Supplement to " Model-free multiple testing for matrix-valued predictors with false discovery control"

Lei Yan
Department of Statistics, Florida State University
and
Xin Zhang
Department of Statistics, Florida State University
and
Zhigen Zhao
Department of Statistics, Operations, and Data Science, Temple University

April 29, 2026

This supplementary material supports the main paper, including detailed technical proofs for our theoretical analyses and additional numerical results. We develop the Gaussian approximation of folding selection subspaces in Section A. Section B lists assumptions required in the theoretical results. Section C includes additional theoretical properties of our procedure. The complete proofs for all theorems and lemmas are provided in Section D, with ancillary lemmas detailed in Section E. Section F contains further simulation studies, and Section G provides details on the EEG real data application and analysis.

A Gaussian approximation of folding selection subspaces

We aim to derive a central limit theorem (CLT) of the sample estimators $\widehat{\mathbf{A}} = \widehat{\Omega}\widehat{\mathrm{E}}[Y\mathbf{X}]$ and $\widehat{\mathbf{B}} = \hat{\Gamma}\widehat{\mathrm{E}}[Y\mathbf{X}^T]$, where $\widehat{\mathrm{E}}[\cdot]$ denotes the sample average over n i.i.d. samples $(Y_i, \mathbf{X}_i)$, and $\widehat{\Omega}, \hat{\Gamma}$ are estimators of the precision matrices. We use the Gaussian approximation theory (Chernozhukov et al. 2013, 2017), combining with techniques in foundational work by (Zhao & Xing 2022), to show $\sqrt{n}\widehat{\mathbf{A}}$ and $\sqrt{n}\widehat{\mathbf{B}}$ behave like the sum of i.i.d. Gaussian matrices. For $j = 1, \dots, p$ and $k = 1, \dots, q$, let $X_{i,jk}$ denote the (j,k)-th element of $\mathbf{X}_i$, $\mathbf{X}_{i,j\cdot}$ and $\mathbf{X}_{i,\cdot k}$ represent the j-th row and k-th column of $\mathbf{X}_i$. We define $z^L_{i,jk}$ for $j = 1, \dots, p, k = 1, \dots q$ as a collection of Gaussian random variables with mean and covariance structure:

$$\mathrm{E}[z^L_{i,jk}] = \Omega^T_{\cdot j}\,\mathrm{E}[Y\mathbf{X}_{\cdot k}], \mathrm{var}(z^L_{i,jk}) = \Omega^T_{\cdot j}\,\mathrm{E}[Y^2\mathbf{X}_{\cdot k}\mathbf{X}^T_{\cdot k}]\Omega_{\cdot j} - \Omega^T_{\cdot j}\,\mathrm{E}[Y\mathbf{X}_{\cdot k}]\,\mathrm{E}[Y\mathbf{X}^T_{\cdot k}]\Omega_{\cdot j},$$

$$\mathrm{cov}(z^L_{i,jk}, z^L_{i,lm}) = \Omega^T_{\cdot j}\,\mathrm{E}[Y^2\mathbf{X}_{\cdot k}\mathbf{X}^T_{\cdot m}]\Omega_{\cdot l} - \Omega^T_{\cdot j}\,\mathrm{E}[Y\mathbf{X}_{\cdot k}]\,\mathrm{E}[Y\mathbf{X}^T_{\cdot m}]\Omega_{\cdot l}.$$

The variables $z^R_{i,kj}$ are defined similarly (Section C of the Supplement). Let $\mathbf{z}^L_{i,\cdot\cdot}$ (or $\mathbf{z}^R_{i,\cdot\cdot}$) denote the $p \times q$ (or $q \times p$) matrix consisting of $z^L_{i,jk}$ (or $z^R_{i,kj}$) where $j = 1, \dots, p$ and $k = 1, \dots, q$. Under $\mathcal{H}^L_j$ and the coverage assumption, the j-th row of $\mathrm{E}[\mathbf{z}^L_{i,\cdot\cdot}]$ is zero, and similarly for columns under $\mathcal{H}^R_k$.

Let $\mathcal{F}^{\mathrm{re}}$ consist of all sets F of the form $F = \{\omega \in \mathbb{R}^{pq} : a_j \le \omega_j \le b_j, j = 1, 2, \dots, pq\}$ for some $-\infty \le a_j \le b_j \le \infty$. Define the maximal approximation errors over the class of hyper-rectangle as

$$\hat{\rho}^L_n(\mathcal{F}^{\mathrm{re}}) = \sup_{F \in \mathcal{F}^{\mathrm{re}}} \left| \Pr(\sqrt{n}\widehat{\mathbf{A}} \in F) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}^L_{i,\cdot\cdot} \in F\right)\right|,$$

$$\hat{\rho}^R_n(\mathcal{F}^{\mathrm{re}}) = \sup_{F \in \mathcal{F}^{\mathrm{re}}} \left| \Pr(\sqrt{n}\widehat{\mathbf{B}} \in F) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}^R_{i,\cdot\cdot} \in F\right)\right|.$$

Bounding these errors $\hat{\rho}^L_n(\mathcal{F}^{\mathrm{re}})$ and $\hat{\rho}^R_n(\mathcal{F}^{\mathrm{re}})$ relies on the statistical accuracy of precision matrices. Several algorithms have been proposed to estimate the precision matrices for matrix and tensor normal distributions (Leng & Tang 2012, Xu et al. 2017, Min et al. 2022). Our analysis requires control over the estimation error in the matrix L_1 norm. These L_1 error bounds can be derived from element-wise maximum norm bounds, which is available for estimators like tensor Lasso (Lyu et al. 2019). The tensor Lasso estimator achieves maximum norm rates of $O_P(\sqrt{\log(p)/(nq)})$ and $O_P(\sqrt{\log(q)/(np)})$ for $\widehat{\Omega}$ and $\hat{\Gamma}$, which are optimal according to (Cai et al. 2016). Let s_1 and s_2 denote the maximum number of nonzero elements in any row of Ω and Γ, respectively. The maximum norm rates imply the convergence rate in matrix L_1 norm is $O_P(s_1\sqrt{\log(p)/(nq)})$ and $O_P(s_2\sqrt{\log(q)/(np)})$ (Corollary C.2 of the Supplement). These rates are key assumptions for the validity of Theorem A.1 below.

Theorem A.1. *Assume Assumptions B.1-B.3 in Section B and* $\|\widehat{\Omega} - \Omega\|_1 = O_P(s_1\sqrt{\log(p)/(nq)}), \|\hat{\Gamma} - \Gamma\|_1 = O_P(s_2\sqrt{\log(q)/(np)})$. *If* $n > C_1 \log^7(pqn)$ *and* $s_1 pq^{1/2} \vee s_2 p^{1/2} q \lesssim n^{1/3} \log^{1/6}(pqn)$, *it holds that*

$$\hat{\rho}_n^L(\mathcal{F}^{\mathrm{re}}), \hat{\rho}_n^R(\mathcal{F}^{\mathrm{re}}) \le C D_{n,pq} + \frac{1}{pq},$$

where $D_{n,pq} = \left(\log^7(pqn)/n\right)^{1/6}$.

Remark A.2. The proof strategy closely follows techniques developed in the seminal work (Zhao & Xing 2022). From this theorem, estimators $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ can be approximated by Gaussian random matrices that preserve mean and covariance structure. The convergence rates of $\widehat{\Omega}$ and $\hat{\Gamma}$ in matrix L_1 norm decide the requirement on n, p, and q such that the result holds. Notably, the condition $(s_1 pq^{1/2} \vee s_2 p^{1/2} q) \lesssim n^{1/3} \log^{1/6}(pqn)$ is substantially weaker than condition $s_1 s_2 pq \lesssim n^{1/3} \log^{1/6}(pqn)$ that would arise if $\mathbf{X}$ were vectorized (Zhao & Xing 2022), highlighting a key advantage of matrix-based method. The reason is the optimal rate in L_1 norm for the precision matrix associated with $\mathrm{vec}(\mathbf{X})$ is $O_P(s_1 s_2 \sqrt{\log(pq)/n})$, which is slower than the rate achieved by separately estimating Ω and Γ.

B Technical assumptions

Assumption B.1. The response variable Y is a sub-Gaussian random variable.

Assumption B.2. There exists positive constant c and C such that the smallest eigenvalue and the largest eigenvalue of $\mathbf{U}$ and $\mathbf{V}$ satisfy $c < \lambda_{\min}(\mathbf{U}) \leq \lambda_{\max}(\mathbf{U}) < C$ and $c < \lambda_{\min}(\mathbf{V}) \leq \lambda_{\max}(\mathbf{V}) < C$.

Assumption B.3. There exists positive constants c and C such that $c < \lambda_{\min}(\text{cov}(Y\mathbf{X}_{\cdot k}, Y\mathbf{X}_{\cdot k})) \leq \lambda_{\max}(\text{cov}(Y\mathbf{X}_{\cdot k}, Y\mathbf{X}_{\cdot k})) < C$ for $k = 1, \ldots q$, and $c < \lambda_{\min}(\text{cov}(Y\mathbf{X}_{j\cdot}, Y\mathbf{X}_{j,\cdot})) \leq \lambda_{\max}(\text{cov}(Y\mathbf{X}_{j\cdot}, Y\mathbf{X}_{j\cdot}))$, $j = 1, \ldots, p$.

The sub-Gaussianity of Y in Assumption B.1 is a standard assumption in high-dimensional statistics for deriving concentration inequalities (Wainwright 2019). Assumption B.2 imposes bounds on the eigenvalues, which are commonly assumed in the estimation of precision and covariance matrices and are widely used in the hypothesis testing literature (Dai et al. 2023*b*, Xing et al. 2023, Zhao & Xing 2022). Assumption B.3 requires a sufficient variation for each row and column of $Y\mathbf{X}$, similar to the distinguishable partition condition in (Zhao & Xing 2022), which guarantees that each slice has adequate variation when applying SIR.

Let θ_j^L (or θ_k^R) be the indicator of whether the hypothesis $\mathcal{H}_j^L$ (or $\mathcal{H}_k^R$) is false. The index set of true null hypotheses for rows is $\mathcal{S}_0^L = \{j : \theta_j^L = 0\}$, and similarly $\mathcal{S}_0^R = \{k : \theta_k^R = 0\}$ for columns. Note that $\mathcal{S}_0^L = \mathcal{I}_L^c$ and $\mathcal{S}_0^R = \mathcal{I}_R^c$, we use $\mathcal{S}_0^L$ and $\mathcal{S}_0^R$ here for ease of representation. Then $\mathcal{S}_1^L$ denotes the complement of $\mathcal{S}_0^L$, and similarly for $\mathcal{S}_1^R$.

Assumption B.4. (a) Let $\mathbf{U}^{jj}$ and $\mathbf{V}^{kk}$ be the covariance matrix of $\mathbf{z}_{i,j\cdot}^L$ and $\mathbf{z}_{i,k\cdot}^R$; $\mathbf{U}^{jj'}$ and $\mathbf{V}^{kk'}$ be the covariance matrix between $\mathbf{z}_{i,j\cdot}^L$ and $\mathbf{z}_{i,j'\cdot}^L$, and $\mathbf{z}_{i,k\cdot}^R$ and $\mathbf{z}_{i,k'\cdot}^R$. We assume that

$$\sum_{j,j'\in\mathcal{S}_0^L} \lambda_{\max}^{1/2}\left((\mathbf{U}^{jj})^{-1/2}\mathbf{U}^{jj'}(\mathbf{U}^{j'j'})^{-1}\mathbf{U}^{j'j}(\mathbf{U}^{jj})^{-1/2}\right) \le C_1 p_0^{\alpha_1},$$
$$\sum_{k,k'\in\mathcal{S}_0^R} \lambda_{\max}^{1/2}\left((\mathbf{V}^{kk})^{-1/2}\mathbf{V}^{kk'}(\mathbf{V}^{k'k'})^{-1}\mathbf{V}^{k'k}(\mathbf{V}^{kk})^{-1/2}\right) \le C_2 q_0^{\alpha_2}$$

holds for positive constants C_1, C_2 and $\alpha_1, \alpha_2 \in (0,2)$. (b) Let $\rho(\cdot,\cdot)$ denote the correlation coefficient of two random variables. We assume that

$$\sum_{j,j'\in\mathcal{S}_0^L}\sum_{k,k'=1}^{q} |\rho(z_{i,jk}^L, z_{i,j'k}^L)|^2 < C_3 p_0^{\alpha_3}, \quad \sum_{k,k'\in\mathcal{S}_0^R}\sum_{j,j'=1}^{p} |\rho(z_{i,kj}^R, z_{i,k'j}^R)|^2 < C_4 q_0^{\alpha_4}$$

holds for positive constants C_3, C_4 and $\alpha_3, \alpha_4 \in (0,2)$.

Assumption B.4 (a) bounds the pairwise canonical correlation between two sets of random vectors, similar to the assumption used by (Zhao & Xing 2022) for vector predictors. Assumption B.4 (b) ensures the sum of pairwise correlations within each row of the two sets of normal random variables, $z_{i,jk}^L$ and $z_{i,kj}^R$, remains bounded. Similar conditions are used in both linear (Xing et al. 2023) and non-linear models (Zhao & Xing 2022), which are designed for vector predictors. However, we consider a more general non-linear model with matrix predictors, where the contribution of j-th row (or k-th column) is associated with a q (or p) dimensional vector. These weak dependence conditions are widely used in the literature to study the asymptotic property of FDR.

Assumption B.5. (a) For any $\epsilon > 0$ and t, the probabilities of the following events goes to 1

$$\frac{1}{p_1}\sum_{j\in\mathcal{S}_1^L} \mathbb{1}(u_j < -t) - \frac{1}{p_0}\sum_{j'\in\mathcal{S}_0^L} \mathbb{1}(u_{j'} < -t) \le \epsilon,$$
$$\frac{1}{q_1}\sum_{k\in\mathcal{S}_1^R} \mathbb{1}(v_k < -t) - \frac{1}{q_0}\sum_{k'\in\mathcal{S}_0^R} \mathbb{1}(v_{k'} < -t) \le \epsilon.$$

(b) As $n, p, q \to \infty$, we have $p_0, q_0 \to \infty$ and $\liminf p_{1,\text{strong}}/p_0, \liminf q_{1,\text{strong}}/q_0 > 0$.

Assumption B.5 (a) states the proportions of negative row and column statistics in $\mathcal{S}_1^L$ and $\mathcal{S}_1^R$ are smaller than the proportions in $\mathcal{S}_0^L$ and $\mathcal{S}_0^R$. This assumption is mild, as u_j and v_k are asymptotically symmetric about zero under the null hypothesis and have positive means under the alternative hypothesis. Assumption B.5 (b) requires the number of rows and columns with strong signals to grow to infinity with p_0 and q_0, similar to the condition used in Dai et al. (2023*b*) for generalized linear models. A key proof argument is $\Pr(\tau_a \le t_a) \to 1$ and $\Pr(\tau_b \le t_b) \to 1$, where t_a and t_b satisfies $G_0^L(t_a) = \frac{ap_{1,\text{strong}}}{(1-a)p_0+p_1}$ and $G_0^R(t_b) = \frac{bq_{1,\text{strong}}}{(1-b)q_0+q_1}$. Assumption B.5 (b) ensures t_a and t_b are bounded as $n, p, q \to \infty$.

C Additional theoretical results

The normal random variables $z_{i,kj}^R$ in Section A have the following mean and covariance structure:

$$\begin{aligned}
&\mathrm{E}[z_{i,kj}^R] = \Gamma_{\cdot k}^T \mathrm{E}[Y\mathbf{X}_{j\cdot}^T],\\
&\mathrm{var}(z_{i,kj}^R) = \Gamma_{\cdot k}^T \mathrm{E}[Y^2\mathbf{X}_{j\cdot}^T\mathbf{X}_{j\cdot}]\Gamma_{\cdot k} - \Gamma_{\cdot k}^T \mathrm{E}[Y\mathbf{X}_{j\cdot}^T]\,\mathrm{E}[Y\mathbf{X}_{j\cdot}]\Gamma_{\cdot k},\\
&\mathrm{cov}(z_{i,kj}^R, z_{i,lm}^R) = \Gamma_{\cdot k}^T \mathrm{E}[Y^2\mathbf{X}_{j\cdot}^T\mathbf{X}_{m\cdot}]\Gamma_{\cdot l} - \Gamma_{\cdot k}^T \mathrm{E}[Y\mathbf{X}_{j\cdot}^T]\,\mathrm{E}[Y\mathbf{X}_{m\cdot}]\Gamma_{\cdot l}.
\end{aligned}$$

Let $\mathbf{A} = \Omega\widehat{\mathrm{E}}[Y\mathbf{X}]$ and $\mathbf{B} = \Gamma\widehat{\mathrm{E}}[Y\mathbf{X}^T]$ denote estimators of FSS using true precision matrices. Define

$$\begin{aligned}
\rho_n^L(\mathcal{F}^{\mathrm{re}}) &= \sup_{F\in\mathcal{F}^{\mathrm{re}}} \left| \Pr(\sqrt{n}\mathbf{A} \in F) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,\cdot\cdot}^L \in F\right)\right|,\\
\rho_n^R(\mathcal{F}^{\mathrm{re}}) &= \sup_{F\in\mathcal{F}^{\mathrm{re}}} \left| \Pr(\sqrt{n}\mathbf{B} \in F) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,\cdot\cdot}^R \in F\right)\right|.
\end{aligned}$$

Lemma C.1. *Assume Assumptions B.1-B.3 hold. Then,*

$$\rho_n^L(\mathcal{F}^{\mathrm{re}}), \rho_n^R(\mathcal{F}^{\mathrm{re}}) \le CD_{n,pq},$$

where $D_{n,pq} = \left(\frac{\log^7(pqn)}{n}\right)^{1/6}$.

