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
A Technical Proofs
A.1 A key Proposition for asymptotic analysis when $R \geq r$ 2
A.2 Proof of Theorem 2.1 2
A.3 Proof of Theorem 3.1 9
A.4 Proof of Theorem 3.2 9
A.4.1 The case $r \geq 1$. 11
A.4.2 The case $r = 0$: there are no factors. 20
A.4.3 Proof of Corollary 3.1. 21
A.5 Proof of Theorem 3.3 21
A.6 Proof of Theorem 3.4 21

∗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 $\nu_{\min}(H)$ and $\nu_{\max}(H)$ respectively denote the minimum and maximum nonzero singular values of $H$. In addition, $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$ denote the projection matrices of a matrix $A$. If $A'A$ is singular, $(A'A)^{-1}$ is replaced with its Moore-Penrose generalized inverse $(A'A)^+$. Let $U$ be the $N \times T$ matrix of $u_{it}$. Recall that $R = \dim(\hat{f}_i)$ and $r = \dim(f_i)$.

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

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

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 $G$ is a $T \times d$ matrix so that $\mathbb{E}(U|G) = 0$, $\frac{1}{T} \|G\|^2 = O_P(1)$, for some fixed dimension $d$, and Assumption 2.1 - 2.4 hold. In addition, for each $K \in \{I_T, M_G\}$, suppose $\lambda_{\min}(\frac{1}{T} F'KF) > c > 0$. Then

(i) $\lambda_{\min}(\frac{1}{T} F'KF) \geq cN^{-1}$ with probability approaching one for some $c > 0$,

(ii) $\|H(\frac{1}{T} F'KF)^{-1}\| = O_P(\frac{N}{N_{\min}})$, and $\|H(\frac{1}{T} F'KF)^{-1}H\| = O_P(1)$.

(iii) $\|H(\frac{1}{T} F'KF)^{-1}H - H(\frac{1}{T} F'K F H)^+ H\| = O_P(\frac{1}{N_{\min}} + \frac{1}{T})$, and $\frac{1}{T} G' (P_{\hat{F}} - P_{F H'}) G = O_P(\frac{1}{N_{\min}} + \frac{1}{T})$.

*Proof.* The proof applies for both $K = I_T$ and $K = M_G$. 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 $\nu_{\min} := \nu_{\min}(H)$, and $\nu_{\max} := \nu_{\max}(H)$.

First, it is easy to see

$$\hat{F} = FH' + E,$$

where $E = (e_1, \cdots, e_T)' = \frac{1}{N}U'W$, which is $T \times R$. Write

$$\Delta := \frac{1}{T} E E'E + \frac{1}{T} H F' K F H' + \frac{1}{T} E'(E - E'E) + \Delta_1$$

where $\Delta_1 = 0$ if $K = I_T$ and $\Delta_1 = -\frac{1}{T} E'P_G E$ if $K = M_G$. 


(i) We have
\[ \frac{1}{T} \hat{F}' \hat{K} \hat{F} = \frac{1}{T} HF'KF' + \Delta. \]

By assumption \( \lambda_{\min}(\frac{1}{T} E U U') \geq c_0 \), so \( \lambda_{\min}(\frac{1}{T} E E' \hat{E}) \geq \lambda_{\min}(\frac{1}{T} E U U') \lambda_{\min}(\frac{1}{N^2} W W') \geq c_0 N^{-1} \) for some \( c_0 > 0 \). In addition, Lemma A.1 shows \( \frac{1}{T} H (E' E - E' E) + \Delta_1 = O_P(\frac{1}{N \sqrt{T}}) \).

Hence \( \| H(E' E - E' E) + \Delta_1 \| \leq \frac{1}{2} \lambda_{\min}(\frac{1}{T} E E') \) with large probability. We now continue the argument conditioning on this event.

Now let \( v \) be the unit vector so that \( v' \frac{1}{T} \hat{F}' \hat{K} \hat{F} v = \lambda_{\min}(\frac{1}{T} \hat{F}' \hat{K} \hat{F}) \) and let
\[ \eta_v^2 := \frac{1}{T} v' HF'KF'v. \]

Because \( v' \frac{1}{T} \hat{F}' \hat{K} \hat{F} v = \eta_v^2 + v' \Delta v \), we have
\[ \lambda_{\min}(\frac{1}{T} \hat{F}' \hat{K} \hat{F}) \geq \eta_v^2 + 2v' \frac{1}{T} HF'KEv + \frac{c_0}{2N}. \]

If \( v'H = 0 \) then \( \lambda_{\min}(\frac{1}{T} \hat{F}' \hat{K} \hat{F}) \geq \frac{c_0}{2N} \). If \( v'H \neq 0 \) then \( \eta_v^2 \neq 0 \) with large probability because \( \frac{1}{T} F'KF \) is positive definite. Now let
\[ X := (\frac{\eta_v^2}{TN})^{-1/2} 2v' \frac{1}{T} HF'KEv, \quad 2v' \frac{1}{T} HF'KEv = X \sqrt{\frac{\eta_v^2}{TN}}. \]

Then
\[ \lambda_{\min}(\frac{1}{T} \hat{F}' \hat{K} \hat{F}) \geq \eta_v^2 + X \sqrt{\frac{\eta_v^2}{TN}} + \frac{c_0}{2N}. \]

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

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
\[ \lambda_{\min}(\frac{1}{T} \hat{F}' \hat{K} \hat{F}) \geq \eta_v^2 - |X| \sqrt{\frac{\eta_v^2}{TN}} + \frac{c_0}{2N} \geq \frac{3}{4} \eta_v^2 + \frac{c_0}{2N} \geq \frac{c_0}{2N}. \]
In both cases, $\lambda_{\min}(\frac{1}{T}\hat{F}'K\hat{F}) > c_0/N$ for some $c_0 > 0$ with overwhelming probability.

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

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

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

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

(ii) Write $\bar{H} := H(\frac{1}{T}F'KF)^{1/2}$ and $S := \frac{N}{T}E'E = \frac{1}{N}W'S_uW$. Then

$$\frac{1}{T}\hat{F}'K\hat{F} = H\bar{H}' + \frac{1}{N}S + \frac{1}{T}HF'KE + \frac{1}{T}E'KFH' + \Delta_2 \quad (A.1)$$

where we proved in (i) that $\|\Delta_2\| = \|\frac{1}{T}(E'E - E'E') + \Delta_1\| = O_P(\frac{1}{\sqrt{NT}})$. Also all eigenvalues of $S$ are bounded away from both zero and infinity. In addition, $\bar{H}$ is a $R \times r$ matrix with $R \geq r$, whose Moore-Penrose generalized inverse is $\bar{H}^+ = (\frac{1}{T}F'KF)^{-1/2}H^+$. Also, $\bar{H}$ is of rank $r$. Let

$$\bar{H}' = U_H(D_H, 0)E'_H$$

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

The proof is divided into several steps.

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

Because $\lambda_{\min}(D_H) \geq cv_{\min}$, for $j = 1, 2$,

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

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

Let $0 < a < \lambda_{\min}(S)$ be a constant. Then $(\bar{H}\bar{H}' + \frac{a}{N}I)^{-1} - (\bar{H}\bar{H}' + \frac{1}{N}S)^{-1}$ is positive definite. (This is because, if both $A_1$ and $A_2 - A_1$ are positive definite, then so is $A_1^{-1} - A_2^{-1}$.) Let $v$ be a unit vector so that $v'\bar{H}'(\bar{H}\bar{H}' + \frac{1}{N}S)^{-1}\bar{H}v = \|\bar{H}'(\bar{H}\bar{H}' + \frac{1}{N}S)^{-1}\bar{H}\|$. Then

$$\|\bar{H}'(\bar{H}\bar{H}' + \frac{1}{N}S)^{-1}\bar{H}\| \leq v'\bar{H}'(\bar{H}\bar{H}' + \frac{a}{N}I)^{-1}\bar{H}v \leq \|\bar{H}'(\bar{H}\bar{H}' + \frac{a}{N}I)^{-1}\bar{H}\|.$$
The right hand side is $O_P(1)$ due to step 1.

Step 3. Show $\|\tilde{H}'(\tilde{H}' + \frac{\alpha}{N} I)^{-1}\| = O_P(\nu_\min^{-1})$.

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

$$
\|\tilde{H}'(\tilde{H}' + \frac{1}{N} S)^{-1}\| \leq \|M\| \|\tilde{H}'(\tilde{H}' + \frac{1}{N} S)^{-1} - M\|
$$

$= (1) \|M\| \|\tilde{H}'(\tilde{H}' + \frac{\alpha}{N} I)^{-1}(\frac{1}{N} S - \frac{\alpha}{N} I)(\tilde{H}' + \frac{1}{N} S)^{-1}\|

\leq \|M\| \|\tilde{H}'(\tilde{H}' + \frac{1}{N} S)^{-1}\|$

$= (2) \|M\||(1 + O_P(1)) = O_P(\nu_\min^{-1})$.

(1) used $A_1^{-1} - A_2^{-1} = A_1^{-1}(A_2 - A_1)A_2^{-1}$; (2) is from: $\|\tilde{H}'(\tilde{H}' + \frac{1}{N} S)^{-1}\| \leq \lambda_\min^{-1}(\frac{1}{N} S) = O_P(N)$.

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

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

$$
\|\tilde{H}'\frac{1}{T}\tilde{F}'K\tilde{F})^{-1} = \|\tilde{H}'A^{-1}(\frac{1}{T}\tilde{F}'K\tilde{F} - A)(\frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\|
$$

$\leq (3) O_P(\nu_\min^{-1}(H)\|(\frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\| = O_P(\sqrt{\frac{N}{T}}).

