Submitted to the Annals of Statistics

SUPPLEMENTARY APPENDIX TO THE PAPER PROJECTED PRINCIPAL COMPONENT ANALYSIS IN FACTOR MODELS

By Jianqing Fan[*], Yuan Liao[†] and Weichen Wang[*]

Princeton University[*] *and University of Maryland*[†]

This document contains all the remaining proofs and technical lemmas. In addition, more detailed justifications of the regularity conditions are provided.

CONTENTS

AMS 2000 subject classifications: Primary 62H25; secondary 62H15.
Keywords and phrases: semi-parametric approximate factor models; high dimensionality; loading matrix modeling; sieve approximation.

APPENDIX B: REMAINING PROOFS FOR SECTION 3

B.1. Technical lemmas.

LEMMA B.1. *(i)* $\max_{k\le K, i\le p}\max_{t\le T}\sum_{s=1}^T|\mathrm{cov}(f_{tk}u_{it}, f_{sk}u_{is})|=O(1)$,
$\max_{i\le p, j\le p}\max_{t\le T}\sum_{s=1}^T|\mathrm{cov}(u_{it}u_{jt}, u_{is}u_{js})|=O(1)$.
$\max_{t\le T}\max_{i\le p}\sum_{s=1}^T|Eu_{it}u_{is}|=O(1)$.
(ii) $\|\mathbf{F}'\mathbf{U}'\|_F=O_P(\sqrt{pT})$, $\|T^{-1}\mathbf{U}\mathbf{U}'-E\mathbf{u}_t\mathbf{u}_t'\|_F=O_P(p/\sqrt{T})$.
(iii) $\|\mathbf{U}'\Phi(\mathbf{X})\|_F^2=O_P(pTJ)$, $\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\|_F^2=O_P(pTJ)$, $\|\mathbf{P}\mathbf{U}\|_F^2=O_P(TJ)$.

PROOF. (i) By Davydov's inequality (Corollary 16.2.4 in Athreya and Hahiri (2006)), there is a constant $C>0$, for all $i\le p, k\le K, t\le T$, $|\mathrm{cov}(f_{tk}u_{it}, f_{sk}u_{is})|\le C\sqrt{\alpha(|t-s|)}$, where $\alpha(t)$ is the α-mixing coefficient of $\{\mathbf{f}_t,\mathbf{u}_t\}$. By the strong mixing condition, $\sum_{t=1}^\infty\sqrt{\alpha(t)}<\infty$. Thus uniformly in T,

$$\max_{k\le K, i\le p}\max_{t\le T}\sum_{s=1}^T|\mathrm{cov}(f_{tk}u_{it}, f_{sk}u_{is})|\le\max_{t\le T}\sum_{s=1}^T C\alpha(|t-s|)^{1/2}<\infty.$$

The other two results follow similarly.
(ii) We have, by part (i),

$$\begin{aligned}
&E\|\mathbf{F}'\mathbf{U}'\|_F^2=E\sum_{k=1}^K\sum_{i=1}^p(\sum_{t=1}^T f_{tk}u_{it})^2=\sum_{k=1}^K\sum_{i=1}^p\mathrm{var}(\sum_{t=1}^T f_{tk}u_{it})\\
=&\sum_{k=1}^K\sum_{i=1}^p[\sum_{t=1}^T\mathrm{var}(f_{tk}u_{it})+\sum_{t\le T, s\le T}\mathrm{cov}(f_{tk}u_{it}, f_{sk}u_{is})]=O(pT)\\
&E\|T^{-1}\mathbf{U}\mathbf{U}'-E\mathbf{u}_t\mathbf{u}_t'\|_F^2=E\sum_{i=1}^p\sum_{j=1}^p(\frac{1}{T}\sum_{t=1}^T u_{it}u_{jt}-Eu_{it}u_{jt})^2=\sum_{i=1}^p\sum_{j=1}^p\mathrm{var}(\frac{1}{T}\sum_{t=1}^T u_{it}u_{jt})\\
=&\sum_{i=1}^p\sum_{j=1}^p\frac{1}{T^2}[\sum_{t=1}^T\mathrm{var}(u_{it}u_{jt})+\sum_{t\le T, s\le T}\mathrm{cov}(u_{it}u_{jt}, u_{is}u_{js})]=O(p^2/T).
\end{aligned}$$

The results follow from Markov inequality.
(iii)

$$E\|\mathbf{U}'\Phi(\mathbf{X})\|_F^2 \;=\; \sum_{t=1}^T\sum_{j=1}^J\sum_{l=1}^d E(\sum_{i=1}^p\phi_j(X_{il})u_{it})^2=\sum_{t=1}^T\sum_{j=1}^J\sum_{l=1}^d\sum_{i=1}^p\sum_{m=1}^p E\phi_j(X_{il})\phi_j(X_{ml})Eu_{it}u_{mt}$$

$$\leq Jdp \max_{j\leq J, l\leq d, i\leq p, m\leq p} |E\phi_j(X_{il})\phi_j(X_{ml})| \sum_{t=1}^T \max_{m\leq p} \sum_{i=1}^p |Eu_{it}u_{mt}|$$
$$\leq Jdp \max_{j\leq J, l\leq d, i\leq p} E\phi_j(X_{il})^2 \sum_{t=1}^T \max_{m\leq p} \sum_{i=1}^p |Eu_{it}u_{mt}| = O(pTJ),$$

where the second inequality is due to the Cauchy Schwarz inequality. On the other hand,

$$E\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\|_F^2 = \sum_{k=1}^K \sum_{j=1}^J \sum_{l=1}^d E(\sum_{i=1}^p \sum_{t=1}^T \phi_j(X_{il})u_{it}f_{tk})^2$$
$$= \sum_{k=1}^K \sum_{j=1}^J \sum_{l=1}^d \mathrm{var}(\sum_{i=1}^p \sum_{t=1}^T \phi_j(X_{il})u_{it}f_{tk})$$
$$= \sum_{k=1}^K \sum_{j=1}^J \sum_{l=1}^d \sum_{i=1}^p \sum_{t=1}^T \mathrm{var}(\phi_j(X_{il})u_{it}f_{tk}) + \sum_{k=1}^K \sum_{j=1}^J \sum_{l=1}^d \sum_{i=1}^p \sum_{t\neq s; t,s\leq T} (E\phi_j(X_{il})^2 f_{tk}f_{sk})Eu_{it}u_{is}$$
$$+ \sum_{k=1}^K \sum_{j=1}^J \sum_{l=1}^d \sum_{i\neq m; i,m\leq p} \sum_{t,s\leq T} (E\phi_j(X_{il})\phi_j(X_{ml})f_{tk}f_{sk})Eu_{it}u_{ms}$$
$$\leq O(pTJ) + O(pTJ) \max_{t\leq T, i\leq p} \sum_{s=1}^T |Eu_{it}u_{is}| + O_P(pTJ)\frac{1}{pT} \sum_{i,m\leq p} \sum_{t,s\leq T} |Eu_{it}u_{ms}| = O(pTJ).$$

Finally, since $\|(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\|_2 = \lambda_{\min}^{-1}(p^{-1}\Phi(\mathbf{X})'\Phi(\mathbf{X}))p^{-1} = O_P(p^{-1})$,

$$\begin{aligned} \|\mathbf{PU}\|_F &= \|\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})'\mathbf{U}\|_F \\ &\leq \|\Phi(\mathbf{X})\|_2 \|(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\|_2 \|\Phi(\mathbf{X})'\mathbf{U}\|_F \\ &= O_P(\sqrt{p}p^{-1}\sqrt{pTJ}) = O_P(\sqrt{TJ}). \end{aligned}$$

□

Lemma B.2. *In the conventional factor model,*
(i) $\|\mathbf{F}'\mathbf{D}_1\|_F = O_P(\sqrt{TJ/p} + TJ/p)$, $\|\mathbf{F}'\mathbf{D}_2\|_F = O_P(J\sqrt{T}/p + \sqrt{J/p})$, $\|\mathbf{F}'\mathbf{D}_3\|_F = O_P(\sqrt{TJ/p})$.
(ii) $\frac{1}{T}\|\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{FM})\|_F = O_P(\sqrt{J/(pT)} + J/p)$.

Proof. The proof of this lemma uses the result $\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{FM}\|_F^2 = O_p(\frac{J}{p})$, which is already proved in Lemma A.3 in the main paper. Note that the proof of Lemma A.3 only requires Lemmas A.2 and B.1, which are self-contained.

(i) Note that $\|\Phi(\mathbf{X})'\mathbf{UF}\|_F = O_P(\sqrt{pTJ})$ and

$$\|\mathbf{PUF}\|_F \leq \|\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\|_2 \|\Phi(\mathbf{X})'\mathbf{UF}\|_F \leq O_P(\sqrt{TJ}).$$

Hence

$$\|\mathbf{F}'\mathbf{D}_1\|_F \le \frac{1}{Tp}\|\mathbf{F}'\mathbf{F}\mathbf{\Lambda}'\mathbf{P}\|(\|\mathbf{PUFM}\|_F + \|\mathbf{PU}(\widehat{\mathbf{F}} - \mathbf{FM})\|_F) = O_P(\sqrt{TJ/p} + TJ/p).$$

$$\|\mathbf{F}'\mathbf{D}_2\|_F \le \frac{1}{Tp}\|\mathbf{F}'\mathbf{U}'\mathbf{P}\|_F(\|\mathbf{PU}(\widehat{\mathbf{F}} - \mathbf{FM})\|_F + \|\mathbf{PUFM}\|_F) = O_P(J/p(1 + \sqrt{JT/p})).$$

$$\|\mathbf{F}'\mathbf{D}_3\|_F \le \frac{1}{Tp}\|\mathbf{F}'\mathbf{U}'\mathbf{P}\|_F\|\mathbf{P}\mathbf{\Lambda}\mathbf{F}'\widehat{\mathbf{F}}\|_F = O_P(\sqrt{TJ/p}).$$

Part (ii) follows from part (i).

□

B.2. Proof of Corollary 3.1.

Proof. Let $\widetilde{\mathbf{V}}$ be an orthogonal matrix whose columns are the eigenvectors of $\mathbf{\Lambda}'\mathbf{\Lambda}$, corresponding to the eigenvalues in a decreasing order. Let $\mathbf{\Sigma}_u$ be the $p \times p$ covariance matrix of $\mathbf{u}_t$. Also let

$$\mathbf{V} = \widetilde{\mathbf{V}}'(\mathbf{\Lambda}'\mathbf{\Lambda})^{1/2}\widehat{\mathbf{D}}^{-1/2}.$$

Since $\widehat{\mathbf{G}}(\mathbf{X}) = \widehat{\mathbf{\Xi}}\widehat{\mathbf{D}}^{1/2}$, by Theorem 3.1, $\frac{1}{p}\|\widehat{\mathbf{\Xi}}\widehat{\mathbf{D}}^{1/2} - \mathbf{P}\mathbf{\Lambda}\|_F^2 = O_P(\frac{J}{pT} + \frac{J^2}{p^2})$. Hence $\frac{1}{p}\|\widehat{\mathbf{\Xi}}\widehat{\mathbf{D}}^{1/2} - \mathbf{\Lambda}\|_F^2 = O_P(\frac{J}{pT} + \frac{J^2}{p^2}) + \frac{1}{p}\|\mathbf{P}\mathbf{\Lambda} - \mathbf{\Lambda}\|_F^2$. By lemma A.2, $\|(\widehat{\mathbf{D}}/p)^{-1}\|_2 = \|\mathbf{K}^{-1}\|_2 = O_P(1)$, which implies

$$\|\widehat{\mathbf{\Xi}} - \mathbf{\Lambda}\widehat{\mathbf{D}}^{-1/2}\|_F^2 = O_P\Big(\frac{J}{pT} + \frac{J^2}{p^2} + \frac{1}{p}\|\mathbf{P}\mathbf{\Lambda} - \mathbf{\Lambda}\|_F^2\Big).$$

On the other hand, define $\bar{\mathbf{\Lambda}} = \mathbf{\Lambda}\widetilde{\mathbf{V}} = (\bar{\mathbf{\Lambda}}_1, ...\bar{\mathbf{\Lambda}}_K)$. Then $\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}}$ is diagonal and $\mathbf{\Lambda}\mathbf{\Lambda}'\bar{\mathbf{\Lambda}}_j = \bar{\mathbf{\Lambda}}_j\|\bar{\mathbf{\Lambda}}_j\|_2^2$, $j = 1, ..., K$. This implies that the columns of $\bar{\mathbf{\Lambda}}(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{-1/2}$ are the eigenvectors of $\mathbf{\Lambda}\mathbf{\Lambda}'$ corresponding to the largest K eigenvalues. In addition, in the factor model, we have the following matrix decomposition: for $\mathbf{\Sigma}_u = \mathrm{cov}(\mathbf{u}_t)$, $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Sigma}_u$. Hence by the same argument of the proof of Proposition 2.2 in Fan et al. (2013),

$$\|\mathbf{\Xi} - \mathbf{\Lambda}\widetilde{\mathbf{V}}(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{-1/2}\|_F = O\Big(\frac{1}{p}\|\mathbf{\Sigma}_u\|_2\Big).$$

Using $\widetilde{\mathbf{V}}\widetilde{\mathbf{V}}' = \mathbf{I}$, we have

$$\begin{aligned}
\|\widehat{\mathbf{\Xi}} - \mathbf{\Xi}\widetilde{\mathbf{V}}'(\mathbf{\Lambda}'\mathbf{\Lambda})^{1/2}\widehat{\mathbf{D}}^{-1/2}\|_F \quad &\le \quad \|\widehat{\mathbf{\Xi}} - \mathbf{\Lambda}\widehat{\mathbf{D}}^{-1/2}\|_F + \|\mathbf{\Lambda}\widehat{\mathbf{D}}^{-1/2} - \mathbf{\Xi}(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{1/2}\widetilde{\mathbf{V}}'\widehat{\mathbf{D}}^{-1/2}\|_F \\
&\quad + \|\mathbf{\Xi}(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{1/2}\widetilde{\mathbf{V}}'\widehat{\mathbf{D}}^{-1/2} - \mathbf{\Xi}\widetilde{\mathbf{V}}'(\mathbf{\Lambda}'\mathbf{\Lambda})^{1/2}\widehat{\mathbf{D}}^{-1/2}\|_F \\
&\le \quad \|\widehat{\mathbf{\Xi}} - \mathbf{\Lambda}\widehat{\mathbf{D}}^{-1/2}\|_F + \|(\mathbf{\Lambda}\widetilde{\mathbf{V}}(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{-1/2} - \mathbf{\Xi})(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{1/2}\widetilde{\mathbf{V}}'\widehat{\mathbf{D}}^{-1/2}\|_F \\
&\quad + \|\mathbf{\Xi}((\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{1/2}\widetilde{\mathbf{V}}' - \widetilde{\mathbf{V}}'(\mathbf{\Lambda}'\mathbf{\Lambda})^{1/2})\widehat{\mathbf{D}}^{-1/2}\|_F.
\end{aligned}$$

On the right hand side, the first term is $O_P(\frac{J}{pT}+\frac{J^2}{p^2}+\frac{1}{p}\|\mathbf{P\Lambda}-\mathbf{\Lambda}\|_F^2)$, as is proved above. Still by $\|(p^{-1}\widehat{\mathbf{D}})^{-1/2}\|_2=O_P(1)$, the second term is bounded by

$$\|\mathbf{\Lambda}\widetilde{\mathbf{V}}(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{-1/2}-\mathbf{\Xi}\|_F\|(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}}/p)^{1/2}\|_F\|\widetilde{\mathbf{V}}\|_2\|(\widehat{\mathbf{D}}/p)^{-1/2}\|_2=O(\|\mathbf{\Sigma}_u\|_2/p).$$

Finally, since $(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{1/2}=\widetilde{\mathbf{V}}'(\mathbf{\Lambda}'\mathbf{\Lambda})^{1/2}\widetilde{\mathbf{V}}$, so $(\bar{\mathbf{\Lambda}}'\bar{\mathbf{\Lambda}})^{1/2}\widetilde{\mathbf{V}}'-\widetilde{\mathbf{V}}'(\mathbf{\Lambda}'\mathbf{\Lambda})^{1/2}=0$, which implies the third term is zero. Hence

$$\|\widehat{\mathbf{\Xi}}-\mathbf{\Xi}\widetilde{\mathbf{V}}'(\mathbf{\Lambda}'\mathbf{\Lambda})^{1/2}\widehat{\mathbf{D}}^{-1/2}\|_F\le O_P\Big(\frac{J}{pT}+\frac{J^2}{p^2}+\frac{1}{p}\|\mathbf{P\Lambda}-\mathbf{\Lambda}\|_F^2\Big)+O\Big(\frac{1}{p}\|\mathbf{\Sigma}_u\|_2\Big).$$

□

APPENDIX C: PROOFS FOR SECTION 4

C.1. Convergence of $\widehat{\mathbf{F}}$. Recall that $\mathbf{K}$ denote the $K\times K$ diagonal matrix consisting of the first K largest eigenvalues of $(pT)^{-1}\mathbf{Y}'\mathbf{PY}$ in descending order. By the definition of eigenvalues, we have

$$\frac{1}{Tp}(\mathbf{Y}'\mathbf{PY})\widehat{\mathbf{F}}=\widehat{\mathbf{F}}\mathbf{K}$$

Let

$$\mathbf{H}=\frac{1}{Tp}\mathbf{B}'\Phi(\mathbf{X})'\Phi(\mathbf{X})\mathbf{B}\mathbf{F}'\widehat{\mathbf{F}}\mathbf{K}^{-1}.$$

We shall show in Lemma C.5 below that $\|\mathbf{H}\|_2=O_P(1)$.

Substituting $\mathbf{Y}=\Phi(\mathbf{X})\mathbf{BF}'+\mathbf{\Gamma F}'+\mathbf{R}(\mathbf{X})\mathbf{F}'+\mathbf{U}$, we have, for $\mathbf{F}$ being the $T\times K$ matrix of f_{tk},

$$\widehat{\mathbf{F}}-\mathbf{FH}=\Big(\sum_{i=1}^{15}\mathbf{A}_i\Big)\mathbf{K}^{-1}\tag{C.1}$$

where

$$\begin{aligned}
\mathbf{A}_1&=\frac{1}{Tp}\mathbf{FB}'\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}},\quad \mathbf{A}_2=\frac{1}{Tp}\mathbf{U}'\Phi(\mathbf{X})\mathbf{BF}'\widehat{\mathbf{F}},\quad \mathbf{A}_3=\frac{1}{Tp}\mathbf{U}'\mathbf{PU}\widehat{\mathbf{F}},\\
\mathbf{A}_4&=\frac{1}{Tp}\mathbf{FB}'\Phi(\mathbf{X})'\mathbf{R}(\mathbf{X})\mathbf{F}'\widehat{\mathbf{F}},\quad \mathbf{A}_5=\frac{1}{Tp}\mathbf{FR}(\mathbf{X})'\Phi(\mathbf{X})\mathbf{BF}'\widehat{\mathbf{F}},\\
\mathbf{A}_6&=\frac{1}{Tp}\mathbf{FR}(\mathbf{X})'\mathbf{PR}(\mathbf{X})\mathbf{F}'\widehat{\mathbf{F}},\quad \mathbf{A}_7=\frac{1}{Tp}\mathbf{FR}(\mathbf{X})'\mathbf{PU}\widehat{\mathbf{F}},\quad \mathbf{A}_8=\frac{1}{Tp}\mathbf{U}'\mathbf{PR}(\mathbf{X})\mathbf{F}'\widehat{\mathbf{F}},\\
\mathbf{A}_9&=\frac{1}{Tp}\mathbf{FB}'\Phi(\mathbf{X})'\mathbf{\Gamma F}'\widehat{\mathbf{F}},\quad \mathbf{A}_{10}=\frac{1}{Tp}\mathbf{F\Gamma}'\Phi(\mathbf{X})\mathbf{BF}'\widehat{\mathbf{F}},\\
\mathbf{A}_{11}&=\frac{1}{Tp}\mathbf{F\Gamma}'\mathbf{PU}\widehat{\mathbf{F}},\quad \mathbf{A}_{12}=\frac{1}{Tp}\mathbf{U}'\mathbf{P\Gamma F}'\widehat{\mathbf{F}},\quad \mathbf{A}_{13}=\frac{1}{Tp}\mathbf{FR}(\mathbf{X})'\mathbf{P\Gamma F}'\widehat{\mathbf{F}},
\end{aligned}$$

$$\mathbf{A}_{14} = \frac{1}{Tp}\mathbf{F}\mathbf{\Gamma}'\mathbf{P}\mathbf{R}(\mathbf{X})\mathbf{F}'\widehat{\mathbf{F}}, \quad \mathbf{A}_{15} = \frac{1}{Tp}\mathbf{F}\mathbf{\Gamma}'\mathbf{P}\mathbf{\Gamma}\mathbf{F}'\widehat{\mathbf{F}}.$$

The proof is divided into two steps. First, we bound $\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2$, and then we prove the convergence for $\frac{1}{T}\|\mathbf{F}\mathbf{H} - \mathbf{F}\|_F^2$. Note that $\frac{1}{T}\|\mathbf{F}\mathbf{H} - \mathbf{F}\|_F^2 \leq \|\mathbf{H} - \mathbf{I}\|_F^2$, hence in the second step we bound $\|\mathbf{H} - \mathbf{I}\|_F^2$. Note that a key step of bounding $\|\mathbf{H} - \mathbf{I}\|_F$ is to show the fast convergence of $\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})$, which will be done in Appendix C.2.

To bound $\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2$, note that there is a constant $C > 0$, so that

$$\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2 \leq C\|\mathbf{K}^{-1}\|_2^2 \sum_{i=1}^{15} \frac{1}{T}\|\mathbf{A}_i\|_F^2.$$

Hence we need to bound $\frac{1}{T}\|\mathbf{A}_i\|_F^2$ for $i = 1, ..., 15$. The following lemma gives the stochastic bounds for individual terms. In this section, $p, J \to \infty$, and T may grow simultaneously with p or stay constant.

Lemma C.1. *(i)* $\frac{1}{T}\|\mathbf{A}_1\|_F^2 = O_P(p^{-1}) = \frac{1}{T}\|\mathbf{A}_2\|_F^2$,
(ii) $\frac{1}{T}\|\mathbf{A}_3\|_F^2 = O_P(J^2/p^2)$, $\frac{1}{T}\|\mathbf{A}_4\|_F^2 = O_P(J^{-\kappa}) = \frac{1}{T}\|\mathbf{A}_5\|_F^2$.
(iii) $\frac{1}{T}\|\mathbf{A}_6\|_F^2 = O_P(J^{-2\kappa})$, $\frac{1}{T}\|\mathbf{A}_7\|_F^2 = O_P(p^{-1}J^{-(\kappa-1)}) = \frac{1}{T}\|\mathbf{A}_8\|_F^2$.
(iv) $\frac{1}{T}\|\mathbf{A}_9\|_F^2 = O_p(\nu_p/p) = \frac{1}{T}\|\mathbf{A}_{10}\|_F^2$, $\frac{1}{T}\|\mathbf{A}_{11}\|_F^2 = O_P(J^2\nu_p/p^2) = \frac{1}{T}\|\mathbf{A}_{12}\|_F^2$.
(v) $\frac{1}{T}\|\mathbf{A}_{13}\|_F^2 = O_P(\nu_p p^{-1}J^{-(\kappa-1)}) = \frac{1}{T}\|\mathbf{A}_{14}\|_F^2$., $\frac{1}{T}\|\mathbf{A}_{15}\|_F^2 = O_P(J\nu_p/p^3)$.

