Supplementary Materials for "Integrative Factor Regression and Its Inference for Multimodal Data Analysis"

Quefeng Li and Lexin Li

University of North Carolina, Chapel Hill and University of California, Berkeley

S1 Proof of Theorem 1

Proof. We first prove the case when σ_ϵ^2 is known, then prove it when σ_ϵ^2 is unknown. Let

$$\begin{aligned}\widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} &= \sigma_\epsilon^{-2}\left\{\frac{1}{n}\sum_{i=1}^n \widehat{\boldsymbol{f}}_{i,m}^{\otimes 2} - \widehat{\boldsymbol{W}}'\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\widehat{\boldsymbol{f}}'_{i,m}\right)\right\},\\ \widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m},\mathbf{0}) &= \frac{1}{n\sigma_\epsilon^2}\sum_{i=1}^n (y_i - \boldsymbol{x}'_{i,-m}\widehat{\boldsymbol{\beta}}_{-m})(\widehat{\boldsymbol{f}}_{i,m} - \widehat{\boldsymbol{W}}'\boldsymbol{x}_{i,-m}).\end{aligned}$$

We first show that, under H_0, when σ_ϵ^2 is known,

$$\widetilde{\boldsymbol{T}}_n = \sqrt{n}\widetilde{\boldsymbol{I}}^{-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m},\mathbf{0}) \xrightarrow{D} N(\mathbf{0},\boldsymbol{I}_{K_m}). \tag{S1}$$

To prove (S1), we show the following two results:

$$\begin{aligned}\sqrt{n}\boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m},\mathbf{0}) &\xrightarrow{D} N(\mathbf{0},\boldsymbol{I}_{K_m}), &\text{(S2)}\\ \widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} &\xrightarrow{P} \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}. &\text{(S3)}\end{aligned}$$

Then given (S2) and (S3), we obtain (S1) by applying Slutsky's Theorem.

To prove (S2), define

$$\ell(\boldsymbol{\beta}^*_{-m},\boldsymbol{\gamma}^*_m) = \frac{1}{2n\sigma_\epsilon^2}\|\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m} - \widehat{\boldsymbol{F}}_m\boldsymbol{\gamma}^*_m\|_2^2.$$

Then, $\nabla_{\boldsymbol{\beta}_{-m}}\ell(\boldsymbol{\beta}^*_{-m},\boldsymbol{\gamma}^*_m) = -(n\sigma_\epsilon^2)^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m} - \widehat{\boldsymbol{F}}_m\boldsymbol{\gamma}^*_m)'\boldsymbol{X}_{-m}$, $\nabla^2_{\boldsymbol{\beta}_{-m}\boldsymbol{\gamma}_m}\ell(\boldsymbol{\beta}^*_{-m},\boldsymbol{\gamma}^*_m) = (n\sigma_\epsilon^2)^{-1}\widehat{\boldsymbol{F}}'_m\boldsymbol{X}_{-m}$, and $\nabla^2_{\boldsymbol{\beta}_{-m}\boldsymbol{\beta}_{-m}}\ell(\boldsymbol{\beta}^*_{-m},\boldsymbol{\gamma}^*_m) = (n\sigma_\epsilon^2)^{-1}\boldsymbol{X}'_{-m}\boldsymbol{X}_{-m}$. Let

$$\boldsymbol{S}(\boldsymbol{\beta}^*_{-m},\mathbf{0}) = \frac{1}{n\sigma_\epsilon^2}\sum_{i=1}^n (y_i - \boldsymbol{x}_{i,-m}\boldsymbol{\beta}^*_{-m})(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m}).$$

Noting that $\boldsymbol{W}^* = \mathrm{E}(\boldsymbol{x}_{i,-m}^{\otimes 2})^{-1}\,\mathrm{E}(\boldsymbol{x}_{i,-m}\boldsymbol{f}'_{i,m})$, we have

$$\begin{aligned}\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0}) = \boldsymbol{S}(\boldsymbol{\beta}^*_{-m}, \boldsymbol{0}) - \frac{1}{n\sigma_\epsilon^2}(\boldsymbol{W}^* - \widehat{\boldsymbol{W}})'\boldsymbol{X}'_{-m}(\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m}) \\ - \frac{1}{n\sigma_\epsilon^2}(\widehat{\boldsymbol{F}}'_m\boldsymbol{X}_{-m} - \widehat{\boldsymbol{W}}'\boldsymbol{X}'_{-m}\boldsymbol{X}_{-m})(\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m}).\end{aligned} \tag{S4}$$

For the second summand on the right-hand-side of (S4), by Lemma 8, for each $k \in [K_m]$,

$$\begin{aligned}&\left|\frac{1}{n\sigma_\epsilon^2}(\boldsymbol{w}^*_k - \widehat{\boldsymbol{w}}_k)'\boldsymbol{X}'_{-m}(\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m})\right| \\ &\le \|\boldsymbol{w}^*_k - \widehat{\boldsymbol{w}}_k\|_1 \left\|\frac{1}{n\sigma_\epsilon^2}\boldsymbol{X}'_{-m}(\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m})\right\|_\infty = \|\boldsymbol{w}^*_k - \widehat{\boldsymbol{w}}_k\|_1 \left\|\frac{1}{n\sigma_\epsilon^2}\boldsymbol{X}'_{-m}\boldsymbol{\epsilon}\right\|_\infty \\ &= O_P\left(s^*_k\left[\sqrt{(\log p_{-m})/n}\{1 \vee (n^{1/4}/\sqrt{p_m})\}\right]\right) O_P\left(\sqrt{(\log p_{-m})/n}\right) = o_P\left(n^{-1/2}\right),\end{aligned} \tag{S5}$$

where $s^*_k = |\operatorname{supp}(\boldsymbol{w}^*_k)|$. For the third summand on the right-hand-side of (S4), let $(\widehat{\boldsymbol{F}}_m)_k$ denote the kth column of $\widehat{\boldsymbol{F}}_m$. By Lemmas 6 and 7, for each $k \in [K_m]$,

$$\begin{aligned}&\left|\frac{1}{n\sigma_\epsilon^2}\{(\widehat{\boldsymbol{F}}_m)'_k\boldsymbol{X}_{-m} - \widehat{\boldsymbol{w}}'_k\boldsymbol{X}'_{-m}\boldsymbol{X}_{-m}\}(\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m})\right| \\ &\le \left\|\frac{1}{n\sigma_\epsilon^2}\{(\widehat{\boldsymbol{F}}_m)'_k\boldsymbol{X}_{-m} - \widehat{\boldsymbol{w}}'_k\boldsymbol{X}'_{-m}\boldsymbol{X}_{-m}\}\right\|_\infty \|\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m}\|_1 \\ &= O_P\left(s^*_{-m}\left\{\sqrt{(\log p_{-m})/n} + 1/\sqrt{p_m}\right\}\right) O_P\left(\sqrt{\frac{\log p_{-m}}{n}}\left(1 \vee \frac{n^{1/4}}{\sqrt{p_m}}\right)\right) \\ &= o_P\left(n^{-1/2}\right).\end{aligned} \tag{S6}$$

Together with the central limit theorem of $\sqrt{n}\boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\boldsymbol{S}(\boldsymbol{\beta}^*_{-m}, \boldsymbol{0})$, we prove (S2).

To prove (S3), since the dimension K_m is fixed, all matrix norms are equivalent. In particular, we show that

$$\|\widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\|_\infty = o_P(1). \tag{S7}$$

Let $(\widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}})_k$ denote the kth row of $\widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}$. By the definition of the information matrix $\boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}$, the identifiability assumption that $\mathrm{E}(\boldsymbol{f}^{\otimes 2}_{i,m}) = \boldsymbol{I}_{K_m}$, and furthermore the fact that $(1/n)\sum_{i=1}^n \widehat{\boldsymbol{f}}^{\otimes 2}_{i,m} = \boldsymbol{I}_{K_m}$, we have

$$\begin{aligned}\|(\widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}})_k\|_\infty &= \sigma_\epsilon^{-2}\left\|\widehat{\boldsymbol{w}}'_k\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\widehat{\boldsymbol{f}}'_{i,m}\right) - \boldsymbol{w}^{*\prime}_k\mathrm{E}(\boldsymbol{x}_{i,-m}\boldsymbol{f}'_{i,m})\right\|_\infty \\ \lesssim \left\|(\widehat{\boldsymbol{w}}_k - \boldsymbol{w}^*_k)'\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\widehat{\boldsymbol{f}}'_{i,m}\right)\right\|_\infty &+ \left\|\boldsymbol{w}^{*\prime}_k\left\{\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\widehat{\boldsymbol{f}}'_{i,m} - \mathrm{E}(\boldsymbol{x}_{i,-m}\boldsymbol{f}'_{i,m})\right\}\right\|_\infty.\end{aligned}$$

For the first term, let $\widehat{f}_{i,m_h}$ be the hth element of $\widehat{\boldsymbol{f}}_{i,m}$, we have

$$\left\|(\widehat{\boldsymbol{w}}_k - \boldsymbol{w}^*_k)'\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\widehat{\boldsymbol{f}}'_{i,m}\right)\right\|_\infty = \max_{h \in [K_m]}\left|(\widehat{\boldsymbol{w}}_k - \boldsymbol{w}^*_k)'\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\widehat{f}_{i,m_h}\right)\right|$$

$$
\le \left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_1 \cdot \max_{h\in[K_m]} \left[\left\| \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\{\widehat{f}_{i,m_h} - \boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^*\} \right\|_\infty \right.
$$

$$
\left. + \left\|\mathrm{E}(\boldsymbol{x}_{i,-m}\boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^*)\right\|_\infty + \left\| \frac{1}{n}\sum_{i=1}^n \left\{\boldsymbol{x}_{i,-m}\boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^* - \mathrm{E}(\boldsymbol{x}_{i,-m}\boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^*)\right\} \right\|_\infty \right]
$$

$$
= O_P\left(s_k^* \sqrt{\frac{\log p_{-m}}{n}} \left(1 \vee \frac{n^{1/4}}{\sqrt{p_m}}\right)\right) = o_P(1),
$$

where $s_k = |\operatorname{supp}(\boldsymbol{w}_k^*)|$, and the second-to-last equality follows from Lemma 8, and the fact that the dominating term in the bracket is $\|\mathrm{E}(\boldsymbol{x}_{i,-m}\boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^*)\|_\infty = O(1)$. For the second term,

$$
\left\|\boldsymbol{w}_k^{*\prime}\left\{\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\widehat{\boldsymbol{f}}_{i,m}' - \mathrm{E}(\boldsymbol{x}_{i,-m}\boldsymbol{f}_{i,m}')\right\}\right\|_\infty = \max_{h\in[K_m]}\left|\boldsymbol{w}_k^{*\prime}\left\{\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\widehat{f}_{i,m_h} - \mathrm{E}(\boldsymbol{x}_{i,-m}f_{i,m_h})\right\}\right|
$$

$$
\le \|\boldsymbol{w}_k^*\|_1 \cdot \max_{h\in[K_m]}\left[\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,-m}\left(\widehat{f}_{i,m_h} - \boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^*\right)\right\|_\infty \right.
$$

$$
\left. + \left\|\frac{1}{n}\sum_{i=1}^n \left\{\boldsymbol{x}_{i,-m}\boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^* - \mathrm{E}(\boldsymbol{x}_{i,-m}\boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^*)\right\}\right\|_\infty\right]
$$

$$
= O_P\left(s_k^* \sqrt{\frac{\log p_{-m}}{n}} \left(1 \vee \frac{n^{1/4}}{\sqrt{p_m}}\right)\right) = o_P(1),
$$

where the second-to-last equality follows from (S31), and the sub-Gaussian assumption on X_{ij} and $\boldsymbol{x}_{i,-m}'\boldsymbol{w}_h^*$, which is implied by (1) and Condition 1.

Next, when σ_ϵ^2 is unknown, we have $\widehat{\boldsymbol{T}}_n - \widetilde{\boldsymbol{T}}_n = \widetilde{\boldsymbol{T}}_n(\sigma_\epsilon/\widehat{\sigma}_\epsilon - 1) = o_P(1)$, which is implied by Condition 5. Then, applying Slutsky's Theorem completes the proof. □

S2 Proof of Theorem 2

Proof. We divide the proof into two main steps.

In Step 1, letting $\boldsymbol{T}_n^* = \sqrt{n}\boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\left\{\boldsymbol{S}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}_m^*) - \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}_m^*\right\}$, and $Q_n^* = (\boldsymbol{T}_n^*)'\boldsymbol{T}_n^*$, we show that $Q_n = Q_n^* + o_P(1)$. First, recall $\widetilde{\boldsymbol{T}}_n$ as defined in (S1), we have

$$
\boldsymbol{T}_n = (\sigma_\epsilon\widehat{\sigma}_\epsilon^{-1})\widetilde{\boldsymbol{T}}_n = \widetilde{\boldsymbol{T}}_n + \frac{\sigma_\epsilon - \widehat{\sigma}_\epsilon}{\widehat{\sigma}_\epsilon}\widetilde{\boldsymbol{T}}_n = \widetilde{\boldsymbol{T}}_n + o_P(1), \tag{S8}
$$

where the last equality follows from Condition 5, and the o_P statement applies to each element of $\boldsymbol{T}_n$. Next, we show that $\widetilde{\boldsymbol{T}}_n = \boldsymbol{T}_n^* + o_P(1)$. Letting $\boldsymbol{T}_n^\dagger = \sqrt{n}\boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\boldsymbol{S}(\boldsymbol{\beta}_{-m}^*, \boldsymbol{0})$, we have that

$$
\widetilde{\boldsymbol{T}}_n = \boldsymbol{T}_n^\dagger + \sqrt{n}\boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\{\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0}) - \boldsymbol{S}(\boldsymbol{\beta}_{-m}^*, \boldsymbol{0})\} + \sqrt{n}\{\widetilde{\boldsymbol{I}}^{-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\}\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0}).
$$

By Lemmas 9 and 10, uniformly for all $\boldsymbol{\beta}^* \in \mathcal{N}$, we have that

$$
\sqrt{n}\boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\{\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0}) - \boldsymbol{S}(\boldsymbol{\beta}_{-m}^*, \boldsymbol{0})\} = o_P(1),
$$

$$\sqrt{n}\{\widetilde{\boldsymbol{I}}^{-1/2}_{\gamma_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^{*^{-1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\}\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \mathbf{0}) = o_P(1).$$

Therefore, $\widetilde{\boldsymbol{T}}_n = \boldsymbol{T}^{\dagger}_n + o_P(1)$. Recall that $\boldsymbol{S}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*_m) = (n\sigma^2_\epsilon)^{-1}\sum_{i=1}^n \epsilon_i \left(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m}\right)$. Henceforth, we have

$$\begin{aligned}
\boldsymbol{T}^{\dagger}_n &= \sqrt{n}\boldsymbol{I}^{*^{-1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\{\boldsymbol{S}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*_m) - \boldsymbol{I}^*_{\gamma_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m\} \\
&\quad + \sqrt{n}\boldsymbol{I}^{*^{-1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\{\boldsymbol{S}(\boldsymbol{\beta}^*_{-m}, \mathbf{0}) - \boldsymbol{S}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*_m) + \boldsymbol{I}^*_{\gamma_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m\} \\
&= \sqrt{n}\boldsymbol{I}^{*^{-1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\{\boldsymbol{S}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*_m) - \boldsymbol{I}^*_{\gamma_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m\} + o_P(1) = \boldsymbol{T}^*_n + o_P(1),
\end{aligned}$$

where the second equality follows from Lemma 11. Therefore, we have $\widetilde{\boldsymbol{T}}_n = \boldsymbol{T}^*_n + o_P(1)$. Together with (S8) and the continuous mapping theorem, we have $Q_n = Q^*_n + o_P(1)$, which completes Step 1.

