DECONFOUNDED-DEBIASED DISTRIBUTED ESTIMATION FOR HIGH-DIMENSIONAL MULTI-SITE DATA WITH HETEROGENEOUS PERVASIVE HIDDEN CONFOUNDERS

Zhaoyang Li, Chen Huang, Guoyou Qin* and Zhongyi Zhu*

Fudan University and York University

Supplementary Materials

S1 Proofs of theorems

We provide the proofs of Theorems 1-4 in this section.

*Corresponding Authors. E-mail: gyqin@fudan.edu.cn; zhuzy@fudan.edu.cn

S1.1 Proof of Theorem 1

Proof of Theorem 1. Recall that $\epsilon^{(k)} = \xi^{(k)} + \Delta^{(k)}$ where $\Delta^{(k)} = H^{(k)}\phi^{(k)} - X^{(k)}b^{(k)}$. By the definition of $\hat{\beta}_j^{(k)}$ in (3.4), we decompose

$$
\begin{aligned}
&\big(\dot{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\beta}_j^{(k)} - \beta_j\big) \\
&= \left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,\cdot}\xi^{(k)} \\
&\quad + \left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,\cdot}\Delta^{(k)} \\
&\quad + \left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{j,\cdot}b^{(k)} \\
&\quad + \left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-\frac{1}{2}}\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{j,\cdot}\big(\hat{\beta}^{(k),\mathrm{dec}} - \beta\big).
\end{aligned}
$$

Let $Z_i^{(k)} = \big(\big(\sigma_\xi^{(k)}\big)^2\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}/n_k^2\big)_{j,j}^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{j,i}\xi_i^{(k)}$, and $\mathsf{E}\big(Z_i^{(k)}|X^{(k)}\big) = 0$ and $\mathsf{Var}\big(Z_i^{(k)}|X^{(k)}\big) = \big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\big)_{j,j}^{-1}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\big)_{j,i}^2 < \infty$ for $i \in [n_k]$. It is obvious that $\big\{Z_i^{(k)}|X^{(k)}\big\}_{i=1}^{n_k}$ are mutually independent, thus $\mathsf{Var}\big(\sum_{i=1}^{n_k} Z_i^{(k)}|X^{(k)}\big) = 1$. By Lemma 1, if $\Delta_{\max} < 1$, we have $\sum_{i=1}^{n_k}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{j,i}^2 \asymp_{\mathrm{P}} n_k^{-1}$ and thus

$$
\max_i\left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-1}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,i}^2 \lesssim_{\mathrm{P}} n_k^{-1}
$$

By Assumption 1, if $\Delta_{\max} < 1$, we have for $\forall a > 0$

$$
\begin{aligned}
&\lim_{n_k\to\infty}\sum_{i=1}^{n_k}\mathsf{E}\big[\big(Z_i^{(k)}\big)^2\mathbb{I}\big\{\big|Z_i^{(k)}\big| > a\big\}|X^{(k)}\big]\\
&\lesssim \lim_{n_k\to\infty}\sum_{i=1}^{n_k}\left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-1}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,i}^2\\
&\qquad\exp\left\{-\left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,i}^{-2}a^2\right\}\\
&\lesssim_{\mathrm{P}} \lim_{n_k\to\infty}\exp\left\{-a^2/\max_i\left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-1}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,i}^2\right\}\\
&=0.
\end{aligned}
$$

Thus, central limit theorem yields that if $\Delta_{\max} < 1$,

$$
\left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,\cdot}\xi^{(k)}|X^{(k)}\xrightarrow{d}\mathcal{N}(0,1).
$$

It follows from Lemma 2 in Guo et al. (2022) that $\mathsf{E}\big\|\Delta^{(k)}\big\|_2^2/n_k = \mathsf{E}\big(\big(\Delta_i^{(k)}\big)^2\big) = \big(\phi^{(k)}\big)^{\mathrm{T}}\big(I_q - \Psi^{(k)}\big(\Sigma_X^{(k)}\big)^{-1}\big(\Psi^{(k)}\big)^{\mathrm{T}}\big)\phi^{(k)} \lesssim_{\mathrm{P}} q_k\sqrt{\log p}/\big(\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)\big)^2$ with probability larger than $1-(\log p)^{-\frac{1}{2}}$. Then Cauchy inequality yields that with probability tending to 1 as $n_k\to\infty$,

$$
\begin{aligned}
&\left|\left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,\cdot}\Delta^{(k)}\right|\\
&\le\left\|\left(\frac{\big(\sigma_\xi^{(k)}\big)^2}{n_k^2}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\right)_{j,j}^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{j,\cdot}\right\|_2\|\Delta^{(k)}\|_2\\
&=\frac{1}{\sigma_\xi^{(k)}}\|\Delta^{(k)}\|_2\\
&\lesssim_{\mathrm{P}}\frac{\sqrt{n_kq_k}(\log p)^{\frac{1}{4}}}{\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)},
\end{aligned}
$$

and it is $o_p(1)$ term if $\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)\gg\sqrt{n_kq_k}(\log p)^{\frac{1}{4}}$.

By Lemma 2 in Guo et al. (2022), $\big\|b^{(k)}\big\|_1\lesssim_{\mathrm{P}} pq_k(\log p)^{\frac{1}{2}}/\big\{\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)\big\}^2$ with probability larger than $1-(\log p)^{-\frac{1}{2}}$. By Lemma 1 and Assumption 5, if $\Delta_{\max} < 1$, Holder's inequality yields that with probability tending to 1 as $n_k \to \infty$,

$$\begin{aligned}
&\left| \left(\frac{\left(\sigma_\xi^{(k)}\right)^2}{n_k^2} \hat{\Theta}^{(k)} \left(X^{(k)}\right)^{\mathrm{T}} \left(Q^{(k)}\right)^4 X^{(k)} \left(\hat{\Theta}^{(k)}\right)^{\mathrm{T}} \right)_{j,j}^{-\frac{1}{2}} \left(\hat{\Theta}^{(k)} \hat{\Sigma}^{(k)}\right)_{j,\cdot} b^{(k)} \right| \\
&\leq \left\| \left(\frac{\left(\sigma_\xi^{(k)}\right)^2}{n_k^2} \hat{\Theta}^{(k)} \left(X^{(k)}\right)^{\mathrm{T}} \left(Q^{(k)}\right)^4 X^{(k)} \left(\hat{\Theta}^{(k)}\right)^{\mathrm{T}} \right)_{j,j}^{-\frac{1}{2}} \left(\hat{\Theta}^{(k)} \hat{\Sigma}^{(k)}\right)_{j,\cdot} \right\|_\infty \left\| b^{(k)} \right\|_1 \\
&= \max_{l \in [p]} \left| \left(\frac{\left(\sigma_\xi^{(k)}\right)^2}{n_k^2} \hat{\Theta}^{(k)} \left(X^{(k)}\right)^{\mathrm{T}} \left(Q^{(k)}\right)^4 X^{(k)} \left(\hat{\Theta}^{(k)}\right)^{\mathrm{T}} \right)_{j,j}^{-\frac{1}{2}} \left(\hat{\Theta}^{(k)} \hat{\Sigma}^{(k)}\right)_{j,l} \right| \left\| b^{(k)} \right\|_1 \\
&\lesssim_{\mathrm{P}} \sqrt{n_k} \Delta_{\max} p q_k (\log p)^{\frac{1}{2}} / \left\{ \lambda_{q_k}^{(k)} \left(\Psi^{(k)}\right) \right\}^2,
\end{aligned}$$

and it is $o_p(1)$ term if $\lambda_{q_k}^{(k)}\left(\Psi^{(k)}\right) \gg \sqrt{\Delta_{\max} p q_k} n_k^{\frac{1}{4}} (\log p)^{\frac{1}{4}}$.

By Corollary 2, Lemma 2, and Proposition 5 in Guo et al. (2022), if $(q_k+1)/n_k \leq \rho_k < 1$ and $\lambda_k \asymp \sqrt{(\log p)/n_k}$, we have

$$\begin{aligned}
\|\hat{\beta}^{(k),\mathrm{dec}} - \beta\|_1 &\lesssim_{\mathrm{P}} \frac{M_k^2 s \lambda_k}{\tau_*} + \frac{\|Q^{(k)} X^{(k)} b^{(k)}\|_2^2}{\lambda_k n_k}, \\
\|Q^{(k)} X^{(k)} b^{(k)}\|_2^2 &\lesssim_{\mathrm{P}} \frac{p q_k \sqrt{\log p}}{\left(\lambda_{q_k}^{(k)}\left(\Psi^{(k)}\right)\right)^2},
\end{aligned}$$

with probability tending to 1. By Lemma 1, if $\Delta_{\max} < 1$, Holder's inequality yields that with probability tending to 1 as $n_k \to \infty$,

$$\begin{aligned}
&\left| \left(\frac{\left(\sigma_\xi^{(k)}\right)^2}{n_k^2} \hat{\Theta}^{(k)} \left(X^{(k)}\right)^{\mathrm{T}} \left(Q^{(k)}\right)^4 X^{(k)} \left(\hat{\Theta}^{(k)}\right)^{\mathrm{T}} \right)_{j,j}^{-\frac{1}{2}} \left(I_p - \hat{\Theta}^{(k)} \hat{\Sigma}^{(k)}\right)_{j,\cdot} \left(\hat{\beta}^{(k),\mathrm{dec}} - \beta\right) \right| \\
&\leq \left\| \left(\frac{\left(\sigma_\xi^{(k)}\right)^2}{n_k^2} \hat{\Theta}^{(k)} \left(X^{(k)}\right)^{\mathrm{T}} \left(Q^{(k)}\right)^4 X^{(k)} \left(\hat{\Theta}^{(k)}\right)^{\mathrm{T}} \right)_{j,j}^{-\frac{1}{2}} \left(I_p - \hat{\Theta}^{(k)} \hat{\Sigma}^{(k)}\right)_{j,\cdot} \right\|_\infty \left\| \hat{\beta}^{(k),\mathrm{dec}} - \beta \right\|_1 \\
&= \max_{l \in [p]} \left| \left(\frac{\left(\sigma_\xi^{(k)}\right)^2}{n_k^2} \hat{\Theta}^{(k)} \left(X^{(k)}\right)^{\mathrm{T}} \left(Q^{(k)}\right)^4 X^{(k)} \left(\hat{\Theta}^{(k)}\right)^{\mathrm{T}} \right)_{j,j}^{-\frac{1}{2}} \left(I_p - \hat{\Theta}^{(k)} \hat{\Sigma}^{(k)}\right)_{j,l} \right| \left\| \hat{\beta}^{(k),\mathrm{dec}} - \beta \right\|_1 \\
&\lesssim_{\mathrm{P}} \sqrt{n_k} \Delta_{\max} \left(\frac{M_k^2 s \lambda_k}{\tau_*} + \frac{\|Q^{(k)} X^{(k)} b^{(k)}\|_2^2}{\lambda_k n_k} \right),
\end{aligned}$$

and it is $o_p(1)$ term if $s \Delta_{\max} M_k^2 \sqrt{\log p} \to 0$ and $\lambda_{q_k}^{(k)}\left(\Psi^{(k)}\right) \gg \sqrt{\Delta_{\max} p q_k}$.