Proof. (i) Because $\|\mathbf{F}\|_F^2 = O_P(T)$, $\|\widehat{\mathbf{F}}\|_F^2 = O_P(T)$. By Lemma C.6 in the supplementary material, $\|\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F^2 = O_P(pT)$. Hence $\frac{1}{T}\|\mathbf{A}_1\|_F^2 = O_P(p^{-1})$. The rate of convergence for $\|\mathbf{A}_2\|$ is obtained in the same way.

(ii) We have $\mathbf{A}_3 = \frac{1}{Tp}\mathbf{U}'\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}$. By Lemma B.1, $\|\mathbf{U}'\Phi(\mathbf{X})\|_F = O_P(\sqrt{pTJ})$. By Assumption 3.3, $\|(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\|_2 = O_P(p^{-1})$. So $\frac{1}{T}\|\mathbf{A}_3\|_F^2 = O_P(J^2/p^2)$.

Note that $\|\Phi(\mathbf{X})\mathbf{B}\|_2 \leq \|\mathbf{G}(\mathbf{X})\|_2 + \|\mathbf{R}(\mathbf{X})\|_2 = O_P(\sqrt{p})$, and $\|\mathbf{R}(\mathbf{X})\|_F^2 = O_P(pJ^{-\kappa})$ Thus

$$\frac{1}{T}\|\mathbf{A}_4\|_F^2 \leq \frac{1}{T^3p^2}\|\mathbf{F}\|_F^4\|\widehat{\mathbf{F}}\|_F^2\|\Phi(\mathbf{X})\mathbf{B}\|_2^2\|\mathbf{R}(\mathbf{X})\|_F^2 = O_P(J^{-\kappa}).$$

Similarly, $\frac{1}{T}\|\mathbf{A}_5\|_F^2 = O_P(J^{-\kappa})$.

(iii) Note that $\|\mathbf{P}\|_2 = \|(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}\Phi(\mathbf{X})'\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}\|_2 = 1$. Hence $\frac{1}{T}\|\mathbf{A}_6\|_F^2 = O_P(J^{-2\kappa})$. In addition,

$$\|\mathbf{A}_7\|_F \leq \frac{1}{Tp}\|\mathbf{F}\|_F\|\widehat{\mathbf{F}}\|_F\|\mathbf{R}(\mathbf{X})\|_F\|\Phi(\mathbf{X})\|_2\|(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\|_2\|\Phi(\mathbf{X})'\mathbf{U}\|_2 = O_P\Big(\sqrt{\frac{TJ}{pJ^\kappa}}\Big).$$

Hence $\frac{1}{T}\|\mathbf{A}_7\|_F^2 = O_P(p^{-1}J^{-(\kappa-1)})$. The rate of convergence for $\mathbf{A}_8$ can be bounded in the same way.

(iv) It follows from Lemma C.6 that $\|\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2 = O_P(J\sum_{i=1}^p \operatorname{var}(\gamma_{ik}))$. Hence $\frac{1}{T}\|\mathbf{A}_9\|_F^2 = O_P(\frac{1}{p^2}\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2) = O_P(\frac{\nu_p}{p})$. The rate for $\mathbf{A}_{10}$ follows similarly. $\frac{1}{T}\|\mathbf{A}_{11}\|_F^2 = O_P(\frac{1}{Tp^4}\|\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2\|\mathbf{U}'\Phi(\mathbf{X})\|_F^2) = O_P(\frac{J^2\nu_p}{p^2})$. Similarly, $\frac{1}{T}\|\mathbf{A}_{12}\|_F^2$ attains the same rate.

(v) $\frac{1}{T}\|\mathbf{A}_{13}\|_F^2 = O_P(\frac{1}{p^2}\|\mathbf{R}(\mathbf{X})\|_F^2\|\mathbf{P}\mathbf{\Gamma}\|_F^2) = O_P(\nu_p p^{-1}J^{-(\kappa-1)})$. The rate for $\mathbf{A}_{14}$ is obtained in the same way. Finally,

$$\frac{1}{T}\|\mathbf{A}_{15}\|_F^2 = O_P\Big(\frac{1}{p^4}\|\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2\Big) = O_P\Big(\frac{J\nu_p}{p^3}\Big).$$

□

The final rate of convergence for $\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2$ is summarized as follows.

Proposition C.1. *As $J = o(\sqrt{p})$ and $\kappa \geq 1$,*

$$\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2 = O_P\Big(\frac{1}{p} + \frac{1}{J^\kappa}\Big). \tag{C.2}$$

Proof. It follows from Lemma C.5 that $\|\mathbf{K}^{-1}\|_2 = O_P(1)$. Then Lemma C.1 and equality (C.1) together imply the desired result.

□

C.2. Upper bounds for $\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})$. In the semi-parametric factor model, we also need to upper bound $\frac{1}{T}\|\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})\|_F$, which is needed to prove the convergence of $\mathbf{H} - \mathbf{I}_K$. Though this term was bounded previously in Lemma B.2 in the regular factor model, the upper bounds can be made sharper in the semi-parametric factor model being considered, where $\mathbf{\Lambda} = \mathbf{G}(\mathbf{X}) + \mathbf{\Gamma}$. The proof below uses the result $\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2 = O_P\Big(\frac{1}{p} + \frac{1}{J^\kappa}\Big)$ achieved in Proposition C.1. Note that we shall not use a simple inequality $\frac{1}{T}\|\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})\|_F \leq \frac{1}{T}\|\mathbf{F}\|_F\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F$, since it is too crude. Instead, by the equality (C.1), we have the following decomposition:

$$\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}) = \sum_{i=1}^{15}\mathbf{F}'\mathbf{A}_i\mathbf{K}^{-1}.$$

To achieve sharp bounds, the fifteen terms on the right hand side can be divided and treated differently in three categories.

Category I: terms $\mathbf{A}_i$: $i =$ 4-6, 9,10, 13-15

For these terms, we shall apply the simple inequality $\|\mathbf{F}'\mathbf{A}_i\|_F \leq \|\mathbf{F}\|_F\|\mathbf{A}_i\|_F$, directly using Lemma C.1.

Category II: terms $\mathbf{A}_i$: $i = \mathbf{1}, \mathbf{7}, \mathbf{11}$

For these terms, we still apply the inequality $\|\mathbf{F}'\mathbf{A}_i\|_F \leq \|\mathbf{F}\|_F\|\mathbf{A}_i\|_F$. But instead of directly using the results in Lemma C.1, we shall further improve the rates of $\frac{1}{T}\|\mathbf{A}_1\|_F^2$, $\frac{1}{T}\|\mathbf{A}_7\|_F^2$ and $\frac{1}{T}\|\mathbf{A}_{11}\|_F^2$. In fact, given $\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2 = O_P\left(\frac{1}{p} + \frac{1}{J^\kappa}\right)$ achieved in Proposition C.1, the rates for these terms can be improved. This will be achieved in Lemma C.2 below.

Category III: terms $\mathbf{A}_i$: $i = \mathbf{2{,}3{,}8{,}12}$

For these terms, we shall directly bound $\|\mathbf{F}'\mathbf{A}_i\|_F$, achieving a shaper bound than that of $\|\mathbf{F}\|_F\|\mathbf{A}_i\|_F$. We do so in Lemma C.3 below.

LEMMA C.2. *(improved convergence rate)*
(i) $\frac{1}{T}\|\mathbf{A}_1\|_F^2 = O_P(p^{-2} + p^{-1}J^{-\kappa} + p^{-1}T^{-1})$,
(ii) $\frac{1}{T}\|\mathbf{A}_3\|_F^2 = O_P(J^2/p^3 + J^4/p^4 + J^{2-\kappa}/p^2 + J^2p^{-2}T^{-1})$,
(iii) $\frac{1}{T}\|\mathbf{A}_7\|_F^2 = O_P(J^{1-\kappa}/p^2 + J^{3-\kappa}/p^3 + J^{1-2\kappa}/p + J^{1-\kappa}p^{-1}T^{-1})$,
(iv) $\frac{1}{T}\|\mathbf{A}_{11}\|_F^2 = O_P(J^2\nu_p p^{-2}(p^{-1}+T^{-1}+J^{-\kappa}))$, *(recall* $\nu_p = \max_{k\leq K}\frac{1}{p}\sum_{i=1}^p \mathrm{var}(\gamma_{ik})$*).*

PROOF. (i) Note that $\frac{1}{T}\|\mathbf{A}_1\|_F^2 \leq \frac{1}{T^3p^2}\|\mathbf{F}\|_F^2\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F^2$.

$$\begin{aligned}
\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F^2 &\leq 2\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})\|_F^2 + 2\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\mathbf{H}\|_F^2 \\
&\leq 2\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\|_F^2\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2 + 2\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\|_F^2\|\mathbf{H}\|_2^2 \\
&\leq O_P(pT)O_P(T/p + TJ^2/p^2 + T/J^\kappa) + O_P(pT)O_P(1) \\
&= O_P(T^2 + T^2J^2/p + T^2p/J^\kappa + pT), \qquad \text{(C.3)}
\end{aligned}$$

where the third inequality follows from Lemmas C.5, C.6 and (C.2). This implies $\frac{1}{T}\|\mathbf{A}_1\|_F^2 = O_P(p^{-2} + J^2/p^3 + p^{-1}J^{-\kappa} + p^{-1}T^{-1})$.

(ii) $\frac{1}{T}\|\mathbf{A}_3\|_F^2 \leq \frac{1}{T^3p^2}\|\mathbf{U}'\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\|_2^2\|\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F^2 = O_P(\frac{J}{T^2p^3})\|\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F^2$, where by (C.2) and Lemma B.1,

$$\begin{aligned}
\|\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F^2 &\leq 2\|\Phi(\mathbf{X})'\mathbf{U}\|_F^2\|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}\|_F^2 + 2\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\mathbf{H}\|_F^2 \\
&\leq O_P(pTJ)(T/p + T/J^\kappa) + O_P(pTJ). \qquad \text{(C.4)}
\end{aligned}$$

This yields the result. Also note $\|\mathbf{P}\mathbf{U}\widehat{\mathbf{F}}\|_F^2 = O_P(\frac{1}{p}\|\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F^2)$.

(iii) We have,

$$\begin{aligned}
\frac{1}{T}\|\mathbf{A}_7\|_F^2 &\leq \frac{1}{T^3p^2}\|\mathbf{F}\mathbf{R}(\mathbf{X})'\|_F^2\|\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\|_2^2\|\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F^2 \\
&= O_P(\frac{1}{T^2p^2J^\kappa})\|\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F^2 \\
&= O_P(J^{1-\kappa}/p^2 + J^{3-\kappa}/p^3 + J^{1-2\kappa}/p + J^{1-\kappa}p^{-1}T^{-1}).
\end{aligned}$$

(iv) Note that $\|\mathbf{F}\mathbf{\Gamma}'\mathbf{P}\mathbf{U}\widehat{\mathbf{F}}\|_F \leq \|\mathbf{F}\mathbf{\Gamma}'\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\|_2\|\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|_F$. Hence the result follows.

□

LEMMA C.3. *Under the assumptions of Theorem 4.1,*
(i) $\frac{1}{T}\|\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})\|_F = O_P(\frac{1}{p} + \frac{1}{\sqrt{pT}} + \frac{1}{J^{\kappa/2}} + \sqrt{\frac{\nu_p}{p}})$.
(ii) $\frac{1}{T}\|\widehat{\mathbf{F}}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})\|_F = O_P(\frac{1}{p} + \frac{1}{\sqrt{pT}} + \frac{1}{J^{\kappa/2}} + \sqrt{\frac{\nu_p}{p}})$.

PROOF. By (C.1), $\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}) = \sum_{i=1}^{15}\mathbf{F}'\mathbf{A}_i\mathbf{V}^{-1}$. We evaluate each term in the sum. For i =1,4-7, 9-11, 13-15, $\frac{1}{T^2}\|\mathbf{F}'\mathbf{A}_i\|_F^2 \leq \frac{1}{T}\|\mathbf{A}_i\|_F^2$, which then follows from Lemmas C.1 and C.2.

The terms involving $\mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_8, \mathbf{A}_{12}$ can be bounded more tightly. Specifically,

$$\begin{aligned}
\frac{1}{T^2}\|\mathbf{F}'\mathbf{A}_2\|_F^2 &= \frac{1}{T^4p^2}\|\mathbf{F}'\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\mathbf{F}'\widehat{\mathbf{F}}\|_F^2 \\
&= O_P(\frac{1}{T^4p^2})O_P(pT)O_P(T^2) = O_P(\frac{1}{pT}), \\
\frac{1}{T^2}\|\mathbf{F}'\mathbf{A}_3\|_F^2 &= \frac{1}{T^4p^2}\|\mathbf{F}'\mathbf{U}'\mathbf{P}\mathbf{U}\widehat{\mathbf{F}}\|_F^2 \\
&\leq \frac{1}{T^4p^4}\|\mathbf{F}'\mathbf{U}'\Phi(\mathbf{X})\|_F^2\|\Phi(\mathbf{X})'\mathbf{U}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})\|_F^2 + \frac{1}{T^4p^4}\|\mathbf{F}'\mathbf{U}'\Phi(\mathbf{X})\|_F^4 \\
&= O_P(\frac{J^2}{Tp^2}(\frac{1}{p} + \frac{1}{T} + \frac{1}{J^\kappa})) = O_P(\frac{J^2}{Tp^3} + \frac{J^2}{Tp^2J^\kappa} + \frac{J^2}{T^2p^2}),
\end{aligned}$$

where we used $\|\Phi(\mathbf{X})'\mathbf{U}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})\|_F^2 = O_P(JT^2 + \frac{pT^2J}{J^\kappa})$, and $\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\|_F^2 = O_P(pTJ)$.

$$\begin{aligned}
\frac{1}{T^2}\|\mathbf{F}'\mathbf{A}_8\|_F^2 &= \frac{1}{T^4p^2}\|\mathbf{F}'\mathbf{U}'\mathbf{P}\mathbf{R}(\mathbf{X})\mathbf{F}'\widehat{\mathbf{F}}\|_F^2 \\
&\leq \frac{1}{T^4p^2}\|\mathbf{F}'\mathbf{U}'\Phi(\mathbf{X})\|_F^2\|(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})'\mathbf{R}(\mathbf{X})\mathbf{F}'\widehat{\mathbf{F}}\|_F^2 \\
&= O_P(\frac{J}{TJ^\kappa p}).
\end{aligned}$$

where we used $\|(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})'\mathbf{R}(\mathbf{X})\mathbf{F}'\widehat{\mathbf{F}}\|_F^2 = O_P(T^2/J^\kappa)$.

$$\frac{1}{T^2}\|\mathbf{F}'\mathbf{A}_{12}\|_F^2 = \frac{1}{T^4p^2}\|\mathbf{F}'\mathbf{U}'\mathbf{P}\mathbf{\Gamma}\mathbf{F}'\widehat{\mathbf{F}}\|_F^2 = O_P(\frac{J^2}{p^3T}p\nu_p)).$$

Hence $\frac{1}{T^2}(\|\mathbf{F}'\mathbf{A}_2\|_F^2 + \|\mathbf{F}'\mathbf{A}_3\|_F^2 + \|\mathbf{F}'\mathbf{A}_8\|_F^2 + \|\mathbf{F}'\mathbf{A}_{12}\|_F^2) = O_P(\frac{1}{pT})$. Combining all these terms, we obtain

$$\frac{1}{T^2}\|\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})\|_F^2 = O_P(1)\sum_{i=1}^{15}\frac{1}{T^2}\|\widehat{\mathbf{F}}'\mathbf{A}_i\|_F^2 = O_P(\frac{1}{p^2} + \frac{1}{pT} + \frac{1}{J^\kappa} + \frac{\nu_p}{p}).$$

(ii) The result follows from the following inequality:

$$\frac{1}{T}\|\widehat{\mathbf{F}}'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})\|_F \le \frac{1}{T}\|\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H}\|_F^2 + \frac{1}{T}\|\mathbf{H}'\mathbf{F}'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})\|_F.$$

□

LEMMA C.4. $\|\mathbf{H}'\mathbf{H}-\mathbf{I}_K\|_F = O_P(\frac{1}{p}+\frac{1}{\sqrt{pT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\nu_p}{p}})$. *Therefore* $\|\mathbf{H}^{-1}\|_2 = O_P(1)$.

PROOF. Note that by the identification condition, $\mathbf{F}'\mathbf{F} = \widehat{\mathbf{F}}'\widehat{\mathbf{F}} = T\mathbf{I}_K$. Hence

$$\mathbf{H}'\mathbf{H} = \frac{1}{T}(\mathbf{F}\mathbf{H})'\mathbf{F}\mathbf{H} = \frac{1}{T}(\mathbf{F}\mathbf{H}-\widehat{\mathbf{F}})'\mathbf{F}\mathbf{H} + \frac{1}{T}\widehat{\mathbf{F}}'(\mathbf{F}\mathbf{H}-\widehat{\mathbf{F}}) + \mathbf{I}_K.$$

This identity implies $\|\mathbf{H}'\mathbf{H}-\mathbf{I}_K\|_F \le \frac{1}{T}\|(\mathbf{F}\mathbf{H}-\widehat{\mathbf{F}})'\mathbf{F}\|_F\|\mathbf{H}\|_2 + \frac{1}{T}\|\widehat{\mathbf{F}}'(\mathbf{F}\mathbf{H}-\widehat{\mathbf{F}})\|_F$, which yields the convergence rate. In addition, it implies $\lambda_{\min}(\mathbf{H}'\mathbf{H}) \ge 1-o_P(1)$. Therefore,

$$\|\mathbf{H}^{-1}\|_2^2 = \lambda_{\max}(\mathbf{H}^{-1}(\mathbf{H}^{-1})') = \lambda_{\max}((\mathbf{H}'\mathbf{H})^{-1}) = \lambda_{\min}^{-1}(\mathbf{H}'\mathbf{H}) = O_P(1).$$

□

C.3. Convergence of loadings. Define $\widehat{\mathbf{B}} = \frac{1}{T}[\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1}\Phi(\mathbf{X})'\mathbf{Y}\widehat{\mathbf{F}}$. Then

$$\widehat{\mathbf{G}}(\mathbf{X}) = \frac{1}{T}\mathbf{P}\mathbf{Y}\widehat{\mathbf{F}} = \Phi(\mathbf{X})\widehat{\mathbf{B}}.$$

Substituting $\mathbf{Y} = \Phi(\mathbf{X})\mathbf{B}\mathbf{F}' + \mathbf{\Gamma}\mathbf{F}' + \mathbf{R}(\mathbf{X})\mathbf{F}' + \mathbf{U}$, and using $\frac{1}{T}\mathbf{F}'\mathbf{F} = \mathbf{I}$,

$$\widehat{\mathbf{B}} = \mathbf{B}\mathbf{H} + \sum_{i=1}^{5}\mathbf{C}_i,$$

where

$$\begin{aligned}
\mathbf{C}_1 &= \frac{1}{T}[\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1}\Phi(\mathbf{X})'\mathbf{R}(\mathbf{X})\mathbf{F}'\widehat{\mathbf{F}}, \quad \mathbf{C}_2 = \frac{1}{T}[\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1}\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\mathbf{H}\\
\mathbf{C}_3 &= \frac{1}{T}[\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1}\Phi(\mathbf{X})'\mathbf{U}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H}), \quad \mathbf{C}_4 = \frac{1}{T}\mathbf{B}\mathbf{F}'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H}),\\
\mathbf{C}_5 &= \frac{1}{T}[\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1}\Phi(\mathbf{X})'\mathbf{\Gamma}\mathbf{F}'\widehat{\mathbf{F}}.
\end{aligned}$$

The proof is divided into two steps: bounding $\|\widehat{\mathbf{B}}-\mathbf{B}\mathbf{H}\|_F^2$ and bounding $\|\mathbf{H}-\mathbf{I}\|^2$. In particular, $\|\widehat{\mathbf{B}}-\mathbf{B}\mathbf{H}\|_F^2 \le C\sum_{i=1}^5\|\mathbf{C}_i\|_F^2$ for some $C>0$. Each individual term is bounded as below:

PROPOSITION C.2. *(i)* $\|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}\|_F^2 = O_P(\frac{J}{p^2} + \frac{J}{pT} + \frac{J}{J^\kappa} + \frac{J\nu_p}{p})$.
(ii) $\frac{1}{p}\|\widehat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})\mathbf{H}\|_F^2 = O_P(\frac{J}{p^2} + \frac{J}{pT} + \frac{J}{J^\kappa} + \frac{J\nu_p}{p})$.
(iii) $\frac{1}{p}\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\mathbf{H}\|^2 = O_P(\frac{1}{J^\kappa} + \frac{J\nu_p}{p} + \frac{J}{p^2} + \frac{1}{T})$.

PROOF. (i) By Lemmas B.1, C.6 and C.3,

$$\|\mathbf{C}_1\|_F^2 = O_P(\frac{1}{J^\kappa}), \quad \|\mathbf{C}_2\|_F^2 = O_P(\frac{J}{Tp}), \quad \|\mathbf{C}_3\|_F^2 = O_P(\frac{J}{p^2} + \frac{J}{pJ^\kappa}),$$
$$\|\mathbf{C}_4\|_F^2 = O_P(\frac{J}{p^2} + \frac{J}{pT} + \frac{J}{J^\kappa} + \frac{J\nu_p}{p}), \quad \|\mathbf{C}_5\|_F^2 = O_P(\frac{J\nu_p}{p}).$$

Therefore, $\|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}\|_F^2 \le O(1)\sum_{i=1}^5 \|\mathbf{C}_i\|_F^2 = O_P(\frac{J}{p^2} + \frac{J}{pT} + \frac{J}{J^\kappa} + \frac{J\nu_p}{p})$.
(ii) Because $\mathbf{G}(\mathbf{X})\mathbf{H} = \Phi(\mathbf{X})\mathbf{B}\mathbf{H} + \mathbf{R}(\mathbf{X})\mathbf{H}$, the result follows from

$$\frac{1}{p}\|\widehat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})\mathbf{H}\|_F^2 \le \frac{2}{p}\|\Phi(\mathbf{X})(\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H})\|_F^2 + \frac{2}{p}\|\mathbf{R}(\mathbf{X})\mathbf{H}\|_F^2 = O_P(\|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}\|_F^2 + J^{-\kappa}).$$

(iii) Substituting $\mathbf{Y} = \Phi(\mathbf{X})\mathbf{B}\mathbf{F}' + \mathbf{\Gamma}\mathbf{F}' + \mathbf{R}(\mathbf{X})\mathbf{F}' + \mathbf{U}$ into $\widehat{\mathbf{\Gamma}} = \frac{1}{T}(\mathbf{I} - \mathbf{P})\mathbf{Y}\widehat{\mathbf{F}}$,

$$\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\mathbf{H} = \sum_{i=1}^{6} \mathbf{D}_i$$

where

$$\begin{aligned}
\mathbf{D}_1 &= \frac{1}{T}(\mathbf{I} - \mathbf{P})\mathbf{\Gamma}\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}), \quad \mathbf{D}_2 = \frac{1}{T}\mathbf{U}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}),\\
\mathbf{D}_3 &= -\mathbf{P}\mathbf{\Gamma}\mathbf{H}, \quad \mathbf{D}_4 = (\mathbf{I} - \mathbf{P})\mathbf{R}(\mathbf{X})(\mathbf{H} + \frac{1}{T}\mathbf{F}'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})),\\
\mathbf{D}_5 &= -\frac{1}{T}\mathbf{P}\mathbf{U}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H}), \quad \mathbf{D}_6 = \frac{1}{T}(\mathbf{I} - \mathbf{P})\mathbf{U}\mathbf{F}\mathbf{H}.
\end{aligned}$$

The result then follows from Lemma C.10. □

C.4. Convergence of H.

