Large-Scale Simultaneous Testing of Cross-Covariance Matrices with Applications to PheWAS

Tianxi Cai[1], T. Tony Cai[2], Katherine Liao[3], and Weidong Liu[4]

[1] *Department of Biostatistics, Harvard T.H Chan School of Public Health*

[2] *Department of Statistics, The Wharton School, University of Pennsylvania*

[3] *Brigham and Women's Hospital*

[4] *Department of Mathematics, Institute of Natural Sciences and MOE-LSC, Shanghai Jiao Tong University*

Supplementary Material

S1. Proofs of the main theorems

S1.1 Proof of Theorems 3.1 and 3.2.

Without loss of generality, we assume $\sigma_{ii} = 1$ for $1 \leq i \leq p$. The proof of Theorems 3.1 and 3.2 mainly relies on the distribution of the test statistic $T_i = n(\hat{\boldsymbol{\sigma}}_i)'\hat{\boldsymbol{\Sigma}}_{Zi}^{-1}\hat{\boldsymbol{\sigma}}_i$ and their tail probabilities. To approximate the distribution of T_i, consider

$$T_i^o = n(\tilde{\boldsymbol{\sigma}}_i)'\boldsymbol{\Sigma}_{Zi}^{-1}\tilde{\boldsymbol{\sigma}}_i = \left\|\sqrt{n}\boldsymbol{\Sigma}_{Zi}^{-1/2}\tilde{\boldsymbol{\sigma}}_i\right\|^2 = \left\|n^{-1/2}\sum_{k=1}^{n}\boldsymbol{\xi}_{ki}\right\|^2,$$

where $\tilde{\boldsymbol{\sigma}}_i = n^{-1}\sum_{k=1}^n(\boldsymbol{Z}_{ki} - \boldsymbol{\sigma}_i)$, $\boldsymbol{\xi}_{ki} = \boldsymbol{\Sigma}_{Zi}^{-1/2}(\boldsymbol{Z}_{ki} - \boldsymbol{\sigma}_i)$, $\boldsymbol{Z}_{ki} = (\boldsymbol{Y}_k - \boldsymbol{\mu}_Y)(X_{ki} - \mu_i)$ and $\|\cdot\|$ denotes the Euclidean norm. Define a truncated version of $\boldsymbol{\xi}_{ki}$,

$$\hat{\boldsymbol{\xi}}_{ki} = \boldsymbol{\xi}_{ki} I\{\|\boldsymbol{\xi}_{ki}\| \le \sqrt{n}/(\log p)^4\} - \mathbb{E}\left[\boldsymbol{\xi}_i I\{\|\boldsymbol{\xi}_{ki}\| \le \sqrt{n}/(\log p)^4\}\right].$$

Then, uniformly in $1 \le i \le p$,

$$\mathbb{P}\left\{\left\|n^{-1/2}\sum_{k=1}^n(\boldsymbol{\xi}_{ki} - \hat{\boldsymbol{\xi}}_{ki})\right\| \ge (\log p)^{-2}\right\} \le n\mathbb{P}\left\{\|\xi_{1i}\| \ge \sqrt{n}/(\log p)^4\right\} = O(p^{-1-\epsilon_1})$$

for some $\epsilon_1 > 0$. By Theorem 1 in Zaïtsev (1987), we have for any $\boldsymbol{x} \in R^d$,

$$\mathbb{P}\left(\left\|n^{-1/2}\sum_{k=1}^n \hat{\xi}_{ki} + \boldsymbol{x}\right\| \ge t\right) \le \mathbb{P}\left\{\|\hat{\boldsymbol{W}} + x\| \ge t - (\log p)^{-2}\right\} + c_{1d}e^{-c_{2d}(\log p)^2}$$
$$\text{and} \quad \mathbb{P}\left(\left\|n^{-1/2}\sum_{k=1}^n \hat{\xi}_{ki} + \boldsymbol{x}\right\| \ge t\right) \ge \mathbb{P}\left\{\|\hat{\boldsymbol{W}} + x\| \ge t + (\log p)^{-2}\right\} - c_{1d}e^{-c_{2d}(\log p)^2},$$

