Statistical inference for large-dimensional tensor factor model by iterative projections
Supplementary material

Matteo Barigozzi[a], Yong He[b], Lingxiao Li[c], Lorenzo Trapani[d,*]

[a]*Department of Economics, Università di Bologna*
[b]*Institute for Financial Studies, Shandong University*
[c]*Department of Data Science and Artificial Intelligence, the Hong Kong Polytechnic University*
[d]*Department of Economics and Management, Università di Pavia*

Appendix A. Technical lemmas

Throughout this appendix, $C_1, C_2, \ldots$ denote some positive constants that do not depend on p_1, p_2, T.

We begin with a series of results which are a direct consequence of Assumptions 1-4, and which will be used throughout the whole paper.

Lemma 2. *We assume that Assumptions 1-4 hold. Then it holds that*

(i)

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{E}\left\|\mathbf{E}_{k,t}\mathbf{A}_k\right\|_F^2 &= O(Tp), \\ \sum_{t=1}^{T} \mathbb{E}\left\|\mathbf{E}_{k,t}\mathbf{B}_k\right\|_F^2 &= O(Tp);\end{aligned}$$

(ii)

$$\begin{aligned}\mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\right\|_F^2 &= O(Tp), \\ \mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{F}_{k,t}^\top\mathbf{A}_k^\top\mathbf{E}_{k,t}\right\|_F^2 &= O(Tp), \\ \mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\right\|_F^2 &= O(Tp), \\ \mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{F}_{k,t}^\top\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{B}_k\right\|_F^2 &= O(Tp);\end{aligned}$$

(iii) for all $1 \le i_k \le p_k$

$$\begin{aligned}\mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\mathbf{e}_{t,k,i\cdot}^\top\right\|^2 &= O(Tp) + O\left((Tp_{-k})^2\right), \\ \mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{e}_{t,k,i\cdot}^\top\right\|^2 &= O(Tp) + O\left((Tp_{-k})^2\right),\end{aligned}$$

*Corresponding author. Email address: `lorenzo.trapani@unipv.it`

and

$$\mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\mathbf{E}_{k,t}^{\top}\right\|_F^2 = O\left(Tp_k^2p_{-k}+T^2p_kp_{-k}^2\right),$$
$$\mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\mathbf{E}_{k,t}^{\top}\mathbf{A}_k\right\|_F^2 = O\left(Tp_k^2p_{-k}+T^2p_kp_{-k}^2\right).$$

Proof. Given that $\{r_k, 1\le k\le K\}$ are fixed, we show the results $r_k=1$ for all $1\le k\le K$, for simplicity and without loss of generality; thus, henceforth $\mathbf{A}_k$ is a p_k-dimensional vector with entries A_{ki}. We begin with part (i); using Assumptions 2(i) and 3(ii), and equations (23)-(24), it follows that

$$\sum_{t=1}^{T}\mathbb{E}\left[\left\|\mathbf{E}_{k,t}^{\top}\mathbf{A}_k\right\|_F^2\right]=\sum_{t=1}^{T}\mathbb{E}\left[\sum_{j=1}^{p_{-k}}\left(\boldsymbol{e}_{k,t,\cdot j}^{\top}\mathbf{A}_k\right)^2\right]=\sum_{t=1}^{T}\sum_{j=1}^{p_{-k}}\mathbb{E}\left[\left(\sum_{i=1}^{p_k}e_{k,t,ij}A_{k,i}\right)^2\right]$$
$$\le \bar{a}_k^2\sum_{t=1}^{T}\sum_{j=1}^{p_{-k}}\sum_{i=1}^{p_k}\sum_{i_1=1}^{p_k}\mathbb{E}\left[e_{k,t,ij}e_{k,t,i_1j}\right]=O\left(Tp\right).$$

The second equation can be proven by similar passages. As far as part (ii) is concerned, it follows directly from Assumption 4(i). Finally, we consider part (iii). Using Assumptions 3(ii)-(iii) and (24), we have

$$\begin{aligned}
&\mathbb{E}\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\mathbf{e}_{t,k,i\cdot}^{\top}\right\|^2\\
=\;&\mathbb{E}\left(\sum_{i_1=1}^{p_k}\left(\sum_{t=1}^{T}\sum_{j=1}^{p_{-k}}e_{t,k,ij}e_{t,k,i_1j}\right)^2\right)\\
\le\;&2\sum_{i_1=1}^{p_k}\mathbb{E}\left(\sum_{t=1}^{T}\sum_{j=1}^{p_{-k}}\left(e_{t,k,ij}e_{t,k,i_1j}-\mathbb{E}\left(e_{t,k,ij}e_{t,k,i_1j}\right)\right)\right)^2+2\sum_{i_1=1}^{p_k}\left(\sum_{t=1}^{T}\sum_{j=1}^{p_{-k}}\mathbb{E}\left(e_{t,k,ij}e_{t,k,i_1j}\right)\right)^2\\
=\;&2\sum_{i_1=1}^{p_k}\sum_{t,s=1}^{T}\sum_{j_1,j_2=1}^{p_{-k}}Cov\left(e_{t,k,ij_1}e_{t,k,i_1j_1},e_{s,k,ij_2}e_{s,k,i_1j_2}\right)\\
&+2\sum_{i_1=1}^{p_k}\sum_{t,s=1}^{T}\sum_{j_1,j_2=1}^{p_{-k}}\left|\mathbb{E}\left(e_{t,k,ij_1}e_{t,k,i_1j_1}\right)\right|\left|\mathbb{E}\left(e_{s,k,ij_2}e_{s,k,i_1j_2}\right)\right|\\
\le\;&c_0Tp+c_1Tp_{-k}\sum_{i_1=1}^{p_k}\sum_{t=1}^{T}\sum_{j_1=1}^{p_{-k}}\left|\mathbb{E}\left(e_{t,k,ij_1}e_{t,k,i_1j_1}\right)\right|\\
\le\;&c_0Tp+c_1\left(Tp_{-k}\right)^2.
\end{aligned}$$

Similarly, using Assumptions 2(i) and 3(ii)-(iii), and (24), it follows that

$$
\begin{aligned}
& \mathbb{E}\left\|\sum_{t=1}^{T} \mathbf{A}_k^{\top} \mathbf{E}_{k,t} \mathbf{e}_{t,k,i\cdot}^{\top}\right\|^2 \\
= & \ \mathbb{E}\left(\sum_{i_1=1}^{p_k} \sum_{t=1}^{T} \sum_{j=1}^{p_{-k}} A_{k,i_1} e_{t,k,ij} e_{t,k,i_1 j}\right)^2 \\
\leq & \ 2\mathbb{E}\left(\sum_{i_1=1}^{p_k} \sum_{t=1}^{T} \sum_{j=1}^{p_{-k}} A_{k,i_1}\left(e_{t,k,ij} e_{t,k,i_1 j} - \mathbb{E}\left(e_{t,k,ij} e_{t,k,i_1 j}\right)\right)\right)^2 + 2\left(\sum_{i_1=1}^{p_k} \sum_{t=1}^{T} \sum_{j=1}^{p_{-k}} A_{k,i_1} \mathbb{E}\left(e_{t,k,ij} e_{t,k,i_1 j}\right)\right)^2 \\
\leq & \ 2 \sum_{i_1,i_2=1}^{p_k} \sum_{t,s=1}^{T} \sum_{j_1,j_2=1}^{p_{-k}} \left|A_{k,i_1}\right|\left|A_{k,i_2}\right| Cov\left(e_{t,k,ij_1} e_{t,k,i_1 j_1}, e_{s,k,ij_2} e_{s,k,i_2 j_2}\right) \\
& +2 \sum_{i_1,i_2=1}^{p_k} \sum_{t,s=1}^{T} \sum_{j_1,j_2=1}^{p_{-k}} \left|A_{k,i_1}\right|\left|A_{k,i_2}\right| \left|\mathbb{E}\left(e_{t,k,ij_1} e_{t,k,i_1 j_1}\right)\right| \left|\mathbb{E}\left(e_{s,k,ij_2} e_{s,k,i_2 j_2}\right)\right| \\
\leq & \ c_0 T p + c_1 p_k \left(T p_{-k}\right)^2 .
\end{aligned}
$$

The proof of the remaining two result is essentially the same, and we omit it to save space. □

We now consider a set of results pertaining to $\widehat{\mathbf{M}}_k$. We will extensively use the following decomposition, which is a direct consequence of (17)

$$
\begin{aligned}
\widehat{\mathbf{M}}_k &= \frac{1}{Tp} \sum_{t=1}^{T} \mathbf{X}_{k,t} \mathbf{X}_{k,t}^{\top} \qquad (A.1) \\
&= \frac{1}{Tp} \sum_{t=1}^{T} \left(\mathbf{A}_k \mathbf{F}_{k,t} \mathbf{B}_k^{\top} + \mathbf{E}_{k,t}\right)\left(\mathbf{A}_k \mathbf{F}_{k,t} \mathbf{B}_k^{\top} + \mathbf{E}_{k,t}\right)^{\top} \\
&= \frac{1}{Tp} \sum_{t=1}^{T} \mathbf{A}_k \mathbf{F}_{k,t} \mathbf{B}_k^{\top} \mathbf{B}_k \mathbf{F}_{k,t}^{\top} \mathbf{A}_k^{\top} + \frac{1}{Tp} \sum_{t=1}^{T} \mathbf{A}_k \mathbf{F}_{k,t} \mathbf{B}_k^{\top} \mathbf{E}_{k,t}^{\top} \\
&\quad + \frac{1}{Tp} \sum_{t=1}^{T} \mathbf{E}_{k,t} \mathbf{B}_k \mathbf{F}_{k,t}^{\top} \mathbf{A}_k^{\top} + \frac{1}{Tp} \sum_{t=1}^{T} \mathbf{E}_{k,t} \mathbf{E}_{k,t}^{\top} \\
&= \mathcal{I} + \mathcal{II} + \mathcal{III} + \mathcal{IV}.
\end{aligned}
$$

