Srivastava	0.286	0.112	0.064	0.053	0.040	0.374	0.080	0.066	0.053	0.040	0.369	0.152	0.062	0.050	0.042	0.161	0.058	0.054	0.048	0.047
100	Φ_α	0.996	0.985	0.899	0.865	0.887	0.996	0.915	0.967	0.930	0.707	0.999	0.989	0.912	0.877	0.898	0.990	0.982	0.899	0.873	0.829
	Schott	0.658	0.330	0.175	0.135	0.118	0.708	0.214	0.195	0.144	0.105	0.728	0.366	0.180	0.137	0.122	0.595	0.291	0.150	0.125	0.099
	Chen	0.566	0.236	0.106	0.080	0.070	0.614	0.141	0.127	0.086	0.057	0.630	0.266	0.112	0.083	0.069	0.506	0.217	0.099	0.084	0.066
	Srivastava	0.576	0.237	0.082	0.059	0.057	0.723	0.127	0.103	0.065	0.043	0.740	0.316	0.085	0.061	0.055	0.251	0.076	0.051	0.051	0.048

Table 1: $\Gamma(10, 1)$ random variables. Model 1-4 Empirical sizes and powers. $\alpha = 0.05$. $n = 60$ and 100. 5000 replications.

n	p	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2					Model 3					Model 4				
		Empirical size																			
60	Φ_α	0.046	0.045	0.047	0.041	0.041	0.043	0.044	0.042	0.039	0.044	0.048	0.043	0.045	0.040	0.038	0.039	0.035	0.035	0.043	0.039
	likelihood	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA
	Schott	0.101	0.101	0.097	0.094	0.089	0.115	0.106	0.102	0.107	0.095	0.131	0.127	0.129	0.119	0.115	0.106	0.100	0.103	0.100	0.100
	Chen	0.066	0.059	0.055	0.054	0.054	0.061	0.056	0.053	0.058	0.051	0.057	0.056	0.057	0.053	0.050	0.093	0.095	0.093	0.101	0.097
	Srivastava	0.061	0.055	0.055	0.051	0.041	0.058	0.053	0.053	0.052	0.041	0.075	0.057	0.058	0.050	0.045	0.046	0.048	0.053	0.049	0.050
100	Φ_α	0.041	0.045	0.040	0.036	0.037	0.038	0.037	0.042	0.033	0.036	0.043	0.034	0.038	0.037	0.034	0.031	0.038	0.033	0.031	0.031
	likelihood	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA
	Schott	0.111	0.101	0.099	0.096	0.095	0.117	0.102	0.103	0.107	0.104	0.144	0.126	0.121	0.120	0.115	0.109	0.107	0.107	0.097	0.100
	Chen	0.060	0.061	0.056	0.054	0.053	0.056	0.058	0.059	0.057	0.046	0.057	0.061	0.053	0.053	0.046	0.098	0.100	0.102	0.094	0.099
	Srivastava	0.053	0.053	0.051	0.055	0.052	0.059	0.060	0.052	0.056	0.052	0.065	0.066	0.053	0.050	0.047	0.051	0.052	0.052	0.048	0.051
		Empirical power																			
60	Φ_α	0.817	0.661	0.418	0.308	0.301	0.799	0.478	0.496	0.360	0.198	0.848	0.690	0.434	0.327	0.315	0.749	0.630	0.420	0.320	0.254
	Schott	0.474	0.283	0.197	0.170	0.153	0.510	0.215	0.215	0.178	0.136	0.522	0.304	0.209	0.171	0.154	0.416	0.243	0.184	0.140	0.129
	Chen	0.295	0.140	0.089	0.070	0.056	0.333	0.100	0.092	0.072	0.047	0.338	0.149	0.091	0.071	0.059	0.264	0.134	0.084	0.068	0.063
	Srivastava	0.299	0.126	0.066	0.056	0.040	0.355	0.083	0.076	0.055	0.040	0.374	0.154	0.073	0.058	0.040	0.171	0.064	0.057	0.048	0.048
100	Φ_α	0.993	0.980	0.874	0.833	0.861	0.994	0.892	0.953	0.906	0.680	0.995	0.988	0.889	0.850	0.872	0.986	0.977	0.871	0.838	0.803
	Schott	0.731	0.419	0.249	0.195	0.166	0.782	0.273	0.273	0.205	0.150	0.785	0.465	0.254	0.204	0.171	0.665	0.356	0.212	0.171	0.129
	Chen	0.561	0.236	0.114	0.085	0.067	0.620	0.129	0.127	0.091	0.063	0.633	0.269	0.113	0.086	0.069	0.498	0.216	0.106	0.085	0.065
	Srivastava	0.570	0.231	0.088	0.063	0.050	0.709	0.122	0.110	0.073	0.051	0.730	0.317	0.092	0.061	0.050	0.253	0.073	0.061	0.055	0.049

