# Supplement to “Learning Latent Factors from Diversified Projections and its Applications to Over-Estimated and Weak Factors”

Jianqing Fan* Yuan Liao†

**Abstract**

This supplement contains all the technical proofs of the main paper.

## Contents



*Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA. jqfan@princeton.edu. His research is supported by NSF grants DMS-1662139 and DMS-1712591.

†Department of Economics, Rutgers University, 75 Hamilton St., New Brunswick, NJ 08901, USA. yuan.liao@rutgers.edu

# A Technical Proofs

Throughout the proofs, we use $C$ to denote a generic positive constant. Recall that $\nu_{\min}(\mathbf{H})$ and $\nu_{\max}(\mathbf{H})$ respectively denote the minimum and maximum nonzero singular values of $\mathbf{H}$. In addition, $\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ and $\mathbf{M_A} = \mathbf{I} - \mathbf{P_A}$ denote the projection matrices of a matrix $\mathbf{A}$. If $\mathbf{A}'\mathbf{A}$ is singular, $(\mathbf{A}'\mathbf{A})^{-1}$ is replaced with its Moore-Penrose generalized inverse $(\mathbf{A}'\mathbf{A})^{+}$. Let $\mathbf{U}$ be the $N \times T$ matrix of $u_{it}$. Recall that $R = \dim(\widehat{\mathbf{f}}_t)$ and $r = \dim(\mathbf{f}_t)$.

We use $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ to respectively denote the operator norm and Frobinus norm. Finally, we define $\|\mathbf{A}\|_\infty$ as follows: if $\mathbf{A}$ is an $N \times K$ matrix with $K = R$ or $r$, then $\|\mathbf{A}\|_\infty = \max_{i\leq N} \|\mathbf{A}_i\|$ where $\mathbf{A}_i$ denotes the $i$ th row of $\mathbf{A}$; if $\mathbf{A}$ is a $K \times N$ matrix with $K = R$ or $r$, then $\|\mathbf{A}\|_\infty = \max_{i\leq N} \|\mathbf{A}_i\|$ where $\mathbf{A}_i$ denotes the $i$ th column of $\mathbf{A}$; if $\mathbf{A}$ is an $N \times N$ matrix, then $\|\mathbf{A}\|_\infty = \max_{i,j\leq N} |A_{ij}|$ where $A_{ij}$ denotes the $(i, j)$ th element of $\mathbf{A}$.

Throughout the proof, all $\mathbb{E}(.)$, $\mathbb{E}(.|.)$ and $\mathrm{Var}(.)$ are calculated conditionally on $\mathbf{W}$.

## A.1 A key Proposition for asymptotic analysis when $R \geq r$

**Proposition A.1.** *Suppose $R \geq r$ and $T, N \to \infty$. Also suppose $\mathbf{G}$ is a $T \times d$ matrix so that $\mathbb{E}(\mathbf{U}|\mathbf{G}) = 0$, $\frac{1}{T}\|\mathbf{G}\|^2 = O_P(1)$, for some fixed dimension $d$, and Assumption 2.1 - 2.4 hold. In addition, for each $\mathbf{K} \in \{\mathbf{I}_T, \mathbf{M_G}\}$, suppose $\lambda_{\min}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F}) > c > 0$. Then*
*(i) $\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}) \geq cN^{-1}$ with probability approaching one for some $c > 0$,*
*(ii) $\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\| = O_P(\nu_{\min}^{-1} + \sqrt{\frac{N}{T}})$, and $\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\mathbf{H}\| = O_P(1)$.*
*(iii) $\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\mathbf{H} - \mathbf{H}'(\mathbf{H}\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F}\mathbf{H}')^{+}\mathbf{H}\| = O_P(\frac{1}{N\nu_{\min}^2} + \frac{1}{T})$, and $\frac{1}{T}\mathbf{G}'(\mathbf{P}_{\widehat{\mathbf{F}}} - \mathbf{P}_{\mathbf{FH}'})\mathbf{G} = O_P(\frac{1}{N\nu_{\min}^2} + \frac{1}{T})$.*

*Proof.* The proof applies for both $\mathbf{K} = \mathbf{I}_T$ and $\mathbf{K} = \mathbf{M_G}$. In addition, the proof depends on results in the later Lemma A.1; the latter is proved independently which does not depend on this proposition. Write $\nu_{\min} := \nu_{\min}(\mathbf{H})$, and $\nu_{\max} := \nu_{\max}(\mathbf{H})$.

First, it is easy to see

$$\widehat{\mathbf{F}} = \mathbf{F}\mathbf{H}' + \mathbf{E}.$$

where $\mathbf{E} = (\mathbf{e}_1, \cdots, \mathbf{e}_T)' = \frac{1}{N}\mathbf{U}'\mathbf{W}$, which is $T \times R$. Write

$$\boldsymbol{\Delta} := \frac{1}{T}\mathbb{E}\,\mathbf{E}'\mathbf{E} + \frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{E} + \frac{1}{T}\mathbf{E}'\mathbf{K}\mathbf{F}\mathbf{H}' + \frac{1}{T}(\mathbf{E}'\mathbf{E} - \mathbb{E}\,\mathbf{E}'\mathbf{E}) + \boldsymbol{\Delta}_1$$

where $\boldsymbol{\Delta}_1 = 0$ if $\mathbf{K} = \mathbf{I}_T$ and $\boldsymbol{\Delta}_1 = -\frac{1}{T}\mathbf{E}'\mathbf{P_G}\mathbf{E}$ if $\mathbf{K} = \mathbf{M_G}$.

(i) We have

$$\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}} = \frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{F}\mathbf{H}' + \mathbf{\Delta}.$$

By assumption $\lambda_{\min}(\frac{1}{T}\mathbb{E}\,\mathbf{U}\mathbf{U}') \geq c_0$, so $\lambda_{\min}(\frac{1}{T}\mathbb{E}\,\mathbf{E}'\mathbf{E}) \geq \lambda_{\min}(\frac{1}{T}\mathbb{E}\,\mathbf{U}\mathbf{U}')\lambda_{\min}(\frac{1}{N^2}\mathbf{W}'\mathbf{W}) \geq c_0 N^{-1}$ for some $c_0 > 0$. In addition, Lemma A.1 shows $\frac{1}{T}(\mathbf{E}'\mathbf{E} - \mathbb{E}\,\mathbf{E}'\mathbf{E}) + \mathbf{\Delta}_1 = O_P(\frac{1}{N\sqrt{T}})$. Hence $\|\frac{1}{T}(\mathbf{E}'\mathbf{E} - \mathbb{E}\,\mathbf{E}'\mathbf{E}) + \mathbf{\Delta}_1\| \leq \frac{1}{2}\lambda_{\min}(\frac{1}{T}\mathbb{E}\,\mathbf{E}'\mathbf{E})$ with large probability. We now continue the argument conditioning on this event.

Now let $\mathbf{v}$ be the unit vector so that $\mathbf{v}'\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}\mathbf{v} = \lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})$ and let

$$\eta_v^2 := \frac{1}{T}\mathbf{v}'\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{F}\mathbf{H}'\mathbf{v}.$$

Because $\mathbf{v}'\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}\mathbf{v} = \eta_v^2 + \mathbf{v}'\mathbf{\Delta}\mathbf{v}$, we have

$$\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}) \geq \eta_v^2 + 2\mathbf{v}'\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{E}\mathbf{v} + \frac{c_0}{2N}.$$

If $\mathbf{v}'\mathbf{H} = 0$ then $\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}) \geq \frac{c_0}{2N}$. If $\mathbf{v}'\mathbf{H} \neq 0$ then $\eta_v^2 \neq 0$ with large probability because $\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F}$ is positive definite. Now let

$$X := (\frac{\eta_v^2}{TN})^{-1/2} 2\mathbf{v}'\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{E}\mathbf{v}, \quad 2\mathbf{v}'\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{E}\mathbf{v} = X\sqrt{\frac{\eta_v^2}{TN}}.$$

Then

$$\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}) \geq \eta_v^2 + X\sqrt{\frac{\eta_v^2}{TN}} + \frac{c_0}{2N}.$$

Suppose for now $X = O_P(1)$, a claim to be proved later. Then consider two cases.

In case 1, $\eta_v^2 \leq 4|X|\sqrt{\frac{\eta_v^2}{TN}}$. Then $|\eta_v| \leq 4|X|\frac{1}{\sqrt{TN}}$ and

$$\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}) \geq \frac{c_0}{2N} - |X||\eta_v|\frac{1}{\sqrt{TN}} \geq \frac{c_0}{2N} - 4|X|^2\frac{1}{TN} \geq \frac{c_0}{4N}$$

where the last inequality holds for $X = O_P(1)$ and as $T \to \infty$, with probability approaching one.

In case 2, $\eta_v^2 > 4|X|\sqrt{\frac{\eta_v^2}{TN}}$, then

$$\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}) \geq \eta_v^2 - |X|\sqrt{\frac{\eta_v^2}{TN}} + \frac{c_0}{2N} \geq \frac{3}{4}\eta_v^2 + \frac{c_0}{2N} \geq \frac{c_0}{2N}.$$

In both cases, $\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}) > c_0/N$ for some $c_0 > 0$ with overwhelming probability.

It remains to argue $X = O_P(1)$. By the assumption $\lambda_{\min}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F}) > c > 0$, we have

$$\eta_v^2 \geq \lambda_{\min}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})\mathbf{v}'\mathbf{H}\mathbf{H}'\mathbf{v} > c\|\mathbf{v}'\mathbf{H}\|^2.$$

In addition, Lemma A.1 shows $\|\frac{1}{T}\mathbf{F}'\mathbf{E}\|^2 = O_P(\frac{1}{TN})$ and $\|\frac{1}{T}\mathbf{G}'\mathbf{E}\|^2 = O_P(\frac{1}{TN})$. With the condition $\frac{1}{T}\|\mathbf{G}\|^2 = O_P(1)$, we reach $\|\frac{1}{T}\mathbf{F}'\mathbf{M}_{\mathbf{G}}\mathbf{E}\|^2 \leq O_P(\frac{1}{TN}) + \|\mathbf{F}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\|^2\|\frac{1}{T}\mathbf{G}'\mathbf{E}\|^2 = O_P(\frac{1}{TN})$. Therefore $\|\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{E}\|^2 = O_P(\frac{1}{TN})$ and consequently,

$$|X|^2 \leq 4TN\eta_v^{-2}\|\mathbf{v}'\mathbf{H}\|^2\|\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{E}\|^2 \leq O_P(1)\eta_v^{-2}\|\mathbf{v}'\mathbf{H}\|^2 \leq O_P(1)c^{-1}\|\mathbf{v}'\mathbf{H}\|^{-2}\|\mathbf{v}'\mathbf{H}\|^2 = O_P(1).$$

(ii) Write $\bar{\mathbf{H}} := \mathbf{H}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{1/2}$ and $\mathbf{S} := \frac{N}{T}\mathbb{E}\,\mathbf{E}'\mathbf{E} = \frac{1}{N}\mathbf{W}'\mathbf{\Sigma}_u\mathbf{W}$. Then

$$\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}} = \bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{1}{N}\mathbf{S} + \frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{E} + \frac{1}{T}\mathbf{E}'\mathbf{K}\mathbf{F}\mathbf{H}' + \mathbf{\Delta}_2 \tag{A.1}$$

where we proved in (i) that $\|\mathbf{\Delta}_2\| = \|\frac{1}{T}(\mathbf{E}'\mathbf{E} - \mathbb{E}\,\mathbf{E}'\mathbf{E}) + \mathbf{\Delta}_1\| = O_P(\frac{1}{N\sqrt{T}})$. Also all eigenvalues of $\mathbf{S}$ are bounded away from both zero and infinity. In addition, $\bar{\mathbf{H}}$ is a $R \times r$ matrix with $R \geq r$, whose Moore-Penrose generalized inverse is $\bar{\mathbf{H}}^+ = (\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}\mathbf{H}^+$. Also, $\bar{\mathbf{H}}$ is of rank $r$. Let

$$\bar{\mathbf{H}}' = \mathbf{U}_{\bar{H}}(\mathbf{D}_{\bar{H}}, 0)\mathbf{E}'_{\bar{H}}$$

be the singular value decomposition (SVD) of $\bar{\mathbf{H}}'$, where 0 is present when $R > r$. Since $\lambda_{\min}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F}) > c > 0$, we have $\lambda_{\min}(\mathbf{D}_{\bar{H}}) \geq c\nu_{\min}$ where $\nu_{\min} := \nu_{\min}(\mathbf{H})$.

The proof is divided into several steps.

Step 1. Show $\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{a}{N}\mathbf{I})^{-j}\bar{\mathbf{H}}\| = O_P(\nu_{\min}^{-(2j-2)})$ for any fixed $a > 0$ and $j = 1, 2$.

Because $\lambda_{\min}(\mathbf{D}_{\bar{H}}) \geq c\nu_{\min}$, for $j = 1, 2$,

$$\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{a}{N}\mathbf{I})^{-j}\bar{\mathbf{H}}\| = \|\mathbf{U}_{\bar{H}}(\mathbf{D}^2_{\bar{H}}(\mathbf{D}^2_{\bar{H}} + \frac{a}{N}\mathbf{I})^{-j}, 0)\mathbf{U}'_{\bar{H}}\| = \|\mathbf{D}^2_{\bar{H}}(\mathbf{D}^2_{\bar{H}} + \frac{a}{N}\mathbf{I})^{-j}\| \leq \|\mathbf{D}_{\bar{H}}^{-2j+2}\|.$$

Step 2. Show $\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{1}{N}\mathbf{S})^{-1}\bar{\mathbf{H}}\| = O_P(1)$.

Let $0 < a < \lambda_{\min}(\mathbf{S})$ be a constant. Then $(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{a}{N}\mathbf{I})^{-1} - (\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{1}{N}\mathbf{S})^{-1}$ is positive definite. (This is because, if both $\mathbf{A}_1$ and $\mathbf{A}_2 - \mathbf{A}_1$ are positive definite, then so is $\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1}$.) Let $\mathbf{v}$ be a unit vector so that $\mathbf{v}'\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{1}{N}\mathbf{S})^{-1}\bar{\mathbf{H}}\mathbf{v} = \|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{1}{N}\mathbf{S})^{-1}\bar{\mathbf{H}}\|$. Then

$$\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{1}{N}\mathbf{S})^{-1}\bar{\mathbf{H}}\| \leq \mathbf{v}'\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{a}{N}\mathbf{I})^{-1}\bar{\mathbf{H}}\mathbf{v} \leq \|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}' + \frac{a}{N}\mathbf{I})^{-1}\bar{\mathbf{H}}\|.$$

The right hand side is $O_P(1)$ due to step 1.

Step 3. Show $\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\|=O_P(\nu_{\min}^{-1})$.