Lemma 3. *We assume that Assumptions 1-4 hold. Then it holds that, as* $\min\{T, p_1, ..., p_K\} \to \infty$

$$
\lambda_j\left(\widehat{\mathbf{M}}_k\right) = \begin{cases} \lambda_j\left(\mathbf{\Sigma}_k\right) + o_P(1) & j \leq r_k \\ O_P\left(\left(T p_{-k}\right)^{-1/2}\right) + O_P\left(p_{-k}^{-1}\right) & j > r_k \end{cases} .
$$

Proof. As shown in equation (A.1), $\widehat{\mathbf{M}}_k = \mathcal{I} + \mathcal{II} + \mathcal{III} + \mathcal{IV}$. We will study the spectral norms of these four terms and show that $\mathcal{I}$ is the dominant one. Firstly, by Assumptions 1 1(ii) and 2(ii), we have

$$
\begin{aligned}
\frac{1}{T p_{-k}} \sum_{t=1}^{T} \mathbf{F}_{k,t} \mathbf{B}_k^{\top} \mathbf{B}_k \mathbf{F}_{kt}^{\top} &\xrightarrow{P} \mathbf{\Sigma}_k, \\
\mathcal{I} &\xrightarrow{P} \frac{\mathbf{A}_k \mathbf{\Sigma}_k \mathbf{A}_k^{\top}}{p_k},
\end{aligned}
$$

while the leading r_k eigenvalues of $p_k^{-1} \mathbf{A}_k \mathbf{\Sigma}_k \mathbf{A}_k^{\top}$ are asymptotically equal to those of $\mathbf{\Sigma}_k$. Hence, $\lambda_j(\mathcal{I}) = \lambda_j\left(\mathbf{\Sigma}_k\right) + o_P(1)$ for $j \leq r_k$ while $\lambda_j(\mathcal{I}) = 0$ for $j > r_k$ because $\text{rank}(\mathcal{I}) \leq r_k$. Secondly, using the Cauchy-Schwartz inequality, Assumption 2(ii) and Lemma 2*(ii)*,

$$
\begin{aligned}
\|\mathcal{II}\|_F &\leq \frac{1}{Tp}\|\mathbf{A}_k\|_F\left\|\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\right\|_F \\
&\lesssim \frac{1}{\sqrt{Tp_{-k}}}\left\|\frac{1}{\sqrt{Tp}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\right\|_F \\
&= O_P\left(\frac{1}{\sqrt{Tp_{-k}}}\right).
\end{aligned}
$$

Exactly the same logic also yields $\|\mathcal{III}\|_F = O_P\left((Tp_{-k})^{-1/2}\right)$. Finally, consider $\mathcal{IV}$, and let $\mathbf{U}_{\mathbf{E}} = (Tp)^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{E}_{k,t}\mathbf{E}_{k,t}^\top\right)$; then, using Assumption 3(iii), it follows that

$$
\begin{aligned}
&\mathbb{E}\|\mathcal{IV}-\mathbf{U}_{\mathbf{E}}\|_F^2 \\
=\ & \frac{1}{(Tp)^2}\sum_{i_1,i_2=1}^{p_k}\mathbb{E}\left(\sum_{t=1}^{T}\sum_{j=1}^{p_{-k}}\left(e_{t,k,i_1j}e_{t,k,i_2j}-\mathbb{E}\left(e_{t,k,i_1j}e_{t,k,i_2j}\right)\right)\right)^2 \\
=\ & \frac{1}{(Tp)^2}\sum_{i_1,i_2=1}^{p_k}\sum_{t,s=1}^{T}\sum_{j_1,j_2=1}^{p_{-k}}Cov\left(e_{t,k,i_1j_1}e_{t,k,i_2j_1},e_{s,k,i_1j_2}e_{s,k,i_2j_2}\right) = O\left(\frac{1}{Tp_{-k}}\right).
\end{aligned}
$$

Further, by Assumption 3*(ii)* and (24)

$$
\begin{aligned}
\|\mathbf{U}_{\mathbf{E}}\|_1 &= \|\mathbf{U}_{\mathbf{E}}\|_\infty \\
&= \frac{1}{Tp}\max_{1\leq i\leq p_k}\sum_{i_1=1}^{p_k}\left|\sum_{t=1}^{T}\sum_{j=1}^{p_{-k}}\mathbb{E}\left(e_{t,k,i_1j}e_{t,k,i_2j}\right)\right| \\
&\leq \max_{1\leq i\leq p_k}\sum_{i_1=1}^{p_k}\sum_{t=1}^{T}\sum_{j=1}^{p_{-k}}\left|\mathbb{E}\left(e_{t,k,i_1j}e_{t,k,i_2j}\right)\right| = O\left(p_k^{-1}\right),
\end{aligned}
$$

which also entails that $\|\mathbf{U}_{\mathbf{E}}\|_F = O\left(p_k^{-1}\right)$. Hence it follows that $\|\mathcal{IV}\|_F = O_P\left((Tp_{-k})^{-1/2}\right) + O\left(p_k^{-1}\right)$. The proof now follows from (repeated applications of) Weyl's inequality. □

Lemma 4. *We assume that Assumptions 1-4 hold. Then it holds that, as* $\min\{T, p_1, ..., p_K\} \to \infty$

$$
\begin{aligned}
\frac{1}{p_k}\left\|\mathcal{II}\widehat{\mathbf{A}}_k\right\|_F^2 &= O_P\left(\frac{1}{Tp_{-k}}\right), \\
\frac{1}{p_k}\left\|\mathcal{III}\widehat{\mathbf{A}}_k\right\|_F^2 &= O_P\left(\frac{1}{Tp_{-k}}\right), \\
\frac{1}{p_k}\left\|\mathcal{IV}\widehat{\mathbf{A}}_k\right\|_F^2 &= O_P\left(\frac{1}{Tp}\right) + O_P\left(\frac{1}{p_k^2}\right) + O_P\left(\frac{1}{Tp_{-k}}+\frac{1}{p_k}\right)\times\frac{1}{p_k}\left\|\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right\|_F^2,
\end{aligned}
$$

where recall that, by (A.1), $\widehat{\mathbf{M}}_k = \mathcal{I} + \mathcal{II} + \mathcal{III} + \mathcal{IV}$.

Proof. We begin by noting that, from equation (A.1) and Lemma 2*(ii)*, it follows that

$$
\begin{aligned}
\frac{1}{p_k}\left\|\mathcal{II}\widehat{\mathbf{A}}_k\right\|_F^2 &= \frac{1}{p_k}\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\widehat{\mathbf{A}}_k\right\|_F^2 \\
&\leq \frac{1}{p_k}\|\mathbf{A}_k\|_F^2\left\|\widehat{\mathbf{A}}_k\right\|_F^2\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\right\|_F^2 \\
&= O_P\left(\frac{1}{Tp_{-k}}\right),
\end{aligned}
$$

having used Assumption 2(ii) and the fact that, by construction, $\left\|\widehat{\mathbf{A}}_k\right\|_F^2 = c_0 p_k$. By the same token, it is easy to see that

$$\begin{aligned}
\frac{1}{p_k}\left\|\mathcal{III}\widehat{\mathbf{A}}_k\right\|_F^2 &= \frac{1}{p_k}\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{k,t}\mathbf{B}_k\mathbf{F}_{k,t}^\top\mathbf{A}_k^\top\widehat{\mathbf{A}}_k\right\|_F^2 \\
&\leq \frac{1}{p_k}\left\|\mathbf{A}_k\right\|_F^2\left\|\widehat{\mathbf{A}}_k\right\|_F^2\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{k,t}\mathbf{B}_k\mathbf{F}_{k,t}^\top\right\|_F^2 \\
&= O_P\left(\frac{1}{Tp_{-k}}\right).
\end{aligned}$$

Finally, using Lemma 2*(iii)* and recalling that $\left\|\widehat{\mathbf{H}}_k\right\|_F^2 = O_P(1)$, it follows that

$$\begin{aligned}
&\frac{1}{p_k}\left\|\mathcal{IV}\widehat{\mathbf{A}}_k\right\|_F^2 \\
=\ & \frac{1}{p_k}\left\|\frac{1}{Tp}\sum_{t=1}^{T}\left(\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k+\mathbf{A}_k\widehat{\mathbf{H}}_k\right)^\top\mathbf{E}_{k,t}\mathbf{E}_{k,t}^\top\right\|_F^2 \\
\leq\ & \frac{1}{p_k}\sum_{i=1}^{p_k}\left\|\frac{1}{Tp}\sum_{t=1}^{T}\widehat{\mathbf{H}}_k^\top\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{e}_{t,k,i\cdot}+\frac{1}{Tp}\sum_{t=1}^{T}\left(\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right)^\top\mathbf{E}_{k,t}\mathbf{e}_{t,k,i\cdot}\right\|_F^2 \\
\leq\ & O_P(1)\frac{2}{p_k}\sum_{i=1}^{p_k}\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{e}_{t,k,i\cdot}\right\|_F^2+\frac{2}{p_k}\sum_{i=1}^{p_k}\left\|\frac{1}{Tp}\sum_{t=1}^{T}\left(\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right)^\top\mathbf{E}_{k,t}\mathbf{e}_{t,k,i\cdot}\right\|_F^2 \\
=\ & O_P\left(\frac{1}{Tp}+\frac{1}{p_k^2}\right)+O_P\left(\frac{1}{Tp_{-k}}+\frac{1}{p_k}\right)\times\frac{1}{p_k}\left\|\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right\|_F^2,
\end{aligned}$$

having used Assumption 3. □

We now report a set of results concerning $\widetilde{\mathbf{M}}_k$. By definition it holds that

$$\begin{aligned}
\widetilde{\mathbf{M}}_k &= \frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^{T}\mathbf{X}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{X}_{k,t}^\top \qquad (A.2)\\
&= \frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^{T}\left(\mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^\top+\mathbf{E}_{k,t}\right)\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\left(\mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^\top+\mathbf{E}_{k,t}\right)^\top \\
&= \frac{1}{Tpp_{-k}}\sum_{t=1}^{T}\mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^\top\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\mathbf{A}_k^\top+\frac{1}{Tpp_{-k}}\sum_{t=1}^{T}\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\mathbf{A}_k^\top \\
&\quad+\frac{1}{Tpp_{-k}}\sum_{t=1}^{T}\mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^\top\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{E}_{k,t}^\top+\frac{1}{Tpp_{-k}}\sum_{t=1}^{T}\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{E}_{k,t}^\top \\
&= \mathcal{V}+\mathcal{VI}+\mathcal{VII}+\mathcal{VIII}.
\end{aligned}$$