In Step 2, we derive the χ^2 approximation of Q^*_n. By definition,

$$\begin{aligned}
&\sqrt{n}\boldsymbol{I}^{*^{-1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\{\boldsymbol{S}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*_m) - \boldsymbol{I}^*_{\gamma_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m\} \\
&= \sqrt{n}\boldsymbol{I}^{*^{-1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\left\{\frac{1}{n\sigma^2_\epsilon}\sum_{i=1}^n \epsilon_i(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m})\right\} - \sqrt{n}\boldsymbol{I}^{*^{1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m \\
&= \sum_{i=1}^n \boldsymbol{\xi}_i - \sqrt{n}\boldsymbol{I}^{*^{1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m,
\end{aligned}$$

where $\boldsymbol{\xi}_i = (\sqrt{n}\sigma^2_\epsilon)^{-1}\boldsymbol{I}^{*^{-1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\epsilon_i(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m})$. By direct calculation, we have that $\mathrm{E}(\boldsymbol{\xi}_i) = \mathbf{0}$, and $\sum_{i=1}^n \mathrm{Var}(\boldsymbol{\xi}_i) = \boldsymbol{I}_{K_m}$. By Conditions 1 and 6, we have

$$\begin{aligned}
\sum_{i=1}^n \mathrm{E}\|\boldsymbol{\xi}\|_2^3 &= \frac{1}{(n\sigma^4_\epsilon)^{3/2}}\mathrm{E}|\epsilon_i|^3\sum_{i=1}^n \mathrm{E}\|\boldsymbol{I}^{*^{-1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m})\|_2^3 \\
&\lesssim \frac{1}{\sqrt{n}}(\mathrm{E}\|\boldsymbol{f}_{i,m}\|_2^3 + \mathrm{E}\|\boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m}\|_2^3) = o(1).
\end{aligned}$$

Then, by Lemma 4, we have that

$$\sup_{\mathcal{C}} |\Pr(\sum_{i=1}^n \boldsymbol{\xi}_i \in \mathcal{C}) - \Pr(\boldsymbol{Z} \in \mathcal{C})| \to 0, \tag{S9}$$

where $\boldsymbol{Z} \sim N(0, K_m)$, and the supremum is taken over all convex sets $\mathcal{C} \in \mathcal{R}^{K_m}$. Consider a special subset $\mathcal{C}_x$ of $\mathcal{C}$, such that $\mathcal{C}_x = \{\boldsymbol{z} \in \mathcal{R}^{K_m} : \|\boldsymbol{z} - \sqrt{n}\boldsymbol{I}^{*^{1/2}}_{\gamma_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m\|_2^2 \le x\}$. It then follows from (S9) that

$$\sup_x |\Pr(Q^*_n \le x) - \Pr(\chi^2(1, h_n) \le x)| = \sup_x |\Pr(\sum_{i=1}^n \boldsymbol{\xi}_i \in \mathcal{C}_x) - \Pr(\boldsymbol{Z} \in \mathcal{C}_x)| \to 0,$$

where $h_n = n\boldsymbol{\gamma}^{*\prime}_m\boldsymbol{I}^*_{\gamma_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m$. Since $Q_n = Q^*_n + o_P(1)$. For any x and $\varepsilon > 0$, we have

$$\begin{aligned}
\Pr(\chi^2(1, h_n) \le x - \varepsilon) + o(1) &\le \Pr(Q^*_n \le x - \varepsilon) + o(1) \le \Pr(Q_n \le x) \\
&\le \Pr(Q^*_n \le x + \varepsilon) + o(1) \le \Pr\left\{\chi^2(1, h_n) \le x + \varepsilon\right\} + o(1).
\end{aligned} \tag{S10}$$

In addition, by Lemma 5, we have

$$\lim_{\varepsilon\to 0}\limsup_{n}\left|\Pr\{\chi^2(1,h_n)\le x+\varepsilon\}-\Pr\{\chi^2(1,h_n)\le x-\varepsilon\}\right|\to 0.$$

Together, we have

$$\sup_x\left|\Pr(Q_n\le x)-\Pr(\chi^2(1,h_n)\le x)\right|\to 0.$$

This completes the proof. □

S3 Proof of Corollary 1

Proof. We only prove (b) when $\phi_{\gamma_m}=1/2$. The proofs of (a) and (c) are similar.

Note that

$$\begin{aligned}&|\Pr(Q_n\le x)-\Pr(\chi^2(K_m,h)\le x)|\\ \le\ &|\Pr(Q_n\le x)-\Pr(\chi^2(K_m,h_{m_n})\le x)|+|\Pr(\chi^2(K_m,h_{m_n})\le x)-\Pr(\chi^2(K_m,h)\le x)|\end{aligned}$$

Then, by Theorem 2, it suffices to prove that

$$\lim_{n\to\infty}\sup_{x>0}|\Pr(\chi^2(K_m,h_{m_n})\le x)-\Pr(\chi^2(K_m,h)\le x)|=0.$$

Let $F(x;k,\lambda)$ denote the cumulative distribution function of the non-central χ^2-distribution with k degrees of freedom and non-centrality parameter λ, and $F(x;k)$ that of the central χ^2-distribution with k degrees of freedom. Then,

$$F(x;k,\lambda)=e^{-\lambda/2}\sum_{j=1}^{\infty}\frac{(\lambda/2)^j}{j!}F(x;k+2j),$$

We have that

$$\begin{aligned}&\Pr(\chi^2(K_m,h_{m_n})\le x)-\Pr(\chi^2(K_m,h)\le x)\\ =\ &e^{-h_{m_n}/2}\left\{F(x;K_m)+\sum_{j=1}^{\infty}\frac{(h_{m_n}/2)^j}{j!}F(x;K_m+2j)\right\}\\ &-e^{-h/2}\left\{F(x;K_m)+\sum_{j=1}^{\infty}\frac{(h/2)^j}{j!}F(x;K_m+2j)\right\}\end{aligned}$$

Then, as $n\to\infty$,

$$\begin{aligned}&\sup_{x>0}|\Pr(\chi^2(K_m,h_{m_n})\le x)-\Pr(\chi^2(K_m,h)\le x)|\\ \le\ &|e^{-h_{m_n}/2}-e^{-h/2}|\sup_{x>0}\left\{F(x;K_m)+\sum_{j=1}^{\infty}\frac{(h_{m_n}/2)^j}{j!}F(x;K_m+2j)\right\}\\ &+e^{-h/2}\sup_{x>0}\left\{\sum_{j=1}^{\infty}\frac{|(h_{m_{n}/2})^j-(h/2)^j|}{j!}F(x;K_m+2j)\right\}\\ \le\ &|e^{-h_{m_n}/2}-e^{-h/2}|e^{h_{m_n}/2}+e^{-h/2}\sum_{j=1}^{\infty}\frac{|(h_{m_n}/2)^j-(h/2)^j|}{j!}\to 0.\end{aligned}$$

This completes the proof. □

S4 Proof of Theorem 3

Proof. We first prove (a) to (c) of the theorem. Our main idea is to show that, by Lemma 2, $\widehat{\boldsymbol{F}}$ is consistent to $\boldsymbol{HF}$ for some nonsingular matrix $\boldsymbol{H}$. Moreover, $\widehat{\boldsymbol{U}}$ is consistent to $\boldsymbol{U}$. Then solving (10) is equivalent to solving the same problem by replacing $(\widehat{\boldsymbol{F}}, \widehat{\boldsymbol{U}})$ with $(\boldsymbol{F}, \boldsymbol{U})$. More precisely, by the Karush-Kuhn-Tucker conditions, any vector $(\widehat{\boldsymbol{\gamma}}_a, \widehat{\boldsymbol{\beta}}_a)$ satisfying the following equations is a solution to (10):

$$\frac{1}{n}\widehat{\boldsymbol{F}}'(\boldsymbol{Y} - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a - \widehat{\boldsymbol{U}}_{T\cup S_a}\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}) = \boldsymbol{0}; \tag{S11}$$

$$\frac{1}{n}\widehat{\boldsymbol{U}}_T'(\boldsymbol{Y} - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a - \widehat{\boldsymbol{U}}_{T\cup S_a}\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}) = \boldsymbol{0}; \tag{S12}$$

$$\frac{1}{n}\widehat{\boldsymbol{U}}_{S_a}'(\boldsymbol{Y} - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a - \widehat{\boldsymbol{U}}_{T\cup S_a}\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}) = \lambda_a \dot{p}(|\widehat{\boldsymbol{\beta}}_{a,S_a}|) I(\widehat{\boldsymbol{\beta}}_{a,S_a} > \boldsymbol{0}); \tag{S13}$$

$$\left\| \frac{1}{n}\widehat{\boldsymbol{U}}_{(T\cup S_a)^c}'(\boldsymbol{Y} - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a - \widehat{\boldsymbol{U}}_{T\cup S_a}\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}) \right\|_\infty < \lambda_a \dot{p}(0+). \tag{S14}$$

where $\dot{p}(\cdot)$ is a vector of first derivatives of $p(\cdot)$, and $I(\cdot)$ is a vector of indicator functions applied to each coordinate of $\widehat{\boldsymbol{\beta}}_{a,S_a}$.

We divide the proof into two main steps. In Step 1, letting $\mathcal{M} = \{(\boldsymbol{\gamma}, \boldsymbol{\beta}) : \|\boldsymbol{\gamma} - \boldsymbol{H}'\boldsymbol{\gamma}^*\|_\infty \le C\delta_n, \|\boldsymbol{\beta} - \boldsymbol{\beta}^*_{T\cup S_a}\|_\infty \le C\delta_n\}$ for some constant C, we show that, with probability tending to 1, there exists a vector $(\widehat{\boldsymbol{\gamma}}_a, \widehat{\boldsymbol{\beta}}_{a,T\cup S_a})$ in $\mathcal{M}$ that satisfies (S11), (S12) and (S13). In Step 2, we set $\widehat{\boldsymbol{\beta}}_a = (\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}, \boldsymbol{0})$, and show that $(\widehat{\boldsymbol{\gamma}}_a, \widehat{\boldsymbol{\beta}}_a)$ satisfies (S14). Together these two steps prove (a) and (b) of the theorem. Then (c) follows from $\|\widehat{\boldsymbol{\beta}}_{a,T\cup S_a} - \boldsymbol{\beta}^*_{T\cup S_a}\|_2 \le \sqrt{t + s_a}\|\widehat{\boldsymbol{\beta}}_{a,T\cup S_a} - \boldsymbol{\beta}^*_{T\cup S_a}\|_\infty$.

For Step 1, by (1), (2), and $\boldsymbol{\zeta} = \boldsymbol{F}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a + (\boldsymbol{U}_{T\cup S_a} - \widehat{\boldsymbol{U}}_{T\cup S_a})\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}$, we have

$$\boldsymbol{Y} - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a - \widehat{\boldsymbol{U}}_{T\cup S_a}\widehat{\boldsymbol{\beta}}_{a,T\cup S_a} = \boldsymbol{U}_{T\cup S_a}(\boldsymbol{\beta}^*_{T\cup S_a} - \widehat{\boldsymbol{\beta}}_{a,T\cup S_a}) + \boldsymbol{\epsilon} + \boldsymbol{\zeta}.$$

Therefore,

$$\begin{aligned}
&\frac{1}{n}\begin{pmatrix} \widehat{\boldsymbol{U}}_T' \\ \widehat{\boldsymbol{U}}_{S_a}' \end{pmatrix}\left(\boldsymbol{Y} - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a - \widehat{\boldsymbol{U}}_{T\cup S_a}\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}\right) \\
= \;& \frac{1}{n}\begin{pmatrix} \widehat{\boldsymbol{U}}_T'\boldsymbol{U}_T & \widehat{\boldsymbol{U}}_T'\boldsymbol{U}_{S_a} \\ \widehat{\boldsymbol{U}}_{S_a}'\boldsymbol{U}_T & \widehat{\boldsymbol{U}}_{S_a}'\boldsymbol{U}_{S_a} \end{pmatrix}\begin{pmatrix} \boldsymbol{\beta}^*_T - \widehat{\boldsymbol{\beta}}_{a,T} \\ \boldsymbol{\beta}^*_{S_a} - \widehat{\boldsymbol{\beta}}_{a,S_a} \end{pmatrix} + \frac{1}{n}\begin{pmatrix} \widehat{\boldsymbol{U}}_T'(\boldsymbol{\epsilon} + \boldsymbol{\zeta}) \\ \widehat{\boldsymbol{U}}_{S_a}'(\boldsymbol{\epsilon} + \boldsymbol{\zeta}) \end{pmatrix} \\
= \;& \boldsymbol{K}_n\begin{pmatrix} \boldsymbol{\beta}^*_T - \widehat{\boldsymbol{\beta}}_{a,T} \\ \boldsymbol{\beta}^*_{S_a} - \widehat{\boldsymbol{\beta}}_{a,S_a} \end{pmatrix} + \frac{1}{n}\begin{pmatrix} \boldsymbol{U}_T'\boldsymbol{\epsilon} \\ \boldsymbol{U}_{S_a}'\boldsymbol{\epsilon} \end{pmatrix} + \frac{1}{n}\begin{pmatrix} (\widehat{\boldsymbol{U}}_T - \boldsymbol{U}_T)'\boldsymbol{\epsilon} + \widehat{\boldsymbol{U}}_T'\boldsymbol{\zeta} \\ (\widehat{\boldsymbol{U}}_{S_a} - \boldsymbol{U}_{S_a})'\boldsymbol{\epsilon} + \widehat{\boldsymbol{U}}_{S_a}'\boldsymbol{\zeta} \end{pmatrix} \\
& - \frac{1}{n}\begin{pmatrix} (\boldsymbol{U}_T - \widehat{\boldsymbol{U}}_T)'\boldsymbol{U}_T & (\boldsymbol{U}_T - \widehat{\boldsymbol{U}}_T)'\boldsymbol{U}_{S_a} \\ (\boldsymbol{U}_{S_a} - \widehat{\boldsymbol{U}}_{S_a})'\boldsymbol{U}_T & (\boldsymbol{U}_{S_a} - \widehat{\boldsymbol{U}}_{S_a})'\boldsymbol{U}_{S_a} \end{pmatrix}\begin{pmatrix} \boldsymbol{\beta}^*_T - \widehat{\boldsymbol{\beta}}_{a,T} \\ \boldsymbol{\beta}^*_{S_a} - \widehat{\boldsymbol{\beta}}_{a,S_a} \end{pmatrix} \\
= \;& \boldsymbol{K}_n\begin{pmatrix} \boldsymbol{\beta}^*_T - \widehat{\boldsymbol{\beta}}_{a,T} \\ \boldsymbol{\beta}^*_{S_a} - \widehat{\boldsymbol{\beta}}_{a,S_a} \end{pmatrix} + \frac{1}{n}\begin{pmatrix} \boldsymbol{U}_T'\boldsymbol{\epsilon} \\ \boldsymbol{U}_{S_a}'\boldsymbol{\epsilon} \end{pmatrix} + \begin{pmatrix} \boldsymbol{R}_T \\ \boldsymbol{R}_{S_a} \end{pmatrix},
\end{aligned} \tag{S15}$$

where

$$\boldsymbol{R}_{T\cup S_a} = \frac{1}{n}\left\{(\widehat{\boldsymbol{U}}_{T\cup S_a} - \boldsymbol{U}_{T\cup S_a})'\boldsymbol{U}_{T\cup S_a}\boldsymbol{\beta}^*_{T\cup S_a} + \widehat{\boldsymbol{U}}_{T\cup S_a}'(\boldsymbol{F}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a) + (\widehat{\boldsymbol{U}}_{T\cup S_a} - \boldsymbol{U}_{T\cup S_a})'\boldsymbol{\epsilon}\right\}.$$