Lemma C.1 is the central limit theory when using the population precision matrices Ω and Γ.

Corollary C.2. *Suppose the same conditions of Theorem 3.9 in (Lyu et al. 2019) hold. Then, the Tlasso estimator satisfies*

$$\|\widehat{\Omega}-\Omega\|_1 = O_P\left(s_1\sqrt{\frac{\log(p)}{nq}}\right), \quad \|\hat{\Gamma}-\Gamma\|_1 = O_P\left(s_2\sqrt{\frac{\log(q)}{np}}\right).$$

We define some quantities

$$V^L(t) = \frac{\#\left\{j : j \in \mathcal{S}_0^L, u_j \le -t\right\}}{p_0}, \quad V'^L(t) = \frac{\#\left\{j : j \in \mathcal{S}_0^L, u_j \ge t\right\}}{p_0},$$
$$V_1^L(t) = \frac{\#\left\{j : j \in \mathcal{S}_1^L, u_j \ge t\right\}}{p_1}, \quad G_0^L(t) = \lim_p \frac{\sum_{j\in\mathcal{S}_0^L} \mathrm{E}[\mathbb{1}\left(u_j \le -t\right)]}{p_0},$$

and

$$\mathrm{FDP}^L(t) := \frac{V'^L(t)}{(V'^L(t)+r_pV_1^L(t))\vee 1/p}, \quad \mathrm{FDP}^L_\infty(t) := \lim_p \frac{V'^L(t)}{(V'^L(t)+r_pV_1^L(t))\vee 1/p},$$
$$\mathrm{FDP}^L_\dagger(t) = \frac{V^L(t)}{(V'^L(t)+r_pV_1^L(t))\vee 1/p}.$$

where $r_p = p_1/p_0$ and $\mathrm{FDP}^L_\infty(t)$ is the pointwise limit of $\mathrm{FDP}^L(t)$. We can similarly define $\mathrm{FDP}^R(t)$, $\mathrm{FDP}^R_\dagger(t)$ and $\mathrm{FDP}^R_\infty(t)$ for the column hypothesis.

Lemma C.3. *If Assumption B.1-B.4 holds and $G_0(t)$ is a continuous functions, we have*

$$\sup_t |V^L(t)-G_0^L(t)| \xrightarrow{P} 0, \quad \sup_t |V'^L(t)-G_0^L(t)| \xrightarrow{P} 0,$$
$$\sup_t |V^R(t)-G_0^R(t)| \xrightarrow{P} 0, \quad \sup_t |V'^R(t)-G_0^R(t)| \xrightarrow{P} 0.$$

Lemma C.4. *If Assumption B.1-B.4 holds and $s_1\log(p)\vee s_2\log(q) = o(\sqrt{n})$, then with* *probability at least* $1 - 1/(pq)$, *for any* t

$$\sum_{j,j' \in \mathcal{S}_0^L} \mathrm{cov}(\mathbb{1}(u_j \geq t), \mathbb{1}(u_{j'} \geq t)) \leq C|\mathcal{S}_0^L|^{\alpha_1},$$

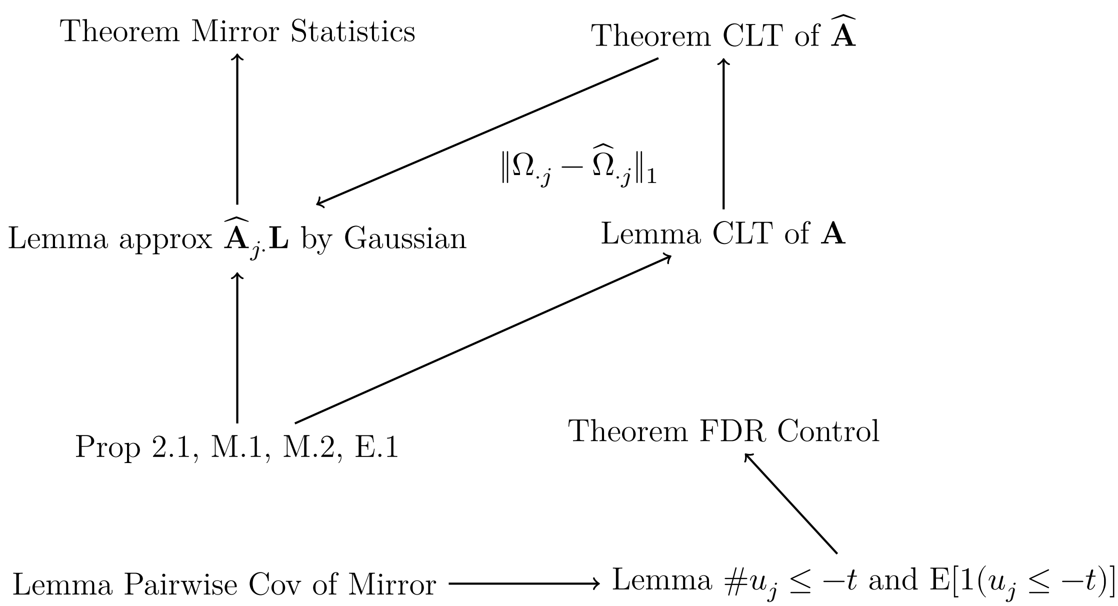
$$\sum_{k,k' \in \mathcal{S}_0^R} \mathrm{cov}(\mathbb{1}(v_k \geq t), \mathbb{1}(v_{k'} \geq t)) \leq C|\mathcal{S}_0^R|^{\alpha_1},$$

where $\alpha_1 \in (0, 2)$.

In Lemma C.4, we bound the sum of pairwise covariances of mirror statistics, which is a key component to establish asymptotic FDR control.

D Proofs

The following diagram shows the relationship between Lemmas and Theorems.



In the proof, we focus on the row statistics u_j and omit the superscript L when possible for simplicity. The results for column statistics v_k follow analogously and are omitted for brevity.

Proof of Lemma 2.2

Proof. Let $\mathbf{x} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}, \tilde{\mathbf{x}} = \mathbf{A}\mathbf{x}$. We have

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{x})}{\partial f(\tilde{\mathbf{x}})} \frac{\partial f(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}} \frac{\partial \tilde{\mathbf{x}}}{\partial \mathbf{x}} = 1 \cdot \mathbf{A}^T \frac{\partial f(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}} = \mathbf{A}^T \frac{\partial f(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}}.$$

Since $\mathbf{X} \sim \mathcal{MN}_{p,q}(\mathbf{0}, \mathbf{U}, \mathbf{V})$, $\operatorname{vec}(\mathbf{X}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V} \otimes \mathbf{U})$. Then we can define $\mathbf{z} = (\mathbf{V} \otimes \mathbf{U})^{-1/2} \operatorname{vec}(\mathbf{X})$. Using the relation above and Lemma E.1 and let $f(\mathbf{x}) = \mathrm{E}[Y|\mathbf{x}]$,

$$\begin{aligned}\mathrm{E}[\nabla f(\operatorname{vec}(\mathbf{X}))] = \mathrm{E}[(\mathbf{V} \otimes \mathbf{U})^{-1/2} \nabla f(\mathbf{z})] &= (\mathbf{V} \otimes \mathbf{U})^{-1/2}\, \mathrm{E}[\mathbf{z} f(\mathbf{z})] \\ &= (\mathbf{V} \otimes \mathbf{U})^{-1/2}\, \mathrm{E}[Y\mathbf{z}] = (\mathbf{V} \otimes \mathbf{U})^{-1}\, \mathrm{E}[Y \operatorname{vec}(\mathbf{X})].\end{aligned}$$

Using property of vec,

$$\mathrm{E}\left[\frac{\partial\, \mathrm{E}[Y|\mathbf{X}]}{\partial \operatorname{vec}(\mathbf{X})}\right] = (\mathbf{V} \otimes \mathbf{U})^{-1}\, \mathrm{E}[Y \operatorname{vec}(\mathbf{X})] \iff \mathrm{E}\left[\frac{\partial\, \mathrm{E}[Y|\mathbf{X}]}{\partial \mathbf{X}}\right] = \mathbf{U}^{-1}\, \mathrm{E}[Y\mathbf{X}]\mathbf{V}^{-1}. \quad (1)$$

Note that $Y|\mathbf{X} \sim Y|\mathbf{A}^T\mathbf{X}, Y|\mathbf{X} \sim Y|\mathbf{X}\mathbf{B}$. Thus

$$\begin{aligned}\operatorname{vec}\left(\frac{\partial\, \mathrm{E}[Y|\mathbf{X}]}{\partial \mathbf{X}}\right) &= \frac{\partial\, \mathrm{E}[Y|\operatorname{vec}(\mathbf{A}^T\mathbf{X})]}{\partial \operatorname{vec}(\mathbf{X})} = \frac{\partial\, \mathrm{E}[Y|\operatorname{vec}(\mathbf{A}^T\mathbf{X})]}{\partial \operatorname{vec}(\mathbf{A}^T\mathbf{X})} \frac{\partial \operatorname{vec}(\mathbf{A}^T\mathbf{X})}{\partial \operatorname{vec}(\mathbf{X})} \\ &= (\mathbf{I}_q \otimes \mathbf{A}) \frac{\partial\, \mathrm{E}[Y|\operatorname{vec}(\mathbf{A}^T\mathbf{X})]}{\partial \operatorname{vec}(\mathbf{A}^T\mathbf{X})} \\ &= \frac{\partial\, \mathrm{E}[Y|\operatorname{vec}(\mathbf{X}\mathbf{B})]}{\partial \operatorname{vec}(\mathbf{X})} = \frac{\partial\, \mathrm{E}[Y|\operatorname{vec}(\mathbf{X}\mathbf{B})]}{\partial \operatorname{vec}(\mathbf{X}\mathbf{B})} \frac{\partial \operatorname{vec}(\mathbf{X}\mathbf{B})}{\partial \operatorname{vec}(\mathbf{X})} \\ &= (\mathbf{B} \otimes \mathbf{I}_p) \frac{\partial\, \mathrm{E}[Y|\operatorname{vec}(\mathbf{X}\mathbf{B})]}{\partial \operatorname{vec}(\mathbf{X}\mathbf{B})}.\end{aligned}$$

Using the property of the vec operator,

$$\frac{\partial\, \mathrm{E}[Y|\mathbf{X}]}{\partial \mathbf{X}} = \mathbf{A} \frac{\partial\, \mathrm{E}[Y|\mathbf{A}^T\mathbf{X}]}{\partial \mathbf{A}^T\mathbf{X}} = \frac{\partial\, \mathrm{E}[Y|\mathbf{X}\mathbf{B}]}{\partial \mathbf{X}\mathbf{B}} \mathbf{B}^T$$

Take expectations of both sides,

$$\begin{aligned}\mathrm{E}\left[\frac{\partial\, \mathrm{E}[Y|\mathbf{X}]}{\partial \mathbf{X}}\right] &= \mathbf{A}\, \mathrm{E}\left[\frac{\partial\, \mathrm{E}[Y|\mathbf{A}^T\mathbf{X}]}{\partial \mathbf{A}^T\mathbf{X}}\right] \Rightarrow \operatorname{span}\left(\mathrm{E}\left[\frac{\partial\, \mathrm{E}[Y|\mathbf{X}]}{\partial \mathbf{X}}\right]\right) \subseteq \operatorname{span}(\mathbf{A}), \\ \mathrm{E}\left[\frac{\partial\, \mathrm{E}[Y|\mathbf{X}]}{\partial \mathbf{X}}\right]^T &= \mathbf{B}\, \mathrm{E}\left[\frac{\partial\, \mathrm{E}[Y|\mathbf{X}\mathbf{B}]}{\partial \mathbf{X}\mathbf{B}}\right]^T \Rightarrow \operatorname{span}\left(\mathrm{E}\left[\frac{\partial\, \mathrm{E}[Y|\mathbf{X}]}{\partial \mathbf{X}}\right]^T\right) \subseteq \operatorname{span}(\mathbf{B}).\end{aligned}$$

Using (1), we have

$$\operatorname{span}(\mathbf{U}^{-1}\, \mathrm{E}[Y\mathbf{X}]\mathbf{V}^{-1}) = \operatorname{span}(\mathbf{U}^{-1}\, \mathrm{E}[Y\mathbf{X}]) \subseteq \operatorname{span}(\mathbf{A}),$$

$$\operatorname{span}(\mathbf{V}^{-1}\, \mathrm{E}[Y\mathbf{X}]^T\mathbf{U}^{-1}) = \operatorname{span}(\mathbf{V}^{-1}\, \mathrm{E}[Y\mathbf{X}]^T) \subseteq \operatorname{span}(\mathbf{B}).$$

□

Proof of Lemma C.1

Proof. By definition, $\mathbf{A}_{\cdot k} = \frac{1}{n}\sum_{i=1}^{n} \Omega \mathbf{X}_{i,\cdot k} Y_i$, where $\widetilde{\mathbf{A}} = (\widetilde{\mathbf{A}}_{\cdot 1}, \dots, \widetilde{\mathbf{A}}_{\cdot q})$. Moreover, we have $\mathbf{A}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \Omega_{j\cdot} \mathbf{X}_{i,\cdot k} Y_i$. Define $\xi_{i,jk} = \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} Y_i$. For given j and k, it is easy to see $\xi_{i,jk}$'s are independent. As a result, we have

$$\mathbf{A}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \xi_{i,jk}, \quad \sqrt{n}\mathbf{A}_{jk} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_{i,jk}.$$

According to Lemma E.2, the mean and covariance structure of $\xi_{i,jk}$ are the same as those of $z_{i,jk}$. According to Proposition 2.1 in (Chernozhukov et al. 2017), it suffices to show that the conditions (M.1), (M.2), and (E.1) hold.

First, we prove the condition (M1), which states

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(\xi_{i,jk}^2) \geq c, \text{ for all } j = 1, \dots, p, k = 1, \dots, q.$$

Note that

$$\begin{aligned}
\mathrm{E}(\xi_{i,jk}^2) &= \Omega_{\cdot j}^T \mathrm{E}(Y^2 \mathbf{X}_{\cdot k}\mathbf{X}_{\cdot k}^T)\Omega_{\cdot j} \\
&\geq \Omega_{\cdot j}^T \mathrm{cov}(Y\mathbf{X}_{\cdot k}, Y\mathbf{X}_{\cdot k})\Omega_{\cdot j} \\
&\geq \|\Omega_{\cdot j}\|_2^2 \lambda_{\min}(\mathrm{E}[Y^2\mathbf{X}_{\cdot k}\mathbf{X}_{\cdot k}^T]) \geq c',
\end{aligned}$$

where $c' > 0$ and the last inequality uses condition (C3).

