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 \( \sigma^2 \) is known, then prove it when \( \sigma^2 \) is unknown. Let

\[
\tilde{I}_{\gamma_m|\beta_m} = \sigma^2 \left\{ \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_{i,m}^2 - \tilde{W}' \left( \frac{1}{n} \sum_{i=1}^{n} x_{i,-m} \tilde{f}_{i,m} \right) \right\}
\]

\[
\tilde{S}(\tilde{\beta}_m, 0) = \frac{1}{n \sigma^2} \sum_{i=1}^{n} (y_i - x_{i,-m} \tilde{\beta}_m) \left( \tilde{f}_{i,m} - \tilde{W}' x_{i,-m} \right).
\]

We first show that, under \( H_0 \), when \( \sigma^2 \) is known,

\[
\tilde{T}_n = \sqrt{n} \tilde{I}_{\gamma_m|\beta_m}^{-1/2} \tilde{S}(\tilde{\beta}_m, 0) \xrightarrow{D} N(0, I_{K_m}). \tag{S1}
\]

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

\[
\sqrt{n} \tilde{I}_{\gamma_m|\beta_m}^{-1/2} \tilde{S}(\tilde{\beta}_m, 0) \xrightarrow{D} N(0, I_{K_m}), \tag{S2}
\]

\[
\tilde{I}_{\gamma_m|\beta_m} \xrightarrow{P} I_{\gamma_m|\beta_m}^*. \tag{S3}
\]

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

To prove (S2), define

\[
\ell(\beta^*_m, \gamma^*_m) = \frac{1}{2n \sigma^2} \| Y - X_{-m} \beta^*_m - \tilde{F}_m \gamma^*_m \|_2^2.
\]

Then, \( \nabla_{\beta_m} \ell(\beta^*_m, \gamma^*_m) = -(n \sigma^2)^{-1} (Y - X_{-m} \beta^*_m - \tilde{F}_m \gamma^*_m)' X_{-m}, \nabla_{\gamma_m} \ell(\beta^*_m, \gamma^*_m) = (n \sigma^2)^{-1} \tilde{F}_m' X_{-m}, \) and \( \nabla_{\beta_m} \ell(\beta^*_m, \gamma^*_m) = (n \sigma^2)^{-1} X_{-m}' X_{-m}. \) Let

\[
S(\beta^*_m, 0) = \frac{1}{n \sigma^2} \sum_{i=1}^{n} (y_i - x_{i,-m} \beta^*_m) (\tilde{f}_{i,m} - W^* x_{i,-m}).
\]
Noting that $W^* = E(x_{i,-m}^{\otimes 2})^{-1} E(x_{i,-m}f_{i,m}')$, we have
\begin{equation}
\tilde{S}(\hat{\beta}_m, 0) = S(\beta_m, 0) - \frac{1}{n\sigma^2}(W^* - \hat{W})'X'_m(Y - X_m\beta_m^*) - \frac{1}{n\sigma^2}(\hat{F}_m X_m - \hat{W}'X'_mX_m)(\hat{\beta}_m - \beta_m^*),
\end{equation}

(S4)

For the second summand on the right-hand-side of (S4), by Lemma 8, for each $k \in [K_m]$,\begin{align*}
\left| \frac{1}{n\sigma^2}(w_k - \hat{w}_k)'X'_m(Y - X_m\beta_m^*) \right| &\leq \|w_k - \hat{w}_k\|_1 \left| \frac{1}{n\sigma^2}X'_m(Y - X_m\beta_m^*) \right|_\infty = \|w_k - \hat{w}_k\|_1 \left| \frac{1}{n\sigma^2}X'_m\epsilon \right|_\infty
\end{align*}
(S5)
\[ = O_P \left( s_k^* \left( \sqrt{(\log p_m)/n} \left\{ 1 \vee (n^{1/4}/\sqrt{p_m}) \right\} \right) \right) O_P \left( \sqrt{(\log p_m)/n} \right) = o_P \left( n^{-1/2} \right),\]

where $s_k^* = |\text{supp}(w_k)|$. For the third summand on the right-hand-side of (S4), let $(\hat{F}_m)_k$ denote the $k$th column of $\hat{F}_m$. By Lemmas 6 and 7, for each $k \in [K_m]$, \begin{align*}
\left| \frac{1}{n\sigma^2}\{(\hat{F}_m)_k X_m - \hat{w}_k'X'_mX_m\}(\hat{\beta}_m - \beta_m^*) \right| &\leq \left| \frac{1}{n\sigma^2}\{(\hat{F}_m)_k X_m - \hat{w}_k'X'_mX_m\} \right| \left\| \hat{\beta}_m - \beta_m^* \right\|_1
\end{align*}
(S6)
\[ = O_P \left( s^*_m \left( \sqrt{(\log p_m)/n} + 1/\sqrt{p_m} \right) \right) O_P \left( \sqrt{(\log p_m)/n} \left( 1 \vee n^{1/4}/\sqrt{p_m} \right) \right) = o_P \left( n^{-1/2} \right).
\]

Together with the central limit theorem of $\sqrt{n}I_{\gamma_m|\beta_m}^{-1/2}S(\beta_m^*, 0)$, we prove (S2).

To prove (S3), since the dimension $K_m$ is fixed, all matrix norms are equivalent. In particular, we show that
\begin{equation}
\|\tilde{I}_{\gamma_m|\beta_m} - I_{\gamma_m|\beta_m}\|_\infty = o_P (1).
\end{equation}
(S7)

Let $(\tilde{I}_{\gamma_m|\beta_m} - I_{\gamma_m|\beta_m})_k$ denote the $k$th row of $\tilde{I}_{\gamma_m|\beta_m} - I_{\gamma_m|\beta_m}$. By the definition of the information matrix $I_{\gamma_m|\beta_m}$, the identifiability assumption that $E(f_{i,m}^{\otimes 2}) = I_{K_m}$, and furthermore the fact that $(1/n)\sum_{i=1}^n \hat{f}_{i,m}^{\otimes 2} = I_{K_m}$, we have
\begin{align*}
\|\tilde{I}_{\gamma_m|\beta_m} - I_{\gamma_m|\beta_m}\|_\infty &\leq \sigma^{-2} \left\| \hat{w}' \left( \frac{1}{n} \sum_{i=1}^n x_{i,-m} \hat{f}_{i,m}' \right) - w_k' E(x_{i,-m}f_{i,m}') \right\|_\infty
\end{align*}
\[ \lesssim \|\hat{w} - w_k^*\|_\infty \left( \frac{1}{n} \sum_{i=1}^n x_{i,-m} \hat{f}_{i,m}' \right) + \left\| w_k^* \left( \frac{1}{n} \sum_{i=1}^n x_{i,-m} \hat{f}_{i,m}' - E(x_{i,-m}f_{i,m}') \right) \right\|_\infty.
\]

For the first term, let $\hat{f}_{i,mh}$ be the $h$th element of $\hat{f}_{i,m}$, we have
\[ \left\| \hat{w} - w_k^* \right\|_\infty \left( \frac{1}{n} \sum_{i=1}^n x_{i,-m} \hat{f}_{i,m}' \right) = \max_{h \in [K_m]} \left\| \hat{w} - w_k^* \right\|_\infty \left( \frac{1}{n} \sum_{i=1}^n x_{i,-m} \hat{f}_{i,mh} \right) \].
where $s_k = |\text{supp}(w_k^*)|$, and the second-to-last equality follows from Lemma 8, and the fact that the dominating term in the bracket is $\|E(x_{i,-m}x_{i,-m}^*w)_h^*\|_{\infty} = O(1)$. For the second term,

$$\|w_k^*\left\{ \frac{1}{n}\sum_{i=1}^{n} x_{i,-m} f_{i,m} - E(x_{i,-m} f_{i,m}) \right\} \|_{\infty} = \max_{h \in [K_m]}\|w_k^*\left\{ \frac{1}{n}\sum_{i=1}^{n} x_{i,-m} f_{i,m} - E(x_{i,-m} f_{i,m}) \right\} \|_{\infty} \leq \|w_k^*\|_1 \cdot \max_{h \in [K_m]}\left[\|\frac{1}{n}\sum_{i=1}^{n} x_{i,-m} f_{i,m} - x_{i,-m}^*w_h^* \|_{\infty} + \left\| \frac{1}{n}\sum_{i=1}^{n} \{x_{i,-m}x_{i,-m}^*w_h^* - E(x_{i,-m}x_{i,-m}^*w_h^*) \} \right\|_{\infty}\right]$$

$$= O_P \left(s^*_k \sqrt{\frac{\log p_m}{n}} \left(1 + \frac{n^{1/4}}{\sqrt{p_m}}\right)\right) = O_P(1)
$$

where the second-to-last equality follows from (S31), and the sub-Gaussian assumption on $X_{ij}$ and $x_{i,-m}^*w_h^*$, which is implied by (1) and Condition 1.