uniformly in $t \in R$ and $1 \le i \le p$, where $\hat{\boldsymbol{W}}$ is a d-dimensional normal random vector with mean zero and covariance matrix $\mathsf{Cov}(\hat{\boldsymbol{\xi}}_{ki})$, c_{1d} and c_{2d} are some constants depending only on d. We have $\|\mathsf{Cov}(\hat{\boldsymbol{\xi}}_{ki}) - \boldsymbol{I}\| \le Cn^{-2\beta}$. Then it is easy to show that

$$\mathbb{P}\left\{\|\hat{\boldsymbol{W}} + x\| \ge t - (\log p)^{-2}\right\} \le \mathbb{P}(\|\boldsymbol{W} + x\| \ge t - 2(\log p)^{-2}) + c_{3d}e^{-c_{4d}n^{2\beta}/(\log p)^4}$$
$$\text{and} \quad \mathbb{P}\left\{\|\hat{\boldsymbol{W}} + x\| \ge t + (\log p)^{-2}\right\} \ge \mathbb{P}(\|\boldsymbol{W} + x\| \ge t + 2(\log p)^{-2}) - c_{3d}e^{-c_{4d}n^{2\beta}/(\log p)^4},$$

where $\boldsymbol{W}$ is the standard normal random vector. Hence, for some $\epsilon_1 > 0$,

$$\mathbb{P}(\|n^{-1/2}\sum_{k=1}^n \boldsymbol{\xi}_{ki} + x\| \ge t) \le \mathbb{P}(\|\boldsymbol{W} + x\| \ge t - 2(\log p)^{-2}) + O(p^{-1-\epsilon_1}),$$
$$\mathbb{P}(\|n^{-1/2}\sum_{k=1}^n \boldsymbol{\xi}_{ki} + x\| \ge t) \ge \mathbb{P}(\|\boldsymbol{W} + x\| \ge t + 2(\log p)^{-2}) - O(p^{-1-\epsilon_1}), \quad \text{(S1.1)}$$

where $O(1)$ is uniformly in $t \in R$ and $1 \le i \le p$. This yields that, for any fixed $\delta > 0$,

$$\mathbb{P}\left\{\max_{1\le i\le p}\left\|n^{-1/2}\sum_{k=1}^n \boldsymbol{\xi}_{ki}\right\|^2 \le (2+\delta)\log p\right\} \to 1.$$

Since $\hat{\boldsymbol{\sigma}}_i = \boldsymbol{\sigma}_i + \tilde{\boldsymbol{\sigma}}_i - (\bar{\boldsymbol{Y}} - \boldsymbol{\mu}_Y)(\bar{X}_i - \mu_{X_i})$, we may write

$$T_i^{1/2} = \sqrt{n}\left\|\hat{\boldsymbol{\Sigma}}_{Zi}^{-1/2}\boldsymbol{\sigma}_i - \hat{\boldsymbol{\Sigma}}_{Zi}^{-1/2}(\bar{\boldsymbol{Y}} - \boldsymbol{\mu}_Y)(\bar{X}_i - \mu_{X_i}) + (\hat{\boldsymbol{\Sigma}}_{Zi}^{-1/2} - \boldsymbol{\Sigma}_{Zi}^{-1/2})\tilde{\boldsymbol{\sigma}}_i + \boldsymbol{\Sigma}_{Zi}^{-1/2}\tilde{\boldsymbol{\sigma}}_i\right\|. \quad \text{(S1.2)}$$

By the proof of Lemma 2 in Cai and Liu (2011), we have for some $C > 0$,

$$\mathbb{P}\Big(\max_{1\le i\le p}|\bar{X}_i - \mu_i| \ge C\sqrt{\frac{\log p}{n}}\Big) \to 0, \qquad \mathbb{P}\Big(\|\bar{\boldsymbol{Y}} - \boldsymbol{\mu}_Y\| \ge C\sqrt{\frac{\log p}{n}}\Big) \to 0, \quad \text{(S1.3)}$$

$$\mathbb{P}\Big(\max_{1\le i\le p}|\tilde{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}_i| \ge C\sqrt{\frac{\log p}{n}}\Big) \to 0, \text{ and } \mathbb{P}\Big(\max_{1\le i\le p}\|\hat{\boldsymbol{\Sigma}}_{Zi} - \boldsymbol{\Sigma}_{Zi}\| \ge C\sqrt{\frac{\log p}{n}}\Big) \to 0. \quad \text{(S1.4)}$$

