Applications of Peter Hall's martingale limit theory to estimating and testing high dimensional covariance matrices

Danning Li, Lingzhou Xue and Hui Zou

Jilin University, Pennsylvania State University and University of Minnesota

Supplementary Material

This supplementary file contains the proofs of the main article.

S1. Proofs

S1.1 Proof of Theorem 1.

The martingale central limit theorem in Hall and Heyde (1980) is the key technical tool to prove Theorems 1. Following Theorems 1, 6 and 7 of Li and Zou (2016), we only need to show that

$$\sum_{h=1}^{\log n} \mathbb{P}(\mathrm{SURE}(k_0+h) - \mathrm{SURE}(k_0) < 0) < n^{-(1+\epsilon)}.$$

For all $1 \le h \le \log n$, we know

$$\mathrm{SURE}(k_0+h) - \mathrm{SURE}(k_0) = \sum_{l=1}^{h} \sum_{|i-j|=k_0+l-1} \{(2a_n - \frac{n+1}{n-1})\tilde{\sigma}_{ij}^2 + 2b_n \tilde{\sigma}_{ii}\tilde{\sigma}_{jj}\}.$$

Define $k_l = k_0 + l - 1$ then

$$\begin{aligned}
&\mathbb{P}(\mathrm{SURE}(k_0+h) - \mathrm{SURE}(k_0) < 0) \\
&\leq \sum_{l=1}^{h} \mathbb{P}(\sum_{|i-j|=k_l} \{(2a_n - \frac{n+1}{n-1})\tilde{\sigma}_{ij}^2 + 2b_n \tilde{\sigma}_{ii}\tilde{\sigma}_{jj}\} < 0) \\
&= \sum_{l=1}^{h} \mathbb{P}(\sum_{i=1}^{p-k_l} \{(2a_n - \frac{n+1}{n-1})\tilde{\sigma}_{i(i+k_l)}^2 + 2b_n \tilde{\sigma}_{ii}\tilde{\sigma}_{(i+k_l)(i+k_l)}\} < 0).
\end{aligned}$$

To simplify our proof, assume $M_n^l = (p-k_l)/(2\log n)$ and $2\log n$ as integrate numbers without loss of generality. Let $i_{s,t} = s + 2(t-1)\log n$. Then,

$$\begin{aligned}
&\mathbb{P}(\sum_{i=1}^{p-k_l} \{(2a_n - \frac{n+1}{n-1})\tilde{\sigma}_{i(i+k_l)}^2 + 2b_n \tilde{\sigma}_{ii}\tilde{\sigma}_{(i+k_l)(i+k_l)}\} < 0) \\
&\leq \sum_{s=1}^{2\log n} \mathbb{P}(\sum_{t=1}^{M_n^l} \{(2a_n - \frac{n+1}{n-1})\tilde{\sigma}_{i_{s,t}(i_{s,t}+k_l)}^2 + 2b_n \tilde{\sigma}_{i_{s,t}i_{s,t}}\tilde{\sigma}_{(i_{s,t}+k_l)(i_{s,t}+k_l)}\} < 0) \\
&= \sum_{s=1}^{2\log n} \mathbb{P}(\sum_{t=1}^{M_n^l} \{(2a_n - \frac{n+1}{n-1})\tilde{\sigma}_{i_{s,t}(i_{s,t}+k_l)}^2 + 2b_n \tilde{\sigma}_{i_{s,t}i_{s,t}}\tilde{\sigma}_{(i_{s,t}+k_l)(i_{s,t}+k_l)}\} < 0).
\end{aligned}$$

For any fixed l and s, $Y_t = (2a_n - \frac{n+1}{n-1})\tilde{\sigma}_{i_{s,t}(i_{s,t}+k_l)}^2 + 2b_n \tilde{\sigma}_{i_{s,t}i_{s,t}}\tilde{\sigma}_{(i_{s,t}+k_l)(i_{s,t}+k_l)}$ are *i.i.d.* with mean $\frac{n-1}{n^2}$ and variance $2(2a_n - \frac{n+1}{n-1})^2 \frac{(n+1)(n-1)}{n^4} + 4b_n^2 \frac{(n-1)^2(2n+1)}{n^4} + 4(2a_n - \frac{n+1}{n-1})b_n \frac{(2+3n)(n-1)}{n^4} = O(\frac{1}{n^2})$. Let $H_n = \frac{7\log p}{n}$. Define $Z_t = Y_t I(|Y_t| < H_n)$ and $V_t = Y_t I(|Y_t| \geq H_n)$. So

$$\begin{aligned}
&\mathbb{P}(\sum_{t=1}^{M_n^l} \{(2a_n - \frac{n+1}{n-1})\tilde{\sigma}_{i_{s,t}(i_{s,t}+k_l)}^2 + 2b_n \tilde{\sigma}_{i_{s,t}i_{s,t}}\tilde{\sigma}_{(i_{s,t}+k_l)(i_{s,t}+k_l)}\} < 0) \\
&\leq \mathbb{P}\big(\sum_{t=1}^{M_n^l} (Z_t - EZ_t) < -M_n^l EY_t/2\big) \\
&+\mathbb{P}\big(\sum_{t=1}^{M_n^l} (V_t - EV_t) < -M_n^l EY_t/2\big). \qquad (1)
\end{aligned}$$