Lemma 5. *We assume that Assumptions 1-4 and (26) hold. Then, as* $\min\{T, p_1, ..., p_K\}\to\infty$*, it holds that*

$$\lambda_j\left(\widetilde{\mathbf{M}}_k\right) = \lambda_j\left(\mathbf{\Sigma}_k\right)+o_P(1),$$

for all $j \leq r_k$.

Proof. Recall that, from equation (A.2), it holds that

$$\widetilde{\mathbf{M}}_k = \mathcal{V}+\mathcal{VI}+\mathcal{VII}+\mathcal{VIII}.$$

We have

$$
\begin{aligned}
\mathcal{V} &= \frac{1}{Tp_kp_{-k}^2}\mathbf{A}_k\left(\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{B}_k^\top\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\right)\mathbf{A}_k^\top \\
&= \frac{1}{Tp_kp_{-k}^2}\mathbf{A}_k\left(\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{B}_k^\top\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}+\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}+\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\right)\mathbf{A}_k^\top \\
&= \frac{1}{Tp_kp_{-k}^2}\mathbf{A}_k\left(\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{B}_k^\top\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\right)\mathbf{A}_k^\top \\
&\quad+\frac{1}{Tp_kp_{-k}^2}\mathbf{A}_k\left(\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{B}_k^\top\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\right)\mathbf{A}_k^\top \\
&\quad+\frac{1}{Tp_kp_{-k}^2}\mathbf{A}_k\left(\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{B}_k^\top\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\right)\mathbf{A}_k^\top \\
&\quad+\frac{1}{Tp_kp_{-k}^2}\mathbf{A}_k\left(\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{B}_k^\top\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\right)\mathbf{A}_k^\top.
\end{aligned}
$$

We begin by noting that

$$\widehat{\mathbf{H}}_{-k}\widehat{\mathbf{H}}_{-k}^\top \xrightarrow{P} \mathbf{I}_{r_{-k}};$$

then, standard passages based on (25-(26), Assumption 1(ii) and Weyl's inequality, we have

$$\lambda_j(\mathcal{V}) = \lambda_j\left(p_k^{-1}\mathbf{A}_k\mathbf{\Sigma}_k\mathbf{A}_k^\top\right)+o_P(1),$$

for $j \le r_k$. Further, since $\mathrm{rank}(\mathcal{V}) \le r_k$, it follows immediately that $\lambda_j(\mathcal{V}) = 0$ for all $j > r_k$. Also, using again (25)-(26) and Lemma 2(ii)

$$
\begin{aligned}
\|\mathcal{VI}\|_F &= \frac{1}{Tp_kp_{-k}^2}\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\mathbf{A}_k^\top\right\|_F \\
&\lesssim \frac{1}{Tp_kp_{-k}}\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\mathbf{F}_{k,t}^\top\mathbf{A}_k^\top\right\|_F \\
&\lesssim \frac{1}{Tp_kp_{-k}}\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\mathbf{F}_{k,t}^\top\right\|_F\|\mathbf{A}_k\|_F+\frac{1}{Tp_kp_{-k}}\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\mathbf{F}_{k,t}^\top\mathbf{A}_k^\top\right\|_F \\
&= O_P\left(m_{-k}^{1/2}\right)+O_P\left((Tp_{-k})^{-1/2}\right) = o_P(1).
\end{aligned}
$$

The same logic also yields

$$\|\mathcal{VII}\|_F = O_P\left(w_{-k}^{1/2}\right)+O_P\left((Tp_{-k})^{-1/2}\right) = o_P(1).$$

Finally, using (25)-(26) and Lemma 2*(i)*

$$
\begin{aligned}
\|\mathcal{VIII}\|_F &= \frac{1}{Tp_k p_{-k}^2}\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k^\top \widehat{\mathbf{B}}_k \mathbf{E}_{k,t}^\top\right\|_F \\
&\lesssim \frac{1}{Tp_k p_{-k}^2}\sum_{t=1}^T \left\|\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\right\|_F^2 \\
&\lesssim \frac{1}{Tp_k p_{-k}^2}\sum_{t=1}^T \left\|\mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\right\|_F^2 + \frac{1}{Tp_k p_{-k}^2}\sum_{t=1}^T \left\|\mathbf{E}_{k,t}\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2 \\
&\lesssim \frac{1}{Tp_k p_{-k}^2}\sum_{t=1}^T \left\|\mathbf{E}_{k,t}\right\|_F^2\left\|\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2 + \frac{1}{Tp_k p_{-k}^2}\sum_{t=1}^T \left\|\mathbf{E}_{k,t}\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2 \\
&= O_P\left(w_{-k}\right) + O_P\left(p_{-k}^{-1}\right) = o_P(1).
\end{aligned}
$$

The proof now follows from Weyl's inequality. □

Lemma 6. *We assume that Assumptions 1-4 and (25)-(26) hold. Then, as* $\min\{T, p_1, ..., p_K\} \to \infty$*, it holds that*

$$
\begin{aligned}
\frac{1}{p_k}\left\|\mathcal{VI}\widetilde{\mathbf{A}}_k\right\|_F^2 &= O_P\left(m_{-k}\right) + O_P\left(\frac{1}{Tp_{-k}}\right), \\
\frac{1}{p_k}\left\|\mathcal{VII}\widetilde{\mathbf{A}}_k\right\|_F^2 &= O_P\left(m_{-k}\right) + O_P\left(\frac{1}{Tp_{-k}}\right), \\
\frac{1}{p_k}\left\|\mathcal{VIII}\widetilde{\mathbf{A}}_k\right\|_F^2 &= O_P\left(\frac{1}{Tp} + \frac{1}{p^2}\right) + O_P\left(\left(\frac{1}{Tp_k} + \frac{1}{p_k^2}\right)w_{-k}^2\right) \\
&\quad + O_P\Big(\frac{1}{p_{-k}^2} + \frac{w_{-k}}{p_{-k}} + w_{-k}^2\Big) \times \frac{1}{p_k}\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2,
\end{aligned}
$$

where recall that, by (A.2), $\widetilde{\mathbf{M}}_k = \mathcal{V} + \mathcal{VI} + \mathcal{VII} + \mathcal{VIII}$.

Proof. We begin by noting that, by Lemma 2*(ii)* and (25)-(26)

$$
\begin{aligned}
\frac{1}{p_k}\left\|\mathcal{VI}\widetilde{\mathbf{A}}_k\right\|_F^2 &= \frac{1}{p_k}\left\|\frac{1}{Tp_k p_{-k}^2}\sum_{t=1}^T \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top \mathbf{B}_k \mathbf{F}_{k,t}^\top \mathbf{A}_k^\top \widetilde{\mathbf{A}}_k\right\|_F^2 \qquad (A.3)\\
&\lesssim \frac{1}{p_k}\left\|\frac{1}{Tp_{-k}}\sum_{t=1}^T \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k \mathbf{F}_{k,t}^\top\right\|_F^2 \\
&\lesssim \frac{1}{p_k}\left\|\frac{1}{Tp_{-k}}\sum_{t=1}^T \mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\mathbf{F}_{k,t}^\top\right\|_F^2 + \frac{1}{p_k}\left\|\frac{1}{Tp_{-k}}\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\mathbf{F}_{k,t}^\top\right\|_F^2 \\
&= O_P\left(m_{-k}\right) + O_P\left(\frac{1}{Tp_{-k}}\right).
\end{aligned}
$$

By the same logic, it can be readily shown that

$$
\frac{1}{p_k}\left\|\mathcal{VII}\widetilde{\mathbf{A}}_k\right\|_F^2 = O_P\left(m_{-k} + \frac{1}{Tp_{-k}}\right).
$$

We now consider the last statement in the lemma; it holds that

$$
\begin{aligned}
\mathcal{VIII}\widetilde{\mathbf{A}}_k &= \frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^T \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{E}_{k,t}^\top\widetilde{\mathbf{A}}_k \\
&= \frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^T \mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}+\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}+\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{E}_{k,t}^\top\widetilde{\mathbf{A}}_k \\
&= \frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\widetilde{\mathbf{A}}_k \\
&\quad +\frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^T \mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\widetilde{\mathbf{A}}_k \\
&\quad +\frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^T \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{E}_{k,t}^\top\widetilde{\mathbf{A}}_k \\
&= \mathcal{VIII}_a+\mathcal{VIII}_b+\mathcal{VIII}_c.
\end{aligned}
$$