Fix any $a>0$. Let $\mathbf{M}=\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{a}{N}\mathbf{I})^{-1}$. By step 1, $\|\mathbf{M}\|=\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{a}{N}\mathbf{I})^{-2}\bar{\mathbf{H}}\|^{1/2}=O_P(\nu_{\min}^{-1})$. So

$$\begin{aligned}
\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\| &\leq \|\mathbf{M}\|+\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}-\mathbf{M}\|\\
&=^{(1)} \|\mathbf{M}\|+\|\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{a}{N}\mathbf{I})^{-1}(\frac{1}{N}\mathbf{S}-\frac{a}{N}\mathbf{I})(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\|\\
&\leq \|\mathbf{M}\|+\frac{C}{N}\|\mathbf{M}\|\|(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\|\\
&\leq^{(2)} \|\mathbf{M}\|(1+O_P(1))=O_P(\nu_{\min}^{-1}).
\end{aligned}$$

(1) used $\mathbf{A}_1^{-1}-\mathbf{A}_2^{-1}=\mathbf{A}_1^{-1}(\mathbf{A}_2-\mathbf{A}_1)\mathbf{A}_2^{-1}$; (2) is from: $\|(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\|\leq\lambda_{\min}^{-1}(\frac{1}{N}\mathbf{S})=O_P(N)$.

Step 4. Show $\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\|=O_P(\nu_{\min}^{-1}+\sqrt{\frac{N}{T}})$.

Let $\mathbf{A}:=\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S}$. By steps 2,3 $\|\bar{\mathbf{H}}\mathbf{A}^{-1}\|=O_P(\nu_{\min}^{-1})$ and $\|\bar{\mathbf{H}}\mathbf{A}^{-1}\bar{\mathbf{H}}\|=O_P(1)$. Now

$$\begin{aligned}
&\|\bar{\mathbf{H}}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}-\bar{\mathbf{H}}'\mathbf{A}^{-1}\|=\|\bar{\mathbf{H}}'\mathbf{A}^{-1}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}-\mathbf{A})(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\|\\
\leq^{(3)}\ & O_P(\frac{\nu_{\max}(\mathbf{H})}{\nu_{\min}(\mathbf{H})\sqrt{TN}})\|(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\|=^{(4)}O_P(\frac{N}{\sqrt{NT}})=O_P(\sqrt{\frac{N}{T}}).
\end{aligned}$$

In (3) we used $\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}-\mathbf{A}=O_P(\frac{1}{N\sqrt{T}}+\|\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{E}\|)=O_P(\frac{1}{N\sqrt{T}}+\frac{\nu_{\max}}{\sqrt{TN}})=O_P(\frac{\nu_{\max}}{\sqrt{TN}})$; in (4) we used $(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}=O_P(N)$ by part (i) and $\nu_{\max}\leq C\nu_{\min}$. Hence

$$\|\bar{\mathbf{H}}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\|\leq O_P(\sqrt{\frac{N}{T}})+\|\bar{\mathbf{H}}\mathbf{A}^{-1}\|=O_P(\nu_{\min}^{-1}+\sqrt{\frac{N}{T}}).$$

Thus $\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\|\leq\|(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}\|\|\bar{\mathbf{H}}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\|$, which leads to the result for $\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\|=O_P(\nu_{\min}^{-1}+\sqrt{\frac{N}{T}})$.

Step 5. show $\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\mathbf{H}=\mathbf{H}'(\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{F}\mathbf{H}'+\frac{1}{N}\mathbf{S})^{-1}\mathbf{H}+O_P(\frac{1}{\nu_{\min}\sqrt{NT}}+\frac{1}{T})$.

Because $\|\bar{\mathbf{H}}\mathbf{A}^{-1}\|=O_P(\nu_{\min}^{-1})$ and $\|\bar{\mathbf{H}}\mathbf{A}^{-1}\bar{\mathbf{H}}\|=O_P(1)$ by step 3, (A.1) implies

$$\begin{aligned}
&\|\bar{\mathbf{H}}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\bar{\mathbf{H}}-\bar{\mathbf{H}}'\mathbf{A}^{-1}\bar{\mathbf{H}}\|=\|\bar{\mathbf{H}}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}}-\mathbf{A})\mathbf{A}^{-1}\bar{\mathbf{H}}\|\\
\leq\ & \|\bar{\mathbf{H}}'\mathbf{A}^{-1}\bar{\mathbf{H}}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{E}(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\bar{\mathbf{H}}\|+\|\bar{\mathbf{H}}'\mathbf{A}^{-1}\frac{1}{T}\mathbf{E}'\mathbf{K}\mathbf{F}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}\bar{\mathbf{H}}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\bar{\mathbf{H}}\|\\
&+\|\bar{\mathbf{H}}'\mathbf{A}^{-1}\mathbf{\Delta}_1(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\bar{\mathbf{H}}\|
\end{aligned}$$

$$\leq \quad O_P(\nu_{\min}^{-1}\frac{1}{N\sqrt{T}}+\frac{1}{\sqrt{NT}})\|(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\bar{\mathbf{H}}\| =^{(5)} O_P(\frac{1}{\sqrt{NT}})O_P(\nu_{\min}^{-1}+\sqrt{\frac{N}{T}}) = O_P(\frac{1}{\nu_{\min}\sqrt{NT}}+\frac{1}{T}).$$

(5) follows from step 4 and $\nu_{\min} \gg N^{-1/2}$. Then due to $\|(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}\| = O_P(1)$,

$$\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\mathbf{H} = \mathbf{H}'(\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{F}\mathbf{H}'+\frac{1}{N}\mathbf{S})^{-1}\mathbf{H} + O_P(\frac{1}{\nu_{\min}\sqrt{NT}}+\frac{1}{T}).$$

In addition, step 3 implies $\|\mathbf{H}'(\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{F}\mathbf{H}'+\frac{1}{N}\mathbf{S})^{-1}\mathbf{H}\| \leq O_P(\nu_{\min}^{-1}\nu_{\max}) = O_P(1)$, so

$$\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\mathbf{H}\| = O_P(1+\frac{1}{\nu_{\min}\sqrt{NT}}+\frac{1}{T}) = O_P(1).$$

(iii) The proof still consists of several steps.

Step 1. $\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\mathbf{H} = \mathbf{H}'(\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{K}\mathbf{F}\mathbf{H}'+\frac{1}{N}\mathbf{S})^{-1}\mathbf{H} + O_P(\frac{1}{\nu_{\min}\sqrt{NT}}+\frac{1}{T})$.

It follows from step 5 of part (ii).

Step 2. show $\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\bar{\mathbf{H}} = \bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}')^{+}\bar{\mathbf{H}} + O_P(\frac{1}{N\nu_{\min}^2})$ where $\bar{\mathbf{H}} = \mathbf{H}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{1/2}$. Write $\mathbf{T} = \bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\bar{\mathbf{H}} - \bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}')^{+}\bar{\mathbf{H}}$. The goal is to show $\|\mathbf{T}\| = O_P(\frac{1}{N\nu_{\min}^2})$. Let $\mathbf{v}$ be the unit vector so that $|\mathbf{v}'\mathbf{T}\mathbf{v}| = \|\mathbf{T}\|$. Define a function, for $d > 0$,

$$g(d) := \mathbf{v}'\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{d}{N}\mathbf{I})^{-1}\bar{\mathbf{H}}\mathbf{v}.$$

Note that there are constants $c, C > 0$ so that $\frac{c}{N} < \lambda_{\min}(\frac{1}{N}\mathbf{S}) \leq \lambda_{\max}(\frac{1}{N}\mathbf{S}) < \frac{C}{N}$. Then we have $g(C) < \mathbf{v}'\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\bar{\mathbf{H}}\mathbf{v} < g(c)$. Hence

$$|\mathbf{v}'\mathbf{T}\mathbf{v}| \leq |g(c) - \mathbf{v}'\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}')^{+}\bar{\mathbf{H}}\mathbf{v}| + |g(C) - \mathbf{v}'\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}')^{+}\bar{\mathbf{H}}\mathbf{v}|.$$

Recall $\bar{\mathbf{H}}' = \mathbf{U}_{\bar{H}}(\mathbf{D}_{\bar{H}}, 0)\mathbf{E}_{\bar{H}}'$ is the SVD of $\bar{\mathbf{H}}'$ and $N^{-1}\lambda_{\min}^{-1}(\mathbf{D}_{\bar{H}}^2) = o_P(1)$. Then for any $d \in \{c, C\}$, as $N \to \infty$, $g(d) = \mathbf{v}'\mathbf{U}_{\bar{H}}\mathbf{D}_{\bar{H}}^2(\mathbf{D}_{\bar{H}}^2+\frac{d}{N}\mathbf{I})^{-1}\mathbf{U}_{\bar{H}}'\mathbf{v} \xrightarrow{P} \mathbf{v}'\mathbf{v} = \mathbf{v}'\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}')^{+}\bar{\mathbf{H}}\mathbf{v}$, where we used $\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}')^{+}\bar{\mathbf{H}} = \mathbf{I}$, easy to see from its SVD. The rate of convergence is

$$\|\mathbf{D}_{\bar{H}}^2(\mathbf{D}_{\bar{H}}^2+\frac{d}{N}\mathbf{I})^{-1} - \mathbf{I}\| \leq \|\mathbf{D}_{\bar{H}}^2(\mathbf{D}_{\bar{H}}^2+\frac{d}{N}\mathbf{I})^{-1}\frac{d}{N}\mathbf{D}_{\bar{H}}^{-2}\| = O_P(\frac{1}{N\nu_{\min}^2}).$$

Hence $|\mathbf{v}'\mathbf{T}\mathbf{v}| = O_P(\frac{1}{N\nu_{\min}^2})$.

Step 3. show $\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\mathbf{H} - \mathbf{H}'(\mathbf{H}\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F}\mathbf{H}')^{+}\mathbf{H}\| = O_P(\frac{1}{N\nu_{\min}^2}+\frac{1}{T})$. By steps 1 and

2,

$$\begin{aligned}
\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\mathbf{H} &= \mathbf{H}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\mathbf{H}+O_P(\frac{1}{\nu_{\min}\sqrt{NT}}+\frac{1}{T}) \\
&= (\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}'+\frac{1}{N}\mathbf{S})^{-1}\bar{\mathbf{H}}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}+O_P(\frac{1}{\nu_{\min}\sqrt{NT}}+\frac{1}{T}) \\
&=^{(6)} (\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}\bar{\mathbf{H}}'(\bar{\mathbf{H}}\bar{\mathbf{H}}')^{+}\bar{\mathbf{H}}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})^{-1/2}+O_P(\frac{1}{N\nu_{\min}^2}+\frac{1}{\nu_{\min}\sqrt{NT}}+\frac{1}{T}) \\
&= \mathbf{H}'(\bar{\mathbf{H}}\bar{\mathbf{H}}')^{+}\mathbf{H}+O_P(\frac{1}{N\nu_{\min}^2}+\frac{1}{T}).
\end{aligned}$$

where (6) is due to $\lambda_{\min}(\frac{1}{T}\mathbf{F}'\mathbf{K}\mathbf{F})>c$ and step 2.

Step 4. show $\frac{1}{T}\mathbf{G}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{G}=\frac{1}{T}\mathbf{G}'\mathbf{P}_{\mathbf{F}\mathbf{H}'}\mathbf{G}+O_P(\frac{1}{N\nu_{\min}^2}+\frac{1}{T})$.

By part (ii) $\|\mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{K}\widehat{\mathbf{F}})^{-1}\|=O_P(\nu_{\min}^{-1}+\sqrt{\frac{N}{T}})$, and that $\frac{1}{T}\mathbf{G}'\mathbf{E}=O_P(\frac{1}{\sqrt{NT}})$,

$$\begin{aligned}
\frac{1}{T}\mathbf{G}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{G} &= \frac{1}{T}\mathbf{G}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{G}+\frac{1}{T}\mathbf{G}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{G}+\frac{1}{T}\mathbf{G}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{G} \\
&\quad +\frac{1}{T}\mathbf{G}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{G} \\
&= \frac{1}{T}\mathbf{G}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{G}+O_P(\frac{1}{T}+\frac{1}{\nu_{\min}\sqrt{NT}}) \\
&= \frac{1}{T}\mathbf{G}'\mathbf{F}\mathbf{H}'(\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}')^{+}\mathbf{H}\mathbf{F}'\mathbf{G}+O_P(\frac{1}{N\nu_{\min}^2}+\frac{1}{T}),
\end{aligned}$$

where the last equality follows from step 3. □

The proof of Lemma A.1 below does not rely on Proposition A.1, as it does not involve $\mathbf{H}$ or $\widehat{\mathbf{F}}$. Also, let $\mathbf{E}=(\mathbf{e}_1,\cdots,\mathbf{e}_T)'=\frac{1}{N}\mathbf{U}'\mathbf{W}$. In addition, we shall use the following inequality $\mathrm{tr}(\mathbf{W}'\mathbf{\Sigma}\mathbf{W})\le R\|\mathbf{W}\|^2\|\mathbf{\Sigma}\|$ for any semipositive definite matrix $\mathbf{\Sigma}$, whose simple proof is as follows: let $\mathbf{v}_i$ be the $i$ th eigenvector of $\mathbf{W}'\mathbf{\Sigma}\mathbf{W}$. Then

$$\mathrm{tr}(\mathbf{W}'\mathbf{\Sigma}\mathbf{W})=\sum_{i=1}^{R}\mathbf{v}_i'\mathbf{W}'\mathbf{\Sigma}\mathbf{W}\mathbf{v}_i\le\|\mathbf{\Sigma}\|\sum_{i=1}^{R}\|\mathbf{W}\mathbf{v}_i\|^2\le\|\mathbf{\Sigma}\|\|\mathbf{W}\|^2R.$$

**Lemma A.1.** *For any $R\ge 1$, ($R$ can be either smaller, equal to or larger than $r$),*
*(i)* $\|\frac{1}{T}\mathbb{E}\,\mathbf{E}'\mathbf{E}\|\le\frac{C}{N}$ *and* $\|\mathbf{E}\|=O_P(\sqrt{\frac{T}{N}})$.
*(ii)* $\mathbb{E}\,\|\frac{1}{T}\mathbf{F}'\mathbf{E}\|^2\le O(\frac{1}{TN})$, $\mathbb{E}\,\|\frac{1}{T}\mathbf{G}'\mathbf{E}\|^2\le O(\frac{1}{TN})$, *here* $\mathbf{G}$ *is defined as in Section 3.1*
*(iii)* $\|\frac{1}{T}(\mathbf{E}'\mathbf{E}-\mathbb{E}\,\mathbf{E}'\mathbf{E})\|\le O_P(\frac{1}{N\sqrt{T}})$, $\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\mathbf{G}}\mathbf{E}\|=O_P(\frac{1}{NT})$.
*(iv)* $\|\frac{1}{N}\mathbf{U}'\mathbf{W}\|\le O_P(\sqrt{\frac{T}{N}})$.

