Activation discovery with FDR control: Application to fMRI data

Nankai University, East China Normal University

Supplementary Material

The Supplementary Material includes the finite-sample theory(Appendix S1), the proof of Theorem 1 and Proposition 1 (Appendix S2), some necessary lemmas (Appendix S3), and additional experiments for supporting data analysis (Appendix S4).

S1 Finite-sample Theory

Here we follow the framework of Du et al. (2021) to implement the finite-sample theory on FDR control. To derive the finite-sample upper bound of FDR, we do a small modification on the thresholding rule (cf. Eq. 1.3),

$$\widehat{L} = \widehat{L}(W_1, \ldots, W_p) = \inf_{L>0} \left\{ \frac{1 + \#\{j : W_j \leq -L\}}{\#\{j : W_j \geq L\} \vee 1} \leq \alpha \right\}. \tag{S1.1}$$

We define $\mathbf{W}_{\widehat{\mathcal{S}}} = (W_{j,\mathrm{SLIP}} : j \in \widehat{\mathcal{S}})^\top$, $\mathbf{W}_{\widehat{\mathcal{S}},-j} = \mathbf{W}_{\widehat{\mathcal{S}}} \setminus W_{j,\mathrm{SLIP}}$, and

$$\Delta_j = |\Pr(W_{j,\mathrm{SLIP}} > 0 \mid |W_{j,\mathrm{SLIP}}|, \mathbf{W}_{\widehat{\mathcal{S}},-j}) - 1/2|. \tag{S1.2}$$

The quantity Δ_j measures the violation of the symmetry property of the jth coordinate under general dependence. As shown in Theorem S1 below, the control of FDR boils down to the extent to which the symmetry properties $\{\Delta_j : j \in \mathcal{A}^c \cap \widehat{\mathcal{S}}\}$ are violated.

Theorem S1. *For any $\alpha \in (0,1)$, the FDR of the SLIP procedure with the modified thresholding rule (S1.1) satisfies*

$$\mathrm{FDR} \leq \min_{0 \leq \epsilon < 1/2} \left\{ \alpha \left(\frac{1+2\epsilon}{1-2\epsilon} \right) + \Pr\left(\max_{\mathcal{A}^c \cap \widehat{\mathcal{S}}} \Delta_j \geq \epsilon \right) \right\}.$$

The upper bound of FDR is hard to analyze, because $\{W_{j,\mathrm{SLIP}}\}$'s are dependent, and $\widehat{\boldsymbol{\beta}}_j$ is not guaranteed symmetric about 0 in a finite sample. We consider the ideal setting: (a) $W_{j,\mathrm{SLIP}}$'s are independent of each other for $j \in \widehat{\mathcal{S}}$, (b) $\mathcal{A} \subseteq \widehat{\mathcal{S}}$, and (c) the distribution of random errors is symmetric. We can show $\Delta_j = 0$ for all $j \in \widehat{\mathcal{S}} \setminus \mathcal{A}$ ((a) ensures the independence while (b) and (c) together ensure the symmetry of $\widehat{\boldsymbol{\beta}}_j$), which induces the precise FDR control at the nominal level α. For the general setting, it is hard to derive the expression of Δ_j because of the involved change structure, which we leave for future research.

Proof of Theorem S1. In this proof, we only consider those indices in $\widehat{\mathcal{S}}$, due to

$$\frac{\#\{j \in \mathcal{A}^c : W_j \geq L\}}{1 \vee \#\{j : W_j \geq L\}} = \frac{\#\{j \in \mathcal{A}^c \cap \widehat{\mathcal{S}} : W_j \geq L\}}{1 \vee \#\{j \in \widehat{\mathcal{S}} : W_j \geq L\}},$$

induced by the fact $W_j = 0$ for $j \notin \widehat{\mathcal{S}}$. So we omit the notation $\widehat{\mathcal{S}}$ and slightly abuse the notation $\mathbf{W} := \mathbf{W}_{\widehat{\mathcal{S}}} = (W_1, \ldots, W_q)$ and $\mathbf{W}_{-j} := \mathbf{W}_{\widehat{\mathcal{S}},-j}$, where $q = |\widehat{\mathcal{S}}|$.

Let $\mathbf{W}^j = (W_1, \ldots, W_{j-1}, |W_j|, W_{j+1}, \ldots, W_q)$ and $L_j = T(\mathbf{W}^j)$. We define the term $R(\epsilon) = \frac{\sum_{j\in\mathcal{A}^c} \mathbb{1}(W_j \geq L, \Delta_j \leq \epsilon)}{1+\sum_{j\in\mathcal{A}^c} \mathbb{1}(W_j \leq -L)}$. Then

$$\frac{\sum_{j\in\mathcal{A}^c} \mathbb{1}(W_j \geq L, \Delta_j \leq \epsilon)}{1 \vee \sum_j \mathbb{1}(W_j \geq L)} = \frac{\sum_{j\in\mathcal{A}^c} \mathbb{1}(W_j \geq L, \Delta_j \leq \epsilon)}{1+\sum_{j\in\mathcal{A}^c} \mathbb{1}(W_j \leq -L)} \cdot \frac{1+\sum_{j\in\mathcal{A}^c} \mathbb{1}(W_j \leq -L)}{1 \vee \sum_j \mathbb{1}(W_j \geq L)}$$
$$\leq \alpha R(\epsilon).$$

Next we derive the upper bound of $\mathbb{E}\{R(\epsilon)\}$. We have

$$\begin{aligned}
\mathbb{E}\{R(\epsilon)\} &= \sum_{j\in\mathcal{A}^c} \mathbb{E}\left\{\frac{\mathbb{1}(W_j \geq L, \Delta_j \leq \epsilon)}{1+\sum_{j\in\mathcal{A}^c} \mathbb{1}(W_j \leq -L)}\right\} \\
&\overset{(a)}{=} \sum_{j\in\mathcal{A}^c} \mathbb{E}\left\{\frac{\mathbb{1}(W_j \geq L_j, \Delta_j \leq \epsilon)}{1+\sum_{k\in\mathcal{A}^c, k\neq j} \mathbb{1}(W_k \leq -L_j)}\right\} \\
&= \sum_{j\in\mathcal{A}^c} \mathbb{E}\left[\mathbb{E}\left\{\frac{\mathbb{1}(W_j \geq L_j, \Delta_j \leq \epsilon)}{1+\sum_{k\in\mathcal{A}^c, k\neq j} \mathbb{1}(W_k \leq -L_j)} \mid |W_j|, \mathbf{W}_{-j}\right\}\right] \\
&= \sum_{j\in\mathcal{A}^c} \mathbb{E}\left[\frac{\Pr(W_j \geq 0 \mid |W_j|, \mathbf{W}_{-j}) \mathbb{1}(|W_j| \geq L_j, \Delta_j \leq \epsilon)}{1+\sum_{k\in\mathcal{A}^c, k\neq j} \mathbb{1}(W_k \leq -L_j)}\right] \\
&\overset{(b)}{\leq} \left(\epsilon + \frac{1}{2}\right) \mathbb{E}\left[\sum_{j\in\mathcal{A}^c} \frac{\mathbb{1}(|W_j| \geq L_j, \Delta_j \leq \epsilon)}{1+\sum_{k\in\mathcal{A}^c, k\neq j} \mathbb{1}(W_k \leq -L_j)}\right] \\
&\leq \left(\epsilon + \frac{1}{2}\right) \left(\mathbb{E}\{R(\epsilon)\} + \mathbb{E}\left[\sum_{j\in\mathcal{A}^c} \frac{\mathbb{1}(W_j \leq -L_j)}{1+\sum_{k\in\mathcal{A}^c, k\neq j} \mathbb{1}(W_k \leq -L_j)}\right]\right) \\
&\overset{(c)}{=} \left(\epsilon + \frac{1}{2}\right) \left(\mathbb{E}\{R(\epsilon)\} + \mathbb{E}\left[\sum_{j\in\mathcal{A}^c} \frac{\mathbb{1}(W_j \leq -L_j)}{1+\sum_{k\in\mathcal{A}^c, k\neq j} \mathbb{1}(W_k \leq -L_k)}\right]\right) \\
&\leq \left(\epsilon + \frac{1}{2}\right) \left[\mathbb{E}\{R(\epsilon)\} + 1\right],
\end{aligned}$$

where the step (a) holds due to the positive threshold value L and the definition of L_j, the step (b) holds by the definition of Δ_j, and the step (c) holds due to the fact that if $W_j \leq -\min(L_j, L_k)$ and $W_k \leq -\min(L_j, L_k)$, then $L_j = L_k$. Then