PROPOSITION C.3. *As $p, T \to \infty$, while T may either grow with p or stay constant,*

$$\|\mathbf{H} - \mathbf{I}_K\|_F = O_P\Big(\frac{1}{p} + \frac{1}{\sqrt{pT}} + \frac{1}{J^{\kappa/2}} + \sqrt{\frac{\nu_p}{p}}\Big)$$

PROOF. Let $\delta_{T,p} = \frac{1}{p} + \frac{1}{\sqrt{pT}} + \frac{1}{J^{\kappa/2}} + \sqrt{\frac{\nu_p}{p}}$. Then by Lemma C.3,

$$\|\frac{1}{T}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})'\mathbf{F}\|_F = O_P(\delta_{T,p}).$$

Hence $\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{F} = \frac{1}{T}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})'\mathbf{F} + \mathbf{H}' = \mathbf{H}' + O_P(\delta_{T,p})$. Therefore

$$\mathbf{H}' = \mathbf{K}^{-1}\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{F}\frac{1}{p}\mathbf{B}'\Phi(\mathbf{X})'\Phi(\mathbf{X})\mathbf{B} = \mathbf{K}^{-1}\mathbf{H}'\frac{1}{p}\mathbf{B}'\Phi(\mathbf{X})'\Phi(\mathbf{X})\mathbf{B} + O_P(\delta_{T,p}).$$

So we have $(\frac{1}{p}\mathbf{B}'\Phi(\mathbf{X})'\Phi(\mathbf{X})\mathbf{B})\mathbf{H} = \mathbf{H}\mathbf{K} + O_P(\delta_{T,p})$. Moreover, $\|\frac{1}{p}\mathbf{G}'\mathbf{G} - \frac{1}{p}\mathbf{B}'\Phi(\mathbf{X})'\Phi(\mathbf{X})\mathbf{B}\|_F = O_P(\frac{1}{J^{\kappa/2}})$. Hence we get

$$(\frac{1}{p}\mathbf{G}'\mathbf{G})\mathbf{H} = \mathbf{H}\mathbf{K} + O_P(\delta_{T,p}).$$

Also by Lemma C.4 in the supplementary material, $\|\mathbf{H}'\mathbf{H} - \mathbf{I}_K\|_F = O_P(\delta_{T,p})$.

Let the elements of $\mathbf{H}$ be h_{ij}, diagonal elements of $\frac{1}{p}\mathbf{G}'\mathbf{G}$ be g_k, diagonal elements of $\mathbf{K}$ be v_k. So in fact we have obtained

$$g_i h_{ij} = v_j h_{ij} + O_P(\delta_{T,p}) \text{ and } (\sum_{k=1}^{K} h_{ki}^2 - 1)^2 = O_P(\delta_{T,p}^2).$$

$\|\mathbf{H}'\mathbf{H} - \mathbf{I}_K\|_F = o_P(1)$ implies that h_{ii} is bounded away from zero with probability approaching one. Hence $g_j h_{jj} = v_j h_{jj} + O_P(\delta_{T,p})$ yields $v_j = g_j + O_P(\delta_{T,p})$. Hence $(g_i - g_j)h_{ij} = O_P(\delta_{T,p})$. We assume $|g_i - g_j|$ is bounded away from zero almost surely when $i \neq j$, thus $h_{ij} = O_P(\delta_{T,p})$ when $i \neq j$.

From the second equation, $\sum_{k=1}^{K} h_{ki}^2 = 1 + O_P(\delta_{T,p})$. Since $\sum_{k=1}^{K} h_{ki}^2 = h_{ii}^2 + O_P(\delta_{T,p}^2)$, we get $h_{ii}^2 = 1 + O_P(\delta_{T,p})$, so $h_{ii} = \pm 1 + O_P(\sqrt{\delta_{T,p}})$. From here, we can assume $h_{ii} = 1 + O_P(\sqrt{\delta_{T,p}})$, otherwise we can always multiply the corresponding columns of $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{G}}(\mathbf{X})$ by -1. Note that $h_{ii} - 1 = \frac{1}{2}(h_{ii}^2 - 1 - (h_{ii} - 1)^2) = O_P(\delta_{T,p})$. Finally,

$$\|\mathbf{H} - \mathbf{I}_K\|_F^2 = \sum_{i \neq j} h_{ij}^2 + \sum_{i=1}^{K}(h_{ii} - 1)^2 = O_P(\delta_{T,p}^2).$$

□

C.5. Proof of Theorem 4.1.

PROOF.

$$\|\mathbf{H} - \mathbf{I}_K\|_F^2 = O_P\Big(\frac{1}{p^2} + \frac{1}{pT} + \frac{1}{J^\kappa} + \frac{\nu_p}{p}\Big)$$

It follows from Propositions C.1, C.2 and C.3 and $J^2 = O(p)$ that

$$\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\|_F^2 = O_P\Big(\frac{1}{p} + \frac{1}{J^\kappa}\Big),$$

$$\frac{1}{p}\|\widehat{\mathbf{G}}(\mathbf{X})-\mathbf{G}(\mathbf{X})\|_F^2 = O_P\Big(\frac{J}{p^2}+\frac{J}{pT}+\frac{J}{J^\kappa}+\frac{J\nu_p}{p}\Big),$$
$$\|\widehat{\mathbf{B}}-\mathbf{B}\|_F^2 = O_P\Big(\frac{J}{p^2}+\frac{J}{pT}+\frac{J}{J^\kappa}+\frac{J\nu_p}{p}\Big),$$
$$\frac{1}{p}\|\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\|^2 = O_P\Big(\frac{1}{J^\kappa}+\frac{J\nu_p}{p}+\frac{J}{p^2}+\frac{1}{T}\Big). \tag{C.5}$$

This also implies, since obviously $\max_k\|\mathbf{b}_k-\widehat{\mathbf{b}}_k\|^2\le\|\widehat{\mathbf{B}}-\mathbf{B}\|_F^2$,

$$\begin{aligned}\max_{k\le K}\sup_{\mathbf{x}\in\mathcal{X}}|g_k(\mathbf{x})-\widehat{g}_k(\mathbf{x})| &\le \sup_{\mathbf{x}\in\mathcal{X}}\|\phi(\mathbf{x})\|\max_k\|\mathbf{b}_k-\widehat{\mathbf{b}}_k\|+\sup_{\mathbf{x}\in\mathcal{X}}|\sum_{l=1}^d R_{kl}(x_l)|\\ &= O_P\left(\frac{J^{1/2}}{p}+\frac{J^{1/2}}{\sqrt{pT}}+\frac{J^{1/2}}{J^{\kappa/2}}+\frac{J^{1/2}\nu_p^{1/2}}{p^{1/2}}\right)\sup_{\mathbf{x}\in\mathcal{X}}\|\phi(\mathbf{x})\|, \end{aligned}\tag{C.6}$$

$\square$

C.6. Further Technical Lemmas. In the lemma below, we prove that $\mathbf{K},\mathbf{K}^{-1}$ and $\mathbf{H}$ all have bounded spectral norms. It is proved in Lemma A.2 that in the general factor model, under Assumption 3.1, $\|\mathbf{K}^{-1}\|_2=O_P(1)$. We now prove that in the semi-parametric factor model with $\mathbf{\Lambda}=\mathbf{G}(\mathbf{X})+\mathbf{\Gamma}$, Assumption 3.1 holds. This then implies $\|\mathbf{K}^{-1}\|_2=O_P(1)$ in the semi-parametric factor model as well.

Lemma C.5. *$\|\mathbf{K}\|_2=O_P(1)$, $\|\mathbf{K}^{-1}\|_2=O_P(1)$ and $\|\mathbf{H}\|_2=O_P(1)$.*

Proof. Note that $\|\mathbf{Y}\|_F^2=O_P(pT)$, and $\|\mathbf{P}\|_2=O_P(1)$. It then follows immediately that $\|\mathbf{K}\|_2=O_P(1)$. To prove $\|\mathbf{K}^{-1}\|_2=O_P(1)$, note that the eigenvalues of $\mathbf{K}$ are the same as those of

$$\mathbf{W}=\frac{1}{Tp}(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}\Phi(\mathbf{X})'\mathbf{Y}\mathbf{Y}'\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2},$$

Write $\mathbf{\Lambda}=\mathbf{G}(\mathbf{X})+\mathbf{\Gamma}$. Substituting $\mathbf{Y}=\mathbf{\Lambda}\mathbf{F}'+\mathbf{U}$, and $\mathbf{F}'\mathbf{F}/T=\mathbf{I}_K$, we have $\mathbf{W}=\sum_{i=1}^4\mathbf{W}_i$, where

$$\begin{aligned}\mathbf{W}_1 &= \frac{1}{p}(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}\Phi(\mathbf{X})'\mathbf{\Lambda}\mathbf{\Lambda}'\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2},\\ \mathbf{W}_2 &= \frac{1}{p}(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}\Phi(\mathbf{X})'(\frac{\mathbf{\Lambda}\mathbf{F}\mathbf{U}'}{T})\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2},\\ \mathbf{W}_3 &= \mathbf{W}_2',\\ \mathbf{W}_4 &= \frac{1}{p}(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}\Phi(\mathbf{X})'\frac{\mathbf{U}\mathbf{U}'}{T}\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}.\end{aligned}$$

Due to Lemma B.1, $\|\mathbf{W}_2\|_2 = O_P(\frac{1}{pT})\|\mathbf{F}\mathbf{U}'\Phi(\mathbf{X})\|_F = O_P(\sqrt{J/(pT)})$, and $\|\mathbf{W}_4\|_2 = O_P(\frac{1}{p^2T})\|\Phi(\mathbf{X})'\mathbf{U}\|_F^2 = O_P(J/p)$. Therefore, for $k = 1, \cdots, K$,

$$|\lambda_k(\mathbf{W}) - \lambda_k(\mathbf{W}_1)| \leq \|\mathbf{W} - \mathbf{W}_1\|_2 = o_P(1).$$

Hence it suffices to prove that the first K eigenvalues of

$$\frac{1}{p}(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}\Phi(\mathbf{X})'\mathbf{\Lambda}\mathbf{\Lambda}'\Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1/2}$$

are bounded away from zero, which are also the first K eigenvalues of $\frac{1}{p}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda} = \frac{1}{p}(\mathbf{G}(\mathbf{X}) + \mathbf{\Gamma})'\mathbf{P}(\mathbf{G}(\mathbf{X}) + \mathbf{\Gamma})$. This was assumed in Assumption 3.1 in the general factor model. In the case $\mathbf{\Lambda} = \mathbf{G}(\mathbf{X}) + \mathbf{\Gamma}$, it can be formally justified. A more formal argument on this is stated and proved in Lemma F.1. We do not present it here since Lemma F.1 can also be used to justify Assumption 3.1. As a result, $\|\mathbf{K}^{-1}\|_2 = O_P(1)$.

We have, $\mathbf{H} = \frac{1}{Tp}(\mathbf{G}(\mathbf{X}) - \mathbf{R}(\mathbf{X}))'(\mathbf{G}(\mathbf{X}) - \mathbf{R}(\mathbf{X}))\mathbf{F}'\widehat{\mathbf{F}}\mathbf{K}^{-1}$. The result then follows from $\|\mathbf{K}^{-1}\|_2 = O_P(1)$ and $\|\mathbf{G}(\mathbf{X})\|_2 = O_P(\sqrt{p})$.

$\square$

Under the assumptions in Section 4, the following lemmas can be proved and are needed.

LEMMA C.6. *(i)* $\|\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F^2 = O_P(pT)$, $\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\|_F^2 = O_P(pT)$.
(ii) $\|\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2 = O_P(Jp\nu_p)$, $\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2 = O_P(p\nu_p)$, $\|\mathbf{P}\mathbf{\Gamma}\|_F^2 = O_P(J\nu_p)$, $\|\mathbf{\Gamma}\|_F^2 = O_P(p\nu_p)$.

PROOF. (iii) Note that $\mathbf{U}'\Phi(\mathbf{X})\mathbf{B} = \mathbf{U}'\mathbf{G}(\mathbf{X}) - \mathbf{U}'\mathbf{R}(\mathbf{X})$, and $\mathbf{X}_i$ and $\mathbf{u}_t$ are independent,

$$\begin{aligned} E\|\mathbf{U}'\mathbf{G}(\mathbf{X})\|_F^2 &= \sum_{t=1}^T\sum_{k=1}^K E(\sum_{i=1}^p g_k(\mathbf{X}_i)u_{it})^2 = \sum_{t=1}^T\sum_{k=1}^K\sum_{j=1}^p\sum_{i=1}^p Eg_k(\mathbf{X}_i)g_k(\mathbf{X}_j)Eu_{it}u_{jt} \\ &\leq pK\max_{k\leq K}Eg_k(\mathbf{X}_i)^2\sum_{t=1}^T\max_{j\leq p}\sum_{i=1}^p|Eu_{it}u_{jt}| = O(pT), \end{aligned}$$

where the inequality is due to the Cauchy Schwarz inequality.

Note that the (i,k) element of $\mathbf{R}(\mathbf{X})$ is $\sum_{l=1}^d R_{kl}(X_{il}) := R_{ik}$, so

$$E\|\mathbf{U}'\mathbf{R}(\mathbf{X})\|_F^2 = \sum_{t=1}^T\sum_{k=1}^K E(\sum_{i=1}^p R_{ik}u_{it})^2 = \sum_{t=1}^T\sum_{k=1}^K\sum_{j=1}^p\sum_{i=1}^p ER_{ik}R_{jk}Eu_{it}u_{jt}$$

$$\leq \quad pK \max_{k\leq K} ER_{ik}^2 \sum_{t=1}^{T} \max_{j\leq p} \sum_{i=1}^{p} |Eu_{it}u_{jt}| = O(pTJ^{-\kappa}),$$

where the inequality is due to the Cauchy Schwarz inequality and that $\max_{k\leq K} ER_{ik}^2 = O(J^{-\kappa})$. Therefore, $\|\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F^2 = O_P(pT)$.

On the other hand, $\mathbf{B}'\Phi(\mathbf{X})'\mathbf{UF} = \mathbf{G}(\mathbf{X})'\mathbf{UF} - \mathbf{R}(\mathbf{X})'\mathbf{UF}$. Here,

$$\begin{aligned}
E\|\mathbf{G}(\mathbf{X})'\mathbf{UF}\|_F^2 &= \sum_{k=1}^{K}\sum_{l=1}^{K} E(\sum_{i=1}^{p}\sum_{t=1}^{T} g_k(\mathbf{X}_i)u_{it}f_{tl})^2 = \sum_{k=1}^{K}\sum_{l=1}^{K} \mathrm{var}(\sum_{i=1}^{p}\sum_{t=1}^{T} g_k(\mathbf{X}_i)u_{it}f_{tl}) \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i=1}^{p} \mathrm{var}(\sum_{t=1}^{T} g_k(\mathbf{X}_i)u_{it}f_{tl}) + \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i\neq j; i,j\leq p} \mathrm{cov}(\sum_{t=1}^{T} g_k(\mathbf{X}_i)u_{it}f_{tl}, \sum_{t=1}^{T} g_k(\mathbf{X}_j)u_{jt}f_{tl}) \\
&= D_1 + D_2, \text{ say.}
\end{aligned}$$

Here, (note that $\mathrm{var}(g_k(\mathbf{X}_i)u_{it}f_{tl})$ and $|Eg_k(\mathbf{X}_i)^2 f_{tl}f_{sl}|$ are bounded uniformly in $k \leq K, l \leq K, t \leq T, s \leq T$)

$$\begin{aligned}
D_1 &= \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i=1}^{p}\sum_{t=1}^{T} \mathrm{var}(g_k(\mathbf{X}_i)u_{it}f_{tl}) + \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i=1}^{p}\sum_{t\neq s; t,s\leq T} \mathrm{cov}(g_k(\mathbf{X}_i)u_{it}f_{tl}, g_k(\mathbf{X}_i)u_{is}f_{sl}) \\
&= O(pT) + \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i=1}^{p}\sum_{t\neq s; t,s\leq T} E(g_k(\mathbf{X}_i)^2 f_{tl}f_{sl})Eu_{it}u_{is} \\
&\leq O(pT) + pTK^2 \max_{k\leq K, l\leq K, t\leq T, s\leq T} |Eg_k(\mathbf{X}_i)^2 f_{tl}f_{sl}| \max_{t\leq T}\max_{i\leq p}\sum_{s=1}^{T} |Eu_{it}u_{is}| = O(pT). \\
D_2 &= \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i\neq j; i,j\leq p}\sum_{t=1}^{T}\sum_{s=1}^{T} E(g_k(\mathbf{X}_i)f_{tl}g_k(\mathbf{X}_j)f_{sl})Eu_{js}u_{it} \\
&\leq K^2 \max_{k\leq K; i,j\leq p; t,s\leq T} |E(g_k(\mathbf{X}_i)f_{tl}g_k(\mathbf{X}_j)f_{sl})| \sum_{i,j\leq p}\sum_{s,t\leq T} |Eu_{it}u_{js}| = O(pT),
\end{aligned}$$

where the last equality follows from $\sum_{i,j\leq p}\sum_{s,t\leq T} |Eu_{it}u_{js}| = O(pT)$. Hence we obtain $\|\mathbf{G}(\mathbf{X})'\mathbf{UF}\|_F^2 = O_P(pT)$. $\|\mathbf{R}(\mathbf{X})'\mathbf{UF}\|_F^2$ can be bounded in the same way. This yields

$$\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{UF}\|_F^2 = O_P(pT).$$

(ii) Because X_{il} and γ_{ik} are independent and $E\gamma_{ik} = 0$,

$$E\|\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2 = \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{d} E(\sum_{i=1}^{p} \phi_j(X_{il})\gamma_{ik})^2 = \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{d} \mathrm{var}(\sum_{i=1}^{p} \phi_j(X_{il})\gamma_{ik})$$

$$
\begin{aligned}
&= \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{d}\sum_{i=1}^{p}\mathrm{var}(\phi_j(X_{il})\gamma_{ik}) + \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{d}\sum_{i\neq s;i,s\leq p}\mathrm{cov}(\phi_j(X_{il})\gamma_{ik},\phi_j(X_{sl})\gamma_{sk})\\
&= \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{d}\sum_{i=1}^{p}E\phi_j(X_{il})^2E\gamma_{ik}^2 + \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{d}\sum_{i\neq s;i,s\leq p}E\phi_j(X_{il})\phi_j(X_{sl})E\gamma_{ik}\gamma_{sk}\\
&= O(Jp\nu_p) + O(Jp)\max_k\max_{s\leq p}\sum_{i\leq p}|E\gamma_{ik}\gamma_{sk}| = O(Jp\nu_p).
\end{aligned}
$$

Also, $\mathbf{B}'\Phi(\mathbf{X})'\mathbf{\Gamma} = \mathbf{G}(\mathbf{X})'\mathbf{\Gamma} - \mathbf{R}(\mathbf{X})'\mathbf{\Gamma}$. Note that

$$
E\|\mathbf{G}(\mathbf{X})'\mathbf{\Gamma}\|_F^2 = \sum_{k=1}^{K}\sum_{l=1}^{K}E(\sum_{i=1}^{p}g_l(\mathbf{X}_i)\gamma_{ik})^2 = O(p\nu_p),
$$

and $E\|\mathbf{R}(\mathbf{X})'\mathbf{\Gamma}\|_F^2 = O_p(J^{-\kappa}p\nu_p)$. Hence $\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2 = O(p\nu_p)$. □

LEMMA C.7. *Let $\mathbf{A}_i$ be as in (C.1) for $i = 1, ..., 15$, then*
(i) $\|\frac{1}{T}\mathbf{U}\mathbf{A}_1\|_F^2 = O_P(\frac{1}{Tp} + \frac{1}{TJ^\kappa} + \frac{1}{T^2})$, $\|\frac{1}{T}\mathbf{U}\mathbf{A}_{15}\|_F^2 = O_P(\frac{J^2\nu_p^2}{Tp})$,
(ii) $\|\frac{1}{T}\mathbf{U}\mathbf{A}_4\|_F^2 = O_P(\frac{p}{J^\kappa T}) = \|\frac{1}{T}\mathbf{U}\mathbf{A}_5\|_F^2$, $\|\frac{1}{T}\mathbf{U}\mathbf{A}_6\|_F^2 = O_P(\frac{p}{J^{2\kappa}T})$
(iii) $\|\frac{1}{T}\mathbf{U}\mathbf{A}_7\|_F^2 = O_P(\frac{J}{J^\kappa Tp} + \frac{J}{J^\kappa T^2} + \frac{J}{J^{2\kappa}T})$
(iv) $\|\frac{1}{T}\mathbf{U}\mathbf{A}_9\|_F^2 = O_P(\frac{\nu_p}{T}) = \|\frac{1}{T}\mathbf{U}\mathbf{A}_{10}\|_F^2$, $\|\frac{1}{T}\mathbf{U}\mathbf{A}_{13}\|_F^2 = O_P(\frac{J\nu_p}{TJ^\kappa}) = \|\frac{1}{T}\mathbf{U}\mathbf{A}_{14}\|_F^2$,
(v) $\|\frac{1}{T}\mathbf{U}\mathbf{A}_{11}\|_F^2 = O_P(\frac{J^2\nu_p}{Tp^2} + \frac{J^2\nu_p}{T^2p} + \frac{J^2\nu_p}{TpJ^\kappa})$.

PROOF. Given (C.3) and (C.4), the proof is straightforward calculations. Summing of all the terms of Lemma C.7 gives $O_P(\frac{1}{Tp} + \frac{1}{T^2} + \frac{p}{J^\kappa T} + \frac{\nu_p}{T})$. □

LEMMA C.8. $\|\mathbf{U}\mathbf{U}'\Phi(\mathbf{X})\|_F^2 = O_P(pT^2 + Tp^2J)$,
and $\|\mathbf{U}\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F^2 = O_P(pT^2 + Tp^2)$.