*Proof.* (i) By the assumption $\|\frac{1}{T}\mathbb{E}\,\mathbf{U}\mathbf{U}'\| = \|\mathbb{E}\,\mathbf{u}_t\mathbf{u}_t'\| \leq \mathbb{E}\,\|\mathbb{E}(\mathbf{u}_t\mathbf{u}_t'|\mathbf{F})\| < C$. Thus

$$\|\frac{1}{T}\mathbb{E}\,\mathbf{E}'\mathbf{E}\| = \frac{1}{N^2}\|\mathbf{W}'\frac{1}{T}\mathbb{E}\,\mathbf{U}\mathbf{U}'\mathbf{W}\| \leq \frac{1}{N^2}\|\mathbf{W}\|^2 \leq \frac{C}{N}.$$

Also, $\mathbb{E}\,\|\mathbf{E}\|^2 \leq \operatorname{tr}\mathbb{E}\,\mathbf{E}'\mathbf{E} \leq R\|\mathbb{E}\,\mathbf{E}'\mathbf{E}\| \leq \frac{CT}{N}$.

(ii) Let $f_{k,t}$ be the $k$ th entry of $\mathbf{f}_t$. By the assumption $\frac{1}{T}\sum_{s=1}^T\sum_{t=1}^T\mathbb{E}\,\|\mathbf{f}_t\|\|\mathbf{f}_s\|\|\mathbb{E}(\mathbf{u}_t\mathbf{u}_s'|\mathbf{F})\| < C$,

$$\begin{aligned}
\mathbb{E}\,\|\frac{1}{T}\mathbf{F}'\mathbf{E}\|^2 &= \frac{1}{T^2N^2}\mathbb{E}\,\|\sum_{t=1}^T\mathbf{W}'\mathbf{u}_t\mathbf{f}_t'\|^2 \leq \sum_{k=1}^r\frac{1}{T^2N^2}\sum_{s=1}^T\sum_{t=1}^T\mathbb{E}\,f_{k,t}f_{k,s}\,\mathbb{E}(\mathbf{u}_s'\mathbf{W}\mathbf{W}'\mathbf{u}_t|\mathbf{F}) \\
&\leq \sum_{k=1}^r\frac{1}{T^2N^2}\sum_{s=1}^T\sum_{t=1}^T\mathbb{E}\,f_{k,t}f_{k,s}\operatorname{tr}\mathbf{W}'\,\mathbb{E}(\mathbf{u}_t\mathbf{u}_s'|\mathbf{F})\mathbf{W} \\
&\leq \sum_{k=1}^r\frac{1}{T^2N^2}\sum_{s=1}^T\sum_{t=1}^T\mathbb{E}\,|f_{k,t}f_{k,s}|\|\mathbf{W}\|_F^2\|\mathbb{E}(\mathbf{u}_t\mathbf{u}_s'|\mathbf{F})\| \\
&\leq \frac{C}{T^2N}\sum_{s=1}^T\sum_{t=1}^T\mathbb{E}\,\|\mathbf{f}_t\|\|\mathbf{f}_s\|\|\mathbb{E}(\mathbf{u}_t\mathbf{u}_s'|\mathbf{F})\| \\
&\leq \frac{C}{TN}.
\end{aligned}$$

Similarly, $\mathbb{E}\,\|\frac{1}{T}\mathbf{G}'\mathbf{E}\|^2 \leq O(\frac{1}{TN})$.

(iii) By the assumption that $\frac{1}{TN^2}\sum_{t,s\leq T}\sum_{i,j,m,n\leq N}|\operatorname{Cov}(u_{it}u_{jt}, u_{ms}u_{ns})| < C$,

$$\begin{aligned}
&\mathbb{E}\,\|\frac{1}{T}(\mathbf{E}'\mathbf{E} - \mathbb{E}\,\mathbf{E}'\mathbf{E})\|^2 \leq \sum_{k,q\leq R}\mathbb{E}(\frac{1}{TN^2}\sum_{t=1}^T\sum_{i,j\leq N}w_{k,i}w_{q,j}(u_{it}u_{jt} - \mathbb{E}\,u_{it}u_{jt}))^2 \\
\leq\;& \frac{C}{TN^2}\frac{1}{TN^2}\sum_{t,s\leq T}\sum_{i,j,m,n\leq N}|\operatorname{Cov}(u_{it}u_{jt}, u_{ms}u_{ns})| \leq \frac{C}{TN^2}.
\end{aligned}$$

Next, by part (ii)

$$\|\frac{1}{T}\mathbf{E}'\mathbf{P}_\mathbf{G}\mathbf{E}\| \;\leq\; \|\frac{1}{T}\mathbf{E}'\mathbf{G}\|^2\|(\frac{1}{T}\mathbf{G}'\mathbf{G})^{-1}\| \leq O_P(\frac{1}{TN}).$$

(iv) $\mathbb{E}\,\|\frac{1}{N}\mathbf{U}'\mathbf{W}\|^2 \leq \frac{1}{N^2}\operatorname{tr}\mathbb{E}\,\mathbf{W}'\mathbf{U}\mathbf{U}'\mathbf{W} \leq \frac{CT}{N^2}\|\mathbf{W}\|_F^2 \leq \frac{CT}{N}$, where we used the assumption that $\|\mathbb{E}\,\mathbf{u}_t\mathbf{u}_t'\| < C$.

□

## A.2 Proof of Theorem 2.1

*Proof.* We shall first show the convergence of $\mathbf{P}_{\widehat{\mathbf{F}}\mathbf{M}} - \mathbf{P}_{\mathbf{F}}$, and then the convergence of $\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{P}_{\mathbf{F}} - \mathbf{P}_{\mathbf{F}}$.

First, from the SVD $\mathbf{H}' = \mathbf{U}_H(\mathbf{D}_H, 0)\mathbf{E}_H'$, it is straightforward to verify that $\mathbf{M}' = \mathbf{U}_H(\mathbf{D}_H^{-1}, 0)\mathbf{E}_H'$. Then from Proposition A.1, $\lambda_{\min}(\frac{1}{T}\mathbf{M}'\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\mathbf{M}) \geq c_0 N^{-1}\lambda_{\min}(\mathbf{D}_H^{-2})$ with large probability. Hence $\mathbf{P}_{\widehat{\mathbf{F}}\mathbf{M}}$ is well defined.

Next, it is easy to see $\mathbf{H}'(\mathbf{H}\mathbf{H}')^{+}\mathbf{H} = \mathbf{I}$ when $R \geq r$. Then $\widehat{\mathbf{F}} = \mathbf{F}\mathbf{H}' + \mathbf{E}$ implies $\widehat{\mathbf{F}}\mathbf{M} - \mathbf{F} = \mathbf{E}(\mathbf{H}\mathbf{H}')^{+}\mathbf{H}$ with $\mathbf{M} = (\mathbf{H}\mathbf{H}')^{+}\mathbf{H}$. Since $\|(\mathbf{H}\mathbf{H}')^{+}\mathbf{H}\| = O_P(\nu_{\min}^{-1})$, we have

$$\frac{1}{\sqrt{T}}\|\widehat{\mathbf{F}}\mathbf{M} - \mathbf{F}\| = O_P(\frac{1}{\sqrt{N}}\nu_{\min}^{-1}), \quad \frac{1}{T}\|\mathbf{F}'(\widehat{\mathbf{F}}\mathbf{M} - \mathbf{F})\| = O_P(\frac{1}{\sqrt{NT}}\nu_{\min}^{-1})$$

where the second statement uses Lemma A.1. Then $\|\frac{1}{T}\mathbf{M}'\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\mathbf{M} - \frac{1}{T}\mathbf{F}'\mathbf{F}\| = O_P(\frac{1}{\sqrt{NT}}\nu_{\min}^{-1} + \frac{1}{N}\nu_{\min}^{-2})$. Thus $(\frac{1}{T}\mathbf{M}'\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\mathbf{M})^{-1} = O_P(1)$ and

$$\|(\frac{1}{T}\mathbf{M}'\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\mathbf{M})^{-1} - (\frac{1}{T}\mathbf{F}'\mathbf{F})^{-1}\| = O_P(\frac{1}{\sqrt{NT}}\nu_{\min}^{-1} + \frac{1}{N}\nu_{\min}^{-2}). \tag{A.2}$$

The triangular inequality then implies $\|\mathbf{P}_{\widehat{\mathbf{F}}\mathbf{M}} - \mathbf{P}_{\mathbf{F}}\| \leq O_P(\frac{1}{\sqrt{N}}\nu_{\min}^{-1})$.

Finally, $\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{P}_{\widehat{\mathbf{F}}\mathbf{M}} = \mathbf{P}_{\widehat{\mathbf{F}}\mathbf{M}}$ gives

$$\|\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{P}_{\mathbf{F}} - \mathbf{P}_{\mathbf{F}}\| \leq \|\mathbf{P}_{\widehat{\mathbf{F}}}(\mathbf{P}_{\mathbf{F}} - \mathbf{P}_{\widehat{\mathbf{F}}\mathbf{M}})\| + \|\mathbf{P}_{\widehat{\mathbf{F}}\mathbf{M}} - \mathbf{P}_{\mathbf{F}}\| \leq O_P(\frac{1}{\sqrt{N}}\nu_{\min}^{-1}).$$

□

## A.3 Proof of Theorem 3.1

*Proof.* Here we assume $R \geq r$. We let $\mathbf{z}_t = (\mathbf{f}_t'\mathbf{H}', \mathbf{g}_t')'$ and $\boldsymbol{\delta} = (\boldsymbol{\alpha}'\mathbf{H}^{+}, \boldsymbol{\beta}')'$. Then $\boldsymbol{\delta}'\mathbf{z}_t = y_{t+h|t}$. First, we have the following expansion

$$\widehat{\boldsymbol{\delta}}'\widehat{\mathbf{z}}_T - \boldsymbol{\delta}'\mathbf{z}_T = (\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\widehat{\mathbf{z}}_T + \boldsymbol{\alpha}'\mathbf{H}^{+}(\widehat{\mathbf{f}}_T - \mathbf{H}\mathbf{f}_T).$$

Now $\widehat{\boldsymbol{\delta}} = (\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})^{-1}\widehat{\mathbf{Z}}'\mathbf{Y}$, where $\mathbf{Y}$ is the $(T-h)\times 1$ vector of $y_{t+h}$, and $\widehat{\mathbf{Z}}$ is the $(T-h)\times\dim(\boldsymbol{\delta})$ matrix of $\widehat{\mathbf{z}}_t$, $t = 1, \cdots, T-h$. Also recall that $\mathbf{e}_t = \widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t = \frac{1}{N}\mathbf{W}'\mathbf{u}_t$. Then

$$\begin{aligned}
\widehat{\mathbf{z}}_T'(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) &= \widehat{\mathbf{z}}_T'(\frac{1}{T}\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})^{-1}\sum_{d=1}^{4} a_d, \text{ where} \\
a_1 &= (\frac{1}{T}\sum_t \varepsilon_t \mathbf{e}_t', 0)', \quad a_2 = \frac{1}{T}\sum_t \mathbf{z}_t\varepsilon_t \\
a_3 &= (-\boldsymbol{\alpha}'\mathbf{H}^{+}\frac{1}{T}\sum_t \mathbf{e}_t\mathbf{e}_t', 0)', \quad a_4 = -\frac{1}{T}\sum_t \mathbf{z}_t\mathbf{e}_t'\mathbf{H}^{+'}\boldsymbol{\alpha}.
\end{aligned}$$

On the other hand, let $\mathbf{G}$ be the $(T-h)\times\dim(\mathbf{g}_t)$ matrix of $\{\mathbf{g}_t : g \leq T-h\}$. We have, by the matrix block inverse formula, for the operator $\mathbf{M}_\mathbf{A} := \mathbf{I} - \mathbf{P}_\mathbf{A}$,

$$(\frac{1}{T}\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})^{-1} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_2' & \mathbf{A}_3 \end{pmatrix}, \quad \text{where } \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{pmatrix} = \begin{pmatrix} (\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{M}_\mathbf{G}\widehat{\mathbf{F}})^{-1} \\ -\mathbf{A}_1\widehat{\mathbf{F}}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1} \\ (\frac{1}{T}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{G})^{-1} \end{pmatrix}.$$

Then $\widehat{\mathbf{z}}_T'(\frac{1}{T}\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})^{-1} = (\widehat{\mathbf{f}}_T'\mathbf{A}_1 + \mathbf{g}_T'\mathbf{A}_2', \widehat{\mathbf{f}}_T'\mathbf{A}_2 + \mathbf{g}_T'\mathbf{A}_3)$. This implies

$$\begin{aligned}
\widehat{\mathbf{z}}_T'(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) &= (\widehat{\mathbf{f}}_T'\mathbf{A}_1 + \mathbf{g}_T'\mathbf{A}_2')\frac{1}{T}\sum_t [\mathbf{e}_t\varepsilon_t - \mathbf{e}_t\mathbf{e}_t'\mathbf{H}^{+'}\boldsymbol{\alpha}] \\
&\quad + (\widehat{\mathbf{f}}_T'\mathbf{A}_1\mathbf{H} + \mathbf{g}_T'\mathbf{A}_2'\mathbf{H})\frac{1}{T}\sum_t [\mathbf{f}_t\varepsilon_t - \mathbf{f}_t\mathbf{e}_t'\mathbf{H}^{+'}\boldsymbol{\alpha}] \\
&\quad + (\widehat{\mathbf{f}}_T'\mathbf{A}_2 + \mathbf{g}_T'\mathbf{A}_3)\frac{1}{T}\sum_t [\mathbf{g}_t\varepsilon_t - \mathbf{g}_t\mathbf{e}_t'\mathbf{H}^{+'}\boldsymbol{\alpha}].
\end{aligned}$$

It is easy to show $\|\frac{1}{T}\sum_t \mathbf{f}_t\varepsilon_t\| + \|\frac{1}{T}\sum_t \mathbf{g}_t\varepsilon_t\| = O_P(\frac{1}{\sqrt{T}})$ and $\|\frac{1}{T}\sum_t \mathbf{e}_t\varepsilon_t\| = O_P(\frac{1}{\sqrt{TN}})$. Also Lemma A.1 gives $\frac{1}{T}\sum_t \mathbf{e}_t\mathbf{e}_t' = \frac{1}{T}\mathbf{E}'\mathbf{E} = O_P(\frac{1}{N})$, $\frac{1}{T}\sum_t \mathbf{f}_t\mathbf{e}_t = \frac{1}{T}\mathbf{F}'\mathbf{E} = O_P(\frac{1}{\sqrt{TN}})$, and $\frac{1}{T}\sum_t \mathbf{g}_t\mathbf{e}_t = \frac{1}{T}\mathbf{F}'\mathbf{E} = O_P(\frac{1}{\sqrt{TN}})$. Together with Lemma A.2,

$$\begin{aligned}
\widehat{\mathbf{z}}_T'(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) &= \|\widehat{\mathbf{f}}_T'\mathbf{A}_1 + \mathbf{g}_T'\mathbf{A}_2\| O_P(\frac{1}{\sqrt{TN}} + \frac{1}{N\nu_{\min}}) \\
&\quad + \|\widehat{\mathbf{f}}_T'\mathbf{A}_1\mathbf{H} + \mathbf{g}_T'\mathbf{A}_2'\mathbf{H}\| O_P(\frac{1}{\sqrt{T}}) + \|\widehat{\mathbf{f}}_T'\mathbf{A}_2 + \mathbf{g}_T'\mathbf{A}_3\| O_P(\frac{1}{\sqrt{T}}) \\
&= O_P(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}\nu_{\min}}).
\end{aligned}$$

Finally, as $\|\mathbf{H}^{+}\| = O_P(\nu_{\min}^{-1})$, $\boldsymbol{\alpha}'\mathbf{H}^{+}(\widehat{\mathbf{f}}_T - \mathbf{H}\mathbf{f}_T) = O_P(\nu_{\min}^{-1})\|\mathbf{e}_T\| = O_P(\nu_{\min}^{-1}N^{-1/2})$.