Finally, we establish the asymptotic distribution of $\hat{\beta}_j^{(k)}$

$$\big(\dot{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\beta}_j^{(k)} - \beta_j\big) \xrightarrow{d} \mathcal{N}(0,1),$$

under Assumptions 1-6 and the above conditions. □

S1.2 Proof of Theorem 2

Proof of Theorem 2. It follows from Corollary 1 that $\lim_{n_k\to\infty} \mathsf{P}(j \in \hat{\mathcal{A}}^{(k)} \mid \beta_j = 0) \leq \alpha$. Thus, for any $a > 0$, there exists an m such that for all $n_k \geq m$,

$$\sum_{k=1}^{K} \mathbb{I}(j \in \hat{\mathcal{A}}^{(k)}) \sim \mathrm{Bin}(K, p_j),\ \ p_j \leq \alpha + a.$$

To prove $\lim_{n_k\to\infty} \mathsf{P}(j \in \hat{\mathcal{A}}) = 0$ for all $j \in \mathcal{A}^{\complement}$, it suffices to prove

$$\lim_{n_k\to\infty} \mathsf{P}\bigg\{\sum_{k=1}^{K} \mathbb{I}(j \in \hat{\mathcal{A}}^{(k)}) > \omega\bigg\} = 0,\ \ \forall j \in \mathcal{A}^{\complement}.$$

Chernoff bound yields that for all $j \in \mathcal{A}^{\complement}$,

$$\mathsf{P}\bigg\{\sum_{k=1}^{K} \mathbb{I}(j \in \hat{\mathcal{A}}^{(k)}) > \omega\bigg\} \leq \exp\Big\{-K \cdot D\Big(\frac{\omega}{K} \parallel \alpha + a\Big)\Big\},$$

where $D(\cdot \parallel \cdot)$ is the Kullback-Leibler divergence. If $\alpha < \omega/K$, we have $\mathsf{P}\{\sum_{k=1}^{K} \mathbb{I}(j \in \hat{\mathcal{A}}^{(k)}) > \omega\} \to 0$ as $n_k, K \to \infty$. Boole inequality yields that if $\alpha < \omega/K$,

$$\mathsf{P}(\hat{\mathcal{A}} \subset \mathcal{A}) \geq 1 - \sum_{j\in\mathcal{A}^{\complement}} \mathsf{P}(j \in \hat{\mathcal{A}}) \to 1,\ \text{ as } n_k, K \to \infty.$$

It follows from Corollary 1 that $\lim_{n_k\to\infty} \mathsf{P}(j \in \hat{\mathcal{A}}^{(k)} \mid \beta_j \neq 0) = 1$. Thus, for any $a > 0$, there exists an m such that for all $n_k \geq m$,

$$\sum_{k=1}^{K} \mathbb{I}(j \in \hat{\mathcal{A}}^{(k)}) \sim \mathrm{Bin}(K, p_j),\ \ p_j > 1 - a.$$

To prove $\lim_{n_k\to\infty} \mathsf{P}(j \notin \hat{\mathcal{A}}) = 0$ for all $j \in \mathcal{A}$, it suffices to prove

$$\lim_{n_k\to\infty} \mathsf{P}\bigg\{\sum_{k=1}^{K} \mathbb{I}(j \in \hat{\mathcal{A}}^{(k)}) \leq \omega\bigg\} = 0,\ \ \forall j \in \mathcal{A}.$$

Hoeffding inequality yields that for all $j \in \mathcal{A}$,

$$\mathsf{P}\bigg\{\sum_{k=1}^{K} \mathbb{I}(j \in \hat{\mathcal{A}}^{(k)}) \le \omega\bigg\} \le \exp\bigg[-\frac{2\{K(1-a)-\omega\}^2}{K}\bigg].$$

For any $a > 0$ and $\omega \in [0, K)$, we have $\mathsf{P}\{\sum_{k=1}^{K} \mathbb{I}(j \in \hat{\mathcal{A}}^{(k)}) \le \omega\} \to 0$ as $n_k, K \to \infty$. Boole inequality yields that

$$\mathsf{P}(\mathcal{A} \subset \hat{\mathcal{A}}) \ge 1 - \sum_{j \in \mathcal{A}} \mathsf{P}(j \notin \hat{\mathcal{A}}) \to 1, \text{ as } n_k, K \to \infty.$$

Thus, if $\alpha < \omega/K$, we establish

$$\lim_{n_k, K \to \infty} \mathsf{P}(\hat{\mathcal{A}} = \mathcal{A}) = 1.$$

□

S1.3 Proof of Theorem 3

Proof of Theorem 3. By Proposition 1, if $s \lesssim \min_{k \in [K]} M_k^{\frac{2}{3}}$, the inverse of $\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}$ exists with probability tending to 1 as $n_k \to \infty$ for all $k \in [K]$. Recall that $\epsilon^{(k)} = \xi^{(k)} + \Delta^{(k)}$ where $\Delta^{(k)} = H^{(k)}\phi^{(k)} - X^{(k)}b^{(k)}$. By the definition of $\hat{\beta}^{(k)}_{\mathcal{A}}$ in (3.4), for any $D \in \mathbb{R}^{l \times s}$ where $DD^{\mathrm{T}} \in \mathbb{R}^{l \times l}$ is positive definite, we decompose

$$\begin{aligned}
& D\bigg\{\sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\bigg\}^{-\frac{1}{2}} \sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big(\hat{\beta}^{(k)}_{\mathcal{A}} - \beta_{\mathcal{A}}\big) \\
&= D\bigg\{\sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\bigg\}^{-\frac{1}{2}} \sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\xi^{(k)} \\
&\quad + D\bigg\{\sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\bigg\}^{-\frac{1}{2}} \sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\Delta^{(k)} \\
&\quad + D\bigg\{\sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\bigg\}^{-\frac{1}{2}} \sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}b^{(k)} \\
&\quad + D\bigg\{\sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\bigg\}^{-\frac{1}{2}} \sum_{k=1}^{K} \big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}\big(\hat{\beta}^{(k),\mathrm{dec}} - \beta\big).
\end{aligned}$$

Let $U^{(k)} = D\big\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big\}^{-\frac{1}{2}}\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big(\hat\Theta^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},\cdot}\xi^{(k)}$, and $\mathsf{E}\big(U^{(k)}|X^{(k)}\big) = \mathbf{0}$ and $\mathsf{Var}\big(U^{(k)}|X^{(k)}\big) = D\big\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big\}^{-\frac{1}{2}}\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big\}^{-\frac{1}{2}}D^{\mathrm{T}}$ for $k \in [K]$. It is obvious that $\big\{U^{(k)}|X^{(k)}\big\}_{k=1}^K$ are mutually independent, thus $\mathsf{Var}\big(\sum_{k=1}^K U^{(k)}|X^{(k)}\big) = DD^{\mathrm{T}}$. By Lemma 1, if $s\Delta_{\max} < 1$, we have

$$
\begin{aligned}
&\left\| D\left\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\right\}^{-\frac{1}{2}}\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat\Theta^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\right\|_2^2 \\
&\le \|D\|_2^2\left\|\left\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\right\}^{-\frac{1}{2}}\right\|_2^2\left\|\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat\Theta^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\right\|_2^2 \\
&\lesssim_{\mathrm{P}} N^{-1}n_k.
\end{aligned}
$$

By Assumption 1, if $s\Delta_{\max} < 1$ and $\max_{k\in[K]} n_k/N \to 0$, we have for $\forall a > 0$

$$
\begin{aligned}
&\lim_{K\to\infty}\frac{1}{\mathrm{tr}(DD^{\mathrm{T}})}\sum_{k=1}^K \mathsf{E}\big[\big\|U^{(k)}\big\|_2^2\mathbb{I}\big\{\big\|U^{(k)}\big\|_2 > a\sqrt{\mathrm{tr}(DD^{\mathrm{T}})}\big\}|X^{(k)}\big] \\
&\le \lim_{K\to\infty}\frac{1}{\mathrm{tr}(DD^{\mathrm{T}})}\sum_{k=1}^K\left\| D\left\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\right\}^{-\frac{1}{2}}\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat\Theta^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\right\|_2^2\mathsf{E}\Bigg[\big\|\xi^{(k)}\big\|_2^2 \\
&\quad\times\mathbb{I}\left\{\big\|\xi^{(k)}\big\|_2 > a\sqrt{\mathrm{tr}(DD^{\mathrm{T}})}\left\| D\left\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\right\}^{-\frac{1}{2}}\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat\Theta^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\right\|_2^{-1}\right\}|X^{(k)}\Bigg] \\
&\lesssim \lim_{K\to\infty}\frac{1}{\mathrm{tr}(DD^{\mathrm{T}})}\sum_{k=1}^K\left\| D\left\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\right\}^{-\frac{1}{2}}\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat\Theta^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\right\|_2^2 \\
&\quad\times\left(n_k + a^2\mathrm{tr}(DD^{\mathrm{T}})\left\| D\left\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\right\}^{-\frac{1}{2}}\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat\Theta^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\right\|_2^{-2}\right) \\
&\quad\times\exp\left\{-a^2\mathrm{tr}(DD^{\mathrm{T}})\left\| D\left\{\sum_{k=1}^K \big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\right\}^{-\frac{1}{2}}\big(\dot\Sigma^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat\Theta^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\right\|_2^{-2}\right\} \\
&\lesssim_{\mathrm{P}} \lim_{K\to\infty}\sum_{k=1}^K N^{-1}n_k^2\exp\{-a^2Nn_k^{-1}\} + a^2\exp\{-a^2Nn_k^{-1}\} \\
&= 0.
\end{aligned}
$$

Central limit theorem yields that if $s\Delta_{\max} < 1$ and $\max_{k\in[K]} n_k/N \to 0$,

$$D\Big\{\sum_{k=1}^{K}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big\}^{-\frac{1}{2}}\sum_{k=1}^{K}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\xi^{(k)}|\big\{X^{(k)}\big\}_{k=1}^{K}\xrightarrow{d}\mathcal{N}(\mathbf{0},I_l).$$

It follows from Lemma 2 in Guo et al. (2022) that $\mathsf{E}\big\|\Delta^{(k)}\big\|_2^2/n_k = \mathsf{E}\big(\big(\Delta_i^{(k)}\big)^2\big) = \big(\phi^{(k)}\big)^{\mathrm{T}}\big(I_{q_k}-\Psi^{(k)}\big(\Sigma_X^{(k)}\big)^{-1}\big(\Psi^{(k)}\big)^{\mathrm{T}}\big)\phi^{(k)} \lesssim_{\mathrm{P}} q_k\sqrt{\log p}/\big(\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)\big)^2$ with probability larger than $1-(\log p)^{-\frac{1}{2}}$. By Lemma 1, if $s\Delta_{\max} < 1$, Cauchy inequality yields that with probability tending to 1 as $K\to\infty$,

$$\begin{aligned}
&\Big\|D\Big\{\sum_{k=1}^{K}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big\}^{-\frac{1}{2}}\sum_{k=1}^{K}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\Delta^{(k)}\Big\|_2\\
&\le \|D\|_2\Big\|\Big\{\sum_{k=1}^{K}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big\}^{-\frac{1}{2}}\Big\|_2\sum_{k=1}^{K}\Big\|\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\Big\|_2\big\|\Delta^{(k)}\big\|_2\\
&\lesssim_{\mathrm{P}} N^{-\frac{1}{2}}\sum_{k=1}^{K}\frac{n_k\sqrt{q_k}(\log p)^{\frac{1}{4}}}{\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)},
\end{aligned}$$

and it is $o_p(1)$ term if $\min_{k\in[K]}\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big) \gg (\sum_{k=1}^{K}n_k\sqrt{q_k})N^{-\frac{1}{2}}(\log p)^{\frac{1}{4}}$.