PROOF. Note that $\|E\mathbf{U}\mathbf{U}'\Phi(\mathbf{X})\|_F^2 = O_P(pT^2)$. On the other hand, let $s_{ik} = \sum_{t=1}^{T}(u_{it}u_{kt} - Eu_{it}u_{kt})$

$$
\begin{aligned}
&E\|(\mathbf{U}\mathbf{U}' - E\mathbf{U}\mathbf{U}')\Phi(\mathbf{X})\|_F^2 = \sum_{i=1}^{p}\sum_{l=1}^{d}\sum_{j=1}^{J}\mathrm{var}(\sum_{k=1}^{p}\phi_j(X_{kl})s_{ik})\\
&= \sum_{i=1}^{p}\sum_{l=1}^{d}\sum_{j=1}^{J}\sum_{k=1}^{p}E\phi_j(X_{kl})^2Es_{ik}^2 + \sum_{i=1}^{p}\sum_{l=1}^{d}\sum_{j=1}^{J}\sum_{k,m\leq p,k\neq m}E\phi_j(X_{kl})\phi_j(X_{ml})\mathrm{cov}(s_{ik},s_{im})\\
&= \sum_{i=1}^{p}\sum_{l=1}^{d}\sum_{j=1}^{J}\sum_{k=1}^{p}E\phi_j(X_{kl})^2\mathrm{var}(\sum_{t=1}^{T}u_{it}u_{kt})
\end{aligned}
$$

$$+\sum_{i=1}^{p}\sum_{l=1}^{d}\sum_{j=1}^{J}\sum_{k,m\leq p,k\neq m}E\phi_j(X_{kl})\phi_j(X_{ml})\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{cov}(u_{it}u_{kt},u_{is}u_{ms})$$
$$=O(JTp^2),$$

by the assumption that $\max_{i\leq p}\sum_{k=1}^{p}\sum_{m=1}^{p}\sum_{t=1}^{T}\sum_{s=1}^{T}|\mathrm{cov}(u_{it}u_{kt},u_{is}u_{ms})| = O(Tp)$. □

LEMMA C.9. *(i)* $\|\frac{1}{T}\mathbf{U}\mathbf{A}_2\|_F^2 = O_P(\frac{1}{T}+\frac{1}{p})$,
(ii) $\|\frac{1}{T}\mathbf{U}\mathbf{A}_3\|_F^2 = O_P(\frac{J}{p^3}+\frac{J}{p^2J^\kappa}+\frac{J^2}{Tp^2}+\frac{J^2}{T^2p}+\frac{J^2}{TpJ^\kappa})$
(iii) $\|\frac{1}{T}\mathbf{U}\mathbf{A}_8\|_F^2 = O_P(\frac{1}{J^\kappa p}+\frac{J}{J^\kappa T})$, $\|\frac{1}{T}\mathbf{U}\mathbf{A}_{12}\|_F^2 = O_P(\frac{J\nu_p}{p^2}+\frac{J^2\nu_p}{pT})$.

PROOF. Given Lemma C.8, the proof is straightforward calculations. Summing all the terms of Lemma C.7 gives $O_P(\frac{1}{T}+\frac{1}{p})$. □

LEMMA C.10. *(i)* $\|\frac{1}{T}(\mathbf{I}-\mathbf{P})\mathbf{\Gamma}\mathbf{F}'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})\|_F^2 = O_P(\nu_p(\frac{1}{p}+\frac{1}{T}+\frac{p}{J^\kappa}+\nu_p))$,
(ii) $\|\frac{1}{T}\mathbf{U}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})\|_F^2 = O_P(\frac{p}{J^\kappa T}+\frac{1}{T}+\frac{1}{p})$,
(iii) $\|\mathbf{P}\mathbf{\Gamma}\mathbf{H}\|_F^2 = O_P(J\nu_p)$,
(iv) $\|(\mathbf{I}-\mathbf{P})\mathbf{R}(\mathbf{X})(\mathbf{H}+\frac{1}{T}\mathbf{F}'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H}))\|_F^2 = O_P(\frac{p}{J^\kappa})$,
(v) $\|\frac{1}{T}\mathbf{P}\mathbf{U}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})\|_F^2 = O_P(\frac{J}{p}+\frac{J}{J^\kappa})$,
(vi) $\|\frac{1}{T}(\mathbf{I}-\mathbf{P})\mathbf{U}\mathbf{F}\mathbf{H}\|_F^2 = O_P(\frac{p}{T})$.

PROOF. (i), (iii)-(v) follow from Lemmas C.6 and C.3. (vi) follows from Cauchy-Schwarz inequality. (ii) follows from Lemmas C.7 and C.9.

□

APPENDIX D: PROOFS FOR SECTION 5

D.1. Proof of Theorem 5.1: Limiting distribution for S_Γ. We first prove the limiting distribution of the standardized S_Γ under H_0^2. One of the key steps is to show that estimating $\mathbf{\Sigma}_u^{-1}$ by $\widehat{\mathbf{\Sigma}}_u^{-1}$ does not affect the asymptotic behavior of the leading term, that is,

$$\mathrm{tr}\Big((\frac{1}{T}\mathbf{U}\mathbf{F})'(\widehat{\mathbf{\Sigma}}_u^{-1}-\mathbf{\Sigma}_u^{-1})(\frac{1}{T}\mathbf{U}\mathbf{F})\Big) = o_P\Big(\frac{\sqrt{p}}{T}\Big). \tag{D.1}$$

However, since p can be either larger or comparable with T, the above result cannot be simply implied from the crude bounds like

$$\|\frac{1}{T}\mathbf{U}\mathbf{F}\|_{\max}\|\frac{1}{T}\mathbf{U}\mathbf{F}\|_1\|\widehat{\mathbf{\Sigma}}_u^{-1}-\mathbf{\Sigma}_u^{-1}\|_1, \quad \|\frac{1}{T}\mathbf{U}\mathbf{F}\|_F^2\|\widehat{\mathbf{\Sigma}}_u^{-1}-\mathbf{\Sigma}_u^{-1}\|_2,$$

even when $\boldsymbol{\Sigma}_u^{-1}, \widehat{\boldsymbol{\Sigma}}_u^{-1}$ are diagonal matrices. As we shall see in Proposition D.2 below, proving (D.1) is highly technically involved. We need to require $p = o(T^2)$, but still allow $p/T \to \infty$.

PROPOSITION D.1. *If p, T, J satisfy $\max\{T^{2/3}, JT^{1/2}, J^2\} = o(p)$ and $\max\{Tp^{1/2}, p\} = o(J^\kappa)$, we have*

$$\mathrm{tr}\Big(\widehat{\boldsymbol{\Lambda}}'(\mathbf{I}-\mathbf{P})\widehat{\boldsymbol{\Sigma}}_u^{-1}(\mathbf{I}-\mathbf{P})\widehat{\boldsymbol{\Lambda}}\Big) = \frac{1}{T^2}\mathrm{tr}\Big(\mathbf{F}'\mathbf{U}'\widehat{\boldsymbol{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}\Big) + o_P\Big(\frac{\sqrt{p}}{T}\Big).$$

PROOF. PROOF. We have,

$$\begin{aligned}
&\mathrm{tr}(\widehat{\boldsymbol{\Lambda}}'(\mathbf{I}-\mathbf{P})\widehat{\boldsymbol{\Sigma}}_u^{-1}(\mathbf{I}-\mathbf{P})\widehat{\boldsymbol{\Lambda}}) \\
&= \mathrm{tr}((\sum_{i=1}^6 \mathbf{D}_i + \frac{1}{T}\mathbf{U}\mathbf{F})'\widehat{\boldsymbol{\Sigma}}_u^{-1}(\sum_{i=1}^6 \mathbf{D}_i + \frac{1}{T}\mathbf{U}\mathbf{F})) \\
&= \mathrm{tr}((\frac{1}{T}\mathbf{U}\mathbf{F})'\widehat{\boldsymbol{\Sigma}}_u^{-1}(\frac{1}{T}\mathbf{U}\mathbf{F})) + \mathrm{tr}((\sum_{i=1}^6 \mathbf{D}_i)'\widehat{\boldsymbol{\Sigma}}_u^{-1}(\sum_{i=1}^6 \mathbf{D}_i)) + 2\mathrm{tr}((\sum_{i=1}^6 \mathbf{D}_i)'\widehat{\boldsymbol{\Sigma}}_u^{-1}\frac{1}{T}\mathbf{U}\mathbf{F}) \\
&= S_1 + S_2 + 2S_3.
\end{aligned}$$

where we define

$$\mathbf{D}_1 = \frac{1}{T}\mathbf{G}\mathbf{F}'(\widehat{\mathbf{F}}-\mathbf{F}), \mathbf{D}_2 = \frac{1}{T}\mathbf{U}(\widehat{\mathbf{F}}-\mathbf{F}),$$

$$\mathbf{D}_3 = -\frac{1}{T}\mathbf{P}\mathbf{G}\mathbf{F}'(\widehat{\mathbf{F}}-\mathbf{F}), \mathbf{D}_4 = -\frac{1}{T}\mathbf{P}\mathbf{U}(\widehat{\mathbf{F}}-\mathbf{F}),$$

$$\mathbf{D}_5 = (\mathbf{I}-\mathbf{P})\mathbf{R}, \mathbf{D}_6 = -\frac{1}{T}\mathbf{P}\mathbf{U}\mathbf{F}.$$

Therefore, it suffices to show $S_2, S_3 = o_P(\sqrt{p}/T)$.

Using the results in Appendix C, we have the following results:

$$\begin{aligned}
\|\mathbf{D}_1 + \mathbf{D}_3\|_F &\le \|\frac{1}{T}(\mathbf{I}-\mathbf{P})\mathbf{G}\mathbf{F}'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})\|_F + \|\frac{1}{T}(\mathbf{I}-\mathbf{P})\mathbf{G}\mathbf{F}'\mathbf{F}(\mathbf{H}-\mathbf{I})\|_F \\
&= O_P(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p}} + \sqrt{\frac{p}{J^\kappa}}),
\end{aligned}$$

where $J = o(\sqrt{p})$ and $\|\mathbf{G}\|_2 = O_P(\sqrt{p})$ by assumption and $\|\mathbf{I}-\mathbf{P}\|_2 = O_P(1)$;

$$\begin{aligned}
\|\mathbf{D}_2 + \mathbf{D}_4\|_F &\le \|\frac{1}{T}(\mathbf{I}-\mathbf{P})\mathbf{U}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})\|_F + \|\frac{1}{T}(\mathbf{I}-\mathbf{P})\mathbf{U}\mathbf{F}(\mathbf{H}-\mathbf{I})\|_F \\
&= O_P(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p}} + \sqrt{\frac{1}{J^\kappa}} + \sqrt{\frac{p}{TJ^\kappa}}),
\end{aligned}$$

which follows from Lemma C.10.

$$\|\mathbf{D}_5\|_F \le \|\mathbf{I}-\mathbf{P}\|_2\|\mathbf{R}\|_F = O_P(\sqrt{\frac{p}{J^\kappa}});$$

$$\|\mathbf{D}_6\|_F \le O_P(\|\frac{1}{T\sqrt{p}}\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\|_F) = O_P(\sqrt{\frac{J}{T}}).$$

Hence,

$$S_2 = O_P(\|\sum_{i=1}^{6}\mathbf{D}_i\|_F^2) = O_P(\frac{J}{T}+\frac{1}{p}+\frac{p}{J^\kappa}).$$

So $S_2 = o_P(\sqrt{p}/T)$ if $J = o(\sqrt{p})$, $T^2 = o(p^3)$ and $T\sqrt{p} = o(J^\kappa)$. The three conditions are satisfied by our assumption of the dimension regime. Then let us work on S_3.

$$S_3 = \sum_{i=1}^{6}\mathrm{tr}(\frac{1}{T}\mathbf{D}_i'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}).$$

We need to bound S_3 term by term, and lemmas D.1-D.5 in the supplementary material are used. Note that under the assumption $J = o(\sqrt{p})$, rate results of in those lemmas could be written in a much simpler form. Lemmas D.1 and D.3 imply the following results.

$$\begin{aligned}\mathrm{tr}(\frac{1}{T}\mathbf{D}_1'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}) &= \frac{1}{T^2}\mathrm{tr}((\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{F}\mathbf{G}'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}) \le \frac{1}{T^2}\|(\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{F}\|_F\|\mathbf{G}'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}\|_F \\ &= O_P(\frac{1}{T}+\frac{1}{\sqrt{pT}}+\sqrt{\frac{p}{TJ^\kappa}}).\end{aligned}$$

$$\begin{aligned}\mathrm{tr}(\frac{1}{T}\mathbf{D}_3'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}) &= \frac{1}{T^2}\mathrm{tr}((\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{F}\mathbf{G}'\mathbf{P}\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}) \le \frac{1}{T^2}\|(\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{F}\|_F\|\mathbf{G}'\mathbf{P}\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}\|_F \\ &\le \frac{1}{T^2}\|(\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{F}\|_F(\|\mathbf{B}'\Phi(\mathbf{X})'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}\|_F + \|\mathbf{R}'\|_F\|\mathbf{P}\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}\|_F) \\ &= O_P(\frac{1}{T}+\frac{1}{\sqrt{pT}}+\sqrt{\frac{p}{TJ^\kappa}}).\end{aligned}$$

$$\begin{aligned}\mathrm{tr}(\frac{1}{T}\mathbf{D}_4'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}) &= \frac{1}{T^2}\mathrm{tr}((\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{U}\mathbf{P}\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}) \\ &\le \frac{1}{T^2p}\|(\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{U}\Phi(\mathbf{X})\|_F\|[\frac{1}{p}\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1}\|\|\Phi(\mathbf{X})'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{U}\mathbf{F}\|_F \\ &= \frac{1}{T^2p}O_P(\sqrt{J}T+\sqrt{\frac{pT^2J}{J^\kappa}})O_P(\sqrt{pTJ}) = O_P(\frac{J}{\sqrt{pT}}+\frac{J}{\sqrt{TJ^\kappa}}).\end{aligned}$$

$$
\begin{aligned}
\operatorname{tr}(\frac{1}{T}\mathbf{D}_5'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}) &= \frac{1}{T}\operatorname{tr}(\mathbf{R}'(\mathbf{I}-\mathbf{P})\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}) \le \frac{\sqrt{K}}{T}(\|\mathbf{R}'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}\|_F + \|\mathbf{R}'\mathbf{P}\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}\|_F) \\
&\le \frac{\sqrt{K}}{T}(\|\mathbf{R}'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}\|_F \\
&+ \|\mathbf{R}'\|_F\|\frac{1}{\sqrt{p}}\Phi(\mathbf{X})\|\|[\frac{1}{p}\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1}\|\|\frac{1}{\sqrt{p}}\Phi(\mathbf{X})'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}\|_F) \\
&= \frac{1}{T}O_P(\sqrt{pTJ^{-\kappa}} + \sqrt{pJ^{-\kappa}}\sqrt{TJ}) = O_P(\sqrt{\frac{pJ}{TJ^{\kappa}}}).
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{tr}(\frac{1}{T}\mathbf{D}_6'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}) &= \frac{1}{T^2}\operatorname{tr}(\mathbf{F}'\mathbf{U}'\mathbf{P}\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}) \le \frac{1}{T^2 p}\|[\frac{1}{p}\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1}\|\|\Phi(\mathbf{X})'\mathbf{UF}\|_F\|\Phi(\mathbf{X})'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}\|_F \\
&= \frac{1}{T^2 p}O_P(pTJ) = O_P(\frac{J}{T}).
\end{aligned}
$$

Lemmas D.4 (iii) and D.5 (iv) in the supplementary material imply the following results.

$$
\begin{aligned}
\operatorname{tr}(\frac{1}{T}\mathbf{D}_2'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}) &= \frac{1}{T^2}\operatorname{tr}((\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{U}'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}) \le \frac{1}{T^2}\sqrt{K}\|(\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{U}'\widehat{\mathbf{\Sigma}}_u^{-1}\mathbf{UF}\|_F \\
&= O_P(\frac{1}{T} + \frac{1}{\sqrt{pT}} + \sqrt{\frac{p}{T^3}} + \sqrt{\frac{p^2}{T^2J^{\kappa}}}) + o_P(\frac{\sqrt{p}}{T}).
\end{aligned}
$$

Combining the above 6 terms together, we derive

$$
S_3 = O_P(\frac{J}{T} + \frac{J}{\sqrt{pT}} + \sqrt{\frac{p}{T^3}} + \sqrt{\frac{pJ}{TJ^{\kappa}}} + \sqrt{\frac{p^2}{T^2J^{\kappa}}}) + o_P(\frac{\sqrt{p}}{T}) = o_P(\frac{\sqrt{p}}{T}),
$$

if $J = o(\sqrt{p})$, $TJ^2 = O(p^2)$ and $\max\{JT, p\} = o(J^{\kappa})$. These three conditions are again satisfied by our assumption of the dimension regime. $\square$

PROPOSITION D.2. *Suppose $p\log^4 p = o(T^2)$, $J\log^3 p = o(T)$ and $p\log^2 p = o(J^{2\kappa})$. Then*

$$
\operatorname{tr}\Big((\frac{1}{T}\mathbf{UF})'(\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})(\frac{1}{T}\mathbf{UF})\Big) = o_P\Big(\frac{\sqrt{p}}{T}\Big).
$$

PROOF. We have, $\operatorname{tr}((\frac{1}{T}\mathbf{UF})'(\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})(\frac{1}{T}\mathbf{UF}))$ equals

$$
\begin{aligned}
& \operatorname{tr}((\frac{1}{T}\mathbf{UF})'\widehat{\mathbf{\Sigma}}_u^{-1}(\mathbf{\Sigma}_u - \widehat{\mathbf{\Sigma}}_u)\mathbf{\Sigma}_u^{-1}(\frac{1}{T}\mathbf{UF})) \\
\le\ & \operatorname{tr}((\frac{1}{T}\mathbf{UF})'(\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})(\mathbf{\Sigma}_u - \widehat{\mathbf{\Sigma}}_u)\mathbf{\Sigma}_u^{-1}(\frac{1}{T}\mathbf{UF}))
\end{aligned}
$$

$$
\begin{aligned}
&+\mathrm{tr}((\frac{1}{T}\mathbf{U}\mathbf{F})'\mathbf{\Sigma}_u^{-1}(\mathbf{\Sigma}_u-\widehat{\mathbf{\Sigma}}_u)\mathbf{\Sigma}_u^{-1}(\frac{1}{T}\mathbf{U}\mathbf{F}))\\
\leq\ &\|\frac{1}{T}\mathbf{U}\mathbf{F}\|_F^2\|\mathbf{\Sigma}_u^{-1}\|_2\|\mathbf{\Sigma}_u^{-1}-\widehat{\mathbf{\Sigma}}_u^{-1}\|_2\|\mathbf{\Sigma}_u-\widehat{\mathbf{\Sigma}}_u\|_2\\
&+\mathrm{tr}((\frac{1}{T}\mathbf{U}\mathbf{F})'\mathbf{\Sigma}_u^{-1}(\mathbf{\Sigma}_u-\widehat{\mathbf{\Sigma}}_u)\mathbf{\Sigma}_u^{-1}(\frac{1}{T}\mathbf{U}\mathbf{F})).
\end{aligned}
$$

Because $\|\mathbf{U}\mathbf{F}\|_F^2=O_P(pT)$, $\|\mathbf{\Sigma}_u^{-1}\|_2=O_P(1)$, $\|\mathbf{\Sigma}_u-\widehat{\mathbf{\Sigma}}_u\|_2=O_P(\sqrt{\frac{\log p}{T}}+J^{-\kappa/2})$, the first term on the right hand side is $O_P(\frac{p\log p}{T^2}+\frac{p}{TJ^\kappa})=o_P(\frac{\sqrt{p}}{T})$, given that $p(\log p)^2=o(T^2)$ and $p=o(J^{2\kappa})$.

The difficulty arises in bounding the second term on the right hand side. We aim to show, $\mathrm{tr}((\frac{1}{T}\mathbf{U}\mathbf{F})'\mathbf{\Sigma}_u^{-1}(\mathbf{\Sigma}_u-\widehat{\mathbf{\Sigma}}_u)\mathbf{\Sigma}_u^{-1}(\frac{1}{T}\mathbf{U}\mathbf{F}))=o_P(T\sqrt{p})$. As we argued before, simple inequalities like Holder's or Cauchy-Schwarz do not work. Note that $\mathbf{\Sigma}_u$ is a diagonal matrix. Write $w_i=\frac{1}{Eu_{it}^2}$,

$$
\begin{aligned}
&\mathrm{tr}((\frac{1}{T}\mathbf{U}\mathbf{F})'\mathbf{\Sigma}_u^{-1}(\mathbf{\Sigma}_u-\widehat{\mathbf{\Sigma}}_u)\mathbf{\Sigma}_u^{-1}(\frac{1}{T}\mathbf{U}\mathbf{F}))=\sum_{i=1}^p\sum_{k=1}^K(\frac{1}{T}\sum_{s=1}^T\widehat{u}_{is}^2-Eu_{is}^2)(\sum_{t=1}^T w_iu_{it}f_{tk})^2\\
=\ &\underbrace{\sum_{i=1}^p\sum_{k=1}^K(\frac{1}{T}\sum_{s=1}^T\widehat{u}_{is}^2-u_{is}^2)(\sum_{t=1}^T w_iu_{it}f_{tk})^2}_{(1)}+\underbrace{\sum_{i=1}^p\sum_{k=1}^K(\frac{1}{T}\sum_{s=1}^T u_{is}^2-Eu_{is}^2)(\sum_{t=1}^T w_iu_{it}f_{tk})^2}_{(2)}
\end{aligned}
$$

Denote by $\boldsymbol{\lambda}_i'$ as the ith row of $\mathbf{\Lambda}$. For part (1), note that $\frac{1}{T}\sum_t\widehat{u}_{it}^2-u_{it}^2=\frac{1}{T}\sum_t(\widehat{u}_{it}-u_{it})^2+\frac{2}{T}\sum_t(\widehat{u}_{it}-u_{it})u_{it}$, and

$$
\widehat{u}_{it}-u_{it}=(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i)'(\widehat{\mathbf{f}}_t-\mathbf{f}_t)+(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i)'\mathbf{f}_t+\boldsymbol{\lambda}_i'(\widehat{\mathbf{f}}_t-\mathbf{f}_t).
$$

On the other hand, $\max_{i\leq p}\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it}-u_{it})^2=O_P(\frac{\log p}{T}+\frac{J^2}{p^2T}+\frac{1}{J^\kappa})$, $\max_{i,k}(\sum_t w_iu_{it}f_{tk})^2=O_P(T\log p)$, hence

$$
\begin{aligned}
(1)\ =\ &\sum_{i=1}^p\sum_{k=1}^K\frac{1}{T}\sum_{s=1}^T(\widehat{u}_{is}-u_{is})^2(\sum_{t=1}^T w_iu_{it}f_{tk})^2+2\sum_{i=1}^p\sum_{k=1}^K\frac{1}{T}\sum_{s=1}^T(\widehat{u}_{is}-u_{is})u_{is}(\sum_{t=1}^T w_iu_{it}f_{tk})^2\\
=\ &O_P(p)O_P(\frac{\log p}{T}+\frac{J^2}{p^2T}+\frac{1}{J^\kappa})O_P(T\log p)+2\sum_{i=1}^p\sum_{k=1}^K\frac{1}{T}\sum_{s=1}^T(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i)'\mathbf{f}_su_{is}(\sum_{t=1}^T w_iu_{it}f_{tk})^2\\
&+2\sum_{i=1}^p\sum_{k=1}^K\frac{1}{T}\sum_{s=1}^T(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i)'(\widehat{\mathbf{f}}_s-\mathbf{f}_s)u_{is}(\sum_{t=1}^T w_iu_{it}f_{tk})^2\\
&+2\sum_{i=1}^p\sum_{k=1}^K\frac{1}{T}\sum_{s=1}^T\boldsymbol{\lambda}_i'(\widehat{\mathbf{f}}_s-\mathbf{f}_s)u_{is}(\sum_{t=1}^T w_iu_{it}f_{tk})^2
\end{aligned}
$$

$$
\begin{aligned}
&= O_P(p\log^2 p + \frac{Tp(\log p)}{J^\kappa}) + \max_i \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| O_P(p) \max_i \|\frac{1}{T}\sum_s \mathbf{f}_s u_{is}\| \max_{i,k}(\sum_t w_i u_{it} f_{tk})^2 \\
&\quad + \max_i \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| O_P(p) \max_i \|\frac{1}{T}\sum_s (\widehat{\mathbf{f}}_s - \mathbf{f}_s) u_{is}\| \max_{i,k}(\sum_t w_i u_{it} f_{tk})^2 \\
&\quad + 2\sum_{i=1}^p \sum_{k=1}^K \frac{1}{T}\sum_{s=1}^T \boldsymbol{\lambda}_i'(\widehat{\mathbf{f}}_s - \mathbf{f}_s) u_{is} (\sum_{t=1}^T w_i u_{it} f_{tk})^2 \\
&= O_P(p\log^2 p + \frac{Tp(\log p)}{J^\kappa} + \sqrt{JpT}(\log p)^{3/2} + p^{3/4}T^{1/4}J^{1/4}(\log p)^{3/2}) \\
&\quad + 2\sum_{k=1}^K \sum_{i=1}^p \frac{1}{T}\sum_{s=1}^T \boldsymbol{\lambda}_i'(\widehat{\mathbf{f}}_s - \mathbf{f}_s) u_{is} (\sum_{t=1}^T w_i u_{it} f_{tk})^2 \\
&= O_P(p\log^2 p + \frac{Tp(\log p)}{J^\kappa} + \sqrt{JpT}(\log p)^{3/2} + p^{3/4}T^{1/4}J^{1/4}(\log p)^{3/2} \\
&\quad + T\sqrt{J}\log p + \frac{TJ\log p}{J^{\kappa/2}}).
\end{aligned}
$$

The last equality follows from Lemma D.6 in the supplementary material. Therefore, part (1) is $o_P(T\sqrt{p})$. It also follows from Lemma D.7 that part (2) is $o_P(T\sqrt{p})$. This finishes the proof.