□

**Lemma A.2.** *For all $R \geq r$, (i) $\|\mathbf{A}_1\widehat{\mathbf{f}}_T\| + \|\mathbf{A}_2\| = O_P(\sqrt{N})$, and*
$\|\mathbf{H}'\mathbf{A}_1\widehat{\mathbf{f}}_T\| + \|\mathbf{H}'\mathbf{A}_2\| + \|\mathbf{A}_2'\widehat{\mathbf{f}}_T\| + \|\mathbf{A}_3\| = O_P(1)$.

*Proof.* First, by Proposition A.1, $\|\mathbf{A}_1\| = O_P(N)$ and $\|\mathbf{A}_1\mathbf{H}\| = O_P(\nu_{\min}^{-1} + \sqrt{\frac{N}{T}})$, and $\frac{1}{T}\mathbf{E}'\mathbf{G} = O_P(\frac{1}{\sqrt{NT}})$

$$
\begin{aligned}
\mathbf{A}_1\widehat{\mathbf{f}}_T &= (\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{M}_{\mathbf{G}}\widehat{\mathbf{F}})^{-1}\mathbf{e}_T + (\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{M}_{\mathbf{G}}\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{f}_T = O_P(\sqrt{N}) \\
\mathbf{H}'\mathbf{A}_1\widehat{\mathbf{f}}_T &= \mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{M}_{\mathbf{G}}\widehat{\mathbf{F}})^{-1}\mathbf{e}_T + \mathbf{H}'(\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{M}_{\mathbf{G}}\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{f}_T = O_P(1) \\
-\mathbf{A}_2 &= \mathbf{A}_1\widehat{\mathbf{F}}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1} = \mathbf{A}_1\mathbf{E}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1} + \mathbf{A}_1\mathbf{H}\mathbf{F}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1} = O_P(\sqrt{\frac{N}{T}} + \nu_{\min}^{-1}) \\
-\mathbf{H}'\mathbf{A}_2 &= \mathbf{H}'\mathbf{A}_1\mathbf{E}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1} + \mathbf{H}'\mathbf{A}_1\mathbf{H}\mathbf{F}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1} = O_P(1) \\
\mathbf{A}_2'\widehat{\mathbf{f}}_T &= \mathbf{A}_2'\mathbf{H}\mathbf{f}_T + \mathbf{A}_2'\mathbf{e}_T = O_P(1).
\end{aligned}
$$

Finally, it follows from Proposition A.1 that $\frac{1}{T}\mathbf{G}'(\mathbf{P}_{\widehat{\mathbf{F}}} - \mathbf{P}_{\mathbf{F}\mathbf{H}'})\mathbf{G} = O_P(\frac{1}{T} + \frac{1}{N\nu_{\min}^2})$. Hence $\|\mathbf{A}_3\| = O_P(1)$ since $\lambda_{\min}(\frac{1}{T}\mathbf{G}'\mathbf{M}_{\mathbf{F}\mathbf{H}'}\mathbf{G}) > c$.

□

## A.4 Proof of Theorem 3.2

Let $\widehat{\boldsymbol{\varepsilon}}_g$, $\widehat{\boldsymbol{\varepsilon}}_y$,$\boldsymbol{\varepsilon}_g$, $\boldsymbol{\varepsilon}_y$, $\mathbf{Y}$, $\mathbf{G}$ and $\boldsymbol{\eta}$ be $T \times 1$ vectors of $\widehat{\varepsilon}_{g,t}$, $\widehat{\varepsilon}_{y,t}$, $\varepsilon_{g,t}$, $\varepsilon_{y,t}$, $y_t$, $\mathbf{g}_t$ and $\eta_t$. Let $\widehat{J}$ denote the index set of components in $\widehat{\mathbf{u}}_t$ that are selected by *either* $\widehat{\boldsymbol{\gamma}}$ *or* $\widehat{\boldsymbol{\theta}}$. Let $\widehat{\mathbf{U}}_{\widehat{J}}$ denote the $N \times |J|_0$ matrix of rows of $\widehat{\mathbf{U}}$ selected by $J$. Then

$$
\widehat{\boldsymbol{\varepsilon}}_y = \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{Y}, \quad \widehat{\boldsymbol{\varepsilon}}_g = \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{G}.
$$

### A.4.1 The case $r \geq 1$.

*Proof.* From Lemma A.7

$$
\begin{aligned}
\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{T}[(\widehat{\boldsymbol{\varepsilon}}_g'\widehat{\boldsymbol{\varepsilon}}_g)^{-1}\widehat{\boldsymbol{\varepsilon}}_g'(\widehat{\boldsymbol{\varepsilon}}_y - \boldsymbol{\varepsilon}_y) + (\widehat{\boldsymbol{\varepsilon}}_g'\widehat{\boldsymbol{\varepsilon}}_g)^{-1}\widehat{\boldsymbol{\varepsilon}}_g'\boldsymbol{\eta} + (\widehat{\boldsymbol{\varepsilon}}_g'\widehat{\boldsymbol{\varepsilon}}_g)^{-1}\widehat{\boldsymbol{\varepsilon}}_g'(\boldsymbol{\varepsilon}_g - \widehat{\boldsymbol{\varepsilon}}_g)\boldsymbol{\beta}] \\
&= O_P(1)\frac{1}{\sqrt{T}}\widehat{\boldsymbol{\varepsilon}}_g'(\widehat{\boldsymbol{\varepsilon}}_y - \boldsymbol{\varepsilon}_y) + O_P(1)\frac{1}{\sqrt{T}}\widehat{\boldsymbol{\varepsilon}}_g'(\boldsymbol{\varepsilon}_g - \widehat{\boldsymbol{\varepsilon}}_g) + O_P(1)\frac{1}{\sqrt{T}}\boldsymbol{\eta}'(\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g) \\
&\quad + (\frac{1}{T}\boldsymbol{\varepsilon}_g'\boldsymbol{\varepsilon}_g)^{-1}\frac{1}{\sqrt{T}}\boldsymbol{\varepsilon}_g'\boldsymbol{\eta} \\
&= \sigma_g^{-2}\frac{1}{\sqrt{T}}\boldsymbol{\varepsilon}_g'\boldsymbol{\eta} + o_P(1) \xrightarrow{d} \mathcal{N}(0, \sigma_g^{-4}\sigma_{\eta g}^2). \qquad \text{(A.3)}
\end{aligned}
$$

In the above, we used the condition that $|J|_0^4+|J|_0^2\log^2 N = o(T)$ , $T|J|_0^4 = o(N^2\min\{1,\nu_{\min}^4|J|_0^4\})$ and $\sqrt{\log N}|J|_0^2 = o(N\nu_{\min}^2)$, whose sufficient conditions are $T|J|_0^4 = o(N^2\min\{1,\nu_{\min}^4|J|_0^4\})$ and $|J|_0^4\log^2 N = o(T)$.

In addition, $\widehat{\sigma}_{\eta,g}^{-1}\widehat{\sigma}_g^2\sqrt{T}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{d}\mathcal{N}(0,1)$, follows from $\widehat{\sigma}_g^2 := \frac{1}{T}\widehat{\boldsymbol{\varepsilon}}_g'\widehat{\boldsymbol{\varepsilon}}_g \xrightarrow{P} \sigma_g^2$.

□

**Proposition A.2.** *Suppose* $T = O(\nu_{\min}^4 N^2\log N)$, $|J|_0^2 T = O(\nu_{\min}^2 N^2\log N)$, $|J|_0^2 = O(N\nu_{\min}^2\log N)$ *and* $|J|_0^2\log N = O(T)$, $|J|_0^2 = o(N)$ *For all* $R\geq r$,
*(i)* $\frac{1}{T}\|\widehat{\mathbf{U}}'\boldsymbol{\theta}-\widehat{\mathbf{U}}'\widetilde{\boldsymbol{\theta}}\|^2 = O_P(|J|_0\frac{\log N}{T})$ *and* $\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_1 = O_P(|J|_0\sqrt{\frac{\log N}{T}})$.
*(ii)* $|\widehat{J}|_0 = O_P(|J|_0)$.

*Proof.* (i) Let $L(\boldsymbol{\theta}) := \frac{1}{T}\sum_{t=1}^T(\mathbf{g}_t - \widehat{\boldsymbol{\alpha}}_g'\widehat{\mathbf{f}}_t - \boldsymbol{\theta}'\widehat{\mathbf{u}}_t)^2 + \tau\|\boldsymbol{\theta}\|_1$,

$$d_t = \boldsymbol{\alpha}_g'\mathbf{f}_t - \widehat{\boldsymbol{\alpha}}_g'\widehat{\mathbf{f}}_t + (\mathbf{u}_t-\widehat{\mathbf{u}}_t)'\boldsymbol{\theta},\quad \boldsymbol{\Delta} = \boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}.$$

Then $\mathbf{g}_t = \boldsymbol{\alpha}_g'\mathbf{f}_t + \boldsymbol{\theta}'\mathbf{u}_t + \varepsilon_{g,t}$, and $L(\widetilde{\boldsymbol{\theta}})\leq L(\boldsymbol{\theta})$ imply

$$\frac{1}{T}\sum_{t=1}^T[(\widehat{\mathbf{u}}_t'\boldsymbol{\Delta})^2 + 2(\varepsilon_{g,t}+d_t)\widehat{\mathbf{u}}_t'\boldsymbol{\Delta}] + \tau\|\widetilde{\boldsymbol{\theta}}\|_1 \leq \tau\|\boldsymbol{\theta}\|_1.$$

It follows from Lemma A.5 that $\|\frac{1}{T}\widehat{\mathbf{U}}\boldsymbol{\varepsilon}_g\|_\infty \leq O_P(\sqrt{\frac{\log N}{T}})$. Also Lemma A.4 implies that

$$\begin{aligned}\|\frac{1}{T}\sum_{t=1}^T d_t\widehat{\mathbf{u}}_t\|_\infty &\leq \|\frac{1}{T}\widehat{\mathbf{U}}\mathbf{E}\mathbf{H}^{+'}\boldsymbol{\alpha}\|_\infty + \|\frac{1}{T}\widehat{\mathbf{U}}\mathbf{E}(\mathbf{H}^{+'}\boldsymbol{\alpha}_g-\widehat{\boldsymbol{\alpha}}_g)\|_\infty + \|\frac{1}{T}\widehat{\mathbf{U}}\mathbf{F}\mathbf{H}'(\mathbf{H}^{+'}\boldsymbol{\alpha}_g-\widehat{\boldsymbol{\alpha}}_g)\|_\infty \\ &\quad + \|\frac{1}{T}\boldsymbol{\theta}'(\widehat{\mathbf{U}}-\mathbf{U})\widehat{\mathbf{U}}'\|_\infty \\ &\leq O_P(|J|_0\sqrt{\frac{\log N}{TN}} + |J|_0\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2} + \nu_{\min}^{-1}\sqrt{\frac{\log N}{TN}} + \frac{|J|_0}{N\nu_{\min}} + \frac{|J|_0}{\nu_{\min}\sqrt{NT}}).\end{aligned}$$

Thus the "score" satisfies $\|\frac{1}{T}\sum_{t=1}^T 2(\boldsymbol{\varepsilon}_{g,t}+d_t)\widehat{\mathbf{u}}_t'\|_\infty \leq \tau/2$ for sufficiently large $C>0$ in $\tau = C\sigma\sqrt{\frac{\log N}{T}}$ with probability arbitrarily close to one, given $T = O(\nu_{\min}^4 N^2\log N)$, $|J|_0^2 T = O(\nu_{\min}^2 N^2\log N)$, $|J|_0^2 = O(N\nu_{\min}^2\log N)$ and $|J|_0^2\log N = O(T)$. Then by the standard argument in the lasso literature,

$$\frac{1}{T}\sum_{t=1}^T(\widehat{\mathbf{u}}_t'\boldsymbol{\Delta})^2 + \frac{\tau}{2}\|\boldsymbol{\Delta}_{J^c}\|_1 \leq \frac{3\tau}{2}\|\boldsymbol{\Delta}_J\|_1.$$

Meanwhile, by the restricted eigenvalue condition and Lemma A.4,

$$\frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{u}}_t'\mathbf{\Delta})^2 \geq \frac{1}{T}\sum_{t=1}^{T}(\mathbf{u}_t'\mathbf{\Delta})^2 - \|\mathbf{\Delta}\|_1^2\|\frac{1}{T}\widehat{\mathbf{U}}\widehat{\mathbf{U}}' - \mathbf{U}\mathbf{U}'\|_\infty \geq \|\mathbf{\Delta}\|_2^2(\phi_{\min} - o_P(1))$$

where the last inequality follows from $|J|_0 O_P(\nu_{\min}^{-2}\frac{1}{N} + \frac{\log N}{T}) = o_P(1)$ (Lemma A.3). From here, the desired convergence results follow from the standard argument in the lasso literature, we omit details for brevity, and refer to, e.g., Hansen and Liao (2018).

(ii) The proof of $|\widehat{J}|_0 = O_P(|J|_0)$ also follows from the standard argument in the lasso literature, we omit details but refer to the proof of Proposition D.1 of Hansen and Liao (2018) and Belloni et al. (2014).

□

**Lemma A.3.** *(i)* $\|\frac{1}{T}\mathbf{E}'\mathbf{U}'\|_\infty = O_P(\sqrt{\frac{\log N}{TN}} + \frac{1}{N})$
*(ii)* $\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\| = O_P(\frac{1}{N})$, $\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\|_\infty = O_P(\sqrt{\frac{\log N}{TN}} + \frac{1}{N})$ ,
*(iii)* $\|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})(\widehat{\mathbf{U}} - \mathbf{U})'\|_\infty + 2\|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})\mathbf{U}'\|_\infty = O_P(\nu_{\min}^{-2}\frac{1}{N} + \frac{\log N}{T})$.
*(iv)* $\|\frac{1}{T}\widehat{\mathbf{U}}\widehat{\mathbf{U}}' - \frac{1}{T}\mathbf{U}\mathbf{U}'\|_\infty = O_P(\nu_{\min}^{-2}\frac{1}{N} + \frac{\log N}{T})$.

