Prior Knowledge Guided Ultra-high Dimensional Variable Screening with Application to Neuroimaging Data

Jie He and Jian Kang

Supplementary Material

S1 Additional Theoretical Results

Proposition 1. To PMS statistics, we have

$$\widehat{\boldsymbol{\beta}}^{\mathrm{PMS}} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}(\mathbf{Y} - \mathbf{X}\boldsymbol{\mu}), \tag{S1.1}$$

where $\boldsymbol{\Omega} = (\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n)^{-1}$.

This results can be straightforwardly derived from the well-known Sherman–Morrison–Woodbury formula. The similar result has been adopted to compute the ridge regression (Lu et al. 2013). The detailed proof of Proposition 1 is in section S3.

Lemma 1. To a positive definite matrix $\boldsymbol{\Lambda}_K$ and a $k \times p$ matrix K, assume that

$$\lambda_{\min}(\boldsymbol{\Lambda}_K) \geq c_k^{-1} n^{-\tau_k}, \lambda_{\max}(\mathbf{K}\mathbf{K}^{\mathrm{T}}) \leq c_{k_1} n^{\tau_{k_1}} \text{ and } \lambda_{\min}(\mathbf{K}\mathbf{K}^{\mathrm{T}}) \geq c_{k_2}^{-1} n^{-\tau_{k_2}},$$

where $c_k, c_{k_1}, c_{k_2}, \tau_k, \tau_{k_1}$ and $\tau_{k_2} > 0$ are constants. Then there exists a $g > 0$, such that

$$(\mathbf{G}_1)_{j,j}/(\mathbf{G}_2)_{j,j} = O(n^g), j = 1, \ldots, k$$

where $(\mathbf{G})_{j,j}$ refers to the jth diagonal element of matrix $\mathbf{G}$, $\mathbf{G}_1$ and $\mathbf{G}_2$ take the forms of

$$\mathbf{G}_1 = \mathbf{X}\mathbf{K}^{\mathrm{T}}(\mathbf{X}\mathbf{K}^{\mathrm{T}}\mathbf{\Lambda}_K\mathbf{K}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n)^{-1}\mathbf{K}\mathbf{X}^{\mathrm{T}}, \mathbf{G}_2 = \mathbf{X}\mathbf{K}^{\mathrm{T}}(\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n)^{-1}\mathbf{K}\mathbf{X}^{\mathrm{T}}.$$

Based on Lemma 1, we obtain the theoretical properties of PMS with prior on selection and group level importance.

S2 Additional Technical Conditions

Additional conditions for Theorem 1.

A1. Let $\mathbf{Z} = \mathbf{X}\mathbf{\Sigma}^{-1/2}$, there are some $c_1 > 1$ and $C_1 > 0$ such that

$$P\left\{\lambda_{\max}\left(p^{-1}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right) > c_1 \text{ or } \lambda_{\min}\left(p^{-1}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right) < c_1^{-1}\right\} \leq \exp\left(-C_1 n\right),$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and largest eigenvalues of a matrix respectively, $\mathbf{\Sigma} = \mathrm{Cov}(\mathbf{x}_i)$ with $\mathbf{x}_i$ be the ith row of $\mathbf{X}$.

A2. For some $c_2, c_3, c_4, c_5, c_6 > 0$ and $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5 \geq 0$, we have

$$\lambda_{\max}\left(\mathbf{\Lambda}\right) \leq c_2 n^{\tau_1}, \lambda_{\min}\left(\mathbf{\Lambda}\right) \geq c_3^{-1} n^{-\tau_2}, \lambda_{\max}\left(\mathbf{\Sigma}\mathbf{\Lambda}\right) \leq c_4 n^{\tau_3},\ \lambda_{\min}\left(\mathbf{\Sigma}\mathbf{\Lambda}\right) \geq c_5^{-1} n^{-\tau_4},$$

$$\text{and } \mathrm{cond}\left(\mathbf{\Sigma}\right) \leq c_6 n^{\tau_5},$$

where $\mathrm{cond}(\cdot) = \lambda_{\max}(\cdot)/\lambda_{\min}(\cdot)$ is the condition number of a matrix.

A3. The random error vector $\boldsymbol{\varepsilon}$ is independent with $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$ and has mean 0 and standard deviation σ. $\boldsymbol{\varepsilon}/\sigma$ has q-exponential tails with some function $q(\cdot)$.

Following are additional regularization conditions for Theorem 2.

B2. There exist some constants $\tau_{s_1}, \tau_{s_2}, c_{s_1}$ and $c_{s_2} > 0$, such that

$$\lambda_{\max}(\mathbf{\Lambda}_{\mathcal{S}}) \leq c_{s_1} n^{\tau_{s_1}} \quad \text{and} \quad \lambda_{\min}(\mathbf{\Lambda}_{\mathcal{S}}) \geq c_{s_2}^{-1} n^{-\tau_{s_2}}.$$

B3. $\mathbf{C}_{\mathcal{S}}$ is a $q \times p$ matrix such that $\mathbf{X}_{\mathcal{S}} = \mathbf{X}\mathbf{C}_{\mathcal{S}}^{\mathrm{T}}$, let $\lambda_j^c = \left(\lambda_{j1}^c, \ldots, \lambda_{jq}^c\right)^{\mathrm{T}}$ be the jth row of matrix $\mathbf{C}_{\mathcal{S}}\mathbf{\Lambda}^{-1/2}$, then to some constant $\epsilon_1 > 0$, we have

$$\sum_{u=1}^{q}\sum_{v\neq u}\left|\lambda_{ju}^c\lambda_{jv}^c\right| = O\left(\sqrt{\log n}n^{\epsilon_1}\right),$$

as well as

$$\sum_{u=1}^{q}\left(\lambda_{ju}^c\right)^2 = O\left(n^{1-2\tau_{s_1}-\tau_{s_2}-3\tau_3-3\tau_4-3\tau_5-2\bar{\gamma}}\right),$$

where τ_3, τ_4 and τ_5 are defined in condition A2.

This part lists some additional conditions for Theorem 3.

C1. To some constants $\bar{c}_8, \bar{c}_9, \tau_6$ and $\tau_7 > 0$, we have

$$\lambda_{\max}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right) \leq \bar{c}_8 n^{\tau_6} \quad \text{and} \quad \lambda_{\min}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right) \geq \left(\bar{c}_9\right)^{-1} n^{-\tau_7}.$$

C3. There exist some constants $c_{b_1}, c_{b_2}, \tau_{b_1}$ and $\tau_{b_2} > 0$, such that

$$\lambda_{\max}(\mathbf{\Lambda}_B) \leq c_{b_1} n^{\tau_{b_1}} \quad \text{and} \quad \lambda_{\max}(\mathbf{\Lambda}_{\mathbf{B}}) \leq c_{b_2} n^{\tau_{b_2}}.$$

C4. Let $\boldsymbol{\lambda}_j^{\mathbf{B}} = \left(\lambda_{j1}^{\mathbf{B}}, \ldots, \lambda_{jm}^{\mathbf{B}}\right)^{\mathrm{T}}$ be the jth row of matrix $\mathbf{B}\mathbf{\Lambda}^{-1/2}$, then to some $\epsilon_2 > 0$, we have

$$\sum_{u=1}^{m}\sum_{v\neq u}\left|\lambda_{ju}^{\mathbf{B}}\lambda_{jv}^{\mathbf{B}}\right| = O\left(n^{\epsilon_2}\sqrt{\log n}\right),$$

and

$$\sum_{u=1}^{m}(\lambda_{ju}^{\mathbf{B}})^2 = O\left(n^{1-3\tau_{b_1}-\tau_{b_2}-2\tau_3-3\tau_4-5\tau_5-4\tau_6-5\tau_7-2\bar{\gamma}_2}\right),$$

where τ_3, τ_4 and τ_5 are defined in condition A2.

S3 Detailed Theoretical Proofs

The proof of Proposition 1

This part lists the detailed proof of the equivalence of two PMS statistics expressions.

Proof: To PMS screening statistics

$$\begin{aligned}\boldsymbol{\nu} =&(\theta\boldsymbol{\Lambda}^{-1}+\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\theta\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}+\mathbf{X}^{\mathrm{T}}\mathbf{Y})\\ =&(\theta\boldsymbol{\Lambda}^{-1}+\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\theta\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}+(\theta\boldsymbol{\Lambda}^{-1}+\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\\ :=&\boldsymbol{\nu}_1+\boldsymbol{\nu}_2.\end{aligned}$$

One the one hand, from the Sherman-Morrison-Woodbury formula

$$(\mathbf{A}+\mathbf{UDV})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{U}(\mathbf{D}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1},$$

let $\mathbf{A}=\theta\boldsymbol{\Lambda}^{-1}$, $\mathbf{U}=\mathbf{X}^{\mathrm{T}}$, $\mathbf{D}=\mathbf{I}_n$ and $\mathbf{V}=\mathbf{X}$, we obtain

$$\begin{aligned}\boldsymbol{\nu}_1 =&(\theta\boldsymbol{\Lambda}^{-1}+\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\theta\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}\\ =&\{\frac{1}{\theta}\boldsymbol{\Lambda}-\frac{1}{\theta}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n)^{-1}\mathbf{X}\boldsymbol{\Lambda}\}\theta\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}\\ =&\boldsymbol{\mu}-\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n)^{-1}\mathbf{X}\boldsymbol{\mu}.\end{aligned}$$

On the other hand,

$$\boldsymbol{\nu}_2=(\theta\boldsymbol{\Lambda}^{-1}+\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}=\{(\theta\mathbf{I}_p+\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\Lambda})\boldsymbol{\Lambda}^{-1}\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}=\boldsymbol{\Lambda}(\theta\mathbf{I}_p+\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\Lambda})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}.$$

From the Sherman-Morrison-Woodbury formula, choosing $\mathbf{A}=\theta\mathbf{I}_p$, $\mathbf{U}=\mathbf{X}^{\mathrm{T}}$, $\mathbf{D}=\mathbf{I}_n$ and $\mathbf{V}=\mathbf{X}\boldsymbol{\Lambda}$, we have

$$\begin{aligned}\theta(\theta\mathbf{I}_p+\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\Lambda})^{-1} =&\theta\{\frac{1}{\theta}\mathbf{I}_p-\frac{1}{\theta}\mathbf{X}^{\mathrm{T}}(\mathbf{I}_n+\frac{1}{\theta}\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}\boldsymbol{\Lambda}\frac{1}{\theta}\mathbf{I}_p\}\\ =&\mathbf{I}_p-\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}\boldsymbol{\Lambda}.\end{aligned}$$

So

$$
\begin{aligned}
\theta(\theta\mathbf{I}_p+\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\Lambda})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} =&\{\mathbf{I}_p-\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}\boldsymbol{\Lambda}\}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\\
=&\mathbf{X}^{\mathrm{T}}\mathbf{Y}-\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\\
=&\mathbf{X}^{\mathrm{T}}\mathbf{Y}-\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}-\theta\mathbf{I}_n)\mathbf{Y}\\
=&\mathbf{X}^{\mathrm{T}}\mathbf{Y}-\mathbf{X}^{\mathrm{T}}\mathbf{Y}+\theta\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{Y}\\
=&\theta\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{Y}.
\end{aligned}
$$

That is to say

$$
(\theta\mathbf{I}_p+\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\Lambda})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}=\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{Y}.
$$

Thus

$$
\boldsymbol{\nu}_2=\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{Y}.
$$

To sum up, we obtain that

$$
\begin{aligned}
\boldsymbol{\nu} =&\boldsymbol{\nu}_1+\boldsymbol{\nu}_2\\
=&\boldsymbol{\mu}-\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n)^{-1}\mathbf{X}\boldsymbol{\mu}+\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}(\theta\mathbf{I}_n+\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{Y}\\
=&\boldsymbol{\mu}+\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n)^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\mu})\\
=&\boldsymbol{\mu}+\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}(\mathbf{Y}-\mathbf{X}\boldsymbol{\mu}),
\end{aligned}
$$

where $\boldsymbol{\Omega}=(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n)^{-1}$. □

Proof of Theorem 1

Proof:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{\mathrm{PMS}} &= \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\mu}) \\
&= \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\mu}) \\
&= \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\mu}) + \boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon} \\
&= \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}) + \boldsymbol{\Lambda}\mathbf{X}^{T}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}.
\end{aligned}$$