We will report our calculations for $r_k = 1$, for simplicity and without loss of generality, for $1 \leq k \leq K$. Consider $\mathcal{VIII}_a$; since $\left\|\widehat{\mathbf{H}}_{-k}\right\|_F = O_P(1)$, it holds that

$$
\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\widetilde{\mathbf{A}}_k\right\|_F^2 = O_P(1)\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\widetilde{\mathbf{A}}_k\right\|_F^2.
$$

Consider now

$$
\begin{aligned}
&\mathbb{E}\left(\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\right\|_F^2\right) \\
&= \|\mathbf{B}_k\|_F^2\,\mathbb{E}\left(\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\mathbf{A}_k^\top\mathbf{E}_{k,t}\right\|_F^2\right) \\
&= \|\mathbf{B}_k\|_F^2\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}}\mathbb{E}\left(\left\|\sum_{t=1}^T \mathbf{B}_k^\top\boldsymbol{e}_{k,t,i\cdot}^\top\boldsymbol{e}_{k,t,\cdot j}^\top\mathbf{A}_k\right\|_F^2\right) \\
&\lesssim p_{-k}^2p_k\mathbb{E}\left(\left\|\sum_{t=1}^T \mathbf{B}_k^\top\boldsymbol{e}_{k,t,i\cdot}^\top\boldsymbol{e}_{k,t,\cdot j}^\top\mathbf{A}_k\right\|_F^2\right).
\end{aligned}
\tag{A.4}
$$

Given that Assumption 2(ii) entails that $\|\mathbf{B}_k\|_F^2 = c_0p_{-k}$, it holds that

$$
\begin{aligned}
&\mathbb{E}\left(\left\|\sum_{t=1}^T \mathbf{B}_k^\top\boldsymbol{e}_{k,t,i\cdot}^\top\boldsymbol{e}_{k,t,\cdot j}^\top\mathbf{A}_k\right\|_F^2\right) \\
&\leq \mathbb{E}\left(\left\|\mathbf{B}_k^\top\sum_{t=1}^T\left(\boldsymbol{e}_{k,t,i\cdot}^\top\boldsymbol{e}_{k,t,\cdot j}^\top-\mathbb{E}\left(\boldsymbol{e}_{k,t,i\cdot}^\top\boldsymbol{e}_{k,t,\cdot j}^\top\right)\right)\mathbf{A}_k\right\|_F^2\right)+\left\|\sum_{t=1}^T\mathbb{E}\left(\mathbf{B}_k^\top\boldsymbol{e}_{k,t,i\cdot}^\top\boldsymbol{e}_{k,t,\cdot j}^\top\mathbf{A}_k\right)\right\|_F^2 \\
&\lesssim \sum_{t,s=1}^T\sum_{i_1,i_2=1}^{p_k}\sum_{j_1,j_2=1}^{p_{-k}}\left|Cov\left(e_{k,t,ij_1}e_{k,t,i_1j},e_{k,s,ij_2}e_{k,s,i_2j}\right)\right|+\left(\sum_{t=1}^T\sum_{i_1=1}^{p_k}\sum_{j_1=1}^{p_{-k}}\left|\mathbb{E}\left(e_{k,t,ij_1}e_{k,t,i_1j}\right)\right|\right)^2 \\
&= O\left(Tp+T^2\right),
\end{aligned}
\tag{A.5}
$$

for all $1 \leq i \leq p_k$ and all $1 \leq j \leq p_{-k}$, where we have used Assumptions 2(ii), 3(ii), and 3(iii). Hence, from (A.4) and (A.5) and Lemma 2*(i)*, and recalling that $\left\|\widetilde{\mathbf{H}}_k\right\|_F^2 = O_P(1)$

$$
\begin{aligned}
&\frac{1}{p_k}\|\mathcal{VIII}_a\|_F^2 \\
&= \frac{1}{p_k}\left\|\frac{1}{Tp_kp_{-k}^2}\left(\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\widetilde{\mathbf{H}}_k + \sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right)\right\|_F^2 \\
&\lesssim \frac{1}{T^2p_k^3p_{-k}^4}\left(\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\mathbf{B}_k\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\right\|_F^2 + \left(\sum_{t=1}^T\left\|\mathbf{E}_{k,t}\mathbf{B}_k\right\|_F^2\right)^2\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2\right) \\
&= O_P\left(\frac{1}{Tp} + \frac{1}{p^2}\right) + O_P\left(\frac{1}{p_kp_{-k}^2}\right)\times\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2. \qquad \text{(A.6)}
\end{aligned}
$$

We now turn to studying $\mathcal{VIII}_b$; it holds that

$$
\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\right\|_F^2 \leq \sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}}\left\|\sum_{t=1}^T e_{k,t,ij}\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\right\|_F^2\left\|\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2; \qquad \text{(A.7)}
$$

using Assumptions 2(ii), 3(ii), and 3(iii), and a similar logic as in the proof of (A.5), it follows that for all $1 \leq i \leq p_k$ and all $1 \leq j \leq p_{-k}$,

$$
\begin{aligned}
&\mathbb{E}\left(\left\|\sum_{t=1}^T e_{k,t,ij}\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\right\|_F^2\right) \\
&\lesssim \sum_{t,s=1}^T\sum_{i_1,i_2=1}^{p_k}\sum_{j_1,j_2=1}^{p_{-k}}\left|Cov\left(e_{k,t,ij}e_{k,t,i_1j_1}, e_{k,s,ij}e_{k,s,i_2j_2}\right)\right| + \left(\sum_{t=1}^T\sum_{i_1=1}^{p_k}\sum_{j_1=1}^{p_{-k}}\left|\mathbb{E}\left(e_{k,t,ij}e_{k,t,i_1j_1}\right)\right|\right)^2 \\
&= O\left(Tp + T^2\right). \qquad \text{(A.8)}
\end{aligned}
$$

Hence, combining (A.7) and (A.8), and by the sufficient condition (26) and Lemma 2*(i)*, it follows that

$$
\begin{aligned}
\frac{1}{p_k}\|\mathcal{VIII}_b\|_F^2 \leq& \frac{1}{p_k}\left\|\frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^T \mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2 \\
&+ \frac{1}{p_k}\left\|\frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^T \mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\widehat{\mathbf{H}}_{-k}^\top\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right\|_F^2 \\
\lesssim& \frac{1}{T^2p_k^3p_{-k}^4}\left\|\sum_{t=1}^T \mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\mathbf{B}_k^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\right\|_F^2 \\
&+ \frac{1}{T^2p_k^3p_{-k}^4}\left(\sum_{t=1}^T\left\|\mathbf{E}_{k,t}\right\|_F^2\right)\left\|\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2\left(\sum_{t=1}^T\left\|\mathbf{E}_{k,t}\mathbf{B}_k\right\|_F^2\right)\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2 \\
=& O_P\left(\left(\frac{1}{Tp} + \frac{1}{p^2}\right)w_{-k}\right) + O_P\left(\frac{w_{-k}}{p}\right)\times\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2, \qquad \text{(A.9)}
\end{aligned}
$$

since $\sum_{t=1}^T\left\|\mathbf{E}_{k,t}\right\|_F^2 = O_P(Tp)$ by the same arguments used in the proof of Lemma 2*(i)*. We conclude by studying

$\mathcal{VIII}_c$. By (26) and parts *(i)* and *(iii)* of Lemma 2, it follows that

$$
\begin{aligned}
&\frac{1}{p_k}\|\mathcal{VIII}_c\|_F^2\\
\leq&\frac{1}{p_k}\left\|\frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^{T}\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{E}_{k,t}^\top\left(\widetilde{\mathbf{A}}_k-\mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right\|_F^2\\
&+\frac{1}{p_k}\left\|\frac{1}{Tp_kp_{-k}^2}\sum_{t=1}^{T}\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{E}_{k,t}^\top\mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2\\
\lesssim&\frac{1}{T^2p_k^3p_{-k}^4}\sum_{t=1}^{T}\left\|\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\right\|_F^2\sum_{t=1}^{T}\left\|\mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\right\|_F^2\left\|\widetilde{\mathbf{A}}_k-\mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2\\
&+\frac{1}{T^2p_k^3p_{-k}^4}\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\mathbf{A}_k^\top\mathbf{E}_{k,t}\right\|_F^2\left\|\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2\\
\leq&\frac{1}{T^2p_k^3p_{-k}^4}\left(\sum_{t=1}^{T}\left\|\mathbf{E}_{k,t}\mathbf{B}_k\right\|_F^2+\sum_{t=1}^{T}\left\|\mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\right\|_F^2\right)\sum_{t=1}^{T}\left\|\mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\right\|_F^2\left\|\widetilde{\mathbf{A}}_k-\mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2\\
&+\frac{1}{T^2p_k^3p_{-k}^4}\left(\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\mathbf{B}_k\mathbf{A}_k^\top\mathbf{E}_{k,t}\right\|_F^2+\sum_{i=1}^{p_k}\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{e}_{k,t,i\cdot}^\top\right\|_F^2\left\|\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2\right)\left\|\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2\\
=&O_P\left(\left(\frac{1}{Tp}+\frac{1}{p^2}\right)w_{-k}\right)+O_P\left(\left(\frac{1}{p_k^2}+\frac{1}{Tp_k}\right)w_{-k}^2\right)+O_P\left(\left(\frac{1}{p_{-k}}+w_{-k}\right)w_{-k}\frac{1}{p_k}\right)\times\left\|\widetilde{\mathbf{A}}_k-\mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2,
\end{aligned}
\tag{A.10}
$$