*Proof.* Let $\widehat{\mathbf{F}} = (\widehat{\mathbf{f}}_1, \cdots, \widehat{\mathbf{f}}_T)'$. In addition, $\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}^+ = -\mathbf{B}\mathbf{H}^+\mathbf{E}'\widehat{\mathbf{F}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} + \mathbf{U}\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} + \mathbf{U}\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}$. Therefore,

$$\begin{aligned}\mathbf{U} - \widehat{\mathbf{U}} &= \widehat{\mathbf{B}}\widehat{\mathbf{F}}' - \mathbf{B}\mathbf{F}' = (\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}^+)\widehat{\mathbf{F}}' + \mathbf{B}\mathbf{H}^+\mathbf{E}' \\ &= -\mathbf{B}\mathbf{H}^+\mathbf{E}'\widehat{\mathbf{F}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}' + \mathbf{U}\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}' + \mathbf{U}\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}' + \mathbf{B}\mathbf{H}^+\mathbf{E}'. \quad \text{(A.4)}\end{aligned}$$

(i) We have

$$\|\frac{1}{T}\mathbf{U}\mathbf{E}\|_\infty \leq \sum_{k\leq r}\max_{i\leq N}|\frac{1}{TN}\sum_t\sum_j(u_{it}u_{jt} - \mathbb{E}\,u_{it}u_{jt})w_{k,j}| + O(\frac{1}{N}) = O_P(\sqrt{\frac{\log N}{TN}} + \frac{1}{N})$$

(ii) By Proposition A.1 , Lemma A.1 , $\nu_{\min} \gg N^{-1/2}$, and $\|\frac{1}{T}\mathbf{F}'\mathbf{U}'\|_\infty = O_P(\sqrt{\frac{\log N}{T}})$

$$\begin{aligned}\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\| &\leq \|\frac{1}{T}\mathbf{E}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{E}\| + \|\frac{2}{T}\mathbf{E}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{E}\| + \|\frac{1}{T}\mathbf{E}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{E}\| \\ &\leq O_P(\frac{1}{N}) \\ \|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\|_\infty &\leq \|\frac{1}{T}\mathbf{E}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{U}'\|_\infty + \|\frac{1}{T}\mathbf{E}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{U}'\|_\infty \\ &\quad + \|\frac{1}{T}\mathbf{E}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{U}'\|_\infty + \|\frac{1}{T}\mathbf{E}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{U}'\|_\infty\end{aligned}$$

$$\leq \quad O_P(\sqrt{\frac{\log N}{TN}} + \frac{1}{N}).$$

(iii) We have $\|\mathbf{H}^+\| = O(\nu_{\min}^{-1})$. Also, $\|\widehat{\mathbf{F}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'\| \leq 1$. In addition, by Lemma A.1, $\|(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'\|^2 = \|(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\| \leq O_P(\frac{N}{T})$ and that $\|\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'\|^2 = \|\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\| = O_P(\frac{1}{T})$. Next, by Lemma A.1, $\|\mathbf{E}\| = O_P(\sqrt{\frac{T}{N}})$, and $\max_i \|\mathbf{b}_i\| < C$. Substitute the expansion (A.4), and by Proposition A.1,

$$\begin{aligned}
& \|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})(\widehat{\mathbf{U}} - \mathbf{U})'\|_\infty + 2\|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})\mathbf{U}'\|_\infty \\
\leq \; & \|\frac{2}{T}\mathbf{B}\mathbf{H}^+\mathbf{E}'\mathbf{U}'\|_\infty + \|\frac{1}{T}\mathbf{B}\mathbf{H}^+\mathbf{E}'\mathbf{E}\mathbf{H}^{+'}\mathbf{B}'\|_\infty + \|\frac{3}{T}\mathbf{U}\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{U}'\|_\infty \\
& + \|\frac{4}{T}\mathbf{B}\mathbf{H}^+\mathbf{E}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{U}'\|_\infty + \|\frac{4}{T}\mathbf{B}\mathbf{H}^+\mathbf{E}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{U}'\|_\infty \\
& + \|(\frac{6}{T}\mathbf{U}\mathbf{E} + \frac{3}{T}\mathbf{U}\mathbf{F}\mathbf{H}')(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{U}'\|_\infty + \|\frac{4}{T}\mathbf{B}\mathbf{H}^+\mathbf{E}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}(\mathbf{H}\mathbf{F}'\mathbf{U}' + \mathbf{E}'\mathbf{U}')\|_\infty \\
& + \|\frac{2}{T}\mathbf{B}\mathbf{H}^+\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\|_\infty + \|\frac{3}{T}\mathbf{B}\mathbf{H}^+\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\mathbf{H}^{+'}\mathbf{B}'\|_\infty \\
\leq \; & \|\frac{C}{T}\mathbf{E}'\mathbf{U}'\|_\infty O_P(\nu_{\min}^{-1}) + \|\frac{C}{T}\mathbf{E}'\mathbf{E}\|O_P(\nu_{\min}^{-2}) + N\|\frac{C}{T}\mathbf{U}\mathbf{E}\|_\infty^2 + N\|\frac{C}{T}\mathbf{E}'\mathbf{E}\|\|\frac{1}{T}\mathbf{E}'\mathbf{U}'\|_\infty O_P(\nu_{\min}^{-1}) \\
& + O_P(\nu_{\min}^{-1})\|\frac{C}{T}\mathbf{E}'\mathbf{E}\|\|(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\|\|\mathbf{F}'\mathbf{U}'\|_\infty + \|\frac{6}{T}\mathbf{U}\mathbf{E}\|_\infty\|(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\|\|\mathbf{F}'\mathbf{U}'\|_\infty \\
& + \|\frac{3}{T}\mathbf{U}\mathbf{F}\|_\infty\|\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\|\|\mathbf{F}'\mathbf{U}'\|_\infty + O_P(\nu_{\min}^{-1})\|\frac{4}{T}\mathbf{E}'\mathbf{F}\|\|\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\|\|\mathbf{F}'\mathbf{U}'\|_\infty \\
& + O_P(\nu_{\min}^{-1})\|\frac{4}{T}\mathbf{E}'\mathbf{F}\|\|\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\|\|\mathbf{E}'\mathbf{U}'\|_\infty + O_P(\nu_{\min}^{-1})\|\frac{C}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\|_\infty + O_P(\nu_{\min}^{-2})\|\frac{C}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\| \\
= \; & O_P(\nu_{\min}^{-2}\frac{1}{N} + \frac{\log N}{T}).
\end{aligned}$$

Also, $\|\frac{1}{T}\widehat{\mathbf{U}}\widehat{\mathbf{U}}' - \frac{1}{T}\mathbf{U}\mathbf{U}'\|_\infty \leq \|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})(\widehat{\mathbf{U}} - \mathbf{U})'\|_\infty + 2\|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})\mathbf{U}'\|_\infty \leq O_P(\nu_{\min}^{-2}\frac{1}{N} + \frac{\log N}{T})$.

□

**Lemma A.4.** *For all $R \geq r$,*
*(i)* $\|\frac{1}{T}\boldsymbol{\theta}'(\widehat{\mathbf{U}} - \mathbf{U})\widehat{\mathbf{U}}'\|_\infty \leq O_P(\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2})|J|_0$.
*(ii)* $\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{F}\| = O_P(\frac{1}{N\nu_{\min}} + \frac{1}{\sqrt{NT}})$, $\|\frac{1}{T}\mathbf{U}\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{F}\|_\infty = O_P(\sqrt{\frac{\log N}{T}} + \frac{1}{N\nu_{\min}})$.
*(iii)* $\|\frac{1}{T}\mathbf{E}'\widehat{\mathbf{U}}'\|_\infty \leq O_P(\sqrt{\frac{\log N}{TN}} + \frac{1}{N\nu_{\min}})$, $\|\frac{1}{T}\mathbf{F}'\widehat{\mathbf{U}}'\|_\infty \leq O_P(\sqrt{\frac{\log N}{T}} + \frac{1}{N\nu_{\min}^2})$,
*(iv)* $\|\frac{1}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{E}\| = |J|_0 O_P(\frac{1}{N} + \frac{1}{\sqrt{NT}})$, $\|\frac{1}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{F}\| = O_P(\sqrt{\frac{|J|_0}{T}})$,
*(v)* $\widehat{\boldsymbol{\alpha}}_g - \mathbf{H}^{+'}\boldsymbol{\alpha}_g = |J|_0 O_P(1 + \sqrt{\frac{N}{T}}) + O_P(\nu_{\min}^{-1})$, $\mathbf{H}'(\widehat{\boldsymbol{\alpha}}_g - \mathbf{H}^{+'}\boldsymbol{\alpha}_g) = O_P(\nu_{\min}^{-1}\frac{|J|_0}{N} + \sqrt{\frac{|J|_0}{T}} + \nu_{\min}^{-2}\frac{1}{N})$.

*Proof.* (i) By Lemma A.3 $\|\frac{1}{T}\boldsymbol{\theta}'(\widehat{\mathbf{U}}-\mathbf{U})\widehat{\mathbf{U}}'\|_\infty \leq \|\boldsymbol{\theta}\|_1\|\frac{1}{T}(\widehat{\mathbf{U}}-\mathbf{U})\widehat{\mathbf{U}}'\|_\infty \leq O_P(\frac{\log N}{T}+\frac{1}{N\nu_{\min}^2})|J|_0$.
(ii) Note $\mathbf{H}'\mathbf{H}^{+'}=\mathbf{I}$, Lemma A.3 shows $\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\| = O_P(\frac{1}{N})$, $\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\|_\infty = O_P(\sqrt{\frac{\log N}{TN}}+\frac{1}{N})$ ,

$$\begin{aligned}
\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{F}\| &\leq \|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\mathbf{H}^{+'}\| + \|\frac{1}{T}\mathbf{E}'\mathbf{E}\mathbf{H}^{+'}\| + \|\frac{1}{T}\mathbf{E}'\mathbf{F}\| = O_P(\frac{1}{N\nu_{\min}}+\frac{1}{\sqrt{NT}}) \\
\|\frac{1}{T}\mathbf{U}\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{F}\|_\infty &\leq \|\frac{1}{T}\mathbf{U}\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\mathbf{H}^{+'}\|_\infty + \|\frac{1}{T}\mathbf{U}\mathbf{E}\mathbf{H}^{+'}\|_\infty + \|\frac{1}{T}\mathbf{U}\mathbf{F}\|_\infty \\
&\leq O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N\nu_{\min}}).
\end{aligned}$$

(iii) By Lemma A.3 $\|\frac{1}{T}\mathbf{E}'\mathbf{U}'\|_\infty = O_P(\sqrt{\frac{\log N}{TN}}+\frac{1}{N})$ and (ii)

$$\begin{aligned}
\|\frac{1}{T}\widehat{\mathbf{U}}\mathbf{E}\|_\infty &\leq \|\frac{1}{T}\mathbf{U}\mathbf{E}\|_\infty + \|\frac{1}{T}(\widehat{\mathbf{U}}-\mathbf{U})\mathbf{E}\|_\infty \\
&\leq \|\frac{1}{T}\mathbf{U}\mathbf{E}\|_\infty + \|\frac{1}{T}\mathbf{B}\mathbf{H}^{+}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\|_\infty + \|\frac{1}{T}\mathbf{U}\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\|_\infty + \|\frac{1}{T}\mathbf{B}\mathbf{H}^{+}\mathbf{E}'\mathbf{E}\|_\infty \\
&\leq O_P(\sqrt{\frac{\log N}{TN}}+\frac{1}{N\nu_{\min}}) \\
\|\frac{1}{T}\widehat{\mathbf{U}}\mathbf{F}\|_\infty &\leq \|\frac{1}{T}\mathbf{U}\mathbf{F}\|_\infty + \|\frac{1}{T}(\widehat{\mathbf{U}}-\mathbf{U})\mathbf{F}\|_\infty \\
&\leq \|\frac{1}{T}\mathbf{U}\mathbf{F}\|_\infty + \|\frac{1}{T}\mathbf{B}\mathbf{H}^{+}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{F}\|_\infty + \|\frac{1}{T}\mathbf{U}\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{F}\|_\infty + \|\frac{1}{T}\mathbf{B}\mathbf{H}^{+}\mathbf{E}'\mathbf{F}\|_\infty \\
&\leq O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N\nu_{\min}^2}).
\end{aligned}$$

(iv) $\frac{1}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{E} = \frac{1}{NT}\boldsymbol{\theta}'(\mathbf{U}\mathbf{U}'-\mathbb{E}\,\mathbf{U}\mathbf{U}')\mathbf{W} + \frac{1}{NT}\boldsymbol{\theta}'\,\mathbb{E}\,\mathbf{U}\mathbf{U}'\mathbf{W}$. So

$$\begin{aligned}
\mathbb{E}\,\|\frac{1}{NT}\boldsymbol{\theta}'(\mathbf{U}\mathbf{U}'-\mathbb{E}\,\mathbf{U}\mathbf{U}')\mathbf{W}\|^2 &= \sum_{k=1}^{R}\frac{1}{N^2T^2}\operatorname{Var}(\sum_{t=1}^{T}\boldsymbol{\theta}'\mathbf{u}_t\mathbf{u}_t'\mathbf{w}_k) \\
&\leq \frac{C}{N^2T^2}\|\boldsymbol{\theta}\|_1^2\max_{j,i\leq N}\sum_{q,v\leq N}\sum_{t,s\leq T}|\operatorname{Cov}(u_{it}u_{qt},u_{js}u_{vs})| \leq \frac{C|J|_0^2}{NT}.
\end{aligned}$$

Also, $\|\frac{1}{NT}\boldsymbol{\theta}'\,\mathbb{E}\,\mathbf{U}\mathbf{U}'\mathbf{W}\| \leq \max_{j\leq N}\sum_k|w_{k,j}|\|\boldsymbol{\theta}\|_1\|\frac{1}{TN}\,\mathbb{E}\,\mathbf{U}\mathbf{U}'\|_1 \leq O(\frac{|J|_0}{N})$. Also,

$$\begin{aligned}
\mathbb{E}\,\|\frac{1}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{F}\|^2 &= \frac{1}{T^2}\operatorname{tr}\mathbb{E}\,\mathbf{F}'\,\mathbb{E}(\mathbf{U}'\boldsymbol{\theta}\boldsymbol{\theta}'\mathbf{U}|\mathbf{F})\mathbf{F} \leq \frac{C}{T}\|\,\mathbb{E}(\mathbf{U}'\boldsymbol{\theta}\boldsymbol{\theta}'\mathbf{U}|\mathbf{F})\|_1 \\
&\leq \frac{C}{T}\max_t\sum_{s=1}^{T}|\,\mathbb{E}(\boldsymbol{\theta}'\mathbf{u}_t\mathbf{u}_s'\boldsymbol{\theta}|\mathbf{F})| \leq \frac{C}{T}\max_t\sum_{s=1}^{T}\|\,\mathbb{E}(\mathbf{u}_t\mathbf{u}_s'|\mathbf{F})\|_1\|\boldsymbol{\theta}\|_1\|\boldsymbol{\theta}\|_\infty \leq \frac{C|J|_0}{T}.
\end{aligned}$$

