Journal of Multivariate Analysis 214 (2026) 105616

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

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

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

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

ARTICLE INFO

AMS 2020 subject classifications:
primary 62H25
secondary 62F12

Keywords:
Least squares
Principal component analysis
Projection estimation
Tensor factor model

ABSTRACT

Tensor Factor Models (TFM) are appealing dimension reduction tools for high-order large-dimensional tensor time series, and have wide applications in economics, finance and medical imaging. In this paper, we propose a projection estimator for the Tucker-decomposition based TFM, and provide its least-square interpretation which parallels to the least-square interpretation of the Principal Component Analysis (PCA) for the vector factor model. The projection technique simultaneously reduces the dimensionality of the signal component and the magnitudes of the idiosyncratic component tensor, thus leading to an increase of the signal-to-noise ratio. We derive a convergence rate of the projection estimator of the loadings and the common factor tensor which are faster than that of the naive PCA-based estimator. Our results are obtained under mild conditions which allow the idiosyncratic components to be weakly cross- and auto- correlated. We also provide a novel iterative procedure based on the eigenvalue-ratio principle to determine the factor numbers. Extensive numerical studies are conducted to investigate the empirical performance of the proposed projection estimators relative to the state-of-the-art ones.

1. Introduction

Tensors, or high-order arrays, emerge ubiquitously in all applied sciences, such as macroeconomic data (Chen et al., 2022b; Barigozzi et al., 2025) and financial data (Han et al., 2022; He et al., 2024a) where typical applications consider higher-order realized moments (Neuberger, 2012; Buckle et al., 2016; Bae and Lee, 2021; Jondeau et al., 2018). They appear also in computer vision data (Panagakis et al., 2021); neuroimaging and functional MRI data (see e.g. Zhou et al., 2013; Ji et al., 2021; Chen et al., 2022a), typically consisting of hundreds of thousands of voxels observed over time; recommender systems (see e.g. Entezari et al., 2021, and the various references therein); data arising in psychometrics (Carroll and Chang, 1970; Douglas Carroll et al., 1980), and chemometrics (see Tomasi and Bro, 2005, and also the references in Acar et al., 2011). For a review of further applications we also refer to Kolda and Bader, 2009 and Bi et al., 2021.

Hence, inference in the context of tensor-valued time series has now become one of the most active research areas in statistics and machine learning. Given that the analysis of tensor-valued data poses significant computational challenges due to high-dimensionality, a natural approach is to model tensor-valued time series through a low-dimensional projection on a space of common factors. Indeed, factor models are one of the most powerful dimension reduction tools in time series for extracting comovements

* Corresponding author.
E-mail address: lorenzo.trapani@unipv.it (L. Trapani).

https://doi.org/10.1016/j.jmva.2026.105616
Received 1 March 2025; Received in revised form 24 January 2026; Accepted 26 January 2026

Available online 5 February 2026

among multi-dimensional features, and, in the past decades, factor models applied to vector-valued time series have been extensively studied - we refer, *inter alia*, to the contributions by Forni et al. (2000), Stock and Watson (2002), Bai (2003), Doz et al. (2012), Fan et al. (2013) and Aït-Sahalia and Xiu (2017). In recent years, the literature has extended the tools developed for the analysis of vector-valued data to the context of matrix-valued data. In particular, we refer to the seminal contribution by Wang et al. (2019), who proposed Matrix Factor Model (MFM) exploiting the double low-rank structure of matrix-valued observations, and to the subsequent paper by Chen and Fan (2023), who propose the so-called α-PCA estimation method. Furthermore, Yu et al. (2022) study a projected version of PCA; Gao et al. (2021) study the Maximum Likelihood approach; and Zhang et al. (2025b) study a hierarchical CP product MFM. In related contributions, He et al. (2023) propose a strong rule to determine whether there is a factor structure of matrix time series; Chen et al. (2020) study the MFM under linear constraints; and Liu and Chen (2019) and He et al. (2024b) study non-linearities in the form of a threshold regression, or a changepoint model, respectively.

In contrast with this plethora of contributions, the statistical analysis of tensor factor models (TFM) is still in its infancy, and most of the recent contributions focus on *i.i.d.* data, or on models with *i.i.d.* noise, thus limiting the applicability to tensor-valued time series. In such a context, a typical approach would be to decompose an observed K -fold tensor time series $\{\mathcal{X}_t \in \mathbb{R}^{p_1\times p_2\times\dots\times p_K},\ t\in\mathbb{Z}\}$ into a sum of a signal, or common, plus a noise, or idiosyncratic, component, denoted as $\{\mathcal{S}_t,\ t\in\mathbb{Z}\}$ and $\{\mathcal{E}_t,\ t\in\mathbb{Z}\}$, respectively:

$$\mathcal{X}_t = \mathcal{S}_t + \mathcal{E}_t. \tag{1}$$

However, whereas in the vector case the possible structure of the signal component is uniquely defined as a product of a loadings matrix times a vector of factors, in the context of a tensor factor model like (1) two possible decompositions exist: the Tucker decomposition, and the CP decomposition (also known as PARAFAC or CANDECOMP decomposition), and we refer to e.g. Lettau (2022) for a recent discussion and a comparison. The choice between the two relies mainly on the kind of data one is dealing with and the problem one is interested in studying. The Tucker decomposition is easier to estimate, less restrictive and more flexible than the CP decomposition. This flexibility comes at a cost, however, since Tucker tensor factor models, as the vector factors models, are not uniquely identified, while CP factor models, if they exist, are uniquely identified. An advantage of Tucker tensor factor models is that they allow for different numbers of factors for each mode, while the CP tensor factor model restricts the number of factors to be the same across modes.

1.1. Contributions of the paper

In this paper we focus on estimation of the tensor factor model based on the Tucker-decomposition, which is defined as:

$$\mathcal{S}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \cdots \times_K \mathbf{A}_K = \sum_{j_1=1}^{r_1}\cdots\sum_{j_K=1}^{r_K} f_{j_1,\dots,j_K,t}\left(\mathbf{a}_{1j_1}\circ\cdots\circ\mathbf{a}_{Kj_K}\right), \tag{2}$$

where $\mathbf{A}_k = (\mathbf{a}_{k1}\dots\mathbf{a}_{kr_k}) \in \mathbb{R}^{p_k\times r_k}$ is the kth mode-wise loading matrix with rows $\mathbf{a}_{kj}^\top$, $\mathcal{F}_t \in \mathbb{R}^{r_1\times\dots\times r_K}$ is the latent common factor tensor with entries $f_{j_1,\dots,j_K,t}$ and such that $r_k \ll p_k$, $(\times_k)$ denotes the mode-k product, and $(\circ)$ denotes the tensor or outer product.

In particular, we consider the Tucker decomposition in (2), adopting the definition of common factors which is typical of econometrics - i.e., we assume that a common factor has a non-negligible impact on almost all series, while the idiosyncratic components can be both weakly cross- and auto-correlated.

We first provide the least squares interpretation of the iterative projection for tensor factor model, which generalizes the least squares interpretation of PCA for vector factor models (Bai, 2003) and of projection estimation for MFM (Yu et al., 2022; He et al., 2024a). Second, we propose a naive estimator of the factor loadings matrices $\{\mathbf{A}_k, 1\le k\le K\}$ based on the eigen-analysis of the mode-wise sample covariance matrix, referred to as initial estimator in the following. Third, we provide a one-step projection method for estimating the factor loadings matrices. Then, we provide the theoretical convergence rates and asymptotic distributions for the initial and the projection estimators of the loadings, and for the tensor of common factors. These indicate that the projection technique improves the estimation accuracy attributed to the increased signal-to-noise ratio due to the use of information about the factors contained in all modes. Last, an iterative procedure is also provided to estimate the factor numbers consistently. In an extensive numerical study our approaches are compared with the existing ones.

1.2. Related literature

Zhang et al. (2025a) propose an iterative projected mode-wise PCA (IPmoPCA) estimation approach. Although, potentially, our projected estimator of the loadings can also be cast into an iterative procedure (see our Algorithm 1 in Section 2.1), we show, however, that our one-step estimator already attains the “optimal” rate of convergence, for the estimated $\mathbf{A}_k$, which would be obtained if $\{\mathbf{A}_j, 1\le j\le K, j\ne k\}$ were known in advance. This conclusion is also reinforced by our Monte Carlo experiments, where we show that our one-step iterative algorithm 1 can get comparable results with the IPmoPCA approach, while being computationally more efficient. Furthermore, differently from Zhang et al. (2025a), we require weaker moment assumptions on the idiosyncratic components, we also thoroughly analyze the asymptotic distribution of the estimated tensor factor, and we provide a theoretical foundation of our estimator based on the least squares approach.

Estimation of (2) has also been studied by Chen et al. (2022b) and Han et al. (2024a, 2022) under the assumption that the idiosyncratic components are weakly cross-correlated but “white”, i.e., with no auto-correlation, and by Lam (2021) and Chen and Lam (2024), who introduce a pre-averaging method before projection and work under the assumption that the idiosyncratic components are both weakly cross-correlated and also serially correlated. While their approach might be convenient in the case of weakly pervasive factors they do not derive any asymptotic distribution and, in the case of strong factors, their rates are comparable to ours (see our Theorem 5). For the CP or PARAFAC decomposition, which is a special case of (2) having the factor tensor diagonal and not considered here, we refer to Han et al. (2024b), Chang et al. (2023), and Babii et al. (2022).

STRUCTURE. The rest of the article is organized as follows. In Section 2, we consider the Tucker decomposition and formulate the estimation of factor loading matrices and the factor score tensor by minimizing the least squares loss under the identifiability condition and give the KKT condition to the optimization problem, from which it naturally leads to a projection estimation algorithm. We also propose an iterative procedure to estimate the factor numbers. In Section 3, we investigate the theoretical properties of the initial estimator and the one-step estimator under mild conditions. We conclude in Section 4. The proofs of the main theorems are in Section 5. In the Supplementary Material we report additional technical details (Appendix A) as well as simulation results (Appendix B) and an empirical application on multi-category import-export network data (Appendix C).

NOTATION. The following notation is used extensively henceforth. For a tensor $\mathcal{X} \in \mathbb{R}^{p_1 \times p_2 \times \cdots \times p_K}$, the mode-$k$ product with a matrix $\mathbf{A} \in \mathbb{R}^{d \times p_k}$, denoting as $\mathcal{X} \times_k \mathbf{A}$, is a tensor of size $p_1 \times \cdots \times p_{k-1} \times d \times p_{k+1} \times \cdots p_K$. Element-wise, $(\mathcal{X} \times \mathbf{A})_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_K} = \sum_{i_k=1}^{p_k} x_{i_i,\ldots,i_k,\ldots,i_K} a_{j,i_k}$. Let also $p = \prod_{k=1}^K p_k$ and $p_{-k} = p/p_k$. The mode-k unfolding matrix of $\mathcal{X}$ is denoted by $\mathrm{mat}_k(\mathcal{X})$ and arranges all p_{-k} mode-k fibers of $\mathcal{X}$ to be the columns to get a $p_k \times p_{-k}$ matrix. For a matrix $\mathbf{A}$, $\mathbf{A}^\top$ is the transpose of $\mathbf{A}$, $\mathrm{Tr}(\mathbf{A})$ is the trace of $\mathbf{A}$, $\|\mathbf{A}\|$ is the spectral norm, $\|\mathbf{A}\|_F$ is the Frobenius norm, and $\|\mathbf{A}\|_{\max}$ is the maximum entry of $\mathbf{A}$ in absolute value. $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$. $\mathbf{I}_k$ represents a k-dimensional identity matrix. For two sequences of random variables X_n and Y_n, $X_n \lesssim Y_n$ means $X_n = O_P(Y_n)$ and $X_n \gtrsim Y_n$ means $Y_n = O_P(X_n)$. When used, the short-hand notation $j \in [K]$ indicates $1 \le j \le K$.

2. Methodology

2.1. Least squares and projection estimation

Let $\{\mathcal{X}_t, 1 \le t \le T\}$ be a $p_1 \times p_2 \times \cdots \times p_K$ sequence of tensor-valued random variables. The corresponding factor model based on the Tucker decomposition is given by

$$\mathcal{X}_t = \mathcal{F}_t \times_{k=1}^K \mathbf{A}_k + \mathcal{E}_t, \tag{3}$$

where: $\mathbf{A}_k$ is the $p_k \times r_k$ loading matrix for mode k, $\mathcal{F}_t$ is the $r_1 \times \cdots \times r_K$ common factor tensor, $\mathcal{E}_t$ is the $p_1 \times \cdots \times p_K$ idiosyncratic component tensor, and $\mathcal{S}_t = \mathcal{F}_t \times_{k=1}^K \mathbf{A}_k$ is the signal or common component tensor.

In order to estimate the loading matrices $\{\mathbf{A}_k, 1 \le k \le K\}$ and the tensor-valued factor $\{\mathcal{F}_t, 1 \le t \le T\}$, we define the quadratic loss function

$$L_1\left(\mathbf{A}_1, \ldots, \mathbf{A}_K, \mathcal{F}_1, \ldots, \mathcal{F}_T\right) := \frac{1}{Tp} \sum_{t=1}^T \left\| \mathcal{X}_t - \mathcal{F}_t \times_{k=1}^K \mathbf{A}_k \right\|_F^2. \tag{4}$$

The least squares estimator of the loadings and of the common factor are then defined as the solutions to the minimisation problem

$$\min_{\substack{\{\mathbf{A}_k, 1\le k\le K\} \\ \{\mathcal{F}_t, 1\le t\le T\}}} L_1\left(\mathbf{A}_1, \ldots, \mathbf{A}_K, \mathcal{F}_1, \ldots, \mathcal{F}_T\right) \tag{5}$$

of this loss. The minimisation problem defined in (5) can be solved by exploiting the following relation, valid for all $1 \le k \le K$ and all $1 \le t \le T$, $\left\| \mathcal{X}_t - \mathcal{F}_t \times_{k=1}^K \mathbf{A}_k \right\|_F^2 = \left\| \mathrm{mat}_k\left(\mathcal{X}_t\right) - \mathbf{A}_k \mathrm{mat}_k\left(\mathcal{F}_t\right) \left(\otimes_{j\in[K]\setminus\{k\}} \mathbf{A}_j\right)^\top \right\|_F^2$, where $\otimes_{j\in[K]\setminus\{k\}} \mathbf{A}_j = \mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1$. For ease of notation, we will henceforth use the following quantities: $\mathbf{B}_k := \otimes_{j\in[K]\setminus\{k\}} \mathbf{A}_j$; $\mathbf{X}_{k,t} := \mathrm{mat}_k\left(\mathcal{X}_t\right)$; and $\mathbf{F}_{k,t} := \mathrm{mat}_k(\mathcal{F}_t)$. Then, for any given $1 \le k \le K$, (4) can be written as

$$\begin{aligned} L_1\left(\mathbf{A}_1, \ldots, \mathbf{A}_K, \mathcal{F}_t\right) &= \frac{1}{Tp} \sum_{t=1}^T \left\| \mathcal{X}_t - \mathcal{F}_t \times_{k=1}^K \mathbf{A}_k \right\|_F^2 = \frac{1}{Tp} \sum_{t=1}^T \left\| \mathbf{X}_{k,t} - \mathbf{A}_k \mathbf{F}_{k,t} \mathbf{B}_k^\top \right\|_F^2 \\ &= \frac{1}{Tp} \sum_{t=1}^T \left[\mathrm{Tr}\left(\mathbf{X}_{k,t}^\top \mathbf{X}_{k,t}\right) - 2\mathrm{Tr}\left(\mathbf{X}_{k,t}^\top \mathbf{A}_k \mathbf{F}_{k,t} \mathbf{B}_k^\top\right) + p\mathrm{Tr}\left(\mathbf{F}_{k,t}^\top \mathbf{F}_{k,t}\right) \right] = L_1\left(\mathbf{A}_k, \mathbf{B}_k, \mathbf{F}_{k,t}\right). \end{aligned} \tag{6}$$

Hence, solving (5) is equivalent to finding $\mathbf{A}_k$, $\mathbf{B}_k$, and $\mathbf{F}_{k,t}$ for all $1 \le k \le K$ and all $1 \le t \le T$, such that they minimize (6). This problem can be solved as follows. If $\mathbf{A}_k$ and $\mathbf{B}_k$ are given, for any $1 \le t \le T$ the first order conditions give the following ordinary least squares solution:

$$\mathbf{F}_{k,t}^* := \frac{1}{p} \mathbf{A}_k^\top \mathbf{X}_{k,t} \mathbf{B}_k, \quad 1 \le t \le T. \tag{7}$$

Substituting $\mathbf{F}_{k,t}^*$ in (6) gives the (concentrated) loss

$$L_1\left(\mathbf{A}_k, \mathbf{B}_k\right) := \frac{1}{Tp} \sum_{t=1}^T \left[\mathrm{Tr}\left(\mathbf{X}_{k,t}^\top \mathbf{X}_{k,t}\right) - \frac{1}{p} \mathrm{Tr}\left(\mathbf{X}_{k,t}^\top \mathbf{A}_k \mathbf{A}_k^\top \mathbf{X}_{k,t} \mathbf{B}_k \mathbf{B}_k^\top\right) \right]. \tag{8}$$

In order to identify the minimizers of (8), we impose the following identifying constraints

$$\frac{\mathbf{A}_k^\top \mathbf{A}_k}{p_k} = \mathbf{I}_{r_k}, \quad \frac{\mathbf{B}_k^\top \mathbf{B}_k}{p_{-k}} = \mathbf{I}_{r_{-k}}, \tag{9}$$

which are $\sum_{k=1}^K r_k(r_k+1)/2$ constraints. These are in agreement of the idea that factors are pervasive along each mode (see Assumption 2*(i)* below).

Hence, minimizing (8) subject to the constraint in (9) is equivalent to minimizing the Lagrangian:

$$\mathcal{L}_1\left(\mathbf{A}_k, \mathbf{B}_k, \boldsymbol{\Theta}, \boldsymbol{\Lambda}\right) := L_1\left(\mathbf{A}_k, \mathbf{B}_k\right) + \mathrm{Tr}\left[\boldsymbol{\Theta}\left(\frac{1}{p_k}\mathbf{A}_k^\top \mathbf{A}_k - \mathbf{I}_{r_k}\right)\right] + \mathrm{Tr}\left[\boldsymbol{\Phi}\left(\frac{1}{p_{-k}}\mathbf{B}_k^\top \mathbf{B}_k - \mathbf{I}_{r_{-k}}\right)\right], \tag{10}$$

where the Lagrange multipliers $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ are symmetric matrices of dimensions $r_k \times r_k$ and $r_{-k} \times r_{-k}$, respectively.

Hence, the minimizers of (10) must solve the following system of equations given by the Karush-Kuhn–Tucker (KKT) conditions:

$$\frac{\partial \mathcal{L}_1}{\partial \mathbf{A}_k} = -\frac{1}{Tp}\sum_{t=1}^T \frac{2}{p}\mathbf{X}_{k,t}\mathbf{B}_k\mathbf{B}_k^\top \mathbf{X}_{k,t}^\top \mathbf{A}_k + \frac{2}{p_k}\mathbf{A}_k\boldsymbol{\Theta} = \mathbf{0}, \qquad \frac{\partial \mathcal{L}_1}{\partial \mathbf{B}_k} = -\frac{1}{Tp}\sum_{t=1}^T \frac{2}{p}\mathbf{X}_{k,t}^\top \mathbf{A}_k\mathbf{A}_k^\top \mathbf{X}_{k,t}\mathbf{B}_k + \frac{2}{p_{-k}}\mathbf{B}_k\boldsymbol{\Phi} = \mathbf{0},$$

In turn, the KKT conditions are equivalent to solving the following linear system:

$$\begin{cases} \left(\frac{1}{Tpp_{-k}}\sum_{t=1}^T \mathbf{X}_{k,t}\mathbf{B}_k\mathbf{B}_k^\top \mathbf{X}_{k,t}^\top\right)\mathbf{A}_k = \mathbf{A}_k\boldsymbol{\Theta}, \\ \left(\frac{1}{Tpp_k}\sum_{t=1}^T \mathbf{X}_{k,t}^\top \mathbf{A}_k\mathbf{A}_k^\top \mathbf{X}_{k,t}\right)\mathbf{B}_k = \mathbf{B}_k\boldsymbol{\Phi}, \end{cases} \text{or} \quad \begin{cases} \mathbf{M}_k\mathbf{A}_k = \mathbf{A}_k\boldsymbol{\Theta}, \\ \mathbf{M}_{-k}\mathbf{B}_k = \mathbf{B}_k\boldsymbol{\Phi}, \end{cases} \tag{11}$$

where we have used the short-hand notation

$$\mathbf{M}_k = \frac{1}{Tpp_{-k}}\sum_{t=1}^T \mathbf{X}_{k,t}\mathbf{B}_k\mathbf{B}_k^\top \mathbf{X}_{k,t}^\top \quad \text{and} \quad \mathbf{M}_{-k} = \frac{1}{Tpp_k}\sum_{t=1}^T \mathbf{X}_{k,t}^\top \mathbf{A}_k\mathbf{A}_k^\top \mathbf{X}_{k,t}.$$

Denoting the largest r_k eigenvalues of $\mathbf{M}_k$ in descending order as $\lambda_{k,1}, \dots, \lambda_{k,r_k}$ and collecting the corresponding normalized eigenvectors into the $p_k \times r_k$ matrix $\mathbf{U}_k = \left(\boldsymbol{u}_{k,1} \cdots \boldsymbol{u}_{k,r_k}\right)$, we have the solutions $\boldsymbol{\Theta}^* := \mathrm{diag}\left(\lambda_{k,1}, \dots, \lambda_{k,r_k}\right)$ and $\mathbf{A}_k^* := \sqrt{p_k}\,\mathbf{U}_k$. Notice that, after replacing $\mathbf{A}_k^*$ into the definition of $\mathbf{F}_{k,t}^*$ in (7), we obtain

$$\frac{1}{T}\sum_{t=1}^T \mathbf{F}_{k,t}^*\mathbf{F}_{k,t}^{*\top} = \frac{1}{Tp^2}\sum_{t=1}^T \mathbf{A}_k^{*\top}\mathbf{X}_{k,t}\mathbf{B}_k\mathbf{B}_k^\top \mathbf{X}_{k,t}^\top \mathbf{A}_k^* = \frac{p_{-k}}{p}\mathbf{A}_k^{*\top}\mathbf{M}_k\mathbf{A}_k^* = \boldsymbol{\Theta}^*,$$

which is diagonal by construction. In this way, for any $1 \le k \le K$ we are implicitly imposing $r_k(r_k-1)/2$ constraints, and, once this is repeated for all $1 \le k \le K$, we impose the remaining $\sum_{k=1}^K r_k(r_k+1)/2$ constraints needed to fully identify the model.