since, by (A.5), $\left\|\sum_{t=1}^T\mathbf{E}_{k,t}\mathbf{B}_k\mathbf{A}_k^\top\mathbf{E}_{k,t}\right\|_F^2=O_P(Tp^2+T^2p)$. Hence, combining (A.6), (A.9) and (A.10), it finally follows that

$$
\begin{aligned}
\frac{1}{p_k}\left\|\mathcal{VIII}\widetilde{\mathbf{A}}_k\right\|_F^2=&O_P\left(\frac{1}{Tp}+\frac{1}{p^2}\right)+O_P\left(\left(\frac{1}{p_k^2}+\frac{1}{Tp_k}\right)w_{-k}^2\right)\\
&+O_P\left(\frac{1}{p_{-k}^2}+\frac{w_{-k}}{p_{-k}}+w_{-k}^2\right)\times\frac{1}{p_k}\left\|\widetilde{\mathbf{A}}_k-\mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2,
\end{aligned}
$$

thus completing the proof. □

Lemma 7. *We assume that Assumptions 1-4 hold and that $\widehat{\mathbf{A}}_k$ is used as projection matrix. Then, as* $\min\{T,p_1,...,p_K\}\to\infty$*, it holds that*

$$
\begin{aligned}
\left\|\frac{1}{p_k}\mathbf{A}_k^\top\left(\widetilde{\mathbf{A}}_k-\mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right\|_F=&O_P\left(\frac{1}{\sqrt{Tp}}\right)+O_P\left(\frac{1}{p}\right)+O_P\left(\sum_{j=1}^{K}\frac{1}{Tp_{-j}}\right)\\
&+O_P\left(\sum_{j=1,j\neq k}^{K}\left(\frac{1}{Tp_j\sqrt{p_{-k}}}+\frac{1}{p_j\sqrt{Tp_k}}+\frac{1}{p_kp_j^2}+\frac{1}{Tp_j^2}\right)\right).
\end{aligned}
$$

Proof. As in the above, we will assume $r_k=1$, $1\leq k\leq K$. By equation (A.2), it holds that

$$
\frac{1}{p_k}\mathbf{A}_k^\top\left(\widetilde{\mathbf{A}}_k-\mathbf{A}_k\widetilde{\mathbf{H}}_k\right)=\frac{1}{p_k}\mathbf{A}_k^\top\left(\mathcal{VI}+\mathcal{VII}+\mathcal{VIII}\right)\widetilde{\mathbf{A}}_k\widetilde{\mathbf{\Lambda}}_k^{-1}.
$$

We know from (29) that $\left\|\widetilde{\Lambda}_k^{-1}\right\|_F = O_P(1)$. It holds that

$$\begin{aligned}
& \left\|\frac{1}{p_k}\mathbf{A}_k^\top \mathcal{VI}\widetilde{\mathbf{A}}_k\right\|_F \\
= & \left\|\frac{1}{Tp^2}\mathbf{A}_k^\top \sum_{t=1}^T \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top \mathbf{B}_k\mathbf{F}_{k,t}^\top \mathbf{A}_k^\top \widetilde{\mathbf{A}}_k\right\|_F \\
\lesssim & \frac{1}{(Tp)^{1/2}}\left\|\frac{1}{T^{1/2}}\sum_{t=1}^T \left(\frac{\mathbf{A}_k}{p_k^{1/2}}\right)^\top \mathbf{E}_{k,t}\left(\frac{\widehat{\mathbf{B}}_k}{p_{-k}^{1/2}}\right)\mathbf{F}_{k,t}^\top\right\|_F = O_P\left((Tp)^{-1/2}\right),
\end{aligned}$$

by Assumption 4(i), recalling that $\|\mathbf{A}_k\|_F = c_0 p_k^{1/2}$ and $\|\mathbf{B}_k\|_F = c_1 p_{-k}^{1/2}$. Similarly

$$\begin{aligned}
& \left\|\frac{1}{p_k}\mathbf{A}_k^\top \mathcal{VII}\widetilde{\mathbf{A}}_k\right\|_F \\
= & \left\|\frac{1}{Tp^2}\mathbf{A}_k^\top \mathbf{A}_k\sum_{t=1}^T \mathbf{B}_k^\top \widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\widetilde{\mathbf{A}}_k\right\|_F \\
\lesssim & \left\|\frac{1}{Tp}\sum_{t=1}^T \widehat{\mathbf{B}}_k^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F + \left\|\frac{1}{Tp}\sum_{t=1}^T \widehat{\mathbf{B}}_k^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right\|_F,
\end{aligned}$$

and

$$\begin{aligned}
& \left\|\frac{1}{Tp}\sum_{t=1}^T \widehat{\mathbf{B}}_k^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right\|_F \\
\lesssim & \left\|\frac{1}{Tp}\sum_{t=1}^T \widehat{\mathbf{H}}_{-k}^\top \mathbf{B}_k^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\right\|_F \left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F + \left\|\frac{1}{Tp}\sum_{t=1}^T \left(\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\right\|_F \left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F \\
= & O_P\left(\sqrt{p_k\widetilde{w}_k}\right)\left(\left\|\frac{1}{Tp}\sum_{t=1}^T \mathbf{B}_k^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\right\|_F + \left\|\frac{1}{Tp}\sum_{t=1}^T \left(\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\right\|_F\right),
\end{aligned}$$

by Theorem 3. Hence, using the same logic as in the passages above

$$\left\|\frac{1}{Tp}\sum_{t=1}^T \widehat{\mathbf{B}}_k^\top \mathbf{E}_{k,t}^\top \mathbf{F}_{k,t}\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right\|_F = O_P\left(\sqrt{p_k\widetilde{w}_k}\right)O_P\left(\frac{1}{\sqrt{Tp}} + \sqrt{\frac{m_{-k}}{p_k}}\right).$$

Finally, the same logic as in the above yields

$$\begin{aligned}
& \left\|\frac{1}{p_k}\mathbf{A}_k^\top \mathcal{VIII}\widetilde{\mathbf{A}}_k\right\|_F \\
= & \left\|\frac{1}{Tp^2}\sum_{t=1}^T \mathbf{A}_k^\top \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top \mathbf{E}_{k,t}^\top \widetilde{\mathbf{A}}_k\right\|_F \\
\lesssim & \frac{1}{p}\sqrt{\frac{1}{Tp}\sum_{t=1}^T \left\|\frac{1}{Tp^2}\sum_{t=1}^T \mathbf{A}_k^\top \mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\right\|_F \times \frac{1}{Tp}\sum_{t=1}^T \left\|\frac{1}{Tp^2}\sum_{t=1}^T \widehat{\mathbf{B}}_k^\top \mathbf{E}_{k,t}^\top \widetilde{\mathbf{A}}_k\right\|_F} \\
\lesssim & \frac{1}{p}\sqrt{(1+p_{-k}w_{-k})^2(1+p_k\widetilde{w}_k)}.
\end{aligned}$$

The desired result now follows from putting all together. □

Appendix B. Simulations

Appendix B.1. Data generation

We investigate the finite sample performance of the proposed iterative projection methods. We compare the performances of initial estimator (IE), the projected estimator (PE), iterative projected mode-wise PCA estimation (IPmoPCA) by Zhang et al. [5], and the Time series Outer-Product Unfolding Procedure (TOPUP) and Time series Inner-Product Unfolding Procedure (TIPUP) with their iteration procedure (iTOPUP and iTIPUP) by Chen et al. [1] in terms of estimating the loading matrices, the common components and the number of factors. The tensor observations are generated following the order 3 tensor factor model:

$$\mathcal{X}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3 + \mathcal{E}_t.$$

We set $r_1 = r_2 = r_3 = 3$, draw the entries of $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ independently from a uniform distribution $\mathcal{U}(-1, 1)$, and let

$$\text{Vec}\,(\mathcal{F}_t) = \phi \times \text{Vec}\,(\mathcal{F}_{t-1}) + \sqrt{1-\phi^2} \times \epsilon_t, \quad \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}\,(\mathbf{0}, \mathbf{I}_{r_1 r_2 r_3})$$

$$\text{Vec}\,(\mathcal{E}_t) = \psi \times \text{Vec}\,(\mathcal{E}_{t-1}) + \sqrt{1-\psi^2} \times \text{Vec}\,(\mathcal{U}_t)$$

where $\mathcal{U}_t$ is drawn from a tensor normal distribution, i.e., $\mathcal{TN}(\mathcal{M}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3)$, which is equivalent to saying that $\text{Vec}(\mathcal{U}_t) \sim \mathcal{N}(\text{Vec}(\mathcal{M}), \mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1)$. In our study, we set $\mathcal{M} = 0$, $\mathbf{\Sigma}_k$ to be the matrix with 1 on the diagonal, and $1/p_k$ on the off-diagonal, for $k = 1, 2, 3$. The parameters ϕ and ψ control for temporal correlations of $\mathcal{F}_t$ and $\mathcal{E}_t$. By setting ϕ and ψ unequal to zero, the common factors have cross-correlations, and the idiosyncratic components have both cross-correlations and weak autocorrelations.

In Appendix B.2 and Appendix B.3, it is assumed that factor numbers are known, and the performance of estimating the number of factors is investigated in Appendix B.4. All the following simulation results are based on 1000 replications.