In (3) we used $\frac{1}{T}\tilde{F}'K\tilde{F} - A = O_P(\frac{1}{\sqrt{NT}} + \frac{1}{T}\tilde{F}'K\tilde{F}) = O_P(\frac{1}{\sqrt{NT}} + \frac{\nu_\max}{\sqrt{NT}}) = O_P(\frac{\nu_\max}{\sqrt{NT}})$; in

(4) we used $(\frac{1}{T}\tilde{F}'K\tilde{F})^{-1} = O_P(N)$ by part (i) and $\nu_\max \leq C\nu_\min$. Hence

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

Thus $\|\tilde{H}'\frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\| \leq \|\frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\|\|\tilde{H}'\frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\|$, which leads to the result for

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

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

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

$$
\|\tilde{H}'\frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\tilde{H}' - \tilde{H}'A^{-1}\tilde{H}\|
$$

$\leq \|\tilde{H}'\frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\tilde{H}' + \frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\tilde{H}\| + \|\tilde{H}'A^{-1}\tilde{H}' - \tilde{H}'A^{-1}\tilde{H}\|

+ \|\tilde{H}'A^{-1}\Delta_1(\frac{1}{T}\tilde{F}'K\tilde{F})^{-1}\tilde{H}\|
$$
\[ \leq O_P(\nu_{\min}^{-1} \frac{1}{N\sqrt{T}} + \frac{1}{\sqrt{NT}}) \| (\frac{1}{T} \hat{F}' \hat{K} \hat{F})^{-1} \hat{H} \| = \left(5\right) O_P(\frac{1}{\sqrt{NT}}) O_P(\nu_{\min}^{-1} + \sqrt{\frac{N}{T}}) = O_P(\frac{1}{\nu_{\min} \sqrt{NT}} + \frac{1}{T}). \]

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

\[ \hat{H}'(\frac{1}{T} \hat{F}' \hat{K} \hat{F})^{-1} \hat{H} = \hat{H}'(\frac{1}{T} \hat{F}' \hat{K} \hat{F} )^{-1} \hat{H} + O_P(\frac{1}{\nu_{\min} \sqrt{NT}} + \frac{1}{T}). \]

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

\[ \| \hat{H}'(\frac{1}{T} \hat{F}' \hat{K} \hat{F} )^{-1} \hat{H} \| = O_P(1 + \frac{1}{\nu_{\min} \sqrt{NT}} + \frac{1}{T}) = O_P(1). \]

(iii) The proof still consists of several steps.

Step 1. \( \hat{H}'(\frac{1}{T} \hat{F}' \hat{K} \hat{F} )^{-1} \hat{H} = \hat{H}'(\frac{1}{T} \hat{F}' \hat{K} \hat{F} )^{-1} \hat{H} + O_P(\frac{1}{\nu_{\min} \sqrt{NT}} + \frac{1}{T}). \)

It follows from step 5 of part (ii).

Step 2. show \( \hat{H}'(\hat{H} \hat{H}' + \frac{1}{N} \hat{S})^{-1} \hat{H} = \hat{H}'(\hat{H} \hat{H}')^+ \hat{H} + O_P(\frac{1}{\nu_{\min} \sqrt{NT}}) \) where \( \hat{H} = \hat{H}'(\frac{1}{T} \hat{F}' \hat{K} \hat{F} )^{1/2} \).

Write \( T = \hat{H}'(\hat{H} \hat{H}' + \frac{1}{N} \hat{S})^{-1} \hat{H} - \hat{H}'(\hat{H} \hat{H}')^+ \hat{H} \). The goal is to show \( \| T \| = O_P(\frac{1}{\nu_{\min} \sqrt{NT}}) \).

Let \( v \) be the unit vector so that \( |v'Tv| = \| T \|. \) Define a function, for \( d > 0, \)

\[ g(d) := v' \hat{H}'(\hat{H} \hat{H}' + \frac{d}{N} I)^{-1} \hat{H} v. \]

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

\[ |v'Tv| \leq |g(c) - v' \hat{H}'(\hat{H} \hat{H}')^+ \hat{H} v| + |g(C) - v' \hat{H}'(\hat{H} \hat{H}')^+ \hat{H} v|. \]

Recall \( \hat{H}' = U_{\hat{H}}(D_{\hat{H}}, 0)E_{\hat{H}}' \) is the SVD of \( \hat{H}' \) and \( N^{-1} \lambda_{\min}^{-1}(D_{\hat{H}}^2) = o_P(1) \). Then for any \( d \in \{c, C\} \), as \( N \to \infty, \)

\[ g(d) = v'U_{\hat{H}}D_{\hat{H}}^2(D_{\hat{H}}^2 + \frac{d}{N} I)^{-1}U_{\hat{H}}'v \to \hat{v}' \hat{v} = v' \hat{H}'(\hat{H} \hat{H}')^+ \hat{H} v, \]

where we used \( \hat{H}'(\hat{H} \hat{H}')^+ \hat{H} = I \), easy to see from its SVD. The rate of convergence is

\[ \| D_{\hat{H}}^2(D_{\hat{H}}^2 + \frac{d}{N} I)^{-1} I \| \leq \| D_{\hat{H}}^2(D_{\hat{H}}^2 + \frac{d}{N} I)^{-1} \| \frac{d}{N} D_{\hat{H}}^{-2} \| = O_P(\frac{1}{\nu_{\min}^2}). \]

Hence \( |v'Tv| = O_P(\frac{1}{\nu_{\min}^2}). \)

Step 3. show \( \| \hat{H}'(\frac{1}{T} \hat{F}' \hat{K} \hat{F})^{-1} \hat{H} - \hat{H}'(\hat{H} \hat{H}' + \frac{1}{T} \hat{F}' \hat{K} \hat{F} )^+ \hat{H} \| = O_P(\frac{1}{\nu_{\min}^2} + \frac{1}{T}). \) By steps 1 and
Lemma A.1. For any $R \geq 1$, $(R$ can be either smaller, equal to or larger than $r$),

(i) $\frac{1}{T} \mathbf{E} \mathbf{E}' \mathbf{E} \leq \frac{C}{N}$ and $\|\mathbf{E}\| = O_P(\sqrt{\frac{T}{N}})$.

(ii) $\|\frac{1}{T} \mathbf{F}' \mathbf{F}\|^{2} \leq O(\frac{1}{TN})$, $\|\frac{1}{T} \mathbf{G}' \mathbf{E}\|^{2} \leq O(\frac{1}{TN})$, here $\mathbf{G}$ is defined as in Section 3.1

(iii) $\|\frac{1}{T} (\mathbf{E}' \mathbf{E} - \mathbf{E} \mathbf{E}' \mathbf{E})\| \leq O_P(\frac{1}{N^{2}T})$, $\|\frac{1}{T} \mathbf{E}' \mathbf{P}_{\mathbf{E}} \mathbf{E}\| = O_P(\frac{1}{NT})$.

(iv) $\|\frac{1}{N} \mathbf{U}' \mathbf{W}\| \leq O_P(\sqrt{\frac{7}{N}})$.
Proof. (i) By the assumption $\|\frac{1}{T} \mathbf{E} \mathbf{U} \mathbf{U}'\| = \| \mathbf{E} \mathbf{u}_t \mathbf{u}_t' \| \leq \mathbf{E} \| \mathbf{u}_t \mathbf{u}_t' | \mathbf{F} \| < C$. Thus
\[
\| \frac{1}{T} \mathbf{E} \mathbf{E}' \| = \frac{1}{T^2} \mathbf{E} \| \mathbf{W}' \frac{1}{T} \mathbf{E} \mathbf{U} \mathbf{U}' \| \leq \frac{1}{N^2} \mathbf{E} \| \mathbf{W} \|^2 \leq \frac{C}{N}.
\]
Also, $\mathbf{E} \| \mathbf{E} \|^2 \leq \mathbf{tr} \mathbf{E} \mathbf{E}' \leq R \| \mathbf{E} \mathbf{E}' \| \leq \frac{C_T}{N}$.

(ii) Let $f_{k,t}$ be the $k$th entry of $\mathbf{f}_t$. By the assumption $\frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T \mathbf{E} \| f_t \| \| f_s \| \| \mathbf{E} (\mathbf{u}_t \mathbf{u}_s' | \mathbf{F}) \| < C$,
\[
\mathbf{E} \| \frac{1}{T} \mathbf{F} \mathbf{E} \|^2 = \frac{1}{T^2 N^2} \mathbf{E} \| \sum_{t=1}^T \mathbf{W}' \mathbf{u}_t \mathbf{f}_t' \|^2 \leq \sum_{k=1}^r \frac{1}{T^2 N^2} \sum_{s=1}^T \sum_{t=1}^T \mathbf{E} f_{k,t} f_{k,s} \mathbf{E} (\mathbf{u}_s \mathbf{W} \mathbf{W}' \mathbf{u}_t | \mathbf{F}) \]
\[
\leq \sum_{k=1}^r \frac{1}{T^2 N^2} \sum_{s=1}^T \sum_{t=1}^T \mathbf{E} f_{k,t} f_{k,s} \mathbf{tr} \mathbf{W}' \mathbf{E} (\mathbf{u}_s \mathbf{u}_s' | \mathbf{F}) \mathbf{W} \]
\[
\leq \frac{C}{T^2 N} \sum_{s=1}^T \sum_{t=1}^T \mathbf{E} \| f_t \| \| f_s \| \| \mathbf{E} (\mathbf{u}_t \mathbf{u}_s' | \mathbf{F}) \|
\leq \frac{C}{T N}.
\]

Similarly, $\mathbf{E} \| \frac{1}{T} \mathbf{G} \mathbf{E} \|^2 \leq O(\frac{1}{T N})$.