$$\mathbb{E}\{R(\epsilon)\} \leq \frac{1+2\epsilon}{1-2\epsilon},$$

which proves the theorem. □

S2 Proof of Theorem 1 and Proposition 1

Before preceding, we give some notations for the subsequent proofs. We use $a_n \lesssim b_n$ to denote that $a_n \leq Cb_n$ for a universal constant $C > 0$, and write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For notational convenience, let $W_j = W_{j,\mathrm{SLIP}}$ for $j = 1, \ldots, p$ and further rewrite $W_j = U_j^{(1)} U_j^{(2)}$, where $U_j^{(1)} = \xi_j^{(1)}/\widehat{\sigma}_{jj}^{1/2}$ and $U_j^{(2)} = \widehat{\beta}_j/\widehat{V}_{jj}^{1/2}$. Let $G(s) = S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}} \Pr(W_j \geq s \mid \mathcal{Z}_1)$ and $G_-(s) = S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}} \Pr(W_j \leq -s \mid \mathcal{Z}_1)$. Recall that $\mathbf{H} = (\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}$ and $V_{jk} = (\mathbf{e}_j)_{\widehat{\mathcal{S}}}^\top \mathrm{Var}(\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{S}}} \mid \mathcal{Z}_1)(\mathbf{e}_k)_{\widehat{\mathcal{S}}}$. Denote $\mathbf{V} = \mathrm{Var}(\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{S}}} \mid \mathcal{Z}_1) = (\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \hat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} \hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \hat{\mathbf{X}} \boldsymbol{\Xi} \hat{\mathbf{X}}^\top \hat{\mathbf{X}}_{\widehat{\mathcal{S}}} (\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \hat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}$ and $\hat{\mathbf{V}} = (\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}$. Let $\mathbf{D}_i$ is the diagonal matrix with $(\mathbf{D}_i)_{jj} = \sqrt{\widehat{\tau}_j^{(2)}/(T_2 - \widehat{\tau}_j^{(2)})}$ if $i > \widehat{\tau}_j^{(2)}$ and $-\sqrt{(T_2 - \widehat{\tau}_j^{(2)})/\widehat{\tau}_j^{(2)}}$ otherwise. Then $\boldsymbol{\xi}^{(2)} = \frac{1}{\sqrt{T_2}} \sum_{i=1}^{T_2} \mathbf{D}_i \mathbf{Z}_i^{(2)}$. We introduce $\tau_j^{*(k)} = \max\{i : \mathbb{E}(Z_{i+1,j}^{(k)}) \neq \mathbb{E}(Z_{i,j}^{(k)})\}$ for $j \in \mathcal{A}$. Let

$$\beta_j^* = \begin{cases} \sqrt{\frac{\tau_j^{*(2)}(T_2 - \tau_j^{*(2)})}{T_2}}\delta_j^*, & j \in \mathcal{A}; \\ 0, & j \notin \mathcal{A}, \end{cases} \quad \text{and } \xi_j^* = \begin{cases} \sqrt{\frac{\tau_j^{*(1)}(T_1 - \tau_j^{*(1)})}{T_1}}\delta_j^*, & j \in \mathcal{A}; \\ 0, & j \in \mathcal{A}. \end{cases}$$

Lemma S2.1. *Suppose Assumptions 1–6 hold. Then for* $j \in \mathcal{A}_*$, $(\delta_j^*)^2|\widehat{\tau}_j^{(2)} - \tau_j^{*(2)}| \leq \Delta_j$, *where* Δ_j*'s satisfy that*

$$\max_{j \in \mathcal{A}_*} \frac{\Delta_j}{|\delta_j^*|\sqrt{T \log \bar{s}_p}} = O(1).$$

Proof. W.l.o.g., we assume that T is divisible by r, where r is the sample-splitting parameter. Recall that $\widehat{\tau}_j^{(2)} = \lfloor T_2 \widehat{\tau}_j^{(1)}/T_1 \rfloor = \lfloor \widehat{\tau}_j^{(1)}/(r-1) \rfloor$, then $|\widehat{\tau}_j^{(2)} - \widehat{\tau}_j^{(1)}/(r-1)| \leq 1$. Similarly, by the definitions of $\tau_j^{*(1)}$ and $\tau_j^{*(2)}$, we have $|\tau_j^{*(2)} - \tau_j^{*(1)}/(r-1)| \leq 1$. Thus, $|\widehat{\tau}_j^{(2)} - \tau_j^{*(2)}| \leq 2 + |\widehat{\tau}_j^{(1)} - \tau_j^{*(1)}|/(r-1)$. Let $\Delta_j = c_\Delta |\delta_j^*| \sqrt{T \log(T p_{1*})}$ for some constant c_Δ, and it suffices to show $(\delta_j^*)^2 |\widehat{\tau}_j^{(1)} - \tau_j^{*(1)}| \leq \Delta_j$. Here for the sake of clarity, we slightly abuse notations. We write $\widehat{\tau}_j^{(1)}, \tau_j^{*(1)}, T_1$ as $\widehat{\tau}_j, \tau_j^*, T$. Then we verify

$$\Pr\left[\bigcup_{j \in \mathcal{A}_*} \left\{(\delta_j^*)^2 |\widehat{\tau}_j - \tau_j^*| > \Delta_j\right\}\right] \to 0.$$

With Bonferroni inequality, we have

$$\begin{aligned}
&\Pr\left[\bigcup_{j \in \mathcal{A}_*} \left\{(\delta_j^*)^2 |\widehat{\tau}_j - \tau_j^*| > \Delta_j\right\}\right] \\
&\leq p_{1*} \max_{j \in \mathcal{A}_*} \Pr\left\{|\widehat{\tau}_j - \tau_j^*| > \Delta_j/(\delta_j^*)^2\right\} \\
&\leq p_{1*} \max_{j \in \mathcal{A}_*} \sum_{\substack{\tau=1 \\ |\tau - \tau_j^*| > \Delta_j/(\delta_j^*)^2}}^{T} \Pr\left\{|N_j(\tau) - N_j(\tau_j^*)| \geq |M_j(\tau_j^*) - M_j(\tau)|\right\} \\
&\leq p_{1*} \max_{j \in \mathcal{A}_*} \sum_{\substack{\tau=1 \\ |\tau - \tau_j^*| > \Delta_j/(\delta_j^*)^2}}^{T} \Pr\left\{N_j(\tau) - N_j(\tau_j^*) \geq M_j(\tau_j^*) - M_j(\tau)\right\}
\end{aligned}$$

where

$$
M_j(\tau) = \begin{cases} \sqrt{\frac{\tau}{T-\tau}} \frac{T-\tau_j^*}{T} \sqrt{T} \delta_j^*, & \tau < \tau_j^*, \\ \sqrt{\frac{T-\tau}{\tau}} \frac{\tau_j^*}{T} \sqrt{T} \delta_j^*, & \tau \geq \tau_j^*, \end{cases}
$$

$$
N_j(\tau) = \sqrt{\frac{\tau(T-\tau)}{T}} \left(\frac{1}{T-\tau} \sum_{i=\tau+1}^{T} \varepsilon_{ij} - \frac{1}{\tau} \sum_{i=1}^{\tau} \varepsilon_{ij} \right),
$$

and the last inequality holds for sufficiently large T. Hereafter, we focus on the probability $\Pr\left\{N_j(\tau) - N_j(\tau_j^*) \geq M_j(\tau_j^*) - M_j(\tau)\right\}$. W.l.o.g., we assume $\delta_j^* > 0$. For $\tau > \tau_j^* + \Delta_j/(\delta_j^*)^2$, by calculation, we have

$$
M_j(\tau_j^*) - M_j(\tau) = \frac{\tau_j^*}{T} \sqrt{T} \delta_j^* \left(\sqrt{\frac{T-\tau_j^*}{\tau_j^*}} - \sqrt{\frac{T-\tau}{\tau}} \right)
$$

$$
\begin{aligned}
N_j(\tau) - N_j(\tau_j^*) = \frac{1}{\sqrt{T}} \Bigg\{ & \left(\sqrt{\frac{T-\tau_j^*}{\tau_j^*}} - \sqrt{\frac{T-\tau}{\tau}} \right) \sum_{i=1}^{\tau_j^*} \varepsilon_{ij} - \left(\sqrt{\frac{\tau_j^*}{T-\tau_j^*}} + \sqrt{\frac{T-\tau}{\tau}} \right) \sum_{i=\tau_j^*+1}^{\tau} \varepsilon_{ij} \\
& + \left(\sqrt{\frac{\tau}{T-\tau}} - \sqrt{\frac{\tau_j^*}{T-\tau_j^*}} \right) \sum_{i=\tau+1}^{T} \varepsilon_{ij} \Bigg\}
\end{aligned}
$$