By (S1.2), (S1.3) and (S1.4), we obtain that

$$\mathbb{P}\Big(\max_{1\le i\le p}\Big|T_i^{1/2} - \Big\|\sqrt{n}\boldsymbol{\Sigma}_{Zi}^{-1/2}\tilde{\boldsymbol{\sigma}}_i + \sqrt{n}\hat{\boldsymbol{\Sigma}}_{Zi}^{-1/2}\boldsymbol{\sigma}_i\Big\|\Big| \ge C\sqrt{\frac{(\log p)^2}{n}}\Big) \to 0. \quad \text{(S1.5)}$$

This, together with the above arguments, implies that the following lemma.

Lemma 1. *We have, as $(n, p) \to \infty$,*

$$\max_{i\in\mathcal{H}_0}\left|\frac{\mathbb{P}(T_i \ge t)}{G(t)} - 1\right| \to 0$$

uniformly in $t \in [0, a_p]$.

Next, define

$$\mathcal{H}_1(c) = \{i : \boldsymbol{\sigma}_i'\boldsymbol{\Sigma}_{Zi}^{-1}\boldsymbol{\sigma}_i \ge c\log p/n\} \quad \text{and} \quad \overline{\mathcal{H}_1(c)} = \{i : \boldsymbol{\sigma}_i'\boldsymbol{\Sigma}_{Zi}^{-1}\boldsymbol{\sigma}_i < c\log p/n\}.$$

For $i \in \mathcal{H}_1(10)$, by (S1.1), (S1.4) and (S1.5), $\mathbb{P}(T_i \ge 2\log p) \to 1$ uniformly in i. On the other hand,

$$\mathbb{P}\Big(\max_{i\in\overline{\mathcal{H}_1(10)}}\Big|T_i^{1/2} - \Big\|\sqrt{n}\boldsymbol{\Sigma}_{Zi}^{-1/2}\tilde{\boldsymbol{\sigma}}_i + \sqrt{n}\boldsymbol{\Sigma}_{Zi}^{-1/2}\boldsymbol{\sigma}_i\Big\|\Big| \ge C\sqrt{\frac{(\log p)^2}{n}}\Big) \to 0. \quad \text{(S1.6)}$$

For $i \in \overline{\mathcal{H}_1(10)} \cap \mathcal{H}_1(c)$ for some $c > 2$, uniformly in i we have

$$\mathbb{P}\left\{\|\boldsymbol{W} + \sqrt{n}\boldsymbol{\Sigma}_{Zi}^{-1/2}\boldsymbol{\sigma}_i\| \ge \sqrt{2\log p + 2(\log p)^{-2}}\right\} \to 1.$$

It follows from (S1.1), (S1.5) and (S1.6) that $\mathbb{P}(T_i \geq 2\log p) \to 1$ uniformly in $i \in \mathcal{H}_1(c)$ for any $c > 2$. Thus, whenever $\mathcal{H}_1(c) \neq \emptyset$, we have

$$\frac{\sum_{i\in\mathcal{H}_1(c)} I\{T_i \geq b_p\}}{\mathrm{Card}\{\mathcal{H}_1(c)\}} \to 1, \quad \text{in probability.} \tag{S1.7}$$

If (3.8) holds, then we have $\mathrm{Card}\{\mathcal{H}_1(c)\}\} \geq (1-\varepsilon)\log p$ for any $\varepsilon > 0$. In this case, $\mathbb{P}(\hat{t} \leq b_p) \to 1$.