Next, when $\sigma^2_\varepsilon$ is unknown, we have $\tilde{T}_n - \tilde{T}_n = \tilde{T}_n(\sigma_\varepsilon/\hat{\sigma}_\varepsilon - 1) = o_P(1)$, 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 $T^*_n = \sqrt{n}I^{1/2}_{\tau_m|\beta_{-m}} \left\{ S(\beta^*, \gamma^*_{\beta_{-m}} - I^{*}_{\gamma_m|\beta_{-m}} \gamma^*_{\beta_{-m}}) \right\}$, and $Q^*_n = (T^*_n)^T T^*_n$, we show that $Q_n = Q^*_n + o_P(1)$. First, recall $\tilde{T}_n$ as defined in (S1), we have

$$T_n = (\sigma_\varepsilon \hat{\sigma}_\varepsilon^{-1})\tilde{T}_n = \tilde{T}_n + \frac{\sigma_\varepsilon - \hat{\sigma}_\varepsilon}{\sigma_\varepsilon} \tilde{T}_n = \tilde{T}_n + o_P(1),$$

(S8) where the last equality follows from Condition 5, and the $o_P$ statement applies to each element of $T_n$. Next, we show that $\tilde{T}_n = T^*_n + o_P(1)$. Letting $\tilde{T}^*_n = \sqrt{n}I^{1/2}_{\tau_m|\beta_{-m}} S(\beta^*, 0)$, we have that

$$\tilde{T}_n = T^*_n + \sqrt{n}I^{*1/2}_{\tau_m|\beta_{-m}} \left\{ \tilde{S}(\beta_{-m}, 0) - S(\beta_{-m}, 0) \right\} + \sqrt{n}(I^{1/2}_{\gamma_m|\beta_{-m}} - I^{1/2}_{\gamma_m|\beta_{-m}})\tilde{S}(\beta_{-m}, 0).$$

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

$$\sqrt{n}I^{*1/2}_{\tau_m|\beta_{-m}} \left\{ \tilde{S}(\beta_{-m}, 0) - S(\beta_{-m}, 0) \right\} = o_P(1).$$
\[ \sqrt{n} \{ \tilde{I}^{-1/2}_{\gamma_m|\beta_m} - I^{-1/2}_{\gamma_m|\beta_m} \} \tilde{S}(\tilde{\beta}_m, 0) = o_P(1). \]

Therefore, \( \tilde{T}_n = T_n^1 + o_P(1). \) Recall that \( S(\beta^*, \gamma_m^*) = (n\sigma^2_n)^{-1} \sum_{i=1}^n \varepsilon_i (f_{i,m} - W^{*'} x_{i,-m}). \) Henceforth, we have

\[
T_n^1 = \sqrt{n} I^{-1/2}_{\gamma_m|\beta_m} \{ S(\beta^*, \gamma_m^*) - I^*_{\gamma_m|\beta_m} \gamma_m^* \}
\]

\[ + \sqrt{n} I^{-1/2}_{\gamma_m|\beta_m} \{ S(\beta^*, \gamma_m^*) - S(\beta^*, \gamma_m^*) + I^*_{\gamma_m|\beta_m} \gamma_m^* \}
\]

\[ = \sqrt{n} I^{-1/2}_{\gamma_m|\beta_m} \{ S(\beta^*, \gamma_m^*) - I^*_{\gamma_m|\beta_m} \gamma_m^* \} + o_P(1) = T_n^* + o_P(1), \]

where the second equality follows from Lemma 11. Therefore, we have \( \tilde{T}_n = T_n^* + o_P(1). \) 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 \( \chi^2 \) approximation of \( Q_n^* \). By definition,

\[
\sqrt{n} I^{-1/2}_{\gamma_m|\beta_m} \{ S(\beta^*, \gamma_m^*) - I^*_{\gamma_m|\beta_m} \gamma_m^* \}
\]

\[ = \sqrt{n} I^{-1/2}_{\gamma_m|\beta_m} \left\{ \frac{1}{n\sigma^2_n} \sum_{i=1}^n \varepsilon_i (f_{i,m} - W^{*'} x_{i,-m}) \right\} - \sqrt{n} I^{-1/2}_{\gamma_m|\beta_m} \gamma_m^*
\]

\[ = \sum_{i=1}^n \xi_i - \sqrt{n} I^{-1/2}_{\gamma_m|\beta_m} \gamma_m^*, \]

where \( \xi_i = (n\sigma^2_n)^{-1} I^{-1/2}_{\gamma_m|\beta_m} \varepsilon_i (f_{i,m} - W^{*'} x_{i,-m}). \) By direct calculation, we have that \( E(\xi_i) = 0, \) and \( \sum_{i=1}^n \text{Var}(\xi_i) = I_{K_m}. \) By Conditions 1 and 6, we have

\[ \sum_{i=1}^n E\|\xi_i\|^2_2 = \frac{1}{(n\sigma^4_n)^{3/2}} E|\varepsilon_i|^{3/2} \sum_{i=1}^n E\|I^{-1/2}_{\gamma_m|\beta_m} (f_{i,m} - W^{*'} x_{i,-m})\|^3_2
\]

\[ \lesssim \frac{1}{\sqrt{n}} (E\|f_{i,m}\|_2^3 + E\|W^{*'} x_{i,-m}\|_2^3) = o(1). \]

Then, by Lemma 4, we have that

\[ \sup_C \left| \Pr \left( \sum_{i=1}^n \xi_i \in C \right) - \Pr(\mathbf{Z} \in C) \right| \to 0, \quad (S9) \]

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

\[ \sup_x |\Pr(Q_n^* \leq x) - \Pr(\chi^2(1, h_n) \leq x)| = \sup_x \left| \Pr \left( \sum_{i=1}^n \xi_i \in C_x \right) - \Pr(\mathbf{Z} \in C_x) \right| \to 0, \]

where \( h_n = n \gamma_m^{*'} I^{-1/2}_{\gamma_m|\beta_m} \gamma_m^*. \) Since \( Q_n = Q_n^* + o_P(1). \) For any \( x \) and \( \varepsilon > 0, \) we have

\[ \Pr(\chi^2(1, h_n) \leq x - \varepsilon) + o(1) \leq \Pr(Q_n \leq x - \varepsilon) + o(1) \leq \Pr(Q_n \leq x) \leq \Pr(Q_n^* \leq x + \varepsilon) + o(1) \leq \Pr \left\{ \chi^2(1, h_n) \leq x + \varepsilon \right\} + o(1). \quad (S10) \]
In addition, by Lemma 5, we have
\[
\lim_{n \to \infty} \lim_{\epsilon \to 0} \sup_{n} \left| \Pr\{\chi^2(1, h_n) \leq x + \epsilon\} - \Pr\{\chi^2(1, h_n) \leq x - \epsilon\} \right| \to 0.
\]
Together, we have
\[
\sup_{x} \left| \Pr(Q_n \leq x) - \Pr(\chi^2(1, h_n) \leq x) \right| \to 0.
\]
This completes the proof. \(\square\)

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
\[
\left| \Pr(Q_n \leq x) - \Pr(\chi^2(K_m, h) \leq x) \right| \leq \left| \Pr(Q_n \leq x) - \Pr(\chi^2(K_m, h_m) \leq x) \right| + \left| \Pr(\chi^2(K_m, h_m) \leq x) - \Pr(\chi^2(K_m, h) \leq x) \right|
\]
Then, by Theorem 2, it suffices to prove that
\[
\lim_{n \to \infty} \sup_{x > 0} \left| \Pr(\chi^2(K_m, h_m) \leq x) - \Pr(\chi^2(K_m, h) \leq x) \right| = 0.
\]
Let \(F(x; k, \lambda)\) denote the cumulative distribution function of the non-central \(\chi^2\)-distribution with \(k\) degrees of freedom and non-centrality parameter \(\lambda\), and \(F(x; k)\) that of the central \(\chi^2\)-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
\[
\Pr(\chi^2(K_m, h_m) \leq x) - \Pr(\chi^2(K_m, h) \leq x) \\
= e^{-h_m/2} \left\{ F(x; K_m) + \sum_{j=1}^{\infty} \frac{(h_m/2)^j}{j!} F(x; K_m + 2j) \right\} \\
- e^{-h/2} \left\{ F(x; K_m) + \sum_{j=1}^{\infty} \frac{(h/2)^j}{j!} F(x; K_m + 2j) \right\}
\]
Then, as \(n \to \infty\),
\[
\sup_{x > 0} \left| \Pr(\chi^2(K_m, h_m) \leq x) - \Pr(\chi^2(K_m, h) \leq x) \right| \\
\leq \left| e^{-h_m/2} - e^{-h/2} \right| \sup_{x > 0} \left\{ F(x; K_m) + \sum_{j=1}^{\infty} \frac{(h_m/2)^j}{j!} F(x; K_m + 2j) \right\} \\
+ e^{-h/2} \sup_{x > 0} \left\{ \sum_{j=1}^{\infty} \frac{|(h_m/2)^j - (h/2)^j|}{j!} F(x; K_m + 2j) \right\} \\
\leq |e^{-h_m/2} - e^{-h/2}| e^{h_m/2} + e^{-h/2} \sum_{j=1}^{\infty} \frac{|(h_m/2)^j - (h/2)^j|}{j!} \to 0.
\]
This completes the proof. \(\square\)
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, \( \hat{F} \) is consistent to \( HF \) for some nonsingular matrix \( H \). Moreover, \( \hat{U} \) is consistent to \( U \). Then solving (10) is equivalent to solving the same problem by replacing \((\hat{F}, \hat{U})\) with \((F, U)\). More precisely, by the Karush-Kuhn-Tucker conditions, any vector \((\hat{\gamma}_a, \hat{\beta}_a)\) satisfying the following equations is a solution to (10):

\[
\begin{align*}
\frac{1}{n} \hat{F}'(Y - \hat{F}\hat{\gamma}_a - \hat{U}_{T\cup S_a}\hat{\beta}_{a,T\cup S_a}) &= 0; \\
\frac{1}{n} \hat{U}'_T(Y - \hat{F}\hat{\gamma}_a - \hat{U}_{T\cup S_a}\hat{\beta}_{a,T\cup S_a}) &= 0; \\
\frac{1}{n} \hat{U}'_{S_a}(Y - \hat{F}\hat{\gamma}_a - \hat{U}_{T\cup S_a}\hat{\beta}_{a,T\cup S_a}) &= \lambda_a \hat{p}(|\hat{\beta}_{a,S_a}|) I(\hat{\beta}_{a,S_a} > 0); \\
\left\| \frac{1}{n} \hat{U}'_{(T\cup S)a}(Y - \hat{F}\hat{\gamma}_a - \hat{U}_{T\cup S_a}\hat{\beta}_{a,T\cup S_a}) \right\|_\infty &= \lambda_a \hat{p}(0+).
\end{align*}
\]

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

We divide the proof into two main steps. In Step 1, letting \( M = \{ (\gamma, \beta) : \|\gamma - H^T \gamma \|_\infty \leq C \delta_n, \|\beta - \beta_{T\cup S_a}\|_\infty \leq C \delta_n \} \) for some constant \( C \), we show that, with probability tending to 1, there exists a vector \((\hat{\gamma}_a, \hat{\beta}_{a,T\cup S_a})\) in \( M \) that satisfies (S11), (S12) and (S13). In Step 2, we set \( \hat{\beta}_a = (\hat{\beta}_{a,T\cup S_a}, 0) \), and show that \((\hat{\gamma}_a, \hat{\beta}_a)\) satisfies (S14). Together these two steps prove (a) and (b) of the theorem. Then (c) follows from \( \|\hat{\beta}_{a,T\cup S_a} - \beta_{T\cup S_a}\|_2 \leq \sqrt{\epsilon + s_a} \|\hat{\beta}_{a,T\cup S_a} - \beta^*_{T\cup S_a}\|_\infty \).