By singular value decomposition, we have $\mathbf{Z} = \mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{V} \in \mathcal{O}(n)$, $\mathbf{D}$ is an $n \times n$ diagonal matrix and $\mathbf{U} \in V_{n,p}$, $\mathcal{O}(n)$ is the orthogonal group and Stiefel manifold $V_{n,p} = \left\{\mathbf{B} \in R^{p\times n} : \mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}_n\right\}$. So

$$\begin{aligned}
&\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda} \\
=&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}\left(\mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda} \\
=&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}\left\{\mathbf{V}\mathbf{D}\left(\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U} + \theta\mathbf{D}^{-2}\right)\mathbf{D}\mathbf{V}^{\mathrm{T}}\right\}^{-1}\mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda} \\
=&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}\mathbf{V}\mathbf{D}^{-1}\left\{\mathbf{U}^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\mathbf{U} + \theta\mathbf{D}^{-2}\right\}^{-1}\mathbf{D}^{-1}\mathbf{V}^{\mathrm{T}}\mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda} \\
=&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\left\{\mathbf{U}^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\mathbf{U} + \theta\mathbf{D}^{-2}\right\}^{-1}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}
\end{aligned}$$

Denote $\mathbf{A} = \left\{\mathbf{U}^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\mathbf{U}\right\}^{1/2}$, we have

$$\begin{aligned}
&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\left\{\mathbf{U}^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\mathbf{U} + \theta\mathbf{D}^{-2}\right\}^{-1}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda} \\
=\;&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \theta\mathbf{D}^{-2}\right)^{-1}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda} \\
=\;&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{A}^{-\mathrm{T}}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \theta\mathbf{D}^{-2}\right)\mathbf{A}^{-1}\mathbf{A}\right\}^{-1}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda} \\
=\;&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\left\{\mathbf{A}^{\mathrm{T}}\left(\mathbf{I}_n + \theta\mathbf{A}^{-T}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)\mathbf{A}\right\}^{-1}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda} \\
=\;&\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}\left(\mathbf{I}_n + \theta\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)^{-1}\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}
\end{aligned}$$

(S3.2)

$$\begin{aligned}
&= \boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}\left(\mathbf{I}_n+\theta\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)^{-1}\mathbf{A}^{-T}\mathbf{U}^{T}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\\
&= \boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}\left\{\mathbf{I}_n+\sum_{k=1}^{\infty}(-\theta)^k\left(\mathbf{A}^{-T}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)^k\right\}\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\\
&= \boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}\mathbf{A}^{-T}\mathbf{U}^{T}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}+\mathbf{M} \qquad \text{(S3.3)}\\
&:= \boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}+\mathbf{M},
\end{aligned}$$

where $\mathbf{H}=\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\left\{\mathbf{U}^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\mathbf{U}\right\}^{-1/2}$. It is obvious that

$$\begin{aligned}
\|\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\| &\le \lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)\le\lambda_{\max}(\mathbf{D}^{-2})\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{A}^{-1})\\
&= \lambda_{\max}(\mathbf{D}^{-2})\lambda_{\max}\left\{(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\right\}\\
&= \left\{\lambda_{\min}\left(\mathbf{D}^2\right)\right\}^{-1}\left\{\lambda_{\min}\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)\right\}^{-1}.
\end{aligned}$$

On the one hand,

$$\begin{aligned}
\lambda_{\min}\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right) &= \lambda_{\min}\left\{\mathbf{U}^{T}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\mathbf{U}\right\}\ge\lambda_{\min}\left(\mathbf{U}^{\mathrm{T}}\mathbf{U}\right)\lambda_{\min}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\\
&= \lambda_{\min}\left(\boldsymbol{\Sigma}\boldsymbol{\Lambda}\right)\ge c_5^{-1}n^{-\tau_4}.
\end{aligned}$$

On the other hand, from *A1*, we have

$$P\left(\lambda_{\min}\left(\mathbf{D}^2\right)<pc_1^{-1}\right)=P\left(p^{-1}\lambda_{\min}\left(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right)<c_1^{-1}\right)<\exp(-C_1n).$$

Thus,

$$P\left(\|\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\|\le c_1c_5p^{-1}n^{\tau_4}\right)\ge 1-\exp(-C_1n).$$

So the necessary condition for (S3.2) is that $\theta\le c_1^{-1}c_5^{-1}pn^{-\tau_4}$, which can ensure that the norm of matrix $\theta\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}$ is smaller than 1.

From the definition of $\mathbf{H}$, we can test

$$\mathbf{H}^{\mathrm{T}}\mathbf{H}=\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}=\mathbf{A}^{-1}\left\{\mathbf{U}^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\mathbf{U}\right\}\mathbf{A}^{-1}=\mathbf{I}_n,$$

which indicates that $\mathbf{H} \in V_{n,p}$. As

$$\lambda_{\max}\left\{\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)^{k}\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{T}\mathbf{\Sigma}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\right\}$$
$$\leq\lambda_{\max}\left\{\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)^{k}\right\}\lambda_{\max}\left(\mathbf{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{\Lambda}^{1/2}\right)=\lambda_{\max}\left\{\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)^{k}\right\}\lambda_{\max}(\mathbf{\Lambda}),$$

thus,

$$\lambda_{\max}\left(\mathbf{M}\right)\leq\lambda_{\max}\left(\mathbf{\Lambda}\right)\sum_{k=1}^{\infty}\theta^{k}\left\{\lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)\right\}^{k}\leq\left\{\theta\lambda_{\max}(\mathbf{\Lambda})\lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)\right\}/\left\{1-\theta\lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)\right\}.$$

$$P\left(\lambda_{\max}(\mathbf{M})>\theta c_1c_2c_5n^{\tau_1+\tau_4}/(p-\theta c_1c_5n^{\tau_4})\right)$$
$$\leq P\left(\left\{\theta\lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)\lambda_{\max}\left(\mathbf{\Lambda}\right)\right\}/\left\{1-\theta\lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)\right\}>\theta c_1c_2c_5n^{\tau_1+\tau_4}/(p-\theta c_1c_5n^{\tau_4})\right)$$
$$\leq P\left(\left\{\theta\lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)c_2n^{\tau_1}\right\}/\left\{1-\theta\lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)\right\}>\theta c_1c_2c_5n^{\tau_1+\tau_4}/(p-\theta c_1c_5n^{\tau_4})\right)$$
$$\leq P\left(\lambda_{\max}\left(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}\right)>c_1c_5n^{\tau_4}/p\right)<\exp(-C_1n).$$

So

$$\max_{i\in\{1,\ldots,p\}}\left|\mathbf{e}_i^{\mathrm{T}}\mathbf{M}\mathbf{\Lambda}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{\beta}\right)\right|^2\leq\lambda_{\max}\left(\mathbf{M}^2\right)\|\mathbf{\Lambda}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{\beta}\right)\|^2\leq\lambda_{\max}\left(\mathbf{M}^2\right)c_7^2n^{2\gamma}/p.$$

Thus

$$P\left(\max_{i\in\{1,\ldots,p\}}\left|\mathbf{e}_i^{\mathrm{T}}\mathbf{M}\mathbf{\Lambda}^{-1}(\boldsymbol{\mu}-\boldsymbol{\beta})\right|>\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}\frac{\theta c_1c_2c_5c_7n^{\tau_1+\tau_3+2\tau_4+2\gamma-\nu-1}}{\sqrt{p}-\theta c_1c_5n^{\tau_4}/\sqrt{p}}\right)<\exp(-C_1n).$$

Let θ satisfying that

$$\frac{\theta c_1c_2c_5c_7n^{\tau_1+\tau_3+2\tau_4+2\gamma-\nu-1}}{\sqrt{p}-\theta c_1c_5n^{\tau_4}/\sqrt{p}}=o(1),$$

above conclusion can be summarized as

$$P\left(\max_{i\in\{1,\ldots,p\}}\left|\mathbf{e}_i^{\mathrm{T}}\mathbf{M}\mathbf{\Lambda}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{\beta}\right)\right|>o(1)\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}\right)<\exp(-C_1n).$$

From the conclusion of Wang and Leng (2016), we know that

$$P\left(\mathbf{b}^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{b} > \tilde{c}_1 \frac{n^{1+(\tau_3+\tau_4)}}{p}\right) < 2\exp(-Cn),$$

for any $p \times 1$ vector $\mathbf{b}$ satisfying $\|\mathbf{b}\| = 1$. So

$$P\left(\bar{\boldsymbol{\lambda}}_i^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\bar{\boldsymbol{\lambda}}_i > \tilde{c}_1\|\bar{\boldsymbol{\lambda}}_i\|^2 \frac{n^{1+(\tau_3+\tau_4)}}{p}\right) < 2\exp(-Cn),$$

where $\bar{\boldsymbol{\lambda}}_i = \left(\bar{\lambda}_{i1}, \ldots, \bar{\lambda}_{ip}\right)^T$ is the ith row of matrix $\boldsymbol{\Lambda}^{1/2}$. From *A4*, we can get

$$P\left(\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\mathbf{e}_i > \bar{c}_1 \frac{n^{1+(\tau_3+\tau_4)+\nu}}{p}\right) < 2\exp(-Cn).$$

In addition, we also know that

$$P\left(\left|\mathbf{e}_i^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_j\right| \leq \frac{M}{\sqrt{\log n}} \frac{n^{1+(\tau_3+\tau_4)-\alpha}}{p}\right) \geq 1 - O\left\{\exp\left(-\frac{Cn^{1-2\alpha}}{2\log n}\right)\right\},$$

for any $0 < \alpha < 1/2$ and $j \neq i; i, j = 1, \ldots, p$. Choosing $\alpha = 2\tau_3 + 2\tau_4 + \gamma$, above result can be summarized as

$$P\left(\left|\mathbf{e}_i^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_j\right| \leq \frac{M}{\sqrt{\log n}} \frac{n^{1-(\tau_3+\tau_4)-\gamma}}{p}\right) \geq 1 - O\left\{\exp\left(-\frac{Cn^{1-4\tau_3-4\tau_4-2\gamma}}{2\log n}\right)\right\}.$$

Thus,

$$\begin{aligned}
\left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\mathbf{e}_j\right| = \left|\bar{\boldsymbol{\lambda}}_i^{T}\mathbf{H}\mathbf{H}^{T}\bar{\boldsymbol{\lambda}}_j\right| &= \left|\sum_{u=1}^{p}\sum_{v=1}^{p}\bar{\lambda}_{iu}\bar{\lambda}_{jv}\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_v\right| \\
&\leq \sum_{u=1}^{p}\sum_{v=1}^{p}\left|\bar{\lambda}_{iu}\bar{\lambda}_{jv}\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_v\right| \\
&= \sum_{u=1}^{p}\sum_{v\neq u}\left|\bar{\lambda}_{iu}\bar{\lambda}_{jv}\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_v\right| + \sum_{u=1}^{p}\left|\bar{\lambda}_{iu}\bar{\lambda}_{ju}\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_u\right| \\
&\leq \frac{M}{\sqrt{\log n}} \frac{n^{1-(\tau_3+\tau_4)-2\gamma}}{p}\bar{c}_3 n^{\nu} + \frac{\bar{c}_1 n^{1+(\tau_3+\tau_4)}}{p} \frac{\bar{c}_4 n^{\nu-2\tau_3-2\tau_4-2\gamma}}{\sqrt{\log n}} \\
&\leq \frac{M^*}{\sqrt{\log n}} \frac{n^{1-(\tau_3+\tau_4)+\nu-2\gamma}}{p},
\end{aligned}$$

where $M^* = \bar{c}_3 M + \bar{c}_1\bar{c}_4$ with probability greater than $1 - O\left\{p^2\exp\left(-\frac{Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right)\right\} - 2p\exp(-C_1 n)$.