By Lemma 2 in Guo et al. (2022), $\big\|b^{(k)}\big\|_1 \lesssim_{\mathrm{P}} pq_k(\log p)^{\frac{1}{2}}/\big\{\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)\big\}^2$ with probability larger than $1-(\log p)^{-\frac{1}{2}}$. By Lemma 1 and Assumption 5, if $s\Delta_{\max} < 1$, Holder's inequality yields that with probability tending to 1 as $K\to\infty$,

$$\begin{aligned}
&\Big\|D\Big\{\sum_{k=1}^{K}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big\}^{-\frac{1}{2}}\sum_{k=1}^{K}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}b^{(k)}\Big\|_2\\
&\le \|D\|_2\Big\|\Big\{\sum_{k=1}^{K}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\Big\}^{-\frac{1}{2}}\Big\|_2\sum_{k=1}^{K}\big\|\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-1}\big\|_2\big\|\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}\big\|_2\big\|b^{(k)}\big\|_2\\
&\lesssim_{\mathrm{P}} \sqrt{s}N^{-\frac{1}{2}}\Delta_{\max}p(\log p)^{\frac{1}{2}}\sum_{k=1}^{K}\frac{n_kq_k}{\big\{\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)\big\}^2},
\end{aligned}$$

and it is $o_p(1)$ term if $\min_{k\in[K]}\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big) \gg \sqrt{\sum_{k=1}^{K}n_kq_k}\sqrt{\Delta_{\max}p}s^{\frac{1}{4}}N^{-\frac{1}{4}}(\log p)^{\frac{1}{4}}$.

By Corollary 2, Lemma 2, and Proposition 5 in Guo et al. (2022), if $(q_k+1)/n_k \le \rho_k < 1$

and $\lambda_k \asymp \sqrt{(\log p)/n_k}$, we have

$$\|\hat{\beta}^{(k),\mathrm{dec}} - \beta\|_1 \lesssim_{\mathrm{P}} \frac{M_k^2 s \lambda_k}{\tau_*} + \frac{\|Q^{(k)}X^{(k)}b^{(k)}\|_2^2}{\lambda_k n_k},$$

$$\|Q^{(k)}X^{(k)}b^{(k)}\|_2^2 \lesssim_{\mathrm{P}} \frac{p q_k \sqrt{\log p}}{\big(\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)\big)^2},$$

with probability tending to 1. By Lemma 1 and Assumption 5, if $s\Delta_{\max} < 1$, Holder's inequality yields that with probability tending to 1 as $K \to \infty$,

$$\begin{aligned}
&\left\| D\left\{\sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\right\}^{-\frac{1}{2}} \sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}\big(\hat{\beta}^{(k),\mathrm{dec}} - \beta\big)\right\|_2 \\
&\le \|D\|_2 \left\|\left\{\sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\right\}^{-\frac{1}{2}}\right\|_2 \sum_{k=1}^K \big\|\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\big\|_2 \big\|\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}\big\|_2 \big\|\hat{\beta}^{(k),\mathrm{dec}} - \beta\big\|_2 \\
&\lesssim_{\mathrm{P}} \sqrt{s} N^{-\frac{1}{2}} \Delta_{\max} \sum_{k=1}^K n_k \left(\frac{M_k^2 s\lambda_k}{\tau_*} + \frac{\|Q^{(k)}X^{(k)}b^{(k)}\|_2^2}{\lambda_k n_k}\right),
\end{aligned}$$

and it is $o_p(1)$ term if $s^{\frac{3}{2}}\Delta_{\max}N^{-\frac{1}{2}}\sqrt{\log p}(\sum_{k=1}^K \sqrt{n_k}M_k^2) \to 0$ and $\min_{k\in[K]} \lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big) \gg \sqrt{\sum_{k=1}^K \sqrt{n_k} q_k}\sqrt{\Delta_{\max}} p s^{\frac{1}{4}} N^{-\frac{1}{4}}$.

For any $D \in \mathbb{R}^{l\times s}$ where $DD^{\mathrm{T}} \in \mathbb{R}^{l\times l}$ is positive definite, we have

$$D\left\{\sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\right\}^{-\frac{1}{2}} \sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\big(\hat{\beta}_{\mathcal{A}}^{(k)} - \beta_{\mathcal{A}}\big) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_l),$$

under Assumptions 1–5 and 9, and the above conditions. On the event $\{\hat{\mathcal{A}} = \mathcal{A}\}$, we have $\hat{\beta}_{\mathcal{A}} = \big\{\sum_{k=1}^K \big(\tilde{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\big\}^{-1}\sum_{k=1}^K \big(\tilde{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\hat{\beta}_{\mathcal{A}}^{(k)}$. Then, it follows from Assumption 7 that

$$\begin{aligned}
&D\left\{\sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\right\}^{\frac{1}{2}}\big(\hat{\beta}_{\mathcal{A}} - \beta_{\mathcal{A}}\big) \\
&= D\left\{\sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\right\}^{\frac{1}{2}}\left\{\sum_{k=1}^K \big(\tilde{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\right\}^{-1}\sum_{k=1}^K \big(\tilde{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\big(\hat{\beta}_{\mathcal{A}}^{(k)} - \beta_{\mathcal{A}}\big) \\
&= D\left\{\sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\right\}^{-\frac{1}{2}}\sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\big(\hat{\beta}_{\mathcal{A}}^{(k)} - \beta_{\mathcal{A}}\big) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_l).
\end{aligned}$$

By Theorem 2, since $\lim_{n_k,K\to\infty}\mathsf{P}(\hat{\mathcal{A}} = \mathcal{A}) = 1$ and $D\big\{\sum_{k=1}^K \big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-1}\big\}^{\frac{1}{2}}\big(\hat{\beta}_{\mathcal{A}} - \beta_{\mathcal{A}}\big)$ is asymptotically normal on the event $\{\hat{\mathcal{A}} = \mathcal{A}\}$, the unconditional asymptotic normality follows. Thus, we establish the asymptotic distribution of $\hat{\beta}_{\mathcal{A}}$

$$D\left\{\sum_{k=1}^{K}\left(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\right)^{-1}\right\}^{\frac{1}{2}}\left(\hat{\beta}_{\mathcal{A}}-\beta_{\mathcal{A}}\right)\xrightarrow{d}\mathcal{N}(\mathbf{0},I_l),$$

under Assumptions 1–5, 7 and 9, and the above conditions. $\square$

S1.4 Proof of Theorem 4

Proof of Theorem 4. It is obvious that $\mathsf{E}(|\mathcal{A}^{\complement}\cap\hat{\mathcal{A}}^{(k)}|)=\mathsf{E}(|\hat{\mathcal{A}}^{(k)}|)-\mathsf{E}(|\mathcal{A}\cap\hat{\mathcal{A}}^{(k)}|)=\bar{s}_k-\mathsf{E}(|\mathcal{A}\cap\hat{\mathcal{A}}^{(k)}|)$. It follows from Assumption 10 that

$$\mathsf{E}(|\mathcal{A}^{\complement}\cap\hat{\mathcal{A}}^{(k)}|)\le\frac{\bar{s}_k}{\left(1+\frac{|\mathcal{A}|}{|\mathcal{A}^{\complement}|}\right)}=\frac{\bar{s}_k|\mathcal{A}^{\complement}|}{p},$$

$$\mathsf{E}(|\mathcal{A}\cap\hat{\mathcal{A}}^{(k)}|)\ge\frac{\bar{s}_k}{\left(1+\frac{|\mathcal{A}^{\complement}|}{|\mathcal{A}|}\right)}=\frac{\bar{s}_k|\mathcal{A}|}{p}.$$

By the exchangeability assumption in Assumption 10, we have

$$\mathsf{P}(j\in\hat{\mathcal{A}}^{(k)}\mid j\in\mathcal{A}^{\complement})=\frac{\mathsf{E}(|\mathcal{A}^{\complement}\cap\hat{\mathcal{A}}^{(k)}|)}{|\mathcal{A}^{\complement}|}\le\frac{\bar{s}_k}{p}\le\frac{s^*}{p},$$

$$\mathsf{P}(j\in\hat{\mathcal{A}}^{(k)}\mid j\in\mathcal{A})=\frac{\mathsf{E}(|\mathcal{A}\cap\hat{\mathcal{A}}^{(k)}|)}{|\mathcal{A}|}\ge\frac{\bar{s}_k}{p}\ge\frac{s_*}{p}.$$

Since observations in each site are independent, when $w\ge Ks^*/p-1$, we have

$$\mathsf{P}(j\in\hat{\mathcal{A}}\mid j\in\mathcal{A}^{\complement})\le 1-F\big(\omega\mid K,\frac{s^*}{p}\big),$$

$$\mathsf{P}(j\in\hat{\mathcal{A}}\mid j\in\mathcal{A})\ge 1-F\big(\omega\mid K,\frac{s_*}{p}\big).$$

Thus, the expected false positive rate (FPR) satisfies

$$\mathsf{E}(\text{FPR})=\mathsf{E}\Big(\frac{|\mathcal{A}^{\complement}\cap\hat{\mathcal{A}}|}{|\mathcal{A}^{\complement}|}\Big)=\sum_{j\in\mathcal{A}^{\complement}}\frac{\mathsf{P}(j\in\hat{\mathcal{A}}\mid j\in\mathcal{A}^{\complement})}{|\mathcal{A}^{\complement}|}\le 1-F\big(\omega\mid K,\frac{s^*}{p}\big),$$

and the expected true positive rate (TPR) satisfies

$$\mathsf{E}(\text{TPR})=\mathsf{E}\Big(\frac{|\mathcal{A}\cap\hat{\mathcal{A}}|}{|\mathcal{A}|}\Big)=\sum_{j\in\mathcal{A}}\frac{\mathsf{P}(j\in\hat{\mathcal{A}}\mid j\in\mathcal{A})}{|\mathcal{A}|}\ge 1-F\big(\omega\mid K,\frac{s_*}{p}\big).$$

$\square$

S2 Proofs of corollaries

This section presents the proofs of Corollaries 1 and 2.