The estimators of $\mathbf{A}_k$ and $\mathbf{F}_{k,t}$ defined above require to know $\mathbf{M}_k$ which, in turn, requires to know the unobservable projection matrix $\mathbf{B}_k$, which depends on $\left\{\mathbf{A}_j, j \neq k\right\}$. A natural solution is to replace each $\mathbf{A}_j$, $j \neq k$, with a consistent initial estimator $\widehat{\mathbf{A}}_j$. While the construction of such initial estimators is extensively discussed in Section 2.2, here we discuss how to proceed once we have such estimators. Defining for short $\widehat{\mathbf{B}}_k := \otimes_{j\in[K]\setminus\{k\}}\widehat{\mathbf{A}}_j$ and $\widehat{\mathbf{Y}}_{k,t} := p_{-k}^{-1}\mathbf{X}_{k,t}\widehat{\mathbf{B}}_k$, and letting

$$\widetilde{\mathbf{M}}_k := \frac{1}{Tp_k}\sum_{t=1}^T \widehat{\mathbf{Y}}_{k,t}\widehat{\mathbf{Y}}_{k,t}^\top, \tag{12}$$

the projected estimator of the loadings $\widetilde{\mathbf{A}}_k$ can be constructed as

$$\widetilde{\mathbf{A}}_k := \sqrt{p_k}\,\widetilde{\mathbf{U}}_k, \tag{13}$$

where $\widetilde{\mathbf{U}}_k = \left(\widetilde{\boldsymbol{u}}_{k,1} \cdots \widetilde{\boldsymbol{u}}_{k,r_k}\right)$ is the $p_k \times r_k$ matrix whose columns are the normalized eigenvectors of $\widetilde{\mathbf{M}}_k$ corresponding to its largest r_k eigenvalues. Then, by iterating the procedure for all $1 \le k \le K$, we obtain the projected loadings estimators $\left\{\widetilde{\mathbf{A}}_k, 1 \le k \le K\right\}$.

We summarize the projection procedure in Algorithm 1 below, which extends, to the tensor case, the algorithm proposed in Yu et al. (2022) for the matrix factor model.

The projection method described above can be implemented recursively by plugging in the newly estimated $\left\{\widetilde{\mathbf{A}}_k, 1 \le k \le K\right\}$ to replace $\left\{\widehat{\mathbf{A}}, 1 \le k \le K\right\}$ in Step 2, and by iterating Steps 2–4, in Algorithm 1. The theoretical analysis (and the computational side) of the recursive solution is challenging; however, simulations show that the projection estimators with one single iteration perform sufficiently well compared with the recursive method. Indeed, as we show in Corollary 1 below, when $T \asymp p_1 \asymp \cdots \asymp p_K$, and for a suitable choice of $\left\{\widehat{\mathbf{A}}, 1 \le k \le K\right\}$, the single-iteration, projected estimator $\widetilde{\mathbf{A}}_k$ converges to $\mathbf{A}_k$ (up to a rotation) at rate $O_P\left(1/\sqrt{Tp_k}\right)$, which is the optimal rate one would obtain if all the other loading matrices were known in advance.

Finally, an estimator of the common factor tensor is obtained by linear projection as:

$$\widetilde{\mathcal{F}}_t := \frac{1}{p}\mathcal{X}_t \times_{k=1}^K \widetilde{\mathbf{A}}_k^\top, \quad 1 \le t \le T. \tag{14}$$

Algorithm 1 Least squares method for estimating the loading spaces

Input: tensor data $\{\mathcal{X}_t, 1 \le t \le T\}$, factor numbers $r_1, \dots, r_K$;
Output: factor loading matrices $\{\widetilde{\mathbf{A}}_k, 1 \le k \le K\}$;
1: obtain the initial estimators $\{\widehat{\mathbf{A}}_k, 1 \le k \le K\}$;
2: define $\widehat{\mathbf{Y}}_{k,t} := p_{-k}^{-1}\mathbf{X}_{k,t}\widehat{\mathbf{B}}_k$, where $\widehat{\mathbf{B}}_k := \otimes_{j\in[K]\setminus\{k\}}\widehat{\mathbf{A}}_j$, $1 \le k \le K$;
3: given $\{\widehat{\mathbf{Y}}_{k,t}, 1 \le k \le K\}$, define $\widetilde{\mathbf{M}}_k := (Tp_k)^{-1}\sum_{t=1}^T \widehat{\mathbf{Y}}_{k,t}(\widehat{\mathbf{Y}}_{k,t})^\top$, set $\widetilde{\mathbf{A}}_k$ as $\sqrt{p_k}$ times the matrix with as columns the first r_k normalized eigenvectors of $\widehat{\mathbf{M}}_k$;
4: Output the projection estimators as $\{\widetilde{\mathbf{A}}_k, 1 \le k \le K\}$.

Notice that, for any given $1 \le k \le K$, we also have: $\widetilde{\mathbf{F}}_{k,t} = \text{mat}_k\left(\widetilde{\mathcal{F}}_t\right) = p^{-1}\widetilde{\mathbf{A}}_k^\top \mathbf{X}_{k,t}\widetilde{\mathbf{B}}_k, \quad 1 \le t \le T$, where $\widetilde{\mathbf{B}}_k := \otimes_{j\in[K]\setminus\{k\}}\widetilde{\mathbf{A}}_j$. It is easy to see that this estimator is the least squares estimator of the mode-k matricization of the common factor tensor in (7), computed using the estimated loadings matrices.

2.2. Initial projection matrices

We now discuss the choice of the initial projection matrices $\widehat{\mathbf{A}}_k$. For the sake of notational simplicity, let $\mathbf{E}_{k,t} := \text{mat}_k\left(\mathcal{E}_t\right)$. Then, for any given $1 \le k \le K$, the mode-k matricization of model (3) is given by

$$\mathbf{X}_{k,t} = \mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^\top + \mathbf{E}_{k,t}. \tag{15}$$

Furthermore, letting $\mathbf{F}_{k,t} = \left(\boldsymbol{f}_{k,t,\cdot 1} \cdots \boldsymbol{f}_{k,t,\cdot r_{-k}}\right)$ and $\mathbf{E}_{k,t} = \left(\boldsymbol{e}_{k,t,\cdot 1} \cdots \boldsymbol{e}_{k,t,\cdot p_{-k}}\right)$, each of the p_{-k} columns of $\mathbf{X}_{k,t}$ follows a vector factor model

$$\boldsymbol{x}_{k,t,\cdot j} = \mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_{k,j\cdot}^\top + \boldsymbol{e}_{k,t,\cdot j} = \mathbf{A}_k\overline{\boldsymbol{f}}_{k,t,\cdot j} + \boldsymbol{e}_{k,t,\cdot j}, \quad 1 \le t \le T, \quad 1 \le j \le p_{-k}. \tag{16}$$

Hence, in order to estimate $\mathbf{A}_k$, we can consider each column of $\mathbf{X}_{k,t}$ as an individual vector-valued time series, and apply the classical PCA estimator for vector time series (Bai, 2003).

Whilst details are in the next sections, here we offer a heuristic preview of the intuition behind PCA in this tensor setting. For any given $1 \le k \le K$ define the scaled mode-wise sample covariance matrix as

$$\widehat{\mathbf{M}}_k := \frac{1}{Tp}\sum_{t=1}^T\sum_{j=1}^{p_{-k}} \boldsymbol{x}_{k,t,\cdot j}\boldsymbol{x}_{k,t,\cdot j}^\top = \frac{1}{Tp}\sum_{t=1}^T \mathbf{X}_{k,t}\mathbf{X}_{k,t}^\top. \tag{17}$$

Under the usual assumption of weak dependence between factors and idiosyncratic components, it approximately holds that

$$\begin{aligned}\widehat{\mathbf{M}}_k &\approx \frac{1}{p_k}\mathbf{A}_k\left(\frac{1}{Tp_{-k}}\sum_{t=1}^T\sum_{j=1}^{p_{-k}}\overline{\boldsymbol{f}}_{k,t,\cdot j}\overline{\boldsymbol{f}}_{k,t,\cdot j}^\top\right)\mathbf{A}_k^\top + \frac{1}{p_k}\frac{1}{Tp_{-k}}\sum_{t=1}^T\sum_{j=1}^{p_{-k}}\boldsymbol{e}_{k,t,\cdot j}\boldsymbol{e}_{k,t,\cdot j}^\top \\ &= \frac{1}{p_k}\mathbf{A}_k\left(\frac{1}{T}\sum_{t=1}^T\sum_{j=1}^{r_{-k}}\mathbf{F}_{k,t,\cdot j}\mathbf{F}_{k,t,\cdot j}^\top\right)\mathbf{A}_k^\top + \frac{1}{Tp_k}\left(\frac{1}{p_{-k}}\sum_{t=1}^T\sum_{j=1}^{p_{-k}}\boldsymbol{e}_{k,t,\cdot j}\boldsymbol{e}_{k,t,\cdot j}^\top\right).\end{aligned} \tag{18}$$

Now, on one hand, the first term on the right-hand-side of (18) converges to a matrix of rank r_k, since the term $T^{-1}\sum_{t=1}^T\sum_{j=1}^{r_{-k}}\boldsymbol{f}_{k,t,\cdot j}\boldsymbol{f}_{k,t,\cdot j}^\top$ converges to a positive definite $r_k \times r_k$ matrix, and since by means of the identifying constraint (9) we imposed pervasiveness of the factors as $p_k \to \infty$. On the other hand, under the usual assumption of weakly cross-correlated idiosyncratic components, the second term on the right-hand-side of (18) is asymptotically negligible, as $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$. As a result, the leading r_k eigenvalues of $\widehat{\mathbf{M}}_k$ dominate over the remaining $p_k - r_k$, and, by Davis-Kahan's $\sin(\Theta)$ theorem (Davis and Kahan, 1970; Yu et al., 2015), the r_k leading eigenvectors of $\widehat{\mathbf{M}}_k$ asymptotically span the same column space as that of the columns of $\mathbf{A}_k$. Therefore, letting $\widehat{\mathbf{U}}_k$ be the $p_k \times r_k$ matrix having as columns the r_k leading normalized eigenvectors of $\widehat{\mathbf{M}}_k$, we define an initial estimator the loadings matrix $\mathbf{A}_k$ as

$$\widehat{\mathbf{A}}_k := \sqrt{p_k}\,\widehat{\mathbf{U}}_k. \tag{19}$$

Similarly, the estimator for $\mathbf{B}_k$ can be naturally chosen as $\widehat{\mathbf{B}}_k := \otimes_{j=1,j\neq k}^K\widehat{\mathbf{A}}_j$.

2.2.1. Alternative initial projection matrices

Some other choices of initial estimates of $\mathbf{A}_k$ are admissible as long as two sufficient conditions stated below in (25)–(26) are fulfilled. In particular, instead of setting the estimator of $\mathbf{B}_k$ as $\widehat{\mathbf{B}}_k$ defined above, we can choose another estimator $\widehat{\mathbf{B}}_k^*$ as $\sqrt{p_{-k}}$ times the matrix with columns being the leading r_{-k} normalized eigenvectors of $\widehat{\mathbf{M}}_{-k} := (Tp)^{-1}\sum_{t=1}^T\mathbf{X}_{k,t}^\top\mathbf{X}_{k,t}$, a choice which is again motivated by the matrix factor model form in (15). We denote the resulted projection estimators, obtained using $\widehat{\mathbf{B}}_k^*$, as $\widetilde{\mathbf{A}}_k^*$. The detailed procedure is given in Algorithm 2.

Algorithm 2 Projected method for estimating the loading spaces

Input: tensor data $\{\mathcal{X}_t, 1 \le t \le T\}$, factor numbers $r_1, \dots, r_K$;
Output: factor loading matrices $\{\widetilde{\mathbf{A}}_k^*, 1 \le k \le K\}$;

1: obtain the initial estimators $\{\widehat{\mathbf{A}}_k, 1 \le k \le K\}$ and $\{\widehat{\mathbf{B}}_k^*, 1 \le k \le K\}$ by $\sqrt{p_{-k}}$ times the matrix with columns being the first r_{-k} normalized eigenvectors of $\widehat{\mathbf{M}}_{-k} := (Tp)^{-1} \sum_{t=1}^T \mathbf{X}_{k,t}^\top \mathbf{X}_{k,t}$;
2: define $\widehat{\mathbf{Y}}_{k,t}^* := p_{-k}^{-1} \mathbf{X}_{k,t} \widehat{\mathbf{B}}_k^*, 1 \le k \le K$;
3: given $\{\widehat{\mathbf{Y}}_{k,t}^*, 1 \le k \le K\}$, define $\widetilde{\mathbf{M}}_k^* := (Tp_k)^{-1} \sum_{t=1}^T \widehat{\mathbf{Y}}_{k,t}^* \widehat{\mathbf{Y}}_{k,t}^{*\top}$, obtain $\widetilde{\mathbf{A}}_k^*$ as $\sqrt{p_k}$ times the matrix with columns being the first r_k normalized eigenvectors of $\widetilde{\mathbf{M}}_k^*$.

2.3. Estimation of factor numbers

We discuss two types of estimators of the numbers of common factors r_k, $1 \le k \le K$, both based on the eigenvalue-ratio principle (Lam and Yao, 2012; Ahn and Horenstein, 2013).

A first estimator can be based on the simple mode-wise sample covariance matrix $\widehat{\mathbf{M}}_k$, whereas a second one can instead be based on the mode-wise sample covariance matrix of the projected data, i.e., $\widetilde{\mathbf{M}}_k$; indeed, the results in He et al. (2023) for matrix-valued time series seem to suggest that using projected data results in better estimators of the number of common factors, especially in small samples.

In particular, we will consider the following family of modified eigenvalue-ratio, criteria

$$\hat{r}_k^{\text{IE-ER}} := \underset{1\le j\le r_{\max}}{\operatorname{argmax}} \frac{\lambda_j\left(\widehat{\mathbf{M}}_k\right)}{\lambda_{j+1}\left(\widehat{\mathbf{M}}_k\right) + \hat{c}\delta_{p_1,\dots,p_K,T}}, \qquad \hat{r}_k^{\text{PE-ER}} := \underset{1\le j\le r_{\max}}{\operatorname{argmax}} \frac{\lambda_j\left(\widetilde{\mathbf{M}}_k\right)}{\lambda_{j+1}\left(\widetilde{\mathbf{M}}_k\right) + \tilde{c}\delta_{p_1,\dots,p_K,T}}, \tag{20}$$

defined for $1 \le k \le K$, where $r_{\max}$ is a predetermined positive constant such that $\max_{1\le k\le K} r_k < r_{\max} < \min\left\{\min_{1\le k\le K} p_k, T\right\}$, the constants $\hat{c}, \tilde{c} \in (0, \infty)$ are user-chosen and

$$\delta_{p_1,\dots,p_K,T} = \frac{1}{\sqrt{Tp_{-k}}} + \frac{1}{p_k}. \tag{21}$$

Under our assumption of pervasive factors (see Assumption 2*(i)* below) and weakly correlated idiosyncratic components, the eigen-gap between the first r_k eigenvalues of $\widehat{\mathbf{M}}_k$ ($\widetilde{\mathbf{M}}_k$) and the remaining ones widens as $p_k \to \infty$. Thus, (20) will take the maximum value at $j = r_k$. The rationale for the extra $\hat{c}\delta_{p_1,\dots,p_K,T}$ (respectively $\tilde{c}\delta_{p_1,\dots,p_K,T}$) term at the denominator of $\hat{r}_k^{\text{IE-ER}}$ (respectively $\hat{r}_k^{\text{PE-ER}}$) is that there is no theoretical guarantee that, for some $j > r_k$, $\lambda_{j+1}\left(\widehat{\mathbf{M}}_k\right)$ (and, similarly, $\lambda_{j+1}\left(\widetilde{\mathbf{M}}_k\right)$), will not be very small, thus artificially inflating the ratio $\lambda_j\left(\widehat{\mathbf{M}}_k\right)/\lambda_{j+1}\left(\widehat{\mathbf{M}}_k\right)$. The presence of $\delta_{p_1,\dots,p_K,T}$ (which is of the same order of magnitude as the upper bound for $\lambda_{j+1}\left(\widehat{\mathbf{M}}_k\right)$ when $j \ge r_k$) serves the purpose of "weighing down" the eigenvalue ratio and avoid such degeneracy.

As far as the choice of $\hat{c}$ and $\tilde{c}$ is concerned, theoretically any positive, finite number is acceptable. In particular, two approaches are possible. On the one hand, one can choose $\hat{c}$ and $\tilde{c}$ adaptively, using different subsamples and choosing the values of $\hat{c}$ and $\tilde{c}$ which offer stable estimates across such subsamples, in a similar spirit to Hallin and Liška (2007) and Alessi et al. (2010). Alternatively, following a similar proposal as in Trapani (2018) and Barigozzi and Trapani (2021), one can use

$$\hat{c} = \sum_{j=1}^{p_k} \lambda_j\left(\widehat{\mathbf{M}}_k\right), \qquad \tilde{c} = \sum_{j=1}^{p_k} \lambda_j\left(\widetilde{\mathbf{M}}_k\right). \tag{22}$$

Further, operationally, in order to compute $\widetilde{\mathbf{M}}_k$ we need to compute $\widehat{\mathbf{B}}_k$ first; in turn, this requires some prior knowlegde of r_{-k} (or, equivalently, of r_j, $j \ne k$). In order to address this, we propose to determine the numbers of factors by the following Algorithm 3.

3. Theoretical results

In this section, we present the main assumptions (Section 3.1), and we then present the asymptotics of the estimated loadings (Sections 3.2 and 3.3), of the estimated common factors and common components (Section 3.4), and of the estimators of the numbers of common factors (Section 3.5).

3.1. Assumptions

The following assumptions are required for our theory, and may be viewed as higher-order extensions of those adopted for large-dimensional matrix factor model by Chen and Fan (2023) and Yu et al. (2022).

Algorithm 3 Projected estimation of the numbers of factor

Input: tensor data $\{\mathcal{X}_t, 1 \le t \le T\}$, maximum number $r_{\max}$, maximum iterative step m
Output: factor numbers $\{\hat{r}_k^{\text{PE-ER}}, 1 \le k \le K\}$

1: initialize: $r_k^{(0)} = r_{\max}, 1 \le k \le K$;
2: given $r_k^{(0)}$, obtain the initial estimators $\{\hat{\mathbf{A}}_k, 1 \le k \le K\}$ and set $\hat{\mathbf{A}}_k^{(0)} = \hat{\mathbf{A}}_k$;
3: for $1 \le s \le m$, compute $\hat{\mathbf{B}}_k^{(s)} = \otimes_{j \in [K] \setminus \{k\}} \hat{\mathbf{A}}_j^{(s-1)}$;
4: compute $\widetilde{\mathbf{M}}_k^{(s)}$, obtain $\hat{r}_k^{(s)}$ for $1 \le k \le K$;
5: update $\hat{\mathbf{A}}_k^{(s)}$ as $\sqrt{p_k}$ times the matrix with as columns the first $\hat{r}_k^{(s)}$ normalized eigenvectors of $\hat{\mathbf{M}}_k$;
6: Repeat steps 3 to 5 until $\hat{r}_k^{(s)} = \hat{r}_k^{(s-1)}$, for all $1 \le k \le K$, or reach the maximum number of iterations;
7: Output the last step estimator as $\{\hat{r}_k^{\text{PE-ER}}, 1 \le k \le K\}$.