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

\[
Y - \hat{F}\hat{\gamma}_a - \hat{U}_{T\cup S_a}\hat{\beta}_{a,T\cup S_a} = U_{T\cup S_a}(\beta^*_{T\cup S_a} - \hat{\beta}_{a,T\cup S_a}) + \epsilon + \zeta.
\]

Therefore,

\[
\begin{align*}
\frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} - \frac{1}{n} \hat{U}'_{S_a} \right) \left( Y - \hat{F}\hat{\gamma}_a - \hat{U}_{T\cup S_a}\hat{\beta}_{a,T\cup S_a} \right) &= \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} U_T - \frac{1}{n} \hat{U}'_{S_a} U_{T\cup S_a} \right) \left( \beta^*_{T} - \hat{\beta}_{a,T}\right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{S_a} \epsilon + \hat{U}'_{S_a} (\epsilon + \zeta) \right) \\
&= \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} U_T - \frac{1}{n} \hat{U}'_{S_a} U_{T\cup S_a} \right) \left( \beta^*_{S_a} - \hat{\beta}_{a,S_a}\right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} \epsilon + \hat{U}'_{S_a} \epsilon \right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} \epsilon + \hat{U}'_{S_a} \epsilon \right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{S_a} \epsilon + \hat{U}'_{S_a} \epsilon \right) \\
&= \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} U_T - \frac{1}{n} \hat{U}'_{S_a} U_{T\cup S_a} \right) \left( \beta^*_{T} - \hat{\beta}_{a,T}\right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} \epsilon + \hat{U}'_{S_a} \epsilon \right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} \epsilon + \hat{U}'_{S_a} \epsilon \right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{S_a} \epsilon + \hat{U}'_{S_a} \epsilon \right) \\
&= \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} U_T - \frac{1}{n} \hat{U}'_{S_a} U_{T\cup S_a} \right) \left( \beta^*_{S_a} - \hat{\beta}_{a,S_a}\right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} \epsilon + \hat{U}'_{S_a} \epsilon \right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{T} \epsilon + \hat{U}'_{S_a} \epsilon \right) + \frac{1}{n} \left( \frac{1}{n} \hat{U}'_{S_a} \epsilon + \hat{U}'_{S_a} \epsilon \right)
\end{align*}
\]

(15)

where

\[
R_{T\cup S_a} = \frac{1}{n} \left\{ (\hat{U}_{T\cup S_a} - U_{T\cup S_a})' U_{T\cup S_a} \beta^*_{T\cup S_a} + \hat{U}'_{T\cup S_a} (F\gamma - \hat{F}\hat{\gamma}_a) + (\hat{U}_{T\cup S_a} - U_{T\cup S_a})' \epsilon \right\}.
\]
By Lemma 12, we have
\[
\|R_{T \cup S_a}\|_\infty = O_P\left(\sqrt{\frac{\log p}{n}} \left(\frac{1}{\sqrt{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}}\right)\right).
\]

In addition, by Condition 1, $U_{ij}\epsilon_i$ is sub-exponential. Therefore, by Bernstein inequality and the union bound, we have
\[
\|n^{-1}U'_{T \cup S_a}\epsilon\|_\infty = O_P\left(\sqrt{(\log p)/n}\right).
\]
Together,
\[
\left\|\frac{1}{n}U'_{T \cup S_a}\epsilon\right\|_\infty + \|R_{T \cup S_a}\|_\infty = O_P\left(\delta_n\right), \text{ where } \delta_n = \sqrt{(\log p)/n}\{1 \lor (n^{1/4}/\sqrt{p_{\min}})\}. \quad (S16)
\]

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

For Step 2, let $\varphi_{T \cup S_a} = (0, \lambda_a\hat{p}(\|\beta_{a,S_a}\|)I(\beta_{a,S_a} > 0))$, we have
\[
in^{-1}\hat{U}'_{(T \cup S_a)\epsilon}(Y - \hat{F}\gamma_a - \hat{U}_{T \cup S_a}\beta_{a,T \cup S_a}) = n^{-1}\hat{U}'_{(T \cup S_a)\epsilon}U_{T \cup S_a}(\beta^*_{T \cup S_a} - \beta_{a,T \cup S_a}) + n^{-1}\hat{U}'_{(T \cup S_a)\epsilon}\epsilon + n^{-1}\hat{U}'_{(T \cup S_a)\epsilon}\zeta = n^{-1}\hat{U}'_{(T \cup S_a)\epsilon}U_{T \cup S_a}K_n^{-1}\{\varphi_{T \cup S_a} - n^{-1}U'_{T \cup S_a}\epsilon - R_{T \cup S_a}\}
\]
\[+ n^{-1}\hat{U}'_{(T \cup S_a)\epsilon}\epsilon + n^{-1}\hat{U}'_{(T \cup S_a)\epsilon}(F\gamma - \hat{F}\gamma_a) + n^{-1}\hat{U}_{(T \cup S_a)\epsilon}(U_{T \cup S_a} - \hat{U}_{T \cup S_a})\beta^*_{T \cup S_a}.\]

By Condition 9 and the sub-Gaussian assumption on $U$, with probability tending to 1, we have $\|n^{-1}U'_{T \cup S_a}U_{T \cup S_a}K_n^{-1}\|_\infty = O(1)$. Therefore, by (S16), we have
\[
\lambda_a^{-1}\|n^{-1}U'_{(T \cup S_a)\epsilon}U_{T \cup S_a}K_n^{-1}\{n^{-1}U'_{T \cup S_a}\epsilon + R_{T \cup S_a}\}\|_\infty \
\leq \lambda_a^{-1}\|n^{-1}U'_{(T \cup S_a)\epsilon}U_{T \cup S_a}K_n^{-1}\|_\infty\|n^{-1}U'_{T \cup S_a}\epsilon + R_{T \cup S_a}\|_\infty \
= O_P\left(\delta_n\lambda_a^{-1}\right) = o_P(1).
\]

Next,
\[
\lambda_a^{-1}\|n^{-1}U'_{(T \cup S_a)\epsilon}U_{T \cup S_a}K_n^{-1}\lambda_a\hat{p}(\|\beta_{a,T \cup S_a}\|)\|_\infty \lesssim \hat{p}(\|\beta_{a,T \cup S_a}\|) < \hat{p}(d_n) < \hat{p}(0+).
\]

Moreover, by Lemma 12,
\[
\lambda_a^{-1}n^{-1}\|\hat{U}'_{(T \cup S_a)\epsilon}\epsilon + \hat{U}'_{(T \cup S_a)\epsilon}(F\gamma - \hat{F}\gamma_a) + \hat{U}_{(T \cup S_a)\epsilon}(U_{T \cup S_a} - \hat{U}_{T \cup S_a})\beta^*_{T \cup S_a}\|_\infty = O_P\left(\delta_n\lambda_a^{-1}\right) = o_P(1).
\]

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

Finally, we prove (d) of the theorem. By Lemma 12, when $p_{\min} \gg n^{3/2}$, $\|R_{T \cup S}\|_\infty = o_P(n^{-1/2})$. Then, it follows from (S12), (S14) and (S15) that
\[
\sqrt{n}(\hat{\beta}_{a,T} - \beta^*_T) = \frac{1}{\sqrt{n}}K_n^{-1}U'_T\epsilon + o_P(1).
\]
\[
\sqrt{n}(\hat{\beta}_{a,S_a} - \beta_{S_a}^{*}) = \frac{1}{\sqrt{n}}K_n^{-1}U_{S_a}^T\epsilon - K_n^{-1}\sqrt{n}\lambda_a\hat{p}(|\hat{\beta}_{a,S_a}|)I(\hat{\beta}_{a,S_a} > 0) + o_P(1).
\]

Since \(\sqrt{n}\lambda_a\hat{p}(|\hat{\beta}_{a,S_a}|) \leq \sqrt{n}\lambda_a\hat{p}(d_n) = o(1)\). Therefore,
\[
K_n^{-1}\sqrt{n}\lambda_a\hat{p}(|\hat{\beta}_{a,S}|) = O_P(\sqrt{n}\lambda_a\hat{p}(d_n)) = o_P(1).
\]

This completes the proof.

\[\square\]

S5 Proof of Theorem 4

Proof. We divide the proof into two main steps.

In Step 1, define \(T_0\) as
\[T_0 = \sigma_{\epsilon}^{-2}(\omega_n + \sqrt{n}h_n)'\Psi^{-1}(\omega_n + \sqrt{n}h_n),\]
where \(h_n = A\beta^* - b, \Psi = A\Omega_T A\), \(\Omega_T\) is the submatrix of \(\Omega_{T \cup S_a} = \Sigma_u^{-1} T \cup S_{a}\) with rows and columns in \(T\), and \(\Sigma_u^{-1} T \cup S_{a}\) is inverse of the submatrix of \(\Sigma_u\) with rows and columns in \(T \cup S_a\), and \(\omega_n = n^{-1/2} (A 0) K_n^{-1} U_{T \cup S_a} \epsilon\). We first show that \(T_w/r = T_0/r + o_P(1)\).