(iii) By the assumption that $\frac{1}{T^2 N^2} \sum_{t,s,T} \sum_{i,j,m,n \leq N} \mathbf{Cov}(u_{it} u_{jt}, u_{ms} u_{ns}) < C$,
\[
\mathbf{E} \| \frac{1}{T} (\mathbf{E} \mathbf{E}' - \mathbf{E} \mathbf{E}' | \mathbf{F}) \|^2 \leq \sum_{k,q \leq R} \mathbf{E} \| (\frac{1}{T^2 N^2} \sum_{t=1}^T \sum_{i,j \leq N} w_{k,i} w_{q,j} (u_{it} u_{jt} - \mathbf{E} u_{it} u_{jt}))^2 \|
\leq \frac{C}{T N^2} \sum_{t,s \leq T} \sum_{i,j,m,n \leq N} \mathbf{Cov}(u_{it} u_{jt}, u_{ms} u_{ns}) \leq \frac{C}{T N^2}.
\]

Next, by part (ii)
\[
\| \frac{1}{T} \mathbf{E} \mathbf{P}_G \mathbf{E} \| \leq \| \frac{1}{T} \mathbf{E} \mathbf{G} \|^2 \| (\frac{1}{T} \mathbf{G} \mathbf{E} | \mathbf{F})^{-1} \| \leq O_P(\frac{1}{T N}).
\]

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

\[\square\]
A.2 Proof of Theorem 2.1

Proof. We shall first show the convergence of $P_{\hat{F}M} - P_F$, and then the convergence of $P_{\hat{F}}P_F - P_F$.

First, from the SVD $H' = U_H(D_H,0)E_H'$, it is straightforward to verify that $M' = U_H(D_H^{-1},0)E_H'$. Then from Proposition A.1, $\lambda_{\min}(\frac{1}{T}M'\hat{F}'\hat{F}M) \geq c_0 N^{-1}\lambda_{\min}(D_H^{-2})$ with large probability. Hence $P_{\hat{F}M}$ is well defined.

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

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

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

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

The triangular inequality then implies $\|P_{\hat{F}M} - P_F\| \leq O_P(\frac{1}{\sqrt{N}}\nu_{\min}^{-1})$.

Finally, $P_{\hat{F}}P_{\hat{F}M} = P_{\hat{F}M}$ gives

$$\|P_{\hat{F}}P_F - P_F\| \leq \|P_{\hat{F}}(P_F - P_{\hat{F}M})\| + \|P_{\hat{F}M} - P_F\| \leq O_P(\frac{1}{\sqrt{N}}\nu_{\min}^{-1}).$$

\[\square\]

A.3 Proof of Theorem 3.1

Proof. Here we assume $R \geq r$. We let $z_t = (f_t'H',g_t')'$ and $\delta = (\alpha'H^+,\beta')'$. Then $\delta'z_t = y_{t+h|t}$. First, we have the following expansion

$${\hat{\delta}'\tilde{z}_T - \delta'z_T} = (\hat{\delta} - \delta)'\tilde{z}_T + \alpha'H^+(f_T - Hf_T).$$
Now \( \hat{\delta} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'Y \), where \( Y \) is the \((T-h) \times 1 \) vector of \( y_{t+h} \), and \( \hat{Z} \) is the \((T-h) \times \text{dim}(\delta) \) matrix of \( \hat{z}_t, t = 1, \cdots, T-h \). Also recall that \( e_t = \hat{f}_t - Hf_t = \frac{1}{\sqrt{N}}W'u_t \). Then

\[
\hat{z}_T'(\delta - \delta) = \hat{z}_T(\frac{1}{T}\hat{Z}'\hat{Z})^{-1} \sum_{d=1}^{4} a_d, \text{ where }
\]

\[
a_1 = \left( \frac{1}{T} \sum_t e_t e_t', 0 \right)', \quad a_2 = \frac{1}{T} \sum_t z_t e_t', \quad a_3 = \left( -\alpha'H + \frac{1}{T} \sum_t e_t e_t', 0 \right)', \quad a_4 = -\frac{1}{T} \sum_t z_t e_t'H'e' \alpha.
\]

On the other hand, let \( G \) be the \((T-h) \times \text{dim}(g_t) \) matrix of \( \{g_t : g \leq T-h\} \). We have, by the matrix block inverse formula, for the operator \( M_A := I - P_A \),

\[
\left( \frac{1}{T}\hat{Z}'\hat{Z} \right)^{-1} = \begin{pmatrix} A_1 & A_2 \\ A'_2 & A_3 \end{pmatrix}, \quad \text{where} \quad \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} (\frac{1}{T}\hat{F}'M_G\hat{F})^{-1} \\ -A_1\hat{F}'G(G'G)^{-1} \\ (\frac{1}{T}G'M_{\hat{F}}G)^{-1} \end{pmatrix}.
\]

Then \( \hat{z}_T'(\frac{1}{T}\hat{Z}'\hat{Z})^{-1} = (\hat{f}_T'A_1 + g_T'A'_2, \hat{f}_T'A_2 + g_T'A_3) \). This implies

\[
\hat{z}_T'(\delta - \delta) = (\hat{f}_T'A_1 + g_T'A'_2) \frac{1}{T} \sum_t [e_t e_t' - e_t e_t'H'e' \alpha]
\]

\[
+ (\hat{f}_T'A_1 H + g_T'A'_2 H) \frac{1}{T} \sum_t [f_t f_t' - f_t f_t'H'e' \alpha]
\]

\[
+ (\hat{f}_T'A_2 + g_T'A_3) \frac{1}{T} \sum_t [g_t g_t' - g_t g_t'H'e' \alpha].
\]

It is easy to show \( \frac{1}{T} \sum_t f_t e_t' \| + \frac{1}{T} \sum_t g_t e_t' \| = O_P(\frac{1}{\sqrt{T}}) \) and \( \frac{1}{T} \sum_t e_t e_t' = O_P(\frac{1}{\sqrt{TN}}) \). Also Lemma A.1 gives \( \frac{1}{T} \sum_t e_t e_t' = \frac{1}{T} F'E = O_P(\frac{1}{\sqrt{TN}}) \), \( \frac{1}{T} \sum_t f_t e_t = \frac{1}{T} F'E = O_P(\frac{1}{\sqrt{TN}}) \), and \( \frac{1}{T} \sum_t g_t e_t = \frac{1}{T} F'E = O_P(\frac{1}{\sqrt{TN}}) \). Together with Lemma A.2,

\[
\hat{z}_T'(\delta - \delta) = \|\hat{f}_T'A_1 + g_T'A'_2\|O_P(\frac{1}{\sqrt{TN}} + \frac{1}{N\nu_{\min}})
\]

\[
+ \|\hat{f}_T'A_1 H + g_T'A'_2 H\|O_P(\frac{1}{\sqrt{T}}) + \|\hat{f}_T'A_2 + g_T'A_3\|O_P(\frac{1}{\sqrt{T}})
\]

\[
= O_P(\frac{1}{\sqrt{T}} + \frac{1}{N\nu_{\min}}).
\]

Finally, as \( \|H'e'\| = O_P(\nu_{\min}^{-1}), \alpha'He'(\hat{f}_T - Hf_T) = O_P(\nu_{\min}^{-1})\|e_T\| = O_P(\nu_{\min}^{-1} N^{-1/2}). \)
Lemma A.2. For all $R \geq r$, (i) $\|A_1 \hat{f}_T\| + \|A_2\| = O_P(\sqrt{N})$, and $\|H' A_1 \hat{f}_T\| + \|H' A_2\| + \|A_3 \hat{f}_T\| + \|A_3\| = O_P(1)$.

Proof. First, by Proposition A.1, $\|A_1\| = O_P(N)$ and $\|A_1 H\| = O_P(\nu_{\min}^{-1} + \sqrt{\frac{N}{T}})$, and $\frac{1}{T} E'G = O_P(\frac{1}{\sqrt{NT}})$

\begin{align*}
A_1 \hat{f}_T &= \left(\frac{1}{T} \hat{F}' M_G \hat{F}\right)^{-1} e_T + \left(\frac{1}{T} \hat{F}' M_G \hat{F}\right)^{-1} Hf_T = O_P(\sqrt{N}) \\
H' A_1 \hat{f}_T &= H'\left(\frac{1}{T} \hat{F}' M_G \hat{F}\right)^{-1} e_T + H'\left(\frac{1}{T} \hat{F}' M_G \hat{F}\right)^{-1} Hf_T = O_P(1) \\
-A_2 &= A_1 \hat{F}'G(G'G)^{-1} = A_1 E'G(G'G)^{-1} - A_1 HF'G(G'G)^{-1} = O_P(\sqrt{\frac{N}{T}} + \nu_{\min}^{-1}) \\
-H' A_2 &= H'A_1 E'G(G'G)^{-1} - H'A_1 HF'G(G'G)^{-1} = O_P(1) \\
A_3 \hat{f}_T &= A_2 Hf_T + A_2 e_T = O_P(1).
\end{align*}

Finally, it follows from Proposition A.1 that $\frac{1}{T} G'(P_{\hat{F}} - P_{F'G})G = O_P(\frac{1}{T} + \frac{1}{N\nu_{\min}^2})$. Hence $\|A_3\| = O_P(1)$ since $\lambda_{\min}(\frac{1}{T} G'M_{FH}G) > c$.

