SUPPLEMENT TO "INFERENCE FOR LOW-RANK MODELS"

VICTOR CHERNOZHUKOV, CHRISTIAN HANSEN, YUAN LIAO, AND YINCHU ZHU

Abstract. This supplement contains all proofs.

Contents

Appendix A. Frobenius-norm matrix convergence and rank consistency

A.1. The convergence of $\|\widetilde{\Theta}_S - \Theta_S\|_F$.

Lemma A.1. *Suppose the n rows of $\mathcal{E} \circ X$ are independent and that each is a $1 \times p$ sub-Gaussian vector. In addition, suppose $\|\frac{1}{n}\mathsf{E}(\mathcal{E} \circ X)(\mathcal{E} \circ X)'\|$ is bounded. Then*

$$\|\mathcal{E} \circ X\| = O_P(\sqrt{n} + \sqrt{p}).$$

Date: This draft: July 13, 2023.

Proof. The eigenvalue-concentration inequality for independent sub-Gaussian random vectors (Theorem 5.39 of [6]) implies

$$\|(\mathcal{E}\circ X)(\mathcal{E}\circ X)' - \mathsf{E}(\mathcal{E}\circ X)(\mathcal{E}\circ X)'\| = O_P(\sqrt{np}+p),$$

which in turn shows

$$\|\mathcal{E}\circ X\|^2 \leq \|\mathsf{E}(\mathcal{E}\circ X)(\mathcal{E}\circ X)'\| + O_P(\sqrt{np}+p) = O_P(n+p).$$

□

To formally state the rate of convergence of the nuclear-norm penalized estimator, let

$$\omega_{np} := \nu\sqrt{J} + \|R\|_{(n)} \asymp (\sqrt{n}+\sqrt{p})\sqrt{J} + \|R\|_{(n)}.$$

Proposition A.1. *For $S \in \{\mathcal{I}, \mathcal{I}^c, \{1, ..., n\}\}$, there is a $J\times J$ rotation matrix H_S satisfying $H_S'H_S = \mathsf{I}$, so that*

$$\|\widetilde{\Theta}_S - \Theta_S\|_F^2 = O_P(\omega_{np}^2), \quad \|\widetilde{V}_S - V_0 H_S\|_F = O_P(J^d \omega_{np}\psi_{np}^{-1}).$$

Proof. The convergence of $\|\widetilde{\Theta}_S - \Theta_S\|_F^2$ follows from the standard arguments in the low-rank literature, e.g., [4]. Moreover, given the convergence (in Frobinus norm) of $\widetilde{\Theta}$, the convergence of $\widetilde{V} - V_0$ follows straightforward by applying the Weyls' theorem for bounding the eigenvalues and the sin-theta theorem for bounding the eigenvectors. The extra "J^d" term in the rate arises from the eigengap ψ_{np}/J^d. We omit details for brevity.

The only thing we would like to emphasize here is that the convergence requires the so-called "restrictive strong convexity", which we shall give a formal prove in the following subsection. □

A.2. **Proof of Restricted strong convexity.** Recall the SVD of Θ_0:

$$\Theta_0 = UDV', \quad U = (U_0, U_c), \quad V = (V_0, V_c).$$

Here (U_c, V_c) are the columns of U, V that correspond to the zero singular values, while (U_0, V_0) denote the columns of U, V associated with the non-zero singular values. In addition, for any $n\times p$ matrix A, let

$$\mathcal{P}(A) = U_c U_c' A V_c V_c' \text{ and } \mathcal{M}(A) = A - \mathcal{P}(A).$$

Here $\mathcal{M}(\cdot)$ can be thought of as the projection matrix onto the columns of U_0 and V_0, which is also the "low-rank" space of Θ_0. $\mathcal{P}(\cdot)$ is then the projection onto the space orthogonal to this low-rank space.

Lemma A.2. *Suppose (i)* $\max_{ij}|x_{ij}| < C$ *and* $\min_{ij}\mathsf{E}x_{ij}^2 > c_0$. *(ii) Either* $x_{ij} = c_j$ *almost surely for some constants* $c_j \neq 0$*, or* x_{ij} *is independent across both* (i,j)*. In addition, define the restricted low-rank set as, for some* $c > 0$*,*

$$\mathcal{C}(c_1, c_2) = \{A \in \mathcal{A} : \|\mathcal{P}(A)\|_{(n)} \leq c_1\|\mathcal{M}(A)\|_{(n)}, \|A\|_F^2 > c_2\sqrt{np}\}.$$

For any $c_1, c_2 > 0$ *there are constants* $\kappa, B > 0$ *so that with probability approaching one, uniformly for* $A \in \mathcal{C}(c_1, c_2)$*,*

$$\|X \circ A\|_F^2 \geq \kappa\|A\|_F^2 - J(n+p)B$$

The same inequality holds when Θ_0 *is replaced with its subsample versions:* $\Theta_{0,\mathcal{I}}$ *and* $\Theta_{0,\mathcal{I}^c}$*.*

Proof. For notational simplicity, write $X(A) = \|X \circ A\|_F^2$. Then for $\mathsf{E}x_{ij}^2 > c_0$,

$$\mathsf{E}X(A) = \sum_{ij} A_{ij}^2 \mathsf{E}x_{ij}^2 \geq c_0\|A\|_F^2.$$

Define an event, for sufficiently large $B > 0$,

$$\mathcal{E}(A) := \{|X(A) - \mathsf{E}X(A)| > 0.5\mathsf{E}X(A) + J(n+p)B\}.$$

We aim to claim $P(\exists A \in \mathcal{C}(c_1, c_2), \mathcal{E}(A)) \to 0$. Once this is proved, then $P(\forall A \in \mathcal{C}(c_1, c_2), \mathcal{E}(A)^c) \to 1$. On $\mathcal{E}(A)^c$, the restricted strong convexity holds for $\kappa = c_0/2$, because:

$$X(A) \geq 0.5\mathsf{E}X(A) - J(n+p)B \geq \kappa\|A\|_F^2 - J(n+p)B.$$

To prove $P(\exists A \in \mathcal{C}(c_1, c_2), \mathcal{E}(A)) \to 0$, we use the standard peeling argument. Let

$$\Gamma_l = \{A \in \mathcal{C}(c_1, c_2) : 2^l v_n \leq \mathsf{E}X(A) \leq 2^{l+1}v_n\},$$

where $v_n = B\sqrt{np}$ and $l \in \mathbb{N}$. We let $c_2 = 2c_0^{-1}B$ in the definition of $\mathcal{C}(c_1, c_2)$.

Step 1: show $\mathcal{C}(c_1, c_2) \subset \cup_{l=1}^{\infty}\Gamma_l$. For $A \in \mathcal{C}(c_1, c_2)$ we have

$$\|A\|_F^2 \geq c_2\sqrt{np} = 2c_0^{-1}B\sqrt{np} = 2c_0^{-1}v_n.$$

Then $\mathsf{E}X(A) \geq c_0\|A\|_F^2 \geq 2v_n$. Hence there is $l \in \mathbb{N}$ so that $A \in \Gamma_l$ as long as $A \in \mathcal{C}(c_1, c_2)$. This shows $\mathcal{C}(c_1, c_2) \subset \cup_{l=1}^{\infty}\Gamma_l$.

Now let

$$\begin{aligned} \mathcal{D}(x) &:= \{A \in \mathcal{C}(c_1, c_2) : \|A\|_F^2 \leq x\} \\ \mathcal{F} &:= \{A : |X(A) - \mathsf{E}X(A)| - J(n+p)B > 0.25 \times 2^{l+1}v_n\}. \end{aligned}$$

and $x_l = c_0^{-1}2^{l+1}v_n$.

Step 2: show $\{A : \mathcal{E}(A)$ is true$\} \cap \Gamma_l \subset \mathcal{D}(x_l) \cap \mathcal{F}$. If $A \in \Gamma_l$ and $\mathcal{E}(A)$ holds, then $A \in \mathcal{F}$ because:

$$|X(A) - \mathsf{E}X(A)| - J(n+p)B > 0.5\mathsf{E}X(A) \geq 0.5 \times 2^l v_n = 0.25 \times 2^{l+1}v_n.$$

Also $\|A\|_F^2 \leq c_0^{-1}\mathsf{E}X(A) \leq c_0^{-1}2^{l+1}v_n$. This implies $A \in \mathcal{D}(x_l)$.

Now let

$$Z(x) := \sup_{A \in \mathcal{D}(x)} |\frac{1}{np}\sum_{ij} x_{ij}^2 A_{ij}^2 - \mathsf{E}x_{ij}^2 A_{ij}^2|.$$

Step 3: bound $\mathsf{E}Z(x)$. For any $A \in \mathcal{D}(x) \subset \mathcal{C}(c_1, c_2)$, $C = 1 + c_1$,

$$\begin{aligned} \|A\|_{(n)} &= \|\mathcal{P}(A) + \mathcal{M}(A)\|_{(n)} \leq (1+c_1)\|\mathcal{M}(A)\|_{(n)} \\ &\leq (1+c_1)\sqrt{\mathsf{rank}(\Theta_0)}\|\mathcal{M}(A)\|_F \leq C\sqrt{J}\|A\|_F \leq C\sqrt{Jx}. \end{aligned}$$

Let ϵ_{ij} be an i.i.d. Rademacher sequence. Let $G = (\epsilon_{ij}x_{ij})_{n \times p}$. Then $\mathsf{E}\|G\| \leq C_1\sqrt{n+p}$ for some constant C_1. Then

$$\begin{aligned} \mathsf{E}Z(x) &\leq^{(i)} 2\mathsf{E} \sup_{A \in \mathcal{D}(x)} |\frac{1}{np}\sum_{ij} x_{ij}^2 A_{ij}^2 \epsilon_{ij}| \leq^{(ii)} C_2\mathsf{E} \sup_{A \in \mathcal{D}(x)} |\frac{1}{np}\sum_{ij} x_{ij}A_{ij}\epsilon_{ij}| \\ &= C_2\mathsf{E} \sup_{A \in \mathcal{D}(x)} |\frac{1}{np}\mathsf{tr}(GA')| \leq C_2\mathsf{E} \sup_{A \in \mathcal{D}(x)} \frac{1}{np}\|G\|\|A\|_{(n)} \\ &\leq C_3\frac{\sqrt{n+p}}{np} \sup_{A \in \mathcal{D}(x)} \|A\|_{(n)} \leq C_4\frac{\sqrt{n+p}}{np}\sqrt{Jx} = 2C_5\frac{\sqrt{J(n+p)}}{\sqrt{npc_0}}\sqrt{\frac{c_0}{32np}x} \\ &\leq \frac{c_0 x}{32np} + \frac{C_5^2}{c_0}\frac{J(n+p)}{np} \leq \frac{c_0 x}{32np} + B\frac{J(n+p)}{np} \end{aligned}$$

where (i) follows from the standard symmetrization argument; (ii) uses the contraction inequality (e.g., (2.3) of [3]), which requires $|x_{ij}| + |A|_{ij} \leq M$ for all (i, j). This holds since $A \in \mathcal{A}$ and by the assumption. The last inequality holds for sufficiently large $B > 0$.

Step 4: bound the tail probability of $Z(x) - \mathsf{E}Z(x)$. The conditions that $|A_{ij}| < M$ (because $A \in \mathcal{A}$) and the independence of x_{ij} over (i, j) allow us to apply the Massart inequality (e.g., Theorem 14.2 of [2]), so

$$P(Z(x) > \mathsf{E}Z(x) + t) \leq \exp(-C_6 npt^2), \quad \forall t > 0,$$

Let $t = \frac{7c_0 x}{32np}$. Then from Step 3,

$$\begin{aligned} & P(Z(x) > \frac{J(n+p)}{np}B + \frac{1}{np}0.25c_0x) \leq P(Z(x) > \mathsf{E}Z(x) - \frac{c_0x}{32np} + \frac{1}{np}0.25c_0x) \\ = \; & P(Z(x) > \mathsf{E}Z(x) + t) \leq \exp(-C_6npt^2) = \exp(-\frac{C_7x^2}{np}). \end{aligned}$$

Step 5: Pealing device. Hence for c' that only depends on c_0, but not on B,

$$\begin{aligned} & P(\exists A \in \mathcal{C}(c_1, c_2), \mathcal{E}(A)) \leq \sum_{l=1}^{\infty} P(A \in \Gamma_l, \mathcal{E}(A)) = \sum_{l=1}^{\infty} P(A \in \mathcal{D}(x_l) \cap \mathcal{F}) \\ \leq \; & \sum_{l=1}^{\infty} P(\sup_{A \in \mathcal{D}(x_l)} |X(A) - \mathsf{E}X(A)| > J(n+p)B + 0.25c_0x_l) \\ = \; & \sum_{l=1}^{\infty} P(Z(x_l) > \frac{J(n+p)}{np}B + \frac{1}{np}0.25c_0x_l) \leq \sum_{l=1}^{\infty} \exp(-\frac{C_7x_l^2}{np}) \\ = \; & \sum_{l=1}^{\infty} \exp(-\frac{C_8 4^{l+1} v_n^2}{np}) = \sum_{l=1}^{\infty} \exp(-C_8 4^{l+1} B^2) \leq \frac{\exp(-16C_8B^2)}{1 - \exp(-16C_8B^2)} < \epsilon \end{aligned}$$

for any $\epsilon > 0$ and sufficiently large B.

□

Appendix B. Proof of Theorems

We use the notation E_I and Var_I to denote the conditional expectation and variance given the subsamples in $S \in \{\mathcal{I}, \mathcal{I}^c\}$. We use I to denote the identity matrix.

B.1. Convergence of $\widehat{\Gamma}_S$.

Recall that columns of H_S are the eigenvectors of $\Gamma_S'\Gamma_S$. Note that $\theta_{ij} = \gamma_i' v_j + r_{ij}$. Throughout the proof, we use v_i and γ_j to denote the true values $\gamma_{i,0}$ and $v_{i,0}$ for notational simplicity.

Lemma B.1. *There are $J \times J$ rotation matrices H and B, so that for each $i \notin S$,*

$$\widehat{\gamma}_i - H'\gamma_i = B^{-1} H_S' \sum_{j=1}^{p} \varepsilon_{ij} x_{ij} v_j + o_P(1).$$

The above $o_P(.)$ is pointwise in i and in $\|.\|$. Also $\|H' - H_S^{-1}\| = O_P(J^d\omega_{np}\psi_{np}^{-1})$, and $\|B\| + \|B^{-1}\| + \|H\| + \|H^{-1}\| = O_P(1)$. Note that both B and H' depend on S.

Proof. First recall that $\widetilde{V}_S$ is estimated using subsamples in S. Note that $\|\widetilde{V}_S - V_0 H_S\| = O_P(J\omega_{np}\psi_{np}^{-1})$ for some rotation matrix H_S. For simplicity of notation, we simply write $\widetilde{v}_j'$ to denote the j th row of $\widetilde{V}_S$. For each $i \notin S$, by definition in Step 3 of the estimation algorithm, and γ_i' as the i th row of Γ_0,

$$\widehat{\gamma}_i - H_S^{-1}\gamma_i = \widehat{B}_i^{-1}\sum_{j=1}^p [y_{ij} - x_{ij}\cdot\gamma_i' H_S^{-1'}\widetilde{v}_j]x_{ij}\widetilde{v}_j,$$

and $\widehat{B}_i = \sum_{j=1}^p x_{ij}^2\widetilde{v}_j\widetilde{v}_j'$. Let $B = H_S'\sum_{j=1}^p(\mathsf{E}x_{ij}^2)v_j v_j' H_S$.

$$\begin{aligned}
\widehat{\gamma}_i - H_S^{-1}\gamma_i &= \widehat{B}_i^{-1}\sum_{j=1}^p[\varepsilon_{ij} + x_{ij}(\theta_{ij} - \gamma_i' H_S^{-1'}\widetilde{v}_j)]x_{ij}\widetilde{v}_j \\
&= B^{-1}H_S'\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j + B^{-1}\sum_{j=1}^p \widetilde{v}_j x_{ij}^2(v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i + \sum_{k=1}^4 \Delta_{i,k}, \\
\Delta_{i,1} &= (\widehat{B}_i^{-1} - B^{-1})\sum_{j=1}^p \widetilde{v}_j x_{ij}^2(v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i, \quad \Delta_{i,2} = \widehat{B}_i^{-1}\sum_{j=1}^p \varepsilon_{ij}x_{ij}(\widetilde{v}_j - H_S' v_j), \\
\Delta_{i,3} &= \widehat{B}_i^{-1}\sum_{j=1}^p x_{ij}^2 r_{ij}\widetilde{v}_j, \quad \Delta_{i,4} = (\widehat{B}_i^{-1} - B^{-1})H_S'\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j. \qquad \text{(B.1)}
\end{aligned}$$

Lemma C.2 shows that $\sum_{k=1}^4 \Delta_{i,k} = o_P(1)$. So

$$\widehat{\gamma}_i - H_S^{-1}\gamma_i = B^{-1}H_S^{-1}\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j + B^{-1}\sum_{j=1}^p \widetilde{v}_j x_{ij}^2(v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i + o_P(1). \quad \text{(B.2)}$$

To bound the second term on the right hand side, we proceed below in different cases on the DGP of x_{ij}, corresponding to conditions (ii-a)-(ii-b) in Assumption 4.3.

Under condition (ii-a) x_{ij}^2 does not vary across $i \le n$. We can write $x_{ij}^2 = x_j^2$, $B = H_S'\sum_{j=1}^p x_j^2 v_j v_j' H_S$ and

$$H' := H_S^{-1} + B^{-1}\sum_{j=1}^p \widetilde{v}_j x_j^2(v_j - \widetilde{v}_j)', \quad \|H' - H_S^{-1}\| = O_P(J^d\omega_{np}\psi_{np}^{-1}).$$

From (B.1), moving $B^{-1}\sum_{j=1}^p \widetilde{v}_j x_j^2 (v_j - \widetilde{v}_j)'\gamma_i$ to the left hand side we have:

$$\widehat{\gamma}_i - H'\gamma_i = B^{-1}H_S^{-1}\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j + \sum_{k=1}^4 \Delta_{i,k} + o_P(1) = B^{-1}H_S^{-1}\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j + o_P(1).$$

The fact that $\|B^{-1}\| = O_P(1)$ follows from the assumption that $\sum_{j=1}^p x_j^2 v_j v_j'$ is bounded away from zero.