(v) Since $\widehat{\boldsymbol{\alpha}}_g = (\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'\mathbf{G}$, simple calculations using Proposition A.1 yield

$$
\begin{aligned}
\widehat{\boldsymbol{\alpha}}_g - \mathbf{H}^{+'}\boldsymbol{\alpha}_g &= (\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'\mathbf{G} - \mathbf{H}^{+'}\boldsymbol{\alpha}_g \\
&= (\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\boldsymbol{\varepsilon}_g - (\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{E}\mathbf{H}^{+'}\boldsymbol{\alpha}_g + (\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{U}'\boldsymbol{\theta} + O_P(\sqrt{\frac{|J|_0}{T}}) \\
&= |J|_0 O_P(1+\sqrt{\frac{N}{T}}) + O_P(\nu_{\min}^{-1}) \\
\mathbf{H}'(\widehat{\boldsymbol{\alpha}}_g - \mathbf{H}^{+'}\boldsymbol{\alpha}_g) &= \mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\boldsymbol{\varepsilon}_g - \mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{E}\mathbf{H}^{+'}\boldsymbol{\alpha}_g + \mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{U}'\boldsymbol{\theta} + O_P(\sqrt{\frac{|J|_0}{T}}) \\
&= O_P(\nu_{\min}^{-1}\frac{|J|_0}{N} + \sqrt{\frac{|J|_0}{T}} + \nu_{\min}^{-2}\frac{1}{N}).
\end{aligned}
$$

□

**Lemma A.5.** *Suppose $|J|_0 = o(N\nu_{\min}^2)$. For any $R \geq r$*
*(i)* $\frac{1}{T}\|\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\boldsymbol{\theta}\|^2 = O_P(\frac{|J|_0^2}{N} + \frac{|J|_0^2}{T} + \frac{|J|_0^{3/2}}{\nu_{\min}N\sqrt{T}})$, $\frac{1}{T}\|\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_g\|^2 = O_P(\frac{1}{T})$,
*(ii)* $\|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})\boldsymbol{\varepsilon}_g\|_\infty = O_P(\frac{\nu_{\min}^{-1}}{\sqrt{NT}} + \frac{\sqrt{\log N}}{T})$, *and* $\|\frac{1}{T}\widehat{\mathbf{U}}\boldsymbol{\varepsilon}_g\|_\infty = O_P(\sqrt{\frac{\log N}{T}}) = \|\frac{1}{T}\widehat{\mathbf{U}}\boldsymbol{\varepsilon}_y\|_\infty$
*(iii)* $\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{U}}_{\hat{J}}\widehat{\mathbf{U}}_{\hat{J}}') > c_0$ *with probability approaching one.* $\frac{1}{T}\|\mathbf{P}_{\widehat{\mathbf{U}}_{\hat{J}}}\boldsymbol{\varepsilon}_g\|^2 = O_P(\frac{|J|_0\log N}{T}) = \frac{1}{T}\|\mathbf{P}_{\widehat{\mathbf{U}}_{\hat{J}}}\boldsymbol{\varepsilon}_y\|^2$.
*(iv)* $\frac{1}{T}\|(\widehat{\mathbf{U}} - \mathbf{U})'\boldsymbol{\theta}\|^2 = O_P(\frac{|J|_0^2+\nu_{\min}^{-2}}{N} + \frac{|J|_0^2}{T} + \frac{\nu_{\min}^{-1}|J|_0^{3/2}}{N\sqrt{T}})$, $\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_y = O_P(\frac{1}{\sqrt{NT}})$ ,
$\frac{1}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_y = O_P(\frac{|J|_0}{T} + \frac{|J|_0}{\sqrt{NT}} + \frac{\nu_{\min}^{-1/2}|J|_0^{3/4}}{\sqrt{N}T^{3/4}})$.

*Proof.* (i) By Lemma A.4 (vi) and Proposition A.1,

$$
\begin{aligned}
\frac{1}{T}\|\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\boldsymbol{\theta}\|^2 &= \frac{1}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{U}'\boldsymbol{\theta} + \frac{2}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{U}'\boldsymbol{\theta} \\
&\quad + \frac{1}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{U}'\boldsymbol{\theta} \\
&\leq O_P(\frac{|J|_0^2}{N} + \frac{|J|_0^2}{T} + \frac{|J|_0^{3/2}}{\nu_{\min}N\sqrt{T}}), \\
\frac{1}{T}\|\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_g\|^2 &= \frac{1}{T}\boldsymbol{\varepsilon}_g'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\boldsymbol{\varepsilon}_g + \frac{2}{T}\boldsymbol{\varepsilon}_g'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\boldsymbol{\varepsilon}_g + \frac{1}{T}\boldsymbol{\varepsilon}_g'\mathbf{F}\mathbf{H}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\boldsymbol{\varepsilon}_g \\
&\leq O_P(\frac{N}{NT}) + O_P(\frac{1}{\sqrt{NT}})\frac{\nu_{\min}^{-1}}{\sqrt{T}} + O_P(\frac{1}{T}) = O_P(\frac{1}{T}).
\end{aligned}
$$

(ii) By (A.4)

$$
\begin{aligned}
\frac{1}{T}(\mathbf{U} - \widehat{\mathbf{U}})\boldsymbol{\varepsilon}_g &= -\frac{1}{T}\mathbf{B}\mathbf{H}^{+}\mathbf{E}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\boldsymbol{\varepsilon}_g - \frac{1}{T}\mathbf{B}\mathbf{H}^{+}\mathbf{E}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\boldsymbol{\varepsilon}_g + \frac{1}{T}\mathbf{U}\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\boldsymbol{\varepsilon}_g \\
&\quad - \frac{1}{T}\mathbf{B}\mathbf{H}^{+}\mathbf{E}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\boldsymbol{\varepsilon}_g - \frac{1}{T}\mathbf{B}\mathbf{H}^{+}\mathbf{E}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\boldsymbol{\varepsilon}_g + \frac{1}{T}\mathbf{U}\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\boldsymbol{\varepsilon}_g
\end{aligned}
$$

$$+\frac{1}{T}\mathbf{UFH}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\boldsymbol{\varepsilon}_g+\frac{1}{T}\mathbf{UFH}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{HF}'\boldsymbol{\varepsilon}_g+\frac{1}{T}\mathbf{BH}^+\mathbf{E}'\boldsymbol{\varepsilon}_g.$$

So by Lemmas A.1 and $\|\frac{1}{T}\mathbf{UE}\|_\infty=O_P(\sqrt{\frac{\log N}{TN}}+\frac{1}{N})$, $\|\frac{1}{T}(\widehat{\mathbf{U}}-\mathbf{U})\boldsymbol{\varepsilon}_g\|_\infty=O_P(\frac{\nu_{\min}^{-1}}{\sqrt{NT}}+\frac{\sqrt{\log N}}{T})$.

Also, with $\|\frac{1}{T}\mathbf{U}\boldsymbol{\varepsilon}_g\|_\infty=O_P(\sqrt{\frac{\log N}{T}})$ we have $\|\frac{1}{T}\widehat{\mathbf{U}}\boldsymbol{\varepsilon}_g\|_\infty=O_P(\sqrt{\frac{\log N}{T}})$. The proof for $\|\frac{1}{T}\widehat{\mathbf{U}}\boldsymbol{\varepsilon}_y\|_\infty$ is the same.

(iii) First, it follows from Lemma A.4 that $\|\frac{1}{T}\widehat{\mathbf{U}}\widehat{\mathbf{U}}'-\frac{1}{T}\mathbf{UU}'\|_\infty\le O_P(\frac{\log N}{T}+\frac{\nu_{\min}^{-2}}{N})$.

Also by Proposition A.2, $|\widehat{J}|_0=O_P(|J|_0)$. Then with probability approaching one,

$$\begin{aligned}
\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{U}}_{\widehat{J}}\widehat{\mathbf{U}}_{\widehat{J}}') &\ge \lambda_{\min}(\frac{1}{T}\mathbf{U}_{\widehat{J}}\mathbf{U}_{\widehat{J}}')-\|\frac{1}{T}\widehat{\mathbf{U}}\widehat{\mathbf{U}}'-\frac{1}{T}\mathbf{UU}'\|_\infty|\widehat{J}|_0\\
&\ge \phi_{\min}-O_P(\frac{\log N}{T}+\frac{\nu_{\min}^{-2}}{N})|J|_0\ge c\\
\frac{1}{T}\|\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\boldsymbol{\varepsilon}_g\|^2 &= \frac{1}{T}\boldsymbol{\varepsilon}_g'\widehat{\mathbf{U}}_{\widehat{J}}'(\widehat{\mathbf{U}}_{\widehat{J}}\widehat{\mathbf{U}}_{\widehat{J}}')^{-1}\widehat{\mathbf{U}}_{\widehat{J}}\boldsymbol{\varepsilon}_g\le\|\frac{1}{T}\boldsymbol{\varepsilon}_g'\widehat{\mathbf{U}}_{\widehat{J}}'\|^2\lambda_{\min}^{-1}(\frac{1}{T}\widehat{\mathbf{U}}_{\widehat{J}}\widehat{\mathbf{U}}_{\widehat{J}}')\\
&\le c\|\frac{1}{T}\boldsymbol{\varepsilon}_g'\widehat{\mathbf{U}}'\|_\infty^2|\widehat{J}|_0\le O_P(\frac{|J|_0\log N}{T}).
\end{aligned}$$

$\frac{1}{T}\|\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\boldsymbol{\varepsilon}_y\|^2$ follows from the same proof.

(iv) Recall that $\|\boldsymbol{\alpha}_g'\|=\|\boldsymbol{\theta}'\mathbf{B}\|<C$. By part (i) and Lemma A.4,

$$\begin{aligned}
\frac{1}{T}\|\boldsymbol{\theta}'(\widehat{\mathbf{U}}-\mathbf{U})\|^2 &\le \frac{1}{T}\|\boldsymbol{\theta}'\mathbf{BH}^+\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\|^2+\frac{1}{T}\|\boldsymbol{\theta}'\mathbf{UP}_{\widehat{\mathbf{F}}}\|^2+\frac{1}{T}\|\boldsymbol{\theta}'\mathbf{BH}^+\mathbf{E}'\|^2\\
&\le O_P(\frac{|J|_0^2+\nu_{\min}^{-2}}{N}+\frac{|J|_0^2}{T}+\frac{\nu_{\min}^{-1}|J|_0^{3/2}}{N\sqrt{T}}).\\
\|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_y\| &\le \|\frac{1}{T}\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\|\|\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_y\|=O_P(\frac{1}{\sqrt{NT}})\\
\frac{1}{T}\boldsymbol{\theta}'\mathbf{UP}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_y &\le \frac{1}{T}\|\boldsymbol{\theta}'\mathbf{UP}_{\widehat{\mathbf{F}}}\|\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_y\|=O_P(\frac{|J|_0}{T}+\frac{|J|_0}{\sqrt{NT}}+\frac{\nu_{\min}^{-1/2}|J|_0^{3/4}}{\sqrt{N}T^{3/4}}).
\end{aligned}$$

□

**Lemma A.6.** *For any $R\ge r$*

*(i)* $\frac{1}{T}\|\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\widehat{\mathbf{U}}'\boldsymbol{\theta}\|^2=O_P(|J|_0\frac{\log N}{T})$, $\frac{1}{T}\|\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{U}'\boldsymbol{\theta}\|^2=O_P(\frac{|J|_0\log N}{T}+\frac{|J|_0^2+\nu_{\min}^{-2}}{N}+\frac{|J|_0^2}{T})$.

*(ii)* $\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}(\widehat{\mathbf{U}}-\mathbf{U})'\boldsymbol{\theta}=|J|_0^2\sqrt{\frac{\log N}{T}}O_P(\frac{\log N}{T}+\frac{1}{N\nu_{\min}^2})$,

$\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{U}'\boldsymbol{\theta}\le O_P(\frac{|J|_0\log N}{T}+\frac{|J|_0+\nu_{\min}^{-1}}{\sqrt{NT}}+\frac{\nu_{\min}^{-1/2}|J|_0^{3/4}}{\sqrt{N}T^{3/4}}+\sqrt{\frac{\log N}{T}}\frac{|J|_0^2}{N\nu_{\min}^2})$,

*(iii)* $\|\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{E}\|=O_P(\sqrt{\frac{|J|_0\log N}{N}}+\frac{\sqrt{T|J|_0}}{N\nu_{\min}})$, $\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{E}=O_P(\frac{|J|_0\log N}{T\sqrt{N}}+\frac{|J|_0\sqrt{\log N}}{N\nu_{\min}\sqrt{T}})$.

*Proof.* (i) First note that $\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\widehat{\mathbf{U}}'\boldsymbol{\theta} = \widehat{\mathbf{U}}'\widehat{\mathbf{m}}$, where

$$\widehat{\mathbf{m}} = (\widehat{m}_1, \cdots, \widehat{m}_N)' = \arg\min_{\mathbf{m}} \|\widehat{\mathbf{U}}'(\boldsymbol{\theta} - \mathbf{m})\| : \quad m_j = 0, \text{ for } j \notin \widehat{J}.$$

Thus by the definition of $\widehat{\mathbf{m}}$, Proposition A.2 and Lemma A.5,

$$\begin{aligned}
\frac{1}{T}\|\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\widehat{\mathbf{U}}'\boldsymbol{\theta}\|^2 &= \frac{1}{T}\|\widehat{\mathbf{U}}'\boldsymbol{\theta} - \widehat{\mathbf{U}}'\widehat{\mathbf{m}}\|^2 \leq \frac{1}{T}\|\widehat{\mathbf{U}}'\boldsymbol{\theta} - \widehat{\mathbf{U}}'\widetilde{\boldsymbol{\theta}}\|^2 \leq O_P(|J|_0\frac{\log N}{T}) \\
\frac{1}{T}\|\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{U}'\boldsymbol{\theta}\|^2 &\leq O_P(\frac{|J|_0\log N}{T}) + \frac{1}{T}\|(\widehat{\mathbf{U}} - \mathbf{U})'\boldsymbol{\theta}\|^2 = O_P(\frac{|J|_0\log N + |J|_0^2}{T} + \frac{|J|_0^2 + \nu_{\min}^{-2}}{N})
\end{aligned}$$

where we used $\frac{\nu_{\min}^{-1}|J|_0^{3/2}}{N\sqrt{T}} = O_P(\frac{|J|_0\log N}{T})$ by our assumption.