By Lemma 12, we have

$$\|\boldsymbol{R}_{T\cup S_a}\|_\infty = O_P\left(\sqrt{\frac{\log p}{n}}\left(\frac{1}{\sqrt{n}}+\frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right).$$

In addition, by Condition 1, $U_{ij}\epsilon_i$ is sub-exponential. Therefore, by Bernstein inequality and the union bound, we have $\|n^{-1}\boldsymbol{U}'_{T\cup S_a}\boldsymbol{\epsilon}\|_\infty = O_P\left(\sqrt{(\log p)/n}\right)$. Together,

$$\|\frac{1}{n}\boldsymbol{U}'_{T\cup S_a}\boldsymbol{\epsilon}\|_\infty + \|\boldsymbol{R}_{T\cup S_a}\|_\infty = O_P\left(\delta_n\right), \text{ where } \delta_n = \sqrt{(\log p)/n}\{1\vee(n^{1/4}/\sqrt{p_{\min}})\}. \quad \text{(S16)}$$

By Condition 7 and the sub-Gaussian assumption on $\boldsymbol{u}$, it holds with probability tending to 1 that $\lambda_{\min}(\boldsymbol{K}_n)$ is bounded away from 0. Then by (S15), there exists $\widehat{\boldsymbol{\beta}}_{a,T}\in\mathcal{M}$ that solves (S12). By the assumption that $\sqrt{n}\lambda_a\dot{p}(d_n)=o(1)$, we have $\lambda_a\dot{p}(|\boldsymbol{\beta}_{a,S_a}|)\le\lambda_a\dot{p}(d_n)=o(\delta_n)$ for all $\boldsymbol{\beta}_{a,S_a}\in\mathcal{M}$. Then, by (S15), there exists $\widehat{\boldsymbol{\beta}}_{a,S_a}\in\mathcal{M}$ that solves (S13). Finally, as we show in Lemma 13, $\|n^{-1}\widehat{\boldsymbol{F}}'(\boldsymbol{Y}-\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\beta}}_a-\widehat{\boldsymbol{F}}\boldsymbol{H}'\boldsymbol{\gamma}^*)\|_\infty = O_P\left(\delta_n\right)$, and thus there exists $\widehat{\boldsymbol{\beta}}_{a,S_a}\in\mathcal{M}$ that solves (S11), which completes Step 1.

For Step 2, let $\boldsymbol{\varphi}_{T\cup S_a} = (\boldsymbol{0},\lambda_a\dot{p}(|\widehat{\boldsymbol{\beta}}_{a,S_a}|)I(\widehat{\boldsymbol{\beta}}_{a,S_a}>\boldsymbol{0}))$, we have

$$\begin{aligned}
& n^{-1}\widehat{\boldsymbol{U}}'_{(T\cup S_a)^c}(\boldsymbol{Y}-\widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a-\widehat{\boldsymbol{U}}_{T\cup S_a}\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}) \\
= \;& n^{-1}\widehat{\boldsymbol{U}}'_{(T\cup S_a)^c}\boldsymbol{U}_{T\cup S_a}(\boldsymbol{\beta}^*_{T\cup S_a}-\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}) + n^{-1}\widehat{\boldsymbol{U}}'_{(T\cup S_a)^c}\boldsymbol{\epsilon} + n^{-1}\widehat{\boldsymbol{U}}'_{(T\cup S_a)^c}\boldsymbol{\zeta} \\
= \;& n^{-1}\boldsymbol{U}'_{(T\cup S_a)^c}\boldsymbol{U}_{T\cup S_a}\boldsymbol{K}_n^{-1}\{\boldsymbol{\varphi}_{T\cup S_a} - n^{-1}\boldsymbol{U}'_{T\cup S_a}\boldsymbol{\epsilon}-\boldsymbol{R}_{T\cup S_a}\} \\
& + n^{-1}\widehat{\boldsymbol{U}}'_{(T\cup S_a)^c}\boldsymbol{\epsilon} + n^{-1}\widehat{\boldsymbol{U}}'_{(T\cup S_a)^c}(\boldsymbol{F}\boldsymbol{\gamma}^*-\widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a) + n^{-1}\widehat{\boldsymbol{U}}_{(T\cup S_a)^c}(\boldsymbol{U}_{T\cup S_a}-\widehat{\boldsymbol{U}}_{T\cup S_a})\boldsymbol{\beta}^*_{T\cup S_a}.
\end{aligned}$$

By Condition 9 and the sub-Gaussian assumption on $\boldsymbol{U}$, with probability tending to 1, we have $\|n^{-1}\boldsymbol{U}'_{(T\cup S_a)^c}\boldsymbol{U}_{T\cup S_a}\boldsymbol{K}_n^{-1}\|_{L_\infty}=O(1)$. Therefore, by (S16), we have

$$\begin{aligned}
& \lambda_a^{-1}\|n^{-1}\boldsymbol{U}'_{(T\cup S_a)^c}\boldsymbol{U}_{T\cup S_a}\boldsymbol{K}_n^{-1}\{n^{-1}\boldsymbol{U}'_{T\cup S_a}\boldsymbol{\epsilon}+\boldsymbol{R}_{T\cup S_a}\}\|_\infty \\
\le \;& \lambda_a^{-1}\|n^{-1}\boldsymbol{U}'_{(T\cup S_a)^c}\boldsymbol{U}_{T\cup S_a}\boldsymbol{K}_n^{-1}\|_{L_\infty}\|n^{-1}\boldsymbol{U}'_{T\cup S_a}\boldsymbol{\epsilon}+\boldsymbol{R}_{T\cup S_a}\|_\infty \\
= \;& O_P\left(\delta_n\lambda_a^{-1}\right) = o_P(1).
\end{aligned}$$

Next,

$$\lambda_a^{-1}\|n^{-1}\boldsymbol{U}'_{(T\cup S_a)^c}\boldsymbol{U}_{T\cup S_a}\boldsymbol{K}_n^{-1}\lambda_a\dot{p}(|\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}|)\|_\infty \lesssim \dot{p}(|\widehat{\boldsymbol{\beta}}_{a,T\cup S_a}|) < \dot{p}(d_n) < \dot{p}(0+).$$

Moreover, by Lemma 12,

$$\begin{aligned}
& \lambda_a^{-1}n^{-1}\|\widehat{\boldsymbol{U}}'_{(T\cup S_a)^c}\boldsymbol{\epsilon}+\widehat{\boldsymbol{U}}'_{(T\cup S_a)^c}(\boldsymbol{F}\boldsymbol{\gamma}^*-\widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)+\widehat{\boldsymbol{U}}_{(T\cup S_a)^c}(\boldsymbol{U}_{T\cup S_a}-\widehat{\boldsymbol{U}}_{T\cup S_a})\boldsymbol{\beta}^*_{T\cup S_a}\|_\infty \\
= \;& O_P\left(\delta_n\lambda_a^{-1}\right) = o_P(1).
\end{aligned}$$

Putting together the above results, we have that $\widehat{\boldsymbol{\beta}}$ satisfies (S14), which completes Step 2.

Finally, we prove (d) of the theorem. By Lemma 12, when $p_{\min}\gg n^{3/2}$, $\|\boldsymbol{R}_{T\cup S}\|_\infty = o_P\left(n^{-1/2}\right)$. Then, it follows from (S12), (S14) and (S15) that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{a,T}-\boldsymbol{\beta}^*_T) \;=\; \frac{1}{\sqrt{n}}\boldsymbol{K}_n^{-1}\boldsymbol{U}'_T\boldsymbol{\epsilon}+o_P\left(1\right).$$

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{a,S_a} - \boldsymbol{\beta}^*_{S_a}) \;=\; \frac{1}{\sqrt{n}}\boldsymbol{K}_n^{-1}\boldsymbol{U}'_{S_a}\boldsymbol{\epsilon} - \boldsymbol{K}_n^{-1}\sqrt{n}\lambda_a\dot{p}(|\widehat{\boldsymbol{\beta}}_{a,S_a}|)I(\widehat{\boldsymbol{\beta}}_{a,S_a} > \mathbf{0}) + o_P\left(1\right).$$

Since $\sqrt{n}\lambda_a\dot{p}(|\widehat{\boldsymbol{\beta}}_{a,S_a}|) \leq \sqrt{n}\lambda_a\dot{p}(d_n) = o(1)$. Therefore,

$$\boldsymbol{K}_n^{-1}\sqrt{n}\lambda_a\dot{p}(|\widehat{\boldsymbol{\beta}}_{a,S}|) = O_P\left(\sqrt{n}\lambda_a\dot{p}(d_n)\right) = o_P\left(1\right).$$

This completes the proof. □

S5 Proof of Theorem 4

Proof. We divide the proof into two main steps.

In Step 1, define T_0 as

$$T_0 = \sigma_\epsilon^{-2}(\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{h}_n)'\boldsymbol{\Psi}^{-1}(\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{h}_n),$$

where $\boldsymbol{h}_n = \boldsymbol{A}\boldsymbol{\beta}^* - \boldsymbol{b}$, $\boldsymbol{\Psi} = \boldsymbol{A}\boldsymbol{\Omega}_T\boldsymbol{A}'$, $\boldsymbol{\Omega}_T$ is the the submatrix of $\boldsymbol{\Omega}_{T\cup S_a} = \boldsymbol{\Sigma}^{-1}_{u,T\cup S_a}$ with rows and columns in T, and $\boldsymbol{\Sigma}^{-1}_{u,T\cup S_a}$ is inverse of the submatrix of $\boldsymbol{\Sigma}_u$ with rows and columns in $T \cup S_a$, and $\boldsymbol{\omega}_n = n^{-1/2}\begin{pmatrix}\boldsymbol{A} & \mathbf{0}\end{pmatrix}\boldsymbol{K}_n^{-1}\boldsymbol{U}'_{T\cup S_a}\boldsymbol{\epsilon}$. We first show that $T_w/r = T_0/r + o_P\left(1\right)$. By Theorem 3, we have

$$\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{\beta}^*_T \\ \widehat{\boldsymbol{\beta}}_{S_a} - \boldsymbol{\beta}^*_{S_a}\end{pmatrix} = \frac{1}{\sqrt{n}}\boldsymbol{K}_n^{-1}\begin{pmatrix}\boldsymbol{U}'_T\boldsymbol{\epsilon} \\ \boldsymbol{U}'_{S_a}\boldsymbol{\epsilon}\end{pmatrix} + \boldsymbol{R}_a,$$

for some remainder term $\boldsymbol{R}_a$ such that $\|\boldsymbol{R}_a\|_2 = o_P\left(1\right)$. Then we have

$$\sqrt{n}\boldsymbol{A}(\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{\beta}^*_T) = \boldsymbol{\omega}_n + \boldsymbol{A}\boldsymbol{R}_{a,T},$$

where $\boldsymbol{R}_{a,T}$ is the subvector of $\boldsymbol{R}_a$ with indices in T. By definition, $\boldsymbol{A}\boldsymbol{\beta}^*_T - \boldsymbol{b} = \boldsymbol{h}_n$. Then,

$$\sqrt{n}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b}) = \boldsymbol{\omega}_n + \boldsymbol{A}\boldsymbol{R}_{a,T} + \sqrt{n}\boldsymbol{h}_n.$$

Let $\boldsymbol{\Psi}_n = \boldsymbol{A}(\boldsymbol{K}_n^{-1})_T\boldsymbol{A}'$, where $(\boldsymbol{K}_n^{-1})_T$ is submatrix of $\boldsymbol{K}_n^{-1}$ with rows and columns in T. We have

$$\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b}) = \boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{\omega}_n + \boldsymbol{A}\boldsymbol{R}_{a,T} + \sqrt{n}\boldsymbol{h}_n). \tag{S17}$$

Next, we bound $\|\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b})\|_2$. By Lemmas 3 and 12, $\|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{A}\|_2 = O_P\left(1\right)$, and $\|\boldsymbol{R}_{a,T}\|_2 = o_P\left(1\right)$. Then it follows that

$$\|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{A}\boldsymbol{R}_{a,T}\|_2 \leq \|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{A}\|_2\|\boldsymbol{R}_{a,T}\|_2 = o_P\left(1\right). \tag{S18}$$

Therefore, $\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b}) = \boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{h}_n) + o_P\left(1\right)$. We further note that,

$$\mathrm{E}\|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{\omega}_n\|_2^2 = \mathrm{tr}\left[\mathrm{E}_{\boldsymbol{u}}\{\boldsymbol{\Psi}_n^{-1/2}\mathrm{E}_\epsilon(\boldsymbol{\omega}_n\boldsymbol{\omega}'_n)\boldsymbol{\Psi}_n^{-1/2}\}\right] = r\sigma_\epsilon^2.$$

Then, by Markov's inequality, $\|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{\omega}_n\|_2 = O_P(\sqrt{r})$. By Lemma 3, $\lambda_{\max}(\boldsymbol{\Psi}_n^{-1}) = O_P(1)$. By Condition 11, $\|\boldsymbol{h}_n\|_2 = O(\sqrt{r/n})$, so that $\|\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{h}_n\|_2 = O_P(\sqrt{r})$. Therefore, $\|\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b})\|_2 = O_P(\sqrt{r})$. By Lemma 3, $\|\boldsymbol{\Psi}_n^{1/2}(\boldsymbol{A}\widehat{\boldsymbol{\Omega}}_T\boldsymbol{A}')^{-1}\boldsymbol{\Psi}_n^{1/2} - \boldsymbol{I}\|_2 = O_P((s_a+t)/\sqrt{n})$, where $\widehat{\boldsymbol{\Omega}}_T$ is the first T rows and columns of the matrix

$$\widehat{\boldsymbol{\Omega}}_{T\cup\widehat{S}_a} = n\begin{pmatrix}\widehat{\boldsymbol{U}}_T'\widehat{\boldsymbol{U}}_T & \widehat{\boldsymbol{U}}_T'\widehat{\boldsymbol{U}}_{\widehat{S}_a}\\ \widehat{\boldsymbol{U}}_{\widehat{S}_a}'\widehat{\boldsymbol{U}}_T & \widehat{\boldsymbol{U}}_{\widehat{S}_a}'\widehat{\boldsymbol{U}}_{\widehat{S}_a}\end{pmatrix}^{-1}.$$

Therefore, under the assumption that $s_a + t = o(n^{1/2})$, we have

$$\begin{aligned}
&\|\{\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b})\}'\{\boldsymbol{\Psi}_n^{1/2}(\boldsymbol{A}\widehat{\boldsymbol{\Omega}}_T\boldsymbol{A}')^{-1}\boldsymbol{\Psi}_n^{1/2} - \boldsymbol{I}\}\{\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b})\}\|_2^2\\
\leq\ & \|\boldsymbol{\Psi}_n^{1/2}(\boldsymbol{A}\widehat{\boldsymbol{\Omega}}_T\boldsymbol{A}')^{-1}\boldsymbol{\Psi}_n - \boldsymbol{I}\|_2\|\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b})\|_2^2\\
=\ & O_P\left(\frac{r(s_a+t)}{\sqrt{n}}\right) = o_P(r). \qquad \text{(S19)}
\end{aligned}$$

Let $T_{w,0} = \widehat{\sigma}_\epsilon^{-2}n(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b})'\boldsymbol{\Psi}_n^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{a,T} - \boldsymbol{b})$. By T_w's definition and (S19), we have $\widehat{\sigma}_\epsilon^2|T_w - T_{w,0}| = o_P(r)$. Condition 5 implies $1/\widehat{\sigma}_\epsilon^2 = O_P(1)$. Therefore, $|T_w - T_{w,0}| = o_P(r)$.