As long as

$$\log p = o\left(\frac{n^{1-4\tau_3-4\tau_4-4\gamma}}{\log n}\right),$$

we can get the conclusion that

$$P\left(\left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\mathbf{e}_j\right| \le \frac{M^*}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)+\nu-2\gamma}}{p}\right) \ge 1 - O\left\{\exp\left(-\frac{Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right)\right\}.$$

Let $\boldsymbol{\eta}(\theta) = \boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X} + \theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}$, we have

$$\eta_i(\theta) = \mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X} + \theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}.$$

Assume that

$$\mathbf{a} = \frac{\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\mathbf{e}_i}{\|\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\mathbf{e}_i\|},$$

then we have

$$\eta_i(\theta) = \|\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\mathbf{e}_i\|\sigma\omega.$$

As

$$\begin{aligned}
&\|\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\mathbf{e}_i\|^2 = \mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-2}\mathbf{X}\boldsymbol{\Lambda}\mathbf{e}_i \\
=&\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1/2}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1/2}\mathbf{X}\boldsymbol{\Lambda}\mathbf{e}_i \\
\le&\lambda_{\max}\left\{\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\right\}\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\mathbf{e}_i \\
\le&\left\{\lambda_{\min}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\right)\right\}^{-1}\mathbf{e}_i^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2} + \mathbf{M}\right)\mathbf{e}_i \\
\le&\left\{\lambda_{\min}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\right)\right\}^{-1}\left\{\mathbf{e}_i^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\right)\mathbf{e}_i + \lambda_{\max}\left(\mathbf{M}\right)\right\}.
\end{aligned}$$

It is obvious that

$$\begin{aligned}
\lambda_{\min}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\right) = \lambda_{\min}\left\{\mathbf{Z}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\mathbf{Z}^{\mathrm{T}}\right\} &\ge \lambda_{\min}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\right)\lambda_{\min}\left(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right) \\
= p\lambda_{\min}\left(\boldsymbol{\Sigma}\boldsymbol{\Lambda}\right)\lambda_{\min}\left(\frac{1}{p}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right) &\ge c_1^{-1}c_5^{-1}pn^{-\tau_4}.
\end{aligned}$$

From previous proof, we know that

$$P\left(\lambda_{\max}(\mathbf{M}) > \frac{\theta c_1 c_2 c_5 n^{\tau_1+\tau_4}}{p - \theta c_1 c_5 n^{\tau_4}}\right) < \exp(-C_1 n),$$

and

$$P\left(\mathbf{e}_i^{\mathrm{T}} \mathbf{\Lambda}^{1/2} \mathbf{H}\mathbf{H}^{\mathrm{T}} \mathbf{\Lambda}^{1/2} \mathbf{e}_i \geq \bar{c}_1 \frac{n^{1+\tau_3+\tau_4+\nu}}{p}\right) \leq 2\exp(-Cn).$$

So,

$$P\left(\left\{\lambda_{\min}\left(\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{\mathrm{T}}\right)\right\}^{-1} \mathbf{e}_i^{\mathrm{T}} \left(\mathbf{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{\Lambda}^{1/2}\right) \mathbf{e}_i > c_1 \bar{c}_1 c_5 \frac{n^{\tau_3+2\tau_4+\nu+1}}{p^2}\right) < 3\exp(-C^* n),$$

and

$$P\left(\left\{\lambda_{\min}\left(\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{\mathrm{T}}\right)\right\}^{-1} \lambda_{\max}(\mathbf{M}) > \frac{n^{\tau_3+2\tau_4+\nu+1}}{p^2} \frac{\theta c_1^2 c_2 c_5^2 n^{\tau_1-\tau_3-\nu-1}}{1-\theta c_1 c_5 n_4^{\tau}/p}\right) < 2\exp(-C^* n),$$

where $C^* = \min\{C, C_1\}$. If θ satisfies

$$\frac{\theta c_1^2 c_2 c_5^2 n^{\tau_1-\tau_3-\nu-1}}{1-\theta c_1 c_5 n^{\tau_4}/p} = o(1),$$

we have

$$P\left(\left\{\lambda_{\min}\left(\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{\mathrm{T}}\right)\right\}^{-1} \lambda_{\max}(\mathbf{M}) > o(1) \frac{n^{\tau_3+2\tau_4+\nu+1}}{p^2}\right) < 2\exp(-C^* n).$$

Thus, above conclusions can be rewritten as

$$P\left(\| \left(\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta \mathbf{I}_n\right)^{-1} \mathbf{X}\mathbf{\Lambda}\mathbf{e}_i\|^2 > \{c_1 \bar{c}_1 c_5 + o(1)\} \frac{n^{\tau_3+2\tau_4+\nu+1}}{p^2}\right) < 5\exp(-C^* n).$$

From Assumption *A3*, we know that

$$P(|\omega| > t) = P\left(\left|\sum_{i=1}^n \frac{a_i \epsilon_i}{\sigma}\right| > t\right) \leq \exp\{1 - q(t)\}.$$

Choosing $t = \dfrac{\sqrt{C^*} n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}$, we have

$$P(|\omega| > t) < \exp\left\{1 - q\left(\frac{\sqrt{C^*} n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}.$$

So

$$P\left(|\eta_i(\theta)| > \frac{\sigma\sqrt{C^*\{c_1\bar{c}_1c_5+o(1)\}}}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}\right)$$
$$< \exp\left\{1-q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}+\exp(-C_1 n).$$

As long as

$$\log p = o\left\{q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\},$$

we have

$$P\left(\max_{i\in\{1,\dots,p\}}|\eta_i(\theta)| > \frac{\sigma\sqrt{C^*\{c_1\bar{c}_1c_5+o(1)\}}}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}\right)$$
$$<p\exp\left\{1-q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}+p\exp(-C_1 n)$$
$$<O\left[\exp\left\{1-\frac{1}{2}q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}+\exp(-\frac{C^*}{2}n)\right].$$

From

$$\widehat{\boldsymbol{\beta}}^{\mathrm{PMS}} = \boldsymbol{\mu}+\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})+\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\boldsymbol{\varepsilon},$$

we have

$$\left|\hat{\beta}_i^{\mathrm{PMS}}\right| = \left|\mu_i+\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})+\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}\right|.$$

As

$$|\mu_i-\beta_i| = \left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right| = \left|\boldsymbol{\lambda}_i^{T}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right| = \left|\sum_{j=1}^{p}\lambda_{ij}\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right|$$
$$\le \max_{j\in\{1,\dots,p\}}\left|\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right|\sum_{j=1}^{p}|\lambda_{ij}| \le \frac{c_7n^{\gamma}}{p}\frac{\bar{c}_5n^{1-(\tau_3+\tau_4)-2\gamma+\nu}}{\sqrt{\log n}} \le \frac{c_7\bar{c}_5}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}.$$

To $i\notin\mathcal{M}_0$, we have

$$|\mu_i| = |\mu_i-\beta_i| \le \frac{c_7\bar{c}_5}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}.$$

Thus,

$$\begin{aligned}
\left|\hat{\beta}_i^{\mathrm{PMS}}\right| &\leq |\mu_i| + \left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\mathbf{X}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right| + \left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}\right| \\
&= |\mu_i| + \left|\mathbf{e}_i^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}+\mathbf{M}\right)\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right| + \left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}\right| \\
&\leq |\mu_i| + \left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}\right| + \left|\mathbf{e}_i^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}+\mathbf{M}\right)\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right| \\
&\quad + \sum_{\substack{j=1 \\ j\neq i}}^{p}\left|\mathbf{e}_i^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}+\mathbf{M}\right)\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right| \\
&\leq \frac{c_7\bar{c}_5}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p} + \frac{\sigma\sqrt{C^*\left\{c_1\bar{c}_1c_5+o(1)\right\}}}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p} + \frac{\bar{c}_1n^{1+(\tau_3+\tau_4)+\nu}}{p}\frac{c_7n^{\gamma}}{p} \\
&\quad + \left\{p\left(\frac{M}{\sqrt{\log n}}+o(1)\right)\frac{n^{1-(\tau_3+\tau_4)-2\gamma+\nu}}{p}\right\}\frac{c_7n^{\gamma}}{p} \\
&= \frac{c^*}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}, \qquad \text{(S3.4)}
\end{aligned}$$

where $c^* = c_7M + c_7\bar{c}_5 + \sigma\sqrt{C^*\left\{c_1\bar{c}_1c_5+o(1)\right\}}$, with probability greater than

$$1 - O\left\{p\exp\left(\frac{-Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right)\right\} - O\left[\exp\left\{1-\frac{1}{2}q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}\right].$$

If only

$$\log p = o\left[\min\left\{\frac{n^{1-4\tau_3-4\tau_4-4\gamma}}{\log n}, q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}\right],$$

the probability of (S3.4) is greater than

$$1 - O\left\{\exp\left(\frac{-Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right)\right\} - O\left[\exp\left\{1-\frac{1}{2}q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}\right].$$

In addition, to $i \in \mathcal{M}_0$, we have

$$|\mu_i| \geq |\beta_i| - \frac{c_7\bar{c}_5}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p} \geq \left(c_8 - \frac{c_7\bar{c}_5}{\sqrt{\log n}}\right)\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p},$$

so we can get that

$$\begin{aligned}\left|\hat{\beta}_i^{\mathrm{PMS}}\right| &\geq |\mu_i| - \left|\mathbf{e}_i^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}+\mathbf{M}\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}\right)\right| - \left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}\right| \\
&\geq |\mu_i| - \left|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\boldsymbol{\epsilon}\right| - \left|\mathbf{e}_i^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}+\mathbf{M}\right)\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}\right)\right| \\
&\quad - \sum_{\substack{j=1\\ j\neq i}}^{p}\left|\mathbf{e}_i^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}+\mathbf{M}\right)\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}\right)\right| \\
&\geq \left(c_8 - \frac{c_7\bar{c}_5}{\sqrt{\log n}}\right)\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p} - \frac{\sigma\sqrt{C^*\left\{c_1\bar{c}_1c_5+o(1)\right\}}}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p} - \frac{\bar{c}_1 n^{1+(\tau_3+\tau_4)+\nu}}{p} \\
&\quad \frac{c_7 n^{\gamma}}{p} - \left[p\left\{\frac{M}{\sqrt{\log n}}+o(1)\right\}\frac{n^{1-(\tau_3+\tau_4)-2\gamma+\nu}}{p}\right]\frac{c_7 n^{\gamma}}{p} \\
&\geq \left(c_8 - \frac{\tilde{c}}{\sqrt{\log n}}\right)\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p} \geq \frac{c_8}{2}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p},\end{aligned}$$

where $\tilde{c} = c_7\bar{c}_5 + \sigma\sqrt{C^*\left\{c_1\bar{c}_1c_5+o(1)\right\}} + c_7M$, with probability greater than

$$1 - O\left\{\exp\left(\frac{-Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right)\right\} - O\left[\exp\left\{1-\frac{1}{2}q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}\right].$$

From above results,

$$\begin{aligned}&P\left(\max_{i\notin\mathcal{M}_0}\left|\hat{\beta}_i^{\mathrm{PMS}}\right| > \frac{c^*}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}\right) \\
\leq &O\left[p\exp\left(\frac{-Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right) + p\exp\left\{1-\frac{1}{2}q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}\right] \\
\leq &O\left[\exp\left(\frac{-Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right) + \exp\left\{1-\frac{1}{2}q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}\right].\end{aligned}$$

Similarly, we have

$$P\left(\min_{i\in\mathcal{M}_0}\left|\hat{\beta}_i^{\mathrm{PMS}}\right| < \frac{c_8}{2}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}\right) \leq O\left[\exp\left(\frac{-Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right) + \exp\left\{1-\frac{1}{2}q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}\right].$$

Choosing $\frac{c^*}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p} < \alpha_n < \frac{c_8}{2}\frac{n^{1-(\tau_3+\tau_4)-\gamma+\nu}}{p}$, we can get the conclusion that

$$P\left(\min_{i\in\mathcal{M}_0}\left|\hat{\beta}_i^{\mathrm{PMS}}\right| > \alpha_n > \max_{i\notin\mathcal{M}_0}\left|\hat{\beta}_i^{\mathrm{PMS}}\right|\right) \geq 1 - O\left[\exp\left(\frac{-Cn^{1-4\tau_3-4\tau_4-4\gamma}}{2\log n}\right) + \exp\left\{1-\frac{1}{2}q\left(\frac{\sqrt{C^*}n^{1/2-3\tau_3/2-2\tau_4-\gamma+\nu/2}}{\sqrt{\log n}}\right)\right\}\right].$$

Let $\xi_1 = 4\tau_3 + 4\tau_4 + 4\gamma$,$\xi_2 = 3\tau_3/2 + 2\tau_4 + \gamma - \nu/2$ and $\tilde{C} = \sqrt{C^*}$, above conclusion can be rewritten as

$$P\left(\min_{i\in\mathcal{M}_0}\left|\hat{\beta}_i^{\mathrm{PMS}}\right| > \alpha_n > \max_{i\notin\mathcal{M}_0}\left|\hat{\beta}_i^{\mathrm{PMS}}\right|\right) \geq 1 - O\left[\exp\left(\frac{-Cn^{1-\xi_1}}{2\log n}\right) + \exp\left\{1 - \frac{1}{2}q\left(\frac{\tilde{C}n^{1/2-\xi_2}}{\sqrt{\log n}}\right)\right\}\right].$$

where $\max\{\xi_1, 2\xi_2\} < 1$.