(ii) Let $\boldsymbol{\Delta} = \boldsymbol{\theta} - \widehat{\mathbf{m}}$. Then $\dim(\boldsymbol{\Delta}) = O_P(|J|_0)$. Also, by Lemma A.4,

$$\boldsymbol{\Delta}'\frac{1}{T}(\widehat{\mathbf{U}}\widehat{\mathbf{U}}' - \mathbf{U}\mathbf{U}')\boldsymbol{\Delta} \leq \|\boldsymbol{\Delta}\|_1^2\|\frac{1}{T}(\widehat{\mathbf{U}}\widehat{\mathbf{U}}' - \mathbf{U}\mathbf{U}')\|_\infty \leq O_P(\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2})\|\boldsymbol{\Delta}\|^2|J|_0.$$

Also, $\|\boldsymbol{\Delta}\|^2 \leq \frac{C}{T}\|\mathbf{U}'\boldsymbol{\Delta}\|^2$ due to the spare eigenvalue condition on $\frac{1}{T}\mathbf{U}\mathbf{U}'$. Then $\widetilde{\boldsymbol{\theta}}_j = 0$ for $j \notin \widehat{J}$ implies $\|\widehat{\mathbf{U}}'\boldsymbol{\Delta}\| \leq \|\widehat{\mathbf{U}}'(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})\|$ and Proposition A.2 implies

$$\begin{aligned}
\|\boldsymbol{\theta} - \widehat{\mathbf{m}}\|_1^2 &\leq |J|_0\|\boldsymbol{\Delta}\|^2 \leq |J|_0\frac{1}{T}\|\mathbf{U}'\boldsymbol{\Delta}\|^2 \leq |J|_0\frac{1}{T}\|\widehat{\mathbf{U}}'\boldsymbol{\Delta}\|^2 + O_P(\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2})\|\boldsymbol{\Delta}\|^2|J|_0 \\
&\leq |J|_0\frac{1}{T}\|\widehat{\mathbf{U}}'\boldsymbol{\theta} - \widehat{\mathbf{U}}'\widetilde{\boldsymbol{\theta}}\|^2 + O_P(\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2})\|\boldsymbol{\Delta}\|^2|J|_0 \\
&\leq \frac{|J|_0^2\log N}{T} + O_P(\frac{|J|_0\log N}{T} + \frac{|J|_0}{N\nu_{\min}^2})\|\boldsymbol{\Delta}\|^2.
\end{aligned}$$

The above implies $\|\boldsymbol{\theta} - \widehat{\mathbf{m}}\|_1^2 \leq O_P(|J|_0^2\frac{\log N}{T})$. Hence by Lemma A.5,

$$\begin{aligned}
\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}(\widehat{\mathbf{U}} - \mathbf{U})'\boldsymbol{\theta} &\leq \|\frac{1}{\sqrt{T}}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\|\|\widehat{\mathbf{U}}(\widehat{\mathbf{U}} - \mathbf{U})'\boldsymbol{\theta}\|_\infty\frac{\sqrt{|J|_0}}{T}\lambda_{\min}^{-1/2}(\frac{1}{T}\widehat{\mathbf{U}}_{\widehat{J}}\widehat{\mathbf{U}}_{\widehat{J}}') \\
&\leq |J|_0^2\sqrt{\frac{\log N}{T}}O_P(\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2}). \\
\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\widehat{\mathbf{U}}'\boldsymbol{\theta} &= \frac{1}{T}\boldsymbol{\varepsilon}_y'\widehat{\mathbf{U}}'(\boldsymbol{\theta} - \widehat{\mathbf{m}}) \leq \|\frac{1}{T}\boldsymbol{\varepsilon}_y'\widehat{\mathbf{U}}'\|_\infty\|\boldsymbol{\theta} - \widehat{\mathbf{m}}\|_1 \leq O_P(\frac{|J|_0\log N}{T}). \\
\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{U}'\boldsymbol{\theta} &\leq \frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\widehat{\mathbf{U}}'\boldsymbol{\theta} + \frac{1}{T}\boldsymbol{\varepsilon}_y'(\widehat{\mathbf{U}} - \mathbf{U})'\boldsymbol{\theta} - \frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}(\widehat{\mathbf{U}} - \mathbf{U})'\boldsymbol{\theta} \\
&\leq O_P(\frac{|J|_0\log N}{T}) + \frac{1}{T}\boldsymbol{\theta}'\mathbf{B}\mathbf{H}^+\mathbf{E}'\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_y + \frac{1}{T}\boldsymbol{\theta}'\mathbf{U}\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_y + \frac{1}{T}\boldsymbol{\theta}'\mathbf{B}\mathbf{H}^+\mathbf{E}'\boldsymbol{\varepsilon}_y \\
&\quad -\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}(\widehat{\mathbf{U}} - \mathbf{U})'\boldsymbol{\theta}
\end{aligned}$$

$$\leq \quad O_P(\frac{|J|_0 \log N}{T} + \frac{|J|_0 + \nu_{\min}^{-1}}{\sqrt{NT}} + \frac{\nu_{\min}^{-1/2}|J|_0^{3/4}}{\sqrt{N}T^{3/4}} + \sqrt{\frac{\log N}{T}} \frac{|J|_0^2}{N\nu_{\min}^2}).$$

(iii) By Lemma A.4,

$$\begin{aligned}
\|\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{E}\| &\leq \|\widehat{\mathbf{U}}_{\widehat{J}}'(\frac{1}{T}\widehat{\mathbf{U}}_{\widehat{J}}\widehat{\mathbf{U}}_{\widehat{J}}')^{-1}\|\frac{1}{T}\|\widehat{\mathbf{U}}\mathbf{E}\|_\infty\sqrt{|J|_0} \leq O_P(\sqrt{\frac{|J|_0 \log N}{N}} + \frac{\sqrt{T|J|_0}}{N\nu_{\min}}) \\
\|\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{E}\| &\leq \|\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\|\|\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{E}\| = O_P(\frac{|J|_0 \log N}{T\sqrt{N}} + \frac{|J|_0\sqrt{\log N}}{N\nu_{\min}\sqrt{T}})
\end{aligned}$$

□

**Lemma A.7.** *For any $R \geq r$,*
*(i)* $\frac{1}{T}\|\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g\|^2 = O_P(\frac{|J|_0^2+|J|_0\log N}{T} + \frac{|J|_0^2+\nu_{\min}^{-2}}{N} + \frac{|J|_0^{3/2}}{\nu_{\min}N\sqrt{T}}) = \frac{1}{T}\|\widehat{\boldsymbol{\varepsilon}}_y - \boldsymbol{\varepsilon}_y\|^2.$
*(ii)* $\frac{1}{T}\boldsymbol{\varepsilon}_y'(\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g) = O_P(\frac{|J|_0 \log N}{T} + \frac{|J|_0+\nu_{\min}^{-1}}{\sqrt{NT}} + \frac{\nu_{\min}^{-1/2}|J|_0^{3/4}}{\sqrt{N}T^{3/4}} + \sqrt{\frac{\log N}{T}}\frac{|J|_0^2}{N\nu_{\min}^2})$*. The same rate applies to* $\frac{1}{T}\boldsymbol{\varepsilon}_g'(\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g)$ , $\frac{1}{T}\boldsymbol{\eta}'(\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g)$, $\frac{1}{T}\boldsymbol{\varepsilon}_g'(\widehat{\boldsymbol{\varepsilon}}_y - \boldsymbol{\varepsilon}_y)$ *and* $\frac{1}{T}\boldsymbol{\varepsilon}_y'(\widehat{\boldsymbol{\varepsilon}}_y - \boldsymbol{\varepsilon}_y)$.
*(iii)* $\frac{1}{T}\widehat{\boldsymbol{\varepsilon}}_g'\widehat{\boldsymbol{\varepsilon}}_g = \frac{1}{T}\boldsymbol{\varepsilon}_g'\boldsymbol{\varepsilon}_g + o_P(1)$.

*Proof.* Note that $\widehat{\boldsymbol{\varepsilon}}_g = \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{G}$ and $\mathbf{G} = \mathbf{F}\boldsymbol{\alpha}_g + \mathbf{U}'\boldsymbol{\theta} + \boldsymbol{\varepsilon}_g$. Also, $\widehat{\mathbf{U}} = \mathbf{X}\mathbf{M}_{\widehat{\mathbf{F}}}$ implies

$$\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{P}_{\widehat{\mathbf{F}}} = 0, \text{ and } \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}.$$

Recall that $\mathbf{H}^+\mathbf{H} = \mathbf{I}$ and $\widehat{\mathbf{F}} = \mathbf{F}\mathbf{H}' + \mathbf{E}$, hence straightforward calculations yield

$$\begin{aligned}
\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g &= \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{U}'\boldsymbol{\theta} - \mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\boldsymbol{\theta} + \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{F}\boldsymbol{\alpha}_g - \mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\boldsymbol{\varepsilon}_g - \mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_g \\
&= \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{U}'\boldsymbol{\theta} - \mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\boldsymbol{\theta} - \mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\boldsymbol{\varepsilon}_g - \mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_g - (\mathbf{I} - \mathbf{P}_{\widehat{\mathbf{F}}} - \mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}})\mathbf{E}\mathbf{H}^{+'}\boldsymbol{\alpha}_g. \quad \text{(A.5)}
\end{aligned}$$

It follows from Lemmas A.5, A.6 that $\frac{1}{T}\|\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g\|^2 = O_P(\frac{|J|_0^2+|J|_0\log N}{T} + \frac{|J|_0^2+\nu_{\min}^{-2}}{N} + \frac{|J|_0^{3/2}}{\nu_{\min}N\sqrt{T}})$. The proof for $\frac{1}{T}\|\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g\|^2$ follows similarly.

(ii) It follows from (A.5) and Lemmas A.5 A.6 that

$$\begin{aligned}
\frac{1}{T}\boldsymbol{\varepsilon}_y'(\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g) &= \frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{U}'\boldsymbol{\theta} - \frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{U}'\boldsymbol{\theta} - \frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\boldsymbol{\varepsilon}_g - \frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_g \\
&\quad -\frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{E}\mathbf{H}^{+'}\boldsymbol{\alpha}_g - \frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{F}}}\mathbf{E}\mathbf{H}^{+'}\boldsymbol{\alpha}_g - \frac{1}{T}\boldsymbol{\varepsilon}_y'\mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{E}\mathbf{H}^{+'}\boldsymbol{\alpha}_g \\
&\leq O_P(\frac{|J|_0 \log N}{T} + \frac{|J|_0 + \nu_{\min}^{-1}}{\sqrt{NT}} + \frac{\nu_{\min}^{-1/2}|J|_0^{3/4}}{\sqrt{N}T^{3/4}} + \sqrt{\frac{\log N}{T}}\frac{|J|_0^2}{N\nu_{\min}^2}).
\end{aligned}$$

The same proof applies to other terms as well.

(iii) It follows from parts (i) that all these terms are $o_P(1)$, given that $|J|_0^2 = o(\min\{T, N\})$, $|J|_0 \log N = o(T)$.

□

### A.4.2 The case $r = 0$: there are no factors.

*Proof.* In this case $\mathbf{x}_t = \mathbf{u}_t$. And we have

$$\widehat{\mathbf{F}} = \frac{1}{N}\mathbf{X}'\mathbf{W} = \frac{1}{N}\mathbf{U}'\mathbf{W} := \mathbf{E}.$$

Then $\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}) = \lambda_{\min}(\frac{1}{T}\mathbf{E}'\mathbf{E}) \geq \frac{c}{N}$ with probability approaching one, still by Lemma A.1. Hence $\frac{1}{T}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}$ is still invertible. In addition, $\widehat{\mathbf{U}} = \mathbf{X}\mathbf{M}_{\widehat{\mathbf{F}}}$ implies $\mathbf{U} - \widehat{\mathbf{U}} = \mathbf{U}\mathbf{P}_{\mathbf{E}}$. Also,

$$\begin{aligned} y_t &= \boldsymbol{\gamma}'\mathbf{u}_t + \boldsymbol{\varepsilon}_{y,t} \\ \mathbf{g}_t &= \boldsymbol{\theta}'\mathbf{u}_t + \boldsymbol{\varepsilon}_{g,t} \\ \boldsymbol{\varepsilon}_{y,t} &= \boldsymbol{\beta}'\boldsymbol{\varepsilon}_{g,t} + \eta_t \end{aligned}$$

Hence $\boldsymbol{\alpha}_g = \boldsymbol{\alpha}_y = 0$. Then $\frac{1}{T}\widehat{\mathbf{F}}'\widehat{\mathbf{F}} = \frac{1}{T}\mathbf{E}'\mathbf{E} = \frac{1}{N^2}\mathbf{W}'\operatorname{Cov}(\mathbf{u}_t)\mathbf{W} + O_P(\frac{1}{N\sqrt{T}})$. Hence with probability approaching one $\lambda_{\min}(\frac{1}{T}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}) \geq cN^{-1}$. In addition, $\widehat{\boldsymbol{\alpha}}_y = (\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbf{U}'\boldsymbol{\gamma} + (\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\boldsymbol{\varepsilon}_y$ implies $\frac{1}{T}\sum_{t=1}^T(\widehat{\boldsymbol{\alpha}}_y'\widehat{\mathbf{f}}_t)^2 = O_P(\frac{|J|_0^2}{N} + \frac{|J|_0^2}{T})$.

As for the "score" $\max_i |\frac{1}{T}\sum_t(\varepsilon_{g,t} + d_t)\widehat{u}_{it}|$ in the proof of Proposition A.2, note that

$$\begin{aligned} \max_{i,j\leq N}|\frac{1}{T}\sum_t(\widehat{u}_{it}\widehat{u}_{jt} - u_{it}u_{jt})| &\leq \frac{3}{T}\|\mathbf{U}\mathbf{P}_{\mathbf{E}}\mathbf{U}'\|_\infty = O_P(\frac{1}{N} + \frac{\log N}{T}) \\ \max_{i\leq N}|\frac{1}{T}\sum_t \widehat{\boldsymbol{\alpha}}_y'\widehat{\mathbf{f}}_t\widehat{u}_{it}| &= O_P(\frac{|J|_0}{N} + \frac{|J|_0\log N}{T}) \\ \max_{i\leq N}|\frac{1}{T}\sum_t \widehat{u}_{it}(\mathbf{u}_t - \widehat{\mathbf{u}}_t)'\boldsymbol{\theta}| &= \frac{1}{T}\|\mathbf{U}\mathbf{P}_{\mathbf{E}}\mathbf{U}'\|_\infty O_P(|J|_0) = O_P(\frac{|J|_0}{N} + \frac{|J|_0\log N}{T}) \\ \max_{i\leq N}|\frac{1}{T}\sum_t \widehat{u}_{it}\varepsilon_{g,t}| &= O_P(\frac{\sqrt{\log N}}{T} + \frac{1}{\sqrt{TN}}). \end{aligned} \tag{A.6}$$

As for the residual, note that $\widehat{\boldsymbol{\varepsilon}}_g = \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{M}_{\mathbf{E}}\mathbf{G}$ and $\mathbf{G} = \mathbf{U}'\boldsymbol{\theta} + \boldsymbol{\varepsilon}_g$. Then

$$\widehat{\boldsymbol{\varepsilon}}_g - \boldsymbol{\varepsilon}_g = \mathbf{M}_{\widehat{\mathbf{U}}_{\widehat{J}}}\mathbf{U}'\boldsymbol{\theta} - \mathbf{P}_{\mathbf{E}}\mathbf{U}'\boldsymbol{\theta} - \mathbf{P}_{\widehat{\mathbf{U}}_{\widehat{J}}}\boldsymbol{\varepsilon}_g - \mathbf{P}_{\mathbf{E}}\boldsymbol{\varepsilon}_g.$$

All the proofs in Section A.4.1 carry over. In fact, all terms involving $\boldsymbol{\alpha}_g, \mathbf{H}$ and $\mathbf{H}^+$ can be set to zero.