Table 2: Mixture random variables. Model 1-4 Empirical sizes and powers. $\alpha = 0.05$. $n = 60$ and 100. 5000 replications.

n	p	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2					Model 3					Model 4				
		Empirical size																			
60	Φ_α	0.048	0.037	0.045	0.043	0.034	0.045	0.041	0.044	0.043	0.035	0.035	0.038	0.039	0.040	0.037	0.038	0.035	0.035	0.036	0.039
	likelihood	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA	1.000	1.000	NA	NA	NA
	Schott	0.092	0.079	0.078	0.071	0.079	0.091	0.083	0.078	0.073	0.080	0.102	0.097	0.093	0.090	0.084	0.105	0.098	0.096	0.103	0.099
	Chen	0.069	0.056	0.057	0.050	0.051	0.063	0.057	0.048	0.043	0.050	0.057	0.052	0.047	0.050	0.048	0.096	0.093	0.092	0.101	0.098
	Srivastava	0.055	0.058	0.055	0.048	0.044	0.056	0.056	0.053	0.048	0.041	0.059	0.056	0.048	0.048	0.040	0.051	0.049	0.046	0.050	0.049
100	Φ_α	0.043	0.038	0.038	0.034	0.038	0.041	0.039	0.039	0.038	0.030	0.042	0.036	0.037	0.031	0.032	0.035	0.032	0.030	0.033	0.033
	likelihood	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA	1.000	1.000	1.000	NA	NA
	Schott	0.089	0.075	0.079	0.080	0.078	0.093	0.081	0.080	0.085	0.079	0.095	0.089	0.089	0.099	0.088	0.100	0.094	0.103	0.094	0.106
	Chen	0.066	0.056	0.054	0.053	0.051	0.064	0.053	0.052	0.053	0.051	0.057	0.050	0.050	0.058	0.049	0.093	0.091	0.101	0.093	0.105
	Srivastava	0.052	0.049	0.053	0.049	0.049	0.055	0.052	0.055	0.051	0.052	0.062	0.056	0.054	0.051	0.048	0.046	0.048	0.052	0.047	0.052
		Empirical power																			
60	Φ_α	0.814	0.661	0.393	0.307	0.279	0.820	0.467	0.506	0.379	0.190	0.847	0.695	0.413	0.334	0.290	0.746	0.638	0.404	0.314	0.231
	Schott	0.421	0.230	0.147	0.121	0.118	0.455	0.166	0.164	0.121	0.105	0.466	0.248	0.149	0.124	0.123	0.368	0.200	0.133	0.118	0.102
	Chen	0.296	0.142	0.079	0.069	0.060	0.322	0.098	0.093	0.066	0.053	0.336	0.157	0.083	0.070	0.062	0.264	0.132	0.082	0.074	0.062
	Srivastava	0.291	0.120	0.056	0.047	0.044	0.365	0.078	0.066	0.052	0.038	0.381	0.149	0.063	0.048	0.045	0.159	0.064	0.045	0.054	0.047
100	Φ_α	0.996	0.985	0.893	0.848	0.863	0.996	0.900	0.958	0.924	0.676	0.997	0.991	0.905	0.867	0.878	0.992	0.980	0.886	0.857	0.813
	Schott	0.679	0.363	0.185	0.144	0.139	0.730	0.228	0.214	0.155	0.109	0.742	0.405	0.187	0.148	0.142	0.615	0.321	0.164	0.128	0.112
	Chen	0.564	0.248	0.100	0.079	0.072	0.614	0.140	0.121	0.087	0.057	0.629	0.279	0.103	0.080	0.074	0.501	0.226	0.097	0.078	0.074
	Srivastava	0.579	0.235	0.084	0.063	0.055	0.723	0.125	0.110	0.064	0.048	0.733	0.320	0.088	0.062	0.059	0.250	0.074	0.055	0.054	0.049

Table 3: $t(12)$ random variables. Model 1-4 Empirical sizes and powers. $\alpha = 0.05$. $n = 60$ and 100. 5000 replications.