Next, we consider the condition (E.1) which states that

$$\mathrm{E}[\exp(|\xi_{i,jk}|/C_n)] \leq 2, \text{ for all } j = 1, \dots, p, k = 1, \dots, q.$$

It is suffices to show $\mathrm{E}[\exp(|\xi_{i,jk}|/C_n)]$ is bounded. Note that

$$\begin{aligned}
\mathrm{E}[\exp(|\xi_{i,jk}|/C_n)] &= 1 + \sum_{h=1}^{\infty} \frac{\mathrm{E}(|\xi_{i,jk}|^h)}{C_n^h h!} = 1 + \sum_{h=1}^{\infty} \frac{\mathrm{E}(|\Omega_{\cdot j}^T \mathbf{X}_{i,\cdot k} Y_i|^h)}{C_n^h h!} \\
&\leq 1 + \sum_{h=1}^{\infty} \frac{\mathrm{E}(|\Omega_{\cdot j}^T \mathbf{X}_{i,\cdot k}|^h |Y_i|^h)}{C_n^h h!} \\
&\leq 1 + \sum_{h=1}^{\infty} \frac{\mathrm{E}(|\Omega_{\cdot j}^T \mathbf{X}_{i,\cdot k}|^{2h})^{1/2}\, \mathrm{E}(|Y_i|^{2h})^{1/2}}{C_n^h h!},
\end{aligned}$$

where the first equality uses $\exp(x) = \sum_{h=0}^{\infty} x^h/h!$ and the second inequality uses Cauchy–Schwarz inequality. Using the fact that $\|\Omega_{\cdot j}\|_2 = \|\Omega \mathbf{e}_j\|_2 \leq \|\Omega\|_2 \leq C$ (condition C2), $\mathrm{E}(|\Omega_{\cdot j}^T \mathbf{X}_{i,\cdot k}|^{2h}) \leq C^{2h}\,\mathrm{E}(|\Omega_{\cdot j}^T \mathbf{X}_{i,\cdot k}/\|\Omega_{\cdot j}\|_2|^{2h})$. Since $\mathbf{X}_{\cdot k} \sim N(\mathbf{0}, \mathbf{V}_{kk}\mathbf{U})$, $\mathbf{X}_{\cdot k}$ is a sub-Gaussian random vector with parameter $\mathbf{V}_{kk}\|\mathbf{U}\|_2$. Using bounds on moment of sub-Gaussian random variables, $\mathrm{E}(|\Omega_{\cdot j}^T \mathbf{X}_{i,\cdot k}/\|\Omega_{\cdot j}\|_2|^{2h}) \leq C_1^h (2h)^h$ and $\mathrm{E}(|Y|^{2h}) \leq C_2^h (2h)^h$ (Assumption B.1). Therefore, we have

$$\mathrm{E}[\exp(|\xi_{i,jk}|/C_n)] \leq 1 + \sum_{h=1}^{\infty} \frac{C^h h^h}{C_n^h h!} \leq 1 + \sum_{h=1}^{\infty} \frac{h^h}{C_n^h h!} \leq 1 + \sum_{h=1}^{\infty} (e/C_n)^h < \infty,$$

where the third inequality holds because $h! \geq (h/e)^h$ and C_n is set to be greater than Ce.

Last, we consider the condition (M.2) which requires that

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(|\xi_{i,jk}|^{2+l}) \leq C_n^l, \text{ for all } j = 1, \ldots, p, k = 1, \ldots, q, l = 1, 2.$$

Note that

$$\sum_{l=1}^{2} \frac{\mathrm{E}(|\xi_{i,jk}|^{2+l})}{C^{l+2}(l+2)!} \leq \mathrm{E}[\exp(|\xi_{i,jk}|/C)] < \infty,$$

where the last inequality holds due to (E.1). We can choose sufficiently large C such that

$$\mathrm{E}(|\xi_{i,jk}|^{2+l}) \leq C, \text{ for all } j = 1, \ldots, p, k = 1, \ldots, q, l = 1, 2.$$

This implies (M.2) holds. Therefore, according to Proposition 2.1 in (Chernozhukov et al. 2017), we have

$$\rho_n(\mathcal{F}^{\mathrm{re}}) \leq C D_{n,pq},$$

where $D_{n,pq} = \left(\frac{\log^7(pqn)}{n}\right)^{1/6}$. □

Proof of Corollary C.2

Proof. This is a direct consequence of Theorem 3.9 in (Lyu et al. 2019) which states that

$$\|\widehat{\Omega} - \Omega\|_{\max} = O_P\left(\sqrt{\frac{\log(p)}{nq}}\right).$$

Since s_1 is the number of the maximum number of non-zeros in any column of Ω, we immediately have

$$\|\widehat{\Omega} - \Omega\|_1 = O_P\left(s_1\sqrt{\frac{\log(p)}{nq}}\right),$$

according to Lemma 7.1 in (Cai et al. 2016). □

Proof of Theorem A.1

Proof. Let $\widehat{\mathbf{A}}_{\cdot k} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\Omega}\mathbf{X}_{i,\cdot k}Y_i$ and $\widehat{\mathbf{B}}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Gamma}\mathbf{X}_{i,j\cdot}^T Y_i$. Moreover, we have $\widehat{\mathbf{A}}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\Omega}_{j\cdot}\mathbf{X}_{i,\cdot k}Y_i$. Note that

$$\begin{aligned}
\sqrt{n}|\widehat{\mathbf{A}}_{jk} - \mathbf{A}_{jk}| &= \left|(\widehat{\Omega}_{\cdot j} - \Omega_{\cdot j})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{X}_{i,\cdot k}Y_i\right| \\
&\leq \|\widehat{\Omega}_{\cdot j} - \Omega_{\cdot j}\|_1 \cdot \frac{1}{\sqrt{n}}\left\|\sum_{i=1}^{n}\mathbf{X}_{i,\cdot k}Y_i\right\|_\infty \\
&= \|\widehat{\Omega}_{\cdot j} - \Omega_{\cdot j}\|_1 \cdot \frac{1}{\sqrt{n}}\|\mathbf{X}^{(k)}\mathbf{Y}\|_\infty,
\end{aligned}$$

where $\mathbf{X}^{(k)} = (\mathbf{X}_{1,\cdot k}, \dots, \mathbf{X}_{n,\cdot k}) \in \mathbb{R}^{p\times n}$ and $\mathbf{Y} = (Y_1, \dots, Y_n)^T$. For $i = 1, \dots, p$, $(\mathbf{X}^{(k)}\mathbf{Y})_i = \sum_{j=1}^{n}\mathbf{X}_{ij}^{(k)}Y_j$. Since $\mathbf{X}_{ij}^{(k)}$ is Gaussian and Y_j is sub-Gaussian, $\mathbf{X}_{ij}^{(k)}Y_j$ is sub-exponential. According to Theorem 2.8.2 in (Vershynin 2018), we have

$$\Pr(|(\mathbf{X}^{(k)}\mathbf{Y})_i| > t) = \Pr\left(\sum_{j=1}^{n}\mathbf{X}_{ij}^{(k)}Y_j > t\right) \leq 2\exp\left[-c\min\left(\frac{t^2}{nC^2}, \frac{t}{C}\right)\right]$$

for $t \geq 0$. Using union bound,

$$\Pr(\|\mathbf{X}^{(k)}\mathbf{Y}\|_\infty > t) = \Pr(\max_{1\leq i\leq p}|(\mathbf{X}^{(k)}\mathbf{Y})_i| > t) \leq 2p\exp\left[-c\min\left(\frac{t^2}{nC^2}, \frac{t}{C}\right)\right].$$

With probability at least $1 - 1/(pq)$, we have

$$\|\mathbf{X}^{(k)}\mathbf{Y}\|_\infty = O(\sqrt{n\log p} \vee \log p).$$

Assume that

$$\|\widehat{\Omega}_{\cdot j} - \Omega_{\cdot j}\|_1 = O_P\left(s_1 q^{-1/2}\sqrt{\frac{\log(p)}{n}}\right).$$

According to (Lyu et al. 2019), their proposed estimator has such rate if $s_1 \log(p) = o(\sqrt{nqp^{-1}})$. Therefore,

$$\sqrt{n}|\widehat{\mathbf{A}}_{jk} - \mathbf{A}_{jk}| = O_P\left(\underbrace{s_1 q^{-1/2}\left\{\frac{\log(p)}{\sqrt{n}} \vee \frac{\log(p)}{\sqrt{n}}\sqrt{\frac{\log(p)}{n}}\right\}}_{\epsilon}\right).$$

Let $\mathcal{E} = \{\|\sqrt{n}\widehat{\mathbf{A}} - \mathbf{A}\|_\infty \leq \epsilon\}$, $\Pr(\mathcal{E}) \geq 1 - 1/(pq)$. For any $F = \{\omega \in \mathbb{R}^{pq} : a_j \leq \omega_j \leq b_j\} \in \mathscr{F}^{\mathrm{re}}$, we define

$$F^{\epsilon+} = \{\omega \in \mathbb{R}^{pq} : a_j - \epsilon \leq \omega_j \leq b_j + \epsilon\}, \quad F^{\epsilon-} = \{\omega \in \mathbb{R}^{pq} : a_j + \epsilon \leq \omega_j \leq b_j - \epsilon\}.$$

Note that

$$\begin{aligned}\Pr(\sqrt{n}\widehat{\mathbf{A}} \in F) &\leq \Pr(\sqrt{n}\mathbf{A} + \sqrt{n}\widehat{\mathbf{A}} - \sqrt{n}\mathbf{A} \in F \mid \mathcal{E}) \Pr(\mathcal{E}) + \Pr(\mathcal{E}^c)\\ &\leq \Pr(\sqrt{n}\mathbf{A} \in F^{\epsilon+}) + \frac{1}{pq}.\end{aligned}$$

Then, we have

$$\begin{aligned}&\Pr(\sqrt{n}\widehat{\mathbf{A}} \in F) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,\cdot} \in F\right)\\ &\leq \Pr(\sqrt{n}\mathbf{A} \in F^{\epsilon+}) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,\cdot} \in F\right) + \frac{1}{pq}.\end{aligned}$$

According to Lemma C.1, we have

$$\Pr(\sqrt{n}\mathbf{A} \in F^{\epsilon+}) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,\cdot} \in F^{\epsilon+}\right) \leq CD_{n,pq}.$$

By the mean value theorem,

$$\left|\Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,\cdot} \in F^{\epsilon+}\right) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,\cdot} \in F\right)\right| \leq O(pq\epsilon) = O(s_1 pq^{1/2}\log(p)/\sqrt{n}),$$

since $n > C\log^7(pqn)$. Therefore, it holds that

$$\Pr(\sqrt{n}\widehat{\mathbf{A}} \in F) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,\cdot} \in F\right) \leq CD_{n,pq} + \frac{1}{pq} \tag{2}$$

as long as $s_1 pq^{1/2}\log(p)/\sqrt{n} = O(D_{n,pq})$. By the definition of $D_{n,pq}$, we require $s_1 pq^{1/2} = O(n^{1/3}\log^{1/6}(pqn))$. Note that if this relationship holds, $s_1\log(p) = o(\sqrt{nqp^{-1}})$ satisfies as well.

Similarly, we have

$$\begin{aligned}
&\Pr(\sqrt{n}\widehat{\mathbf{A}} \in F) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_{i,\cdot\cdot} \in F\right) \\
&\geq \Pr(\sqrt{n}\mathbf{A} \in F^{\epsilon-}) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_{i,\cdot\cdot} \in F\right) - \frac{1}{pq} \geq -CD_{n,pq} - \frac{1}{pq}. \qquad (3)
\end{aligned}$$

Combining (2) and (3), we know that $\hat{\rho}_n(\mathscr{F}^{\mathrm{re}}) \leq CD_{n,pq} + \frac{1}{pq}$.

$\square$

Proof of Theorem 3.1

Proof. Note that

$$\begin{aligned}
\Pr(u_j < -t) &= \Pr\left(\sum_{l=1}^{q}\sqrt{n/2}\widehat{\mathbf{A}}_{jl}^{(1)}\widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{n/2}\right) \\
&= \mathrm{E}_{\widehat{\mathbf{A}}_{j\cdot}^{(2)}}\left[\Pr\left(\sum_{l=1}^{q}\sqrt{n/2}\widehat{\mathbf{A}}_{jl}^{(1)}\widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{n/2} \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right)\right].
\end{aligned}$$

According to Lemma E.3 and the independence of $\widehat{\mathbf{A}}_{j\cdot}^{(1)}$ and $\widehat{\mathbf{A}}_{j\cdot}^{(2)}$, it holds that

$$\sup_t\left|\Pr\left(\sum_{l=1}^{q}\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{jl}^{(1)}\widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{\frac{n}{2}}\right) - \Pr\left(\sum_{l=1}^{q}\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{I}_1}z_{i,jl}^{(1)}\widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{\frac{n}{2}}\right)\right| \leq C\frac{s_1\log(p)}{\sqrt{nq}},$$

when conditioning on $\widehat{\mathbf{A}}_{jl}^{(2)}$. Under the null hypothesis, $z_{i,jl}$ are normally distributed with zero means. Then,

$$\Pr\left(\sum_{l=1}^{q}\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{I}_1}z_{i,jl}^{(1)}\widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{\frac{n}{2}}\right) = \Pr\left(\sum_{l=1}^{q}\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{I}_1}z_{i,jl}^{(1)}\widehat{\mathbf{A}}_{jl}^{(2)} > t\sqrt{\frac{n}{2}}\right),$$

which implies

$$\sup_t\left|\Pr\left(\sum_{l=1}^{q}\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{jl}^{(1)}\widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{\frac{n}{2}}\right) - \Pr\left(\sum_{l=1}^{q}\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{I}_1}z_{i,jl}^{(1)}\widehat{\mathbf{A}}_{jl}^{(2)} > t\sqrt{\frac{n}{2}}\right)\right| \leq C\frac{s_1\log(p)}{\sqrt{nq}}. \qquad (4)$$

Note that

$$
\begin{aligned}
&\sup_t \left|\Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} > t\sqrt{\frac{n}{2}}\right) - \Pr\left(\sum_{l=1}^q \sqrt{\frac{2}{n}} \sum_{i \in \mathcal{I}_1} z_{i,jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} > t\sqrt{\frac{n}{2}}\right)\right| \\
&= \sup_t \left|1 - \Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} < t\sqrt{\frac{n}{2}}\right) - 1 + \Pr\left(\sum_{l=1}^q \sqrt{\frac{2}{n}} \sum_{i \in \mathcal{I}_1} z_{i,jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} < t\sqrt{\frac{n}{2}}\right)\right| \\
&= \sup_t \left|\Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} < t\sqrt{\frac{n}{2}}\right) - \Pr\left(\sum_{l=1}^q \sqrt{\frac{2}{n}} \sum_{i \in \mathcal{I}_1} z_{i,jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} < t\sqrt{\frac{n}{2}}\right)\right| \leq C\frac{s_1 \log(p)}{\sqrt{nq}}.
\end{aligned}
$$

Combining with (4), we have

$$
\sup_t \left|\Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{\frac{n}{2}} \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right) - \Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} > t\sqrt{\frac{n}{2}} \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right)\right| \leq C\frac{s_1 \log(p)}{\sqrt{nq}}.
$$

Therefore,

$$
\begin{aligned}
&\sup_t \left|\Pr(u_j < -t) - \Pr(u_j > t)\right| \\
&= \sup_t \left|\Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{\frac{n}{2}}\right) - \Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} > t\sqrt{\frac{n}{2}}\right)\right| \\
&\leq \sup_t \mathrm{E}_{\widehat{\mathbf{A}}_{j\cdot}^{(2)}} \left[\left|\Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{\frac{n}{2}} \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right) - \Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} > t\sqrt{\frac{n}{2}} \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right)\right|\right] \\
&\leq \mathrm{E}_{\widehat{\mathbf{A}}_{j\cdot}^{(2)}} \left[\sup_t \left|\Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} < -t\sqrt{\frac{n}{2}} \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right) - \Pr\left(\sum_{l=1}^q \sqrt{\frac{n}{2}} \widehat{\mathbf{A}}_{jl}^{(1)} \widehat{\mathbf{A}}_{jl}^{(2)} > t\sqrt{\frac{n}{2}} \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right)\right|\right] \\
&\leq C\frac{s_1 \log(p)}{\sqrt{nq}}.
\end{aligned}
$$

□

Proof of Lemma C.3

Proof. Note that

$$
\Pr\left(\sup_{t \in \mathbb{R}} |V(t) - G_0(t)| > \epsilon\right) \leq \Pr\left(\sup_{t \in \mathbb{R}} V(t) - G_0(t) > \epsilon\right) + \Pr\left(\inf_{t \in \mathbb{R}} V(t) - G_0(t) < -\epsilon\right).
$$

For any $\epsilon \in (0,1)$ and $N_\epsilon = \lceil 2/\epsilon \rceil$, let $-\infty = a_0^p < a_1^p < \cdots < a_{N_\epsilon}^p = \infty$ such that $G_0(a_{l-1}^p) - G_0(a_l^p) \leq \epsilon/2$ for $l = 1, \ldots, N_\epsilon$. Such a sequence exists since $G_0(t)$ is continuous with respect to $t \in \mathbb{R}$. Note that

$$
\sup_{t \in \mathbb{R}} V(t) - G_0(t) \leq \max_{l \in [N_\epsilon]} \sup_{t \in [a_{l-1}^p, a_l^p]} V(t) - G_0(t).
$$

Then applying union bound, we have

$$\Pr\left(\sup_{t\in\mathbb{R}} V(t) - G_0(t) > \epsilon\right) \le \Pr\left(\bigcup_{l=1}^{N_\epsilon} \sup_{t\in[a_{l-1}^p, a_l^p)} V(t) - G_0(t) > \epsilon\right)$$
$$\le \sum_{l=1}^{N_\epsilon} \Pr\left(\sup_{t\in[a_{l-1}^p, a_l^p)} V(t) - G_0(t) > \epsilon\right).$$