This completes the proof of sure screening. □

Proof of Lemma 1

Proof: First of all, from the property of minimize eigenvalue, we know that

$$\begin{aligned}\lambda_{\min}(\mathbf{G}_2) &= \lambda_{\min}\{\mathbf{X}\mathbf{K}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n)^{-1}\mathbf{K}\mathbf{X}^{\mathrm{T}}\} \\ &\geq \lambda_{\min}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\min}(\mathbf{K}^{\mathrm{T}}\mathbf{K})\{\lambda_{\max}(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{\mathrm{T}}) + \theta\}^{-1} \\ &\geq \frac{\lambda_{\min}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\min}(\mathbf{K}^{\mathrm{T}}\mathbf{K})}{\lambda_{\max}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\max}(\boldsymbol{\Lambda}) + \theta} \geq c\frac{\lambda_{\min}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\min}(\mathbf{K}^{\mathrm{T}}\mathbf{K})}{\lambda_{\max}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\max}(\boldsymbol{\Lambda})}.\end{aligned}$$

Similarly, form the property of maximize eigenvalue, we also obtain that

$$\begin{aligned}\lambda_{\max}(\mathbf{G}_1) &= \lambda_{\max}\{\mathbf{X}\mathbf{K}^{\mathrm{T}}(\mathbf{X}\mathbf{K}^{\mathrm{T}}\boldsymbol{\Lambda}_K\mathbf{K}\mathbf{X}^{\mathrm{T}} + \theta\mathbf{I}_n)^{-1}\mathbf{K}\mathbf{X}^{\mathrm{T}}\} \\ &\leq \lambda_{\max}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\max}(\mathbf{K}^{\mathrm{T}}\mathbf{K})\{\lambda_{\min}(\mathbf{X}\mathbf{K}^{\mathrm{T}}\boldsymbol{\Lambda}_K\mathbf{K}\mathbf{X}^{\mathrm{T}}) + \theta\} \leq \frac{\lambda_{\max}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\max}(\mathbf{K}^{\mathrm{T}}\mathbf{K})}{\lambda_{\min}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\min}(\mathbf{K}^{\mathrm{T}}\mathbf{K})\lambda_{\min}(\boldsymbol{\Lambda}_K)}.\end{aligned}$$

It is obvious that

$$\begin{aligned}\frac{(\mathbf{G}_1)_{jj}}{(\mathbf{G}_2)_{jj}} &\leq \frac{\lambda_{\max}(\mathbf{G}_1)}{\lambda_{\min}(\mathbf{G}_2)} \leq c^{-1}\frac{\lambda_{\max}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\max}(\mathbf{K}^{\mathrm{T}}\mathbf{K})}{\lambda_{\min}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\min}(\mathbf{K}^{\mathrm{T}}\mathbf{K})\lambda_{\min}(\boldsymbol{\Lambda}_K)}\frac{\lambda_{\max}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\max}(\boldsymbol{\Lambda})}{\lambda_{\min}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\min}(\mathbf{K}^{\mathrm{T}}\mathbf{K})} \\ &= c^{-1}\{\mathrm{cond}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\}^2\mathrm{cond}(\mathbf{K}^{\mathrm{T}}\mathbf{K})\{\lambda_{\min}(\mathbf{K}^{\mathrm{T}}\mathbf{K})\}^{-1}\{\lambda_{\min}(\boldsymbol{\Lambda}_K)\}^{-1}\lambda_{\max}(\boldsymbol{\Lambda}) \\ &\leq c^* n^{\tau_1+2\tau_5+\tau_k+\tau_{k_1}+2\tau_{k_2}} \leq C^* n^g.\end{aligned}$$

This completes the proof of Lemma 1. □

Proof of Theorem 2

Proof: Assume that $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$, where $\mathcal{S}_1 \subset \mathcal{M}_0$, $\mathcal{S}_2 \cap \mathcal{M}_0 = \varnothing$. In addition, assume that $\mathcal{S}_3 = \mathcal{S}^c \cap \mathcal{M}_0$, then we have

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{X}_{\mathcal{S}_1}\boldsymbol{\beta}_{\mathcal{S}_1} + \mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3} + \boldsymbol{\epsilon} = \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}} + \mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3} + \boldsymbol{\epsilon},$$

then from the expression of $\tilde{\boldsymbol{\mu}}_{\mathcal{S}}$, when $q \leq n$,

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_{\mathcal{S}} =& (\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{Y} \\ =& (\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}} + (\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3} + (\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\epsilon}.\end{aligned}$$

To any $j \in \mathcal{S}$, we have

$$\begin{aligned}\hat{\beta}_{\mathcal{S},j}^{\mathrm{PMS}} =& \mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}} + \mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3} + \mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\epsilon} \\ =& \beta_j + \mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3} + \mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\epsilon}.\end{aligned}$$

Thus we have

$$|\hat{\beta}_{\mathcal{S},j}^{\mathrm{PMS}} - \beta_j| \leq |\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}| + |\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\epsilon}|.$$

On the one hand, to $|\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}|$, we have

$$\begin{aligned}&|\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}|^2 \\ =& \mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}^{\mathrm{T}}\mathbf{X}_{\mathcal{S}_3}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{e}_j \\ \leq& \|\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}\|^2\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}^2\mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{e}_j \\ \leq& \|\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}\|^2\lambda_{\max}(\boldsymbol{\Omega}_{\mathcal{S}})\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{e}_j \\ \leq& \|\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}\|^2\lambda_{\max}(\boldsymbol{\Omega}_{\mathcal{S}})\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{1/2}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-2}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{1/2}\mathbf{e}_j \\ \leq& \|\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}\|^2\lambda_{\max}(\boldsymbol{\Omega}_{\mathcal{S}})\lambda_{\max}\{(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-2}\}\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}}\mathbf{e}_j \\ \leq& \|\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}\|^2\lambda_{\max}(\boldsymbol{\Omega}_{\mathcal{S}})\{\lambda_{\min}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})\}^{-2}\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}}\mathbf{e}_j.\end{aligned}$$

Let $\mathbf{C}_\mathcal{S}$ be a $q \times p$ matrix such that $\mathbf{X}_\mathcal{S} = \mathbf{X}\mathbf{C}_\mathcal{S}^\mathrm{T}$, it is obvious that only one element in each row of $\mathbf{C}_\mathcal{S}$ is equal to 1 and 0 else, and we could demonstrate that $\mathbf{C}_\mathcal{S}\mathbf{C}_\mathcal{S}^\mathrm{T} = \mathbf{I}_q$. So,

$$\begin{aligned}\lambda_{\max}(\boldsymbol{\Omega}_\mathcal{S})\{\lambda_{\min}(\mathbf{X}_\mathcal{S}^\mathrm{T}\boldsymbol{\Omega}_\mathcal{S}\mathbf{X}_\mathcal{S})\}^{-2} \leq& \lambda_{\max}(\boldsymbol{\Omega}_\mathcal{S})\{\lambda_{\min}(\mathbf{X}_\mathcal{S}^\mathrm{T}\mathbf{X}_\mathcal{S})\lambda_{\min}(\boldsymbol{\Omega}_\mathcal{S})\}^{-2}\\ \leq& \mathrm{cond}(\boldsymbol{\Omega}_\mathcal{S})\{\lambda_{\min}(\mathbf{X}_\mathcal{S}^\mathrm{T}\mathbf{X}_\mathcal{S})\}^{-2}\{\lambda_{\min}(\boldsymbol{\Omega}_\mathcal{S})\}^{-1}.\end{aligned}$$

As

$$\{\lambda_{\min}(\boldsymbol{\Omega}_\mathcal{S})\}^{-1} = \{\lambda_{\min}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T} + \theta\mathbf{I}_n)^{-1}\}^{-1} = \lambda_{\max}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T} + \theta\mathbf{I}_n) = \lambda_{\max}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T}) + \theta,$$

and

$$\begin{aligned}\mathrm{cond}(\boldsymbol{\Omega}_\mathcal{S}) =& \frac{\lambda_{\max}\{(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S} + \theta\mathbf{I}_n)^{-1}\}}{\lambda_{\min}\{(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T} + \theta\mathbf{I}_n)^{-1}\}} = \frac{\lambda_{\max}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T} + \theta\mathbf{I}_n)}{\lambda_{\min}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T} + \theta\mathbf{I}_n)} = \frac{\lambda_{\max}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T}) + \theta}{\lambda_{\min}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T}) + \theta}\\ \leq& \frac{\lambda_{\max}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T})}{\lambda_{\min}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T})} \leq \frac{\lambda_{\max}(\mathbf{X}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T})\lambda_{\max}(\boldsymbol{\Lambda}_\mathcal{S})}{\lambda_{\min}(\mathbf{X}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T})\lambda_{\min}(\boldsymbol{\Lambda}_\mathcal{S})} \leq \frac{\lambda_{\max}(\mathbf{C}_\mathcal{S}^\mathrm{T}\mathbf{C}_\mathcal{S})\lambda_{\max}(\mathbf{Z}\mathbf{Z}^\mathrm{T})\lambda_{\max}(\boldsymbol{\Sigma})\lambda_{\max}(\boldsymbol{\Lambda}_\mathcal{S})}{\lambda_{\min}(\mathbf{C}_\mathcal{S}^\mathrm{T}\mathbf{C}_\mathcal{S})\lambda_{\min}(\mathbf{Z}\mathbf{Z}^\mathrm{T})\lambda_{\min}(\boldsymbol{\Sigma})\lambda_{\min}(\boldsymbol{\Lambda}_\mathcal{S})}\\ =& \mathrm{cond}(\mathbf{Z}\mathbf{Z}^\mathrm{T})\mathrm{cond}(\boldsymbol{\Sigma})\mathrm{cond}(\boldsymbol{\Lambda}_\mathcal{S}) \leq c_1^2 c_6 c_{s_1} c_{s_2} n^{\tau_5+\tau_{s_1}+\tau_{s_2}},\end{aligned}$$

so

$$\begin{aligned}\lambda_{\max}(\boldsymbol{\Omega}_\mathcal{S})\{\lambda_{\min}(\mathbf{X}_\mathcal{S}^\mathrm{T}\boldsymbol{\Omega}_\mathcal{S}\mathbf{X}_\mathcal{S})\}^{-2} \leq& \mathrm{cond}(\boldsymbol{\Omega}_\mathcal{S})\{\lambda_{\min}(\mathbf{C}_\mathcal{S}\mathbf{C}_\mathcal{S}^\mathrm{T})\lambda_{\min}(\mathbf{X}^\mathrm{T}\mathbf{X})\}^{-2}\{\lambda_{\max}(\mathbf{X}_\mathcal{S}\boldsymbol{\Lambda}_\mathcal{S}\mathbf{X}_\mathcal{S}^\mathrm{T}) + \theta\}\\ \leq& \mathrm{cond}(\boldsymbol{\Omega}_\mathcal{S})\{\lambda_{\min}(\mathbf{X}^\mathrm{T}\mathbf{X})\}^{-2}\lambda_{\max}(\mathbf{X}\mathbf{X}^\mathrm{T})\lambda_{\max}(\boldsymbol{\Lambda}_\mathcal{S}) + \theta\mathrm{cond}(\boldsymbol{\Omega}_\mathcal{S})\{\lambda_{\min}(\mathbf{X}^\mathrm{T}\mathbf{X})\}^{-2}\\ =& \mathrm{cond}(\boldsymbol{\Omega}_\mathcal{S})\mathrm{cond}(\mathbf{X}\mathbf{X}^\mathrm{T})\{\lambda_{\min}(\mathbf{X}^\mathrm{T}\mathbf{X})\}^{-1}\lambda_{\max}(\boldsymbol{\Lambda}_\mathcal{S}) + \theta\mathrm{cond}(\boldsymbol{\Omega}_\mathcal{S})\{\lambda_{\min}(\mathbf{X}^\mathrm{T}\mathbf{X})\}^{-2}\\ \leq& c_1^5 c_6^3 c_{s_1}^2 c_{s_2}\frac{n^{3\tau_5+2\tau_{s_1}+\tau_{s_2}}}{p} + \theta c_1^4 c_6^3 c_{s_1} c_{s_2}\frac{n^{3\tau_5+\tau_{s_1}+\tau_{s_2}}}{p^2} \leq 2c_1^5 c_6^3 c_{s_1}^2 c_{s_2}\frac{n^{3\tau_5+2\tau_{s_1}+\tau_{s_2}}}{p}.\end{aligned}$$