5.2 Simulation results for small samples

When the sample size is small, say $n = 30$, as is explained in the discussion section in the paper, the critical value derived from the asymptotic distribution is not sufficiently accurate and modification is thus needed. It is shown in this section that the proposed test with the modified critical value also performs well. As we can see from Section 5 in the paper that Chen and Li (2011)'s test performs similarly as Schott (2007)'s test. Hence to save computation cost, we considered three test statistics in this section, the proposed test, Schott (2007)'s test and Srivastava and Yanagihara (2010)'s test. We consider the same models as in the paper but with $\boldsymbol{D} = \boldsymbol{I}$ for the first three models and $\boldsymbol{O} = \boldsymbol{I}$ in the fourth model under the normal distribution. To evaluate the power of the tests, let $\boldsymbol{U} = (u_{kl})$ be a matrix with $2\lfloor K/2 \rfloor$ random nonzero entries. The locations of the $\lfloor K/2 \rfloor$ nonzero entries are selected randomly from the upper triangle of $\boldsymbol{U}$, each with magnitude 0.9. The other $\lfloor K/2 \rfloor$ nonzero entries in the lower triangle are determined by symmetry.

We consider two settings here. In the first case, the number of nonzero entries of the difference of two covariance matrices stays the same and we choose $K = 32$. In the other case, the number of nonzero entries increases as p grows and we choose $K = \lfloor p/4 \rfloor$. We use the following four pairs of covariance matrices $(\boldsymbol{\Sigma}_1^{(i)}, \boldsymbol{\Sigma}_2^{(i)})$, $i = 1, 2, 3$ and 4, to compare the power of the tests, where $\boldsymbol{\Sigma}_1^{(i)} = \boldsymbol{\Sigma}^{(i)}$ and $\boldsymbol{\Sigma}_2^{(i)} = \boldsymbol{\Sigma}^{(i)} + \boldsymbol{U}$.

The dimension p still varies over the values 50, 100, 200, 400 and 800. Because the sample size is small, we use the "normal cut off" method discussed in Section 6 in the paper to choose the critical value through simulation (20000 replications are used). The nominal significant level for all the tests is set at $\alpha = 0.05$. The actual sizes and powers for the four models, reported in Table 4, are estimated from 5000 replications.

It can be seen from Table 4 that the estimated sizes of our test Φ_α and Srivastava and Yanagihara (2010)'s test are close to the nominal level 0.05 in all the cases. Schott (2007)'s test has nominal level close to 0.05 in the first three models, while having size distortion in the fourth model.

The power results in Table 4 show that the proposed test has much higher power than the other tests in all settings. In the first case, the number of nonzero off-diagonal entries of $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2$ does not change when p grows, so the estimated powers of all tests tend to decrease when the dimension p increases. It can be seen in Table 4 that the powers of Schott (2007) and Srivastava and Yanagihara (2010)'s tests decrease extremely fast as p grows. In the second case, the number of nonzero entries increases as p grows. The proposed test has much higher power than the other tests no matter how p varies.

Simulations for Gamma distribution $\Gamma(10, 1)$ are also carried in this small sample case. Similar phenomena as those in the Gaussian case are observed. The results are summarized in Table 5.

References

[1] Cai, T. and Liu, W.D. (2011), Adaptive thresholding for sparse covariance matrix estimation, *Journal of the American Statistical Association*, 106,672-684.

[2] Chen, S. and Li, J. (2011), Two sample tests for high dimensional covariance matrices. Technical report.

[3] Jing, B.Y., Shao, Q.M. and Wang, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables, *The Annals of probability* 31, 2167-2215.

[4] Schott, J. R. (2007), A test for the equality of covariance matrices when the dimension is large relative to the sample sizes.

