# High-Dimensional Time Series Analysis: Imputation and Testing Kronecker Product Structure on Tensor Factor Models, Main Effect Factor Models, and Spatial Autoregressive Models

Zetai Cen

Department of Statistics

London School of Economics and Political Science

This thesis is submitted for the degree of

*Doctor of Philosophy*

June 2025

江上人常驻，月下青鸟迟。

To my loving family.

# Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).



I confirm that Chapters 3–5 are jointly co-authored with my supervisor, Prof. Clifford Lam. Chapter 6 is jointly co-authored with Dr. Yudong Chen (University of Warwick) and Prof. Clifford Lam. A version of Chapter 3 has been accepted for publication in *Journal of Econometrics*; see the online version as Cen and Lam (2025b). Chapters 4, 5, and 6 have been submitted for publication in adapted versions of Cen and Lam (2025a), Lam and Cen (2024), and Cen et al. (2024), respectively.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I declare that my thesis consists of 43,603 words.

Zetai Cen

June 2025

# Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Professor Clifford Lam, for his heartfelt assistance and support throughout this journey. My research would have been impossible without his dedicated involvement – where my understanding faltered, his wisdom built bridges. Though never a gifted student, I remain profoundly indebted to my supervisor for his patient guidance that illuminated paths in statistics when I stumbled in darkness. I would also like to thank my co-supervisor, Professor Wicher Bergsma, for his consistent encouragement and invaluable advice. Although our conversations often unfolded amidst coffee breaks, they never failed to spark moments of enlightenment.

One could not wish for a more supportive collaborator than Yudong Chen. I am truly thankful to him for his helpful suggestions and sincere devotion. I certainly benefit from our numerous discussions, which are necessary to make our joint work possible. Such an experience would undoubtedly be a cornerstone in my future journey. My first acquaintance with Shakeel Gavioli-Akilagun dates back to my undergraduate study when he delivered my time series analysis seminar. His countless and selfless support has been influential to every aspect of my study. To both Yudong and Shakeel, I once again express my appreciation, which is too emotionally sophisticated to articulate.

It is extremely fortunate for me to take part in the series of weekly meetings led by Tengyao Wang and Yining Chen, by which I have been continuously inspired. This experience constitutes a meaningful part of my journey, and I sincerely thank every member in the meeting. Since my undergraduate studies at LSE, I have been deeply influenced by the Department of Statistics. I hereby express my gratitude to Gelly Mitrodima for organising an event that initiated me to open up my research on factor models and to Qiwei Yao for his recognition of the corresponding outcome and immense support since then. I would like to extend my appreciation to Irini Moustaki and Angelos Dassios for supporting me as my undergraduate and postgraduate mentors, to Xinghao Qiao for his guidance on reviewing literature, to Chengchun Shi for his trust in me being a journal referee, and to James Abdey, Piotr Fryzlewicz, and Yufei Zhang for offering me teaching opportunities.

It is also my fortune to have all the heart-warming faculties and colleagues in the department. Special thanks to every professional services staff member for their administrative support. It would be impossible not to mention my cohorts, Zezhun Chen, Weilin Chen, Wenyu Cheng, Qin Fang, Tao Ma, Sakina Hansen, Sixing Hao, Ziqing Ho, Liyuan Hu, Camilo Cárdenas Hurtado, Sahoko Ishida, Gabriel Wallin, Kerry, Sze Ming Lee, Mingwei Lin, Kaixin Liu, Xinyi Liu, Olga Lupova, Seyedpouya Mirrezaeiroudaki, Motonori Oka, Alexandros Pavlis, Zhichao Shen, Pietro Maria Sparago, Di Su, Pingfan Su, Mohammad Fadil Sumun, Ke Sun, Yutong Wang, Xiaoyi Wen, Lin Yang, Shuhan Yang, Xuzhi Yang, Yilin Zhang, Zeng Zhang, Kaifang Zhou, Jin Zhu, Xianghe Zhu, and many others.

Science is passed down and developed from generation to generation. This was never concrete to me until the IMS Annual Meeting in 2022. I thank the forever touching speech of a gentleman, and I am grateful to all other lovely scholars for their books, encouragement, and sincere advice.

My gratitude is extended to all my friends for their constant companionship, especially to Jiahao Cui, Zhenjing He, Ying Li, Yingzhi Xu, Tianwen Wang, Mengqi Zhu, and Zifeng Xu for our friendship in London, and to Junyu Lu for his guidance in Asian histories and material cultures, from which I find a peace of mind.

Last but not least, I would like to thank my family for being my steadfast anchor over all these years. Deepest appreciation to my dearest parents and my cat, Abo, for all their understanding, encouragement, and unconditional love, which I have to spend the rest of my life to repay.

# Abstract


High-dimensional time series are increasingly ubiquitous, which leads to an urgent need for statistical methodologies correspondingly. The emergence of tensor time series, where data are arranged as general tensors (e.g. vectors, matrices) at each timestamp, poses new challenges to researchers and practitioners. This thesis sheds light on time series analysis from factor modelling to spatiotemporal analysis.

First, we explore how to estimate the factor structure in a tensor factor model with missing data and weak factors. With a rank estimator proposed, we introduce an imputation procedure by leveraging all estimators and discuss how to perform practical inference. We elaborate on the performance of our method with two real data examples on portfolio returns and national economic indicators, respectively.

We also attempt to answer a fundamental question on tensor factor modelling: can we test if a factor structure is violated on a given tensor time series while preserved on the flattened series? Generally put, we are interested to understand whether the factor structure is mode-related or not. We formulate the testing problem and provide a residual test with theoretical guarantees, followed by extensive data examples.

For matrix time series, we design a factor model with time-varying main effects in addition to a common component to disentangle row and column information of the observed matrix. It assumes a more general structure than the prevalent matrix factor model with Tucker decomposition in the common component governing only the "joint" effect. We establish theories for statistical inference and propose a test on the necessity of our model. We apply our model to study a set of taxi traffic data and discover an "hour" effect within.

Lastly, we contribute to the field of spatial econometrics by presenting a spatial autoregressive model with time-varying spatial weights, featuring the spill-over effects among cross-sectional units contemporaneously in the observed vector time series. We circumvent the difficulty of selecting spatial weight matrices by penalised estimation. A set of industrial profits is analysed through our approach.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Large dimensional panel data is easier to obtain than ever thanks to quickly evolving computational technology and more diverse platforms. Due to the nature of time series data, methodologies particularly designed for independent data could give misleading conclusions. This has motivated an active field of research to understand the dynamics of observed time series in the presence of temporal and cross-sectional dependence.

One prominent approach is factor modelling which attributes the dependence within data sets to only a few factors, i.e., the observed time series is assumed to be driven by latent factors with a much smaller dimension (e.g. Chamberlain and Rothschild, 1983; Stock and Watson, 1998; Bai and Ng, 2002; Onatski, 2009; Lam et al., 2011; Zhang et al., 2024a). This is often promising due to large data dimensions or the nature of data per se. While factor analyses have been well developed for vector-valued time series, researchers nowadays have opened up to studying generally multi-way time series, namely tensor-valued time series, which can be seen as a generalisation of vector time series. Through exploiting the tensor structure of observed data, instead of stacking them into vectors, more in-depth analyses can be available to retrieve mode-wise information. For instance, Chen et al. (2022a) studies the monthly import-export volume of products among different countries by proposing a tensor factor model based on Tucker decomposition; Lettau (2022) applies both CP and Tucker decompositions to study mutual fund characteristics; Guan (2024) adapts CP decomposition in immunological and clinical studies; Liu et al. (2022) applies tensor PCA to study the spatiotemporal patterns in human brains, to name but a few works.

Another useful methodology among others is spatiotemporal modelling, which reads cross-sectional dependence as different kinds of spatial relations (e.g. Anselin, 1988; LeSage and Pace, 2009). In other words, it is desirable to learn the interaction between individual units and their neighbourhoods. One general form to describe spatial dependence is spatial autoregressive models. A large body of literature focuses on how to specify appropriate spatial weight matrices, which can be difficult in practice. See e.g., the development in Sun (2016), where a spatial autoregressive model with nonparametric spatial weights is leveraged to analyse national economic growths; Lam and Souza (2020) and Higgins and Martellosio (2023) propose to use a linear combination of multiple spatial weight matrices with constant coefficients.

This thesis contains two parts, respectively addressing the two fields discussed above. The first part (Chapter 3–5) consists of three projects spanning important areas of factor modelling for matrix- and general tensor-valued time series, whereas the second (Chapter 6) concerns a project on spatial autoregressive models. In particular, the contribution of those projects on factor modelling can be viewed from three different aspects, as summarised below.

1. Application – Chapter 3 develops a missing value imputation scheme based on tensor factor models. The main contributions are twofold. To the best of our knowledge, it is new to impute general tensor time series while all existing literatures focus on vector time series. Moreover, we propose a consistent estimator on the number of factors under missingness, which is also new to the community of factor models for missing data.

2. Testing – Chapter 4 introduces a tensor reshape operator and tests the Kronecker product structure in the loading matrix for factor models. It is a first in the literature to formulate a (series of) testing problems on the validity of Tucker-decomposition tensor factor models, which is arguably the most fundamental problem for higher-order tensor factor models. We hope our dedication enlightens a mindset to understand tensor factor models.

3. Modelling – Chapter 5 proposes a matrix factor model with time-varying main effects to greatly enhance interpretability. Although certain efforts have been made on matrix/tensor factor modelling, it still remains challenging to effectively utilise the information along each mode. Our presented model aims to address this.

On our project related to spatiotemporal analysis, Chapter 6 introduces a spatial autoregressive model with time-varying spatial correlation coefficients. It incorporates multiple spatial weight matrices through their linear combinations with varying coefficients and hence encompasses existing literature on changepoint/threshold spatial autoregressive models as special cases. With adaptive LASSO estimators, the flexibility of our model is further enhanced.

From the above overview, we may conclude the main theme of our developed models as "nested models" in the sense that traditional models are nested in our proposed frameworks. The rest of this thesis is organised as follows. Chapter 2 introduces the (tensor) notations, which are solidified in subsequent surveys on the existing literature of factor models. We review recent works on spatial autoregressive models separately in Chapter 6. Chapter 3 discusses our imputation procedure for a tensor time series with very mild observational patterns. In Chapter 4, we design a test on the existence of Kronecker product structure for general tensor factor models, where we formally define the concept of factor models with Kronecker product structure and pinpoint the equivalence of Tucker-decomposition tensor factor models under a reshape operator along any selection of modes. Chapter 5 proposes a matrix-valued (i.e., order-2 tensor) factor model with time-varying main effects, and more importantly, we lay down a test on the necessity of our model. Finally, Chapter 6 develops a framework of spatial autoregressive models that can be versatile in variable selection and change point detection.

# Chapter 2

# Literature Review

## 2.1 Notations and Tensor Basics

Throughout this thesis and unless otherwise specified, we use the lower-case or capital letter, bold lower-case letter, bold capital letter, and calligraphic letter, i.e., $x$ or $X$, $\mathbf{x}$, $\mathbf{X}$, $\mathcal{X}$, to denote a scalar, a vector, a matrix, and a tensor (introduced later), respectively. We also use $x_i, X_{i,j}, \mathbf{X}_{i\cdot}, \mathbf{X}_{\cdot i}$ to denote, respectively, the $i$-th element of $\mathbf{x}$, the $(i,j)$-th element of $\mathbf{X}$, the $i$-th row (as a column vector) of $\mathbf{X}$, and the $i$-th column of $\mathbf{X}$. We use $a \asymp b$ to denote $a = O(b)$ and $b = O(a)$, while $a \asymp_P b$ to denote $a = O_P(b)$ and $b = O_P(a)$. Hereafter, given a positive integer $m$, define $[m] := \{1, \ldots, m\}$. We use $\mathbf{1}_m$ to denote a vector of ones of length $m$, $\mathbf{0}$ a vector of conformable length, and $\mathbf{I}_m$ an $m \times m$ identity matrix. The $i$-th largest eigenvalue (resp. singular value) of $\mathbf{X}$ is denoted by $\lambda_i(\mathbf{X})$ (resp. $\sigma_i(\mathbf{X})$). We use $\mathbf{X}'$ (resp. $\mathbf{x}'$) to denote the transpose of $\mathbf{X}$ (resp. $\mathbf{x}$), and $\text{diag}(\mathbf{X})$ to denote a diagonal matrix with the diagonal elements of $\mathbf{X}$, while $\text{diag}(\{x_1, \ldots, x_n\})$ represents the diagonal matrix with $\{x_1, \ldots, x_n\}$ on the diagonal. A random variable $X$ is sub-Gaussian with variance proxy $\sigma^2$, denoted as $X \sim \text{subG}(\sigma^2)$, if $\mathbb{E}[\exp\{s(X - \mathbb{E}[X])\}] \le \exp(s^2\lambda^2/2)$ for all $s \in \mathbb{R}$. A random variable $X$ is sub-exponential with parameter $\lambda$, denoted as $X \sim \text{subE}(\lambda)$, if $\mathbb{E}[\exp\{s(X - \mathbb{E}[X])\}] \le \exp(s^2\lambda^2/2)$ for all $|s| \le 1/\lambda$.

**Norm notations**: Sets are also denoted by calligraphic letters. For a given set, we denote by $|\cdot|$ its cardinality. We use $\|\cdot\|$ to denote the spectral norm of a matrix or the $L_2$ norm of a vector, and $\|\cdot\|_F$ to denote the Frobenius norm of a matrix or a tensor. We use $\|\cdot\|_{\max}$ to denote the maximum absolute value of the elements in a vector, a matrix or a tensor. The notations $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the $L_1$- and $L_\infty$-norm of a matrix respectively, defined by $\|\mathbf{X}\|_1 := \max_j \sum_i |X_{i,j}|$ and $\|\mathbf{X}\|_\infty := \max_i \sum_j |X_{i,j}|$. For $q > 0$, we define the $L_q$-norm of a given real-valued random variable $x$ as $\|x\|_q := (\mathbb{E}|x|^q)^{1/q}$. Without loss of generality, we always assume the eigenvalues of a matrix are arranged by descending orders, and so are their corresponding eigenvectors.

**Tensor-related notations**: For the rest of this section, we briefly introduce the notations and operations for tensor data, which will be sufficient for this thesis. For more details on tensor manipulations, readers are referred to Kolda and Bader (2009). To begin with, a vector and a matrix are respectively order-1 and order-2 tensors. In general, an *order-$K$ tensor* is a multidimensional array with $K$ ways, denoted by $\mathcal{X} = (X_{i_1,\ldots,i_K}) \in \mathbb{R}^{I_1\times\cdots\times I_K}$. Its $k$-th way is termed as *mode-$k$*, $I_k$ as the *mode-$k$ dimension*, and a column vector $(X_{i_1,\ldots,i_{k-1},i,i_{k+1},\ldots,i_K})_{i\in[I_k]}$ as one of its *mode-$k$ fibers*. We denote by $\text{mat}_k(\mathcal{X}) \in \mathbb{R}^{I_k\times I_{\text{-}k}}$ (or sometimes $\mathbf{X}_{(k)}$, with $I_{\text{-}k} := (\prod_{j=1}^K I_j)/I_k$) the *mode-$k$ unfolding/matricization* of $\mathcal{X}$, defined by placing all mode-$k$ fibers into a matrix, see Figure 2.1 for an illustration (figure from Tao et al. (2019)). We use $\mathbf{vec}(\cdot)$ to denote the vectorisation of a matrix or the vectorisation of the mode-1 unfolding of a tensor. The *refolding/tensorisation* of a vector $\mathbf{x} \in \mathbb{R}^{I_1\cdots I_K}$ on $\{I_1,\ldots,I_K\}$ is defined to be an order-$K$ tensor $\text{FOLD}(\mathbf{x}, \{I_1,\ldots,I_K\}) \in \mathbb{R}^{I_1\times\cdots\times I_K}$ such that $\mathbf{x} = \mathbf{vec}\{\text{FOLD}(\mathbf{a}, \{I_1,\ldots,I_K\})\}$. The *refolding/tensorisation* of a matrix $\mathbf{X} \in \mathbb{R}^{I_k\times I_{\text{-}k}}$ on $\{I_1,\ldots,I_K\}$ along mode-$k$ is defined to be $\text{FOLD}_k(\mathbf{X}, \{I_1,\ldots,I_K\}) \in \mathbb{R}^{I_1\times\cdots\times I_K}$ such that $\mathbf{X} = \text{mat}_k\{\text{FOLD}_k(\mathbf{X}, \{I_1,\ldots,I_K\})\}$. The $\text{RESHAPE}(\cdot,\cdot)$ operator is only involved in Chapter 4 and hence introduced in Section 4.2.

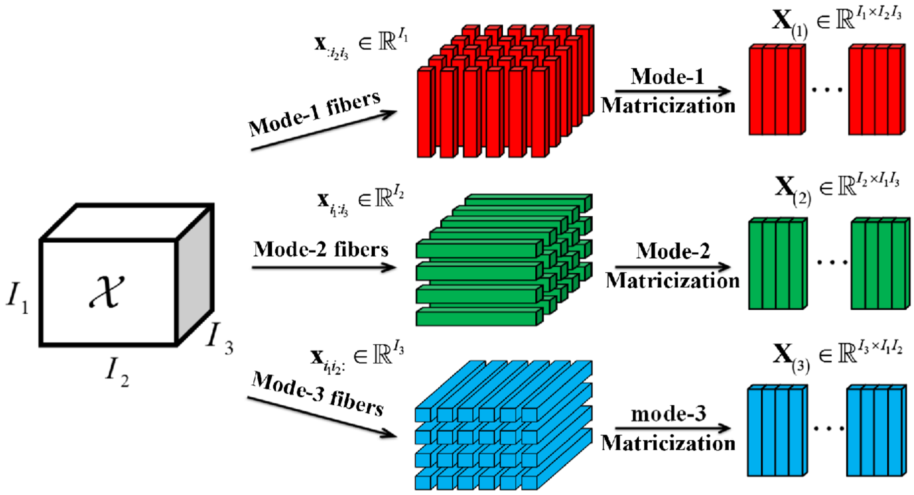


Figure 2.1: Illustration of the mode-$k$ fibers and its corresponding unfolding matrix.

**Product notations**: We use $*$ to denote the Hadamard product (i.e., element-wise product), $\circ$ the tensor outer product defined between an order-$K$ tensor $\mathcal{X}$ and an order-$L$ tensor $\mathcal{Y}$ as an order-$(K+L)$ tensor such that $(\mathcal{X}\circ\mathcal{Y})_{i_1,\ldots,i_K,j_1,\ldots,j_L} := X_{i_1,\ldots,i_K} Y_{j_1,\ldots,j_L}$, $\mathcal{X}\times_k \mathbf{A}$ the mode-$k$ product defined between a tensor $\mathcal{X}$ and a conformable matrix $\mathbf{A}$ as

$$\text{mat}_k(\mathcal{X}\times_k \mathbf{A}) := \mathbf{A}\,\text{mat}_k(\mathcal{X}),$$

$\odot$ the Khatri–Rao product defined between $\mathbf{X} \in \mathbb{R}^{a\times b}$ and $\mathbf{Y} \in \mathbb{R}^{c\times b}$ as a $ac\times b$ matrix

$$\mathbf{X}\odot\mathbf{Y} := \big(\mathbf{X}_{\cdot 1}\otimes\mathbf{Y}_{\cdot 1},\ldots,\mathbf{X}_{\cdot b}\otimes\mathbf{Y}_{\cdot b}\big),$$

and $\otimes$ the Kronecker product defined between $\mathbf{X} \in \mathbb{R}^{a\times b}$ and $\mathbf{Y} \in \mathbb{R}^{c\times d}$ as a $ac \times bd$ matrix

$$\mathbf{X} \otimes \mathbf{Y} := \begin{pmatrix} X_{1,1}\mathbf{Y} & \dots & X_{1,b}\mathbf{Y} \\ \vdots & \ddots & \vdots \\ X_{a,1}\mathbf{Y} & \dots & X_{a,b}\mathbf{Y} \end{pmatrix}.$$

By convention, the total Kronecker product for an index set is computed in descending order.

**Useful properties related to products**: All the above products are bilinear and associative. For the mode-$k$ product, distinct mode products are commutative in the sense that for $n \neq m$,

$$\mathcal{X} \times_n \mathbf{A} \times_m \mathbf{B} = \mathcal{X} \times_m \mathbf{B} \times_n \mathbf{A},$$

where we do not distinguish the special case when the matrix $\mathbf{A}$ or $\mathbf{B}$ appears as a vector, as long as during the mode products, we maintain the order of $\mathcal{X}$ even if it is degenerate, i.e., some mode dimensions are 1. For identical mode products,

$$\mathcal{X} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{X} \times_n (\mathbf{BA}).$$

Over the Kronecker product, transpose and inverse are distributive. More importantly, it holds for any conformable matrices and tensors that

$$\textbf{vec}(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A})\textbf{vec}(\mathbf{X}), \tag{2.1}$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{X} \otimes \mathbf{Y}) = (\mathbf{AX}) \otimes (\mathbf{BY}), \tag{2.2}$$

$$\text{mat}_k(\mathcal{X} \times_{j=1}^{K} \mathbf{A}_j) = \mathbf{A}_k \, \text{mat}_k(\mathcal{X})(\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)'. \tag{2.3}$$

## 2.2 Preliminaries: Tensor Decompositions

On the aspect of storage and computation, useful structures are often imposed on the data of interest. Tensor, a mathematical object, has been developed since its initial introduction in the 19th century as a representation and computational tool; see Favier (2019) for the historical background. Similar to matrices, general order tensors can be represented by elements in certain forms, namely different decompositions (also coined in some literature as "factorisations" or "formats"). Throughout the development, there are two fundamental decompositions, which are the *CP decomposition* (e.g. Hitchcock, 1927; Carroll and Chang, 1970; Harshman, 1970; Kiers, 2000) and the *Tucker decomposition* (e.g. Tucker, 1963; De Lathauwer et al., 2000).

Although CP decomposition dates back as early as in Hitchcock (1927), we first focus on Tucker decomposition, which is the form of factor models used in Chapter 3–5. Simply put, Tucker decomposition is a generalisation of matrix singular value decomposition (SVD) to order-3 tensors when originally proposed by Tucker (1963), and further to general order tensors by Wansbeek et al. (1986) using vectorisation and De Lathauwer et al. (2000) using mode products. Therefore, Tucker decomposition is also known as higher-order SVD (De Lathauwer et al., 2000). Formally, Tucker decomposition writes an order-$K$ tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_K}$ as

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}_1 \times_2 \cdots \times_K \mathbf{A}_K, \tag{2.4}$$

where $\mathcal{G} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is the core factor with ranks $\{r_k\}_{k \in [K]}$ and $\mathbf{A}_k \in \mathbb{R}^{I_k \times r_k}$ ($k \in [K]$) are the factor matrices. Notice first (2.4) is trivial if $\mathcal{G} = \mathcal{X}$ and every $\mathbf{A}_k = \mathbf{I}_{I_k}$. On the other hand, the $\prod_{k=1}^K I_k$ entries of $\mathcal{X}$ can be represented by $(\prod_{k=1}^K r_k + \sum_{k=1}^K I_k r_k)$ entries according to (2.4). This effectively reduces the amount of data when the tensor order is high and $r_k \ll I_k$. Hence it is often, if not always, assumed in the context of factor analysis that $\mathcal{G}$'s size is much smaller than $\mathcal{X}$'s, so that $\mathcal{X}$ admits a low-rank structure. The study of Tucker decomposition remains active in modern research. For instance, Zhang and Xia (2018) studies the optimality for tensor SVD, whereas Zhang (2019) proposes a "Cross" measurement scheme to efficiently recover tensor data.

The element-wise representation for a tensor decomposition is generally helpful to bridge different products and hence tensor decompositions. Such a representation for (2.4) is

$$X_{i_1,\ldots,i_K} = \sum_{j_1=1}^{r_1} \cdots \sum_{j_K=1}^{r_K} G_{j_1,\ldots,j_K} A_{1,i_1,j_1} \ldots A_{K,i_K,j_K}, \tag{2.5}$$

which used the fact that the $(i_1, \ldots, i_K)$-th element of $\mathcal{X}$ corresponds to the element on the $i_k$-th row of $\text{mat}_k(\mathcal{X})$ for all $k \in [K]$, followed by a simple induction argument. Immediately from (2.5), Tucker decomposition can be equivalently represented by outer products that

$$\mathcal{X} = \sum_{j_1=1}^{r_1} \cdots \sum_{j_K=1}^{r_K} G_{j_1,\ldots,j_K} (\mathbf{A}_{1,\cdot j_1} \circ \cdots \circ \mathbf{A}_{K,\cdot j_K}). \tag{2.6}$$

Suppose $\mathcal{G}$ in (2.4) is (super)diagonal, i.e., $r \equiv r_1 = \cdots = r_K$ and $G_{j_1,\ldots,j_K} \neq 0$ only if $j \equiv j_1 = \cdots = j_K$. This special case of Tucker decomposition boils down to the CP decomposition. Historically, CP decomposition has many names (even "CP" can be read as "CANDECOMP/PARAFAC" or "Canonical Polyadic"), see Table 3.1 in Kolda and Bader (2009) for some of them; we simply use "CP decomposition" in this thesis. It is convenient to represent CP decomposition directly by outer products such that for an order-$K$ tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_K}$,

$$\mathcal{X} = \sum_{j=1}^{r} G_{j,\ldots,j} (\mathbf{A}_{1,\cdot j} \circ \cdots \circ \mathbf{A}_{K,\cdot j}), \tag{2.7}$$

which indeed coincides with a special case of (2.6).