Proof of Theorem 5.1 for S_Γ

By Propositions D.1 and D.2 we have $S = \frac{1}{T^2}\mathrm{tr}(\mathbf{F}'\mathbf{U}'\boldsymbol{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}) + o_P(\frac{\sqrt{p}}{T})$. So it suffices to argue the normality for

$$
\frac{1}{T^2}\mathrm{tr}(\mathbf{F}'\mathbf{U}'\boldsymbol{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}) = [\mathrm{vec}(\frac{1}{T}\mathbf{U}\mathbf{F})]'(\boldsymbol{\Sigma}_u^{-1}\otimes \mathbf{I}_K)[\mathrm{vec}(\frac{1}{T}\mathbf{U}\mathbf{F})]
$$

Let

$$
\mathbf{M}_t = (\mathbf{f}_t' u_{1t}, \cdots, \mathbf{f}_t' u_{pt})', \quad (pK)\times 1.
$$

Then it is easy to see $\mathrm{vec}(\mathbf{U}\mathbf{F}) = \sum_{t=1}^T \mathbf{M}_t$ and $\mathbf{W} = \boldsymbol{\Sigma}_u^{-1}\otimes\mathbf{I}_K = [\frac{1}{T}\sum_{t=1}^T \mathrm{cov}(\mathbf{M}_t|\mathbf{F})]^{-1}$, because $\frac{1}{T}\sum_t \mathbf{f}_t\mathbf{f}_t' = \mathbf{I}_K$. Denote $\mathbf{x} = T^{-1/2}\mathbf{W}^{1/2}\sum_t \mathbf{M}_t$, so $TS = \mathbf{x}'\mathbf{x} + o_P(\sqrt{p})$ and $\mathbf{x}|\mathbf{F} \sim N(0, \mathbf{I}_{pK})$. Therefore $\mathbf{x}'\mathbf{x}$, conditioning on $\mathbf{F}$, is chi-square distributed with degree of freedom pK, and

$$
\frac{\mathbf{x}'\mathbf{x} - pK}{\sqrt{2pK}}|\mathbf{F} \to^d N(0,1),
$$

almost surely in $\mathbf{F}$. We now need to prove the convergence unconditional on $\mathbf{F}$.

The conditional normality means for any t, let z_t be the tth quantile of standard normal, then almost surely in $\mathbf{F}$,

$$
P(\frac{\mathbf{x}'\mathbf{x} - pK}{\sqrt{2pK}} < t|\mathbf{F}) \to z_t
$$

Since the probability is always bounded by one, by dominated convergence theorem,

$$EP(\frac{\mathbf{x}'\mathbf{x}-pK}{\sqrt{2pK}}<t|\mathbf{F})=P(\frac{\mathbf{x}'\mathbf{x}-pK}{\sqrt{pK}}<t)\to z_t,$$

which holds for any fixed t. Hence we have shown $(\mathbf{x}'\mathbf{x}-pK)/\sqrt{2pK}\to^d N(0,1)$. Hence

$$\frac{TS-pK}{\sqrt{2pK}}=\frac{\mathbf{x}'\mathbf{x}-pK}{\sqrt{2pK}}+o_P(\sqrt{\frac{p}{2pK}})\to^d N(0,1).$$

□

D.2. Proof of Theorem 5.1: Limiting distribution for S_G.

PROOF. Write $\mathbf{W}_1=(\frac{1}{p}\widetilde{\mathbf{\Lambda}}'\widetilde{\mathbf{\Lambda}})^{-1}$ and $\mathbf{\Omega}=\mathrm{cov}(\boldsymbol{\gamma}_i)^{-1}$. Under $H_0:\mathbf{G}(\mathbf{X})=0$, $\mathbf{Y}=\mathbf{\Gamma}\mathbf{F}'+\mathbf{U}$. Substituting into the test statistic and by Lemma D.8,

$$\frac{1}{T^2}\mathrm{tr}(\mathbf{W}_1\widetilde{\mathbf{F}}'\mathbf{Y}'\mathbf{P}\mathbf{Y}\widetilde{\mathbf{F}})=\mathrm{tr}(\mathbf{\Omega}\mathbf{\Gamma}'\mathbf{P}\mathbf{\Gamma})+o_P(1).$$

Now define $\mathbf{Z}=\Phi(\mathbf{X})'\mathbf{\Gamma}$, Then $\mathrm{tr}(\mathbf{\Omega}\mathbf{\Gamma}'\mathbf{P}\mathbf{\Gamma})=\mathrm{tr}(\mathbf{Z}'(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\mathbf{Z}\mathbf{\Omega})$. Let $\mathbf{A}=(E\phi(\mathbf{X}_i)\phi(\mathbf{X}_i)')^{-1}$. We then have $\mathrm{tr}(\mathbf{Z}'(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\mathbf{Z}\mathbf{\Omega})=\frac{1}{p}\mathrm{tr}(\mathbf{Z}'\mathbf{A}\mathbf{Z}\mathbf{\Omega})+o_P(1)$. Also define $\mathbf{z}=\frac{1}{\sqrt{p}}(\mathbf{\Omega}^{1/2}\otimes\mathbf{A}^{1/2})\mathrm{vec}(\mathbf{Z})$. Using the formula $\mathrm{tr}(\mathbf{Z}'\mathbf{A}\mathbf{Z}\mathbf{\Omega})=\mathrm{vec}(\mathbf{Z})'(\mathbf{\Omega}\otimes\mathbf{A})\mathrm{vec}(\mathbf{Z})$, we then reach $\frac{1}{p}\mathrm{tr}(\mathbf{Z}'\mathbf{A}\mathbf{Z}\mathbf{\Omega})=\mathbf{z}'\mathbf{z}$. Hence

$$\frac{1}{T^2}\mathrm{tr}(\mathbf{W}_1\widetilde{\mathbf{F}}'\mathbf{Y}'\mathbf{P}\mathbf{Y}\widetilde{\mathbf{F}})=\mathbf{z}'\mathbf{z}+o_P(1).$$

Further define $\mathbf{h}(\boldsymbol{\gamma}_i)=\frac{1}{\sqrt{p}}(\mathbf{\Omega}^{1/2}\otimes\mathbf{A}^{1/2})\mathrm{vec}(\phi(\mathbf{X}_i)\boldsymbol{\gamma}_i')$, then $\mathbf{z}'\mathbf{z}=(\sum_{i=1}^p\mathbf{h}(\boldsymbol{\gamma}_i))'\sum_{i=1}^p\mathbf{h}(\boldsymbol{\gamma}_i)$. Hence it suffices to show

$$\frac{(\sum_{i=1}^p\mathbf{h}(\boldsymbol{\gamma}_i))'\sum_{i=1}^p\mathbf{h}(\boldsymbol{\gamma}_i)-JdK}{\sqrt{2JdK}}\to^d N(0,1).$$

Note that the left hand side equals M_1+M_2, where

$$M_1=[\sum_i\mathbf{h}(\boldsymbol{\gamma}_i)'\mathbf{h}(\boldsymbol{\gamma}_i)-JdK]/\sqrt{2JdK},\quad M_2=2\sum_{i<j}\mathbf{h}(\boldsymbol{\gamma}_i)'\mathbf{h}(\boldsymbol{\gamma}_j)/\sqrt{2JdK}.$$

Note that $\mathbf{h}(\boldsymbol{\gamma}_i)'\mathbf{h}(\boldsymbol{\gamma}_i)=\frac{1}{p}\mathrm{tr}(\boldsymbol{\gamma}_i\phi(\mathbf{X}_i)'\mathbf{A}\phi(\mathbf{X}_i)\boldsymbol{\gamma}_i'\mathbf{\Omega})$. Hence

$$E\sum_i\mathbf{h}(\boldsymbol{\gamma}_i)'\mathbf{h}(\boldsymbol{\gamma}_i)\quad=\quad E\frac{1}{p}\sum_i\mathrm{tr}(\phi(\mathbf{X}_i)'\mathbf{A}\phi(\mathbf{X}_i)\mathrm{tr}(\mathbf{\Omega}E\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i'))$$

$$= \quad K\frac{1}{p}\sum_i \mathrm{tr}(\mathbf{A}E\phi(\mathbf{X}_i)\phi(\mathbf{X}_i)') = JdK,$$

implying $EM_1 = 0$. We now consider $\mathrm{var}(\sum_i \mathbf{h}(\boldsymbol{\gamma}_i)'\mathbf{h}(\boldsymbol{\gamma}_i))$. Let $\tilde{\boldsymbol{\gamma}}_i = \boldsymbol{\Omega}^{1/2}\boldsymbol{\gamma}_i = (\tilde{\gamma}_{i1}, ..., \tilde{\gamma}_{iK})'$ and $a_i = \frac{1}{p}\phi(\mathbf{X}_i)'\mathbf{A}\phi(\mathbf{X}_i)$. Then

$$\begin{aligned}
&\mathrm{var}(\sum_i \mathbf{h}(\boldsymbol{\gamma}_i)'\mathbf{h}(\boldsymbol{\gamma}_i)) = \mathrm{var}(\sum_i \mathrm{tr}(\tilde{\boldsymbol{\gamma}}_i a_i \tilde{\boldsymbol{\gamma}}_i')) = \mathrm{var}(\sum_i \sum_{k\le K} \tilde{\gamma}_{ik}^2 a_i)\\
&= \sum_{k\le K} \mathrm{var}(\sum_i \tilde{\gamma}_{ik}^2 a_i) + 2\sum_{k_1<k_2} \mathrm{cov}(\sum_i \tilde{\gamma}_{ik_1}^2 a_i, \sum_i \tilde{\gamma}_{ik_2}^2 a_i)\\
&= \sum_k p\mathrm{var}(\tilde{\gamma}_{ik}^2 a_i^2) + 2\sum_{k_1<k_2}\sum_{i\le p}\sum_{j\le p} \mathrm{cov}(\tilde{\gamma}_{ik_1}^2 a_i, \tilde{\gamma}_{jk_2}^2 a_j)\\
&= \sum_k p\mathrm{var}(\tilde{\gamma}_{ik}^2 a_i^2) + 2\sum_{k_1<k_2}\sum_{i\le p} \mathrm{cov}(\tilde{\gamma}_{ik_1}^2 a_i, \tilde{\gamma}_{ik_2}^2 a_i) = o(1).
\end{aligned}$$

This implies $M_1 = o_P(1)$.

We now show that $M_2 \to^d N(0,1)$. To do this, we apply Theorem 1 of Hall (1984) by showing that its hypotheses are satisfied. Let $H_p(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{2}{JdK}}\mathbf{h}(\mathbf{x})'\mathbf{h}(\mathbf{y})$, and

$$\begin{aligned}
G_p(\mathbf{x}, \mathbf{y}) &= E[H_p(\boldsymbol{\gamma}_1, \mathbf{x})H_p(\boldsymbol{\gamma}_1, \mathbf{y})]\\
&= \frac{2}{JdK}\mathbf{h}(\mathbf{x})'E\mathbf{h}(\boldsymbol{\gamma}_1)\mathbf{h}(\boldsymbol{\gamma}_1)'\mathbf{h}(\mathbf{y}) = \sqrt{\frac{2}{JdKp^2}}H_p(\mathbf{x}, \mathbf{y})
\end{aligned}$$

where the third equality is due to $E\mathbf{h}(\boldsymbol{\gamma}_1)\mathbf{h}(\boldsymbol{\gamma}_1)' = \mathbf{I}/p$. Note that $E[H_p(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)|\boldsymbol{\gamma}_1, \mathbf{X}] = 0$ and that

$$\begin{aligned}
EH_p(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)^2 &= 2E\mathbf{h}(\boldsymbol{\gamma}_1)'\mathbf{h}(\boldsymbol{\gamma}_2)\mathbf{h}(\boldsymbol{\gamma}_2)'\mathbf{h}(\boldsymbol{\gamma}_1)/(JdK)\\
&= 2E\mathbf{h}(\boldsymbol{\gamma}_1)'\mathbf{h}(\boldsymbol{\gamma}_1)/(JdKp) = 2/p^2.
\end{aligned}$$

On the other hand,

$$\frac{1}{p}EH_p(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)^4/\{E[H_p(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)^2]\}^2 = \frac{1}{p}(\frac{p^2}{JdK})^2 E(\mathbf{h}(\boldsymbol{\gamma}_1)'\mathbf{h}(\boldsymbol{\gamma}_1))^4 = o(1).$$

and

$$E[G_p(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)^2]/\{E[H_p(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)^2]\}^2 = o(1).$$

Therefore, by Theorem 1 of Hall (1984),

$$M_2 = \sum_{i<j} H_p(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j)|\mathbf{X} \to^d N(0,1).$$

□

D.3. Further Technical Lemmas.

LEMMA D.1. *For any p by p positive definite matrix $\mathbf{\Sigma}_u^{-1}$ with bounded l_1 norm,*
(i) $\|\mathbf{G}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F = O_P(\sqrt{pT})$;
(ii) $\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F = O_P(\sqrt{pT})$;
(iii) $\|\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F = O_P(\sqrt{pTJ})$;
(iv) $\|\mathbf{R}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F = O_P(\sqrt{pTJ^{-\kappa}})$.

PROOF. Note that we have derived the same rates for the case $\mathbf{\Sigma}_u^{-1} = \mathbf{I}$ in Lemmas B.1 and C.6. Here we only need to define $\tilde{\mathbf{G}} = \mathbf{\Sigma}_u^{-1}\mathbf{G}$, $\tilde{\Phi}(\mathbf{X}) = \mathbf{\Sigma}_u^{-1}\Phi(\mathbf{X})$, $\tilde{\mathbf{R}} = \mathbf{\Sigma}_u^{-1}\mathbf{R}$ and prove $\tilde{\mathbf{G}}, \tilde{\Phi}(\mathbf{X}), \tilde{\mathbf{R}}$ possess the same properties as $\mathbf{G}, \Phi(\mathbf{X}), \mathbf{R}$. The results follow using the same argument as in Lemmas B.1 and C.6 given that $\mathbf{\Sigma}_u^{-1}$ has bounded row sums. Hence we omit the proofs to avoid repetitions.

□

LEMMA D.2. *Under H_0, $\|\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1}\|_2 = O_P(\sqrt{\frac{\log p}{T}} + \frac{1}{J^{\kappa/2}})$.*

PROOF. Note that under H_0, $\boldsymbol{\lambda}_i = \mathbf{g}(\mathbf{X}_i)$. By definition of $\widehat{\boldsymbol{\lambda}}_i$,
$\widehat{\boldsymbol{\lambda}}_i = \frac{1}{T}\sum_{t=1}^T(\mathbf{f}_t'\boldsymbol{\lambda}_i + u_{it})\widehat{\mathbf{f}}_t = \frac{1}{T}\sum_{t=1}^T(\widehat{\mathbf{f}}_t - \mathbf{f}_t)\mathbf{f}_t'\boldsymbol{\lambda}_i + \boldsymbol{\lambda}_i + \frac{1}{T}\sum_{t=1}^T u_{it}\widehat{\mathbf{f}}_t$. Hence

$$\max_{i\le p}\|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| = O_P(\sqrt{\frac{\log p}{T}}).$$

Because $\widehat{u}_{it} - u_{it} = (\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i)'\widehat{\mathbf{f}}_t + \boldsymbol{\lambda}_i'(\widehat{\mathbf{f}}_t - \mathbf{f}_t)$, and $\max_i \|\frac{1}{T}\sum_t u_{it}\mathbf{f}_t\| = O_P(\sqrt{\frac{\log p}{T}})$, we have $\max_{i\le p}\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it} - u_{it})^2 = O_P(\frac{\log p}{T} + \frac{J^2}{p^2T} + \frac{1}{J^\kappa})$, which also implies

$$\max_{i\le p}|\frac{1}{T}\sum_t u_{it}^2 - \widehat{u}_{it}^2| = O_P(\sqrt{\frac{\log p}{T}} + J/(p\sqrt{T}) + J^{-\kappa/2}).$$

In addition, $\max_{i\le p}|\frac{1}{T}\sum_t u_{it}^2 - Eu_{it}^2| = O_P(\sqrt{\frac{\log p}{T}})$. We have $\max_{i\le p}|\frac{1}{T}\sum_t \widehat{u}_{it}^2 - Eu_{it}^2| = O_P(\sqrt{\frac{\log p}{T}})$. Therefore, $\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\|_2 = \max_{i\le p}|\frac{1}{T}\sum_t \widehat{u}_{it}^2 - Eu_{it}^2| = O_P(\sqrt{\frac{\log p}{T}} + \frac{1}{J^{\kappa/2}})$. Finally, we have the desired result because $\lambda_{\min}(\mathbf{\Sigma}_u)$ is bounded away from zero, implying

$$\|\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1}\|_2 \le \|\mathbf{\Sigma}_u^{-1}(\mathbf{\Sigma}_u - \widehat{\mathbf{\Sigma}}_u)\widehat{\mathbf{\Sigma}}_u^{-1}\|_2 = O_P(\sqrt{\frac{\log p}{T}} + \frac{1}{J^{\kappa/2}})$$

given that $\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_u^{-1})$ is bounded away from zero with probability approaching one. $\square$

LEMMA D.3. *If $JT^{1/2}+J^2=o(p)$, $\sqrt{p}\log p=o(T)$, and $p=o(J^{\kappa})$,*
(i) $\|\mathbf{G}'(\widehat{\boldsymbol{\Sigma}}_u^{-1}-\boldsymbol{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F=O_P(\sqrt{pT})$;
(ii) $\|\mathbf{B}'\Phi(\mathbf{X})'(\widehat{\boldsymbol{\Sigma}}_u^{-1}-\boldsymbol{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F=O_P(\sqrt{pT})$;
(iii) $\|\Phi(\mathbf{X})'(\widehat{\boldsymbol{\Sigma}}_u^{-1}-\boldsymbol{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F=O_P(\sqrt{pTJ})$;
(iv) $\|\mathbf{R}'(\widehat{\boldsymbol{\Sigma}}_u^{-1}-\boldsymbol{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F=O_P(\sqrt{pTJ^{-\kappa}})$.

PROOF. The proofs for terms (ii)-(iv) are very similar to that of (i), so we only prove (i) to avoid repetitions.