Let $\mathcal{F} = \{\max_{i \in \mathcal{Z}_1} \|\varepsilon_i^{(1)}\|_\infty \leq A_T\}$, where $A_T = T^{1/\theta+\zeta} m_{p1}$ for small $\zeta > 0$. It is easy to see $P(\mathcal{F}) \to 1$. With Lemma S3.2, for some uniform constant C for $j \in \mathcal{A}_*$ related to c_τ in Assumption 1 and $c_{\overline{\sigma}}$ in Assumption 5, we have

$$
\begin{aligned}
&\Pr\left\{N_j(\tau) - N_j(\tau_j^*) \geq M_j(\tau_j^*) - M_j(\tau) \mid \mathcal{F}\right\} \\
\leq &\Pr\left[\sqrt{T}\{N_j(\tau) - N_j(\tau_j^*)\} \geq \frac{\tau_j^*}{T} T\delta_j^* \left(\sqrt{\frac{T-\tau_j^*}{\tau_j^*}} - \sqrt{\frac{T-\tau}{\tau}}\right) \mid \mathcal{F}\right] \\
\leq &\Pr\left[\sqrt{T}\{N_j(\tau) - N_j(\tau_j^*)\} \geq \sqrt{C\sigma_{jj}} T\delta_j^* \frac{\Delta_j}{T(\delta_j^*)^2} \mid \mathcal{F}\right] \\
\leq &\ \exp\left\{-\frac{C' \log(Tp_{1*})}{4 + 2/3\sigma_{jj}^{-1/2} \cdot A_T \sqrt{C' \log(Tp_{1*})/T}}\right\},
\end{aligned}
$$

where $C' = Cc_\Delta^2 > 4$ fulfilled by some sufficiently large c_Δ. By Assumption 4, we have $A_T\sqrt{\log(Tp_{1*})/T} \to 0$ as $T, p \to \infty$. Similar arguments give also for $\tau < \tau_j^* - \Delta_j/(\delta_j^*)^2$ and $\delta_j^* < 0$. Thus,

$$
\Pr\left[\bigcup_{j\in\mathcal{A}_*} \left\{(\delta_j^*)^2|\widehat{\tau}_j - \tau_j^*| > \Delta_j\right\}\right] = o(1).
$$

The conclusion above shows the term $\Delta_j = c_\Delta|\delta_j^*|\sqrt{T\log(Tp_{1*})}$ for $j \in \mathcal{A}_*$ are valid, which means $(\delta_j^*)^2|\widehat{\tau}_j - \tau_j^*|$ can be asymptotically bounded by such Δ_j's uniformly for $j \in \mathcal{A}_*$. It is easy to see

$$
\max_{j\in\mathcal{A}_*} \frac{\Delta_j}{|\delta_j^*|\sqrt{T\log \bar{s}_p}} = O(1),
$$

by Assumption 3. Then the lemma follows.

□

Lemma S2.2. *Suppose Assumptions 1–6 hold. Then as $T \to \infty$,*

$$
\Pr\left\{|\widehat{\beta}_j - \beta_j^*|/(\widehat{V}_{jj})^{1/2} > C\sqrt{\log \bar{s}_p} \mid \mathcal{Z}_1\right\} = o(1/\bar{s}_p)
$$

holds uniformly for $j \in \widehat{\mathcal{S}}$, where C is some positive constant.

Proof. Recall that $\beta_j = \mathbb{E}(\xi_j^{(2)} \mid \mathcal{Z}_1)$ for $j \in \mathcal{A}$ and $\beta_j = 0$ otherwise. It can be verified that

$$\beta_j = \begin{cases} \frac{1}{\sqrt{T_2}}(T_2 - \tau_j^{*(2)})\sqrt{\frac{\widehat{\tau}_j^{(2)}}{T_2 - \widehat{\tau}_j^{(2)}}}\delta_j^*, & j \in \mathcal{A},\ \widehat{\tau}_j^{(2)} \leq \tau_j^{*(2)}, \\ \frac{1}{\sqrt{T_2}}\tau_j^{*(2)}\sqrt{\frac{T_2 - \widehat{\tau}_j^{(2)}}{\widehat{\tau}_j^{(2)}}}\delta_j^*, & j \in \mathcal{A},\ \widehat{\tau}_j^{(2)} > \tau_j^{*(2)}, \\ 0, & j \notin \mathcal{A}. \end{cases}$$

Note that $\widehat{\beta}_j - \beta_j^* = \left(\widehat{\beta}_j - \beta_j\right) + (\beta_j - \beta_{\mathcal{A}_*,j}) + \left(\beta_{\mathcal{A}_*,j} - \beta_j^*\right)$, where $\beta_{\mathcal{A}_*,j} = \beta_j$ for $j \in \mathcal{A}_*$ and 0 otherwise. It can be further verified that suppose $j \in \mathcal{A}_*,\ \widehat{\tau}_j^{(2)} \leq \tau_j^{*(2)}$, by Assumption 1, it can be concluded that $|\beta_{\mathcal{A}_*,j} - \beta_j^*| \lesssim T_2^{-1/2}\Delta_j/|\delta_j^*|$, where Δ_j is defined in Lemma S2.1. Similarly, bound can be derived for $j \in \mathcal{A}_*,\ \widehat{\tau}_j^{(2)} > \tau_j^{*(2)}$. And with Assumption 1, we have $|\beta_j - \beta_{\mathcal{A}_*,j}| \asymp \sqrt{T}|\delta_j^*|$ for $j \in \mathcal{A}\backslash\mathcal{A}_*$ and 0 otherwise.

For $j \in \widehat{\mathcal{S}}$, we have

$$\widehat{\beta}_j - \beta_j = \mathbf{e}_j^\top(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}\left\{\boldsymbol{\xi}^{(2)} - \mathbb{E}(\boldsymbol{\xi}^{(2)} \mid \mathcal{Z}_1)\right\} + \mathbf{e}_j^\top(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c}\boldsymbol{\beta}_{\widehat{\mathcal{S}}^c}$$

Consider the decomposition

$$\begin{aligned}(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c}\boldsymbol{\beta}_{\widehat{\mathcal{S}}^c} &= (\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}})^{-1}\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}^c}\boldsymbol{\beta}_{\widehat{\mathcal{S}}^c} + (\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}\left(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c} - \mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}^c}\right)\boldsymbol{\beta}_{\widehat{\mathcal{S}}^c} \\ &+ \left\{(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} - (\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}})^{-1}\right\}\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}^c}\boldsymbol{\beta}_{\widehat{\mathcal{S}}^c}.\end{aligned}$$

By Lemma S3.6, S3.7, and the results from the proof of Lemma S3.1, we have

$$\|\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c} - \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c}\|_2 = \|(\widehat{\boldsymbol{\Xi}}^{-1} - \boldsymbol{\Xi}^{-1})_{\widehat{\mathcal{S}},\widehat{\mathcal{S}}^c}\|_2 \le \|\widehat{\boldsymbol{\Xi}}^{-1} - \boldsymbol{\Xi}^{-1}\|_2 = o_p\left\{\frac{1}{u_p(p_1 - p_{1*})}\right\},$$

$$\|(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} - (\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})^{-1}\|_2 = O_p(1)\|\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}} - \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}}\|_2 = o_p\left\{\frac{1}{u_p(p_1 - p_{1*})}\right\},$$

and

$$\left\|(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c} - \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c})\right\|_2 \le \left\|\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}\right\|_2 \left\|\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c} - \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c}\right\|_2 = o_p\left\{\frac{1}{u_p(p_1 - p_{1*})}\right\},$$

then

$$\max_{i,j}\left\{(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c} - \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c})\right\}_{ij} = o_p\left\{\frac{1}{u_p(p_1 - p_{1*})}\right\}$$

$$\max_{i,j}\left\{(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} - (\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})^{-1}\right\}_{ij} = o_p\left\{\frac{1}{u_p(p_1 - p_{1*})}\right\}.$$

By the Assumption 6, we have

$$\max_{i,j}\left[\left\{(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} - (\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})^{-1}\right\}\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c}\right]_{ij} = o_p\left(\frac{1}{p_1 - p_{1*}}\right),$$

and

$$\left\|(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}\left(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c} - \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c}\right)\boldsymbol{\beta}_{\widehat{\mathcal{S}}^c}\right\|_\infty = o_p(\sqrt{\log \bar{s}_p}/u_p),$$

$$\left\|\left\{(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} - (\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})^{-1}\right\}\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c}\boldsymbol{\beta}_{\widehat{\mathcal{S}}^c}\right\|_\infty = o_p(\sqrt{\log \bar{s}_p}).$$

Thus, with Assumption 6,

$$\|(\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}^c}\boldsymbol{\beta}_{\widehat{\mathcal{S}}^c}\|_\infty = O_p\left(\sqrt{\log \bar{s}_p}\right).$$

So for $j \in \widehat{\mathcal{S}}$, there exists a constant c_β,

$$|\widehat{\beta}_j - \beta_j| \le \left|\frac{1}{\sqrt{T_2}}\sum_{i=1}^{T_2}\mathbf{e}_j^\top \mathbf{H}\mathbf{D}_i\boldsymbol{\varepsilon}_i^{(2)}\right| + c_\beta\sqrt{\log \bar{s}_p}. \tag{S2.1}$$