By Theorem 3, we have
\[
\sqrt{n}(\hat{\beta}_{a,T} - \beta_T^{*}) = \frac{1}{\sqrt{n}}K_n^{-1} U_T'\epsilon + R_a,
\]
for some remainder term \(R_a\) such that \(\|R_a\|_2 = o_P(1)\). Then we have
\[
\sqrt{n}A(\hat{\beta}_{a,T} - \beta_T^{*}) = \omega_n + AR_{a,T},
\]
where \(R_{a,T}\) is the subvector of \(R_a\) with indices in \(T\). By definition, \(A\beta_T^{*} - b = h_n\). Then,
\[
\sqrt{n}(A\hat{\beta}_{a,T} - b) = \omega_n + AR_{a,T} + \sqrt{n}h_n.
\]

Let \(\Psi_n = A(K_n^{-1})_T A\), where \((K_n^{-1})_T\) is submatrix of \(K_n^{-1}\) with rows and columns in \(T\). We have
\[
\sqrt{n}\Psi_n^{-1/2}(A\hat{\beta}_{a,T} - b) = \Psi_n^{-1/2}(\omega_n + AR_{a,T} + \sqrt{n}h_n). \quad (S17)
\]

Next, we bound \(\|\sqrt{n}\Psi_n^{-1/2}(A\hat{\beta}_{a,T} - b)\|_2\). By Lemmas 3 and 12, \(\|\Psi_n^{-1/2}A\|_2 = O_P(1)\), and \(\|R_{a,T}\|_2 = o_P(1)\). Then it follows that
\[
\|\Psi_n^{-1/2}AR_{a,T}\|_2 \leq \|\Psi_n^{-1/2}A\|_2\|R_{a,T}\|_2 = o_P(1) \quad (S18).
\]

Therefore, \(\sqrt{n}\Psi_n^{-1/2}(A\hat{\beta}_{a,T} - b) = \Psi_n^{-1/2}(\omega_n + \sqrt{n}h_n) + o_P(1)\). We further note that,
\[
E\|\Psi_n^{-1/2}\omega_n\|_2^2 = \text{tr}\left[ E_u\{\Psi_n^{-1/2}E_u(\omega_n\omega_n')\Psi_n^{-1/2}\} \right] = r\sigma_{\epsilon}^2.
\]

8
Then, by Markov’s inequality, \( \| \Psi_n^{-1/2} \omega_n \|_2 = O_P(\sqrt{r}) \). By Lemma 3, \( \lambda_{\text{max}}(\Psi_n^{-1}) = O_P(1) \). By Condition 11, \( \| h_n \|_2 = O(\sqrt{r/n}) \), so that \( \| \sqrt{n} \Psi_n^{-1/2} h_n \|_2 = O_P(\sqrt{r}) \). Therefore, \( \| \sqrt{n} \Psi_n^{-1/2} (A\hat{\alpha}_T - b) \|_2 = O_P(\sqrt{r}) \). By Lemma 3, \( \| \Psi_n^{1/2} (A\hat{\Omega}_T A')^{-1} \Psi_n^{1/2} - I \|_2 = O_P((s_a + t)/\sqrt{n}) \), where \( \hat{\Omega}_T \) is the first \( T \) rows and columns of the matrix

\[
\hat{\Omega}_{T \cup S_a} = n \left( \hat{U}_T' \hat{U}_T \right)^{-1} \hat{U}_{S_a}' \hat{U}_{S_a}.
\]

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

\[
\| \{ \sqrt{n} \Psi_n^{-1/2} (A\hat{\alpha}_T - b) \} \{ \Psi_n^{1/2} (A\hat{\alpha}_T A')^{-1} \Psi_n^{1/2} - I \} \{ \sqrt{n} \Psi_n^{-1/2} (A\hat{\alpha}_T - b) \}^2 \leq \| \Psi_n^{1/2} (A\hat{\alpha}_T A')^{-1} \Psi_n - I \|_2 \sqrt{n} \Psi_n^{-1/2} (A\hat{\alpha}_T - b) \|_2^2 = O_P \left( \frac{r(s_a + t)}{\sqrt{n}} \right) = o_P(r).
\]

(S19)

Let \( T_{w,0} = \hat{\sigma}_\epsilon^{-2} n (A\hat{\alpha}_T - b)' \Psi_n^{-1} (A\hat{\alpha}_T - b) \). By \( T_w \)'s definition and (S19), we have \( \hat{\sigma}_\epsilon^2 |T_w - T_{w,0}| = o_P(r) \). Condition 5 implies \( 1/\hat{\sigma}_\epsilon^2 = O_P(1) \). Therefore, \( |T_w - T_{w,0}| = o_P(r) \).

Next, we show that \( |T_{w,0} - T_0| = o_P(r) \). Let \( T_{w,1} = \hat{\sigma}_\epsilon^{-2} \| \Psi_n^{-1/2} \omega_n + \sqrt{n} \Psi_n^{-1/2} h_n \|_2 \). By (S17) and (S18), we have

\[
\hat{\sigma}_\epsilon^2 |T_{w,0} - T_0| = \| \Psi_n^{-1/2} \omega_n + \sqrt{n} \Psi_n^{-1/2} h_n + o_P(1) \|_2^2 = \| \Psi_n^{-1/2} \omega_n + \sqrt{n} \Psi_n^{-1/2} h_n \|_2^2 + o_P(1) + o_P(\sigma_\epsilon^{-1} (\omega_n + \sqrt{n} h_n)) = \| \Psi_n^{-1/2} \omega_n + \sqrt{n} \Psi_n^{-1/2} h_n \|_2^2 + o_P(1) + o_P(r) = \hat{\sigma}_\epsilon^2 |T_{w,1}| + (\omega_n + \sqrt{n} h_n)'(\Psi_n^{-1} - \Psi_0^{-1})(\omega_n + \sqrt{n} h_n) + o_P(r).
\]

By Lemma 3, we have that \( \| \Psi_n^{-1} - \Psi_0^{-1} \|_2 = o_P(1) \). Since \( \| \omega_n \|_2 \leq \| \Psi_n^{1/2} \|_2 \| \Psi_n^{-1/2} \omega_n \|_2 = O_P(\sqrt{r}) \), and by Condition 11, \( \| \sqrt{n} h_n \|_2 = O_P(\sqrt{r}) \). Therefore, \( \| \omega_n + \sqrt{n} h_n \|_2 = O_P(\sqrt{r}) \).

Considering that \( 1/\hat{\sigma}_\epsilon^2 = O(1) \), we have \( T_{w,0} = T_{w,1} + o_P(r) \). Finally, we have

\[
|T_{w,1} - T_0| = \frac{|\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2|}{\hat{\sigma}_\epsilon^2} \| \Psi_n^{-1/2} \omega_n + \sqrt{n} \Psi_n^{-1/2} h_n \|_2 = o_P(r).
\]

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 \( \chi^2 \) approximation holds for \( T_0 \). Recall the definition of \( T_0 \), which can be written as \( T_0 = \sigma_\epsilon^{-2} \| \Psi_n^{-1/2} \omega_n + \sqrt{n} \Psi_n^{-1/2} h_n \|_2^2 \). By the definition of \( \omega_n \),

\[
\sigma_\epsilon^{-1} \Psi_n^{-1/2} \omega_n = \sum_{i=1}^n \frac{1}{\sqrt{n_\epsilon}} \Psi_n^{-1/2} (A \ 0) \Omega_{i \cup S_a} U_{i \cup S_a} \epsilon_i = \sum_{i=1}^n \xi_i.
\]

By direct calculation, we have \( \sum_{i=1}^n \text{Var}(\xi_i) = I_r \). Because of the sub-Gaussian assumption on \( \epsilon \) in Condition 1, we have \( E|\epsilon|^3 < \infty \). Then,

\[
r^{1/4} \sum_{i=1}^n E|\xi_i|^3 = \frac{r^{1/4}}{(n \sigma_\epsilon^2)^{3/2}} \sum_{i=1}^n E\| \Psi_n^{-1/2} (A \ 0) \Omega_{i \cup S_a} U_{i \cup S_a} \epsilon_i \|_2^3
\]

9
Together, we have

\[ \frac{\nu}{\sqrt{n}} \rightarrow \mathcal{N}(0,1), \]

In addition, by Lemma 5, we have

\[ \frac{\nu}{\sqrt{n}} \rightarrow \mathcal{N}(0,1), \]

where

\[ \nu = \sqrt{n} \mathbb{E} \left[ \frac{\psi^{-1/2}(\mathbf{A} - \mathbf{0}) \Omega_{\mathcal{T} \cup \mathcal{S}_n} \mathbf{U}_{i, \mathcal{T} \cup \mathcal{S}_n}}{\mathcal{L}} \right] \]

This completes Step 2.

Combining the results of Step 1 and Step 2 completes the proof. \qed
S6  Consistency of \( \hat{\sigma}_\epsilon^2 \)

Recall the estimator for \( \sigma_\epsilon \) in Section 3, \( \hat{\sigma}_\epsilon^2 = n^{-1} \sum_{i=1}^{n} (y_i - x_i' \tilde{\beta})^2 \), where \( \tilde{\beta} = \text{argmin}_\beta (2n)^{-1} \sum_{i=1}^{n} (y_i - x_i' \beta)^2 + \lambda_\epsilon \| \beta \|_1 \). The next proposition shows that \( \hat{\sigma}_\epsilon^2 \) is a consistent estimator of \( \sigma_\epsilon^2 \), and thus \( \hat{\sigma}_\epsilon^2 \) satisfies Condition 5.