Assumption 1. It holds that: (i) for all $1 \le t \le T$, $\mathbb{E}\left(\text{Vec}\left(\mathcal{F}_t\right)\right) = 0$ and $\mathbb{E}\left(\left\|\text{Vec}\left(\mathcal{F}_t\right)\right\|^4\right) \le c$ for some $c < \infty$ independent of t; (ii) for all $1 \le k \le K$, as $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$, $T^{-1}\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{F}_{k,t}^\top \xrightarrow{P} \boldsymbol{\Sigma}_k$, where $\boldsymbol{\Sigma}_k$ is an $r_k \times r_k$ positive definite matrix with finite, distinct eigenvalues and spectral decomposition $\boldsymbol{\Sigma}_k = \boldsymbol{\Gamma}_k\boldsymbol{\Lambda}_k\boldsymbol{\Gamma}_k^\top$; (iii) the factor numbers $\left\{r_k, 1 \le k \le K\right\}$ are fixed integers as $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$.

Assumption 2. It holds that: (i) for all $1 \le k \le K$, $\left\|\mathbf{A}_k\right\|_{\max} \le \bar{a}_k$ for some $\bar{a}_k < \infty$ independent of p_k; (ii) $\left\|p_k^{-1}\mathbf{A}_k^\top\mathbf{A}_k - \mathbf{I}_k\right\|_F \to 0$ as $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$.

Assumption 1 imposes finite fourth moments of the mode-k unfolding factor matrices for all $1 \le k \le K$, and it ensures that the second-order sample moments converge to positive definite matrices $\boldsymbol{\Sigma}_k$, which are also assumed to have distinct eigenvalues to ensure the identifiability of eigenvectors. This assumption is typical in vector factor models, and it can be compared e.g. with Assumption A in Bai (2003).

As far as Assumption 2 is concerned, this is also a standard requirement in the context of vector (and matrix) factor models, and we refer to e.g. Assumption B in Bai (2003) for comparison. In the assumption, the common factors are assumed to be "strong" or pervasive - see, e.g., the recent contribution by Bai and Ng (2023) for a treatment, and useful insights, on the notion of strong versus weak common factors and the case of vector data. We would like to point out that extensions to the case of "weak" common factors - where $\left\|\mathbf{A}_k\right\|_F^2 = c_0 p_k^{\alpha_k}$ for some $0 < \alpha_k < 1$ - are in principle possible, at the price of more complicated algebra, even in the context of tensor-valued time series; He et al. (2023) offer a comprehensive treatment for the case of matrix-valued time series.

Assumption 3. It holds that: (i) for all $1 \le t \le T$, $1 \le i_k \le p_k$ and $1 \le k \le K$, $\mathbb{E}\left(e_{t,i_1,\dots,i_K}\right) = 0$ and $\mathbb{E}\left(e_{t,i_1,\dots,i_K}\right)^4 \le c$ for some $c < \infty$ independent of t and i_k; (ii) for all $1 \le k \le K$, and all p_k and T, it holds that $(Tp)^{-1}\sum_{t,s=1}^T\sum_{i,l=1}^{p_k}\sum_{j,h=1}^{p_{-k}}\left|\mathbb{E}\left(e_{t,k,lj}e_{s,k,ih}\right)\right| \le c$, for some $c < \infty$ independent of k, p_k and T; (iii) for all $1 \le i, l_1 \le p_k$, $1 \le j, h_1 \le p_{-k}$, $1 \le k \le K$, and all p_k and T, it holds that

$$\sum_{s=1}^T\sum_{l_2=1}^{p_k}\sum_{h_2=1}^{p_{-k}}\left|\text{Cov}\left(e_{t,k,ij}e_{t,k,l_1j}, e_{s,k,ih_2}e_{s,k,l_2h_2}\right)\right| \le c, \quad \sum_{s=1}^T\sum_{l_2=1}^{p_k}\sum_{h_2=1}^{p_{-k}}\left|\text{Cov}\left(e_{t,k,ij}e_{t,k,ih_1}, e_{s,k,l_2j}e_{s,k,l_2h_2}\right)\right| \le c,$$
$$\sum_{s=1}^T\sum_{l_2=1}^{p_k}\sum_{h_2=1}^{p_{-k}}\left|\text{Cov}\left(e_{t,k,ij}e_{t,k,l_1h_1}, e_{s,k,ij}e_{s,k,l_2h_2}\right)\right| + \left|\text{Cov}\left(e_{t,k,l_1j}e_{t,k,ih_1}, e_{s,k,l_2j}e_{s,k,ih_2}\right)\right| \le c,$$

for some $c < \infty$ independent of k, p_k and T.

According to Assumption 3, weak serial and cross-sectional correlation are allowed along each mode; in particular, the summability conditions in part *(ii)* of the assumptions can be verified under more primitive assumptions of weak dependence, at least across t; He et al. (2023) study the case of stationary causal processes approximable by an m-dependent sequence, in the case of matrix-valued time series. Essentially the same extension can be studied in this context. We note that the following equations are nested within the assumption

$$\sum_{s=1}^T\sum_{l=1}^{p_k}\sum_{h=1}^{p_{-k}}\left|\mathbb{E}\left(e_{t,k,ij}e_{s,k,lh}\right)\right| \le c, \quad \sum_{l=1}^{p_k}\sum_{h=1}^{p_{-k}}\left|\mathbb{E}\left(e_{t,k,lj}e_{t,k,ih}\right)\right| \le c, \tag{23}$$

$$\frac{1}{p}\sum_{i,l=1}^{p_k}\sum_{j,h=1}^{p_{-k}}\left|\mathbb{E}\left(e_{t,k,lj}e_{t,k,ih}\right)\right| \le c, \quad \sum_{l=1}^{p_k}\left|\mathbb{E}\left(e_{t,k,lj}e_{t,k,ij}\right)\right| \le c, \tag{24}$$

for all $1 \le i \le p_k$, $1 \le j \le p_{-k}$ and $1 \le t \le T$. Part *(iii)* of the assumption controls the second-order correlation among the elements of mode-k unfolding matrices of the noise tensors, and it implies existence and summability of all 4th order cumulants of the process $\left\{\mathbf{e}_{k,t}, t \in \mathbb{Z}\right\}$ - in turn, this is a necessary and sufficient condition for $(Tp_{-k})^{-1}\sum_{t=1}^T \mathbf{e}_{k,t}\mathbf{e}_{k,t}^\top \xrightarrow{P} p_{-k}^{-1}\mathbb{E}\left(\mathbf{e}_{k,t}\mathbf{e}_{k,t}^\top\right)$, as $T \to \infty$

(see e.g. Hannan, 1970, pp. 209–211). Essentially the same set of assumptions is required in the context of matrix factor models (see Yu et al., 2022 and He et al., 2023)

Assumption 4. It holds that: (i) for all $1 \le t \le T$, $1 \le k \le K$ and any couple of deterministic vectors $\mathbf{v}$ of dimension p_k and $\mathbf{w}$ of dimension p_{-k} such that $\|\mathbf{v}\| = 1$ and $\|\mathbf{w}\| = 1$, $\mathbb{E}\left\| T^{-1/2} \sum_{t=1}^{T} \mathbf{F}_{k,t} \left(\mathbf{v}^\top \mathbf{E}_{k,t} \mathbf{w}\right)\right\|_F^2 \le c$, for some $c < \infty$ independent of k and T; (ii) letting $\zeta_{i_1,\dots,i_K} = \text{Vec}\left(T^{-1/2} \sum_{t=1}^{T} \mathcal{F}_t e_{t,i_1,\dots,i_K}\right)$, then for all $1 \le k \le K$, $1 \le i_k \le p_k$ and all p_k

$$\left\| \sum_{h=1,h\neq k}^{K} \sum_{i_h'=1}^{p_h} \mathbb{E}\left(\zeta_{i_1,\dots,i_K} \otimes \zeta_{i_1',\dots,i_{k-1}',i_k,i_{k+1}',\dots,i_K'}\right)\right\|_{\max} \le c,$$

$$\left\| \sum_{h=1,h\neq k}^{K} \sum_{i_h'=1}^{p_h} \sum_{l=1,l\neq \ell}^{K} \sum_{j_l'=1}^{p_l} Cov\left(\zeta_{i_1,\dots,i_K} \otimes \zeta_{j_1,\dots,j_K}, \zeta_{i_1',\dots,i_{k-1}',i_k,i_{k+1}',\dots,i_K'} \otimes \zeta_{j_1',\dots,j_{\ell-1}',j_\ell,j_{\ell+1}',\dots,j_K'}\right)\right\|_{\max} \le c,$$

for some $c < \infty$ independent of k, p_k, p_ℓ and i_k, j_ℓ.

Assumption 4 is automatically satisfied if the idiosyncratic component and the factor tensor processes are mutually independent, but it allows for possible correlation between the common factors and the idiosyncratic component. Indeed, given that $\mathbf{v}^\top \mathbf{E}_{k,t} \mathbf{w}$ is a random variable with zero mean and bounded variance (because of Assumption 3), part *(i)* of Assumption 4 states that that the correlation between $\{\mathbf{F}_{k,t}\}$ and $\{\mathbf{v}^\top \mathbf{E}_{k,t} \mathbf{w}\}$ is also weak. By the same token, part *(ii)* controls the high-order dependence between the common factors and the idiosyncratic components.

Assumption 5. It holds that: (i) for all $1 \le i \le p_k$ and all $1 \le k \le K$, as $\min\{T, p_1, \dots, p_K\} \to \infty$, $(Tp_{-k})^{-1/2} \sum_{t=1}^{T} \mathbf{F}_{k,t} \mathbf{B}_k^\top \mathbf{e}_{k,t,i\cdot}^\top \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \mathbf{V}_{ki}\right)$, where $\mathbf{V}_{ki} = \lim_{\min\{T,p_1,\dots,p_K\}\to\infty} (Tp_{-k})^{-1} \sum_{t,s=1}^{T} \mathbb{E}\left(\mathbf{F}_{k,t} \mathbf{B}_k^\top \mathbf{e}_{k,t,i\cdot}^\top \mathbf{e}_{k,s,i\cdot} \mathbf{B}_k \mathbf{F}_{k,s}^\top\right)$ is a positive definite $r_k \times r_k$ matrix with $\|\mathbf{V}_{ki}\|_F < \infty$; (ii) for all $1 \le t \le T$ and all $1 \le k \le K$, as $\min\{T, p_1, \dots, p_K\} \to \infty$, $p^{-1/2}\left(\mathbf{B}_k \otimes \mathbf{A}_k\right)^\top \text{Vec}\left(\mathbf{E}_{k,t}\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \mathbf{W}_{kt}\right)$, where $\mathbf{W}_{kt} = \lim_{\min\{T,p_1,\dots,p_K\}\to\infty} p^{-1}\left(\mathbf{B}_k \otimes \mathbf{A}_k\right)^\top \mathbb{E}\left\{\text{Vec}\left(\mathbf{E}_{k,t}\right)\text{Vec}\left(\mathbf{E}_{k,t}\right)^\top\right\}\left(\mathbf{B}_k \otimes \mathbf{A}_k\right)$ is a positive definite $r \times r$ matrix with $\|\mathbf{W}_{kt}\|_F < \infty$.

Assumption 5 is required only to derive the limiting laws of the estimated loadings and common factors, without being required to derive rates. In principle, it could be derived from more primitive assumptions on the dependence structure of the data, as, e.g., strong mixing.

We now study the asymptotic properties of the "initial" estimator of the loadings, $\widehat{\mathbf{A}}_k$, defined in (19), of the projection based one $\widetilde{\mathbf{A}}_k$, defined in (13), and of the estimated factor tensor defined in (14). We also study the consistency of the estimators of the numbers of factors defined in Section 2.3.

3.2. Asymptotic properties of the initial estimator of the loadings

The following theorem establishes the convergence rate of the initial projection estimators discussed in Section 2.2 and defined in (19). Henceforth, we use the notation

$$w_k := \frac{1}{p_k^2} + \frac{1}{Tp_{-k}}.$$

Theorem 1. *We assume that Assumptions 1–4 are satisfied. Then, for any given $1 \le k \le K$, there exists an $r_k \times r_k$ invertible matrix $\widehat{\mathbf{H}}_k$ such that, as $\min\{T, p_1, \dots, p_K\} \to \infty$, it holds that $\widehat{\mathbf{H}}_k \widehat{\mathbf{H}}_k^\top \xrightarrow{P} \mathbf{I}_{r_k}$, and $p_k^{-1}\left\|\widehat{\mathbf{A}}_k - \mathbf{A}_k \widehat{\mathbf{H}}_k\right\|_F^2 = O_P\left(w_k\right)$.*

The theoretical convergence rates for the initial estimators incorporates the results for matrix-valued case, studied in Yu et al. (2022), as a special case. We note that essentially the same result is obtained when considering a vector factor model, and the result in the theorem can be contrasted with Theorem 2 in Bai (2003), where results are reported for the unit-specific estimates of the loadings - in such a case, p_k is the only cross-sectional dimension (denoted as N in Bai, 2003), and the other dimensions p_{-k} are equal to 1.

The next theorem presents the asymptotic distributions of the initial estimators of the loadings, and again it can be read in conjunction with Theorem 2 in Bai (2003).

Theorem 2. *We assume that Assumptions 1–5 are satisfied. Then, for any given $1 \le i \le p_k$ and $1 \le k \le K$, as $\min\{T, p_1, \dots, p_K\} \to \infty$*

(i) if $Tp_{-k} = o\left(p_k^2\right)$, then it holds that $\sqrt{Tp_{-k}}\left(\widehat{\mathbf{A}}_{k,i\cdot}^\top - \widehat{\mathbf{H}}_k^\top \mathbf{A}_{k,i\cdot}^\top\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \mathbf{\Lambda}_k^{-1} \mathbf{\Gamma}_k^\top \mathbf{V}_{ki} \mathbf{\Gamma}_k \mathbf{\Lambda}_k^{-1}\right)$, where $\mathbf{\Gamma}_k$ and $\mathbf{\Lambda}_k$ are defined in Assumption 1(iii) and $\mathbf{V}_{ki}$ is defined in Assumption 5;

(ii) if $Tp_{-k} \gtrsim p_k^2$, then it holds that $\left\|\widehat{\mathbf{A}}_{k,i\cdot}^\top - \widehat{\mathbf{H}}_k^\top \mathbf{A}_{k,i\cdot}^\top\right\|_F = O_P\left(p_k^{-1}\right)$.

3.3. Asymptotic properties of the projection-based estimators

We now turn to studying the projection-based estimator defined in (13). As can be expected, the properties of $\widetilde{\mathbf{A}}_k$ are bound to depend on the properties of the initial projection matrices.

We begin by stating sufficient conditions on the rates of convergence of the initial estimators which guarantees the consistency of the projection-based estimator.

Sufficient Conditions. Consider a generic initial estimators $\left\{\widehat{\mathbf{A}}_k^{(0)},\, 1 \le k \le K\right\}$, and define the corresponding $\widehat{\mathbf{B}}_k^{(0)} := \otimes_{j\in[K]\setminus\{k\}}\widehat{\mathbf{A}}_j^{(0)}$. For any given $1 \le k \le K$, there exist $r_{-k} \times r_{-k}$ matrices $\widehat{\mathbf{H}}_{-k}$ such that as $\min\{T, p_1, \dots, p_k\} \to \infty$, $\widehat{\mathbf{H}}_{-k}\widehat{\mathbf{H}}_{-k}^\top \xrightarrow{p} \mathbf{I}_{r_{-k}}$ and

$$\frac{1}{p_{-k}}\left\|\widehat{\mathbf{B}}_k^{(0)} - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2 = O_P\left(w_{-k}\right), \tag{25}$$

$$\frac{1}{p_k}\left\|\frac{1}{Tp_{-k}}\sum_{s=1}^{T}\mathbf{E}_{k,s}\left(\widehat{\mathbf{B}}_k^{(0)} - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\mathbf{F}_{k,s}^\top\right\|_F^2 = O_P\left(m_{-k}\right), \tag{26}$$

with $w_{-k}, m_{-k} \to 0$.

In essence, the sufficient conditions presented above require the initial estimator to have an estimation error that vanishes "sufficiently fast". We will see that any estimator which satisfies (25)–(26) results in a projection-based estimate, $\widetilde{\mathbf{A}}_k$, which consistently estimates the space spanned by $\mathbf{A}_k$. In particular, as shown in Lemma 1 below, the PCA estimator defined in (19) satisfies these sufficient conditions.

The following theorem presents the rate of convergence of the projected estimators $\widetilde{\mathbf{A}}_k$ defined in (13) when computed using any set of initial estimators satisfying the sufficient conditions in (25)–(26).

Henceforth, we use the notation

$$\widetilde{w}_k := \frac{1}{Tp_{-k}} + \frac{1}{p^2} + w_{-k}\left(\frac{1}{Tp_k} + \frac{1}{p_k^2}\right) + m_{-k}.$$

Theorem 3. *We assume that Assumptions 1–4 are satisfied, and that the projected estimators $\left\{\widetilde{\mathbf{A}}_k, 1 \le k \le K\right\}$ are computed using an initial estimator satisying (25)–(26). Then, for any given $1 \le k \le K$, there exists an $r_k \times r_k$ invertible matrix $\widetilde{\mathbf{H}}_k$ such that, as $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$, it holds that $\widetilde{\mathbf{H}}_k^\top\widetilde{\mathbf{H}}_k \xrightarrow{P} \mathbf{I}_{r_k}$, and $p_k{}^{-1}\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2 = O_P\left(\widetilde{w}_k\right)$.*

Theorem 3 illustrates how the convergence rates of the projection estimators depend on the convergence rates of the initial estimators, and it is a general and rather abstract result. In order to make it more practically useful, we now discuss the impact of using the PCA estimator $\widehat{\mathbf{A}}_k$ on the properties of $\widetilde{\mathbf{A}}_k$.

We begin by showing that, when using $\left\{\widehat{\mathbf{A}}_k, 1 \le k \le K\right\}$ as the initial estimators, these satisfy the sufficient conditions (25)–(26).

Lemma 1. *We assume that Assumptions 1–4 are satisfied. Then the initial estimators $\left\{\widehat{\mathbf{A}}_k, 1 \le k \le K\right\}$ defined in (19) satisfy (25)–(26) with $\widehat{\mathbf{H}}_{-k} = \otimes_{j\in[K]\setminus\{k\}}\widehat{\mathbf{H}}_k$ and*

$$w_{-k} = \sum_{j=1, j\neq k}^{K}\left(\frac{1}{p_j^2} + \frac{1}{Tp_{-j}}\right), \quad m_{-k} = \sum_{j=1, j\neq k}^{K}\left(\frac{1}{Tp_j^2} + \frac{1}{T^2p_{-j}^2}\right).$$

As a consequence of Lemma 1, we can prove the following version of Theorem 3.

Corollary 1. *We assume that Assumptions 1–4 are satisfied, and that the and that the projected estimators $\left\{\widetilde{\mathbf{A}}_k, 1 \le k \le K\right\}$ are computed using $\left\{\widehat{\mathbf{A}}_k, 1 \le k \le K\right\}$ defined in (19). Then, $\left\{\widetilde{\mathbf{A}}_k, 1 \le k \le K\right\}$ satisfy Theorem 3 with*

$$\widetilde{w}_k = \frac{1}{Tp_{-k}} + \frac{1}{p^2} + \sum_{j=1, j\neq k}^{K}\left(\frac{1}{Tp_j^2} + \frac{1}{T^2p_{-j}^2} + \frac{1}{p_k^2p_j^4}\right).$$

Corollary 1 states that if we choose - as initial estimators - the PCA based $\widehat{\mathbf{A}}_k$ defined in (19), then the one-step projection estimator $\widetilde{\mathbf{A}}_k$ might achieve faster convergence rates than $\widehat{\mathbf{A}}_k$. In particular, by comparing the convergence rates of $\widehat{\mathbf{A}}_k$ in Theorem 1, and the ones of $\widetilde{\mathbf{A}}_k$ in Corollary 1 , we see that the latter might improve on the former, especially when p_k is "small".

Finally, the following theorem presents the asymptotic distributions of the projected estimators of the loadings.