(i) We have $\|\mathbf{G}'(\widehat{\boldsymbol{\Sigma}}_u^{-1}-\boldsymbol{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F\leq a+b$, where

$$a=\|\mathbf{G}'\boldsymbol{\Sigma}_u^{-1}(\widehat{\boldsymbol{\Sigma}}_u-\boldsymbol{\Sigma}_u)(\widehat{\boldsymbol{\Sigma}}_u^{-1}-\boldsymbol{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F=O_P(\frac{p\log p}{\sqrt{T}})=O_P(\sqrt{pT})$$

and for $\mathbf{A}'=\mathbf{G}(\mathbf{X})'\boldsymbol{\Sigma}_u^{-1}=(a_{ki})$ and $\mathbf{E}=\boldsymbol{\Sigma}_u^{-1}\mathbf{U}=(e_{it})$,

$$b=\|\mathbf{G}'\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\Sigma}_u-\widehat{\boldsymbol{\Sigma}}_u)\boldsymbol{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F=\|\mathbf{A}'(\boldsymbol{\Sigma}_u-\widehat{\boldsymbol{\Sigma}}_u)\mathbf{E}\mathbf{F}\|_F.$$

Write $\sigma_i=Eu_{it}^2,\widehat{\sigma}_i=\frac{1}{T}\sum_t\widehat{u}_{it}^2,s_i=\frac{1}{T}\sum_t u_{it}^2$. Then $Eb^2\leq c_1+c_2$, where

$$c_1=2\sum_{k=1}^K\sum_{j=1}^K(\sum_{i=1}^p a_{ik}(\sigma_i-s_i)\sum_{t=1}^T e_{it}f_{tj})^2$$

$$c_2=2\sum_{k=1}^K\sum_{j=1}^K(\sum_{i=1}^p a_{ik}(\widehat{\sigma}_i-s_i)\sum_{t=1}^T e_{it}f_{tj})^2.$$

Note that $\boldsymbol{\Sigma}_u$ is diagonal, hence $\{e_{it}\}$ is i.i.d. Gaussian across both (i,t), and is independent of $\{a_{ki},f_{tj}\}$. Thus

$$\begin{aligned}
Ec_1 &= 2\sum_{k=1}^K\sum_{j=1}^K\mathrm{var}(\sum_{i=1}^p a_{ik}(\sigma_i-s_i)\sum_{t=1}^T e_{it}f_{tj})\\
&= \frac{2}{T^2}\sum_{k=1}^K\sum_{j=1}^K\mathrm{var}(\sum_{i=1}^p a_{ik}\sum_{s=1}^T(u_{is}^2-Eu_{is}^2)\sum_{t=1}^T e_{it}f_{tj})\\
&= \frac{2}{T^2}\sum_{k=1}^K\sum_{j=1}^K\sum_{i=1}^p\mathrm{var}(a_{ik}\sum_{s=1}^T(u_{is}^2-Eu_{is}^2)\sum_{t=1}^T e_{it}f_{tj})\\
&\quad+\frac{2}{T^2}\sum_{k=1}^K\sum_{j=1}^K\sum_{i=1}^p\sum_{i\neq h\leq p}\mathrm{cov}(a_{ik}\sum_{s=1}^T(u_{is}^2-Eu_{is}^2)\sum_{t=1}^T e_{it}f_{tj},a_{ik}\sum_{s=1}^T(u_{hs}^2-Eu_{hs}^2)\sum_{t=1}^T e_{ht}f_{tj})
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2}{T^2}\sum_{k=1}^{K}\sum_{j=1}^{K}\sum_{i=1}^{p}\mathrm{var}(a_{ik}\sum_{s=1}^{T}(u_{is}^2-Eu_{is}^2)\sum_{t=1}^{T}e_{it}f_{tj}) \quad \text{(the second term is zero)}\\
&= \frac{2}{T^2}\sum_{k=1}^{K}\sum_{j=1}^{K}\sum_{i=1}^{p}\sum_{s=1}^{T}\mathrm{var}(a_{ik}(u_{is}^2-Eu_{is}^2)\sum_{t=1}^{T}e_{it}f_{tj})\\
&\quad +\frac{2}{T^2}\sum_{k=1}^{K}\sum_{j=1}^{K}\sum_{i=1}^{p}\sum_{s\neq m\leq T}\mathrm{cov}(a_{ik}(u_{is}^2-Eu_{is}^2)\sum_{t=1}^{T}e_{it}f_{tj}, a_{ik}(u_{im}^2-Eu_{im}^2)\sum_{t=1}^{T}e_{it}f_{tj})\\
&= O(pT)+\frac{2}{T^2}\sum_{k=1}^{K}\sum_{j=1}^{K}\sum_{i=1}^{p}\sum_{s\neq m, t\neq s,m, l\neq s,m\leq T}\mathrm{cov}(a_{ik}(u_{is}^2-Eu_{is}^2)e_{it}f_{tj}, a_{ik}(u_{im}^2-Eu_{im}^2)e_{il}f_{lj})\\
&= O(pT).
\end{aligned}
$$

Because $c_1 \geq 0$, hence $c_1 = O_P(pT)$. Using the same proof of term (1) in Lemma D.2, it can be proved that $c_2 = O_P(pT)$. Hence $b = O_P(\sqrt{pT})$. This implies $\|\mathbf{G}'(\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})\mathbf{UF}\|_F = O_P(\sqrt{pT})$.

□

LEMMA D.4. *(i)* $\|\Phi(\mathbf{X})'\mathbf{UU}'\mathbf{\Sigma}_u^{-1}\mathbf{UF}\|_F = O_P(\sqrt{pTJ}(p+T+\sqrt{pT}))$;
(ii) $\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{UU}'\mathbf{\Sigma}_u^{-1}\mathbf{UF}\|_F = O_P(\sqrt{pT}(p+T+\sqrt{pT}))$;
(iii) $\|(\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{UF}\|_F = O_P(T+\sqrt{T^3/p}+\sqrt{pT}+pTJ^{-\kappa/2})$.

PROOF. (i) As before, we need to calculate the order of $E[\|\Phi(\mathbf{X})'\mathbf{UU}'\mathbf{\Sigma}_u^{-1}\mathbf{UF}\|_F^2]$. To do this, we expand the matrix multiplication into element sums. Write $\mathbf{\Sigma}_u^{-1} = (w_{ss})_{s\leq p}$.

$$
E\|\Phi(\mathbf{X})'\mathbf{UU}'\mathbf{\Sigma}_u^{-1}\mathbf{UF}\|_F^2 = \sum_{k\leq K}\sum_{j\leq J}\sum_{l\leq d}E(\sum_{t\leq T}\sum_{s\leq p}\sum_{a\leq T}\sum_{b\leq p}f_{tk}u_{st}w_{ss}u_{sa}u_{ba}\phi_j(x_{bl}))^2
$$

$$
= O(J\sum_{t1,t2,a1,a2\leq T}\sum_{s1,s2,b1,b2\leq p}E(u_{s1,t1}u_{s2,t2}w_{s1,s1}w_{s2,s2}u_{s1,a1}u_{s2,a2}u_{b1,a1}u_{b2,a2})),
$$

since $E(f_{k,t1}f_{k,t2}\phi_j(x_{b1,l})\phi_j(x_{b2,l})) = O(1)$. $E(u_{s1,t1}u_{s2,t2}u_{s1,a1}u_{s2,a2}u_{b1,a1}u_{b2,a2})$ could be expressed as the summation, over all the possible ways of constructing three pairs by the six u's, of the product of three expectations of each pair. In total, there are 15 terms. The first term is

$$
\sum_{t1,t2,a1,a2=1}^{T}\sum_{s1,s2,b1,b2=1}^{p}|w_{s1,s1}w_{s2,s2}E(u_{s1,t1}u_{s2,t2})E(u_{s1,a1}u_{s2,a2})E(u_{b1,a1}u_{b2,a2})|
$$

$E(u_{s1,t1}u_{s2,t2})$ is not zero only when $s1 = s2$ and $t1 = t2$, so the above term equals $O_P(pT^2\sum_{s\leq p}|w_{s,s}w_{s,s}|) = O_P(p^2T^2)$. The second and third terms are

$$\sum_{t1,t2,a1,a2=1}^{T}\sum_{s1,s2,b1,b2=1}^{p}|w_{s1,s1}w_{s2,s2}E(u_{s1,t1}u_{s2,t2})E(u_{s1,a1}u_{b1,a1})E(u_{s2,a2}u_{b2,a2})| = O(pT^3),$$

$$\sum_{t1,t2,a1,a2=1}^{T}\sum_{s1,s2,s2,b1,b2=1}^{p}|w_{s1,s1}w_{s2,s2}E(u_{s1,t1}u_{s1,a1})E(u_{s2,t2}u_{s2,a2})E(u_{b1,a1}u_{b2,a2})| = O(p^3T).$$

All the other terms can be bounded similarly. For simplicity, details are omitted. Therefore, $E\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F^2 = O(pTJ(p^2+T^2+pT))$ which gives the conclusion.

(ii) Similar to (i), if we regard $\mathbf{B}'\Phi(\mathbf{X})'$ as a whole, it is equal to $\mathbf{G}'-\mathbf{R}'$. Since $\mathbf{G}'$ is K by p, so the only difference for the rate of $E\|\mathbf{G}'\mathbf{U}\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F^2$ is that we sum over K in stead of Jd, which is the row dimension of $\Phi(\mathbf{X})'$. All others being the same, we could easily obtain $E\|\mathbf{G}'\mathbf{U}\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F^2 = O(pT(p^2+T^2+pT))$. Since $\mathbf{R}$ is small compared to $\mathbf{G}$, we will get the same rate for $E\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F^2$, which leads to the result.

(iii) $\|(\widehat{\mathbf{F}}-\mathbf{F})'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \|(\mathbf{H}-\mathbf{I})'\mathbf{F}'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F + \|(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F$. The first term is bounded by

$$\|(\mathbf{H}-\mathbf{I})'\mathbf{F}'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \|\mathbf{H}-\mathbf{I}\|_F\|\mathbf{\Sigma}_u^{-1}\|\|\mathbf{U}\mathbf{F}\|_F^2 = O_P(T+\sqrt{pT}+pTJ^{-\kappa/2}).$$

Now let us bound the second term in the following. Again we decompose $\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H}$ into eight term as in $A_1 \sim A_8$ in (C.1) (note that under H_0, $A_i = 0$ for $i = 9, ..., 15$).

$$\|\mathbf{A}_1'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \frac{1}{Tp}\|\widehat{\mathbf{F}}'\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F\|\mathbf{F}'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \frac{1}{Tp}\|\widehat{\mathbf{F}}'\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F\|\mathbf{\Sigma}_u^{-1}\|\|\mathbf{U}\mathbf{F}\|_F^2.$$

By Lemmas B.1 and C.6 $\|\mathbf{U}\mathbf{F}\|_F^2 = O_P(pT)$ and $\|\widehat{\mathbf{F}}'\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F \le \|(\widehat{\mathbf{F}}-\mathbf{F})'\|_F\|\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F + \|\mathbf{F}'\mathbf{U}'\Phi(\mathbf{X})\mathbf{B}\|_F = O_P(\sqrt{pT})$, so $\|\mathbf{A}_1'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F = O_P(\sqrt{pT})$.

$$\|\mathbf{A}_2'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \frac{1}{Tp}\|\widehat{\mathbf{F}}'\mathbf{F}\|_F\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F = O_P(T+\sqrt{T^3/p}+\sqrt{pT}).$$

$$\|\mathbf{A}_3'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \frac{1}{Tp}\|\widehat{\mathbf{F}}'\mathbf{U}'\Phi\|_F\|(\Phi'\Phi)^{-1}\|\|\Phi'\mathbf{U}\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F = O_P(J+TJ/p+\sqrt{\frac{TJ^2}{p}}).$$

$$\|\mathbf{A}_4'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \frac{1}{Tp}\|\widehat{\mathbf{F}}'\mathbf{F}\|_F\|\mathbf{R}'\|_F\|\Phi\mathbf{B}\|_F\|\mathbf{\Sigma}_u^{-1}\|\|\mathbf{U}\mathbf{F}\|_F^2 = O_P(pTJ^{-\kappa/2}),$$

and $\|\mathbf{A}_5'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F$ can be bounded exactly by the same way.

$$\|\mathbf{A}_6'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \frac{1}{Tp}\|\widehat{\mathbf{F}}'\mathbf{F}\|_F\|\mathbf{P}\|\|\mathbf{R}'\|_F^2\|\mathbf{\Sigma}_u^{-1}\|\|\mathbf{U}\mathbf{F}\|_F^2 = O_P(pTJ^{-\kappa}).$$

$$\|\mathbf{A}_7'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \frac{1}{Tp}\|\widehat{\mathbf{F}}'\|_F\|\mathbf{U}'\Phi\|_F\|(\Phi'\Phi)^{-1}\|\|\Phi'\|\|\mathbf{R}\|_F\|\mathbf{\Sigma}_u^{-1}\|\|\mathbf{U}\mathbf{F}\|_F^2 = O_P(\sqrt{pT^2J^{1-\kappa}}).$$

$$\begin{aligned}\|\mathbf{A}_8'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F &\le \frac{1}{Tp}\|\widehat{\mathbf{F}}'\mathbf{F}\|_F\|\mathbf{R}'\|_F\|\Phi\|\|(\Phi'\Phi)^{-1}\|\|\Phi'\mathbf{U}\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \\ &= O_P(\sqrt{\frac{JT}{pJ^{\kappa}}}(p+T+\sqrt{pT})).\end{aligned}$$

Combining all the terms above, we conclude for the second term that

$$\|(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F \le \sum_{i=1}^{8}\|\mathbf{A}_i'\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U}\mathbf{F}\|_F = O_P(T+\sqrt{T^3/p}+\sqrt{pT}+pTJ^{-\kappa/2}).$$

Together with the rate of the first term, the proof is complete. □

LEMMA D.5. *Suppose $J\log p = o(T)$, then*
(i) $\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{U}'\|_F = O_P(p\sqrt{TJ} + T\sqrt{p})$,
(ii) $\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{U}'(\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F = o_P(\sqrt{\frac{p^4T^3}{J}} + \sqrt{p^3T^2J^{\kappa}})$;
(iii) $\|\mathbf{B}'\Phi(\mathbf{X})'\mathbf{U}\mathbf{U}'(\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F = o_P(Tp^{3/2})$;
(iv) $\|(\widehat{\mathbf{F}} - \mathbf{F})'\mathbf{U}'(\widehat{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})\mathbf{U}\mathbf{F}\|_F = o_P(T\sqrt{p})$.

PROOF. $\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{U}'\|_F \le \|\Phi(\mathbf{X})'(\mathbf{U}\mathbf{U}' - E\mathbf{U}\mathbf{U}')\|_F + \|\Phi(\mathbf{X})'E\mathbf{U}\mathbf{U}'\|_F$. The first term on the right hand side is $O_P(p\sqrt{TJ})$, while the second term is $O_P(T\sqrt{p})$. Parts (ii)(iii) follow from part (i) and Lemmas D.2. Part (iv) follows from (ii)(iii) by exactly the same arguments as D.4 (iii). □

LEMMA D.6. *For all $k \le K$, and $w_i = \frac{1}{Eu_{it}^2}$, under H_0,*

$$\sum_{i=1}^{p}\sum_{s=1}^{T}\boldsymbol{\lambda}_i'(\widehat{\mathbf{f}}_s-\mathbf{f}_s)u_{is}(\sum_{t=1}^{T}w_iu_{it}f_{tk})^2 = O_P(T^2\sqrt{J}\log p+T\sqrt{JTp}\log p+T^2J\log p/J^{\kappa/2}).$$

PROOF. We have

$$\sum_{i=1}^{p}\sum_{s=1}^{T}(\widehat{\mathbf{f}}_s - \mathbf{f}_s)'\boldsymbol{\lambda}_iu_{is}(\sum_{t=1}^{T}w_iu_{it}f_{tk})^2 = \underbrace{\frac{1}{Tp}\sum_{i=1}^{p}\sum_{s=1}^{T}u_{is}\mathbf{f}_s'\mathbf{\Lambda}'\mathbf{P}\mathbf{U}\widehat{\mathbf{F}}\mathbf{K}_1^{-1}\boldsymbol{\lambda}_i(\sum_{t=1}^{T}w_iu_{it}f_{tk})^2}_{(1)}$$

$$+\underbrace{\frac{1}{Tp}\sum_{i=1}^{p}\sum_{s=1}^{T}u_{is}\mathbf{u}_s'\mathbf{PU}\widehat{\mathbf{F}}\mathbf{K}_1^{-1}\boldsymbol{\lambda}_i(\sum_{t=1}^{T}w_iu_{it}f_{tk})^2}_{(2)}+\underbrace{\frac{1}{Tp}\sum_{i=1}^{p}\sum_{s=1}^{T}u_{is}\mathbf{u}_s'\mathbf{P}\boldsymbol{\Lambda}\mathbf{F}'\widehat{\mathbf{F}}\mathbf{K}_1^{-1}\boldsymbol{\lambda}_i(\sum_{t=1}^{T}w_iu_{it}f_{tk})^2}_{(3)}$$

Note that $\|\mathbf{PU}\widehat{\mathbf{F}}\|_F=O_P(\sqrt{TJ}+TJ/\sqrt{p})$ and $\max_i\|\sum_{s=1}^T u_{is}\mathbf{f}_s\|=O_P(\sqrt{T\log p})$. Hence part (1) is $O_P((\sqrt{pJ}+\sqrt{T}J)T(\log p)^{3/2})$. Also, $E\sum_{s=1}^T\|\mathbf{u}_s'\Phi(\mathbf{X})\|^2=O(TpJ)$, by (C.4), $\|\frac{1}{p}\Phi(\mathbf{X})'\mathbf{U}\widehat{\mathbf{F}}\|^2=O_P(T^2J/p^2+T^2J^3/p^3+T^2J/(pJ^\kappa)+TJ/p)$. Hence part (2) is $O_P((T^2J/\sqrt{p}+T^2J^2/p+T^2J/(J^{\kappa/2})+T^{3/2}J)\log p)$. Part (3) is dominated by

$$O_P(T\log p)\frac{1}{Tp}\sum_{i=1}^{p}\|\sum_{s=1}^{T}u_{is}\mathbf{u}_s'\mathbf{P}\|\|\boldsymbol{\Lambda}\mathbf{F}'\widehat{\mathbf{F}}\mathbf{K}_1^{-1}\|_F\|\boldsymbol{\lambda}_i\|\le O_P(T\log p)\frac{1}{p}\sum_{i=1}^{p}\|\sum_{s=1}^{T}u_{is}\mathbf{u}_s'\Phi(\mathbf{X})\|.$$

Also,

$$\begin{aligned}
&E\|\sum_s u_{is}\mathbf{u}_s'\Phi(\mathbf{X})\|^2=\sum_{k=1}^{J}\sum_{l=1}^{d}E(\sum_s\sum_j u_{js}u_{is}\phi_k(X_{jl}))^2\\
&\le 2\sum_{k=1}^{J}\sum_{l=1}^{d}E(\sum_s\sum_j(u_{js}u_{is}-Eu_{js}u_{is})\phi_k(X_{jl}))^2+2\sum_{k=1}^{J}\sum_{l=1}^{d}E(\sum_s\sum_j(Eu_{js}u_{is})\phi_k(X_{jl}))^2\\
&\le 2\sum_{k=1}^{J}\sum_{l=1}^{d}\mathrm{var}(\sum_s\sum_j(u_{js}u_{is}-Eu_{js}u_{is})\phi_k(X_{jl}))+2\sum_{k=1}^{J}\sum_{l=1}^{d}T^2(Eu_{is}^2)^2E\phi_k(X_{il})^2\\
&\le O(T^2J)+2\sum_{k=1}^{J}\sum_{l=1}^{d}\sum_{s=1}^{T}\mathrm{var}(\sum_j(u_{js}u_{is}-Eu_{js}u_{is})\phi_k(X_{jl}))\\
&+2\sum_{k=1}^{J}\sum_{l=1}^{d}\sum_{s\ne t}\sum_{m=1}^{p}\sum_{j=1}^{p}E((u_{js}u_{is}-Eu_{js}u_{is})(u_{mt}u_{it}-Eu_{mt}u_{it}))E(\phi_k(X_{jl})\phi_k(X_{ml}))\\
&=O(T^2J)+2\sum_{k=1}^{J}\sum_{l=1}^{d}\sum_{s=1}^{T}\sum_{j=1}^{p}\mathrm{var}(u_{js}u_{is}-Eu_{js}u_{is})\phi_k(X_{jl}))\\
&+2\sum_{k=1}^{J}\sum_{l=1}^{d}\sum_{s=1}^{T}\sum_{j\ne m}\mathrm{cov}((u_{js}u_{is}-Eu_{js}u_{is})\phi_k(X_{jl}),(u_{ms}u_{is}-Eu_{ms}u_{is})\phi_k(X_{ml}))\\
&=O(T^2J+JTp)+4\sum_{k=1}^{J}\sum_{l=1}^{d}\sum_{s=1}^{T}\sum_{j\ne i}\mathrm{cov}(u_{js}u_{is},u_{is}u_{is})E(\phi_k(X_{jl})\phi_k(X_{il}))=O(T^2J+JTp).
\end{aligned}$$

where the involved $O(\cdot)$ terms are independent of i. Hence

$$E\frac{1}{p}\sum_{i=1}^{p}\|\sum_{s=1}^{T}u_{is}\mathbf{u}_s'\Phi(\mathbf{X})\| \leq \frac{1}{p}\sum_{i=1}^{p}(E\|\sum_{s=1}^{T}u_{is}\mathbf{u}_s'\Phi(\mathbf{X})\|^2)^{1/2} = O(T\sqrt{J}+\sqrt{JTp}).$$

This yields (3) $= O_P((T^2\sqrt{J}+T\sqrt{JTp})\log p)$. Combining (1)-(3), we obtain the result.

Lemma D.7. *For each $k \leq K$, $w_i = \frac{1}{Eu_{it}^2}$,*

$$\sum_{i=1}^{p}(\frac{1}{T}\sum_{s=1}^{T}u_{is}^2 - Eu_{is}^2)(\sum_{t=1}^{T}w_iu_{it}f_{tk})^2 = O_P(\sqrt{pT}+p).$$

Proof. Let $Z = \sum_{i=1}^{p}(\frac{1}{T}\sum_{s=1}^{T}u_{is}^2 - Eu_{is}^2)(\sum_{t=1}^{T}w_iu_{it}f_{tk})^2$. We aim to find $EZ^2 = (EZ)^2 + \mathrm{var}(Z)$.

$$\begin{aligned}
EZ &= \sum_{i=1}^{p}\mathrm{cov}(\frac{1}{T}\sum_{s=1}^{T}u_{is}^2, (\sum_{t=1}^{T}w_iu_{it}f_{tk})^2) = \sum_{i=1}^{p}\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{s=1}^{T}\mathrm{cov}(u_{is}^2, w_iu_{it}f_{tk}w_iu_{il}f_{lk})\\
&= 2\sum_{i=1}^{p}\frac{1}{T}\sum_{l=1}^{T}\sum_{s=1}^{T}\mathrm{cov}(u_{is}^2, w_iu_{is}f_{sk}w_iu_{il}f_{lk}) + 2\sum_{i=1}^{p}\frac{1}{T}\sum_{t\neq s}\sum_{l=1}^{T}\sum_{s=1}^{T}\mathrm{cov}(u_{is}^2, w_iu_{it}f_{tk}w_iu_{il}f_{lk})\\
&= 2\sum_{i=1}^{p}\frac{1}{T}\sum_{l=1}^{T}\mathrm{cov}(u_{il}^2, w_i^2u_{il}^2f_{lk}^2) + 2\sum_{i=1}^{p}\frac{1}{T}\sum_{t\neq s}\sum_{s=1}^{T}\mathrm{cov}(u_{is}^2, w_i^2u_{it}f_{tk}u_{is}f_{sk})\\
&= O(p) + 2\sum_{i=1}^{p}\frac{1}{T}\sum_{t\neq s}\sum_{s=1}^{T}E(u_{is}^3w_i^2u_{it})E(f_{tk}f_{sk}) = O(p).
\end{aligned}$$

Let $v_{is} = u_{is}^2 - Eu_{is}^2$, then

$$\begin{aligned}
\mathrm{var}(Z) &= \sum_{i=1}^{p}T\mathrm{var}(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}(u_{is}^2 - Eu_{is}^2)(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_iu_{it}f_{tk})^2)\\
&\quad + \sum_{i\neq j}\mathrm{cov}((\frac{1}{T}\sum_{s=1}^{T}u_{is}^2 - Eu_{is}^2)(\sum_{t=1}^{T}w_iu_{it}f_{tk})^2, (\frac{1}{T}\sum_{s=1}^{T}u_{js}^2 - Eu_{js}^2)(\sum_{t=1}^{T}w_ju_{jt}f_{tk})^2)\\
&\leq \underbrace{\sum_{i=1}^{p}TE((\frac{1}{\sqrt{T}}\sum_{s=1}^{T}v_{is})^2(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_iu_{it}f_{tk})^4)}_{(1)}
\end{aligned}$$

$$+\underbrace{\sum_{i\neq j}\frac{1}{T}\sum_{s=1}^{T}\frac{1}{T}\sum_{l=1}^{T}\sum_{q=1}^{T}\sum_{t=1}^{T}\sum_{n=1}^{T}\sum_{m=1}^{T} w_i^2 w_j^2 \mathrm{cov}(v_{is}u_{iq}f_{qk}u_{it}f_{tk}, v_{jl}u_{jm}f_{mk}u_{jn}f_{nk})}_{(2)}.$$

We have $(1) = O(pT)$. As for part (2), note that when $i \neq j$,

$$\mathrm{cov}(v_{is}u_{iq}f_{qk}u_{it}f_{tk}, v_{jl}u_{jm}f_{mk}u_{jn}f_{nk}) = E(v_{is}u_{iq}u_{it})E(v_{jl}u_{jm}u_{jn})\mathrm{cov}(f_{mk}f_{nk}, f_{qk}f_{tk})$$

Hence $(2) = \sum_{i\neq j}\frac{1}{T}\sum_{s=1}^{T}\frac{1}{T}\sum_{l=1}^{T} w_i^2 w_j^2 E(v_{is}u_{is}^2)E(v_{jl}u_{jl}^2)\mathrm{cov}(f_{lk}^2, f_{sk}^2) = O(p^2)$ (all other terms in the summation are zero). Thus $\mathrm{var}(Z) = O(pT + p^2)$, implying $EZ^2 = O(pT + p^2)$, and thus $Z = O_P(\sqrt{pT} + p)$. □

LEMMA D.8. *For $\widetilde{\mathbf{F}}$ as the regular PCA estimator of the factors,*
(i) $\|\frac{1}{T}\mathbf{W}_1(\widetilde{\mathbf{F}} - \mathbf{F})'\mathbf{F}\mathbf{\Gamma}'\mathbf{P}\mathbf{\Gamma}\|_F = o_P(1)$
(ii) $\|\frac{1}{T^2}\mathbf{W}_1\widetilde{\mathbf{F}}'\mathbf{F}\mathbf{\Gamma}'\mathbf{P}\mathbf{U}\widetilde{\mathbf{F}}\|_F = o_P(1)$, $\|\frac{1}{T^2}\mathbf{W}_1\widetilde{\mathbf{F}}'\mathbf{U}'\mathbf{P}\mathbf{U}\widetilde{\mathbf{F}}\|_F = o_P(1)$,
(iii) $\|(\mathbf{\Omega} - \mathbf{W}_1)\mathbf{\Gamma}'\mathbf{P}\mathbf{\Gamma}\|_F = o_P(1)$.