Then by Bernstein inequality, we have

$$\begin{aligned}
& \mathbb{P}\Big(\sum_{t=1}^{M_n^l}(Z_t - EZ_t) < -M_n^l EY_t/2\Big) \\
\leq\ & \exp\Big(-\frac{\frac{1}{8}(M_n^l EY_t)^2}{\sum var(Y_t) + \frac{2}{3}H_n M_n^l EY_t}\Big) \\
\leq\ & \exp(-M_n^l/(C\log p)).
\end{aligned}$$

Now we bound the second term in (1).

$$\begin{aligned}
& \mathbb{P}(|Y_t| \geq H_n) \\
\leq\ & \mathbb{P}(|2a_n - \frac{n+1}{n-1}|\tilde{\sigma}^2_{i_{s,t}(i_{s,t}+k_l)} > H_n - 2b_n\tilde{\sigma}_{i_{s,t}i_{s,t}}\tilde{\sigma}_{(i_{s,t}+k_l)(i_{s,t}+k_l)}) \\
\leq\ & \mathbb{P}(|\tilde{\sigma}_{i_{s,t}(i_{s,t}+k_l)}| \geq \sqrt{6\log p/n}) + \mathbb{P}(|\tilde{\sigma}_{i_{s,t}i_{s,t}} - \frac{n-1}{n}| \geq \sqrt{6\log p/n}) \\
=\ & O(p^{-4}).
\end{aligned}$$

So $EV_t = O(\frac{1}{n^2p^4})$ and

$$\mathbb{P}\Big(\sum_{t=1}^{M_n^l}(V_t - EV_t) < -M_n^l EY_t/2\Big) \leq \mathbb{P}(\max_{1\leq t\leq M_n^l}|Y_t| \geq H_n) = O(p^{-3})$$

Now we can conclude that

$$\mathbb{P}(\text{SURE}(k_0+h) - \text{SURE}(k_0) < 0) \leq C(\log n)^2(\exp(-M_n^l/(C\log p)) + p^{-3}) \leq Cn^{-2}.$$

So we show that SURE is consistent.

S1.2 Proof of Theorem 2.

Since $Y_m = 0$, we use Theorem 1 of Hall (1984) to derive the central limit theorem. For ease of notation, we follow the similar notation as in

Hall (1984). Define

$$G_n(x,y) = EH_n(Z_1,x)H_n(Z_1,y). \tag{2}$$

Then

$$\begin{aligned} EG_n(Z_1,Z_2)^2 &= \frac{1}{n^8}\sum_{1\le i_1,i_2,i_3,i_4,i_5,i_6,i_7,i_8\le p} 2^4\omega^{(k_0)}_{i_1i_2}\omega^{(k_0)}_{i_3i_4}\omega^{(k_0)}_{i_5i_6}\omega^{(k_0)}_{i_7i_8} \\ &\quad (\sigma_{i_1i_3}\sigma_{i_2i_4}+\sigma_{i_1i_4}\sigma_{i_2i_3})(\sigma_{i_5i_7}\sigma_{i_6i_8}+\sigma_{i_5i_8}\sigma_{i_6i_7}) \\ &\quad (\sigma_{i_1i_5}\sigma_{i_2i_6}+\sigma_{i_1i_6}\sigma_{i_2i_5})(\sigma_{i_3i_7}\sigma_{i_4i_8}+\sigma_{i_3i_8}\sigma_{i_4i_7}) \\ &\le C\frac{tr(\Sigma^8)}{n^8} \end{aligned}$$

By the definition of $H_n(Z_1,Z_2)$, it is easy to see that,

$$EH_n(Z_1,Z_2)^4 \le C\frac{(tr(\Sigma^2))^4}{n^8}$$

and

$$EH_n(Z_1,Z_2)^2 = \frac{1}{n^4}\sum_{1\le i_1,i_2,i_3,i_4\le p} 4\omega^{(k_0)}_{i_1i_2}\omega^{(k_0)}_{i_3i_4}(\sigma_{i_1i_3}\sigma_{i_2i_4}+\sigma_{i_1i_4}\sigma_{i_2i_3})^2.$$

So we have

$$\begin{aligned} &\mathrm{Var}_n(k_0) \\ =\ & E\frac{n(n-1)}{2}H_n(Z_1,Z_2)^2 \\ =\ & \frac{2(n-1)}{n^3}\sum_{1\le i_1,i_2,i_3,i_4\le p}\omega^{(k_0)}_{i_1i_2}\omega^{(k_0)}_{i_3i_4}(\sigma_{i_1i_3}\sigma_{i_2i_4}+\sigma_{i_1i_4}\sigma_{i_2i_3})^2 \\ \ge\ & \frac{2(n-1)}{n^3}tr(\Sigma^2)(p-2k_0). \end{aligned}$$

It is easy to see the conditions in Hall (1984) are satisfied as follow:

$$\frac{EG_n(Z_1,Z_2)^2+\frac{EH_n(Z_1,Z_2)^4}{n}}{(EH_n(Z_1,Z_2)^2)^2}\to 0.$$

Then $(\mathrm{Var}_n(k_0))^{-1/2}(S_n^2(k_0)-ES_n^2(k_0))\to N(0,1)$. By the convergence rate of the martingale central limit theorem from Haeusler (1988) and the detail proofs in Hall (1984), we have

$$\begin{aligned}&\sup_t|P(\frac{S_n^2(k_0)-ES_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}}\le t)-\Phi(t)|\\ \le\;& C\left(\frac{EG_n(Z_1,Z_2)^2}{(EH_n(Z_1,Z_2)^2)^2}\right)^{2/5}\\ &+C\left(\frac{H_n(Z_1,Z_2)^4}{n(EH_n(Z_1,Z_2)^2)^2}\right)^{1/5}\le Cn^{-1/5}.\end{aligned}$$

S1.3 Proof of Theorem 3.