Theorem 4. *We assume that Assumptions 1–5 are satisfied, and that the projected estimators $\left\{\widetilde{\mathbf{A}}_k, 1 \le k \le K\right\}$ are computed using $\left\{\widehat{\mathbf{A}}_k, 1 \le k \le K\right\}$ defined in (19). Then, for any given $1 \le i \le p_k$ and $1 \le k \le K$, as $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$*

(i) *if $K \le 2$ and $Tp_{-k} = o\left(\min\left\{p^2, T^2p_{-j}^2, Tp_j^2, p_k^2p_j^2\right\}\right)$, for all $j \neq k$, then it holds that*

$$\sqrt{Tp_{-k}}\left(\widetilde{\mathbf{A}}_{k,i\cdot}^\top - \widetilde{\mathbf{H}}_k^\top\mathbf{A}_{k,i\cdot}^\top\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \mathbf{\Lambda}_k^{-1}\mathbf{\Gamma}_k^\top\mathbf{V}_{ki}\mathbf{\Gamma}_k\mathbf{\Lambda}_k^{-1}\right),$$

with $\mathbf{\Gamma}_k$ *and* $\mathbf{\Lambda}_k$ *defined in Assumption 1(iii) and* $\mathbf{V}_{ki}$ *defined in Assumption 5;*

(ii) *if* $K \leq 2$ *and* $Tp_{-k} \gtrsim \min\left\{p^2, T^2p_{-j}^2, Tp_j^2, p_k^2p_j^2\right\}$*, for all* $j \neq k$*, or if* $K \geq 3$*, then it holds that* $\left\|\widetilde{\mathbf{A}}_{k,i\cdot}^\top - \widetilde{\mathbf{H}}_k^\top \mathbf{A}_{k,i\cdot}^\top\right\| = O_P\left(p^{-1} + \sum_{j=1,j\neq k}^K \left(T^{-1/2}p_j^{-1} + T^{-1}p_{-j}^{-1} + p_k^{-1}p_j^{-2}\right)\right)$.

Comparing part (i) of the theorem with part (i) of Theorem 2 (which holds under the same assumptions and restrictions on the relative rate of divergence of the dimensions $T, p_1, \dots, p_K$ as they pass to infinity), it emerges that the projected estimator $\widetilde{\mathbf{A}}_{k,i\cdot}$ has the same consistency rate of the initial estimator $\widehat{\mathbf{A}}_{k,i\cdot}$ and it is equally efficient. Indeed, due to its iterative nature, $\widetilde{\mathbf{A}}_{k,i\cdot}$ behaves similarly to a classical one-step estimator (see, e.g., Lehmann and Casella, 2006, Theorem 4.3). We would like to emphasize that this result holds only for vector- or matrix-valued time series - i.e. when $K \leq 2$. Indeed, it is easy to see that if $K > 2$, then the required constraint between the rates of divergence of T and p_k, $1 \leq k \leq K$, is never satisfied. For the case of higher-order tensors, i.e. $K \geq 3$, comparing the results in part (ii) of Theorem 2 with part (ii) of Theorem 4, it emerges that $\widetilde{\mathbf{A}}_{k,i\cdot}$ has a faster rate of convergence - this result can be understood by noting that, unlike the initial estimator $\widehat{\mathbf{A}}_{k,i\cdot}$, the projected estimator $\widetilde{\mathbf{A}}_{k,i\cdot}$ it is built by using all the information contained in all other modes. In this case however no asymptotic distribution can be derived under the present set of assumptions.

However, if we consider $\widehat{\mathbf{B}}_k^*$ as initial estimators (see Section 2.2.1) we can derive also asymptotic normality. Let

$$\widetilde{\omega}_k^* = \frac{1}{Tp_{-k}} + \frac{1}{p^2} + \frac{1}{T^2p_k^2}.$$

The following theorem summarizes the asymptotics (rates and limiting distribution) of $\widetilde{\mathbf{A}}_k^*$.

Theorem 5. *We assume that Assumptions 1–4 are satisfied. Then, as* $\min\{T, p_1, \dots, p_k\} \to \infty$*, it holds that, for all* $1 \leq k \leq K$ $p_k^{-1}\left\|\widetilde{\mathbf{A}}_k^* - \mathbf{A}_k\widetilde{\mathbf{H}}_k^*\right\|_F^2 = O_P\left(\widetilde{\omega}_k^*\right)$*, where* $\widetilde{\mathbf{H}}_k^*$ *is defined in (53). Further, if Assumption 5 is also satisfied, and if* $Tp_{-k} = o\left(\min\left\{p^2, T^2p_k^2\right\}\right)$*, then it holds that* $\sqrt{Tp_{-k}}\left(\widetilde{\mathbf{A}}_{k,i}^{*\top} - \widetilde{\mathbf{H}}_k^{*\top}\mathbf{A}_{k,i}^\top\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \mathbf{\Lambda}_k^{-1}\mathbf{\Gamma}_k\mathbf{V}_{k,i}\mathbf{\Gamma}_k^\top\mathbf{\Lambda}_k^{-1}\right)$*, where* $\mathbf{\Lambda}_k$*,* $\mathbf{\Gamma}_k$ *and* $\mathbf{V}_{k,i}$ *are defined in Theorem 4.*

In order to appreciate the differences in the convergence rates between $\widetilde{\mathbf{A}}_k$ and $\widetilde{\mathbf{A}}_k^*$, let us assume for simplicity that that $p_1 \asymp \cdots \asymp p_K \asymp \overline{p}$. When $K = 2$, then $\widetilde{\omega}_k = O_P(\overline{p}^{-4}) + (T\overline{p}^{-1})$ and $\widetilde{\omega}_k^* = \widetilde{\omega}_k$, since $\widehat{\mathbf{B}}_k^* = \widehat{\mathbf{B}}_k$ in this case. Similarly, when $K = 3$, $\widetilde{\omega}_k = O_P\left(\overline{p}^{-6} + (T\overline{p}^2)^{-1}\right)$ and the equality $\widetilde{\omega}_k^* = \widetilde{\omega}_k$ still holds. When $K > 3$, $\widetilde{\omega}_k = O_P\left(\overline{p}^{-6} + (T\overline{p}^2)^{-1}\right)$, and $\widetilde{\omega}_k^* = O_P\left((\overline{p}^{-2K}) + (T\overline{p}^{-2}) + (T\overline{p}^{(K-1)})^{-1}\right)$. This shows that when K is large (and, also, when T is large), the estimator $\widetilde{\mathbf{A}}_k^*$ will have faster converge rate than $\widetilde{\mathbf{A}}_k$. Note however that, in order to obtain $\widehat{\mathbf{B}}_k^*$, one needs to perform the eigen-decomposition of a very large scale, $p_{-k} \times p_{-k}$, covariance matrix, with computational complexity $O\left(p_{-k}^3\right)$. Conversely, $\widetilde{\mathbf{A}}_k$ requires calculating, as initial estimators, the loading matrices $\widehat{\mathbf{A}}_k$ separately; this requires performing the eigen-decompositions of $p_k \times p_k$ covariance matrices, and then calculating Kronecker product of $(K-1)$ small scale matrices of dimension $p_k \times r_k$, which has (lower) computational complexity $O\left(\sum_{j\neq k} p_j^3 + p_{-k} \times r_{-k}\right)$.

Finally, we notice that in the case of strong factors as considered in this paper, Chen and Lam (2024) have a (squared) consistency rate which is $\min(T^2, p^2, Tp_{-k})$ which is comparable to the rate $\min(T^2p_k^2, p^2, Tp_{-k})$ in Theorem 5.

3.4. Asymptotic properties of the factors and common component

As long as the loading matrices are determined, the factor $\mathcal{F}_t$ can be estimated easily by $\widetilde{\mathcal{F}}_t = p^{-1}\mathcal{F}_t \times_{k=1}^K \widetilde{\mathbf{A}}_k^\top$, where the projected estimators $\left\{\widetilde{\mathbf{A}}_k, 1 \leq k \leq K\right\}$ are computed using $\left\{\widehat{\mathbf{A}}_k, 1 \leq k \leq K\right\}$ defined in (19). The next theorem provides the consistency of the asymptotic properties of the estimated factors.

Theorem 6. *We assume that Assumptions 1–5 are satisfied. Then, for any given* $1 \leq i \leq p_k$ *and* $1 \leq k \leq K$*, as* $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$

(i) *if* $p = o\left(\min\left\{Tp_{-k}^2, Tp_kp_j^2, Tp_{-k}p_j^2, p_k^2p_j^4\right\}\right)$*, for all* $j \neq k$*, then it holds that*

$$\sqrt{p}\left(\mathrm{Vec}\left(\widetilde{\mathbf{F}}_{t,k}\right) - \left(\widetilde{\mathbf{H}}_{-k} \otimes \widetilde{\mathbf{H}}_k\right)^{-1}\mathrm{Vec}\left(\mathbf{F}_{t,k}\right)\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{H}_{-k} \otimes \mathbf{H}_k\right)^\top \mathbf{W}_{kt}\left(\mathbf{H}_{-k} \otimes \mathbf{H}_k\right)\right),$$

with $\mathbf{H}_k = P\text{-}\lim_{\min\{T,p_1,\dots,p_K\}\to\infty}\widetilde{\mathbf{H}}_k$ *and* $\mathbf{H}_{-k} = \otimes_{j\in[K]/\{k\}}\mathbf{H}_j$*, and* $\mathbf{W}_{kt}$ *defined in Assumption 5;*

(ii) *if* $p \gtrsim \min\left\{Tp_{-k}^2, Tp_kp_j^2, Tp_{-k}p_j^2, p_k^2p_j^4\right\}$*, for all* $j \neq k$*, then it holds that*

$$\left\|\widetilde{\mathcal{F}}_t - \mathcal{F}_t \times_1 \widetilde{\mathbf{H}}_1^{-1} \times_2 \cdots \times_K \widetilde{\mathbf{H}}_K^{-1}\right\| = O_P\left(\sum_{k=1}^K\left(\frac{1}{\sqrt{Tp_{-k}}} + \sum_{j\neq k}\left(\frac{1}{\sqrt{Tp_k}p_j} + \frac{1}{\sqrt{Tp_{-k}}p_j} + \frac{1}{p_kp_j^2}\right)\right)\right).$$

The theorem states the asymptotic distribution of the estimated factor tensor, which holds for any finite tensor order K and it can be compared with Theorem 1 in Bai (2003) which holds for $K = 1$, i.e. for the vector factor model. Two comments are worth making. First, notice that an explicit expression for the matrix $\mathbf{H}_k$ can be derived from the definition of $\widetilde{\mathbf{H}}_k$ given in (35) in Section 5. However, we prefer not to write those expressions explicitly to avoid introducing further notation. Second, a similar result can be derived for the factor tensor estimators $\widehat{\mathcal{F}}_t = p^{-1}\mathcal{F}_t \times_{k=1}^{K} \widehat{\mathbf{A}}_k^\top$ and $\widetilde{\mathcal{F}}_t^* = p^{-1}\mathcal{F}_t \times_{k=1}^{K} \widetilde{\mathbf{A}}_k^{*\top}$, which are estimated using the initial estimator of the loadings $\left\{\widehat{\mathbf{A}}_k, 1 \le k \le K\right\}$ or the estimator of the loadings $\left\{\widetilde{\mathbf{A}}_k^*, 1 \le k \le K\right\}$, respectively. These results are omitted for brevity.

Finally, the estimated common component is given by $\widetilde{\mathcal{S}}_t = \widetilde{\mathcal{F}}_t \times_{k=1}^{K} \widetilde{\mathbf{A}}_k$. The next theorem provides consistency of the estimated common component.

Theorem 7. *We assume that Assumptions 1–5 are satisfied. Then, for any given* $1 \le i_k \le p_k$, $1 \le k \le K$, *and* $1 \le t \le T$, *as* $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$

$$\left|\widetilde{\mathcal{S}}_{t,i_1\cdots i_K} - \mathcal{S}_{t,i_1\cdots i_K}\right| = O_P\left(\frac{1}{\sqrt{p}} + \sum_{k=1}^{K}\left(\frac{1}{\sqrt{Tp_k}} + \frac{1}{\sqrt{Tp_{-k}}} + \sum_{j\neq k}\frac{1}{p_k p_j^2}\right)\right).$$

The consistency rate depends of the rates for $\{\widetilde{\mathbf{A}}_k,\ 1 \le k \le K\}$ in Theorem 4 and for $\widetilde{\mathcal{F}}_t$ in Theorem 6. Two other estimators of the common component can be considered: $\widehat{\mathcal{S}}_t = \widehat{\mathcal{F}}_t \times_{k=1}^{K} \widehat{\mathbf{A}}_k$ and $\widetilde{\mathcal{S}}_t^* = \widetilde{\mathcal{F}}_t^* \times_{k=1}^{K} \widetilde{\mathbf{A}}_k^*$, and the proof of their consistency is straightforward and thus it is omitted. The proof of asymptotic normality for $\widetilde{\mathcal{S}}_t$, $\widehat{\mathcal{S}}_t$, and $\widetilde{\mathcal{S}}_t^*$ is also omitted for brevity. Indeed, it can be easily proved in a similar way as in Bai (2003). Notice that, while asymptotic normality of $\widehat{\mathcal{S}}_t$ and $\widetilde{\mathcal{S}}_t^*$ would hold for any finite tensor order K, for $\widetilde{\mathcal{S}}_t$ it would hold only if $K \le 2$, since only in that case we can prove asymptotic normality of the corresponding estimated loadings.

3.5. Asymptotic properties of the estimators for factor numbers

In this section, we establish the consistency of the proposed estimators of factor numbers. We first focus on the estimators based on the simple mode-wise sample covariance matrix $\widehat{\mathbf{M}}_k$. The following theorem shows that the $\hat{r}_k^{\text{IE-ER}}$ defined in (20) are consistent.

Theorem 8. *We assume that Assumptions 1–4 are satisfied, and that* $r_k \ge 1$, *for all* $1 \le k \le K$, *with* $r_{\max} \ge \max_{1\le k\le K} r_k$. *Then, for all* $1 \le k \le K$, *as* $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$, *it holds that* $\Pr\left(\hat{r}_k^{\text{IE-ER}} = r_k\right) \to 1$.

In Algorithm 3, we choose to calculate eigenvalue-ratio of the projection version $\widetilde{\mathbf{M}}_k$ rather than the initial version $\widehat{\mathbf{M}}_k$, as $\widetilde{\mathbf{M}}_k$ is more accurate. The consistency of the estimator outputted by the iterative algorithm is guaranteed by the following theorem.

Theorem 9. *We assume that Assumptions 1–4 are satisfied, and that* $r_k \ge 1$, *for all* $1 \le k \le K$, *with* $r_{\max} \ge \max_{1\le k\le K} r_k$. *If, for all* $j \ne k$, *it holds that* $r_j \le \hat{r}_j^{(s-1)} \le r_{\max}$, *for some* s *in Algorithm 3, then, for all* $1 \le k \le K$, *as* $\min\left\{T, p_1, \dots, p_K\right\} \to \infty$, *it holds that* $\Pr\left(\hat{r}_k^{(s)} = r_k\right) \to 1$ *thereby,* $\Pr\left(\hat{r}_k^{\text{PE-ER}} = r_k\right) \to 1$.

The theorem states that as long as we choose $r_k^{(0)}$ larger than real r_k at the beginning, the iterative method will yield consistent estimators of the number of factors. The algorithm is computationally very fast because it has a large probability to stop within few steps.

4. Concluding remarks

Tensor factor model is a powerful tool for dimension reduction of high-order tensors and is drawing growing attention in the last few years. In this paper, we propose a projection estimation method for Tucker-decomposition based Tensor Factor Model (TFM). We also provide the least squares interpretation of the iterative projection for TFM, which parallels to the least squares interpretation of traditional PCA for vector factor models (Fan et al., 2013) and of Projection Estimation for matrix factor models (He et al., 2024a). We establish the theoretical properties for the one-step iteration projection estimators, and faster convergence rates are achieved by the projection technique compared with the naive tensor PCA method.

5. Technical details

We report all proofs assuming, for simplicity and with no loss of generality, that $r_k = 1$, $1 \le k \le K$. We refer to the Supplementary Material for the additional lemmas.

Proof of Theorem 1. We state some preliminary facts. From From (A.1) in Lemma 2 and define $\widehat{\mathbf{\Lambda}}_k$ as the diagonal matrix containing the largest r_k eigenvalues of $\widehat{\mathbf{M}}_k$, viz. $\widehat{\mathbf{\Lambda}}_k = \text{diag}(\lambda_1(\widehat{\mathbf{M}}_k), \dots, \lambda_{r_k}(\widehat{\mathbf{M}}_k))$. Also recall that by definition $\widehat{\mathbf{A}}_k = \sqrt{p_k}\,\widehat{\mathbf{U}}_k$, where $\widehat{\mathbf{U}}_k$ has as columns the normalized eigenvectors corresponding to the r_k largest eigenvalues of $\widehat{\mathbf{M}}_k$ - see (19). Then, by definition $\widehat{\mathbf{A}}_k\widehat{\mathbf{\Lambda}}_k = \widehat{\mathbf{M}}_k\widehat{\mathbf{A}}_k$. Define now

$$\widehat{\mathbf{H}}_k = \frac{1}{Tp}\left(\sum_{t=1}^{T} \mathbf{F}_{k,t}^\top \widehat{\mathbf{B}}_k^\top \widehat{\mathbf{B}}_k \mathbf{F}_{k,t}^\top\right)\mathbf{A}_k^\top \widehat{\mathbf{A}}_k \widehat{\mathbf{\Lambda}}_k^{-1}. \tag{27}$$

Then it is easy to see that

$$\widehat{\mathbf{A}}_k - \mathbf{A}_k\widehat{\mathbf{H}}_k = (\mathcal{II} + \mathcal{III} + \mathcal{IV})\widehat{\mathbf{A}}_k\widehat{\mathbf{\Lambda}}_k^{-1}, \tag{28}$$

where $\mathcal{II}$, $\mathcal{III}$ and $\mathcal{IV}$ are defined in (A.1). Recall that: by Assumption 1(ii) $T^{-1}\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{F}_{k,t}^{\top}\xrightarrow{P}\boldsymbol{\Sigma}_k$, as $T\to\infty$; by Assumption 2(ii) and the definition of $\mathbf{B}_k$ it readily follows that $p_{-k}^{-1}\mathbf{B}_k^{\top}\mathbf{B}_k\to\mathbf{I}_{r_{-k}}$, as $p_{-k}\to\infty$; and by Assumption 2(ii) again, and the definition of $\widehat{\mathbf{A}}_k$, $\|\mathbf{A}_k\|_F^2\asymp\|\widehat{\mathbf{A}}_k\|_F^2\asymp p_k$, as $p_k\to\infty$. Also, by Lemma 3, as $\min\{T,p_1,\ldots,p_k\}\to\infty$, $\lambda_j(\widehat{\mathbf{M}}_k)$ converge to some positive constants for all $j\le r_k$; hence

$$\left\|\widehat{\boldsymbol{\Lambda}}_k\right\|_F=O_P(1),\quad\left\|\widehat{\boldsymbol{\Lambda}}_k^{-1}\right\|_F=O_P(1).\tag{29}$$

Putting all together, it follows that

$$\left\|\widehat{\mathbf{H}}_k\right\|_F=O_P(1).\tag{30}$$

Using now (28) and (29), it holds that

$$\begin{aligned}\frac{1}{p_k}\left\|\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right\|_F^2&=\frac{1}{p_k}\left\|(\mathcal{II}+\mathcal{III}+\mathcal{IV})\widehat{\mathbf{A}}_k\widehat{\boldsymbol{\Lambda}}_k^{-1}\right\|_F^2\lesssim\frac{1}{p_k}\left(\left\|\mathcal{II}\widehat{\mathbf{A}}_k\right\|_F^2+\left\|\mathcal{III}\widehat{\mathbf{A}}_k\right\|_F^2+\left\|\mathcal{IV}\widehat{\mathbf{A}}_k\right\|_F^2\right)\\&=O_P\left(\frac{1}{Tp_{-k}}+\frac{1}{p_k^2}\right)+o_p\left(\frac{1}{p_k}\left\|\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right\|_F^2\right).\end{aligned}\tag{31}$$

We note that the last term in (31) can be neglected: indeed, $p_k^{-1}\left\|\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right\|_F^2$ is bounded byt $O_P(1)$ because $\|\mathbf{A}_k\|_F^2\asymp p_k$, $\|\widehat{\mathbf{A}}_k\|_F^2\asymp p_k$, and $\|\widehat{\mathbf{H}}_k\|_F=O_P(1)$ as shown in (30). In order to complete the proof, it remains to show that $\widehat{\mathbf{H}}_k^{\top}\widehat{\mathbf{H}}_k\xrightarrow{P}\mathbf{I}_{r_k}$. By Cauchy-Schwartz inequality and (31), it holds that

$$\begin{aligned}\left\|\frac{1}{p_k}\mathbf{A}_k^{\top}\left(\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right)\right\|_F^2&\le\frac{\|\mathbf{A}_k\|_F^2}{p_k}\frac{\left\|\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right\|_F^2}{p_k}=o_P(1),\\\left\|\frac{1}{p_k}\widehat{\mathbf{A}}_k^{\top}\left(\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right)\right\|_F^2&\le\frac{\left\|\widehat{\mathbf{A}}_k\right\|_F^2}{p_k}\frac{\left\|\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right\|_F^2}{p_k}=o_P(1).\end{aligned}$$