PROOF. By Lemma C.6, $\|\Phi(\mathbf{X})'\mathbf{\Gamma}\|_F^2 = O_P(Jp)$. In addition, there is a $K \times K$ matrix $\bar{\mathbf{H}}$, so that $\|\bar{\mathbf{H}}\|_2 = O_P(1)$ and

$$\|\Phi(\mathbf{X})'\mathbf{U}\widetilde{\mathbf{F}}\|_F^2 \leq 2\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\bar{\mathbf{H}}\|_F^2 + 2\|\Phi(\mathbf{X})'\mathbf{U}\|_F^2\|\tilde{\mathbf{F}} - \mathbf{F}\bar{\mathbf{H}}\|_F^2.$$

By Lemma B.1, $\|\Phi(\mathbf{X})'\mathbf{U}\mathbf{F}\bar{\mathbf{H}}\|_F^2 = O_P(pTJ)$; $\|\Phi(\mathbf{X})'\mathbf{U}\|_F^2 = O_P(pTJ)$. In addition, for the PCA estimator $\tilde{\mathbf{F}}$, it is well known that $\|\tilde{\mathbf{F}} - \mathbf{F}\bar{\mathbf{H}}\|_F^2 = O_P(1 + T/p)$ (e.g., see Bai (2003); Fan et al. (2013)). Hence $\|\Phi(\mathbf{X})'\mathbf{U}\widetilde{\mathbf{F}}\|_F^2 = O_P(pTJ)(T/p + 1)$. Also, $\|\mathbf{W}_1\|_2 = O_P(1)$ and $J\|\mathbf{\Omega} - \mathbf{W}_1\|_F = o_P(1)$. Hence straightforward calculations yield the result.

□

APPENDIX E: PROOFS FOR SECTION 6

Let $\lambda_k(\mathbf{A})$ denote the kth largest eigenvalue of matrix $\mathbf{A}$ and $\sigma_k(\mathbf{A})$ the kth singular value of $\mathbf{A}$. We write $\mathbf{G} = \mathbf{G}(\mathbf{X})$. The following proofs use the technical lemmas E.4-E.8.

E.1. Behavior of $\lambda_k(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT))$ for $k \leq K$.

LEMMA E.1. *Suppose $J = o(p)$, for $k = 1, \cdots, K-1$,*

$$\frac{\lambda_k(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT))}{\lambda_{k+1}(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT))} = O_P(1).$$

PROOF. The result follows immediately from Lemma C.5.

E.2. Behavior of $\lambda_k(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT))$ for $k > K$.

LEMMA E.2. *Let* $\mathbf{B} = \frac{1}{pT}\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X}) - \Big[\frac{1}{p}\Phi(\mathbf{X})'\mathbf{\Sigma}_u\Phi(\mathbf{X})\Big]$. *Suppose at least one of the two sets of assumptions holds:*

(1) the sub-Gaussian condition in condition (iii) and condition (iv) of Assumption 6.1, $J = o(T)$ *or*

(2) $\mathbf{\Sigma}_u$ *is diagonal, with* $\max_{i\le p, j\le J, l\le d} E\phi_j(X_{il})^4 < \infty$, *and* $J = o(\sqrt{T})$.

Then as $T, p \to \infty$, $\|\mathbf{B}\|_2 = o_P(1)$.

PROOF. (1) Each column of $\mathbf{A} := \Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}/\sqrt{p}$ is denoted by $\mathbf{A}_t = \Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{e}_t/\sqrt{p}$. By assumption, $\mathbf{A}_t$ is sub-Gaussian conditional on $\mathbf{X}$ with bounded sub-Gaussian norm. Specifically the tail inequality of $\mathbf{e}_t$ in Assumption 6.1 implies that $E[\exp(t\mathbf{v}'\mathbf{e}_t)] \le \exp(Ct^2)$ for any $t > 0, \|\mathbf{v}\| = 1$ (Vershynin, 2010). Therefore,

$$
\begin{aligned}
E[\exp(t\mathbf{v}'\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{e}_t/\sqrt{p})|\mathbf{X}] &\le \exp(Ct^2\|\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})\mathbf{v}\|^2/p) \\
&\le \exp(Ct^2\|\Phi(\mathbf{X})'\mathbf{\Sigma}_u\Phi(\mathbf{X})/p\|^2) \le \exp(CC_2 d_{\max} t^2)\,,
\end{aligned}
$$

where the last inequality holds due to condition (iv) of Assumption 6.1. This implies that $\mathbf{A}_t$ is indeed sub-Gaussian conditioning on $\mathbf{X}$.

Then by Vershynin (2010), also cited in Lemma E.4, conditioning on $\mathbf{X}$,

$$
P\Big(\|\mathbf{B}\|_2 \le C_1\sqrt{\frac{Jd}{T}}\Big|\mathbf{X}\Big) \ge 1 - \exp(-C_2 J)\,,
$$

for constants C_1, C_2 independent of $\mathbf{X}$ since

$$
\|\mathbf{A}_t\|_{\phi_2} := \sup_{\|\mathbf{v}\|=1}\sup_{m\ge1} m^{-1/2}E(|\mathbf{v}'\mathbf{A}_t|^m)^{1/m} < \sqrt{CC_2 d_{\max}}
$$

(which follows from Lemma 5.5 of Vershynin (2010)). Taking extra expectation with respect to $\mathbf{X}$, we have

$$
\|\mathbf{B}\|_2 = O_P\Big(\sqrt{\frac{J}{T}}\Big) = o_P(1)\,.
$$

(2) Write $\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X}) = (a_{il})_{p\times Jd}$. By assumption, we have $\max_{il} Ea_{il}^4 < \infty$. Also note that

$$
\mathrm{var}[\frac{1}{pT}\sum_{i=1}^p\sum_{j=1}^p\sum_{t=1}^T a_{ik}a_{jl}(e_{it}e_{jt} - Ee_{it}e_{jt})]
$$

$$= E\text{var}[\frac{1}{pT}\sum_{i=1}^p\sum_{j=1}^p\sum_{t=1}^T a_{ik}a_{jl}(e_{it}e_{jt}-Ee_{it}e_{jt})|\mathbf{X}]$$
$$= \sum_{t=1}^T E\text{var}[\frac{1}{pT}\sum_{i=1}^p\sum_{j=1}^p a_{ik}a_{jl}(e_{it}e_{jt}-Ee_{it}e_{jt})|\mathbf{X}] = \sum_{t=1}^T \text{var}[\frac{1}{pT}\sum_{i=1}^p\sum_{j=1}^p a_{ik}a_{jl}(e_{it}e_{jt}-Ee_{it}e_{jt})]$$
$$= \frac{1}{T}\text{var}[\frac{1}{p}\sum_{i=1}^p\sum_{j=1}^p a_{ik}a_{jl}(e_{it}e_{jt}-Ee_{it}e_{jt})].$$

Hence

$$E\|\mathbf{B}\|_F^2 = E\sum_{k=1}^{Jd}\sum_{l=1}^{Jd}[\frac{1}{pT}\sum_{i=1}^p\sum_{j=1}^p\sum_{t=1}^T a_{ik}a_{jl}(e_{it}e_{jt}-Ee_{it}e_{jt})]^2$$
$$= \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\text{var}[\frac{1}{pT}\sum_{i=1}^p\sum_{j=1}^p\sum_{t=1}^T a_{ik}a_{jl}(e_{it}e_{jt}-Ee_{it}e_{jt})]$$
$$= \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\text{var}[\sum_{i=1}^p\sum_{j=1}^p a_{ik}a_{jl}(e_{it}e_{jt}-Ee_{it}e_{jt})]$$
$$= \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{j=1}^p\sum_{m=1}^p\sum_{n=1}^p \text{cov}[a_{ik}a_{jl}(e_{it}e_{jt}-Ee_{it}e_{jt}), a_{mk}a_{nl}(e_{mt}e_{nt}-Ee_{mt}e_{nt})]$$
$$= \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{j=1}^p\sum_{m=1}^p\sum_{n=1}^p \text{cov}[a_{ik}a_{jl}e_{it}e_{jt}, a_{mk}a_{nl}e_{mt}e_{nt}]$$
$$= \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{n=1}^p \text{cov}[a_{ik}a_{il}e_{it}^2, a_{nk}a_{nl}e_{nt}^2] + \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{m\neq n}^p\sum_{n=1}^p Ea_{ik}a_{il}a_{mk}a_{nl}Ee_{it}^2e_{mt}e_{nt}$$
$$+ \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{j\neq i}^p\sum_{m=1}^p\sum_{n=1}^p Ea_{ik}a_{jl}a_{mk}a_{nl}Ee_{it}e_{jt}e_{mt}e_{nt}$$
$$= O(\frac{J^2}{T}) + \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{j\neq i}^p\sum_{n=1}^p Ea_{ik}a_{jl}a_{ik}a_{nl}Ee_{it}^2e_{jt}e_{nt}$$
$$+ \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{j\neq i}^p\sum_{m\neq i}^p Ea_{ik}a_{jl}a_{mk}a_{nl}Ee_{it}^2Ee_{jt}e_{mt}$$
$$= O(\frac{J^2}{T}) + \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{j\neq i}^p Ea_{ik}a_{jl}a_{ik}a_{jl}Ee_{it}^2e_{jt}^2$$
$$+ \sum_{k=1}^{Jd}\sum_{l=1}^{Jd}\frac{1}{Tp^2}\sum_{i=1}^p\sum_{j\neq i}^p Ea_{ik}a_{jl}a_{jk}a_{nl}Ee_{it}^2Ee_{jt}^2 = O(\frac{J^2}{T}).$$

Hence $\|\mathbf{B}\|_F^2 = O_P(1)$ when $J^2 = o(T)$. Hence $\|\mathbf{B}\|_2 = o_P(1)$.

LEMMA E.3. *Under assumptions of Theorem 6.1, for J such that $2K < Jd$, there exist constants c and C such that for $j = 1, 2, \cdots, Jd - 2K$,*

$$c + o_P(1) \leq p\lambda_{K+j}(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT)) \leq C + o_P(1),$$

where c, C and $o_P(1)$ are uniform in $j \leq Jd - 2K$.

PROOF. By the assumptions, there exist c_1 and c_2 such that uniformly in T and p,

$$\lambda_1(\mathbf{\Sigma}_u) \leq c_1, \lambda_1(\mathbf{M}) \leq c_1,$$
$$\lambda_p(\mathbf{\Sigma}_u) \geq c_2, \lambda_T(\mathbf{M}) \geq c_2.$$

From Lemma E.6, we obtain

$$\begin{aligned}
\lambda_1(\mathbf{U}'\mathbf{P}\mathbf{U}/T) &= \lambda_1(\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\mathbf{P}\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{M}/T) \\
&\leq \lambda_1(\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})/(pT))\lambda_1([p^{-1}\Phi(\mathbf{X})'\Phi(\mathbf{X})]^{-1})\lambda_1(\mathbf{M}) \\
&\leq c_1 d_{\min}^{-1}\lambda_1(\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})/(pT)),
\end{aligned}$$

and from Lemma E.8,

$$\begin{aligned}
\lambda_{Jd}(\mathbf{U}'\mathbf{P}\mathbf{U}/T) &= \lambda_{Jd}(\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\mathbf{P}\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{M}/T) \geq \lambda_{Jd}(\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\mathbf{P}\mathbf{\Sigma}_u^{1/2}\mathbf{E}/T)\lambda_T(\mathbf{M}) \\
&\geq c_2 d_{\max}^{-1}\lambda_{Jd}(\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})/(pT)).
\end{aligned}$$

Hence, eigenvalues of $\mathbf{U}'\mathbf{P}\mathbf{U}/T$ are bounded by those of $\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})/(pT)$.

We decompose $\mathbf{Y}'\mathbf{P}\mathbf{Y}$ by

$$\mathbf{Y}'\mathbf{P}\mathbf{Y} = (\mathbf{F}\mathbf{\Lambda}' + \mathbf{U}')\mathbf{P}(\mathbf{\Lambda}\mathbf{F}' + \mathbf{U}) = \tilde{\mathbf{F}}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda}\tilde{\mathbf{F}}' + \mathbf{U}'\mathbf{P}[I - \mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}]\mathbf{P}\mathbf{U},$$

where $\mathbf{\Lambda} = \mathbf{G} + \mathbf{\Gamma}$ and $\tilde{\mathbf{F}} = \mathbf{F} + \mathbf{U}'\mathbf{P}\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda})^{-1}$. So from Lemma E.5,

$$\begin{aligned}
\lambda_{K+j}(\mathbf{Y}'\mathbf{P}\mathbf{Y}) &\leq \lambda_{K+1}(\tilde{\mathbf{F}}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda}\tilde{\mathbf{F}}') + \lambda_j(\mathbf{U}'\mathbf{P}[I - \mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}]\mathbf{P}\mathbf{U}) \\
&= \lambda_j(\mathbf{U}'\mathbf{P}[I - \mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}]\mathbf{P}\mathbf{U}) \leq \lambda_j(\mathbf{U}'\mathbf{P}\mathbf{U}),
\end{aligned}$$

where the last inequality is due to Lemma E.6. From the decomposition we also notice that

$$\begin{aligned}
\lambda_{K+j}(\mathbf{Y}'\mathbf{P}\mathbf{Y}) &\geq \lambda_{K+j}(\mathbf{U}'\mathbf{P}[\mathbf{I} - \mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}]\mathbf{P}\mathbf{U}) \\
&= \lambda_{K+j}(\mathbf{U}'\mathbf{P}[\mathbf{I} - \mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}]\mathbf{P}\mathbf{U}) + \lambda_{K+1}(\mathbf{U}'\mathbf{P}\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}\mathbf{P}\mathbf{U}) \\
&\geq \lambda_{2K+j}(\mathbf{U}'\mathbf{P}\mathbf{U}).
\end{aligned}$$

The last equality is because $\mathbf{U}'\mathbf{P}\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}\mathbf{P}\mathbf{U}$ has rank K and the last inequality is again due to Lemma E.6.

Therefore, for $j \leq Jd - 2K$,

$$c_2 d_{\max}^{-1}\lambda_{Jd}(\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})/(pT)) \leq \lambda_{Jd}(\mathbf{U}'\mathbf{P}\mathbf{U}/T)$$

$$\leq p\lambda_{K+j}(\mathbf{Y}\mathbf{P}\mathbf{Y}'/(pT)) \leq \lambda_1(\mathbf{U}'\mathbf{P}\mathbf{U}/T) \leq c_1 d_{\min}^{-1}\lambda_1(\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})/(pT)).$$

Now by Lemma E.2,

$$\left\|\frac{1}{pT}\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X}) - \frac{1}{p}\Phi(\mathbf{X})'\mathbf{\Sigma}_u\Phi(\mathbf{X})\right\| = o_P(1)\,.$$

The maximal and minimal eigenvalues of $\Phi(\mathbf{X})'\mathbf{\Sigma}_u\Phi(\mathbf{X})/p$ are bounded from above and below. So $c + o_P(1) \leq \lambda_{Jd}(\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})/(pT)) \leq \lambda_1(\Phi(\mathbf{X})'\mathbf{\Sigma}_u^{1/2}\mathbf{E}\mathbf{E}'\mathbf{\Sigma}_u^{1/2}\Phi(\mathbf{X})/(pT)) \leq C + o_P(1)$. Finally we have,

$$c + o_P(1) \leq p\lambda_{K+j}(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT)) \leq C + o_P(1).$$

□

Define

$$\Theta_k = \frac{\lambda_k(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT))}{\lambda_{k+1}(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT))}.$$

Proof of Theorem 6.1

PROOF. By Lemma E.1, uniformly for $k = 1, ..., K-1$, $\Theta_k = O_P(1)$. By Lemma E.3, uniformly for $k = K+1, ..., Jd-K$,

$$c/p - o_P(1/p) < \lambda_k(\mathbf{Y}'\mathbf{P}\mathbf{Y}/(pT)) < C/p + o_P(1/p).$$

Hence uniformly for $k = K+1, ..., Jd-K-1$, $\Theta_k = O_P(1)$. When $k = K$,

$$\Theta_K \geq \tilde{c}p - o_P(1)$$

for some $\tilde{c} > 0$. Since $K < Jd/2$, we have $Jd/2 \leq Jd - K - 1$. Define

$$A = \{k \leq Jd/2 : k \neq K\}.$$

Then for any $\epsilon > 0$, there is $C > 0$, $P(\max_{k\in A}\Theta_k > C) < \epsilon$. Then

$$\begin{aligned} P(\widehat{K} \neq K) &\leq P(\max_{k\in A}\Theta_k \geq \Theta_K) \\ &\leq P(\Theta_K \leq \max_{k\in A}\Theta_k \leq C) + P(\max_{k\in A}\Theta_k > C) < \epsilon. \end{aligned}$$

It implies $P(\widehat{K} \neq K) \to 0$ because $\epsilon > 0$ is arbitrary.

□

E.3. Further Technical Lemmas.

LEMMA E.4. **(Vershynin, 2010)** *Let* $\mathbf{A}$ *be an* n *by* N *matrix whose columns* $\mathbf{A}_i$ *are independent* n*-dimensional sub-Gaussian random vectors with covariance* $\mathbf{\Sigma}$. *Then for every* $t \geq 0$, *with probability at least* $1 - 2\exp(-ct^2)$ *one has*

$$\left\| \frac{1}{N}\mathbf{A}\mathbf{A}' - \mathbf{\Sigma} \right\| \leq \max\{\delta, \delta^2\} \ \ \textit{where} \ \ \delta = C\sqrt{\frac{n}{N}} + \frac{t}{\sqrt{N}}.$$

Here $C, c > 0$ *depends only on the sub-Gaussian norm* $\max_i \|\mathbf{A}_i\|_{\phi_2}$ *of the rows.*

LEMMA E.5. **(Weyl inequality)** *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are two* $n \times n$ *symmetric matrices,*

$$\lambda_{i+j-1}(\mathbf{A} + \mathbf{B}) \leq \lambda_i(\mathbf{A}) + \lambda_j(\mathbf{B}),$$

for every $i, j \geq 1$ *and* $i + j - 1 \leq n$.

LEMMA E.6. **(Ahn and Horenstein, 2013)** *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $n \times n$ *positive definite and positive semidefinite matrices respectively, then*

$$\lambda_i(\mathbf{AB}) \leq \lambda_j(\mathbf{A})\lambda_k(\mathbf{B}), \quad \textit{for } j + k \leq i + 1$$

$$\lambda_i(\mathbf{AB}) \geq \lambda_j(\mathbf{A})\lambda_k(\mathbf{B}), \quad \textit{for } j + k \geq i + n.$$

In addition, if both $\mathbf{A}$ *and* $\mathbf{B}$ *are positive definite,*

$$\lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{A} + \mathbf{B}), \quad \textit{for } i = 1, 2, \ldots, n.$$

LEMMA E.7. **(Weyl's Theorem)** *Let* $\{\lambda_i\}_{i=1}^N$ *be the eigenvalues of* $\mathbf{\Sigma}$ *in descending order and* $\{\xi_i\}_{i=1}^N$ *be their associated eigenvectors. Correspondingly, let* $\{\widehat{\lambda}_i\}_{i=1}^N$ *be the eigenvalues of* $\widehat{\mathbf{\Sigma}}$ *in descending order and* $\{\widehat{\boldsymbol{\xi}}_i\}_{i=1}^N$ *be their associated eigenvectors. Then for all* $i \leq N$, $|\widehat{\lambda}_i - \lambda_i| \leq \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|$.

LEMMA E.8.

$$\lambda_k(\mathbf{AB}) \geq \lambda_k(\mathbf{A})\lambda_n(\mathbf{B}),$$

where $\mathbf{A}, \mathbf{B}$ *are* $n \times n$ *positive semidefinite and positive definite matrix respectively and* $k \leq n$ *(a special case of Lemma (E.6));*

$$\lambda_k(\mathbf{AB}) \geq \sigma_k(\mathbf{A})\sigma_m(\mathbf{B}),$$

where $\mathbf{A}$ *is* $n \times m$ *matrix of non-negative singular values,* $\mathbf{B}$ *is* $m \times n$ *matrix of full row rank, and* $k \leq m \leq n$ *(an easy extension of Lemma (E.6));*

$$\sigma_k(\mathbf{AB}) \geq \lambda_k(\mathbf{A})\sigma_m(\mathbf{B}),$$

where $\mathbf{A}$ *is* $n \times n$ *positive semidefinite matrix,* $\mathbf{B}$ *is* $n \times m$ *matrix of full column rank, and* $k \leq m \leq n$.