Since both $V(t)$ and $G_0(t)$ are non-increasing functions, the following holds

$$\sup_{t\in[a_{l-1}^p, a_l^p)} V(t) - G_0(t) \le V(a_{l-1}^p) - G_0(a_l^p) \le V(a_{l-1}^p) - G_0(a_{l-1}^p) + \epsilon/2$$

for any $l \in [N_\epsilon]$. According to Lemma C.4, $\sigma^2 := \mathrm{var}(V(t)) = 1/p_0^2 \sum_{j,j'\in\mathcal{S}_0^L} \mathrm{cov}(\mathbb{1}(u_j \le -t), \mathbb{1}(u_{j'} \le -t)) \le 1/p_0^2 C|\mathcal{S}_0^L|^{\alpha_1}$. Then, according to Chebyshev's inequality,

$$\Pr\left(\sup_{t\in[a_{l-1}^p, a_l^p]} V(t) - G_0(t) > \epsilon\right) \le \Pr\left(V(a_{l-1}^p) - G_0(a_{l-1}^p) > \frac{\epsilon}{2\sigma}\sigma\right)$$
$$\le \Pr\left(|V(a_{l-1}^p) - G_0(a_{l-1}^p)| > \frac{\epsilon}{2\sigma}\sigma\right)$$
$$\le \frac{4\sigma^2}{\epsilon^2} \le \frac{4C|\mathcal{S}_0^L|^{\alpha_1}}{p_0^2\epsilon^2} = \frac{4C}{|\mathcal{S}_0^L|^{2-\alpha_1}\epsilon^2}.$$

It follows that

$$\Pr\left(\sup_{t\in\mathbb{R}} V(t) - G_0(t) > \epsilon\right) \le \frac{4CN_\epsilon}{|\mathcal{S}_0^L|^{2-\alpha_1}\epsilon^2} \to 0 \text{ as } p \to \infty.$$

Similarly, we have

$$\Pr\left(\inf_{t\in\mathbb{R}} V(t) - G_0(t) < -\epsilon\right) \le \Pr\left(\bigcup_{l=1}^{N_\epsilon} \inf_{t\in[a_{l-1}^p, a_l^p)} V(t) - G_0(t) < -\epsilon\right)$$
$$\le \sum_{l=1}^{N_\epsilon} \Pr\left(\inf_{t\in[a_{l-1}^p, a_l^p)} V(t) - G_0(t) < -\epsilon\right)$$
$$\le \sum_{l=1}^{N_\epsilon} \Pr\left(\sup_{t\in[a_{l-1}^p, a_l^p)} G_0(t) - V(t) > \epsilon\right)$$
$$\le \sum_{l=1}^{N_\epsilon} \Pr\left(G_0(a_l^p) - V(a_l^p) > \epsilon/2\right)$$
$$\le \sum_{l=1}^{N_\epsilon} \Pr\left(|V(a_l^p) - G_0(a_l^p)| > \epsilon/2\right)$$
$$\le \frac{4CN_\epsilon}{|\mathcal{S}_0^L|^{2-\alpha_1}\epsilon^2} \to 0 \text{ as } p \to \infty.$$

This concludes the proof that $\sup_t |V(t) - G_0(t)| \xrightarrow{P} 0$. Using the symmetric property of u_j for $j \in \mathcal{S}_0^L$, we can show $\sup_t |V_0'(t) - G_0(t)| \xrightarrow{P} 0$. $\square$

Proof of Lemma C.4

Proof. Let $z_{i,jk}^{(1)}$ and $z_{i,jk}^{(2)}$ be two sets of normal random variables corresponding to the two subsets after the data splitting, $\tilde{u}_j = \frac{4}{n^2} \sum_{i \in \mathcal{I}_1} \mathbf{z}_{i,j\cdot}^{(1)} \sum_{i \in \mathcal{I}_2} (\mathbf{z}_{i,j\cdot}^{(2)})^T$. By definition,

$$\operatorname{cov}\left(\mathbb{1}(u_j \geq t), \mathbb{1}(u_{j'} \geq t)\right) = \Pr(u_j \geq t, u_{j'} \geq t) - \Pr(u_j \geq t) \Pr(u_{j'} \geq t).$$

Note that $\Pr(u_j \geq t, u_{j'} \geq t) = \Pr(u_j \geq t) + \Pr(u_{j'} \geq t) + \Pr(u_j < t, u_{j'} < t) - 1$. Using Lemma E.4 and E.5,

$$\sup_t \left| \operatorname{cov}\left(\mathbb{1}(u_j \geq t), \mathbb{1}(u_{j'} \geq t)\right) - \operatorname{cov}\left(\mathbb{1}(\tilde{u}_j \geq t), \mathbb{1}(\tilde{u}_{j'} \geq t)\right) \right| \leq C \frac{s_1 \log(p)}{\sqrt{nq}}.$$

Therefore, we only need to consider the Gaussian approximation $\tilde{u}_j$.

By the law of total covariance, we have

$$\begin{aligned}
&\sum_{j,j' \in \mathcal{S}_0^L} \operatorname{cov}\left(\mathbb{1}(\tilde{u}_j \geq t), \mathbb{1}(\tilde{u}_{j'} \geq t)\right) \\
&= \sum_{j,j' \in \mathcal{S}_0^L} \mathrm{E}\left[\operatorname{cov}\left(\mathbb{1}(\tilde{u}_j \geq t), \mathbb{1}(\tilde{u}_{j'} \geq t) \mid \mathbf{z}^{(2)}\right)\right] \\
&+ \sum_{j,j' \in \mathcal{S}_0^L} \operatorname{cov}\left(\mathrm{E}\left[\mathbb{1}(\tilde{u}_j \geq t) \mid \mathbf{z}^{(2)}\right], \mathrm{E}\left[\mathbb{1}(\tilde{u}_{j'} \geq t) \mid \mathbf{z}^{(2)}\right]\right) \\
&\triangleq I_1 + I_2.
\end{aligned}$$

We first bound I_1. The maximum correlation coefficient between two random variables η_1 and η_2 is

$$\rho_{\max}(\eta_1, \eta_2) = \sup_{g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2} \mathrm{E}[g_1(\eta_1) g_2(\eta_2)],$$

where $\mathcal{G}_j = \{g_j : \mathrm{E}[g_j(\eta_j)] = 0, \operatorname{var}(g_j(\eta_j)) = 1\}, j = 1, 2$. For given $\mathbf{z}^{(2)}$, $\operatorname{var}(\mathbb{1}(\tilde{u}_j \geq t)) \leq$

$1, j \in \mathcal{S}_0^L$. Then, we can bound each term in I_1 as

$$
\begin{aligned}
&\operatorname{cov}\left(\mathbb{1}(\tilde{u}_j \geq t), \mathbb{1}(\tilde{u}_{j'} \geq t) \mid \mathbf{z}^{(2)}\right) \\
&= \mathrm{E}[\{\mathbb{1}(\tilde{u}_j \geq t) - \Pr(\tilde{u}_j \geq t)\}\{\mathbb{1}(\tilde{u}_{j'} \geq t) - \Pr(\tilde{u}_{j'} \geq t)\} \mid \mathbf{z}^{(2)}] \\
&\leq \rho_{\max}(\tilde{u}_j, \tilde{u}_{j'} \mid \mathbf{z}^{(2)}).
\end{aligned}
$$

Condition on $\mathbf{z}^{(2)}$, $(\tilde{u}_j, \tilde{u}_{j'})$ follows a bivariate normal distribution. For bivariate normal distributions, the maximum correlation coefficient is the absolute value of the ordinary correlation. Therefore,

$$
I_1 \leq \sum_{j,j' \in \mathcal{S}_0^L} \mathrm{E}\left[\rho_{\max}(\tilde{u}_j, \tilde{u}_{j'} \mid \mathbf{z}^{(2)})\right] = \sum_{j,j' \in \mathcal{S}_0^L} \mathrm{E}\left[\left|\operatorname{cor}(\tilde{u}_j, \tilde{u}_{j'} \mid \mathbf{z}^{(2)})\right|\right].
$$

We calculate the covariance of $\tilde{u}_j$ and $\tilde{u}_{j'}$ for $j, j' \in \mathcal{S}_0^L$.

$$
\begin{aligned}
\operatorname{cov}(\tilde{u}_j, \tilde{u}_{j'} \mid \mathbf{z}^{(2)}) &= \mathrm{E}[\tilde{u}_j \tilde{u}_{j'} \mid \mathbf{z}^{(2)}] - \mathrm{E}[\tilde{u}_j \mid \mathbf{z}^{(2)}]\,\mathrm{E}[\tilde{u}_{j'} \mid \mathbf{z}^{(2)}] \\
&= \frac{16}{n^4}\,\mathrm{E}\left[\sum_{i \in \mathcal{I}_1} \mathbf{z}_{i,j\cdot}^{(1)} \sum_{i \in \mathcal{I}_2} (\mathbf{z}_{i,j\cdot}^{(2)})^T \sum_{i \in \mathcal{I}_1} \mathbf{z}_{i,j'\cdot}^{(1)} \sum_{i \in \mathcal{I}_2} (\mathbf{z}_{i,j'\cdot}^{(2)})^T \mid \mathbf{z}^{(2)}\right] \\
&= \frac{16}{n^4} \sum_{k,k'=1}^{q} \mathrm{E}\left[\sum_{i \in \mathcal{I}_1} z_{i,jk}^{(1)} \sum_{i \in \mathcal{I}_2} z_{i,jk}^{(2)} \sum_{i \in \mathcal{I}_1} z_{i,j'k'}^{(1)} \sum_{i \in \mathcal{I}_2} z_{i,j'k'}^{(2)} \mid \mathbf{z}^{(2)}\right] \\
&= \frac{4}{n^2} \sum_{k,k'=1}^{q} \sum_{i \in \mathcal{I}_2} z_{i,jk}^{(2)} \sum_{i \in \mathcal{I}_2} z_{i,j'k'}^{(2)}\,\mathrm{E}\left[z_{1,jk} z_{1,j'k'}\right] \\
&= \frac{4}{n^2} \sum_{k,k'=1}^{q} \sum_{i \in \mathcal{I}_2} z_{i,jk}^{(2)} \sum_{i \in \mathcal{I}_2} z_{i,j'k'}^{(2)} \operatorname{cov}\left(z_{1,jk} z_{1,j'k'}\right).
\end{aligned}
$$

Therefore, by condition (C3)

$$
\mathrm{E}\left[\operatorname{cov}(\tilde{u}_j, \tilde{u}_{j'} \mid \mathbf{z}^{(2)})\right] = \sum_{k,k'=1}^{q} \operatorname{cov}^2(z_{1,jk}, z_{1,j'k'}) \leq C \sum_{k,k'=1}^{q} \operatorname{cor}^2(z_{1,jk}, z_{1,j'k'}).
$$

By condition (C4),

$$
I_1 \leq C \sum_{j,j' \in \mathcal{S}_0^L} \sum_{k,k'=1}^{q} \operatorname{cor}^2(z_{1,jk}, z_{1,j'k'}) \leq C_3 p_0^{\alpha_3}.
$$

We proceed to bound I_2. Note that $\mathrm{E}[\mathbb{1}(\tilde{u}_j \geq t) \mid \mathbf{z}^{(2)}] = \Pr(\tilde{u}_j \geq t \mid \mathbf{z}^{(2)}) = f(\mathbf{z}_{\cdot,j\cdot}^{(2)})$ and

$\mathrm{E}[\mathbb{1}(\tilde{u}_{j'} \geq t) \mid \mathbf{z}^{(2)}] = g(\mathbf{z}^{(2)}_{\cdot,j'\cdot})$, which are functions of Gaussian random vectors. Then

$$
\begin{aligned}
I_2 &= \sum_{j,j' \in \mathcal{S}_0^L} \operatorname{cov}\left(\mathrm{E}\left[\mathbb{1}(\tilde{u}_j \geq t) \mid \mathbf{z}^{(2)}\right], \mathrm{E}\left[\mathbb{1}(\tilde{u}_{j'} \geq t) \mid \mathbf{z}^{(2)}\right]\right) \\
&= \sum_{j,j' \in \mathcal{S}_0^L} \operatorname{cov}(f(\mathbf{z}^{(2)}_{\cdot,j\cdot}), g(\mathbf{z}^{(2)}_{\cdot,j'\cdot})) \\
&= \sum_{j,j' \in \mathcal{S}_0^L} \mathrm{E}[\{f(\mathbf{z}^{(2)}_{\cdot,j\cdot}) - \mathrm{E}[f(\mathbf{z}^{(2)}_{\cdot,j\cdot})]\}\{g(\mathbf{z}^{(2)}_{\cdot,j'\cdot}) - \mathrm{E}[g(\mathbf{z}^{(2)}_{\cdot,j'\cdot})]\}].
\end{aligned}
$$

We claim that

$$
\left|\mathrm{E}[\{f(\mathbf{z}^{(2)}_{\cdot,j\cdot}) - \mathrm{E}[f(\mathbf{z}^{(2)}_{\cdot,j\cdot})]\}\{g(\mathbf{z}^{(2)}_{\cdot,j'\cdot}) - \mathrm{E}[g(\mathbf{z}^{(2)}_{\cdot,j'\cdot})]\}]\right| \leq C\rho_{\max}(\operatorname{vec}(\mathbf{z}^{(2)}_{\cdot,j\cdot}), \operatorname{vec}(\mathbf{z}^{(2)}_{\cdot,j'\cdot})). \quad (5)
$$

The absolute correlation between two $nq/2$ dimensional centered Gaussian random vectors η_1 and η_2 is defined as

$$
\rho_{\max}(\eta_1, \eta_2) = \sup_{\alpha_1,\alpha_2 \in \mathbb{R}^{nq/2}} \frac{|\,\mathrm{E}[\alpha_1^T \eta_1 \eta_2^T \alpha_2]|}{\left(\mathrm{E}[|\alpha_1^T \eta_1|^2]\right)^{1/2} \left(\mathrm{E}[|\alpha_2^T \eta_2|^2]\right)^{1/2}}.
$$

We apply Theorem 3.4 in (Veraar 2009) to show (5). First, note that

$$
\left(\mathrm{E}\left[\left|f(\mathbf{z}^{(2)}_{\cdot,j\cdot}) - \mathrm{E}[f(\mathbf{z}^{(2)}_{\cdot,j\cdot})]\right|^2\right]\right)^{1/2} \leq 2, \quad \left(\mathrm{E}\left[\left|g(\mathbf{z}^{(2)}_{\cdot,j'\cdot}) - \mathrm{E}[g(\mathbf{z}^{(2)}_{\cdot,j'\cdot})]\right|^2\right]\right)^{1/2} \leq 2
$$

since $f(\mathbf{z}^{(2)}_{\cdot,j\cdot})$ and $g(\mathbf{z}^{(2)}_{\cdot,j'\cdot})$ are probabilities. Second, it is easy to see

$$
\mathrm{E}\left[f(\mathbf{z}^{(2)}_{\cdot,j\cdot}) - \mathrm{E}[f(\mathbf{z}^{(2)}_{\cdot,j\cdot})]\right] = 0, \quad \mathrm{E}\left[g(\mathbf{z}^{(2)}_{\cdot,j'\cdot}) - \mathrm{E}[g(\mathbf{z}^{(2)}_{\cdot,j'\cdot})]\right] = 0.
$$

Then, conditions of Theorem 3.4 in (Veraar 2009) are satisfied and we have (5). Let rows of $\mathbf{z}^{(2)}_{\cdot,j\cdot} \in \mathbb{R}^{n/2 \times q}$ be $\mathbf{z}_{1,j\cdot}, \dots, \mathbf{z}_{n/2,j\cdot}$. For $j \in \mathcal{S}_0^L$, $\mathbf{z}^{(2)}_{\cdot,j\cdot} \sim \mathcal{MN}(\mathbf{0}, \mathbf{I}_{n/2}, \mathbf{U}^{jj})$, where $\mathbf{U}^{jj}$ is the covariance matrix of $\mathbf{z}_{1,j\cdot}$. Then, $\operatorname{vec}(\mathbf{z}^{(2)}_{\cdot,j\cdot}) \sim \mathcal{N}(\mathbf{0}, \mathbf{U}^{jj} \otimes \mathbf{I}_{n/2})$ and $\operatorname{vec}(\mathbf{z}^{(2)}_{\cdot,j'\cdot}) \sim \mathcal{N}(\mathbf{0}, \mathbf{U}^{j'j'} \otimes \mathbf{I}_{n/2})$. According to the result of canonical correlation analysis, $\rho_{\max}(\operatorname{vec}(\mathbf{z}^{(2)}_{\cdot,j\cdot}), \operatorname{vec}(\mathbf{z}^{(2)}_{\cdot,j'\cdot}))$ is the square root of the largest eigenvalue of

$$
\begin{aligned}
&(\mathbf{U}^{jj} \otimes \mathbf{I}_{n/2})^{-1/2}(\mathbf{U}^{jj'} \otimes \mathbf{I}_{n/2})(\mathbf{U}^{j'j'} \otimes \mathbf{I}_{n/2})^{-1}(\mathbf{U}^{j'j} \otimes \mathbf{I}_{n/2})(\mathbf{U}^{jj} \otimes \mathbf{I}_{n/2})^{-1/2} \\
&= ((\mathbf{U}^{jj})^{-1/2} \otimes \mathbf{I}_{n/2})(\mathbf{U}^{jj'} \otimes \mathbf{I}_{n/2})((\mathbf{U}^{j'j'})^{-1} \otimes \mathbf{I}_{n/2})(\mathbf{U}^{j'j} \otimes \mathbf{I}_{n/2})((\mathbf{U}^{jj})^{-1/2} \otimes \mathbf{I}_{n/2}) \\
&= (\mathbf{U}^{jj})^{-1/2}\mathbf{U}^{jj'}(\mathbf{U}^{j'j'})^{-1}\mathbf{U}^{j'j}(\mathbf{U}^{jj})^{-1/2} \otimes \mathbf{I}_{n/2}
\end{aligned}
$$