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

**Proof.** Letting \( H_x = n^{-1} \sum_{i=1}^{n} x_i^{(2)} \) and \( \Delta = \tilde{\beta} - \beta^* \), we have
\[
\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2 = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2 - \sigma_\epsilon^2 = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2 - \sigma_\epsilon^2 \leq \frac{2}{n} \sum_{i=1}^{n} \epsilon_i x_i.
\] (S21)

By the sub-Gaussian assumption on \( \epsilon_i \) in Condition 1, it follows from the standard concentration result (e.g., Ning and Liu, 2017, Lemma H.2) that
\[
\frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2 - \sigma_\epsilon^2 = O_P \left( \sqrt{\frac{(\log n)}{n}} \right).
\]

Condition 3 implies that \( \lambda_{\min}(E(x_i^{(2)})) \geq \lambda_{\min}(E(u_i^{(2)})) > c \). Then, it follows from Raskutti et al. (2011, Proposition 1) that the restricted eigenvalue condition holds for \( H_x \). Then, following a similar argument as in the proof of Ning and Liu (2017, Lemma B.3), we have
\[
\Delta' H_x \Delta = O_P \left( s^*(\log p)/n \right).
\] (S22)

Moreover, \( \epsilon_i x_{ij} = \epsilon_i (\sum_{k=1}^{K} \lambda_{jk} f_{ik} + u_{ij}) \), where \( \lambda_{jk} \) is the \((j,k)\)th element of \( \Lambda \). 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 \( |n^{-1} \sum_{i=1}^{n} \epsilon_i x_i|_\infty = O_P(\sqrt{\log p}/n) \). By the well-known estimation error of the Lasso estimator (e.g., Negahban et al., 2012, Corollary 2), it holds that \( \| \Delta \|_1 = O_P \left( s^* \sqrt{\log p}/n \right) \). Therefore,
\[
\left| \frac{2}{n} \Delta' \sum_{i=1}^{n} \epsilon_i x_i \right| \leq \| \Delta \|_1 \left| \frac{2}{n} \sum_{i=1}^{n} \epsilon_i x_i \right|_\infty = O_P \left( s^* (\log p)/n \right).
\] (S23)

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

\qed

S7  Proof of Proposition 1

**Proof.** To prove (a), consider regressing \( y \) using all but the \( m \)th modality. Letting \( P_{-m} = X_{-m}(X'_{-m} X_{-m})^{-1} X'_{-m} \), then \( \tilde{Y}_{-m} = P_{-m} Y \). Letting \( \xi = X_m \beta^*_{m} + \epsilon \), then
\[
\| Y - \tilde{Y}_{-m} \|^2 = Y' (I_n - P_{-m}) Y = \xi' (I_n - P_{-m}) \xi
\]
\[
= \epsilon' (I_n - P_{-m}) \epsilon + 2 (X_m \beta^*_{m})' (I_n - P_{-m}) \epsilon + (X_m \beta^*_{m})' (I_n - P_{-m}) X_m \beta^*_{m}.
\]
Taking the expectation on both sides, and noting that,
\[
\begin{align*}
\mathbb{E}\{e'(I_n - \mathbf{P}_m)e\} &= (n-p_m)\sigma^2_e, \\
\mathbb{E}\{2(X_m\beta^*_m)'(I_n - \mathbf{P}_m)e\} &= 0, \\
\mathbb{E}\{(X_m\beta^*_m)'(I_n - \mathbf{P}_m)X_m\beta^*_m\} &= \mathbb{E}_{x_m|x_m}[(X_m\beta^*_m)'(I_n - \mathbf{P}_m)X_m\beta^*_m]\]
\end{align*}
\]
Moreover,
\[
\begin{align*}
\mathbb{E}_{x_m}|x_m\mathbb{E}[(I_n - \mathbf{P}_m)\sigma^2_{m|-m}] = (n-p_m)\sigma^2_{m|-m}.
\end{align*}
\]

To prove (b), note that \(\mathbb{E}(\|Y - \hat{Y}\|^2) = (n-p)\sigma^2_e\) when regressing \(y\) on all data modalities. Then a direct calculation proves (b).

To prove (c), by factor decomposition, we have \(x_m = \Lambda_m f + u_m\) and \(x_m = \Lambda_m f + u_m\). Then
\[
\begin{align*}
\begin{pmatrix} u_m \\ x_m \end{pmatrix} &\sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{u_m} & \Sigma_{u_m} \\ \Sigma_{u_m} & \Sigma_{x_m} \end{pmatrix} \right),
\end{align*}
\]
where \(\Sigma_{x_m} = \Lambda_m \Lambda_m' + \Sigma_{u_m}\). Consequently,
\[
\begin{align*}
\mathbb{E}(u_m|x_m) = \Sigma_{u_m}^{-1} \Sigma_{x_m} x_m, \quad \text{Var}(u_m|x_m) = \Sigma_{u_m} - \Sigma_{u_m} \Sigma_{x_m}^{-1} \Sigma_{u_m}.
\end{align*}
\]
As \(\Lambda_m f = x_m - u_m\), then \(f = D_m^{-1} \Lambda_m'(x_m - u_m)\), where \(D_m = \Lambda_m \Lambda_m\). Then,
\[
\begin{align*}
x'_m \beta^*_m &= f' \Lambda_m' \beta^*_m + u'_m \beta^*_m = (x_m - u_m)' C_m \beta^*_m + u'_m \beta^*_m,
\end{align*}
\]
where \(C_m = \Lambda_m D_m^{-1} \Lambda_m'\). Therefore, we have
\[
\begin{align*}
\mathbb{E}(x'_m \beta^*_m|x_m) &= \{x_m - \mathbb{E}(u_m|x_m)\}' C_m \beta^*_m = x'_m (I - \Sigma_{u_m} \Sigma_{x_m}^{-1})' C_m \beta^*_m. \quad (S24)
\end{align*}
\]
Moreover,
\[
\begin{align*}
\text{Var}(x'_m \beta^*_m|x_m) &= \beta^*_m \{C_m \text{Var}(u_m|x_m) C_m + \Sigma_{u_m}\} \beta^*_m \\
&= \beta^*_m \{C_m (\Sigma_{u_m} - \Sigma_{u_m} \Sigma_{x_m}^{-1} \Sigma_{u_m}) C_m + \Sigma_{u_m}\} \beta^*_m. \quad (S25)
\end{align*}
\]
Then, by Woodbury matrix identity, we have
\[
\begin{align*}
\Sigma_{x_m}^{-1} &= \Sigma_{u_m}^{-1} - \Sigma_{u_m}^{-1} \Lambda_m (I_K + \Lambda_m' \Sigma_{u_m} \Lambda_m)^{-1} \Lambda_m' \Sigma_{u_m}^{-1}.
\end{align*}
\]
Then we have,
\[
\begin{align*}
\Sigma_{u_m} - \Sigma_{u_m} \Sigma_{x_m}^{-1} \Sigma_{u_m} &= \Lambda_m (I_K + \Lambda_m' \Sigma_{u_m} \Lambda_m)^{-1} \Lambda_m', \\
C'_m (\Sigma_{u_m} - \Sigma_{u_m} \Sigma_{x_m}^{-1} \Sigma_{u_m}) C_m &= \Lambda_m (I_K + \Lambda_m' \Sigma_{u_m} \Lambda_m)^{-1} \Lambda_m'.
\end{align*}
\]
Plugging these equalities into (S25) gives
\[
\begin{align*}
\text{Var}(x'_m \beta^*_m|x_m) &= \beta^*_m \{\Lambda_m (I_K + \Lambda_m' \Sigma_{u_m} \Lambda_m)^{-1} \Lambda_m' + \Sigma_{u_m}\} \beta^*_m,
\end{align*}
\]
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 $H_m \in \mathcal{R}^{K_m \times K_m}$, such that

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

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

(c) $\| I_{K_m} - H_m H_m' \|_2 = O_P \left( 1/\sqrt{n} + 1/\sqrt{p_m} \right)$.

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

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 $F = (F_1, \ldots, F_M)$, $\hat{F} = (\hat{F}_1, \ldots, \hat{F}_M)$, where $\hat{F}_m$ is obtained by running PCA on the $m$th modality, $H = \text{diag}(H_1, \ldots, H_M)$, $U = (U_1, \ldots, U_m)$, $\hat{U} = (\hat{U}_1, \ldots, \hat{U}_M)$, and $p_{\min} = \min_{m \in [M]} p_m$. Then the following results hold:

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

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

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

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) $\| \Psi_n^{-1/2} A \|_2 = O_P (1)$.

(b) $\| \sqrt{n} \Psi_n^{-1/2} h_n \|_2 = O_P (\sqrt{r})$.

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

(d) $\| \Psi_n^{-1} - \Psi^{-1} \|_2 = O_P (1)$.

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

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

$$\| \Psi - \Psi_n \|_2 = \| A (\Omega_T - (K_n^{-1})_T) A' \|_2 \lesssim \| \Omega_T - (K_n^{-1})_T \|_2 \leq \| \Sigma_{u,T \cup S_a} - K_n^{-1} \|_2,$$

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

$$\| K_n - \Sigma_{u,T \cup S_a} \|_2 \leq (t + s_a) \| K_n - \Sigma_{u,T \cup S_a} \|_\infty = o_P (1).$$
Consequently,
\[ \| \Psi - \Psi_n \|_2 \lesssim \| \Sigma_{a,T \cup S_n}^{-1} - K_n^{-1} \|_2 \leq \| K_n^{-1} \|_2 \| \Sigma_{a,T \cup S_n}^{-1} \|_2 \| K_n - \Sigma_{a,T \cup S_n} \|_2 = o_p(1). \]

This completes the proof. \(\square\)

**Lemma 4.** Let \(\{X_i\}_{i=1}^n\) denote independent \(p\)-dimensional random vectors, with \(E(X_i) = 0\) and \(\sum_{i=1}^n \text{Var}(X_i) = I_p\). Let \(Z\) denote a \(p\)-dimensional multivariate normal vector, with mean \(0\) and covariance matrix \(I_p\). Then,
\[ \sup_C \left| P \left( \sum_{i=1}^n X_i \in C \right) - P(Z \in C) \right| \leq c_0 p^{1/4} \sum_{i=1}^n E(\|X_i\|_2), \]
for some constant \(c_0\), where the supremum is taken over all convex subsets in \(\mathbb{R}^p\).