Under null hypothesis, we have that $S^2=\sum_{1\le l<m\le n}H^2(Z_m,Z_l)$. Then $ES^2=\mathrm{Var}_n(k_0)$ and we have

$$P(|\frac{S^2}{\mathrm{Var}_n(k_0)}-1|>\epsilon)\le \mathrm{Var}(S^2)/(\mathrm{Var}_n(k_0))^2\to 0.$$

It is easy to see that

$$S^2-\mathrm{Var}_n(k_0)=\sum_{1\le l<m\le n}(H_n^2(Z_m,Z_l)-EH_n^2(Z_m,Z_l))$$

is a U statistic. The dominate term of the variance of S^2 is $\frac{n(n-1)^2}{2}Eg(X_1)^2$, where

$$\begin{aligned}g(X_1)\;&=\;\frac{1}{n^4}\sum_{1\le i_1,i_2,i_3,i_4\le p}4\omega_{i_1i_2}^{(k_0)}\omega_{i_3i_4}^{(k_0)}\{(z_{1i_1}z_{1i_2}-\sigma_{i_1i_2})(z_{1i_3}z_{1i_4}-\sigma_{i_3i_4})-\\ &\quad(\sigma_{i_1i_3}\sigma_{i_2i_4}+\sigma_{i_1i_4}\sigma_{i_2i_3})\}(\sigma_{i_1i_3}\sigma_{i_2i_4}+\sigma_{i_1i_4}\sigma_{i_2i_3}).\end{aligned}$$

Then

$$Eg(X_1)^2 = \frac{1}{n^8}\sum_{1\le i_1,\ldots,i_4\le p} 16\omega^{(k_0)}_{i_1i_2}\omega^{(k_0)}_{i_3i_4}\omega^{(k_0)}_{i_5i_6}\omega^{(k_0)}_{i_7i_8}\{(z_{1i_1}z_{1i_2}-\sigma_{i_1i_2})(z_{1i_3}z_{1i_4}-\sigma_{i_3i_4}) - (\sigma_{i_1i_3}\sigma_{i_2i_4}+\sigma_{i_1i_4}\sigma_{i_2i_3})\}\{(z_{1i_5}z_{1i_6}-\sigma_{i_5i_6})(z_{1i_7}z_{1i_8}-\sigma_{i_7i_8}) - (\sigma_{i_5i_7}\sigma_{i_6i_8}+\sigma_{i_5i_8}\sigma_{i_6i_7})\}(\sigma_{i_1i_3}\sigma_{i_2i_4}+\sigma_{i_1i_4}\sigma_{i_2i_3})(\sigma_{i_5i_7}\sigma_{i_6i_8}+\sigma_{i_5i_8}\sigma_{i_6i_7}).$$

So we have $\mathrm{Var}(S^2)\le C\frac{n(n-1)^2}{n^8}[tr(\Sigma^2)]^4$, and combine with $(\mathrm{Var}_n(k_0))^2 \ge \frac{1}{Cn^4}p^4$, then $S^2/\mathrm{Var}_n(k_0)\to 1$ in probability.

S1.4 Proof of Theorem 5.

Define $M_n(k_0)=\max_{|i-j|\ge h} n|\tilde{\sigma}_{ij}|^2$. Also define the marginal distribution functions of $S_n^2(k_0)$ and $M_n(k_0)$ as $P_{S_n}(z) = P\big(\frac{S_n^2(k_0)-ES_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}}\le z\big)$, and $P_{M_n}(y)=P\big(M_n(k_0)-4\log p+\log\log p\le y\big)$. Moreover, we introduce their joint distribution function as

$$P_{S_n,M_n}(z,y) = P\Big(\{\frac{S_n^2(k_0)-ES_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}}\le z\}\cap\{M_n(k_0)-4\log p+\log\log p\ge y\}\Big).$$

Lemma 1 is useful to prove Theorem 5, and its proof is given in the next subsection.

Lemma 1. *Assume the same conditions of Theorem 5. Under* $\mathbf{H}_0$, *for any* z *and* y

$$P_{S_n,M_n}(z,y)\to\Phi(z)\left(1-e^{\frac{-1}{\sqrt{8\pi}}e^{\frac{-y}{2}}}\right). \tag{3}$$

Now, given Lemma 1, from the proof of Theorem 4 in Cai and Jiang (2010) and the definition of Z_i's, we know that $|nL_n^2 - M_n(k_0)|\to 0$ in probability. Combining Theorem 2 and 3, we know $Q_n^2 - S_n^2(k_0) \to 0$ in probability too. Therefore, as long as Lemma 1 is proved, we complete the proof of Theorem 5. ■

S1.5 Proof of Lemma 1.