Therefore, according to the condition (C4)

$$\begin{aligned}
I_2 &= \sum_{j,j' \in \mathcal{S}_0^L} \mathrm{E}[\{f(\mathbf{z}^{(2)}_{\cdot,j\cdot}) - \mathrm{E}[f(\mathbf{z}^{(2)}_{\cdot,j\cdot})]\}\{g(\mathbf{z}^{(2)}_{\cdot,j'\cdot}) - \mathrm{E}[g(\mathbf{z}^{(2)}_{\cdot,j'\cdot})]\}] \\
&\leq \sum_{j,j' \in \mathcal{S}_0^L} C\lambda_{\max}^{1/2}\left((\mathbf{U}^{jj})^{-1/2}\mathbf{U}^{jj'}(\mathbf{U}^{j'j'})^{-1}\mathbf{U}^{j'j}(\mathbf{U}^{jj})^{-1/2}\right) \\
&\leq C_1 p_0^{\alpha_1}.
\end{aligned}$$

□

Proof of Theorem 3.2

Proof. Let

$$V_1'(t) = \frac{\#\left\{j : j \in \mathcal{S}_1^L, u_j \leq -t\right\}}{p_1}.$$

Define

$$\mathrm{FDP}_{\dagger}^L(t) = \frac{V(t)}{(V_{\prime}(t) + r_p V_1(t)) \vee 1/p}, \quad \overline{\mathrm{FDP}}(t) = \frac{G_0(t)}{(G_0(t) + r_p V_1(t)) \vee 1/p}.$$

Recall that

$$\widehat{\mathrm{FDP}}^L(t) = \frac{V(t) + r_p V_1'(t)}{(V_{\prime}(t) + r_p V_1(t)) \vee 1/p}, \quad \tau_a = \inf\{t : \widehat{\mathrm{FDP}}^L(t) < a\}.$$

The proof is based on the following cutoff:

$$\tau_a^* = \inf\{t : \mathrm{FDP}_{\dagger}^L(t) \leq a\}.$$

Since $\mathrm{FDP}_{\dagger}^L(t) \leq \widehat{\mathrm{FDP}}(t)^L$, $\tau_a^* \leq \tau_a$, which implies the empirical cutoff overestimate τ_a^*. In the following, we assume the pointwise limit $\mathrm{FDP}_{\infty}^L(t)$ of $\mathrm{FDP}^L(t)$ exists for all $t > 0$ and there is a constant $t_a > 0$ such that $\mathrm{FDP}_{\infty}^L(t_a) \leq a$. We first show that for any $\epsilon \in (0, a)$, we have

$$\Pr(\tau_a^* \leq t_{a-\epsilon}) \geq 1 - \epsilon,$$

where $t_{a-\epsilon} > 0$ and $\mathrm{FDP}_{\infty}^L(t_{a-\epsilon}) \leq a - \epsilon$. Such $t_{a-\epsilon}$ exists if $\mathrm{FDP}_L^{\infty}(t)$ is left continuous at a. By Lemma C.3, for any fixed $t \in \mathbb{R}$, we have

$$|\,\mathrm{FDP}_{\dagger}^L(t) - \mathrm{FDP}^L(t)| \xrightarrow{P} 0.$$

It follows that for any fixed $t \in \mathbb{R}$,

$$|\operatorname{FDP}_{\dagger}^{L}(t) - \operatorname{FDP}_{\infty}^{L}(t)| \le |\operatorname{FDP}_{\dagger}^{L}(t) - \operatorname{FDP}^{L}(t)| + |\operatorname{FDP}^{L}(t) - \operatorname{FDP}_{\infty}^{L}(t)| \xrightarrow{P} 0.$$

Then for any ϵ, $\Pr(|\operatorname{FDP}_{\dagger}^{L}(t_{a-\epsilon}) - \operatorname{FDP}_{\infty}^{L}(t_{a-\epsilon})| \le \epsilon) \ge 1-\epsilon$ for large enough p. According to the definition of τ_a^*, we have

$$\begin{aligned}
\Pr(\tau_a^* \le t_{a-\epsilon}) &\ge \Pr(\operatorname{FDP}_{\dagger}^{L}(t_{a-\epsilon}) \le a) \\
&\ge \Pr(|\operatorname{FDP}_{\dagger}^{L}(t_{a-\epsilon}) - \operatorname{FDP}_{\infty}^{L}(t_{a-\epsilon})| \le \epsilon) \\
&\ge 1-\epsilon,
\end{aligned}$$

for large enough p. Condition on the event $\tau_a^* \le t_{a-\epsilon}$, we have

$$\begin{aligned}
\limsup_{p\to\infty} \mathrm{E}[\operatorname{FDP}^{L}(\tau_a^*)] &= \limsup_{p\to\infty} \mathrm{E}[\operatorname{FDP}^{L}(\tau_a^*)\mathbb{1}(\tau_a^* \le t_{a-\epsilon})] + \limsup_{p\to\infty} \mathrm{E}[\operatorname{FDP}^{L}(\tau_a^*)\mathbb{1}(\tau_a^* > t_{a-\epsilon})] \\
&\le \limsup_{p\to\infty} \mathrm{E}[\operatorname{FDP}^{L}(\tau_a^*)\mathbb{1}(\tau_a^* \le t_{a-\epsilon})] + \epsilon \\
&\le \limsup_{p\to\infty} \mathrm{E}[|\operatorname{FDP}^{L}(\tau_a^*) - \overline{\operatorname{FDP}}(\tau_a^*)|\mathbb{1}(\tau_a^* \le t_{a-\epsilon})] \\
&+ \limsup_{p\to\infty} \mathrm{E}[|\operatorname{FDP}_{\dagger}^{L}(\tau_a^*) - \overline{\operatorname{FDP}}(\tau_a^*)|\mathbb{1}(\tau_a^* \le t_{a-\epsilon})] \\
&+ \limsup_{p\to\infty} \mathrm{E}[\operatorname{FDP}_{\dagger}^{L}(\tau_a^*)\mathbb{1}(\tau_a^* \le t_{a-\epsilon})] + \epsilon \\
&\le \limsup_{p\to\infty} \mathrm{E}\left[\sup_{0<t\le t_{a-\epsilon}} |\operatorname{FDP}^{L}(t) - \overline{\operatorname{FDP}}(t)|\right] \\
&+ \limsup_{p\to\infty} \mathrm{E}\left[\sup_{0<t\le t_{a-\epsilon}} |\operatorname{FDP}_{\dagger}^{L}(t) - \overline{\operatorname{FDP}}(t)|\right] \\
&+ \limsup_{p\to\infty} \mathrm{E}[\operatorname{FDP}_{\dagger}^{L}(\tau_a^*)] + \epsilon.
\end{aligned}$$

According to Lemma C.3,

$$|\operatorname{FDP}^{L}(t) - \overline{\operatorname{FDP}}(t)| \xrightarrow{P} 0, \quad |\operatorname{FDP}_{\dagger}^{L}(t) - \overline{\operatorname{FDP}}(t)| \xrightarrow{P} 0,$$

for any fixed $t \in \mathbb{R}$. Then the first two terms are 0 by the dominated convergence theorem. For the last term, we have $\operatorname{FDP}_{\dagger}^{L}(\tau_a^*) \le a$ by definition. Therefore,

$$\limsup_{p\to\infty} \mathrm{E}[\operatorname{FDP}^{L}(\tau_a^*)] \le a.$$

$\square$

Proof of Theorem 3.3

Proof. This proof is a stronger result based on $\tau_a = \inf\{t : \widehat{\text{FDP}}(t)^L < a\}$, but requires additional assumptions. We aim to show

$$\limsup_{p\to\infty} \text{E}[\text{FDP}^L(\tau_a)] \leq a.$$

A key step in bounding $\limsup_{p\to\infty} \text{E}[\text{FDP}^L(\tau_a^*)]$ is conditioning on $\tau_a^* \leq t_{a-\epsilon}$. Since $\tau_a^* \leq \tau_a$, $\Pr(\tau_a^* \leq t_{a-\epsilon}) \geq 1-\epsilon$ does not imply $\Pr(\tau_a \leq t_{a-\epsilon}) \geq 1-\epsilon$. Therefore, we need to find t_a such that

$$\Pr(\tau_a \leq t_a) \to 1.$$

It suffices to show with probability approaching one,

$$\widehat{\text{FDP}}^L(t_a) \leq a.$$

According to Lemma E.4, $\frac{n}{2}u_j$ can be approximate by $\frac{2}{n}\sum_{i\in\mathcal{J}_1} \mathbf{z}_{i,j\cdot}^{(1)} \sum_{i\in\mathcal{J}_2} (\mathbf{z}_{i,j\cdot}^{(2)})^T$. Then, for $j \in \mathcal{S}_1^L$, we have

$$\frac{n}{2}\,\text{E}(u_j) = \frac{2}{n}\sum_{i\in\mathcal{J}_1} \text{E}(\mathbf{z}_{i,j\cdot}^{(1)}) \sum_{i\in\mathcal{J}_2} \text{E}(\mathbf{z}_{i,j\cdot}^{(2)})^T = \frac{n}{2}\,\text{E}(\mathbf{z}_{i,j\cdot})\,\text{E}(\mathbf{z}_{i,j\cdot})^T > 0.$$

This shows u_j has a positive mean if $j \in \mathcal{S}_1^L$. Therefore, we can assume for $t > 0$

$$\frac{1}{p_1}\sum_{j\in\mathcal{S}_1^L} \mathbb{1}(u_j < -t) \leq \frac{1}{p_0}\sum_{j'\in\mathcal{S}_0^L} \mathbb{1}(u_j < -t) + \epsilon.$$

Then, we have as $n, p \to \infty$,

$$\widehat{\text{FDP}}^L(t) = \frac{V(t) + r_p V_1'(t)}{(V_{\prime}(t) + r_p V_1(t)) \vee 1/p} \leq \frac{V(t) + r_p V(t) + r_p\epsilon}{(V_{\prime}(t) + r_p V_1(t)) \vee 1/p} \leq (1+r_p)\overline{\text{FDP}}(t) + C\epsilon,$$

where the last inequality uses Lemma C.3. It remains to bound $\overline{\text{FDP}}(t)$. We introduce a signal strength concept similar to (Dai et al. 2023*b*). Let $\mathcal{S}_{1,\text{strong}}^L$ denote the largest subset of $\mathcal{S}_1^L$ such that

$$\frac{1}{p_{1,\text{strong}}} \sum_{j \in \mathcal{S}_{1,\text{strong}}^L} \mathbb{1}(u_j > t) \xrightarrow{P} 1, \text{ for } t_a > 0,$$

where $p_{1,\text{strong}} = |\mathcal{S}_{1,\text{strong}}^L|$ and t_a is defined in the following. We further assume that

$$p_0 \to \infty, \liminf_{n,p\to\infty} \frac{p_{1,\text{strong}}}{p_0} > 0. \tag{6}$$

Since $G_0(t)$ is continuous and $G_0(-\infty) = 1$ and $G_0(-\infty) = 0$, intermediate value theorem guarantees that there exist a t_a such that $G_0(t_a) = \frac{ap_{1,\text{strong}}}{(1-a)p_0+p_1}$. Without loss of generality, we assume $ap_1 < (1-a)p_0$ since otherwise we can select all rows and the FDR is under control. The condition (6) implies $0 < G_0(t_a) < 1$, which means t_a is bounded. Then, we have

$$\overline{\text{FDP}}(t_a) = \frac{G_0(t_a)}{G_0(t_a) + 1/p_0 \sum_{j\in\mathcal{S}_1^L} \mathbb{1}(u_j \geq t_a)} \leq \frac{G_0(t_a)}{G_0(t_a) + p_{1,\text{strong}}/p_0} + o_p(1) \leq \frac{a}{1+r_p} + o_p(1).$$

Therefore,

$$\widehat{\text{FDP}}^L(t_a) \leq (1+r_p)\overline{\text{FDP}}(t) + \epsilon \leq a + \epsilon + o_p(1).$$

Then we can condition on $\tau_a \leq t_a$:

$$\begin{aligned}
\limsup_{p\to\infty} \mathrm{E}[\text{FDP}^L(\tau_a)] &= \limsup_{p\to\infty} \mathrm{E}[\text{FDP}^L(\tau_a)\mathbb{1}(\tau_a \leq t_a)] + \limsup_{p\to\infty} \mathrm{E}[\text{FDP}^L(\tau_a)\mathbb{1}(\tau_a > t_a)] \\
&\leq \limsup_{p\to\infty} \mathrm{E}[\text{FDP}^L(\tau_a)\mathbb{1}(\tau_a \leq t_a)] + \epsilon \\
&\leq \limsup_{p\to\infty} \mathrm{E}[|\,\text{FDP}^L(\tau_a) - \overline{\text{FDP}}(\tau_a)|\mathbb{1}(\tau_a \leq t_a)] \\
&+ \limsup_{p\to\infty} \mathrm{E}[|\,\text{FDP}_\dagger^L(\tau_a) - \overline{\text{FDP}}(\tau_a)|\mathbb{1}(\tau_a \leq t_a)] \\
&+ \limsup_{p\to\infty} \mathrm{E}[\text{FDP}_\dagger^L(\tau_a)\mathbb{1}(\tau_a \leq t_a)] + \epsilon \\
&\leq \limsup_{p\to\infty} \mathrm{E}\left[\sup_{0<t\leq t_a} |\,\text{FDP}^L(t) - \overline{\text{FDP}}(t)|\right] \\
&+ \limsup_{p\to\infty} \mathrm{E}\left[\sup_{0<t\leq t_a} |\,\text{FDP}_\dagger^L(t) - \overline{\text{FDP}}(t)|\right] \\
&+ \limsup_{p\to\infty} \mathrm{E}[\text{FDP}_\dagger^L(\tau_a)] + \epsilon.
\end{aligned}$$

According to lemma C.3,

$$|\operatorname{FDP}^L(t) - \overline{\operatorname{FDP}}(t)| \xrightarrow{P} 0, \quad |\operatorname{FDP}_{\dagger}^L(t) - \overline{\operatorname{FDP}}(t)| \xrightarrow{P} 0,$$

for any fixed $t \in \mathbb{R}$. Then the first two terms are 0 by the dominated convergence theorem. For the last term, we have $\operatorname{FDP}_{\dagger}^L(\tau_a) \leq \widehat{\operatorname{FDP}}^L(\tau_a) \leq a$ by definition. Therefore,

$$\limsup_{p\to\infty} \mathrm{E}[\operatorname{FDP}^L(\tau_a)] \leq a.$$

□

Proof of Corollary 3.6

Proof. Let $M_{jk} = (u_j, v_k), j = 1, \dots, p, k = 1, \dots, q$. For real numbers t_1, t_2, $M_{jk} > (t_1, t_2)$ means $u_j > t_1$ and $v_k > t_2$. Then, the element-wise FDP is

$$\operatorname{FDP}(t_1, t_2) = \frac{\#\{(j,k) : M_{jk} > (t_1, t_2), j \in \mathcal{S}_0^L \text{ or } k \in \mathcal{S}_0^R\}}{\#\{(j,k) : M_{jk} > (t_1, t_2)\}}$$

Note that

$$\begin{aligned}
&\#\{(j,k) : M_{jk} > (t_1, t_2), j \in \mathcal{S}_0^L \text{ or } k \in \mathcal{S}_0^R\} \\
&= \#\{j : u_j > t_1, j \in \mathcal{S}_0^L\} \times \#\{k : v_k > t_2\} + \#\{k : v_k > t_2, k \in \mathcal{S}_0^R\} \times \#\{j : u_j > t_1\} \\
&\quad - \#\{j : u_j > t_1, j \in \mathcal{S}_0^L\} \times \#\{k : v_k > t_2, k \in \mathcal{S}_0^R\}
\end{aligned}$$

and

$$\#\{(j,k) : M_{jk} > (t_1, t_2)\} = \#\{j : u_j > t_1\} \times \#\{k : v_k > t_2\}.$$

Therefore, we have

$$\operatorname{FDP}(\tau_a^*, \tau_b^*) = \operatorname{FDP}^L(\tau_a^*) + \operatorname{FDP}^R(\tau_b^*) - \operatorname{FDP}^L(\tau_a^*) \operatorname{FDP}^R(\tau_b^*). \tag{7}$$

Using (7) and Theorem 3.2, the element-wise FDR satisfies

$$\limsup_{n,p,q\to\infty} \operatorname{FDR}(\tau_a^*, \tau_b^*) \leq a + b - ab.$$

□

E Ancillary Lemmas

Lemma E.1 ((Bellec & Zhang 2021))**.** *If* $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$*, then*

$$\mathrm{E}[\mathbf{z} f(\mathbf{z})] = \mathrm{E}[\nabla f(\mathbf{z})], \quad \mathrm{E}[\mathbf{z}\mathbf{z}^T f(\mathbf{z})] = \mathrm{E}[f(\mathbf{z})]\mathbf{I}_p + \mathrm{E}[\nabla^2 f(\mathbf{z})]$$

where $\nabla f(\mathbf{z}) = \frac{\partial f(\mathbf{z})}{\partial \mathbf{z}}, \nabla^2 f(\mathbf{z}) = \frac{\partial f(\mathbf{z})}{\partial \mathbf{z} \partial \mathbf{z}^T}$.