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

**Lemma 5.** Let \(\chi^2(r, \gamma)\) denote a \(\chi^2\) random variable with \(r\) degrees of freedom and the non-centrality parameter \(\gamma\). Then,
\[ \lim_{\epsilon \to 0^+} \sup_{r \geq 1, \gamma \geq 0} |P(\chi^2(r, \gamma) \leq x + r \epsilon) - P(\chi^2(r, \gamma) \leq x - r \epsilon)| \to 0. \]

**Proof.** The result follows from Lemma S7 of Shi et al. (2019). \(\square\)

**Lemma 6.** Suppose Conditions 1–3 hold, and \(\lambda_1 \asymp \sqrt{(\log p_m)/n + 1/\sqrt{p_m}}\). Then,
\[ \| \hat{\beta}_m - \beta_m^* \|_1 = O_p \left( s_{m}^* \left\{ \sqrt{(\log p_m)/n + 1/\sqrt{p_m}} \right\} \right). \]

**Proof.** Recall that
\[ (\hat{\beta}_m, \hat{\gamma}_m) = \arg\min_{(\beta_m, \gamma_m)} \frac{1}{2n} \sum_{i=1}^n (y_i - x_{i,m}' \beta_m - \tilde{f}_{i,m} \gamma_m)^2 + \lambda_1 \| \beta \|_1. \] (S26)

By Lemma 1, there exists a nonsingular matrix \(H_m \in \mathcal{R}^{K_m \times K_m}\) such that \(\tilde{F}_m = \tilde{F}_m H_m\) is a consistent estimator of \(F_m\). We note that (S26) is equivalent to
\[ (\tilde{\beta}_m, \tilde{\gamma}_m) = \arg\min_{(\beta_m, \gamma_m)} \frac{1}{2n} \sum_{i=1}^n (y_i - x_{i,m}' \beta_m - \tilde{f}_{i,m} \gamma_m)^2 + \lambda_1 \| \beta \|_1, \] (S27)

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

Let \(Q = (X_m, F_m)\), \(\tilde{Q} = (X_m, \tilde{F}_m)\), and \(\tilde{Q} = Q \tilde{H}\), where \(\tilde{H} = \text{diag}(I_{p_m}, H_m) \in \mathcal{R}^{q_m \times q_m}\) is a block-diagonal matrix, and \(q_m = p_m + K_m\). Let \(\vartheta = (\beta^*_m, \gamma^*_m) \in \mathcal{R}^{q-m}\), \(\tilde{\vartheta} = (\tilde{\beta}_m, \tilde{\gamma}_m)' \in \mathcal{R}^{q-m}\) denote the solution of (S26), \(\tilde{\vartheta} = \tilde{H}^{-1}, \tilde{\vartheta} \in \mathcal{R}^{q-m}\) and \(\tilde{\vartheta}^* = \)
\[ \mathbf{H}^{-1}(\mathbf{\beta}'_m, \mathbf{\gamma}'_m) \in \mathcal{R}^{q-m}. \] By direct calculation, we can verify that \( \mathbf{\hat{\beta}}_m = \mathbf{\hat{\theta}}_{[p-m]} = \mathbf{\hat{\theta}}_{[p-m]}, \)

where \( \mathbf{\hat{\theta}} = (\mathbf{\hat{\beta}}_m, \mathbf{\hat{\gamma}}_m)' \) solves (S27). Then, it follows that

\[
\| \mathbf{\hat{\beta}}_m - \mathbf{\beta}'_m \|_1 = \| \mathbf{\hat{\theta}}_{[p-m]} - \mathbf{\hat{\theta}}_{[p-m]} \|_1 \leq \| \mathbf{\hat{\theta}} - \mathbf{\hat{\theta}}^* \|_1.
\]

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

To bound \( \| \nabla \ell(\mathbf{\hat{\theta}}) \|_\infty, \) we aim to show that

\[
\left\| \nabla \ell(\mathbf{\hat{\theta}}) \right\|_\infty = \left\| \frac{1}{n} \sum_{i=1}^n \mathbf{\tilde{Q}}_i \mathbf{\hat{\epsilon}}_i \right\|_\infty = O_P \left( \sqrt{\frac{\log p_m}{n}} + \frac{1}{\sqrt{p_n}} \right), \quad \text{(S28)}
\]

where \( \mathbf{\tilde{Q}}_i \) is the \( i \)th row of \( \mathbf{\tilde{Q}}. \) Indeed,

\[
\left\| \frac{1}{n} \sum_{i=1}^n \mathbf{\tilde{Q}}_i \mathbf{\hat{\epsilon}}_i \right\|_\infty \leq \max_{j \in [q-m]} \left| \frac{1}{n} \sum_{i=1}^n Q_{ij} \mathbf{\hat{\epsilon}}_i \right| + \max_{j \in [q-m]} \left| \frac{1}{n} \sum_{i=1}^n (\mathbf{\tilde{Q}}_{ij} - Q_{ij}) \mathbf{\hat{\epsilon}}_i \right|.
\]

By Condition 1 and Bernstein inequality,

\[
P \left( \left| \frac{1}{n} \sum_{i=1}^n Q_{ij} \mathbf{\hat{\epsilon}}_i \right| > C \sqrt{\frac{\log q_m}{n}} \right) \leq q_m^{-2} \text{ for all } j \in [q-m].
\]

Then, by the union bound,

\[
\max_{j \in [q-m]} \left| \frac{1}{n} \sum_{i=1}^n Q_{ij} \mathbf{\hat{\epsilon}}_i \right| = O_P \left( \sqrt{(\log q_m)/n} \right) = O_P \left( \sqrt{(\log p_m)/n} \right), \quad \text{(S29)}
\]

where the last equality is due to the fact that, since \( K_m \) is fixed, \( q_m \approx p_m. \) We choose to present the results using \( p_m \) in order to unify the presentation. Then by Cauchy-Schwartz inequality,

\[
\max_{j \in [q-m]} \left| \frac{1}{n} \sum_{i=1}^n (\mathbf{\tilde{Q}}_{ij} - Q_{ij}) \mathbf{\hat{\epsilon}}_i \right| \leq \max_{j \in [q-m]} \left( \frac{1}{n} \sum_{i=1}^n (\mathbf{\tilde{Q}}_{ij} - Q_{ij})^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{\hat{\epsilon}}_i^2 \right)^{1/2}
\]

If we use the \( m \)th modality to obtain \( \mathbf{\hat{f}}_{i,m}, \) it follows from Lemma 1 that

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

where \( (\mathbf{\hat{F}}_m \mathbf{H}_m)_{ik} \) is the \( (i,k) \)th element of \( \mathbf{\hat{F}}_m \mathbf{H}_m \) and \( f_{i,m_k} \) is the \( k \)th element of \( f_{i,m}. \)

This implies that

\[
\max_{j \in [q-m]} \left( \frac{1}{n} \sum_{i=1}^n (\mathbf{\tilde{Q}}_{ij} - Q_{ij})^2 \right)^{1/2} = O_P \left( 1/\sqrt{n} + 1/\sqrt{q_m} \right) = O_P \left( 1/\sqrt{n} + 1/\sqrt{p_m} \right).
\]
Since \((n^{-1} \sum_{i=1}^{n} \epsilon_i^2)^{1/2} = O_P(1)\), we have
\[
\max_{j \in [m-n]} \left| \frac{1}{n} \sum_{i=1}^{n} (\tilde{Q}_{ij} - Q_{ij}) \epsilon_i \right| = O_P \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p_m}} \right). \tag{S30}
\]
Combining (S29) and (S30) together proves (S28).

To check the restricted eigenvalue condition, it follows from Condition 3 and the factor decomposition (1) that \(\lambda_{\min}(E(x_{,m}^{(2)}) \geq \lambda_{\min}(E(u_{,m}^{(2)}) > c\). In addition, the sub-Gaussian assumptions on \(f_{-m}^{*}\) and \(u_{-m}^{*}\) imply that \(x_{-m}^{*}\) is also sub-Gaussian. Since \(\nabla^2_{\beta_{S_{-m}}} \ell(\theta) = n^{-1} \sum_{i=1}^{n} x_{,i,S_{-m}}^{(2)}\), 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. \(\square\)

**Lemma 7.** Suppose Conditions 1–3 hold. Then,
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} x_{i,-m}(\hat{f}_{i,mk} - f_{i,-m}^* w_k^*) \right\|_{\infty} = O_P \left( \frac{\log p_m}{n} \left( 1 + \frac{n^{1/4}}{\sqrt{p_m}} \right) \right). \tag{S31}
\]

**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 \(C\), \(w_k^*\) is in the feasible set with high probability; i.e.,
\[
\frac{1}{n} \left\| \sum_{i=1}^{n} x_{i,-m}(\hat{f}_{i,mk} - f_{i,-m}^* w_k^*) \right\|_{\infty} \leq C\sqrt{(\log p_m)/n \{ 1 \vee (n^{1/4}/\sqrt{p_m}) \}}.
\]
We have
\[
\frac{1}{n} \sum_{i=1}^{n} x_{i,-m}(\hat{f}_{i,mk} - f_{i,-m}^* w_k^*) = \frac{1}{n} \sum_{i=1}^{n} x_{i,-m}(\hat{f}_{i,mk} - f_{i,mk}^\dagger) + \frac{1}{n} \sum_{i=1}^{n} x_{i,-m}(f_{i,mk}^\dagger - x_{i,-m}^* w_k^*),
\]
where \(f_{i,mk}^\dagger = H_m f_{i,m}\) for some non-singular matrix \(H_m \in \mathcal{R}_{K_m \times K_m}\), and \(f_{i,mk}^\dagger\) is the \(k\)th element of \(f_{i,mk}^\dagger\). The sub-Gaussian assumption in Condition 1 implies that \(X_{ij,f_{i,mk}}^\dagger\) and \(X_{ij,x_{i,-m}^* w_k^*}\) are sub-exponential. Then, by Bernstein inequality and the union bound,
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} x_{i,-m}(f_{i,mk}^\dagger - x_{i,-m}^* w_k^*) \right\|_{\infty} = O_P \left( \frac{\log p_m}{n} \right).
\]
On the other hand, we have
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} x_{i,-m}(\hat{f}_{i,mk} - f_{i,mk}^\dagger) \right\|_{\infty} \leq \left( \max_{i \leq n} |\hat{f}_{i,mk} - f_{i,mk}^\dagger| \right) \left\| \frac{1}{n} \sum_{i=1}^{n} x_{i,-m} \right\|_{\infty}
\]
16
\[ \leq \left( \max_{i \leq n} \| \hat{f}_{i,m} - f^\dagger_{i,m} \|_2 \right) \left\| \frac{1}{n} \sum_{i=1}^{n} x_{i,-m} \right\|_\infty = O_P \left( \frac{1}{\sqrt{n}} + \frac{n^{1/4}}{\sqrt{p_m}} \right) O_P \left( \sqrt{\frac{\log p_m}{n}} \right), \]