Next, we show that $|T_{w,0} - T_0| = o_P(r)$. Let $T_{w,1} = \widehat{\sigma}_\epsilon^{-2}\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{\Psi}^{-1/2}\boldsymbol{h}_n\|_2$. By (S17) and (S18), we have

$$\begin{aligned}
\widehat{\sigma}_\epsilon^2 T_{w,0} &= \|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{h}_n + o_P(1)\|_2^2\\
&= \|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{h}_n\|_2^2 + o_P(1) + o_P\left(\boldsymbol{\Psi}_n^{-1/2}(\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{h}_n)\right)\\
&= \|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{h}_n\|_2^2 + o_P(1) + o_P(r)\\
&= \widehat{\sigma}_\epsilon^2 T_{w,1} + (\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{h}_n)'(\boldsymbol{\Psi}_n^{-1} - \boldsymbol{\Psi}^{-1})(\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{h}_n) + o_P(r).
\end{aligned}$$

By Lemma 3, we have that $\|\boldsymbol{\Psi}_n^{-1} - \boldsymbol{\Psi}^{-1}\|_2 = o_P(1)$. Since $\|\boldsymbol{\omega}_n\|_2 \leq \|\boldsymbol{\Psi}_n^{1/2}\|_2\|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{\omega}_n\|_2 = O_P(\sqrt{r})$, and by Condition 11, $\|\sqrt{n}\boldsymbol{h}_n\|_2 = O_P(\sqrt{r})$. Therefore, $\|\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{h}_n\|_2 = O_P(\sqrt{r})$. Considering that $1/\widehat{\sigma}_\epsilon^2 = O(1)$, we have $T_{w,0} = T_{w,1} + o_P(r)$. Finally, we have

$$|T_{w,1} - T_0| = \frac{|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2|}{\widehat{\sigma}_\epsilon^2\sigma_\epsilon^2}\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{\Psi}^{-1/2}\boldsymbol{h}_n\|_2 = o_P(r).$$

Therefore $|T_{w,0} - T_0| = o_P(r)$.

Combining the results $|T_w - T_{w,0}| = o_P(r)$ and $|T_{w,0} - T_0| = o_P(r)$ completes Step 1.

In Step 2, we show that the χ^2 approximation holds for T_0. Recall the definition of T_0, which can be written as $T_0 = \sigma_\epsilon^{-2}\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n + \sqrt{n}\boldsymbol{\Psi}^{-1/2}\boldsymbol{h}_n\|_2^2$. By the definition of $\boldsymbol{\omega}_n$,

$$\sigma_\epsilon^{-1}\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n = \sum_{i=1}^n \frac{1}{\sqrt{n\sigma_\epsilon^2}}\boldsymbol{\Psi}^{-1/2}\begin{pmatrix}\boldsymbol{A} & \boldsymbol{0}\end{pmatrix}\boldsymbol{\Omega}_{T\cup S_a}\boldsymbol{U}_{i,T\cup S_a}\epsilon_i = \sum_{i=1}^n \boldsymbol{\xi}_i.$$

By direct calculation, we have $\sum_{i=1}^n \mathrm{Var}(\boldsymbol{\xi}_i) = \boldsymbol{I}_r$. Because of the sub-Gaussian assumption on ϵ in Condition 1, we have $E|\epsilon|^3 < \infty$. Then,

$$r^{1/4}\sum_{i=1}^n \mathrm{E}\|\boldsymbol{\xi}_i\|_2^3 = \frac{r^{1/4}}{(n\sigma_\epsilon^2)^{3/2}}\sum_{i=1}^n \mathrm{E}\|\boldsymbol{\Psi}^{-1/2}\begin{pmatrix}\boldsymbol{A} & \boldsymbol{0}\end{pmatrix}\boldsymbol{\Omega}_{T\cup S_a}\boldsymbol{U}_{i,T\cup S_a}\epsilon_i\|_2^3$$

$$
\begin{aligned}
&= \frac{r^{1/4}}{(n\sigma_\epsilon^2)^{3/2}} \mathrm{E}|\epsilon|^3 \sum_{i=1}^n \mathrm{E}\|\boldsymbol{\Psi}^{-1/2} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \boldsymbol{\Omega}_{T\cup S_a} \boldsymbol{U}_{i,T\cup S_a}\|_2^3 \\
&\lesssim \frac{r^{1/4}}{(n\sigma_\epsilon^2)^{3/2}} \sum_{i=1}^n \mathrm{E}\{\|\boldsymbol{\Psi}^{-1/2} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \boldsymbol{\Omega}_{T\cup S_a}^{1/2}\|_2^3 \|\boldsymbol{\Omega}_{T\cup S_a}^{1/2} \boldsymbol{U}_{i,T\cup S_a}\|_2^3\} \\
&\lesssim \frac{r^{1/4}}{(n\sigma_\epsilon^2)^{3/2}} \sum_{i=1}^n \mathrm{E}|\boldsymbol{U}_{i,T\cup S_a}' \boldsymbol{\Omega}_{T\cup S_a} \boldsymbol{U}_{i,T\cup S_a}|^{3/2}, \\
&\lesssim \frac{r^{1/4}}{n^{1/2}} \mathrm{E}|\boldsymbol{u}_{T\cup S_a}' \boldsymbol{\Omega}_{T\cup S_a} \boldsymbol{u}_{T\cup S_a}|^{3/2} = o(1),
\end{aligned}
$$

where the third-to-last relation is due to the fact that

$$
\|\boldsymbol{\Psi}^{-1/2} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \boldsymbol{\Omega}_{T\cup S_a}^{1/2}\|_2 = \lambda_{\max} \left[\mathrm{tr} \left\{ \boldsymbol{\Psi}^{-1/2} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \boldsymbol{\Omega}_{T\cup S_a} \begin{pmatrix} \boldsymbol{A}' \\ \boldsymbol{0}' \end{pmatrix} \boldsymbol{\Psi}^{-1/2} \right\} \right] = r,
$$

and the last equality follows from Condition 12. Then by Lemma 4, we have

$$
\sup_{\mathcal{C}} \left| \Pr\{\sigma_\epsilon^{-1} \boldsymbol{\Psi}^{-1/2} \boldsymbol{\omega}_n \in \mathcal{C}\} - \Pr(\boldsymbol{Z} \in \mathcal{C}) \right| \to 0, \tag{S20}
$$

where $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}_r)$ and the supremum is taken over all convex sets $\mathcal{C} \in \mathcal{R}^r$.

Consider a special subset $\mathcal{C}_x$ of $\mathcal{C}$, where

$$
\mathcal{C}_x = \{\boldsymbol{z} \in \mathcal{R}^r : \|\boldsymbol{z} + \sqrt{n/\sigma_\epsilon^2} \boldsymbol{\Psi}^{-1/2} \boldsymbol{h}_n\|_2^2 \le x\}.
$$

It follows from (S20) that

$$
\sup_x \left| \Pr\{(n\sigma_\epsilon^2)^{-1/2} \boldsymbol{\Psi}^{-1/2} \boldsymbol{\omega}_n \in \mathcal{C}_x\} - \Pr(\boldsymbol{Z} \in \mathcal{C}_x) \right| = \sup_x \left| \Pr(T_0 \le x) - \Pr(\chi^2(r, \nu_n) \le x) \right| \to 0,
$$

where $\nu_n = n\sigma_\epsilon^{-2} \boldsymbol{h}_n' \boldsymbol{\Psi}^{-1} \boldsymbol{h}_n$.

Consider any statistic $T^* = T_0 + o_P(r)$. For any x and $\varepsilon > 0$, we have

$$
\begin{aligned}
\Pr(\chi^2(r, \nu_n) &\le x - r\varepsilon) + o(1) \le \Pr(T_0 \le x - r\varepsilon) + o(1) \le \Pr(T^* \le x) \\
&\le \Pr(T_0 \le x + r\varepsilon) + o(1) \le \Pr\left\{\chi^2(r, \nu_n) \le x + r\varepsilon\right\} + o(1).
\end{aligned}
$$

In addition, by Lemma 5, we have

$$
\lim_{\varepsilon \to 0} \limsup_n \left| \Pr\{\chi^2(r, \nu_n) \le x + r\varepsilon\} - \Pr\{\chi^2(r, \nu_n) \le x - r\varepsilon\} \right| \to 0.
$$

Together, we have

$$
\sup_x \left| \Pr(T^* \le x) - \Pr(\chi^2(r, \nu_n) \le x) \right| \to 0.
$$

This completes Step 2.

Combining the results of Step 1 and Step 2 completes the proof. □

S6 Consistence of $\widehat{\sigma}_\epsilon^2$

Recall the estimator for σ_ϵ in Section 3, $\widehat{\sigma}_\epsilon^2 = n^{-1}\sum_{i=1}^n (y_i - \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}})^2$, where $\widetilde{\boldsymbol{\beta}} = \operatorname{argmin}_{\boldsymbol{\beta}} (2n)^{-1} \sum_{i=1}^n (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \lambda_\epsilon \|\boldsymbol{\beta}\|_1$. The next proposition shows that $\widehat{\sigma}_\epsilon^2$ is a consistent estimator of σ_ϵ^2, and thus $\widehat{\sigma}_\epsilon^2$ satisfies Condition 5.

Proposition S1. *Suppose $\boldsymbol{x}_m$ satisfies the factor decomposition in (1) for $m \in [M]$, and Conditions 1 and 3 hold. Suppose $s^*(\log p)/n = o(1)$, where $s^* = |\operatorname{supp}(\boldsymbol{\beta}^*)|$, and $\lambda_\epsilon = C\sqrt{(\log p)/n}$, where C is a positive constant. Then $\widehat{\sigma}_\epsilon^2 = \sigma_\epsilon^2 + o_P(1)$.*

Proof. Letting $\boldsymbol{H}_x = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i^{\otimes 2}$ and $\widehat{\boldsymbol{\Delta}} = \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$, we have

$$\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2 = \frac{1}{n}\sum_{i=1}^n \epsilon_i^2 - \sigma_\epsilon^2 + \widehat{\boldsymbol{\Delta}}'\boldsymbol{H}_x\widehat{\boldsymbol{\Delta}} - \widehat{\boldsymbol{\Delta}}'\left(\frac{2}{n}\sum_{i=1}^n \epsilon_i \boldsymbol{x}_i\right). \tag{S21}$$

By the sub-Gaussian assumption on ϵ_i in Condition 1, it follows from the standard concentration result (e.g., Ning and Liu, 2017, Lemma H.2) that

$$\frac{1}{n}\sum_{i=1}^n \epsilon_i^2 - \sigma_\epsilon^2 = O_P\left(\sqrt{(\log n)/n}\right).$$

Condition 3 implies that $\lambda_{\min}(\mathrm{E}(\boldsymbol{x}_i^{\otimes 2})) \geq \lambda_{\min}(\mathrm{E}(\boldsymbol{u}_i^{\otimes 2})) > c$. Then, it follows from Raskutti et al. (2011, Proposition 1) that the restricted eigenvalue condition holds for $\boldsymbol{H}_x$. Then, following a similar argument as in the proof of Ning and Liu (2017, Lemma B.3), we have

$$\widehat{\boldsymbol{\Delta}}'\boldsymbol{H}_x\widehat{\boldsymbol{\Delta}} = O_P\left(s^*(\log p)/n\right). \tag{S22}$$

Moreover, $\epsilon_i x_{ij} = \epsilon_i(\sum_{k=1}^K \lambda_{jk} f_{ik} + u_{ij})$, where λ_{jk} is the (j,k)th element of $\boldsymbol{\Lambda}$. It follows from Conditions 1 and 3 that x_{ij} is also sub-Gaussian. Therefore, $\epsilon_i x_{ij}$ is sub-exponential. This implies that $\|n^{-1}\sum_{i=1}^n \epsilon_i \boldsymbol{x}_i\|_\infty = O_P(\sqrt{(\log p)/n})$. By the well-known estimation error of the Lasso estimator (e.g., Negahban et al., 2012, Corollary 2), it holds that $\|\widehat{\boldsymbol{\Delta}}\|_1 = O_P\left(s^*\sqrt{(\log p)/n}\right)$. Therefore,

$$\left|(2/n)\widehat{\boldsymbol{\Delta}}'\sum_{i=1}^n \epsilon_i \boldsymbol{x}_i\right| \leq \|\widehat{\boldsymbol{\Delta}}\|_1 \|(2/n)\sum_{i=1}^n \epsilon_i \boldsymbol{x}_i\|_\infty = O_P\left(s^*(\log p)/n\right). \tag{S23}$$

Putting (S21), (S22) and (S23) together completes the proof. □

S7 Proof of Proposition 1

Proof. To prove (a), consider regressing y using all but the mth modality. Letting $\boldsymbol{P}_{-m} = \boldsymbol{X}_{-m}(\boldsymbol{X}_{-m}'\boldsymbol{X}_{-m})^{-1}\boldsymbol{X}_{-m}'$, then $\widehat{\boldsymbol{Y}}_{-m} = \boldsymbol{P}_{-m}\boldsymbol{Y}$. Letting $\boldsymbol{\xi} = \boldsymbol{X}_m\boldsymbol{\beta}_m^* + \boldsymbol{\epsilon}$, then

$$\begin{aligned}\|\boldsymbol{Y} - \widehat{\boldsymbol{Y}}_{-m}\|_2^2 &= \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{Y} = \boldsymbol{\xi}'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{\xi} \\ &= \boldsymbol{\epsilon}'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{\epsilon} + 2(\boldsymbol{X}_m\boldsymbol{\beta}_m^*)'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{\epsilon} + (\boldsymbol{X}_m\boldsymbol{\beta}_m^*)'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{X}_m\boldsymbol{\beta}_m^*.\end{aligned}$$

Taking the expectation on both sides, and noting that,

$$\begin{aligned}
\mathrm{E}\{\boldsymbol{\epsilon}'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{\epsilon}\} &= (n - p_{-m})\sigma_\epsilon^2, \\
\mathrm{E}\{2(\boldsymbol{X}_m\boldsymbol{\beta}_m^*)'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{\epsilon}\} &= 0, \\
\mathrm{E}\{(\boldsymbol{X}_m\boldsymbol{\beta}_m^*)'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{X}_m\boldsymbol{\beta}_m^*\} &= \mathrm{E}_{\boldsymbol{x}_{-m}}[\mathrm{E}_{\boldsymbol{x}_m|\boldsymbol{x}_{-m}}\{(\boldsymbol{X}_m\boldsymbol{\beta}_m^*)'(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\boldsymbol{X}_m\boldsymbol{\beta}_m^*\}] \\
&= \mathrm{E}_{\boldsymbol{x}_{-m}}[\mathrm{tr}\{(\boldsymbol{I}_n - \boldsymbol{P}_{-m})\sigma_{m|-m}^2\}] = (n - p_{-m})\sigma_{m|-m}^2.
\end{aligned}$$

To prove (b), note that $\mathrm{E}\|\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\|_2^2 = (n-p)\sigma_\epsilon^2$ when regressing y on all data modalities. Then a direct calculation proves (b).