Under condition (ii-b) x_{ij} is independent across $i \le n$ and is weakly dependent across $j \le p$. Let $B = H_S' \sum_{j=1}^p (\mathsf{E}x_{ij}^2) v_j v_j' H_S$. Define

$$\begin{aligned} \Delta_{i,5} &= B^{-1}\sum_{j=1}^p \widetilde{v}_j (x_{ij}^2 - \mathsf{E}x_{ij}^2)(v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i \\ H' &:= H_S^{-1} + B^{-1}\sum_{j=1}^p \widetilde{v}_j(\mathsf{E}x_{ij}^2)(v_j - \widetilde{v}_j)', \quad \|H' - H_S^{-1}\| = O_P(J^d\omega_{np}\psi_{np}^{-1}). \end{aligned}$$

where it is the assumption that $\mathsf{E}x_{ij}^2$ is stationary in i (does not vary across i). From (B.2), moving $B^{-1}\sum_{j=1}^p \widetilde{v}_j(\mathsf{E}x_{ij}^2)(v_j - \widetilde{v}_j)'\gamma_i$ to the left hand side we have:

$$\widehat{\gamma}_i - H'\gamma_i = B^{-1}H_S^{-1}\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j + \sum_{k=1}^5 \Delta_{i,k}.$$

Lemma C.2 bounds $\Delta_{i,1} \sim \Delta_{i,5}$.

Under condition (ii-c) In this case we focus on $x_{ij} \in \{0,1\}$. Let $\mathcal{B}_i = \{j : x_{ij} = 1\}$. Then $B = H_S' \sum_{j\in\bar{\mathcal{B}}}^p v_j v_j' H_S$. The fact $\|B^{-1}\| = O_P(1)$ follows from the assumption. In addition, let

$$\Delta_{i,6} := B^{-1}\sum_{j=1}^p \widetilde{v}_j x_{ij}^2 (v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i - B^{-1}\sum_{j\in\bar{\mathcal{B}}} \widetilde{v}_j (v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i.$$

We define

$$H' := H_S^{-1} + B^{-1}\sum_{j\in\bar{\mathcal{B}}} \widetilde{v}_j (v_j - \widetilde{v}_j)', \quad \|H' - H_S^{-1}\| = O_P(J^d\omega_{np}\psi_{np}^{-1}).$$

From (B.1), moving $B^{-1}\sum_{j\in\bar{\mathcal{B}}} \widetilde{v}_j (v_j - \widetilde{v}_j)'\gamma_i$ to the left hand side we have:

$$\widehat{\gamma}_i - H'\gamma_i = B^{-1}H_S^{-1}\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j + \sum_{k=1}^4 \Delta_{i,k} + \Delta_{i,6} = B^{-1}H_S^{-1}\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j + o_P(1).$$

Lemma C.2 shows that in this case $\Delta_{i,1} \sim \Delta_{i,4}, \Delta_{i,6} = o_P(1)$.

□

B.2. **Convergence of $\widehat{V}_{\mathcal{I}}$.** Let $L_{j,\mathcal{I}} = H'\sum_{i\notin\mathcal{I}} x_{ij}^2\gamma_i\gamma_i'H$.

As the vector g can be either sparse (e.g., $g = (1,0...0)'$) or dense (e.g., $g = (1,...,1)'/p$), sometimes we shall use the following inequality to bound terms of the form: $\sum_{j=1}^p X_j L_{j,\mathcal{I}} g_j$:

Let $\mathcal{G} = \{j = 1,...,p : g_j \neq 0\}$ and $|\mathcal{G}|$ be its cardinality.

$$\sum_{j=1}^p X_j L_{j,\mathcal{I}} g_j \leq \sqrt{\sum_{j\in\mathcal{G}} L_{j,\mathcal{I}}^2}\sqrt{\sum_{j=1}^p X_j^2 g_j^2},$$

which is shaper than the usual bound $L := \sqrt{\sum_{j=1}^p L_{j,\mathcal{I}}^2}\sqrt{\sum_{j=1}^p X_j^2 g_j^2}$ when g is sparse, and reaches about the same order of L when g is dense.

Lemma B.2. *Given a $p\times 1$ fixed vector g, for the rotation matrix H in Lemma B.1 (which depends on the sample $\mathcal{I}$),*

$$\widehat{V}_{\mathcal{I}}'g - H^{-1}V_0'g = \sum_{j=1}^p\sum_{i\notin\mathcal{I}} L_{j,\mathcal{I}}^{-1}H'\gamma_i\varepsilon_{ij}x_{ij}g_j + O_P(\xi_{np})$$

where ξ_{np} is defined in (B.5). The above convergence is in $\|.\|$.

Proof. Write $\widehat{V}_{\mathcal{I}} = (\widehat{v}_1,...,\widehat{v}_p)'$. Then for $\widehat{L}_{j,\mathcal{I}} = \sum_{i\notin\mathcal{I}} x_{ij}^2\widehat{\gamma}_i\widehat{\gamma}_i'$,

$$\widehat{v}_j - H^{-1}v_j = \widehat{L}_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}[y_{ij} - x_{ij}\cdot\widehat{\gamma}_i'H^{-1}v_j]x_{ij}\widehat{\gamma}_i.$$

Therefore,

$$\widehat{V}_{\mathcal{I}}'g - H^{-1}V_0'g = \sum_{j=1}^p(\widehat{v}_j - H^{-1}v_j)g_j = \sum_{j=1}^p\sum_{i\notin\mathcal{I}} L_{j,\mathcal{I}}^{-1}H'\gamma_i\varepsilon_{ij}x_{ij}g_j + \sum_{d=1}^4\Delta_d, \quad \text{(B.3)}$$

where

$$\Delta_1 = \sum_{j=1}^p\widehat{L}_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}} x_{ij}^2 r_{ij}\widehat{\gamma}_i g_j, \quad \Delta_2 = \sum_{j=1}^p[\widehat{L}_{j,\mathcal{I}}^{-1} - L_{j,\mathcal{I}}^{-1}]H'\sum_{i\notin\mathcal{I}}\gamma_i\varepsilon_{ij}x_{ij}g_j,$$

$$\Delta_3 = \sum_{j=1}^p\widehat{L}_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}(\widehat{\gamma}_i - H'\gamma_i)\varepsilon_{ij}x_{ij}g_j, \quad \Delta_4 = \sum_{j=1}^p\widehat{L}_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}} x_{ij}^2(\gamma_i'H - \widehat{\gamma}_i')H^{-1}v_j\widehat{\gamma}_i g_j.$$

(B.4)

By Lemmas C.4-C.6,

$$
\begin{aligned}
\sum_{d=1}^{4}\|\Delta_d\|^2 &= O_P(\xi_{np}^2)\\
\xi_{np}^2 &:= \left(J^{2+2d+4b}\omega_{np}^2 \mathsf{E}\max_{j\leq p}\|v_j\|^2 + npJ^{2+2b}\max_{ij} r_{ij}^2 + \psi_{np}^{-2}J^{2+4d+4b}\omega_{np}^4\right)\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2\psi_{np}^{-2}\\
&+\left(\omega_{np}^4\psi_{np}^{-2}J^{3+4d+8b}+J^{2+2b}\right)\sum_{j=1}^{p}\|v_jg_j\|^2|\mathcal{G}|\psi_{np}^{-2}\\
&+O_P\left(\omega_{np}^4\psi_{np}^{-4}nJ^{2+4d+6b}+\omega_{np}^2\psi_{np}^{-2}J^{2+2d+6b}+J^{2b+1}\max_{ij}|r_{ij}|^2 n\right)\psi_{np}^{-2}\|g\|^2|\mathcal{G}|\\
&+O_P\left(1+nJ^{2d+1}\omega_{np}^2\psi_{np}^{-2}+nJ\sum_{j=1}^{p}\mathsf{E}\|v_j\|^4\right)\omega_{np}^2\psi_{np}^{-4}\|g\|^2J^{1+2d+2b}. \qquad \text{(B.5)}
\end{aligned}
$$

where

$$\mu_{np}^2 := np^{-2}\omega_{np}^2\psi_{np}^{-4}\|g\|^2J^{4+2d+2b}[\sum_{\mathcal{B}_i\Delta\bar{\mathcal{B}}}1]^2.$$

□

B.3. **Asymptotic normality of $\widehat{\theta}_i'g$.** The asymptotic normality follows from Proposition B.1 below, whose condition (B.6) is verified by Lemma C.7 in the case of either sparse or dense g, and the primitive condition in Assumption 4.6.

Proposition B.1. *Let ξ_{np} be as defined in (B.5). Suppose*

$$|r_i'g| + J^b\|u_i\|\psi_{np}\xi_{np} = o_P(p^{-1/2}+\|u_i\|\|g\|). \qquad \text{(B.6)}$$

Then for a fixed $i\leq n$,

$$\frac{\widehat{\theta}_i'g-\theta_i'g}{\sqrt{s_{np,1}^2+s_{np,2}^2}} \to^d N(0,1),$$

where, with $L_j=\sum_{i=1}^n x_{ij}^2\gamma_i\gamma_i'$ and $\bar{B}=\sum_{j=1}^p(\mathsf{E}x_{ij}^2)v_jv_j'$,

$$
\begin{aligned}
s_{np,1}^2 &:= \sum_{j=1}^{p}\sum_{t=1}^{n}\mathsf{Var}(\varepsilon_{tj}|\Theta,X)[\gamma_i'L_j^{-1}\gamma_t]^2x_{tj}^2g_j^2\\
s_{np,2}^2 &:= \sum_{j=1}^{p}\mathsf{Var}(\varepsilon_{ij}|\Theta,X)x_{ij}^2[v_j'B^{-1}V_0'g]^2.
\end{aligned}
$$

Proof. Note that for fixed $i\leq n$, $\widehat{\theta}_{\mathcal{I},i}'=\widehat{\gamma}_i'\widehat{V}_{\mathcal{I}}'$ and $\theta_i'=\gamma_i'V_0'+r_i'$, where r_i' denotes the i th row of R, the low-rank approximation error. By Lemmas B.1, B.2, for

$\bar{L}_{j,\mathcal{I}} = \sum_{i\notin\mathcal{I}} x_{ij}^2\gamma_i\gamma_i'$,

$$\begin{aligned}(\widehat{\theta}_{\mathcal{I},i}-\theta_i)'g &= (\widehat{\gamma}_i'\widehat{V}_{\mathcal{I}}'-\gamma_i'V_0')g - r_i'g\\ &= \gamma_i'\sum_{j=1}^p\sum_{t\notin\mathcal{I}}\bar{L}_{j,\mathcal{I}}^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j+\sum_{j=1}^p\varepsilon_{ij}x_{ij}v_j'\bar{B}^{-1}V_0'g+\mathcal{R}\\ &= 2\gamma_i'\sum_{j=1}^p\sum_{t\notin\mathcal{I}}L_j^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j+\sum_{j=1}^p\varepsilon_{ij}x_{ij}v_j'\bar{B}^{-1}V_0'g+\mathcal{R}\\ \mathcal{R} &:= o_P(p^{-1/2}+\|u_i\|\|g\|)+O_P(J^{1+b}\|g\|\psi_{np}^{-1})+\xi_{np}O_P(J^b\|u_i\|\psi_{np}+J^{1/2})-r_i'g\end{aligned}$$

where we used $\|H'-H_S^{-1}\|=(J^d\omega_{np}\psi_{np}^{-1})$ and $L_j=\sum_{i=1}^n x_{ij}^2\gamma_i\gamma_i'$

$$\|\sum_{j=1}^p\sum_{t\notin\mathcal{I}}L_{j,\mathcal{I}}^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j\|=O_P(J^{1/2+b}\|g\|\psi_{np}^{-1}),\quad \|\sum_{j=1}^p\varepsilon_{ij}x_{ij}v_j'B^{-1}\|=O_P(\sqrt{J}).$$

$$\begin{aligned}&|\gamma_i'\sum_{j=1}^p\sum_{t\notin\mathcal{I}}(2L_j^{-1}-\bar{L}_{j,\mathcal{I}}^{-1})\gamma_t\varepsilon_{tj}x_{tj}g_j| = \sqrt{O_P(J)\|u_i\|^2\sum_{j=1}^p\|\frac{1}{2}L_j-\bar{L}_{j,\mathcal{I}}\|^2g_j^2\psi_{np}^{-4}}\\ =\ & o_P(\|u_i\|\|g\|).\end{aligned}$$

Similarly, once we switch $\mathcal{I}$ and $\mathcal{I}^c$, $(\widehat{\theta}_{\mathcal{I}^c,i}-\theta_i)g$ has a similar expansion. Hence

$$\begin{aligned}\widehat{\theta}_i'g-\theta_i'g &= \gamma_i'\sum_{j=1}^p\sum_{t\notin\mathcal{I}}L_j^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j+\gamma_i'\sum_{j=1}^p\sum_{t\notin\mathcal{I}^c}L_j^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j\\ &\quad+\sum_{j=1}^p\varepsilon_{ij}x_{ij}v_j'\bar{B}^{-1}V_0'g+\mathcal{R}\\ &= \gamma_i'\sum_{j=1}^p\sum_{t=1}^n L_j^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j+\sum_{j=1}^p\varepsilon_{ij}x_{ij}v_j'\bar{B}^{-1}V_0'g+\mathcal{R}+O_P(\|u_i\|^2\|g\|)\end{aligned}$$

where $\sum_{t\notin\mathcal{I}}+\sum_{t\notin\mathcal{I}^c}=\sum_{t=1}^n+1\{t=i\}$, and $\gamma_i'\sum_{j=1}^p 1\{t=i\}L_j^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j = O_P(\|u_i\|^2\|g\|)$. Also note that $\|V_0'g\|\asymp p^{-1/2}$ and $J^{1+b}\|g\|\psi_{np}^{-1}=o(p^{-1/2})$.

Next, we verify the Lindeberg condition for the first two leading terms of $\widehat{\theta}_i'g-\theta_i'g$. First, we emphasize that the defined $\mathcal{X}_{t,1}$ and $\mathcal{X}_{j,2}$ (as defined below) do not depend on the sample $\mathcal{I}$ or $\mathcal{I}^c$. Let

$$\begin{aligned}\mathcal{X}_{t,1} &:= \gamma_i'\sum_{j=1}^p L_j^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j,\quad \mathcal{X}_{j,2}=\varepsilon_{ij}x_{ij}v_j'\bar{B}^{-1}V_0'g\\ s_{np,1}^2 &:= \mathsf{Var}(\sum_{t=1}^n\mathcal{X}_{t,1}|\Theta,X)=\sum_{j=1}^p\sum_{t=1}^n\mathsf{Var}(\varepsilon_{tj}|\Theta,X)[\gamma_i'L_j^{-1}\gamma_t]^2x_{tj}^2g_j^2\end{aligned}$$

$$
\begin{aligned}
&\geq c\min_{j\leq p}\psi_{\min}(\sum_{t=1}^n x_{tj}^2\gamma_t\gamma_t')\psi_{\min}(L_j^{-2})\|g\|^2\|\gamma_i\|^2\\
&\geq c\min_{j\leq p}\psi_{\min}(\sum_{t=1}^n x_{tj}^2u_tu_t')\psi_{\max}^{-2}(\sum_{t=1}^n x_{tj}^2u_tu_t')\|g\|^2\|u_i\|^2\geq c\|u_i\|^2\|g\|^2.\\
s_{np,2}^2 &:= \mathsf{Var}(\sum_{j=1}^p \mathcal{X}_{j,2}|\Theta,X)=\sum_{j=1}^p\mathsf{Var}(\varepsilon_{ij}|\Theta,X)x_{ij}^2[v_j'\bar{B}^{-1}V_0'g]^2\\
&\geq c\|g'V_0\|^2\min_i\psi_{\min}(\sum_{j=1}^p x_{ij}^2v_jv_j')\psi_{\max}^{-2}(\bar{B})\geq cp^{-1}.
\end{aligned}
$$

Write $\mathsf{E}_|(.) := \mathsf{E}(.|\Theta,X)$. We first bound $\sum_{t=1}^n\mathsf{E}_|\mathcal{X}_{t,1}^4$ and $\sum_{j=1}^p\mathsf{E}_|\mathcal{X}_{j,2}^4$. Write $\mathcal{Y}_{t,j} := \gamma_i'L_j^{-1}\gamma_t\varepsilon_{tj}x_{tj}g_j$, then $\mathcal{X}_{t,1}=\sum_{j=1}^p\mathcal{Y}_{t,j}$. Due to conditional independence,

$$
\begin{aligned}
\sum_{t=1}^n\mathsf{E}_|(\mathcal{X}_{t,1}^4) &= \sum_{t=1}^n\mathsf{E}_|(\sum_{j=1}^p\mathcal{Y}_{t,j})^4\leq\sum_{t=1}^n\sum_{j=1}^p\mathsf{E}_|\mathcal{Y}_{t,j}^4+3\sum_{t=1}^n[\sum_{j=1}^p\mathsf{E}_|\mathcal{Y}_{t,j}^2]^2\\
&\leq C\|u_i\|^4\sum_{t=1}^n\|u_t\|^4[\sum_{j=1}^p g_j^4+\|g\|^4]\\
\sum_{j=1}^p\mathsf{E}_|(\mathcal{X}_{j,2}^4) &\leq C\sum_{j=1}^p\|v_j\|^4\|V_0'g\|^4\leq C\sum_{j=1}^p\|v_j\|^4p^{-2}.
\end{aligned}
$$

Now for any $\epsilon>0$, by Cauchy-Schwarz and Markov inequalities and

$$
\sum_{j=1}^p\|v_j\|^4=o_P(1),\quad \sum_{i=1}^n\|u_i\|^4=o_P(1)
$$

we have

$$
\begin{aligned}
\frac{1}{s_{np,1}^2}\sum_{t=1}^n\mathsf{E}_|\mathcal{X}_{t,1}^2 1\{|\mathcal{X}_{t,1}|>\epsilon s_{np,1}\} &\leq \frac{1}{s_{np,1}^2\epsilon}\sqrt{\sum_{t=1}^n\mathsf{E}_|\mathcal{X}_{t,1}^4}\\
&\leq \frac{C}{\epsilon}\sqrt{\sum_{t=1}^n\|u_t\|^4}\sqrt{\frac{\sum_{j=1}^p g_j^4}{\|g\|^4}+1}=o_P(1).\\
\frac{1}{s_{np,2}^2}\sum_{j=1}^p\mathsf{E}_|\mathcal{X}_{j,1}^2 1\{|\mathcal{X}_{j,1}|>\epsilon s_{np,2}\} &\leq Cp\sqrt{\sum_{j=1}^p\|v_j\|^4p^{-2}}=o_P(1).
\end{aligned}
$$

In addition,

$$
\frac{1}{\|u_i\|\|g\|p^{-1/2}}\mathsf{Cov}(\sum_{t=1}^n\mathcal{X}_{t,1},\sum_{j=1}^p\mathcal{X}_{j,2}|\Theta,X)\leq C\|V_0\|_F\|u_i\|=o_P(1).
$$