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

Lemma 8. Suppose Conditions 1–3 hold, and \( \lambda_2 \simeq \sqrt{(\log p_m)/n} \{1 \vee (n^{1/4}/\sqrt{p_m}) \} \). Then,

\[ \| \hat{w}_k - w_k^* \|_1 = O_P \left( s_k^* \left\{ \sqrt{\frac{\log p_m}{n}} \left( 1 \vee \frac{n^{1/4}}{\sqrt{p_m}} \right) \right\} \right). \]

Proof. Recall that, for the \( k \)th column of \( W \), we solve

\[ \hat{w}_k = \text{argmin} \| w_k \|_1, \text{ such that } \left\| \frac{1}{n} \sum_{i=1}^{n} x_{i,-m}(\hat{f}_{i,m} - x_{i,-m}w_k) \right\|_\infty \leq \lambda_2, \]

where \( \hat{f}_{i,m} \) is the \( k \)th element of \( \hat{f}_{i,m} \). Let \( S_k = \text{supp}(w_k^*) \), where \( w_k^* \) is the \( k \)th column of \( W^* = E(x_{i,-m}^{\otimes 2})^{-1}E(x_{i,-m}f_{i,m}) \). Then, we have \( \| w_{S_k}^* \|_1 \geq \| \hat{w}_{S_k}^* \|_1 + \| \hat{w}_{S_k}^* \|_1 \). By the triangle inequality, we have \( \| \hat{w}_{S_k}^* \|_1 \geq \| w_{S_k}^* \|_1 - \| \hat{w}_{S_k} - w_{S_k}^* \|_1 \). Let \( \hat{\Delta}_k = \hat{w}_k - w_k^* \). Then, by noting that \( \| w_{S_k}^* \|_1 = 0 \), we have \( \| \hat{\Delta}_{S_k} \|_1 \geq \| \Delta_{S_k} \|_1 \). It follows from Lemma 7 that

\[ \left\| \frac{1}{n} \sum_{i=1}^{n} x_{i,-m}(\hat{f}_{i,m} - x_{i,-m}w_k) \right\|_\infty = O_P \left( \sqrt{\frac{\log p_m}{n}} \left( 1 \vee \frac{n^{1/4}}{\sqrt{p_m}} \right) \right). \]

Denote \( H_x = n^{-1} \sum_{i=1}^{n} x_{i,-m}^{\otimes 2} \), \( H_x f = n^{-1} \sum_{i=1}^{n} x_{i,-m}\hat{f}_{i,m} \), and \( \Delta_k = \hat{w}_k - w_k^* \). Then,

\[ \| H_x \hat{\Delta}_k \|_\infty \leq \| H_x f - H_x \hat{w}_k \|_\infty + \| H_x f - H_x w_k^* \|_\infty = O_P \left( \sqrt{\frac{\log p_m}{n}} \left( 1 \vee \frac{n^{1/4}}{\sqrt{p_m}} \right) \right). \]

Together with \( \| \hat{\Delta}_k \|_1 \leq 2 \| \Delta_{S_k} \|_1 \leq 2 \sqrt{s_k^*} \| \hat{\Delta}_k \|_2 \), we have

\[ \hat{\Delta}_k H_x \hat{\Delta}_k \leq \| \hat{\Delta}_k \|_1 \| H_x \hat{\Delta}_k \|_\infty = O_P \left( \| \hat{\Delta}_k \|_1 \sqrt{\frac{\log p_m}{n}} \left( 1 \vee \frac{n^{1/4}}{\sqrt{p_m}} \right) \right) \]

\[ = O_P \left( \| \hat{\Delta}_k \|_2 \sqrt{s_k^* \log p_m / n} \left( 1 \vee \frac{n^{1/4}}{\sqrt{p_m}} \right) \right). \] (S32)

Note that Condition 1 and (1) imply that \( x_{-m} \) is sub-Gaussian. In addition, Condition 3 implies that \( \lambda_{\min}(E(x_{i,-m}^{\otimes 2})) > c \). Then, it follows from Proposition 1 of Raskutti et al. (2011) that the restricted eigenvalue condition holds for \( H_x \) with high probability, i.e.,

\[ \Delta_k H_x \hat{\Delta}_k \geq \kappa \| \hat{\Delta}_k \|_2^2 \]

for some \( \kappa > 0 \) and all \( \hat{\Delta}_k \), such that \( \| \Delta_{S_k} \|_1 \leq \| \Delta_{S_k} \|_1 \). This result, together with (S32), implies that

\[ \| \hat{\Delta}_k \|_1 \leq 2 \sqrt{s_k^*} \| \hat{\Delta}_k \|_2 = O_P \left( s_k^* \sqrt{\frac{\log p_m}{n}} \left( 1 \vee \frac{n^{1/4}}{\sqrt{p_m}} \right) \right). \]

This completes the proof. \( \square \)
Lemma 9. Suppose the conditions of Theorem 2 hold. Then, uniformly for all $\beta^* \in \mathcal{N}$,

$$\|\sqrt{n}(\hat{I}_{\gamma_m|\beta_m}^{-1/2} - I_{\gamma_m|\beta_m}^{-1/2})\hat{S}(\hat{\beta}_m, 0)\|_2 = o_P(1).$$

Proof. First, we note that

$$\|\hat{I}_{\gamma_m|\beta_m}^{-1/2} - I_{\gamma_m|\beta_m}^{-1/2}\|_2 = \|I_{\gamma_m|\beta_m}^{-1/2} (I_{\gamma_m|\beta_m}^{1/2} - \hat{I}_{\gamma_m|\beta_m}^{1/2}) I_{\gamma_m|\beta_m}^{-1/2}\|_2 \leq \|I_{\gamma_m|\beta_m}^{-1/2} - \hat{I}_{\gamma_m|\beta_m}^{-1/2}\|_2 \leq \|I_{\gamma_m|\beta_m}^{-1/2} - \hat{I}_{\gamma_m|\beta_m}^{-1/2}\|_{1/2} \leq \sqrt{K_m}\|I_{\gamma_m|\beta_m}^{-1/2} - \hat{I}_{\gamma_m|\beta_m}^{-1/2}\|_{1/2} = o_P(1),$$

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

$$\|\sqrt{n}(\hat{I}_{\gamma_m|\beta_m}^{-1/2} - I_{\gamma_m|\beta_m}^{-1/2})\hat{S}(\hat{\beta}_m, 0)\|_2 \leq \|\hat{I}_{\gamma_m|\beta_m}^{-1/2} - I_{\gamma_m|\beta_m}^{-1/2}\|_2 \|\sqrt{n}\hat{S}(\hat{\beta}_m, 0)\|_2 = o_P(1).$$

This completes the proof. $\square$

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

$$\sqrt{n}I_{\gamma_m|\beta_m}^{-1/2} \{\hat{S}(\hat{\beta}_m, 0) - S(\beta^*_m, 0)\} = o_P(1).$$

Proof. By (S4), we have that

$$S(\beta^*_m, 0) - \hat{S}(\hat{\beta}_m, 0) = \frac{1}{n\sigma^2_\epsilon}(W^* - \hat{W})'X'_m(Y - X_m\beta^*_m) + \frac{1}{n\sigma^2_\epsilon}(F'_mX_m - \hat{W}'X'_mX_m)(\hat{\beta}_m - \beta^*_m) \equiv I + II.$$

For the term $I$, under $H_a$, $Y - X_m\beta^*_m = X_m\beta^*_m + \epsilon$. Therefore,

$$\|\frac{1}{n\sigma^2_\epsilon}X'_m(Y - X_m\beta^*_m)\|_\infty = \|\frac{1}{n\sigma^2_\epsilon}X'_m(X_m\beta^*_m + \epsilon)\|_\infty \leq \|\frac{1}{n\sigma^2_\epsilon}X'_m\epsilon\|_\infty + \|\frac{1}{n\sigma^2_\epsilon}X'_m(F_m\gamma_m + U_m\beta^*_m)\|_\infty = O_P\left(\sqrt{\frac{\log p_m}{n}}\right),$$

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

$$|n\sigma^2_\epsilon^{-1}(w_k^* - \hat{w}_k)'X'_m(Y - X_m\beta^*_m)| \leq \|w_k^* - \hat{w}_k\|_1|n\sigma^2_\epsilon^{-1}(Y - X_m\beta^*_m)|_\infty = o_P(1/\sqrt{n}).$$

(S33)

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

18
For the term $II$, we bound $\|\hat{\beta}_m - \beta^*_m\|_1$ and $\|(n\sigma^2)_{\epsilon}^{-1}\hat{f}'_m X_m - \hat{W}' X'_m X_m\|_\infty$ respectively. To bound $\|\hat{\beta}_m - \beta^*_m\|_1$, under $H_a$, we only need to replace $\epsilon_i$ in (S28) with $\epsilon_i + U_{i,m}\beta^*_m$. Due to the sub-Gaussian assumption of $U_{i,m}$ in Condition 1, and the fact that $\epsilon_i$ and $U_{i,m}$ are uncorrelated, the bounds we have established in (S29) and (S30) still hold. Therefore, uniformly for all $\beta^* \in \mathcal{N}$, $\|\hat{\beta}_m - \beta^*_m\|_1$ has the same upper bound as the one established in Lemma 6. In addition, the bound we have established for $\|(n\sigma^2)_{\epsilon}^{-1}\hat{f}'_m X_m - \hat{W}' X'_m X_m\|_\infty$ also holds uniformly for all $\beta^* \in \mathcal{N}$. 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. \hfill \square