In essence, CP decomposition can be understood as a further data-sparse structure on top of Tucker decomposition. To see this, consider Tucker decomposition and apply (2.1) and (2.3) on the mode-1 unfolding of (2.4), then we arrive at

$$\mathbf{vec}(\mathcal{X}) = (\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_1)\mathbf{vec}(\mathcal{G}),$$

where $(\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_1)$ has $\prod_{k=1}^K r_k$ columns in general. Under CP decomposition in (2.7), the above core factor $\mathcal{G}$ is diagonal and hence we have

$$\begin{aligned}\mathbf{vec}(\mathcal{X}) &= \big(\mathbf{A}_{K,\cdot 1} \otimes \cdots \otimes \mathbf{A}_{1,\cdot 1}, \ldots, \mathbf{A}_{K,\cdot r} \otimes \cdots \otimes \mathbf{A}_{1,\cdot r}\big) \begin{pmatrix} G_{1,\ldots,1} \\ \vdots \\ G_{r,\ldots,r} \end{pmatrix} \\ &= (\mathbf{A}_K \odot \cdots \odot \mathbf{A}_1) \begin{pmatrix} G_{1,\ldots,1} \\ \vdots \\ G_{r,\ldots,r} \end{pmatrix},\end{aligned}$$

where $(\mathbf{A}_K \odot \cdots \odot \mathbf{A}_1)$ has $r$ columns only, regardless of the order of $\mathcal{X}$. The discussion above also suggests that the mode interaction within Tucker decomposition is governed through Kronecker products, while that within CP decomposition is through Khatri–Rao products. It is worth to point out that this specific structure of CP decomposition is less of our interest for dimension reduction when the tensor order and the core factor rank are assumed fixed. One scenario where Tucker decomposition is particularly considered is the testing problem in Chapter 4. We would restrict the test on tensor factor models in the form of Tucker decomposition only, although the test can also be directly extended to the form of CP decomposition. In a nutshell, Tucker decomposition features a general structure for dimension reduction, while CP decomposition holds more stringently but allows us to potentially further exploit its structure.

## 2.3 Factor Modelling

### 2.3.1 Factor models

Due to the surge of big data, dependence across measurements is often inevitable and hence ubiquitous. In the presence of multivariate or high-dimensional data, low-rank structure is often seen, and ignoring it could lead to inefficient use of data and inaccurate conclusions. Therefore, people have endeavoured to exploit such structures with either known or latent factors through *factor analyses* which, stemming from the earliest work by Spearman (1904) to investigate psychical activities, have been useful tools in multivariate analyses with widespread applications in psychology (e.g. Spearman, 1927; Bartlett, 1950; McCrae and John, 1992), biology (e.g. Hirzel et al., 2002; Hochreiter et al., 2006), economics and finance (e.g. Chamberlain and Rothschild, 1983; Fama and French, 1993; Stock and Watson, 2002a,b), etc.

Throughout this thesis, we focus on (latent) *factor models* which we define as the factor analyses for time series data. In factor models, there are two frameworks whose developments are relatively independent of each other until recent decades. The first framework branches from the literature of financial econometrics and was motivated by the study of pricing models CAPM (Sharpe, 1964) and APT (Ross, 1972). Tracing back, Ross (1976) first introduced the *exact/strict factor model* as an alternative pricing model, which was later generalised by Chamberlain and Rothschild (1983) proposing the *approximate factor model*. See also e.g. Bai and Ng (2002) and Bai (2003). To illustrate, suppose $\mathbf{y}_t \in \mathbb{R}^d$ is observed at each timestamp $t \in [T]$, then a (vector) factor model assumes that the time series admits a decomposition

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}\mathbf{f}_t + \mathbf{e}_t, \tag{2.8}$$

where $\boldsymbol{\mu}$ is a vector accounting for the mean of $\mathbf{y}_t$, $\mathbf{A} \in \mathbb{R}^{d\times r}$ is the *factor loading* with the *number of factors* (or rank) $r \ll d$, $\mathbf{f}_t \in \mathbb{R}^r$ is the zero-mean *core factor* (or factor score), and $\mathbf{e}_t$ is the idiosyncratic noise. We refer to $\mathbf{A}\mathbf{f}_t$ the *common component* of the factor model. Generally, the time series can be demeaned by its sample mean, so that $\boldsymbol{\mu} = \mathbf{0}$ which is assumed for the remaining discussion on vector factor models. The model (2.8) is characterised by its noise covariance $\boldsymbol{\Sigma}_e := \mathbb{E}[\mathbf{e}_t\mathbf{e}_t']$. An exact factor model assumes $\boldsymbol{\Sigma}_e$ to be diagonal, while an approximate factor model allows for weak cross-sectional dependence in noise in the sense that $\boldsymbol{\Sigma}_e$ has uniformly bounded entries, i.e., $\boldsymbol{\Sigma}_e$ can have nonzero sparsely on its off-diagonal.

The main idea in a factor model is that there exists a small number of factors driving all dynamics of the series of interest. Notice the decomposition (2.8) always exists if we relax $r \ll d$, but it is only useful with a low rank $r$ to reduce the dimensionality. As a simple example, consider the problem of covariance estimation given (2.8) being an exact factor model. Assume for simplicity that any entry in $\mathbf{f}_t$ is uncorrelated with $\mathbf{e}_t$, then we can read $\boldsymbol{\Sigma}_y := \mathbb{E}[\mathbf{y}_t\mathbf{y}_t']$ as

$$\boldsymbol{\Sigma}_y = \mathbf{A}\boldsymbol{\Sigma}_f\mathbf{A}' + \boldsymbol{\Sigma}_e, \tag{2.9}$$

where $\boldsymbol{\Sigma}_f := \mathbb{E}[\mathbf{f}_t\mathbf{f}_t']$. Now we only need to estimate $[dr + r(r+1)/2 + d]$ parameters, rather than $d(d+1)/2$ parameters in $\boldsymbol{\Sigma}_y$, hence achieving dimension reduction. For a more general treatment under approximate factor models, see e.g. Fan et al. (2013).

The second framework branches from the statistics literature, where researchers sought dimension reduction in vector time series on both time and frequency domains. See Peña and Box (1987), Pan and Yao (2008), Lam et al. (2011), Lam and Yao (2012), and the references within. The difference between the two frameworks mainly lies in the temporal dependence among the idiosyncratic noise. The noise can be serially correlated in the first framework, but set as white noise in the second with the core factor accounting for all serial correlation. With such modification, the second framework allows for a greater cross-sectional dependence in noise by relaxing the uniform boundedness of $\mathbf{\Sigma}_e$. Another significant consequence of the second framework is to allow for naturally leveraging nonzero-lag autocovariance matrices in parameter estimation (e.g. Lam et al., 2011; Lam and Yao, 2012), see (2.12) in Section 2.3.2.

**Remark 2.1** *Both aforementioned frameworks are represented by static loading matrices as in* (2.8)*, and hence they are all referred to as* static factor models*. On the other hand, Geweke (1977) proposes* dynamic factor models *by considering dynamic loading matrices in the form of* $\mathbf{A}(L) := \sum_{w=0}^{\infty} \mathbf{A}_w L^w$*, where* $L$ *is the lag operator. There is a large body of literature on this topic (e.g. Forni et al., 2000; Forni and Lippi, 2001; Stock and Watson, 2005; Doz et al., 2011; Hallin and Lippi, 2013), see Barigozzi and Hallin (2024) for an overview.*

### 2.3.2 Estimation: factor loading

For either framework in Section 2.3.1, (2.8) (with $\boldsymbol{\mu} = \mathbf{0}$) can be equivalently written in a matrix form $\mathbf{Y} = \mathbf{AF} + \mathbf{E}$, where $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$, and $\mathbf{F}$ and $\mathbf{E}$ are similarly defined. Even with an identification (or sometimes normalisation) on the loading matrix that $\mathbf{A}'\mathbf{A}/d = \mathbf{I}_r$, $\mathbf{A}$ and $\mathbf{F}$ are identifiable only up to some rotation matrix $\mathbf{H}$ due to the indeterminacy incurred by the multiplication $\mathbf{AF} = (\mathbf{AH})(\mathbf{H}'\mathbf{F})$. However, the common component and the factor loading space $\mathcal{M}(\mathbf{A})$ can be identified, where $\mathcal{M}(\mathbf{A})$ denotes the column space of $\mathbf{A}$.

The above identification can be taken advantage of to facilitate estimation of the loading matrix which, once obtained, leads to straightforward estimation of the core factor. Recall that the number of factors $r$ is also unknown, and hence there are two main steps in estimating the factor structure in (2.8) – factor loading estimation and rank estimation.

First, let $r$ be specified (known a priori or estimated). Consider (2.8) as an approximate factor model, we may use asymptotic principal components to estimate the factor loading and core factor at the same time (e.g. Connor and Korajczyk, 1986; Stock and Watson, 1998; Bai and Ng, 2002), by solving a least squares problem

$$V(r) = \min_{\mathbf{F},\mathbf{A}} \frac{1}{Td} \|\mathbf{Y} - \mathbf{AF}\|_F^2 \quad \text{subj. to} \quad \mathbf{A}'\mathbf{A}/d = \mathbf{I}_r. \tag{2.10}$$

Concentrating out the core factor in (2.10), we can show the optimisation problem is equivalent to maximise $\text{tr}\{\mathbf{A}'(\mathbf{YY}')\mathbf{A}\}$, which gives a set of *least-squares-type* estimators (i.e. solutions of some least squares problems)

$$\widehat{\mathbf{A}} = \sqrt{d}\big[\boldsymbol{\gamma}_1(\widehat{\mathbf{\Sigma}}_y), \dots, \boldsymbol{\gamma}_r(\widehat{\mathbf{\Sigma}}_y)\big], \quad \widehat{\mathbf{F}} = \widehat{\mathbf{A}}'\mathbf{Y}/d, \tag{2.11}$$

where $\widehat{\mathbf{\Sigma}}_y := \mathbf{Y}\mathbf{Y}'/T$ is the sample version of $\mathbf{\Sigma}_y$ in (2.9), and $\boldsymbol{\gamma}_j(\cdot)$ denotes the eigenvector corresponding to the $j$-th largest eigenvalue of a matrix. The estimation is closely related to eigenanalysis, as discussed at the end of this subsection. The least-squares-type estimators are also adapted under different settings, see for instance, Onatski (2012) and Bai and Ng (2023) develop the estimators for weak factor models.

As discussed in Section 2.3.1, if the idiosyncratic noise is a white noise process, the factor structure can be estimated from a different perspective (e.g. Lam et al., 2011). With (2.8) ($\boldsymbol{\mu} = \mathbf{0}$) and notations therein, given a pre-specified positive integer $h_0$, we define

$$\widehat{\mathbf{\Sigma}}_y(h) = \frac{1}{T-h}\sum_{t=1}^{T-h}\mathbf{y}_{t+h}\mathbf{y}_t', \quad \widehat{\mathbf{M}} = \sum_{h=1}^{h_0}\widehat{\mathbf{\Sigma}}_y(h)\widehat{\mathbf{\Sigma}}_y(h)'. \tag{2.12}$$

The matrix $\widehat{\mathbf{\Sigma}}_y(h)$ is the sample autocovariance matrix at lag $h$, and $\widehat{\mathbf{M}}$ can be regarded as a matrix accumulating the autocovariance information (from lag 1) up to lag $h_0$. A factor loading estimator (and hence the corresponding core factor estimator) is then feasible by the principal component analysis (PCA) and constructed similarly as in (2.11), with $\widehat{\mathbf{\Sigma}}_y$ replaced by $\widehat{\mathbf{M}}$. In essence, we now leverage the autocovariance structure rather than the covariance structure in (2.11). There are also extensions using both covariance and autocovariance structures, such as Zhang et al. (2024a). One shortcoming of using autocovariance matrices (at nonzero lags) is the difficulty in constructing asymptotic distributions for the loading or core factor estimators, which is less challenging for (2.11); see Bai (2003), Bai and Ng (2023), etc.

Estimators obtained directly from PCA on some covariance-like matrices are referred to as the *PCA-type* estimators. Notably, if we replace $\widehat{\mathbf{M}}$ in (2.12) by $\widehat{\mathbf{\Sigma}}_y$, the two types of estimators are equivalent (sometimes up to a scale transformation). They remain so until we consider factor models for higher-order tensor time series; see Section 2.3.4. Recent developments include, for instance, Barigozzi and Cho (2020) where a PCA-type estimator scaled by the eigenvalues of the sample covariance matrix is proposed to remedy the potentially over-estimated number of factors, and He et al. (2022a) where a PCA-type estimator based on the spatial Kendall's tau matrix is proposed to address heavy-tailedness under the elliptical distribution framework.

### 2.3.3 Estimation: number of factors

Except for performing hypothesis tests (e.g. Pan and Yao, 2008; Onatski, 2009) or incorporating both eigenvalues and eigenvectors (e.g. Freyaldenhoven, 2022, to discover local factors), literatures on determining the number of factors use eigenvalue information in mainly three forms, beyond which there are e.g. bootstrap-based estimators (Yu et al., 2024b). Extended from the discussion in Section 4.1 in Freyaldenhoven (2022), we briefly summarise the main types of estimators in the following, under the setup of (2.8) so that we only focus on static factor models for vector time series.

1. *Threshold-based estimators* select the number of factors by thresholding the eigenvalue distribution. They usually appear in literatures as certain information criteria dated back to Bai and Ng (2002) where the number of factors is estimated as

$$\widehat{r} = \underset{1\le k\le r_{\max}}{\arg\min} \Big\{ V(k) + k\cdot g(T,d) \Big\}, \tag{2.13}$$

where $V(k)$ is from (2.10), $g(T,d)$ is some information criteria, and $r_{\max}$ is some positive integers (such as $\lfloor d/2 \rfloor$ in practice) representing the upper bound for searching the number of factors. Since $V(k)$ virtually requires $\mathbf{AF}$ to be the best rank-$k$ approximation of $\mathbf{Y}$, we can rewrite $V(k)$ and hence (2.13) is equivalent to

$$\begin{aligned}\widehat{r} &= \underset{1\le k\le r_{\max}}{\arg\min} \Big\{ \frac{1}{Td}\Big( \|\mathbf{Y}\|_F^2 - \sum_{j=1}^{k} \lambda_j(\mathbf{YY}') \Big) + k\cdot g(T,d) \Big\} \\ &= \underset{1\le k\le r_{\max}}{\arg\max} \Big\{ \sum_{j=1}^{k} \lambda_j(\widehat{\boldsymbol{\Sigma}}_y) - kd\cdot g(T,d) \Big\} \\ &= \max_{1\le k\le r_{\max}} \Big\{ k \mid \lambda_k(\widehat{\boldsymbol{\Sigma}}_y) \ge d\cdot g(T,d) \Big\},\end{aligned} \tag{2.14}$$

which effectively sets a lower bound on the sample covariance matrix eigenvalues which should be sufficiently inflated by all $r$ factors. The estimator in (2.13) (or with $V(k)$ replaced by $\log(V(k))$) is shown consistent by Bai and Ng (2002); see also the discussion in Section 4.2 in Bai and Ng (2019).

   Fan et al. (2022) proposes a scale-invariant estimator similar to (2.14) with $\widehat{\boldsymbol{\Sigma}}_y$ replaced by the sample correlation matrix and the threshold value replaced correspondingly. For more threshold-based estimators, see e.g. Li et al. (2017a) and Su and Wang (2017).

2. *Difference-based estimators* identify the number of factors with sufficiently large eigengaps, i.e., differences between adjacent eigenvalues. Their motivation is from threshold-based estimators based on the eigenvalues of sample covariance matrices, which typically requires $r$ eigenvalues to grow proportionally with $d$ while the others remain bounded. This should display on the scree plot a cut-off, whereas a gradual decrease is often seen empirically (see e.g. Figure 1 in Freyaldenhoven (2019)), suggesting the existence of "less pervasive" factors. To formalise this, let

$$\|\mathbf{A}_{\cdot j}\|^2 \asymp d^{\delta_j}, \tag{2.15}$$

where the *factor strength* $\delta_j \in (0,1]$ can be heterogeneous across factors $j \in [r]$. We have *pervasive/strong factors* if $\delta_j = 1$ or *weak factors* otherwise, and we do not seek further classification of weakness in this thesis. An immediate example is (2.10) where all factors are strong.

To handle weak factors, Onatski (2010), assuming a similar fashion as $\delta_j > 0$, proposes the ED (edge distribution) estimator resulting from an iterative algorithm based on

$$\widehat{r}(\xi) = \max_{1 \le k \le r_{\max}} \left\{ k \mid \lambda_k(\widehat{\boldsymbol{\Sigma}}_y) - \lambda_{k+1}(\widehat{\boldsymbol{\Sigma}}_y) \ge \xi \right\},$$

where $\xi$ is a calibrated constant, and the other notations are borrowed from (2.14). See also Kapetanios (2010) using eigenvalue differences to construct test statistics.

3. *Ratio-based estimators* are constructed by the ratio between certain functions of (in most literatures, consecutive) eigenvalues. For example, Lam and Yao (2012) uses $\widehat{\mathbf{M}}$ defined in (2.12) and for arbitrary weak factors (with technical restrictions), introduce

$$\widehat{r} = \arg\min_{1 \le k \le r_{\max}} \left\{ \frac{\lambda_{k+1}(\widehat{\mathbf{M}})}{\lambda_k(\widehat{\mathbf{M}})} \right\}, \tag{2.16}$$

with only $\mathbb{P}(\widehat{r} \ge r) \to 1$ constructed. Such estimator is further developed in e.g. Li et al. (2017b) and Zhang et al. (2024a), where consistent estimators are proposed.

On the other hand, Ahn and Horenstein (2013) independently proposes the ER (eigenvalue ratio) estimator similar to (2.16) except for $\widehat{\mathbf{M}}$ replaced by $\widehat{\boldsymbol{\Sigma}}_y$, together with the GR (growth ratio) estimator which is also ratio-based. Both were shown to cope with weak factors by simulations, but consistency was only obtained under strong factors.

### 2.3.4 Higher-order tensor factor models: from matrices to beyond

With the advancement of statistical analyses for large dimensional panel data over the past decade, researchers also open up more to time series data with higher order, namely, tensor time series such that a tensor is observed at each timestamp; see Section 2.1 for an introduction to tensor. A prominent example, compared with vector time series, would be order-2 tensor time series, i.e., matrix(-valued) time series. Chen et al. (2021), Wu and Bi (2023), and Zhang (2024) propose autoregressive and moving-average models for matrix time series. More recently, Yu et al. (2024a) proposes matrix generalised autoregressive conditional heteroscedasticity models, and Han et al. (2024b) proposes a decorrelation scheme to transform matrix time series into blocks of cross-uncorrelated submatrices. See Tsay (2024) for a comprehensive review of matrix time series analysis.

Beyond matrix time series, general order tensor time series is also more of interest. Examples include different tensor autoregressive models proposed by Li and Xiao (2021) which extends Chen et al. (2021) and is based on CP decomposition, and by Wang et al. (2024) with a low-rank structure using "generalised inner products" (or sometimes "contracted tensor inner products" in the community of tensor regression), respectively.

As promised, we focus on factor modelling which, for matrix or general order tensor time series, is a subject still in its infancy. As vector factor models achieve dimension reduction in the form of singular value decomposition, it should not be surprising that higher-order tensor factor models are closely related to tensor decompositions discussed in Section 2.2. For all existing literatures and potential future works, we pinpoint the key theme in higher-order tensor factor modelling – **mode interaction**, which is related to the flexibility of various frameworks and naturally arises as there is more than one mode in higher-order tensors.

Compared to matrix/tensor factor models based on Tucker decompositions, those based on CP decompositions are only studied to a limited extent. Generally speaking, there exist mainly two independent series of approaches to address estimation/inference on factor models with the common component governed by CP decomposition, as follows.

1. For matrix time series, Chang et al. (2023) develops a generalised eigenanalysis, where as Chang et al. (2024) allows for rank-deficit loadings and proposes a non-orthogonal joint diagonalisation scheme.

2. For general order tensor time series, Han et al. (2024c) proposes the HOPE (high-order projection estimator) initialised by a composite PCA estimator, which is further investigated by Chen et al. (2024b) using the sample covariance matrix.

See also Guan (2024) among others, where covariates can be taken into the loadings and parameters are estimated by an EM approach. For the rest of this subsection, we discuss Tucker-decomposition factor models.

For matrix time series, factor models are first studied in Wang et al. (2019) based on Tucker decomposition such that for each observed $\mathbf{Y}_t \in \mathbb{R}^{d_1 \times d_2}$ ($t \in [T]$), we can write

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}_1 \mathbf{F}_t \mathbf{A}_2' + \mathbf{E}_t, \tag{2.17}$$

where $\boldsymbol{\mu}$ is the mean matrix, $\mathbf{F}_t \in \mathbb{R}^{r_1 \times r_2}$ is the zero-mean core factor matrix, $\mathbf{A}_1$ and $\mathbf{A}_2$ are the row and column loadings, and $\mathbf{E}_t$ is the noise. When $\mathbf{Y}_t$ degenerates to a vector $\mathbf{y}_t$, the model boils down to (2.8). Hence (2.17) can be understood as a matrix analogy of the vector factor model, with the common component decomposed using Tucker decomposition (2.4), and we refer to (2.17) the Tucker-decomposition matrix factor model (MFM).

To estimate the factor structure in (2.17), Wang et al. (2019) follows the idea of Lam et al. (2011) by leveraging the sample autocovariance matrices respectively for the row and column dimensions. Suppose $\mathbf{Y}_t$ has been appropriately demeaned, then the row loading estimator $\widehat{\mathbf{A}}_1$

is a PCA-type estimator constructed by the eigenvectors corresponding to the first $r_1$ largest eigenvalues of $\widehat{\mathbf{M}}_1$ defined akin to (2.12) (but based on tensor outer products, c.f. Equation (16) in Chen et al. (2022a)) as

$$\widehat{\boldsymbol{\Sigma}}_1(h,i,j) = \frac{1}{T-h}\sum_{t=1}^{T-h}\mathbf{Y}_{t,\cdot i}\mathbf{Y}'_{t+h,\cdot j}, \quad \widehat{\mathbf{M}}_1 = \sum_{h=1}^{h_0}\sum_{i=1}^{d_2}\sum_{j=1}^{d_2}\widehat{\boldsymbol{\Sigma}}_1(h,i,j)\widehat{\boldsymbol{\Sigma}}_1(h,i,j)'. \tag{2.18}$$

The column loading estimator $\widehat{\mathbf{A}}_2$ is obtained similarly but using the columns of $\mathbf{Y}_t'$ instead of $\mathbf{Y}_t$. With $\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2$, the core factor can be estimated by $\widehat{\mathbf{F}}_t = \widehat{\mathbf{A}}_1'\mathbf{Y}_t\widehat{\mathbf{A}}_2$. In contrast, Chen and Fan (2023) uses the sample covariance matrix, yet different from the estimator (2.11) only based on second moment information, also incorporates first moment information by weighting the sample mean. For example, denote $\bar{\mathbf{Y}} = \sum_{t=1}^{T}\mathbf{Y}_t/T$ the sample mean and $\alpha \in [-1,\infty)$ some hyper-parameter to be tuned, their namely $\alpha$-PCA estimator for the row loading is obtained as the eigenvectors (corresponding to the first $r_1$ largest eigenvalues) of the matrix

$$(1+\alpha)\bar{\mathbf{Y}}\bar{\mathbf{Y}}' + \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})'.$$

On the other hand, Yu et al. (2022a) proposes the PE (projection estimator). To ease discussion, let $\boldsymbol{\mu} = \mathbf{0}$ for the rest of this section. To start up, PE requires initial estimators for the row and column loadings (in fact, either one suffices); then the row/column loading estimators are recursively updated by a PCA-type estimator using the data projected on the most updated column/row loading estimator. With some stopping criteria, the resulting estimators are PE, and the core factor estimator is again direct. He et al. (2024b) discovers that PE is nothing else but a least-squares-type estimator corresponding to the least squares problem

$$\min_{\mathbf{A}_1,\mathbf{A}_2,\mathbf{F}_t}\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{Y}_t - \mathbf{A}_1\mathbf{F}_t\mathbf{A}_2'\|_F^2 \quad \text{subj. to} \quad \mathbf{A}_1'\mathbf{A}_1/d_1 = \mathbf{I}_{r_1}, \quad \mathbf{A}_2'\mathbf{A}_2/d_2 = \mathbf{I}_{r_2}. \tag{2.19}$$

The optimisation problem is non-convex to all parameters, but convex to individual parameter given others. Eventually, we can show that $\mathbf{A}_1$ (resp. $\mathbf{A}_2$) as the solution should be the eigenvector matrix corresponding to the first $r_1$ (resp. $r_2$) largest eigenvalues of the matrix

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_t\mathbf{A}_2\mathbf{A}_2'\mathbf{Y}_t' \quad \Big(\text{resp.} \quad \frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_t'\mathbf{A}_1\mathbf{A}_1'\mathbf{Y}_t\Big).$$

Thus, we may start with sufficiently good estimators and iterate, which is only non-trivial for tensor time series with order at least two. To estimate (2.17), Xu et al. (2024) also suggests a quasi maximum likelihood approach encompassing PE as a special case. Instead of the square loss in (2.19), He et al. (2024b) uses Huber loss on the Frobenius norm to alleviate heavy-tailedness, which is further studied in He et al. (2023b) by using element-wise Huber loss. More on robust factor modelling, He et al. (2022b) generalises He et al. (2022a) (for vector time series) to matrix factor models by proposing row and column matrix Kendall's tau.