Let $\mathcal{F} = \{\max_{i=1,\ldots,T_2} \|\mathbf{H}\mathbf{D}_i\boldsymbol{\varepsilon}_i^{(2)}\|_\infty \leq A_T\}$, where $A_T = T_2^{1/\theta+\epsilon_1} m_{p2}$ for some small $\epsilon_1 > 0$. It is easy to see $\Pr(\mathcal{F} \mid \mathcal{Z}_1) \to 1$. Note that $\max_{j=1,\ldots,p} V_{jj}/\widehat{V}_{jj} = 1 + o(1)$. By the fact that

$$|\beta_{\mathcal{A}_*,j} - \beta_j^*| = O(\sqrt{\log \bar{s}_p}), \text{ uniformly for } j \in \mathcal{A}_*,$$

$$|\beta_j - \beta_{\mathcal{A}_*,j}| = O(\sqrt{\log \bar{s}_p}), \text{ uniformly for } j \in \mathcal{A}\backslash\mathcal{A}_*,$$

with Lemma S2.1 and Assumption 2, we have

$$\begin{aligned}
&\Pr\left\{|\widehat{\beta}_j - \beta_j^*|/(\widehat{V}_{jj})^{1/2} > C\sqrt{\log \bar{s}_p} \text{ for some } j \in \widehat{\mathcal{S}} \mid \mathcal{Z}_1, \mathcal{F}\right\} \\
&\leq \bar{s}_p \max_{j\in\widehat{\mathcal{S}}} \Pr\left(|\widehat{\beta}_j - \beta_j| + |\beta_j - \beta_{\mathcal{A}_*,j}| + |\beta_{\mathcal{A}_*,j} - \beta_j^*| > C\sqrt{\widehat{V}_{jj}\log \bar{s}_p} \mid \mathcal{Z}_1, \mathcal{F}\right) \\
&\leq 2\bar{s}_p \max_{j\in\widehat{\mathcal{S}}} \Pr\left(\frac{1}{\sqrt{T_2}}\sum_{i=1}^{T_2} \mathbf{e}_j^\top \mathbf{H}\mathbf{D}_i\boldsymbol{\varepsilon}_i^{(2)} > \sqrt{C'V_{jj}\log \bar{s}_p} \mid \mathcal{Z}_1, \mathcal{F}\right),
\end{aligned}$$

where $C' > 4$ is some constant fulfilled by some sufficiently large C. By Lemma S3.2,

$$\begin{aligned}
&\Pr\left(\frac{1}{\sqrt{T_2}}\sum_{i=1}^{T_2} \mathbf{e}_j^\top \mathbf{H}\mathbf{D}_i\boldsymbol{\varepsilon}_i^{(2)} > \sqrt{C'V_{jj}\log \bar{s}_p} \mid \mathcal{Z}_1, \mathcal{F}\right) \\
&\leq \exp\left(-\frac{C'\log \bar{s}_p}{2 + 2/3\sqrt{C'/V_{jj}} \cdot A_T\sqrt{\log \bar{s}_p/T_2}}\right) \\
&= o(1/\bar{s}_p^2).
\end{aligned}$$

The conclusion follows. □

Lemma S2.3. *Suppose Assumptions 1–6 hold, we have*

$$\frac{G(s)}{G_-(s)} - 1 \to 0$$

uniformly for $0 \leq s \leq s^*$*, where* s^* *satisfies* $G_-(s^*) = \alpha p_{1*}/(2S_0)$.

Proof. Let $M_T = C\sqrt{\log \bar{s}_p}$, where C is specified in Lemma S2.2. Then

$$\begin{aligned}
\frac{G(s)}{G_-(s)} - 1 &= \frac{\sum_{j\in\widehat{\mathcal{S}}\backslash\mathcal{A}} \{\Pr(W_j \geq s \mid \mathcal{Z}_1) - \Pr(W_j \leq -s \mid \mathcal{Z}_1)\}}{S_0 G_-(s)} \\
&= \frac{\sum_{j\in\widehat{\mathcal{S}}\backslash\mathcal{A}} \left\{\Pr(W_j \geq s, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1) - \Pr(W_j \leq -s, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)\right\}}{S_0 G_-(s)} \\
&\quad + \frac{\sum_{j\in\widehat{\mathcal{S}}\backslash\mathcal{A}} \left\{\Pr(W_j \geq s, |U_j^{(2)}| > M_T \mid \mathcal{Z}_1) - \Pr(W_j \leq -s, |U_j^{(2)}| > M_T \mid \mathcal{Z}_1)\right\}}{S_0 G_-(s)} \\
&:= I + II.
\end{aligned}$$

By Lemma S2.2 and the fact that $S_0 G_-(s) \geq \alpha p_{1*}/2$, we have $II = o(1)$.

For $j \in \widehat{\mathcal{S}}\backslash\mathcal{A}$, $U_j^{(2)} = \frac{1}{\sqrt{T_2}} \sum_{i=1}^{T_2} \mathbf{e}_j^\top \mathbf{H}\mathbf{D}_i \boldsymbol{\varepsilon}_i^{(2)} / \sqrt{\widehat{V}_{jj}}$ (cf. Eq. (S2.1)). By Assumption 4, it can be shown that for some constant $0 < C' < 1$ (in Lemma S3.3), $M_T \leq C' \{2\log(1/L_T)\}^{1/2}$, where $L_T \lesssim T_2^{1-\theta/2} m_{p2}^{\theta}$. Hence, Lemma S3.3 entails that

$$\frac{\Pr(W_j \geq s, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)}{\Pr(U_j^{(1)} Z \geq s, |Z| \leq M_T \mid \mathcal{Z}_1)} \to 1, \tag{S2.2}$$

where $Z \sim N(0,1)$ is independent of $U_j^{(1)}$. Similarly,

$$\frac{\Pr(W_j \leq -s, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)}{\Pr(U_j^{(1)} Z \leq -s, |Z| \leq M_T \mid \mathcal{Z}_1)} \to 1.$$

By noting that $\Pr(U_j^{(1)} Z \leq -s, |Z| \leq M_T \mid \mathcal{Z}_1) = \Pr(U_j^{(1)} Z \geq s, |Z| \leq M_T \mid \mathcal{Z}_1)$, we conclude that $I = o(1)$. The conclusion follows. □

Lemma S2.4. *Suppose Assumptions 1–6 hold. Then conditional on* $\mathcal{Z}_1$*,* *we have*

$$\sup_{0\le s\le s^*}\left|\frac{S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}}\mathbb{I}(W_j\ge s)}{G(s)}-1\right|=o_p(1),$$

$$\sup_{0\le s\le s^*}\left|\frac{S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}}\mathbb{I}(W_j\le -s)}{G_-(s)}-1\right|=o_p(1),$$

where s^* is specified as in Lemma S2.3.

Proof. We only prove the first one, and the second one follows similarly.

From the proof of Lemma S2.3, we have

$$G(s)=S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}}\Pr(W_j\ge s,|U_j^{(2)}|\le M_T\mid \mathcal{Z}_1)\left\{1+o(1)\right\}:=\tilde{G}(s)\left\{1+o(1)\right\}.$$

Similarly, conditional $\mathcal{Z}_1$, we have

$$S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}}\mathbb{I}(W_j\ge s)=S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}}\mathbb{I}(W_j\ge s,|U_j^{(2)}|\le M_T)\left\{1+o_p(1)\right\}.$$

Thus, it suffices to show

$$\sup_{0\le s\le s^*}\left|\frac{S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}}\mathbb{I}(W_j\ge s,|U_j^{(2)}|\le M_T)}{\tilde{G}(s)}-1\right|=o_p(1).$$

By Lemma S2.3 and the fact $G(0)+G_-(0)=1$, we have $G(0)=G_-(0)\to 1/2$. Let $z_0=\alpha p_{1*}/(2S_0)<z_1<\cdots<z_{h_T}=1/2$, where $z_i=z_0+\beta_T\exp(i^{\varsigma})/(2S_0)$ and $h_T=[\log\{(S_0-\alpha p_{1*})/\beta_T\}]^{1/\varsigma}$ with $\beta_T/p_{1*}\to 0$ for $0<\varsigma<1$. Introduce $s_0=s^*>s_1>\cdots>s_{h_T}=0$ such that $\tilde{G}(s_i)=z_i$. By the construction, we have $\tilde{G}(s_i)/\tilde{G}(s_{i+1})=1+o(1)$ uniformly for $i=0,1,\ldots,h_T$. Thus, it suffices to show

$$D_T:=\sup_{0\le i\le h_T}\left|\frac{S_0^{-1}\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}}\mathbb{I}(W_j\ge s_i,|U_j^{(2)}|\le M_T)}{\tilde{G}(s_i)}-1\right|=o_p(1).$$

Let $\mathcal{R}_j = \{k \in \widehat{\mathcal{S}} \backslash \mathcal{A} : |R_{jk}| \geq C(\log T_2)^{-2-\nu}\}$, where $C > 0$ is a constant, we have

$$\mathbb{E}\left(\left[\sum_{j\in\widehat{\mathcal{S}}\backslash\mathcal{A}}\left\{\mathbb{I}(W_j \geq s_i, |U_j^{(2)}| \leq M_T) - \Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)\right\}\right]^2 \mid \mathcal{Z}_1\right)$$

$$= \sum_{j\in\widehat{\mathcal{S}}\backslash\mathcal{A}}\sum_{k\in\mathcal{R}_j}\Big\{\Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T, W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1)$$

$$-\Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)\Pr(W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1)\Big\}$$

$$+ \sum_{j\in\widehat{\mathcal{S}}\backslash\mathcal{A}}\sum_{k\notin\mathcal{R}_j}\Big\{\Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T, W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1)$$

$$-\Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)\Pr(W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1)\Big\}$$

$$:= III + IV.$$

By Assumption 5, we have

$$III \leq \sum_{j\in\widehat{\mathcal{S}}\backslash\mathcal{A}}\sum_{k\in\mathcal{R}_j}\Pr\left(W_j \geq s_i, |U_j^{(2)}| \leq M_T, W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1\right)$$

$$\leq \sum_{j\in\widehat{\mathcal{S}}\backslash\mathcal{A}}\sum_{k\in\mathcal{R}_j}\Pr\left(W_j \geq s_i, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1\right)$$

$$\leq r_p S_0 \tilde{G}(s_i).$$

By Lemma 1 in Supplement Material of Cai and Liu (2016), we can get

$$\left|\frac{\Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T, W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1)}{\Pr(U_j^{(1)}Z_1 \geq s_i, |Z_1| \leq M_T \mid \mathcal{Z}_1)\Pr(V_k^{(1)}Z_2 \geq s_i, |Z_2| \leq M_T \mid \mathcal{Z}_1)} - 1\right| \leq C(\log T_2)^{-1-\nu_1},$$

where $Z_1, Z_2 \overset{\text{iid}}{\sim} N(0,1)$ are independent of $U_j^{(1)}, V_k^{(1)}$, $b > 0$ and $\nu_1 =$