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

$$S(\beta^*, \gamma^*_m) - S(\beta^*_m, 0) - I_{\gamma_m|\beta_m}(\gamma^*_m) = O_P \left( n^{-1/2} \right).$$

**Proof.** By definition, we have

$$\sqrt{n} \{ S(\beta^*, \gamma^*_m) - S(\beta^*_m, 0) - I_{\gamma_m|\beta_m}(\gamma^*_m) \}$$

$$= \frac{1}{\sqrt{n\sigma^2_{\epsilon}}} \sum_{i=1}^{n} \{ x'^*_i \beta^*_m(f_{i,m} - W^*x_{i,-m}) - I_{\gamma_m|\beta_m}(\gamma^*_m) \}$$

$$+ \frac{1}{\sqrt{n\sigma^2_{\epsilon}}} \sum_{i=1}^{n} I_{i,m}(f_{i,m} - W^*x_{i,-m}) \equiv I + II.$$ 

For the term $I$, its $k$th element equals

$$\frac{1}{\sqrt{n\sigma^2_{\epsilon}}} \sum_{i=1}^{n} [(f_{i,m} - w^*_k x_{i,-m}) f^*_{i,m} x'_{i,m} \gamma^*_{m} - E \{(f_{i,m} - w^*_k x_{i,-m}) f^*_{i,m} x'_{i,m} \gamma^*_{m} \}].$$

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

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

$$\frac{1}{\sqrt{n\sigma^2_{\epsilon}}} \sum_{i=1}^{n} U'_{i,m} \beta^*_m(f_{i,m} - w^*_k x_{i,-m}) = O_P \left( \|b_m\|_2 \sqrt{\log n} \right) = O_P(1).$$

Combining the results for the terms $I$ and $II$ completes the proof. \hfill \square

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

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

where $R = \sum_{i=1}^{n} \sum_{j=1}^{p} (\hat{U}_i U \beta^* + \hat{U}_j (F \gamma^* - \hat{F}_a \gamma^*) + (\hat{U}_i - U) \epsilon_j)$. 

19
Proof. First, we bound \( n^{-1} \tilde{U}' (\mathcal{F} \gamma^* - \hat{F} \gamma_a) \). By Lemma 2, \( \| \tilde{U} - U \|_\infty = o_P(1) \). Therefore, \[
\begin{align*}
\| \frac{1}{n} \tilde{U}' (\mathcal{F} \gamma^* - \hat{F} \gamma_a) \|_\infty & \lesssim \| \frac{1}{n} U' (\mathcal{F} \gamma^* - \hat{F} \gamma_a) \|_\infty \\
& \leq \| \frac{1}{n} U' (\mathcal{F} \gamma^* - \mathcal{F} \mathcal{H} \mathcal{H}' \gamma^*) \|_\infty + \| \frac{1}{n} U' (\mathcal{F} \mathcal{H} \mathcal{H}' \gamma^* - \hat{F} \gamma_a) \|_\infty \\
& \leq \| \frac{1}{n} U' \mathcal{F} (I_K - \mathcal{H} \mathcal{H}') \gamma^* \|_\infty + \| \frac{1}{n} U' (\mathcal{F} \mathcal{H} \mathcal{H}' \gamma^* - \hat{F} \gamma_a) \|_\infty .
\end{align*}
\] (S34)

Let \( \tilde{\gamma} = (I_K - \mathcal{H} \mathcal{H}') \gamma^* \). Then, \( \| \tilde{\gamma} \|_2 \leq \| I_K - \mathcal{H} \mathcal{H}' \|_2 \| \gamma^* \|_2 \). By Lemma 2, \( \| I_K - \mathcal{H} \mathcal{H}' \|_2 = O_P \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p_{\min}}} \right) \).

Since \( \| \gamma^* \|_2^2 \lesssim \text{Var}(f' \gamma^*) \lesssim \sigma_y^2 = O(1) \), we have \( \| \tilde{\gamma} \|_\infty \leq \| \tilde{\gamma} \|_2 = O_P \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p_{\min}}} \right) \). Then,
\[
\| \frac{1}{n} U' \mathcal{F} (I_K - \mathcal{H} \mathcal{H}') \gamma^* \|_\infty = \max_{j \in [p]} \left| \frac{1}{n} \sum_{i=1}^{n} U_{ij} \sum_{k=1}^{K} f_{ik} \tilde{\gamma}_{ik} \right| .
\]

By Condition 1, \( K \) is fixed, and \( \| \tilde{\gamma} \|_2 = O_P \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p_{\min}}} \right) \), we have that \( \sum_{k=1}^{K} f_{ik} \tilde{\gamma}_{ik} \) is sub-Gaussian with variance bounded by \( O \left( \frac{1}{n} + \frac{1}{p_{\min}} \right) \). Moreover, \( U_{ij} \) is also sub-Gaussian and uncorrelated with \( f_{ik} \) for any \( k \in [K] \). Then, \( U_{ij} \sum_{k=1}^{K} f_{ik} \tilde{\gamma}_{ik} \) is sub-exponential. Applying Bernstein inequality, we have
\[
\| \frac{1}{n} U' \mathcal{F} (I_K - \mathcal{H} \mathcal{H}') \gamma^* \|_\infty = O_P \left( \frac{\log p}{n} \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p_{\min}}} \right) \right) .
\] (S35)

On the other hand, we have
\[
\| \frac{1}{n} U' (\mathcal{F} \mathcal{H} \mathcal{H}' \gamma^* - \hat{F} \gamma_a) \|_\infty \leq \| \frac{1}{n} U' (\mathcal{F} \mathcal{H} - \hat{F}) \mathcal{H}' \gamma^* \|_\infty + \| \frac{1}{n} U' \hat{F} (\mathcal{H}' \gamma^* - \hat{\gamma}_a) \|_\infty .
\] (S36)

For the first term in (S36), we have
\[
\begin{align*}
\| \frac{1}{n} U' (\mathcal{F} \mathcal{H} - \hat{F}) \mathcal{H}' \gamma^* \|_\infty & \leq \left( \max_{i \in [n]} |(\mathcal{F} \mathcal{H} - \hat{F})_{ik} (\mathcal{H}' \gamma^*)_k | \right) \| \frac{1}{n} \sum_{i=1}^{n} U_{i} \|_\infty \\
& \leq \| \mathcal{H}' \gamma^* \|_2 \left( \max_{i \in [n]} \| \hat{f}_i - \mathcal{H} f_i \|_2 \right) \| \frac{1}{n} \sum_{i=1}^{n} U_{i} \|_\infty \\
& = O_P \left( \frac{\log p}{n} \right) O_P \left( \frac{1}{\sqrt{n}} + \frac{n^{1/4}}{\sqrt{p_{\min}}} \right),
\end{align*}
\] (S37)

where the last equality is implied by Lemma 1, and the fact that \( \| \mathcal{H}' \gamma^* \|_2 \leq \| \mathcal{H}' \|_2 \| \gamma^* \|_2 = O(1) \). For the second term in (S36), we have
\[
\| \frac{1}{n} U' \hat{F} (\mathcal{H}' \gamma^* - \hat{\gamma}_a) \|_\infty \lesssim \| \frac{1}{n} U' (\mathcal{H}' \gamma^* - \hat{\gamma}_a) \|_\infty \leq \left( \max_{k \in [K]} |(\mathcal{H} \gamma^* - \hat{\gamma}_a)_k | \right) \sum_{k=1}^{K} \| \frac{1}{n} U_{i} f_{ik} \|_\infty
\]
Since $\hat{\gamma}_a \in \mathcal{M}$, we have $\max_{k \in [K]} |(H^* \gamma - \hat{\gamma}_a)_k| = O_P(\delta_n)$. By Condition 1, for each $k \in [K]$, $U_{ij} f_{ik}$ is sub-exponential. Then $\|n^{-1} U_{ij} f_{ik}\|_\infty = O_P\left(\sqrt{(\log p)/n}\right)$. Since $K$ is fixed, $\sum_{k=1}^K n^{-1} U_{ij} f_{ik} = O_P\left(\sqrt{(\log p)/n}\right)$. Then,

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

Therefore,

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

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

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

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

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

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

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

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

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

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

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

$$\left\| \frac{1}{n} \hat{F}'(Y - \hat{U}\beta_a - \hat{F}H^* \gamma) \right\|_\infty \leq \left\| \frac{1}{n} F' \{F \gamma - \hat{F}H^* \gamma + U\beta^* - \hat{U}\beta_a + \epsilon\} \right\|_\infty$$

$$\leq \left\| \frac{1}{n} F' \{F \gamma - FHH^* \gamma\} \right\|_\infty + \left\| \frac{1}{n} F'(FH - \hat{F})H^* \right\|_\infty$$

$$+ \left\| \frac{1}{n} F' \hat{U}\beta^* \right\|_\infty + \left\| \frac{1}{n} F'\hat{\beta}_a \right\|_\infty + \left\| \frac{1}{n} F'\epsilon \right\|_\infty.$$

Similar to (S35), we have

$$\left\| \frac{1}{n} F' \{F \gamma - FHH^* \gamma\} \right\|_\infty = \left\| \frac{1}{n} F' F(I_K - HH^*) \gamma^* \right\|_\infty = O_P\left(\sqrt{\log K/n} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p_{\text{min}}}}\right)\right).$$

21
Figure S1: The heat map of the correlation matrix of the multimodal neuroimaging data.

Similar to (S37), we have

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

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

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