Also, $\mathcal{R} = o_P(p^{-1/2} + \|u_i\|\|g\|)$, given condition (B.6). Thus we have achieved

$$\widehat{\theta}_i' g - \theta_i' g = \sum_{t=1}^{n} \mathcal{X}_{t,1} + \sum_{j=1}^{p} \mathcal{X}_{j,2} + o_P(p^{-1/2} + \|u_i\|\|g\|) \tag{B.7}$$

where $\sum_{t=1}^n \mathcal{X}_{t,1}/s_{np,1}$ and $\sum_{j=1}^p \mathcal{X}_{j,2}/s_{np,1}$ are asymptotically normal and independent. Hence we can apply the argument of the proof of Theorem 3 in [1] to conclude that, conditionally on (Θ, X),

$$\frac{\sum_{t=1}^n \mathcal{X}_{t,1} + \sum_{j=1}^p \mathcal{X}_{j,2}}{\sqrt{s_{np,1}^2 + s_{np,2}^2}} \to^d N(0,1).$$

Also, $o_P(p^{-1/2} + \|u_i\|\|g\|)/\sqrt{s_{np,1}^2 + s_{np,2}^2} = o_P(1)$. This implies conditionally on (Θ, X),

$$\mathcal{Z} := \frac{\widehat{\theta}_i' g - \theta_i' g}{\sqrt{s_{np,1}^2 + s_{np,2}^2}} \to^d N(0,1).$$

It remains to argue that the conditional weak convergence holds unconditionally. Let Φ denote the standard normal cumulative distribution function. Fix any $x \in \mathbb{R}$, let $f(\Theta, X) := P(\mathcal{Z} < x|\Theta, X)$. Then $f(\Theta, X) \to \Phi(x)$, pointwise in (Θ, X). By the dominated convergence theorem and that f is dominated by 1,

$$P(\mathcal{Z} < x) = \mathsf{E} f(\Theta, X) \to \Phi(x).$$

This holds for any x, thus proves the weak convergence of $\mathcal{Z}$ unconditionally. □

B.4. **Proof of Theorem 4.2: efficiency bound.**

Lemma B.3. *Let Z be a random variable with sub-Gaussian norm bounded by r. If $r \leq 1/3$, then $|\mathsf{E}\exp(Z) - 1| \leq cr$, where $c > 0$ is an absolute constant.*

Proof. By assumption, $[\mathsf{E}|Z|^j]^{1/j}/\sqrt{j} \leq r$ for any $j \geq 1$. Thus, $\mathsf{E}|Z|^j \leq r^j j^{j/2}$ for any $j \geq 1$. By Taylor's expansion, $\exp(Z) - 1 = \sum_{j=1}^\infty \frac{Z^j}{j!}$. Hence,

$$\begin{aligned} \mathsf{E}\left|\exp(Z) - 1\right| &= \sum_{j=1}^\infty \frac{\mathsf{E}|Z|^j}{j!} \leq \sum_{j=1}^\infty \frac{r^j j^{j/2}}{j!} \overset{\text{(i)}}{\leq} \sum_{j=1}^\infty \frac{r^j j^{j/2}}{(j/e)^j} \\ &= \sum_{j=1}^\infty (er)^j j^{-j/2} \leq \sum_{j=1}^\infty (er)^j \overset{\text{(ii)}}{=} \frac{er}{1-er} \overset{\text{(iii)}}{\leq} \frac{er}{1-e/3}, \end{aligned}$$

where (i) follows by $j! \geq (j/e)^j$ and (ii) and (iii) follow by $r \leq 1/3 < 1/e$. Thus, the result holds with $c = e/(1-e/3)$. $\square$

Proof of Theorem 4.2. Let $\Theta = \Gamma V'$ be an arbitrary point in $\mathcal{M}$, where $\Gamma \in \mathbb{R}^{n\times J}$ and $V \in \mathbb{R}^{p\times J}$. We partition $\Gamma = \begin{pmatrix} \Gamma_1' \\ \Gamma_{-1}' \end{pmatrix} \in \mathbb{R}^{n\times J}$ with $\Gamma_1 \in \mathbb{R}^J$ and $\Gamma_{-1} \in \mathbb{R}^{J\times(n-1)}$. By assumption, σ is bounded. Since X_{1j} is bounded, we have that μ_f is also bounded. It follows that $\kappa_1 \leq \sqrt{\mu_f}/(2\sigma) \leq \kappa_2$ for some constants $\kappa_1, \kappa_2 > 0$.

Consider $\tilde{\Theta} = \tilde{\Gamma} V'$, where $\tilde{\Gamma} = \begin{pmatrix} \tilde{\Gamma}_1' \\ \Gamma_{-1}' \end{pmatrix} \in \mathbb{R}^{n\times J}$ with $\tilde{\Gamma}_1 = \Gamma_1 + qV'g$ and $q = t/\|V'g\|$. Here, $t \in (0, \kappa_1)$ is a fixed but arbitrary constant. Recall $s_*^2(\Theta, f, \sigma) = \sigma^2 \mu_f^{-1} \|V'g\|^2$ does not depend on Γ. So $s_*^2(\tilde{\Theta}, f, \sigma) = s_*^2(\Theta, f, \sigma)$. Let $\Delta = \tilde{\Theta} - \Theta = (\tilde{\Gamma} - \Gamma)V'$. Notice that

$$\Delta = \begin{pmatrix} q(V'g)'V' \\ 0 \end{pmatrix}$$

and

$$\|\Delta\|_F^2 = q^2 \|(V'g)'V'\|_F^2 = q^2 \|V'g\|^2 = t^2.$$

Let $\phi_{(\Theta,f,\sigma)}(Y,X)$ and $\phi_{(\tilde{\Theta},f,\sigma)}(Y,X)$ be the density of the data (Y,X) under (Θ, f, σ) and $(\tilde{\Theta}, f, \sigma)$, respectively; we notice that

$$\phi_{(\Theta,f,\sigma)}(Y,X) = (2\pi)^{-np/2} \sigma^{-np} \exp\left(-\frac{1}{2\sigma^2}\|Y - X \circ \Theta\|_F^2\right) \cdot \prod_{i=1}^n \prod_{j=1}^p f(X_{ij}).$$

Then by the asymptotic unbiasedness, we have

$$\begin{aligned} \mathsf{E}_{(\tilde{\Theta},f,\sigma)} T(Y,X) - \mathsf{E}_{(\Theta,f,\sigma)} T(Y,X) &= h(\tilde{\Theta}) - h(\Theta) + o(s_*(\Theta,f,\sigma)) + o(s_*(\tilde{\Theta},f,\sigma)) \\ &\overset{\text{(i)}}{=} q\|V'g\|^2 + o(s_*(\Theta,f,\sigma)) \\ &= t\|V'g\| + o(s_*(\Theta,f,\sigma)), \end{aligned}$$

where (i) follows by $s_*^2(\tilde{\Theta}, f, \sigma) = s_*^2(\Theta, f, \sigma)$. On the other hand,

$$\begin{aligned}
&\mathsf{E}_{(\tilde{\Theta},f,\sigma)}T(Y,X) - \mathsf{E}_{(\Theta,f,\sigma)}T(Y,X) \\
&= \mathsf{E}_{(\Theta,f,\sigma)}[T(Y,X) - h(\Theta)]\left(\frac{\phi_{(\tilde{\Theta},f,\sigma)}(Y,X)}{\phi_{(\Theta,f,\sigma)}(Y,X)} - 1\right) \\
&\leq \sqrt{\mathsf{Var}_{(\Theta,f,\sigma)}[T(Y,X)]} \times \sqrt{\mathsf{E}_{(\Theta,f,\sigma)}\left(\frac{\phi_{(\tilde{\Theta},f,\sigma)}(Y,X)}{\phi_{(\Theta,f,\sigma)}(Y,X)} - 1\right)^2} \\
&= \sqrt{\mathsf{Var}_{(\Theta,f,\sigma)}[T(Y,X)]} \times \sqrt{\mathsf{E}_{(\Theta,f,\sigma)}\left(\frac{\phi_{(\tilde{\Theta},f,\sigma)}(Y,X)}{\phi_{(\Theta,f,\sigma)}(Y,X)}\right)^2 - 1}.
\end{aligned}$$

It follows that

$$\mathsf{Var}_{(\Theta,f,\sigma)}[T(Y,X)] \geq \frac{\left[t\|V'g\|^2 + o(s_*(\Theta,f,\sigma))\right]^2}{\mathsf{E}_{(\Theta,f,\sigma)}\left(\frac{\phi_{(\tilde{\Theta},f,\sigma)}(Y,X)}{\phi_{(\Theta,f,\sigma)}(Y,X)}\right)^2 - 1}. \tag{B.8}$$

The rest of the proof proceeds in two steps.

Step 1: compute $\mathsf{E}_{(\Theta,f,\sigma)}\left(\frac{\phi_{(\tilde{\Theta},f,\sigma)}(Y,X)}{\phi_{(\Theta,f,\sigma)}(Y,X)}\right)^2$. Under (Θ, f, σ), $\varepsilon = Y - X \circ \Theta$ is a matrix with entries following $N(0, \sigma^2)$. By the explicit formulas above on $\phi_{(\Theta,f,\sigma)}(Y,X)$ and $\phi_{(\tilde{\Theta},f,\sigma)}(Y,X)$, we have that

$$\begin{aligned}
\mathsf{E}_{(\Theta,f,\sigma)}\left(\frac{\phi_{(\tilde{\Theta},f,\sigma)}(Y,X)}{\phi_{(\Theta,f,\sigma)}(Y,X)}\right)^2 &= \mathsf{E}_{(\Theta,f,\sigma)}\exp\left(-\frac{1}{\sigma^2}\|Y - X\circ\tilde{\Theta}\|_F^2 + \frac{1}{\sigma^2}\|Y - X\circ\Theta\|_F^2\right) \\
&\overset{\text{(i)}}{=} \mathsf{E}_{(\Theta,f,\sigma)}\exp\left(\sigma^{-2}\sum_{j=1}^p X_{1j}^2\Delta_{1j}^2\right) \\
&= \exp\left(\sigma^{-2}\sum_{j=1}^p \mu_f\Delta_{1j}^2\right) \times \mathsf{E}_{(\Theta,f,\sigma)}\exp\left(\sigma^{-2}\sum_{j=1}^p [X_{1j}^2 - \mu_f]\Delta_{1j}^2\right),
\end{aligned}$$

where (i) follows by the fact that under (Θ, f, σ), $\varepsilon_{i,j}$ is i.i.d $N(0, \sigma^2)$ conditional on X.

Notice that under (Θ, f, σ), X_{1j}^2 is i.i.d, bounded and has mean μ_f. By Hoeffding's inequality and equivalent bounds for sub-Gaussian distributions (e.g., Proposition 5.10 and Lemma 5.5 in Vershynin [6]), the variable $\sigma^{-2}\sum_{j=1}^p[X_{1j}^2 - \mu_f]\Delta_{1j}^2$ has sub-Gaussian norm bounded by $\kappa_3\sigma^{-2}\sqrt{\sum_{j=1}^p \Delta_{1j}^4}$, where $\kappa_3 > 0$ is a constant that only depends on the bound of X_{1j}^2. We observe that

$$\sum_{j=1}^{p}\Delta_{1j}^4 = \sum_{j=1}^{p}[q(V'g)'v_j]^4 = q^4\left(\max_{1\le j\le p}[(V'g)'v_j]^2\right)\sum_{j=1}^{p}[(V'g)'v_j]^2$$
$$= q^4\left(\max_{1\le j\le p}[(V'g)'v_j]^2\right)\|V'g\|^2 \le q^4\left(\|V'g\|^2\max_{1\le j\le p}\|v_j\|^2\right)\|V'g\|^2$$
$$\lesssim q^4\|V'g\|^4(Jp^{-1}) \lesssim t^4Jp^{-1} = o(1).$$

By Lemma B.3, we have

$$\mathsf{E}_{(\Theta,f,\sigma)}\exp\left(\sigma^{-2}\sum_{i=1}^{n}\sum_{j=1}^{p}(X_{ij}^2-\mu_f)\Delta_{ij}^2\right) = 1+o(1).$$

Moreover,

$$\sigma^{-2}\sum_{j=1}^{p}\mu_f\Delta_{1j}^2 = \mu_f\sigma^{-2}q^2(V'g)'V'V(V'g) = \mu_f\sigma^{-2}q^2\|V'g\|^2 = \mu_f\sigma^{-2}t^2.$$

By the above three displays, we have that

$$\mathsf{E}_{(\Theta,f,\sigma)}\left(\frac{\phi_{(\tilde\Theta,f,\sigma)}(Y,X)}{\phi_{(\Theta,f,\sigma)}(Y,X)}\right)^2 = \exp(\mu_f\sigma^{-2}t^2)\left(1+o(1)\right). \tag{B.9}$$

Step 2: derive the final result.

Since t is fixed and $s_*(\Theta,f,\sigma)\asymp\|V'g\|$, it follows by (B.8) and (B.9) that

$$\mathsf{Var}_{(\Theta,f,\sigma)}[T(Y,X)] \ge \frac{\left[t\|V'g\|+o(s_*(\Theta,f,\sigma))\right]^2}{\mathsf{E}_{(\Theta,f,\sigma)}\left(\frac{\phi_{(\tilde\Theta,f,\sigma)}(Y,X)}{\phi_{(\Theta,f,\sigma)}(Y,X)}\right)^2-1}$$
$$= \frac{\left[t\|V'g\|+o(s_*(\Theta,f,\sigma))\right]^2}{\exp(\mu_f\sigma^{-2}t^2)\left(1+o(1)\right)-1}$$
$$= \frac{t^2\|V'g\|^2+o(s_*^2(\Theta,f,\sigma))}{\exp(\mu_f\sigma^{-2}t^2)-1+o(1)\cdot\exp(\mu_f\sigma^{-2}t^2)}$$
$$= \frac{t^2\|V'g\|^2(1+o(1))}{\exp(\mu_f\sigma^{-2}t^2)-1+o(1)\cdot\exp(\mu_f\sigma^{-2}t^2)}.$$

Consider $r(a)=\exp(a)$. Notice that the second derivative is $r''(a)=\exp(a)$, which is increasing in a. We have that for any $a>0$, $|r(a)-r(0)-r'(0)a|\le\exp(a)a^2/2$. Thus, for any $a>0$,

$$\exp(a)-1\le r'(0)a+\exp(a)a^2/2 = a+\exp(a)a^2/2.$$

Since $t \in (0, \kappa_1)$ with $\kappa_1 \leq \sqrt{\mu_f}/(2\sigma)$, we have $\mu_f \sigma^{-2} t^2 \in (0, 1/4)$ and thus

$$\exp(\mu_f \sigma^{-2} t^2) - 1 \leq \mu_f \sigma^{-2} t^2 + \exp(1/4) \mu_f^2 \sigma^{-4} t^4 / 2.$$

Therefore, we have

$$\begin{aligned} \mathsf{Var}_{(\Theta,f,\sigma)}[T(Y,X)] &\geq \frac{t^2 \|V'g\|^2 (1+o(1))}{\exp(\mu_f \sigma^{-2} t^2) - 1 + o(1) \cdot \exp(\mu_f \sigma^{-2} t^2)} \\ &\geq \frac{t^2 \|V'g\|^2 (1+o(1))}{\mu_f \sigma^{-2} t^2 + \exp(1/4) \mu_f^2 \sigma^{-4} t^4 / 2 + o(1) \cdot \exp(1/4)} \\ &= \frac{\|V'g\|^2 (1+o(1))}{\left[\mu_f \sigma^{-2} + \exp(1/4) \mu_f^2 \sigma^{-4} t^2 / 2\right] (1+o(1))}. \end{aligned}$$

In other words,

$$\begin{aligned} \frac{\mathsf{Var}_{(\Theta,f,\sigma)}[T(Y,X)]}{s_*^2(\Theta,f,\sigma)} &\geq \frac{1+o(1)}{\sigma^2 \mu_f^{-1} \left[\mu_f \sigma^{-2} + \exp(1/4) \mu_f^2 \sigma^{-4} t^2 / 2\right] (1+o(1))} \\ &\overset{\text{(i)}}{\geq} \frac{1+o(1)}{\left[1 + 8 \exp(1/4) \kappa_2^4 t^2\right] (1+o(1))}, \end{aligned}$$

where (i) follows by $\mu_f \sigma^{-2} \leq 4\kappa_2^2$. We take the limit and obtain

$$\liminf_{n,p\to\infty} \frac{\mathsf{Var}_{(\Theta,f,\sigma)}[T(Y,X)]}{s_*^2(\Theta,f,\sigma)} \geq \frac{1}{1 + 8\exp(1/4)\kappa_2^4 t^2}.$$

Notice that the left-hand side does not depend on the choice of t. Since the above bound holds for any $t \in (0, \kappa_1)$, we can choose small t and obtain

$$\liminf_{n,p\to\infty} \frac{\mathsf{Var}_{(\Theta,f,\sigma)}[T(Y,X)]}{s_*^2(\Theta,f,\sigma)} \geq 1.$$

The proof is complete. □

B.5. **Proof of Theorem 4.3: minimax rate.**

Proof. We provide proofs for both cases: sparse g ($g_1 = 1$ and $g_j = 0$ for $j \geq 2$) as well as dense g. The arguments and notations for these two cases are independent.

Case 1: sparse g

Fix any $\Theta_* \in \mathcal{S}$ with entries $\{\theta_{ij,*}\}$ such that $\mathsf{rank}\Theta_* \leq J-1$ and $\max_{1\leq i\leq n} \max_{1\leq j\leq p} |\theta_{ij,*}| \leq c_1/2$. (This is always feasible because we can simply choose Θ_* to be the zero matrix.)

Let $\kappa > 0$ be a constant satisfying the following

$$\kappa \leq \min \left\{ c_1/2,\ c_3/\sqrt{1-c_2} \right\}, \tag{B.10}$$

Let Θ_{**} be a matrix with entries $\{\theta_{ij,**}\}$ defined as follows

$$\theta_{ij,**} = \begin{cases} \theta_{ij,*} + \kappa & \text{if } (i,j) = 1 \\ \theta_{1j,*} & \text{otherwise.} \end{cases}$$

Since Θ_* and Θ_{**} only differ in the $(1,1)$ entry, it follows that $\mathsf{rank}\Theta_{**} \leq \mathsf{rank}\Theta_* + 1 \leq J-1+1 = J$. Since $|\theta_{11,**}| = |\theta_{11,*}+\kappa| \leq |\theta_{11,*}|+\kappa \leq c_1/2+\kappa \leq c_1$. Therefore, $\Theta_{**} \in \mathcal{S}$.