To prove (c), by factor decomposition, we have $\boldsymbol{x}_m = \boldsymbol{\Lambda}_m\boldsymbol{f} + \boldsymbol{u}_m$ and $\boldsymbol{x}_{-m} = \boldsymbol{\Lambda}_{-m}\boldsymbol{f} + \boldsymbol{u}_{-m}$. Then

$$\begin{pmatrix}\boldsymbol{u}_{-m}\\ \boldsymbol{x}_{-m}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{u_{-m}} & \boldsymbol{\Sigma}_{u_{-m}}\\ \boldsymbol{\Sigma}_{u_{-m}} & \boldsymbol{\Sigma}_{x_{-m}}\end{pmatrix}\right),$$

where $\boldsymbol{\Sigma}_{x_{-m}} = \boldsymbol{\Lambda}_{-m}\boldsymbol{\Lambda}'_{-m} + \boldsymbol{\Sigma}_{u_{-m}}$. Consequently,

$$\mathrm{E}(\boldsymbol{u}_{-m}|\boldsymbol{x}_{-m}) = \boldsymbol{\Sigma}_{u_{-m}}\boldsymbol{\Sigma}_{x_{-m}}^{-1}\boldsymbol{x}_{-m}, \quad \mathrm{Var}(\boldsymbol{u}_{-m}|\boldsymbol{x}_{-m}) = \boldsymbol{\Sigma}_{u_{-m}} - \boldsymbol{\Sigma}_{u_{-m}}\boldsymbol{\Sigma}_{x_{-m}}^{-1}\boldsymbol{\Sigma}_{u_{-m}}.$$

As $\boldsymbol{\Lambda}_{-m}\boldsymbol{f} = \boldsymbol{x}_{-m} - \boldsymbol{u}_{-m}$, then $\boldsymbol{f} = \boldsymbol{D}_{-m}^{-1}\boldsymbol{\Lambda}'_{-m}(\boldsymbol{x}_{-m} - \boldsymbol{u}_{-m})$, where $\boldsymbol{D}_{-m} = \boldsymbol{\Lambda}'_{-m}\boldsymbol{\Lambda}_{-m}$. Then,

$$\boldsymbol{x}'_m\boldsymbol{\beta}_m^* = \boldsymbol{f}'\boldsymbol{\Lambda}'_m\boldsymbol{\beta}_m^* + \boldsymbol{u}'_m\boldsymbol{\beta}_m^* = (\boldsymbol{x}_{-m} - \boldsymbol{u}_{-m})'\boldsymbol{C}_m\boldsymbol{\beta}_m^* + \boldsymbol{u}'_m\boldsymbol{\beta}_m^*,$$

where $\boldsymbol{C}_m = \boldsymbol{\Lambda}_{-m}\boldsymbol{D}_{-m}^{-1}\boldsymbol{\Lambda}'_m$. Therefore, we have

$$\mathrm{E}(\boldsymbol{x}'_m\boldsymbol{\beta}_m^*|\boldsymbol{x}_{-m}) = \{\boldsymbol{x}_{-m} - \mathrm{E}(\boldsymbol{u}_{-m}|\boldsymbol{x}_{-m})\}'\boldsymbol{C}_m\boldsymbol{\beta}_m^* = \boldsymbol{x}'_{-m}(\boldsymbol{I} - \boldsymbol{\Sigma}_{u_{-m}}\boldsymbol{\Sigma}_{x_{-m}}^{-1})'\boldsymbol{C}_m\boldsymbol{\beta}_m^*. \tag{S24}$$

Moreover,

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{x}'_m\boldsymbol{\beta}_{-m}^*|\boldsymbol{x}_{-m}) &= \boldsymbol{\beta}_m^{*\prime}\{\boldsymbol{C}'_m\mathrm{Var}(\boldsymbol{u}_{-m}|\boldsymbol{x}_{-m})\boldsymbol{C}_m + \boldsymbol{\Sigma}_{u_m}\}\boldsymbol{\beta}_m^* \\
&= \boldsymbol{\beta}_m^{*\prime}\{\boldsymbol{C}'_m(\boldsymbol{\Sigma}_{u_{-m}} - \boldsymbol{\Sigma}_{u_{-m}}\boldsymbol{\Sigma}_{x_{-m}}^{-1}\boldsymbol{\Sigma}_{u_{-m}})\boldsymbol{C}_m + \boldsymbol{\Sigma}_{u_m}\}\boldsymbol{\beta}_m^*. \qquad \text{(S25)}
\end{aligned}$$

Then, by Woodbury matrix identity, we have

$$\boldsymbol{\Sigma}_{x_{-m}}^{-1} = \boldsymbol{\Sigma}_{u_{-m}}^{-1} - \boldsymbol{\Sigma}_{u_{-m}}^{-1}\boldsymbol{\Lambda}_{-m}(\boldsymbol{I}_K + \boldsymbol{\Lambda}'_{-m}\boldsymbol{\Sigma}_{u_{-m}}^{-1}\boldsymbol{\Lambda}_{-m})^{-1}\boldsymbol{\Lambda}'_{-m}\boldsymbol{\Sigma}_{u_{-m}}^{-1}.$$

Then we have,

$$\begin{aligned}
\boldsymbol{\Sigma}_{u_{-m}} - \boldsymbol{\Sigma}_{u_{-m}}\boldsymbol{\Sigma}_{x_{-m}}^{-1}\boldsymbol{\Sigma}_{u_{-m}} &= \boldsymbol{\Lambda}_{-m}(\boldsymbol{I}_K + \boldsymbol{\Lambda}'_{-m}\boldsymbol{\Sigma}_{u_{-m}}^{-1}\boldsymbol{\Lambda}_{-m})^{-1}\boldsymbol{\Lambda}'_{-m}, \\
\boldsymbol{C}'_m(\boldsymbol{\Sigma}_{u_{-m}} - \boldsymbol{\Sigma}_{u_{-m}}\boldsymbol{\Sigma}_{x_{-m}}^{-1}\boldsymbol{\Sigma}_{u_{-m}})\boldsymbol{C}_m &= \boldsymbol{\Lambda}_m(\boldsymbol{I}_K + \boldsymbol{\Lambda}'_{-m}\boldsymbol{\Sigma}_{u_{-m}}^{-1}\boldsymbol{\Lambda}_{-m})^{-1}\boldsymbol{\Lambda}'_m.
\end{aligned}$$

Plugging these equalities into (S25) gives

$$\mathrm{Var}(\boldsymbol{x}'_m\boldsymbol{\beta}_{-m}^*|\boldsymbol{x}_{-m}) = \boldsymbol{\beta}_m^{*\prime}\{\boldsymbol{\Lambda}_m(\boldsymbol{I}_K + \boldsymbol{\Lambda}'_{-m}\boldsymbol{\Sigma}_{u_{-m}}^{-1}\boldsymbol{\Lambda}_{-m})^{-1}\boldsymbol{\Lambda}'_m + \boldsymbol{\Sigma}_{u_m}\}\boldsymbol{\beta}_m^*,$$

which completes the proof of (c). □

S8 Additional technical lemmas

Lemma 1. *Suppose Conditions 1–3 hold. For any $m \in [M]$, there exists a nonsingular matrix $\boldsymbol{H}_m \in \mathcal{R}^{K_m \times K_m}$, such that*

(a) $\max_{k\in[K_m]}(1/n)\sum_{i=1}^{n}|(\widehat{\boldsymbol{F}}_m\boldsymbol{H}_m)_{ik}-f_{i,m_k}|^2 = O_P\left(1/n+1/p_m\right)$, *where* $(\widehat{\boldsymbol{F}}_m\boldsymbol{H}_m)_{ik}$ *is the* (i,k)*th element of* $\widehat{\boldsymbol{F}}_m\boldsymbol{H}_m$, *and* f_{i,m_k} *is the* k*th element of* $\boldsymbol{f}_{i,m}$.

(b) $\max_{i\in[n]}\|\widehat{\boldsymbol{f}}_{i,m}-\boldsymbol{H}_m\boldsymbol{f}_{i,m}\|_2 = O_P\left(1/\sqrt{n}+n^{1/4}/\sqrt{p_m}\right)$.

(c) $\|\boldsymbol{I}_{K_m}-\boldsymbol{H}_m\boldsymbol{H}_m'\|_2 = O_P\left(1/\sqrt{n}+1/\sqrt{p_m}\right)$.

(d) $\max_{i\in[n],j\in[p_m]}|\widehat{u}_{ij}-u_{ij}| = O_P\left(\sqrt{(\log n)/n}+n^{1/4}/\sqrt{p_m}\right)$.

Proof. The results of (a) and (b) follow from Lemma C.9 of Fan et al. (2013). The results of (c) and (d) follow from Lemmas 3 and A.3 of Li et al. (2018). □

Lemma 2. *Let $\boldsymbol{F} = (\boldsymbol{F}_1,\ldots,\boldsymbol{F}_M)$, $\widehat{\boldsymbol{F}} = (\widehat{\boldsymbol{F}}_1,\ldots,\widehat{\boldsymbol{F}}_M)$, where $\widehat{\boldsymbol{F}}_m$ is obtained by running PCA on the mth modality, $\boldsymbol{H} = \mathrm{diag}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_M)$, $\boldsymbol{U} = (\boldsymbol{U}_1,\ldots,\boldsymbol{U}_m)$, $\widehat{\boldsymbol{U}} = (\widehat{\boldsymbol{U}}_1,\ldots,\widehat{\boldsymbol{U}}_M)$, and $p_{\min} = \min_{m\in[M]}p_m$. Then the following results hold:*

(a) $\max_{i\in[n]}\|\widehat{\boldsymbol{f}}_i-\boldsymbol{H}\boldsymbol{f}_i\|_2 = O_P\left(1/\sqrt{n}+n^{1/4}/\sqrt{p_{\min}}\right)$.

(b) $\|\boldsymbol{I}_K-\boldsymbol{H}\boldsymbol{H}'\|_2 = O_P\left(1/\sqrt{n}+1/\sqrt{p_{\min}}\right)$.

(c) $\max_{i\in[n],j\in[p]}|\widehat{u}_{ij}-u_{ij}| = O_P\left(\sqrt{(\log n)/n}+n^{1/4}/\sqrt{p_{\min}}\right)$.

Proof. The results follow given Lemma 1 and the fact that the convergence rate depends on the slowest one among all M modalities. □

Lemma 3. *Suppose the conditions of Theorem 4 hold. Then the following results hold:*

(a) $\|\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{A}\|_2 = O_P(1)$.

(b) $\|\sqrt{n}\boldsymbol{\Psi}_n^{-1/2}\boldsymbol{h}_n\|_2 = O_P(\sqrt{r})$.

(c) $\|\boldsymbol{\Psi}_n^{1/2}(\boldsymbol{A}\widehat{\boldsymbol{\Omega}}_T\boldsymbol{A}')^{-1}\boldsymbol{\Psi}_n^{1/2}-\boldsymbol{I}\|_2 = O_P((t+s_a)/\sqrt{n})$.

(d) $\|\boldsymbol{\Psi}_n^{-1}-\boldsymbol{\Psi}^{-1}\|_2 = o_P(1)$.

Proof. The results of (a)–(c) follow from (5.4), (5.5) and (5.6) of Shi et al. (2019).

For (d), we have $\boldsymbol{\Psi}_n^{-1}-\boldsymbol{\Psi}^{-1} = \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_n)\boldsymbol{\Psi}^{-1}$. Therefore, $\|\boldsymbol{\Psi}_n^{-1}-\boldsymbol{\Psi}^{-1}\|_2 \asymp \|\boldsymbol{\Psi}-\boldsymbol{\Psi}_n\|_2$. Moreover,

$$\|\boldsymbol{\Psi}-\boldsymbol{\Psi}_n\|_2 = \|\boldsymbol{A}\{\boldsymbol{\Omega}_T-(\boldsymbol{K}_n^{-1})_T\}\boldsymbol{A}'\|_2 \lesssim \|\boldsymbol{\Omega}_T-(\boldsymbol{K}_n^{-1})_T\|_2 \le \|\boldsymbol{\Sigma}_{u,T\cup S_a}^{-1}-\boldsymbol{K}_n^{-1}\|_2,$$

where $\boldsymbol{\Sigma}_{u,T\cup S_a}$ is the submatrix of $\boldsymbol{\Sigma}_u$ with rows and columns in $T\cup S_a$. By the sub-Gaussian assumption on $\boldsymbol{u}$, we have $\|\boldsymbol{K}_n-\boldsymbol{\Sigma}_{u,T\cup S_a}\|_\infty = O_P\left(\sqrt{\{\log(t+s_a)\}/n}\right)$. Then,

$$\|\boldsymbol{K}_n-\boldsymbol{\Sigma}_{u,T\cup S_a}\|_2 \le (t+s_a)\|\boldsymbol{K}_n-\boldsymbol{\Sigma}_{u,T\cup S_a}\|_\infty = o_P(1).$$

Consequently,

$$\|\boldsymbol{\Psi} - \boldsymbol{\Psi}_n\|_2 \lesssim \|\boldsymbol{\Sigma}_{u,T\cup S_a}^{-1} - \boldsymbol{K}_n^{-1}\|_2 \leq \|\boldsymbol{K}_n^{-1}\|_2 \|\boldsymbol{\Sigma}_{u,T\cup S_a}^{-1}\|_2 \|\boldsymbol{K}_n - \boldsymbol{\Sigma}_{u,T\cup S_a}\|_2 = o_P\left(1\right).$$

This completes the proof. □

Lemma 4. *Let $\{\boldsymbol{X}_i\}_{i=1}^n$ denote independent p-dimensional random vectors, with* $\mathrm{E}(\boldsymbol{X}_i) = \boldsymbol{0}$ *and* $\sum_{i=1}^n \mathrm{Var}(\boldsymbol{X}_i) = \boldsymbol{I}_p$. *Let $\boldsymbol{Z}$ denote a p-dimensional multivariate normal vector, with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}_p$. Then,*

$$\sup_C \left| P\left(\sum_{i=1}^n \boldsymbol{X}_i \in C\right) - P(\boldsymbol{Z} \in C)\right| \leq c_0 p^{1/4} \sum_{i=1}^n \mathrm{E}(\|\boldsymbol{X}_i\|_2^3),$$

for some constant c_0, where the supremum is taken over all convex subsets in $\mathcal{R}^p$.