To $\mathbf{e}_j^\mathrm{T}\mathbf{X}_\mathcal{S}^\mathrm{T}\boldsymbol{\Omega}\mathbf{X}_\mathcal{S}\mathbf{e}_j$, we have

$$\begin{aligned}\mathbf{e}_j^\mathrm{T}\mathbf{X}_\mathcal{S}^\mathrm{T}\boldsymbol{\Omega}\mathbf{X}_\mathcal{S}\mathbf{e}_j =& \mathbf{e}_j^\mathrm{T}\mathbf{C}_\mathcal{S}\mathbf{X}^\mathrm{T}(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^\mathrm{T} + \theta\mathbf{I}_n)^{-1}\mathbf{X}\mathbf{C}_\mathcal{S}^\mathrm{T}\mathbf{e}_j\\ =& \mathbf{e}_j^\mathrm{T}\mathbf{C}_\mathcal{S}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}^\mathrm{T}(\mathbf{Z}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}^\mathrm{T} + \theta\mathbf{I}_n)^{-1}\mathbf{Z}\boldsymbol{\Sigma}^{1/2}\mathbf{C}_\mathcal{S}^\mathrm{T}\mathbf{e}_j\\ =& \mathbf{e}_j^\mathrm{T}\mathbf{C}_\mathcal{S}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^\mathrm{T}(\mathbf{V}\mathbf{D}\mathbf{U}^\mathrm{T}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^\mathrm{T} + \theta\mathbf{I}_n)^{-1}\mathbf{V}\mathbf{D}\mathbf{U}^\mathrm{T}\boldsymbol{\Sigma}^{1/2}\mathbf{C}_\mathcal{S}^\mathrm{T}\mathbf{e}_j\end{aligned}$$

$$
\begin{aligned}
=&\mathbf{e}_j^{\mathrm{T}}\mathbf{C}_{\mathcal{S}}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}\{\mathbf{V}\mathbf{D}(\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}+\theta\mathbf{D}^{-2})\mathbf{D}\mathbf{V}^{\mathrm{T}}\}^{-1}\mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\mathbf{C}_{\mathcal{S}}^{\mathrm{T}}\mathbf{e}_j\\
=&\mathbf{e}_j^{\mathrm{T}}\mathbf{C}_{\mathcal{S}}\boldsymbol{\Sigma}^{1/2}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}+\theta\mathbf{D}^{-2})^{-1}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\mathbf{C}_{\mathcal{S}}^{\mathrm{T}}\mathbf{e}_j\\
=&\mathbf{e}_j^{\mathrm{T}}\mathbf{C}_{\mathcal{S}}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}+\theta\mathbf{D}^{-2})^{-1}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{C}_{\mathcal{S}}^{\mathrm{T}}\mathbf{e}_j\\
=&\mathbf{e}_j^{\mathrm{T}}\mathbf{C}_{\mathcal{S}}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}(\mathbf{I}_n+\theta\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^{-1}\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{C}_{\mathcal{S}}^{\mathrm{T}}\mathbf{e}_j,
\end{aligned}
$$

similar with the proof in Theorem 1, by Taylor expansion, we have

$$
\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X}_{\mathcal{S}}\mathbf{e}_j=\mathbf{e}_j^{\mathrm{T}}\mathbf{C}_s\boldsymbol{\Lambda}^{-1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\mathbf{C}_{\mathcal{S}}^{\mathrm{T}}\mathbf{e}_j+\mathbf{e}_j^{\mathrm{T}}\mathbf{M}_1\mathbf{e}_j,
$$

where

$$
\mathbf{H}=\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\{\mathbf{U}^{\mathrm{T}}(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2})\mathbf{U}\}^{-1/2},
$$

and

$$
\mathbf{M}_1=\sum_{k=1}^{\infty}(-\theta)^k\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}.
$$

On the one hand,

$$
\mathbf{e}_j^{\mathrm{T}}\mathbf{C}_{\mathcal{S}}\boldsymbol{\Lambda}^{-1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\mathbf{c}_{\mathcal{S}}^{\mathrm{T}}\mathbf{e}_j=(\boldsymbol{\lambda}_j^c)^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\lambda}_j^c\le\sum_{u=1}^{p}\sum_{v\neq u}|\lambda_{ju}^c\lambda_{jv}^c\mathbf{e}_u^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_v|+\sum_{u=1}^{p}(\lambda_{ju}^c)^2\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_u.
$$

From conclusion in Wang and Leng (2016), we know that

$$
\Pr\left(|\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_v|\le\bar{c}_1\frac{n^{1+(\tau_3+\tau_4)-\alpha}}{p}\right)\ge1-O\left\{\exp\left(-\frac{Cn^{1-2\alpha}}{2\log n}\right)\right\},\quad \Pr\left(\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_u>\tilde{c}_1\frac{n^{1+(\tau_3+\tau_4)}}{p}\right)<2\exp(-Cn),
$$

choosing $\alpha=3\tau_3+3\tau_4+3\tau_5+2\tau_{s_1}+\tau_{s_2}+2\bar{\gamma}_1+\epsilon_1-1$, we have

$$
\begin{aligned}
&\mathbf{e}_j^{\mathrm{T}}\mathbf{C}_{\mathcal{S}}\boldsymbol{\Lambda}^{-1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\mathbf{c}_{\mathcal{S}}^{\mathrm{T}}\mathbf{e}_j\\
\le&\frac{M}{\sqrt{\log n}}\frac{n^{2-2\tau_3-2\tau_4-3\tau_5-2\tau_{s_1}-\tau_{s_2}-2\bar{\gamma}_1-\epsilon_1}}{p}\sum_{u=1}^{p}\sum_{v\neq u}|\lambda_{ju}^c\lambda_{jv}^c|+\tilde{c}_1\frac{n^{1+(\tau_3+\tau_4)}}{p}\sum_{u=1}^{p}(\lambda_{ju}^c)^2\\
\le&M^*\tilde{c}_1\frac{n^{2-2\tau_3-2\tau_4-3\tau_5-2\tau_{s_1}-\tau_{s_2}-2\bar{\gamma}_1}}{p}+\tilde{c}_2\frac{n^{1+(\tau_3+\tau_4)}}{p}n^{1-3\tau_3-3\tau_4-3\tau_5-2\tau_{s_1}-\tau_{s_2}-2\bar{\gamma}_1}\le M_1^*\frac{n^{2-2\tau_3-2\tau_4-3\tau_5-2\tau_{s_1}-\tau_{s_2}-2\bar{\gamma}_1}}{p},
\end{aligned}
$$

where $M_1^* = M^*\tilde{c}_1 + \tilde{c}_2$ with probability greater than

$$1 - O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-6\tau_5-4\tau_{s_1}-2\tau_{s_2}-4\bar{\gamma}_1-2\epsilon_1}}{2\log n}\right) + \exp(-Cn)\right\}.$$

On the other hand, as

$$\begin{aligned}
&\lambda_{\max}(\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2}\mathbf{\Sigma}\mathbf{U}\mathbf{A}^{-1}(-\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{\Sigma}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{-1/2}) \\
\le&\lambda_{\max}\{(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\}\lambda_{\max}(\mathbf{\Lambda}^{-1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{\Lambda}^{-1/2}) \le \lambda_{\max}\{(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\}\lambda_{\max}(\mathbf{\Lambda}^{-1}).
\end{aligned}$$

thus

$$\lambda_{\max}(\mathbf{M}_1) \le \lambda_{\max}(\mathbf{\Lambda}^{-1})\sum_{k=1}^{\infty}\theta^k\{\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})\}^k \le \frac{\theta\lambda_{\max}(\mathbf{\Lambda}^{-1})\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})}{1-\theta\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})}.$$

W can obtain that

$$\begin{aligned}
\Pr\left(\lambda_{\max}(\mathbf{M}_1) > \frac{\theta c_1c_3c_5n^{\tau_2+\tau_4}}{p-\theta c_1c_5n^{\tau_4}}\right) \le&\Pr\left(\frac{\theta\lambda_{\max}(\mathbf{\Lambda}^{-1})\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})}{1-\theta\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})} > \frac{\theta c_1c_3c_5n^{\tau_2+\tau_4}}{p-\theta c_1c_5n^{\tau_4}}\right) \\
\le&\Pr\left(\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}) > \frac{c_1c_5n^{\tau_4}}{p}\right) \le \exp(-C_1n),
\end{aligned}$$

which indicates that

$$\mathbf{e}_j^{\mathrm{T}}\mathbf{M}_1\mathbf{e}_j \le \lambda_{\max}(\mathbf{M}_1) \le \frac{\theta c_1c_3c_5n^{\tau_2+\tau_4}}{p-\theta c_1c_5n^{\tau_4}},$$

with probability greater than $1-\exp(-C_1n)$. Choosing θ, such that

$$\frac{\theta c_1c_3c_5n^{\tau_2+2\tau_3+3\tau_4+3\tau_5+2\tau_{s_1}+\tau_{s_2}+2\bar{\gamma}_1-2}}{1-\theta c_1c_4n^{\tau_4}/p} = o(1),$$

we can deduce that

$$\Pr\left(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\mathbf{\Omega}\mathbf{X}_{\mathcal{S}}\mathbf{e}_j \le (M_1^*+o(1))\frac{n^{2-2\tau_3-2\tau_4-3\tau_5-2\tau_{s_1}-\tau_{s_2}-2\bar{\gamma}_1}}{p}\right) \ge 1-O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-6\tau_5-4\tau_{s_1}-2\tau_{s_2}-4\bar{\gamma}_1-2\epsilon_1}}{2\log n}\right)\right\}.$$

In addition, from condition B3, we know that

$$\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\mathbf{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}}\mathbf{e}_j \le c_g(M_1^*+o(1))\frac{n^{2-2\tau_3-2\tau_4-3\tau_5-2\tau_{s_1}-\tau_{s_2}-2\bar{\gamma}_1+g}}{p},$$

with probability greater than $1 - O\left\{\exp\left(-\dfrac{Cn^{3-6\tau_3-6\tau_4-6\tau_5-4\tau_{s_1}-2\tau_{s_2}-4\bar{\gamma}_1-2\epsilon_1}}{2\log n}\right)\right\}$.

To sum up, we can prove that

$$\Pr\left(|\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}_3}\boldsymbol{\beta}_{\mathcal{S}_3}|^2 > \frac{K_1^* n^{2-2\tau_3-2\tau_4+g}}{p^2}\right) \geq 1 - O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-6\tau_5-4\tau_{s_1}-2\tau_{s_2}-4\bar{\gamma}_1-2\epsilon_1}}{2\log n}\right)\right\},$$

where $K_1^* = 2c_1^5 c_6^3 c_{s_1} c_{s_2} c_g \bar{c}_1^2\{M_1^* + o(1)\}$.

To the noise part, we have

$$\begin{aligned}
\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\|^2 =& \mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}^2\mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{e}_j \\
\leq& \lambda_{\max}(\boldsymbol{\Omega}_{\mathcal{S}})\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{1/2}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-2}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{1/2}\mathbf{e}_j \\
\leq& \lambda_{\max}(\boldsymbol{\Omega}_{\mathcal{S}})\lambda_{\max}\{(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-2}\}\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{e}_j.
\end{aligned}$$

From previous proof, we know that

$$\begin{aligned}
&\Pr\left(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}}\mathbf{e}_j \leq M_2^* \frac{n^{2-2\tau_3-2\tau_4-3\tau_5-2\tau_{s_1}-\tau_{s_2}-2\bar{\gamma}_1+g}}{p}\right) \\
\geq& 1 - O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-6\tau_5-4\tau_{s_1}-2\tau_{s_2}-4\bar{\gamma}_1-2\epsilon_1}}{2\log n}\right)\right\},
\end{aligned}$$

where $M_2^* = c_g\{M_1 + o(1)\}$, as well as

$$\Pr\left(\lambda_{\max}(\boldsymbol{\Omega}_{\mathcal{S}})\{\lambda_{\min}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})\}^{-2} \leq 2c_1^5 c_6^3 c_{s_1} c_{s_2} \frac{n^{3\tau_5+2\tau_{s_1}+\tau_{s_2}}}{p}\right) \geq 1 - \exp(-C_1 n).$$