$\min\{\nu, 1/2\}$. Then with Eq. (S2.2), we have

$$\left|\frac{\Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T, W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1)}{\Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)\Pr(W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1)} - 1\right| \leq C(\log T_2)^{-1-\nu_1}.$$

Hence,

$$\begin{aligned} IV &\leq C(\log T_2)^{-1-\nu_1} \sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}} \sum_{k\notin\mathcal{R}_j} \Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)\Pr(W_k \geq s_i, |V_k^{(2)}| \leq M_T \mid \mathcal{Z}_1) \\ &\leq C(\log T_2)^{-1-\nu_1} S_0^2 \tilde{G}^2(s_i). \end{aligned}$$

Consequently, for $\forall\varepsilon > 0$, by Markov inequality,

$$\begin{aligned} &\Pr(D_T > \varepsilon \mid \mathcal{Z}_1) \\ \leq &\frac{1}{\varepsilon^2} \sum_{i=0}^{h_T} \frac{\mathbb{E}\left(\left[\sum_{j\in\widehat{\mathcal{S}}\setminus\mathcal{A}} \left\{\mathbb{I}(W_j \geq s_i, |U_j^{(2)}| \leq M_T) - \Pr(W_j \geq s_i, |U_j^{(2)}| \leq M_T \mid \mathcal{Z}_1)\right\}\right]^2 \mid \mathcal{Z}_1\right)}{S_0^2 \tilde{G}^2(s_i)} \\ \leq &\frac{1}{\varepsilon^2} \left\{C(\log T_2)^{-1-\nu_1} h_T + r_p \sum_{i=0}^{h_T} \frac{1}{S_0 \tilde{G}(s_i)}\right\}. \end{aligned}$$

Let ς be arbitrarily close to 1, then $(\log T_2)^{-1-\nu_1} h_T = o(1)$. And

$$\sum_{i=0}^{h_T} \frac{1}{S_0 \tilde{G}(s_i)} = \frac{1}{\alpha p_{1*}} + \sum_{i=1}^{h_T} \frac{1}{\alpha p_{1*} + \beta_T \exp(i^{\varsigma})} \lesssim \beta_T^{-1},$$

where we can choose β_T arbitrarily large as long as $\beta_T/p_{1*} \to 0$. Thus, $D_T = o_p(1)$, fulfilled by Assumption 5. □

Lemma S2.5. *Suppose Assumptions 1–6 hold. Then as $T, p \to \infty$,*

$$\Pr\left(|\xi_j^{(1)} - \xi_j^*|/(\widehat{\sigma}_{jj}^{1/2})^{1/2} > \sqrt{C\log(T_1 p_{1*})}\right) = o(1/(T_1 p_{1*}))$$

hold uniformly in $\mathcal{A}_*$, where C is some positive constant.

Proof. Without loss of generality, we assume that $\delta_j^* > 0$. For each $j = 1, \ldots, p$, denote

$$N_j = \max_{\lfloor T_1 \varrho \rfloor < k \le T_1 - \lfloor T_1 \varrho \rfloor} \sqrt{\frac{k(T_1 - k)}{T_1}} \left| \frac{1}{T_1 - k} \sum_{i=k+1}^{T_1} \varepsilon_{ij}^{(1)} - \frac{1}{k} \sum_{i=1}^{k} \varepsilon_{ij}^{(1)} \right|.$$

For $j \in \mathcal{A}_*$ we have $|N_j - \xi_j^*| \le |\xi_j^{(1)}| \le N_j + \xi_j^*$. By Assumption 2, for sufficiently large T, $\xi_j^* - N_j \le \xi_j^{(1)} \le N_j + \xi_j^*$. By using similar arguments as in the proof of Lemma S2.2, the conclusion follows. □

Lemma S2.6. *Under Assumptions 1–6, if the threshold t_p is selected such that $t_p/\sqrt{\log T_1} \to c$ for some $c \ge 0$, then* $\Pr(\mathcal{A}_* \subseteq \widehat{\mathcal{S}}) \to 1$ *as $T, p \to \infty$.*

Proof. The Lemma S2.6 is indeed a corollary of Lemma S2.5. We have $\log \bar{s}_p \asymp \log(T_1 p_{1*})$ by Assumption 3. For $j \in \mathcal{A}_*$, $\xi_j^* \asymp \sqrt{T}|\delta_j^*|$ and $\sqrt{\log \bar{s}_p}/(\sqrt{T}|\delta_j^*|) \to 0$. Thus, the conclusion follows. □

Proof of Theorem 1 Recall that

$$L = \inf\left\{ s > 0 : \frac{\#\{j : W_j \le -s\}}{\#\{j : W_j \ge s\} \vee 1} \le \alpha \right\}.$$

We prove that $L \le s^*$ so that Lemma S2.3 and S2.4 can be utilized.

First, we show that

$$\Pr(W_j \le s^*, j \in \mathcal{A}_*) \to 0. \tag{S2.3}$$

Recall that $U_j^{(1)} = \xi_j^{(1)}/(\widehat{\sigma}_{jj}^{1/2})^{1/2}$, $U_j^{(2)} = \widehat{\beta}_j/(\widehat{V}_{jj})^{1/2}$. We have

$$
\begin{aligned}
&\Pr(W_j \leq s^* \text{ for some } j \in \mathcal{A}_*) \\
\leq\ & p_{1*} \max_{j \in \mathcal{A}_*} \Pr\Big\{\xi_j^{(1)}\widehat{\beta}_j/(\widehat{\sigma}_{jj}^{1/2}\widehat{V}_{jj})^{1/2} - \xi_j^*\beta_j^*/(\widehat{\sigma}_{jj}^{1/2}\widehat{V}_{jj})^{1/2} \\
&\qquad\qquad \leq s^* - \xi_j^*\beta_j^*/(\widehat{\sigma}_{jj}^{1/2}\widehat{V}_{jj})^{1/2}\Big\} \\
\leq\ & p_{1*} \max_{j \in \mathcal{A}_*} \Pr\Big(|\xi_j^{(1)} - \xi_j^*||\widehat{\beta}_j - \beta_j^*| + |\xi_j^*||\widehat{\beta}_j - \beta_j^*| + |\beta_j^*||\xi_j^{(1)} - \xi_j^*| \geq \xi_j^*\beta_j^*\{1 + o(1)\}\Big) \\
\leq\ & 2p_{1*} \max_{j \in \mathcal{A}_*} \Pr\Big(|\xi_j^{(1)} - \xi_j^*| \geq |\xi_j^*|/4\Big) + 2p_{1*} \max_{j \in \mathcal{A}_*} \Pr\Big(|\widehat{\beta}_j - \beta_j^*| \geq |\beta_j^*|/4\Big) \\
:=\ & 2V + 2VI.
\end{aligned}
$$

Under Assumption 2, 3, by Lemma S2.5, $V = o(1)$ and by Lemma S2.2, $VI = o(1)$. This implies

$$
\alpha p_{1*} \leq \alpha \sum_j \mathbb{I}(W_j \geq s^*). \tag{S2.4}
$$

Then we show

$$
\sum_j \mathbb{I}(W_j \leq -s^*) \leq \alpha p_{1*} \tag{S2.5}
$$

hold. We do decomposition:

$$
\sum_j \mathbb{I}(W_j \leq -s^*) = \sum_{j \notin \mathcal{A}_*} \mathbb{I}(W_j \leq -s^*) + \sum_{j \in \mathcal{A}_*} \mathbb{I}(W_j \leq -s^*)
$$

From Equation (S2.3), we have $\sum_{j \in \mathcal{A}_*} \mathbb{I}(W_j \leq -s^*) = o_p(1)$. Recall that s^* satisfies $G_-(s^*) = \alpha p_{1*}/(2S_0)$. By using similar arguments as in the proof of Lemma S2.4, we have

$$
\sum_{j \notin \mathcal{A}_*} \mathbb{I}(W_j \leq -s^*) \leq \left(\frac{1}{2} + \frac{p_1 - p_{1*}}{2S_0}\right)\{1 + o(1)\}\alpha p_{1*} \leq \alpha p_{1*}
$$

hold if $(p_1 - p_{1*})/S_0 \to 0$, which is fulfilled by Assumption 3.