Proof. The result follows from Lemma S6 of Shi et al. (2019), which was originally given in Theorem 1 of Bentkus (2005). □

Lemma 5. *Let $\chi^2(r,\gamma)$ denote a χ^2 random variable with r degrees of freedom and the non-centrality parameter γ. Then,*

$$\lim_{\epsilon\to 0+} \sup_{r\geq 1,\gamma\geq 0} |P(\chi^2(r,\gamma) \leq x + r\varepsilon) - P(\chi^2(r,\gamma) \leq x - r\varepsilon)| \to 0.$$

Proof. The result follows from Lemma S7 of Shi et al. (2019). □

Lemma 6. *Suppose Conditions 1–3 hold, and $\lambda_1 \asymp \sqrt{(\log p_{-m})/n} + 1/\sqrt{p_m}$. Then,*

$$\|\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m}\|_1 = O_P\left(s^*_{-m}\left\{\sqrt{(\log p_{-m})/n} + 1/\sqrt{p_m}\right\}\right).$$

Proof. Recall that

$$(\widehat{\boldsymbol{\beta}}_{-m}, \widehat{\boldsymbol{\gamma}}_m) = \underset{(\boldsymbol{\beta}_{-m},\boldsymbol{\gamma}_m)}{\operatorname{argmin}} \frac{1}{2n}\sum_{i=1}^n (y_i - \boldsymbol{x}'_{i,-m}\boldsymbol{\beta}_{-m} - \widehat{\boldsymbol{f}}'_{i,m}\boldsymbol{\gamma}_m)^2 + \lambda_1\|\boldsymbol{\beta}\|_1. \tag{S26}$$

By Lemma 1, there exists a nonsingular matrix $\boldsymbol{H}_m \in \mathcal{R}^{K_m\times K_m}$ such that $\widetilde{\boldsymbol{F}}_m = \widehat{\boldsymbol{F}}_m\boldsymbol{H}_m$ is a consistent estimator of $\boldsymbol{F}_m$. We note that (S26) is equivalent to

$$(\widetilde{\boldsymbol{\beta}}_{-m}, \widetilde{\boldsymbol{\gamma}}_m) = \underset{(\boldsymbol{\beta}_{-m},\boldsymbol{\gamma}_m)}{\operatorname{argmin}} \frac{1}{2n}\sum_{i=1}^n (y_i - \boldsymbol{x}'_{i,-m}\boldsymbol{\beta}_{-m} - \widetilde{\boldsymbol{f}}'_{i,m}\boldsymbol{\gamma}_m)^2 + \lambda_1\|\boldsymbol{\beta}\|_1, \tag{S27}$$

where $\widetilde{\boldsymbol{f}}'_{i,m}$ is the ith row of $\widetilde{\boldsymbol{F}}_m$. Then, solving (S27) is equivalent as replacing $\widetilde{\boldsymbol{f}}_{i,m}$ with $\boldsymbol{f}_{i,m}$, which becomes a standard M-estimation problem.

Let $\boldsymbol{Q} = (\boldsymbol{X}_{-m}, \boldsymbol{F}_m)$, $\widehat{\boldsymbol{Q}} = (\boldsymbol{X}_{-m}, \widehat{\boldsymbol{F}}_m)$, and $\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}\widetilde{\boldsymbol{H}}$, where $\widetilde{\boldsymbol{H}} = \mathrm{diag}(\boldsymbol{I}_{p_{-m}}, \boldsymbol{H}_m) \in \mathcal{R}^{q_{-m}\times q_{-m}}$ is a block-diagonal matrix, and $q_{-m} = p_{-m} + K_m$. Let $\boldsymbol{\vartheta} = (\boldsymbol{\beta}'_{-m}, \boldsymbol{\gamma}'_m)' \in \mathcal{R}^{q_{-m}}$, $\widehat{\boldsymbol{\vartheta}} = (\widehat{\boldsymbol{\beta}}'_{-m}, \widehat{\boldsymbol{\gamma}}'_m)' \in \mathcal{R}^{q_{-m}}$ denote the solution of (S26), $\widetilde{\boldsymbol{\vartheta}} = \widetilde{\boldsymbol{H}}^{-1}\widehat{\boldsymbol{\vartheta}} \in \mathcal{R}^{q_{-m}}$ and $\widetilde{\boldsymbol{\vartheta}}^* =$

$\widetilde{\boldsymbol{H}}^{-1}(\boldsymbol{\beta}^{*\prime}_{-m}, \boldsymbol{\gamma}^{*\prime}_{m})' \in \mathcal{R}^{q_{-m}}$. By direct calculation, we can verify that $\widehat{\boldsymbol{\beta}}_{-m} = \widehat{\boldsymbol{\vartheta}}_{[p_{-m}]} = \widetilde{\boldsymbol{\vartheta}}_{[p_{-m}]}$, where $\widetilde{\boldsymbol{\vartheta}} = (\widetilde{\boldsymbol{\beta}}'_{-m}, \widetilde{\boldsymbol{\gamma}}'_{m})'$ solves (S27). Then, it follows that

$$\|\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m}\|_1 = \|\widetilde{\boldsymbol{\vartheta}}_{[p_{-m}]} - \widetilde{\boldsymbol{\vartheta}}^*_{[p_{-m}]}\|_1 \leq \|\widetilde{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^*\|_1.$$

To bound $\|\widetilde{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^*\|_1$, we turn to bound $\|\nabla \ell(\boldsymbol{\vartheta})\|_\infty$, and check the restricted eigenvalue condition on $\nabla^2 \ell(\boldsymbol{\vartheta})$, where $\ell(\boldsymbol{\vartheta}) = (2n)^{-1} \sum_{i=1}^n (y_i - \boldsymbol{x}'_{i,-m}\boldsymbol{\beta}_{-m} - \widetilde{\boldsymbol{f}}'_{i,m}\boldsymbol{\gamma}_m)^2$.

To bound $\|\nabla \ell(\boldsymbol{\vartheta})\|_\infty$, we aim to show that

$$\Big\|\nabla \ell(\boldsymbol{\vartheta})\Big\|_\infty = \Big\|\frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{Q}}_i \epsilon_i\Big\|_\infty = O_P\left(\sqrt{\frac{\log p_{-m}}{n}} + \frac{1}{\sqrt{p_m}}\right), \tag{S28}$$

where $\widetilde{\boldsymbol{Q}}_i$ is the ith row of $\widetilde{\boldsymbol{Q}}$. Indeed,

$$\Big\|\frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{Q}}_i \epsilon_i\Big\|_\infty \leq \max_{j\in[q_{-m}]}\left|\frac{1}{n}\sum_{i=1}^n Q_{ij}\epsilon_i\right| + \max_{j\in[q_{-m}]}\left|\frac{1}{n}\sum_{i=1}^n (\widetilde{Q}_{ij} - Q_{ij})\epsilon_i\right|.$$

By Condition 1 and Bernstein inequality,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n Q_{ij}\epsilon_i\right| > C\sqrt{\frac{\log q_{-m}}{n}}\right) \leq q_{-m}^{-2} \text{ for all } j \in [q_{-m}].$$

Then, by the union bound,

$$\max_{j\in[q_{-m}]}\left|\frac{1}{n}\sum_{i=1}^n Q_{ij}\epsilon_i\right| = O_P\left(\sqrt{(\log q_{-m})/n}\right) = O_P\left(\sqrt{(\log p_{-m})/n}\right), \tag{S29}$$

where the last equality is due to the fact that, since K_m is fixed, $q_{-m} \asymp p_{-m}$. We choose to present the results using p_{-m} in order to unify the presentation. Then by Cauchy-Schwatz inequality,

$$\max_{j\in[q_{-m}]}\left|\frac{1}{n}\sum_{i=1}^n (\widetilde{Q}_{ij} - Q_{ij})\epsilon_i\right| \leq \max_{j\in[q_{-m}]}\left(\frac{1}{n}\sum_{i=1}^n (\widetilde{Q}_{ij} - Q_{ij})^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \epsilon_i^2\right)^{1/2}$$

If we use the mth modality to obtain $\widehat{\boldsymbol{f}}_{i,m}$, it follows from Lemma 1 that

$$\max_{k\in[K_m]} (1/n)\sum_{i=1}^n |(\widehat{\boldsymbol{F}}_m\boldsymbol{H}_m)_{ik} - f_{i,m_k}|^2 = O_P\left(1/n + 1/p_m\right),$$

where $(\widehat{\boldsymbol{F}}_m\boldsymbol{H}_m)_{ik}$ is the (i,k)th element of $\widehat{\boldsymbol{F}}_m\boldsymbol{H}_m$ and f_{i,m_k} is the kth element of $\boldsymbol{f}_{i,m}$. This implies that

$$\max_{j\in[q_{-m}]}\left(\frac{1}{n}\sum_{i=1}^n (\widetilde{Q}_{ij} - Q_{ij})^2\right)^{1/2} = O_P\left(1/\sqrt{n} + 1/\sqrt{q_m}\right) = O_P\left(1/\sqrt{n} + 1/\sqrt{p_m}\right).$$

Since $(n^{-1}\sum_{i=1}^{n}\epsilon_i^2)^{1/2} = O_P(1)$, we have

$$\max_{j\in[q_{-m}]}\left|\frac{1}{n}\sum_{i=1}^{n}(\widetilde{Q}_{ij}-Q_{ij})\epsilon_i\right| = O_P\left(1/\sqrt{n}+1/\sqrt{p_m}\right). \tag{S30}$$

Combining (S29) and (S30) together proves (S28).

To check the restricted eigenvalue condition, it follows from Condition 3 and the factor decomposition (1) that $\lambda_{\min}(\mathrm{E}(\boldsymbol{x}_{-m}^{\otimes 2})) \geq \lambda_{\min}(\mathrm{E}(\boldsymbol{u}_{-m}^{\otimes 2})) > c$. In addition, the sub-Gaussian assumptions on $\boldsymbol{f}_{-m}$ and $\boldsymbol{u}_{-m}$ imply that $\boldsymbol{x}_{-m}$ is also sub-Gaussian. Since $\nabla^2_{\boldsymbol{\beta}_{S_{-m}}\boldsymbol{\beta}_{S_{-m}}}\ell(\boldsymbol{\vartheta}) = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_{i,S_{-m}}^{\otimes 2}$, where $S_{-m} = \{j\in[p_{-m}] : \beta_j^* \neq 0\}$. Then, it follows from Proposition 1 of Raskutti et al. (2011) that the restricted eigenvalue condition holds with high probability.

Given that both (S28) and the restricted eigenvalue condition hold, the rest of the proof follows the standard arguments of the high-dimensional M-estimator (Negahban et al., 2012, Theorem 1). A relevant proof in the context of factor model can be found in Theorem 4.2 of Fan et al. (2016). This completes the proof. □

Lemma 7. *Suppose Conditions 1–3 hold. Then,*

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(\widehat{f}_{i,m_k}-\boldsymbol{x}_{i,-m}'\widehat{\boldsymbol{w}}_k)\right\|_\infty = O_P\left(\sqrt{\frac{\log p_{-m}}{n}}\left(1\vee\frac{n^{1/4}}{\sqrt{p_m}}\right)\right).$$

Proof. By definition, it suffices to show that, if we choose $\lambda_2 = C\sqrt{(\log p_{-m})/n}\{1\vee(n^{1/4}/\sqrt{p_m})\}$ for some large enough constant C, $\boldsymbol{w}_k^*$ is in the feasible set with high probability; i.e.,

$$\frac{1}{n}\left\|\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(\widehat{f}_{i,m_k}-\boldsymbol{x}_{i,-m}'\boldsymbol{w}_k^*)\right\|_\infty \leq C\sqrt{(\log p_{-m})/n}\{1\vee(n^{1/4}/\sqrt{p_m})\}. \tag{S31}$$

We have

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(\widehat{f}_{i,m_k}-\boldsymbol{x}_{i,-m}'\boldsymbol{w}_k^*) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(\widehat{f}_{i,m_k}-f_{i,m_k}^\dagger)+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(f_{i,m_k}^\dagger-\boldsymbol{x}_{i,-m}'\boldsymbol{w}_k^*),$$

where $\boldsymbol{f}_{i,m}^\dagger = \boldsymbol{H}_m\boldsymbol{f}_{i,m}$ for some non-singular matrix $\boldsymbol{H}_m \in \mathcal{R}^{K_m\times K_m}$, and $f_{i,m_k}^\dagger$ is the kth element of $\boldsymbol{f}_{i,m}^\dagger$. The sub-Gaussian assumption in Condition 1 implies that $X_{ij}f_{i,m_k}^\dagger$ and $X_{ij}\boldsymbol{x}_{i,-m}'\boldsymbol{w}_k^*$ are sub-exponential. Then, by Bernstein inequality and the union bound,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(f_{i,m_k}^\dagger-\boldsymbol{x}_{i,-m}'\boldsymbol{w}_k^*)\right\|_\infty = O_P\left(\sqrt{\frac{\log p_{-m}}{n}}\right).$$

On the other hand, we have

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(\widehat{f}_{i,m_k}-f_{i,m_k}^\dagger)\right\|_\infty \leq \left(\max_{i\leq n}|\widehat{f}_{i,m_k}-f_{i,m_k}^\dagger|\right)\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}\right\|_\infty$$

$$\leq \left(\max_{i\leq n}\|\widehat{\boldsymbol{f}}_{i,m}-\boldsymbol{f}^{\dagger}_{i,m}\|_2\right)\Big\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}\Big\|_\infty = O_P\left(\frac{1}{\sqrt{n}}+\frac{n^{1/4}}{\sqrt{p_m}}\right)O_P\left(\sqrt{\frac{\log p_{-m}}{n}}\right),$$

where the last equality follows from Lemma 1. This completes the proof. □

Lemma 8. *Suppose Conditions 1–3 hold, and* $\lambda_2 \asymp \sqrt{(\log p_{-m})/n}\{1\vee(n^{1/4}/\sqrt{p_m})\}$*. Then,*

$$\|\widehat{\boldsymbol{w}}_k-\boldsymbol{w}^*_k\|_1 = O_P\left(s^*_k\left\{\sqrt{\frac{\log p_{-m}}{n}}\left(1\vee\frac{n^{1/4}}{\sqrt{p_m}}\right)\right\}\right).$$

Proof. Recall that, for the kth column of $\boldsymbol{W}$, we solve

$$\widehat{\boldsymbol{w}}_k = \text{argmin}\|\boldsymbol{w}_k\|_1, \quad \text{such that} \quad \Big\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(\widehat{f}_{i,m_k}-\boldsymbol{x}'_{i,-m}\boldsymbol{w}_k)\Big\|_\infty \leq \lambda_2,$$

where $\widehat{f}_{i,m_k}$ is the kth element of $\widehat{\boldsymbol{f}}_{i,m}$. Let $S_k = \text{supp}(\boldsymbol{w}^*_k)$, where $\boldsymbol{w}^*_k$ is the kth column of $\boldsymbol{W}^* = \text{E}(\boldsymbol{x}^{\otimes 2}_{i,-m})^{-1}\text{E}(\boldsymbol{x}_{i,-m}\boldsymbol{f}'_{i,m})$. Then, we have $\|\boldsymbol{w}^*_{S_k}\|_1 \geq \|\widehat{\boldsymbol{w}}_{S_k}\|_1 + \|\widehat{\boldsymbol{w}}_{S^c_k}\|_1$. By the triangle inequality, we have $\|\widehat{\boldsymbol{w}}_{S^c_k}\|_1 \geq \|\boldsymbol{w}^*_{S_k}\|_1 - \|\widehat{\boldsymbol{w}}_{S_k}-\boldsymbol{w}^*_{S_k}\|_1$. Let $\widehat{\boldsymbol{\Delta}}_k = \widehat{\boldsymbol{w}}_k-\boldsymbol{w}^*_k$. Then, by noting that $\|\boldsymbol{w}^*_{S^c_k}\|_1 = 0$, we have $\|\widehat{\boldsymbol{\Delta}}_{S_k}\|_1 \geq \|\widehat{\boldsymbol{\Delta}}_{S^c_k}\|_1$. It follows from Lemma 7 that