So we have

$$\begin{aligned}
&\Pr\left(\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\| \leq K_2^* \frac{n^{1-\tau_3-\tau_4-\bar{\gamma}_1+g/2}}{p}\right) \\
\geq& 1 - O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-6\tau_5-4\tau_{s_1}-2\tau_{s_2}-4\bar{\gamma}_1-2\epsilon_1}}{2\log n}\right)\right\},
\end{aligned}$$

where $K_2^* = \sqrt{2c_1^5 c_6^3 c_{s_1} c_{s_2} M_2^*}$. Choosing $t = \sqrt{C_1^*}\dfrac{n^{\bar{\gamma}_1}}{\sqrt{\log n}}$, then from inequality

$$\Pr(|\omega| > t) < \exp\left\{1 - q(\frac{\sqrt{C_1^* n^{\bar{\gamma}_1}}}{\sqrt{\log n}})\right\},$$

we have

$$\Pr\left(|\mathbf{e}_j^{\mathrm{T}}(\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\mathbf{X}_{\mathcal{S}})^{-1}\mathbf{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\epsilon}| \leq \frac{\sigma K_3^*}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)+g/2}}{p}\right) \geq 1 - O\left[\exp\left\{1 - q(\frac{\sqrt{C_1^* n^{\bar{\gamma}_1}}}{\sqrt{\log n}})\right\}\right],$$

where $K_3^* = K_2^*\sqrt{C_1^*}$. To sum up, we have

$$\begin{aligned}\Pr\left(|\hat{\beta}_{\mathcal{S},j}^{\mathrm{PMS}} - \beta_j| \leq K^*\frac{n^{1-(\tau_3+\tau_4)+g/2}}{p}\right) \geq & 1 - O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-6\tau_5-4\tau_{s_1}-2\tau_{s_2}-4\bar{\gamma}_1-2\epsilon_1}}{2\log n}\right)\right\} \\ & - O\left[\exp\left\{1 - q(\frac{\sqrt{C_1^* n^{\bar{\gamma}_1}}}{\sqrt{\log n}})\right\}\right],\end{aligned}$$

where $K^* = \sqrt{K_1^*} + \frac{\sigma\sqrt{K_2}}{\sqrt{\log n}}$. □

Proof of Theorem 3

Proof: Denoting $\mathbf{B}^*$ as the true $m \times p$ group indicator matrix, we have

$$\mathbf{Y} = \mathbf{X}(\mathbf{B}^*)^{\mathrm{T}}\bar{\boldsymbol{\beta}} + \boldsymbol{\epsilon} = \mathbf{X}\mathbf{B}^{\mathrm{T}}\bar{\boldsymbol{\beta}} - \mathbf{X}\left\{\mathbf{B}^{\mathrm{T}} - (\mathbf{B}^*)^{\mathrm{T}}\right\}\bar{\boldsymbol{\beta}} + \boldsymbol{\epsilon},$$

where $\bar{\boldsymbol{\beta}} = (\beta_1, \ldots, \beta_m)^{\mathrm{T}}$ is the true parameters in each group. So

$$\begin{aligned}\bar{\boldsymbol{\nu}} &= \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{Y} = \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\left[\mathbf{X}\mathbf{B}^{\mathrm{T}}\bar{\boldsymbol{\beta}} - \mathbf{X}\left\{\mathbf{B}^{\mathrm{T}} - (\mathbf{B}^*)^{\mathrm{T}}\right\}\bar{\boldsymbol{\beta}} + \boldsymbol{\epsilon}\right] \\ &= \bar{\boldsymbol{\beta}} - \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\left\{\mathbf{B}^{\mathrm{T}} - (\mathbf{B}^*)^{\mathrm{T}}\right\}\bar{\boldsymbol{\beta}} + \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\boldsymbol{\epsilon},\end{aligned}$$

which indicates that

$$\left|\bar{\nu}_j - \bar{\beta}_j\right| \leq \left|\mathbf{e}_j^{\mathrm{T}}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\left\{\mathbf{B}^{\mathrm{T}} - (\mathbf{B}^*)^{\mathrm{T}}\right\}\bar{\boldsymbol{\beta}}\right| + \left|\mathbf{e}_j^{\mathrm{T}}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\boldsymbol{\epsilon}\right|. \tag{S3.5}$$

To the first part in inequality (S3.5), we have

$$\begin{aligned}&\left|\mathbf{e}_j^{\mathrm{T}}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\left\{\mathbf{B}^{\mathrm{T}} - (\mathbf{B}^*)^{\mathrm{T}}\right\}\bar{\boldsymbol{\beta}}\right|^2 \\ =&\mathbf{e}_j^{\mathrm{T}}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\left\{\mathbf{B}^{\mathrm{T}} - (\mathbf{B}^*)^{\mathrm{T}}\right\}\bar{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}^{\mathrm{T}}\left(\mathbf{B} - \mathbf{B}^*\right)\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1}\mathbf{e}_j\end{aligned}$$

$$
\begin{aligned}
&\leq \| \left(\mathbf{B}-\mathbf{B}^*\right)\bar{\boldsymbol{\beta}}\|^2 \mathbf{e}_j^{\mathrm{T}} \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1} \mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}} \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-1} \mathbf{e}_j \\
&\leq \| \left(\mathbf{B}-\mathbf{B}^*\right)\bar{\boldsymbol{\beta}}\|^2 \lambda_{\max}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right) \mathbf{e}_j^{\mathrm{T}} \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{1/2} \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-3} \\
&\quad \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{1/2} \mathbf{e}_j \\
&\leq \| \left(\mathbf{B}-\mathbf{B}^*\right)\bar{\boldsymbol{\beta}}\|^2 \lambda_{\max}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right) \lambda_{\max}\left\{\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-3}\right\} \mathbf{e}_j^{\mathrm{T}} \left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right) \mathbf{e}_j .
\end{aligned}
$$

It is obvious that

$$
\begin{aligned}
&\lambda_{\max}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{B}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right) \lambda_{\max}\left\{\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-3}\right\} \\
\leq & \lambda_{\max}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right) \left\{\lambda_{\max}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)\right\}^2 \left\{\lambda_{\max}\left(\boldsymbol{\Omega}_{\mathbf{B}}\right)\right\}^2 \left\{\lambda_{\min}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right) \lambda_{\min}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right) \lambda_{\min}\left(\boldsymbol{\Omega}_{\mathbf{B}}\right)\right\}^{-3} \\
= & \mathrm{cond}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right) \left\{\mathrm{cond}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)\right\}^2 \left\{\mathrm{cond}\left(\boldsymbol{\Omega}_{\mathbf{B}}\right)\right\}^2 \left\{\lambda_{\min}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)\right\}^{-2} \left\{\lambda_{\min}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)\right\}^{-1} \left\{\lambda_{\min}\left(\boldsymbol{\Omega}_{\mathbf{B}}\right)\right\}^{-1} .
\end{aligned}
$$

As

$$
\begin{aligned}
\left\{\lambda_{\min}\left(\boldsymbol{\Omega}_{\mathbf{B}}\right)\right\}^{-1} &= \left\{\lambda_{\min}\left(\mathbf{X}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Lambda}_{\mathbf{B}}\mathbf{B}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right)^{-1}\right\}^{-1} = \lambda_{\max}\left(\mathbf{X}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Lambda}_{\mathbf{B}}\mathbf{B}\mathbf{X}^{\mathrm{T}}+\theta\mathbf{I}_n\right) \\
&= \lambda_{\max}\left(\mathbf{X}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Lambda}_{\mathbf{B}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\right)+\theta \leq \lambda_{\max}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right) \lambda_{\max}(\mathbf{B}^{\mathrm{T}}\mathbf{B}) \lambda_{\max}\left(\boldsymbol{\Lambda}_{B}\right)+\theta,
\end{aligned}
$$

similarly, we have

$$
\mathrm{cond}(\boldsymbol{\Omega}_{\mathbf{B}}) \leq \mathrm{cond}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})\mathrm{cond}(\boldsymbol{\Sigma})\mathrm{cond}(\mathbf{B}^{\mathrm{T}}\mathbf{B})\mathrm{cond}(\boldsymbol{\Lambda}_{\mathbf{B}}).
$$

So we get

$$
\begin{aligned}
&\lambda_{\max}\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right) \lambda_{\max}\left\{\left(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\right)^{-3}\right\} \\
\leq & \left\{\mathrm{cond}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)\right\}^2 \left\{\mathrm{cond}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)\right\}^3 \left\{\mathrm{cond}\left(\boldsymbol{\Omega}_{\mathbf{B}}\right)\right\}^2 \left\{\lambda_{\min}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)\right\}^{-1} \lambda_{\max}\left(\boldsymbol{\Lambda}_{\mathbf{B}}\right) \\
& + \theta \mathrm{cond}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right) \left\{\mathrm{cond}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)\right\}^2 \left\{\mathrm{cond}\left(\boldsymbol{\Omega}_{B}\right)\right\}^2 \left\{\lambda_{\min}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)\right\}^{-2} \left\{\lambda_{\min}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)\right\}^{-1} \\
\leq & \left\{\mathrm{cond}\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)\right\}^4 \left\{\mathrm{cond}\left(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right)\right\}^5 \left\{\mathrm{cond}(\boldsymbol{\Sigma})\right\}^5 \left\{\mathrm{cond}(\boldsymbol{\Lambda}_{B})\right\}^2 \left\{\lambda_{\min}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\right\}^{-1} \lambda_{\max}(\boldsymbol{\Lambda}_{B}) \\
& + \theta\{\mathrm{cond}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\}^3\{\mathrm{cond}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})\}^4\{\mathrm{cond}(\boldsymbol{\Sigma})\}^4\{\mathrm{cond}(\boldsymbol{\Lambda}_{B})\}^2\{\lambda_{\min}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\}^{-2}\{\lambda_{\min}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\}^{-1}
\end{aligned}
$$

$$\leq c_1^{10} c_6^5 \bar{c}_8^{-4} \bar{c}_9^{-5} c_{b_1}^3 c_{b_2}^2 n^{5\tau_5+4\tau_6+5\tau_7+3\tau_{b_1}+2\tau_{b_2}} + \theta c_1^9 c_6^5 \bar{c}_8^{-3} \bar{c}_9^{-4} c_{b_1}^2 c_{b_2}^2 \frac{n^{5\tau_5+3\tau_6+5\tau_7+2\tau_{b_1}+2\tau_{b_2}}}{p}$$
$$\leq 2c_1^{10} c_6^5 \bar{c}_8^{-4} \bar{c}_9^{-5} c_{b_1}^3 c_{b_2}^2 n^{5\tau_5+4\tau_6+5\tau_7+3\tau_{b_1}+2\tau_{b_2}}.$$

with probability greater than $1 - \exp(-C_1 n)$. In addition,

$$\begin{aligned}
\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j =& \mathbf{e}_j^{\mathrm{T}}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}^{\mathrm{T}} + \theta\mathbf{I}_n)^{-1}\mathbf{Z}\boldsymbol{\Sigma}^{1/2}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j \\
=& \mathbf{e}_j^{\mathrm{T}}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}(\mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}} + \theta\mathbf{I}_n)^{-1}\mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j \\
=& \mathbf{e}_j^{\mathrm{T}}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}\{\mathbf{V}\mathbf{D}(\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U} + \theta\mathbf{D}^{-2})\mathbf{D}\mathbf{V}^{\mathrm{T}}\}^{-1}\mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j \\
=& \mathbf{e}_j^{\mathrm{T}}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}\mathbf{U} + \theta\mathbf{D}^{-2})^{-1}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j \\
=& \mathbf{e}_j^{\mathrm{T}}\mathbf{B}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}\left\{\mathbf{I}_n + \sum_{k=1}^{\infty}(-\theta)^k(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\right\}\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}^{1/2} \\
& \boldsymbol{\Lambda}^{-1/2}\mathbf{B}\mathbf{e}_j \\
=& \mathbf{e}_j^{\mathrm{T}}\mathbf{B}\boldsymbol{\Lambda}^{-1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\mathbf{B}\mathbf{e}_j + \mathbf{e}_j^{\mathrm{T}}\mathbf{M}_2\mathbf{e}_j,
\end{aligned}$$

where $\mathbf{M}_2 = \sum_{k=1}^{\infty}(-\theta)^k\mathbf{B}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{B}^{\mathrm{T}}$.

First of all,

$$\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\boldsymbol{\Lambda}^{-1/2}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\mathbf{B}\mathbf{e}_j = \boldsymbol{\lambda}_j^{\mathbf{B}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\lambda}_j^{\mathbf{B}} \leq \sum_{u=1}^{p}\sum_{v\neq u}|\lambda_{ju}^{\mathbf{B}}\lambda_{jv}^{\mathbf{B}}\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_v| + \sum_{u=1}^{p}(\lambda_{ju}^{\mathbf{B}})^2\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_u.$$

Choosing $\alpha = 3\tau_3 + 3\tau_4 + 5\tau_5 + 4\tau_6 + 5\tau_7 + 3\tau_{b_1} + 2\tau_{b_2} + 2\bar{\gamma}_2 + \epsilon_2 - 1$, from

$$\Pr\left(|\mathbf{e}_u\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_v| > \frac{M}{\sqrt{\log n}}\frac{n^{1+(\tau_3+\tau_4)-\alpha}}{p}\right) \geq 1 - O\left\{\exp\left(-\frac{Cn^{1-2\alpha}}{2\log n}\right)\right\}, \quad \Pr\left(\mathbf{e}_u^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{e}_u > \tilde{c}_1\frac{n^{1+(\tau_3+\tau_4)}}{p}\right) < 2\exp(-Cn),$$

we know that

$$\begin{aligned}
& \mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j \\
\leq & \frac{M}{\sqrt{\log n}}\frac{n^{2-2\tau_3-2\tau_4-5\tau_5-4\tau_6-5\tau_7-3\tau_{b_1}-2\tau_{b_2}-2\bar{\gamma}_2-\epsilon_2}}{p}\sum_{u=1}^{p}\sum_{v\neq u}|\lambda_{ju}^{\mathbf{B}}\lambda_{jv}^{\mathbf{B}}| + \tilde{c}_1\frac{n^{1+(\tau_3+\tau_4)}}{p}\sum_{u=1}^{p}(\lambda_{ju}^{\mathbf{B}})^2 \\
\leq & \tilde{K}_1\frac{n^{2-2\tau_3-2\tau_4-5\tau_5-4\tau_6-5\tau_7-3\tau_{b_1}-\tau_{b_2}-2\bar{\gamma}_2}}{p},
\end{aligned}$$

with probability greater than

$$1 - O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-10\tau_5-8\tau_6-10\tau_7-6\tau_{b_1}-4\tau_{b_2}-4\bar{\gamma}_2-2\epsilon_2}}{2\log n}\right)\right\},$$

where $\tilde{K}_1 = Mc^2_{\bar{\gamma}_2} + \tilde{c}_1 c_{\epsilon_2}$. As

$$\begin{aligned}&\lambda_{\max}\{\mathbf{B}\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2}\mathbf{\Sigma}^{1/2}\mathbf{U}\mathbf{A}^{-1}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\mathbf{A}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{\Sigma}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{B}^{\mathrm{T}}\}\\ \leq&\lambda_{\max}\{(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\}\lambda_{\max}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\max}(\mathbf{\Lambda}^{-1})\lambda_{\max}(\mathbf{H}\mathbf{H}^{\mathrm{T}}),\end{aligned}$$

thus

$$\begin{aligned}\lambda_{\max}(\mathbf{M}_2) \leq&\lambda_{\max}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\max}(\mathbf{\Lambda}^{-1})\sum_{k=1}^{\infty}\theta^k\{\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})^k\}\\ \leq&\frac{\theta\lambda_{\max}(\mathbf{\Lambda}^{-1})\lambda_{\max}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})}{1-\theta\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1})}.\end{aligned}$$