Extended from (2.17), Chen et al. (2020) incorporates pre-specified constraint matrices to enhance interpretation; Chen and Chen (2022) adopts (2.17) to analyse transport network, with the possible scenario $\mathbf{A}_1 = \mathbf{A}_2$ considered; Liu and Chen (2022) studies threshold MFM with regime changes governed by an observed threshold variable; Chen et al. (2024a) applies kernel estimation to study time-varying MFM; Kong et al. (2024) proposes quantile MFM; He et al. (2024a) studies sequential detection on the changes of factor loadings; etc.

All the above matrix factor models employ only *two-way* structures (not to be confused with the inconsistent use of "two-way" over different works below) such that the mode interaction (between row and column here) is governed simultaneously by one term, i.e., the common component. Researchers also introduce different structures to exploit the matrix nature of the data. In particular, *one-way* structures are often included so that rows or columns independently contribute to the data matrix. Given a zero-mean time series with $\mathbf{Y}_t \in \mathbb{R}^{d_1 \times d_2}$ observed at each timestamp, Gao and Tsay (2023) assumes the noise has both one-way and two-way structures partially shared by the same source that drives the factor $\mathbf{F}_t$, i.e.,

$$\mathbf{Y}_t = \mathbf{A}_1 \begin{pmatrix} \mathbf{F}_t & \mathbf{E}_{1,t} \\ \mathbf{E}_{2,t} & \mathbf{E}_{3,t} \end{pmatrix} \mathbf{A}_2', \tag{2.20}$$

where $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$ are noise components, $\mathbf{A}_1, \mathbf{A}_2$ are square matrices, and $\mathbf{F}_t \in \mathbb{R}^{r_1 \times r_2}$ is the low-rank core factor with $r_1 \ll d_1, r_2 \ll d_2$; the estimation relies on PCA-type estimators followed by projection which exploits the model structure. Yuan et al. (2023) proposes the 2w-DFM (two-way dynamic factor model) such that each observed $\mathbf{Y}_t$ can be decomposed as

$$\mathbf{Y}_t = \mathbf{A}_1 \mathbf{F}_{1,t}' + \mathbf{F}_{2,t} \mathbf{A}_2' + \mathbf{E}_t, \tag{2.21}$$

where $\mathbf{F}_{1,t} \in \mathbb{R}^{r_1 \times d_2}$ and $\mathbf{F}_{2,t} \in \mathbb{R}^{d_1 \times r_2}$ are two factor series with some unknown autoregressive dynamics; all parameters are estimated in a two-step procedure based on quasi likelihood. At the same time, He et al. (2023a) tests the existence of the two-way structure as in (2.17) against the existence of only one-way factor structure as in (2.21), or just pure noise, through a randomisation scheme based on the spectrum of the sample covariance matrices. Zhang et al. (2024c) proposes the RaDFaM (rank-decomposition-based matrix factor model) incorporating both one-way and two-way structures such that

$$\mathbf{Y}_t = \mathbf{A}_1 \mathbf{F}_t \mathbf{A}_2' + \mathbf{A}_1 \mathbf{F}_{1,t}' + \mathbf{F}_{2,t} \mathbf{A}_2' + \mathbf{E}_t,$$

which can be either viewed as a combination of the Tucker-decomposition MFM and 2w-DFM, or a similar form as (2.20) except for the absence of the particular structure implied by $\mathbf{E}_{3,t}$.

Frameworks or techniques from MFM can often be adapted to general order tensor factor models (TFM). For example, Chen et al. (2022a) proposes the Tucker-decomposition TFM to decompose each zero-mean order-$K$ tensor $\mathcal{Y}_t \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ into

$$\mathcal{Y}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \cdots \times_K \mathbf{A}_K + \mathcal{E}_t, \tag{2.22}$$

where $\mathcal{F}_t \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is an order-$K$ core tensor, each $\mathbf{A}_k$ ($k \in [K]$) is the mode-$k$ factor loading, and $\mathcal{E}_t$ has no serial correlation as in Lam et al. (2011). Define $d = \prod_{k=1}^{K} d_k$ and $d_{\text{-}k} = d/d_k$. By generalising from Wang et al. (2019), Chen et al. (2022a) proposes the TOPUP (time series outer-product unfolding procedure) to estimate each mode-$k$ loading by PCA-type estimators based on $\text{mat}_1(\mathcal{Y}_{\text{TOPUP},k})$, where $\mathcal{Y}_{\text{TOPUP},k} \in \mathbb{R}^{d_k \times d_{\text{-}k} \times d_k \times d_{\text{-}k} \times h_0}$ is an order-5 tensor formed by stacking all order-4 tensor ($h \in [h_0]$)

$$(\mathcal{Y}_{\text{TOPUP},k})_{\cdots\cdot h} = \frac{1}{T-h} \sum_{t=1}^{T-h} \text{mat}_k(\mathcal{Y}_t) \circ \text{mat}_k(\mathcal{Y}_{t+h})$$

in its last mode. To see that for $K = 2$ (i.e., matrix time series), the TOPUP estimator coincides with the PCA-type estimator by Wang et al. (2019), we may read $\widehat{\mathbf{M}}_1$ in (2.18) as a block-matrix product with each block $\widehat{\mathbf{\Sigma}}_1(h, i, j)$ so that

$$\text{mat}_1(\mathcal{Y}_{\text{TOPUP},1})\, \text{mat}_1(\mathcal{Y}_{\text{TOPUP},1})' = \widehat{\mathbf{M}}_1.$$

Besides, Chen et al. (2022a) also proposes the TIPUP (time series inner-product unfolding procedure) based on $\text{mat}_1(\mathcal{Y}_{\text{TIPUP},k})$, where $\mathcal{Y}_{\text{TIPUP},k} \in \mathbb{R}^{d_k \times d_k \times h_0}$ is an order-3 tensor formed by stacking all matrix ($h \in [h_0]$)

$$(\mathcal{Y}_{\text{TIPUP},k})_{\cdot\cdot h} = \frac{1}{T-h} \sum_{t=1}^{T-h} \text{mat}_k(\mathcal{Y}_t)\text{mat}_k(\mathcal{Y}_{t+h})' \tag{2.23}$$

in its last mode. See Remark 6 in Chen et al. (2022a) for the comparison between TOPUP and TIPUP. Iterative procedures by projection (as in Yu et al. (2022a) for matrix time series) are numerically demonstrated and later theoretically studied in Han et al. (2024a).

Almost concurrently, Barigozzi et al. (2023b) and Zhang et al. (2024b) independently propose the similar iterative projection estimator by extending Yu et al. (2022a) and He et al. (2024b) from MFM to TFM; recall that the framework originates from approximate (vector) factor models (Bai and Ng, 2002; Bai, 2003). The initial estimator is a PCA-type estimator based on sample covariance matrices, i.e., (2.23) with $h = 0$. Instead of projecting on the space of loading estimators, Chen and Lam (2024b) proposes a pre-averaging estimator by projecting on random sets of fibres to achieve superior results in the presence of weak factors.

Other developments include Chen et al. (2024c) on semi-parametric TFM, Barigozzi et al. (2023a) on robust TFM which directly extends the Huber regression in He et al. (2024b), and Barigozzi et al. (2024) on tail-robust TFM by element-wise truncation followed by iterative projection, etc. There are also lots of studies on rank estimation for tensor (time series). Besides those methods proposed within some aforementioned works, see e.g. Yokota et al. (2017), Lam (2021), and Han et al. (2022).

# Chapter 3

# Tensor Time Series Imputation through Tensor Factor Modelling

## 3.1 Introduction

In large time series analysis, a less addressed topic is the treatment of missing data, in particular, imputation of missing data and the corresponding inferences. While there are numerous data-centric methods in various scientific fields for imputing multivariate time series data (see Chapon et al. (2023) for environmental time series, Kazijevs and Samad (2023) for health time series, Zhao et al. (2023) and Zhang et al. (2021) for using deep-learning related architectures for imputations, to name but a few), almost none of them address statistically how accurate their methods are, and all of them are not for higher-order tensor time series. We certainly can line up the variables in a tensor time series to make it a longitudinal panel, but in doing so we lose special structures and insights that can be utilised for forecasting and interpretation of the data. More importantly, transforming a moderate sized tensor to a vector means the length of the vector can be much larger than the sample size, creating curse of dimensionality.

For imputing large panel of time series with statistical analyses, Bai and Ng (2021) define the concept of "Tall" and "Wide" blocks of data and propose an iterative "TW" algorithm in imputing missing values in a large panel, while Cahan et al. (2023) improve the TW algorithm to a "Tall-Project" algorithm so that there is no iterations needed. Both papers use factor modelling for the imputations, and derive rates of convergence when all factors are pervasive and the number of factors known. Asymptotic normality for rows of estimated factor loadings and the corresponding practical inferences are developed as well. Xiong and Pelger (2023) also base their imputations on a factor model for a large panel of time series with pervasive factors and number of factors known, and build a method for imputing missing values under very general missing patterns, with asymptotic normality and inferences also developed.

To the best of our knowledge, for tensor time series with order larger than 1 (i.e., at least matrix-valued), there are no theoretical analyses on the imputation performance. Imputation methodologies developed on tensor time series are also scattered around very different applications. See Chen et al. (2022b) on traffic tensor data and Pan et al. (2021) for RNA-sequence tensor data for instance.

In view of all the above, as a first in the literature, we aim to develop a tensor imputation method accompanied by theoretical analyses in this chapter. Like Cahan et al. (2023), we use factor modelling for tensor time series as a basis for our imputation method. Unlike Cahan et al. (2023), Bai and Ng (2021) or Xiong and Pelger (2023) though, we develop a method that can consistently estimate the number of factors, or the core tensor rank, in a Tucker-decomposition factor model for the tensor time series with missing values. Our method can be considered a combination of Barigozzi et al. (2023b) for the tensor factor model, and Xiong and Pelger (2023) for the imputation methodology with general missingness. In Section 3.2, we introduce two motivating examples and our methodology at the same time. One is the Fama–French portfolio return data with missing entries, to be analysed in Section 3.4.2. The other is a set of monthly and quarterly OECD economic indicators, with missingness naturally occurring for the quarterly recorded indicators relative to the monthly ones. We analyse this set of OECD data in Section 3.4.3.

As a further contribution, we also allow factors to be weak. A weak factor corresponds to a column in a factor loading matrix being sparse, or approximately sparse. This implies that not all units in a tensor has dynamics contributed by all the factors inside the core tensor. In Chen and Lam (2024b), they allow for weak factors and discovers that there are potentially weak factors in the NYC taxi traffic data. We prove consistency of our imputations under general missingness, and develop asymptotic normality and practical inferences for rows of factor loading matrix estimators, with rates of convergence in all consistency results spelt out.

The rest of this chapter is organised as follows. Section 3.2 presents the Fama–French portfolio returns data and the OECD data as two motivating examples, before describing the tensor factor model and the imputation methodology we employ. Section 3.3 lays down the main assumptions for this chapter, with consistent estimation and rates of convergence of all factor loading matrix estimators and imputed values presented. Specifically, asymptotic normality and the estimators of the corresponding covariance matrices for practical inferences are introduced as well in Section 3.3.3, before our proposed ratio-based estimators for the number of factors in Section 3.3.5. Section 3.4 presents extensive simulation results, together with analyses for the Fama–French portfolio return data and the OECD economic data. Section 3.5 contains all the proofs of theorems and propositions. Our method is available in the R package `tensorMiss`, which leveraged the `Rcpp` package to greatly boost computational speed.

## 3.2 Motivating Examples and the Imputation Procedure

We first describe two motivating data examples in Section 3.2.1 and 3.2.2, before presenting our imputation procedure for a general order-$K$ mean-zero tensor $\mathcal{Y}_t = (\mathcal{Y}_{t,i_1,\ldots,i_K}) \in \mathbb{R}^{d_1\times\cdots\times d_K}$ ($t \in [T]$). The two data examples are analysed in detail in Section 3.4.2 and 3.4.3 respectively.

### 3.2.1 Example: Fama–French portfolio returns

This is a set of Fama–French portfolio returns data with missingness. Stocks are categorised into ten levels of market equity (ME) and ten levels of book-to-equity ratio (BE) which is the book equity for the last fiscal year divided by the end-of-year ME. At the end of June each year, both ME and BE use NYSE deciles as breakpoints, with stocks of NYSE, AMEX and NASDAQ firms allocated accordingly. Moreover, the stocks in each of the $10 \times 10$ categories form exactly two portfolios, one being value weighted, and the other of equal weight. Hence, there are two sets of 10 by 10 portfolios with their time series to be studied. We use monthly data from January 1974 to June 2021, so that $T = 570$, and for both value weighted and equal weighted portfolios we have each of our data set as an order-2 tensor $\mathcal{X}_t \in \mathbb{R}^{10\times 10}$ for $t \in [570]$. For more details, please visit
https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_100_port_sz.html.

If no stocks are allocated to a category (i.e., intersection of ME and BE categorisation) at a timestamp, the corresponding return data is unavailable and hence missing. It is reasonable to argue the missingness might depend on the rows of the loading matrix, i.e., extreme categories tend to contain fewer stocks, but independent of latent factors and noise. The total number of missing entries is 161 and hence the percentage of missing is $161/(10\times 10\times 570) = 0.28\%$ for both the value weighted and equal weighted series. However, the irregular missing pattern here can be harmful if we are after a complete case analysis. For full observation after a timestamp, we may only start from July 2009 and hence $74.7\%$ of the data would be ditched. On the other hand, we might ditch four categories to obtain a complete data set but lose the potential insights on the return series of the four categories.

### 3.2.2 Example: OECD economic indicators

In this example, we study a group of economic indicators for a selection of countries obtained from the Organization for Economic Co-operation and Development (OECD). The data consists of monthly/quarterly observations of 11 economic indicators: current account balance as percentage of GDP (CA-GDP), consumer price index (CP), merchandise exports (EX), merchandise imports (IM), short-term interest rates (IR3TIB), long-term interest rates (IRLT), interbank rates (IRSTCI), producer price index (PP), production volume (PRVM), retail trade volume (TOVM) and unit labour cost (ULC). They are observed for 17 countries: Belgium (BEL), Canada (CAN), Denmark (DNK), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Italy (ITA), Luxembourg (LUX), Netherlands (NLD), Norway (NOR), Portugal (PRT), Spain (ESP), Sweden (SWE), Switzerland (CHE), United Kingdom (GBR) and United States (USA), with data spanning from January 1971 to December 2023. We correspond respectively rows and columns to countries and indicators, so that we have our data as an order-2 tensor $\mathcal{Y}_t \in \mathbb{R}^{17\times 11}$ for $t \in [636]$. For more details, see key short-term economic indicators available at https://data.oecd.org/.

The data is naturally missing for three reasons: unavailable indicator records for some countries at early time periods, quarterly indicators are only available at the end of each quarter, and are sometimes unrecorded. Similar to the Fama–French data, we suppose the missing pattern is dependent on the loading matrices by arguing that relatively less important indicators are only available quarterly. The percentage of missing data is $26.2\%$, which leads to significantly inefficient use of data if we hope to analyse a set of complete data. The fact that the data is observed at least quarterly in the long run ensures the existence of a lower bound on the proportion of available data, which in turn satisfies Assumption (O1) in Section 3.3.

### 3.2.3 The model and the imputation procedure

**The Model**: Suppose the order-$K$ mean zero tensor $\mathcal{Y}_t$ is modelled by

$$\mathcal{Y}_t = \mathcal{C}_t + \mathcal{E}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \cdots \times_K \mathbf{A}_K + \mathcal{E}_t, \quad t \in [T], \tag{3.1}$$

where $\mathcal{C}_t$ is the common component and $\mathcal{E}_t$ the error tensor. The core tensor is $\mathcal{F}_t \in \mathbb{R}^{r_1\times\cdots\times r_K}$, and each mode-$k$ factor loading matrix $\mathbf{A}_k$ has dimension $d_k \times r_k$. See Barigozzi et al. (2023b) amongst others using the same tensor factor model. Using the QR decomposition, if we can decompose $\mathbf{A}_k = \mathbf{Q}_k\mathbf{Z}_k^{1/2}$ (see Assumption (L1) in Section 3.3.1 for details), then (3.1) can be written as

$$\begin{aligned} \mathcal{Y}_t &= \mathcal{F}_{Z,t} \times_1 \mathbf{Q}_1 \times_2 \cdots \times_K \mathbf{Q}_K + \mathcal{E}_t, \quad t \in [T], \quad \text{where} \\ \mathcal{F}_{Z,t} &:= \mathcal{F}_t \times_1 \mathbf{Z}_1^{1/2} \times_2 \cdots \times_K \mathbf{Z}_K^{1/2}. \end{aligned} \tag{3.2}$$

Model (3.1) is an extension to the usual time series factor model ($K = 1$):

$$\mathcal{Y}_t = \text{mat}_1(\mathcal{Y}_t) = \text{mat}_1(\mathcal{F}_t \times_1 \mathbf{A}_1) + \text{mat}_1(\mathcal{E}_t) = \mathbf{A}_1\text{mat}_1(\mathcal{F}_t) + \text{mat}_1(\mathcal{E}_t) = \mathbf{A}_1\mathcal{F}_t + \mathcal{E}_t,$$

and also for a matrix-valued time series factor model ($K = 2$):

$$\mathcal{Y}_t = \text{mat}_1(\mathcal{Y}_t) = \mathbf{A}_1\text{mat}_1(\mathcal{F}_t)\mathbf{A}_2' + \text{mat}_1(\mathcal{E}_t) = \mathbf{A}_1\mathcal{F}_t\mathbf{A}_2' + \mathcal{E}_t.$$

**The Imputation Procedure**: We only observe partial data. Define the missingness tensor $\mathcal{M}_t = (\mathcal{M}_{t,i_1,\ldots,i_K}) \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with

$$\mathcal{M}_{t,i_1,\ldots,i_K} = \begin{cases} 1, & \text{if } \mathcal{Y}_{t,i_1,\ldots,i_K} \text{ is observed;} \\ 0, & \text{otherwise.} \end{cases}$$

Our aim is to recover the value for the common component $\mathcal{C}_{t,i_1,\ldots,i_K}$ if $\mathcal{M}_{t,i_1,\ldots,i_K} = 0$. Assuming first the number of factors $r_k$ is known for all modes, we want to obtain the estimators of the factor loading matrices, $\widehat{\mathbf{Q}}_k$ for $k \in [K]$, and then the estimated core tensor series $\widehat{\mathcal{F}}_{Z,t}$ for $t \in [T]$. See (3.2) for the definition of $\mathbf{Q}_k$ and $\mathcal{F}_{Z,t}$. We can then estimate the common components at time $t$ by

$$\widehat{\mathcal{C}}_t = \widehat{\mathcal{F}}_{Z,t} \times_1 \widehat{\mathbf{Q}}_1 \times_2 \cdots \times_K \widehat{\mathbf{Q}}_K. \tag{3.3}$$

With (3.3), we can impute $\mathcal{Y}_t$ using

$$\widetilde{\mathcal{Y}}_{t,i_1,\ldots,i_K} = \begin{cases} \mathcal{Y}_{t,i_1,\ldots,i_K}, & \text{if } \mathcal{M}_{t,i_1,\ldots,i_K} = 1; \\ \widehat{\mathcal{C}}_{t,i_1,\ldots,i_K}, & \text{if } \mathcal{M}_{t,i_1,\ldots,i_K} = 0. \end{cases}$$

We leave the discussion of estimating $r_k$ to Section 3.3.5. See Section 3.2.4 in how to obtain $\widehat{\mathbf{Q}}_k$ and Section 3.2.5 in how to obtain $\widehat{\mathcal{F}}_{Z,t}$.

### 3.2.4 Estimation of factor loading matrices

In this chapter, we make use of the following notation:

$$\psi_{k,ij,h} := \Big\{ t \in [T] \mid \text{mat}_k(\mathcal{M}_t)_{ih} \text{mat}_k(\mathcal{M}_t)_{jh} = 1 \Big\}. \tag{3.4}$$

Hence $\psi_{k,ij,h}$ is the set of time periods where both the $i$-th and $j$-th entries of the $h$-th mode-$k$ fibre are observed, $i, j \in [d_k]$, $h \in [d_{\text{-}k}]$ with $d_{\text{-}k} := d_1 \cdots d_K / d_k$.

Inspired by Xiong and Pelger (2023) for a vector time series panel, our method relies on the reconstruction of the mode-$k$ sample covariance matrix $\mathbf{S}_k$, defined for $i, j \in [d_k]$,

$$(S_k)_{ij} := \frac{1}{T} \sum_{t=1}^{T} \text{mat}_k(\mathcal{Y}_t)'_{i\cdot} \text{mat}_k(\mathcal{Y}_t)_{j\cdot} = \sum_{h=1}^{d_{\text{-}k}} \frac{1}{T} \sum_{t=1}^{T} \text{mat}_k(\mathcal{Y}_t)_{ih} \text{mat}_k(\mathcal{Y}_t)_{jh}. \tag{3.5}$$

With missing entries characterised by $\mathcal{M}_t$ and $\psi_{k,ij,h}$ in (3.4), we can generalise the above to

$$(\widehat{S}_k)_{ij} = \sum_{h=1}^{d_{\text{-}k}} \Big\{ \frac{1}{|\psi_{k,ij,h}|} \sum_{t \in \psi_{k,ij,h}} \text{mat}_k(\mathcal{Y}_t)_{ih} \text{mat}_k(\mathcal{Y}_t)_{jh} \Big\}. \tag{3.6}$$

Intuitively, the cross-covariance between unit $i$ and $j$ at the $h$-th mode-$k$ fibre is estimated inside the curly bracket in (3.6) using only the corresponding available data. PCA can now be performed on $\widehat{\mathbf{S}}_k$, and $\widehat{\mathbf{Q}}_k$ is obtained as the first $r_k$ eigenvectors of $\widehat{\mathbf{S}}_k$.

### 3.2.5 Estimation of the core tensor series

With $\widehat{\mathbf{Q}}_k$ available (which is estimating the factor loading space of $\mathbf{Q}_k$, with $\widehat{\mathbf{Q}}_k$ having orthonormal columns), we can estimate $\mathcal{F}_{Z,t}$ (equivalently $\textbf{vec}(\mathcal{F}_{Z,t})$) by observing from (3.2),

$$\textbf{vec}\big(\mathcal{Y}_t\big) = \mathbf{Q}_{\otimes}\textbf{vec}\big(\mathcal{F}_{Z,t}\big) + \textbf{vec}\big(\mathcal{E}_t\big), \quad \text{where } \mathbf{Q}_{\otimes} := \mathbf{Q}_K \otimes \cdots \otimes \mathbf{Q}_1.$$

If $\mathbf{Q}_{\otimes}$ is known, then the least squares estimator of $\textbf{vec}\big(\mathcal{F}_{Z,t}\big)$ is given by

$$\textbf{vec}\big(\mathcal{F}_{Z,t}\big) = (\mathbf{Q}_{\otimes}'\mathbf{Q}_{\otimes})^{-1}\mathbf{Q}_{\otimes}'\textbf{vec}\big(\mathcal{Y}_t\big) = \Big(\sum_{j=1}^{d}\mathbf{Q}_{\otimes,j\cdot}\mathbf{Q}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d}\mathbf{Q}_{\otimes,j\cdot}[\textbf{vec}\big(\mathcal{Y}_t\big)]_j\Big).$$

With missing data, using the missingness tensor $\mathcal{M}_t$, the above can be generalised to

$$\textbf{vec}\big(\widehat{\mathcal{F}}_{Z,t}\big) = \Big(\sum_{j=1}^{d}[\textbf{vec}\big(\mathcal{M}_t\big)]_j\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d}[\textbf{vec}\big(\mathcal{M}_t\big)]_j\widehat{\mathbf{Q}}_{\otimes,j\cdot}[\textbf{vec}\big(\mathcal{Y}_t\big)]_j\Big). \tag{3.7}$$

## 3.3 Assumptions and Theoretical Results

### 3.3.1 Assumptions

We present our assumptions for consistent imputation and estimation of factor loading matrices, with the corresponding theoretical results presented afterwards.

(O1) (Observation patterns).
*1. $\mathcal{M}_t$ is independent of $\mathcal{F}_s$ and $\mathcal{E}_s$ for any $t, s \in [T]$.*
*2. Given $\mathcal{M}_t$ with $t \in [T]$, for any $k \in [K], i, j \in [d_k], h \in [d_{\text{-}k}]$, there exists a constant $\psi_0$ such that with probability going to 1, we have*

$$\frac{|\psi_{k,ij,h}|}{T} \geq \psi_0 > 0.$$

(M1) (Alpha mixing). *The vector processes $\{\textbf{vec}\big(\mathcal{F}_t\big)\}$ and $\{\textbf{vec}\big(\mathcal{E}_t\big)\}$ are $\alpha$-mixing, respectively. A vector process $\{\mathbf{x}_t : t = 0, \pm 1, \dots\}$ is $\alpha$-mixing if, for some $\gamma > 2$, the mixing coefficients satisfy*

$$\sum_{h=1}^{\infty}\alpha(h)^{1-2/\gamma} < \infty,$$

*where* $\alpha(h) = \sup_\tau \sup_{A\in\mathcal{H}_{-\infty}^\tau, B\in\mathcal{H}_{\tau+h}^\infty} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ *and* $\mathcal{H}_\tau^s$ *is the* $\sigma$*-field generated by* $\{\mathbf{x}_t : \tau \le t \le s\}$.

(F1) (Time series in $\mathcal{F}_t$). *There is* $\mathcal{X}_{f,t}$ *the same dimension as* $\mathcal{F}_t$*, such that we may write* $\mathcal{F}_t = \sum_{q\ge 0} a_{f,q}\mathcal{X}_{f,t-q}$. *The time series* $\{\mathcal{X}_{f,t}\}$ *has i.i.d. elements with mean* 0 *and variance* 1*, with uniformly bounded fourth order moments. The coefficients* $a_{f,q}$ *are such that* $\sum_{q\ge 0} a_{f,q}^2 = 1$ *and* $\sum_{q\ge 0} |a_{f,q}| \le c$ *for some constant* $c$.