Define $y_n = 4\log p - \log(\log p) + y$, $W_0 = \{(i,j) : 1 \le i < j \le p, |i-j| \ge k_0\}$ and $W_1 = \{(i,j) : i \in \Gamma_{p,\delta}, |i-j| \ge k_0\} \cup \{(i,j) : j \in \Gamma_{p,\delta}, |i-j| \ge k_0\}$. For easy of notation, we rearrange the distinct indices in any ordering such that $W = \{(i_l, j_l) : 1 \le l \le q = \mathrm{card}(W_0 \setminus W_1), (i_l, j_l) \in W_0 \setminus W_1\}$ and $W' = \{(i_l, j_l) : q < l \le \mathrm{card}(W_0), (i_l, j_l) \in W_1\}$. Define $V_l = (\mathrm{Var}_n(k_0))^{-1/2}\{\hat{\sigma}^2_{i_l j_l} - \sum_{m=1}^{n-1} \frac{(z_{mi_l} z_{mj_l})^2}{n^2}\}$, $q_1 = \mathrm{card}(W_1)$ and $M_n'(k_0) = \max_{1 \le l \le q} |\hat{V}_l|^2$, where $\hat{V}_l = \frac{1}{\sqrt{n}} \sum_{m=1}^{n-1} Y_{ml}$ and

$$Y_{ml} \;=\; z_{mi_k} z_{mj_k} I(|z_{mi_k} z_{mj_k}| \le \tau_n) - E z_{mi_k} z_{mj_k} I(|z_{mi_k} z_{mj_k}| \le \tau_n)$$

with $\tau_n = 8\log(p)$. Then, we have

$$S_n^2(k_0) = \sum_{l=1}^{q+q_1} V_l$$

and

$$|M_n'(k_0) - M_n(k_0)| \to 0$$

in probability. Equivalently, we define the joint distribution as

$$P_{S_n,M_n'}(z,y) = P\big(\{\max_{l=1,\cdots,q}|\hat{V}_l| > \sqrt{y_n}\}\cap\{\sum_{l=1}^{q+q_1}V_l \le z\}\big)$$
$$= P\big(\cup_{l=1}^{q}[\{|\hat{V}_l| > \sqrt{y_n}\}\cap\{\sum_{l=1}^{q+q_1}V_l \le z\}]\big),$$

where we used the fact that $\{\max_{l=1,\cdots,q}|\hat{V}_l| > \sqrt{y_n}\} = \{\cup_{l=1}^{q}[|\hat{V}_l| > \sqrt{y_n}]\}$ in the second equality. Let $B_l = \{|\hat{V}_l| > \sqrt{y_n}\}\cap\{\sum_{l=1}^{q+q_1}V_l \le z\}$. Then, we have $P_{S_n,M_n'}(z,y) = P(\cup_{l=1}^{q}B_l)$. By using Bonferroni inequality, for any fixed even number $d < [q/2]$, we know that

$$\sum_{s=1}^{d}(-1)^{s-1}\sum_{1\le l_1<\ldots<l_s\le q}P(\cap_{t=1}^{s}B_{l_t}) \le P_{S_n,M_n'}(z,y) \le \sum_{s=1}^{d-1}(-1)^{s-1}\sum_{1\le l_1<\ldots<l_s\le q}P(\cap_{t=1}^{s}B_{l_t}) \tag{4}$$

and also that

$$H_d \le P\big(\cup_{l=1}^{q}\{|\hat{V}_l| > \sqrt{y_n}\}\big) \le H_{d-1} \tag{5}$$

where $H_d = \sum_{s=1}^{d}(-1)^{s-1}\sum_{1\le l_1<\ldots<l_s\le q}P(\cap_{t=1}^{s}\{|\hat{V}_l| > \sqrt{y_n}\})$. Let $\Upsilon_n = n^{-1/5}$. We define two index sets $I = \{(i_{l_t}, j_{l_t}), 1 \le t \le d)\}$ and $W_I = \{(i,j), |i-s| < k_0 \,\text{or}\, |j-t| < k_0 \,\text{or}\, |i-s| < k_0 \,\text{or}\, |j-t| < k_0\,, (s,t) \in I \,\text{and}\, (i,j) \in W\}$. The cardinality of W_I is no greater than $2d(p(2k_0-1)-(2k_0-1)2k_0)$. By construction, $\{|\hat{V}_l|, (i_l, j_l) \in I\}$ and $\{V_{l'}, (i_{l'}, j_{l'}) \in (W\cup W')/W_I\}$ are independent. Using the fact that $\sum_{l=1}^{q+q_1}V_l = \sum_{(i_l,j_l)\in W_I}V_l + \sum_{(i_l,j_l)\in W_I}V_l$, we have

$$P\big(\cap_{t=1}^d B_{l_t}\big) \le P\big(\cap_{t=1}^d\{|\hat{V}_l| > \sqrt{y_n}\}\big)P\big(\sum_{(i_l,j_l)\in(W\cup W')\setminus W_I} V_l \le z+\Upsilon_n\big)+P\big(\big|\sum_{(i_l,j_l)\in W_I} V_l\big| \ge \Upsilon_n\big)$$

and

$$P\big(\cap_{t=1}^d B_{l_t}\big) \ge P\big(\cap_{t=1}^d\{|\hat{V}_l| > \sqrt{y_n}\}\big)P\big(\sum_{(i_l,j_l)\in(W\cup W')\setminus W_I} V_l \le z-\Upsilon_n\big)-P\big(\big|\sum_{(i_l,j_l)\in W_I} V_l\big| \ge \Upsilon_n\big).$$

From Thereorem 3, we obtain that

$$|P\big(\sum_{(i_l,j_l)\in(W\cup W')\setminus W_I} V_l \le z \pm \Upsilon_n\big) - P\big(\sum_{l=1}^{q+q_1} V_l \le z\big)| \le C\Upsilon_n.$$