Now with these distributional properties of T_i, we return to the proof of Theorems 3.1 and 3.2. When $\hat{t}$ in (2.5) exists, by the continuity of $G(t)$ and the monotonicity of the indicator function,

$$G(\hat{t}) = \frac{\alpha \max\{\sum_{1\leq i\leq p} I(T_i \geq \hat{t}), 1\}}{p}$$

and hence

$$\mathrm{FDP} = \alpha \frac{\sum_{i\in\mathcal{H}_0} I(T_i \geq \hat{t})}{pG(\hat{t})}.$$

If $\hat{t}$ in (2.5) does not exist, then $\{FDP \geq \varepsilon\} \subseteq \{\max_{i\in\mathcal{H}_0} T_i \geq a_p\}$. Note that, by (S1.1) and (S1.5),

$$\mathbb{P}(\max_{i\in\mathcal{H}_0} T_i \geq a_p) \leq 2pG(a_p - 3(\log p)^{-1}) + O(p^{-\epsilon_1}) = O((\log p)^{-1/2}).$$

To prove Theorems 3.1 and 3.2, it suffices to show that

$$\sup_{0\leq t\leq b_p} \Big| \frac{\sum_{i\in\mathcal{H}_0} I(T_i \geq t)}{p_0 G(t)} - 1 \Big| \to 0 \quad \text{in probability.}$$

Let $b_p' = b_p + (\log p)^{-2}$. By (S1.5), it is enough to prove that

$$\sup_{0\leq t\leq b_p'} \Big| \frac{\sum_{i\in\mathcal{H}_0} I\{T_i^o \geq t\}}{p_0 G(t)} - 1 \Big| \to 0 \quad \text{in probability.}$$

By the proof of Lemma 6.3 in Liu (2013), we only need to show that the following lemma.

Lemma 2. *We have, for any $\varepsilon > 0$,*

$$\sup_{0\le t\le b_p'} \mathbb{P}\Big(\Big|\frac{\sum_{i\in\mathcal{H}_0}[I\{T_i^o \ge t\} - \mathbb{P}(T_i^o \ge t)]}{p_0 G(t)}\Big| \ge \varepsilon\Big) = o(1) \tag{S1.8}$$

and

$$\int_0^{b_p'} \mathbb{P}\Big(\Big|\frac{\sum_{i\in\mathcal{H}_0}[I\{T_i^o \ge t\} - \mathbb{P}(T_i^o \ge t)]}{p_0 G(t)}\Big| \ge \varepsilon\Big) dt = o(v_p), \tag{S1.9}$$

where $v_p = 1/\log\log p$.

To prove Lemma 2, define

$$\begin{aligned} \mathcal{B}_1 &= \{(i,j) : i \in \mathcal{H}_0, j \in \mathcal{H}_0, (i,j) \in \mathcal{A}(\varepsilon), i \neq j\}, \\ \text{and} \quad \mathcal{B}_2 &= \{(i,j) : i \in \mathcal{H}_0, j \in \mathcal{H}_0, (i,j) \notin \mathcal{A}(\varepsilon), i \neq j\}. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}\Big(\sum_{i\in\mathcal{H}_0}[I\{T_i^o \ge t\} - \mathbb{P}(T_i^o \ge t)]\Big)^2 &= \sum_{(i,j)\in\mathcal{B}_1} \Big[\mathbb{P}(T_i^o \ge t, T_j^o \ge t) - \mathbb{P}(T_i^o \ge t)\mathbb{P}(T_j^o \ge t)\Big] \\ &+ \sum_{(i,j)\in\mathcal{B}_2} \Big[\mathbb{P}(T_i^o \ge t, T_j^o \ge t) - \mathbb{P}(T_i^o \ge t)\mathbb{P}(T_j^o \ge t)\Big] \\ &+ \sum_{i\in\mathcal{H}_0} \Big[\mathbb{P}(T_i^o \ge t) - (\mathbb{P}(T_i^o \ge t))^2\Big]. \end{aligned} \tag{S1.10}$$