Combining Equation (S2.4) and (S2.5), the inequality $L \leq s^*$ holds. Thereby, with Lemma S2.3 and S2.4, with probability one,

$$\frac{\sum_{j \notin \mathcal{A}} \mathbb{I}(W_j \geq L)}{\sum_{j \notin \mathcal{A}} \mathbb{I}(W_j \leq -L)} \to 1$$

By the definition of FDP, we have

$$\text{FDP} = \frac{\sum_{j \notin \mathcal{A}} \mathbb{I}(W_j \geq L)}{\sum_j \mathbb{I}(W_j \geq L) \vee 1} = \frac{\sum_j \mathbb{I}(W_j \leq -L)}{\sum_j \mathbb{I}(W_j \geq L) \vee 1} \times \frac{\sum_{j \notin \mathcal{A}} \mathbb{I}(W_j \geq L)}{\sum_j \mathbb{I}(W_j \leq -L)} \leq \alpha R(L)$$

where

$$R(L) = \frac{\sum_{j \notin \mathcal{A}} \mathbb{I}(W_j \geq L)}{\sum_j \mathbb{I}(W_j \leq -L)} \leq \frac{\sum_{j \notin \mathcal{A}} \mathbb{I}(W_j \geq L)}{\sum_{j \notin \mathcal{A}} \mathbb{I}(W_j \leq -L)}.$$

Thus, $\limsup_{T,p\to\infty} \text{FDP} \leq \alpha + o_p(1)$. Then for any $\epsilon > 0$,

$$\text{FDR} \leq (1+\epsilon)\alpha R(L) + \Pr\left(\text{FDP} \geq (1+\epsilon)\alpha R(L)\right)$$

which proves the rest part of the theorem. □

Proof of Proposition 1. This corollary is directly derived by Eq. (S2.3) and $L \leq s^*$.

S3 Necessary lemmas

Lemma S3.1. *Suppose Assumptions 5 and 6 hold, then*

$$\frac{1}{2c_{\overline{\kappa}}} \leq \liminf_{T\to\infty} \lambda_{\min}(\widehat{\mathbf{V}}) \leq \limsup_{T\to\infty} \lambda_{\max}(\widehat{\mathbf{V}}) \leq \frac{2}{c_{\underline{\kappa}}},$$

and

$$\frac{1}{4c_{\overline{\kappa}}} \leq \liminf_{T\to\infty} \lambda_{\min}(\mathbf{V}) < \limsup_{T\to\infty} \lambda_{\max}(\mathbf{V}) \leq \frac{4}{c_{\underline{\kappa}}}.$$

Proof. Firstly, we show some conclusions on the matrix $\hat{\mathbf{\Xi}}$. By Assumption 6, we have

$$\mathbf{\Xi} - \omega_p \mathbf{I}_p \leqslant \hat{\mathbf{\Xi}} = \mathbf{J} \circ \hat{\mathbf{\Sigma}} = \mathbf{\Xi} + \mathbf{J} \circ (\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}) \leqslant \mathbf{\Xi} + \omega_p \mathbf{I}_p$$

due to the unit diagonal elements of $\mathbf{J}$ and Lemma S3.4, and

$$(\mathbf{\Xi} + \omega_p \mathbf{I}_p)^{-1} \leqslant \hat{\mathbf{\Xi}}^{-1} \leqslant (\mathbf{\Xi} - \omega_p \mathbf{I}_p)^{-1}$$

By Woodbury matrix identity, we have

$$(\mathbf{\Xi} + \omega_p \mathbf{I}_p)^{-1} = \mathbf{\Xi}^{-1} - \mathbf{\Xi}^{-1} \left(\frac{1}{\omega_p} \mathbf{I}_p + \mathbf{\Xi}^{-1} \right)^{-1} \mathbf{\Xi}^{-1}.$$

It is easy to see

$$\mathbf{\Xi}^{-1} \left(\frac{1}{\omega_p} \mathbf{I}_p + \mathbf{\Xi}^{-1} \right)^{-1} \mathbf{\Xi}^{-1} \geqslant 0$$

and

$$\begin{aligned}
\lambda_{\max} \left\{ \mathbf{\Xi}^{-1} \left(\frac{1}{\omega_p} \mathbf{I}_p + \mathbf{\Xi}^{-1} \right)^{-1} \mathbf{\Xi}^{-1} \right\} &\leq \lambda_{\max} \left\{ \left(\frac{1}{\omega_p} \mathbf{I}_p + \mathbf{\Xi}^{-1} \right)^{-1} \right\} / \lambda_{\min}^2(\mathbf{\Xi}) \\
&\leq \frac{\omega_p}{\lambda_{\min}^2(\mathbf{\Sigma}) \{1 + \omega_p / \lambda_{\max}(\mathbf{\Sigma})\}} \to 0.
\end{aligned}$$

Similarly, we have

$$\lambda_{\max} \left\{ \mathbf{\Xi}^{-1} \left(\frac{1}{\omega_p} \mathbf{I}_p - \mathbf{\Xi}^{-1} \right)^{-1} \mathbf{\Xi}^{-1} \right\} \leq \frac{\omega_p}{\lambda_{\min}^2(\mathbf{\Sigma}) - \omega_p \lambda_{\min}(\mathbf{\Sigma})} \to 0.$$

Thus, $\|\hat{\mathbf{\Xi}}^{-1} - \mathbf{\Xi}^{-1}\|_2 = o_p \left\{ \frac{1}{u_p(p_1 - p_{1*})} \right\}$.

After that, we show $\|\hat{\boldsymbol{\Xi}}^{-1/2}\boldsymbol{\Xi}\hat{\boldsymbol{\Xi}}^{-1/2} - \mathbf{I}_p\|_2 = o_p(1)$. With $\hat{\boldsymbol{\Xi}} - \varpi\mathbf{I}_p \leqslant \boldsymbol{\Xi} \leqslant \hat{\boldsymbol{\Xi}} + \omega_p\boldsymbol{I}_p$, we have

$$\hat{\boldsymbol{\Xi}}^{-1/2}\boldsymbol{\Xi}\hat{\boldsymbol{\Xi}}^{-1/2} \leqslant \mathbf{I}_p + \omega_p\hat{\boldsymbol{\Xi}}^{-1} \leqslant \{1+o_p(1)\}\mathbf{I}_p + \omega_p\boldsymbol{\Xi}^{-1} \leqslant \left\{1 + \frac{\omega_p}{\lambda_{\min}(\boldsymbol{\Sigma})} + o_p(1)\right\}\mathbf{I}_p.$$

Similarly, we have $\hat{\boldsymbol{\Xi}}^{-1/2}\boldsymbol{\Xi}\hat{\boldsymbol{\Xi}}^{-1/2} \geqslant \left\{1 - \frac{\omega_p}{\lambda_{\min}(\boldsymbol{\Sigma})} - o_p(1)\right\}\mathbf{I}_p$. Then

$$\|\hat{\boldsymbol{\Xi}}^{-1/2}\boldsymbol{\Xi}\hat{\boldsymbol{\Xi}}^{-1/2} - \mathbf{I}_p\|_2 \leq \frac{\omega_p}{\lambda_{\min}(\boldsymbol{\Sigma})} + o_p(1) = o_p(1).$$

Next, we show the eigenvalues of $\widehat{\mathbf{V}}_{\widehat{\mathcal{S}}} = \left(\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}\right)^{-1}$ can be bounded. By the decomposition below,

$$\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\hat{\mathbf{X}}_{\widehat{\mathcal{S}}} = \left(\widehat{\boldsymbol{\Xi}}^{-1}\right)_{\widehat{\mathcal{S}},\widehat{\mathcal{S}}} = \mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}} + \{\widehat{\boldsymbol{\Xi}}^{-1} - \boldsymbol{\Xi}^{-1}\}_{\widehat{\mathcal{S}},\widehat{\mathcal{S}}},$$

we have

$$\begin{aligned}
\lambda_{\max}(\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}) &\leq \lambda_{\max}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}}) + \lambda_{\max}\left\{(\widehat{\boldsymbol{\Xi}}^{-1} - \boldsymbol{\Xi}^{-1})_{\widehat{\mathcal{S}},\widehat{\mathcal{S}}}\right\} \\
&\leq \lambda_{\max}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}}) + \lambda_{\max}\left(\widehat{\boldsymbol{\Xi}}^{-1} - \boldsymbol{\Xi}^{-1}\right) \\
&\leq \lambda_{\max}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}}) + \|\widehat{\boldsymbol{\Xi}}^{-1} - \boldsymbol{\Xi}^{-1}\|_2 \\
&\leq 2\lambda_{\max}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}})
\end{aligned}$$

and similarly, $\lambda_{\min}(\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}) \geq (1/2)\lambda_{\min}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}})$ for sufficiently large T, where the inequalities hold due to Lemma S3.5. Thus,

$$\frac{1}{2c_{\overline{\kappa}}} \leq \frac{1}{2\lambda_{\max}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}})} \leq \lambda_{\min}(\widehat{\mathbf{V}}) \leq \lambda_{\max}(\widehat{\mathbf{V}}) \leq \frac{2}{\lambda_{\min}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top\mathbf{X}_{\widehat{\mathcal{S}}})} \leq \frac{2}{c_{\underline{\kappa}}}$$