S2.1 Proof of Corollary 1

Proof of Corollary 1. We start with the proof of the asymptotic validity of individual hypothesis testings. By Theorem 1, we have for any $a > 0$,

$$\begin{aligned}
&\lim_{n_k\to\infty} \sup_{\beta_j\in\mathcal{H}_{0,j}(s)} \mathsf{P}_\beta\big(P_j^{(k)} \le \alpha\big) \\
&= \lim_{n_k\to\infty} \sup_{\beta\in\mathbb{R}^p} \big\{\mathsf{P}_\beta\big(P_j^{(k)} \le \alpha\big) : \|\beta\|_0 \le s, \beta_j = 0\big\} \\
&= \lim_{n_k\to\infty} \sup_{\beta\in\mathbb{R}^p} \Big\{\mathsf{P}_\beta\Big(\big(\tilde{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\big|\hat{\beta}_j^{(k)} - \beta_j\big| > \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big) : \|\beta\|_0 \le s\Big\} \\
&\le \lim_{n_k\to\infty} \sup_{\beta\in\mathbb{R}^p} \left\{\mathsf{P}_\beta\Big(\big(\dot{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\big|\hat{\beta}_j^{(k)} - \beta_j\big| > (1-a)\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big) + \mathsf{P}_\beta\left(\left|\frac{\hat{\sigma}_\xi^{(k)}}{\sigma_\xi^{(k)}} - 1\right| \ge a\right) : \|\beta\|_0 \le s\right\} \\
&= \lim_{n_k\to\infty} \sup_{\beta\in\mathbb{R}^p} \left\{1 - \Phi\Big((1-a)\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big) + \Phi\Big((a-1)\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big) + \mathsf{P}_\beta\left(\left|\frac{\hat{\sigma}_\xi^{(k)}}{\sigma_\xi^{(k)}} - 1\right| \ge a\right) : \|\beta\|_0 \le s\right\}.
\end{aligned}$$

It follows from Assumption 7 that $\lim_{n_k\to\infty} \mathsf{P}\big(\big|\hat{\sigma}_\xi^{(k)}/\sigma_\xi^{(k)} - 1\big| \ge a\big) = 0$. For any $a > 0$, we establish

$$\lim_{n_k\to\infty} \sup_{\beta_j\in\mathcal{H}_{0,j}(s)} \mathsf{P}_\beta\big(P_j^{(k)} \le \alpha\big) \le \alpha.$$

Next, we establish the asymptotic power lower bounds on individual hypothesis testings. By Lemma 1, if $\Delta_{\max} < 1$, we have with probability tending to 1

$$\Big|\big(\dot{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\nu\Big| \asymp_{\mathrm{P}} \sqrt{n_k}\nu.$$

By Theorem 1, if $\Delta_{\max} < 1$, we have for some constant $C > 0$ and any $a > 0$,

$$
\begin{aligned}
&\lim_{n_k\to\infty} \inf_{\beta_j\in\mathcal{H}_{1,j}(s,\nu)} \frac{\mathsf{P}_\beta\big(P_j^{(k)} \le \alpha\big)}{G\big(\alpha, \Delta_n^{(k)}\big)} \\
&= \lim_{n_k\to\infty} \frac{1}{G\big(\alpha, \Delta_n^{(k)}\big)} \inf_{\beta\in\mathbb{R}^p} \Big\{ \mathsf{P}_\beta\big(P_j^{(k)} \le \alpha\big) : \|\beta\|_0 \le s, |\beta_j| \ge \nu \Big\} \\
&\ge \lim_{n_k\to\infty} \frac{1}{G\big(\alpha, \Delta_n^{(k)}\big)} \inf_{\beta\in\mathbb{R}^p} \Big\{ \mathsf{P}_\beta\Big(\big(\tilde{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\big|\hat{\beta}_j^{(k)} - \beta_j + \nu\big| > \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big) : \|\beta\|_0 \le s, \Big|\big(\dot{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\nu\Big| \asymp_{\mathrm{P}} \sqrt{n_k}\nu \Big\} \\
&\ge \lim_{n_k\to\infty} \frac{1}{G\big(\alpha, \Delta_n^{(k)}\big)} \inf_{\beta\in\mathbb{R}^p} \Bigg\{ \mathsf{P}_\beta\Big(\big(\tilde{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\big|\hat{\beta}_j^{(k)} - \beta_j + \nu\big| > \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) + 2a + a\Big|\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big|\Big) \\
&\qquad - \mathsf{P}_\beta\left(\left|\frac{\hat{\sigma}_\xi^{(k)}}{\sigma_\xi^{(k)}} - 1\right| \ge a\right) : \|\beta\|_0 \le s, \Big|\big(\dot{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\nu\Big| \asymp_{\mathrm{P}} \sqrt{n_k}\nu \Bigg\} \\
&= \lim_{n_k\to\infty} \frac{1}{G\big(\alpha, \Delta_n^{(k)}\big)} \inf_{\beta\in\mathbb{R}^p} \Bigg\{ \mathsf{P}_\beta\Big(\Big|\big(\tilde{\Sigma}_{j,j}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\beta}_j^{(k)} - \beta_j\big) + C\sqrt{n_k}\nu\Big| > \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) + 2a + a\Big|\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big|\Big) \\
&\qquad - \mathsf{P}_\beta\left(\left|\frac{\hat{\sigma}_\xi^{(k)}}{\sigma_\xi^{(k)}} - 1\right| \ge a\right) : \|\beta\|_0 \le s \Bigg\}.
\end{aligned}
$$

It follows from Assumption 7 that $\lim_{n_k\to\infty} \mathsf{P}\big(\big|\hat{\sigma}_\xi^{(k)}/\sigma_\xi^{(k)} - 1\big| \ge a\big) = 0$. For some constant $C > 0$ and any $a > 0$, we have

$$
\lim_{n_k\to\infty} \inf_{\beta_j\in\mathcal{H}_{1,j}(s,\nu)} \frac{\mathsf{P}_\beta\big(P_j^{(k)} \le \alpha\big)}{G\big(\alpha, \Delta_n^{(k)}\big)} \ge \lim_{n_k\to\infty} \inf_{\beta_j\in\mathcal{H}_{1,j}(s,\nu)} \frac{G\big(\alpha, C\sqrt{n_k}\nu\big)}{G\big(\alpha, \Delta_n^{(k)}\big)} = 1
$$

if $\Delta_n^{(k)} \asymp \sqrt{n_k}\nu$ and $\Delta_{\max} < 1$. □

S2.2 Proof of Corollary 2

Proof of Corollary 2. By Proposition 1, if $s \lesssim M_k^{\frac{2}{3}}$, the inverse of $\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}$ exists with probability tending to 1 as $n_k \to \infty$ for $k \in [K]$. Recall that $\epsilon^{(k)} = \xi^{(k)} + \Delta^{(k)}$ where $\Delta^{(k)} = H^{(k)}\phi^{(k)} -$

$X^{(k)}b^{(k)}$. By the definition of $\hat{\beta}_{\mathcal{A}}^{(k)}$ in (3.4), for any $u \in \mathbb{R}^s$ where $\|u\|_2 = 1$, we decompose

$$\begin{aligned}
&u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\beta}_{\mathcal{A}}^{(k)} - \beta_{\mathcal{A}}\big)\\
&= u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\xi^{(k)} + u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\Delta^{(k)}\\
&\quad + u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}b^{(k)} + u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}\big(\hat{\beta}^{(k),\mathrm{dec}} - \beta\big).
\end{aligned}$$

Let $W_i^{(k)} = u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\Big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\Big)_{\mathcal{A},i}\xi_i^{(k)}$, and $\mathsf{E}\big(W_i^{(k)}|X^{(k)}\big) = 0$ and $\mathsf{Var}\big(W_i^{(k)}|X^{(k)}\big) = \big(\sigma_\xi^{(k)}\big)^2 u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\Big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},i}\Big(\big(Q^{(k)}\big)^2X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}\Big)_{i,\mathcal{A}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}u/n_k^2$ for $i \in [n_k]$. It is obvious that $\big\{W_i^{(k)}|X^{(k)}\big\}_{i=1}^{n_k}$ are mutually independent, thus $\mathsf{Var}\big(\sum_{i=1}^{n_k} W_i^{(k)}|X^{(k)}\big) = 1$. By Lemma 1, if $s\Delta_{\max} < 1$, we have $\sum_{i=1}^{n_k}\big\|\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\big\|_2^2 \lesssim_{\mathrm{P}} sn_k^{-1}$ and thus

$$\begin{aligned}
&\max_{i\in[n_k]}\Big(u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\Big)^2\\
&\le \max_{i\in[n_k]}\|u^{\mathrm{T}}\|_2^2\big\|\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big\|_2^2\big\|\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\big\|_2^2\\
&\lesssim_{\mathrm{P}} sn_k^{-1}.
\end{aligned}$$

By Assumption 1, if $s\Delta_{\max} < 1$ and $s \ll n_k$, we have for $\forall a > 0$

$$
\begin{aligned}
&\lim_{n_k\to\infty}\sum_{i=1}^{n_k}\mathsf{E}\big[\big(W_i^{(k)}\big)^2\mathbb{I}\big\{\big|W_i^{(k)}\big|>a\big\}|X^{(k)}\big]\\
&=\lim_{n_k\to\infty}\sum_{i=1}^{n_k}\mathsf{E}\Big[\Big(u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\Big)^2\big(\xi_i^{(k)}\big)^2\\
&\qquad\times\mathbb{I}\Big\{\Big|u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\Big|\big|\xi_i^{(k)}\big|>a\Big\}|X^{(k)}\Big]\\
&=\lim_{n_k\to\infty}\sum_{i=1}^{n_k}\Big(u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\Big)^2\mathsf{E}\Big[\big(\xi_i^{(k)}\big)^2\\
&\qquad\times\mathbb{I}\Big\{\big|\xi_i^{(k)}\big|>\Big|u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\Big|^{-1}a\Big\}|X^{(k)}\Big]\\
&\lesssim\lim_{n_k\to\infty}\sum_{i=1}^{n_k}\Big(u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\Big)^2\\
&\qquad\times\exp\Big\{-a^2\Big(u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\Big)^{-2}\Big\}\\
&\lesssim\lim_{n_k\to\infty}\exp\Big\{-a^2/\max_{i\in[n_k]}\Big(u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2/n_k\big)_{\mathcal{A},i}\Big)^2\Big\}\\
&=0.
\end{aligned}
$$