Combining (5) and the above inequalities, we have

$$\begin{aligned}
P\big(\cup_{l=1}^q B_{l_t}\big) &\le H_{d-1}P\big(\sum_{l=1}^{q+q_1} V_l \le z\big) + CH_{d-1}\Upsilon_n + q^d \max_{I\in W} P\big(\big|\sum_{(i_l,j_l)\in W_I} V_l\big| \ge \Upsilon_n\big)\\
&\le P\big(\cup_{l=1}^q [\{\hat{V}_l > \sqrt{y_n}\}\big)P\big(\{\sum_{l=1}^{q+q_1} V_l \le z\}]\big) + |H_d - H_{d-1}|\\
&\quad +C\Upsilon_n + q^d \max_{I\in W} P\big(\big|\sum_{(i_l,j_l)\in W_I} V_l\big| \ge \Upsilon_n\big),
\end{aligned}$$

while we used the triangle inequality and Bonferroni inequality in the second inequality. Following the same idea in Cai et. al. (2013), we define $|a|_{\min} = \min_{1\le i\le d}|a_i|$ for any vector $a \in \mathcal{R}^d$. For any d, we have

$$\begin{aligned}
|H_{d-1} - H_d| &= \sum_{1\le l_1<\ldots<l_d\le q} P(\cap_{s=1}^d\{|\hat{V}_{ls}| \ge \sqrt{y_n}\})\\
&\le \sum_{1\le l_1<\ldots<l_d\le q} P(|\frac{\sum_{m=1}^{n-1}(Y_{ml_s}, 1 \le s \le d)}{\sqrt{n}}|_{\min} \ge \sqrt{y_n})
\end{aligned}$$

For any $1 \le l_1 < \ldots < l_d \le q$, by Theorem 1 in Zaitsev (1987),

$$P(|\frac{\sum_{m=1}^{n-1}(Y_{ml_s}, 1 \le s \le d)}{\sqrt{n}}|_{\min} \ge \sqrt{y_n}) \le P(|\mathrm{N}_d|_{\min} \ge \sqrt{y_n} - \epsilon_n(\log p)^{-1/2}) + C_1 d^{5/2} \exp\big(-\frac{n^{1/2}\epsilon_n}{C_2 d^3 \tau_n (\log(p))^{1/2}}\big)$$

where $\mathrm{N}_d = (N_{l_1}, \ldots, N_{l_d})$ is a normal vector with $E\mathrm{N}_d = 0$ and $cov(\mathrm{N}_d) = cov((Y_{1l_s}, 1 \le s \le d))$. By Lemma 5 in Cai et al (2013), we have

$$\sum_{1 \le l_1 < \ldots < l_d \le q} P(|\mathrm{N}_d|_{\min} \ge \sqrt{y_n} - \epsilon_n(\log p)^{-1/2}) \le \frac{1}{d!}\big(\frac{1}{\sqrt{8\pi}}\exp(-\frac{y}{2})\big)^d(1+o(1)).$$

So we have

$$|H_d - H_{d-1}| \le \frac{1}{d!}\big(\frac{1}{\sqrt{8\pi}}\exp(-\frac{y}{2})\big)^d(1+o(1)), + C_1 q^d d^{5/2} \exp\big(-\frac{n^{1/2}\epsilon_n}{C_2 d^3 \tau_n (\log(p))^{1/2}}\big).$$

We will show the following claim

$$P\big(\big|\sum_{(i_l, j_l) \in W_I} V_l\big| \ge \Upsilon_n\big) \le C e^{-cn^{1/5}}. \tag{6}$$

Take $\epsilon_n = (\log p)^{-1/2}$ and with claim(6), we have

$$P\big(\cup_{l=1}^q B_{l_t}\big) \le P\big(\cup_{l=1}^q [\{\hat{V}_l > y_n\}\big) P\big(\{\sum_{l=1}^{q+q_1} V_l \le z\}]\big) + C\Upsilon_n + C\frac{1}{d!}\big(\frac{1}{\sqrt{8\pi}}\exp(-\frac{y}{2})\big)^d$$

Similarly, we also have

$$P\big(\cup_{l=1}^q B_{l_t}\big) \ge P\big(\cup_{l=1}^q [\{\hat{V}_l > y_n\}\big) P\big(\{b_n(\sum_{l=1}^{q+q_1} V_l - a_n) \le z\}]\big) - C\Upsilon_n - C\frac{1}{d!}\big(\frac{1}{\sqrt{8\pi}}\exp(-\frac{y}{2})\big)^d.$$

Then let $d \to \infty$, for fixed y and z, we have

$$P_{S_n, M_n'}(z, y) \to \Phi(z)\left(1 - e^{\frac{-1}{\sqrt{8\pi}}e^{\frac{-y}{2}}}\right).$$