Note now that, by construction, $p_k^{-1}\widehat{\mathbf{A}}_k^{\top}\widehat{\mathbf{A}}_k=\mathbf{I}_{r_k}$; also, by Assumption 2(ii), $p_k^{-1}\mathbf{A}_k^{\top}\mathbf{A}_k\to\mathbf{I}_{r_k}$. Hence, by (31) $\mathbf{I}_{r_k}=(p_k)^{-1}\widehat{\mathbf{A}}_k^{\top}\mathbf{A}_k\widehat{\mathbf{H}}_k+o_P(1)=\widehat{\mathbf{H}}_k^{\top}\widehat{\mathbf{H}}_k+o_P(1)$, which concludes the proof. □

Proof of Theorem 2. Based on equation (A.1) in Lemma 2, for all $1\le i\le p_k$, it holds that

$$\widehat{\mathbf{A}}_{k,i\cdot}^{\top}-\widehat{\mathbf{H}}_k^{\top}\mathbf{A}_{k,i\cdot}^{\top}=\frac{1}{Tp}\sum_{t=1}^{T}\widehat{\boldsymbol{\Lambda}}_k^{-1}\left(\widehat{\mathbf{A}}_k^{\top}\mathbf{E}_{k,t}\mathbf{B}_k\mathbf{F}_{k,t}^{\top}\mathbf{A}_{k,i\cdot}^{\top}+\widehat{\mathbf{A}}_k^{\top}\mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^{\top}e_{k,t,i\cdot}^{\top}+\widehat{\mathbf{A}}_k^{\top}\mathbf{E}_{k,t}e_{k,t,i\cdot}^{\top}\right).$$

Using now: Eq. (29) and Assumptions 2(i) and 4(i), it follows that

$$\left\|\frac{1}{Tp}\sum_{t=1}^{T}\widehat{\boldsymbol{\Lambda}}_k^{-1}\widehat{\mathbf{A}}_k^{\top}\mathbf{E}_{k,t}\mathbf{B}_k\mathbf{F}_{k,t}^{\top}\mathbf{A}_{k,i\cdot}^{\top}\right\|_F\lesssim\frac{1}{\sqrt{Tp}}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\widehat{\mathbf{A}}_k}{p_k^{1/2}}\right)^{\top}\mathbf{E}_{k,t}\left(\frac{\mathbf{B}_k}{p_{-k}^{1/2}}\right)\mathbf{F}_{k,t}^{\top}\right\|_F=O_P\left(\frac{1}{\sqrt{Tp}}\right).\tag{32}$$

By the same token, and using Assumptions 2(i) and 4(ii), it holds that

$$\left\|\frac{1}{Tp}\sum_{t=1}^{T}\widehat{\boldsymbol{\Lambda}}_k^{-1}\widehat{\mathbf{A}}_k^{\top}\mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^{\top}e_{k,t,i\cdot}^{\top}\right\|_F\le\left\|\widehat{\boldsymbol{\Lambda}}_k^{-1}\right\|_F\left\|\frac{1}{p_k}\widehat{\mathbf{A}}_k^{\top}\mathbf{A}_k\right\|_F\left|\frac{1}{Tp_{-k}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^{\top}e_{k,t,i\cdot}^{\top}\right|=O_P\left(\frac{1}{\sqrt{Tp_{-k}}}\right).\tag{33}$$

Moreover, by the same arguments used above, and using Lemma 2(iii), we have

$$\begin{aligned}\left\|\frac{1}{Tp}\sum_{t=1}^{T}\widehat{\boldsymbol{\Lambda}}_k^{-1}\widehat{\mathbf{A}}_k^{\top}\mathbf{E}_{k,t}e_{k,t,i\cdot}^{\top}\right\|_F&\lesssim\left\|\frac{1}{Tp}\sum_{t=1}^{T}\left(\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right)^{\top}\mathbf{E}_{k,t}e_{k,t,i\cdot}^{\top}\right\|_F+\left\|\frac{1}{Tp}\sum_{t=1}^{T}\widehat{\mathbf{H}}_k^{\top}\mathbf{A}_k^{\top}\mathbf{E}_{k,t}e_{k,t,i\cdot}^{\top}\right\|_F\\&\lesssim\left\|\widehat{\mathbf{A}}_k-\mathbf{A}_k\widehat{\mathbf{H}}_k\right\|_F\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{k,t}e_{k,t,i\cdot}^{\top}\right\|_F+\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{A}_k^{\top}\mathbf{E}_{k,t}e_{k,t,i\cdot}^{\top}\right\|_F\\&=O_P\left(\frac{1}{p_k}\right)+O_P\left(\frac{1}{\sqrt{Tp}}\right).\end{aligned}\tag{34}$$

Therefore, as $\min\{T,p_1,\ldots,p_K\}\to\infty$ under the restriction $Tp_{-k}=o\left(p_k^2\right)$, it follows that

$$\sqrt{Tp_{-k}}\left(\widehat{\mathbf{A}}_{k,i\cdot}^{\top}-\widehat{\mathbf{H}}_k^{\top}\mathbf{A}_{k,i\cdot}^{\top}\right)=\widehat{\boldsymbol{\Lambda}}_k^{-1}\frac{\widehat{\mathbf{A}}_k^{\top}\mathbf{A}_k}{p_k}\frac{1}{\sqrt{Tp_{-k}}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^{\top}e_{k,t,i\cdot}^{\top}+o_P(1).$$

By Lemma 3, it holds that as $\min\{T,p_1,\ldots,p_K\}\to\infty$, $\widehat{\boldsymbol{\Lambda}}_k\xrightarrow{P}\boldsymbol{\Lambda}_k$; further, from the proof of Theorem 1, it holds that $\widehat{\mathbf{H}}_k=p_k^{-1}\mathbf{A}_k^{\top}\widehat{\mathbf{A}}_k+o_P(1)$; using Assumption 1(ii), the definition of $\widehat{\mathbf{H}}_k$ in (27), and recalling that $\boldsymbol{\Sigma}_k$ has spectral decomposition $\boldsymbol{\Sigma}_k=\boldsymbol{\Gamma}_k\boldsymbol{\Lambda}_k\boldsymbol{\Gamma}_k^{\top}$, we have $\widehat{\mathbf{H}}_k=\boldsymbol{\Gamma}_k\boldsymbol{\Lambda}_k\boldsymbol{\Gamma}_k^{\top}\widehat{\mathbf{H}}_k\boldsymbol{\Lambda}_k^{-1}+o_P(1)$. By rearranging the last equation and recalling that $\boldsymbol{\Gamma}_k^{\top}\boldsymbol{\Gamma}_k=\mathbf{I}_{r_k}$, $\boldsymbol{\Gamma}_k^{\top}\widehat{\mathbf{H}}_k\boldsymbol{\Lambda}_k=\boldsymbol{\Lambda}_k\boldsymbol{\Gamma}_k^{\top}\widehat{\mathbf{H}}_k+o_P(1)$.

We note that $\boldsymbol{\Lambda}_k$ is diagonal with distinct entries; hence, $\boldsymbol{\Gamma}_k^\top \widehat{\mathbf{H}}_k$ must be asymptotically diagonal. Furthermore, since $\widehat{\mathbf{H}}_k^\top \widehat{\mathbf{H}}_k = \left(\boldsymbol{\Gamma}_k^\top \widehat{\mathbf{H}}_k\right)^\top \boldsymbol{\Gamma}_k^\top \widehat{\mathbf{H}}_k = \mathbf{I}_{r_k} + o_P(1)$, the diagonal entries of $\boldsymbol{\Gamma}_k^\top \widehat{\mathbf{H}}_k$ must be asymptotically 1 or -1, and we can always choose the column signs of $\widehat{\mathbf{A}}_k$ in such a way that $\boldsymbol{\Gamma}_k^\top \widehat{\mathbf{H}}_k = \mathbf{I}_{r_k} + o_P(1)$. Therefore, it finally follows that $p_k{}^{-1}\widehat{\mathbf{A}}_k^\top \mathbf{A}_k = \widehat{\mathbf{H}}_k^\top + o_P(1) = \boldsymbol{\Gamma}_k^\top + o_P(1)$. Consequently, by Slutsky's theorem and Assumption 5(i), as $\min\{T, p_1, \dots, p_K\} \to \infty$ under the restriction $Tp_{-k} = o\left(p_k^2\right)$, it holds that $\sqrt{Tp_{-k}}\left(\widehat{\mathbf{A}}_{k,i\cdot}^\top - \widehat{\mathbf{H}}_k^\top \mathbf{A}_{k,i\cdot}^\top\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Gamma}_k^\top \mathbf{V}_{ki}\boldsymbol{\Gamma}_k \boldsymbol{\Lambda}_k^{-1}\right)$, which concludes the proof of part (i) of the theorem. Whenever $Tp_{-k} \gtrsim p_k^2$, it immediately follows from (32)–(34) that $\|\widehat{\mathbf{A}}_{k,i\cdot}^\top - \widehat{\mathbf{H}}_k^\top \mathbf{A}_{k,i\cdot}^\top\| = O_P\left(p_k^{-1}\right)$, proving part (ii). □

Proof of Theorem 3. The proof follows a similar logic to the proof of Theorem 1. Given (A.2) in Lemma 4, define the diagonal matrix containing the first r_k eigenvalues of $\widetilde{\mathbf{M}}_k$ as $\widetilde{\boldsymbol{\Lambda}}_k = diag\left\{\lambda_1\left(\widetilde{\mathbf{M}}_k\right), \dots, \lambda_{r_k}\left(\widetilde{\mathbf{M}}_k\right)\right\}$. By definition, $\widetilde{\mathbf{A}}_k := \sqrt{p_k}\,\widetilde{\mathbf{U}}_k$, where the columns of $\widetilde{\mathbf{U}}_k$ are orthonormal basis of the eigenvectors corresponding to the r_k largest eigenvalues of $\widetilde{\mathbf{M}}_k$. Hence $\widetilde{\mathbf{A}}_k \widetilde{\boldsymbol{\Lambda}}_k = \widetilde{\mathbf{M}}_k \widetilde{\mathbf{A}}_k$. Define now

$$\widetilde{\mathbf{H}}_k := \frac{1}{Tp_k p_{-k}^2}\left(\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{B}_k^\top \widehat{\mathbf{B}}_k \widehat{\mathbf{B}}_k^\top \mathbf{B}_k \mathbf{F}_{k,t}^\top\right)\mathbf{A}_k^\top \widetilde{\mathbf{A}}_k \widetilde{\boldsymbol{\Lambda}}_k^{-1}. \tag{35}$$

It is easy to see, in the light of (A.2) that

$$\widetilde{\mathbf{A}}_k - \mathbf{A}_k \widetilde{\mathbf{H}}_k = (\mathcal{VI} + \mathcal{VII} + \mathcal{VIII})\widetilde{\mathbf{A}}_k \widetilde{\boldsymbol{\Lambda}}_k^{-1}. \tag{36}$$

Note now that, using Assumption 2(ii) and the definitions of $\widetilde{\mathbf{A}}_k$ in (13) and $\widehat{\mathbf{B}}_k$ following from (19) respectively, itthat $\|\mathbf{A}_k\|_F^2 \asymp \left\|\widetilde{\mathbf{A}}_k\right\|_F^2 \asymp p_k$ and $\|\mathbf{B}_k\|_F^2 \asymp \left\|\widehat{\mathbf{B}}_k\right\|_F^2 \asymp p_{-k}$ as $p_{-k}, p_k \to \infty$. Assumption 1(ii) states that $T^{-1}\sum_{t=1}^T \mathbf{F}_{k,t}\mathbf{F}_{k,t}^\top \xrightarrow{P} \boldsymbol{\Sigma}_k$, as $T \to \infty$; further, by Lemma 3, the diagonal entries of $\widetilde{\boldsymbol{\Lambda}}_k$ converge to some positive constants as $\min\{T, p_1, \dots, p_k\} \to \infty$. Hence

$$\left\|\widetilde{\boldsymbol{\Lambda}}_k\right\|_F = O_P(1), \quad \left\|\widetilde{\boldsymbol{\Lambda}}_k^{-1}\right\|_F = O_P(1). \tag{37}$$

Therefore, we have $\|\widetilde{\mathbf{H}}_k\|_F = O_P(1)$. Combining (36) and (37) and using Lemma 3, it holds that

$$\begin{aligned}\frac{1}{p_k}\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2 &= \frac{1}{p_k}\left\|(\mathcal{VI} + \mathcal{VII} + \mathcal{VIII})\widetilde{\mathbf{A}}_k \widetilde{\boldsymbol{\Lambda}}_k^{-1}\right\|_F^2 \le \frac{1}{p_k}\left(\left\|\mathcal{VI}\widetilde{\mathbf{A}}_k\right\|_F^2 + \left\|\mathcal{VII}\widetilde{\mathbf{A}}_k\right\|_F^2 + \left\|\mathcal{VIII}\widetilde{\mathbf{A}}_k\right\|_F^2\right)\\ &= O_P\left(\frac{1}{Tp_{-k}} + \frac{1}{p^2} + w_{-k}^2\left(\frac{1}{p_k^2} + \frac{1}{Tp_k}\right) + m_{-k}\right) + o_P\left(\frac{1}{p_k}\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2\right).\end{aligned} \tag{38}$$

By the same logic as in the proof of Theorem 1, the last term in (38) can be neglected. In order to complete the proof, it remains to show that $\widetilde{\mathbf{H}}_k^\top \widetilde{\mathbf{H}}_k \xrightarrow{P} \mathbf{I}_{r_k}$. Using, as in the proof of Theorem 1, the Cauchy-Schwartz inequality and (38), it holds that

$$\begin{aligned}\left\|\frac{1}{p_k}\mathbf{A}_k^\top\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right\|_F^2 &\le \frac{\|\mathbf{A}_k\|_F^2}{p_k}\frac{\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2}{p_k} = o_P(1),\\ \left\|\frac{1}{p_k}\widetilde{\mathbf{A}}_k^\top\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)\right\|_F^2 &\le \frac{\left\|\widetilde{\mathbf{A}}_k\right\|_F^2}{p_k}\frac{\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2}{p_k} = o_P(1),\end{aligned}$$

since, as noted above, $\|\mathbf{A}_k\|_F^2 \asymp \left\|\widetilde{\mathbf{A}}_k\right\|_F^2 \asymp p_k$. Recalling that $p_k^{-1}\widetilde{\mathbf{A}}_k^\top \widetilde{\mathbf{A}}_k = \mathbf{I}_{r_k}$ and $p_k^{-1}\mathbf{A}_k^\top \mathbf{A}_k \to \mathbf{I}_{r_k}$, then from (38) $\mathbf{I}_{r_k} = p_k{}^{-1}\widetilde{\mathbf{A}}_k^\top \mathbf{A}_k \widetilde{\mathbf{H}}_k + o_P(1) = \widetilde{\mathbf{H}}_k^\top \widetilde{\mathbf{H}}_k + o_P(1)$, which concludes the proof. □

Proof of Lemma 1. Consider part (i) of the lemma, and note that

$$\begin{aligned}\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k} &= \widehat{\mathbf{A}}_K \otimes \cdots \otimes \widehat{\mathbf{A}}_{k+1} \otimes \widehat{\mathbf{A}}_{k-1} \otimes \cdots \otimes \widehat{\mathbf{A}}_1 - \left(\otimes_{j=1,j\neq k}^K \mathbf{A}_j\right)\left(\otimes_{j=1,j\neq k}^K \widehat{\mathbf{H}}_j\right)\\ &= \left(\widehat{\mathbf{A}}_K - \mathbf{A}_K\widehat{\mathbf{H}}_K + \mathbf{A}_K\widehat{\mathbf{H}}_K\right) \otimes \cdots \otimes \left(\widehat{\mathbf{A}}_{k+1} - \mathbf{A}_{k+1}\widehat{\mathbf{H}}_{k+1} + \mathbf{A}_{k+1}\widehat{\mathbf{H}}_{k+1}\right)\\ &\quad \otimes \left(\widehat{\mathbf{A}}_{k-1} - \mathbf{A}_{k-1}\widehat{\mathbf{H}}_{k-1} + \mathbf{A}_{k-1}\widehat{\mathbf{H}}_{k-1}\right) \otimes \cdots \otimes \left(\widehat{\mathbf{A}}_1 - \mathbf{A}_1\widehat{\mathbf{H}}_1 + \mathbf{A}_1\widehat{\mathbf{H}}_1\right) - \left(\otimes_{j=1,j\neq k}^K\left(\mathbf{A}_j\widehat{\mathbf{H}}_j\right)\right)\\ &= \left(\widehat{\mathbf{A}}_K - \mathbf{A}_K\widehat{\mathbf{H}}_K\right) \otimes \left(\otimes_{j\neq k,K}\widehat{\mathbf{A}}_j\right) + \mathbf{A}_K\widehat{\mathbf{H}}_K \otimes \left(\widehat{\mathbf{A}}_{K-1} - \mathbf{A}_{K-1}\widehat{\mathbf{H}}_{K-1}\right) \otimes \left(\otimes_{j\neq k,K-1,K}\widehat{\mathbf{A}}_j\right)\\ &\quad + \cdots + \otimes_{j\neq 1,k}\left(\mathbf{A}_j\widehat{\mathbf{H}}_j\right) \otimes \left(\widehat{\mathbf{A}}_1 - \mathbf{A}_1\widehat{\mathbf{H}}_1\right).\end{aligned}$$

Hence, using Theorem 1, Assumption 2(ii), and the definition of $\widehat{\mathbf{A}}_\ell$ in (19), it holds that

$$\begin{aligned}\frac{1}{p_{-k}}\left\|\widehat{\mathbf{B}}_k - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right\|_F^2 &\lesssim \frac{1}{p_{-k}}\sum_{j=1,j\neq k}^K\left\|\left(\otimes_{\ell=k+1,\ell\neq k}^K \mathbf{A}_\ell\widehat{\mathbf{H}}_\ell\right) \otimes \left(\widehat{\mathbf{A}}_j - \mathbf{A}_j\widehat{\mathbf{H}}_j\right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1}\widehat{\mathbf{A}}_\ell\right)\right\|_F^2\\ &\lesssim \frac{1}{p_{-k}}\sum_{j=1,j\neq k}^K\left\|\left(\otimes_{\ell=k+1,\ell\neq k}^K \mathbf{A}_\ell\right) \otimes \left(\widehat{\mathbf{A}}_j - \mathbf{A}_j\widehat{\mathbf{H}}_j\right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1}\mathbf{A}_\ell\right)\right\|_F^2\end{aligned}$$

$$\lesssim \frac{1}{p_{-k}} \sum_{j=1, j\neq k}^{K} \left(\prod_{\ell=1,\ell\neq j,k}^{K} \|\mathbf{A}_\ell\|_F^2 \right) \left\| \widehat{\mathbf{A}}_j - \mathbf{A}_j \widehat{\mathbf{H}}_j \right\|_F^2$$

$$\lesssim \sum_{j=1,j\neq k}^{K} \frac{1}{p_j} \left\| \widehat{\mathbf{A}}_j - \mathbf{A}_j \widehat{\mathbf{H}}_j \right\|_F^2 = O_P\left(\sum_{j=1,j\neq k}^{K} w_j \right)$$