We now compare the Kullback–Leibler divergence bewteen $P_{(\Theta_*,\rho,\sigma)}$ and $P_{(\Theta_{**},\rho,\sigma)}$:

$$KL(P_{(\Theta_*,\rho,\sigma)}, P_{(\Theta_{**},\rho,\sigma)}) = \int \left(\log \frac{dP_{(\Theta_*,\rho,\sigma)}}{dP_{(\Theta_{**},\rho,\sigma)}} \right) dP_{(\Theta_*,\rho,\sigma)}.$$

By the formula of Gaussian densities, we have that

$$\frac{dP_{(\Theta_*,\rho,\sigma)}}{dP_{(\Theta_{**},\rho,\sigma)}} = \frac{\exp\left(-\frac{1}{2}(y_{11} - \theta_{1j,*}x_{11})^2 \sigma_{11}^{-2}\right)}{\exp\left(-\frac{1}{2}(y_{11} - \theta_{1j,**}x_{11})^2 \sigma_{11}^{-2}\right)}.$$

Therefore,

$$\begin{aligned} KL(P_{(\Theta_*,\rho,\sigma)}, P_{(\Theta_{**},\rho,\sigma)}) &= \int \left(\log \frac{dP_{(\Theta_*,\rho,\sigma)}}{dP_{(\Theta_{**},\rho,\sigma)}} \right) dP_{(\Theta_*,\rho,\sigma)} \\ &= \mathsf{E}_{(\Theta_*,\rho,\sigma)} \left(-\frac{1}{2} \sum_{j=1}^{p} \left[(y_{11} - \theta_{11,*}x_{11})^2 - (y_{11} - \theta_{11,**}x_{11})^2 \right] \sigma_{1j}^{-2} \right) \\ &= \mathsf{E}_{(\Theta_*,\rho,\sigma)} \left(\left[\frac{1}{2}\kappa^2 x_{11} - (y_{11} - \theta_{11,*}x_{11})\kappa x_{11} \right] \sigma_{11}^{-2} \right) \\ &= \frac{1}{2}\kappa^2 \rho_1 \sigma_{11}^{-2} \leq \frac{1}{2}\kappa^2 (1-c_2) c_3^{-2} \overset{\text{(i)}}{\leq} 1/2, \end{aligned}$$

where (i) follows by (B.10).

We notice that $|h(\Theta_*) - h(\Theta_{**})| = \kappa$. By Theorem 2.2 in Tsybakov [5] and Equation (2.9) therein, we have that

$$\inf_T \sup_{\Theta \in \mathcal{S}} P_{(\Theta,\rho,\sigma)} \left(|T - h(\Theta)| > \kappa \right) \geq \max \left(\frac{1}{4}\exp(-1/2), \frac{1-\sqrt{(1/2)/2}}{2} \right) = 1/4.$$

Case 2: dense g

Fix any $\Theta_* \in \mathcal{S}$ with entries $\{\theta_{ij,*}\}$ such that $\mathsf{rank}\Theta_* \leq J-1$ and $\max_{1\leq i\leq n}\max_{1\leq j\leq p}|\theta_{ij,*}| \leq c_1/2$. (This is always feasible because we can simply choose Θ_* to be the zeor matrix.)

Let $\kappa > 0$ be a constant satisfying the following

$$\kappa \leq c_1 c_5/2 \quad \textit{and} \quad \frac{\kappa^2(1-c_2)}{2c_3^2 c_5^2} \leq 1/2. \tag{B.11}$$

Define $\Delta_j = \kappa p^{-3/2}/g_j$. Let Θ_{**} be a matrix with entries $\{\theta_{ij,**}\}$ defined as follows

$$\theta_{ij,**} = \begin{cases} \theta_{ij,*} & \text{if } i \neq 1 \\ \theta_{1j,*} + \Delta_j & \text{if } i = 1. \end{cases}$$

Since Θ_* and Θ_{**} only differ in the first row, it follows that $\mathsf{rank}\Theta_{**} \leq \mathsf{rank}\Theta_* + 1 \leq J - 1 + 1 = J$. Since

$$\begin{aligned} \max_{1\leq j\leq p}|\theta_{1j,**}| = \max_{1\leq j\leq p}|\theta_{1j,*} + \Delta_j| &\leq \max_{1\leq j\leq p}|\theta_{1j,*}| + \max_{1\leq j\leq p}|\Delta_j| \\ &\leq \frac{c_1}{2} + \frac{\kappa p^{-3/2}}{\min_j|g_j|} \leq \frac{c_1}{2} + \frac{\kappa p^{-3/2}}{c_5/p} \overset{\text{(i)}}{\leq} c_1, \end{aligned}$$

where (i) follows by $\kappa \leq c_1 c_5 p^{1/2}/2$ (due to (B.11) and $p \geq 1$). Therefore, $\Theta_{**} \in \mathcal{S}$.

We now compare the Kullback–Leibler divergence bewteen $P_{(\Theta_*,\rho,\sigma)}$ and $P_{(\Theta_{**},\rho,\sigma)}$:

$$KL(P_{(\Theta_*,\rho,\sigma)}, P_{(\Theta_{**},\rho,\sigma)}) = \int \left(\log \frac{dP_{(\Theta_*,\rho,\sigma)}}{dP_{(\Theta_{**},\rho,\sigma)}}\right) dP_{(\Theta_*,\rho,\sigma)}.$$

By the formula of Gaussian densities, we have that

$$\frac{dP_{(\Theta_*,\rho,\sigma)}}{dP_{(\Theta_{**},\rho,\sigma)}} = \frac{\exp\left(-\frac{1}{2}\sum_{j=1}^p (y_{1j} - \theta_{1j,*}x_{1j})^2 \sigma_{1j}^{-2}\right)}{\exp\left(-\frac{1}{2}\sum_{j=1}^p (y_{1j} - \theta_{1j,**}x_{1j})^2 \sigma_{1j}^{-2}\right)}.$$

Therefore,

$$\begin{aligned}
KL(P_{(\Theta_*,\rho,\sigma)}, P_{(\Theta_{**},\rho,\sigma)}) &= \int \left(\log \frac{dP_{(\Theta_*,\rho,\sigma)}}{dP_{(\Theta_{**},\rho,\sigma)}}\right) dP_{(\Theta_*,\rho,\sigma)} \\
&= \mathsf{E}_{(\Theta_*,\rho,\sigma)} \left(-\frac{1}{2}\sum_{j=1}^{p} \left[(y_{1j}-\theta_{1j,*}x_{1j})^2 - (y_{1j}-\theta_{1j,**}x_{1j})^2\right] \sigma_{1j}^{-2}\right) \\
&= \mathsf{E}_{(\Theta_*,\rho,\sigma)} \left(\sum_{j=1}^{p} \left[\frac{1}{2}\Delta_j^2 x_{1j} - (y_{1j}-\theta_{1j,*}x_{1j})\Delta_j x_{1j}\right] \sigma_{1j}^{-2}\right) \\
&= \frac{\kappa^2 p^{-3}}{2}\sum_{j=1}^{p} \sigma_{1j}^{-2}\rho_j g_j^{-2} \le \frac{\kappa^2(1-c_2)}{2c_3^2 c_5^2} \overset{\text{(i)}}{\le} 1/2,
\end{aligned}$$

where (i) follows by (B.11).

We notice that

$$|h(\Theta_*) - h(\Theta_{**})| = \left|\sum_{j=1}^{p} g_j \Delta_j\right| = \kappa p^{-1/2}.$$

By Theorem 2.2 in Tsybakov [5] and Equation (2.9) therein, we have that

$$\inf_T \sup_{\Theta\in\mathcal{S}} P_{(\Theta,\rho,\sigma)}\left(|T - h(\Theta)| > \kappa p^{-1/2}\right) \ge \max\left(\frac{1}{4}\exp(-1/2), \frac{1-\sqrt{(1/2)/2}}{2}\right) = 1/4.$$

The proof is complete. □

B.6. **Proof of Theorem 4.4.**

Proof. We write $\widehat{\mu}_j^2 = \widehat{\mu}_{j,i}^2$ for simplicity because we fix i of interest. Note that $g_j y_{ij} x_{ij} = g_j x_{ij}^2 \theta_{ij} + g_j e_{ij}$, where $e_{ij} = x_{ij}\varepsilon_{ij}$. Let $\mu_j^2 = \mathsf{E}x_{kj}^2$.

$$\begin{aligned}
\widehat{h_i(\Theta)} - \theta_i' g &= \sum_{j=1}^{p} \mu_j^{-2} g_j [x_{ij}\varepsilon_{ij} + (x_{ij}^2 - \mu_j^2)\theta_{ij}] + (a) ... + (d) \\
(a) &= \sum_{j=1}^{p} (\widehat{\mu}_j^{-2} - \mu_j^{-2})\mu_j^{-2}(\mu_j^2 - \widehat{\mu}_j^2) g_j x_{ij}^2 \theta_{ij} \\
(b) &= \sum_{j=1}^{p} \mu_j^{-4}(\mu_j^2 - \widehat{\mu}_j^2) g_j x_{ij}^2 \theta_{ij} \\
(c) &= \sum_{j=1}^{p} (\widehat{\mu}_j^{-2} - \mu_j^{-2})\mu_j^{-2}(\mu_j^2 - \widehat{\mu}_j^2) g_j e_{ij}
\end{aligned}$$

$$(d) = \sum_{j=1}^p \mu_j^{-4}(\mu_j^2 - \widehat{\mu}_j^2) g_j e_{ij}.$$

We now bound each term.

For (a)(c), note $\max_j |\widehat{\mu}_j^2 - \mu_j^2| = \max_j |\frac{1}{n-1}\sum_{k\neq i} x_{kj}^2 - \mathsf{E}x_{kj}^2| = O_P(\sqrt{\frac{\log p}{n}})$. Hence $\max_j \widehat{\mu}_j^2 = O_P(1)$, and $\max_j |\widehat{\mu}_j^{-2} - \mu_j^{-2}| = O_P(\sqrt{\frac{\log p}{n}})$. Hence

$$\begin{aligned} (a) &\leq O_P(\frac{\log p}{n}) \max_j |\mu_j^{-2} x_{ij} g_j| \sum_{j=1}^p |\theta_{ij}| \leq O_P(\frac{\log p}{n}) \\ (c) &\leq O_P(\frac{\log p}{n}) \max_j |\mu_j^{-2}| \sum_j |g_j e_{ij}| \leq O_P(\frac{\log p}{n}). \end{aligned}$$

For (b)(d), let $v_{kj} = x_{kj}^2 - \mathsf{E}x_{kj}^2$ and $c_j = \mu_j^{-4}\theta_{ij}$. Then v_{kj} is independent over k.

$$\begin{aligned} \mathsf{E}(b)^2 &= \mathsf{E}\left(\frac{1}{n-1}\sum_{j=1}^p \sum_{k\neq i} v_{kj} g_j x_{ij}^2 c_j\right)^2 \\ &\leq O(\frac{1}{n^2}) \sum_{j,l\leq p}\sum_{k\neq i} g_j g_l \mathsf{E}v_{kj}v_{kl}\mathsf{E}c_j c_l \mathsf{E}x_{il}^2 x_{ij}^2 \\ &\leq O(\frac{1}{np^2}) \max_{i\leq n} \sum_{j,l\leq p} |\mathsf{Cov}(x_{ij}^2, x_{il}^2)| = O(\frac{1}{np}) \\ \mathsf{E}(d)^2 &= \mathsf{E}\left(\frac{1}{n-1}\sum_{j=1}^p \sum_{k\neq i} v_{kj} g_j x_{ij}\mu_j^{-4}\varepsilon_{ij}\right)^2 \\ &\leq O(\frac{1}{np^2}) \max_{k\leq n} \sum_{j,l\leq p} |\mathsf{Cov}(x_{kl}^2, x_{kj}^2)\mathsf{E}x_{ij}\varepsilon_{ij}x_{il}\varepsilon_{il}| = O(\frac{1}{np}). \end{aligned}$$

Hence provided that $\sqrt{p}\log p = o(n)$, we have

$$\sqrt{p}[\widehat{h_i(\Theta)} - \theta_i' g] = \sum_{j=1}^p Z_j + o_P(1)$$

where $Z_j = \sqrt{p}\mu_j^{-2} g_j W_{ij}$ with $W_{ij} = x_{ij}\varepsilon_{ij} + x_{ij}^2\theta_{ij}$.

We now verify the Lindeberg condition. Suppose $\mu_j > c > 0$. Let $s_n^2 = \sum_{j=1}^p \mathsf{Var}(Z_j)$. Note that $\mathsf{E}Z_j^4 = O(p^{-2})\mathsf{E}W_{ij}^4 = O(p^{-2})$. Also $\mathsf{Var}(Z_j) \leq \sqrt{\mathsf{E}Z_j^4} = O(p^{-1})$. Also $\mathsf{Var}(Z_j) \geq cp^{-1}$. Hence $s_n^{-1} = O(1)$. Then for any $\epsilon > 0$,

$$s_n^{-2}\sum_j \mathsf{E}(Z_j - \mathsf{E}Z_j)^2 1\{|Z_j - \mathsf{E}Z_j| > \epsilon s_n\} \leq \epsilon^{-1} s_n^{-3} \sum_j \sqrt{\mathsf{E}(Z_j - \mathsf{E}Z_j)^4}\sqrt{\mathsf{Var}(Z_j)} = O(p^{-1/2}).$$

Thus $s_n^{-1}\sqrt{p}[\widehat{h_i(\Theta)} - \theta_i' g] \to^d N(0,1)$. In particular, $s_n^2 = p\sum_{j=1}^p g_j^2 \mu_j^{-4} \mathsf{Var}(W_{ij})$. □

B.7. **Proof of Theorem 5.2: the Treatment Effect Study.**

B.7.1. *Proof of Lemma 5.1.* We now prove Lemma 5.1 which verifies the SSV and incoherence condition in the treatment effect study. For ease of readability, this lemma is restated below.

Lemma B.4. *(i) The minimum nonzero singular value ψ_{np}^2 for $\Theta_0(m)$ can be taken as*

$$\psi_{np}^2 \asymp J^{-1}\sum_{i=1}^n \sum_{j=1}^p h_{j,m}(\eta_i)^2, \quad m = 0, 1,$$

which means $\psi_J(\Theta_0(m)) \geq c\psi_{np}^2$ for this choice of ψ_{np}.

(ii) The incoherence Assumption 4.4 holds.

(iii) The low-rank approximation error satisfies $\|R(m)\|_{(n)} \leq CJ^{-(a-1)}\max\{p, n\}$.

Proof. Fix $m \in \{0, 1\}$, we drop the dependence on m for notational simplicity.

(i) SSV: Let $\psi_j(A)$ be the j th largest singular value of A. Then $\psi_j(\Theta_0) = \psi_j(A_m)$. Now for any $\epsilon > 0$, with probability approaching one,

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^p h_j(\eta_i)^2 &= \|\Theta\|_F^2 \leq (1+\epsilon)\|\Theta_0\|_F^2 = (1+\epsilon)\sum_{k=1}^J \psi_k^2(\Theta_0) \\
&\leq (1+\epsilon)J\psi_1^2(A) \leq^{(1)} C(1+\epsilon)J^{2b+1}\psi_J^2(A) \\
&= C(1+\epsilon)J^{2b+1}\psi_J^2(\Theta_0) \\
\sum_{i=1}^n \sum_{j=1}^p h_j(\eta_i)^2 &= \|\Theta\|_F^2 \geq (1-\epsilon)\|\Theta_0\|_F^2 = (1-\epsilon)\sum_{k=1}^J \psi_k^2(\Theta_0) \\
&\geq (1-\epsilon)J\psi_J^2(\Theta_0).
\end{aligned}$$

where (1) follows from Assumption 5.1.

(ii) Incoherence: We derive the singular vectors and singular values of Θ_0. Note that we can write

$$\Theta = \underbrace{\Phi\Lambda'}_{\Theta_0} + R.$$

where Φ is the $n \times J$ matrix of ϕ_i; Λ is the $p \times J$ matrix of λ_j; R is the $n \times p$ matrix of r_{ij}. Write $S_\Lambda = \frac{1}{p}\Lambda'\Lambda$, $S_\Phi = \frac{1}{n}\Phi'\Phi$, $A = S_\Phi^{1/2} S_\Lambda S_\Phi^{1/2}$. Also let G_Φ be a $J \times J$ matrix whose columns are the eigenvectors of A, and T be the diagonal matrix of corresponding eigenvalues. Let $H_\Phi := S_\Phi^{-1/2} G_\Phi$, it can be verified that

$$\Theta_0\Theta_0'\Phi H_\Phi = pn\Phi H_\Phi T, \qquad \frac{1}{n}(\Phi H_\Phi)'\Phi H_\Phi = \mathsf{I}.$$

This shows that the columns of ΦH_Φ are the left singular-vectors of Θ_0; the eigenvalues of A equal the the first J eigenvalues of $\Theta_0'\Theta_0$. Similarly, we can define $H_\Lambda = S_\Lambda^{-1/2} G_\Lambda$, where G_Λ is a $J \times J$ matrix whose columns are the eigenvectors of $S_\Lambda^{1/2} S_\Phi S_\Lambda^{1/2}$. Hence we have

$$U_0 = n^{-1/2}\Phi H_\Phi, \quad V_0 = p^{-1/2}\Lambda H_\Lambda. \tag{B.12}$$

The $h_j(\cdot)$ function belongs to the Hilber space with uniformly-bounded L^2 norm, meaning that $\max_{j\le p}\|h_j\|_{L_2}^2 = \max_{j\le p}\sum_{k=1}^\infty \lambda_{j,k}^2 < \infty$. Thus

$$\begin{aligned}
\mathsf{E}\max_{j\le p}\|v_j\|^2 &\le \max_{j\le p}\|\lambda_j\|^2\psi_{\min}^{-1}(\frac{1}{p}\Lambda'\Lambda)p^{-1} \\
&\le \max_{j\le p}\sum_{k=1}^\infty \lambda_{j,k}^2 p^{-1}\psi_{\min}^{-1}(\frac{1}{p}\Lambda'\Lambda) \le O(p^{-1}) \\
\sum_{i=1}^n \|u_i\|^4 &\le \sum_{i=1}^n \|\phi_i\|^4\psi_{\min}^{-2}(S_\Phi) \le \sum_{i=1}^n \|\Phi_i\|^4\psi_{\min}^{-2}(\frac{1}{n}\Phi'\Phi)n^{-2} \\
&\le \frac{1}{n}\sum_{i=1}^n \|\Phi_i\|^4 O_P(J^2n^{-1}) = o_P(1) \\
\mathsf{E}\max_{i\le n}\|u_i\|^2 &\le \mathsf{E}\max_{i\le n}\|\Phi_i\|^2\psi_{\min}^{-1}(S_\Phi) \\
&\le n^{-1}J\sup_\eta\max_{j\le J}|\phi_j(\eta)|^2\mathsf{E}\psi_{\min}^{-1}(S_\Phi) = O(Jn^{-1}).
\end{aligned}$$