□

A.4 Proof of Theorem 3.2

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

\begin{align*}
\hat{\varepsilon}_y &= M_{\hat{U}_J} M_{\hat{F}} Y, \\
\hat{\varepsilon}_g &= M_{\hat{U}_J} M_{\hat{F}} G.
\end{align*}

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

Proof. From Lemma A.7

\begin{align*}
\sqrt{T}(\hat{\beta} - \beta) &= \sqrt{T}[(\hat{\varepsilon}_y \hat{\varepsilon}_g)^{-1}\hat{\varepsilon}_g (\varepsilon_y - \varepsilon_g) + (\hat{\varepsilon}' \hat{\varepsilon}_g)^{-1}\hat{\varepsilon}_g \eta + (\hat{\varepsilon}' \hat{\varepsilon}_g)^{-1}\hat{\varepsilon}_g (\varepsilon_g - \hat{\varepsilon}_g) \beta] \\
&= O_P(1) \frac{1}{\sqrt{T}} \hat{\varepsilon}_g (\varepsilon_y - \varepsilon_g) + O_P(1) \frac{1}{\sqrt{T}} \hat{\varepsilon}_g (\varepsilon_g - \hat{\varepsilon}_g) + O_P(1) \frac{1}{\sqrt{T}} \eta' (\hat{\varepsilon}_g - \varepsilon_g) \\
&\quad + \frac{1}{T} \epsilon_g \epsilon_g^{-1} \frac{1}{\sqrt{T}} \epsilon_g \eta \\
&= \sigma_g^{-2} \frac{1}{\sqrt{T}} \hat{\varepsilon}_g \eta + o_P(1) \xrightarrow{d} \mathcal{N}(0, \sigma_g^{-4} \sigma_{\eta g}^2).
\end{align*}

(A.3)
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, $\hat{\sigma}_{g}^{-1} \hat{\sigma}_{g}^{2} \sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, 1)$, follows from $\hat{\sigma}_{g}^{2} := \frac{1}{T} \hat{\epsilon}_{g}^{\prime} \hat{\epsilon}_{g} \xrightarrow{P} \sigma_{g}^{2}$.

\[\square\]

**Proposition A.2.** Suppose $T = O(\nu_{\min}^2 N^2 \log N)$, $|J|_0^2 T = O(\nu_{\min}^2 N^2 \log N)$, $|J|_0^2 = O(N \nu_{\min}^2 \log N)$ and $|J|_0^2 \log N = O(T)$, $|J|_0^2 = o(N)$ For all $R \geq r$,

\[(i) \frac{1}{T} \|\hat{U}' \theta - \tilde{U}' \tilde{\theta}\|^2 = O_P(\frac{|J|_0 \log N}{T})\] and $\|\tilde{\theta} - \theta\|_1 = O_P(\|J\|_0 \sqrt{\frac{\log N}{T}})$.

\[(ii) |\hat{J}|_0 = O_P(|J|_0).\]

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

$$d_t = \alpha'_g \hat{f}_t - \hat{\alpha}_g' \hat{f}_t + (u_t - \hat{u}_t)' \theta, \quad \Delta = \theta - \hat{\theta}.$$ Then $g_t = \alpha'_g \hat{f}_t + \theta' \hat{u}_t + \varepsilon_{g,t}$, and $L(\theta) \leq L(\hat{\theta})$ imply

$$\frac{1}{T} \sum_{t=1}^{T} [(\hat{u}'_t \Delta)^2 + 2(\varepsilon_{g,t} + d_t)\hat{u}'_t \Delta] + \tau \|\tilde{\theta}\|_1 \leq \tau \|\theta\|_1.$$ It follows from Lemma A.5 that $\frac{1}{T} \|\hat{U} \varepsilon_g\|_\infty \leq O_P(\sqrt{\frac{\log N}{T}})$. Also Lemma A.4 implies that

$$\frac{1}{T} \sum_{t=1}^{T} d_t \hat{u}_t \|_\infty \leq \frac{1}{T} \|\hat{U} H' \alpha\|_\infty + \frac{1}{T} \|\hat{U} E(H' \alpha - \hat{\alpha})\|_\infty + \frac{1}{T} \|\hat{U} F H'(H' \alpha - \hat{\alpha})\|_\infty$$

$$+ \frac{1}{T} \|\theta' (\hat{U} - U) \hat{U}'\|_\infty$$

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

Thus the “score” satisfies $\frac{1}{T} \sum_{t=1}^{T} (\varepsilon_{g,t} + d_t) \hat{u}'_t \|_\infty \leq \tau/2$ 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,

$$\frac{1}{T} \sum_{t=1}^{T} (\hat{u}'_t \Delta)^2 + \frac{\tau}{2} \|\Delta_{J^c}\|_1 \leq \frac{3\tau}{2} \|\Delta_J\|_1.$$
Meanwhile, by the restricted eigenvalue condition and Lemma A.4,

\[
\frac{1}{T} \sum_{t=1}^{T} (\hat{u}_t' \Delta)^2 \geq \frac{1}{T} \sum_{t=1}^{T} (u_t' \Delta)^2 - \|\Delta\|^2 \| \frac{1}{T} \hat{U} \hat{U}' - UU' \|_{\infty} \geq \|\Delta\|^2 (\phi_{\min} - o_P(1))
\]

where the last inequality follows from \(|J|_0 O_P(\nu_{\min}^{-2} \frac{1}{N} + \log \frac{N}{T}) = o_P(1)\) (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 \(|\hat{J}|_0 = O_P(\|J\|_0)\) 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) \(\| \frac{1}{T} E' U' \|_{\infty} = O_P(\sqrt{\frac{\log N}{TN}} + \frac{1}{N})\)

(ii) \(\frac{1}{T} E' \hat{P} F E = O_P(\frac{1}{N})\), \(\frac{1}{T} E' \hat{P} U' \|_{\infty} = O_P(\sqrt{\frac{\log N}{TN}} + \frac{1}{N})\),

(iii) \(\frac{1}{T} (\hat{U} - U)(\hat{U} - U)' \|_{\infty} + 2 \frac{1}{T} (\hat{U} - U)' \|_{\infty} = O_P(\nu_{\min}^{-2} \frac{1}{N} + \log \frac{N}{T})\).

(iv) \(\| \frac{1}{T} \hat{U} \hat{U}' - \frac{1}{T} UU' \|_{\infty} = O_P(\nu_{\min}^{-2} \frac{1}{N} + \log \frac{N}{T})\).

Proof. Let \(\hat{F} = (\hat{f}_1, \ldots, \hat{f}_T)'\). In addition, \(\hat{B} - BH^+ = -BH^+ E' \hat{F} (\hat{F}' \hat{F})^{-1} + UE(\hat{F}' \hat{F})^{-1} + UFH' (\hat{F}' \hat{F})^{-1}\). Therefore,

\[
U - \hat{U} = \hat{B} \hat{F}' - BF' = (\hat{B} - BH^+) \hat{F}' + BH^+ E' = -BH^+ E' \hat{F}' (\hat{F}' \hat{F})^{-1} \hat{F}' + UE(\hat{F}' \hat{F})^{-1} \hat{F}' + UFH' (\hat{F}' \hat{F})^{-1} \hat{F}' + BH^+ E'. \tag{A.4}
\]

(i) We have

\[
\| \frac{1}{T} UE \|_{\infty} \leq \sum_{k \leq r} \max_{i \leq N} \| \frac{1}{T} \sum_{t=1}^{T} (u_{it} u_{jt} - E u_{it} u_{jt}) w_{k,j} \| + O(\frac{1}{N}) = O_P(\sqrt{\frac{\log N}{TN}} + \frac{1}{N})
\]

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

\[
\| \frac{1}{T} E' \hat{P} E \| \leq \| \frac{1}{T} E' E (\hat{F}' \hat{F})^{-1} E' E \| + \| \frac{2}{T} E' E (\hat{F}' \hat{F})^{-1} HF' E \| + \| \frac{1}{T} E' FH' (\hat{F}' \hat{F})^{-1} HF' E \|
\leq O_P(\frac{1}{N})
\]

\[
\| \frac{1}{T} E' \hat{P} U' \|_{\infty} \leq \| \frac{1}{T} E' E (\hat{F}' \hat{F})^{-1} E' U' \|_{\infty} + \| \frac{1}{T} E' E (\hat{F}' \hat{F})^{-1} HF' U' \|_{\infty}
+ \| \frac{1}{T} E' FH' (\hat{F}' \hat{F})^{-1} E' U' \|_{\infty} + \| \frac{1}{T} E' FH' (\hat{F}' \hat{F})^{-1} HF' U' \|_{\infty}
\]

13
\[ \leq O_P\left(\sqrt{\frac{\log N}{TN}} + \frac{1}{N}\right). \]