For $(i,j) \in \mathcal{B}_2$, we have by Lemma 3 below,

$$\mathbb{P}(T_i^o \ge t, T_j^o \ge t) = (1 + A_n)\mathbb{P}(T_i^o \ge t)\mathbb{P}(T_j^o \ge t) \tag{S1.11}$$

uniformly for $0 \le t \le b_p'$, where $|A_n| \le C(\log p)^{-1-\gamma}$. For $(i,j) \in \mathcal{B}_1$, we have by Lemma 3, for any $\delta > 0$,

$$\mathbb{P}(T_i^o \ge t, T_j^o \ge t) \le C(t+1)^{-1}\exp(-t/(1 + \rho_{ij}^* + \delta)) \tag{S1.12}$$

uniformly in $0 \le t \le b_p^{'}$. Submitting (S1.11) and (S1.12) into (S1.10), we obtain

$$\mathbb{E}\Big(\sum_{i\in\mathcal{H}_0}[I\{T_i^o \ge t\} - \mathbb{P}(T_i^o \ge t)]\Big)^2 \le C(\sum_{(i,j)\in\mathcal{A}(\varepsilon)} e^{-\frac{t}{1+\rho_{ij}^*+\delta}}(1+t)^{-1} + A_n p^2 G^2(t) + pG(t))$$

uniformly in $0 \le t \le b_p^{'}$. Note that, by (C1) and letting δ be sufficiently small,

$$\sum_{(i,j)\in\mathcal{A}(\varepsilon)} \int_0^{b_p^{'}} \exp\Big(\frac{\rho_{ij}^*+\delta}{1+\rho_{ij}^*+\delta}t\Big)dt = o(p^2 v_p).$$

This, together with $\int_0^{b_p^{'}} 1/G(t)dt = O(p(\log p)^{-1/2})$, proves (S1.9). (S1.8) can be proved similarly. This concludes the proof of Theorem 2.

Lemma 3. *(i). We have for any* $\delta > 0$,

$$\mathbb{P}(T_i^o \ge t, T_j^o \ge t) \le C(t+1)^{-1}\exp(-t/(1+\rho_{ij}^*+\delta))$$

uniformly in $0 \le t \le b_p^{'}$ *and* $(i,j) \in \mathcal{B}_1$. *(ii). We have*

$$\mathbb{P}(T_i^o \ge t, T_j^o \ge t) = (1+A_n)\mathbb{P}(T_i^o \ge t)\mathbb{P}(T_j^o \ge t)$$

uniformly in $0 \le t \le b_p^{'}$ *and* $(i,j) \in \mathcal{B}_2$, *where* $|A_n| \le C(\log p)^{-1-\gamma}$ *for some* $\gamma > 0$.

To prove Lemma 3, we need the following lemma which comes from Lemma 6.2 in Liu (2013). Let $\boldsymbol{\eta}_k = (\eta_{k1}, \eta_{k2})'$ are independent and identically distributed 2-dimensional random vectors with mean zero.

Lemma 4. *Suppose that* $p \le cn^r$ *and* $\mathbb{E}\|\boldsymbol{\eta}_1\|^{2br+2+\epsilon} < \infty$ *for some fixed* $c > 0$, $r > 0$, $b > 0$ *and* $\epsilon > 0$. *Assume that* $\mathsf{Var}(\eta_{11}) = \mathsf{Var}(\eta_{12}) = 1$ *and* $|\mathsf{Cov}(\eta_{11}, \eta_{12})| \le \delta$ *for some* $0 \le \delta < 1$. *Then we have*

$$\mathbb{P}\Big(|\sum_{k=1}^n \eta_{k1}| \ge t\sqrt{n}, |\sum_{k=1}^n \eta_{k2}| \ge t\sqrt{n}\Big) \le C(t+1)^{-2}\exp(-t^2/(1+|\mathsf{Cov}(\eta_{11},\eta_{12})|))$$

uniformly for $0 \le t \le \sqrt{b\log p}$, *where* C *only depends on* $c, b, r, \epsilon, \delta$.