(L1) (Factor strength). *We assume for* $k \in [K]$, $\mathbf{A}_k$ *is of full column rank and independent of factors and errors series. Furthermore, as* $d_k \to \infty$,

$$\mathbf{Z}_k^{-1/2}\mathbf{A}_k'\mathbf{A}_k\mathbf{Z}_k^{-1/2} \to \mathbf{\Sigma}_{A,k}, \tag{3.8}$$

*where* $\mathbf{Z}_k = \text{diag}(\mathbf{A}_k'\mathbf{A}_k)$ *and* $\mathbf{\Sigma}_{A,k}$ *is positive definite with all eigenvalues bounded away from 0 and infinity. We assume* $(\mathbf{Z}_k)_{jj} \asymp d_k^{\alpha_{k,j}}$ *for* $j \in [r_k]$*, and* $1/2 < \alpha_{k,r_k} \le \cdots \le \alpha_{k,2} \le \alpha_{k,1} \le 1$.

With Assumption (L1), we can denote $\mathbf{Q}_k := \mathbf{A}_k\mathbf{Z}_k^{-1/2}$ and hence $\mathbf{Q}_k'\mathbf{Q}_k \to \mathbf{\Sigma}_{A,k}$. We need $\alpha_{k,j} > 1/2$ so that the ratio-based estimator of the number of factors in Section 3.3.5 works.

(E1) (Decomposition of $\mathcal{E}_t$). *We assume* $K$ *is constant, and*

$$\mathcal{E}_t = \mathcal{F}_{e,t} \times_1 \mathbf{A}_{e,1} \times_2 \cdots \times_K \mathbf{A}_{e,K} + \mathbf{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t, \tag{3.9}$$

*where* $\mathcal{F}_{e,t}$ *is an order-*$K$ *tensor with dimension* $r_{e,1} \times \cdots \times r_{e,K}$*, containing independent elements with mean* 0 *and variance* 1*. The order-*$K$ *tensor* $\boldsymbol{\epsilon}_t \in \mathbb{R}^{d_1\times\cdots\times d_K}$ *contains independent mean zero elements with unit variance, with the two time series* $\{\boldsymbol{\epsilon}_t\}$ *and* $\{\mathcal{F}_{e,t}\}$ *being independent. The order-*$K$ *tensor* $\mathbf{\Sigma}_\epsilon$ *contains the standard deviations of the corresponding elements in* $\boldsymbol{\epsilon}_t$*, and has elements uniformly bounded.*

*Moreover, for each* $k \in [K]$, $\mathbf{A}_{e,k} \in \mathbb{R}^{d_k\times r_{e,k}}$ *is such that* $\|\mathbf{A}_{e,k}\|_1 = O(1)$*. That is,* $\mathbf{A}_{e,k}$ *is (approximately) sparse.*

(E2) (Time series in $\mathcal{E}_t$). *There is* $\mathcal{X}_{e,t}$ *the same dimension as* $\mathcal{F}_{e,t}$*, and* $\mathcal{X}_{\epsilon,t}$ *the same dimension as* $\boldsymbol{\epsilon}_t$*, such that* $\mathcal{F}_{e,t} = \sum_{q\ge 0} a_{e,q}\mathcal{X}_{e,t-q}$ *and* $\boldsymbol{\epsilon}_t = \sum_{q\ge 0} a_{\epsilon,q}\mathcal{X}_{\epsilon,t-q}$*, with* $\{\mathcal{X}_{e,t}\}$ *and* $\{\mathcal{X}_{\epsilon,t}\}$ *independent of each other, and each time series has independent elements with mean* 0 *and variance* 1 *with uniformly bounded fourth order moments. Both* $\{\mathcal{X}_{e,t}\}$ *and* $\{\mathcal{X}_{\epsilon,t}\}$ *are independent of* $\{\mathcal{X}_{f,t}\}$ *from (F1).*

*The coefficients* $a_{e,q}$ *and* $a_{\epsilon,t}$ *are such that* $\sum_{q\ge 0} a_{e,q}^2 = \sum_{q\ge 0} a_{\epsilon,q}^2 = 1$ *and for some constant* $c$ *it holds that* $\sum_{q\ge 0} |a_{e,q}|, \sum_{q\ge 0} |a_{\epsilon,q}| \le c$.

(R1) (Further rate assumptions). *We assume that, with* $d := \prod_{k=1}^K d_k$ *and* $g_s := \prod_{k=1}^K d_k^{\alpha_{k,1}}$,

$$dg_s^{-2}T^{-1}d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})+1},\ dg_s^{-1}T^{-1}d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})},\ dg_s^{-1}d_k^{\alpha_{k,1}-\alpha_{k,r_k}-1/2} = o(1).$$

Assumption (O1) means that the missing mechanism is independent of the factors and the noise series, which is also assumed in Xiong and Pelger (2023) for the purpose of identification. It also means that the missing pattern can depend on the $K$ factor loading matrices, allowing for a wide variety of missing patterns that can vary over time and units in different dimensions. Condition 2 of (O1) implies that the number of time periods that any two individual units are both observed are at least proportional to $T$, which simplifies proofs and presentations, and is also used in Xiong and Pelger (2023). Assumption (M1) is a standard assumption in vector time series factor models, which facilitates proofs using central limit theorem for time series without losing too much generality. Assumption (F1), (E1) and (E2) are exactly the corresponding assumptions in Chen and Lam (2024b), allowing for serial correlations in the factor series, and serial and cross-sectional dependence within and among the error tensor fibres. These three assumptions facilitate the proof of asymptotic normality in Section 3.3.3, and boil down to similar assumptions in Chen and Fan (2023) for matrix time series and in Barigozzi et al. (2023b) for general tensor time series (see Proposition 3.2 in Section 3.5 for the technical details). Together with Assumption (M1), we implicitly restrict the general linear processes in (F1) and (E2) to be, for instance, of short rather than long dependence.

Assumption (L1) is quite different from assumptions in other existing works on factor models, in the sense that we allow for the existence of weak factors alongside the pervasive ones. Chen and Lam (2024b) adapted the same assumption, which allows each column of $\mathbf{A}_k$ to be completely dense (i.e., a pervasive factor) or sparse to a certain extent. A diagonal entry in $\mathbf{Z}_k$ then records how dense a column really is, and the corresponding strength of factors defined. Assumption (L1) is similar to, yet technically more general than, Assumption 1(iii) in Onatski (2012) which requires $\boldsymbol{\Sigma}_{A,k}$ to be diagonal while the normalisation on the factor series is essentially the same as ours. If all factors are pervasive, (3.8) can be read as $d_k^{-1}\mathbf{A}_k'\mathbf{A}_k \to \boldsymbol{\Sigma}_{A,k}$ which is akin to Assumption 3 of Chen and Fan (2023) for $K = 2$. Modelling with weak factors is closer to reality, and empirical evidence can be found in economics and finance, etc. For instance, apart from a pervasive market factor, there can be weaker sector factors in a large selection of stock returns (Trzcinka, 1986). More recent work on factor models specifically focuses on weak factors with real data examples confirming the existence of weak factors, such as Freyaldenhoven (2022) and Chen and Lam (2024b).

Finally, Assumption (R1) gives the technical rates needed for the proof of various theorems in this chapter because of the existence of weak factors. If all factors are pervasive (i.e., $\alpha_{k,j} = 1$), then the conditions are automatically satisfied. Suppose $K = 2$, $T \asymp d_1 \asymp d_2$ and the strongest factors are all pervasive (i.e., $\alpha_{k,1} = 1$), then we need $\alpha_{k,r_k} > 1/2$ for (R1) to be satisfied. This condition is the same as the one remarked right after we stated Assumption (L1). A factor with $\alpha_{k,j}$ close to 0.5 presents a significantly weak factor with only more than $d_k^{1/2}$ of elements are nonzero in the corresponding column of $\mathbf{A}_k$.

**Remark 3.1** *To see Assumptions (F1) and (E1) do not imply (M1), a simple counterexample which satisfies (F1) and (E1) but not (M1) can be an appropriate moving average process not satisfying $\alpha$-mixing, see e.g. Sidorov (2010). Rather than delving into specific classes of processes that are $\alpha$-mixing, (M1) should be general enough to facilitate theoretical results to be smoothly spelt out. More generally, unless a Gaussian innovation process is assumed in the linear processes, showing (M1) by additional assumptions is unnecessarily complicated to be pursued. As discussed in Section 15.3 in Davidson (2021): "[...] allowing more general distributions for the innovations yields surprising results. Contrary to what might be supposed, having the $\theta_j$ tend to zero even at an exponential rate is not sufficient by itself for strong mixing [...]", where $\theta_j$ is the coefficient in the linear process; see also Theorem 15.9 in Davidson (2021) for a fairly general result which requires certain non-trivial smoothness conditions on the innovations' p.d.f.'s and decays on the coefficients for a univariate linear process to be $\alpha$-mixing.*

**Remark 3.2** *With the missing entries imputed by the estimated common components $\widehat{\mathcal{C}}_{t,i_1,\ldots,i_K}$, we have a completed data set which could be used for re-estimation and hence re-imputation. The convergence could be shown empirically to be accelerated by such a procedure. The rate improvement would be from the difference between $T$ and $\psi_0 T$, where $\psi_0$ is the lowest proportion of observation among all entries from Assumption (O1). We omit the lengthy proofs as eventually the rates only differ by a constant, but we note here that re-imputation can indeed improve our imputation, which is essentially credited to the more observations used when we have an initially good imputation.*

*Note that even if we adopt the iterative projection estimator from e.g. Barigozzi et al. (2023b), the initial imputation inherits the rate from our procedure and hence there would be no improvement on the theoretical rate. To potentially obtain a better rate, we should investigate how to adapt the iterative projection estimator to cope with missing data, which is non-trivial both on imputation procedure and theoretical derivation. Since it is not the main concern here, we leave this as a future direction as it is certainly a worthy yet technical extension.*

### 3.3.2 Consistency: factor loadings and imputed values

We present consistency results in this subsection. For $k \in [K], j \in [d_k]$, define

$$\mathbf{H}_{k,j} := \widehat{\mathbf{D}}_k^{-1} \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{k,i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{k,ij,h}|} \sum_{t \in \psi_{k,ij,h}} \Big\{ \sum_{m=1}^{r_{\text{-}k}} \Lambda_{k,hm} \text{mat}_k(\mathcal{F}_{Z,t})_{\cdot m} \Big\}' \mathbf{Q}_{k,i\cdot} \Big\{ \sum_{m=1}^{r_{\text{-}k}} \Lambda_{k,hm} \text{mat}_k(\mathcal{F}_{Z,t})_{\cdot m} \Big\}', \tag{3.10}$$

$$\mathbf{H}_k^a := \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{D}}_k^{-1} \widehat{\mathbf{Q}}_k' \mathbf{Q}_k \text{mat}_k(\mathcal{F}_{Z,t}) \mathbf{\Lambda}_k' \mathbf{\Lambda}_k \text{mat}_k(\mathcal{F}_{Z,t})', \tag{3.11}$$

where $\widehat{\mathbf{D}}_k := \widehat{\mathbf{Q}}_k' \widehat{\mathbf{S}}_k \widehat{\mathbf{Q}}_k$ is a diagonal matrix of eigenvalues of $\widehat{\mathbf{S}}_k$ defined in (3.6). Hence $\mathbf{H}_{k,j} = \mathbf{H}_k^a$ if there are no missing entries, i.e., $|\psi_{k,ij,h}| = T$ for each $k \in [K], i, j \in [d_k]$ and $h \in [d_{\text{-}k}]$. Furthermore, each $\mathbf{H}_{k,j}$ and $\mathbf{H}_k^a$ can be shown asymptotically bounded and invertible (see Theorem 3.1 with Lemma 3.3 in Section 3.5).

We first present a consistency result for the factor loading matrix estimator $\widehat{\mathbf{Q}}_k$ of $\mathbf{Q}_k$. In particular, our theoretical rates are shown in the presence of potential weak factors. To compare with results in similar literature, we will end this subsection with a simplified result. Readers interested in the rates under only pervasive factors can go straight to Corollary 3.1.

**Theorem 3.1** *Under Assumptions (O1), (M1), (F1), (L1), (E1), (E2) and (R1), for any* $k \in [K]$,

$$\frac{1}{d_k} \sum_{j=1}^{d_k} \big\| \widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_{k,j} \mathbf{Q}_{k,j\cdot} \big\|^2 = O_P \bigg\{ d_k^{2(\alpha_{k,1} - \alpha_{k,r_k}) - 1} \Big( \frac{1}{T d_{\text{-}k}} + \frac{1}{d_k} \Big) \frac{d^2}{g_s^2} \bigg\} = o_P(1),$$

*where* $g_s$ *is defined in Assumption (R1). Furthermore, define* $\eta := 1 - \psi_0$ *with* $\psi_0$ *from Assumption (O1). We have* $\mathbf{H}_k^a$ *is asymptotically invertible with* $\|\mathbf{H}_k^a\|_F = O_P(1)$, *and*

$$\frac{1}{d_k} \sum_{j=1}^{d_k} \big\| \widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{k,j\cdot} \big\|^2 = \frac{1}{d_k} \big\| \widehat{\mathbf{Q}}_k - \mathbf{Q}_k \mathbf{H}_k^{a\prime} \big\|_F^2$$
$$= O_P \Bigg( d_k^{2(\alpha_{k,1} - \alpha_{k,r_k}) - 1} \bigg[ \Big( \frac{1}{T d_{\text{-}k}} + \frac{1}{d_k} \Big) \frac{d^2}{g_s^2} + \min \Big\{ \frac{1}{T}, \frac{\eta^2}{(1-\eta)^2} \Big\} \bigg] \Bigg) = o_P(1).$$

The proof of the theorem is relegated to Section 3.5. The two results in Theorem 3.1 coincide with each other if $\eta = 0$, i.e., there are no missing values.

We present the two results in the theorem to highlight the difficulty of obtaining consistency when there are missing values. Since a factor loading matrix is not uniquely defined, in the second result in Theorem 3.1 we are estimating how close $\widehat{\mathbf{Q}}_k$ is to a version of $\mathbf{Q}_k$ in Frobenius norm, namely $\mathbf{Q}_k \mathbf{H}_k^a$, which is still defining the same factor loading space as $\mathbf{Q}_k$ does. With missing data, such a feat is complicated, in the sense that for the $j$-th row of $\mathbf{Q}_k$, $\widehat{\mathbf{Q}}_{k,j\cdot}$, there corresponds an $\mathbf{H}_{k,j}$ different from $\mathbf{H}_k^a$ in general, so that $\widehat{\mathbf{Q}}_{k,j\cdot}$ is close to $\mathbf{H}_{k,j}\mathbf{Q}_{k,j\cdot}$. The extra rate $\min(1/T, \eta^2/(1-\eta)^2)$ in the second result is essentially measuring how close each $\mathbf{H}_{k,j}$ is to $\mathbf{H}_k^a$. See Lemma 3.3 in Section 3.5 as well.

**Theorem 3.2** *Let all the assumptions in Theorem 3.1 hold, and define*

$$g_\eta := \min\Big\{\frac{1}{T}, \frac{\eta^2}{(1-\eta)^2}\Big\}, \quad g_w := \prod_{k=1}^K d_k^{\alpha_{k,r_k}}.$$

*Suppose we further have* $d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}} = o(d_{\text{-}k})$*, then we have the following:*
*1. The error of the estimated factor series has rate*

$$\begin{aligned}&\big\|\mathbf{vec}(\widehat{\mathcal{F}}_{Z,t}) - \big(\mathbf{H}_K^{a'}\otimes\cdots\otimes\mathbf{H}_1^{a'}\big)^{-1}\mathbf{vec}(\mathcal{F}_{Z,t})\big\|^2\\ &= O_P\Big(\max_{k\in[K]}\Big\{T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1} + d^2g_s^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1} + g_\eta g_s d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big\} + \frac{d}{g_w}\Big).\end{aligned}$$

*2. For any* $k\in[K], i_k\in[d_k], t\in[T]$*, the squared individual imputation error is*

$$(\widehat{\mathcal{C}}_{t,i_1,\ldots,i_K} - \mathcal{C}_{t,i_1,\ldots,i_K})^2 = \frac{d}{g_w}\cdot O_P\Big\{\frac{1}{Td}\sum_{t=1}^T\sum_{i_1,\ldots,i_K=1}^{d_1,\ldots,d_K}(\widehat{\mathcal{C}}_{t,i_1,\ldots,i_K} - \mathcal{C}_{t,i_1,\ldots,i_K})^2\Big\}.$$

*3. The average imputation error is given by*

$$\begin{aligned}&\frac{1}{Td}\sum_{t=1}^T\sum_{i_1,\ldots,i_K=1}^{d_1,\ldots,d_K}(\widehat{\mathcal{C}}_{t,i_1,\ldots,i_K} - \mathcal{C}_{t,i_1,\ldots,i_K})^2\\ &= O_P\Big(\max_{k\in[K]}\Big\{T^{-1}d_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1} + dg_s^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1} + d^{-1}g_\eta g_s d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big\} + \frac{1}{g_w}\Big).\end{aligned}$$

The proof can be found in Section 3.5, which utilises some rates from the proof of Theorem 3.3 (without the need for extra rate restrictions like Theorem 3.3 though). The complication of missing data comes explicitly from the rate $g_\eta$. The average squared imputation error in result 3 improves upon individual squared error in result 2 when weak factors exist, with degree of improvements larger when the difference in strength of factors is larger.

Our rate can be considered a generalisation of approximate factor models to a general order tensor, with general factor strengths and missing data, see the comparison of our results with others' below Corollary 3.1. Such generalisations have certainly revealed that when there are weak factors, especially when the strongest and weakest factor strengths are quite different, those rates of convergence greatly suffer.

**Corollary 3.1** *(Simplified Theorem 3.1 and 3.2 under pervasive factors). Let Assumption (O1), (M1), (F1), (L1), (E1) and (E2) hold. If all factors are pervasive such that* $\alpha_{k,j} = 1$ *for all* $k \in [K], j \in [r_k]$*, then with the renormalised loading and core factor estimators defined as* $\widehat{\mathbf{A}}_k = \sqrt{d_k}\,\widehat{\mathbf{Q}}_k$ *and* $\widehat{\mathcal{F}}_t = \widehat{\mathcal{F}}_{Z,t}/\sqrt{d}$*, we have the following:*

*1. The (renormalised) loading estimator is consistent such that for any* $k \in [K]$*,*

$$\frac{1}{d_k}\sum_{j=1}^{d_k}\big\|\widehat{\mathbf{A}}_{k,j\cdot} - \mathbf{H}_{k,j}\mathbf{A}_{k,j\cdot}\big\|^2 = O_P\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k}\Big) = o_P(1),$$

$$\frac{1}{d_k}\sum_{j=1}^{d_k}\big\|\widehat{\mathbf{A}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{A}_{k,j\cdot}\big\|^2 = O_P\Big\{\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k} + \min\Big(\frac{1}{T}, \frac{\eta^2}{(1-\eta)^2}\Big)\Big\} = o_P(1).$$

*2. The (renormalised) core factor estimator is consistent such that for any* $t \in [T]$*,*

$$\big\|\mathbf{vec}(\widehat{\mathcal{F}}_t) - \big(\mathbf{H}_K^{a'}\otimes\cdots\otimes\mathbf{H}_1^{a'}\big)^{-1}\mathbf{vec}(\mathcal{F}_t)\big\|^2 = O_P\Big\{\max_{k\in[K]}\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k^2}\Big) + \min\Big(\frac{1}{T}, \frac{\eta^2}{(1-\eta)^2}\Big)\Big\}.$$

*3. The imputation is consistent both for each entry and on average (with the same rate), such that for any* $k \in [K], i_k \in [d_k], t \in [T]$*,*

$$(\widehat{\mathcal{C}}_{t,i_1,\ldots,i_K} - \mathcal{C}_{t,i_1,\ldots,i_K})^2 = O_P\Big\{\max_{k\in[K]}\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k^2}\Big) + \min\Big(\frac{1}{T}, \frac{\eta^2}{(1-\eta)^2}\Big) + \frac{1}{d}\Big\}.$$

When $K = 1$ with missing data, result 1 has rate $1/\min(d_1, T)$, which is the same as the rate in Theorem 1 of Xiong and Pelger (2023). If $K = 2$ and $\eta = 0$ (i.e, no missing values), result 1 has rate $1/\min(d_k, Td_{\text{-}k})$, which is consistent with Theorem 1 of Chen and Fan (2023), for example. For a general order-$K$ tensor without missing data (i.e., $\eta = 0$), our Lemma 3.5 in Section 3.5 states that

$$\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{Q}_{k,j\cdot}\|^2 = O_P\Big(\frac{1}{Td} + \frac{1}{d_k^3}\Big), \quad \text{implying} \quad \frac{1}{d_k}\|\widehat{\mathbf{A}}_k - \mathbf{A}_k\mathbf{H}_k^a\|_F^2 = O_P\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k^2}\Big),$$

which aligns with Theorem 3.1 of Barigozzi et al. (2023b). Note that the rate in Lemma 3.5 is improved at no cost if all factors are pervasive from the proof of the lemma. For more technical details, see (4.40) in the proof of Lemma 4.5 where the decomposition therein shows how the rate of convergence for the loading estimators can be improved.

If $K \geq 2$ and $\eta = 0$, result 3 has rate

$$\max_{k\in[K]}\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k^2}\Big) + \frac{1}{d} \asymp \frac{1}{\min(Td_{\text{-}1}, \ldots, Td_{\text{-}K}, d_1^2, \ldots, d_K^2)}.$$

This rate is the same as the result in Theorem 4 of Chen and Fan (2023) for $K = 2$, which is a rate for estimating the common component. On the other hand, if $\eta$ is a constant and $K = 1$, then result 3 becomes $d_1^{-1}+T^{-1} \asymp 1/\min(d_1, T)$, which is the same rate as result 3 of Theorem 2 in Xiong and Pelger (2023).

**Remark 3.3** *From the discussion below Corollary 3.1, our result would boil down to that in Xiong and Pelger (2023), but we point out that the rate of convergence, e.g. for the imputation, suffers from large portion of missing data due to the rate* $\min(1/T, \eta^2/(1-\eta)^2)$ *which can be neglected for* $K = 1$, *i.e., vector time series where* $1/Td_{\text{-}k} \equiv 1/T$. *This is also manifested throughout the derivation of theoretical results, where the improved rate* $1/Td_{\text{-}k}$ *from other modes in a tensor factor models is undermined by the essentially the general missing pattern considered in our setup. That said, we should expect to recover the classical rates on common component estimators in tensor factor models if the observational pattern is regular, which then becomes unrealistic and less useful for a general tensor time series. For example, we manage to generalize the TALL-WIDE imputation algorithm in Bai and Ng (2021) for tensor time series, also mentioned in Section 3.4.1, but this track of procedure is even hardly valid when the data is missing at random. Hence it is celebrating to see how our procedure is capable of imputing on even pessimistically observed tensor data. To further highlight the advantage of our method, we refer to the comparison between our method and an iterative vectorisation-based algorithm in Section 3.4.1.*

### 3.3.3 Inference on the factor loadings

We establish asymptotic normality of the factor loadings for inference purpose. In Section 3.3.4 we present the covariance matrix estimator for practical use of our asymptotic normality result. First, we define

$$\mathbf{H}_k^{a,*} := \text{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}))^{1/2} \cdot \mathbf{D}_k^{-1/2}\Upsilon_k'\mathbf{Z}_k^{1/2}, \tag{3.12}$$

where $\mathbf{D}_k := \text{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k})\,\text{diag}\{\lambda_1(\mathbf{A}_k'\mathbf{A}_k), \dots, \lambda_{r_k}(\mathbf{A}_k'\mathbf{A}_k)\}$, and $\Upsilon_k$ is the eigenvector matrix of $\text{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}) \cdot g_s^{-1} d_k^{\alpha_{k,1}-\alpha_{k,r_k}}\, \mathbf{Z}_k^{1/2}\boldsymbol{\Sigma}_{A,k}\mathbf{Z}_k^{1/2}$. It turns out $\mathbf{H}_k^{a,*}$ is the probability limit of $\mathbf{H}_k^a$ defined in (3.11). Before presenting our results, we need three additional assumptions.