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}(\widehat{f}_{i,m_k}-\boldsymbol{x}'_{i,-m}\widehat{\boldsymbol{w}}_k)\Big\|_\infty = O_P\left(\sqrt{\frac{\log p_{-m}}{n}}\left(1\vee\frac{n^{1/4}}{\sqrt{p_m}}\right)\right).$$

Denote $\boldsymbol{H}_x = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}^{\otimes 2}_{i,-m}$, $\boldsymbol{H}_{xf} = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_{i,-m}\widehat{f}_{i,m_k}$, and $\boldsymbol{\Delta}_k = \widehat{\boldsymbol{w}}_k-\boldsymbol{w}^*_k$. Then,

$$\Big\|\boldsymbol{H}_x\widehat{\boldsymbol{\Delta}}_k\Big\|_\infty \leq \Big\|\boldsymbol{H}_{xf}-\boldsymbol{H}_x\widehat{\boldsymbol{w}}_k\Big\|_\infty + \Big\|\boldsymbol{H}_{xf}-\boldsymbol{H}_x\boldsymbol{w}^*_k\Big\|_\infty = O_P\left(\sqrt{\frac{\log p_{-m}}{n}}\left(1\vee\frac{n^{1/4}}{\sqrt{p_m}}\right)\right).$$

Together with $\|\widehat{\boldsymbol{\Delta}}_k\|_1 \leq 2\|\widehat{\boldsymbol{\Delta}}_{S_k}\|_1 \leq 2\sqrt{s^*_k}\|\widehat{\boldsymbol{\Delta}}_k\|_2$, we have

$$\begin{aligned}\widehat{\boldsymbol{\Delta}}_k\boldsymbol{H}_x\widehat{\boldsymbol{\Delta}}_k &\leq \|\widehat{\boldsymbol{\Delta}}_k\|_1\Big\|\boldsymbol{H}_x\widehat{\boldsymbol{\Delta}}_k\Big\|_\infty = O_P\left(\|\widehat{\boldsymbol{\Delta}}_k\|_1\sqrt{\frac{\log p_{-m}}{n}}\left(1\vee\frac{n^{1/4}}{\sqrt{p_m}}\right)\right)\\ &= O_P\left(\|\widehat{\boldsymbol{\Delta}}_k\|_2\sqrt{\frac{s^*_k\log p_{-m}}{n}}\left(1\vee\frac{n^{1/4}}{\sqrt{p_m}}\right)\right). \qquad \text{(S32)}\end{aligned}$$

Note that Condition 1 and (1) imply that $\boldsymbol{x}_{-m}$ is sub-Gaussian. In addition, Condition 3 implies that $\lambda_{\min}(\text{E}(\boldsymbol{x}^{\otimes 2}_{i,-m})) > c$. Then, it follows from Proposition 1 of Raskutti et al. (2011) that the restricted eigenvalue condition holds for $\boldsymbol{H}_x$ with high probability, i.e., $\widehat{\boldsymbol{\Delta}}_k\boldsymbol{H}_x\widehat{\boldsymbol{\Delta}}_k \geq \kappa\|\widehat{\boldsymbol{\Delta}}_k\|^2_2$ for some $\kappa > 0$ and all $\widehat{\boldsymbol{\Delta}}_k$, such that $\|\widehat{\boldsymbol{\Delta}}_{S^c_k}\|_1 \leq \|\widehat{\boldsymbol{\Delta}}_{S_k}\|_1$. This result, together with (S32), implies that

$$\|\widehat{\boldsymbol{\Delta}}_k\|_1 \leq 2\sqrt{s^*_k}\|\widehat{\boldsymbol{\Delta}}_k\|_2 = O_P\left(s^*_k\sqrt{\frac{\log p_{-m}}{n}}\left(1\vee\frac{n^{1/4}}{\sqrt{p_m}}\right)\right).$$

This completes the proof. □

Lemma 9. *Suppose the conditions of Theorem 2 hold. Then, uniformly for all $\boldsymbol{\beta}^* \in \mathcal{N}$,*

$$\|\sqrt{n}(\widetilde{\boldsymbol{I}}^{-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}})\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0})\|_2 = o_P(1).$$

Proof. First, we note that

$$\begin{aligned}
&\|\widetilde{\boldsymbol{I}}^{-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\|_2 = \|\boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}(\boldsymbol{I}^{*1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \widetilde{\boldsymbol{I}}^{1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}})\widetilde{\boldsymbol{I}}^{-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\|_2 \\
\lesssim\ &\|\boldsymbol{I}^{*1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \widetilde{\boldsymbol{I}}^{1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\|_2 \le \|\boldsymbol{I}^{*}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\|_2^{1/2} \le \|\boldsymbol{I}^{*}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\|_1^{1/2} \\
\le\ &\sqrt{K_m}\|\boldsymbol{I}^{*}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \widetilde{\boldsymbol{I}}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\|_\infty^{1/2} = o_P(1),
\end{aligned}$$

where $\lesssim$ follows from Condition 6, and the last equality follows from (S7) and that K_m is fixed. Lemma 10 implies that $\|\sqrt{n}\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0})\|_2 = o_P(1)$. Therefore,

$$\|\sqrt{n}(\widetilde{\boldsymbol{I}}^{-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}})\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0})\|_2 \le \|\widetilde{\boldsymbol{I}}^{-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}} - \boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\|_2\|\sqrt{n}\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0})\|_2 = o_P(1).$$

This completes the proof. □

Lemma 10. *Suppose the conditions of Theorem 2 hold. Then, uniformly for all $\boldsymbol{\beta}^* \in \mathcal{N}$,*

$$\sqrt{n}\boldsymbol{I}^{*-1/2}_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\{\widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0}) - \boldsymbol{S}(\boldsymbol{\beta}^*_{-m}, \boldsymbol{0})\} = o_P(1).$$

Proof. By (S4), we have that

$$\begin{aligned}
\boldsymbol{S}(\boldsymbol{\beta}^*_{-m}, \boldsymbol{0}) - \widetilde{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}_{-m}, \boldsymbol{0}) = &\frac{1}{n\sigma_\epsilon^2}(\boldsymbol{W}^* - \widehat{\boldsymbol{W}})'\boldsymbol{X}'_{-m}(\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m}) \\
&+ \frac{1}{n\sigma_\epsilon^2}(\widehat{\boldsymbol{F}}'_m\boldsymbol{X}_{-m} - \widehat{\boldsymbol{W}}'\boldsymbol{X}'_{-m}\boldsymbol{X}_{-m})(\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m}) \equiv I + II.
\end{aligned}$$

For the term I, under H_a, $\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m} = \boldsymbol{X}_m\boldsymbol{\beta}^*_m + \boldsymbol{\epsilon}$. Therefore,

$$\begin{aligned}
&\|\frac{1}{n\sigma_\epsilon^2}\boldsymbol{X}'_{-m}(\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m})\|_\infty = \|\frac{1}{n\sigma_\epsilon^2}\boldsymbol{X}'_{-m}(\boldsymbol{X}_m\boldsymbol{\beta}^*_m + \boldsymbol{\epsilon})\|_\infty \\
\le\ &\|\frac{1}{n\sigma_\epsilon^2}\boldsymbol{X}'_{-m}\boldsymbol{\epsilon}\|_\infty + \|\frac{1}{n\sigma_\epsilon^2}\boldsymbol{X}'_{-m}(\boldsymbol{F}_m\boldsymbol{\gamma}^*_m + \boldsymbol{U}_m\boldsymbol{\beta}^*_m)\|_\infty = O_P\left(\sqrt{\frac{\log p_{-m}}{n}}\right),
\end{aligned}$$

where the last equation follows from the sub-Gaussian assumption in Condition 1, and an application of Bernstein inequality. A careful inspection of the proof of Lemma 8 shows that the lemma still holds under H_a. Therefore, following the same argument as in (S5), for each $k \in [K_m]$, we have

$$\begin{aligned}
|(n\sigma_\epsilon^2)^{-1}(\boldsymbol{w}^*_k - \widehat{\boldsymbol{w}}_k)'\boldsymbol{X}'_{-m}(\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m})| &\le \|\boldsymbol{w}^*_k - \widehat{\boldsymbol{w}}_k\|_1\|(n\sigma_\epsilon^2)^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{-m}\boldsymbol{\beta}^*_{-m})\|_\infty \\
&= o_P\left(1/\sqrt{n}\right).
\end{aligned} \tag{S33}$$

Therefore, $I = o_P(1/\sqrt{n})$.

For the term II, we bound $\|\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m}\|_1$ and $\|(n\sigma_\epsilon^2)^{-1}\widehat{\boldsymbol{F}}'_m\boldsymbol{X}_{-m} - \widehat{\boldsymbol{W}}'\boldsymbol{X}'_{-m}\boldsymbol{X}_{-m}\|_\infty$ respectively. To bound $\|\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m}\|_1$, under H_a, we only need to replace ϵ_i in (S28) with $\epsilon_i + \boldsymbol{U}_{i,m}\boldsymbol{\beta}^*_m$. Due to the sub-Gaussian assumption of $\boldsymbol{U}_{i,m}$ in Condition 1, and the fact that ϵ_i and $\boldsymbol{U}_{i,m}$ are uncorrelated, the bounds we have established in (S29) and (S30) still hold. Therefore, uniformly for all $\boldsymbol{\beta}^* \in \mathcal{N}$, $\|\widehat{\boldsymbol{\beta}}_{-m} - \boldsymbol{\beta}^*_{-m}\|_1$ has the same upper bound as the one established in Lemma 6. In addition, the bound we have established for $\|(n\sigma_\epsilon^2)^{-1}\widehat{\boldsymbol{F}}'_m\boldsymbol{X}_{-m} - \widehat{\boldsymbol{W}}'\boldsymbol{X}'_{-m}\boldsymbol{X}_{-m}\|_\infty$ also holds uniformly for all $\boldsymbol{\beta}^* \in \mathcal{N}$. Then, it follows from (S33) that $II = o_P(1/\sqrt{n})$.

Combining the bounds of the terms I, II, and Condition 6 completes the proof. □

Lemma 11. *Suppose the conditions of Theorem 2 hold. Then, uniformly for all* $\boldsymbol{\beta}^* \in \mathcal{N}$,

$$\boldsymbol{S}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*_m) - \boldsymbol{S}(\boldsymbol{\beta}^*_{-m}, \boldsymbol{0}) - \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m = o_P\left(n^{-1/2}\right).$$

Proof. By definition, we have

$$\begin{aligned}
&\sqrt{n}\{\boldsymbol{S}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*_m) - \boldsymbol{S}(\boldsymbol{\beta}^*_{-m}, \boldsymbol{0}) - \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m\} \\
=& \frac{1}{\sqrt{n}\sigma_\epsilon^2}\sum_{i=1}^n \boldsymbol{x}'_{i,m}\boldsymbol{\beta}^*_m(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m}) - \boldsymbol{I}^*_{\boldsymbol{\gamma}_m|\boldsymbol{\beta}_{-m}}\boldsymbol{\gamma}^*_m \\
=& \frac{1}{\sqrt{n}\sigma_\epsilon^2}\sum_{i=1}^n [(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m})\boldsymbol{f}'_{i,m}\boldsymbol{\gamma}^*_m - \mathrm{E}\{(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m})\boldsymbol{f}'_{i,m}\boldsymbol{\gamma}^*_m\}] \\
&+ \frac{1}{\sqrt{n}\sigma_\epsilon^2}\sum_{i=1}^n \boldsymbol{U}'_{i,m}\boldsymbol{\beta}^*_m(\boldsymbol{f}_{i,m} - \boldsymbol{W}^{*\prime}\boldsymbol{x}_{i,-m}) \equiv I + II.
\end{aligned}$$

For the term I, its kth element equals

$$\frac{1}{\sqrt{n}\sigma_\epsilon^2}\sum_{i=1}^n \left[\left(f_{i,m_k} - \boldsymbol{w}^{*\prime}_k\boldsymbol{x}_{i,-m}\right)\boldsymbol{f}'_{i,m}\boldsymbol{\gamma}^*_m - \mathrm{E}\left\{\left(\boldsymbol{f}_{i,m} - \boldsymbol{w}^{*\prime}_k\boldsymbol{x}_{i,-m}\right)\boldsymbol{f}'_{i,m}\boldsymbol{\gamma}^*_m\right\}\right].$$

By the sub-Gaussian assumptions on f_{i,m_k} and $\boldsymbol{w}^{*\prime}_k\boldsymbol{x}_{i,-m}$ in Condition 1, and the standard concentration inequality (e.g., Ning and Liu, 2017, Lemma H.2), the kth element of I is bounded by $O_P\left(\|\boldsymbol{c}_{m_n}\|_2\sqrt{\log n}\right) = o_P(1)$ for all $k \in [K_m]$, and this bound is uniform for all $\boldsymbol{\beta}^* \in \mathcal{N}$.

For the term II, its kth element satisfies that, uniformly for all $\boldsymbol{\beta}^* \in \mathcal{N}$,

$$\frac{1}{\sqrt{n}\sigma_\epsilon^2}\sum_{i=1}^n \boldsymbol{U}'_{i,m}\boldsymbol{\beta}^*_m(f_{i,m_k} - \boldsymbol{w}^{*\prime}_k\boldsymbol{x}_{i,-m}) = O_P\left(\|\boldsymbol{b}_{m_n}\|_2\sqrt{\log n}\right) = o_P(1).$$

Combining the results for the terms I and II completes the proof. □

Lemma 12. *Suppose the conditions of Theorem 3 hold. Then,*

$$\|\boldsymbol{R}\|_\infty = O_P\left(\sqrt{\frac{\log p}{n}}\left(\frac{1}{\sqrt{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right),$$

where $\boldsymbol{R} = n^{-1}\left\{(\widehat{\boldsymbol{U}} - \boldsymbol{U})'\boldsymbol{U}\boldsymbol{\beta}^* + \widehat{\boldsymbol{U}}'(\boldsymbol{F}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\gamma}_a) + (\widehat{\boldsymbol{U}} - \boldsymbol{U})'\boldsymbol{\epsilon}\right\}$.