By letting $w_{-k} = \sum_{j=1,j\neq k}^{K} w_j$, it follows immediately that the sufficient condition in (25) is satisfied. We now turn to part (b) of the lemma. Recall that by equation (A.1), it holds that $\widehat{\mathbf{M}}_k = \mathcal{I} + \mathcal{II} + \mathcal{III} + \mathcal{IV}$; further, from (31), we have $\left(\widehat{\mathbf{A}}_k - \mathbf{A}_k \widehat{\mathbf{H}}_k\right) = (\mathcal{II} + \mathcal{III} + \mathcal{IV}) \widehat{\mathbf{A}}_k \widehat{\mathbf{\Lambda}}_k^{-1}$. Thus, using (29) and (39), we obtain

$$\frac{1}{p_k} \left\| \frac{1}{T p_{-k}} \sum_{s=1}^{T} \mathbf{E}_{k,s} \left(\widehat{\mathbf{B}}_k - \mathbf{B}_k \widehat{\mathbf{H}}_{-k} \right) \mathbf{F}_{k,s}^\top \right\|_F^2 \tag{39}$$

$$\lesssim \frac{1}{p_k} \sum_{j=1,j\neq k}^{K} \left\| \frac{1}{T} \sum_{s=1}^{T} \mathbf{E}_{k,s} \left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K} \frac{1}{p_\ell} \mathbf{A}_\ell \widehat{\mathbf{H}}_\ell \right) \otimes \frac{1}{p_j} \left(\widehat{\mathbf{A}}_j - \mathbf{A}_j \widehat{\mathbf{H}}_j \right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1} \frac{1}{p_\ell} \widehat{\mathbf{A}}_\ell \right) \right) \mathbf{F}_{k,s}^\top \right\|_F^2$$

$$\lesssim \frac{1}{p_k} \sum_{j=1,j\neq k}^{K} \left\| \frac{1}{T} \sum_{s=1}^{T} \mathbf{E}_{k,s} \left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K} \frac{1}{p_\ell} \mathbf{A}_\ell \right) \otimes \frac{1}{p_j} \left(\widehat{\mathbf{A}}_j - \mathbf{A}_j \widehat{\mathbf{H}}_j \right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1} \frac{1}{p_\ell} \mathbf{A}_\ell \right) \right) \mathbf{F}_{k,s}^\top \right\|_F^2$$

$$\lesssim \frac{1}{p_k} \sum_{j=1,j\neq k}^{K} \left\| \frac{1}{T p_{-k}} \sum_{s=1}^{T} \mathbf{E}_{k,s} \left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K} \mathbf{A}_\ell \right) \otimes \left((\mathcal{II} + \mathcal{III} + \mathcal{IV}) \widehat{\mathbf{A}}_j \widehat{\mathbf{\Lambda}}_j^{-1} \right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1} \mathbf{A}_\ell \right) \right) \mathbf{F}_{k,s}^\top \right\|_F^2$$

$$\lesssim \frac{1}{p_k} \sum_{j=1,j\neq k}^{K} \left\| \frac{1}{T p_{-k}} \sum_{s=1}^{T} \mathbf{E}_{k,s} \left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K} \mathbf{A}_\ell \right) \otimes \left(\mathcal{II} \widehat{\mathbf{A}}_j \right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1} \mathbf{A}_\ell \right) \right) \mathbf{F}_{k,s}^\top \right\|_F^2$$

$$+ \frac{1}{p_k} \sum_{j=1,j\neq k}^{K} \left\| \frac{1}{T p_{-k}} \sum_{s=1}^{T} \mathbf{E}_{k,s} \left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K} \mathbf{A}_\ell \right) \otimes \left(\mathcal{III} \widehat{\mathbf{A}}_j \right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1} \mathbf{A}_\ell \right) \right) \mathbf{F}_{k,s}^\top \right\|_F^2$$

$$+ \frac{1}{p_k} \sum_{j=1,j\neq k}^{K} \left\| \frac{1}{T p_{-k}} \sum_{s=1}^{T} \mathbf{E}_{k,s} \left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K} \mathbf{A}_\ell \right) \otimes \left(\mathcal{IV} \widehat{\mathbf{A}}_j \right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1} \mathbf{A}_\ell \right) \right) \mathbf{F}_{k,s}^\top \right\|_F^2 = \mathcal{XII} + \mathcal{XIII} + \mathcal{XIV}.$$

Consider $\mathcal{XII}$; by the definition of $\mathcal{II}$ in (A.1), and using Lemmas 2(ii) and 4, and Theorem 1, it follows that

$$\mathcal{XII} \lesssim \frac{1}{T^2 p_k p_{-k}^2} \sum_{j=1,j\neq k}^{K} \left\| \frac{1}{Tp} \sum_{t=1}^{T} \mathbf{F}_{j,t} \mathbf{B}_j^\top \mathbf{E}_{j,t}^\top \widehat{\mathbf{A}}_j \right\|_F^2 \left\| \sum_{s=1}^{T} \mathbf{E}_{k,s} \mathbf{B}_k \mathbf{F}_{k,s}^\top \right\|_F^2 \tag{40}$$

$$\lesssim \frac{1}{T^4 p_k^3 p_{-k}^4} \sum_{j=1,j\neq k}^{K} \left(\left\| \frac{1}{Tp} \sum_{t=1}^{T} \mathbf{F}_{j,t} \mathbf{B}_j^\top \mathbf{E}_{j,t}^\top \mathbf{A}_j \right\|_F^2 + \left\| \frac{1}{Tp} \sum_{t=1}^{T} \mathbf{F}_{j,t} \mathbf{B}_j^\top \mathbf{E}_{j,t}^\top \right\|_F^2 \left\| \widehat{\mathbf{A}}_j - \mathbf{A}_j \widehat{\mathbf{H}}_j \right\|_F^2 \right) \left\| \sum_{s=1}^{T} \mathbf{E}_{k,s} \mathbf{B}_k \mathbf{F}_{k,s}^\top \right\|_F^2$$

$$= O_P\left(\frac{1}{T^2 p_k p_{-k}^2} \sum_{j=1,j\neq k}^{K} \left(1 + \frac{p_j}{T p_{-j}} + \frac{1}{p_j} \right) \right).$$

Turning to $\mathcal{XIII}$, recall that we have assumed $r_k = 1$, and therefore $\mathbf{F}_{k,t}$ is a scalar, $\mathbf{A}_k$ is a p_k -dimensional vector with entries A_{ki}, $\mathbf{B}_k$ is a p_{-k} -dimensional vector with entries B_{ki}, and $\zeta_{i_1,\dots,i_K} = \text{Vec}\left(T^{-1/2} \sum_{t=1}^{T} \mathcal{F}_t e_{t,i_1,\dots,i_K} \right)$, is also a scalar. Let now $h\left(m_1,\dots,m_N\right) = m_1 + \left(m_2 - 1\right) I_1 + \left(m_3 - 1\right) I_1 I_2 + \cdots + \left(m_N - 1\right) I_1 I_2 \dots I_{N-1}$, where $\left(m_1,\dots,m_N\right) \in \mathbb{Z}^{I_1 \times \cdots \times I_N}$, and define further $n_{k,i_1,\dots,i_{k-1},i_{k+1},\dots,i_K} = h\left(i_1,\dots,i_{k-1},i_{k+1},\dots,i_K\right)$, with $1 \le i_j \le p_j$ - then it is easy to see that $n_{k,i_1,\dots,i_{k-1},i_{k+1},\dots,i_K}$ runs from 1, when $i_j = 1$ for all j, to p_{-k} when $i_j = p_j$ for all j. Hence, using Assumption 4(ii), it follows that

$$\mathbb{E}\left(\left\| \sum_{s=1}^{T} \mathbf{E}_{k,s} \left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K} \mathbf{A}_\ell \right) \otimes \left(\frac{1}{Tp} \sum_{t=1}^{T} \mathbf{E}_{j,t} \mathbf{B}_j \mathbf{F}_{j,t}^\top \right) \otimes \left(\otimes_{\ell=1,\ell\neq k}^{j-1} \mathbf{A}_\ell \right) \right) \mathbf{F}_{k,s}^\top \right\|_F^2 \right)$$

$$= \frac{1}{p^2} \sum_{i_k=1}^{p_k} \mathbb{E}\left(\frac{1}{T} \sum_{t,s=1}^{T} \sum_{i_1=1}^{p_1} \cdots \sum_{i_{k-1}=1}^{p_{k-1}} \sum_{i_{k+1}=1}^{p_{k+1}} \cdots \sum_{i_K=1}^{p_K} e_{s,i_1,\dots,i_K} \left(\prod_{\ell=1,\ell\neq j,k}^{K} A_{\ell,i_\ell} \right) \left(\mathbf{e}_{j,t,i_j \cdot} \mathbf{B}_j \right) \mathbf{F}_{k,s} \mathbf{F}_{k,t} \right)^2$$

$$= \frac{1}{p^2} \sum_{i_k=1}^{p_k} \mathbb{E}\left(\frac{1}{T} \sum_{t,s=1}^{T} \sum_{i_1=1}^{p_1} \cdots \sum_{i_{k-1}=1}^{p_{k-1}} \sum_{i_{k+1}=1}^{p_{k+1}} \cdots \sum_{i_K=1}^{p_K} \sum_{h_1=1}^{p_1} \cdots \sum_{h_{j-1}=1}^{p_{j-1}} \sum_{h_{j+1}=1}^{p_{j+1}} \cdots \sum_{h_K=1}^{p_K} e_{s,i_1,\dots,i_K} \right.$$

$$\left. \cdot \left(\prod_{\ell=1,\ell\neq j,k}^{K} A_{\ell,i_\ell} \right) e_{s,h_1,\dots,i_j,\dots,h_K} B_{j,n_j,h_1,\dots,h_{j-1},h_{j+1},\dots,h_K} \mathbf{F}_{k,s} \mathbf{F}_{k,t} \right)^2$$

$$
\begin{aligned}
&\lesssim \frac{1}{p^2}\sum_{i_k=1}^{p_k}\mathbb{E}\left(\sum_{i_1=1}^{p_1}\cdots\sum_{i_{k-1}=1}^{p_{k-1}}\sum_{i_{k+1}=1}^{p_{k+1}}\cdots\sum_{i_K=1}^{p_K}\sum_{h_1=1}^{p_1}\cdots\sum_{h_{j-1}=1}^{p_{j-1}}\sum_{h_{j+1}=1}^{p_{j+1}}\cdots\sum_{h_K=1}^{p_K}\zeta_{i_1,\dots,i_K}\zeta_{h_1,\dots,i_j,\dots,h_K}-\mathbb{E}\left(\zeta_{i_1,\dots,i_K}\zeta_{h_1,\dots,i_j,\dots,h_K}\right)\right)^2\\
&\quad+\frac{1}{p^2}\sum_{i_k=1}^{p_k}\mathbb{E}\left(\sum_{i_1=1}^{p_1}\cdots\sum_{i_{k-1}=1}^{p_{k-1}}\sum_{i_{k+1}=1}^{p_{k+1}}\cdots\sum_{i_K=1}^{p_K}\sum_{h_1=1}^{p_1}\cdots\sum_{h_{j-1}=1}^{p_{j-1}}\sum_{h_{j+1}=1}^{p_{j+1}}\cdots\sum_{h_K=1}^{p_K}\mathbb{E}\left(\zeta_{i_1,\dots,i_K}\zeta_{h_1,\dots,i_j,\dots,h_K}\right)\right)^2\\
&=\frac{1}{p^2}\sum_{i_k=1}^{p_k}\sum_{i_1,i_1'=1}^{p_1}\cdots\sum_{\substack{i_{k-1}=1\\ i_{k-1}'=1}}^{p_{k-1}}\sum_{\substack{i_{k+1}=1\\ i_{k+1}'=1}}^{p_{k+1}}\cdots\sum_{\substack{i_K=1\\ i_K'=1}}^{p_K}\sum_{\substack{h_1=1\\ h_1'=1}}^{p_1}\cdots\sum_{\substack{h_{j-1}=1\\ h_{j-1}'=1}}^{p_{j-1}}\sum_{\substack{h_{j+1}=1\\ h_{j+1}'=1}}^{p_{j+1}}\cdots\sum_{\substack{h_K=1\\ h_K'=1}}^{p_K}\mathrm{Cov}\left(\zeta_{i_1,\dots,i_K}\zeta_{h_1,\dots,i_j,\dots,h_K},\zeta_{i_1',\dots,i_K'}\zeta_{h_1',\dots,i_j,\dots,h_K'}\right)\\
&\quad+\frac{1}{p^2}\sum_{i_k=1}^{p_k}\mathbb{E}\left(\sum_{i_1=1}^{p_1}\cdots\sum_{i_{k-1}=1}^{p_{k-1}}\sum_{i_{k+1}=1}^{p_{k+1}}\cdots\sum_{i_K=1}^{p_K}\sum_{h_1=1}^{p_1}\cdots\sum_{h_{j-1}=1}^{p_{j-1}}\sum_{h_{j+1}=1}^{p_{j+1}}\cdots\sum_{h_K=1}^{p_K}\mathbb{E}\left(\zeta_{i_1,\dots,i_K}\zeta_{h_1,\dots,i_j,\dots,h_K}\right)\right)^2\\
&\lesssim\frac{1}{p_j}+\frac{1}{p_k}.
\end{aligned}
$$

Hence, from the definition of $\mathcal{III}$ in (A.1), by Theorem 1 it holds that

$$
\begin{aligned}
\mathcal{XIII}&\lesssim\frac{1}{p_k}\sum_{j=1,j\neq k}^{K}\left\|\frac{1}{Tp_{-k}}\sum_{s=1}^{T}\mathbf{E}_{k,s}\left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K}\mathbf{A}_\ell\right)\otimes\left(\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{B}_j\mathbf{F}_{j,t}^\top\right)\otimes\left(\otimes_{\ell=1,\ell\neq k}^{j-1}\mathbf{A}_\ell\right)\right)\mathbf{F}_{k,s}^\top\right\|_F^2\times\left\|\mathbf{A}_j^\top\widehat{\mathbf{A}}_j\right\|_F^2\\
&=O_P\left(\frac{1}{T^2p_kp_{-k}^2}\sum_{j=1,j\neq k}^{K}\left(p_j+\frac{p_j^2}{p_k}\right)\right).
\end{aligned}
\tag{41}
$$

We finally consider $\mathcal{XIV}$; recalling Lemma 2(iii), it holds that

$$
\begin{aligned}
\mathcal{XIV}&\lesssim\frac{1}{p_k}\sum_{\substack{j=1\\ j\neq k}}^{K}\left\|\frac{1}{Tp_{-k}}\sum_{s=1}^{T}\mathbf{E}_{k,s}\left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K}\mathbf{A}_\ell\right)\otimes\left(\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\widehat{\mathbf{A}}_j\right)\otimes\left(\otimes_{\ell=1,\ell\neq k}^{j-1}\mathbf{A}_\ell\right)\right)\mathbf{F}_{k,s}^\top\right\|_F^2\\
&\lesssim\frac{1}{p_k}\left\|\frac{1}{Tp_{-k}}\sum_{s=1}^{T}\mathbf{E}_{k,s}\mathbf{F}_{k,s}\right\|_F^2\left(\sum_{j=1,j\neq k}^{K}\left[\left(\prod_{\ell=1,\ell\neq j,k}^{K}\|\mathbf{A}_\ell\|_F^2\right)\left(\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\mathbf{A}_j\right\|_F^2+\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\right\|_F^2\left\|\widehat{\mathbf{A}}_j-\mathbf{A}_j\widehat{\mathbf{H}}_j\right\|_F^2\right)\right]\right)\\
&=O_P\left(\frac{1}{T^2p^2}\sum_{j=1,j\neq k}^{K}\left(p_j^2+Tp_{-j}^2\right)\right).
\end{aligned}
\tag{42}
$$

Indeed, to offer more detail on (42), we note that, using Assumption 4 and the same logic as above

$$
\mathbb{E}\left\|\frac{1}{Tp_{-k}}\sum_{s=1}^{T}\mathbf{E}_{k,s}\mathbf{F}_{k,s}\right\|_F^2=\left(\frac{1}{Tp_{-k}}\right)^2\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}}\mathbb{E}\left(\sum_{s=1}^{T}e_{k,t,ij}\mathbf{F}_{k,s}\right)^2\leq c_0\frac{p_k}{Tp_{-k}},
$$

whence

$$
\left\|\frac{1}{Tp_{-k}}\sum_{s=1}^{T}\mathbf{E}_{k,s}\mathbf{F}_{k,s}\right\|_F^2=O_P\left(\frac{p_k}{Tp_{-k}}\right).
$$

Moreover, we know from Lemma 2(iii) that

$$
\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\mathbf{A}_j\right\|_F^2=O_P\left(\frac{1}{p_k}\right)+O_P\left(\frac{1}{Tp_{-k}}\right),
$$

and, by Assumption 2(i), that $\|\mathbf{A}_\ell\|_F^2=c_0p_\ell$. Noting that $\prod_{\ell=1,\ell\neq j,k}^{K}p_\ell=p/(p_jp_k)$, we have

$$
\begin{aligned}
&\frac{1}{p_k}\left\|\frac{1}{Tp_{-k}}\sum_{s=1}^{T}\mathbf{E}_{k,s}\mathbf{F}_{k,s}\right\|_F^2\left(\sum_{j=1,j\neq k}^{K}\left[\left(\prod_{\ell=1,\ell\neq j,k}^{K}\|\mathbf{A}_\ell\|_F^2\right)\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\mathbf{A}_j\right\|_F^2\right]\right)\\
&\quad=\frac{1}{p_k}\frac{p_k}{Tp_{-k}}\left(\sum_{j=1,j\neq k}^{K}\left(\prod_{\ell=1,\ell\neq j,k}^{K}p_\ell\right)\left(\frac{1}{p_j}+\frac{1}{Tp_{-j}}\right)\right)O_P(1)\\
&\quad=\frac{1}{Tp_{-k}}\left(\sum_{j=1,j\neq k}^{K}\left(\frac{p}{p_kp_j^2}+\frac{1}{Tp_k}\right)\right)O_P(1)=O_P\left(\frac{1}{pT^2}\right)+O_P\left(\sum_{j=1,j\neq k}^{K}\frac{1}{p_j^2T}\right).
\end{aligned}
\tag{43}
$$

Also, using the facts that

$$
\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\right\|_F^2=O_P\left(\frac{1}{Tp_{-j}}\right)+O_P\left(\frac{1}{p_j^2}\right),\qquad\left\|\widehat{\mathbf{A}}_j-\mathbf{A}_j\widehat{\mathbf{H}}_j\right\|_F^2=O_P\left(p_j\widehat{w}_j\right),
$$

we have

$$\frac{1}{p_k}\left\|\frac{1}{Tp_{-k}}\sum_{s=1}^{T}\mathbf{E}_{k,s}\mathbf{F}_{k,s}\right\|_F^2\left(\sum_{j=1,j\neq k}^{K}\left[\left(\prod_{\ell=1,\ell\neq j,k}^{K}\|\mathbf{A}_\ell\|_F^2\right)\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\right\|_F^2\left\|\widehat{\mathbf{A}}_j-\mathbf{A}_j\widehat{\mathbf{H}}_j\right\|_F^2\right]\right)$$

$$=\frac{1}{p_k}\frac{p_k}{Tp_{-k}}\left(\sum_{j=1,j\neq k}^{K}\left(\prod_{\ell=1,\ell\neq j,k}^{K}p_\ell\right)\left(\frac{1}{Tp_{-j}}+\frac{1}{p_j^2}\right)p_j\widehat{w}_j\right)O_P(1)=\frac{1}{T}\left(\sum_{j=1,j\neq k}^{K}\left(\frac{1}{Tp_{-j}}+\frac{1}{p_j^2}\right)\widehat{w}_j\right), \tag{44}$$

which is dominated. Then, (42) follows from combining (43) and (44). The desired result obtains by combining (40), (41) and (42), this showing that the sufficient condition (26) is satisfied with $m_{-k}=\sum_{j=1,j\neq k}^{K}\left(T^{-2}p_{-j}^{-2}+T^{-1}p_j^{-2}\right)$. □

Proof of Corollary 1. The proof follows directly from Theorem 3 and Lemma 1. □

Proof of Theorem 4. Based on Eq. (36), for any $1\le i\le p_k$ it holds that

$$\widetilde{\mathbf{A}}_{k,i}^\top-\widetilde{\mathbf{H}}_k\mathbf{A}_{k,i}^\top=\frac{1}{Tp_kp_{-k}^2}\widetilde{\mathbf{\Lambda}}_k^{-1}\widetilde{\mathbf{A}}_k^\top\sum_{t=1}^{T}\left(\mathbf{A}_k\mathbf{F}_{k,t}\mathbf{B}_k^\top\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{e}_{t,k,i\cdot}^\top+\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{B}_k\mathbf{F}_{k,t}^\top\mathbf{A}_{k,i\cdot}^\top+\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\widehat{\mathbf{B}}_k^\top\mathbf{e}_{t,k,i\cdot}^\top\right)$$
$$=\mathcal{XV}+\mathcal{XVI}+\mathcal{XVII}.$$

Using Theorem 1, it holds that

$$\mathcal{XV}=\widetilde{\mathbf{\Lambda}}_k^{-1}\frac{\widetilde{\mathbf{A}}_k^\top\mathbf{A}_k}{p_k}\frac{1}{Tp_{-k}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\frac{\mathbf{B}_k^\top\widehat{\mathbf{B}}_k}{p_{-k}}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}+\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{e}_{t,k,i\cdot}^\top$$
$$=\widetilde{\mathbf{\Lambda}}_k^{-1}\frac{\widetilde{\mathbf{A}}_k^\top\mathbf{A}_k}{p_k}\frac{1}{Tp_{-k}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\left(\mathbf{I}_{r_{-k}}+O_P\left(w_k^{1/2}\right)\right)\mathbf{B}_k^\top\mathbf{e}_{t,k,i\cdot}^\top+\widetilde{\mathbf{\Lambda}}_k^{-1}\frac{\widetilde{\mathbf{A}}_k^\top\mathbf{A}_k}{p_k}\frac{\mathbf{B}_k^\top\widehat{\mathbf{B}}_k}{p_{-k}}\frac{1}{Tp_{-k}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top\mathbf{e}_{t,k,i\cdot}^\top.$$