(L2) (Eigenvalues). *For any* $k \in [K]$, *the eigenvalues of the* $r_k \times r_k$ *matrix* $\boldsymbol{\Sigma}_{A,k}\mathbf{Z}_k$ *from Assumption (L1) are distinct.*

(AD1) *Define* $\omega_B := d_{\text{-}k}^{-1} d_k^{2\alpha_{k,r_k}-3\alpha_{k,1}} g_s^2$ *and the following,*

$$\boldsymbol{\Xi}_{k,j} := \operatorname*{plim}_{T,d_1,\dots,d_K\to\infty} \text{Var}\Big\{ \sum_{i=1}^{d_k} \mathbf{Q}_{k,i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{k,ij,h}|} \sum_{t\in\psi_{k,ij,h}} \text{mat}_k(\mathcal{E}_t)_{jh}(\mathbf{A}_{\text{-}k})_{h\cdot}' \text{mat}_k(\mathcal{F}_t)'\mathbf{A}_{k,i\cdot} \Big\},$$

*then we assume* $T\omega_B \cdot \left\|\mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\boldsymbol{\Xi}_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}\right\|_F$ *is of constant order.*

(AD2) *Define the filtration* $\mathcal{G}^T := \sigma(\cup_{s=1}^T \mathcal{G}_s)$ *with* $\mathcal{G}_s := \sigma(\{\mathcal{M}_{t,i_1,\ldots,i_K} | t \leq s\}, \mathbf{A}_1, \ldots, \mathbf{A}_K)$*, and*

$$\Delta_{F,k,ij,h} := \frac{1}{|\psi_{k,ij,h}|}\sum_{t\in\psi_{k,ij,h}} \mathrm{mat}_k(\mathcal{F}_t)\mathbf{v}_{k,h}\mathbf{v}_{k,h}'\mathrm{mat}_k(\mathcal{F}_t)' - \frac{1}{T}\sum_{t=1}^T \mathrm{mat}_k(\mathcal{F}_t)\mathbf{v}_{k,h}\mathbf{v}_{k,h}'\mathrm{mat}_k(\mathcal{F}_t)',$$

*where* $\mathbf{v}_{k,h} := [\otimes_{l\in[K]\setminus\{k\}}\mathbf{A}_l]_{h\cdot}$*. With* $\mathbf{Q}_k$ *being the normalised mode-k factor loading defined below Assumption (L1), we have for every* $k \in [K], j \in [d_k]$*, for a function* $h_{k,j} : \mathbb{R}^{r_k} \to \mathbb{R}^{r_k \times r_k}$*,*

$$\sqrt{Td_k^{\alpha_{k,r_k}}} \cdot \mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\mathbf{A}_{k,i\cdot}'\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\mathbf{A}_{k,j\cdot}$$
$$\to \mathcal{N}(\mathbf{0}, \mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}h_{k,j}(\mathbf{A}_{k,j\cdot})(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}) \quad \mathcal{G}^T\text{-stably.}$$

Assumption (AD1) guarantees a part of the covariance matrix of the asymptotic normality in Theorem 3.3 is of constant order. It can be regarded as a lower bound condition which is necessary for the dominance of a certain term involved in the asymptotic normality. Since we show the upper bound of $T\omega_B \cdot \left\|\mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\boldsymbol{\Xi}_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}\right\|_F$ is of constant order in the proof of Theorem 3.3, this assumption is not particularly strong.

Assumption (AD2) is required since the missing data creates a discrepancy term $\Delta_{F,k,ij,h}$ as defined in the assumption. This assumption is also parallel to Assumption G3.5 in Xiong and Pelger (2023). We demonstrate how this assumption is satisfied with Assumption (O1), (F1), (L1) and two additional but simpler assumptions in Proposition 3.1 in Section 3.3.6.

**Theorem 3.3** *Let all the assumptions under Theorem 3.2 hold, in addition to (L2), (AD1) and (AD2) above. With* $r_k$ *fixed and* $d_k, T \to \infty$ *for* $k \in [K]$*, suppose* $Td_{\text{-}k} = o(d_k^{\alpha_{k,1}+\alpha_{k,r_k}})$*, then*

$$\sqrt{Td_k^{\alpha_{k,r_k}}} \cdot (\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{Q}_{k,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}(Td_k^{\alpha_{k,r_k}} \cdot \boldsymbol{\Xi}_{k,j} + h_{k,j}(\mathbf{A}_{j\cdot}))(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}).$$

*Furthermore, if* $Td^{-1}g_s^2 g_\eta d_k^{1+\alpha_{k,1}-3\alpha_{k,r_k}} = o(1)$ *is also satisfied, then*

$$\sqrt{T\omega_B} \cdot (\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{Q}_{k,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, T\omega_B \cdot \mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\boldsymbol{\Xi}_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}).$$

If all factors are pervasive, the rate condition $Td_{\text{-}k} = o(d_k^{\alpha_{k,1}+\alpha_{k,r_k}})$ reduces to $Td_{\text{-}k} = o(d_k^2)$, which is equivalent to the condition needed for asymptotic normality in Bai (2003) for $K = 1$ and Chen and Fan (2023) for $K = 2$. The first asymptotic normality result is compatible to Theorem 2.1 of Xiong and Pelger (2023) when all factors are pervasive. In their Theorem 2.1, the $\Gamma_{\Lambda,j}^{obs}$ is in fact of rate $N^{-1}$, so that the normalising rate is $\sqrt{TN}$, which is exactly $\sqrt{Td_1}$ in our first result when $K = 1$.

Suppose all factors are pervasive. The rate condition $Td^{-1}g_s^2 g_\eta d_k^{1+\alpha_{k,1}-3\alpha_{k,r_k}} = o(1)$ is automatically satisfied when there is no missing data, i.e., $\eta = 0$ so that $g_\eta = 0$. If so, the rate of convergence is $\sqrt{T\omega_B} = \sqrt{Td}$, which is compatible to Theorem 2.1, Theorem 2.2 of Chen and Fan (2023) and Theorem 3.2 of Barigozzi et al. (2023b) (after our normalisation to their factor loading matrices). The condition is also satisfied when there is only finite number of missing data, so that $\eta \asymp T^{-1}$ and $g_\eta \asymp T^{-2}$, and $d_1, d_2 = o(T)$ for $K = 2$.

**Remark 3.4** *We do not establish asymptotic normality for the estimated factor series and common components. The reason is that for tensor with $K > 1$, the decomposition in the estimated factor series and the common components cannot be dominated by terms that are asymptotically normal. This is also the reason why Chen and Fan (2023) does not include asymptotic normality for the estimated factor series and common components. Barigozzi et al. (2023b) constructs asymptotic normality for the core factor built upon their projection estimator $\widetilde{\mathcal{F}}_t$, which is sensible as the projecting loading estimator already has an improved rate. In comparisons, the rate of any PCA-type estimators, such as the one in Chen and Fan (2023) for matrix data and the one in our case for general tensors, is insufficient for a potentially asymptotically Gaussian term to be dominating. The main goal of this chapter is to impute missing entries, and existing methods on tensor factor models using Tucker decomposition should be able to be applied with all missing entries replaced by the consistent imputations.*

### 3.3.4 Estimation of the asymptotic covariance matrix

In order to carry out inferences for the factor loadings using Theorem 3.3, we need to estimate the asymptotic covariance matrix for $\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{k,j\cdot}$. To this end, we use the heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West, 1987) based on $\{\widehat{\mathbf{Q}}_k, \text{mat}_k(\widehat{\mathcal{C}}_t), \text{mat}_k(\widehat{\mathcal{E}}_t)\}_{t\in[T]}$, where

$$
\begin{aligned}
\text{mat}_k(\widehat{\mathcal{C}}_t) &:= (\widehat{\mathbf{Q}}_k)\text{mat}_k(\widehat{\mathcal{F}}_{Z,t})(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_{k+1} \otimes \widehat{\mathbf{Q}}_{k-1} \otimes \cdots \otimes \widehat{\mathbf{Q}}_1)', \\
\text{mat}_k(\widehat{\mathcal{E}}_t) &:= \text{mat}_k(\mathcal{Y}_t) - \text{mat}_k(\widehat{\mathcal{C}}_t).
\end{aligned}
$$

With a tuning parameter $\beta$ that $\beta \to \infty$, $\beta/(Td_k^{\alpha_{k,r_k}})^{1/4} \to 0$, we define two HAC estimators

$$
\begin{aligned}
\widehat{\mathbf{\Sigma}}_{HAC} &:= \mathbf{D}_{k,0,j} + \sum_{\nu=1}^{\beta}\Big(1 - \frac{\nu}{1+\beta}\Big)\Big(\mathbf{D}_{k,\nu,j} + \mathbf{D}'_{k,\nu,j}\Big), \\
\widehat{\mathbf{\Sigma}}_{HAC}^{\Delta} &:= \mathbf{D}_{k,0,j}^{\Delta} + \sum_{\nu=1}^{\beta}\Big(1 - \frac{\nu}{1+\beta}\Big)\Big(\mathbf{D}_{k,\nu,j}^{\Delta} + (\mathbf{D}_{k,\nu,j}^{\Delta})'\Big), \text{ where}
\end{aligned}
$$

$$\begin{aligned}\mathbf{D}_{k,\nu,j} &:= \sum_{t=1+\nu}^{T} \Big\{ \sum_{i=1}^{d_k} \Big( \frac{1}{T} \sum_{s=1}^{T} \widehat{\mathbf{D}}_k^{-1} \widehat{\mathbf{Q}}_k' \widehat{\mathbf{C}}_{(k),s} \widehat{\mathbf{C}}_{(k),s,i\cdot} \Big) \sum_{h=1}^{d_{\text{-}k}} \frac{\mathbb{1}\{t \in \psi_{k,ij,h}\}}{|\psi_{k,ij,h}|} \widehat{\mathbf{E}}_{(k),t,jh} \widehat{\mathbf{C}}_{(k),t,ih} \Big\} \\ &\quad \cdot \Big\{ \sum_{i=1}^{d_k} \Big( \frac{1}{T} \sum_{s=1}^{T} \widehat{\mathbf{D}}_k^{-1} \widehat{\mathbf{Q}}_k' \widehat{\mathbf{C}}_{(k),s} \widehat{\mathbf{C}}_{(k),s,i\cdot} \Big) \sum_{h=1}^{d_{\text{-}k}} \frac{\mathbb{1}\{t-\nu \in \psi_{k,ij,h}\}}{|\psi_{k,ij,h}|} \widehat{\mathbf{E}}_{(k),t-\nu,jh} \widehat{\mathbf{C}}_{(k),t-\nu,ih} \Big\}', \\ \mathbf{D}^{\Delta}_{k,\nu,j} &:= \sum_{t=1+\nu}^{T} \Big\{ \sum_{i=1}^{d_k} \Big( \frac{1}{T} \sum_{s=1}^{T} \widehat{\mathbf{D}}_k^{-1} \widehat{\mathbf{Q}}_k' \widehat{\mathbf{C}}_{(k),s} \widehat{\mathbf{C}}_{(k),s,i\cdot} \Big) \cdot \sum_{h=1}^{d_{\text{-}k}} \Big( \frac{\mathbb{1}\{t \in \psi_{k,ij,h}\}}{|\psi_{k,ij,h}|} \widehat{\mathbf{C}}_{(k),t,ih} \widehat{\mathbf{C}}_{(k),t,jh} \\ &\quad - \frac{1}{T} \widehat{\mathbf{C}}_{(k),t,ih} \widehat{\mathbf{C}}_{(k),t,jh} \Big) \Big\} \cdot \Big\{ \sum_{i=1}^{d_k} \Big( \frac{1}{T} \sum_{s=1}^{T} \widehat{\mathbf{D}}_k^{-1} \widehat{\mathbf{Q}}_k' \widehat{\mathbf{C}}_{(k),s} \widehat{\mathbf{C}}_{(k),s,i\cdot} \Big) \\ &\quad \cdot \sum_{h=1}^{d_{\text{-}k}} \Big( \frac{\mathbb{1}\{t-\nu \in \psi_{k,ij,h}\}}{|\psi_{k,ij,h}|} \widehat{\mathbf{C}}_{(k),t-\nu,ih} \widehat{\mathbf{C}}_{(k),t-\nu,jh} - \frac{1}{T} \widehat{\mathbf{C}}_{(k),t-\nu,ih} \widehat{\mathbf{C}}_{(k),t-\nu,jh} \Big) \Big\}',\end{aligned}$$

where $\widehat{\mathbf{C}}_{(k),s} := \text{mat}_k(\widehat{\mathcal{C}}_s)$ and $\widehat{\mathbf{E}}_{(k),s} := \text{mat}_k(\widehat{\mathcal{E}}_s)$.

**Theorem 3.4** *Let all the assumptions under Theorem 3.2 hold, in addition to (L2), (AD1) and (AD2) above. With $r_k$ fixed and $d_k, T \to \infty$ for $k \in [K]$, suppose also the rate for the individual common component imputation error in result 2 of Theorem 3.2 is $o(1)$, together with $Td_{\text{-}k} = o(d_k^{\alpha_{k,1}+\alpha_{k,r_k}})$ and $d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}[(Td_{\text{-}k})^{-1} + d_k^{-1}]d^2 g_s^{-2} = o(1)$. Then*

1. $\widehat{\mathbf{D}}_k^{-1}\widehat{\boldsymbol{\Sigma}}_{HAC}\widehat{\mathbf{D}}_k^{-1}$ *is consistent for* $\mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\boldsymbol{\Xi}_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}$;
2. $\widehat{\mathbf{D}}_k^{-1}\widehat{\boldsymbol{\Sigma}}^{\Delta}_{HAC}\widehat{\mathbf{D}}_k^{-1}$ *is consistent for* $(Td_k^{\alpha_{k,r_k}})^{-1}\mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}h_{k,j}(\mathbf{A}_{k,j\cdot})(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}$;
3. $(\widehat{\boldsymbol{\Sigma}}_{HAC} + \widehat{\boldsymbol{\Sigma}}^{\Delta}_{HAC})^{-1/2}\widehat{\mathbf{D}}_k(\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{Q}_{k,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_{r_k})$.

The extra rate assumption $d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}[(Td_{\text{-}k})^{-1} + d_k^{-1}]d^2 g_s^{-2} = o(1)$ makes sure that we have Frobenius norm consistency for $\widehat{\mathbf{Q}}_k$ from Theorem 3.1. The imputation error from result 2 of Theorem 3.2 also has rate going to 0 when are all factors are pervasive, for instance. With result 3 in particular, we can perform inferences on any rows of $\widehat{\mathbf{Q}}_k$. Practical performances of result 3 is demonstrated in Section 3.4.1. The reason that we need two HAC estimators is that similar to Theorem 3.1, there is a component for missing data, arising from the fact that $\mathbf{H}_{k,j}$ is different from $\mathbf{H}_k^a$ for each $j \in [d_k]$ in general.

### 3.3.5 Estimation of number of factors

The reconstructed mode-$k$ sample covariance matrix $\widehat{\mathbf{S}}_k$ is in fact estimating a complete-sample version of a matrix $\mathbf{R}_k^*$, where

$$\mathbf{R}_k^* := \frac{1}{T}\sum_{t=1}^{T} \mathbf{Q}_k \text{mat}_k(\mathcal{F}_{Z,t})\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k \text{mat}_k(\mathcal{F}_{Z,t})'\mathbf{Q}_k', \tag{3.13}$$

and $\mathcal{F}_{Z,t}$ and $\mathbf{\Lambda}_k$ are defined in (3.2). It turns out that $\lambda_j(\widehat{\mathbf{S}}_k) \asymp_P \lambda_j(\mathbf{R}_k^*)$ for $j \in [r_k]$, and

$$\lambda_j(\mathbf{R}_k^*) \asymp_P d_k^{\alpha_{k,j}-\alpha_{k,1}} g_s, \quad g_s := \prod_{k=1}^{K} d_k^{\alpha_{k,1}} \text{ as defined in (R1).}$$

We have the following theorem.

**Theorem 3.5** *Let Assumption (O1), (M1), (F1), (L1), (E1), (E2) and (R1) hold. Moreover, assume*

$$\begin{cases} dg_s^{-1} d_k^{\alpha_{k,1}-\alpha_{k,r_k}} [(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}] = o(d_k^{\alpha_{k,j+1}-\alpha_{k,j}}), & j \in [r_k - 1] \textit{ with } r_k \geq 2; \\ dg_s^{-1} [(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}] = o(1), & r_k = 1. \end{cases}$$

*Then $\widehat{r}_k$ is a consistent estimator of $r_k$, where*

$$\widehat{r}_k := \arg\min_{\ell} \left\{ \frac{\lambda_{\ell+1}(\widehat{\mathbf{S}}_k) + \xi}{\lambda_\ell(\widehat{\mathbf{S}}_k) + \xi}, \ \ell \in \big[\lfloor d_k/2 \rfloor\big] \right\}, \quad \xi \asymp d\big[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}\big]. \tag{3.14}$$

The extra rate assumption is satisfied, for instance, when all factors corresponding to $\mathbf{A}_k$ are pervasive. An eigenvalue-ratio estimator is considered in Lam and Yao (2012) and Ahn and Horenstein (2013), while a perturbed eigenvalue ratio estimator is considered in Pelger (2019). However, all of these estimators are for a vector time series factor model. Our estimator $\widehat{r}_k$ in (3.14) extracts eigenvalues from $\widehat{\mathbf{S}}_k$, which is not necessarily positive semi-definite. The addition of $\xi$ can make $\widehat{\mathbf{S}}_k + \xi \mathbf{I}_{d_k}$ positive semi-definite, while stabilizing the estimator. We naturally assume that $r_k < d_k/2$, which is a very reasonable assumption for all applications of factor models. In fact, our recommended choice of $\xi$ is

$$\xi = \frac{1}{5} d\big[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}\big].$$

The requirement $\xi \asymp d\big[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}\big]$ ensures $\xi = o_P(\lambda_{r_k}(\widehat{\mathbf{S}}_k))$ from our rate assumption in the theorem. Our simulations in Section 3.4.1 suggest that this proposal works well.

### 3.3.6 *How Assumption (AD2) can be implied

This section details how Assumption (AD2) can be implied from simpler assumptions. Readers can skip this part and go straight to the next section for a more integral reading experience. We begin by presenting a proposition.

**Proposition 3.1** *Let Assumption (O1), (F1), (L1) hold. For a given $k \in [K], j \in [d_k]$, assume:*

1. *The mode-$k$ factor is strong enough such that $\alpha_{k,r_k} > 4/5$, and $d_k^{\alpha_{k,1}-\alpha_{k,r_k}} T^{-\epsilon/2} = o(1)$ with some $\epsilon \in (0,1)$.*

2. *There exists some* $\psi_{k,ij}$ *such that* $\psi_{k,ij} = \psi_{k,ij,h}$ *for any* $i \in [d_k], h \in [d_{\text{-}k}]$. *Furthermore, there exists* $\omega_{\psi,k,j}$ *such that for any* $t \in [T]$, *as* $d_k, T \to \infty$,

$$d_k^{-2} \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} \Big( \frac{T \cdot \mathbb{1}\{t \in \psi_{k,ij}\}}{|\psi_{k,ij}|} - 1 \Big) \Big( \frac{T \cdot \mathbb{1}\{t \in \psi_{k,lj}\}}{|\psi_{k,lj}|} - 1 \Big) \xrightarrow{p} \omega_{\psi,k,j}.$$

*With the above, Assumption (AD2) is satisfied. Moreover, condition 2 can be replaced by missing at random over all elements such that* $\mathbb{P}(\mathcal{M}_{t,i_1,\dots,i_K} = 1)$ *is the same for any* $t, i_1, \dots, i_K$.

Condition 1 and 2 in Proposition 3.1 are on the factor strength and missingness pattern, respectively. Condition 1 is trivially satisfied if all factors are pervasive. If the data is also missing at random, Proposition 3.1 holds.

**Remark 3.5** *Condition 2 can be satisfied by assuming that in* $\text{mat}_k(\mathcal{Y}_t)$, *all the elements are missing at random over rows with probability* $1 - p_0$, *and meanwhile missing dependently over columns such that* $\psi_{k,ij,1} = \cdots = \psi_{k,ij,d_{\text{-}k}}$ *(which still allows the pattern to be different to certain extent over columns). We then have for each* $t \in [T]$, *as* $d_k, T \to \infty$,

$$\begin{aligned} &d_k^{-2} \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} \Big( \frac{T \cdot \mathbb{1}\{t \in \psi_{k,ij}\}}{|\psi_{k,ij}|} - 1 \Big) \Big( \frac{T \cdot \mathbb{1}\{t \in \psi_{k,lj}\}}{|\psi_{k,lj}|} - 1 \Big) \\ &= d_k^{-2} \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} \Big( \frac{T^2 \cdot \mathbb{1}\{t \in \psi_{k,ij}\} \cdot \mathbb{1}\{t \in \psi_{k,lj}\}}{|\psi_{k,ij}| \cdot |\psi_{k,lj}|} - \frac{T \cdot \mathbb{1}\{t \in \psi_{k,ij}\}}{|\psi_{k,ij}|} - \frac{T \cdot \mathbb{1}\{t \in \psi_{k,lj}\}}{|\psi_{k,lj}|} + 1 \Big) \\ &\xrightarrow{p} p_0^{-1} - 1, \end{aligned}$$

*which is* $\omega_{\psi,k,j}$. *Similar to Assumption S3.2 in Xiong and Pelger (2023), the value of* $\omega_{\psi,k,j}$ *can be regarded as a measure of missingness complexity. It is a parameter related to the variance of the stable convergence, and tends to increase when there is a larger portion of data missing.*

## 3.4 Numerical Results

### 3.4.1 Simulations

We demonstrate the empirical performance of our estimators in this subsection. Note that we do not have comparisons to other imputation methods since to the best of our knowledge, there are no other general imputation methods available for $K > 1$ apart from tensor completion methods for very specific applications as mentioned in the introduction. However, we will make comparison with an alternative approach to impute tensor time series combining Xiong and Pelger (2023) and Chen and Lam (2024b). Under different missing patterns which will be described later, we investigate the performance of the factor loading matrix estimators, the imputation, and the estimator of the number of factors. We also demonstrate asymptotic normality as described in Theorem 3.3, followed by an example plot of a statistical power function using result 3 of Theorem 3.4. Throughout this subsection, each simulation experiment of a particular setting is repeated 1000 times, unless stated otherwise.

For the data generating process, we use model (3.1) together with Assumption (E1), (E2) and (F1). More precisely, the elements in $\mathcal{F}_t$ are independent standardised AR(5) with AR coefficients 0.7, 0.3, -0.4, 0.2, and -0.1. The elements in $\mathcal{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are generated similarly, but their AR coefficients are (-0.7, -0.3, -0.4, 0.2, 0.1) and (0.8, 0.4, -0.4, 0.2, -0.1) respectively. The standard deviation of each element in $\boldsymbol{\epsilon}_t$ is generated by i.i.d. $|\mathcal{N}(0,1)|$.

For each $k \in [K]$, each factor loading matrix $\mathbf{A}_k$ is generated independently with $\mathbf{A}_k = \mathbf{U}_k\mathbf{B}_k$, where each entry of $\mathbf{U}_k \in \mathbb{R}^{d_k \times r_k}$ is i.i.d. $\mathcal{N}(0,1)$, and $\mathbf{B}_k \in \mathbb{R}^{r_k \times r_k}$ is diagonal with the $j$-th diagonal entry being $d_k^{-\zeta_{k,j}}$, $0 \leq \zeta_{k,j} \leq 0.5$. Pervasive (strong) factors have $\zeta_{k,j} = 0$, while weak factors have $0 < \zeta_{k,j} \leq 0.5$. Each entry of $\mathbf{A}_{e,k} \in \mathbb{R}^{d_k \times r_{e,k}}$ is i.i.d. $\mathcal{N}(0,1)$, but has independent probability of 0.95 being set exactly to 0. We set $r_{e,k} = 2$ for all $k \in [K]$ throughout this subsection.

To investigate the performance with missing data, we consider four missing patterns:

- (M-i) Random missing with probability 0.05.
- (M-ii) Random missing with probability 0.3.
- (M-iii) The missing entries have index $(t, i_1, \ldots, i_K)$, where

$$0.5T \leq t \leq T, \quad 1 \leq i_k \leq 0.5d_k \text{ for all } k \in [K].$$

- (M-iv) Conditional random missing such that the unit with index $j$ along mode-1 is missing with probability 0.2 if $(\mathbf{A}_1)_{j,1} \geq 0$, and with probability 0.5 if $(\mathbf{A}_1)_{j,1} < 0$.

To test how robust our imputation is under heavy-tailed distribution, we consider two distributions for the innovation process in generating $\mathcal{F}_t$, $\mathcal{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$: 1) i.i.d. $\mathcal{N}(0,1)$; 2) i.i.d. $t_3$.

**Accuracy in the factor loading matrix estimators and imputations**

For both the factor loading matrix estimators and the imputations, since our procedure for vector time series ($K = 1$) is essentially the same as that in Xiong and Pelger (2023), we show here only the performance for $K = 2, 3$. We use the column space distance

$$\mathcal{D}(\mathbf{Q}, \widehat{\mathbf{Q}}) = \left\| \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}' - \widehat{\mathbf{Q}}(\widehat{\mathbf{Q}}'\widehat{\mathbf{Q}})^{-1}\widehat{\mathbf{Q}}' \right\|$$

for any given $\mathbf{Q}, \widehat{\mathbf{Q}}$, which is a commonly used measure in the literature. For measuring the imputation accuracy, we report the relative mean squared errors (MSE) defined by

$$\text{relative MSE}_{\mathcal{S}} = \frac{\sum_{j\in\mathcal{S}}(\widehat{C}_j - C_j)^2}{\sum_{j\in\mathcal{S}} C_j^2}, \tag{3.15}$$

where $\mathcal{S}$ either denotes the set of all missing, all observed, or all available units.

We consider the following simulation settings:

(Ia) $K = 2, T = 100, d_1 = d_2 = 40, r_1 = 1, r_2 = 2$. All factors are pervasive with $\zeta_{k,j} = 0$ for all $k, j$. All innovation processes in constructing $\mathcal{F}_t, \mathcal{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are i.i.d. standard normal, and missing pattern is (M-i).

(Ib) Same as (Ia), but one factor is weak with $\zeta_{k,1} = 0.2$ for all $k \in [K]$.

(Ic) Same as (Ia), but all innovation processes are i.i.d. $t_3$, and all factors are weak with $\zeta_{k,j} = 0.2$ for all $k, j$.

(Id) Same as (Ic), but $T = 200, d_1 = d_2 = 80$.

(Ie) $K = 3, T = 80, d_1 = d_2 = d_3 = 20, r_1 = r_2 = r_3 = 2$. All factors are pervasive with $\zeta_{k,j} = 0$ for all $k, j$. All innovation processes in constructing $\mathcal{F}_t, \mathcal{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are i.i.d. standard normal, and missing pattern is (M-i).

(If) Same as (Ie), but all factors are weak with $\zeta_{k,j} = 0.2$ for all $k, j$.