Proof of Lemma 3. We first prove (i). Let

$$T_i^o(\boldsymbol{\alpha}) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\alpha}^{'} \boldsymbol{\xi}_{ki}.$$

For any $\|\boldsymbol{\alpha}\| = 1$ and $\|\boldsymbol{\beta}\| = 1$, we have, for $i \in \mathcal{H}_0$ and $j \in \mathcal{H}_0$,

$$|\mathsf{Cov}(T_i^o(\boldsymbol{\alpha}), T_j^o(\boldsymbol{\beta}))| \leq \rho_{ij}^*.$$

Let $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q$ satisfying $\|\boldsymbol{\alpha}_j\|_2 = 1$. For any $\|\boldsymbol{\alpha}\| = 1$, there exists $\boldsymbol{\alpha}_j$ such that $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_j\| \leq c_q$, where $c_q \to 0$ as $q \to \infty$ uniformly in $\boldsymbol{\alpha}$ and $1 \leq j \leq q$. Then

$$\left| (T_i^o)^{1/2} - \max_{1 \leq j \leq q} |T_i^o(\boldsymbol{\alpha}_j)| \right| \leq c_q (T_i^o)^{1/2}.$$

So we have

$$(T_i^o)^{1/2} \leq (1 - c_q)^{-1} \max_{1 \leq j \leq q} |T_i^o(\boldsymbol{\alpha}_j)|.$$

It follows from Lemma 4 that

$$\begin{aligned} \mathbb{P}(T_i^o \geq t, T_j^o \geq t) &\leq \sum_{k=1}^{q} \sum_{l=1}^{q} \mathbb{P} \left\{ |T_i^o(\boldsymbol{\alpha}_k)| \geq \sqrt{t}(1 - c_q), |T_i^o(\boldsymbol{\alpha}_l)| \geq \sqrt{t}(1 - c_q) \right\} \\ &\leq C(t+1)^{-1} e^{-t/(1+\rho_{ij}^* + \delta)} \end{aligned}$$

for any $\delta > 0$ by letting q sufficiently large. This proves (i).

To prove (ii), we first note that, using the similar arguments for (S1.1) and Theorem 1 in Zaïtsev (1987),

$$\mathbb{P}(T_i^o \geq t, T_j^o \geq t) \leq \mathbb{P}(\|\hat{\boldsymbol{W}}_1\|^2 \geq t^{'}, \|\hat{\boldsymbol{W}}_2\|^2 \geq t^{'}) + c_{5d} \exp(-c_{6d} n^{2\beta} / (\log p)^4),$$

$$\mathbb{P}(T_i^o \geq t, T_j^o \geq t) \geq \mathbb{P}(\|\hat{\boldsymbol{W}}_1\|^2 \geq t^{''}, \|\hat{\boldsymbol{W}}_2\|^2 \geq t^{''}) - c_{5d} \exp(-c_{6d} n^{2\beta} / (\log p)^4),$$

where $t^{'} = (\sqrt{t} - (\log p)^{-2})^2$ and $t^{''} = (\sqrt{t} + (\log p)^{-2})^2$, and $(\hat{\boldsymbol{W}}_1^{'}, \hat{\boldsymbol{W}}_2^{'})'$ is the normal random vector with mean zero and covariance matrix $\mathsf{Cov}(\hat{\boldsymbol{\xi}}_{kij})$, where $\hat{\boldsymbol{\xi}}_{kij} = (\hat{\boldsymbol{\xi}}_{ki}^{'}, \hat{\boldsymbol{\xi}}_{kj}^{'})'$. We have

$$\|\mathsf{Cov}(\hat{\boldsymbol{\xi}}_{kij}) - \boldsymbol{I}\| \leq C(\log p)^{-2-\varepsilon}$$

for some $\varepsilon > 0$. By the density of multivariate normal random vector,

$$\mathbb{P}(\|\hat{\boldsymbol{W}}_1\|^2 \geq t', \|\hat{\boldsymbol{W}}_2\|^2 \geq t') = (1 + A_n)[G(t)]^2.$$

Similar equation holds when t' is replaced by t''. This proves (ii).

S1.2 Proof of Theorems 3.3 and 3.4.

By the proof of (S1.7), for $t \sim 2(1-\theta)\log p$,

$$\frac{\sum_{i \in \mathcal{H}_1(c)} I(T_i \geq t)}{m_1(c)} \to 1 \tag{S1.13}$$

in probability. Then, for $t \sim 2(1-\theta)\log p$,

$$\frac{\sum_{i=1}^{p} I(T_i \geq t)}{p} \geq (1 + o(1))p^{-1+\theta}$$

with probability tending to one. So $\mathbb{P}\{0 \leq \hat{t} \leq G^{-1}(\alpha p^{-1+\theta}/2)\} \to 1$. Hence, $\widehat{PO} \to 1$ in probability. Theorem 3.3 follows immediately by letting $b_p = G^{-1}(\alpha p^{-1+\theta}/2)$ in the proof of Theorem 3.2.