Central limit theorem yields that if $s\Delta_{\max} < 1$ and $s \ll n_k$,

$$
u^{\mathrm{T}}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big)^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\xi^{(k)}|X^{(k)}\xrightarrow{d}\mathcal{N}(0,1).
$$

It follows from Lemma 2 in Guo et al. (2022) that $\mathsf{E}\big\|\Delta^{(k)}\big\|_2^2/n_k = \mathsf{E}\big(\big(\Delta_i^{(k)}\big)^2\big) = \big(\phi^{(k)}\big)^{\mathrm{T}}(I_{q_k}-\Psi^{(k)}\big(\Sigma_X^{(k)}\big)^{-1}\big(\Psi^{(k)}\big)^{\mathrm{T}})\phi^{(k)} \lesssim_{\mathrm{P}} q_k\sqrt{\log p}/\big(\lambda_{q_k}^{(k)}\big(\Psi^{(k)}\big)\big)^2$ with probability larger than

$1-(\log p)^{-\frac{1}{2}}$. Then Cauchy inequality yields that with probability tending to 1 as $n_k \to \infty$,

$$
\begin{aligned}
&\left|u^{\mathrm{T}}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\Delta^{(k)}\right|\\
&\le \|u^{\mathrm{T}}\|_2\left\|\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\Delta^{(k)}\right\|_2\\
&\le \left\|\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\Big(\frac{1}{n_k}\hat{\Theta}^{(k)}\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^2\Big)_{\mathcal{A},\cdot}\right\|_2\|\Delta^{(k)}\|_2\\
&= \frac{1}{\sigma^{(k)}_{\xi}}\|\Delta^{(k)}\|_2\\
&\lesssim_{\mathrm{P}} \frac{\sqrt{n_k q_k}(\log p)^{\frac{1}{4}}}{\lambda^{(k)}_{q_k}\big(\Psi^{(k)}\big)},
\end{aligned}
$$

and it is $o_p(1)$ term if $\lambda^{(k)}_{q_k}\big(\Psi^{(k)}\big) \gg \sqrt{n_k q_k}(\log p)^{\frac{1}{4}}$.

By Lemma 2 in Guo et al. (2022), $\big\|b^{(k)}\big\|_1 \lesssim_{\mathrm{P}} p q_k (\log p)^{\frac{1}{2}}/\big\{\lambda^{(k)}_{q_k}\big(\Psi^{(k)}\big)\big\}^2$ with probability larger than $1-(\log p)^{-\frac{1}{2}}$. By Lemma 1 and Assumption 5, if $s\Delta_{\max} < 1$, Holder's inequality yields that with probability tending to 1 as $n_k \to \infty$,

$$
\begin{aligned}
&\big|u^{\mathrm{T}}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}b^{(k)}\big|\\
&\le \big\|u^{\mathrm{T}}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}\big\|_{\max}\big\|b^{(k)}\big\|_1\\
&\le \max_{j\in[p]}\big\|u^{\mathrm{T}}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\big\|_1\big\|\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},j}\big\|_{\max}\big\|b^{(k)}\big\|_1\\
&\le \max_{j\in[p]}\sqrt{s}\|u^{\mathrm{T}}\|_2\big\|\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\big\|_2\big\|\big(\hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},j}\big\|_{\max}\big\|b^{(k)}\big\|_1\\
&\lesssim_{\mathrm{P}} \sqrt{s n_k}\Delta_{\max} p q_k (\log p)^{\frac{1}{2}}/\big\{\lambda^{(k)}_{q_k}\big(\Psi^{(k)}\big)\big\}^2,
\end{aligned}
$$

and it is $o_p(1)$ term if $\lambda^{(k)}_{q_k}\big(\Psi^{(k)}\big) \gg \sqrt{\Delta_{\max} p q_k}\, n_k^{\frac{1}{4}}(\log p)^{\frac{1}{4}} s^{\frac{1}{4}}$.

By Corollary 2, Lemma 2, and Proposition 5 in Guo et al. (2022), if $(q_k+1)/n_k \le \rho_k < 1$ and $\lambda_k \asymp \sqrt{(\log p)/n_k}$, we have

$$
\begin{aligned}
\|\hat{\beta}^{(k),\mathrm{dec}} - \beta\|_1 &\lesssim_{\mathrm{P}} \frac{M_k^2 s \lambda_k}{\tau_*} + \frac{\|Q^{(k)}X^{(k)}b^{(k)}\|_2^2}{\lambda_k n_k},\\
\|Q^{(k)}X^{(k)}b^{(k)}\|_2^2 &\lesssim_{\mathrm{P}} \frac{p q_k\sqrt{\log p}}{\big(\lambda^{(k)}_{q_k}\big(\Psi^{(k)}\big)\big)^2},
\end{aligned}
$$

with probability tending to 1. By Lemma 1 and Assumption 5, if $s\Delta_{\max} < 1$, Holder's inequality yields that with probability tending to 1 as $n_k \to \infty$,

$$
\begin{aligned}
&\left|u^{\mathrm{T}}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}\big(\hat{\beta}^{(k),\mathrm{dec}} - \beta\big)\right| \\
&\le \left\|u^{\mathrm{T}}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},\cdot}\right\|_{\max}\|\hat{\beta}^{(k),\mathrm{dec}} - \beta\|_1 \\
&\le \max_{j\in[p]}\left\|u^{\mathrm{T}}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\right\|_1\left\|\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},j}\right\|_{\max}\|\hat{\beta}^{(k),\mathrm{dec}} - \beta\|_1 \\
&\le \max_{j\in[p]}\sqrt{s}\|u^{\mathrm{T}}\|_2\left\|\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\right\|_2\left\|\big(I_p - \hat{\Theta}^{(k)}\hat{\Sigma}^{(k)}\big)_{\mathcal{A},j}\right\|_{\max}\|\hat{\beta}^{(k),\mathrm{dec}} - \beta\|_1 \\
&\lesssim_{\mathrm{P}} \sqrt{sn_k}\Delta_{\max}\left(\frac{M_k^2 s\lambda_k}{\tau_*} + \frac{\|Q^{(k)}X^{(k)}b^{(k)}\|_2^2}{\lambda_k n_k}\right),
\end{aligned}
$$

and it is $o_p(1)$ term if $s^{\frac{3}{2}}\Delta_{\max}M_k^2\sqrt{\log p} \to 0$ and $\lambda^{(k)}_{q_k}\big(\Psi^{(k)}\big) \gg \sqrt{\Delta_{\max}pq_k}s^{\frac{1}{4}}$.

For any $u \in \mathbb{R}^s$ where $\|u\|_2 = 1$, we establish the asymptotic distribution of $\hat{\beta}^{(k)}_{\mathcal{A}}$

$$
u^{\mathrm{T}}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big)^{-\frac{1}{2}}\big(\hat{\beta}^{(k)}_{\mathcal{A}} - \beta_{\mathcal{A}}\big) \xrightarrow{d} \mathcal{N}(0,1),
$$

under Assumptions 1–5, 8 and the above conditions. □

S3 Proof of proposition

We provide the proof of Proposition 1 in this section.

S3.1 Proof of Proposition 1

Proof of Proposition 1. Let $\mathrm{Null}(A) = \{a \in \mathbb{R}^p | Aa = \mathbf{0}\}$ be the null space of a matrix $A \in \mathbb{R}^{n\times p}$. It is obvious that $Q^{(k)}$ is positive definite. Then, $\mathrm{Null}\big(\big(Q^{(k)}\big)^2X^{(k)}\big)$ is the same as $\mathrm{Null}\big(Q^{(k)}X^{(k)}\big)$. To prove the positive definite $\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}$, it is equivalent to proving for any nonzero vector $v \in \mathbb{R}^p$

$$
v^{\mathrm{T}}\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}v = \left\|\big(Q^{(k)}\big)^2X^{(k)}\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}_{\cdot,\mathcal{A}}v\right\|_2^2,
$$

and then

$$\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}_{\cdot,\mathcal{A}} v \notin \mathrm{Null}\big(Q^{(k)} X^{(k)}\big).$$

Let $\mathcal{U}^{(k)} = \{u \in \mathbb{R}^p : u \neq \mathbf{0}, B \subseteq [p], \mid B \mid \leq s, \|u_{B^{\complement}}\|_1 \leq C M_k \|u_B\|_1\}$. By Assumption 4, we have for $\forall u \neq \mathbf{0}$

$$u \notin \mathrm{Null}\big(Q^{(k)} X^{(k)}\big).$$

Thus, the proof of the positive definite $\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}$ is equivalent to that of $\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}_{\cdot,\mathcal{A}} v \in \mathcal{U}^{(k)}$. It is obvious that $\mathcal{A} \subseteq [p]$ and $\mid \mathcal{A} \mid = s$. It follows from Lemma 1 that

$$\|\big(\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}_{\cdot,\mathcal{A}} v\big)_{\mathcal{A}^{\complement}}\|_1 \leq s \lambda^{(k)}_{\max}\big(\hat{\Theta}^{(k)}_{\mathcal{A},\mathcal{A}^{\complement}}\big)\|v\|_1 \lesssim_{\mathrm{P}} s^2 \Delta_{\max}\|v\|_1 \lesssim_{\mathrm{P}} s^{\frac{3}{2}} \lambda^{(k)}_{\min}\big(\hat{\Theta}^{(k)}_{\mathcal{A},\mathcal{A}}\big)\|v\|_2 \leq s^{\frac{3}{2}} \|\big(\big(\hat{\Theta}^{(k)}\big)^{\mathrm{T}}_{\cdot,\mathcal{A}} v\big)_{\mathcal{A}}\|_1.$$

Thus, for all $k \in [K]$, if $s \lesssim M_k^{\frac{2}{3}}$, we have the positive definite $\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}$ with probability tending to 1. □

S4 Lemma and proof

This section presents Lemma 1 and its proof.

S4.1 Lemma 1

Lemma 1. *Under Assumption 1-6, if* $(q_k + 1)/n_k \leq \rho_k < 1$ *and* $\Delta_{\max} < 1$*, for all* $k \in [K]$*, we have*

$$\big|\dot{\Sigma}^{(k)}_{j,j}\big| \asymp_{\mathrm{P}} n_k^{-1}$$

with probability tending to 1*. If* $(q_k + 1)/n_k \leq \rho_k < 1$ *and* $s\Delta_{\max} < 1$*, for all* $k \in [K]$*, we have*

$$\lambda^{(k)}_{\max}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big) \lesssim_{\mathrm{P}} n_k^{-1}$$

$$\lambda^{(k)}_{\min}\big(\dot{\Sigma}^{(k)}_{\mathcal{A},\mathcal{A}}\big) \gtrsim_{\mathrm{P}} n_k^{-1}$$

with probability tending to 1.