(Ig) Same as (If), but $T = 200, d_1 = d_2 = d_3 = 40$.

Settings (Ia)–(Id) have $K = 2$, and settings (Ie)–(Ig) have $K = 3$. They all have missing pattern (M-i), but we have considered all settings with missing patterns (M-ii)–(M-iv), with performance of the factor loading matrix estimators very similar to those with missing pattern (M-i). Hence we are only presenting the results for settings (Ia)–(Ig) in Figure 3.1 for the missing pattern (M-i). The imputation results for the above settings are collected in Table 3.1, together with those under different missing patterns.

We can see from Figure 3.1 that the factor loading matrix estimators perform worse when there are weak factors or when the distribution of the innovation processes is fat-tailed. However, larger dimensions ameliorate the worsen performance. The increase in the loading space distance from $k = 1$ to $k = 2$ in settings (Ia)–(Id) is due to more factors along mode-2, which naturally incurs more errors compared to smaller $r_k$. In comparison, the loading space error shown in the right panel of Figure 3.1 are in line for all modes due to the same number of factors along each mode.

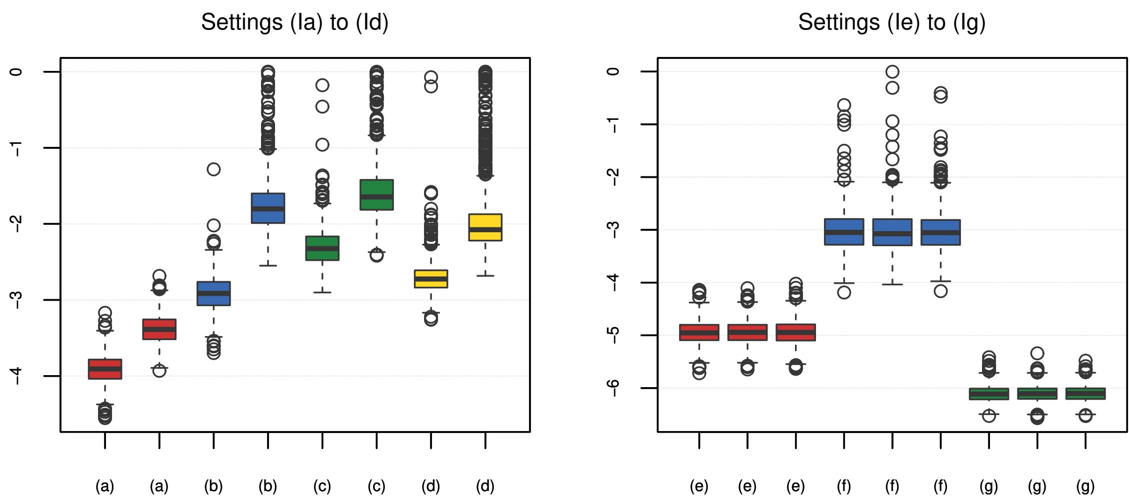


Figure 3.1: Plot of the column space distance $\mathcal{D}(\mathbf{Q}_k, \widehat{\mathbf{Q}}_k)$ (in log-scale) for $k \in [K]$ for missing pattern (M-i), with $K = 2$ on the left panel and $K = 3$ on the right. The horizontal axis is indexed from (a) to (g) to represent Settings (Ia) to (Ig), with the $k$-th boxplot of each setting corresponding to the $k$-th factor loading matrix $\mathbf{Q}_k$ therein. Performance on other missing patterns are very similar, and are omitted.

From Table 3.1, we can see that missing pattern (M-iii) is uniformly more difficult in all settings for imputation. This is understandable as there is a large block of data missing in setting (M-iii), so that we obtain less information towards the "centre" of the missing block. This is also the reason why under (M-iii), the imputation performance for the missing set is worse than the observed set, unlike for other missing patterns where all imputation performances are close.

Random missing in (M-i) and (M-ii) are relatively easier for our imputation procedure to handle. Note that if the TALL-WIDE algorithm in Bai and Ng (2021) were to be extended to the case for $K > 1$, it can handle missing pattern (M-iii), but not (M-i) and (M-ii). The design of our method allows us to handle a wider variety of missing patterns, including random missingness. We want to stress that we have attempted to generalise the TALL-WIDE algorithm to impute high-order time series data for comparisons, yet the method is almost impossible to use in tensor data. The generalisation is also too complicated, and hence is not shown here.

**Performance for the estimation of the number of factors**

We now demonstrate the performance of our ratio estimator $\widehat{r}_k$ in (3.14) for estimating $r_k$ for $K = 1, 2, 3$. For each $k \in [K]$, we set the value of $\xi$ in Theorem 3.5 as $\xi = d[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}]/5$. We have tried a wide range of values other than $1/5$ for $\xi$ in all settings, but $1/5$ is working the best in vast majority of settings; see simulation results on the sensitivity of different $\xi$ in Table 3.4. Hence we do not recommend treating it as a tuning parameter in this section for

| Setting | | K=2 | | | | K=3 | | |
|---|---|---|---|---|---|---|---|---|
| Missing Pattern | $\mathcal{S}$ | (Ia) | (Ib) | (Ic) | (Id) | (Ie) | (If) | (Ig) |
| (M-i) | obs | .002 | .020 | .066 | .039 | 2.61 | 120 | .293 |
| | miss | .002 | .020 | .066 | .039 | 2.63 | 121 | .294 |
| | all | .002 | .020 | .066 | .039 | 2.61 | 120 | .293 |
| (M-ii) | obs | .003 | .025 | .079 | .045 | 5.97 | 154 | .702 |
| | miss | .003 | .025 | .079 | .045 | 6.06 | 155 | .703 |
| | all | .003 | .025 | .079 | .045 | 6.00 | 154 | .702 |
| (M-iii) | obs | .004 | .025 | .079 | .048 | 6.64 | 136 | 1.75 |
| | miss | .009 | .036 | .107 | .061 | 14.7 | 164 | 4.02 |
| | all | .005 | .026 | .083 | .050 | 7.19 | 138 | 1.89 |
| (M-iv) | obs | .004 | .027 | .086 | .047 | 7.75 | 173 | .888 |
| | miss | .004 | .028 | .088 | .047 | 8.49 | 179 | .964 |
| | all | .004 | .027 | .086 | .047 | 8.00 | 175 | .914 |

Table 3.1: Relative MSE for settings (Ia) to (Ig), reported for $\mathcal{S}$ as the set containing respectively observed (obs), missing (miss), and all (all) units. For $K = 3$, all results presented are multiplied by $10^4$.

saving computational time.

We present the results under a fully observed scenario and a missing data scenario for each of the following setting:

(IIa) $K = 1, T = d_1 = 80, r_1 = 2$. All factors are strong with $\zeta_{1,j} = 0$ for all $j$. All innovation processes involved are i.i.d. $\mathcal{N}(0, 1)$. We try missing patterns (M-ii), (M-iii) and (M-iv).

(IIb) Same as (IIa), but $\zeta_{1,1} = 0.1$ and we only try missing pattern (M-ii).

(IIc) Same as (IIb), but factors are weak with $\zeta_{1,1} = 0.1$ and $\zeta_{1,2} = 0.15$.

(IId) Same as (IIc), but $T = 160$.

(IIIa) $K = 2, T = d_1 = d_2 = 40, r_1 = 2, r_2 = 3$. For all $k, j$, we set $\zeta_{k,j} = 0$. All innovation processes involved are i.i.d. $\mathcal{N}(0, 1)$, and we only try missing pattern (M-ii).

(IIIb) Same as (IIIa), but all factors are weak with $\zeta_{k,j} = 0.1$ for all $k, j$.

(IIIc) Same as (IIIb), but $T = d_1 = d_2 = 80$.

(IVa) $K = 3, T = d_1 = d_2 = d_3 = 20, r_1 = 2, r_2 = 3, r_3 = 4$. For all $k, j$, we set $\zeta_{k,j} = 0$. All innovation processes involved are i.i.d. $\mathcal{N}(0, 1)$, and we only try missing pattern (M-ii).

(IVb) Same as (IVa), but all innovation processes are i.i.d. $t_3$.

(IVc) Same as (IVa), but $T = 40$.

Since estimating the number of factors with missing data is new to the literature, it is of interest to explore the accuracy of the estimator under different missing patterns. Hence we explore different missing patterns in setting (IIa). Extensive experiments (not shown here) on the imputation accuracy using misspecified number of factors show that underestimation is harmful, while slight overestimation hardly worsen the performance of the imputations. Thus, for each of the above settings, we also compare the performance using re-imputation and iTIP-ER by Han et al. (2022), where the re-imputation is done by using both $\widehat{r}_k$ and $\widehat{r}_k + 1$ to avoid information loss due to underestimating the number of factors, see Table 3.2 and Table 3.3.

| Setting (IIa) (True $r_1 = 2$) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Missing Pattern | $\widehat{r}$ | $\widehat{r}_{\text{re},0}$ | $\widehat{r}_{\text{re},1}$ | $\widehat{r}_{\text{iTIP,re},0}$ | $\widehat{r}_{\text{iTIP,re},1}$ | $\widehat{r}_{\text{full}}$ | $\widehat{r}_{\text{iTIP,full}}$ |
| | Mean$_{(\text{SD})}$ | | | | | | |
| (M-ii) | $1.98_{(.13)}$ | $1.98_{(.13)}$ | $2.00_{(.06)}$ | $1.97_{(.18)}$ | $1.97_{(.22)}$ | | |
| (M-iii) | $1.92_{(.27)}$ | $1.93_{(.26)}$ | $1.97_{(.20)}$ | $1.90_{(.30)}$ | $1.92_{(.31)}$ | $1.99_{(.10)}$ | $1.92_{(.28)}$ |
| (M-iv) | $1.98_{(.14)}$ | $1.98_{(.14)}$ | $2.01_{(.08)}$ | $1.97_{(.17)}$ | $1.98_{(.24)}$ | | |
| | Correct Proportion | | | | | | |
| (M-ii) | .982 | .982 | .996 | .967 | .949 | | |
| (M-iii) | .921 | .93 | .96 | .901 | .898 | .99 | .917 |
| (M-iv) | .979 | .979 | .993 | .97 | .943 | | |

Table 3.2: Results for setting (IIa). Each column reports the mean and SD (subscripted, in bracket) of the estimated number of factors over 1000 replications, followed by the correct proportion of the estimates. The estimator $\widehat{r}$ is our proposed estimator; $\widehat{r}_{\text{re},0}$ and $\widehat{r}_{\text{re},1}$ are similar but used imputed data where the imputation is done using the number of factors as $\widehat{r}$ and $\widehat{r}+1$, respectively; $\widehat{r}_{\text{iTIP,re},0}$ and $\widehat{r}_{\text{iTIP,re},1}$ are iTIP-ER on imputed data (using $\widehat{r}$ and $\widehat{r}+1$ respectively); $\widehat{r}_{\text{full}}$ and $\widehat{r}_{\text{iITP,full}}$ are our estimator and iTIP-ER on fully observed data (in green), respectively.

From both Table 3.2 and 3.3, it is easy to see that our proposed method generally gives more accurate estimates than iTIP-ER, and it is clear that the re-imputation estimate is at least as good as the initial estimate. In fact, $\widehat{r}_{\text{re},1}$ outperforms $\widehat{r}_{\text{full}}$ which is based on full observation.

**Sensitivity for the tuning parameter in Theorem 3.5**

We provide some simulation results on the performance of our number of factor estimator $\widehat{r}_k$ relative to the choice of $\xi$ in Theorem 3.5. For demonstration purpose, we adapt the general setup depicted in Section 3.4 to generate the loading matrices, factor and noise series, except that only $\mathcal{N}(0,1)$ is used to generate the innovation process. On the dimension of data, we consider order $K = 1, 2, 3$ such that

| Correct Proportion | | | | | | | |
|---|---|---|---|---|---|---|---|
| Setting | $\widehat{r}$ | $\widehat{r}_{\text{re},0}$ | $\widehat{r}_{\text{re},1}$ | $\widehat{r}_{\text{iTIP,re},0}$ | $\widehat{r}_{\text{iTIP,re},1}$ | $\widehat{r}_{\text{full}}$ | $\widehat{r}_{\text{iITP,full}}$ |
| $K = 1$ (True $r_1 = 2$) | | | | | | | |
| (IIb) | .556 | .556 | .886 | .526 | .765 | .633 | .53 |
| (IIc) | .626 | .626 | .762 | .594 | .668 | .67 | .539 |
| (IId) | .791 | .791 | .817 | .794 | .837 | .812 | .767 |
| $K = 2$ (True $(r_1, r_2) = (2, 3)$) | | | | | | | |
| (IIIa) | 1 | 1 | 1 | .995 | .995 | 1 | .994 |
| (IIIb) | .978 | .978 | .987 | .985 | .989 | .981 | .986 |
| (IIIc) | .999 | .999 | 1 | 1 | .996 | .999 | 1 |
| $K = 3$ (True $(r_1, r_2, r_3) = (2, 3, 4)$) | | | | | | | |
| (IVa) | 1 | 1 | 1 | .987 | .987 | 1 | .988 |
| (IVb) | .996 | .996 | .999 | .991 | .991 | 1 | .991 |
| (IVc) | 1 | 1 | 1 | .999 | 1 | 1 | 1 |

Table 3.3: Results for settings (II), (III), and (IV), excluding (IIa). Refer to Table 3.2 for the definitions of different estimators. The missing pattern concerned in all settings is (M-ii).

- $K = 1$: $T = 160$, $d_1 = 80$;
- $K = 2$: $T = d_1 = d_2 = 40$;
- $K = 3$: $T = d_1 = d_2 = d_3 = 20$;

where we assume two pervasive factors on each mode, i.e. true $r_k = 2$ and $\zeta_{k,j} = 0$ for all $k, j$.

To show the robustness of our choice of $\xi$, each setting is repeated 1000 times under four missing patterns: fully observed, (M-i), (M-ii) and (M-iii). See Section 3.4 for the description of these missing patterns.

As the dimension and the number of factors along each tensor mode is the same (within any setting), it suffices to study the correct proportion of $\widehat{r}_1 = r_1 = 2$. The result for different values of $\xi \in \{0.002, 0.02, 0.2, 2, 20\}$ is shown in Table 3.4. It is clear from the results that relatively small values of $\xi$ should help to estimate the number of factors consistently. In particular, $\xi$ ranges from $0.02$ to $0.2$ should work sufficiently well.

### Asymptotic normality

We present the asymptotic normality results for $K = 1, 2, 3$ respectively. When the data is a vector time series ($K = 1$), our approach is similar to Xiong and Pelger (2023), but their proposed covariance estimator for the asymptotic normality includes information at lag 0 only

| Correct Proportion of $\widehat{r}_1 = r_1 = 2$ | | | | | |
|---|---|---|---|---|---|
| Missing Pattern | $\xi = 0.002$ | $\xi = 0.02$ | $\xi = 0.2$ | $\xi = 2$ | $\xi = 20$ |
| $K = 1$ ($T = 160,\ d_1 = 80$) | | | | | |
| Fully observed | .999 | 999 | 1 | .986 | .909 |
| (M-i) | .999 | .999 | 1 | .985 | .906 |
| (M-ii) | .997 | .997 | .995 | .985 | .903 |
| (M-iii) | .99 | .989 | .987 | .934 | .791 |
| $K = 2$ ($T = d_1 = d_2 = 40$) | | | | | |
| Fully observed | 1 | 1 | 1 | .992 | .843 |
| (M-i) | 1 | 1 | 1 | .993 | .842 |
| (M-ii) | 1 | 1 | 1 | .986 | .842 |
| (M-iii) | .994 | .995 | .996 | .972 | .815 |
| $K = 3$ ($T = d_1 = d_2 = d_3 = 20$) | | | | | |
| Fully observed | .999 | .999 | .997 | .967 | .732 |
| (M-i) | .999 | .999 | .997 | .967 | .73 |
| (M-ii) | .998 | .998 | .996 | .959 | .718 |
| (M-iii) | .995 | .997 | .995 | .95 | .695 |

Table 3.4: Results of correct proportion for the number of factor estimator $\widehat{r}_k$ relative to the choice of $\xi$ in Theorem 3.5 on mode-1 in 1000 replications.

(i.e., the estimator of the asymptotic variance of the loading estimator), while we use the HAC-type estimator facilitating more serial information. For all $K$ considered, we present the result on $(\widehat{\mathbf{Q}})_{11}$, with the parameter $\beta$ of our HAC-type estimator set as $\lfloor \frac{1}{5}(Td_1)^{1/4} \rfloor$. We use (M-i) as the missing pattern for all settings.

The data generating process is similar to the ones for assessing the factor loading matrix estimators and imputations, but the parameters are slightly adjusted. All elements in $\mathcal{F}_t$, $\mathcal{F}_{e,t}$, and $\boldsymbol{\epsilon}_t$ are now independent standardised AR(1) with AR coefficients 0.05, and we use i.i.d. $\mathcal{N}(0,1)$ as the innovation process. We stress that we include contemporary and serial dependence among the noise variables following Assumption (E1) and (E2), while most existing literature demonstrating asymptotic normality display results only for i.i.d. Gaussian noise.

We assume all factors are pervasive in this subsection. For all $K = 1, 2, 3$, given $d_1$, we set $T, d_i = d_1/2$, $i \neq 1$. We generate a two-factor model for $K = 1$, and a one-factor model for $K = 2, 3$. For the settings $(K, d_1) = (1, 1000)$, $(2, 400)$ and $(3, 160)$, we consider $(\widehat{\mathbf{\Sigma}}_{HAC} + \widehat{\mathbf{\Sigma}}_{HAC}^{\Delta})^{-1/2}\widehat{\mathbf{D}}_1(\widehat{\mathbf{Q}}_{1,1\cdot} - \mathbf{H}_1^a\mathbf{Q}_{1,1\cdot})$. In particular, we plot the histograms of the first and second entry in Figure 3.2, whereas the corresponding QQ plots are presented in Figure 3.3.

The plots in Figure 3.2 provide empirical support to Theorem 3.3 and result 3 of Theorem 3.4. For $K = 3$, there are some heavy-tail issues, as seen in the bump at the right tail in the histogram (confirmed by its corresponding QQ plot). The QQ plot for $K = 2$ also hints on this, but the tail is thinned as the dimension increases. Our simulation is similar to that in Chen and Fan (2023) for $K = 2$, but we allow partial data unobserved and we generalise to any tensor order $K$. We remark that the convergence rate of the HAC-type estimator is not completely satisfactory, such that relatively large dimension is needed, and it becomes less feasible for some applications. We leave the improvements of the HAC-type estimator to future work.

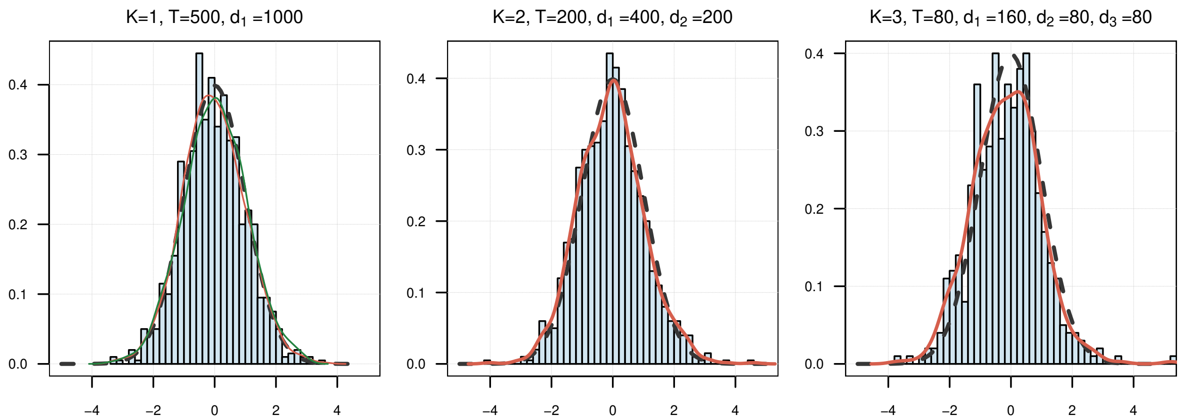


Figure 3.2: Histograms of the first entry of $(\widehat{\mathbf{\Sigma}}_{HAC}+\widehat{\mathbf{\Sigma}}^{\Delta}_{HAC})^{-1/2}\widehat{\mathbf{D}}_1(\widehat{\mathbf{Q}}_{1,1\cdot}-\mathbf{H}_1^a\mathbf{Q}_{1,1\cdot})$. In each panel, the curve (in red) is the empirical density, and the other curve (in green) in the left panel depicts the empirical density of the second entry of $(\widehat{\mathbf{\Sigma}}_{HAC}+\widehat{\mathbf{\Sigma}}^{\Delta}_{HAC})^{-1/2}\widehat{\mathbf{D}}_1(\widehat{\mathbf{Q}}_{1,1\cdot}-\mathbf{H}_1^a\mathbf{Q}_{1,1\cdot})$. The density curve for $\mathcal{N}(0,1)$ (in black, dotted) is also superimposed on each histogram.

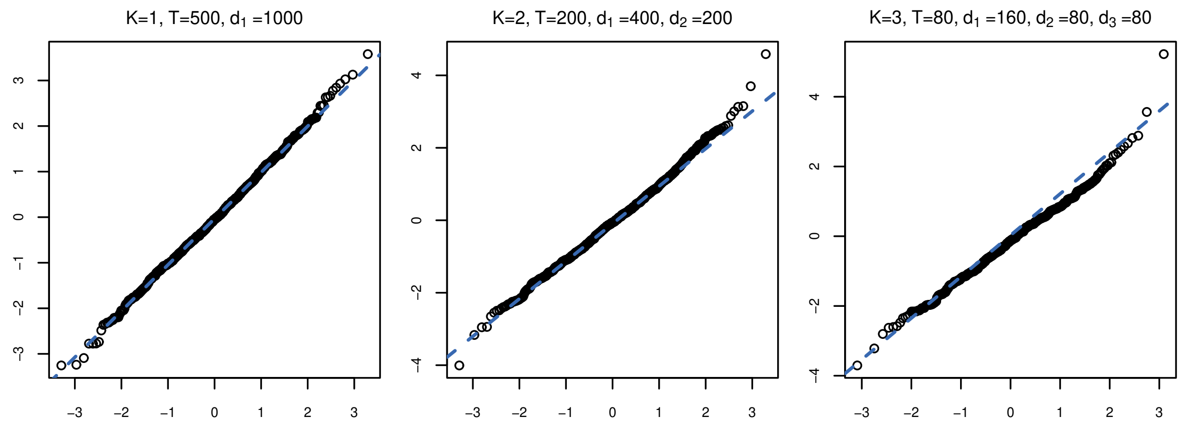


Figure 3.3: QQ plots of the first entry of $(\widehat{\mathbf{\Sigma}}_{HAC}+\widehat{\mathbf{\Sigma}}^{\Delta}_{HAC})^{-1/2}\widehat{\mathbf{D}}_1(\widehat{\mathbf{Q}}_{1,1\cdot}-\mathbf{H}_1^a\mathbf{Q}_{1,1\cdot})$. The horizontal and vertical axes are theoretical and empirical quantiles respectively.

Lastly, we demonstrate an example of statistical testing for the above one-factor model for $K = 2$. More precisely, we want to test the null hypothesis $\mathcal{H}_0 : \mathbf{Q}_{1,11} = 0$ with a two-sided test. A $5\%$ significance level is used so that we reject the null if $(\widehat{\mathbf{\Sigma}}_{HAC} + \widehat{\mathbf{\Sigma}}^{\Delta}_{HAC})^{-1/2}\widehat{\mathbf{D}}_1\widehat{\mathbf{Q}}_{1,11}$ is not in $[-1.96, 1.96]$. Each experiment is repeated 400 times and the power function for $\mathbf{Q}_{1,11}$ ranging from $-0.02$ to $0.02$ is presented in Figure 3.4. The power function is approximately symmetric, and suggests that our test can successfully reject the null if the true value for $\mathbf{Q}_{1,11}$ is away from 0. When $\mathbf{Q}_{1,11} = 0$, the false positive probability is $7.25\%$ which is slightly higher than the designated size of test. This is due to the slow convergence of the HAC estimators, and an increase in dimensions would improve this.

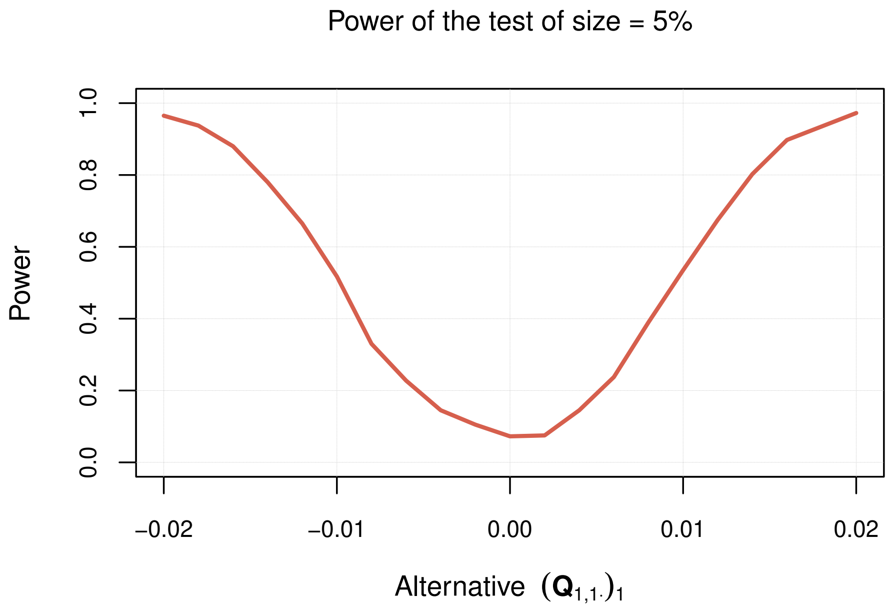


Figure 3.4: Statistical power of testing the null $\mathcal{H}_0 : (\mathbf{Q}_{1,1\cdot})_1 = \mathbf{Q}_{1,11} = 0$ against the two-sided alternative. The null is rejected when $|(\widehat{\mathbf{\Sigma}}_{HAC} + \widehat{\mathbf{\Sigma}}^{\Delta}_{HAC})^{-1/2}\widehat{\mathbf{D}}_1\widehat{\mathbf{Q}}_{1,11}| > 1.96$.