In addition, in the case $R = r = 0$, the setting/estimators are the same as in Belloni et al. (2014). $\square$

### A.4.3 Proof of Corollary 3.1.

*Proof.* The corollary immediately follows from Theorem 3.2. If there exist a pair $(r, R)$ that violate the conclusion of the corollary, then it also violates the conclusion of Theorem 3.2. This finishes the proof. $\square$

## A.5 Proof of Theorem 3.3

*Proof.* In the proof of Theorem 3.3 we assume $R \geq r$.

(i) When $r > 0$, by Lemma A.3,

$$\max_{i,j\leq N} |\frac{1}{T}\sum_t (\widehat{u}_{it}\widehat{u}_{jt} - u_{it}u_{jt})| \leq \|\frac{1}{T}\widehat{\mathbf{U}}\widehat{\mathbf{U}}' - \frac{1}{T}\mathbf{U}\mathbf{U}'\|_\infty \leq O_P(\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2}).$$

When $r = 0$ and $R > 0$, by (A.6), $\max_{i,j\leq N} |\frac{1}{T}\sum_t (\widehat{u}_{it}\widehat{u}_{jt} - u_{it}u_{jt})| \leq O_P(\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2})$.

In both cases, part (i) implies, for $\nu_{\min}^2 \gg \frac{1}{\sqrt{N}}$ or $\nu_{\min}^2 \gg \frac{1}{N}\sqrt{\frac{T}{\log N}}$,

$$\begin{aligned}
\max_{i,j\leq N} |s_{u,ij} - \mathbb{E}\, u_{it}u_{jt}| &\leq \max_{i,j\leq N} |\frac{1}{T}\sum_t \widehat{u}_{it}\widehat{u}_{jt} - u_{it}u_{jt}| + \max_{i,j\leq N} |\frac{1}{T}\sum_t u_{it}u_{jt} - \mathbb{E}\, u_{it}u_{jt}| \\
&\leq O_P(\sqrt{\frac{\log N}{T}} + \frac{1}{N\nu_{\min}^2}) = O_P(\sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}).
\end{aligned}$$

where $\max_{i,j\leq N} |\frac{1}{T}\sum_t u_{it}u_{jt} - \mathbb{E}\, u_{it}u_{jt}| = O_P(\sqrt{\frac{\log N}{T}})$.

Given this convergence, the convergence of $\widehat{\mathbf{\Sigma}}_u$ and $\widehat{\mathbf{\Sigma}}_u^{-1}$ in (ii)(iii) then follows from the same proof of Theorem A.1 of Fan et al. (2013). We thus omit it for brevity. Finally, the case $r = R = 0$ is the usual case of sparse thresholding as in Bickel and Levina (2008). $\square$

## A.6 Proof of Theorem 3.4

*Proof.* First note that when $R = r$, by (A.2)

$$\|(\frac{1}{T}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} - (\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}')^{-1}\| \leq O_P(\frac{1}{N} + \frac{\nu_{\max}(\mathbf{H})}{\sqrt{TN}})\frac{1}{\nu_{\min}^4(\mathbf{H})}.$$

Also by the proof of Theorem 2.1 for $\|(\frac{1}{T}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\| + \|(\frac{1}{T}\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}')^{-1}\| \leq \frac{c}{\nu^2_{\min}(\mathbf{H})}$. Because $\mathbf{P}_{\widehat{\mathbf{F}}} - \mathbf{P}_{\mathbf{G}} = \mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}' + \mathbf{F}\mathbf{H}'[(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} - (\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}')^{-1}]\mathbf{H}\mathbf{F}' + \widehat{\mathbf{F}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'$, we have

$$\begin{aligned}
\|\mathbf{P}_{\widehat{\mathbf{F}}} - \mathbf{P}_{\mathbf{G}}\|_F^2 &= \operatorname{tr}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}'(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{E} + \operatorname{tr}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{E} \\
&\quad +2\operatorname{tr}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}'[(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} - (\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}')^{-1}]\mathbf{H}\mathbf{F}'\mathbf{E} \\
&\quad +\operatorname{tr}[(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} - (\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}')^{-1}]\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}'[(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} - (\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}')^{-1}]\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}' \\
&\quad +2\operatorname{tr}\mathbf{F}\mathbf{H}'[(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} - (\mathbf{H}\mathbf{F}'\mathbf{F}\mathbf{H}')^{-1}]\mathbf{H}\mathbf{F}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}' \\
&\quad +2\operatorname{tr}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{E}'\mathbf{E} \\
&\quad +2\operatorname{tr}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{E}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\mathbf{H}\mathbf{F}'\mathbf{E} \\
&= 2\operatorname{tr}\mathbf{H}'^{-1}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{H}^{-1}\mathbf{E}'\mathbf{E} + O_P(\frac{1}{TN\nu^2_{\min}} + \frac{1}{N^2\nu^4_{\min}} + \frac{1}{N\sqrt{NT}\nu^3_{\min}}).
\end{aligned}$$

Write $X := 2\operatorname{tr}\mathbf{H}'^{-1}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{H}^{-1}\mathbf{E}'\mathbf{E} = \operatorname{tr}(\mathbf{A}\frac{1}{T}\mathbf{E}'\mathbf{E})$ and $\mathbf{A} := 2\mathbf{H}'^{-1}(\frac{1}{T}\mathbf{F}'\mathbf{F})^{-1}\mathbf{H}^{-1}$. Now

$$\text{MEAN} = \mathbb{E}(X|\mathbf{F},\mathbf{W}) = \operatorname{tr}\mathbf{A}\frac{1}{N^2}\mathbf{W}'(\mathbb{E}\,\mathbf{u}_t\mathbf{u}_t'|\mathbf{F})\mathbf{W} = \operatorname{tr}\mathbf{A}\frac{1}{N^2}\mathbf{W}'\mathbf{\Sigma}_u\mathbf{W}.$$

We note that $\operatorname{Var}(X|\mathbf{F}) = \frac{1}{TN^2}\sigma^2$ and that $N\sqrt{T}\frac{(X-\text{MEAN})}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$ due to the serial indepence of $\mathbf{u}_t\mathbf{u}_t'$ conditionally on $\mathbf{F}$ and that $\mathbb{E}\,\|\frac{1}{\sqrt{N}}\mathbf{W}'\mathbf{u}_t\|^4 < C$. In addition, Lemma A.8 below shows that with $\widehat{\text{MEAN}} = \operatorname{tr}\widehat{\mathbf{A}}\frac{1}{N^2}\mathbf{W}'\widehat{\mathbf{\Sigma}}_u\mathbf{W}$, and $\widehat{\mathbf{A}} = 2(\frac{1}{T}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}$, we have

$$(\widehat{\text{MEAN}} - \text{MEAN})N\sqrt{T} = o_P(1).$$

Also, the same lemma shows $\widehat{\sigma}^2 \xrightarrow{P} \sigma^2$. As a result

$$\frac{\|\mathbf{P}_{\widehat{\mathbf{F}}} - \mathbf{P}_{\mathbf{G}}\|_F^2 - \widehat{\text{MEAN}}}{\frac{1}{N\sqrt{T}}\widehat{\sigma}} = \frac{X - \text{MEAN}}{\frac{1}{N\sqrt{T}}\sigma} + o_P(1) \xrightarrow{d} \mathcal{N}(0,1).$$

given that $\sigma > 0$, $\sqrt{T} = o(N)$.

□

**Lemma A.8.** *Suppose $R = r$. Let $g_{NT} := \nu^{-2}_{\min}\frac{1}{N} + \frac{\log N}{T}$.*
*(i)* $\widehat{\text{MEAN}} - \text{MEAN} = O_P(\frac{g^2_{NT}}{N^2\nu^2_{\min}})\sum_{\sigma_{u,ij}\neq 0} 1 + O_P(\frac{1}{N^2\nu^4_{\min}} + \frac{1}{N\sqrt{NT}\nu^3_{\min}})$.
*(ii)* $\widehat{\sigma}^2 \xrightarrow{P} \sigma^2$.

*Proof.* By lemma A.3,

$$\max_{ij}|\frac{1}{T}\sum_t u_{it}(\widehat{u}_{jt} - u_{jt})| \leq O_P(g_{NT}).$$

(i) Recall $\mathbf{A} := 2\mathbf{H}'^{-1}(\frac{1}{T}\mathbf{F}'\mathbf{F})^{-1}\mathbf{H}^{-1}$. Note that $\|\mathbf{A}\| = O_P(\frac{1}{\nu^2_{\min}(\mathbf{H})})$. We now bound $\frac{1}{N}\mathbf{W}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{W}$. For simplicity we focus on the case $r = R = 1$ and hard-thresholding estimator. The proof of SCAD thresholding follows from the same argument. We have

$$\frac{1}{N}\mathbf{W}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{W} = \frac{1}{N}\sum_{\sigma_{u,ij}=0} w_i w_j \widehat{\sigma}_{u,ij} + \frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} w_i w_j(\widehat{\sigma}_{u,ij} - \sigma_{u,ij}) := a_1 + a_2.$$

Term $a_1$ satisfies: for any $\epsilon > 0$, when $C$ in the threshold is large enough,

$$\mathbb{P}(a_1 > (NT)^{-2}) \leq \mathbb{P}(\max_{\sigma_{u,ij}=0} |\widehat{\sigma}_{u,ij}| \neq 0) \leq \mathbb{P}(|s_{u,ij}| > \tau_{ij}, \text{ for some } \sigma_{u,ij} = 0) < \epsilon.$$

Thus $a_1 = O_P((NT)^{-2})$. The main task is to bound $a_2 = \frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} w_i w_j(\widehat{\sigma}_{u,ij} - \sigma_{u,ij})$.

$$\begin{aligned} a_2 &= a_{21} + a_{22}, \\ a_{21} &= \frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} w_i w_j \frac{1}{T}\sum_t(\widehat{u}_{it}\widehat{u}_{jt} - u_{it}u_{jt}) \\ a_{22} &= \frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} w_i w_j \frac{1}{T}\sum_t(u_{it}u_{jt} - \mathbb{E}\, u_{it}u_{jt}). \end{aligned}$$

Now for $\omega_{NT} := \sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}$, by part (i),

$$\begin{aligned} a_{21} &= \frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} w_i w_j \frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})(\widehat{u}_{jt} - u_{jt}) + \frac{2}{N}\sum_{\sigma_{u,ij}\neq 0} w_i w_j \frac{1}{T}\sum_t u_{it}(\widehat{u}_{jt} - u_{jt}) \\ &\leq [\max_i \frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})^2 + \max_{ij}|\frac{1}{T}\sum_t u_{it}(\widehat{u}_{jt} - u_{jt})|]\frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} 1 \\ &\leq O_P(g^2_{NT})\frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} 1. \end{aligned}$$

As for $a_{22}$, due to $\frac{1}{N}\sum_{\sigma_{u,mn}\neq 0}\sum_{\sigma_{u,ij}\neq 0}|\operatorname{Cov}(u_{it}u_{jt}, u_{mt}u_{nt})| < C$ and serial independence,

$$\begin{aligned} \operatorname{Var}(a_{22}) &\leq \frac{1}{N^2T^2}\sum_{s,t\leq T}\sum_{\sigma_{u,mn}\neq 0}\sum_{\sigma_{u,ij}\neq 0}|\operatorname{Cov}(u_{it}u_{jt}, u_{ms}u_{ns})| \\ &\leq \frac{1}{N^2T}\sum_{\sigma_{u,mn}\neq 0}\sum_{\sigma_{u,ij}\neq 0}|\operatorname{Cov}(u_{it}u_{jt}, u_{mt}u_{nt})| \leq O(\frac{1}{NT}). \end{aligned}$$

Together $a_2 = O_P(g_{NT}^2)\frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} 1 + O_P(\frac{1}{\sqrt{NT}})$. Therefore

$$\frac{1}{N}\mathbf{W}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{W} = O_P(g_{NT}^2)\frac{1}{N}\sum_{\sigma_{u,ij}\neq 0} 1 + O_P(\frac{1}{\sqrt{NT}}).$$

This implies

$$\begin{aligned} |\widehat{\text{MEAN}} - \text{MEAN}| &\leq \frac{C}{N}\|\mathbf{A}\|\|\frac{1}{N}\mathbf{W}'(\mathbf{\Sigma}_u - \widehat{\mathbf{\Sigma}}_u)\mathbf{W}\| + O_P(\frac{1}{N})\|\mathbf{A} - 2(\frac{1}{T}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\| \\ &\leq O_P(\frac{g_{NT}^2}{N^2\nu_{\min}^2})\sum_{\sigma_{u,ij}\neq 0} 1 + O_P(\frac{1}{N^2\nu_{\min}^4} + \frac{1}{N\sqrt{NT}\nu_{\min}^3}). \end{aligned}$$

(ii) First, note that $|\sigma^2 - f(\mathbf{A},\mathbf{V})| \to 0$ by the assumption. In addition, it is easy to show that $\|\widehat{\mathbf{A}} - \mathbf{A}\| = o_P(1)$ and $\|\widehat{\mathbf{V}} - \mathbf{V}\| \leq \frac{1}{N}\|\mathbf{W}\|^2\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\| = o_P(1)$. Since $f(\mathbf{A},\mathbf{V})$ is continuous in $(\mathbf{A},\mathbf{V})$ due to the property of the normality of $\mathbf{Z}_t$, we have $|f(\mathbf{A},\mathbf{V}) - f(\widehat{\mathbf{A}},\widehat{\mathbf{V}})| = o_P(1)$. Hence $|f(\widehat{\mathbf{A}},\widehat{\mathbf{V}}) - \sigma^2| = o_P(1)$. This finishes the proof since $\widehat{\sigma}^2 := f(\widehat{\mathbf{A}},\widehat{\mathbf{V}})$.

□

# References


Belloni, A., Chernozhukov, V. and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* **81** 608–650.

Bickel, P. and Levina, E. (2008). Covariance regularization by thresholding. *Annals of Statistics* **36** 2577–2604.

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

Hansen, C. and Liao, Y. (2018). The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications. *Econometric Theory* 1–45.