S1.3 Proof of Proposition 2.

Under the condition in Theorem 3.2 that $m_1(c) \geq \log p$ for some $c > 2$, the proof of Theorem 3.2 shows that $\mathbb{P}(\hat{t}_{BH} \leq b_p) \to 1$. So $\mathbb{P}(\hat{t}_{BH} = \hat{t}) \to 1$. This indicates that $\mathrm{FDR}_{BH} - \mathrm{FDR} = o(1)$ and $\mathrm{FDP}_{BH} - \mathrm{FDP} = o_{\mathbb{P}}(1)$. The proposition is proved.

S1.4 Proof of Proposition 1.

Suppose (3.13) does not hold. So there is a sequence $(n_k, p_k) \to \infty$ as $k \to \infty$ and $p_k \leq n_k^{\beta}$ such that

$$\mathbb{P}(\mathrm{FDP}_{BH} \leq \zeta) \to 1$$

for some $0 < \zeta < 1$ as $k \to \infty$. Let $\hat{p}_0$ denote the number of wrong rejections by BH method. So we have $\mathbb{P}(\hat{p}_0 \leq \zeta|\mathcal{H}_1|/(1-\zeta)) \to 1$ as $k \to \infty$. Write $p^{'} = [\zeta|\mathcal{H}_1|/(1-\zeta)]$ and let $\mathrm{p}_{(1),\mathcal{H}_0} \leq \cdots \leq \mathrm{p}_{(|\mathcal{H}_0|),\mathcal{H}_0}$ be the ordered p-values of $\{\mathrm{p}_i, i \in \mathcal{H}_0\}$. By the definition of BH method, we have

$$\mathbb{P}(\mathrm{p}_{(p'),\mathcal{H}_0} \geq \alpha/p_k) \to 1 \tag{S1.14}$$

as $k \to \infty$.

We next show that, for any $\gamma > 0$,

$$\liminf_{(n,p)\to\infty} \mathbb{P}(\mathrm{p}_{(p'),\mathcal{H}_0} < \gamma/p) > 0. \tag{S1.15}$$

Let $T_{(1),\mathcal{H}_0} \geq \cdots \geq T_{(|\mathcal{H}_0|),\mathcal{H}_0}$ be the ordered values of $\{T_i, i \in \mathcal{H}_0\}$ and $T^o_{(1),\mathcal{H}_0} \geq \cdots \geq T^o_{(|\mathcal{H}_0|),\mathcal{H}_0}$ be the ordered values of $\{T^o_i, i \in \mathcal{H}_0\}$. To prove (S1.15), it is enough to show that

$$\liminf_{(n,p)\to\infty} \mathbb{P}(T_{(p'),\mathcal{H}_0} > G^{-1}(\gamma/p)) > 0. \tag{S1.16}$$

By the proof of Theorem 3.1, we can easily show that

$$\mathbb{P}\Big(\max_{i\in\mathcal{H}_0} |T_i^{1/2} - (T_i^o)^{1/2}| \geq C\sqrt{\frac{(\log p)^2}{n}}\Big) \to 0.$$

Thus, we only need to show that

$$\liminf_{(n,p)\to\infty} \mathbb{P}(T^o_{(p'),\mathcal{H}_0} \geq x_{np}) > 0, \tag{S1.17}$$

where $x_{np} = G^{-1}(\gamma/p) + C\sqrt{\frac{(\log p)^2}{n}}$. Write

$$\mathbb{P}(T^o_{(p'),\mathcal{H}_0} \geq x_{np}) = \mathbb{P}(\cup^*_{i_1<\cdots<i_{p'}}\{T^o_{i_1} \geq x_{np}, \ldots, T^o_{i_{p'}} \geq x_{np}\}),$$

where the notation $\cup^*_{i_1<\cdots<i_{p'}}$ denotes the union of all $i_1 < \cdots < i_{p'}$ with $i_k \in \mathcal{H}_0$, $1 \leq k \leq p^{'}$. Then we have