(iii) We have \(\|H^+\| = O(\nu_{\min}^{-1}).\) Also, \(\|\hat{F}^T\hat{F}\|^{-1} \leq 1.\) In addition, by Lemma A.1, \(\|\hat{F}^T\hat{F}\|^2 = \|\hat{F}^T\hat{F}\|^{-1} \leq O_P(\frac{N}{T})\) and that \(\|H^T\hat{F}\|^{-1} \leq O_P(\frac{1}{T}).\) Next, by Lemma A.1, \(\|E\| = O_P(\sqrt{\frac{T}{N}}),\) and \(\max_i \|b_i\| < C.\) Substitute the expansion (A.4), and by Proposition A.1,

\[
\begin{align*}
\|\frac{1}{T}(\hat{U} - U)(\hat{U} - U)'\|_\infty + 2\|\frac{1}{T}(\hat{U} - U)'\|_\infty & \leq \frac{1}{2}B\hat{H}E'U'\|_\infty + \frac{1}{T}B\hat{H}E'\hat{E}H'B'\|_\infty + \frac{3}{T}UE(\hat{F}^T)^{-1}E'U'\|_\infty \\
& \leq \frac{C}{T}E'U'\|_\infty O_P(\nu_{\min}^{-1}) + \frac{C}{T}E'\|_\infty O_P(\nu_{\min}^{-1}) + \frac{C}{T}E'\|_\infty O_P(\nu_{\min}^{-1}) + \frac{C}{T}E'\|_\infty O_P(\nu_{\min}^{-1}) + \frac{C}{T}E'\|_\infty O_P(\nu_{\min}^{-1}) + \frac{C}{T}E'\|_\infty O_P(\nu_{\min}^{-1}) + \frac{C}{T}E'\|_\infty O_P(\nu_{\min}^{-1}) + \frac{C}{T}E'\|_\infty O_P(\nu_{\min}^{-1})
\end{align*}
\]

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

Lemma A.4. For all \(R \geq r,\)

(i) \(\|\frac{1}{T}\theta(\hat{U} - U)\|_\infty \leq O_P(\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2})\|J\|_0.\)

(ii) \(\|\frac{1}{T}\hat{F}P'F\| = O_P(\frac{1}{N\nu_{\min}} + \frac{1}{\sqrt{NT}}), \|\frac{1}{T}UP'F\|_\infty = O_P(\sqrt{\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2}}).\)

(iii) \(\|\frac{1}{T}\theta UE\|_\infty \leq O_P(\sqrt{\frac{\log N}{T} + \frac{1}{N\nu_{\min}^2}}), \|\frac{1}{T}\theta UF\|^2 \leq O_P(\frac{1}{N\nu_{\min}^2}).\)

(iv) \(\|\frac{1}{T}\theta UE\| = \|J\|_0 O_P(\frac{1}{N} + \frac{1}{\sqrt{NT}}), \|\frac{1}{T}\theta UF\| = O_P(\frac{|J|_0}{T}).\)

(v) \(\hat{\alpha}_g - H^\top \alpha_g = |J|_0 O_P(1 + \frac{\sqrt{N}}{T}) + O_P(\nu_{\min}^{-1}), H^\top(\hat{\alpha}_g - H^\top \alpha_g) = O_P(\nu_{\min}^{-1} |J|_0 T + \sqrt{\frac{|J|_0}{T}} + \nu_{\min}^{-1} N).\)
Proof. (i) By Lemma A.3 \( \| \frac{1}{T} \theta' (\hat{\mathbf{U}} - \mathbf{U}) \hat{\mathbf{U}}' \|_{\infty} \leq \| \theta' \|_1 \| \frac{1}{T} (\hat{\mathbf{U}} - \mathbf{U}) \hat{\mathbf{U}}' \|_{\infty} \leq O_P \left( \frac{\log N}{T} + \frac{1}{N \nu_{\min}} \right) |J|_0. \)

(ii) Note \( \mathbf{H}' \mathbf{H}'^+ = \mathbf{I}, \) Lemma A.3 shows \( \| \frac{1}{T} \mathbf{E}' \mathbf{P}_f \mathbf{E} \| = O_P \left( \frac{1}{N} \right), \) \( \| \frac{1}{T} \mathbf{E}' \mathbf{P}_f \mathbf{U}' \|_{\infty} = O_P \left( \frac{\sqrt{\log N}}{TN} + \frac{1}{N} \right). \)

\[
\begin{align*}
\| \frac{1}{T} \mathbf{E}' \mathbf{P}_f \mathbf{F} \| & \leq \| \frac{1}{T} \mathbf{E}' \mathbf{P}_f \mathbf{E} \mathbf{H}'^+ \| + \| \frac{1}{T} \mathbf{E}' \mathbf{E} \mathbf{H}'^+ \| + \| \frac{1}{T} \mathbf{E}' \mathbf{F} \| = O_P \left( \frac{1}{N \nu_{\min}} + \frac{1}{\sqrt{NT}} \right) \\
\| \frac{1}{T} \mathbf{U} \mathbf{P}_f \mathbf{F} \|_{\infty} & \leq \| \frac{1}{T} \mathbf{U} \mathbf{P}_f \mathbf{E} \mathbf{H}'^+ \|_{\infty} + \| \frac{1}{T} \mathbf{U} \mathbf{E} \mathbf{H}'^+ \|_{\infty} + \| \frac{1}{T} \mathbf{U} \mathbf{F} \|_{\infty} \\
& \leq O_P \left( \sqrt{\frac{\log N}{T}} + \frac{1}{N \nu_{\min}} \right).
\end{align*}
\]

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

\[
\begin{align*}
\| \frac{1}{T} \hat{\mathbf{E}} \|_{\infty} & \leq \| \frac{1}{T} \mathbf{E} \|_{\infty} + \| \frac{1}{T} (\hat{\mathbf{U}} - \mathbf{U}) \mathbf{E} \|_{\infty} \\
& \leq \| \frac{1}{T} \mathbf{E} \|_{\infty} + \| \frac{1}{T} \mathbf{B} \hat{\mathbf{H}}^+ \mathbf{E} \mathbf{P}_f \mathbf{E} \|_{\infty} + \| \frac{1}{T} \mathbf{U} \mathbf{P}_f \mathbf{E} \|_{\infty} + \| \frac{1}{T} \mathbf{B} \hat{\mathbf{H}}^+ \mathbf{E} \|_{\infty} \\
& \leq O_P \left( \sqrt{\frac{\log N}{TN}} + \frac{1}{N \nu_{\min}} \right) \\
\| \frac{1}{T} \hat{\mathbf{F}} \|_{\infty} & \leq \| \frac{1}{T} \mathbf{F} \|_{\infty} + \| \frac{1}{T} (\hat{\mathbf{U}} - \mathbf{U}) \mathbf{F} \|_{\infty} \\
& \leq \| \frac{1}{T} \mathbf{F} \|_{\infty} + \| \frac{1}{T} \mathbf{B} \hat{\mathbf{H}}^+ \mathbf{E} \mathbf{P}_f \mathbf{F} \|_{\infty} + \| \frac{1}{T} \mathbf{U} \mathbf{P}_f \mathbf{F} \|_{\infty} + \| \frac{1}{T} \mathbf{B} \hat{\mathbf{H}}^+ \mathbf{E} \mathbf{F} \|_{\infty} \\
& \leq O_P \left( \sqrt{\frac{\log N}{T}} + \frac{1}{N \nu_{\min}^2} \right).
\end{align*}
\]

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

\[
\begin{align*}
\mathbb{E} \left\| \frac{1}{NT} \theta' (\mathbf{U} \mathbf{U}' - \mathbf{E} \mathbf{U} \mathbf{U}') \mathbf{W} \right\|^2 & = \mathbb{E} \left( \sum_{k=1}^{R} \sum_{t=1}^{T} \mathbf{P}_k \mathbf{E} \mathbf{F} \right) \\
& \leq \frac{C}{N^2 T^2} \| \theta \|_1^2 \max_{j \leq N} \sum_{q, u \leq N} \sum_{t, s \leq T} \mid \mathbb{C} \mathbb{O} \left( u_{it} u_{qt}, u_{js} u_{us} \right) \mid \leq \frac{C |J|_0^2}{NT}.
\end{align*}
\]

Also, \( \| \frac{1}{NT} \theta' \mathbf{E} \mathbf{U} \mathbf{U}' \mathbf{W} \| \leq \max_{j \leq N} \sum_{k} |w_{k,j}| \| \theta' \|_1 \| \frac{1}{T} \mathbf{E} \mathbf{U} \mathbf{U}' \|_{1} \leq O \left( \frac{|J|_0}{NT} \right). \)

\[
\begin{align*}
\mathbb{E} \| \frac{1}{T} \theta' \mathbf{U} \mathbf{F} \|_{\infty}^2 & = \frac{1}{T^2} \text{tr} \mathbf{F} \mathbf{E} (\mathbf{U}' \theta' \mathbf{U} \mathbf{F}) \mathbf{F} \leq \frac{C}{T} \| \mathbf{E} (\mathbf{U}' \theta' \mathbf{U} \mathbf{F}) \mathbf{F} \|_{1} \\
& \leq \frac{C}{T} \max_{s=1}^{T} \sum_{t} \| \mathbb{E} (\theta' u_{it} u_{it}', \theta' \mathbf{F}) \|_{1} \| \theta' \|_{\infty} \leq \frac{C |J|_0}{T}.
\end{align*}
\]
(v) Since $\hat{\alpha}_g = (\hat{F}'\hat{F})^{-1}\hat{F}'G$, simple calculations using Proposition A.1 yield

$$\hat{\alpha}_g - H^{+}\alpha_g = (∩\hat{F}'\hat{F})^{-1}\hat{F}'G - H^{+}\alpha_g$$

$$= (∩\hat{F}'\hat{F})^{-1}E'\varepsilon_g - (∩\hat{F}'\hat{F})^{-1}E'EH^{+}\alpha_g + (∩\hat{F}'\hat{F})^{-1}E'U'\theta + O_P\left(\sqrt{\frac{|J|_0}{T}}\right)$$

$$= |J|_0O_P(1 + \sqrt{\frac{N}{T}}) + O_P(\nu_{\min}^{-1})$$

$$H'(\hat{\alpha}_g - H^{+}\alpha_g) = H'(∩\hat{F}'\hat{F})^{-1}E'\varepsilon_g - H'(∩\hat{F}'\hat{F})^{-1}E'EH^{+}\alpha_g + H'(∩\hat{F}'\hat{F})^{-1}E'U'\theta + O_P\left(\sqrt{\frac{|J|_0}{T}}\right)$$

$$= O_P(\nu_{\min}^{-1} |J|_0 + \sqrt{\frac{|J|_0}{T}} + \nu_{\min}^{-2} \frac{1}{N}).$$