Finally, with the decomposition to $\mathbf{V}_{\widehat{\mathcal{S}}}$,

$$\mathbf{V} = (\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\hat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} + (\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\hat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1}\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\left(\hat{\boldsymbol{\Xi}}^{-1/2}\boldsymbol{\Xi}\hat{\boldsymbol{\Xi}}^{-1/2} - \mathbf{I}_p\right)\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}(\hat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top\hat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1},$$

we have

$$\frac{1}{4c_{\overline{\kappa}}} \leq \frac{1}{4\lambda_{\max}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})} \leq \lambda_{\min}(\mathbf{V}) \leq \lambda_{\max}(\mathbf{V}) \leq \frac{4}{\lambda_{\min}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})} \leq \frac{4}{c_{\underline{\kappa}}}.$$

□

Lemma S3.2 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_n$ be independent centered random variables almost surely bounded by $A < \infty$ in absolute value. Let $\sigma^2 = n^{-1} \sum_{i=1}^n \mathbb{E}(X_i^2)$. Then for all $x > 0$,*

$$\Pr\left(\sum_{i=1}^n X_i \geq x\right) \leq \exp\left(-\frac{x^2}{2n\sigma^2 + 2Ax/3}\right).$$

Lemma S3.3 (Moderate deviation for the independent sum)**.** *Suppose that $X_1, \ldots, X_n$ are independent random variables with mean zero, satisfying $\mathbb{E}(|X_j|^{2+\delta}) < \infty$ for $j = 1, \ldots, n$. Let $B_n = \sum_{i=1}^n \mathbb{E}(X_i^2)$. Then as $n \to \infty$*

$$\frac{\Pr(\sum_{i=1}^n X_i > x\sqrt{B_n})}{1 - \Phi(x)} \to 1,$$

uniformly in the domain $0 \leq x \leq C\{2\log(1/L_n)\}^{1/2}$, where $L_n = B_n^{-1-\delta/2} \sum_{i=1}^n \mathbb{E}(|X_i|^{2+\delta})$ and C is a constant satisfying $0 < C < 1$.

Lemma S3.4. *(Schott, 2016, Theorem 8.21) Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times m$ symmetric matrices. If $\mathbf{A}$ and $\mathbf{B}$ are non-negative definite, then the ith the largest eigenvalues of $\mathbf{A} \circ \mathbf{B}$ satisfies*

$$\lambda_{\min}(\mathbf{A}) \left\{\min_{1 \leq i \leq m} B_{i,i}\right\} \leq \lambda_i(\mathbf{A} \circ \mathbf{B}) \leq \lambda_{\max}(\mathbf{A}) \left\{\max_{1 \leq i \leq m} B_{i,i}\right\},$$

where $B_{i,i}$ is the ith diagonal element of $\mathbf{B}$.

Lemma S3.5. *(Horn and Johnson, 1985, p.189) Let* $\mathbf{A}$ *be an* $n \times n$ *Hermite matrix with* $\lambda_1(\mathbf{A}) \leq \ldots \leq \lambda_n(\mathbf{A})$, *let* r *be an integer with* $1 \leq r \leq n$, *and let* $\mathbf{A}_r$ *denote any* r*-by-*r *principal sub-matrix of* $\mathbf{A}$ *(obtained by deleting* $n-r$ *rows and the corresponding columns from* $\mathbf{A}$*). For each integer* k *such that* $1 \leq k \leq r$ *we have*

$$\lambda_k(\mathbf{A}) \leq \lambda_k(\mathbf{A}_r) \leq \lambda_{k+n-r}(\mathbf{A}).$$

Lemma S3.6. *(Fan et al., 2013, Lemma 2) Suppose that* $\mathbf{A}$ *and* $\mathbf{B}$ *are symmetric semi-positive definite matrices, and* $\lambda_{\min}(\mathbf{B}) > c_T$ *for a sequence* $c_T > 0$. *If* $\|\mathbf{A} - \mathbf{B}\| = o_p(c_T)$, *then* $\lambda_{\min}(\mathbf{A}) > c_T/2$, *and*

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\| = O_p(c_T^{-2})\|\mathbf{A} - \mathbf{B}\|.$$

Lemma S3.7. *(Thompson, 1972, Theorem 1) Let* A *be an* $m \times n$ *matrix with singular values* $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_{\min(m,n)}$. *Let* B *be a* $p \times S$ *sub-matrix of* A *with singular values* $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_{\min(p,S)}$. *Then*

$$\begin{aligned} &\alpha_i \geq \beta_i, && \text{for} \quad i = 1, 2, \ldots, \min(p, S) \\ &\beta_i \geq \alpha_{i+(m-p)+(n-S)}, && \text{for} \quad i \leq \min(p + S - m, p + S - n) \end{aligned}$$

S4 Additional experiments

S4.1 Synthetic data

We add some additional synthetic data experiments to show that the advantages of the SLIP against the classical Benjamini-Hochberg procedure are **not** limited to the chosen scenarios of Section 5.1.

We consider five procedures (SLIP-indep, SLIP-thresh, SLIP-lasso, BH-asymp, BH-simul), set the nominal FDR level $\alpha = 20\%$, and conducted 500 replications to estimate the FDR and power of each procedure as in the main text. First, we conduct a simulation study in the setting $\mathbf{\Sigma} = \mathbf{I}$ and more extreme random errors–(i) scaled bimodal distribution from $\mathcal{N}(2,1)/2 + \mathcal{N}(-2,1)/2$ with $\sigma = 1$ and (ii) exponential distribution with the parameter $\lambda = 1$. We set the proportion of activated data sequences $p_1 = \lfloor 0.15p \rfloor$, whose indices are randomly chosen from $\{1, \ldots, p\}$. For each activated data sequence, the activation time τ_j^* is randomly sampled from $\{\lfloor T\varrho \rfloor + 1, \ldots, T - 1 - \lfloor T\varrho \rfloor\}$ with $\varrho = 0.05$, and the change magnitude δ_j^* is firstly uniformly sampled from the interval $[\delta - 0.1, \delta + 0.1]$ with $\delta > 0.1$ and then its sign is flipped with probability 0.5, where δ is a parameter controlling the signal strength. Here, we assume the covariance is unknown and estimate the covariance matrix by Lee and Lee (2021). The result is presented in Figure S1.

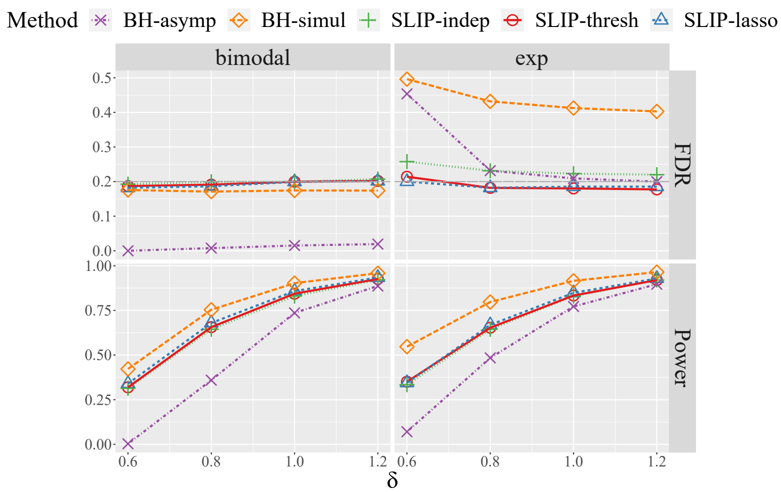

Figure S1: Empirical FDR and power of the SLIP and BH methods when $(T, p) = (120, 800)$ under the independent scenario with $\mathbf{\Sigma} = \mathbf{I}$.

We can see that the BH-simul outperforms the SLIP procedures for the bimodal normal mixture random errors, as in the normal case of Figure 1. However, for the exponentially-distributed random errors, which are asymmetric, the BH-simul severely losses the control of FDR because the calculation of component-wise p-values is based on the normal distribution. The SLIP-indep exactly controls the FDR for the bimodal case, which is explained in the finite-sample theory. The SLIP-lasso and SLIP-thresh maintain comparable power to the BH-simul for the bimodal case and control the FDR well in both cases.

We have shown the impact of the dependence under the t-distributed

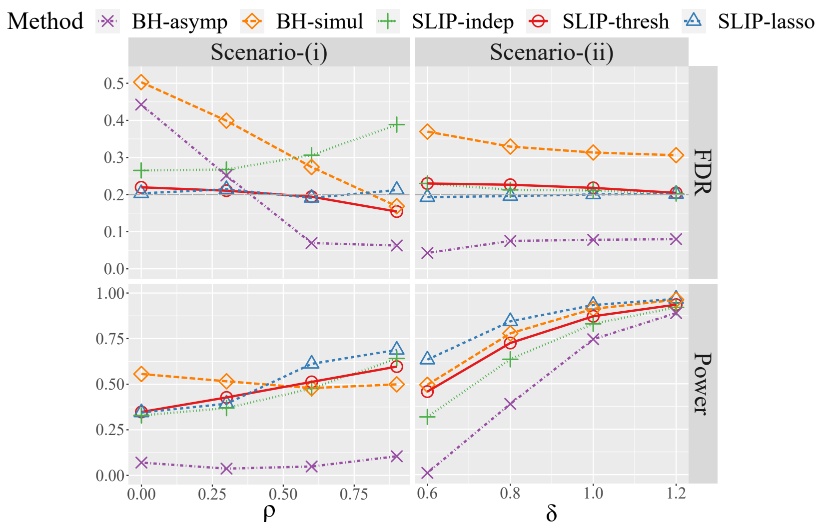

Figure S2: Empirical FDR and power of the SLIP and BH methods where the settings are the same as in Figure 2, except that the random errors are exponentially distributed (asymmetric).