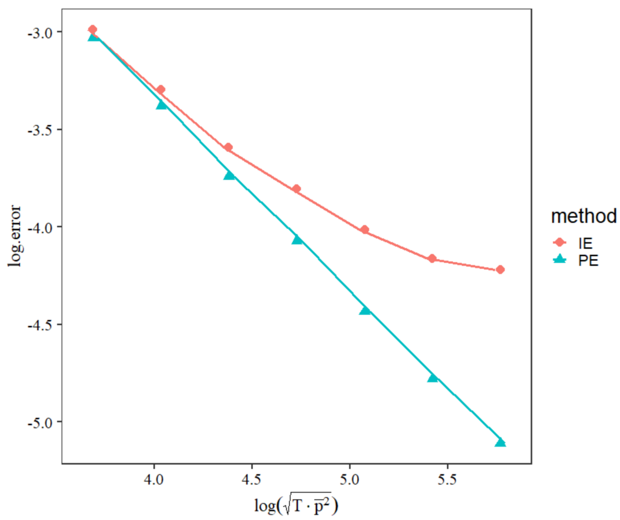

Fig. B.1: Mean log error for estimating the loading matrices over 1000 replications for $p_1 = 40$, $p_2 = p_3 = \bar{p} = 10$, $T \in (16, 32, 64, 128, 256, 512, 1024)$.

Appendix B.2. Verifying the convergence rates for loading spaces

We compare the performance of PE, IE, IPmoPCA, TOPUP, TIPUP, iTOPUP, and iTIPUP (with $h_0 = 1$ in their implementation) in estimating loading spaces. We consider the following three settings:

Setting A: $p_1 = p_2 = p_3 = 10$, $\phi = 0.1$, $\psi = 0.1$, $T \in (20, 50, 100, 200)$

Table B.1: Averaged estimation errors and standard errors of loading spaces for Settings A, B and D over 1000 replications. "PE": projection estimation method."IE": initial estimation method. "TOPUP": Time series Outer-Product Unfolding Procedure with $h_0 = 1$. "TIPUP": Time series Inner-Product Unfolding Procedure with $h_0 = 1$. "iTOPUP": Iterative Time series Outer-Product Unfolding Procedure with $h_0 = 1$. "iTIPUP": Iterative Time series Inner-Product Unfolding Procedure with $h_0 = 1$. "IPmoPCA": iterative projected mode-wise PCA estimation.

Evaluation	p_1	p_2	p_3	T	PE	IE	IPmoPCA	TOPUP	TIPUP	iTOPUP	iTIPUP
$\mathcal{D}(\widehat{\mathbf{A}}_1, \mathbf{A}_1)$	10	10	10	20	0.0444	0.1970	0.0476	0.2065	0.4711	0.0677	0.3564
				50	0.0322	0.1783	0.0286	0.1851	0.4191	0.0544	0.3039
				100	0.0248	0.1728	0.0219	0.1776	0.3426	0.0502	0.2357
				200	0.0202	0.1713	0.0175	0.1732	0.2695	0.0477	0.1519
	100	10	10	20	0.0424	0.0410	0.0401	0.0486	0.3541	0.0584	0.3495
				50	0.0266	0.0259	0.0252	0.0348	0.2963	0.0476	0.2941
				100	0.0189	0.0186	0.0179	0.0291	0.2328	0.0435	0.2350
				200	0.0133	0.0133	0.0126	0.0257	0.1526	0.0405	0.1548
	20	20	20	20	0.0203	0.0512	0.0199	0.0536	0.3391	0.0294	0.2551
				50	0.0127	0.0456	0.0125	0.0478	0.2656	0.0237	0.1927
				100	0.0091	0.0450	0.0090	0.0467	0.1817	0.0218	0.1363
				200	0.0064	0.0445	0.0063	0.0452	0.0978	0.0204	0.0758
$\mathcal{D}(\widehat{\mathbf{A}}_2, \mathbf{A}_2)$	10	10	10	20	0.0474	0.1873	0.0430	0.1943	0.4634	0.0647	0.3456
				50	0.0327	0.1843	0.0290	0.1899	0.4198	0.0554	0.3066
				100	0.0250	0.1736	0.0221	0.1781	0.3388	0.0505	0.2344
				200	0.0205	0.1773	0.0177	0.1782	0.2744	0.0483	0.1530
	100	10	10	20	0.0129	0.1778	0.0125	0.1821	0.4170	0.0192	0.2039
				50	0.0082	0.1746	0.0079	0.1763	0.3518	0.0159	0.1510
				100	0.0059	0.1728	0.0057	0.1732	0.2797	0.0148	0.0958
				200	0.0042	0.1642	0.0040	0.1633	0.2162	0.0134	0.0492
	20	20	20	20	0.0203	0.0514	0.0199	0.0542	0.3388	0.0293	0.2551
				50	0.0127	0.0461	0.0125	0.0481	0.2752	0.0237	0.2007
				100	0.0091	0.0450	0.0089	0.0466	0.1812	0.0220	0.1375
				200	0.0064	0.0430	0.0063	0.0439	0.0999	0.0205	0.0783
$\mathcal{D}(\widehat{\mathbf{A}}_3, \mathbf{A}_3)$	10	10	10	20	0.0482	0.1927	0.0440	0.1992	0.4642	0.0654	0.3539
				50	0.0325	0.1821	0.0289	0.1880	0.4208	0.0550	0.3049
				100	0.0252	0.1716	0.0222	0.1757	0.3443	0.0507	0.2350
				200	0.0204	0.1691	0.0175	0.1712	0.2587	0.0478	0.1515
	100	10	10	20	0.0128	0.1748	0.0125	0.1797	0.4146	0.0189	0.2050
				50	0.0081	0.1685	0.0078	0.1707	0.3436	0.0158	0.1500
				100	0.0059	0.1719	0.0057	0.1732	0.2770	0.0147	0.0944
				200	0.0042	0.1681	0.0041	0.1676	0.2169	0.0138	0.0492
	20	20	20	20	0.0203	0.0506	0.0199	0.0534	0.3404	0.0295	0.2617
				50	0.0128	0.0462	0.0126	0.0484	0.2618	0.0238	0.1945
				100	0.0090	0.0435	0.0089	0.0450	0.1750	0.0219	0.1328
				200	0.0064	0.0425	0.0063	0.0433	0.0977	0.0205	0.0772

Setting B: $p_1 = 100,\ p_2 = p_3 = 10,\ \phi = 0.1,\ \psi = 0.1,\ T \in (20, 50, 100, 200)$

Setting C: $p_1 = p_2 = p_3 = 15,\ \phi = 0.1,\ \psi = 0.1,\ T \in (20, 50, 100, 200)$

Setting D: $p_1 = p_2 = p_3 = 20,\ \phi = 0.1,\ \psi = 0.1,\ T \in (20, 50, 100, 200)$

Setting E: $p_1 = p_2 = p_3 = 30,\ \phi = 0.1,\ \psi = 0.1,\ T \in (20, 50, 100, 200)$

Setting F: $p_1 = 40,\ p_2 = p_3 = 10,\ \phi = 0.1,\ \psi = 0.1,\ T \in (16, 32, 64, 128, 256, 512, 1024)$

Due to the identifiability issue of factor model, the performance of the candidates methods is evaluated by comparing the distance between the estimated loading space and the true loading space, which is

$$\mathcal{D}(\widehat{\mathbf{A}}_k, \mathbf{A}_k) = \left(1 - \frac{1}{r_k}\operatorname{Tr}\left(\widehat{\mathbf{Q}}_k\widehat{\mathbf{Q}}_k^\top \mathbf{Q}_k\mathbf{Q}_k^\top\right)\right)^{1/2},\ k = 1, 2, 3,$$

where $\mathbf{Q}_k$ and $\widehat{\mathbf{Q}}_k$ are the left singular-vector matrices of the true loading matrix $\mathbf{A}_k$ and its estimator $\widehat{\mathbf{A}}_k$. The distance is always in the interval $[0, 1]$. When $\mathbf{A}_k$ and $\widehat{\mathbf{A}}_k$ span the same space, the distance $\mathcal{D}(\widehat{\mathbf{A}}_k, \mathbf{A}_k)$ is equal to 0, while is equal to 1 when the two spaces are orthogonal.

Table B.1 shows the averaged estimation errors under Settings A, B and D for tensor normal distribution. Table B.1 shows that all these methods benefit from the increase in dimensions and sample size T. For fixed sample size T, when p_ks are of the same order, PE performs better than all other procedures in estimating loading except IPmoPCA (IPmoPCA adopts a multi-step projection technique of PE). When the tensor dimensions are not balanced, that is, one mode dimension is much higher than the other mode dimensions and sample size T, the estimation effect of PE for this dimension is comparable to that of IE, which is consistent with the conclusion of Theorem 1 and Corollary 1 that when p_k is in high order and $p_j \asymp \bar{p}$ for all $j \neq k$, the convergence rates of IE and PE are both $1/(T\bar{p}^2)$. Figure B.1

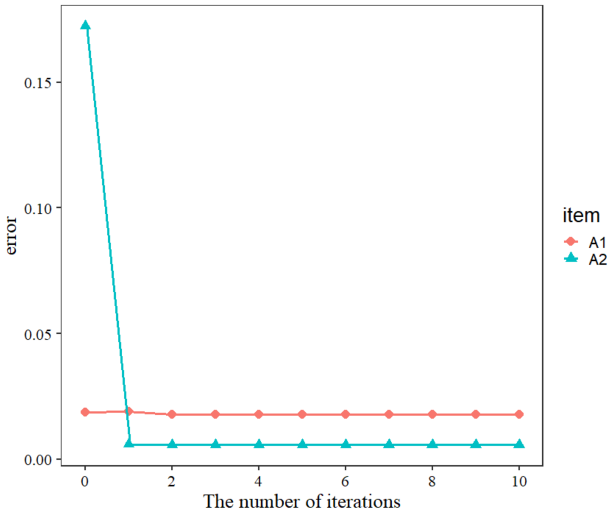

Fig. B.2: Mean estimation error at each step of the recursive procedure over 1000 replications for $T = p_1 = 100,\ p_2 = p_3 = 10$.

plots mean log error of PE and IE under setting F, which clearly shows that PE and IE have different convergence rates. Figure B.1 shows that the log error of the PE method for estimating $\mathbf{A}_1$ is almost linear to $\ln(\sqrt{T\bar{p}^2})$ with slope -1 while $p_2 = p_3 = \bar{p}$, which also matches the rate in Corollary 1. For the IE method, the log error initially decreases as $\ln(\sqrt{Tp_2p_3})$ increases and then tends to be constant. According to Theorem 1, this is because when $\ln(\sqrt{T\bar{p}^2})$ is large enough, the error of IE only depends on p_1. This shows that PE is a more accurate estimation method compared IE.