Since $|M'_n(k_0) - M_n(k_0)| \to 0$ in probability, we obtain the desired result. Now it only remains to prove the claim (6). Let set M contain all the distinct variables with subindex appear in set I and set $Q = \{i, |i - M| \le k_0 - 1\}$, then $L = card(Q)$ and $L \le 2d(2k_0 - 1)$. Without loss of generality, we only consider the first L rows from sample covariance Σ. We can bound

$$\begin{aligned} & P\Big(\sum_{(i_l, j_l) \in W_I} V_l \ge \Upsilon_n \Big) \\ \le & \sum_{i=1}^{L} P\Big(\sum_{1 \le m \le i - k_0 + 1, i + k_0 - 1 \le m \le p} \Big\{ \frac{(\sum_{k=1}^{n-1} z_{ki} z_{km})^2}{n} - \frac{\sum_{k=1}^{n-1} z_{ki}^2 z_{km}^2}{n} \Big\} \ge \frac{\mathrm{Var}_n^{1/2} n \Upsilon_n}{L} \Big) \\ = & \sum_{i=1}^{L} EP^i\Big(\sum_{1 \le m \le i - k_0 + 1, i + k_0 - 1 \le m \le p} \Big\{ \frac{(\sum_{k=1}^{n-1} z_{ki} z_{km})^2}{n} - \frac{\sum_{k=1}^{n-1} z_{ki}^2}{n} \Big\} \ge \frac{\mathrm{Var}_n^{1/2} n \Upsilon_n}{L} \Big). \end{aligned}$$

Without loss of generality, choose $i = 1$ and assume $(p - k_0)/k_0$ is a integer,

$$\sum_{m=k_0}^{p} \Big\{ \frac{(\sum_{k=1}^{n-1} z_{k1} z_{km})^2}{n} - \frac{\sum_{k=1}^{n-1} z_{k1}^2}{n} \Big\}$$

can be rewritten as

$$\sum_{l=0}^{k_0 - 1} \sum_{j=0}^{\frac{p-k_0}{k_0} - 1} \Big\{ \frac{(\sum_{k=1}^{n-1} z_{k1} z_{k(l + k_0 + jk_0)})^2}{n} - \frac{\sum_{k=1}^{n-1} z_{k1}^2}{n} \Big\}$$

Since Σ is banded matrix with bandwidth k_0,

$$\begin{aligned} & P^1\Big(\sum_{m=k_0}^{p} \Big\{ \frac{(\sum_{k=1}^{n-1} z_{k1} z_{km})^2}{n} - \frac{\sum_{k=1}^{n-1} z_{k1}^2}{n} \Big\} \ge \frac{\mathrm{Var}_n^{1/2} n \Upsilon_n}{L} \Big) \\ \le & \sum_{l=0}^{h-1} P^1\Big(\sum_{j=0}^{\frac{p-k_0}{k_0} - 1} \Big\{ \frac{(\sum_{k=1}^{n-1} z_{k1} z_{k(l + k_0 + jk_0)})^2}{n} - \frac{\sum_{k=1}^{n-1} z_{k1}^2}{n} \Big\} \ge \frac{\mathrm{Var}_n^{1/2} n \Upsilon_n}{Lh} \Big) \end{aligned}$$

Set $q_n = n^{1/3}$ and $\mu_n = E^1\big(\frac{\sum_{k=1}^{n-1} z_{k1}z_{k2}}{\sqrt{n}}\big)^2 I\big(\frac{(\sum_{k=1}^{n-1} z_{k1}z_{k2})^2}{n} \leq q_n\big)$. Define

$$\begin{aligned} y_m &= \Big(\frac{\sum_{k=1}^{n-1} z_{k1}z_{km}}{\sqrt{n}}\Big)^2 I\Big(\frac{(\sum_{k=1}^{n-1} z_{k1}z_{km})^2}{n} \leq q_n\Big) - \mu_n \\ z_m &= \Big(\frac{\sum_{k=1}^{n-1} z_{k1}z_{km}}{\sqrt{n}}\Big)^2 I\Big(\frac{(\sum_{k=1}^{n-1} z_{k1}z_{km})^2}{n} > q_n\Big) + \mu_n - \frac{\sum_{k=1}^{n-1} z_{k1}^2}{n} \end{aligned}$$

for all $m \geq 1$. Define

$$T_n^i = \{|\frac{\sum_{k=1}^{n-1} z_{ki}^2}{n} - 1| \leq \epsilon n^{-1/3}, |\frac{\sum_{k=1}^{n-1} z_{ki}^4}{n} - EZ_{ki}^4| \leq \epsilon n^{-1/3}, \max_{1\leq k\leq n-1} |z_{ki}| \leq n^{1/6}\}.$$

Use the inequality $P(U+V \geq u+v) \leq P(U \geq u) + P(V \geq v)$ to obtain

$$\begin{aligned} & P^1\Big(\sum_{j=0}^{\frac{p-k_0}{k_0}-1} \Big\{\frac{(\sum_{k=1}^{n-1} z_{k1}z_{k(k_0+jk_0)})^2}{n} - \frac{\sum_{k=1}^{n-1} z_{k1}^2}{n}\Big\} \geq \frac{\mathrm{Var}_n^{1/2} n\Upsilon_n}{Lk_0}\Big) I_{T_n^1} \\ \leq\ & P^1\Big(\sum_{j=0}^{\frac{p-k_0}{k_0}-1} y_{(k_0+jk_0)} \geq \frac{\mathrm{Var}_n^{1/2} n\Upsilon_n}{2Lk_0}\Big) I_{T_n^1} + P^1\Big(\sum_{j=0}^{\frac{p-k_0}{k_0}-1} z_{(k_0+jk_0)} \geq \frac{\mathrm{Var}_n^{1/2} n\Upsilon_n}{2Lh}\Big) I_{T_n^1} \\ :=\ & A_n + B_n \end{aligned} \tag{7}$$