(symmetric) random errors in Figure 2. To show the robustness of the SLIP procedure, we rerun the simulations in Figure 2, keeping all settings unchanged except replacing the t-distributed random errors with the exponentially distributed (asymmetric) ones. The result is shown in Figure S2. We can see that the performance of procedures in Figure S2 coincides with that in Figure 2. The SLIP procedures (SLIP-lasso and SLIP-thresh) still outperform the BH procedures, especially the SLIP-lasso.

S4.2 Real-data analysis

We included the ROIs' identifiers in the embedded data of the R package SLIP (`SLIP::fmri.data$regions`). We take one of the identifiers, "2001xlyr", as an example to illustrate the naming convention. The number "2001" indicates the region number in the automated anatomical labeling (AAL) template[1], "2001xl" represents that this region lies in the left[2] part of the partition along x-axis to the region "2001", "2001xlyr" represents that this region lies in the right part of the partition along y-axis to the region "2001xl", and so on. Then the identifier reflects the relationship with the AAL template and the partition path.

We list the specific identifiers of the ROIs detected by the SLIP-thresh and SLIP-lasso in Table S1 and collect other slices of the brain maps in Figure S3 and S4.

References

Cai, T. T. and W. Liu (2016). Large-scale multiple testing of correlations. *Journal of the American Statistical Association 111*(513), 229–240.

Du, L., X. Guo, W. Sun, and C. Zou (2021). False discovery rate control under general depen-

[1] The information of the total 116 AAL regions (brain: 90, cerebellum: 26) can be found at `http://rfmri.org/comment/1910#comment-1910`.

[2] "Left" means the smaller coordinate values while "right" means the larger values.

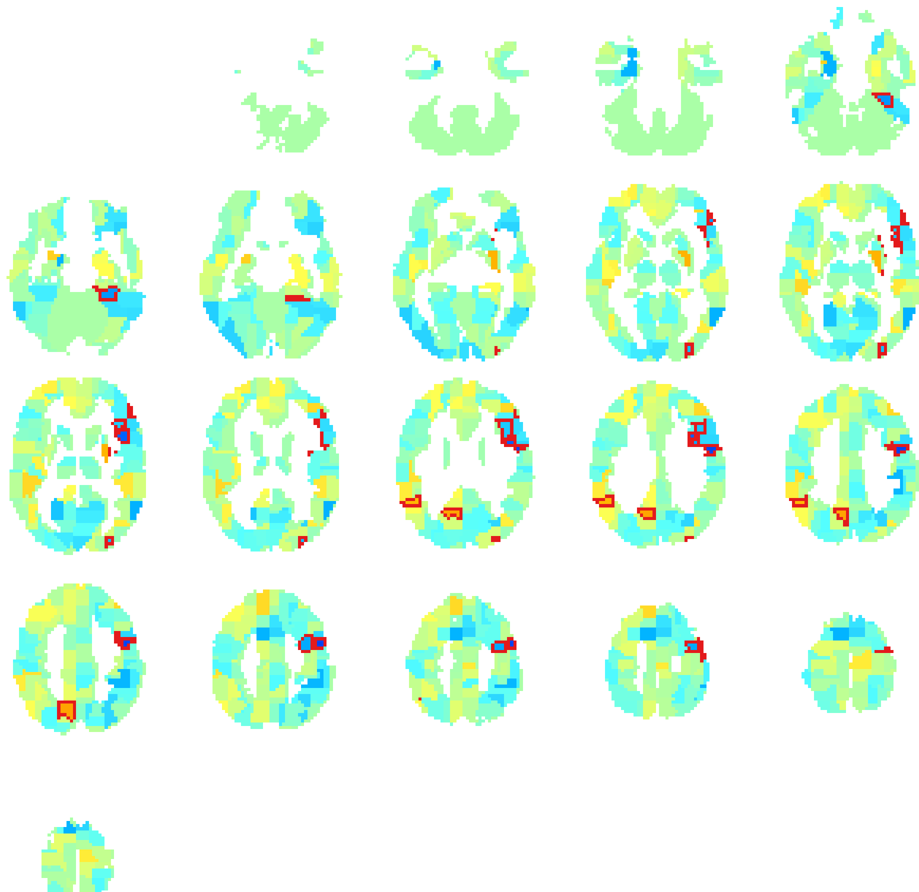

Figure S3: Activations in the brain detected by the SLIP-thresh. The activated regions are marked with red margins.

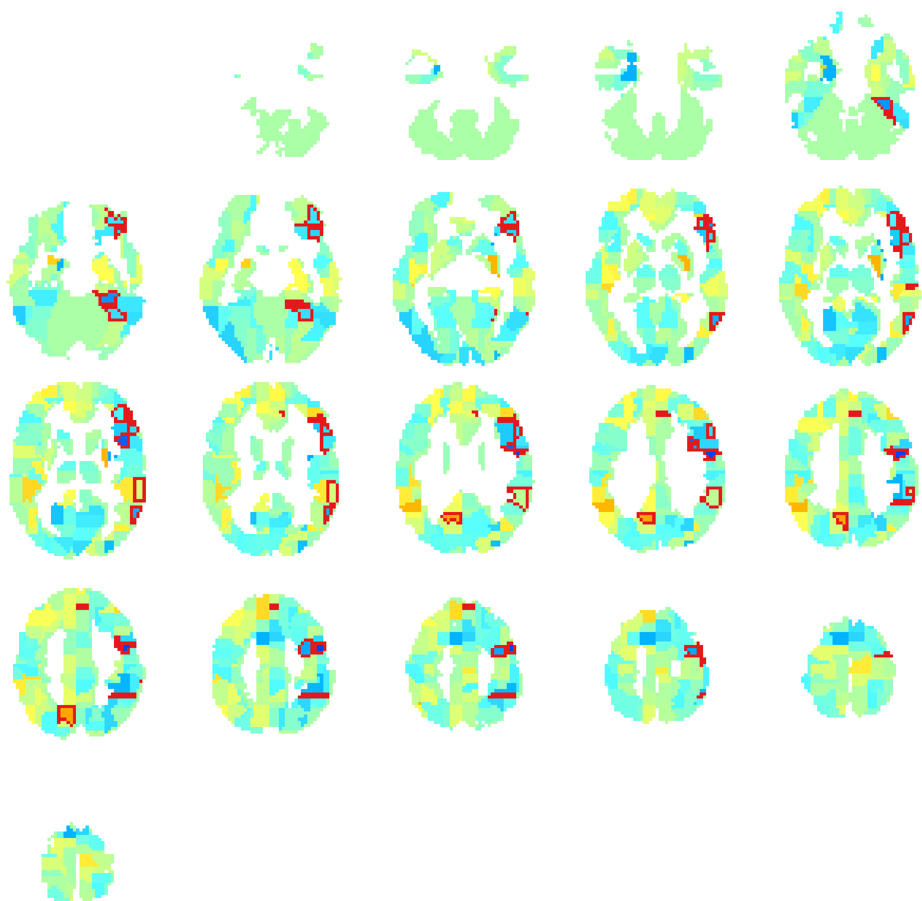

Figure S4: Activations in the brain detected by the SLIP-lasso. The activated regions are marked with red margins.

Procedure	Discoveries
SLIP-thresh	2001xlyr, 2001xryl, 2301xl, 2311ylxl, 2311yrxr, 3001, 5201xlylxr, 5401yryl, 6222xlyr, 6302ylzl
SLIP-lasso	2001xlyr, 2001xryl, 2301xl, 2311ylxr, 2311yrxl, 2311yrxr, 2321ylxr, 2321yr, 2601ylylyr, 5401ylyr, 5401yryl,6201ylyr, 6211yl, 6302ylzl, 8111ylxr, 8201ylylxr

Table S1: Discoveries of the SLIP-thresh and SLIP-lasso. The numbers in the front of the names of discoveries indicate the different ROI regions.

dence by symmetrized data aggregation. *Journal of the American Statistical Association*, 1–15. doi:10.1080/01621459.2021.1945459.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

Horn, R. A. and C. R. Johnson (Eds.) (1985). *Matrix Analysis*. USA: Cambridge University Press.

Lee, K. and J. Lee (2021). Estimating large precision matrices via modified Cholesky decomposition. *Statistica Sinica 31*, 173–196.

Schott, J. (2016). *Matrix Analysis for Statistics*. Wiley Series in Probability and Statistics. Wiley.

Thompson, R. (1972). Principal submatrices ix: interlacing inequalities for singular values of submatrices. *Linear Algebra and its Applications 5*(1), 1–12.