PROOF. We prove the third one by Courant-Fischer min-max theorem Tao (2012). By the theorem, for a n by n matrix,

$$\lambda_k(\mathbf{A}) = \inf_{\dim(\mathbf{V})=n-k+1} \sup_{\mathbf{v}\in\mathbf{V}:\|\mathbf{v}\|_2=1} \mathbf{v}'\mathbf{A}\mathbf{v},$$

where the inf is taken over all $\mathbf{V}$, a subspace of $\mathbb{R}^m$ such that $\dim(\mathbf{V}) = m - k + 1$. Therefore,

$$\sup_{\mathbf{v}\in\mathbf{V}} \frac{\mathbf{v}'\mathbf{B}'\mathbf{A}'\mathbf{A}\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{v}} = \sup_{\mathbf{v}\in\mathbf{V}} \frac{\mathbf{v}'\mathbf{B}'\mathbf{A}'\mathbf{A}\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{B}'\mathbf{B}\mathbf{v}} \frac{\mathbf{v}'\mathbf{B}'\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{v}}.$$

$$\geq \sup_{\mathbf{v}\in\mathbf{V}} \frac{\mathbf{v}'\mathbf{B}'\mathbf{A}'\mathbf{A}\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{B}'\mathbf{B}\mathbf{v}}(\sigma_m(\mathbf{B}))^2 = \sup_{\mathbf{v}\in\tilde{\mathbf{V}}} \frac{\mathbf{v}'\mathbf{A}'\mathbf{A}\mathbf{v}}{\mathbf{v}'\mathbf{v}}(\sigma_m(\mathbf{B}))^2,$$

where $\tilde{\mathbf{V}} = \mathbf{B}\mathbf{V} \subset \mathbb{R}^n$ and $\dim(\tilde{\mathbf{V}}) = m - k + 1$. Take inf with respect to $\tilde{\mathbf{V}}$,

$$\sigma_k(\mathbf{AB}) \geq \sqrt{\lambda_{k+n-m}(\mathbf{A}'\mathbf{A})}\sigma_m(\mathbf{B}) \geq \lambda_k(\mathbf{A})\sigma_m(\mathbf{B}).$$

$\square$

APPENDIX F: JUSTIFICATIONS OF SOME OF THE ASSUMPTIONS

In this section, we provide sufficient conditions and prove Assumptions 3.1 and 3.4(iii) imposed in the paper. In addition, we present more details about Example 3.1 in the main paper.

F.1. Justifications of Assumptions 3.1 and 3.4. Recall Assumption 3.1:

Assumption 3.1: There are positive constants $c_{\min}$ and $c_{\max}$ such that with probability approaching one,

$$c_{\min} < \lambda_{\min}(p^{-1}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda}) < \lambda_{\max}(p^{-1}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda}) < c_{\max}.$$

This condition requires the loading matrix not be "projected off" on the space spanned by the basis functions of the covariates. A sufficient condition is given in the following lemma.

LEMMA F.1. *Consider the semi-parametric factor model* $\mathbf{\Lambda} = \mathbf{G}(\mathbf{X}) + \mathbf{\Gamma}$. *Assumption 3.1 holds as long as there are constants* $c_1, c_2 > 0$ *such that with probability approaching one (as* $p, J \to \infty$*),*

$$c_1 < \lambda_{\min}(\frac{1}{p}\mathbf{G}(\mathbf{X})'\mathbf{G}(\mathbf{X})) \le \lambda_{\max}(\frac{1}{p}\mathbf{G}(\mathbf{X})'\mathbf{G}(\mathbf{X})) < c_2.$$

PROOF. For simplicity, we write $\mathbf{G} = \mathbf{G}(\mathbf{X})$, $\mathbf{\Phi} = \mathbf{\Phi}(\mathbf{X})$ and $\mathbf{R} = \mathbf{R}(\mathbf{X})$.

$$\frac{1}{p}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda} - \frac{1}{p}\mathbf{G}'\mathbf{G} = \frac{1}{p}\mathbf{G}'\mathbf{P}\mathbf{\Gamma} + \frac{1}{p}\mathbf{\Gamma}'\mathbf{P}\mathbf{G} + \frac{1}{p}\mathbf{\Gamma}'\mathbf{P}\mathbf{\Gamma} - \frac{1}{p}\mathbf{G}'(\mathbf{I} - \mathbf{P})\mathbf{G}.$$

We now show that all the terms on the right hand side are stochastically negligible. Note that $\mathbf{G} = \mathbf{\Phi}\mathbf{B} + \mathbf{R}$. By Lemma C.6,

$$\frac{1}{p}\mathbf{G}'\mathbf{P}\mathbf{\Gamma} = \frac{1}{p}\mathbf{B}'\mathbf{\Phi}'\mathbf{\Gamma} + \frac{1}{p}\mathbf{R}'\mathbf{P}\mathbf{\Gamma} = O_P(\sqrt{\frac{\nu_p}{pJ^{\kappa-1}}}).$$

Similarly, $\|\frac{1}{p}\mathbf{\Gamma}'\mathbf{P}\mathbf{G}\|_F = O_P(\sqrt{\frac{\nu_p}{pJ^{\kappa-1}}})$. $\frac{1}{p}\mathbf{\Gamma}'\mathbf{P}\mathbf{\Gamma} = \frac{1}{p}\|\mathbf{P}\mathbf{\Gamma}\|_F^2 = O_P(\frac{J\nu_p}{p})$. Finally, $\frac{1}{p}\mathbf{G}'(\mathbf{I} - \mathbf{P})\mathbf{G} = \frac{1}{p}\mathbf{R}'(\mathbf{I} - \mathbf{P})\mathbf{R} = O_P(J^{-\kappa})$. Define the event $A = \{\|\frac{1}{p}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda} - \frac{1}{p}\mathbf{G}'\mathbf{G}\|_F < c_1/2\}$. It then implies, A occurs with probability approaching one.

$$\begin{aligned} & P(\lambda_{\min}(\frac{1}{p}\mathbf{\Lambda}'\mathbf{P}'\mathbf{\Lambda}) > c_1/3) \ge P(\lambda_{\min}(\frac{1}{p}\mathbf{G}'\mathbf{G}) - \|\frac{1}{p}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda} - \frac{1}{p}\mathbf{G}'\mathbf{G}\|_F > \frac{c_1}{3}) \\ \ge\ & P(\lambda_{\min}(\frac{1}{p}\mathbf{G}'\mathbf{G}) > \frac{5c_1}{6}, A) \ge 1 - P(A^c) - P(\lambda_{\min}(\frac{1}{p}\mathbf{G}'\mathbf{G}) \le \frac{5c_1}{6}) = 1 - o(1). \end{aligned}$$

In addition,

$$\begin{aligned} & P(\lambda_{\max}(\frac{1}{p}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda}) < 2c_2) \ge P(\lambda_{\max}(\frac{1}{p}\mathbf{G}'\mathbf{G}) + \|\frac{1}{p}\mathbf{\Lambda}'\mathbf{P}\mathbf{\Lambda} - \frac{1}{p}\mathbf{G}'\mathbf{G}\|_F < 2c_2) \\ \ge\ & P(\lambda_{\max}(\frac{1}{p}\mathbf{G}'\mathbf{G}) < \frac{3c_2}{2}, A) \ge 1 - P(A^c) - P(\lambda_{\min}(\frac{1}{p}\mathbf{G}'\mathbf{G}) \le \frac{3c_2}{2}) = 1 - o(1). \end{aligned}$$

□

Assumption 3.4 (iii): Weak dependence: there is $C > 0$ so that

$$\max_{j \le p} \sum_{i=1}^{p} |Eu_{it}u_{jt}| < C, \quad \frac{1}{pT}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{t=1}^{T}\sum_{s=1}^{T} |Eu_{it}u_{js}| < C,$$

$$\max_{i \le p} \frac{1}{pT}\sum_{k=1}^{p}\sum_{m=1}^{p}\sum_{t=1}^{T}\sum_{s=1}^{T} |\mathrm{cov}(u_{it}u_{kt}, u_{is}u_{ms})| < C.$$

We verify this condition in the case when $\{u_{it}\}_{i\leq p}$ are mostly cross-sectionally independent, and serially independent. For $j=1,...,p$, define

$$A(j)=\{i\leq p: u_{it} \text{ is NOT independent of } u_{jt}\}.$$

Let $|A(i)|_0$ be the cardinality of $A(j)$. We assume $|A(i)|_0$ is bounded as $p\to\infty$, which also implies that each row of $\mathbf{\Sigma}_u$ is a sparse vector.

LEMMA F.2. *Suppose $\{\mathbf{u}_t\}_{t\leq T}$ are independent across t, and there is $C_1>0$, $\max_{j\leq p} Eu_{jt}^4<C_1$. Suppose as $p\to\infty$,*

$$\max_{j\leq p}|A(j)|_0<C_1.$$

Then Assumption 3.4(iii) is satisfied.

PROOF. First of all, $\max_{j\leq p}\sum_{i=1}^p|Eu_{it}u_{jt}|\leq C_1\max_{j\leq p}|A(j)|_0<C$. Secondly,

$$\frac{1}{pT}\sum_{i=1}^p\sum_{j=1}^p\sum_{t=1}^T\sum_{s=1}^T|Eu_{it}u_{js}|=\frac{1}{p}\sum_{i=1}^p\sum_{j=1}^p|Eu_{it}u_{jt}|\leq\max_{j\leq p}\sum_{i=1}^p|Eu_{it}u_{jt}|<C.$$

Finally, uniformly in $i\leq p$,

$$\begin{aligned}
&\frac{1}{pT}\sum_{k=1}^p\sum_{m=1}^p\sum_{t=1}^T\sum_{s=1}^T|\mathrm{cov}(u_{it}u_{kt},u_{is}u_{ms})|=\frac{1}{p}\sum_{k=1}^p\sum_{m=1}^p|\mathrm{cov}(u_{it}u_{kt},u_{it}u_{mt})|\\
=\;&\frac{1}{p}\sum_{k=1}^p\sum_{m\in A(i)}|\mathrm{cov}(u_{it}u_{kt},u_{it}u_{mt})|+\frac{1}{p}\sum_{k\in A(i)}\sum_{m\notin A(i)}|E(u_{it}^2u_{kt}u_{mt})|\\
&+\frac{1}{p}\sum_{k\notin A(i)}\sum_{m\notin A(i)}|Eu_{it}^2Eu_{kt}u_{mt}|\\
\leq\;& C|A(i)|_0+\frac{1}{p}\sum_{k\notin A(i)}\sum_{m\in A(k)}|Eu_{it}^2Eu_{kt}u_{mt}|+\frac{1}{p}\sum_{k\notin A(i)}\sum_{m\notin A(k)}|Eu_{it}^2Eu_{kt}Eu_{mt}|\\
=\;& C|A(i)|_0+\frac{1}{p}\sum_{k\notin A(i)}\sum_{m\in A(k)}|Eu_{it}^2Eu_{kt}u_{mt}|\leq C|A(i)|_0+C\max_{k\leq p}|A(k)|_0.
\end{aligned}$$

□

F.2. More details about Example 3.1. First of all, let us recall the setups of this example:

$$y_{it}=\beta f_t+\gamma_i f_t+u_{it}.$$

The projection in this case is averaging over i, which yields

$$\bar{y}_{\cdot t} = \beta f_t + \bar{\gamma}_{\cdot} f_t + \bar{u}_{\cdot t},$$

where $\bar{y}_{\cdot t}$, $\bar{\gamma}_{\cdot}$ and $\bar{u}_{\cdot t}$ denote the averages of their corresponding quantities over i. We obtain estimators

$$\widehat{\beta} = \left(\frac{1}{T}\sum_{t=1}^{T} \bar{y}_{\cdot t}^2\right)^{1/2}, \quad \text{and} \quad \widehat{f}_t = \bar{y}_{\cdot t}/\hat{\beta}.$$

Assumptions in Section 4 are all easy to verify. For simplicity, we shall assume u_{it}'s are i.i.d. in both i, t, γ_i's are i.i.d. in i and f_t's are i.i.d. in t. First of all, Assumption 4.1 (ii) holds as long as $\frac{1}{p}(\beta \mathbf{1}_p)'\beta\mathbf{1}_p = \beta^2 \neq 0$. Assumption 4.2 is satisfied since $\{\gamma_i\}$'s are i.i.d with $\nu_p = \mathrm{var}(\gamma_i)$. Finally, Assumption 4.3 is satisfied since $g(x) = \beta$ for all x. Once we set $J = 1$ and $\phi(x) = 1$, the approximation error in condition (ii) is zero. Hence we can ignore the term $J^{-\kappa}$. It then follows from Theorem 4.1 that

$$\text{(F.1)} \qquad \frac{1}{T}\sum_{t=1}^{T}(\widehat{f}_t - f_t)^2 = O_P(\frac{1}{p}), \quad |\widehat{\beta} - \beta|^2 = O_P(\frac{1}{p^2} + \frac{1}{pT} + \frac{\mathrm{var}(\gamma_i)}{p}).$$

To intuitively understand that these rates are sharp, we directly analyze the expansions of $\widehat{\beta}$ and $\widehat{f}_t$. Note that

$$\frac{1}{T}\sum_{t=1}^{T} \bar{y}_{\cdot t}^2 = \beta^2 + \Delta, \quad \Delta = 2\beta\bar{\gamma}_{\cdot} + \bar{\gamma}_{\cdot}^2 + \frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2 + \frac{2}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t} f_t(\beta + \bar{\gamma}_{\cdot}).$$

Here $\bar{\gamma}_{\cdot} = O_P(\sqrt{\frac{\mathrm{var}(\gamma_i)}{p}})$, $\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2 = O_P(\mathrm{var}(\bar{u}_{\cdot,t})) = O_P(\frac{1}{p})$, and $\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t} f_t = O_P(\sqrt{\mathrm{var}(\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t} f_t)}) = O_P(\sqrt{\frac{1}{T}\mathrm{var}(\bar{u}_{\cdot t}) E f_t^2}) = O_P(\frac{1}{\sqrt{pT}})$. Combining these yields:

$$|\widehat{\beta}^2 - \beta^2| = \Delta = O_P(\sqrt{\frac{\mathrm{var}(\gamma_i)}{p}} + \frac{1}{p} + \frac{1}{\sqrt{pT}}) = |\widehat{\beta} - \beta|.$$

Moreover,

$$\frac{\bar{y}_{\cdot t}}{\widehat{\beta}} - f_t = f_t(\frac{\beta}{\widehat{\beta}} - 1) + (\bar{\gamma}_{\cdot} f_t + \bar{u}_{\cdot t})\frac{1}{\widehat{\beta}}.$$

This implies

$$\frac{1}{T}\sum_{t=1}^{T}(\widehat{f}_t - f_t)^2 = (\frac{\beta}{\widehat{\beta}} - 1)^2 + \frac{1}{T}\sum_{t=1}^{T}(\bar{\gamma}_{\cdot} f_t + \bar{u}_{\cdot t})^2\frac{1}{\widehat{\beta}^2} + \frac{2}{T}\sum_{t=1}^{T} f_t(\bar{\gamma}_{\cdot} f_t + \bar{u}_{\cdot t})\frac{1}{\widehat{\beta}}(\frac{\beta}{\widehat{\beta}} - 1).$$

Note that $(\frac{\beta}{\widehat{\beta}}-1)=O_P(|\beta-\widehat{\beta}|)$, $\frac{1}{T}\sum_{t=1}^T(\bar{\gamma}_\cdot f_t+\bar{u}_{\cdot t})^2=O_P(\frac{1}{p})$, and $\frac{2}{T}\sum_{t=1}^T f_t(\bar{\gamma}_\cdot f_t+\bar{u}_{\cdot t})=\sqrt{\frac{\mathrm{var}(\gamma_i)}{p}}+\frac{1}{\sqrt{pT}}$. So

$$\frac{1}{T}\sum_{t=1}^T(\widehat{f}_t-f_t)^2=O_P(|\beta-\widehat{\beta}|^2)+O_P(\frac{1}{p})+(\sqrt{\frac{\mathrm{var}(\gamma_i)}{p}}+\frac{1}{\sqrt{pT}})O_P(|\widehat{\beta}-\beta|)=O_P(\frac{1}{p}),$$

which also matches with (F.1).

APPENDIX G: TIME-VARIANT COVARIATES

We provide heuristic discussions on the case when the covariates are time dependent. For simplicity, consider $d=1$. Let $\phi(x_{it})$ be $J\times 1$ dimensional basis functions. Now let $\mathbf{Z}_t$ be $p\times J$ matrix of $\phi(x_{it})$. Write $\mathbf{y}_t=(y_{1t},...,y_{pt})'$ and $\mathbf{x}_t=(x_{1t},...,x_{pt})'$. Projecting $\mathbf{y}_t$ onto the space generated by $\mathbf{x}_t$, we obtain the projected vector $\widehat{\mathbf{y}}_t=\mathbf{P}_t\mathbf{y}_t$, where $\mathbf{P}_t=\mathbf{Z}_t(\mathbf{Z}_t'\mathbf{Z}_t)^{-1}\mathbf{Z}_t'$. Here $\widehat{\mathbf{y}}_t$ nonparametrically approximates $E(\mathbf{y}_t|x_t)$. Let the projected data matrix be $\widehat{\mathbf{Y}}=(\widehat{\mathbf{y}}_1,...,\widehat{\mathbf{y}}_T)$, a $p\times T$ matrix.

We estimate the factors using the projected PCA based on the projected data. The columns of $\widehat{\mathbf{F}}$ are the leading eigenvectors of $\frac{1}{pT}\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}}$.

The key assumption on the projection property is the follows:

Assumption G.1. *Let $m_{it}=E(\boldsymbol{\lambda}_i'\mathbf{f}_t|x_{it})$, and let $\mathbf{M}=(m_{it})_{p\times T}$ be a $p\times T$ matrix. With probability approaching one, the first K eigenvalues of $\frac{1}{pT}\mathbf{M}'\mathbf{M}$ are bounded away from both zero and infinity, as $p\to\infty$ (J,T may either grow or stay constant, but $T>K$).*

This assumption requires $\boldsymbol{\lambda}_i'\mathbf{f}_t$ be not "projected off" by $E(.|x_{it})$, and are strongly associated with the covariates. Technically, it guarantees that the first K eigenvalues of $\frac{1}{pT}\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}}$ are bounded away from both zero and infinity with a probability arbitrarily close to one.

Let $\mathbf{V}$ be a $K\times K$ diagonal matrix, consisting of the first K eigenvalues of $\frac{1}{pT}\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}}$. Then we have

$$\frac{1}{pT}\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}}\widehat{\mathbf{F}}=\widehat{\mathbf{F}}\mathbf{V}.$$

This also implies, for $t=1,...,T$, $\frac{1}{pT}\widehat{\mathbf{y}}_t'\sum_{s=1}^T\widehat{\mathbf{y}}_s\widehat{\mathbf{f}}_s'=\widehat{\mathbf{f}}_t'\mathbf{V}$. Note that $\mathbf{y}_t=\boldsymbol{\Lambda}\mathbf{f}_t+\mathbf{u}_t$. Thus

$$\widehat{\mathbf{y}}_t=\mathbf{P}_t\mathbf{y}_t=\mathbf{P}_t\boldsymbol{\Lambda}\mathbf{f}_t+\mathbf{P}_t\mathbf{u}_t.$$

Let

$$\mathbf{H}_t=\frac{1}{pT}\mathbf{V}^{-1}\sum_{s=1}^T\widehat{\mathbf{f}}_s\mathbf{f}_s'\boldsymbol{\Lambda}'\mathbf{P}_s\mathbf{P}_t\boldsymbol{\Lambda}$$

It then can be verified that

$$\widehat{\mathbf{f}}_t' - \mathbf{f}_t'\mathbf{H}_t' = \frac{1}{pT}\mathbf{f}_t'\mathbf{\Lambda}'\mathbf{P}_t\sum_{s=1}^T \mathbf{P}_s\mathbf{u}_s\widehat{\mathbf{f}}_s'\mathbf{V}^{-1} + \frac{1}{pT}\mathbf{u}_t'\mathbf{P}_t\sum_{s=1}^T \mathbf{P}_s\mathbf{\Lambda}\mathbf{f}_s\widehat{\mathbf{f}}_s'\mathbf{V}^{-1} + \frac{1}{pT}\mathbf{u}_t'\mathbf{P}_t\sum_{s=1}^T \mathbf{P}_s\mathbf{u}_s\widehat{\mathbf{f}}_s'\mathbf{V}^{-1}.$$

Under regularity conditions, it can be proved that

$$\frac{1}{T}\sum_{t=1}^T \|\widehat{\mathbf{f}}_t - \mathbf{H}_t\mathbf{f}_t\|^2 = O_P(\frac{J}{p}).$$

We note that when x_{it} depends on t, the estimated factor $\widehat{\mathbf{f}}_t$ based on the projected PCA consistently estimates a transformed $\mathbf{f}_t$, where the transformation matrix $\mathbf{H}_t$ also depends on t through the projection matrix $\mathbf{P}_t$. The rate of convergence does not depend on T, and hence T can still be finite for the consistency (of estimating the transformed factors).

Under additional suitable smoothness conditions on the projection matrix $\mathbf{P}_t$ over t, it is possible to estimate $\mathbf{f}_t$ directly with the same identification conditions as those in the main paper. The key step is to "smooth" $\mathbf{H}_t$, such that the transformation matrix is approximately independent of t. This can be achieved by the following assumption:

ASSUMPTION G.2. *For each $s \leq T$, there is a time-independent $p \times p$ matrix $\bar{\mathbf{P}}_s$ and $h_T \to 0$ such that*

$$\frac{1}{Th_T}\sum_{t=1}^T \|\mathbf{P}_t - \bar{\mathbf{P}}_s\|_2^2 K\left(\frac{t-s}{h_T}\right) = o_P(1).$$

Here $K(\cdot)$ is a regular kernel function satisfying $K(t) \to 0$ as $|t| \to \infty$, and h_T is the bandwidth. This assumption essentially means $\mathbf{P}_t$ is approximately independent of t in each local window. Note that $\mathbf{H}_t = \mathbf{M}\mathbf{P}_t\mathbf{\Lambda}$, where $\mathbf{M} = \frac{1}{pT}\mathbf{V}^{-1}\sum_{s=1}^T \widehat{\mathbf{f}}_s\mathbf{f}_s'\mathbf{\Lambda}'\mathbf{P}_s$. Hence the transformation matrix $\mathbf{H}_t$ is approximately locally constant. With the estimated factors, we can proceed to estimate the loadings, and investigate the convergence properties as in the main paper. Essentially, this corresponds to applying the projected PCA on the local windows of the data, based on kernel smoothing.

All the above arguments can be made rigorous, though detailed derivations are out of the scope of this paper, and may be investigated in the future.

REFERENCES

Athreya, K. and Hahiri, S. (2006). *Measure Theory and Probability Theory.* The first edition ed. Springer, New York,.

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

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.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis* **14** 1–16.

Tao, T. (2012). *Topics in Random Matrix Theory.* The first edition ed. American Mathematical Society, Volume 132.

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

Address:
Department of ORFE,
Sherrerd Hall, Princeton University,
Princeton, NJ 08544, USA.
E-mail: jqfan@princeton.edu
weichenw@princeton.edu

Address:
Department of Mathematics,
University of Maryland.
College Park, MD 20742, USA. E-mail: yuanliao@umd.edu