for any $n \geq 1$. Since $p > n$, then $\frac{\sqrt{2}pn^{-1/5}}{2Cd(2k_0-1)k_0} \leq \frac{\mathrm{Var}_n^{1/2} n\Upsilon_n}{Lk_0} \leq C\frac{\sqrt{2}pn^{-1/5}}{2d(2k_0-1)k_0}$ and $\frac{\sqrt{2}pn^{-1/5}}{2d(2k_0-1)k_0} \gg p^{1-\epsilon}n^{-1/5} = p^{4/5-\epsilon}$, we can bound A_n as follow

$$A_n \tag{8}$$

$$\leq 4\cdot\exp\Big\{-\frac{\big(\frac{\sqrt{2}pn^{-1/5}}{4Cd(2k_0-1)k_0}\big)^2}{(p-1)(3\sum_{k=1}^{n-1} z_{k1}^4 + \sum_{1\leq k\neq l\leq n-1} z_{k1}^2 z_{l1}^2)/n^2 + q_n C\frac{\sqrt{2}pn^{-1/5}}{12d(2k_0-1)k_0}}\Big\} I_{T_n^1}$$

$$\leq 4\cdot\exp\Big\{-\frac{p^{1-\epsilon}n^{-11/15}}{3C}\Big\}. \tag{9}$$

Define $b_2 = z_{k2}I(|z_{k2}| \leq n^{1/6}) - Ez_{k2}I(|x_{k2}| \leq n^{1/6})$ and $b_3 = z_{k2}I(|z_{k2}| >$

$n^{1/6}) - Ez_{k2}I(|z_{k2}| > n^{1/6})$. By Bernstein's inequality, we have

$$\begin{aligned}
& P^1\Big(\frac{(\sum_{k=1}^{n-1} z_{k1}z_{k2})^2}{n} \geq q_n\Big) I_{T_n^1} \\
\leq\ & 2P^1\Big(\sum_{k=1}^{n-1} z_{k1}b_2 \geq n^{2/3}/2\Big) I_{T_n^1} + 2P^1\Big(\sum_{k=1}^{n-1} z_{k1}b_3 \geq n^{2/3}/2\Big) I_{T_n^1} \\
\leq\ & 2\exp\Big\{-\frac{n^{4/3}}{8\sum_{k=1}^{n-1} z_{k1}^2 + 8/3 \max|z_{k1}| n^{5/6}}\Big\} I_{T_n^1} \\
& +2P(\max_{1\leq k\leq n} |x_{k2}| > n^{1/6}) \leq Ce^{-n^{1/3}/C}. \qquad (10)
\end{aligned}$$

Since

$$\begin{aligned}
& |\mu_n - \frac{\sum_{k=1}^{n-1} z_{k1}^2}{n}| I_{T_n^1} \\
\leq\ & C\frac{\sum_{k=1}^{n-1} z_{k1}^2 + \sum_{k_1\neq k_2} z_{k_1 1}z_{k_2 1}}{n} I_{T_n^1} \\
& [P^1\Big(\frac{(\sum_{k=1}^{n-1} z_{k1}z_{k2})^2}{n} \geq q_n\Big) I_{T_n^1}]^{1/2} \leq Ce^{-n^{1/3}/C},
\end{aligned}$$

then we conclude $B_n \leq pP^1\big(\frac{(\sum_{k=1}^{n} x_{k1}x_{k2})^2}{n} \geq q_n\big) I_{T_n^1}$. We have $P((T_n^i)^c) \leq Ce^{-n^{1/3}/C}$ from Li and Xue (2015). By using (7), (8) and (10), we show the claim. ■

S1.6 Proof of Theorem 6.

The first part is a direct conclusion from Theorem 5. It is enough to prove the second part only. To simplify notation, we let $SL_n = Q_n^2 + (nL_n^2 -$

$4\log p + \log\log p) \geq c_\alpha$. It is obvious that

$$
\begin{aligned}
& \inf_{\Sigma\in\mathcal{G}_1\cup\mathcal{G}_2} P(TS=1) \\
= & \inf_{\Sigma\in\mathcal{G}_1\cup\mathcal{G}_2} P(SL_n \geq c_\alpha) \\
\geq & \min(\inf_{\Sigma\in\mathcal{G}_1} P(SL_n > c_\alpha), \inf_{\Sigma\in\mathcal{G}_2} P(SL_n > c_\alpha)).
\end{aligned}
$$

Recall that the threshold c_α is the α upper quantile of $\Phi \star F$. On the one hand, we have the simple probability bound that

$$
\begin{aligned}
\inf_{\Sigma\in\mathcal{G}_1} P(SL_n \geq c_\alpha) \geq & \inf_{\Sigma\in\mathcal{G}_1} P(nL_n - 4\log p + \log\log p \geq \frac{1}{2}\log p + c_\alpha) \\
& - \sup_{\Sigma\in\mathcal{G}_1} P(Q_n^2 \leq -\frac{1}{2}\log p).
\end{aligned}
$$

Let $\mathrm{V}_n(k_0) = var(S_n^2(k_0))$. When relaxing the null hypothesis, using martingale central limit theorem, we still have that $\mathrm{V}_n(k_0)^{-\frac{1}{2}}(S_n^2(k_0) - ES_n^2(k_0))$ converges to $N(0,1)$ as $n \to \infty$. While we also know that $\mathrm{V}_n(k_0) = \mathrm{Var}_n(k_0) + \mathrm{V}'_n(k_0)$, where

$$
\mathrm{V}'_n(k_0) = \frac{4(n-1)(n-2)^2}{n^4} \sum_{1\leq i,j,s,t\leq p} \omega_{ij}^{(k_0)}\omega_{st}^{(k_0)}\sigma_{ij}\sigma_{st}(\sigma_{is}\sigma_{jt} + \sigma_{it}\sigma_{js}).
$$