Proof. First, we bound $n^{-1}\widehat{\boldsymbol{U}}'(\boldsymbol{F}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)$. By Lemma 2, $\|\widehat{\boldsymbol{U}} - \boldsymbol{U}\|_\infty = o_P(1)$. Therefore,

$$\begin{aligned}
\Big\|\frac{1}{n}\widehat{\boldsymbol{U}}'(\boldsymbol{F}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)\Big\|_\infty &\lesssim \Big\|\frac{1}{n}\boldsymbol{U}'(\boldsymbol{F}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)\Big\|_\infty \\
\le\ & \Big\|\frac{1}{n}\boldsymbol{U}'(\boldsymbol{F}\boldsymbol{\gamma}^* - \boldsymbol{F}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\gamma}^*)\Big\|_\infty + \Big\|\frac{1}{n}\boldsymbol{U}'(\boldsymbol{F}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)\Big\|_\infty \\
\le\ & \Big\|\frac{1}{n}\boldsymbol{U}'\boldsymbol{F}(\boldsymbol{I}_K - \boldsymbol{H}\boldsymbol{H}')\boldsymbol{\gamma}^*\Big\|_\infty + \Big\|\frac{1}{n}\boldsymbol{U}'(\boldsymbol{F}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)\Big\|_\infty. \qquad \text{(S34)}
\end{aligned}$$

Let $\widetilde{\boldsymbol{\gamma}} = (\boldsymbol{I}_K - \boldsymbol{H}\boldsymbol{H}')\boldsymbol{\gamma}^*$. Then, $\|\widetilde{\boldsymbol{\gamma}}\|_2 \le \|\boldsymbol{I}_K - \boldsymbol{H}\boldsymbol{H}'\|_2\|\boldsymbol{\gamma}^*\|_2$. By Lemma 2,

$$\|\boldsymbol{I}_K - \boldsymbol{H}\boldsymbol{H}'\|_2 = O_P\left(1/\sqrt{n} + 1/\sqrt{p_{\min}}\right).$$

Since $\|\boldsymbol{\gamma}^*\|_2^2 \lesssim \mathrm{Var}(\boldsymbol{f}'\boldsymbol{\gamma}^*) \lesssim \sigma_y^2 = O(1)$, we have $\|\widetilde{\boldsymbol{\gamma}}\|_\infty \le \|\widetilde{\boldsymbol{\gamma}}\|_2 = O_P\left(1/\sqrt{n} + 1/\sqrt{p_{\min}}\right)$. Then,

$$\Big\|\frac{1}{n}\boldsymbol{U}'\boldsymbol{F}(\boldsymbol{I}_K - \boldsymbol{H}\boldsymbol{H}')\boldsymbol{\gamma}^*\Big\|_\infty = \max_{j\in[p]}\Big|\frac{1}{n}\sum_{i=1}^n U_{ij}\sum_{k=1}^K f_{ik}\widetilde{\gamma}_k\Big|.$$

By Condition 1, K is fixed, and $\|\widetilde{\boldsymbol{\gamma}}\|_2 = O_P\left(1/\sqrt{n} + 1/\sqrt{p_{\min}}\right)$, we have that $\sum_{k=1}^K f_{ik}\widetilde{\gamma}_k$ is sub-Gaussian with variance bounded by $O(1/n + 1/p_{\min})$. Moreover, U_{ij} is also sub-Gaussian and uncorrelated with f_{ik} for any $k \in [K]$. Then, $U_{ij}\sum_{k=1}^K f_{ik}\widetilde{\gamma}_k$ is sub-exponential. Applying Bernstein inequality, we have

$$\Big\|\frac{1}{n}\boldsymbol{U}'\boldsymbol{F}(\boldsymbol{I}_K - \boldsymbol{H}\boldsymbol{H}')\boldsymbol{\gamma}^*\Big\|_\infty = O_P\left(\sqrt{\frac{\log p}{n}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p_{\min}}}\right)\right). \qquad \text{(S35)}$$

On the other hand, we have

$$\Big\|\frac{1}{n}\boldsymbol{U}'(\boldsymbol{F}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)\Big\|_\infty \le \Big\|\frac{1}{n}\boldsymbol{U}'(\boldsymbol{F}\boldsymbol{H} - \widehat{\boldsymbol{F}})\boldsymbol{H}'\boldsymbol{\gamma}^*\Big\|_\infty + \Big\|\frac{1}{n}\boldsymbol{U}'\widehat{\boldsymbol{F}}(\boldsymbol{H}'\boldsymbol{\gamma}^* - \widehat{\boldsymbol{\gamma}}_a)\Big\|_\infty. \quad \text{(S36)}$$

For the first term in (S36), we have

$$\begin{aligned}
\Big\|\frac{1}{n}\boldsymbol{U}'(\boldsymbol{F}\boldsymbol{H} - \widehat{\boldsymbol{F}})\boldsymbol{H}'\boldsymbol{\gamma}^*\Big\|_\infty &\le \left(\max_{i\in[n]}\sum_{k=1}^K |(\boldsymbol{F}\boldsymbol{H} - \widehat{\boldsymbol{F}})_{ik}(\boldsymbol{H}'\boldsymbol{\gamma}^*)_k|\right)\Big\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{U}_i\Big\|_\infty \\
\le\ & \|\boldsymbol{H}'\boldsymbol{\gamma}^*\|_2\left(\max_{i\in[n]}\|\widehat{\boldsymbol{f}}_i - \boldsymbol{H}\boldsymbol{f}_i\|_2\right)\Big\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{U}_i\Big\|_\infty \\
=\ & O_P\left(\sqrt{\frac{\log p}{n}}\right)O_P\left(\frac{1}{\sqrt{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}}\right), \qquad \text{(S37)}
\end{aligned}$$

where the last equality is implied by Lemma 1, and the fact that $\|\boldsymbol{H}'\boldsymbol{\gamma}^*\|_2 \le \|\boldsymbol{H}'\|_2\|\boldsymbol{\gamma}^*\|_2 = O(1)$. For the second term in (S36), we have

$$\Big\|\frac{1}{n}\boldsymbol{U}'\widehat{\boldsymbol{F}}(\boldsymbol{H}'\boldsymbol{\gamma}^* - \widehat{\boldsymbol{\gamma}}_a)\Big\|_\infty \lesssim \Big\|\frac{1}{n}\boldsymbol{U}'\boldsymbol{F}(\boldsymbol{H}'\boldsymbol{\gamma}^* - \widehat{\boldsymbol{\gamma}}_a)\Big\|_\infty \le \left(\max_{k\in[K]}|(\boldsymbol{H}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{\gamma}}_a)_k|\right)\sum_{k=1}^K\Big\|\frac{1}{n}\boldsymbol{U}_i f_{ik}\Big\|_\infty$$

Since $\widehat{\boldsymbol{\gamma}}_a \in \mathcal{M}$, we have $\max_{k\in[K]} |(\boldsymbol{H}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{\gamma}}_a)_k| = O_P(\delta_n)$. By Condition 1, for each $k \in [K]$, $U_{ij} f_{ik}$ is sub-exponential. Then $\|n^{-1}\boldsymbol{U}_i f_{ik}\|_\infty = O_P\left(\sqrt{(\log p)/n}\right)$. Since K is fixed, $\sum_{k=1}^K \|n^{-1}\boldsymbol{U}_i f_{ik}\|_\infty = O_P\left(\sqrt{(\log p)/n}\right)$. Then,

$$\left\|\frac{1}{n}\boldsymbol{U}'\widehat{\boldsymbol{F}}(\boldsymbol{H}'\boldsymbol{\gamma}^* - \widehat{\boldsymbol{\gamma}}_a)\right\|_\infty = O_P(\delta_n)\, O_P\left(\sqrt{\frac{\log p}{n}}\right) = O_P\left(\sqrt{\frac{\log p}{n}}\left(\frac{1}{\sqrt{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right).$$

Therefore,

$$\left\|\frac{1}{n}\boldsymbol{U}'(\boldsymbol{F}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)\right\|_\infty = O_P\left(\sqrt{\frac{\log p}{n}}\left(\frac{1}{\sqrt{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right). \tag{S38}$$

Then, it follows from (S34), (S35) and (S38) that

$$\left\|\frac{1}{n}\widehat{\boldsymbol{U}}'(\boldsymbol{F}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\gamma}}_a)\right\|_\infty = O_P\left(\sqrt{\frac{\log p}{n}}\left(\frac{1}{\sqrt{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right). \tag{S39}$$

For $\|n^{-1}(\widehat{\boldsymbol{U}} - \boldsymbol{U})'\boldsymbol{\epsilon}\|_\infty$, we have

$$\|n^{-1}(\widehat{\boldsymbol{U}} - \boldsymbol{U})'\boldsymbol{\epsilon}\|_\infty \le \left(\max_{ij} |\widehat{u}_{ij} - u_{ij}|\right)\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i\right| = O_P\left(\frac{1}{\sqrt{n}}\left(\sqrt{\frac{\log n}{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right). \tag{S40}$$

where the last equality follows from Lemma 2, and $n^{-1}\sum_{i=1}^n \epsilon_i = O_P(1/\sqrt{n})$, since ϵ_i is sub-Gaussian. Similarly, we have

$$\|n^{-1}(\widehat{\boldsymbol{U}} - \boldsymbol{U})'\boldsymbol{U}\boldsymbol{\beta}\|_\infty = O_P\left(\frac{1}{\sqrt{n}}\left(\sqrt{\frac{\log n}{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right). \tag{S41}$$

Combining (S39), (S40) and (S41) completes the proof. □

Lemma 13. *Suppose the conditions of Theorem 3 hold. Then,*

$$\left\|\frac{1}{n}\widehat{\boldsymbol{F}}'(\boldsymbol{Y} - \widehat{\boldsymbol{U}}\widehat{\boldsymbol{\beta}}_a - \widehat{\boldsymbol{F}}\boldsymbol{H}'\boldsymbol{\gamma}^*)\right\|_\infty = O_P(\delta_n).$$

Proof. The proof is similar to Lemma 12. We outline the key steps here. We have

$$\begin{aligned}
&\left\|\frac{1}{n}\widehat{\boldsymbol{F}}'(\boldsymbol{Y} - \widehat{\boldsymbol{U}}\widehat{\boldsymbol{\beta}}_a - \widehat{\boldsymbol{F}}\boldsymbol{H}'\boldsymbol{\gamma}^*)\right\|_\infty \lesssim \left\|\frac{1}{n}\boldsymbol{F}'\{\boldsymbol{F}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{F}}\boldsymbol{H}'\boldsymbol{\gamma}^* + \boldsymbol{U}\boldsymbol{\beta}^* - \widehat{\boldsymbol{U}}\widehat{\boldsymbol{\beta}}_a + \boldsymbol{\epsilon}\}\right\|_\infty \\
\le\ & \left\|\frac{1}{n}\boldsymbol{F}'\{\boldsymbol{F}\boldsymbol{\gamma}^* - \boldsymbol{F}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\gamma}^*\}\right\|_\infty + \left\|\frac{1}{n}\boldsymbol{F}'(\boldsymbol{F}\boldsymbol{H} - \widehat{\boldsymbol{F}})\boldsymbol{H}'\boldsymbol{\gamma}^*\right\|_\infty \\
&+ \left\|\frac{1}{n}\boldsymbol{F}'\boldsymbol{U}\boldsymbol{\beta}^*\right\|_\infty + \left\|\frac{1}{n}\boldsymbol{F}'\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\beta}}_a\right\|_\infty + \left\|\frac{1}{n}\boldsymbol{F}'\boldsymbol{\epsilon}\right\|_\infty.
\end{aligned}$$

Similar to (S35), we have

$$\left\|\frac{1}{n}\boldsymbol{F}'\{\boldsymbol{F}\boldsymbol{\gamma}^* - \boldsymbol{F}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{\gamma}^*\}\right\|_\infty = \left\|\frac{1}{n}\boldsymbol{F}'\boldsymbol{F}(\boldsymbol{I}_K - \boldsymbol{H}\boldsymbol{H}')\boldsymbol{\gamma}^*\right\|_\infty = O_P\left(\sqrt{\frac{\log K}{n}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p_{\min}}}\right)\right).$$

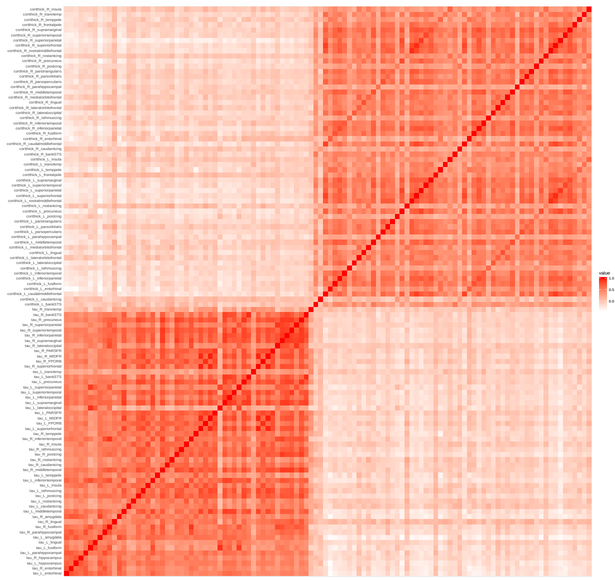

Figure S1: The heat map of the correlation matrix of the multimodal neuroimaging data.

Similar to (S37), we have

$$\left\|\frac{1}{n}\boldsymbol{F}'(\boldsymbol{F}\boldsymbol{H}-\widehat{\boldsymbol{F}})\boldsymbol{H}'\boldsymbol{\gamma}^*\right\|_\infty = O_P\left(\sqrt{\frac{\log K}{n}}\left(\frac{1}{\sqrt{n}}+\frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right).$$

Note that $\|\boldsymbol{\beta}^*\|_2^2 \lesssim \mathrm{Var}(\boldsymbol{x}'\boldsymbol{\beta}^*) \leq \sigma_y^2 = O(1)$, and $\lambda_{\max}(\boldsymbol{\Sigma}_u) = O(1)$. Therefore, $\boldsymbol{U}_i'\boldsymbol{\beta}^*$ is sub-Gaussian with bounded variance. Since f_{ik} is sub-Gaussian, $f_{ik}\boldsymbol{U}_i'\boldsymbol{\beta}^*$ is exponential. Then by Bernstein inequality, we have that $\|n^{-1}\boldsymbol{F}'\boldsymbol{U}\boldsymbol{\beta}^*\|_\infty = O_P\left(\sqrt{(\log K)/n}\right)$. Similarly, $\|n^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{U}}\boldsymbol{\beta}^*\|_\infty = O_P\left(\sqrt{(\log K)/n}\right)$. Finally, $f_{ik}\epsilon_i$ is sub-exponential, then $\|n^{-1}\boldsymbol{F}'\boldsymbol{\epsilon}\|_\infty = O_P\left(\sqrt{(\log K)/n}\right)$. Combining these results completes the proof. □

S9 Additional numerical results

Figure S1 shows the heat map of the correlation matrix of the multimodal neuroimaging data analyzed in Section 7.3. It is seen that some covariates are highly correlated.

References

Bentkus, V. (2005). A Lyapunov-type bound in $\mathcal{R}^d$. *Theory of Probability & Its Applications* **49,** 311–323.

Fan, J., Ke, Y., and Wang, K. (2016). Factor-adjusted regularized model selection. *arXiv:1612.08490* .

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

Li, Q., Cheng, G., Fan, J., and Wang, Y. (2018). Embracing the blessing of dimensionality in factor models. *Journal of the American Statistical Association* **113,** 380–389.

Negahban, S. N., Ravikumar, P., Wainwright, M. J., and Yu, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statistical Science* **27,** 538–557.

Ning, Y. and Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics* **45,** 158–195.

Raskutti, G., Wainwright, M. J., and Yu, B. (2011). Minimax rates of estimation for high-dimensional linear regression over L_q-balls. *IEEE Transactions on Information Theory* **57,** 6976–6994.

Shi, C., Song, R., Chen, Z., and Li, R. (2019). Linear hypothesis testing for high dimensional generalized linear models. *The Annals of Statistics, to appear.*