By Lemma 1, it follows that

$$\left\|\frac{1}{Tp_{-k}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)^\top e_{k,t,i\cdot}^\top\right\|_F=O_P\left(\sum_{j=1,j\neq k}^{K}\left(\frac{1}{p_j\sqrt{T}}+\frac{1}{Tp_{-j}}\right)\right).$$

Hence, using Theorems 1 and 3, Eq. (37) and noting that, by Lemma 2(ii)

$$\widetilde{\mathbf{\Lambda}}_k^{-1}\frac{\widetilde{\mathbf{A}}_k^\top\mathbf{A}_k}{p_k}\frac{1}{Tp_{-k}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^\top e_{k,t,i\cdot}^\top=O_P\left(\frac{1}{\sqrt{Tp_{-k}}}\right),$$

it follows that

$$\mathcal{XV}=\widetilde{\mathbf{\Lambda}}_k^{-1}\frac{\widetilde{\mathbf{A}}_k^\top\mathbf{A}_k}{p_k}\frac{1}{Tp_{-k}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^\top e_{k,t,i\cdot}^\top+o_P\left(\frac{1}{\sqrt{Tp_{-k}}}\right)+O_P\left(\sum_{j=1,j\neq k}^{K}\left(\frac{1}{p_j\sqrt{T}}+\frac{1}{Tp_{-j}}\right)\right). \tag{45}$$

Using now by Theorem 1, Assumption 2(i) and (37), we now turn to showing that

$$\|\mathcal{XVI}\|_F\lesssim\left\|\frac{1}{Tp}\sum_{t=1}^{T}\widetilde{\mathbf{A}}_k^\top\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\mathbf{F}_{k,t}^\top\right\|_F\lesssim\left\|\widetilde{\mathbf{A}}_k-\mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\mathbf{F}_{k,t}^\top\right\|_F+\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\mathbf{F}_{k,t}^\top\right\|_F \tag{46}$$
$$=O_P\left(\frac{1}{\sqrt{Tp}}+\frac{1}{Tp_{-k}}+\sum_{j=1,j\neq k}^{K}\left(\frac{1}{Tp_{-j}}+\frac{1}{\sqrt{Tp_kp_j}}+\frac{1}{Tp_j\sqrt{p_{-k}}}+\frac{1}{Tp_j^2}\right)\right).$$

In order to prove (46), we begin by noting that by Lemmas 2(ii) and 1, it follows that

$$\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\mathbf{F}_{k,t}^\top\right\|_F\lesssim\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\mathbf{B}_k\mathbf{F}_{k,t}^\top\right\|_F+\left\|\sum_{t=1}^{T}\mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\mathbf{F}_{k,t}^\top\right\|_F=O_P\left(Tp+\sum_{j=1,j\neq k}^{K}\left(p_j^2+Tp_{-j}^2\right)\right).$$

We also note that

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

Recalling that $\left\|\widehat{\mathbf{H}}_{-k}\right\|_F^2=O_P(1)$, Assumption 4(i) readily entails that

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

Further (recall the notation in (A.1))

$$\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\left(\widehat{\mathbf{B}}_k-\mathbf{B}_k\widehat{\mathbf{H}}_{-k}\right)\mathbf{F}_{k,t}^\top\right\|_F^2 \lesssim \sum_{j=1,j\neq k}^{K}\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K}\mathbf{A}_\ell\right)\otimes\left(\mathcal{II}\widehat{\mathbf{A}}_j\right)\otimes\left(\otimes_{\ell=1,\ell\neq k}^{j-1}\mathbf{A}_\ell\right)\right)\mathbf{F}_{k,t}^\top\right\|_F^2$$
$$+\sum_{j=1,j\neq k}^{K}\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K}\mathbf{A}_\ell\right)\otimes\left(\mathcal{III}\widehat{\mathbf{A}}_j\right)\otimes\left(\otimes_{\ell=1,\ell\neq k}^{j-1}\mathbf{A}_\ell\right)\right)\mathbf{F}_{k,t}^\top\right\|_F^2$$
$$+\sum_{j=1,j\neq k}^{K}\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\left(\left(\otimes_{\ell=j+1,\ell\neq k}^{K}\mathbf{A}_\ell\right)\otimes\left(\mathcal{IV}\widehat{\mathbf{A}}_j\right)\otimes\left(\otimes_{\ell=1,\ell\neq k}^{j-1}\mathbf{A}_\ell\right)\right)\mathbf{F}_{k,t}^\top\right\|_F^2$$
$$=\mathcal{XVIII}+\mathcal{XIX}+\mathcal{XX}.$$

Upon inspecting the proof of Lemma 1, it can be shown that

$$\mathcal{XVIII}\lesssim\sum_{j=1,j\neq k}^{K}\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{B}_k\mathbf{F}_{k,t}^\top\right\|_F^2\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{F}_{j,t}\mathbf{B}_j^\top\mathbf{E}_{j,t}^\top\widehat{\mathbf{A}}_j\mathbf{F}_{k,t}^\top\right\|_F^2=O_P\left(\sum_{j=1,j\neq k}^{K}\left(1+\frac{p_j}{Tp_{-j}}\right)\right),$$
$$\mathcal{XIX}\lesssim T^2p_kp_{-k}^2\left\|\mathbf{A}_k\right\|_F^2\mathcal{XIII}=O_P\left(\sum_{j=1,j\neq k}^{K}\left(p_j^2+p_jp_k\right)\right),$$

and

$$\mathcal{XX}\lesssim\sum_{j=1,j\neq k}^{K}\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{F}_{k,t}\right\|_F^2\left(\prod_{\ell=1,\ell\neq j,k}^{K}\left\|\mathbf{A}_\ell\right\|_F^2\right)\left(\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\mathbf{A}_j\right\|_F^2+\left\|\frac{1}{Tp}\sum_{t=1}^{T}\mathbf{E}_{j,t}\mathbf{E}_{j,t}^\top\right\|_F^2\left\|\widehat{\mathbf{A}}_j-\mathbf{A}_j\widehat{\mathbf{H}}_j\right\|_F^2\right)$$
$$=O_P\left(\sum_{j=1,j\neq k}^{K}\frac{1}{p_k}\left(Tp_{-j}^2+\frac{p_j^2}{T}\right)\right).$$

Putting all together, we have

$$\left\|\sum_{t=1}^{T}\mathbf{A}_k^\top\mathbf{E}_{k,t}\widehat{\mathbf{B}}_k\mathbf{F}_{k,t}^\top\right\|_F^2=O_P\left(Tp+\sum_{j=1,j\neq k}^{K}\left(p_j^2+p_jp_k+\frac{Tp_{-j}^2}{p_k}\right)\right),$$

whence we finally obtain (46).

Finally we consider $\mathcal{XVII}$. Following the same passages as in the proof of Lemma 6, we obtain

$$\|\mathcal{XVI}\|_F\lesssim\frac{1}{Tp}+\frac{1}{p^2}+\left(\frac{1}{p_{-k}^2}+\frac{w_{-k}}{p_{-k}}\right)\widetilde{w}_k+\left(\frac{1}{Tp}+\frac{1}{p^2}\right)w_{-k}+\left(\frac{1}{Tp_k}+\frac{1}{p_k^2}\right)w_{-k}^2+\left(\frac{1}{p_{-k}}+w_{-k}\right)w_{-k}\widetilde{w}_k$$
$$=O_P\left(\frac{1}{p}\right)+O_P\left(\sum_{j=1,j\neq k}^{K}\frac{1}{p_kp_j^2}\right)+o_P\left(\frac{1}{\sqrt{Tp_{-k}}}\right)+o_P\left(\frac{1}{Tp_{-j}}\right)+o_P\left(\frac{1}{\sqrt{Tp_j}}\right). \quad (47)$$

Now, consider the restriction

$$Tp_{-k}=o\left(\min\left(p^2,T^2p_{-j}^2,Tp_j^2,p_k^2p_j^4\right)\right), \quad (48)$$

for all $j\neq k$ and as $\min\{T,p_1,\dots,p_K\}\to\infty$, and notice that if $K\geq 3$ this cannot be satisfied. Indeed, if $K=3$, we would need $Tp_2p_3=o(Tp_2^2)$ but also and $Tp_2p_3=o(Tp_3^2)$, which would create a contradiction.

Hence let us consider first the case $K\leq 2$, and by combining (45)–(47), as $\min\{T,p_1,\dots,p_K\}\to\infty$ under the restriction (48) we have

$$\sqrt{Tp_{-k}}\left(\widetilde{\mathbf{A}}_{k,i\cdot}^\top-\widetilde{\mathbf{H}}_k^\top\mathbf{A}_{k,i\cdot}^\top\right)=\widetilde{\boldsymbol{\Lambda}}_k^{-1}\frac{\widetilde{\mathbf{A}}_k^\top\mathbf{A}_k}{p_k}\frac{1}{\sqrt{Tp_{-k}}}\sum_{t=1}^{T}\mathbf{F}_{k,t}\mathbf{B}_k^\top\boldsymbol{e}_{k,t,i\cdot}^\top+o_P(1).$$

We know from Lemma 5 that $\widetilde{\boldsymbol{\Lambda}}_k\xrightarrow{P}\boldsymbol{\Lambda}_k$; further, recall that, by Assumption 1(ii), $\boldsymbol{\Sigma}_k$ has spectral decomposition $\boldsymbol{\Sigma}_k=\boldsymbol{\Gamma}_k\boldsymbol{\Lambda}_k\boldsymbol{\Gamma}_k^\top$. Using the definition of $\widetilde{\mathbf{H}}_k$ in (35), Theorems 1 and 3 entail $\widetilde{\mathbf{H}}_k=\boldsymbol{c}_k\boldsymbol{\Lambda}_k\boldsymbol{\Gamma}_k^\top\widetilde{\mathbf{H}}_k\boldsymbol{\Lambda}_k^{-1}+o_P(1)$. By rearranging the last equation, and since $\boldsymbol{\Gamma}_k^\top\boldsymbol{\Gamma}_k=\mathbf{I}_{r_k}$, $\boldsymbol{\Gamma}_k^\top\widetilde{\mathbf{H}}_k\boldsymbol{\Lambda}_k=\boldsymbol{\Lambda}_k\boldsymbol{\Gamma}_k^\top\widetilde{\mathbf{H}}_k+o_P(1)$. Hence, $\boldsymbol{\Gamma}_k^\top\widetilde{\mathbf{H}}_k=\mathbf{I}_{r_k}+o_P(1)$. Therefore, from Theorem 3, $p_k^{-1}\widetilde{\mathbf{A}}_k^\top\mathbf{A}_k=\boldsymbol{\Gamma}_k^\top+o_P(1)$. Using now Slutsky's theorem, and Assumption 5(i), it readily follows that $\sqrt{Tp_{-k}}\left(\widetilde{\mathbf{A}}_{k,i\cdot}^\top-\widetilde{\mathbf{H}}_k^\top\mathbf{A}_{k,i\cdot}^\top\right)\xrightarrow{D}\mathcal{N}\left(\mathbf{0},\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Gamma}_k^\top\mathbf{V}_{ki}\boldsymbol{\Gamma}_k\boldsymbol{\Lambda}_k^{-1}\right)$, which concludes the proof of part (i).

Finally, consider the case in which $K\leq 2$ and we impose the restriction

$$Tp_{-k}\gtrsim\min\left\{p^2,T^2p_{-j}^2,Tp_j^2,p_k^2p_j^2\right\}, \quad (49)$$

for all $j\neq k$ and as $\min\{T,p_1,\dots,p_K\}\to\infty$, and notice also that if $K\geq 3$, then (49) always hold. Then, the proof of part (ii) follows from (45)–(47) under (49). This completes the proof. □

Proof of Theorem 6. By definition, it holds that

$$\begin{aligned}\widetilde{\mathcal{F}}_t &= \frac{1}{p}\mathcal{X}_t \times_1 \widetilde{\mathbf{A}}_1^\top \times_2 \dots \times_K \widetilde{\mathbf{A}}_K^\top = \frac{1}{p}\left(\mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \dots \times_K \mathbf{A}_K + \mathcal{E}_t\right) \times_1 \widetilde{\mathbf{A}}_1^\top \times_2 \dots \times_K \widetilde{\mathbf{A}}_K^\top \\ &= \frac{1}{p}\mathcal{F}_t \times_1 \widetilde{\mathbf{A}}_1^\top\left(\mathbf{A}_1 - \widetilde{\mathbf{A}}_1\widetilde{\mathbf{H}}_1^{-1} + \widetilde{\mathbf{A}}_1\widetilde{\mathbf{H}}_1^{-1}\right) \times_2 \dots \times_K \widetilde{\mathbf{A}}_K^\top\left(\mathbf{A}_K - \widetilde{\mathbf{A}}_K\widetilde{\mathbf{H}}_K^{-1} + \widetilde{\mathbf{A}}_K\widetilde{\mathbf{H}}_K^{-1}\right) \\ &\quad + \frac{1}{p}\mathcal{E}_t \times_1 \left(\widetilde{\mathbf{A}}_1 - \mathbf{A}_1\widetilde{\mathbf{H}}_1 + \mathbf{A}_1\widetilde{\mathbf{H}}_1\right)^\top \times_2 \dots \times_K \left(\widetilde{\mathbf{A}}_K - \mathbf{A}_K\widetilde{\mathbf{H}}_K + \mathbf{A}_K\widetilde{\mathbf{H}}_K\right)^\top . \end{aligned} \tag{50}$$

Then we have

$$\begin{aligned}\left\|\widetilde{\mathcal{F}}_t - \mathcal{F}_t \times_1 \widetilde{\mathbf{H}}_1^{-1} \times_2 \dots \times_K \widetilde{\mathbf{H}}_K^{-1}\right\|_F &\lesssim \sum_{k=1}^K \left\|\frac{1}{p_k}\widetilde{\mathbf{A}}_k^\top\left(\mathbf{A}_k - \widetilde{\mathbf{A}}_k\widetilde{\mathbf{H}}_k^{-1}\right)\mathbf{F}_{k,t}\right\|_F + \sum_{k=1}^K \left\|\frac{1}{p}\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)^\top \mathbf{E}_{k,t}\mathbf{B}_k\right\|_F \\ &\quad + \sum_{k=1}^K \left\|\frac{1}{p}\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{B}_k\right\|_F .\end{aligned}$$

The first term follows immediately by using Lemma 7; similarly, it also follows that

$$\begin{aligned}\left\|\frac{1}{p}\left(\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right)^\top \mathbf{E}_{k,t}\mathbf{B}_k\right\|_F^2 &\le \frac{1}{p^2}\left\|\widetilde{\mathbf{A}}_k - \mathbf{A}_k\widetilde{\mathbf{H}}_k\right\|_F^2 \left\|\mathbf{E}_{k,t}\mathbf{B}_k\right\|_F^2 \\ &\lesssim \frac{1}{p p_{-k}}\widetilde{\omega}_k \left(\sum_{i=1}^{p_k}\sum_{h=1}^{r_{-k}}\left(\sum_{j=1}^{p_{-k}} e_{k,t,ij}B_{k,jh}\right)^2\right) \\ &\lesssim \frac{1}{p p_{-k}}\widetilde{\omega}_k \sum_{i=1}^{p_k}\sum_{j_1,j_2=1}^{p_{-k}}\left|e_{k,t,ij_1}e_{k,t,ij_2}\right| = O_P(1)\frac{1}{p_{-k}}\widetilde{\omega}_k,\end{aligned}$$

having used (24) in the last passage. Also, it holds that

$$\left\|\frac{1}{p}\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{B}_k\right\|_F^2 = \frac{1}{p^2}\sum_{h=1}^{r_k}\sum_{g=1}^{r_{-k}}\left(\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}} A_{k,ih}B_{k,jg}e_{k,t,ij}\right)^2 \lesssim \frac{1}{p^2}\sum_{i_1,i_2=1}^{p_k}\sum_{j_1,j_2=1}^{p_{-k}}\left|e_{k,t,ij_1}e_{k,t,ij_2}\right| = O_P(1)\frac{1}{p},$$

again by (24). Hence, we have

$$\begin{aligned}\left\|\widetilde{\mathcal{F}}_t - \mathcal{F}_t \times_1 \widetilde{\mathbf{H}}_1^{-1} \times_2 \dots \times_K \widetilde{\mathbf{H}}_K^{-1}\right\|_F &\lesssim O_P(1)\left(\frac{1}{\sqrt{p}} + \sum_{k=1}^K\left(z_k + \sqrt{\frac{\widetilde{\omega}_k}{p_{-k}}}\right)\right) \\ &= O_P\left(\frac{1}{\sqrt{p}} + \sum_{k=1}^K\left(\frac{1}{\sqrt{Tp_k}} + \frac{1}{\sqrt{Tp_{-k}}}\sum_{j=1,j\neq k}^K \frac{1}{p_k p_j^2}\right)\right),\end{aligned} \tag{51}$$

where we have used the short-hand notation (based on Lemma 7)

$$z_k = O_P\left(\frac{1}{\sqrt{Tp}}\right) + O_P\left(\frac{1}{p}\right) + O_P\left(\sum_{j=1}^K \frac{1}{Tp_{-j}}\right) + O_P\left(\sum_{j=1,j\neq k}^K\left(\frac{1}{Tp_j\sqrt{p_{-k}}} + \frac{1}{p_j\sqrt{Tp_k}} + \frac{1}{p_k p_j^2} + \frac{1}{Tp_j^2}\right)\right).$$

Now, if for all $1 \le k \le K$

$$p = o\left(\min\left\{Tp_{-k}^2, Tp_kp_j^2, Tp_{-k}p_j^2, p_k^2p_j^4\right\}\right) \tag{52}$$

for all $j \neq k$ and as $\min\{T, p_1, \dots, p_K\} \to \infty$, then from (50) we have $\sqrt{p}\left(\widetilde{\mathcal{F}}_t - \mathcal{F}_t \times_{k=1}^K \widetilde{\mathbf{H}}_k^{-1}\right) = p^{-1/2}\mathcal{E}_t \times_{k=1}^K \left(\mathbf{A}_k\widetilde{\mathbf{H}}_k\right)^\top + o_P(1)$, from which it follows that for all $1 \le k \le K$, $\sqrt{p}\left(\widetilde{\mathbf{F}}_{t,k} - \widetilde{\mathbf{H}}_k^{-1}\mathbf{F}_{t,k}\left(\widetilde{\mathbf{H}}_{-k}^{-1}\right)^\top\right) = p^{-1/2}\widetilde{\mathbf{H}}_k^\top\mathbf{A}_k^\top\mathbf{E}_{k,t}\mathbf{B}_k\widetilde{\mathbf{H}}_{-k} + o_P(1)$, where $\widetilde{\mathbf{H}}_{-k} = \otimes_{j\in[K]/\{k\}}\widetilde{\mathbf{H}}_j$ and so $\widetilde{\mathbf{H}}_{-k}^{-1} = \otimes_{j\in[K]/\{k\}}\widetilde{\mathbf{H}}_j^{-1}$. And therefore, $\sqrt{p}\left(\mathrm{Vec}\left(\widetilde{\mathbf{F}}_{t,k}\right) - \left(\widetilde{\mathbf{H}}_{-k}\otimes\widetilde{\mathbf{H}}_k\right)^{-1}\mathrm{Vec}\left(\mathbf{F}_{t,k}\right)\right) = p^{-1/2}\left(\widetilde{\mathbf{H}}_{-k}\otimes\widetilde{\mathbf{H}}_k\right)^\top\left(\mathbf{B}_k\otimes\mathbf{A}_k\right)^\top\mathrm{Vec}\left(\mathbf{E}_{k,t}\right) + o_P(1)$. The proof of part (i) follows from Slutsky's theorem, and Assumption 5(ii). The proof of part (ii) follows from (51) when condition (52) does not hold. □

Proof of Theorem 5. We begin by noting that

$$\widetilde{\mathbf{H}}_k^* = \widetilde{\mathbf{\Lambda}}_k^{-1}\left(\frac{\widetilde{\mathbf{A}}_k^\top\mathbf{A}_k}{p_k}\right)\frac{1}{T}\sum_{t=1}^T \mathbf{F}_{k,t}\left(\frac{\mathbf{B}_k^\top\widehat{\mathbf{B}}_k^*}{p_{-k}}\right)\left(\frac{\mathbf{B}_k^\top\widehat{\mathbf{B}}_k^*}{p_{-k}}\right)^\top \mathbf{F}_{k,t}^\top. \tag{53}$$