S4.2 Proof of Lemma 1

Proof of Lemma 1. It follows from (2.1) that $\big|\lambda_i^{(k)}\big(X^{(k)}\big) - \lambda_i^{(k)}\big(E^{(k)}\big)\big| \leq \big\|H^{(k)}\Psi^{(k)}\big\|_2$ for $i \in [n_k]$. By Assumption 2 and 6, we have $\lambda_i^{(k)}\big(X^{(k)}\big) \gg \sqrt{n_k}$ for $i \leq q_k$ and $\lambda_i^{(k)}\big(X^{(k)}\big) \asymp \sqrt{n_k}$ for $i > q_k$. The condition $(q_k+1)/n_k \leq \rho_k$ yields that $\lambda_{\lfloor \rho_k n_k \rfloor}^{(k)}\big(X^{(k)}\big) \asymp \sqrt{n_k}$. Thus, the singular values of $\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}$ are

$$\lambda_i^{(k)}\Big(\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}\Big) = \begin{cases} \big(\lambda_{\lfloor \rho_k n_k \rfloor}^{(k)}\big(X^{(k)}\big)\big)^4 / \big(\lambda_i^{(k)}\big(X^{(k)}\big)\big)^2 \to 0 \ (n_k \to \infty) & \text{if } i \leq q_k \\ \big(\lambda_i^{(k)}\big(X^{(k)}\big)\big)^2 \asymp n_k & \text{if } i > q_k \end{cases}.$$

It implies that for $\big(X^{(k)}\big)^{\mathrm{T}}\big(Q^{(k)}\big)^4 X^{(k)}$, the first q_k directional vectors $\{V_{\cdot,i}^{(k)}\}_{i=1}^{q_k}$ are removed as $n_k \to \infty$, and the vectors from directions q_k+1 to n_k, $\{V_{\cdot,i}^{(k)}\}_{i=q_k+1}^{n_k}$, are retained. Similarly, we have positive constants $c_{\min}^{(k)} > 0$ and $c_{\max}^{(k)} > 0$ such that

$$0 < c_{\min}^{(k)} \leq \lambda_{\min}^{(k)}\big(\hat{\Sigma}^{(k)}\big) \leq \lambda_{\max}^{(k)}\big(\hat{\Sigma}^{(k)}\big) \leq c_{\max}^{(k)} < \infty.$$

By Assumption 5, we have

$$\big|1 - \hat{\Theta}_{j,\cdot}^{(k)}\hat{\Sigma}_{\cdot,j}^{(k)}\big| \lesssim_{\mathrm{P}} \Delta_{\max},$$

$$\big\|I_s - \hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\hat{\Sigma}_{\cdot,\mathcal{A}}^{(k)}\big\|_2 \leq s\big\|I_s - \hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\hat{\Sigma}_{\cdot,\mathcal{A}}^{(k)}\big\|_{\max} \lesssim_{\mathrm{P}} s\Delta_{\max}.$$

Then, with probability tending to 1 as $n_k \to \infty$, we have

$$1 - \Delta_{\max} \leq \big|\hat{\Theta}_{j,\cdot}^{(k)}\hat{\Sigma}_{\cdot,j}^{(k)}\big| \leq 1 + \Delta_{\max},$$

$$1 - s\Delta_{\max} \leq \lambda_{\min}^{(k)}\big(\hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\hat{\Sigma}_{\cdot,\mathcal{A}}^{(k)}\big) \leq \lambda_{\max}^{(k)}\big(\hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\hat{\Sigma}_{\cdot,\mathcal{A}}^{(k)}\big) \leq 1 + s\Delta_{\max}.$$

We derive

$$1 - \Delta_{\max} \lesssim_{\mathrm{P}} \lambda^{(k)}\big(\hat{\Theta}_{j,\cdot}^{(k)}\big) = \frac{\big|\hat{\Theta}_{j,\cdot}^{(k)}\hat{\Sigma}_{\cdot,j}^{(k)}\big|}{\lambda^{(k)}\big(\hat{\Sigma}_{\cdot,j}^{(k)}\big)} \lesssim_{\mathrm{P}} 1 + \Delta_{\max},$$

$$\lambda_{\min}^{(k)}\big(\hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\big) \geq \frac{\lambda_{\min}^{(k)}\big(\hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\hat{\Sigma}_{\cdot,\mathcal{A}}^{(k)}\big)}{\lambda_{\max}^{(k)}\big(\hat{\Sigma}_{\cdot,\mathcal{A}}^{(k)}\big)} \gtrsim_{\mathrm{P}} 1 - s\Delta_{\max}, \quad \lambda_{\max}^{(k)}\big(\hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\big) \leq \frac{\lambda_{\max}^{(k)}\big(\hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\hat{\Sigma}_{\cdot,\mathcal{A}}^{(k)}\big)}{\lambda_{\min}^{(k)}\big(\hat{\Sigma}_{\cdot,\mathcal{A}}^{(k)}\big)} \lesssim_{\mathrm{P}} 1 + s\Delta_{\max}.$$

If $\Delta_{\max} < 1$, we have with probability tending to 1 as $n_k \to \infty$,

$$0 < c_{\min}^{(k)} \leq \lambda^{(k)}\big(\hat{\Theta}_{j,\cdot}^{(k)}\big) < c_{\max}^{(k)} < \infty.$$

Thus, if $\Delta_{\max} < 1$, we have with probability tending to 1 as $n_k \to \infty$,

$$\big|\dot{\Sigma}_{j,j}^{(k)}\big| \asymp_{\mathrm{P}} n_k^{-1}.$$

If $s\Delta_{\max} < 1$, we have with probability tending to 1 as $n_k \to \infty$,

$$0 < a_{\min}^{(k)} \leq \lambda_{\min}^{(k)}\big(\hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\big) \leq \lambda_{\max}^{(k)}\big(\hat{\Theta}_{\mathcal{A},\cdot}^{(k)}\big) \leq a_{\max}^{(k)} < \infty.$$

Thus, if $s\Delta_{\max} < 1$, we have with probability tending to 1 as $n_k \to \infty$,

$$\lambda_{\max}^{(k)}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big) \lesssim_{\mathrm{P}} n_k^{-1},$$

$$\lambda_{\min}^{(k)}\big(\dot{\Sigma}_{\mathcal{A},\mathcal{A}}^{(k)}\big) \gtrsim_{\mathrm{P}} n_k^{-1}.$$

□

S5 Additional simulation experiments

In this section, we provide the optimal parameter values and detailed simulation configurations. Furthermore, we introduce two global estimation methods and compare them with the proposed distributed method.

S5.1 Optimal ω values and simulation configurations

We provide the optimal ω values for three distributed estimation methods in Table 1, corresponding to all simulation cases depicted in Figures 4(a), 4(b), 5(a), and 5(b). Besides, detailed simulation configurations of Figures 5(a) and 5(b) are given in Tables 2 and 3.

Table 1: Optimal ω values of three distributed estimation methods corresponding to four simulation cases.

ω	Method	$K = 2$	$K = 5$	$K = 10$	$K = 20$
Homogeneous small-sample case	DDD	1	2	3	4
	DLD	1	3	5	8
	LD	1	4	6	10
Homogeneous large-sample case	DDD	1	2	2	3
	DLD	1	3	3	5
	LD	1	3	6	8
Heterogeneous small-sample case	DDD	1	2	3	5
	DLD	1	3	5	9
	LD	1	4	6	11
Heterogeneous large-sample case	DDD	1	2	3	4
	DLD	1	3	4	6
	LD	1	3	5	7

Table 2: Detailed simulation configurations of Figure 5(a).

$K = 2$	$n_1 = 80$, $n_2 = 120$
$K = 5$	$n_1 = 80$, $n_2 = 90$, $n_3 = 100$, $n_4 = 110$, $n_5 = 120$
$K = 10$	$n_1, n_2 = 80$, $n_3, n_4 = 90$, $n_5, n_6 = 100$, $n_7, n_8 = 110$, $n_9, n_{10} = 120$
$K = 20$	$n_{k\in[4]} = 80$, $n_{k\in[8]\setminus[4]} = 90$, $n_{k\in[12]\setminus[8]} = 100$, $n_{k\in[16]\setminus[12]} = 110$, $n_{k\in[20]\setminus[16]} = 120$

Table 3: Detailed simulation configurations of Figure 5(b).

$K = 2$	$n_1 = 400$, $n_2 = 600$
$K = 5$	$n_1 = 400$, $n_2 = 450$, $n_3 = 500$, $n_4 = 550$, $n_5 = 600$
$K = 10$	$n_1, n_2 = 400$, $n_3, n_4 = 450$, $n_5, n_6 = 500$, $n_7, n_8 = 550$, $n_9, n_{10} = 600$
$K = 20$	$n_{k\in[4]} = 400$, $n_{k\in[8]\setminus[4]} = 450$, $n_{k\in[12]\setminus[8]} = 500$, $n_{k\in[16]\setminus[12]} = 550$, $n_{k\in[20]\setminus[16]} = 600$

S5.2 Introduction to two global estimations

We provide a detailed description of the global deconfounded-debiased (DDG) estimation and the global deconfounded lasso (DLG) estimation methods in Algorithm 1. The DLG estimation is an intermediate step in the DDG estimation procedure. While the DDG estimation simultaneously corrects for hidden confounding bias and high-dimensional estimation bias, the DLG estimation corrects only for hidden confounding bias. Both global estimation methods are applicable to cross-site data that can be either homogeneous or heterogeneous. Under these two scenarios, the methods differ only in the construction of the spectral transformation matrix and the estimation of the random error variance, while all other steps remain identical. In the heterogeneous scenario, the spectral transformation matrix is constructed to incorporate the heterogeneity of hidden confounders across sites, and the random error variance is estimated to account for cross-site heteroscedasticity.

Next, we compare the proposed deconfounded-debiased distributed (DDD) estimation method with the DDG method. Notably, both methods are capable of correcting hidden confounding bias and high-dimensional estimation bias in both homogeneous and heterogeneous settings, thereby achieving high accuracy in coefficient estimation. The key differences between the two methods are primarily twofold:

- The DDG method is a global estimation procedure that requires pooling sample data

Algorithm 1: Global deconfounded-debiased (DDG) and global deconfounded lasso (DLG) estimations

Input: Data $\{X^{(k)}, Y^{(k)}\}_{k=1}^{K}$; parameters ρ or $\{\rho_k\}_{k=1}^{K}$

Output: DDG estimate $\hat{\beta}^{\mathrm{DDG}}$; asymptotic variance estimate $\tilde{\Sigma}$; p-values $\{P_j\}_{j=1}^{p}$; DLG estimate $\hat{\beta}^{\mathrm{DLG}}$

1. Pool all samples from all sites and define $X = \big((X^{(1)})^{\mathrm{T}}, \ldots, (X^{(K)})^{\mathrm{T}}\big)^{\mathrm{T}} \in \mathbb{R}^{N\times p}$, $Y = \big((Y^{(1)})^{\mathrm{T}}, \ldots, (Y^{(K)})^{\mathrm{T}}\big)^{\mathrm{T}} \in \mathbb{R}^{N}$, and $\xi = \big((\xi^{(1)})^{\mathrm{T}}, \ldots, (\xi^{(K)})^{\mathrm{T}}\big)^{\mathrm{T}} \in \mathbb{R}^{N}$.
2. Construct the spectral transformation matrix Q.