Lemma A.5. Suppose $|J|_0 = o(N\nu_{\min}^2)$. For any $r \geq R$

(i) $\frac{1}{T}\|P_{\hat{F}}U'\theta\|^2 = O_P\left(\frac{|J|_0^2}{N} + \frac{|J|_0^2}{\nu_{\min}N}\right)$, $\frac{1}{T}\|P_{\hat{F}}\varepsilon_g\|^2 = O_P\left(\frac{1}{T}\right)$,

(ii) $\frac{1}{T}\|U'\varepsilon_g\|_\infty = O_P\left(\frac{1}{T}\right)$, and $\|1/\sqrt{T}\Upsilon\|_\infty = O_P\left(\frac{1}{T}\right)$

(iii) $\lambda_{\min}(\frac{1}{3}\Upsilon_j\Upsilon_j') > c_0$ with probability approaching one. $\frac{1}{T}\|P_{\Upsilon_j}\varepsilon_g\|^2 = O_P\left(\frac{1}{T}\right)$

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

$$\frac{1}{T}\|P_{\hat{F}}U'\theta\|^2 = \frac{1}{T}\theta'UE(∩\hat{F}')^{-1}E'U'\theta + \frac{2}{T}\theta'UE(∩\hat{F}')^{-1}HF'U'\theta - \frac{1}{T}\theta'U\hat{F}'(∩\hat{F}')^{-1}HF'U'\theta$$

$$\leq O_P\left(\frac{|J|_0^2}{N} + \frac{|J|_0^2}{\nu_{\min}N}\right).$$

$$\frac{1}{T}\|P_{\hat{F}}\varepsilon_g\|^2 = \frac{1}{T}\varepsilon_g'UE(∩\hat{F}')^{-1}E'\varepsilon_g + \frac{2}{T}\varepsilon_g'UE(∩\hat{F}')^{-1}HF'\varepsilon_g + \frac{1}{T}\varepsilon_g'FH(∩\hat{F}')^{-1}HF'\varepsilon_g$$

$$\leq O_P\left(\frac{N}{NT}\right) + O_P\left(\frac{1}{T}\right) + O_P\left(\frac{1}{T}\right).$$

(ii) By (A.4)

$$\frac{1}{T}(U - \hat{U})\varepsilon_g = -\frac{1}{T}B\hat{F}'E'U\hat{F}'^{-1}E'\varepsilon_g - \frac{1}{T}B\hat{F}'E'F'H(∩\hat{F}')^{-1}E'\varepsilon_g + \frac{1}{T}UE(∩\hat{F}')^{-1}E'\varepsilon_g$$

$$- \frac{1}{T}B\hat{F}'E'F'H(∩\hat{F}')^{-1}HF'\varepsilon_g + \frac{1}{T}B\hat{F}'E'F'H(∩\hat{F}')^{-1}HF'\varepsilon_g + \frac{1}{T}UE(∩\hat{F}')^{-1}HF'\varepsilon_g$$

16
\[ \frac{1}{T} \text{U} \text{F} \text{H}' (\hat{\text{F}}' \hat{\text{F}})^{-1} \text{E}' \varepsilon_g + \frac{1}{T} \text{U} \text{F} \text{H}' (\hat{\text{F}}' \hat{\text{F}})^{-1} \text{H} \text{F}' \varepsilon_g + \frac{1}{T} \text{B} \text{H}^+ \text{E}' \varepsilon_g. \]

So by Lemmas A.1 and \( \| \frac{1}{T} \text{UE} \|_\infty = O_P(\sqrt{\frac{\log N}{T^2}} + \frac{1}{N}) \), \( \| \frac{1}{T} (\hat{\text{U}} - \text{U}) \varepsilon_g \|_\infty = O_P(\sqrt{\frac{\log N}{T}}) \).

Also, with \( \| \frac{1}{T} \text{UE}_g \|_\infty = O_P(\sqrt{\frac{\log N}{T}}) \) we have \( \| \frac{1}{T} \hat{\text{U}} \varepsilon_g \|_\infty = O_P(\sqrt{\frac{\log N}{T}}) \). The proof for \( \| \frac{1}{T} \hat{\text{U}} \varepsilon_g \|_\infty \) is the same.

(iii) First, it follows from Lemma A.4 that \( \| \frac{1}{T} \hat{\text{U}} \hat{\text{U}}' - \frac{1}{T} \text{U}
\text{U}' \|_\infty \leq O_P(\log \frac{N}{T} + \frac{\nu^{-2}}{N}). \)

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

\[ \lambda_{\min} (\frac{1}{T} \hat{\text{U}}_j \hat{\text{U}}_{j}' ) \geq \lambda_{\min} (\frac{1}{T} \text{U}_j \text{U}_{j}' ) - \| \frac{1}{T} \hat{\text{U}} \hat{\text{U}}' - \frac{1}{T} \text{U}
\text{U}' \|_\infty |\hat{J}|_0 \geq \phi_{\min} - O_P\left( \frac{\log N}{T} + \frac{\nu^{-2}}{N} \right). \]

\[ \frac{1}{T} \| \text{P} \hat{\text{U}}_j \varepsilon_g \|_2^2 = \frac{1}{T} \| \varepsilon_g \hat{\text{U}}_j \hat{\text{U}}_{j}' - \hat{\text{U}} \hat{\text{U}}' \varepsilon_g \|_2^2 \leq \| \varepsilon_g \hat{\text{U}}_j \hat{\text{U}}_{j}' \|_2^2 \lambda_{\min} (\frac{1}{T} \hat{\text{U}} \hat{\text{U}}') \leq c \| \varepsilon_g \hat{\text{U}}' \|_\infty^2 |\hat{J}|_0 \leq O_P\left( \frac{|J|_0 \log N}{T} \right). \]

\( \frac{1}{T} \| \text{P} \hat{\text{U}}_j \varepsilon_g \|_2^2 \) follows from the same proof.

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

\[ \frac{1}{T} \| \theta' (\hat{\text{U}} - \text{U}) \|_2^2 \leq \frac{1}{T} \| \theta' \text{B} \text{H}^+ \text{E}' \text{P} \|_2^2 + \frac{1}{T} \| \theta' \text{U} \text{P} \|_2^2 + \frac{1}{T} \| \theta' \text{B} \text{H}^+ \text{E}' \|_2^2 \leq O_P\left( \frac{|J|_0^2 + \nu^{-2}}{N} + \frac{|J|_0^2}{N \sqrt{T}} \right). \]

\[ \| \frac{1}{T} \theta' \text{P} \varepsilon_y \|_2 \leq \| \frac{1}{T} \theta' \text{P} \|_2 \| \text{P} \varepsilon_y \| = O_P\left( \frac{1}{\sqrt{NT}} \right). \]

\[ \frac{1}{T} \theta' \text{U} \text{P} \varepsilon_y \leq \frac{1}{T} \| \theta' \text{U} \text{P} \|_2 \| \varepsilon_y \| = O_P\left( \frac{|J|_0}{T} + \frac{|J|_0}{\sqrt{NT}} + \frac{\nu^{-1/2} |J|_0^3/4}{\sqrt{NT^{3/4}}} \right). \]

\[ \frac{1}{T} \| \text{M} \hat{\text{U}}_j \theta' \|_2^2 = O_P\left( |J|_0 \frac{\log N}{T} \right), \frac{1}{T} \| \text{M} \hat{\text{U}}_j \theta' \|_2^2 = O_P\left( |J|_0 \frac{\log N}{T} + \frac{|J|_0^3 + \nu^{-2}}{N \nu_{\min}} + \frac{|J|_0^2}{T} \right). \]

\[ \frac{1}{T} \varepsilon_y' \text{P} \hat{\text{U}}_j (\hat{\text{U}} - \text{U})' \theta = |J|_0 \frac{\log N}{T} O_P\left( \frac{\log N}{T} + \frac{1}{N \nu_{\min}} \right), \]

\[ \frac{1}{T} \varepsilon_y' \text{M} \hat{\text{U}}_j \text{U}' \theta \leq O_P\left( \frac{|J|_0 \log N}{T} + \frac{|J|_0 + \nu^{-1}}{\sqrt{NT}} + \frac{\nu^{-1/2} |J|_0^3/4}{\sqrt{NT^{3/4}}} + \frac{\log N}{T} \frac{|J|_0^3}{N \nu_{\min}} + \frac{\nu^{-1/2} |J|_0^3/4}{\sqrt{NT^{3/4}}} \right), \]

\[ \| \text{P} \hat{\text{U}}_j \| = O_P\left( \sqrt{\frac{|J|_0 \log N}{N}} + \frac{\sqrt{T} |J|_0}{N \nu_{\min}} \right), \frac{1}{T} \varepsilon_y' \text{P} \hat{\text{U}}_j \text{E} = O_P\left( \frac{|J|_0 \log N}{T \sqrt{N}} + \frac{|J|_0 \log N}{N \nu_{\min} \sqrt{T}} \right). \]
Proof. (i) First note that \( \mathbf{P}_{\tilde{U}_j} \hat{\mathbf{U}}' \theta = \hat{\mathbf{U}}' \tilde{\mathbf{m}} \), where

\[
\tilde{\mathbf{m}} = (\tilde{m}_1, \ldots, \tilde{m}_N)' = \arg \min_{\mathbf{m}} \| \hat{\mathbf{U}}'(\theta - \mathbf{m}) \| : \ m_j = 0, \ \text{for} \ j \notin \tilde{J}.
\]

Thus by the definition of \( \tilde{\mathbf{m}} \), Proposition A.2 and Lemma A.5,

\[
\begin{align*}
\frac{1}{T} \| \mathbf{M}_{\tilde{U}_j} \hat{\mathbf{U}}' \theta \|^2 & = \frac{1}{T} \| \hat{\mathbf{U}}' \theta - \hat{\mathbf{U}}' \tilde{\mathbf{m}} \|^2 \leq \frac{1}{T} \| \hat{\mathbf{U}}' \theta - \hat{\mathbf{U}}' \hat{\theta} \|^2 \leq O_P(\frac{|J_0|}{T} \log N) \\
\frac{1}{T} \| \mathbf{M}_{\tilde{U}_j} \mathbf{U}' \theta \|^2 & \leq O_P(\frac{|J_0| \log N}{T} \log \log \log N \log T(N))^2 + \frac{1}{T} \| \mathbf{(U - U')}\theta \|^2 = O_P(\frac{|J_0| \log N + |J_0|^2}{N})
\end{align*}
\]

where we used \( \frac{\nu_{\min}^{-1}|J_0|}{N \log T} = O_P(\frac{|J_0| \log N}{T}) \) by our assumption.