Lemma E.2. *Let* $\xi_{i,jk} = Y_i \Omega_{j\cdot} \mathbf{X}_{i,\cdot k}$*. Then*

- $\mathrm{E}(\xi_{i,jk}) = \Omega_{\cdot j}^T \mathrm{E}(Y \mathbf{X}_{\cdot k})$;
- $\mathrm{var}(\xi_{i,jk}) = \Omega_{\cdot j}^T \mathrm{E}(Y^2 \mathbf{X}_{\cdot k} \mathbf{X}_{\cdot k}^T) \Omega_{\cdot j} - \Omega_{\cdot j}^T \mathrm{E}(Y \mathbf{X}_{\cdot k}) \mathrm{E}(Y \mathbf{X}_{\cdot k}^T) \Omega_{\cdot j}$;
- $\mathrm{cov}(\xi_{i,jk}, \xi_{i,lm}) = \Omega_{\cdot j}^T \mathrm{E}(Y^2 \mathbf{X}_{\cdot k} \mathbf{X}_{\cdot m}^T) \Omega_{\cdot l} - \Omega_{\cdot j}^T \mathrm{E}(Y \mathbf{X}_{\cdot k}) \mathrm{E}(Y \mathbf{X}_{\cdot m}^T) \Omega_{\cdot l}$.

Proof. First, $\mathrm{E}(\xi_{i,jk}) = \mathrm{E}(\Omega_{j\cdot} \mathbf{X}_{\cdot k} Y) = \Omega_{\cdot j}^T \mathrm{E}(Y \mathbf{X}_{\cdot k})$ since Ω is symmetric. Next, consider the second moment $\mathrm{E}(\xi_{i,jk}^2)$ which can be written as

$$\begin{aligned} \mathrm{E}(\xi_{i,jk}^2) &= \mathrm{E}(\Omega_{\cdot j}^T \mathbf{X}_{i,\cdot k} Y_i \cdot Y_i \mathbf{X}_{i,\cdot k}^T \Omega_{\cdot j}) = \Omega_{\cdot j}^T \mathrm{E}(Y_i^2 \mathbf{X}_{i,\cdot k} \mathbf{X}_{i,\cdot k}^T) \Omega_{\cdot j} \\ &= \Omega_{\cdot j}^T \mathrm{E}(Y^2 \mathbf{X}_{\cdot k} \mathbf{X}_{\cdot k}^T) \Omega_{\cdot j}. \end{aligned}$$

Then, we have

$$\mathrm{var}(\xi_{i,jk}) = \mathrm{E}(\xi_{i,jk}^2) - \mathrm{E}(\xi_{i,jk})^2 = \Omega_{\cdot j}^T \mathrm{E}(Y^2 \mathbf{X}_{\cdot k} \mathbf{X}_{\cdot k}^T) \Omega_{\cdot j} - \Omega_{\cdot j}^T \mathrm{E}(Y \mathbf{X}_{\cdot k}) \mathrm{E}(Y \mathbf{X}_{\cdot k}^T) \Omega_{\cdot j}.$$

Finally, we compute the covariance

$$\mathrm{cov}(\xi_{i,jk}, \xi_{i,lm}) = \mathrm{E}(\xi_{i,jk} \xi_{i,lm}) - \mathrm{E}(\xi_{i,jk}) \mathrm{E}(\xi_{i,lm}).$$

Note that

$$\mathrm{E}(\xi_{i,jk} \xi_{i,lm}) = \mathrm{E}(Y_i \Omega_{\cdot j}^T \mathbf{X}_{i,\cdot k} Y_i \mathbf{X}_{i,\cdot m}^T \Omega_{\cdot l}) = \Omega_{\cdot j}^T \mathrm{E}(Y_i^2 \mathbf{X}_{i,\cdot k} \mathbf{X}_{i,\cdot m}^T) \Omega_{\cdot l} = \Omega_{\cdot j}^T \mathrm{E}(Y^2 \mathbf{X}_{\cdot k} \mathbf{X}_{\cdot m}^T) \Omega_{\cdot l}.$$

Therefore,

$$\mathrm{cov}(\xi_{i,jk}, \xi_{i,lm}) = \Omega_{\cdot j}^T \mathrm{E}(Y^2 \mathbf{X}_{\cdot k} \mathbf{X}_{\cdot m}^T) \Omega_{\cdot l} - \Omega_{\cdot j}^T \mathrm{E}(Y \mathbf{X}_{\cdot k}) \mathrm{E}(Y \mathbf{X}_{\cdot m}^T) \Omega_{\cdot l}.$$

□

Lemma E.3. *Assume Assumptions B.1&B.2 hold. Let $\mathbf{L} \in \mathbb{R}^q$ be any fixed vector such that $\|\mathbf{L}\|_1 = 1$ and $\mathbf{z}_{i,j\cdot} \in \mathbb{R}^q$ be a row vector consisting of $z_{i,jk}$. If $s_1 \log(p) = o\left(\sqrt{nqp^{-1}}\right)$, we have*

$$\sup_t \left| \Pr(\sqrt{n}\widehat{\mathbf{A}}_{j\cdot}\mathbf{L} \le t) - \Pr(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{z}_{i,j\cdot}\mathbf{L} \le t) \right| \le C \frac{s_1 \log(p)}{\sqrt{nq}},$$

for any $j = 1, \dots, p$.

Proof. First, we consider the population form

$$\widetilde{\mathbf{A}}_{j\cdot}\mathbf{L} = \frac{1}{n} \sum_{i=1}^{n} \Omega_{j\cdot} \sum_{k=1}^{q} \mathbf{X}_{i,\cdot k} Y_i l_k.$$

Let $\xi_{i,j} = \sum_{k=1}^{q} \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} Y_i l_k$, then

$$\widetilde{\mathbf{A}}_{j\cdot}\mathbf{L} = \frac{1}{n} \sum_{i=1}^{n} \xi_{i,j}, \quad \sqrt{n}\widetilde{\mathbf{A}}_{j\cdot}\mathbf{L} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{i,j}.$$

To apply the Berry-Esseen Theorem, we focus on bounds on $\mathrm{E}(\xi_{i,j}^2)$ and $\mathrm{E}(|\xi_{i,j}|^3)$. We first show $\mathrm{E}(\xi_{i,j}^2) \ge c$ for all $j = 1, \dots, p$. Suppose $\mathrm{E}(\xi_{i,j}^2) = 0$, we have $\xi_{i,j} = Y_i \sum_{k=1}^{q} l_k \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} = 0$ almost surely, which means $Y_i = 0$ or $\sum_{k=1}^{q} l_k \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} = 0$ almost surely. Since $Y \neq 0$ and $\sum_{k=1}^{q} l_k \Omega_{j\cdot}\mathbf{X}_{i,\cdot k}$ is a normal random variable with mean zero. Therefore, $\xi_{i,j} = 0$ implies $\sum_{k=1}^{q} l_k \Omega_{j\cdot}\mathbf{X}_{i,\cdot k}$ has zero variance, which contradicts the fact that both $\mathbf{U}$ and $\mathbf{V}$ are non-degenerate (condition C2).

Then we compute the upper bound on the moment of ξ_{ij}. Note that

$$\mathrm{E}(|\xi_{ij}|^h) \le \mathrm{E}\left(\left| \sum_{k=1}^{q} \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} l_k \right|^h |Y_i|^h \right) \le \mathrm{E}\left(\left| \sum_{k=1}^{q} \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} l_k \right|^{2h} \right)^{1/2} \mathrm{E}(|Y_i|^{2h})^{1/2}.$$

Since $\mathbf{X}_{i,\cdot k} \sim N(0, V_{kk}\mathbf{U})$, $\Omega_{j\cdot}\mathbf{X}_{i,\cdot k}/\|\Omega_{j\cdot}\|_2$ is sub-Gaussian with parameter $V_{kk}\|\mathbf{U}\|_2$. By Lemma E.6, we have $\sum_{k=1}^{q} \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} l_k$ is sub-Gaussian with parameter $(\sum_{k=1}^{q} |l_k| \|\Omega_{j\cdot}\|_2 \sqrt{V_{kk}\|\mathbf{U}\|_2})^2$. Note that $(\sum_{k=1}^{q} |l_k| \|\Omega_{j\cdot}\|_2 \sqrt{V_{kk}\|\mathbf{U}\|_2})^2 = \|\mathbf{U}\|_2 (\sum_{k=1}^{q} |l_k| \|\Omega_{j\cdot}\|_2 \sqrt{V_{kk}})^2 \le C_1 \|\mathbf{U}\|_2 \|\Omega\|_2^2 \|\mathbf{L}\|_1^2 \le C_2$ since $\|\mathbf{L}\|_1 = 1$. Then, we have $\mathrm{E}(|\sum_{k=1}^{q} \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} l_k|^{2h}) \le C_1^h h^h$ and $\mathrm{E}(|Y|^{2h}) \le C_2^h h^h$ (Assumption B.1). Therefore, $\mathrm{E}(|\xi_{ij}|^h) \le C h^h$.

Using the above results, we obtain

$$c \le \mathrm{E}[\xi_{i,j}^2] \le C_1\|\mathbf{L}\|_1^2 = C_1, \quad \mathrm{E}[\xi_{i,j}^2]^{3/2} \le \mathrm{E}[|\xi_{i,j}|^3] \le C_2\|\mathbf{L}\|_1^3 = C_2.$$

Immediately, we have

$$\sup_t \left| \Pr(\sqrt{n}\widetilde{\mathbf{A}}_{j\cdot}\mathbf{L} \le t) - \Pr(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,j\cdot}\mathbf{L} \le t) \right| \le C\frac{1}{\sqrt{n}}.$$

Using the same argument as in the proof of Theorem A.1, we have

$$\begin{aligned}|\sqrt{n}\widehat{\mathbf{A}}_{j\cdot}\mathbf{L} - \sqrt{n}\widetilde{\mathbf{A}}_{j\cdot}\mathbf{L}| &= |\sum_{k=1}^q \sqrt{n}(\widehat{A}_{jk} - \widetilde{A}_{jk})l_k| \\ &\le \sum_{k=1}^q |\sqrt{n}(\widehat{A}_{jk} - \widetilde{A}_{jk})l_k| = O_P\left(s_1 q^{-1/2}\frac{\log(p)}{\sqrt{n}}\right).\end{aligned}$$

According to (Lyu et al. 2019), the desired convergence rate of $\widehat{\Omega}$ holds if $s_1\log(p) = o(\sqrt{nqp^{-1}})$. Since it is a one-dimensional random variable, we have

$$\left|\Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,j\cdot}\mathbf{L} \le t+\epsilon\right) - \Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,j\cdot}\mathbf{L} \le t\right)\right| \le O(\epsilon) = O(s_1 q^{-1/2}\log(p)/\sqrt{n}).$$

Then, if $s_1 p^{1/2}\log(p)/\sqrt{nq} = o(1)$, we have

$$\sup_t \left| \Pr(\sqrt{n}\widehat{\mathbf{A}}_{j\cdot}\mathbf{L} \le t) - \Pr(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,j\cdot}\mathbf{L} \le t) \right| \le C\frac{s_1\log(p)}{\sqrt{nq}}.$$

Proof using (Chernozhukov et al. 2017). For condition (M.1), we need to show $\mathrm{E}(\xi_{i,j}^2) \ge c$ for all $j = 1, \dots, p$. Suppose $\mathrm{E}(\xi_{i,j}^2) = 0$, we have $\xi_{i,j} = Y_i\sum_{k=1}^q l_k\Omega_{j\cdot}\mathbf{X}_{i,\cdot k} = 0$ almost surely, which means $Y_i = 0$ or $\sum_{k=1}^q l_k\Omega_{j\cdot}\mathbf{X}_{i,\cdot k} = 0$ almost surely. Since $Y \ne 0$ and $\sum_{k=1}^q l_k\Omega_{j\cdot}\mathbf{X}_{i,\cdot k}$ is a normal random variable with mean zero. Therefore, $\xi_{i,j} = 0$ implies $\sum_{k=1}^q l_k\Omega_{j\cdot}\mathbf{X}_{i,\cdot k}$ has zero variance, which contradicts the fact that both $\mathbf{U}$ and $\mathbf{V}$ are non-degenerate.

Second, for condition (E.1), we have

$$\begin{aligned}
\mathrm{E}[\exp(|\xi_{i,j}|/C_n)] &= 1 + \sum_{h=1}^{\infty} \frac{\mathrm{E}(|\xi_{i,j}|^h)}{C_n^h h!} = 1 + \sum_{h=1}^{\infty} \frac{\mathrm{E}(|\sum_{k=1}^{q} \Omega_{j\cdot} \mathbf{X}_{i,\cdot k} Y_i l_k|^h)}{C_n^h h!} \\
&\leq 1 + \sum_{h=1}^{\infty} \frac{\mathrm{E}(|\sum_{k=1}^{q} \Omega_{j\cdot} \mathbf{X}_{i,\cdot k} l_k|^h |Y_i|^h)}{C_n^h h!} \\
&\leq 1 + \sum_{h=1}^{\infty} \frac{\mathrm{E}(|\sum_{k=1}^{q} \Omega_{j\cdot} \mathbf{X}_{i,\cdot k} l_k|^{2h})^{1/2}\, \mathrm{E}(|Y_i|^{2h})^{1/2}}{C_n^h h!}.
\end{aligned}$$

Since $\mathbf{X}_{i,\cdot k} \sim N(0, V_{kk}\mathbf{U})$, $\Omega_{j\cdot}\mathbf{X}_{i,\cdot k}/\|\Omega_{j\cdot}\|_2$ is sub-Gaussian with parameter $V_{kk}\|\mathbf{U}\|_2$. By Lemma E.6, we have $\sum_{k=1}^{q} \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} l_k$ is sub-Gaussian with parameter $(\sum_{k=1}^{q} |l_k| \|\Omega_{j\cdot}\|_2 \sqrt{V_{kk}\|\mathbf{U}\|_2})^2$. Note that $(\sum_{k=1}^{q} |l_k| \|\Omega_{j\cdot}\|_2 \sqrt{V_{kk}\|\mathbf{U}\|_2})^2 = \|\mathbf{U}\|_2 (\sum_{k=1}^{q} |l_k| \|\Omega_{j\cdot}\|_2 \sqrt{V_{kk}})^2 \leq C_1 \|\mathbf{U}\|_2 \|\Omega\|_2^2 \|\mathbf{L}\|_1^2 \leq C_2$ since $\|\mathbf{L}\|_1 = 1$. Then, we have $\mathrm{E}(|\sum_{k=1}^{q} \Omega_{j\cdot}\mathbf{X}_{i,\cdot k} l_k|^{2h}) \leq C_1^h h^h$ and $\mathrm{E}(|Y|^{2h}) \leq C_2^h h^h$. Since $h! \geq (h/e)^h$, it follows that

$$\mathrm{E}[\exp(|\xi_{i,j}|/C_n)] \leq 1 + \sum_{h=1}^{\infty} \frac{C^h h^h}{C_n^h h!} \leq 1 + \sum_{h=1}^{\infty} \frac{(Ce)^h}{C_n^h} < \infty,$$

if $C_n > Ce$.