(iii) Sieve error:

$$\|R\|_{(n)} \le \|R\|_F(p\vee n)^{1/2} \le C(p\vee n)^{3/2}J^{-a}.$$

□

B.7.2. *Proof of Lemma 5.3.* In this lemma, we verify the following condition:

$$\begin{aligned}
\psi_1(\frac{1}{np}\Theta\Theta')/\psi_J(\frac{1}{np}\Theta\Theta') &\le O_P(J^\alpha) \\
\min_{k=1\ldots J-1}\psi_k(\frac{1}{np}\Theta\Theta') - \psi_{k+1}(\frac{1}{np}\Theta\Theta') &\ge cJ^{-(\alpha+1)}.
\end{aligned}$$

Proof. Since and h_j are independently generated from the Gaussian process prior, we have

$$\max_{i,l}\left|\frac{1}{p}\sum_{j=1}^{p} h_j(\eta_i)h_j(\eta_l)-K(\eta_i,\eta_l)\right| = O_P(\sqrt{\frac{\log n}{p}}).$$

Hence for $r_J := \sup_{\eta_1,\eta_2}|\sum_{k>J}\nu_k\phi_k(\eta_1)\phi_k(\eta_2)|$,

$$\|\frac{1}{np}\Theta\Theta'-\frac{1}{n}\Phi D_\lambda\Phi'\| \leq O_P(\sqrt{\frac{\log n}{p}})+r_J.$$

Thus

$$\max_{j\leq J}|\psi_j(\frac{1}{np}\Theta\Theta')-\psi_j(\frac{1}{n}\Phi D_\lambda\Phi')| \leq \underbrace{\|\frac{1}{np}\Theta\Theta'-\frac{1}{n}\Phi D_\lambda\Phi'\|}_{O_P(\sqrt{\frac{\log n}{p}}+r_J)}.$$

We now show that the eigenvalues of $\frac{1}{n}\Phi D_\lambda\Phi'$ are approximately D_λ. In fact, the top J eigenvalues equal those of $(\frac{1}{n}\Phi'\Phi)^{1/2}D_\lambda(\frac{1}{n}\Phi'\Phi)^{1/2}$. Because η_i are i.i.d Unif$[0,1]$, and ϕ_j are eigenfunctions of the operator T, so $\int\phi_{k_1}(\eta)\phi_{k_2}(\eta)d\eta = 1\{k_1=k_2\}$, showing $\mathsf{E}\frac{1}{n}\Phi'\Phi=I$. Hence $\|\frac{1}{n}\Phi'\Phi-I\|=O_P(\frac{J}{\sqrt{n}})$. This implies

$$\max_{j\leq J}|\psi_j(\frac{1}{n}\Phi D_\lambda\Phi')-\nu_j| \leq \|(\frac{1}{n}\Phi'\Phi)^{1/2}D_\lambda(\frac{1}{n}\Phi'\Phi)^{1/2}-D_\lambda\| \leq \|D_\lambda\|O_P(\frac{J}{\sqrt{n}}).$$

Together, for $w_n := \sqrt{\frac{\log n}{p}}+r_J+\|D_\lambda\|\frac{J}{\sqrt{n}}$,

$$\max_{j\leq J}|\psi_j(\frac{1}{np}\Theta\Theta')-\nu_j| \leq O_P(w_n)=o_P(\nu_J/J).$$

This implies

$$\frac{\psi_1(\frac{1}{np}\Theta\Theta')}{\psi_J(\frac{1}{np}\Theta\Theta')} \leq \frac{O_P(w_n)}{\nu_J-O_P(w_n)}+\frac{2\nu_1}{\nu_J} \leq O_P(J^\alpha).$$

Also, for any $k\leq J-1$, the mean value theorem implies

$$\psi_k(\frac{1}{np}\Theta\Theta')-\psi_{k+1}(\frac{1}{np}\Theta\Theta') \geq M(k^{-\alpha}-(k+1)^{-\alpha})-O_P(w_n) \geq \alpha MJ^{-\alpha-1}-O_P(w_n).$$

The lower bound also holds (up to a constant) for $\psi_k(\frac{1}{\sqrt{np}}\Theta)-\psi_{k+1}(\frac{1}{\sqrt{np}}\Theta)$ because $\psi_1(\frac{1}{\sqrt{np}}\Theta)=O_P(1)$.

□

B.7.3. *Proof of Theorem 5.2.*

Proof. Let the columns of $V_0(m) = (v_j(m) : j \leq p)$ be the right singular vectors of $\Theta_0(m)$ and $\bar{B}(0) = \sum_{j \in T_0} x_{ij}(0) v_j(0) v_j(0)'$ and $\bar{B}(1) = \sum_{j \in T_1} x_{ij}(1) v_j(1) v_j(1)'$.

Note that condition (12) implies Assumption 4.3 (ii-c) when $\{1, ..., p\}$ is replaced with T_0 and for $x_{ij} = x_{ij}(0)$. Hence (B.7) holds, for $g_0 = \frac{1}{p_0}(1, ..., 1)'$, on T_0 and $m = 0$. We then have

$$\begin{aligned} \widehat{\theta}_i(0)' g_0 - \theta_i(0)' g_0 &= \sum_{j \in T_0} e_{ij} x_{ij}(0) v_j(0)' \bar{B}(0)^{-1} V_0(0)' g_0 + o_P(p_0^{-1/2}) \\ &:= \sum_{j \in T_0} e_{ij} \zeta_{ij}(0) + o_P(p_0^{-1/2}). \end{aligned}$$

Similarly, for $g_1 = \frac{1}{p_1}(1, ..., 1)'$, on T_1 and $m = 1$, we have

$$\widehat{\theta}_i(1)' g_1 - \theta_i(1)' g_1 = \sum_{j \in T_1} e_{ij} \zeta_{ij}(1) + o_P(p_1^{-1/2}).$$

Together, we reach

$$\begin{aligned} \widehat{\tau}_i - \tau_i &= \widehat{\theta}_i(1)' g_1 - \widehat{\theta}_i(0)' g_0 - \left[\frac{1}{p} \sum_{j=1}^{p} \theta_{ij}(1) - \frac{1}{p} \sum_{j=1}^{p} \theta_{ij}(0) \right] \\ &= \widehat{\theta}_i(1)' g_1 - \theta_i(1)' g_1 - [\widehat{\theta}_i(0)' g_0 - \theta_i(0)' g_0] \\ &\quad + \theta_i(1)' g_1 - \frac{1}{p} \sum_{j=1}^{p} \theta_{ij}(1) + \frac{1}{p} \sum_{j=1}^{p} \theta_{ij}(0) - \theta_i(0)' g_0 \\ &= \sum_{j \in T_1} e_{ij} \zeta_{ij}(1) - \sum_{j \in T_0} e_{ij} \zeta_{ij}(0) + o_P(p_0^{-1/2} + p_1^{-1/2}) \\ &= \sum_{j=1}^{p} e_{ij} \Delta_{ij} + o_P(\min\{p_0, p_1\}^{-1/2}) \\ \Delta_{ij} &:= \zeta_{ij}(1) 1\{j \in T_1\} - \zeta_{ij}(0) 1\{j \in T_0\}, \\ \bar{s}_{np,i}^2 &:= \sum_{j \in T_0} \mathsf{Var}(e_{ij} | X, g(\eta)) \zeta_{ij}(0)^2 + \sum_{j \in T_1} \mathsf{Var}(e_{ij} | X, g(\eta)) \zeta_{ij}(1)^2 \end{aligned}$$

Note that e_{ij} is independent over j conditioning on $x_{ij}(m)$ and $\Theta(m)$, $m = 0, 1$. We verify the Lindeberg condition: for any $\epsilon > 0$,

$$\begin{aligned} &\frac{1}{\bar{s}_{np,i}^2} \sum_{j=1}^{p} \mathsf{E}[e_{ij}^2 \Delta_{ij}^2 1\{|e_{ij} \Delta_{ij}| > \epsilon \bar{s}_{np,i}\} | X, g(\eta)] \leq \frac{1}{\bar{s}_{np,i}^4 \epsilon^2} \sum_{j=1}^{p} \mathsf{E}[e_{ij}^4 \Delta_{ij}^4 | X, g(\eta)] \\ \leq\ & \frac{C}{\bar{s}_{np,i}^4} \sum_{j \in T_0} \zeta_{ij}(0)^4 + \frac{C}{\bar{s}_{np,i}^4} \sum_{j \in T_1} \zeta_{ij}(1)^4 \leq \frac{C}{\bar{s}_{np,i}^4 p_0^2} \sum_{j \in T_0} \|v_j(0)\|^4 + \frac{C}{\bar{s}_{np,i}^4 p_1^2} \sum_{j \in T_1} \|v_j(1)\|^4 \end{aligned}$$

$$\leq \frac{CJ^2}{\bar{s}_{np,i}^4 p_0^3} + \frac{CJ^2}{\bar{s}_{np,i}^4 p_1^3} = o_P(1).$$

Thus with the condition that $\bar{s}_{np,i}^2 \min\{p_0, p_1\} \geq c$ for some $c > 0$ with probability approaching one, we have $\bar{s}_{np,i}^{-1}(\widehat{\tau}_i - \tau_i) \to^d N(0, 1)$.

□

Appendix C. Technical Lemmas

Lemma C.1. *Suppose* $\sum_{j=1}^p \mathsf{E}\|v_j\|^4 = o(1)$ *and* $\min_{i\leq n} \|\sum_{j=1}^p v_j v_j' x_{ij}^2\| \geq c$. *Let* $\widehat{B}_i = \sum_{j=1}^p x_{ij}^2 \widetilde{v}_j \widetilde{v}_j'$ *and* $B = H_S' \sum_{j=1}^p (\mathsf{E}x_{ij}^2) v_j v_j' H_S$. *Then*

(i) $\max_{i\leq n} \|\widehat{B}_i^{-1}\| = O_P(1)$.

(ii) Consider the DGP of x_{ij}, *corresponding to conditions (ii-a)-(ii-b) in Assumption 4.3.*

For each fixed $i \leq n$,
$\|\widehat{B}_i^{-1} - B^{-1}\| = O_P(J^d \omega_{np} \psi_{np}^{-1} + \sqrt{\sum_{j=1}^p \mathsf{E}\|v_j\|^4})$, *and*
$\sum_{i\notin\mathcal{I}} \|\widehat{B}_i^{-1} - B^{-1}\|^2 = O_P(nJ^{2d}\omega_{np}^2 \psi_{np}^{-2} + n\sum_{j=1}^p \mathsf{E}\|v_j\|^4)$.
In case (ii-c), For each fixed $i \leq n$,
$\|\widehat{B}_i^{-1} - B^{-1}\| = O_P(J^d \omega_{np} \psi_{np}^{-1} + Jp^{-1}\sum_{j\in\mathcal{B}_i \triangle \bar{\mathcal{B}}} 1)$, *and*
$\sum_{i\notin\mathcal{I}} \|\widehat{B}_i^{-1} - B^{-1}\|^2 = O_P(nJ^{2d}\omega_{np}^2 \psi_{np}^{-2}) + O_P(nJ^2 p^{-2})[\sum_{j\in\mathcal{B}_i \triangle \bar{\mathcal{B}}} 1]^2$.

Proof. (i) Note that for $\max_{ij} |x_{ij}| = O_P(1)$,

$$\begin{aligned}
&\max_{i\leq n} \|\widehat{B}_i - H_S' \sum_{j=1}^p x_{ij}^2 v_j v_j' H_S\| \leq O_P(1)[\max_{i\leq n} \|\sum_{j=1}^p x_{ij}^2 (\widetilde{v}_j - H_S' v_j) v_j'\| \\
&+ \max_{i\leq n} \sum_{j=1}^p x_{ij}^2 \|\widetilde{v}_j - H_S' v_j\|^2] \\
\leq\ & O_P(1)\|\widetilde{V}_S - V_0 H_S\| = O_P(J^d \omega_{np} \psi_{np}^{-1}) = o_P(1).
\end{aligned}$$

On the other hand, by the assumption $\min_{i\leq n} \|\sum_{j=1}^p v_j v_j' x_{ij}^2\| \geq c$ almost surely. We have $\max_{i\leq n} \|\widehat{B}_i^{-1}\| \leq \max_{i\leq n} \|(\sum_{j=1}^p x_{ij}^2 v_j v_j')^{-1}\| + o_P(1) = O_P(1)$.

To prove (ii), we proceed below corresponding to conditions (ii-a)-(ii-b) in Assumption 4.3.

Under condition (ii-a) x_{ij}^2 does not vary across $i \leq n$. We can write $x_{ij}^2 = x_j^2$. In this case $B = H_S' \sum_{j=1}^p x_j^2 v_j v_j' H_S$.

We have

$$\|\widehat{B}_i^{-1}-B^{-1}\|\leq O_P(1)\|\widehat{B}_i-H_S'\sum_{j=1}^p x_{ij}^2v_jv_j'H_S\|\leq O_P(J^d\omega_{np}\psi_{np}^{-1}).$$

In addition, $\sum_{i\notin\mathcal{I}}\|\widehat{B}_i-B\|^2\leq O_P(nJ^{2d}\omega_{np}^2\psi_{np}^{-2})$.

Under condition (ii-b) x_{ij} is independent across $i\leq n$ and is weakly dependent across $j\leq p$. In this case $B=H_S'\sum_{j=1}^p(\mathsf{E}x_{ij}^2)v_jv_j'H_S$.

We now bound $\sum_{j=1}^p(x_{ij}^2-\mathsf{E}x_{ij}^2)v_jv_j'$ for fixed $i\leq n$. By the assumption x_{ij}^2 is weakly dependent across j. Also $\mathsf{E}((x_{ij}^2-\mathsf{E}x_{ij}^2)|\Theta_0)=0$. Let $X_i^2=(x_{i1}^2,...,x_{ip}^2)'$. $\max_i\|\mathsf{Var}(X_i^2|\Theta_0)\|^2<C$. Let $V_{k_1,k_2}=p\times 1$ vector of $v_{j,k_1}v_{j,k_2}$. Then with the assumption $\sum_{j=1}^p\mathsf{E}\|v_j\|^4=o(1)$,

$$\mathsf{E}\|\sum_{j=1}^p(x_{ij}^2-\mathsf{E}x_{ij}^2)v_jv_j'\|_F^2=\sum_{k_1,k_2\leq J}\mathsf{E}\mathsf{Var}[\sum_{j=1}^p x_{ij}^2v_{j,k_1}v_{j,k_2}|\Theta_0]$$

$$=\sum_{k_1,k_2\leq J}\mathsf{E}V_{k_1,k_2}'\mathsf{Var}[X_i^2|\Theta_0]V_{k_1,k_2}\leq C\sum_{j=1}^p\mathsf{E}\|v_j\|^4\leq o(1).$$

Hence $\|\widehat{B}_i^{-1}-B^{-1}\|\leq O_P(1)\|\widehat{B}_i-B\|\leq O_P(J^d\omega_{np}\psi_{np}^{-1}+\sqrt{\sum_{j=1}^p\mathsf{E}\|v_j\|^4})$.

In addition, by the proof of part (i),

$$\sum_{i\notin\mathcal{I}}\|\widehat{B}_i-B\|^2\leq O_P(nJ^{2d}\omega_{np}^2\psi_{np}^{-2})+\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p(x_{ij}^2-\mathsf{E}x_{ij}^2)v_jv_j'\|^2.$$

By a similar argument of part (ii), $\mathsf{E}\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p(x_{ij}^2-\mathsf{E}x_{ij}^2)v_jv_j'\|^2\leq n\sum_{j=1}^p\mathsf{E}\|v_j\|^4$. Together, $\|B\|^2=O_P(1)$ and $\max\|\widehat{B}_i^{-1}\|=O_P(1)$ so

$$\begin{aligned}\sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1})\|^2 &\leq \|B^{-1}\|^2\max\|\widehat{B}_i^{-1}\|^2\sum_{i\notin\mathcal{I}}\|\widehat{B}_i-B\|^2\\ &= O_P(nJ^{2d}\omega_{np}^2\psi_{np}^{-2}+n\sum_{j=1}^p\mathsf{E}\|v_j\|^4).\end{aligned}$$

Under condition (ii-c) There is a common $\bar{\mathcal{B}}$. In this case we restrict to $x_{ij}\in\{0,1\}$, and let $B=H_S'\sum_{j\in\bar{\mathcal{B}}}^p v_jv_j'H_S$. Let $\mathcal{B}_i=\{j:x_{ij}=1\}$.

We have

$$\|\widehat{B}_i^{-1}-B^{-1}\| \leq O_P(1)\|\widehat{B}_i-H_S'\sum_{j=1}^p x_{ij}^2v_jv_j'H_S\|+O_P(1)\|\sum_{j\in\mathcal{B}_i}v_jv_j'-\sum_{j\in\bar{\mathcal{B}}}v_jv_j'\|$$

$$\begin{aligned}&\leq O_P(J^d\omega_{np}\psi_{np}^{-1})+O_P(1)\max_{j\leq p}\|v_j\|^2\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}1\\&\leq O_P(J^d\omega_{np}\psi_{np}^{-1})+O_P(Jp^{-1})\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}1.\end{aligned}$$

In addition,

$$\begin{aligned}\sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1})\|^2 &\leq O_P(1)\sum_{i\notin\mathcal{I}}\|\widehat{B}_i-B\|^2\\&= O_P(nJ^{2d}\omega_{np}^2\psi_{np}^{-2})+O_P(nJ^2p^{-2})[\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}1]^2\end{aligned}$$

□

Lemma C.2. *For $d=1\sim 4$, consider $\Delta_{i,d}$ as defined in (B.1). In addition, let*

$$\Delta_{i,5} := B^{-1}\sum_{j=1}^p\widetilde{v}_j(x_{ij}^2-\mathsf{E}x_{ij}^2)(v_j-H_S^{-1\prime}\widetilde{v}_j)'\gamma_i$$

Also, consider the three different cases on the DGP of x_{ij}, corresponding to conditions (ii-a)-(ii-c) in Assumption 4.3. Then:
(i) Under conditions either (ii-a) or (ii-b), for each $i\notin\mathcal{I}$, and $d=1\sim 4$, $\|\Delta_{i,d}\|=o_P(1)$ and $\sum_{i\notin\mathcal{I}}\|\Delta_{i,d}\|^2=O_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2)$.
(ii) Under condition (ii-b), for each $i\notin\mathcal{I}$, $\|\Delta_{i,5}\|=o_P(1)$ and $\sum_{i\notin\mathcal{I}}\|\Delta_{i,5}\|^2=O_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2)$.
(iii) Under conditions (ii-c), for each for each $i\notin\mathcal{I}$, and $d=1\sim 4$, 6, $\|\Delta_{i,d}\|=o_P(1)$ and $\sum_{i\notin\mathcal{I}}\|\Delta_{i,d}\|^2=O_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2)$.