**Comparison with an iterative vectorisation-based approach**

We compare our proposed tensor factor-based imputation method with the following procedure.

Iterative vectorisation-based imputation

1. Given an order-$K$ tensor with missing entries, $\mathcal{Y}_t \in \mathbb{R}^{d_1\times\cdots\times d_K}$ for $t \in [T]$, obtain $\mathbf{y}_t = \textbf{vec}(\mathcal{Y}_t) \in \mathbb{R}^d$ for all timestamps. Impute the vector time series $\{\mathbf{y}_t\}_{t\in[T]}$ by Xiong and Pelger (2023) and denote by the tensorised imputation data $\{\widehat{\mathcal{Y}}_{\text{vec},t}\}_{t\in[T]}$.

2. Replace missing entries in $\mathcal{Y}_t$ by corresponding entries in $\widehat{\mathcal{Y}}_{\text{vec},t}$. On the resulting series, estimate the loading matrices, core factors and hence the common components by Chen and Lam (2024b). Denote the series of estimated common components by $\{\widehat{\mathcal{Y}}_{\text{preavg},t}\}_{t\in[T]}$.

3. Iterate from step 2, except that the missing entries in $\mathcal{Y}_t$ are replaced by entries of $\widehat{\mathcal{Y}}_{\text{preavg},t}$ from the previous iteration.

The above algorithm is a natural way of leveraging the vector imputation of Xiong and Pelger (2023) to tensor time series, and the iteration step is akin to Appendix A of Stock and Watson (2002b). For demonstration, all innovation processes in constructing $\mathcal{F}_t, \mathcal{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are i.i.d. $\mathcal{N}(0,1)$, and all factors are pervasive. In particular, the following settings are considered:

(Va) $K = 2, T = 20, d_1 = d_2 = 40, r_1 = r_2 = 2$, and missing pattern is (M-ii).

(Vb) Same as (Va), except that the missing pattern is (M-iii).

(Vc) $K = 3, T = 10, d_1 = d_2 = d_3 = 10, r_1 = r_2 = r3 = 2$, and missing pattern is (M-ii).

(Vd) Same as (Vc), except that the missing pattern is (M-iii).

The results for settings (Va) to (Vd) are shown in Figure 3.5. From both panels, our proposed method (in dashed lines) performs better than the direct vectorised imputation. One intuition can be the following. Suppose we have a matrix-valued time series $\mathbf{Y}_t \in \mathbb{R}^{d_1\times d_2}$, and assume $d_1 \asymp d_2$ and the data is asymptotically observed with the rate $\eta \asymp 1/\sqrt{Td_1}$. According to Corollary 3.1, the squared imputation error has rate $1/(Td_1) + 1/d_1^2$. In comparison, if we choose to vectorise the data and impute, the squared error rate is $1/T + 1/d_1^2$ which is inflated.

The performance of the vectorisation-based imputation can be further improved by iterative imputation in the context of tensor data. However, Figure 3.5 demonstrates the low efficiency of such iterative method if the missing pattern is unbalanced to a certain extent. We also point out that the computation time of the initial vectorised imputations can be significantly larger than the our proposed method if the order of the data is large. In fact, the computational complexity (given the number of factors) of direct vectorised imputation is (ignoring the cost of vectorisation and unfolding) $O(Td^2 + d^3)$, while our proposed method is $O(K \max_{k\in[K]}\{Tdd_k + d_k^3\})$, which can be of significantly smaller order than $d^3$.

### 3.4.2 Real data analysis: Fama–French portfolio returns

We analyse the set of Fama–French portfolio returns data described in Section 3.2.1. With sufficient observed samples of each category along its time series, Assumption (O1) in Section 3.3 can be satisfied and our imputation approach is applicable under such missing pattern. Since

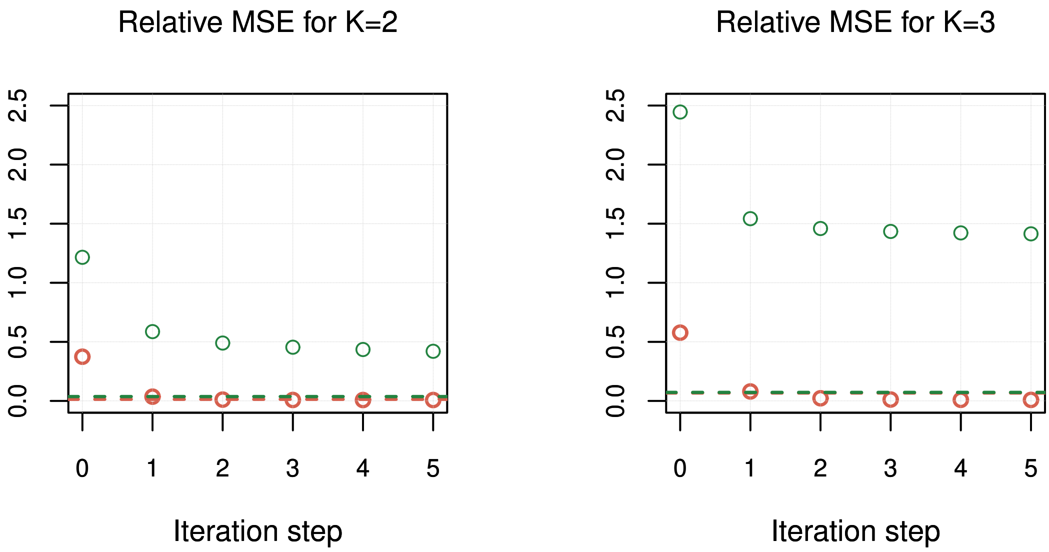


Figure 3.5: Plot of the relative MSE for Setting (Va) to (Vd), averaged over 1000 replications. Setting (Va), (Vb), (Vc) and (Vd) are represented by symbols in *red on left panel*, *green on left panel*, *red on right panel* and *green on right panel*, respectively. Dashed lines denote our tensor-based approach (without iteration); points denote the iterative vectorisation-based method with step 0 corresponding to the initial imputation.

the market factor is pervasive in financial returns, we remove the market effect by modelling the data with CAPM as

$$\mathbf{vec}(\mathcal{X}_t) = \mathbf{vec}(\bar{\mathcal{X}}) + \boldsymbol{\beta}(r_t - \bar{r}) + \mathbf{vec}(\mathcal{Y}_t),$$

where $\mathbf{vec}(\mathcal{X}_t) \in \mathbb{R}^{100}$ is the vectorised returns at time $t$, $\mathbf{vec}(\bar{\mathcal{X}})$ is the sample mean of $\mathbf{vec}(\mathcal{X}_t)$ based on all observed data, $\boldsymbol{\beta}$ is the coefficient vector to be estimated, $r_t$ is the return of the NYSE composite index at time $t$, $\bar{r}$ is the sample mean of $r_t$, and $\mathbf{vec}(\mathcal{Y}_t)$ is the CAPM residual. We compute the sample mean using only the observed data, and more sophisticated methods could be studied in the future. The least squares solution is

$$\widehat{\boldsymbol{\beta}} = \frac{\sum_{t=1}^{T}(r_t - \bar{r})\{\mathbf{vec}(\mathcal{X}_t) - \mathbf{vec}(\bar{\mathcal{X}})\}}{\sum_{t=1}^{T}(r_t - \bar{r})^2}.$$

Hence for the rest of this subsection, we focus on the matrix series $\{\widehat{\mathcal{Y}}_t\}_{t\in[570]}$ with $\widehat{\mathcal{Y}}_t \in \mathbb{R}^{10\times 10}$, constructed from the estimated CAPM residual $\{\mathbf{vec}(\mathcal{X}_t) - \mathbf{vec}(\bar{\mathcal{X}}) - \widehat{\boldsymbol{\beta}}(r_t - \bar{r})\}_{t\in[570]}$.

To estimate the rank of the core factors, we first use our proposed rank estimator to obtain initial estimates $(\widehat{r}_1, \widehat{r}_2) = (1, 1)$ for both series, followed by re-estimating the rank based on

| | initial | | Miss-ER | | BCorTh | | iTIP-ER | | RTFA-ER | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\widehat{r}_1$ | $\widehat{r}_2$ | $\widehat{r}_1$ | $\widehat{r}_2$ | $\widehat{r}_1$ | $\widehat{r}_2$ | $\widehat{r}_1$ | $\widehat{r}_2$ | $\widehat{r}_1$ | $\widehat{r}_2$ |
| Value Weighted | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 |
| Equal Weighted | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 |

Table 3.5: Rank estimators for Fama–French portfolios. Miss-ER represents the rank re-estimated by our proposed eigenvalue-ratio estimator for missing data.

the imputed series using $(\widehat{r}_1 + r_*, \widehat{r}_2 + r_*)$ with some pre-specified integer $r_*$ to capture any omitted weak factors. We have seen in Table 3.2 and Table 3.3 where such rank re-estimation with $r_* = 1$ is stable and accurate. However, factors can be empirically too weak to detect in the initial estimation under various missing patterns. According to previous studies by e.g. Wang et al. (2019), we choose $r_* = 3$ here to ensure sufficient information of factors is carried in the imputation, at the cost of including more noise. For re-estimation, in addition to our eigenvalue-ratio estimator, we also experiment BCorTh by Chen and Lam (2024b), iTIP-ER by Han et al. (2022) and RTFA-ER by He et al. (2022b). The results are presented in Table 3.5. To ease demonstration, we use $(2, 2)$ as the core factor rank for both series hereafter.

| | ME1 | ME2 | ME3 | ME4 | ME5 | ME6 | ME7 | ME8 | ME9 | ME10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Factor 1 | -15 | -14 | -9 | -7 | -6 | -3 | -1 | 0 | 2 | 3 |
| Factor 2 | 5 | 3 | -3 | -6 | -7 | -9 | -10 | -11 | -10 | -10 |
| | **BE1** | **BE2** | **BE3** | **BE4** | **BE5** | **BE6** | **BE7** | **BE8** | **BE9** | **BE10** |
| Factor 1 | 2 | -1 | -2 | -3 | -5 | -6 | -7 | -8 | -10 | -18 |
| Factor 2 | 16 | 12 | 9 | 7 | 5 | 3 | 3 | 2 | 1 | -7 |

Table 3.6: Estimated loading matrices $\widehat{\mathbf{Q}}_1$ and $\widehat{\mathbf{Q}}_2$ for the value weighted portfolio series, after varimax rotation and scaling (entries rounded to the nearest integer). Magnitudes larger than 9 are in red to highlight units with heavy loadings. All null hypotheses of a row of $\mathbf{Q}_1$ or $\mathbf{Q}_2$ being zero (see (3.16)) are rejected at $5\%$ significance level.

With the chosen rank, we perform imputation which is further refined by re-imputation. The results are similar on the two portfolio series, so we only present the one for the value weighted series. The estimated loading matrices are presented in Table 3.6, after a varimax rotation and scaling. It is clear from the entries in red that on the size factor (i.e., ME loading), ME1 and ME2 form one group ("small size") and ME7 to ME10 form the other ("large size"). On the book-to-equity factor (i.e., BE loading), BE1 and BE2 form a group and BE9 and BE10 form the other, which can be interpreted as "undervalued" and "overvalued" respectively. This grouping effect is similarly seen in Table 9 and 10 in Wang et al. (2019).

Moreover, we apply our Theorem 3.3 and Theorem 3.4 to test if any rows of the loading matrices are zero. For each $k \in [2], i \in [10]$, we test

$$\mathcal{H}_0 : \mathbf{Q}_{k,i\cdot} = \mathbf{0}, \quad \mathcal{H}_1 : \mathbf{Q}_{k,i\cdot} \neq \mathbf{0}. \tag{3.16}$$

The above can be tested since $\mathbf{H}_k^a \mathbf{Q}_{k,i\cdot} = \mathbf{0}$ under the null, and no matter what varimax rotations we use, it retains its meaning. For instance, if $\mathbf{Q}_{1,i\cdot} = \mathbf{0}$, then it means the $i$-th category of the row factor (here, the $i$-th Market Equity category) is useless in explaining any data variability.

It turns out that at 5% significance level, we cannot reject any null hypotheses for $\mathbf{Q}_{1,i\cdot} = \mathbf{0}$ or $\mathbf{Q}_{2,i\cdot} = \mathbf{0}$, meaning that individual market equity and book-to-equity ratio categories are tested to be meaningful in explaining some variations of the data. We remark that, since the dimensions of our data are not very large, the accuracy of the asymptotic normality and the HAC estimators are weakened, and there can be false positives as a result.

Lastly, two imputation examples for the category (ME10, BE10) are displayed in Figure 3.6. From the timestamps on which the portfolio series is observed, we see that the estimated series (in green) does capture some patterns of fluctuations on the true CAPM residual series (in red) and hence can be a good reference for the CAPM residual of portfolios consisted of large size, overvalued stocks. This is certainly more revealing than a naive imputation using zeros or local means. From the above discussions, the estimated factors can be potentially used to replace the Fama—French size factor (SMB) and book-to-equity factor (HML) in a Fama—French factor model for asset pricing, factor trading etc., with a further sophisticated analysis of the data.

### 3.4.3 Real data analysis: OECD economic indicators for countries

We analyse the OECD economic data described in Section 3.2.2. After investigating the estimated number of factors (Table 3.7) in a similar re-imputation approach as in Section 3.4.2, we decide to use $(\widehat{r}_1, \widehat{r}_2) = (3, 3)$ for the rest of this section due to the potentially weak factors suggested by iTIP-ER and RTFA-ER. The estimated loading matrices for countries are presented in Table 3.8 after a varimax rotation and scaling, with entries highlighted in red to facilitate interpretation. The first factor is mainly formed by European countries except the Northern European ones which, together with Canada, form the third factor. Such regional grouping effects are also confirmed in the second factor which mainly consists of the United States, and the fact that Germany loads also heavily on this factor suggests their similar economic patterns as large economic entities. For the estimated loading for indicators reported in Table 3.9, CP, PRVM and TOVM form the first factor ("consumption factor"), PP and ULC form the second ("production factor"), and EX and IM form the third ("international trade factor").

Moreover, we apply Theorem 3.3 and Theorem 3.4 to test if a particular row in the two factor loading matrices is zero, meaning that if a country (if a row in $\mathbf{Q}_1$ is $\mathbf{0}$) or an economic indicator (if a row in $\mathbf{Q}_2$ is $\mathbf{0}$) cannot explain any variations in the data. The meaning here

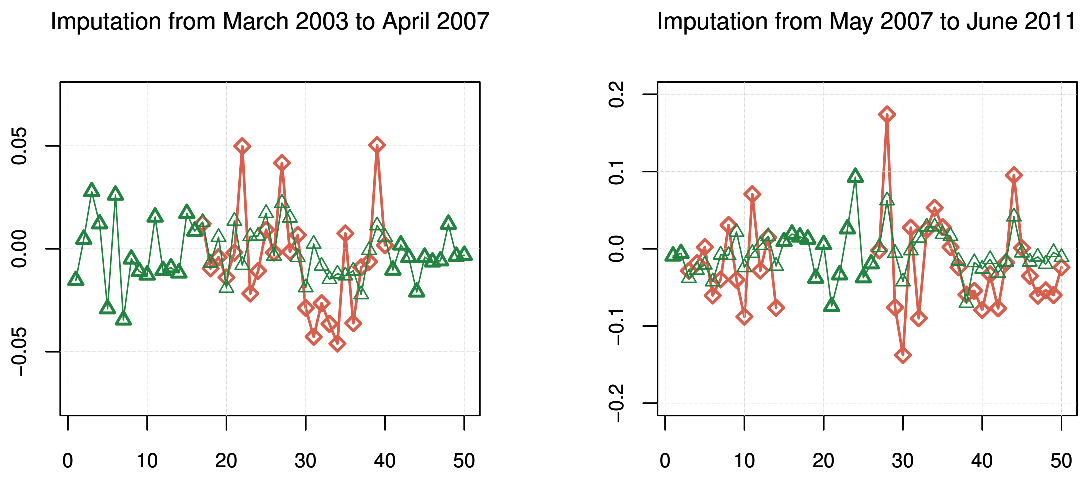


Figure 3.6: Two 50-day examples for the value weighted series in the category (ME10, BE10), with horizontal axis of both panels indexed by each day of the selected period. Green triangles denote the estimated series and red squares denote the observed true series. Bold symbols represent the imputed series which consists of the observed series whenever available and the estimated series otherwise.

| | initial | | Miss-ER | | BCorTh | | iTIP-ER | | RTFA-ER | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\widehat{r}_1$ | $\widehat{r}_2$ | $\widehat{r}_1$ | $\widehat{r}_2$ | $\widehat{r}_1$ | $\widehat{r}_2$ | $\widehat{r}_1$ | $\widehat{r}_2$ | $\widehat{r}_1$ | $\widehat{r}_2$ |
| OECD | 1 | 1 | 1 | 1 | 1 | 2 | 4 | 5 | 3 | 3 |

Table 3.7: Rank estimators for economic indicators. Refer to Table 3.5 for the definitions of different estimators.

is independent of the varimax rotation performed. For each $k \in [2], i \in [d_k], j \in [3]$ with $(d_1, d_2) = (17, 11)$, we form the hypothesis

$$\mathcal{H}_0 : \mathbf{Q}_{k,i\cdot} = \mathbf{0}, \quad \mathcal{H}_1 : \mathbf{Q}_{k,i\cdot} \neq \mathbf{0}. \tag{3.17}$$

Similar to the Fama–French data analysis, all null hypotheses of a row of $\mathbf{Q}_1$ or $\mathbf{Q}_2$ being zero are rejected at 5% significance level. It means that all individual country and economic indicator are tested to be meaningful categories in explaining some variations of the data. Similar to a reminder in Section 3.4.2, there could be false positives due to the fact that the dimension of the data is not very large.

In Figure 3.7, we present two examples of the imputed series overlaid on the observed series. One panel plots ULC of the United States and the other plots PP of the United Kingdom.

| | BEL | CAN | DNK | FIN | FRA | DEU | GRC | ITA | LUX | NLD | NOR | PRT | ESP | SWE | CHE | GBR | USA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -10 | 4 | -3 | -7 | -8 | -8 | -9 | -7 | -1 | -2 | 1 | -1 | -10 | 0 | -15 | -13 | 1 |
| 2 | -1 | -6 | 2 | 5 | -2 | -12 | 7 | -1 | 2 | -7 | 0 | 2 | 1 | -2 | 0 | -1 | -24 |
| 3 | 1 | -12 | -8 | -5 | -2 | 3 | -6 | -4 | -11 | -6 | -12 | -11 | -2 | -10 | 5 | 4 | -1 |

Table 3.8: Estimated loading matrix $\widehat{\mathbf{Q}}_1$ on three country factors for the OECD data, after varimax rotation and scaling (entries rounded to the nearest integer). Magnitudes larger than 9 are in red to highlight units with heavy loadings. All null hypotheses of a row of $\mathbf{Q}_1$ being zero (see (3.17)) are rejected at $5\%$ significance level.

| | CA-GDP | CP | EX | IM | IR3TIB | IRLT | IRSTCI | PP | PRVM | TOVM | ULC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | -20 | 1 | 3 | 0 | 0 | 0 | 0 | -20 | -11 | -1 |
| 2 | 0 | -6 | 2 | 3 | 1 | 1 | 1 | 20 | 1 | 9 | 18 |
| 3 | 0 | 9 | 18 | 22 | -2 | -2 | -2 | 1 | -4 | -2 | -1 |

Table 3.9: Estimated loading matrix $\widehat{\mathbf{Q}}_2$ on three indicator factors for OECD data, after varimax rotation and scaling (entries rounded to the nearest integer). Magnitudes larger than 9 are in red to highlight units with heavy loadings. All null hypotheses of a row of $\mathbf{Q}_2$ being zero (see (3.17)) are rejected at $5\%$ significance level.

ULC is a quarterly observed index and the peak pattern in-between each reported timestamp suggests potentially high labour cost in the United States from 1971 to 1975. The PP data in our OECD data is unavailable for the United Kingdom until December 2008. Our imputation implies a gradual increase of the PP before the data is reported, which is reasonable by the impact of the financial crisis. Lastly, we compare between our tensor imputation (matrix imputation for this example) and the vectorised imputation using Xiong and Pelger (2023). We use different models to perform imputations whose results are summarised in Table 3.10 similar to Wang et al. (2019), except that the reported residual sum of squares are computed on the observed entries. Although we require a larger number of factors in general for matrix models, the imputation by matrix models with less parameters can perform better than those by vector models with a much larger number of parameters. This is consistent with the conclusion of Table 11 in Wang et al. (2019).

## 3.5 Proof of Theorems and Auxiliary Results

From Section 3.2.4, $\widehat{\mathbf{Q}}_k$ contains the eigenvectors corresponding to the first $r_k$ largest eigenvalues of $\widehat{\mathbf{S}}_k$. Hence with $\widehat{\mathbf{D}}_k$ an $r_k \times r_k$ diagonal matrix containing all the eigenvalues of $\widehat{\mathbf{S}}_k$ (WLOG from the largest on the top-left element to the smallest on the bottom right element), we have $\widehat{\mathbf{S}}_k\widehat{\mathbf{Q}}_k = \widehat{\mathbf{Q}}_k\widehat{\mathbf{D}}_k$, so that

$$\widehat{\mathbf{Q}}_k = \widehat{\mathbf{S}}_k\widehat{\mathbf{Q}}_k\widehat{\mathbf{D}}_k^{-1}. \tag{3.18}$$

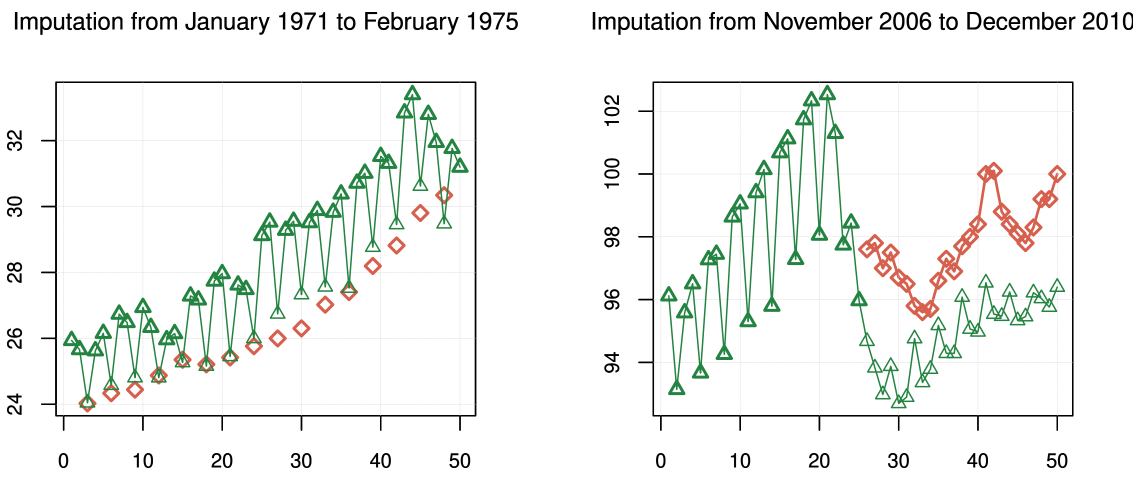


Figure 3.7: 50-day examples for unit labour cost of the United States (left panel) and production price index of the United Kingdom (right panel), with horizontal axis of both panels indexed by each day of the selected period. Refer to Figure 3.6 for the explanations of different symbols.

| | Factor | RSS | # factors | # parameters |
|---|---|---|---|---|
| Matrix model | (3,3) | 7,087,373 | 9 | 84 |
| Matrix model | (4,4) | 4,542,956 | 16 | 112 |
| Matrix model | (5,5) | 3,066,851 | 25 | 140 |
| Matrix model | (6,6) | 1,973,321 | 36 | 168 |
| Vector model | 2 | 8,240,976 | 2 | 374 |
| Vector model | 3 | 3,954,554 | 3 | 561 |
| Vector model | 4 | 2,093,001 | 4 | 748 |

Table 3.10: Comparison of different models for the OECD data. The total sum of squares of the observation is 324,402,709.

To simplify notations, hereafter we fix $k$ and only focus on the mode-$k$ unfolded data. Define

$$\begin{aligned} &\widehat{\mathbf{D}} := \widehat{\mathbf{D}}_k, \ \mathbf{Y}_t := \text{mat}_k(\mathcal{Y}_t), \ \widehat{\mathbf{S}} := \widehat{\mathbf{S}}_k, \ \psi_{ij,h} := \psi_{k,ij,h}, \ \mathbf{Q} := \mathbf{Q}_k, \\ &\mathbf{\Lambda} := \mathbf{\Lambda}_k, \ \mathbf{F}_{Z,t} := \text{mat}_k(\mathcal{F}_{Z,t}), \ \mathbf{E}_t := \text{mat}_k(\mathcal{E}_t), \ \mathbf{H}_j := \mathbf{H}_{k,j}, \ \mathbf{H}^a := \mathbf{H}_k^a, \end{aligned} \tag{3.19}$$

where $\mathbf{H}_{k,j}$ and $\mathbf{H}_k^a$ are defined in (3.10) and (3.11) respectively, and similarly to all respective hat versions of the above.