It can be seen from Table B.1 that, as one-step iteration and multi-step iteration versions, PE and IPmoPCA have similar performance under any setting. And when the dimensions p_ks are large enough, they behave almost exactly the same. We consider whether the multi-step iteration will further reduce the estimation error and propose the following iterative estimation Algorithm 4. Figure B.2 plots the estimation error of each step of the 10-step iterative estimators of $\mathbf{A}_1$ and $\mathbf{A}_2$ under Setting B. Since $\mathbf{A}_2$ and $\mathbf{A}_3$ are symmetric, the estimation result of $\mathbf{A}_3$ is omitted here. The estimation error of $\mathbf{A}_2$ decreases significantly in the first step, but does not decrease further in the subsequent multi-step iterations. The error in the estimation of $\mathbf{A}_1$ has been almost constant, which is due to the fact that the size of estimation error of $\mathbf{A}_k$ are all $O_p(1/\sqrt{T\bar{p}^2})$ for any step while $p_j = \bar{p}$ for all $j \neq k$ and $p^2 = O(T\bar{p}^2)$. This shows that our one-step iterative algorithm 1 can get comparable results with the IPmoPCA by Zhang et al. [5] while has much less computational burden.

Algorithm 4 Iterative estimation for loading spaces

Input: tensor datas $\{\mathcal{X}_t, 1 \le t \le T\}$, factor numbers $r_1, \cdots, r_K$

Output: factor loading matrices $\{\widehat{\mathbf{A}}_k^{(s)}, 1 \le k \le K, 0 \le s \le S\}$

1: obtain the initial estimators $\{\widehat{\mathbf{A}}_k, 1 \le k \le K\}$, let $\widehat{\mathbf{A}}_k^{(0)} = \widehat{\mathbf{A}}_k$;

2: project to reduce dimensions by defining $\widehat{\mathbf{Y}}_{k,t}^{(s)} := \frac{1}{p_{-k}}\mathbf{X}_{k,t}\widehat{\mathbf{B}}_k^{(s)}$, where $\widehat{\mathbf{B}}_k^{(s)} = \otimes_{j\in[K]/\{k\}}\widehat{\mathbf{A}}_j^{(s-1)}, 1 \le k \le K$;

3: given $\{\widehat{\mathbf{Y}}_{k,t}^{(s)}, 1 \le k \le K\}$, define $\widetilde{\mathbf{M}}_k^{(s)} := (Tp_k)^{-1}\sum_{t=1}^T \widehat{\mathbf{Y}}_{k,t}^{(s)}(\widehat{\mathbf{Y}}_{k,t}^{(s)})^\top$, set $\widehat{\mathbf{A}}_k^{(s)}$ as $\sqrt{p_k}$ times the matrix with columns being the first r_k eigenvectors of $\widehat{\mathbf{M}}_k^{(s)}$;

4: Repeat steps 2 to 3 until up to the maximum number of iterations $s = S$. Output the projection estimators as $\{\widehat{\mathbf{A}}_k^{(s)}, 1 \le k \le K, 0 \le s \le S\}$.

Appendix B.3. Verifying the convergence rates for common components

We compare the performance of PE, IE, TOPUP, TIPUP, iTOPUP and iTIPUP (with $h_0 = 1$ in their implementation) in common components under Setting A, B, D. We use the mean squared error to evaluate the performance of

Table B.2: Averaged estimation errors and standard errors of common components for Settings A, B and D over 1000 replications. “PE”“IE”“TOPUP”“TIPUP”“iTOPUP”“iTIPUP” are the same as in Table B.2

p_1	p_2	p_3	T	PE	IE	IPmoPCA	TOPUP	TIPUP	iTOPUP	iTIPUP
10	10	10	20	0.032144	0.082885	0.031372	0.088068	0.463034	0.036216	0.287679
			50	0.029511	0.078503	0.029042	0.082402	0.402929	0.033516	0.239876
			100	0.028914	0.075327	0.028576	0.077963	0.282279	0.032770	0.161073
			200	0.028364	0.075043	0.028087	0.076372	0.170435	0.032172	0.083502
100	10	10	20	0.004583	0.032746	0.004404	0.035071	0.377769	0.006369	0.190216
			50	0.003459	0.030847	0.003388	0.032154	0.280331	0.005202	0.133925
			100	0.003069	0.030864	0.003033	0.031784	0.171590	0.004768	0.079095
			200	0.002887	0.029564	0.002869	0.029921	0.083551	0.004511	0.031586
20	20	20	20	0.004394	0.009253	0.004361	0.009900	0.308279	0.005508	0.182627
			50	0.003792	0.008399	0.003779	0.008847	0.229848	0.004817	0.119517
			100	0.003592	0.008126	0.003585	0.008432	0.119257	0.004587	0.062612
			200	0.003486	0.007941	0.003483	0.008077	0.033438	0.004446	0.020716

different estimated procedure, i.e.,

$$\text{MSE} = \frac{1}{Tp}\sum_{t=1}^{T}\left\|\widehat{\mathcal{S}}_t - \mathcal{S}_t\right\|_F^2 .$$

Table B.2 shows the averaged estimation errors under Settings A, B and D. Table B.2 shows that all these methods benefit from the increase in dimensions and sample size T, and PE performs to IPmoPCA and better than all other producers in estimating common components.

Table B.3: The frequencies of exact estimation of the numbers of factors under Settings A, C, D and E over 1000 replications. “\” means the data is too large to calculate.

$p_1 = p_2 = p_3$	T	PE-ER	IE-ER	TCorTh	TOP-ER	TIP-ER	iTOP-ER	iTIP-ER
10	20	0.395	0.049	0.000	0.048	0.002	0.776	0.046
	50	0.424	0.071	0.034	0.065	0.003	0.789	0.097
	100	0.450	0.069	0.139	0.058	0.013	0.811	0.302
	200	0.456	0.077	0.318	0.076	0.029	0.796	0.577
15	20	0.932	0.309	0.125	0.288	0.007	0.992	0.169
	50	0.947	0.343	0.596	0.329	0.023	0.993	0.321
	100	0.944	0.346	0.822	0.340	0.061	0.991	0.644
	200	0.948	0.362	0.920	0.347	0.134	0.999	0.940
20	20	0.997	0.675	0.628	0.662	0.022	1.000	0.215
	50	0.999	0.731	0.941	0.700	0.067	1.000	0.434
	100	1.000	0.728	0.990	0.705	0.153	1.000	0.766
	200	1.000	0.724	0.994	0.691	0.325	1.000	0.975
30	20	1.000	0.970	0.988	\	0.052	\	0.321
	50	1.000	0.982	1.000	\	0.138	\	0.573
	100	1.000	0.981	1.000	\	0.346	\	0.839
	200	1.000	0.981	1.000	\	0.650	\	0.991

Appendix B.4. Estimating the numbers of factors

We investigate the empirical performances of the proposed projected iteration procedure based on eigenvalue-ratio (PE-ER), along with the initial version (IE-ER), total mode-k correlation thresholding (TCorTh) in Lam [3], information-criterion/eigenvalue-ratio method that used TOPUP/ TIPUP/ iTOPUP/ iTIPUP in Han et al. [2], abbreviated as TOP-IC/ TIP-IC/ iTOP-IC/ iTIP-IC/ TOP-ER/ TIP-ER/ iTOP-ER/ iTIP-ER respectively, while we fixed $h_0 = 1$ in their implementation.

First of all, we note that the accuracy of IC method is zero under any setting, so IC method is not an appropriate method to estimate the number of factors. We will ignore the IC method in the following discussion. Table B.3 presents the frequencies of exact estimation over 1000 replications under Settings A, C, D and E by different methods. We set $r_{\max} = 8$ for all estimation procedures. The performance of all three methods improve as the dimension and sample size increase. Among them, PE-ER, IE-ER, TOP-ER and ITOP-ER are more sensitive to the increase of dimension, while TCorTH, TIP-ER and iTIP-ER are more sensitive to the increase of sample size.

Table B.3 also shows that the iterative algorithm generally outperforms the non-iterative algorithm, and TOP-ER and iTOP-ER outperform TIP-ER and iTIP-ER, respectively, under this setting. The accuracy of PE-ER is second only to iTOP-ER. Note that iTOP-ER require much longer run time and large storage space, since the TOPUP procedure needs to deal with a large high-order tensor, such as $\mathbb{R}^{p_1\times p_2\times p_3\times p_1\times p_2\times p_3\times 1}$ while $h_0 = 1$. Therefore, iTOP-ER may have some difficulties in practical calculation.