Proof. (i) **For term $\Delta_{i,1}$.**

Lemma C.1 shows $\|\widehat{B}_i^{-1}-B^{-1}\|=O_P(J^d\omega_{np}\psi_{np}^{-1}+\sqrt{\sum_{j=1}^p\mathsf{E}\|v_j\|^4})$ and $\|\widehat{B}_i^{-1}\|=O_P(1)$. Also let $X_i^2=(x_{i1}^2,...,x_{ip}^2)'$. Note $\|\gamma_i\|\leq O_P(J^b\psi_{np})\|u_i\|$.

$$\|\sum_{j=1}^p\widetilde{v}_jx_{ij}^2(v_j-H_S^{-1\prime}\widetilde{v}_j)'\gamma_i\|\leq\|\gamma_i\|\|\widetilde{V}'\mathsf{diag}(X_i^2)(\widetilde{V}H_S^{-1}-V)\|\leq\|u_i\|O_P(J^{b+g}\omega_{np}).$$

Hence $\|\Delta_{i,1}\|=O_P(J^d\omega_{np}\psi_{np}^{-1}+\sqrt{\sum_{j=1}^p\mathsf{E}\|v_j\|^4})\|u_i\|\omega_{np}J^{b+g}=o_P(1)$.

In addition, note that $\sum_{i\notin\mathcal{I}}\|\gamma_i\|^2=O_P(J^{1+2b}\psi_{np}^2)$. Hence

$$\sum_{i\notin\mathcal{I}}\|\Delta_{i,1}\|^2 \leq O_P(1)\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p\widetilde{v}_jx_{ij}^2(v_j-H_S^{-1\prime}\widetilde{v}_j)'\gamma_i\|^2$$

$$\leq O_P(1)\sum_{i\notin\mathcal{I}}\|\gamma_i\|^2\|\widetilde{V}'\mathsf{diag}(X_i^2)(\widetilde{V}_S-VH_S)\|^2=O_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2).$$

For term $\Delta_{i,2}$.

By assumption ε_{ij} is independent across $i\leq n$. Because $\widetilde{V}_{\mathcal{I}}$ is estimated using subsamples excluding i, $\mathsf{E}(\varepsilon_{ij}|\widetilde{V}_{\mathcal{I}},\Theta_0,X)=0$. Let $\widetilde{P}_{ik}$ be $p\times 1$ vector of $x_{ij}(\widetilde{v}_{j,k}-v_{j,k})$. By the assumption $\mathsf{Var}[\varepsilon_i|X,\Theta_0]<C$,

$$\begin{aligned}\mathsf{E}_I(\|\sum_{j=1}^p\varepsilon_{ij}x_{ij}(\widetilde{v}_j-H_S'v_j)\|^2|\Theta_0)&=\sum_{k\leq J}\mathsf{E}_I\widetilde{P}_{ik}'\mathsf{Var}_I[\varepsilon_i|\widetilde{V}_{\mathcal{I}},X,\Theta_0]\widetilde{P}_{ik}\\ \leq\quad \mathsf{E}_I(\sum_{j=1}^p x_{ij}^2\|\widetilde{v}_j-H_S'v_j\|^2|\Theta_0)&\leq\|\widetilde{V}_{\mathcal{I}}-VH_S|_F^2=O_P(J^{2d}\omega_{np}^2\psi_{np}^{-2})=o_P(1).\end{aligned}\tag{C.1}$$

Hence $\|\Delta_{i,2}\|=o_P(1)$. In addition, $\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p\varepsilon_{ij}x_{ij}(\widetilde{v}_j-v_j)\|^2=O_P(nJ^{2d}\omega_{np}^2\psi_{np}^{-2})$. implies $\sum_{i\notin\mathcal{I}}\|\Delta_{i,2}\|^2\leq O_P(nJ^{2d}\omega_{np}^2\psi_{np}^{-2})=O_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2)$.

For term $\Delta_{i,3}$.

$$\begin{aligned}\max_{i\leq n}\|\Delta_{i,3}\|&\leq O_P(1)\|\sum_{j=1}^p r_{ij}\widetilde{v}_j\|\leq\max_{ij}|r_{ij}|\sqrt{pJ}=o_P(1)\\ \sum_{i\notin\mathcal{I}}\|\Delta_{i,3}\|^2&\leq n\max_i\|\Delta_{i,3}\|^2\leq npJ\max_{ij}r_{ij}^2=o_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2).\end{aligned}\tag{C.2}$$

For term $\Delta_{i,4}$.

Note $\mathsf{E}\|\sum_{j=1}^p\varepsilon_{ij}x_{ij}v_j\|^2\leq C\sum_{j=1}^p x_{ij}^2\|v_j\|^2\leq CJ$. So

$$\begin{aligned}\Delta_{i4}&=O_P(\sqrt{J})(J^d\omega_{np}\psi_{np}^{-1}+\sqrt{\sum_{j=1}^p\mathsf{E}\|v_j\|^4})=o_P(1).\\ \sum_{i\notin\mathcal{I}}\|\Delta_{i,4}\|^2&\leq O_P(1)\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p\varepsilon_{ij}x_{ij}v_j\|^2\leq O_P(nJ)=o_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2).\end{aligned}$$

(ii) **For term $\Delta_{i,5}$ under Assumption 4.3 (ii-b).**

First, $\|\gamma_i\|\leq O_P(J^b\psi_{np})\|u_i\|$. So

$$\begin{aligned}&\|\sum_{j=1}^p(\widetilde{v}_j-H_S'v_j)(v_j-H_S^{-1'}\widetilde{v}_j)'\gamma_i(x_{ij}^2-\mathsf{E}x_{ij}^2)\|\\ \leq\;& O_P(\|\gamma_i\|)\|\widetilde{V}_{\mathcal{I}}-VH_S\|_F^2=O_P(J^{b+2d}\omega_{np}^2\psi_{np}^{-1}\|u_i\|)=o_P(1).\end{aligned}$$

Next, by the assumption x_{ij}^2 is independent across i. Let $\widetilde{P}_{ki}$ denote $p\times 1$ vector of $v_{j,k}(v_j-\widetilde{v}_j)'\gamma_i$ and recall $X_i^2=(\dot{m}_{i1}^2,...,\dot{m}_{ip}^2)'$. Since $i\notin\mathcal{I}$, $\widetilde{V}_{\mathcal{I}}$ is estimated using subsample in $\mathcal{I}$, and x_{ij} is independent of Θ_0, we have that

$$\begin{aligned}&\mathsf{E}_I\|\sum_{j=1}^p v_j(v_j-H_S^{-1\prime}\widetilde{v}_j)'\gamma_i(x_{ij}^2-\mathsf{E}x_{ij}^2)\|^2=\sum_{k\leq J}\mathsf{E}_I\mathsf{Var}_I[\widetilde{P}_{ki}'X_i^2|\Theta_0]\\ \leq\ & (\mathsf{E}\max_{j\leq p}\|v_j\|^2)\|\widetilde{V}_{\mathcal{I}}-VH_S\|_F^2\|\gamma_i\|^2.\end{aligned}$$

This makes $\|\sum_{j=1}^p v_j(v_j-H_S^{-1\prime}\widetilde{v}_j)'\gamma_i(x_{ij}^2-\mathsf{E}x_{ij}^2)\|=O_P(\sqrt{\mathsf{E}\max_{j\leq p}\|v_j\|^2})\omega_{np}\|u_i\|J^{g+b}$. The last term is assumed to be $o_P(1)$. Hence $\|\Delta_{i,5}\|=o_P(1)$.

In addition,

$$\begin{aligned}\sum_{i\notin\mathcal{I}}\|\Delta_{i,5}\|^2\quad &\leq\quad C\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p(\widetilde{v}_j-H_S'v_j)(x_{ij}^2-\mathsf{E}x_{ij}^2)(v_j-H_S^{-1\prime}\widetilde{v}_j)'\gamma_i\|^2\\ &\quad +C\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p v_j(x_{ij}^2-\mathsf{E}x_{ij}^2)(v_j-H_S^{-1\prime}\widetilde{v}_j)'\gamma_i\|^2\\ &\leq\quad C\sum_{i\notin\mathcal{I}}\|\gamma_i\|^2\|(\widetilde{V}_{\mathcal{I}}-VH_S)'\mathsf{diag}\{X_i^2-\mathsf{E}X_i^2\}(\widetilde{V}_{\mathcal{I}}-VH_S)\|^2\\ &\quad +O_P(1)\sum_{i\notin\mathcal{I}}\|\gamma_i\|^2(\mathsf{E}\max_{j\leq p}\|v_j\|^2)\|\widetilde{V}_{\mathcal{I}}-VH_S\|_F^2\\ &\leq\quad O_P(J^{1+2b+4g}\omega_{np}^4\psi_{np}^{-2}+J^{1+2d+2b}\omega_{np}^2\mathsf{E}\max_{j\leq p}\|v_j\|^2)=O_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2).\end{aligned}\tag{C.3}$$

(iii) We now focus on condition (ii-c) of Assumption 4.3.

For terms $\Delta_{i,2},\Delta_{i,3}$.

These two terms are the same as under Assumption 4.3 (ii-a) and (ii-b), so their bounds are the same as that of part (i).

For terms $\Delta_{i,1},\Delta_{i,4}$.

By Lemma C.1, $\|\widehat{B}_i^{-1}-B^{-1}\|=O_P(J^d\omega_{np}\psi_{np}^{-1}+Jp^{-1}\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}1)$. Also

$$\|\sum_{j=1}^p\widetilde{v}_jx_{ij}^2(v_j-H_S^{-1\prime}\widetilde{v}_j)'\gamma_i\|\leq\|\gamma_i\|\|\widetilde{V}'\mathsf{diag}(X_i^2)(\widetilde{V}H_S^{-1}-V)\|\leq\|u_i\|O_P(\omega_{np}J^{b+g}).$$

Hence $\|\Delta_{i,1}\|=O_P(J^d\omega_{np}\psi_{np}^{-1}+Jp^{-1}\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}1)\|u_i\|\omega_{np}J^{b+g}=o_P(1)$.

Also, $\sum_{i\notin\mathcal{I}}\|\Delta_{i,1}\|^2\leq O_P(J^{1+2d+2b}\omega_{np}^2)=o_P(\psi_{np}^2)$.

Finally, $\mathsf{E}\|\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j\|^2 \leq C\sum_{j=1}^p x_{ij}^2\|v_j\|^2 \leq CJ$. So

$$\Delta_{i4} = O_P(\sqrt{J})(J^d\omega_{np}\psi_{np}^{-1} + Jp^{-1}\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}} 1) = o_P(1).$$

$$\sum_{i\notin\mathcal{I}}\|\Delta_{i,4}\|^2 \leq O_P(1)\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j\|^2 \leq O_P(nJ) = o_P(\psi_{np}^2).$$

For term $\Delta_{i,6}$.

Uniformly in i,

$$\begin{aligned}
|\Delta_{i,6}| &= |B^{-1}\sum_{j\in\mathcal{B}_i}\widetilde{v}_j(v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i - B^{-1}\sum_{j\in\bar{\mathcal{B}}}\widetilde{v}_j(v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i| \\
&\leq O_P(1)\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}\|\widetilde{v}_j(v_j - H_S^{-1'}\widetilde{v}_j)'\gamma_i\| \\
&\leq O_P(1)\|\gamma_i\|\|\widetilde{V}_S - V_0H_S\|^2 + O_P(1)\|\gamma_i\|\|\widetilde{V}_S - V_0H_S\|\max_j\|v_j\|\sqrt{\sum_j 1\{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}\}} \\
&\leq O_P(J^{2d}\omega_{np}^2\psi_{np}^{-2})\|\gamma_i\| + O_P(J^d\omega_{np}\psi_{np}^{-1})\|\gamma_i\|\max_j\|v_j\|\sqrt{\sum_j 1\{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}\}} \\
&\leq O_P(J^{2d+b}\omega_{np}^2\psi_{np}^{-1}\sqrt{Jn^{-1}}) + O_P(J^{g+b+1}\omega_{np}\sqrt{(np)^{-1}})\sqrt{\sum_j 1\{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}\}} = o_P(1),
\end{aligned} \tag{C.4}$$

where we used $\|\gamma_i\| = O_P(J^b\psi_{np})\|u_i\| = O_P(J^b\psi_{np}\sqrt{Jn^{-1}})$ and $\max_j\|v_j\| = O_P(\sqrt{Jp^{-1}})$ from Assumption 4.4. Finally,

$$\begin{aligned}
\sum_{i\notin\mathcal{I}}\|\Delta_{i,6}\|^2 &\leq \max_i\|\Delta_{i,6}\|^2 n \\
&\leq O_P(\omega_{np}^4\psi_{np}^{-2}J^{1+4d+2b} + \omega_{np}^2J^{2+2d+2b}p^{-1}\sum_j 1\{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}\}) \\
&= O_P(J^{1+2d+2b}\omega_{np}^2).
\end{aligned}$$

□

Lemma C.3. *(i)* $\|\widehat{\Gamma}_S - \Gamma_{0,S}H\|_F^2 = O_P(J^{1+2d+2b}\omega_{np}^2) = o_P(\psi_{np}^2)$ *where* $\Gamma_{0,S}$ *is the submatrix of* Γ_0 *corresponding to the rows in* $\mathcal{S}$ *(i.e., sample splitting).*
(ii) $\max_{j\leq p}\|\widehat{L}_{j,\mathcal{S}}^{-1}\| = O_P(\psi_{np}^{-2})$. *Recall that* $L_j = \sum_i x_{ij}^2\gamma_i\gamma_i'$
(iii) $\sum_{j\in\mathcal{G}}\|\widehat{L}_{j,\mathcal{S}}^{-1} - L_{j,\mathcal{S}}^{-1}\|^2 = O_P(\omega_{np}^2J^{1+2d+4b}\psi_{np}^{-6}|\mathcal{G}|)$. *Here* $\mathcal{G} = \{j : g_j \neq 0\}$.

Proof. (i) First not that $\|\Gamma_{0,S}\|^2 = O_P(J^{2b}\psi_{np}^2)$. It follows from (B.1) and Lemma C.2

$$\begin{aligned}\|\widehat{\Gamma}_S - \Gamma_{0,S}H\|_F^2 &\leq \sum_{i\notin S}\|B^{-1}H_S'\sum_{j=1}^p \varepsilon_{ij}x_{ij}v_j\|^2 + \sum_{k=1}^5\sum_{i\notin\mathcal{I}}\|\Delta_{i,k}\|^2 \\ &= O_P(\omega_{np}^2 J^{1+2d+2b}) = o_P(\psi_{np}^2).\end{aligned}$$

(ii) Now set $S=\mathcal{I}$. Write $\mathsf{diag}\{X_j^2\}$ be $|\mathcal{I}|_0\times|\mathcal{I}|_0$ diagonal matrix of x_{ij}^2. Then

$$\widehat{L}_{j,\mathcal{I}} = \widehat{\Gamma}_{\mathcal{I}}'\mathsf{diag}\{X_j^2\}\widehat{\Gamma}_{\mathcal{I}}, \quad L_{j,\mathcal{I}} = H'\Gamma_{0,\mathcal{I}}'\mathsf{diag}\{X_j^2\}\Gamma_{0,\mathcal{I}}H.$$

$$\begin{aligned}\max_{j\leq p}\|\widehat{L}_{j,\mathcal{I}} - L_{j,\mathcal{I}}\| &\leq 2\max_{j\leq p}\|(\widehat{\Gamma}_{\mathcal{I}} - \Gamma_{0,\mathcal{I}}H)'\mathsf{diag}\{X_j^2\}\Gamma_{0,\mathcal{I}}H\| \\ &\quad + \max_{j\leq p}\|(\widehat{\Gamma}_{\mathcal{I}} - \Gamma_{0,\mathcal{I}}H)'\mathsf{diag}\{X_j^2\}(\widehat{\Gamma}_{\mathcal{I}} - \Gamma_{0,\mathcal{I}}H)\| \\ &= O_P(\omega_{np}J^{1/2+d+2b}\psi_{np} + \omega_{np}^2 J^{1+2d+2b}) = o_P(\psi_{np}^2).\end{aligned} \tag{C.5}$$

Also, $\min_j \psi_{\min}(L_{j,\mathcal{I}}) \geq c\psi_{\min}(\Theta_0)^2 \geq c\psi_{np}^2$. Thus $\min_j \psi_{\min}(\widehat{L}_{j,\mathcal{I}}) \geq (c - o_P(1))\psi_{np}^2$.

(iii) $\sum_{j\in\mathcal{G}}\|\widehat{L}_{j,\mathcal{I}} - L_{j,\mathcal{I}}\|^2 \leq O_P(J^{1+2d+4b}\omega_{np}^2\psi_{np}^2|\mathcal{G}|)$. So

$$\sum_{j\in\mathcal{G}}\|\widehat{L}_{j,\mathcal{I}}^{-1} - L_{j,\mathcal{I}}^{-1}\|^2 \leq \max_j\|\widehat{L}_{j,\mathcal{I}}^{-2}\|\|L_{j,\mathcal{I}}^{-2}\|\sum_{j\in\mathcal{G}}\|\widehat{L}_{j,\mathcal{I}} - L_{j,\mathcal{I}}\|^2 = O_P(\omega_{np}^2\psi_{np}^{-6}|\mathcal{G}|J^{1+2d+4b}).$$

□

In lemmas below, Δ_d, $d = 1\sim 4$ are defined in (B.4).

Lemma C.4. $\|\Delta_1\| + \|\Delta_2\| = O_P(\omega_{np}\psi_{np}^{-1}J^{1+d+3b} + J^{b+1/2}\max_{ij}|r_{ij}|\sqrt{n})\psi_{np}^{-1}\|g\|\sqrt{|\mathcal{G}|}$.

Proof. First $\mathsf{E}\sum_{j=1}^p\|\sum_{i\notin\mathcal{I}}H_S'\gamma_i\varepsilon_{ij}x_{ij}g_j\|^2 \leq CJ^{2b+1}\psi_{np}^2\|g\|^2$.