So

$$\Pr\left(\lambda_{\max}(\mathbf{M}_2) > \frac{\theta c_1c_3c_5\bar{c}_8 n^{\tau_2+\tau_4+\tau_6}}{p-\theta c_1c_5n^{\tau_4}}\right) \leq \Pr\left(\lambda_{\max}(\mathbf{A}^{-\mathrm{T}}\mathbf{D}^{-2}\mathbf{A}^{-1}) > \frac{c_1c_5n^{\tau_4}}{p}\right) < \exp(-C_1n),$$

which indicates that

$$\mathbf{e}_j^{\mathrm{T}}\mathbf{M}_2\mathbf{e}_j \leq o(1)\frac{n^{2-2\tau_3-2\tau_4-5\tau_5-4\tau_6-5\tau_7-3\tau_{b_1}-\tau_{b_2}-2\bar{\gamma}_2}}{p^2}$$

with probability greater than $1-\exp(-C_1n)$ if θ satisfying that

$$\frac{\theta c_1c_3c_5n^{\tau_2+2\tau_3+3\tau_4+5\tau_5+5\tau_6+5\tau_7+3\tau_{b_1}+\tau_{b_2}+2\bar{\gamma}_2-2}}{1-\theta c_1c_5n\tau_4/p} = o(1).$$

To sum up, we obtain that

$$\begin{aligned}&\Pr\left(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{\Omega}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j \leq \{\tilde{K}_1+o(1)\}\frac{n^{2-2\tau_3-2\tau_4-5\tau_5-4\tau_6-5\tau_7-3\tau_{b_1}-2\tau_{b_2}-2\bar{\gamma}_2}}{p}\right)\\ &\geq 1-O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-10\tau_5-8\tau_6-10\tau_7-6\tau_{b_1}-4\tau_{b_2}-4\bar{\gamma}_2-2\epsilon_2}}{2\log n}\right)\right\}.\end{aligned}$$

From condition C3, we have

$$\Pr\left(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j \le c_b\{\tilde{K}_1+o(1)\}\frac{n^{2-2\tau_3-2\tau_4-5\tau_5-4\tau_6-5\tau_7-3\tau_{b_1}-2\tau_{b_2}-2\bar{\gamma}_2+b}}{p}\right)$$
$$\ge 1-O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-10\tau_5-8\tau_6-10\tau_7-6\tau_{b_1}-4\tau_{b_2}-4\bar{\gamma}_2-2\epsilon_2}}{2\log n}\right)\right\}.$$

To the first part, we obtain the conclusion that

$$\Pr\left(\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\{\mathbf{B}^{\mathrm{T}}-(\mathbf{B}^*)^{\mathrm{T}}\}\bar{\boldsymbol{\beta}}\| > \sqrt{\tilde{K}_2}\frac{n^{1-(\tau_3+\tau_4)+b/2}}{p}\right)$$
$$\ge 1-O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-10\tau_5-8\tau_6-10\tau_7-6\tau_{b_1}-4\tau_{b_2}-4\bar{\gamma}_2-2\epsilon_2}}{2\log n}\right)\right\},$$

where $\tilde{K}_2 = 2c_1^{10}c_6^5\bar{c}_8^{-4}\bar{c}_9^{-5}c_{b_1}^3c_{b_2}^2c_b\{\tilde{K}_1+o(1)\}$.

To the noise part, we have

$$\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\|^2 = \mathbf{e}_j^{\mathrm{T}}(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}(\boldsymbol{\Omega}_{\mathbf{B}})^2\mathbf{X}\mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{e}_j$$
$$\le\lambda_{\max}(\boldsymbol{\Omega}_{\mathbf{B}})\mathbf{e}_j^{\mathrm{T}}(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{e}_j \le \lambda_{\max}(\boldsymbol{\Omega}_{\mathbf{B}})\lambda_{\max}\{(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}})^{-2}\}\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j$$
$$=\lambda_{\max}(\boldsymbol{\Omega}_{\mathbf{B}})\{\lambda_{\min}(\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}})\}^{-2}\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j$$
$$\le\lambda_{\max}(\boldsymbol{\Omega}_{\mathbf{B}})\{\lambda_{\min}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\min}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\lambda_{\min}(\boldsymbol{\Omega}_{\mathbf{B}})\}^{-2}\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j$$
$$=\mathrm{cond}(\boldsymbol{\Omega}_{\mathbf{B}})\{\lambda_{\min}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\min}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-2}\{\lambda_{\min}(\boldsymbol{\Omega}_{\mathbf{B}})\}^{-1}\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j$$
$$\le\mathrm{cond}(\boldsymbol{\Omega}_{\mathbf{B}})\{\lambda_{\min}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\min}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-2}\lambda_{\max}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\max}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\max}(\boldsymbol{\Lambda}_{\mathbf{B}})\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j+$$
$$\theta\mathrm{cond}(\boldsymbol{\Omega}_{\mathbf{B}})\{\lambda_{\min}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\min}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-2}\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j$$
$$\le2\mathrm{cond}(\boldsymbol{\Omega}_{\mathbf{B}})\{\lambda_{\min}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\min}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-2}\lambda_{\max}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\lambda_{\max}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\max}(\boldsymbol{\Lambda}_{\mathbf{B}})\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j$$
$$=2\mathrm{cond}(\boldsymbol{\Omega}_{\mathbf{B}})\mathrm{cond}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\mathrm{cond}(\mathbf{X}\mathbf{X}^{\mathrm{T}})\{\lambda_{\min}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\lambda_{\min}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-1}\lambda_{\max}(\boldsymbol{\Lambda}_B)\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j$$
$$\le2c_1^5c_6^3\bar{c}_8^{-2}\bar{c}_9^{-3}c_{b_1}^2c_{b_2}\frac{n^{3\tau_5+2\tau_6+3\tau_7+2\tau_{b_1}+\tau_{b_2}}}{p}\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathbf{B}}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j.$$

In addition, as

$$\Pr\left(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{\Omega_B}\mathbf{X}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j \le c_b\{\tilde{K}_1+o(1)\}\frac{n^{2-2\tau_3-2\tau_4-5\tau_5-4\tau_6-5\tau_7-3\tau_{b_1}-2\tau_{b_2}-2\bar{\gamma}_2+b}}{p}\right) \ge 1-O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-10\tau_5-8\tau_6-10\tau_7-6\tau_{b_1}-4\tau_{b_2}-4\bar{\gamma}_2-2\epsilon_2}}{2\log n}\right)\right\}.$$

we obtain

$$\|\mathbf{e}_j^{\mathrm{T}}(\mathbf{B}\mathbf{X}^{\mathrm{T}}\mathbf{\Omega}_B\mathbf{X}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\mathbf{\Omega_B}\| \le \sqrt{\tilde{K}_3}\frac{n^{1-\tau_3-\tau_4-\tau_5-\tau_6-\tau_7-\tau_{b_1}/2-\tau_{b_2}/2-\bar{\gamma}_2+b/2}}{p},$$

with probability greater than $1-O\left\{\exp\left(-\frac{Cn^{3-6\tau_3-6\tau_4-10\tau_5-8\tau_6-10\tau_7-6\tau_{b_1}-4\tau_{b_2}-4\bar{\gamma}_2-2\epsilon_2}}{2\log n}\right)\right\}$, where $\tilde{K}_3 = 2c_1^5c_6^3\bar{c}_8^2\bar{c}_9^3b_{b_1}^2c_{b_2}c_b\{\tilde{K}_1+o(1)\}$.

Choosing $t = \frac{\bar{C}n^{\tau_{b_1}/2+\tau_{b_2}/2+\tau_5+\tau_6+\tau_7+\bar{\gamma}_2}}{\sqrt{\log n}}$, from

$$P(|\omega|>t) < \exp\left\{1-q\left(\frac{\bar{C}n^{\tau_{b_1}/2+\tau_{b_2}/2+\tau_5+\tau_6+\tau_7+\bar{\gamma}_2}}{\sqrt{\log n}}\right)\right\},$$

we have

$$P\left(|\mathbf{e}_j^{\mathrm{T}}(\mathbf{B}\mathbf{X}^{\mathrm{T}}\mathbf{\Omega_B}\mathbf{X}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{B}\mathbf{X}^{\mathrm{T}}\mathbf{\Omega_B}\boldsymbol{\epsilon}| > \frac{\sigma\sqrt{\bar{C}\tilde{K}_3}}{\sqrt{\log n}}\frac{n^{1-(\tau_3+\tau_4)+b/2}}{p}\right)$$

$$\ge 1-O\left[\exp\left(-\frac{\bar{C}n^{3-6\tau_{b_1}-4\tau_{b_2}-6\tau_{b_3}-10\tau_5-8\tau_6-10\tau_7-4\bar{\gamma}_2-2\epsilon_2}}{2\log n}\right)-\exp\left\{1-q\left(\frac{\bar{C}n^{\tau_{b_1}/2+\tau_{b_2}/2+\tau_5+\tau_6+\tau_7+\bar{\gamma}_2}}{\sqrt{\log n}}\right)\right\}\right].$$

That is to say

$$P\left(|\bar{\nu}_j-\bar{\beta}_j| \le \tilde{K}\frac{n^{1-\tau_3-\tau_4+b/2}}{p}\right) \ge 1-O\left\{\exp\left(-\frac{\bar{C}n^{3-6\tau_3-6\tau_4-10\tau_5-8\tau_6-10\tau_7-6\tau_{b_1}-4\tau_{b_2}-4\bar{\gamma}_2-2\epsilon_2}}{2\log n}\right)\right\} \exp\left\{1-q\left(\frac{\bar{C}n^{\tau_{b_1}/2+\tau_{b_2}/2+\tau_5+\tau_6+\tau_7+\bar{\gamma}_2}}{\sqrt{\log n}}\right)\right\},$$

where $\tilde{K} = \sqrt{\tilde{K}_2} + \frac{\sigma\sqrt{\bar{C}\tilde{K}_3}}{\sqrt{\log n}}$. □

The Proof of Theorem 4:

Proof: To finite K and $|\mathcal{M}_0|$, it is obvious that

$$\left\{\min_{j\in\mathcal{M}_0}\max_{k\in\{1,\ldots,K\}}|\hat{\beta}_j^{(k)}| > \max_{j\notin\mathcal{M}_0}\max_{k\in\{1,\ldots,K\}}|\hat{\beta}_j^{(k)}|\right\} = \left\{\max_{k\in\{1,\ldots,K\}}\min_{j\in\mathcal{M}_0}|\hat{\beta}_j^{(k)}| > \max_{k\in\{1,\ldots,K\}}\max_{j\notin\mathcal{M}_0}|\hat{\beta}_j^{(k)}|\right\},$$

for simplicity, we denote

$$A_k = \left\{\min_{j\in\mathcal{M}_0}|\hat{\beta}_j^{(k)}| > \max_{j\notin\mathcal{M}_0}|\hat{\beta}_j^{(k)}|\right\}, k = 1\ldots, K.$$

Then, to combined PMS statistics, we have

$$P\left(\min_{j\in\mathcal{M}_0}|\hat{\beta}_j^{\text{CPMS}}| > \max_{j\notin\mathcal{M}_0}|\hat{\beta}_j^{\text{CPMS}}|\right) \geq P\left(\bigcap_{k=1}^{K}A_k\right) \geq 1 - P\left(\bigcup_{k=1}^{K}\bar{A}_k\right) \geq 1 - \sum_{k=1}^{K}\left\{1 - P\left(A_k\right)\right\}.$$