$$\mathbb{P}(T^o_{(p'),\mathcal{H}_0} \geq x_{np})$$

$$\geq \sum_{i_1<\cdots<i_{p'}}^{*} \mathbb{P}(T^o_{i_1} \geq x_{np}, \ldots, T^o_{i_{p'}} \geq x_{np})$$
$$- \sum_{i_1<\cdots<i_{p'}}^{*} \sum_{\substack{j_1 < \cdots < j_{p'} \\ (j_1, \cdots, j_{p'}) \neq (i_1, \cdots, i_{p'})}}^{*} \mathbb{P}(T^o_{i_1} \geq x_{np}, \ldots, T^o_{i_{p'}} \geq x_{np}, T^o_{j_1} \geq x_{np}, \ldots, T^o_{j_{p'}} \geq x_{np}),$$

where the notation $\sum_{\cdots}^{*}$ denotes the sum for all $i_1 < \cdots < i_{p'}$ with $i_k \in \mathcal{H}_0$, $1 \leq k \leq p^{'}$. By the proof of Lemma 7.2 and the assumptions that $\mathbf{\Sigma}$ is diagonal, it is easy to show that

$$\mathbb{P}(T^o_{i_1} \geq x_{np}, \ldots, T^o_{i_d} \geq x_{np}) = (1 + o(1))[G(x_{np})]^d$$

for any distinct $i_1, \ldots, i_d \in \mathcal{H}_0$ and fixed d. This implies that

$$\begin{aligned} \sum_{i_1<\cdots<i_{p'}}^{*} \mathbb{P}(T^o_{i_1} \geq x_{np}, \ldots, T^o_{i_{p'}} \geq x_{np}) &= (1 + o(1))C^{p^{'}}_{|\mathcal{H}_0|}(\gamma/p_k)^{p^{'}} \\ &= (1 + o(1))\frac{\gamma^{p^{'}}}{p^{'}!}. \end{aligned}$$

Let s denote the number of indices of the set $\{i_1, \ldots, i_{p'}, j_1, \ldots, j_{p'}\}$. Then we have $p^{'}+1 \leq s \leq 2p^{'}$. Note that the number of pairs $(i_1, \ldots, i_{p'}, j_1, \ldots, j_{p'})$ with $|\{i_1, \ldots, i_{p'}, j_1, \ldots, j_{p'}\}| = s$ is no more than $O(C^{p^{'}}_{|\mathcal{H}_0|}|\mathcal{H}_0|^{s-p^{'}}) = O(|\mathcal{H}_0|^s)$. Also, when $|\{i_1, \ldots, i_{p'}, j_1, \ldots, j_{p'}\}| = s$, we have

$$\mathbb{P}(T^o_{i_1} \geq x_{np}, \ldots, T^o_{i_{p'}} \geq x_{np}, T^o_{j_1} \geq x_{np}, \ldots, T^o_{j_{p'}} \geq x_{np}) = (1 + o(1))(\gamma/p_k)^s,$$

which implies that

$$\sum_{i_1<\cdots<i_{p'}}^{*} \sum_{\substack{j_1 < \cdots < j_{p'} \\ (j_1, \cdots, j_{p'}) \neq (i_1, \cdots, i_{p'})}}^{*} \mathbb{P}(T^o_{i_1} \geq x_{np}, \ldots, T^o_{i_{p'}} \geq x_{np}, T^o_{j_1} \geq x_{np}, \ldots, T^o_{j_{p'}} \geq x_{np}) \leq C\gamma^s.$$

Combining the above arguments, we have

$$\mathbb{P}(T^o_{(p'),\mathcal{H}_0} \geq x_{np}) \geq (1 + o(1))\frac{\gamma^{p^{'}}}{p^{'}!} - C\gamma^{p^{'}+1} \geq C\gamma^{p^{'}}$$

for small γ. This implies (S1.15), which is contradict with (S1.14). The proof is complete.

References

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

Liu, W. (2013), Gaussian graphical model estimation with false discovery rate control. *Annals of Statistics*, 41, 2948–2978.

Zaïtsev, 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.