[5] Srivastava, M. S. and Yanagihara, H. (2010), Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis*, 101, 1319-1329.

p	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
	Model 1					Model 2					Model 3					Model 4				
	Empirical size																			
Φ_α	0.046	0.051	0.050	0.053	0.056	0.053	0.046	0.050	0.053	0.050	0.049	0.051	0.053	0.049	0.048	0.038	0.050	0.043	0.045	0.049
Schott	0.059	0.057	0.052	0.052	0.049	0.058	0.060	0.055	0.050	0.050	0.055	0.049	0.050	0.045	0.052	0.094	0.093	0.096	0.095	0.096
Srivastava	0.055	0.043	0.043	0.040	0.030	0.048	0.048	0.045	0.036	0.027	0.047	0.044	0.034	0.034	0.023	0.056	0.052	0.053	0.049	0.051
	Empirical power ($K = 32$)																			
Φ_α	0.485	0.307	0.212	0.145	0.094	0.506	0.332	0.210	0.138	0.095	0.640	0.387	0.246	0.135	0.096	0.464	0.381	0.218	0.151	0.126
Schott	0.301	0.152	0.090	0.069	0.058	0.439	0.197	0.112	0.076	0.062	0.681	0.322	0.154	0.085	0.069	0.135	0.103	0.098	0.096	0.097
Srivastava	0.121	0.058	0.045	0.039	0.025	0.057	0.052	0.046	0.037	0.023	0.594	0.165	0.077	0.043	0.023	0.073	0.065	0.055	0.049	0.050
	Empirical power ($K = \lfloor p/4 \rfloor$)																			
Φ_α	0.249	0.246	0.288	0.316	0.354	0.279	0.263	0.286	0.303	0.328	0.353	0.316	0.331	0.299	0.279	0.303	0.268	0.296	0.428	0.487
Schott	0.124	0.120	0.127	0.122	0.118	0.156	0.150	0.158	0.145	0.152	0.244	0.225	0.243	0.233	0.233	0.108	0.099	0.100	0.099	0.098
Srivastava	0.062	0.054	0.049	0.046	0.040	0.054	0.053	0.047	0.036	0.029	0.134	0.120	0.128	0.107	0.075	0.063	0.057	0.056	0.058	0.055

Table 4: $N(0,1)$ random variables. Model 1-4 Empirical sizes and powers. $\alpha = 0.05$. $n = 30$. 5000 replications.

p	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
	Model 1					Model 2					Model 3					Model 4				
	Empirical size																			
Φ_α	0.042	0.043	0.045	0.048	0.048	0.043	0.046	0.042	0.045	0.043	0.044	0.040	0.044	0.039	0.044	0.040	0.043	0.038	0.034	0.043
Schott	0.073	0.068	0.061	0.067	0.066	0.078	0.073	0.066	0.067	0.073	0.079	0.083	0.073	0.075	0.077	0.095	0.100	0.103	0.098	0.093
Srivastava	0.060	0.049	0.042	0.039	0.024	0.057	0.047	0.041	0.037	0.026	0.060	0.052	0.042	0.033	0.024	0.051	0.054	0.052	0.052	0.051
	Empirical power ($K = 32$)																			
Φ_α	0.496	0.329	0.218	0.154	0.122	0.506	0.336	0.220	0.151	0.109	0.623	0.400	0.250	0.134	0.091	0.470	0.397	0.234	0.157	0.141
Schott	0.355	0.181	0.114	0.087	0.080	0.494	0.247	0.130	0.102	0.086	0.745	0.409	0.198	0.118	0.100	0.149	0.113	0.104	0.094	0.093
Srivastava	0.119	0.068	0.050	0.040	0.024	0.063	0.051	0.048	0.037	0.025	0.588	0.159	0.084	0.037	0.023	0.079	0.065	0.056	0.043	0.049
	Empirical power ($K = \lfloor p/4 \rfloor$)																			
Φ_α	0.260	0.262	0.313	0.357	0.417	0.280	0.287	0.306	0.339	0.389	0.342	0.315	0.328	0.311	0.303	0.305	0.264	0.304	0.432	0.492
Schott	0.148	0.146	0.146	0.150	0.154	0.192	0.195	0.188	0.194	0.192	0.304	0.300	0.294	0.298	0.299	0.116	0.107	0.106	0.097	0.096
Srivastava	0.065	0.061	0.053	0.049	0.032	0.060	0.052	0.046	0.039	0.027	0.138	0.117	0.138	0.106	0.076	0.063	0.056	0.060	0.053	0.054

Table 5: $\Gamma(10,1)$ random variables. Model 1-4 Empirical sizes and powers. $\alpha = 0.05$. $n = 30$. 5000 replications.