Third, for condition (M.2), note that

$$\sum_{l=1}^{2} \frac{\mathrm{E}(|\xi_{i,j}|^{2+l})}{C_n^{l+2}(l+2)!} \leq \mathrm{E}[\exp(|\xi_{i,j}|/C_n)] < \infty.$$

Letting $C_n > Ce$ sufficiently large, we have

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(|\xi_{i,j}|^{2+l}) \leq C_n^l, l = 1, 2.$$

According to Proposition 2.1 in (Chernozhukov et al. 2017),

$$\sup_t \left| \Pr(\sqrt{n}\widetilde{\mathbf{A}}_{j\cdot}\mathbf{L} \leq t) - \Pr(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{z}_{i,j\cdot}\mathbf{L} \leq t) \right| \leq CD_{n,1}.$$

Using the same argument as in the proof of Theorem A.1, we have

$$\begin{aligned}
|\sqrt{n}\widehat{\mathbf{A}}_{j\cdot}\mathbf{L} - \sqrt{n}\widetilde{\mathbf{A}}_{j\cdot}\mathbf{L}| &= |\sum_{k=1}^{q} \sqrt{n}(\widetilde{A}_{jk} - \widetilde{A}_{jk}) l_k| \\
&\leq \sum_{k=1}^{q} |\sqrt{n}(\widetilde{A}_{jk} - \widetilde{A}_{jk}) l_k| = O_P\left(s_1 q^{-1/2} \frac{\log(p)}{\sqrt{n}}\right).
\end{aligned}$$

Since it is a one-dimensional random variable, we have

$$\left|\Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_{i,j\cdot}\mathbf{L}\le t+\epsilon\right)-\Pr\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_{i,j\cdot}\mathbf{L}\le t\right)\right|\le O(\epsilon)=O(s_1q^{-1/2}\log(p)/\sqrt{n}).$$

Then the following holds

$$\sup_t\left|\Pr(\sqrt{n}\widehat{\mathbf{A}}_{j\cdot}\mathbf{L}\le t)-\Pr(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_{i,j\cdot}\mathbf{L}\le t)\right|\le CD_{n,1}=C\frac{\log^{7/6}(n)}{n^{1/6}}$$

if we require $s_1q^{-1/2}\log(p)/\sqrt{n}=O(D_{n,1})$, which implies $s_1\log(p)=O(q^{1/2}n^{1/3}\log^{7/6}(n))$. On the other hand, we require $s_1\log(p)=o(\sqrt{nqp^{-1}})$ to achieve the desired convergence rate of $\widehat{\Omega}$ (Lyu et al. 2019). It is easy to check $s_1\log(p)=o(\sqrt{nqp^{-1}})$ is a stronger condition. □

Lemma E.4. *Assume Assumption B.1&B.2 hold and $s_1\log(p)=o(\sqrt{nqp^{-1}})$, let $z_{i,jk}^{(1)}$ and $z_{i,jk}^{(2)}$ denote the normal random variables corresponding to the two subsets satisfying the mean and covariance structure. Then, with probability at least $1-1/(pq)$,*

$$\sup_t\left|\Pr\left(\frac{n}{2}u_j\le t\right)-\Pr\left(\frac{2}{n}\sum_{i\in\mathcal{I}_1}\mathbf{z}_{i,j\cdot}^{(1)}\sum_{i\in\mathcal{I}_2}(\mathbf{z}_{i,j\cdot}^{(2)})^T\le t\right)\right|\le C\frac{s_1\log(p)}{\sqrt{nq}},$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ are the sample indices of two subsets.

Proof. Note that

$$\frac{n}{2}u_j=\frac{n}{2}\sum_{l=1}^{q}\widehat{A}_{jl}^{(1)}\widehat{A}_{jl}^{(2)}=\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{j\cdot}^{(1)}\mathbf{L},\quad \mathbf{L}=\sqrt{\frac{n}{2}}(\widehat{A}_{j1}^{(2)},\cdots,\widehat{A}_{jq}^{(2)})^T.$$

Then we have

$$\Pr\left(\frac{n}{2}u_j\le t\right)=\Pr\left(\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{j\cdot}^{(1)}\mathbf{L}\le t\right).$$

According to Lemma E.3,

$$\begin{aligned}
&\sup_t \left|\Pr\left(\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{j\cdot}^{(1)}\mathbf{L} \le t\right) - \Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}^{(1)}\mathbf{L} \le t\right)\right| \\
&= \sup_t \left|\mathrm{E}_{\widehat{\mathbf{A}}_{j\cdot}^{(2)}}\left[\Pr\left(\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{j\cdot}^{(1)}\mathbf{L} \le t \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right) - \Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}^{(1)}\mathbf{L} \le t \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right)\right]\right| \\
&\le \mathrm{E}_{\widehat{\mathbf{A}}_{j\cdot}^{(2)}}\left[\sup_t \left|\Pr\left(\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{j\cdot}^{(1)}\mathbf{L} \le t \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right) - \Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}^{(1)}\mathbf{L} \le t \mid \widehat{\mathbf{A}}_{j\cdot}^{(2)}\right)\right|\right] \\
&\le C\frac{s_1\log(p)}{\sqrt{nq}}. \qquad (8)
\end{aligned}$$

Similarly, we can condition on $\sqrt{2/n}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}^{(1)}$ and apply Lemma E.3 again to get

$$\sup_t \left|\Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}^{(1)}\mathbf{L} \le t\right) - \Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}^{(1)}\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_2}(\mathbf{z}_{i,j\cdot}^{(2)})^T \le t\right)\right| \le C\frac{s_1\log(p)}{\sqrt{nq}}.$$

Adding the above result with (8) completes the proof. □

Lemma E.5. *Assume Assumption B.1&B.2 hold and* $s_1\log(p) = o(\sqrt{nqp^{-1}})$*, then with probability at least* $1-1/(pq)$*,*

$$\begin{aligned}
&\sup_t \left|\Pr\left(\frac{n}{2}u_j \le t, \frac{n}{2}u_{j'} \le t\right) - \right. \\
&\qquad \left.\Pr\left(\frac{2}{n}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}^{(1)}\sum_{i\in\mathcal{J}_2}(\mathbf{z}_{i,j\cdot}^{(2)})^T \le t, \frac{2}{n}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j'\cdot}^{(1)}\sum_{i\in\mathcal{J}_2}(\mathbf{z}_{i,j'\cdot}^{(2)})^T \le t\right)\right| \le C\frac{s_1\log(p)}{\sqrt{nq}}.
\end{aligned}$$

Proof. Using the same argument as in Lemma E.3, we can show that

$$\begin{aligned}
&\sup_{t_1,t_2}\left|\Pr(\sqrt{n}\widehat{\mathbf{A}}_{j\cdot}\mathbf{L}_1 \le t_1, \sqrt{n}\widehat{\mathbf{A}}_{j'\cdot}\mathbf{L}_2 \le t_2) - \right. \\
&\qquad \left.\Pr(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,j\cdot}\mathbf{L}_1 \le t_1, \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{z}_{i,j'\cdot}\mathbf{L}_2 \le t_2)\right| \le C\frac{s_1\log(p)}{\sqrt{nq}}, \qquad (9)
\end{aligned}$$

where $\mathbf{L}_1, \mathbf{L}_2 \in \mathbb{R}^q$ are fixed vectors with unit ℓ_1 norm.

Note that

$$\begin{aligned}
\frac{n}{2}u_j &= \frac{n}{2}\sum_{l=1}^q \widehat{A}_{jl}^{(1)}\widehat{A}_{jl}^{(2)} = \sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{j\cdot}^{(1)}\mathbf{L}_1, \quad \mathbf{L}_1 = \sqrt{\frac{n}{2}}(\widehat{A}_{j1}^{(2)}, \cdots, \widehat{A}_{jq}^{(2)})^T \\
\frac{n}{2}u_{j'} &= \frac{n}{2}\sum_{l=1}^q \widehat{A}_{j'l}^{(1)}\widehat{A}_{j'l}^{(2)} = \sqrt{\frac{n}{2}}\widehat{\mathbf{A}}_{j'\cdot}^{(1)}\mathbf{L}_2, \quad \mathbf{L}_2 = \sqrt{\frac{n}{2}}(\widehat{A}_{j'1}^{(2)}, \cdots, \widehat{A}_{j'q}^{(2)})^T
\end{aligned}$$

Then we have

$$\Pr\left(\frac{n}{2}u_j \le t, \frac{n}{2}u_{j'} \le t\right) = \Pr\left(\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}^{(1)}_{j\cdot}\mathbf{L}_1 \le t, \sqrt{\frac{n}{2}}\widehat{\mathbf{A}}^{(1)}_{j'\cdot}\mathbf{L}_2 \le t\right).$$

Similarly, we condition on $\widehat{\mathbf{A}}^{(2)}$ and use (9)

$$\begin{aligned}
&\sup_t\left|\Pr\left(\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}^{(1)}_{j\cdot}\mathbf{L}_1 \le t, \sqrt{\frac{n}{2}}\widehat{\mathbf{A}}^{(1)}_{j'\cdot}\mathbf{L}_2 \le t\right) - \Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}\mathbf{L}_1 \le t, \sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j'\cdot}\mathbf{L}_2 \le t\right)\right| \\
&= \sup_t\left|\mathrm{E}\left[\Pr\left(\sqrt{\frac{n}{2}}\widehat{\mathbf{A}}^{(1)}_{j\cdot}\mathbf{L}_1 \le t, \sqrt{\frac{n}{2}}\widehat{\mathbf{A}}^{(1)}_{j'\cdot}\mathbf{L}_2 \le t \mid \widehat{\mathbf{A}}^{(2)}\right) - \right.\right. \\
&\qquad\qquad \left.\left.\Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}\mathbf{L}_1 \le t, \sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j'\cdot}\mathbf{L}_2 \le t \mid \widehat{\mathbf{A}}^{(2)}\right)\right]\right| \\
&\le C\frac{s_1\log(p)}{\sqrt{nq}}. \qquad (10)
\end{aligned}$$

Then we can condition on $\sqrt{2/n}\mathbf{z}_{i,\cdot\cdot}$ and apply (9) again

$$\begin{aligned}
&\sup_t\left|\Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}\mathbf{L}_1 \le t, \sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j'\cdot}\mathbf{L}_2 \le t\right) - \right. \\
&\qquad \left.\Pr\left(\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j\cdot}\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_2}\mathbf{z}_{i,j\cdot} \le t, \sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_1}\mathbf{z}_{i,j'\cdot}\sqrt{\frac{2}{n}}\sum_{i\in\mathcal{J}_2}\mathbf{z}_{i,j'\cdot} \le t\right)\right| \le C\frac{s_1\log(p)}{\sqrt{nq}}.
\end{aligned}$$

Combing this result with (10) completes the proof. $\square$

Lemma E.6. *If* $X_i \sim \mathrm{subG}(\sigma_i^2), i = 1, \dots n$, $\sum_{i=1}^n a_iX_i \sim \mathrm{subG}((\sum_{i=1}^n |a_i|\sigma_i)^2)$, *where* $a_i \in \mathbb{R}$ *are constants.*

Proof. If $X_i \sim \mathrm{subG}(\sigma_i^2)$, then $\forall a_i \in \mathbb{R}$, $a_iX_i \sim \mathrm{subG}((|a_i|\sigma_i)^2)$, which follows from the definition

$$\mathrm{E}[\exp(sa_iX_i)] \le \exp(\sigma_i^2s^2a_i^2/2) = \exp((|a_i|\sigma_i)^2s^2/2).$$

Next, we aim to show $\sum_{i=1}^n X_i \sim \mathrm{subG}((\sum_{i=1}^n \sigma_i)^2)$. Let $p_i = \sum_{i=1}^n \sigma_i/\sigma_i$, then $p_i \in (1, \infty]$

and $\sum_{i=1}^n 1/p_i = 1$. By Hölder's inequality, we have

$$\begin{aligned}
\mathrm{E}\left[\exp\left(s\sum_{i=1}^n X_i\right)\right] &= \mathrm{E}\left[\prod_{i=1}^n \exp(sX_i)\right] \le \prod_{i=1}^n \mathrm{E}[\exp(sX_ip_i)]^{1/p_i} \\
&\le \prod_{i=1}^n \exp(s^2p_i^2\sigma_i^2/2)^{1/p_i} = \prod_{i=1}^n \exp(s^2\sigma_i^2p_i/2) \\
&= \exp\left(\frac{s^2}{2}\sum_{i=1}^n \sigma_i^2 p_i\right) = \exp\left(\frac{s^2}{2}\left(\sum_{i=1}^n \sigma_i\right)^2\right).
\end{aligned}$$

Therefore, we have shown that $\sum_{i=1}^n a_iX_i \sim \mathrm{subG}((\sum_{i=1}^n |a_i|\sigma_i)^2)$. Note that we do not assume X_i are independent. With independence, we can easily show that $\sum_{i=1}^n a_iX_i \sim \mathrm{subG}(\sum_{i=1}^n a_i^2\sigma_i^2)$.

□

F Additional simulation results

We first report the column selection results for Model 3 in Table S.1. The performance closely mirrors that of row selection due to the symmetry in the model setup: $\mathbf{U} = \mathbf{V}$, $p = q$, and $p_1 = q_1$. The column selection results for Models 1-2 are also similar to row selection and therefore omitted to avoid redundancy. Given this symmetry, the main text focuses on row selection. Unless stated otherwise, the target FDR level is set to 0.2. All simulations are conducted on a high-performance compute cluster, utilizing a single node with 25 CPUs, each featuring 8 cores and 3.9GB RAM.

In both the simulations and the real data analysis, we estimate the row and column covariance matrices $\mathbf{U}$ and $\mathbf{V}$ using the iterative maximum likelihood method of Dutilleul (1999). The algorithm updates the estimator of $\mathbf{U}$ while holding $\mathbf{V}$ fixed, and vice-versa. For completeness, we summarize the procedure in Algorithm S.1.

Algorithm S.1: MLE Algorithm for the Matrix Normal Distribution

Input : n independent samples $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, and threshold ϵ.

Output: MLE $\widehat{U}$, and $\widehat{V}$.

Compute $\widehat{\mathbf{M}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i$;

Set $\widehat{\mathbf{U}} = I_p, \widehat{\mathbf{V}} = \mathbf{I}_q$;

repeat

    Save previous estimate: $\widehat{\mathbf{U}}_{\text{old}} = \widehat{\mathbf{U}}, \widehat{\mathbf{V}}_{\text{old}} = \widehat{\mathbf{V}}$;

    $\widehat{\mathbf{U}} = \frac{1}{nq}\sum_{i=1}^{n}(\mathbf{X}_i - \widehat{\mathbf{M}})\widehat{\mathbf{V}}^{-1}(\mathbf{X}_i - \widehat{\mathbf{M}})^T$;

    $\widehat{\mathbf{V}} = \frac{1}{np}\sum_{i=1}^{n}(\mathbf{X}_i - \widehat{\mathbf{M}})^T\widehat{\mathbf{U}}^{-1}(\mathbf{X}_i - \widehat{\mathbf{M}})$;

until $||\widehat{\mathbf{U}} - \widehat{\mathbf{U}}_{old}||_F < \epsilon$ *and* $||\widehat{\mathbf{V}} - \widehat{\mathbf{V}}_{old}||_F < \epsilon$;

For reference, the explicit definition of Model 3 in the main text is

$$Y = \sum_{j=1}^{\lfloor p_1/2\rfloor}\sum_{k=1}^{\lfloor q_1/2\rfloor}\exp(0.5X_{jk}) + \sum_{\substack{j=p+1\\-\lfloor p_1/2\rfloor}}^{p}\sum_{\substack{k=q+1\\-\lfloor q_1/2\rfloor}}^{q}\exp(0.5X_{jk})$$
$$+\,\mathrm{sign}(\sum_{j=1}^{\lfloor p_1/2\rfloor}\sum_{\substack{k=q+1\\-\lfloor q_1/2\rfloor}}^{q}X_{jk}) + \mathrm{sign}(\sum_{\substack{j=p+1\\-\lfloor p_1/2\rfloor}}^{p}\sum_{k=1}^{\lfloor q_1/2\rfloor}X_{jk}).$$

Models 1-3 in the main text are all regression models with a continuous response. In this section, we additionally consider a binary response model.

Model 4. The response $Y \sim \text{Bernoulli}(\pi)$. Let the conditional distribution of $\mathbf{X}|Y$ be normal. The conditional means are $\mathrm{E}[\mathbf{X}|Y = 0] = \mathbf{0}_{p\times q}$ and $\mathrm{E}[\mathbf{X}|Y = 1]$ has a block structure with $\mu\mathbf{I}_{p_1\times q_1}$ in the active top-left region and zeros elsewhere. The conditional variances are $\mathrm{var}(X_{ij}|Y = 0) = \sigma^2$ for active elements and 1 otherwise, and $\mathrm{var}(X_{ij}|Y = 1) = \tau^2$ for active elements and 1 otherwise. Using Baye's theorem, the conditional probability $\Pr(Y = 1|\mathbf{X})$ is a function of the submatrix $\mathbf{X}_{1:p_1,1:p_1}$. We set $\pi = 0.5, \mu = 0.5, \sigma^2 = 1, \tau^2 = 2$.

For row selection with $(n, p, p_1, q, q_1) = (2000, 50, 10, 50, 10)$, our method achieves $(\mathrm{FDR}\,\%, \mathrm{Power}\,\%) = (18.3, 100)$, whereas MVSA+BH yields (0, 36.7). For element selection with $(n, p, p_1, q, q_1) = (1000, 20, 6, 20, 6)$, the results are: Proposed (19.3, 91.2); DRC+BH (72.2, 21.9); HSIC+BH (19.1, 100); MTA (55.0, 3.0); Knockoff-SIR (31.5, 9.9); Knockoff-Lasso (12.4, 15.8); MVSA+BH (0.0, 7.5). These findings indicate good FDR control and high power for our method under this classification model.