From the conclusion of Theorem 1, to each $k \in \{1,\ldots,K\}$, we have

$$1 - P(A_k) = O\left\{\exp\left(-C_k\frac{n^{1-\xi_3^{(k)}}}{2\log n}\right) + \exp\left(1 - \frac{1}{2}q\left(\frac{\tilde{C}_k n^{1/2-\xi_4^{(k)}}}{\sqrt{\log n}}\right)\right)\right\},$$

with $0 < \xi_3^{(k)}, 2\xi_4^{(k)} < 1$. Choosing $\tilde{\xi}_3 = \max_{k\in\{1,\ldots,K\}}\{\xi_3^{(k)}\}$ and $\tilde{\xi}_4 = \max_{k\in\{1,\ldots,K\}}\{\xi_4^{(k)}\}$, we can finally prove that

$$\begin{aligned}P\left(\min_{j\in\mathcal{M}_0}|\hat{\beta}_j^{\text{CPMS}}| > \max_{j\notin\mathcal{M}_0}|\hat{\beta}_j^{\text{CPMS}}|\right) &\geq 1 - KO\left\{\exp\left(-C_k\frac{n^{1-\tilde{\xi}_3}}{2\log n}\right) + \exp\left(1 - \frac{1}{2}q\left(\frac{\tilde{C}_k n^{1/2-\tilde{\xi}_4}}{\sqrt{\log n}}\right)\right)\right\}\\ &= 1 - O\left\{\exp\left(-C'\frac{n^{1-\tilde{\xi}_3}}{2\log n}\right) + \exp\left(1 - \frac{1}{2}q\left(\frac{\tilde{C}' n^{1/2-\tilde{\xi}_4}}{\sqrt{\log n}}\right)\right)\right\}.\end{aligned}$$

That completes the proof. □

S4 Additional Simulation Studies

S4.1 Compound Symmetry Case

In this setting, we assigned a standard normal distribution to the marginal distribution of all covariates. And the covariance matrix had a compound symmetry structure with correlation

$\rho = 0.5$. We chose sample size $n = 200$, dimension of features $p = 10,000$, and signal-to-noise ratio $R^2 = 0.5, 0.9$. The coefficient vector was specified as $\boldsymbol{\beta} = (3, 3, 3, 3, 3, -7.5, 0, \ldots, 0)^{\mathrm{T}}$, which indicates the number of active features was 6.

For PMS, we chose $\boldsymbol{\Lambda} = \mathbf{I}_p$, a case without any prior correlation information. To incorporate prior mean information, we considered four specific cases. Case I was a similar case with Kang et al. (2017). Under this setting, we divided all 10,000 features into 357 groups with group labels $g_j = \sum_{g=1}^{357} gI\{28g - 27 \leq j \leq 28g\}, g = 1, \ldots, 357$, as well as $g_j = 357, j = 9997, \ldots, 10000$. Thus, the first 356 groups included 28 features while the 357th group had 32. We introduced that partition structure to PartS, which allocated all 6 active features into the same group. At the same time, we designed $\mathcal{S}_{\mathrm{I}} = \{j : g_j = 1\}$ as the prior selected set to the PMS method with a total of 28 variables. For these cases (designed as II-IV), we chose the prior selected set as $\mathcal{S}_{\mathrm{II}} = \{2, 6\}$, $\mathcal{S}_{\mathrm{III}} = \{6, 8\}$ and $\mathcal{S}_{\mathrm{IV}} = \{2, 8\}$. $\mathcal{S}_{\mathrm{II}}$ consisted of 2 active variables, while $\mathcal{S}_{\mathrm{III}}$ and $\mathcal{S}_{\mathrm{IV}}$ only contained 1 active variable, a prime active one and a subordinate active one respectively. To apply the PartS method, we randomly partitioned all features into 357 groups and assigned two features in set $\mathcal{S}_j, j \in \{\mathrm{II}, \mathrm{III}, \mathrm{IV}\}$ into group 1. We also considered the PMS group method in cases II-IV, with the corresponding groups denoted as $\mathcal{G}_1 = \{s_{j_1}\}$, $\mathcal{G}_2 = \{s_{j_2}\}$ and $\mathcal{G}_3 = \{j : j \notin \mathcal{S}_j\}$ with $\mathcal{S}_j = \{s_{j_1}, s_{j_2}\}$, $j \in \{\mathrm{II}, \mathrm{III}, \mathrm{IV}\}$. In addition to PartS, we also compared the various PMS methods with SIS, HOLP and CIS. The simulation results are summarized in Table 1.

S4.2 Evaluation of Random decoupling

Using the same simulation design with previous section S4.1, we compensated the simulation results of choosing the thresholding parameter by the random decoupling method. The results are summarized in Table 2, which indicate that random decoupling is an effective method to

Table 1: Screening accuracy for predictors with the compound symmetry correlation structure.

	$R^2 = 0.5$			$R^2 = 0.9$		
Method	FPR	FNR	Model Size	FPR	FNR	Model Size
SIS	72 (97)	183 (100)	5037 (1945, 8609)	9 (26)	149 (56)	5073 (2095, 8329)
HOLP	31 (35)	86 (100)	839 (359, 1925)	1 (3)	3 (20)	39 (16, 120)
CIS	23 (22)	0 (0)	657 (267, 710)	0 (1)	0 (0)	9 (7, 19)
PartS-I	43 (38)	91 (114)	912 (380, 1663)	1 (1)	0 (0)	23 (14, 38)
PartS-II	35 (42)	81 (102)	901 (440, 1888)	2 (4)	5 (29)	58 (22, 168)
PartS-III	41 (47)	97 (110)	976 (513, 2510)	3 (7)	8 (35)	72 (24, 228)
PartS-IV	39 (43)	91 (107)	952 (474, 2467)	2 (5)	9 (38)	66 (23, 229)
PMS-selection-I	0 (0)	0 (0)	6 (6, 7)	0 (0)	0 (0)	6 (6, 6)
PMS-selection-II	12 (20)	33 (71)	335 (103, 827)	0 (0)	0 (0)	6 (6, 7)
PMS-selection-III	23 (29)	65 (94)	593 (228, 1392)	0 (1)	2 (17)	11 (8, 23)
PMS-selection-IV	22 (29)	66 (90)	659 (238, 1558)	1 (1)	3 (20)	22 (11, 64)
PMS-group-II	7 (15)	27 (66)	194 (65, 550)	0 (0)	0 (0)	6 (6, 6)
PMS-group-III	12 (20)	38 (76)	281 (128, 886)	0 (0)	0 (0)	7 (7, 7)
PMS-group-IV	21 (27)	66 (88)	579 (222, 1398)	1 (1)	1 (12)	21(10, 55)

PMS method in selecting a suitable threshold value.

Table 2: Selection accuracy for random decoupling.

	$R^2 = 0.5$				$R^2 = 0.9$			
Method	PIT	FPR*	FNR*	Model Size	PIT	FPR*	FNR*	Model Size
SIS	680 (468)	677 (211)	53 (78)	6979 (5463, 8557)	965 (184)	929 (81)	6 (31)	9603 (9061, 9854)
HOLP	255 (437)	35 (4)	188 (145)	356 (328, 381)	870 (337)	25 (4)	23 (59)	256 (228, 288)
CIS	390 (489)	43 (2)	141 (139)	436 (424, 449)	990 (100)	36 (2)	2 (17)	363 (351, 376)
PartS	285 (453)	41 (3)	198 (161)	418 (403, 438)	1000 (0)	36 (3)	0(0)	370 (350, 391)
PMS-selection	355 (480)	18 (4)	150 (139)	185 (159, 211)	995 (71)	4 (2)	1 (12)	42 (32, 52)
PMS-group	420 (495)	13 (4)	128 (130)	134 (114, 159)	990 (100)	1 (1)	2 (17)	13 (10, 17)

Note: "PIT" refers to the estimated probability of including all true predictors in the top n selected predictors multiplied by 1000, "FPR*" and "FNR*" respectively refer to the false positive and false negative rates multiplied by 1000. The others are same with Table 1 in the main text.

S4.3 Evaluation of CPMS

In this part, we test the performance of CPMS method based on additional numerical studies for the imaging regression. The true signal is same with section 5.2 in the main text. However, the prior information is different, which is summarized in Figure 1. For prior covariance matrix, we set $\mathbf{\Lambda} = (\lambda_{ij}) = (\exp\{-0.3\|\mathbf{s}_i - \mathbf{s}_j\|_2^2\})$. Here we only considered the results of PMS with group level by setting all features in set "PTP" as one group and the others as another group. Firstly, we applied the PMS method to Case I–VI respectively. To apply CPMS, we combine prior information from I and II, III and IV, and V and VI respectively, which are denoted as "CPMS-I", "CPMS-II" and "CPMS-III". All simulation results based on $(n, p) = (200, 10000)$ are summarized in Table 3. From these results, we could find that incorporating prior information from different sources by CPMS can improve the screening results.

Table 3: Screening accuracy for the CPMS method under imaging regression case.

	$R^2 = 0.5$			$R^2 = 0.9$		
Method	FPR	FNR	Model Size	FPR	FNR	Model Size
PMS-I	250 (66)	474 (178)	3763 (2911, 4935)	219 (43)	284 (28)	3576 (3176, 4102)
PMS-II	242 (69)	533 (219)	3996 (3453, 4578)	226 (25)	537 (194)	3396 (3092, 3684)
CPMS-I	237 (76)	402 (123)	3818 (3079, 4404)	200 (32)	373 (109)	2869 (2674, 3095)
PMS-III	305 (92)	503 (187)	4765 (3867, 5608)	290 (36)	443 (145)	4053 (3641, 4514)
PMS-IV	211 (45)	358 (163)	3425 (3006, 3908)	186 (29)	227 (7)	2896 (2671, 3191)
CPMS-II	196 (74)	289 (103)	3458 (2843, 4286)	111 (45)	174 (77)	2647 (2045, 3148)
PMS-V	104 (67)	206 (184)	2111 1350, 2934)	43(32)	41 (79)	1167 (805, 1541)
PMS-VI	93 (73)	183 (199)	1638 (1061, 2445)	31(31)	25 (76)	700 (505, 1089)
CPMS-III	0(0)	0 (0)	217 (217, 217)	0(0)	0(0)	217 (217, 217)

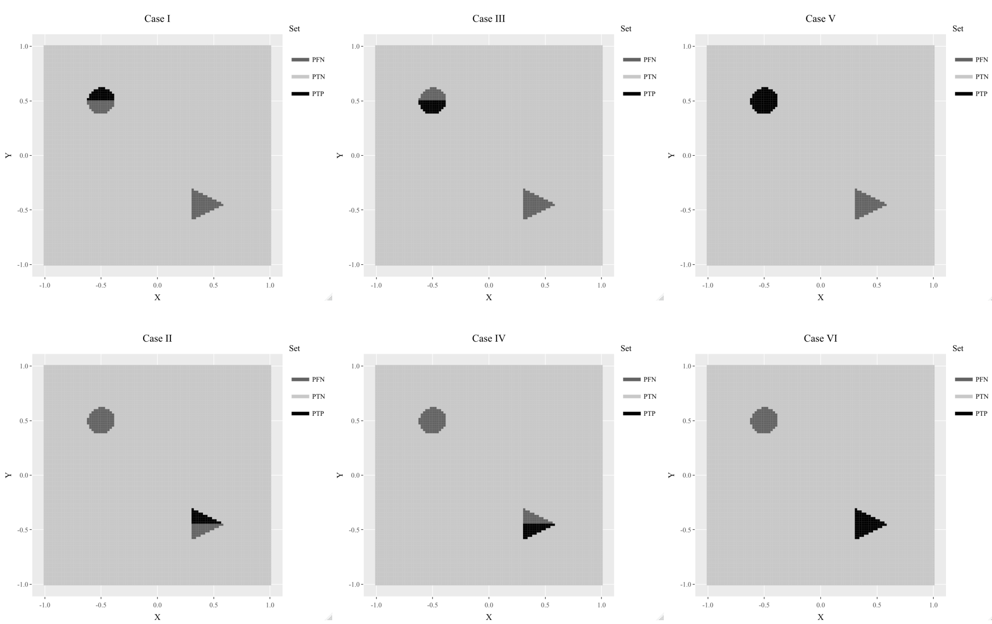

Figure 1: Case plots in CPMS study.

Bibliography

Kang, J., Hong, H. G., and Li, Y. (2017), "Partition-based ultrahigh-dimensional variable screening," *Biometrika*, 104, 785–800.

Lu, Y., Dhillon, P., Foster, D. P., and Ungar, L. (2013), "Faster ridge regression via the subsampled randomized hadamard transform," in *Advances in neural information processing systems*, pp. 369–377.

Wang, X. and Leng, C. (2016), "High dimensional ordinary least squares projection for screening variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 589–611.