   (Homogeneous case): Let U be the left singular matrix of X and $\lambda_{\lfloor \rho N\rfloor}$ be the ρ-quantile of the singular values of X for a pre-specified parameter $\rho \in (0,1)$.

$$Q = USU^{\mathrm{T}},\ S = \mathrm{diag}(S_{1,1}, \ldots, S_{N,N}),\ S_{i,i} = \begin{cases} \lambda_{\lfloor \rho N\rfloor}/\lambda_i & \text{if } i \le \lfloor \rho N\rfloor \\ 1 & \text{otherwise} \end{cases}.$$

   (Heterogeneous case): Let $U^{(k)}$ be the left singular matrix of $X^{(k)}$ and $\lambda^{(k)}_{\lfloor \rho_k n_k\rfloor}$ be the ρ_k-quantile of the singular values of $X^{(k)}$ for a pre-specified parameter $\rho_k \in (0,1)$.

$$Q = \mathrm{diag}\big(Q^{(1)}, \ldots, Q^{(K)}\big),\ Q^{(k)} = U^{(k)} S^{(k)} \big(U^{(k)}\big)^{\mathrm{T}},$$

$$S^{(k)} = \mathrm{diag}\big(S^{(k)}_{1,1}, \ldots, S^{(k)}_{n_k,n_k}\big),\ S^{(k)}_{i,i} = \begin{cases} \lambda^{(k)}_{\lfloor \rho_k n_k\rfloor}/\lambda^{(k)}_i & \text{if } i \le \lfloor \rho_k n_k\rfloor \\ 1 & \text{otherwise} \end{cases}.$$

3. Compute the DLG estimate $\hat{\beta}^{\mathrm{DLG}} = \underset{\beta\in\mathbb{R}^p}{\arg\min}\Big\{\|QY - QX\beta\|_2^2/2N + \lambda\sum_{j=1}^{p}\|QX_{\cdot,j}\|_2|\beta_j|/\sqrt{N}\Big\}$, where λ is a regularization parameter selected via cross-validation.
4. Obtain the estimate $\hat{\Theta}$ of the precision matrix $\Theta = \{\mathrm{Cov}((QX)_{i,\cdot})\}^{-1}$.
5. Compute the DDG estimate $\hat{\beta}^{\mathrm{DDG}} = \hat{\beta}^{\mathrm{DLG}} + \hat{\Theta}(QX)^{\mathrm{T}}(QY - QX\hat{\beta}^{\mathrm{DLG}})/N$.
6. Obtain the estimate of the random error variance

$$\hat{\mathrm{Var}}(\xi) = \begin{cases} \hat{\sigma}_\xi^2 I_N & \text{(homogeneous case)} \\ \mathrm{diag}\Big(\big(\hat{\sigma}_\xi^{(1)}\big)^2 I_{n_1}, \ldots, \big(\hat{\sigma}_\xi^{(K)}\big)^2 I_{n_K}\Big) & \text{(heterogeneous case)} \end{cases}.$$

7. Compute the asymptotic variance estimate of $\hat{\beta}^{\mathrm{DDG}}$ as $\tilde{\Sigma} = \hat{\Theta}X^{\mathrm{T}}Q^2\hat{\mathrm{Var}}(\xi)Q^2X\hat{\Theta}^{\mathrm{T}}/N^2$.
8. Compute the p-value for $\hat{\beta}_j^{\mathrm{DDG}}$ as $P_j = 2\{1 - \Phi((\tilde{\Sigma}_{j,j})^{-\frac{1}{2}}|\hat{\beta}_j^{\mathrm{DDG}}|)\}$.

from all sites for subsequent estimation. In contrast, the DDD method adopts a distributed estimation framework which constructs the distributed estimator by transmitting and aggregating local estimates. At the cost of a slight loss in estimation accuracy, the DDD method is well suited to scenarios where sharing site-level sample data is prohibited.

- The DDG method performs variable selection through hypothesis testing. However, as indicated by Corollary 1, the Type I error rate is non-zero, which leads to poor variable selection performance, especially in excluding variables with truly zero coefficients. Conversely, the DDD method incorporates majority voting for variable selection when aggregating local estimates. According to Theorem 2, its variable selection is consistent.

S6 Additional real data application

We present selected variables identified by the deconfounded-debiased distributed estimation across varying thresholds ω in Table 4. Table 5 lists proteins shared between the deconfounded-debiased distributed estimation and each global estimation method. Enriched pathways for proteins identified by the global deconfounded-debiased estimation and the global lasso estimation are given in Figures 1 and 2.

S7 Additional justifications for assumptions

We provide additional justifications for assumptions in this section.

Table 4: Results of selected variables by the deconfounded-debiased distributed estimation under different thresholds ω.

ω	Selected variables
$\omega = 1$	Age; Disintegrin and metalloproteinase domain-containing protein 9; Adhesion G protein-coupled receptor E1; Angiotensinogen; Protein AMBP; Alpha-amylase 2B; Brevican core protein; Calcium-dependent secretion activator 1; C-C motif chemokine 20; CD2-associated protein; CUB domain-containing protein 1; Cadherin-3; C-type lectin domain family 4 member A; C-type lectin domain family 4 member C; Carboxypeptidase A1; Complement receptor type 1; Decorin; Dermatopontin; Dual specificity protein phosphatase 13 isoform A; Endothelin-converting enzyme 1; Zinc phosphodiesterase ELAC protein 1; Tumor necrosis factor ligand superfamily member 6; Fc receptor-like protein 2; Peptidyl-prolyl cis-trans isomerase FKBP14; Trans-acting T-cell-specific transcription factor GATA-3; Glial fibrillary acidic protein; Hepatitis A virus cellular receptor 1; Protein HEG homolog 1; Hephaestin; Hyaluronidase-1; Insulin-like growth factor 1 receptor; Inosine triphosphate pyrophosphatase; Kallikrein-8; Latent-transforming growth factor beta-binding protein 2; Cation-dependent mannose-6-phosphate receptor; Myoglobin; MORN repeat-containing protein 4; Myosin light chain 3; Nucleus accumbens-associated protein 1; Neurofilament light polypeptide; NFU1 iron-sulfur cluster scaffold homolog, mitochondrial; Phosphofurin acidic cluster sorting protein 2; Phospholipase A2; Dihydropteridine reductase; Rab9 effector protein with kelch motifs; Replication factor C subunit 4; RGM domain family member B; Protein S100-A4; Neuroserpin; Seizure 6-like protein 2; Organic solute transporter subunit beta; SPARC-related modular calcium-binding protein 1; Sphingomyelin phosphodiesterase 3; Transforming growth factor beta-2 proprotein; Transmembrane protease serine 11D; Tumor necrosis factor receptor superfamily member 21; Vasodilator-stimulated phosphoprotein; Z-DNA-binding protein 1
$\omega = 2$	Age; Latent-transforming growth factor beta-binding protein 2; Neurofilament light polypeptide; Organic solute transporter subunit beta
$\omega = 3$	Age; Latent-transforming growth factor beta-binding protein 2
$\omega \geq 4$	-

Table 5: Proteins shared between the deconfounded-debiased distributed estimation and each global estimation method.

Global estimation	Proteins
Global deconfounded-debiased	Alpha-amylase 2B; Glial fibrillary acidic protein; Neurofilament light polypeptide
Global lasso	Angiotensinogen; Alpha-amylase 2B; Brevican core protein; CUB domain-containing protein 1; Cadherin-3; Carboxypeptidase A1; Glial fibrillary acidic protein; Hepatitis A virus cellular receptor 1; Protein HEG homolog 1; Latent-transforming growth factor beta-binding protein 2; Cation-dependent mannose-6-phosphate receptor; Neurofilament light polypeptide; Rab9 effector protein with kelch motifs; Replication factor C subunit 4; RGM domain family member B; Organic solute transporter subunit beta; Transforming growth factor beta-2 proprotein

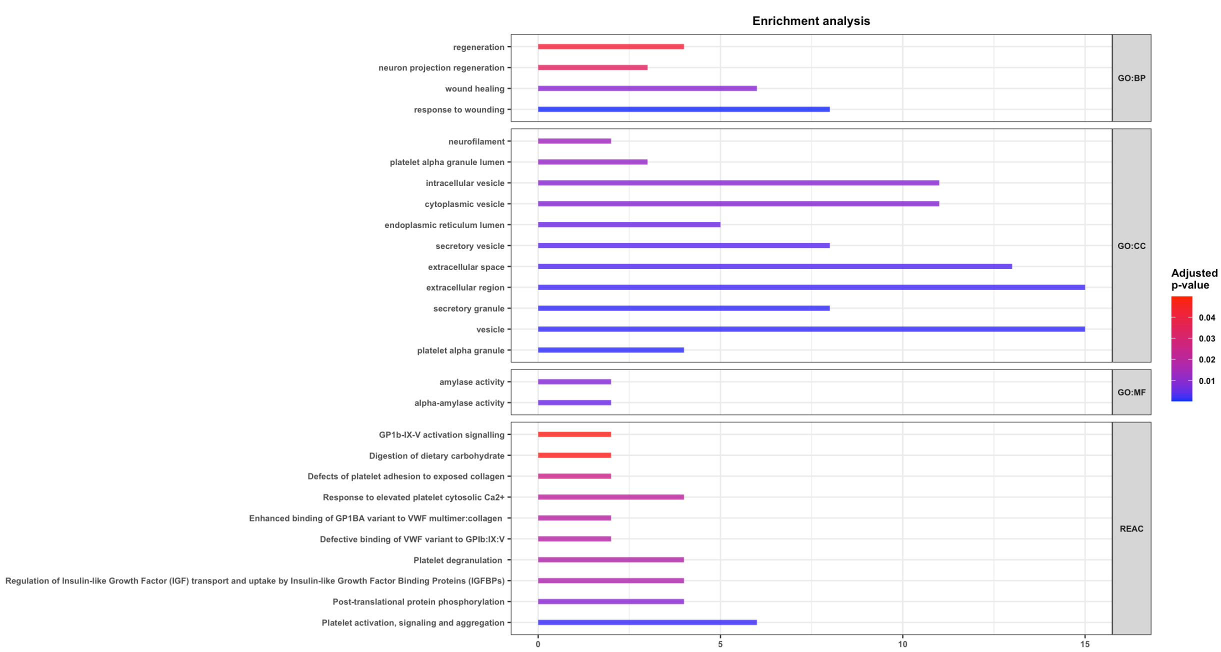

Figure 1: Enriched pathways for proteins identified by the global deconfounded-debiased estimation in enrichment analysis.