Appendix C. Real data analysis

We analyzed multi-category import-export network data analyzed in [1]. The data contains total monthly imports and exports for 15 commodity categories between 22 European and American countries from January 2014 to December 2019, with missing values for the export from any country to itself. More detailed information on data, countries and commodity categories can be found in Chen et al. [1]. Similar to [1], we set the volume of each country's exports to zero and take a three-month rolling average of the data to eliminate the impact of unusual large transactions or shipping delays. At this point, this data can be viewed as a $22 \times 22 \times 15 \times 70$ fourth-order tensor. Each element $x_{i,j,k,t}$ is the three-month average total export volume of country i to country j at time t for category k goods. We will model the multi-category import-export network data according to the following model. Let $\mathcal{X}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3 + \mathcal{E}_t$, where $\mathcal{X}_t \in \mathbb{R}^{p_1\times p_2\times p_3}$ is the observed tensor at time t, $\mathcal{F}_t \in \mathbb{R}^{r_1\times r_2\times r_3}$ is the factor tensor, $\{\mathbf{A}_k \in \mathbb{R}^{p_k\times r_k},\ k = 1, 2, 3\}$ are the loading matrices, for $t = 1, 2, \cdots, 70$, $p_1 = p_2 = 22,\ p_3 = 15$, $\{r_k, k = 1, 2, 3\}$ to be determined.

We first use a rolling-validation procedure as in Wang et al. [4] to compare the methods mentioned in Appendix B.2. To implement the rolling-validation procedure, we add additional November and December 2013 data before making the three-month moving average and normalize the final data. For each year t from 2017 to 2019, we repeatedly use the n (bandwidth) year observations prior to t to estimate the loading matrices and then used loadings to estimate the observations and the corresponding residuals for the 12 months of the year. Specifically, let $\mathcal{X}_t^i$ and $\widehat{\mathcal{X}}_t^i$ be the observed and estimated import-export tensor of month i in year t, and further define the mean squared error as:

$$\mathrm{MSE}_t = \frac{1}{12\times 22\times 22\times 15}\sum_{i=1}^{12}\left\|\widehat{\mathcal{X}}_t^i - \mathcal{X}_t^i\right\|_F^2 .$$

Table C.4: The averaged MSE of rolling validation for the three-month moving average data. $12n$ is the sample size of the training set. $r_1 = r_2 = r_3 = r$ is the number of factors. "PE": projection estimation method."IE": initial estimation method. "TOPUP": Time series Outer-Product Unfolding Procedure with $h_0 = 1$. "TIPUP": Time series Inner-Product Unfolding Procedure with $h_0 = 1$. "iTOPUP": Iterative Time series Outer-Product Unfolding Procedure with $h_0 = 1$. "iTIPUP": Iterative Time series Inner-Product Unfolding Procedure with $h_0 = 1$. "IPmoPCA": iterative projected mode-wise PCA estimation.

n	r	PE	IE	IPmoPCA	TIPUP	iTIPUP	TOPUP	iTOPUP
1	3	0.717165	0.719191	0.717929	0.719222	0.717905	0.719204	0.717891
2	3	0.358891	0.359821	0.359287	0.359839	0.359309	0.359794	0.359284
3	3	0.239331	0.239936	0.239582	0.239945	0.239595	0.239909	0.239571
1	4	0.711102	0.712812	0.711038	0.712861	0.711067	0.712864	0.711088
2	4	0.355766	0.356680	0.355675	0.356697	0.355685	0.356598	0.355629
3	4	0.237183	0.237880	0.237054	0.237888	0.237066	0.237798	0.237037
1	5	0.697179	0.699275	0.693501	0.698985	0.693601	0.698836	0.693544
2	5	0.348440	0.349485	0.346284	0.349321	0.346368	0.349145	0.346376
3	5	0.232304	0.233041	0.230282	0.232942	0.230299	0.232837	0.230415
1	6	0.672986	0.675783	0.672979	0.675047	0.672926	0.673810	0.672323
2	6	0.336668	0.338309	0.336630	0.338103	0.336587	0.337334	0.335946
3	6	0.224361	0.225838	0.224242	0.225780	0.224165	0.225522	0.223770

Table C.4 compares the mean values of MSE of various estimation methods for different combinations of bandwidth n and factor number $r_1 = r_2 = r_3 = r$. The estimation errors of PE, iTIOPUP and iTIPUP methods are very close, and the estimation performance of projected methods is better than that of non-prejected methods.

For processed $22 \times 22 \times 15 \times 70$ fourth-order tensor data, our PE-ER method suggests $r_1 = 1,\ r_2 = 3,\ r_3 = 1$, while TCorTh suggests $r_1 = 3,\ r_2 = 1,\ r_3 = 3$, iTOP-ER suggests $r_1 = 3,\ r_2 = 4,\ r_3 = 1$, iTIP-ER suggests $r_1 = 2,\ r_2 = 2,\ r_3 = 1$ and other methods all suggest $r_1 = r_2 = r_3 = 1$. For better illustration, we take $r_1 = r_2 = 4,\ r_3 = 6$ as same as in [1].

Table C.5: Estimated loading matrix $\mathbf{A}_3$ for category fiber.

	Animal	Vegetable	Food	Mineral	Chemicals	Plastics	Leather	Wood	Textiles	Footwear	Stone	Metals	Machinery	Transport	Misc
1	0	-3	1	0	1	-2	0	2	-1	0	-1	0	**-29**	0	-6
2	0	1	0	**30**	0	0	0	3	-1	0	2	1	0	0	1
3	1	3	0	0	**-29**	-1	0	0	1	0	2	-2	0	0	-8
4	1	0	4	0	-1	-3	0	2	0	0	-1	2	0	**29**	2
5	6	5	5	0	1	19	1	5	5	1	-1	17	1	0	-9
6	2	2	4	-2	2	-2	1	4	2	1	**29**	0	-1	0	0

Table C.6: Estimated loading matrix $\mathbf{A}_1$ for export fiber.

	BE	BU	CA	DK	FI	FR	DE	GR	HU	IS	IR	IT	MX	NO	PO	PT	ES	SE	CH	ER	US	UK
1	1	0	-1	0	0	-1	-4	0	0	0	2	-1	**-30**	0	0	0	1	0	1	0	1	-1
2	-1	0	2	0	0	-3	1	0	-2	0	2	-3	-1	-1	-3	0	-1	-1	-2	-1	**-29**	-1
3	-1	0	**-30**	0	0	0	2	0	0	0	-1	0	0	-1	0	0	0	0	-1	0	-2	-1
4	6	0	1	2	1	8	24	0	1	0	8	6	-3	1	3	1	5	2	5	2	-1	5

Table C.7: Estimated loading matrix $\mathbf{A}_2$ for import fiber.

	BE	BU	CA	DK	FI	FR	DE	GR	HU	IS	IR	IT	MX	NO	PO	PT	ES	SE	CH	ER	US	UK
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**-30**	0
2	1	0	0	1	0	-1	-8	0	1	0	-1	2	**-28**	0	2	0	1	1	4	0	0	-5
3	-3	-1	-1	-3	-2	-16	2	0	-5	0	1	-8	2	-2	-5	-2	-8	-5	1	-4	0	-19
4	9	0	0	0	0	1	20	1	-1	0	5	7	-2	0	3	0	2	-1	17	1	0	-4

Table C.5 shows an estimator of $\mathbf{A}_3$ under tensor factor model using PE procedure. For better interpretation, the loading matrix is rotated using the varimax procedure and all numbers are multiplied by 30 and then truncated to integers for clearer viewing. Table C.5 shows that there is a group structure of these six factors, that each factor is called a "condensed product group" in Chen et al. [1]. Factor 1, 2, 3, 4 and 6 can be interpreted as Machinery & Electrical factor, Mineral Products factor, Chemicals & Allied Industries factor, Transportation factor and Stone & Glass factor, because these factors are mainly loaded on these commodity categories. Factor 5 can be viewed as a mixing factor, with Plastics & Rubbers and Metals as main load.

Table C.6 and C.7 show estimators of $\mathbf{A}_1$ and $\mathbf{A}_2$. The loading matrices are also rotated via varimax procedure. [1] proposed that there are some virtual import hubs and export hubs in the import-export data factor model. When exporting a commodity, the exporting country first puts the commodity into the virtual export hub, then the commodities are exported from the export hub to the virtual import hub, and finally taken out of the import by the importing countries. Table C.6 shows that Mexico, the United States of America and Canada heavily load on virtual export hubs E1, E2 and E3, respectively. European countries mainly load on export hub E4, and Germany occupies an important position. This can be seen in Table C.7 that the United States of America and Mexico heavily load on virtual import hubs I1 and I2. European countries mainly load on import hub I3 and I4, while I3 is mainly loaded by Western European countries France and Britain and I4 is mainly loaded by Central European countries Germany and Switzerland.

References

[1] R. Chen, D. Yang, C.-H. Zhang, Factor models for high-dimensional tensor time series, Journal of the American Statistical Association 117 (2022) 94–116.

[2] Y. Han, C. H. Zhang, R. Chen, Rank determination in tensor factor model, Electronic Journal of Statistics 16 (2022) 1726–1803.

[3] C. Lam, Rank determination for time series tensor factor model using correlation thresholding. LSE, London, Technical Report, UK, Working Paper, 2021.

[4] D. Wang, X. Liu, R. Chen, Factor models for matrix-valued high-dimensional time series, Journal of Econometrics 208 (2019) 231–248.

[5] X. Zhang, G. Li, C. C. Liu, J. Guo, Tucker tensor factor models: matricization and mode-wise pca estimation, Science China Mathematics (2025) 1–52.