By Lemma C.3, note $\sum_{j=1}^p|g_j| \leq \|g\|\sqrt{|\mathcal{G}|}$,

$$\begin{aligned}\|\Delta_1\| &\leq \max_{ij}\|\widehat{L}_{j,\mathcal{I}}^{-1}r_{ij}x_{ij}^2\|\sum_{j=1}^p|g_j|\sum_{i\notin\mathcal{I}}\|\widehat{\gamma}_i\| \leq J^b\max_{ij}|r_{ij}|\sqrt{nJ}\psi_{np}^{-1}\|g\|\sqrt{|\mathcal{G}|}. \\ \|\Delta_2\| &\leq \left(\sum_{k\in\mathcal{G}}\|\widehat{L}_{k,\mathcal{I}}^{-1} - L_{k,\mathcal{I}}^{-1}\|^2\sum_{j=1}^p\|\sum_{i\notin\mathcal{I}}H_S'\gamma_i\varepsilon_{ij}x_{ij}g_j\|^2\right)^{1/2} \leq O_P(\omega_{np}\psi_{np}^{-2}|\mathcal{G}|^{1/2}J^{1+d+3b}\|g\|).\end{aligned}$$

□

Lemma C.5. *Suppose* $\mathsf{E}(\varepsilon_{ij}|X,\Theta) = 0$. *Then*

$$\|\Delta_3\|^2 \leq \sum_{j\in\mathcal{G}}\|v_j\|^2\|g\|^2\psi_{np}^{-4}n^2 + \omega_{np}^4\psi_{np}^{-6}nJ^{2+4d+6b}\|g\|^2|\mathcal{G}| + O_P(J^{1+2b+2d}\psi_{np}^{-4}\|g\|^2\omega_{np}^2).$$

$$+O_P\left(J^{2d}\omega_{np}^2\psi_{np}^{-2}+\sum_{j=1}^p\mathsf{E}\|v_j\|^4+J^2p^{-2}[\sum_{\mathcal{B}_i\Delta\bar{\mathcal{B}}}1]^2\right)\omega_{np}^2\psi_{np}^{-4}\|g\|^2nJ^{2d+2+2b}.$$

Proof. Let $S=\mathcal{I}$. We now bound $\|\Delta_3\|$. First, $\sum_{j=1}^p\sum_{i\notin\mathcal{I}}\varepsilon_{ij}^2x_{ij}^2g_j^2=O_P(n\|g\|^2)$.

$$\begin{aligned}\|\Delta_3\|^2 &\leq \|\sum_{j=1}^p\widehat{L}_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}(\widehat{\gamma}_i-H'\gamma_i)\varepsilon_{ij}x_{ij}g_j\|^2\leq\|\sum_{j=1}^pL_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}(\widehat{\gamma}_i-H'\gamma_i)\varepsilon_{ij}x_{ij}g_j\|^2\\ &\quad+\sum_{j\in\mathcal{G}}\|\widehat{L}_{j,\mathcal{I}}^{-1}-L_{j,\mathcal{I}}^{-1}\|^2\|\widehat{\Gamma}_{\mathcal{I}}-\Gamma_{0,\mathcal{I}}H\|_F^2\sum_{j=1}^p\sum_{i\notin\mathcal{I}}\varepsilon_{ij}^2x_{ij}^2g_j^2\\ &\leq \|\sum_{j=1}^pL_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}B^{-1}\sum_{k=1}^pv_k\varepsilon_{ik}\varepsilon_{ij}x_{ij}x_{ik}g_j\|^2+\|\sum_{j=1}^pL_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\sum_{k=1}^5\Delta_{i,k}\varepsilon_{ij}x_{ij}g_j\|^2\\ &\quad+O_P(\omega_{np}^4\psi_{np}^{-6}nJ^{2+4d+6b}\|g\|^2|\mathcal{G}|).\end{aligned}$$

The first term is, let $L_{j,\mathcal{I}}^{-1}B^{-1}v_k=(a_{jk,1},...,a_{jk,J})'$. Then

$$\begin{aligned}&\mathsf{E}\|\sum_{j=1}^pL_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}B^{-1}\sum_{k=1}^pv_k\varepsilon_{ik}\varepsilon_{ij}x_{ij}x_{ik}g_j\|^2\leq\sum_{t=1}^J\mathsf{E}[\sum_{j=1}^p\sum_{i\notin\mathcal{I}}\sum_{k=1}^pa_{jk,t}\varepsilon_{ik}\varepsilon_{ij}x_{ij}x_{ik}g_j]^2\\ \leq&\sum_{t=1}^J\sum_{j=1}^p\sum_{i\notin\mathcal{I}}\sum_{k=1}^pg_jg_k\mathsf{E}a_{jk,t}x_{ij}^2x_{ik}^2a_{kj,t}\mathsf{E}\varepsilon_{ik}^2\mathsf{E}\varepsilon_{ij}^2\\ &+\sum_{t=1}^J\sum_{j=1}^p\sum_{i\notin\mathcal{I}}\sum_{j'=1}^p\sum_{i'\neq i}\mathsf{E}a_{jj,t}x_{ij}^2g_j(\mathsf{E}\varepsilon_{ij}^2)a_{j'j',t}\dot{m}_{i'j'}^2g_{j'}\mathsf{E}\varepsilon_{i'j'}^2\\ \leq&\mathsf{E}[\sum_{j=1}^p\|v_jg_j\|]^2O_P(\psi_{np}^{-4}n^2)\leq\mathsf{E}\sum_{j\in\mathcal{G}}\|v_j\|^2\|g\|^2O_P(\psi_{np}^{-4}n^2).\end{aligned}$$

We now bound the second term. Note

$$\mathsf{E}\|\sum_{j=1}^p\varepsilon_{ij}x_{ij}g_jL_{j,\mathcal{I}}^{-1}\|_F^2=O(\|g\|^2\psi_{np}^{-4}J).$$

(i) Lemma C.1 shows
$\sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1}\|^2=O_P(nJ^{2d}\omega_{np}^2\psi_{np}^{-2}+n\sum_{j=1}^p\mathsf{E}\|v_j\|^4+nJ^2p^{-2}[\sum_{\mathcal{B}_i\Delta\bar{\mathcal{B}}}1]^2)$.
Then

$$\begin{aligned}d &:= \sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1}\|^2\|\widetilde{V}_{\mathcal{I}}-V_{0,\mathcal{I}}H_S\|^2\\ &= O_P\left(J^{2d}\omega_{np}^2\psi_{np}^{-2}+\sum_{j=1}^p\mathsf{E}\|v_j\|^4+J^2p^{-2}[\sum_{\mathcal{B}_i\Delta\bar{\mathcal{B}}}1]^2\right)\omega_{np}^2\psi_{np}^{-2}J^{2d}n.\end{aligned}$$

Next, let b'_{ju} be the u th row of $L_{j,\mathcal{I}}^{-1}$ and $a_{ij}=\sum_{k=1}^p \widetilde{v}_k x_{ik}^2(v_k-\widetilde{v}_k H_S^{-1'})'\gamma_i\varepsilon_{ij}x_{ij}g_j$. Then

$$
\begin{aligned}
&\|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\Delta_{i,1}\varepsilon_{ij}x_{ij}g_j\|^2\\
\leq\ & \|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}(\widehat{B}_i^{-1}-B^{-1})\sum_{k=1}^p \widetilde{v}_k x_{ik}^2(v_k-\widetilde{v}_k H_S^{-1'})'\gamma_i\varepsilon_{ij}x_{ij}g_j\|^2\\
\leq\ & \sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1}\|^2\sum_{u=1}^J\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p a_{ij}b'_{ju}\|_F^2 \leq O_P(d)\sum_{i\notin\mathcal{I}}\|\gamma_i\|^2\|\sum_{j=1}^p\varepsilon_{ij}x_{ij}g_jL_{j,\mathcal{I}}^{-1}\|_F^2\\
\leq\ & O_P\left(J^{2d}\omega_{np}^2\psi_{np}^{-2}+\sum_{j=1}^p\mathsf{E}\|v_j\|^4+J^2p^{-2}[\sum_{\mathcal{B}_i\Delta\bar{\mathcal{B}}}1]^2\right)\omega_{np}^2\psi_{np}^{-4}\|g\|^2nJ^{2d+2+2b}.
\end{aligned}
$$

(ii) Let $a_i=\sum_{k=1}^p\varepsilon_{ik}x_{ik}(\widetilde{v}_k-H'_Sv_k)$, $b_{ij}=L_{j,\mathcal{I}}^{-1}\varepsilon_{ij}x_{ij}g_j$. First, by (C.1),

$$
\begin{aligned}
&\sum_{i\notin\mathcal{I}}\|a_i\|^2=O_P(nJ^{2d}\omega_{np}^2\psi_{np}^{-2}).\\
&\|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\Delta_{i,2}\varepsilon_{ij}x_{ij}g_j\|^2\leq\|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\widehat{B}_i^{-1}\sum_{k=1}^p\varepsilon_{ik}x_{ik}(\widetilde{v}_k-H'_Sv_k)\varepsilon_{ij}x_{ij}g_j\|^2\\
\leq\ & \sum_{i\notin\mathcal{I}}\|a_i\|^2\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p b_{ij}\|^2\max_i\|\widehat{B}_i^{-1}\|^2\leq O_P(\omega_{np}^4\psi_{np}^{-6}nJ^{2+4d+6b}\|g\|^2|\mathcal{G}|).
\end{aligned}
$$

(iii) Note that $\mathsf{E}(\varepsilon_{ij}|\widetilde{V}_{\mathcal{I}},X,\Theta)=0$. Let $a_{jk,d}$ be the d th element of $L_{j,\mathcal{I}}^{-1}B^{-1}\widetilde{v}_k$.

$$
\begin{aligned}
&\|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\Delta_{i,3}\varepsilon_{ij}x_{ij}g_j\|^2\leq\|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}(\widehat{B}_i^{-1}-B^{-1})\sum_{k=1}^p\widetilde{v}_kx_{ik}^2r_{ik}\varepsilon_{ij}x_{ij}g_j\|^2\\
&+\|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}B^{-1}\sum_{k=1}^p\widetilde{v}_kx_{ik}^2r_{ik}\varepsilon_{ij}x_{ij}g_j\|^2\\
\leq\ & J\sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1}\|^2\sum_{i\notin\mathcal{I}}\sum_{k=1}^p x_{ik}^4r_{ik}^2\|\sum_{j=1}^p\varepsilon_{ij}x_{ij}g_jL_{j,\mathcal{I}}^{-1}\|_F^2\\
&+O_P(1)\mathsf{E}_I\|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}B^{-1}\sum_{k=1}^p\widetilde{v}_kx_{ik}^2r_{ik}\varepsilon_{ij}x_{ij}g_j\|^2\\
\leq\ & (J\sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1}\|^2+1)\|R\|_F^2\|g\|^2\psi_{np}^{-4}J\\
\leq\ & \left(J^{1+2d}n\omega_{np}^2\psi_{np}^{-2}+nJ^2\mathsf{E}\max_{j\leq p}\|v_j\|^2+nJ^3p^{-2}[\sum_{\mathcal{B}_i\Delta\bar{\mathcal{B}}}]^2+1\right)\|R\|_F^2\|g\|^2\psi_{np}^{-4}J.
\end{aligned}
$$

We note that this term is dominated by other terms. In particular,

$$\begin{aligned} & \left(nJ^3p^{-2}[\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}} 1]^2 \right) \|R\|_F^2\|g\|^2\psi_{np}^{-4}J \\ = \;& o_P\left(\frac{np}{(n+p)J^3+J\|R\|_{(n)}^2}\right)^2 \|R\|_F^2\|g\|^2\psi_{np}^{-4}nJ^4p^{-2} \\ \leq \;& o_P\left(\frac{n}{J^2}\right)\|R\|_F^2\|g\|^2\psi_{np}^{-4} \text{ dominated by the (v) term below.} \end{aligned}$$

(iv) Let b_{ju}' be the u th row of $L_{j,\mathcal{I}}^{-1}$ and $a_{ij}=\sum_{k=1}^p \varepsilon_{ik}x_{ik}v_k\varepsilon_{ij}x_{ij}g_j$, and $f:=\sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1}\|^2$, $c_{iutqkj}=v_{k,q}x_{ij}x_{ik}g_jb_{ju,t}$

$$\begin{aligned} & \|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\Delta_{i,4}\varepsilon_{ij}x_{ij}g_j\|^2 \leq \|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}(\widehat{B}_i^{-1}-B^{-1})\sum_{k=1}^p\varepsilon_{ik}x_{ik}v_k\varepsilon_{ij}x_{ij}g_j\|^2 \\ \leq \;& f\sum_{u=1}^J\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^p a_{ij}b_{ju}'\|_F^2 \leq O_P(f)\sum_{t,q,u\leq J}\sum_{i\notin\mathcal{I}}\mathsf{E}[|\sum_{j=1}^p\sum_{k=1}^p\varepsilon_{ik}\varepsilon_{ij}c_{ikjutq}|^2|X,\Theta] \\ \leq \;& O_P(f)\sum_{t,q,u\leq J}\sum_{i\notin\mathcal{I}}\sum_{k=1}^p\sum_{j=1}^p c_{iutqkj}^2 + O_P(f)\sum_{t,q,u\leq J}\sum_{i\notin\mathcal{I}}(\sum_{k=1}^p c_{iutqkk})^2 \\ \leq \;& O_P(fn)\sum_{j=1}^p Jg_j^2\|L_j^{-1}\|_F^2 \leq \text{the bound for } \|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\Delta_{i,1}\varepsilon_{ij}x_{ij}g_j\|^2. \end{aligned}$$

(v) Let $a_{jk,t}$ be the t th element of $L_{j,\mathcal{I}}^{-1}B^{-1}\widetilde{v}_k$. Because $\mathsf{E}(\varepsilon_{ij}|\Theta,X,\mathcal{I})=0$, and $\mathsf{E}(x_{ik}^2-\mathsf{E}x_{ik}^2|\Theta,\mathcal{I})=0$. Let A_{ijt} be p-dimensional vector of $a_{jk,t}(v_k-\widetilde{v}_k)'\gamma_i$. Let $e_k'=(0,..,0,1,0...)$. Then $\max_{k\leq p}\|\widetilde{v}_k\|=\max_k\|e_k'\widetilde{V}\|\leq\max_k\|e_k\|=1$. Under Assumption 4.3 (ii-b),

$$\begin{aligned} & \|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\Delta_{i,5}\varepsilon_{ij}x_{ij}g_j\|^2 \\ \leq \;& \|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}B^{-1}\sum_{k=1}^p\widetilde{v}_k(x_{ik}^2-\mathsf{E}x_{ik}^2)(v_k-H_S^{-1'}\widetilde{v}_k)'\gamma_i\varepsilon_{ij}x_{ij}g_j\|^2 \\ \leq \;& O_P(1)\sum_{t\leq J}\sum_{i\notin\mathcal{I}}\sum_{j=1}^p\mathsf{E}_I\mathsf{Var}_I[\sum_{k=1}^p a_{jk,t}(x_{ik}^2-\mathsf{E}x_{ik}^2)(v_k-H_S^{-1'}\widetilde{v}_k)'\gamma_i\varepsilon_{ij}|\Theta,X]x_{ij}^2g_j^2 \\ \leq \;& O_P(1)\sum_{t\leq J}\sum_{i\notin\mathcal{I}}\sum_{j=1}^p\mathsf{E}_IA_{ijt}'\mathsf{Var}_I[X_i^2|\Theta]A_{ijt}g_j^2 \end{aligned}$$

$$\begin{aligned}&\leq O_P(1)\sum_{i\notin\mathcal{I}}\sum_{j=1}^{p}\mathsf{E}_I\sum_{k=1}^{p}\|a_{jk}\|^2[(v_k-H_S^{-1'}\widetilde{v}_k)'\gamma_i]^2g_j^2\\&\leq O_P(J^{1+2b}\psi_{np}^{-2})\|g\|^2\max_k\|\widetilde{v}_k\|^2\|V-\widetilde{V}H_S^{-1}\|^2\leq O_P(J^{1+2b+2d}\psi_{np}^{-4}\|g\|^2\omega_{np}^2).\end{aligned}$$

(vi) Recall $\Delta_{i,6}=B^{-1}\sum_{j=1}^p\widetilde{v}_jx_{ij}^2(v_j-H_S^{-1'}\widetilde{v}_j)'\gamma_i-B^{-1}\sum_{j\in\bar{\mathcal{B}}}\widetilde{v}_j(v_j-H_S^{-1'}\widetilde{v}_j)'\gamma_i$. Because $\mathsf{E}(\varepsilon_{ij}|\Theta,X,\mathcal{I})=0$, we have $\mathsf{E}(\varepsilon_{ij}|\Theta,X,\mathcal{I},\Delta_{i,6})=0$. So by the bound for $\max_i\Delta_{i,6}$ in (C.4),

$$\begin{aligned}&\|\sum_{j=1}^{p}L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}\Delta_{i,6}\varepsilon_{ij}x_{ij}g_j\|^2\leq O_P(1)\sum_{j=1}^{p}\sum_{i\notin\mathcal{I}}\|L_{j,\mathcal{I}}^{-1}\|^2\|\Delta_{i,6}\|^2g_j^2x_{ij}^2\mathsf{Var}(\varepsilon_{ij}|\Theta,X,\mathcal{I})\\&\leq O_P(n\psi_{np}^{-4}\|g\|^2)\max_i\|\Delta_{i,6}\|^2\leq O_P(J^{1+2b+2d}\psi_{np}^{-4}\|g\|^2\omega_{np}^2).\end{aligned}$$

Putting together we obtain the desired result.