The proof of the first part follows from showing that

$$\frac{1}{p_{-k}}\left\|\widehat{\mathbf{B}}_k^* - \mathbf{B}_k\widehat{\mathbf{H}}_{-k}^*\right\|_F^2 = O_P\left(\frac{1}{Tp_k}\right) + O_P\left(\frac{1}{p_{-k}^2}\right),$$

with $\widehat{\mathbf{H}}^*_{-k} = \otimes_{j=1,j\neq k}^{K} \widehat{\mathbf{H}}^*_k$, using the same logic as in the remainder of the proofs, and subsequently from a direct application of Corollary 3.1 in Yu et al. (2022). The second part also follows immediately, by the same passages as in the proof of Theorem 4, mutatis mutandis using $\widehat{\mathbf{B}}^*_k$ instead of $\widehat{\mathbf{B}}_k$. □

Proof of Theorem 7. Recall the notation $\widetilde{S}_{t,i_1,\dots,i_K} = \widetilde{\mathcal{F}}_t \times_1 \widetilde{\mathbf{A}}_{1,i_1\cdot} \times_2 \dots \times_K \widetilde{\mathbf{A}}_{K,i_K\cdot}$, and $S_{t,i_1,\dots,i_K} = \mathcal{F}_t \times_1 \mathbf{A}_{1,i_1\cdot} \times_2 \dots \times_K \mathbf{A}_{K,i_K\cdot}$. We have from Theorems 4 and 6

$$
\begin{aligned}
\left|\widetilde{S}_{t,i_1,\dots,i_K} - S_{t,i_1,\dots,i_K}\right| \lesssim & \sum_{k=1}^{K} \left\| \left(\widetilde{\mathbf{A}}_{k,i_k\cdot} - \mathbf{A}_{k,i_k\cdot}\widetilde{\mathbf{H}}_k\right)^\top \widetilde{\mathbf{F}}_{k,t}\mathbf{B}_{k,j_k\cdot}\widetilde{\mathbf{H}}_{-k}\right\|_F \\
& + \left\| \left(\widetilde{\mathcal{F}}_t - \mathcal{F}_t \times_1 \widetilde{\mathbf{H}}_1^{-1} \times_2 \dots \times_K \widetilde{\mathbf{H}}_K^{-1}\right) \times_1 \left(\mathbf{A}_{1,i_1\cdot}\widetilde{\mathbf{H}}_1\right)^\top \times_2 \dots \times_K \left(\mathbf{A}_{K,i_K\cdot}\widetilde{\mathbf{H}}_K\right)^\top \right\|_F \\
\lesssim & O_P(1)\left(\sum_{k=1}^{K}\sqrt{\widetilde{w}_k} + \frac{1}{\sqrt{p}} + \sum_{k=1}^{K}\left(z_k + \sqrt{\frac{\widetilde{w}_k}{p_{-k}}}\right)\right) \\
= & O_P\left(\frac{1}{\sqrt{p}} + \sum_{k=1}^{K}\left(\frac{1}{\sqrt{Tp_k}} + \frac{1}{\sqrt{Tp_{-k}}}\sum_{j=1,j\neq k}\frac{1}{p_k p_j^2}\right)\right),
\end{aligned}
$$

whence the result obtains. □

Proof of Theorem 8. By Lemma 3 it holds that, for all $j \leq r_k$, it holds that

$$
\lambda_j\left(\widehat{\mathbf{M}}_k\right) = \lambda_j\left(\boldsymbol{\Sigma}_k\right) + o_P(1), \tag{54}
$$

with $\lambda_j\left(\boldsymbol{\Sigma}_k\right) > 0$. Moreover, Lemma 4 and standard arguments based on Weyl's inequality entail that, for all $j > r_k$

$$
\lambda_j\left(\widehat{\mathbf{M}}_k\right) = O_P\left(\frac{1}{\sqrt{Tp_{-k}}} + \frac{1}{p_k}\right). \tag{55}
$$

Hence, using elementary arguments, (54) entails

$$
\max_{1\leq j\leq r_k-1} \frac{\lambda_j\left(\widehat{\mathbf{M}}_k\right)}{\lambda_{j+1}\left(\widehat{\mathbf{M}}_k\right) + \widehat{c}\delta_{p_1,\dots,p_K,T}} \leq \max_{1\leq j\leq r_k-1} \frac{\lambda_j\left(\widehat{\mathbf{M}}_k\right)}{\lambda_{j+1}\left(\widehat{\mathbf{M}}_k\right)} = O_P(1),
$$

and, by (55) and the definition of $\delta_{p_1,\dots,p_K,T}$, it also follows that

$$
\max_{r_k+1\leq j\leq r_{\max}} \frac{\lambda_j\left(\widehat{\mathbf{M}}_k\right)}{\lambda_{j+1}\left(\widehat{\mathbf{M}}_k\right) + \widehat{c}\delta_{p_1,\dots,p_K,T}} \leq \max_{1\leq j\leq r_k-1} \frac{\lambda_j\left(\widehat{\mathbf{M}}_k\right)}{\widehat{c}\delta_{p_1,\dots,p_K,T}} = O_P(1).
$$

Finally, combining (54) and (55), there exists a $c_0 > 0$ such that

$$
\Pr\left(\frac{\lambda_{r_k}\left(\widehat{\mathbf{M}}_k\right)}{\lambda_{r_k+1}\left(\widehat{\mathbf{M}}_k\right) + \widehat{c}\delta_{p_1,\dots,p_K,T}} \geq c_0\delta^{-1}_{p_1,\dots,p_K,T}\lambda_{r_k}\left(\widehat{\mathbf{M}}_k\right)\right) = 1.
$$

The final result obtains readily upon noting that $\delta^{-1}_{p_1,\dots,p_K,T}\lambda_{r_k}\left(\widehat{\mathbf{M}}_k\right) \xrightarrow{P} \infty$, as $\min\left\{p_1,\dots,p_K,T\right\} \to \infty$. We conclude by showing that, when choosing $\widehat{c}$ and $\widetilde{c}$ according to (22), it holds that

$$
\widehat{c}^L + o_P(1) \leq \widehat{c} \leq \widehat{c}^U + o_P(1), \quad \widetilde{c}^L + o_P(1) \leq \widetilde{c} \leq \widetilde{c}^U + o_P(1), \tag{56}
$$

which, in turn, automatically entails that the theorem holds when using $\widehat{c}$ and $\widetilde{c}$. We focus only on the first inequality in (56) to save space, since the proof of the second one is basically the same. We begin by noting that $\widehat{c} \geq \lambda_1\left(\widehat{\mathbf{M}}_k\right) = c_0 + o_P(1)$, where $c_0 > 0$, as an immediate consequence of Lemma 3. Also, setting $r_k = 1$ for simplicity

$$
\sum_{j=1}^{p_k}\lambda_j\left(\widehat{\mathbf{M}}_k\right) = \frac{1}{Tp}\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}} x^2_{k,t,ij} \leq 2\left(\frac{1}{Tp}\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}}\sum_{t=1}^{T} A^2_{ki}B^2_{kj}F^2_{k,t} + \frac{1}{Tp}\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}}\sum_{t=1}^{T} e^2_{k,t,ij}\right).
$$

It holds that

$$
\mathbb{E}\left(\frac{1}{Tp}\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}}\sum_{t=1}^{T} A^2_{ki}B^2_{kj}F^2_{k,t}\right) \leq \frac{1}{Tp}\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}}\sum_{t=1}^{T} A^2_{ki}B^2_{kj}\mathbb{E}\left(F^2_{k,t}\right) \leq c_0,
$$

having used Assumptions 1 and 2; further, by Assumption 3(i), it follows immediately that $\sum_{i=1}^{p_k}\sum_{j=1}^{p_{-k}}\sum_{t=1}^{T} e^2_{k,t,ij} = O_P(Tp)$, whence the upper bound for $\sum_{j=1}^{p_k}\lambda_j\left(\widehat{\mathbf{M}}_k\right)$. □

Proof of Theorem 9. We begin by noting that, if $\hat{r}_j^{(s-1)} > r_j$ for all $j \in [K]\backslash k$, we can and we will assume $\hat{r}_j^{(s-1)} = r_j + 1$, without loss of generality because $r_{\max}$ is a constant. Hence, by definition, $\hat{\mathbf{A}}_j^{(s)} = \left(\hat{\mathbf{A}}_j, \hat{\gamma}_j\right)$, where $\hat{\mathbf{A}}_j$ is the PCA estimator computed using the eigenvectors of $\hat{\mathbf{M}}_k^{(s)}$ corresponding to its r_j largest eigenvalues (using the true number of common factors r_j), and $p_j^{-1/2}\hat{\gamma}_j$ is the $\left(r_j+1\right)$th eigenvector of $\hat{\mathbf{M}}_k^{(s)}$; by construction, $\hat{\mathbf{A}}_j^\top \hat{\gamma}_j = \mathbf{0}$. Therefore, it follows that

$$\begin{aligned}\widetilde{\mathbf{M}}_k^{(s)} &= \frac{1}{Tp}\sum_{t=1}^T \mathbf{X}_{k,t}\hat{\mathbf{B}}_k^{(s)}\left(\hat{\mathbf{B}}_k^{(s)}\right)^\top \mathbf{X}_{k,t}^\top = \frac{1}{Tp}\sum_{t=1}^T \mathbf{X}_{k,t}\left(\otimes_{j\neq k}\left(\hat{\mathbf{A}}_j, \hat{\gamma}_j\right)\right)\left(\otimes_{j\neq k}\left(\hat{\mathbf{A}}_j, \hat{\gamma}_j\right)\right)^\top \mathbf{X}_{k,t}^\top \\ &= \frac{1}{Tp}\sum_{t=1}^T \mathbf{X}_{k,t}\left(\otimes_{j\neq k}\left(\hat{\mathbf{A}}_j\hat{\mathbf{A}}_j^\top + \hat{\gamma}_j\hat{\gamma}_j^\top\right)\right)\mathbf{X}_{k,t}^\top = \widetilde{\mathbf{M}}_k + \frac{1}{Tp}\sum_{t=1}^T \mathbf{X}_{k,t}\left(\otimes_{j\neq k}\left(\hat{\gamma}_j\hat{\gamma}_j^\top\right)\right)\mathbf{X}_{k,t}^\top,\end{aligned}$$

where the second matrix is non-negative definite. Using Weyl's inequality, it is easy to see that

$$\lambda_j\left(\widetilde{\mathbf{M}}_k^{(s)}\right) \geq \lambda_j\left(\widetilde{\mathbf{M}}_k\right) + \lambda_{\min}\left(\frac{1}{Tp}\sum_{t=1}^T \mathbf{X}_{k,t}\left(\otimes_{j\neq k}\left(\hat{\gamma}_j\hat{\gamma}_j^\top\right)\right)\mathbf{X}_{k,t}^\top\right) \geq \lambda_j\left(\widetilde{\mathbf{M}}_k\right),$$

for all $j \leq r_k$. Thus, by Lemma 3, $\lambda_j\left(\widetilde{\mathbf{M}}_k^{(s)}\right) \geq \lambda_j\left(\boldsymbol{\Sigma}_k\right) + o_P(1)$, for all $j \leq r_k$. Also note that, on account of Assumption 2(ii), the matrix $\left(\hat{\mathbf{A}}_j\hat{\mathbf{A}}_j^\top + \hat{\gamma}_j\hat{\gamma}_j^\top\right)$ is the idempotent, and therefore so is $\mathbf{I}_{r_j} - \left(\hat{\mathbf{A}}_j\hat{\mathbf{A}}_j^\top + \hat{\gamma}_j\hat{\gamma}_j^\top\right)$. Hence

$$\hat{\mathbf{M}}_k - \widetilde{\mathbf{M}}_k^{(s)} = \frac{1}{Tp}\sum_{t=1}^T \mathbf{X}_{k,t}\left(\otimes_{j\neq k}\left[\mathbf{I}_{r_j} - \left(\hat{\mathbf{A}}_j\hat{\mathbf{A}}_j^\top + \hat{\gamma}_j\hat{\gamma}_j^\top\right)\right]\right)\mathbf{X}_{k,t}^\top,$$

is a non-negative definite matrix. Hence it is immediate to see that, for all j $\lambda_j\left(\hat{\mathbf{M}}_k\right) - \lambda_j\left(\widetilde{\mathbf{M}}_k^{(s)}\right) \geq 0$. Thus, by Lemma 4, it follows that

$$\lambda_j\left(\widetilde{\mathbf{M}}_k^{(s)}\right) = O_P\left(\frac{1}{\sqrt{Tp_{-k}}} + \frac{1}{p_k}\right),$$

for all $j > r_k$. From hereon, the proof is the same as that of Theorem 8. □

Acknowledgments

This work is supported by National Key R&D Program of China (No. 2023YFA1008701).
The authors are listed in alphabetical order and all authors contributed equally to this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2026.105616.

References

Acar, E., Dunlavy, D.M., Kolda, T.G., Mørup, M., 2011. Scalable tensor factorizations for incomplete data. Chemometr. Intell. Lab. Syst. 106 (1), 41–56.

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227.

Aït-Sahalia, Y., Xiu, D., 2017. Using principal component analysis to estimate a high dimensional factor model with high frequency data. J. Econometrics 201, 388–399.

Alessi, L., Barigozzi, M., Capasso, M., 2010. Improved penalization for determining the number of factors in approximate static factor models. Statist. Probab. Lett. 80, 1806–1813.

Babii, A., Ghysels, E., Pan, J., 2022. Tensor principal component analysis. arXiv preprint arXiv:2212.12981.

Bae, K., Lee, S., 2021. Realized higher-order comoments. Quant. Finance 21, 421–429.

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

Bai, J., Ng, S., 2023. Approximate factor models with weaker loadings. J. Econometrics (ISSN: 0304-4076) 235 (2), 1893–1916.

Barigozzi, M., Cavaliere, G., Moramarco, G., 2025. Factor network autoregressions. J. Bus. Econom. Statist. 43 (4), 1105–1118. http://dx.doi.org/10.1080/07350015.2025.2476695.

Barigozzi, M., Trapani, L., 2021. Testing for common trends in nonstationary large datasets. J. Bus. Econom. Statist. 1–16.

Bi, X., Tang, X., Yuan, Y., Zhang, Y., Qu, A., 2021. Tensors in statistics. Annu. Rev. Stat. Appl. 8, 345–368.

Buckle, M., Chen, J., Williams, J.M., 2016. Realised higher moments: theory and practice. Eur. J. Financ. 22, 1272–1291.

Carroll, J.D., Chang, J.-J., 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika 35 (3), 283–319.

Chang, J., He, J., Yang, L., Yao, Q., 2023. Modelling matrix time series via a tensor CP-decomposition. J. R. Stat. Soc. Ser. B Stat. Methodol. 85 (1), 127–148.

Chen, E.Y., Fan, J., 2023. Statistical inference for high-dimensional matrix-variate factor models. J. Amer. Statist. Assoc. 118 (542), 1038–1055.

Chen, H., Guo, Y., He, Y., Ji, J., Liu, L., Shi, Y., Wang, Y., Yu, L., Zhang, X., 2022a. Simultaneous differential network analysis and classification for Matrix-variate data with application to Brain Connectivity. Biostatistics 23, 967–989.

Chen, W., Lam, C., 2024. Rank and factor loadings estimation in time series tensor factor model by Pre-averaging. Ann. Statist. 52, 364–391.

Chen, E.Y., Tsay, R.S., Chen, R., 2020. Constrained factor models for high-dimensional matrix-variate time series. J. Amer. Statist. Assoc. 115 (530), 775–793.

Chen, R., Yang, D., Zhang, C.-H., 2022b. Factor models for High-Dimensional tensor time series. J. Amer. Statist. Assoc. 117 (537), 94–116. http://dx.doi.org/10.1080/01621459.2021.1912757.

Davis, C., Kahan, W.M., 1970. The rotation of eigenvectors by a perturbation. III. SIAM J. Numer. Anal. 7 (1), 1–46.

Douglas Carroll, J., Pruzansky, S., Kruskal, J.B., 1980. CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. Psychometrika 45, 3–24.
Doz, C., Giannone, D., Reichlin, L., 2012. A quasi–maximum likelihood approach for large, approximate dynamic factor models. Rev. Econ. Stat. 94 (4), 1014–1024.
Entezari, N., Papalexakis, E.E., Wang, H., Rao, S., Prasad, S.K., 2021. Tensor-based complementary product recommendation. In: 2021 IEEE International Conference on Big Data (Big Data). IEEE, pp. 409–415.
Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (4), 603–680.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: Identification and estimation. Rev. Econ. Stat. 82 (4), 540–554.
Gao, Z., Yuan, C., Jing, B., Huang, W., Guo, J., 2021. A two-way factor model for high-dimensional matrix data. arXiv preprint arXiv:2103.07920.
Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. J. Amer. Statist. Assoc. 102 (478), 603–617.
Han, Y., Chen, R., Yang, D., Zhang, C., 2024a. Tensor factor model estimation by iterative projection. Ann. Statist. 52, 2641–2667.
Han, Y., Yang, D., Zhang, C.-H., Chen, R., 2024b. CP factor model for dynamic tensors. J. R. Stat. Soc. Ser. B Stat. Methodol. 86 (5), 1383–1413.
Han, Y., Zhang, C.H., Chen, R., 2022. Rank determination in tensor factor model. Electron. J. Stat. 16 (1), 1726–1803.
Hannan, E.J., 1970. Multiple Time Series. John Wiley & Sons.
He, Y., Kong, X., Trapani, L., Yu, L., 2023. One-way or Two-way factor model for matrix sequences? J. Econometrics 235 (1), 1981–2004.
He, Y., Kong, X.B., Trapani, L., Yu, L., 2024b. Online Change-point detection for Matrix-valued time series with latent two-way factor structure. Ann. Statist. 52, 1646–1670.
He, Y., Kong, X., Yu, L., Zhang, X., Zhao, C., 2024a. Matrix factor analysis: From least squares to iterative projection. J. Bus. Econom. Statist. 42, 322–334.
Ji, J., He, Y., Liu, L., Xie, L., 2021. Brain connectivity alteration detection via matrix-variate differential network model. Biometrics 77, 1409–1421.
Jondeau, E., Jurczenko, E., Rockinger, M., 2018. Moment component analysis: An illustration with international stock markets. J. Bus. Econom. Statist. 36, 576–598.
Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Rev. 51 (3), 455–500.
Lam, C., 2021. Rank Determination for Time Series Tensor Factor Model Using Correlation Thresholding. Technical Report, Working Paper, LSE, London, UK.
Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: inference for the number of factors. Ann. Statist. 40 (2), 694–726.
Lehmann, E.L., Casella, G., 2006. Theory of Point Estimation. Springer Science & Business Media.
Lettau, M., 2022. High Dimensional Factor Models with an Application to Mutual Fund Characteristics. Technical Report, National Bureau of Economic Research.
Liu, X., Chen, E., 2019. Helping effects against curse of dimensionality in threshold factor models for matrix time series. arXiv preprint arXiv:1904.07383.
Neuberger, A., 2012. Realized skewness. Rev. Financ. Stud. 25, 3423–3455.
Panagakis, Y., Kossaifi, J., Chrysos, G.G., Oldfield, J., Nicolaou, M.A., Anandkumar, A., Zafeiriou, S., 2021. Tensor methods in computer vision and deep learning. Proc. IEEE 109 (5), 863–890.
Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97 (460), 1167–1179.
Tomasi, G., Bro, R., 2005. PARAFAC and missing values. Chemometr. Intell. Lab. Syst. 75 (2), 163–180.
Trapani, L., 2018. A randomized sequential procedure to determine the number of factors. J. Amer. Statist. Assoc. 113 (523), 1341–1349.
Wang, D., Liu, X., Chen, R., 2019. Factor models for matrix-valued high-dimensional time series. J. Econometrics 208 (1), 231–248.
Yu, L., He, Y., Kong, X., Zhang, X., 2022. Projected estimation for large-dimensional matrix factor models. J. Econometrics 229, 201–217.
Yu, Y., Wang, T., Samworth, R.J., 2015. A useful variant of the Davis-Kahan theorem for statisticians. Biometrika 102 (2), 315–323.
Zhang, X., Li, G., Liu, C.C., Guo, J., 2025a. Tucker tensor factor models: matricization and mode-wise PCA estimation. Sci. China Math. 1–52.
Zhang, X., Liu, C.C., Guo, J., Yuen, K., Welsh, A.H., 2025b. Modeling and learning on high-dimensional matrix-variate sequences. J. Amer. Statist. Assoc. 120 (549), 419–434.
Zhou, H., Li, L., Zhu, H., 2013. Tensor regression with applications in neuroimaging data analysis. J. Amer. Statist. Assoc. 108 (502), 540–552.