(ii) Let \( \Delta = \theta - \tilde{\mathbf{m}} \). Then \( \dim(\Delta) = O_P(\frac{|J_0|}{T}) \). Also, by Lemma A.4,

\[
\frac{1}{T} \| \hat{\mathbf{U}}' \Delta \| \leq 1 \| \mathbf{U}' \Delta \| \leq \| \hat{\mathbf{U}}'(\theta - \tilde{\theta}) \| \text{ and Proposition A.2 implies}
\]

\[
\| \theta - \tilde{\mathbf{m}} \|_1 \leq \| J_0 \| \Delta \|^2 \leq J_0 \| \mathbf{U}' \Delta \|^2 \leq J_0 \| \hat{\mathbf{U}}' \Delta \|^2 + O_P(\frac{\log N}{T} + \frac{1}{N \nu_{\min}^2}) \| \Delta \|^2 |J_0|
\]

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

\[
\| \theta - \tilde{\mathbf{m}} \|_1^2 \leq O_P(\frac{|J_0| \log N}{T}). \text{ Hence by Lemma A.5,}
\]

\[
\frac{1}{T} \varepsilon_y' \mathbf{P}_{\tilde{U}_j} \hat{\mathbf{U}}' \theta = \| \frac{1}{\sqrt{T}} \varepsilon_y' \mathbf{P}_{\tilde{U}_j} \| \| \hat{\mathbf{U}}(\mathbf{U} - \mathbf{U}')\theta \|_{\infty} \frac{\sqrt{|J_0|}}{T} \nu_{\min}^{-1/2}(\frac{1}{T} \mathbf{U}_j \hat{\mathbf{U}}_j')
\]

\[
\leq \frac{1}{T} \varepsilon_y' \mathbf{M}_{\tilde{U}_j} \hat{\mathbf{U}}' \theta = \frac{1}{T} \varepsilon_y' \hat{\mathbf{U}}'(\theta - \tilde{\mathbf{m}}) \leq \frac{1}{T} \varepsilon_y' \hat{\mathbf{U}}'\| \mathbf{U}' \Delta \| \| \theta - \tilde{\mathbf{m}} \|_1 \leq O_P(\frac{|J_0| \log N}{T}).
\]

\[
\frac{1}{T} \varepsilon_y' \mathbf{M}_{\tilde{U}_j} \hat{\mathbf{U}}' \theta = \frac{1}{T} \varepsilon_y' \hat{\mathbf{U}}'(\theta - \tilde{\mathbf{m}}) \mathbf{P}_{\tilde{U}_j} \| \hat{\mathbf{U}}'(\theta - \tilde{\mathbf{m}}) \| \| \mathbf{U}' \Delta \| \| \theta - \tilde{\mathbf{m}} \|_1 \leq O_P(\frac{|J_0| \log N}{T} + \frac{1}{T} \theta' \mathbf{B} \mathbf{H}^+ \mathbf{E}' \mathbf{P}_F \varepsilon_y + \frac{1}{T} \theta' \mathbf{U} \mathbf{P}_F \varepsilon_y + \frac{1}{T} \theta' \mathbf{B} \mathbf{H}^+ \mathbf{E}' \varepsilon_y)
\]

\[
\leq O_P(\frac{|J_0| \log N}{T}) + \frac{1}{T} \theta' \mathbf{B} \mathbf{H}^+ \mathbf{E}' \mathbf{P}_F \varepsilon_y + \frac{1}{T} \theta' \mathbf{U} \mathbf{P}_F \varepsilon_y + \frac{1}{T} \theta' \mathbf{B} \mathbf{H}^+ \mathbf{E}' \varepsilon_y
\]

\[
- \frac{1}{T} \varepsilon_y' \mathbf{P}_{\tilde{U}_j} \hat{\mathbf{U}}'(\mathbf{U} - \mathbf{U}')\theta
\]

18
\[ \leq O_P \left( \frac{|J_0| \log N}{T} + \frac{|J_0| + \nu_\min^1}{\sqrt{NT}} + \frac{\nu_\min^{-1/2} |J_0|^{3/4}}{\sqrt{NT^{3/4}}} + \sqrt{\frac{\log N}{T}} \frac{|J_0|^2}{N \nu_\min^2} \right). \]

(iii) By Lemma A.4,
\[
\|P_{\hat{U}} E\| \leq \frac{1}{T} \epsilon_y' \left( \frac{1}{T} \hat{U} \hat{U}' \right)^{-1} \frac{1}{T} \|\hat{U} E\| \leq O_P \left( \frac{|J_0| \log N}{T \sqrt{N}} + \frac{|J_0|^3}{T \sqrt{N}} \right).
\]

\[
\|\frac{1}{T} \epsilon_y' P_{\hat{U}} E\| \leq O_P \left( \frac{|J_0| \log N}{T \sqrt{N}} + \frac{|J_0|^3}{T \sqrt{N}} \right).
\]

Lemma A.7. For any \( R \geq r \),
\[
\frac{1}{T} \|\epsilon_y - \epsilon_g\|^2 = O_P \left( \frac{|J_0| + |J_0| \log N}{T} + \frac{|J_0|^2 + \nu_\min^2}{\sqrt{NT}} + \frac{|J_0|^{3/2}}{\sqrt{NT^{3/4}}} \right).
\]

(i) The same proof applies to other terms as well.

Proof. Note that \( \hat{\epsilon}_g = M_{\hat{U}} \hat{\theta} \) and \( \hat{\theta} = \hat{U} \hat{\theta} \). Also, \( \hat{U} = X \hat{M}_F \) implies \( P_{\hat{U}} P_F = 0 \), and \( M_{\hat{U}} M_F = M_F - P_{\hat{U}} \).

Recall that \( H^+ H = I \) and \( \hat{F} = FH' + E \), hence straightforward calculations yield
\[
\hat{\epsilon}_g - \epsilon_g = M_{\hat{U}} U' \theta - P_F U' \theta + M_{\hat{U}} M_F \alpha_g - P_{\hat{U}} \epsilon_g - P_F \epsilon_g
\]
\[
= M_{\hat{U}} U' \theta - P_F U' \theta - P_{\hat{U}} \epsilon_g - P_F \epsilon_g - (I - P_F - P_{\hat{U}}) E H^+ \alpha_g. \quad (A.5)
\]

It follows from Lemmas A.5, A.6 that \( \frac{1}{T} \|\hat{\epsilon}_g - \epsilon_g\|^2 = O_P \left( \frac{|J_0|^2 + |J_0| \log N}{T} + \frac{|J_0|^2 + \nu_\min^2}{\sqrt{NT}} + \frac{|J_0|^{3/2}}{\sqrt{NT^{3/4}}} \right).
\]

The proof for \( \frac{1}{T} \|\hat{\epsilon}_g - \epsilon_g\|^2 \) follows similarly.

(ii) It follows from (A.5) and Lemmas A.5 A.6 that
\[
\frac{1}{T} \epsilon_y' (\hat{\epsilon}_g - \epsilon_g) = \frac{1}{T} \epsilon_y' U' \theta - \frac{1}{T} \epsilon_y' P_F U' \theta - \frac{1}{T} \epsilon_y' P_{\hat{U}} \epsilon_g - \frac{1}{T} \epsilon_y' P_F \epsilon_g
\]
\[
- \frac{1}{T} \epsilon_y' E H^+ \alpha_g - \frac{1}{T} \epsilon_y' P_F E H^+ \alpha_g - \frac{1}{T} \epsilon_y' P_{\hat{U}} E H^+ \alpha_g
\]
\[
\leq O_P \left( \frac{|J_0| \log N}{T} + \frac{|J_0| + \nu_\min^1}{\sqrt{NT}} + \frac{\nu_\min^{-1/2} |J_0|^{3/4}}{\sqrt{NT^{3/4}}} + \sqrt{\frac{\log N}{T}} \frac{|J_0|^2}{N \nu_\min^2} \right).
\]

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 \(|J|_0^2 = o(\min\{T, N\})\), \(|J|_0 \log N = o(T)\).

\[
\frac{1}{T} \sum_{i,j} \frac{1}{T} \sum_{t} (\hat{u}_{it} \hat{y}_{jt} - u_{it} u_{jt}) \right. \\
= \frac{1}{T} \sum_{i} \frac{1}{T} \sum_{t} \hat{\alpha}_y \hat{f}_{it} \hat{u}_{it} \\
= \frac{1}{T} \sum_{i} \frac{1}{T} \sum_{t} \hat{u}_{it} (u_{it} - \hat{u}_t)^\prime \theta \\
= \frac{1}{T} \sum_{i} \frac{1}{T} \sum_{t} \hat{u}_{it} \varepsilon_{g,t} \\
= O_P(\frac{|I|_0^2}{N} + \frac{|I|_0 \log N}{T}) \\
\]

As for the residual, note that \(\hat{\varepsilon}_g = M_{U_j} M_E G\) and \(G = U' \theta + \varepsilon_g\). Then

\[
\hat{\varepsilon}_g - \varepsilon_g = M_{U_j} U' \theta - P_E U' \theta - P_{\hat{U}_j} \varepsilon_g - P_E \varepsilon_g.
\]

All the proofs in Section A.4.1 carry over. In fact, all terms involving \(\alpha_g, H\) and \(H^+\) 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).