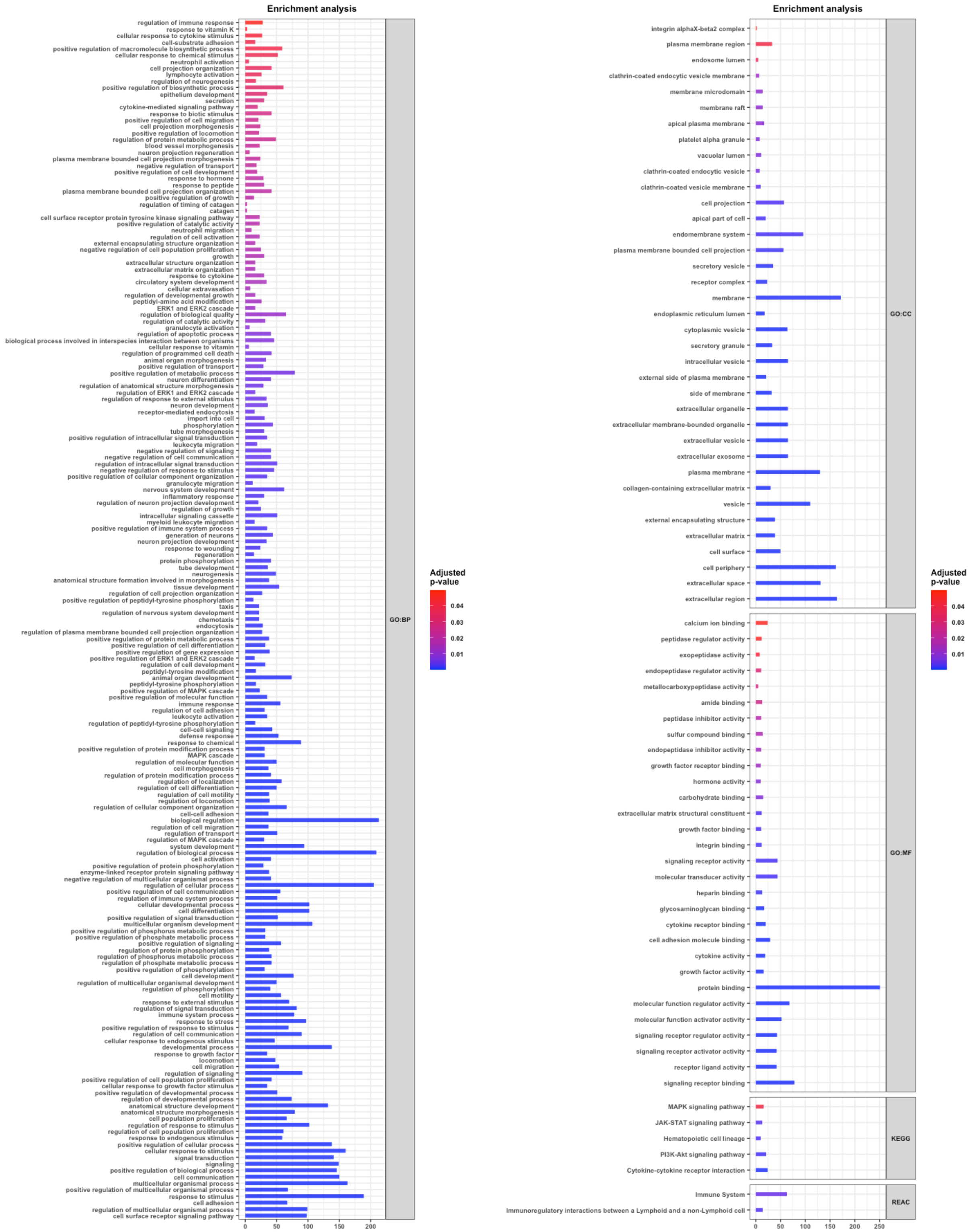

Figure 2: Enriched pathways for proteins identified by the global lasso estimation in enrichment analysis.

S7.1 Additional justification for Assumption 3

We offer additional justification for the reasonableness of Assumption 3 in typical data scenarios. Assumption 3 restricts the ℓ_2 norm of the hidden confounding coefficient $\phi^{(k)}$ at each site to grow at most at a controlled rate. This is a standard technical condition in the theoretical analysis of deconfounding methods (Guo et al., 2022). It ensures that the hidden confounding bias does not dominate the primary signal β, guaranteeing the identifiability and consistent estimation of the parameter of interest. Assumption 3 is reasonable in real-world scenarios for two main reasons. First, the bound scales with the number of hidden confounders q_k, which is typically much smaller than the dimension p (e.g., a few unobserved protein detection batch factors). Second, the effects of hidden confounders on the outcome Y are intrinsically bounded by physical or biological mechanisms and do not grow explosively with the dimension p. If the individual effects are $O(1)$, their cumulative effect $\left\|\phi^{(k)}\right\|_2^2$ is naturally of order $O(q_k)$, which readily satisfies the required bound $O(q_k\sqrt{\log p})$. The $\sqrt{\log p}$ term is a standard conservative scaling factor in high-dimensional statistics to accommodate dimensionality and noise (Bühlmann and Van De Geer, 2011), further ensuring that this assumption is reasonable in typical high-dimensional data scenarios.

S7.2 Additional justification for the identical β assumption across sites

It is reasonable to assume that after fully adjusting for all site-specific hidden confounders, the coefficients β of $X_{i,\cdot}^{(k)}$ on $Y_i^{(k)}$ are identical across all sites. This assumption rests on the fact that β represents the true linear effect of X on Y, an objective relationship that holds universally across sites and is not altered by between-site heterogeneity in hidden confounders. As shown in (2.1), in each site, once all hidden confounders are controlled for, the coefficient β indeed recovers this true linear effect. Hence, all sites share the identical coefficient β after fully adjusting for all heterogeneous hidden confounders across all sites. Conversely, if hidden confounders are not controlled within each site, the coefficient of $X_{i,\cdot}^{(k)}$ on $Y_i^{(k)}$ becomes $\beta + b^{(k)}$, where the perturbed coefficient $b^{(k)}$ reflects the hidden confounding. In this scenario, the estimand shifts to the composite effect $\beta + b^{(k)}$, which no longer corresponds to the true linear effect of X on Y.

Next, we further explain this assumption from a super-population perspective. Let $P^*(X, H, Y)$ denote the joint distribution of (X, H, Y) in the super-population, where H encompasses all site-specific hidden confounders. Decomposing this joint distribution yields

$$P^*(X, H, Y) = P^*(Y|X, H)P^*(X, H).$$

Consider the following linear regression model for (X, H, Y):

$$\begin{aligned} Y &= X\beta + H\phi + \xi, \\ X &= H\Psi + E, \end{aligned} \tag{S7.1}$$

where β is a global parameter in the super-population that represents the true linear effect of X on Y after fully adjusting for all site-specific hidden confounders H. Samples from each site can be regarded as independent draws from the super-population via stratified sampling, with each site corresponding to a stratum. Thus, the joint distribution within site k is given by $P_k(X, H, Y) = P^*(X, H, Y|S = k)$, where S denotes the site indicator. According to (2.1), the joint distribution can be factorized as

$$P_k(X, H, Y) = P_k(Y|X, H)P_k(X|H)P_k(H) = P_k(Y|X, H)P_k(E)P_k(H).$$

Under the covariate shift assumption, the marginal distributions $P_k(E)$ and $P_k(H)$ differ across sites, i.e., $P_k(E) \neq P^*(E)$ and $P_k(H) \neq P^*(H)$. However, the conditional distribution remains invariant across sites, i.e., $P_k(Y|X,H) = P^*(Y|X,H)$ for all $k \in [K]$. Therefore, as implied by (2.1), after adjusting for all hidden confounders, the true linear effect of X on Y in each site can be consistently represented by the global parameter β. This justifies our assumption that, after fully adjusting for all site-specific hidden confounders, the coefficients β of $X_{i,\cdot}^{(k)}$ on $Y_i^{(k)}$ are identical across all sites.

S8 Discussion on information leakage

Regarding the potential for information leakage resulting from the aggregation of local estimators in our method, we provide an in-depth analysis from two key perspectives.

First, our method faces multi-level information leakage risks, including feature reconstruction, subgroup attribute inference, and individual identity identification.

- The local estimators transmitted to the central site comprise coefficient estimates and asymptotic variance estimates. Inherently, the local coefficient estimates preserve the Fisher information regarding the global parameters, while the local asymptotic variance estimates encapsulate the second-order moment information of the local design matrix. Consequently, the central site can leverage these variance structures to inversely reconstruct the geometric properties of the local feature distribution.

- Local estimators inevitably absorb systematic biases from the local heterogeneous data. By horizontally comparing local estimates across different sites, the central site can infer the statistical characteristics of specific local datasets. For instance, an abnormal shift in a local coefficient estimate may reveal that the incidence rate of a specific disease at a given hospital is significantly higher than the global average.

- In the finite-sample setting, local estimators remain highly sensitive to individual data points via influence functions. If a specific sample possesses extreme characteristics (e.g., a rare medical history), its inclusion or exclusion will substantially alter the local estimates. Consequently, the central site can infer the presence of specific extreme samples within the local dataset by observing anomalous fluctuations in a given local estimator.

- Although the central site lacks direct access to the raw data, the aggregated local estimators embed parametric projections of the local data. By leveraging these projections, the central site can launch attribute inference attacks to infer the distributional characteristics of a specific local dataset. Furthermore, by detecting anomalies in fitting residuals, it could even execute membership inference attacks to determine whether a sensitive record participated in the local training.

Second, our method mitigates information leakage through two-round communication, majority voting, and deconfounding techniques.

- Within the mathematical framework of database reconstruction attacks, the effective convergence of an attack heavily relies on the accumulation of extensive query constraints. By employing two round communications, our method fundamentally limits the volume of cumulative information extractable from local data, thereby severing the attack path that attempts to reconstruct individual records through massive cumulative queries.

- The central site aggregates only the local estimators that pass the majority voting screening. This voting-integrated aggregation acts as an information bottleneck, preventing the central site from accessing the complete coefficient vector of any single site.

- The local estimates transmitted to the central site are deconfounded. Consequently, the parameter projections received by the central site are stripped of the structural influence of local hidden confounders. This inherently obfuscates the original data structure, significantly increasing the difficulty of directly reconstructing local covariates.

In summary, while our method mitigates information leakage, it cannot theoretically preclude the potential for information leakage. Therefore, when applied to highly sensitive data, our method necessitates the integration of privacy-enhancing technologies, such as differential privacy.

Bibliography

Bühlmann, P. and S. Van De Geer (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Guo, Z., D. Ćevid, and P. Bühlmann (2022). Doubly debiased lasso: High-dimensional inference under hidden confounding. *Annals of statistics 50*(3), 1320.