Since $p \gg n$, we can conclude that $\mathrm{V}'_n(k_0)/\mathrm{Var}_n(k_0) \to 0$. So relax null assumption, we still have $\mathrm{Var}_n(k_0)^{-\frac{1}{2}}(S_n^2(k_0) - ES_n^2(k_0))$ converges to the standard normal distribution as $n \to \infty$. In general

$$
ES^2 = \sum_{m=2}^{n-1}\sum_{l=1}^{m-1} \frac{2(n-1)}{n^3} \sum_{1\leq i_1,i_2,i_3,i_4\leq p} \omega_{i_1i_2}^{(k_0)}\omega_{i_3i_4}^{(k_0)}(\sigma_{i_1i_3}\sigma_{i_2i_4} + \sigma_{i_1i_4}\sigma_{i_2i_3} + \sigma_{i_1i_2}\sigma_{i_3i_4})^2,
$$

It is not hard to show that $S^2/ES^2 \to 1$ in probability. Further we have $C_1 \le ES^2/\mathrm{Var}_n(k_0) \le C_2$, where C_1 and C_2 are constants. We shall show that $\inf_{\Sigma\in\mathcal{G}_1} P(nL_n - 4\log p + \log\log p \ge \frac{1}{2}\log p + c_\alpha) \to 1$ and $\sup_{\Sigma\in\mathcal{G}_1} P(\frac{S_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}} \le -\frac{1}{2\sqrt{C_2}}\log p) \to 0$ as n diverges to infinity. By $ES_n^2(k_0) = \sum_{|i-j|\ge k_0}\sigma_{ij}^2$, we have

$$\begin{aligned} & P(\frac{S_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}} \le -\frac{1}{2\sqrt{C_2}}\log p) \\ \le\ & P\big(\frac{S_n^2(k_0) - ES_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}} \le -\frac{1}{2\sqrt{C_2}}\log p - \frac{ES_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}}\big) \\ \to\ & 0. \end{aligned}$$

In the meantime, we also have

$$\begin{aligned} & \inf_{\Sigma\in\mathcal{G}_1} P(nL_n^2 - 4\log p + \log\log p \ge \frac{1}{2}\log p + c_\alpha) \\ & \ge \inf_{\Sigma\in\mathcal{G}_1} P(\max|\sigma_{ij}| - \max_{ij}|\hat{\sigma}_{ij} - \sigma_{ij}| \ge \sqrt{(\frac{9}{2}\log p - \log\log p)/n}) \\ & \ge 1 - \sup_{\Sigma\in\mathcal{G}_1} p(\max_{ij}|\hat{\sigma}_{ij} - \sigma_{ij}| \ge (C - \frac{9}{2})\sqrt{\log p/n}). \end{aligned}$$

Thus, $\inf_{\Sigma\in\mathcal{G}_1} P(nL_n - 4\log p + \log\log p \ge \frac{1}{2}\log p + c_\alpha) \to 1$, when $n \to \infty$. We immediately obtain that $\inf_{\Sigma\in\mathcal{G}_1} P(SL_n \ge c_\alpha) \to 1$.

On the other hand, we use the simple probability bound again to obtain

$$\begin{aligned} & \inf_{\Sigma\in\mathcal{G}_2} P(SL_n > c_\alpha) \\ & \ge \inf_{\Sigma\in\mathcal{G}_2} P(Q_n^2 \ge 4\log p + c_\alpha) - \sup_{\Sigma\in\mathcal{G}_2} P(nL_n - 4\log p + \log\log p \le -4\log p). \end{aligned}$$

It is obvious that $\sup_{\Sigma\in\mathcal{G}_2} P(nL_n - 4\log p + \log\log p \le -4\log p) = 0$. Moreover, as $n \to \infty$, since $\frac{ES_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}} \gg \log p$

$$\begin{aligned} & \inf_{\Sigma \in \mathcal{G}_2} P\left(\frac{S_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}} \geq \frac{4 \log p + c_\alpha}{\sqrt{C_1}}\right) \\ = \quad & P\left(\frac{S_n^2(k_0) - ES_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}} \geq \frac{4 \log p}{\sqrt{C_1}} - \frac{ES_n^2(k_0)}{\sqrt{\mathrm{Var}_n(k_0)}}\right) \\ & \to 1 \end{aligned}$$

Thus, we obtain that $\inf_{\Sigma \in \mathcal{G}_2} P(SL_n \geq c_\alpha) \to 1$. Now we get the conclusion.

References

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.

Cai, T., Liu, W. and Xia, Y. (2013) , Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings, *JASA 108*, 265-277.

Haeusler, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. *The Annals of Probability 16*, 275-299.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal. 14*, 1-16.

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, Academic Press.

Li, D. and Xue, L. (2015). Joint limiting laws for high-dimensional independence tests. *arXiv:1512.08819.*

Li, D. and Zou, H. (2016). SURE information criteria for large covariance matrix estimation and their asymptotic properties. *IEEE Transactions on Information Theory 62*, 2153-2169.

Zaitsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions. *Probability Theory and Related Fields 74*, 535-566.