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.

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,

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

When $r = 0$ and $R > 0$, by (A.6),

$$\max_{i,j \leq N} \left| \frac{1}{T} \sum_t (\hat{u}_{it} \hat{u}_{jt} - u_{it} u_{jt}) \right| \leq O_P \left( \frac{\log N}{T} + \frac{1}{N \nu_{\min}^2} \right).$$

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

$$\max_{i,j \leq N} \left| s_{u,ij} - E u_{it} u_{jt} \right| \leq \max_{i,j \leq N} \left| \frac{1}{T} \sum_t \hat{u}_{it} \hat{u}_{jt} - u_{it} u_{jt} \right| + \max_{i,j \leq N} \left| \frac{1}{T} \sum_t u_{it} u_{jt} - E u_{it} u_{jt} \right| \leq O_P \left( \frac{\sqrt{\log N}}{T} + \frac{1}{N \nu_{\min}^2} \right) = O_P \left( \frac{\sqrt{\log N}}{T} + \frac{1}{\sqrt{N}} \right),$$

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

Given this convergence, the convergence of $\hat{\Sigma}_u$ and $\hat{\Sigma}_u^{-1}$ 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).

A.6 Proof of Theorem 3.4

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

$$\left\| \left( \frac{1}{T} \hat{F}' \hat{F} \right)^{-1} - \left( \frac{1}{T} HH'F'F \right)^{-1} \right\| \leq O_P \left( \frac{1}{N} + \frac{\nu_{\max}(H)}{\sqrt{T N}} \right) \frac{1}{\nu_{\min}^4(\mathbf{H})}.$$
Also by the proof of Theorem 2.1 for \( \| (\frac{1}{T} \hat{F} \hat{F})^{-1} \| + \| (\frac{1}{T} H F' F H')^{-1} \| \leq \frac{c}{\nu_{\min}^2(H)} \). Because

\[
P_{\hat{F}} - P_G = E(\hat{F} \hat{F})^{-1} H F' + F H'[(\hat{F} \hat{F})^{-1} - (H F' F H')^{-1}] H F' + \hat{F} (\hat{F} \hat{F})^{-1} E',
\]

we have

\[
\| P_{\hat{F}} - P_G \|_{F}^2 = \text{tr}(\hat{F} \hat{F})^{-1} H F' H' (\hat{F} \hat{F})^{-1} E' E + \text{tr}(\hat{F} \hat{F})^{-1} E' E
\]

\[
+ 2 \text{tr}(\hat{F} \hat{F})^{-1} H F' H'[(\hat{F} \hat{F})^{-1} - (H F' F H')^{-1}] H F' E
\]

\[
+ \text{tr}[(\hat{F} \hat{F})^{-1} - (H F' F H')^{-1}] H F' (\hat{F} \hat{F})^{-1} F'
\]

\[
+ 2 \text{tr}(\hat{F} \hat{F})^{-1} H F' E (\hat{F} \hat{F})^{-1} E' E
\]

\[
+ 2 \text{tr}(\hat{F} \hat{F})^{-1} H F' E (\hat{F} \hat{F})^{-1} H F'
\]

\[
= 2 \text{tr} H^{-1} (F' F)^{-1} H^{-1} E' E + O_P(\frac{1}{T N \nu_{\min}^2} + \frac{1}{N^2 \nu_{\min}^4} + \frac{1}{N^{\sqrt{T} \nu_{\min}^3}}).
\]

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

\[
\text{MEAN} = E(X \mid F, W) = \text{tr}(A \frac{1}{N^2} W' (E u_i u_i' \mid F) W) = \text{tr}(A \frac{1}{N^2} W' \Sigma_u W).
\]

We note that \( \text{Var}(X \mid F) = \frac{1}{T N^2} \sigma^2 \) and that \( N \sqrt{T} \frac{(X - \text{MEAN})}{\sigma} \overset{d}{\rightarrow} \mathcal{N}(0, 1) \) due to the serial indepence of \( u_i u_i' \mid F \) and that \( E \| \frac{1}{\sqrt{N}} W' u_i \|^4 < C \). In addition, Lemma A.8 below shows that with \( \text{MEAN} = \text{tr}(\frac{1}{N^2} W' \tilde{\Sigma_u} W) \), and \( \tilde{\Sigma_u} = 2(\frac{1}{T} \hat{F} \hat{F})^{-1} \), we have

\[
(\text{MEAN} - \text{MEAN}) N \sqrt{T} = o_P(1).
\]

Also, the same lemma shows \( \hat{\sigma}^2 \overset{P}{\rightarrow} \sigma^2 \). As a result

\[
\frac{\| P_{\hat{F}} - P_G \|_{F}^2 - \text{MEAN}}{N \sqrt{T} \hat{\sigma}} = \frac{X - \text{MEAN}}{\frac{1}{N \sqrt{T} \sigma}} + o_P(1) \overset{d}{\rightarrow} \mathcal{N}(0, 1).
\]

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

\[ \square \]

**Lemma A.8.** Suppose \( R = r \). Let \( g_{NT} := \nu_{\min}^{-2} \frac{1}{N} + \frac{\log N}{T} \).

(i) \( \text{MEAN} - \text{MEAN} = O_P(\frac{g_{NT}}{N \nu_{\min}^2}) \sum_{\sigma_u, ij \neq 0} 1 + O_P(\frac{1}{N^2 \nu_{\min}^4} + \frac{1}{N^{\sqrt{T} \nu_{\min}^3}}) \).

(ii) \( \hat{\sigma}^2 \overset{P}{\rightarrow} \sigma^2 \).

**Proof.** By lemma A.3,

\[
\max_{ij} \left| \frac{1}{T} \sum_t u_{it}(\tilde{u}_{jt} - u_{jt}) \right| \leq O_P(g_{NT}).
\]

22
(i) Recall \( A := 2H^{-1}(\frac{1}{T}F^T)H^{-1} \). Note that \( \|A\| = O_P(\frac{1}{\min(T)}) \). We now bound 
\[ \frac{1}{N}W'(\tilde{\Sigma}_u - \Sigma_u)W. \]
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

\[ \frac{1}{N}W'(\tilde{\Sigma}_u - \Sigma_u)W = \frac{1}{N} \sum_{\sigma_{u,ij} = 0} w_iw_j \tilde{\sigma}_{u,ij} + \frac{1}{N} \sum_{\sigma_{u,ij} \neq 0} w_iw_j(\tilde{\sigma}_{u,ij} - \sigma_{u,ij}) := a_1 + a_2. \]

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

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

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

\[ a_2 = a_{21} + a_{22}, \]
\[ a_{21} = \frac{1}{N} \sum_{\sigma_{u,ij} \neq 0} w_iw_j \frac{1}{T} \sum_t (\tilde{u}_{it}\tilde{u}_{jt} - u_{it}u_{jt}) \]
\[ a_{22} = \frac{1}{N} \sum_{\sigma_{u,ij} \neq 0} w_iw_j \frac{1}{T} \sum_t (u_{it}u_{jt} - E u_{it}u_{jt}). \]

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

\[ a_{21} = \frac{1}{N} \sum_{\sigma_{u,ij} \neq 0} w_iw_j \frac{1}{T} \sum_t (\tilde{u}_{it} - u_{it})(\tilde{u}_{jt} - u_{jt}) + \frac{2}{N} \sum_{\sigma_{u,ij} \neq 0} w_iw_j \frac{1}{T} \sum_t u_{it}(\tilde{u}_{jt} - u_{jt}) \]
\[ \leq \max_{i} \frac{1}{T} \sum_t (\tilde{u}_{it} - u_{it})^2 + \max_{ij} \frac{1}{T} \sum_t u_{it}(\tilde{u}_{jt} - u_{jt}) \frac{1}{N} \sum_{\sigma_{u,ij} \neq 0} 1 \]
\[ \leq O_P(g^2_{NT}) \frac{1}{N} \sum_{\sigma_{u,ij} \neq 0} 1. \]

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

\[ \text{Var}(a_{22}) \leq \frac{1}{N^2T^2} \sum_{s,t \leq T} \sum_{\sigma_{u,mm} \neq 0} \sum_{\sigma_{u,ij} \neq 0} |Cov(u_{it}u_{jt}, u_{ms}u_{ns})| \]
\[ \leq \frac{1}{N^2T} \sum_{\sigma_{u,mm} \neq 0} \sum_{\sigma_{u,ij} \neq 0} |Cov(u_{it}u_{jt}, u_{mt}u_{nt})| \leq O\left(\frac{1}{NT}\right). \]
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

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

This implies

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

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

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.