In Table 3, we report results to demonstrate robustness to non-normal distribution for Model 3. We summarize the results for Models 1 and 2 in Table S.2. Overall, the conclusions are consistent across Models 1–3.

Furthermore, our procedure is substantially faster than competitors. Table S.4 shows that our method runs 180-380 times faster than MVSA+BH for row/column selection. In the element-wise setting (Table S.5), our method is the fastest among those compared, running approximately 5 times faster than MTA and up to 1500 times faster than DRC+BH.

G Details of real data analysis

The EEG data analyzed is from Temple University Hospital (TUH), publicly available online at https://isip.piconepress.com/projects/tuh_eeg/index.shtml. The license information of this dataset is not provided on the website. Before analysis, the data needs to be preprocessed. For subjects with seizure events, we extract the segment between the seizure start and end times. For control subjects, the middle half of the recording duration (from 1/4 to 3/4) is selected. Then, signals with absolute values exceeding 1000 are removed to eliminate artifacts. Finally, all recordings are truncated to match the shortest duration (2.5 seconds) across subjects. With a sampling frequency of 20Hz, this results in data matrices of 19 electrodes by 50 time points for each subject. To account for potential effects from available covariates (sex and age), we first regress the preprocessed EEG signals against age and sex using ordinary least squares (Li & Zhang 2017). When splitting the data, we use stratified splits by response. Each split preserves the original

Table S.1: Model 3: Row- and column-wise FDR and power (tuples (FDR, Power) in percent) for $(n, p, p_1, q, q_1) = (2000, 50, 10, 50, 10)$.

Test	Covariance	FDR Level	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$
Row	AR	0.2	(20.4, 100)	(21.9, 100)	(19.6, 96.5)
	AR	0.15	(13.2, 100)	(15.2, 99.9)	(13.0, 94.3)
	AR	0.1	(12.6, 100)	(13.8, 99.8)	(10.3, 90.8)
	AR	0.05	(6.0, 100)	(7.8, 99.8)	(6.7, 89.9)
	CS	0.2	(19.1, 100)	(20.6, 99.4)	(18.9, 86.7)
	CS	0.15	(15.0, 100)	(15.3, 98.9)	(14.3, 85.3)
	CS	0.1	(13.5, 100)	(14.3, 98.5)	(11.3, 81.2)
	CS	0.05	(8.1, 100)	(7.0, 98.2)	(8.5, 80.0)
Column	AR	0.2	(18.2, 100)	(19.9, 99.9)	(21.0, 95.4)
	AR	0.15	(12.8, 100)	(15.1, 99.9)	(15.5, 94.3)
	AR	0.1	(11.7, 100)	(13.8, 99.8)	(11.8, 90.9)
	AR	0.05	(5.3, 100)	(8.7, 99.8)	(8.0, 90.0)
	CS	0.2	(19.1, 100)	(21.3, 99.3)	(19.7, 89.6)
	CS	0.15	(14.5, 100)	(14.8, 98.8)	(13.7, 85.2)
	CS	0.1	(12.9, 100)	(13.0, 97.8)	(9.6, 81.2)
	CS	0.05	(5.9, 100)	(7.5, 97.4)	(7.6, 80.2)

Table S.2: Scalability and robustness for Model 1&2 under AR covariance (tuples are $(\mathrm{FDR}, \mathrm{Power})$ in percent).

(n, p)	Distribution	Element-wise test			Row-wise test		
		$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$
				Model 1			
(2000, 50)	Uniform	(26.1, 100)	(27.1, 99.8)	(18.3, 80.0)	(21.2, 100)	(22.3, 100)	(19.0, 95.3)
(2000, 50)	Laplace	(24.9, 100)	(26.3, 99.9)	(25.3, 82.3)	(20.2, 100)	(23.0, 100)	(19.6, 95.9)
(2000, 50)	Logistic	(24.7, 100)	(26.0, 99.8)	(20.8, 80.7)	(18.7, 100)	(22.8, 100)	(22.4, 95.5)
(2000, 50)	Normal	(25.3, 100)	(23.1, 99.9)	(20.9, 81.9)	(21.2, 100)	(17.4, 100)	(17.8, 95.3)
(4000, 100)	Normal	(24.4, 100)	(26.2, 100)	(24.5, 97.5)	(18.6, 100)	(17.9, 100)	(19.7, 99.5)
				Model 2			
(2000, 50)	Uniform	(24.0, 100)	(26.2, 99.9)	(18.8, 79.9)	(19.8, 100)	(21.8, 100)	(18.4, 94.7)
(2000, 50)	Laplace	(24.8, 100)	(27.1, 99.8)	(24.4, 81.5)	(20.5, 100)	(22.8, 100)	(20.4, 95.4)
(2000, 50)	Logistic	(25.4, 100)	(28.1, 99.8)	(21.0, 79.4)	(20.0, 100)	(22.7, 100)	(21.8, 95.3)
(2000, 50)	Normal	(24.6, 100)	(22.0, 99.8)	(21.1, 81.7)	(21.1, 100)	(16.6, 100)	(17.2, 95.2)
(4000, 100)	Normal	(23.2, 100)	(24.8, 100)	(24.8, 97.7)	(17.8, 100)	(18.6, 100)	(20.3, 99.3)

Table S.3: Row-wise FDR and power (tuples are $(\text{FDR}, \text{Power})$ in percent) for Models 1-–3 under CS when $(n, p, p_1, q, q_1) = (2000, 50, 10, 50, 10)$.

Model	FDR Level	Method	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$
Model 1	0.2	Proposed	(22.7, 100)	(20.4, 99.6)	(18.1, 90.4)
		MVSA+BH	(15.5, 98.9)	(14.6, 61.1)	(16.7, 15.7)
	0.15	Proposed	(16.6, 100)	(16.4, 99.3)	(13.3, 86.1)
		MVSA+BH	(11.7, 98.6)	(10.0, 54.5)	(12.5, 12.1)
	0.1	Proposed	(14.1, 100)	(14.8, 98.9)	(9.7, 83.2)
		MVSA+BH	(8.5, 98.2)	(6.2, 48.2)	(9.7, 8.8)
	0.05	Proposed	(7.5, 100)	(8.5, 98.8)	(7.3, 82.1)
		MVSA+BH	(5.1, 96.6)	(3.3, 36.2)	(7.0, 4.9)
Model 2	0.2	Proposed	(22.3, 100)	(19.9 99.6)	(18.9 89.8)
		MVSA+BH	(15.0, 99.1)	(15.0, 62.0)	(16.2, 16.0)
	0.15	Proposed	(16.4, 100)	(16.1, 99.4)	(13.1, 86.2)
		MVSA+BH	(11.5, 98.9)	(10.6, 55.1)	(12.6, 12.4)
	0.1	Proposed	(13.8, 100)	(14.6, 98.9)	(9.8, 83.2)
		MVSA+BH	(8.6, 98.4)	(7.1, 47.9)	(9.6, 9.2)
	0.05	Proposed	(7.7, 100)	(9.1, 98.8)	(7.5, 82.4)
		MVSA+BH	(4.8, 96.5)	(3.9, 37.4)	(6.8, 5.0)
Model 3	0.2	Proposed	(19.1, 100)	(20.6 99.4)	(18.9 86.7)
		MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
	0.15	Proposed	(15.0, 100)	(15.3, 98.9)	(14.3, 85.3)
		MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
	0.1	Proposed	(13.5, 100)	(14.3, 98.5)	(11.3, 81.2)
		MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
	0.05	Proposed	(8.1, 100)	(7.0, 98.2)	(8.5, 80.0)
		MVSA+BH	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)

Table S.4: Computation time (seconds) for row selection with $(n, p, p_1, q, q_1) = (2000, 50, 10, 50, 10)$. Entries are averaged for $\rho = 0.4$.

Model	AR		CS	
	Our	MVSA+BH	Our	MVSA+BH
Model 1	2.0	454.7	2.5	452.0
Model 2	2.0	769.1	2.5	770.6
Model 3	2.0	765.1	2.4	759.4

Table S.5: Computation time (seconds) for element-wise selection with $(n, p, p_1, q, q_1) = (1000, 20, 6, 20, 6)$. Entries are averaged for $\rho = 0.4$.

Model		Our	DRC+BH	HSIC-BH	MTA	Knockoff-SIR	Knockoff-Lasso	MVSA+BH
Model 1	AR	0.1	154.9	120.9	1.4	10.2	8.8	24.9
	CS	0.2	157.2	124.1	1.4	10.0	8.5	25.0
Model 2	AR	0.2	156.0	125.8	1.4	10.1	9.5	25.5
	CS	0.2	158.2	127.3	1.4	9.9	9.0	25.5
Model 3	AR	0.2	157.9	71.3	0.9	6.3	15.2	25.1
	CS	0.2	154.5	83.2	1.0	6.2	11.9	25.1

patient–control proportion by dividing the two groups evenly. Then, our Algorithm 1 is applied to the residuals to select important time points and electrodes associated with seizure status.

We use the multiple data splitting procedure of Dai et al. (2023*a*) to mitigate the instability issue that can arise from a single random split. For completeness, we briefly summarize the procedure. Let m denote the number of independent random data splits. For $\ell = 1, \dots, p$, let $\widehat{S}_{\mathrm{row}}^{(\ell)}$ be the set of selected rows. For row j, the inclusion rate $\hat{I}_j$ is computed as

$$\hat{I}_j = \frac{1}{m} \sum_{\ell=1}^{m} \frac{\mathbb{1}(j \in \widehat{S}_{\mathrm{row}}^{(\ell)})}{|\widehat{S}_{\mathrm{row}}^{(\ell)})| \vee 1}.$$

Next, sort the inclusion rates in non-decreasing order, $0 \leq \hat{I}_{(1)} \leq \hat{I}_{(2)} \leq \cdots \leq \hat{I}_{(p)}$. We choose the largest $h \in \{1, \dots, p\}$ such that $\hat{I}_{(1)} + \cdots \hat{I}_{(h)} \leq a$, where a is the target FDR level. Finally, the selected rows are $\{j : \hat{I}_j > \hat{I}_h\}$. The same steps are applied to select columns.

In the main text, we focus on analyses in the time domain. Here, we provide a complementary analysis in the frequency domain. For each subject and each electrode, we center the time-domain signal and compute the discrete Fourier transform (DFT). The corresponding periodogram is computed as the squared modulus of the DFT (Shumway & Stoffer 2005). With 50 time points sampled at 20 Hz, the data yields a frequency resolution of $20/50 = 0.4$ Hz and a Nyquist frequency of 10 Hz. After removing the 0 Hz component, the transformed predictor has dimensions of 19 electrodes $\times$ 25 frequencies.

We then apply our selection procedure. The four selected electrodes (F3, P7, O1, P4) are visualized in Figure S.1. They are spatially close to electrodes selected with time-domain data (C3, P3, P4) and are consistent with prior neurophysiological findings. O1 and P4 are located over the occipital and parietal lobe regions, which are involved in various seizure onset (Ristić et al. 2012). P7 targets the posterior temporal region, which is sensitive to temporal lobe epilepsy. F3 captures frontal lobe activity, and it is a common seizure origin in frontal lobe epilepsy (García-López et al. 2021).

The selected frequencies (4.4, 6.8, 8.4, 8.8, 9.2 Hz) fall in the Theta (4-8 Hz) and Alpha (8-12 Hz) bands. Fox et al. (2022) suggests that rhythmic Theta activity can be a marker of epileptogenicity in temporal lobe epilepsy. Our selection of 4.4 and 6.8 Hz aligns with this fact. In addition, the selected electrode P7 is located in the temporal lobe. The selection of frequencies in the 8.4-9.2 Hz range is consistent with findings in Abela et al. (2019), which states that the slow alpha rhythm associates with epilepsy patients. Notably, Abela et al. (2019) also indicates that this slower alpha occurred over frontal and occipital regions, which include the selected electrodes F3 and O1.

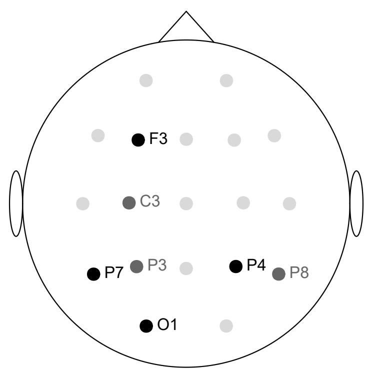


Figure S.1: The selected electrodes using transformed (original) EEG signals are black (gray) circles.

References

Abela, E., Pawley, A. D., Tangwiriyasakul, C., Yaakub, S. N., Chowdhury, F. A., Elwes, R. D., Brunnhuber, F. & Richardson, M. P. (2019), 'Slower alpha rhythm associates with poorer seizure control in epilepsy', *Annals of clinical and translational neurology* **6**(2), 333–343.

Bellec, P. C. & Zhang, C.-H. (2021), 'Second-order Stein: SURE for SURE and other applications in high-dimensional inference', *The Annals of Statistics* **49**(4), 1864–1903.

Cai, T. T., Liu, W. & Zhou, H. H. (2016), 'Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation', *The Annals of Statistics* **44**(2), 455 – 488.
URL: *https://doi.org/10.1214/13-AOS1171*

Chernozhukov, V., Chetverikov, D. & Kato, K. (2013), 'Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors', *The Annals of Statistics* **41**(6), 2786 – 2819.
URL: *https://doi.org/10.1214/13-AOS1161*

Chernozhukov, V., Chetverikov, D. & Kato, K. (2017), 'Central limit theorems and bootstrap in high dimensions', *The Annals of Probability* **45**(4), 2309 – 2352.

Dai, C., Lin, B., Xing, X. & Liu, J. S. (2023*a*), 'False discovery rate control via data splitting', *Journal of the American Statistical Association* **118**(544), 2503–2520.

Dai, C., Lin, B., Xing, X. & Liu, J. S. (2023*b*), 'A scale-free approach for false discovery rate control in generalized linear models', *Journal of the American Statistical Association* **118**(543), 1551–1565.

Dutilleul, P. (1999), 'The mle algorithm for the matrix normal distribution', *Journal of statistical computation and simulation* **64**(2), 105–123.

Fox, J., Samudra, N., Johnson, M., Humayun, M. J. & Abou-Khalil, B. W. (2022), 'Temporal intermittent rhythmic theta activity (tirta): A marker of epileptogenicity?', *Eneurologicalsci* **29**, 100433.

García-López, B., Sueiras-Gil, M., Gómez-Menéndez, A. I., Vázquez-Sánchez, F., Lloria-Gil, M. C., González-Bernal, J. J., González-Santos, J., Santamaría-Pelaéz, M., Soto-Cámara, R. & Wesenberg Kjaer, T. (2021), 'Morphological description of frontal eeg interictal and ictal discharges in an adult cohort of 175 patients', *Journal of Clinical Medicine* **10**(6), 1219.

Leng, C. & Tang, C. Y. (2012), 'Sparse matrix graphical models', *Journal of the American Statistical Association* **107**(499), 1187–1200.

Li, L. & Zhang, X. (2017), 'Parsimonious tensor response regression', *Journal of the American Statistical Association* **112**(519), 1131–1146.

Lyu, X., Sun, W. W., Wang, Z., Liu, H., Yang, J. & Cheng, G. (2019), 'Tensor graphical model: Non-convex optimization and statistical inference', *IEEE transactions on pattern analysis and machine intelligence* **42**(8), 2024–2037.

Min, K., Mai, Q. & Zhang, X. (2022), 'Fast and separable estimation in high-dimensional tensor gaussian graphical models', *Journal of Computational and Graphical Statistics* **31**(1), 294–300.

Ristić, A. J., Alexopoulos, A. V., So, N., Wong, C. & Najm, I. M. (2012), 'Parietal lobe epilepsy: the great imitator among focal epilepsies', *Epileptic Disorders* **14**(1), 22–31.

Shumway, R. H. & Stoffer, D. S. (2005), 'Time series analysis and its applications (springer texts in statistics)'.

Veraar, M. (2009), 'Correlation inequalities and applications to vector-valued gaussian random variables and fractional brownian motion', *Potential Analysis* **30**(4), 341–370.

Vershynin, R. (2018), *High-dimensional probability: An introduction with applications in data science*, Vol. 47, Cambridge university press.

Wainwright, M. J. (2019), *High-dimensional statistics: A non-asymptotic viewpoint*, Vol. 48, Cambridge university press.

Xing, X., Zhao, Z. & Liu, J. S. (2023), 'Controlling false discovery rate using gaussian mirrors', *Journal of the American Statistical Association* **118**(541), 222–241.

Xu, P., Zhang, T. & Gu, Q. (2017), Efficient Algorithm for Sparse Tensor-variate Gaussian Graphical Models via Gradient Descent, *in* A. Singh & J. Zhu, eds, 'Proceedings of the 20th International Conference on Artificial Intelligence and Statistics', Vol. 54 of *Proceedings of Machine Learning Research*, PMLR, pp. 923–932.

Zhao, Z. & Xing, X. (2022), 'On the testing of multiple hypothesis in sliced inverse regression', *arXiv preprint arXiv:2210.05873* .