□

Lemma C.6. *Suppose* $\mathsf{E}(\varepsilon_{ij}|X,\Theta)=0$. *Then*

$$\begin{aligned}\|\Delta_4\|^2\ &\leq\ O_P\left(J^{2+2d+2b}\omega_{np}^2\mathsf{E}\max_{j\leq p}\|v_j\|^2+npJ^2\max_{ij}r_{ij}^2+\psi_{np}^{-2}J^{2+4d+2b}\omega_{np}^4\right)\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2\psi_{np}^{-2}J^{2b}\\&\quad+\left(\omega_{np}^4\psi_{np}^{-2}J^{3+4d+8b}+J^{2+2b}\right)\sum_{j=1}^{p}\|v_jg_j\|^2|\mathcal{G}|\psi_{np}^{-2}.\end{aligned}$$

Proof. First,

$$\begin{aligned}&\sum_{j=1}^{p}\sum_{i\notin\mathcal{I}}\|H^{-1}v_j\widehat{\gamma}_i'x_{ij}^2g_j\|^2\leq\sum_{i\notin\mathcal{I}}\|\widehat{\gamma}_i'\|^2\sum_{j=1}^{p}\|v_j\|^2g_j^2=O_P(J^{1+2b}\psi_{np}^2)\sum_{j=1}^{p}\|v_j\|^2g_j^2\\&\mathsf{E}\|\sum_{i\notin\mathcal{I}}x_{ij}^2\sum_{k=1}^{p}\gamma_i\varepsilon_{ik}x_{ik}v_k'\|^2\leq O(J^{2+2b}\psi_{np}^2),\quad\mathsf{E}\|\sum_{k=1}^{p}\varepsilon_{ik}x_{ik}H^{-1'}v_k\|^2=O(J).\end{aligned}$$

Next, define and bound: (using Lemmas C.2 C.3)

$$\begin{aligned}I\ &:=\ \|\sum_{j=1}^{p}L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}}x_{ij}^2B^{-1}\sum_{k=1}^{p}\varepsilon_{ik}x_{ik}v_k'H^{-1}v_j\widehat{\gamma}_ig_j\|^2\\&\leq\ \sum_{j\in G}\|L_{j,\mathcal{I}}^{-1}B^{-1}\|^2\|\sum_{i\notin\mathcal{I}}x_{ij}^2\sum_{k=1}^{p}\gamma_i\varepsilon_{ik}x_{ik}v_k'\|^2\|H^{-1}\|^2\sum_{j\in\mathcal{G}}\|v_jg_j\|^2\\&\quad+\sum_{i\notin\mathcal{I}}\|\sum_{j=1}^{p}x_{ij}^2g_jL_{j,\mathcal{I}}^{-1}B^{-1}v_j\|^2\|\sum_{k=1}^{p}\varepsilon_{ik}x_{ik}H^{-1'}v_k\|^2\sum_{i\notin\mathcal{I}}\|(\widehat{\gamma}_i-H'\gamma_i)\|^2\\&\leq\ O_P(|\mathcal{G}|\psi_{np}^{-2}J^{2+2b}\sum_{j\in G}\|v_jg_j\|^2+n\psi_{np}^{-4}\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2J^{2+2d+2b}\omega_{np}^2.\end{aligned}$$

$$
\begin{aligned}
II &:= \sum_{d=1}^{6}\|\sum_{j=1}^{p} L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}} x_{ij}^2\Delta_{i,d}' H^{-1}v_j(\widehat{\gamma}_i - H'\gamma_i)g_j\|^2 \\
&\leq O_P(1)[\sum_{d=1}^{6}\sum_{i\notin\mathcal{I}}\|\Delta_{i,d}\|^2]\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2\sum_{i\notin\mathcal{I}}\|(\widehat{\gamma}_i - H'\gamma_i)\|^2\max_j\|L_{j,\mathcal{I}}^{-1}\|^2 \\
&\leq O_P(J^{2+4d+4b}\omega_{np}^4\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2\psi_{np}^{-4}). \\
III &:= \sum_{j\in\mathcal{G}}\|\widehat{L}_{j,\mathcal{I}}^{-1} - L_{j,\mathcal{I}}^{-1}\|^2\sum_{i\notin\mathcal{I}}\|\gamma_i'H - \widehat{\gamma}_i'\|^2\sum_{j=1}^{p}\sum_{i\notin\mathcal{I}}\|H^{-1}v_j\widehat{\gamma}_i' x_{ij}^2 g_j\|^2 \\
&= O_P(\omega_{np}^4\psi_{np}^{-4}|\mathcal{G}|J^{3+4d+8b}\sum_{j=1}^{p}\|v_j\|^2 g_j^2). \\
\|\Delta_4\|^2 &= \|\sum_{j=1}^{p}\widehat{L}_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}} x_{ij}^2(\gamma_i'H - \widehat{\gamma}_i')H^{-1}v_j\widehat{\gamma}_i g_j\|^2 \\
&\leq I + II + III + \sum_{d=1}^{6}\|\sum_{j=1}^{p} L_{j,\mathcal{I}}^{-1}\sum_{i\notin\mathcal{I}} H'\gamma_i x_{ij}^2 v_j' H^{-1'}\Delta_{i,d} g_j\|^2 \\
&\leq (\omega_{np}^4\psi_{np}^{-2}J^{3+4d+8b} + J^{2+2b})\sum_{j=1}^{p}\|v_j g_j\|^2|\mathcal{G}|\psi_{np}^{-2} + \sum_{j\in\mathcal{G}}\|v_j\|^2\|g\|^2\psi_{np}^{-4}J^{2+4d+4b}\omega_{np}^4 \\
&\quad + \sum_{d=1}^{6}\|\sum_{j=1}^{p} L_{j,\mathcal{I}}^{-1}H'\sum_{i\notin\mathcal{I}} g_j\gamma_i x_{ij}^2 v_j' H^{-1'}\Delta_{i,d}\|^2.
\end{aligned}
$$

We now bound the last term on the right hand side. Recall that

$$
\sum_{i\notin\mathcal{I}}\|(\widehat{B}_i^{-1} - B^{-1})\|^2 = O_P(J^{2d}n\omega_{np}^2\psi_{np}^{-2} + n\sum_{j=1}^{p}\mathsf{E}\|v_j\|^4 + nJ^2p^{-2}[\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}1]^2).
$$

(i) Recall $\sum_{i\notin\mathcal{I}}\|\gamma_i\|^2 = O_P(J^{1+2b}\psi_{np}^2)$.

$$
\begin{aligned}
&\|\sum_{j=1}^{p} L_{j,\mathcal{I}}^{-1}H'\sum_{i\notin\mathcal{I}} g_j\gamma_i x_{ij}^2 v_j' H^{-1'}\Delta_{i,1}\|^2 \\
\leq\ & O_P(1)\sum_{i\notin\mathcal{I}}\|(\widehat{B}_i^{-1} - B^{-1})\|^2\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2 J\|\widetilde{V}_S - V_0H_S\|_F^2\sum_{i\notin\mathcal{I}}\|\gamma_i\|^2\max_j\|L_{j,\mathcal{I}}^{-1}H'\|^2 \\
\leq\ & O_P(\omega_{np}^2\psi_{np}^{-2}J^{2d} + \sum_{j=1}^{p}\mathsf{E}\|v_j\|^4 + J^2p^{-2}[\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}1]^2)\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2 n\omega_{np}^2\psi_{np}^{-4}J^{2d+1+2b}.
\end{aligned}
$$

We note that the term involving $\sum_{j\in\mathcal{B}_i\triangle\bar{\mathcal{B}}}1$ is dominated by other terms so can be ignored.

(ii) As $\mathsf{E}_I(\varepsilon_{ik}|\widehat{B}_i,\widetilde{V},\Theta,X)=0$ for $i\notin\mathcal{I}$, ($\widehat{B}_i,\widetilde{V}$ are estimated using data in $\mathcal{I}$),

$$
\begin{aligned}
&\|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}H'\sum_{i\notin\mathcal{I}} g_j\gamma_i x_{ij}^2 v_j' H^{-1'}\Delta_{i,2}\|^2 \\
\leq\ & \|\sum_{j\in\mathcal{G}} g_j L_{j,\mathcal{I}}^{-1}H'\sum_{i\notin\mathcal{I}}\gamma_i x_{ij}^2 v_j' H^{-1'}\widehat{B}_i^{-1}\sum_{k=1}^p \varepsilon_{ik}x_{ik}(\widetilde{v}_k-H_S'v_k)\|^2 \\
\leq\ & O_P(1)\sum_{j\in\mathcal{G}}\|g_jL_{j,\mathcal{I}}^{-1}H'\|^2\sum_{t=1}^J\sum_{j\in\mathcal{G}}\mathsf{Var}_I[\sum_{k=1}^p\sum_{i\notin\mathcal{I}}\gamma_{i,t}x_{ij}^2v_j'H^{-1'}\widehat{B}_i^{-1}\varepsilon_{ik}x_{ik}(\widetilde{v}_k-H_S'v_k)] \\
\leq\ & O_P(1)\sum_{j\in\mathcal{G}}\|g_jL_{j,\mathcal{I}}^{-1}H'\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2\sum_{i\notin\mathcal{I}}\|\gamma_i\|^2\|\widetilde{V}_S-V_0H_S\|_F^2 \\
\leq\ & O_P(\psi_{np}^{-4}\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2J^{1+2b+2d}\omega_{np}^2)
\end{aligned}
$$

(iii) By (C.2)(C.3) $\sum_{d=3,5}\sum_i\|\Delta_{i,d}\|^2=O_P(a_n)$, (sharper bound than the conclusion in Lemmas C.2), where

$$
a_n:=J^{1+2b+4g}\omega_{np}^4\psi_{np}^{-2}+J^{1+2d+2b}\omega_{np}^2\mathsf{E}\max_{j\leq p}\|v_j\|^2+npJ\max_{ij}r_{ij}^2.
$$

Hence

$$
\begin{aligned}
&\sum_{d=3,5}\|\sum_{i\notin\mathcal{I}}\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}H'g_j\gamma_i x_{ij}^2v_j'H^{-1'}\Delta_{i,d}\|^2 \\
\leq\ & \sum_{i\notin\mathcal{I}}\|\gamma_i\|^2\sum_{j=1}^p\|L_{j,\mathcal{I}}^{-1}H'g_j\|^2\sum_{j\in\mathcal{G}}\|v_j'\|^2\sum_{d=3,5}\sum_{i\notin\mathcal{I}}\|\Delta_{i,d}\|^2 \\
\leq\ & O_P(J^{1+2b}\psi_{np}^{-2}\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2)\sum_{d=3,5}\sum_{i\notin\mathcal{I}}\|\Delta_{i,d}\|^2\leq O_P(a_n)J^{1+2b}\psi_{np}^{-2}\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2.
\end{aligned}
$$

This is the leading term.

(iv) $\mathsf{E}_I(\varepsilon_{ik}|H,\widehat{B}_i,\widetilde{V},\Theta,X)=0$,

$$
\begin{aligned}
&\|\sum_{i\notin\mathcal{I}}\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}H'g_j\gamma_i x_{ij}^2v_j'H^{-1'}\Delta_{i,4}\|^2 \\
\leq\ & \|\sum_{j=1}^p L_{j,\mathcal{I}}^{-1}H'g_j\sum_{k=1}^p\sum_{i\notin\mathcal{I}}\gamma_i x_{ij}^2v_j'H^{-1'}(\widehat{B}_i^{-1}-B^{-1})\varepsilon_{ik}x_{ik}v_k\|^2 \\
\leq\ & O_P(1)\sum_{j=1}^p\|L_{j,\mathcal{I}}^{-1}H'g_j\|^2\sum_{j\in\mathcal{G}}\sum_{t=1}^J\mathsf{Var}_I[\sum_{k=1}^p\sum_{i\notin\mathcal{I}}\gamma_{i,t}x_{ij}^2v_j'H^{-1'}(\widehat{B}_i^{-1}-B^{-1})v_k\varepsilon_{ik}x_{ik}] \\
\leq\ & O_P(1)\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2\sum_{i\notin\mathcal{I}}\|\widehat{B}_i^{-1}-B^{-1}\|^2\|\gamma_i\|^2J\psi_{np}^{-4}
\end{aligned}
$$

$$\leq \quad O_P(1)(\max_i \|u_i\|^2 n J^{2d}\omega_{np}^2\psi_{np}^{-2} + n\max_i \|u_i\|^2 \sum_{j=1}^p \mathsf{E}\|v_j\|^4)J^{2b}\psi_{np}^{-2}\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2.$$

This is dominated by other terms. Also the last term involving $\Delta_{i,6}$, depending on $\bar{\mathcal{B}}\Delta\mathcal{B}_i$, is dominated by other terms. So together leads to the desired result.

□

We now verify (B.6) using more primitive conditions.

Lemma C.7. *Suppose* $\max_{ij}|r_{ij}|^2(p\vee n)^3J^3 = o(1)$. *Then*

$$|r_i'g| + J^b\|u_i\|\psi_{np}\xi_{np} = o_P(p^{-1/2} + \|u_i\|\|g\|).$$

holds true when either case (I) or case (II) below holds:

(I) sparse g: $\|g\|\leq C$,
$(p\vee n)^{3/4}J^{1+d+2b} = o(\psi_{np})$ *and* $J^{3+2d+6b} = o(\min(\sqrt{p}, p/\sqrt{n}))$.
(II) dense g: $\max_{j\leq p}|g_j|\leq Cp^{-1}$.
$(p\vee n)^{3/4}J^{5/4+d+2b} + J^{7/2+5b+2d}pn^{-1/2} = o(\psi_{np})$, *and* $J^{3+3b+g} = o(\min(\sqrt{n},\sqrt{p}))$.

Proof. First,

$$\begin{aligned}
\xi_{np}^2 &:= \sum_{d=1}^6 a_d + \mu_{np}^2\\
a_1 &= \left(\omega_{np}^4\psi_{np}^{-2}J^{3+4d+8b} + J^{2+2b}\right)\sum_{j=1}^p\|v_jg_j\|^2|\mathcal{G}|\psi_{np}^{-2}\\
a_{25} &= \left(J^{2+2d+4b}\omega_{np}^2\mathsf{E}\max_{j\leq p}\|v_j\|^2 + \psi_{np}^{-2}J^{2+4d+4b}\omega_{np}^4\right)\|g\|^2\sum_{j\in\mathcal{G}}\|v_j\|^2\psi_{np}^{-2}\\
a_3 &= \left(\omega_{np}^2\psi_{np}^{-2}nJ^{2d}+1\right)\psi_{np}^{-4}\|g\|^2|\mathcal{G}|\omega_{np}^2J^{2+2d+6b}\\
a_4 &= \left(1 + nJ^{2d+1}\omega_{np}^2\psi_{np}^{-2} + nJ\sum_{j=1}^p\mathsf{E}\|v_j\|^4\right)\omega_{np}^2\psi_{np}^{-4}\|g\|^2J^{1+2d+2b}\\
a_5 &= \left(pJ\sum_{j\in\mathcal{G}}\|v_j\|^2 + |\mathcal{G}|\right)n\max_{ij}|r_{ij}|^2\|g\|^2\psi_{np}^{-2}J^{2b+1}\\
\mu_{np}^2 &:= np^{-2}\omega_{np}^2\psi_{np}^{-4}\|g\|^2J^{4+2d+2b}[\sum_{\mathcal{B}_i\Delta\bar{\mathcal{B}}}1]^2.
\end{aligned}$$

Therefore, we aim to show

$$|r_i'g|^2 + J^{2b}\|u_i\|^2\psi_{np}^2 a_5 = o_P(p^{-1} + \|u_i\|^2\|g\|^2) \quad \text{(C.6)}$$

$$J^{2b}\|u_i\|^2\psi_{np}^2(a_1 + ... + a_4) = o_P(p^{-1} + \|u_i\|^2\|g\|^2) \quad \text{(C.7)}$$

$$J^b\|u_i\|\psi_{np}\mu_{np} = o_P(p^{-1/2} + \|u_i\|\|g\|). \quad \text{(C.8)}$$

Under $\max_{ij}|r_{ij}|^2(p\vee n)^2J^{3+4b} = o(1)$, (C.6) holds. In particular, $pJ\sum_{j\in\mathcal{G}}\|v_j\|^2 + |\mathcal{G}| \leq O_P(pJ^2)$ and

$$J^{2b}\psi_{np}^2 a_5 = o_P(\|g\|^2). \quad \text{(C.9)}$$

Proof of (C.7). We consider two cases.

CASE I: sparse g. In this case, $\|g\| + |\mathcal{G}| < C$ and we use bound: $\sum_{j\in\mathcal{G}}\|v_j\|^2 \leq \max_{j\in\mathcal{G}}\|v_j\|^2|\mathcal{G}| \leq O_P(J/p)$. We have the following bounds

$$\begin{aligned} a_1 + a_2 &\leq O_P\left(\omega_{np}^2\psi_{np}^{-2}J^{2d+4b} + p^{-1}\right)J^{4+2d+4b}p^{-1}\omega_{np}^2\psi_{np}^{-2}\|g\|^2 \\ a_3 &= O_P\left(\omega_{np}^2\psi_{np}^{-2}nJ^{2d} + 1\right)\psi_{np}^{-4}\omega_{np}^2J^{2+2d+6b}\|g\|^2 \\ a_4 &= O_P\left(1 + nJ^{2d+1}\omega_{np}^2\psi_{np}^{-2} + nJ^3p^{-1}\right)\omega_{np}^2\psi_{np}^{-4}J^{1+2d+2b}\|g\|^2 \end{aligned}$$

So $J^{2b}\|u_i\|^2\psi_{np}^2(a_1 + ... + a_4) = o_P(\|u_i\|^2\|g\|^2)$ holds as long as: $(p\vee n)^{3/4}J^{1+d+2b} = o(\psi_{np})$ and $J^{3+2d+6b} \ll \min(\sqrt{p}, p/\sqrt{n})$.

CASE II: dense g. In this case $\max_{j\leq p}|g_j| \leq Cp^{-1}$. $\|g\|^2 \leq Cp^{-1}$, and $|\mathcal{G}| = O(p)$. We use the bounds $\sum_{j=1}^p\|v_jg_j\|^2 \leq CJ\|g\|^2/p$, $\sum_{j\in\mathcal{G}}\|v_j\|^2 \leq J$. Then

$$a_1 + ... + a_4 \leq \left(\omega_{np}^2\psi_{np}^{-2}J^{2+2d+4b} + J^2p^{-1} + pn\omega_{np}^2\psi_{np}^{-4}J^{2d+2b} + \psi_{np}^{-2}J^{2b}p\right)J^{2+2d+4b}\|g\|^2\psi_{np}^{-2}\omega_{np}^2.$$

Hence $J^{2b}\|u_i\|^2\psi_{np}^2(a_1 + ... + a_4) = o_P(p^{-1})$ holds as long as: $J^{7/2+5b+2d}pn^{-1/2} = o(\psi_{np})$, $J^{3+3b+g} = o(\min(\sqrt{n}, \sqrt{p}))$ and $(p\vee n)^{3/4}J^{5/4+d+2b} = o(\psi_{np})$.

Proof of (C.8). We divide into two cases.

CASE I: sparse g.

We have $J^b\|u_i\|\psi_{np}\mu_{np} = o_P(\|u_i\|\|g\|)$.

CASE II: dense g.

We have $\|u_i\|\psi_{np}\mu_{np} = o_P(p^{-1/2})$.

These hold under the condition

$$\max_{i\leq n}\sum_{j=1}^p 1\{j \in \bar{\mathcal{B}} \triangle \mathcal{B}_i\} = o_P\left(\frac{\min\{n, p, \psi_{np}\}p}{(n+p)J + \|R\|_{(n)}^2}\right)J^{-(2+d+2b)}.$$

□

References

[1] Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171.

[2] Bühlmann, P. and van de Geer, S. (2011). *Statistics for high-dimensional data, methods, theory and applications*. The first edition ed. Springer, New York.

[3] Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008*, vol. 2033. Springer Science & Business Media.

[4] Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics* **39** 1069–1097.

[5] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer-Verlag New York.

[6] Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027* .

Department of Economics, MIT, Cambridge, MA 02139
Email address: vchern@mit.edu

Booth School of Business, University of Chicago, Chicago, IL 60637
Email address: Christian.Hansen@chicagobooth.edu

Department of Economics, Rutgers University, New Brunswick, NJ 08901
Email address: yuan.liao@rutgers.edu

Department of Economics, Brandeis University, 415 South St, Waltham, MA 02453
Email address: yinchuzhu@brandeis.edu