Before proving any theorems, we present and prove the following Proposition first.

**Proposition 3.2** *Let Assumption (E1), (E2) and (F1) hold. Then*

1. *there exists a constant $c > 0$ so that for any $k \in [K], t \in [T], i_k \in [d_k]$ and $h \in [d_{\text{-}k}]$, we have $\mathbb{E}\mathcal{E}_{t,i_1,\ldots,i_K} = 0$, $\mathbb{E}\mathcal{E}^4_{t,i_1,\ldots,i_K} \le c$, and*

$$\sum_{j=1}^{d_k}\sum_{l=1}^{d_{\text{-}k}} \Big|\mathbb{E}[\text{mat}_k(\mathcal{E}_t)_{ih}\text{mat}_k(\mathcal{E}_t)_{jl}]\Big| \le c,$$
$$\sum_{l=1}^{d_{\text{-}k}}\sum_{s\in\psi_{k,ij,l}} \Big|\text{Cov}\Big(\text{mat}_k(\mathcal{E}_t)_{ih}\text{mat}_k(\mathcal{E}_t)_{jh}, \text{mat}_k(\mathcal{E}_s)_{il}\text{mat}_k(\mathcal{E}_s)_{jl}\Big)\Big| \le c;$$

2. *there exists a constant $c > 0$ so that for any $k \in [K], i, j \in [d_k]$, and any deterministic vectors $\mathbf{u} \in \mathbb{R}^{r_k}$ and $\mathbf{v} \in \mathbb{R}^{r_{\text{-}k}}$ with constant magnitudes,*

$$\mathbb{E}\left(\frac{1}{d_{\text{-}k}^{1/2}}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{k,ij,h}|^{1/2}}\sum_{t\in\psi_{k,ij,h}}\text{mat}_k(\mathcal{E}_t)_{jh}\mathbf{u}'\text{mat}_k(\mathcal{F}_t)\mathbf{v}\right)^2 \le c;$$

3. *for any $k \in [K], i, j \in [d_k], h \in [d_{\text{-}k}]$,*

$$\frac{1}{|\psi_{k,ij,h}|}\sum_{t\in\psi_{k,ij,h}}\text{mat}_k(\mathcal{F}_t)\text{mat}_k(\mathcal{F}_t)', \quad \frac{1}{T}\sum_{t=1}^{T}\text{mat}_k(\mathcal{F}_t)\text{mat}_k(\mathcal{F}_t)' \xrightarrow{p} \mathbf{\Sigma}_k := r_{\text{-}k}\mathbf{I}_{r_k},$$

*with the number of factors $r_k$ fixed as $\min\{T, d_1, \ldots, d_K\} \to \infty$. For each $t \in [T]$, all elements in $\mathcal{F}_t$ are independent of each other, with mean 0 and unit variance.*

Our consistency results in Theorem 3.1 can be proved assuming the three implied results from Proposition 3.2, on top of Assumption (O1), (M1), (L1) and (R1). Result 1 from Proposition 3.2 can be a stand alone assumption on the weak correlation of the noise $\mathcal{E}_t$ across different dimensions and times, while result 2 can be on the weak dependence between the factor $\mathcal{F}_t$ and the noise $\mathcal{E}_t$. Finally, result 3 can be a stand alone assumption on the factors $\mathcal{F}_t$.

***Proof of Proposition 3.2.*** We have $\mathbb{E}(\mathcal{E}_t) = 0$ from Assumption (E1). Next we want to show that for any $k \in [K], t \in [T]$ and $i_k \in [d_k]$, $\mathbb{E}\mathcal{E}^4_{t,i_1,\ldots,i_K}$ is bounded uniformly. From (3.9), each entry in $\mathcal{E}_t$ is a sum of two parts: a linear combination of the elements in $\mathcal{F}_{e,t}$, and the corresponding entry in $\boldsymbol{\epsilon}_t$. By Assumption (E2), we have

$$\mathbb{E}[(\boldsymbol{\epsilon}_t)^4_{i_1,\ldots,i_K}] = \mathbb{E}\Big\{\Big[\sum_{q\ge0}a_{\epsilon,q}(\mathcal{X}_{\epsilon,t-q})_{i_1,\ldots,i_K}\Big]^4\Big\} \le \Big(\sum_{q\ge0}|a_{\epsilon,q}|\Big)^4\sup_t\mathbb{E}\Big[(\mathcal{X}_{\epsilon,t})^4_{i_1,\ldots,i_K}\Big] \le C,$$

where $C > 0$ is a generic constant. It holds similarly that $\mathbb{E}[(\mathcal{F}_{e,t})^4_{i_1,\ldots,i_K}] \le C$ uniformly on all indices. With this, defining $\mathbf{A}^{(*m)} := \mathbf{A} * \cdots * \mathbf{A}$ (element-wise $m$-th power),

$$\begin{aligned}
&\|\mathbb{E}\mathcal{E}_t^{(*4)}\|_{\max} = \|\mathbb{E}[\text{mat}_k(\mathcal{E}_t)]^{(*4)}\|_{\max} \\
&\leq 8\big(\|\mathbb{E}[\mathbf{A}_{e,k}\text{mat}_k(\mathcal{F}_{e,t})\mathbf{A}'_{e,\text{-}k}]^{(*4)}\|_{\max} + \|\boldsymbol{\Sigma}_\epsilon^{(*4)}\|_{\max} \cdot \|\mathbb{E}\boldsymbol{\epsilon}_t^{(*4)}\|_{\max}\big) \\
&\leq 8\big(\big\|\mathbf{A}_{e,k}\big\|_\infty^4 \cdot \big\|\mathbf{A}_{e,\text{-}k}\big\|_\infty^4 \cdot \big\|\mathbb{E}\mathcal{F}_{e,t}^{(*4)}\big\|_{\max} + \|\boldsymbol{\Sigma}_\epsilon^{(*4)}\|_{\max} \cdot \|\mathbb{E}\boldsymbol{\epsilon}_t^{(*4)}\|_{\max}\big) \\
&= 8\Big(\prod_{j=1}^K \big\|\mathbf{A}_{e,j}\big\|_\infty^4 \cdot \big\|\mathbb{E}\mathcal{F}_{e,t}^{(*4)}\big\|_{\max} + \|\boldsymbol{\Sigma}_\epsilon^{(*4)}\|_{\max} \cdot \|\mathbb{E}\boldsymbol{\epsilon}_t^{(*4)}\|_{\max}\Big) \leq C,
\end{aligned}$$

where $C > 0$ is again a generic constant, and we used Assumption (E1) in the last line and the fact that $r_j$ is a constant for $j \in [K]$. This is equivalent to $\mathbb{E}\mathcal{E}^4_{t,i_1,\ldots,i_K} \leq c$ for some constant $c$.

With (3.9) in Assumption (E1), we have

$$\text{mat}_k(\mathcal{E}_t) = \mathbf{A}_{e,k}\text{mat}_k(\mathcal{F}_{e,t})\mathbf{A}'_{e,\text{-}k} + \text{mat}_k(\boldsymbol{\Sigma}_\epsilon) * \text{mat}_k(\boldsymbol{\epsilon}_t),$$

where $\mathbf{A}_{e,\text{-}k} := \mathbf{A}_{e,K} \otimes \cdots \otimes \mathbf{A}_{e,k+1} \otimes \mathbf{A}_{e,k-1} \otimes \cdots \otimes \mathbf{A}_{e,1}$. Each mode-$k$ noise fibre $\mathbf{e}_{t,k,l}$ for $l \in [d_{\text{-}k}]$ can then be decomposed as

$$\mathbf{e}_{t,k,l} := \mathbf{A}_{e,k}\text{mat}_k(\mathcal{F}_{e,t})\mathbf{A}_{e,\text{-}k,l\cdot} + \boldsymbol{\Sigma}_{\epsilon,k,l}^{1/2}\boldsymbol{\epsilon}_{t,k,l}, \tag{3.20}$$

where $\boldsymbol{\Sigma}_{\epsilon,k,l} = \text{diag}((\text{mat}_k(\boldsymbol{\Sigma}_\epsilon))_{\cdot l}(\text{mat}_k(\boldsymbol{\Sigma}_\epsilon))'_{\cdot l})$, and $\boldsymbol{\epsilon}_{t,k,l}$ contains independent elements each with mean 0 and variance 1.

Given $h \in [d_{\text{-}k}], i \in [d_k]$, from (3.20) and Assumption (E1) and (E2), we have

$$\sum_{l\neq h}^{d_{\text{-}k}} \sum_{j=1}^{d_k} \Big|\mathbb{E}[\text{mat}_k(\mathcal{E}_t)_{ih}\text{mat}_k(\mathcal{E}_t)_{jl}]\Big| \leq \|\mathbf{A}_{e,\text{-}k,h\cdot}\|\|\mathbf{A}_{e,\text{-}k,l\cdot}\|\|\mathbf{A}_{e,k}\|_1\|\mathbf{A}_{e,k}\|_\infty = O(1).$$

Moreover,

$$\begin{aligned}
&\sum_{j=1}^{d_k} \Big|\mathbb{E}[\text{mat}_k(\mathcal{E}_t)_{ih}\text{mat}_k(\mathcal{E}_t)_{jh}]\Big| \leq \|\text{Cov}(\mathbf{e}_{t,k,h}, \mathbf{e}_{t,k,h})\|_1 \\
&\leq \|\mathbf{A}_{e,\text{-}k,h\cdot}\|^2\|\mathbf{A}_{e,k}\|_1\|\mathbf{A}_{e,k}\|_\infty + \|\boldsymbol{\Sigma}_{\epsilon,k,h}\|_1 = O(1),
\end{aligned}$$

where the last equality is from Assumption (E1).

To finish the proof of the first result in the Proposition, fix indices $k, t, i, j, h$. From Assumption (E2) and (3.20), we have

$$[\text{mat}_k(\mathcal{E}_t)]_{il} = \sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,i\cdot}\text{mat}_k(\mathcal{X}_{e,t-q})\mathbf{A}_{e,\text{-}k,l\cdot} + [\text{mat}_k(\boldsymbol{\Sigma}_\epsilon)]_{il}\sum_{q\geq 0} a_{\epsilon,q}\text{mat}_k(\mathcal{X}_{\epsilon,t-q})_{il}. \tag{3.21}$$

Hence when $l \neq h$, from the independence between $\{\mathcal{X}_{e,t}\}$ and $\{\mathcal{X}_{\epsilon,t}\}$ and the independence of the elements within $\{\mathcal{X}_{\epsilon,t}\}$ in Assumption (E2), we have for any $s \in \psi_{k,ij,l}$,

$$\begin{aligned}
&\mathrm{Cov}\Big(\mathrm{mat}_k(\mathcal{E}_t)_{ih}\mathrm{mat}_k(\mathcal{E}_t)_{jh}, \mathrm{mat}_k(\mathcal{E}_s)_{il}\mathrm{mat}_k(\mathcal{E}_s)_{jl}\Big) \\
&= \mathrm{Cov}\Big\{\Big(\sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,i\cdot}\mathrm{mat}_k(\mathcal{X}_{e,t-q})\mathbf{A}_{e,\text{-}k,h\cdot}\Big)\Big(\sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,j\cdot}\mathrm{mat}_k(\mathcal{X}_{e,t-q})\mathbf{A}_{e,\text{-}k,h\cdot}\Big), \\
&\Big(\sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,i\cdot}\mathrm{mat}_k(\mathcal{X}_{e,s-q})\mathbf{A}_{e,\text{-}k,l\cdot}\Big)\Big(\sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,j\cdot}\mathrm{mat}_k(\mathcal{X}_{e,s-q})\mathbf{A}_{e,\text{-}k,l\cdot}\Big)\Big\}.
\end{aligned} \tag{3.22}$$

By (E2), all mixed covariance terms are zero except for $\mathrm{Cov}\big(\mathrm{mat}_k(\mathcal{X}_{e,t-q})^2_{nm}, \mathrm{mat}_k(\mathcal{X}_{e,t-q})^2_{nm}\big)$ for all $q \geq 0, n \in [r_{e,k}], m \in [r_{e,\text{-}k}]$, with coefficient $a^2_{e,q}a^2_{e,q+|t-s|}A^2_{e,k,in}A^2_{e,k,jn}A^2_{e,\text{-}k,hm}A^2_{e,\text{-}k,lm}$. Thus we have

$$\begin{aligned}
&\sum_{l=1,l\neq h}^{d_{\text{-}k}} \Big|\mathrm{Cov}\Big(\mathrm{mat}_k(\mathcal{E}_t)_{ih}\mathrm{mat}_k(\mathcal{E}_t)_{jh}, \mathrm{mat}_k(\mathcal{E}_s)_{il}\mathrm{mat}_k(\mathcal{E}_s)_{jl}\Big)\Big| \\
&= \sum_{l=1,l\neq h}^{d_{\text{-}k}} \Big|\sum_{n=1}^{r_{e,k}}\sum_{m=1}^{r_{e,\text{-}k}}\sum_{q\geq 0} a^2_{e,q}a^2_{e,q+|t-s|}A^2_{e,k,in}A^2_{e,k,jn}A^2_{e,\text{-}k,hm}A^2_{e,\text{-}k,lm} \\
&\qquad\qquad \cdot \mathrm{Cov}\big(\mathrm{mat}_k(\mathcal{X}_{e,t-q})^2_{nm}, \mathrm{mat}_k(\mathcal{X}_{e,t-q})^2_{nm}\big)\Big| \\
&= O\Big\{\sum_{q\geq 0} a^2_{e,q}a^2_{e,q+|t-s|}\Big(\sum_{n=1}^{r_{e,k}} A^2_{e,k,in}A^2_{e,k,jn}\Big)\Big(\sum_{m=1}^{r_{e,\text{-}k}} A^2_{e,\text{-}k,hm}\sum_{l=1,l\neq h}^{d_{\text{-}k}} A^2_{e,\text{-}k,lm}\Big)\Big\} = \sum_{q\geq 0} O(a^2_{e,q}a^2_{e,q+|t-s|}),
\end{aligned}$$

where we use Assumption (E2) in the second last equality, and (E1) in the last. Consequently,

$$\begin{aligned}
&\sum_{l=1,l\neq h}^{d_{\text{-}k}}\sum_{s\in\psi_{k,ij,l}} \Big|\mathrm{Cov}\Big(\mathrm{mat}_k(\mathcal{E}_t)_{ih}\mathrm{mat}_k(\mathcal{E}_t)_{jh}, \mathrm{mat}_k(\mathcal{E}_s)_{il}\mathrm{mat}_k(\mathcal{E}_s)_{jl}\Big)\Big| \\
&= \sum_{q\geq 0}\sum_{s=1}^{T} O(a^2_{e,q}a^2_{e,q+|t-s|}) = O(1),
\end{aligned}$$

where the last equality uses Assumption (E2). Now consider lastly $l = h$. All arguments starting from (3.21) follow exactly, except the following term is added in (3.22):

$$\begin{aligned}
&\sum_{q\geq 0} a^2_{\epsilon,q}a^2_{\epsilon,q+|t-s|}\Sigma_{\epsilon,k,h,ii}\Sigma_{\epsilon,k,h,jj} \\
&\cdot \mathrm{Cov}\Big(\mathrm{mat}_k(\mathcal{X}_{\epsilon,t-q})_{ih}\mathrm{mat}_k(\mathcal{X}_{\epsilon,t-q})_{jh}, \mathrm{mat}_k(\mathcal{X}_{\epsilon,s-q})_{ih}\mathrm{mat}_k(\mathcal{X}_{\epsilon,s-q})_{jh}\Big) = O\Big(\sum_{q\geq 0} a^2_{\epsilon,q}a^2_{\epsilon,q+|t-s|}\Big),
\end{aligned}$$

which is $O(1)$ and we used again Assumption (E2) in the last line. Finally,

$$\sum_{s\in\psi_{k,ij,h}} \left|\mathrm{Cov}\Big(\mathrm{mat}_k(\mathcal{E}_t)_{ih}\mathrm{mat}_k(\mathcal{E}_t)_{jh}, \mathrm{mat}_k(\mathcal{E}_s)_{ih}\mathrm{mat}_k(\mathcal{E}_s)_{jh}\Big)\right| = O(1).$$

This completes the proof of result 1 in the Proposition.

To prove the second result, fix $k \in [K], i, j \in [d_k]$ and deterministic vectors $\mathbf{u} \in \mathbb{R}^{r_k}$ and $\mathbf{v} \in \mathbb{R}^{r_{\text{-}k}}$ with $\|\mathbf{u}\|, \|\mathbf{v}\| = O(1)$. Note that

$$\mathbb{E}[\mathrm{mat}_k(\mathcal{F}_t)\mathbf{v}\mathbf{v}'\mathrm{mat}_k(\mathcal{F}_s)'] = \mathbf{v}'\mathbf{v}(r_{\text{-}k}\sum_{q\geq 0} a_{f,q}a_{f,q+|t-s|})\mathbf{I}_{r_k},$$

as the series $\{\mathcal{X}_{f,t}\}$ has i.i.d. elements from Assumption (F1). Similarly, from (3.20) and Assumption (E1) and (E2),

$$\begin{aligned}
&\mathrm{Cov}(\mathrm{mat}_k(\mathcal{E}_t)_{jh}, \mathrm{mat}_k(\mathcal{E}_s)_{jl}) \\
&= \mathbb{E}[\mathbf{A}'_{e,k,j\cdot}\mathrm{mat}_k(\mathcal{F}_{e,t})\mathbf{A}_{e,\text{-}k,h\cdot}\mathbf{A}'_{e,\text{-}k,l\cdot}\mathrm{mat}_k(\mathcal{F}_{e,s})'\mathbf{A}_{e,k,j\cdot}] + \mathbb{E}[\boldsymbol{\epsilon}'_{t,k,h}(\Sigma_{\epsilon,k,h,j\cdot}\Sigma'_{\epsilon,k,l,j\cdot})^{1/2}\boldsymbol{\epsilon}_{s,k,l}] \\
&= \mathbf{A}'_{e,\text{-}k,l\cdot}\mathbf{A}_{e,\text{-}k,h\cdot}\cdot\|\mathbf{A}_{e,k,j\cdot}\|^2\cdot\sum_{q\geq 0} a_{e,q}a_{e,q+|t-s|} + \mathbb{1}_{\{h=l\}}\cdot\Sigma_{\epsilon,k,h,jj}\sum_{q\geq 0} a_{\epsilon,q}a_{\epsilon,q+|t-s|}.
\end{aligned}$$

Hence if we fix $h \in [d_{\text{-}k}], t \in \psi_{k,ij,h}$, then together with Assumption (E2), we have

$$\begin{aligned}
&\sum_{l=1}^{d_{\text{-}k}}\sum_{s\in\psi_{k,ij,l}}\frac{1}{|\psi_{k,ij,l}|}\cdot\mathbb{E}\Big[\mathrm{mat}_k(\mathcal{E}_t)_{jh}\mathbf{u}'\mathrm{mat}_k(\mathcal{F}_t)\mathbf{v}\cdot\mathrm{mat}_k(\mathcal{E}_s)_{jl}\mathbf{v}'\mathrm{mat}_k(\mathcal{F}_s)'\mathbf{u}\Big] \\
&= \sum_{l=1}^{d_{\text{-}k}}\sum_{s\in\psi_{k,ij,l}}\frac{1}{|\psi_{k,ij,l}|}\cdot\mathrm{Cov}(\mathrm{mat}_k(\mathcal{E}_t)_{jh}, \mathrm{mat}_k(\mathcal{E}_s)_{jl})\cdot\mathbb{E}\Big[\mathbf{u}'\mathrm{mat}_k(\mathcal{F}_t)\mathbf{v}\mathbf{v}'\mathrm{mat}_k(\mathcal{F}_s)'\mathbf{u}\Big] \\
&= \sum_{l=1}^{d_{\text{-}k}}\frac{1}{|\psi_{k,ij,l}|}\Big\{O(\mathbf{A}'_{e,\text{-}k,l\cdot}\mathbf{A}_{e,\text{-}k,h\cdot}\cdot\|\mathbf{A}_{e,k,j\cdot}\|^2)\cdot\sum_{q\geq 0}\sum_{p\geq 0}\sum_{s\in\psi_{k,ij,l}} a_{e,q}a_{e,q+|t-s|}a_{f,p}a_{f,p+|t-s|} \\
&\quad + O(\mathbb{1}_{\{h=l\}}\cdot\Sigma_{\epsilon,k,h,jj})\cdot\sum_{q\geq 0}\sum_{p\geq 0}\sum_{s\in\psi_{k,ij,l}} a_{\epsilon,q}a_{\epsilon,q+|t-s|}a_{f,p}a_{f,p+|t-s|}\Big\} \\
&= \sum_{l=1}^{d_{\text{-}k}}\frac{1}{|\psi_{k,ij,l}|}\cdot O(\mathbf{A}'_{e,\text{-}k,l\cdot}\mathbf{A}_{e,\text{-}k,h\cdot}\cdot\|\mathbf{A}_{e,k,j\cdot}\|^2 + \mathbb{1}_{\{h=l\}}\cdot\Sigma_{\epsilon,k,h,jj}) = O\Big(\frac{1}{T}\Big),
\end{aligned} \tag{3.23}$$

where for the second last equality, we argue for the first term in the second last line only, as the second term could be shown similarly:

$$\sum_{q\geq 0}\sum_{p\geq 0}\sum_{s\in\psi_{k,ij,h}} a_{e,q}a_{e,q+|t-s|}a_{f,p}a_{f,p+|t-s|} = \sum_{q\geq 0}\sum_{p\geq 0} a_{e,q}a_{f,p}\sum_{s\in\psi_{k,ij,h}} a_{e,q+|t-s|}a_{f,p+|t-s|}$$
$$\leq \sum_{q\geq 0}\sum_{p\geq 0}|a_{e,q}||a_{f,p}|\Big(\sum_{s\in\psi_{k,ij,h}} a^2_{e,q+|t-s|}\Big)^{1/2}\Big(\sum_{s\in\psi_{k,ij,h}} a^2_{f,p+|t-s|}\Big)^{1/2} \leq \sum_{q\geq 0}\sum_{p\geq 0}|a_{e,q}||a_{f,p}| \leq c^2,$$

where the constant $c$ is from Assumption (F1) and (E2). Finally,

$$\mathbb{E}\Big\{\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{k,ij,h}}\frac{1}{d_{\text{-}k}\cdot|\psi_{k,ij,h}|}\,\text{mat}_k(\mathcal{E}_t)_{jh}\mathbf{u}'\text{mat}_k(\mathcal{F}_t)\mathbf{v}\Big\}^2$$
$$= \frac{1}{d^2_{\text{-}k}}\sum_{h,l=1}^{d_{\text{-}k}}\sum_{t\in\psi_{k,ij,h}}\sum_{s\in\psi_{k,ij,l}}\frac{1}{|\psi_{k,ij,h}||\psi_{k,ij,l}|}\mathbb{E}\Big(\text{mat}_k(\mathcal{E}_t)_{jh}\mathbf{u}'\text{mat}_k(\mathcal{F}_t)\mathbf{v}\,\text{mat}_k(\mathcal{E}_s)_{jl}\mathbf{v}'\text{mat}_k(\mathcal{F}_s)'\mathbf{u}\Big)$$
$$= \frac{1}{d^2_{\text{-}k}T}\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{k,ij,h}} O\Big(\frac{1}{T}\Big) = O\Big(\frac{1}{d_{\text{-}k}T}\Big),$$

which then implies result 2 of the Proposition.

Finally, we prove result 3 of the Proposition. From Assumption (F1), we have $\mathbb{E}[\mathcal{F}_t] = 0$. Next, for any $t \in [T]$, it is direct from Assumption (F1) that all elements in $\mathcal{F}_t$ are independent. Moreover,

$$\mathbb{E}[\text{mat}_k(\mathcal{F}_t)\text{mat}_k(\mathcal{F}_t)'] = \mathbb{E}\Big\{\Big(\sum_{q\geq 0} a_{f,q}\text{mat}_k(\mathcal{X}_{f,t-q})\Big)\Big(\sum_{q\geq 0} a_{f,q}\text{mat}_k(\mathcal{X}_{f,t-q})'\Big)\Big\}$$
$$= \sum_{q\geq 0} a^2_{f,q}\mathbb{E}[\text{mat}_k(\mathcal{X}_{f,t-q})\,\text{mat}_k(\mathcal{X}_{f,t-q})'] = \Big(\sum_{q\geq 0} a^2_{f,q}\Big)\cdot r_{\text{-}k}\mathbf{I}_{r_k} = r_{\text{-}k}\mathbf{I}_{r_k},$$

where we use Assumption (F1) in the last line. To complete the proof, without loss of generality, consider the variance of the $j$-th diagonal element of $\text{mat}_k(\mathcal{F}_t)\text{mat}_k(\mathcal{F}_t)'$. From Assumption

(F1), we have

$$
\begin{aligned}
&\mathrm{Var}\Big\{\frac{1}{T}\sum_{t=1}^{T}[\mathrm{mat}_k(\mathcal{F}_t)]'_{j\cdot}[\mathrm{mat}_k(\mathcal{F}_t)]_{j\cdot}\Big\}\\
&=\frac{1}{T^2}\mathrm{Var}\Big\{\sum_{t=1}^{T}\Big(\sum_{q\geq 0}a_{f,q}[\mathrm{mat}_k(\mathcal{X}_{f,t-q})]'_{j\cdot}\Big)\Big(\sum_{q\geq 0}a_{f,q}[\mathrm{mat}_k(\mathcal{X}_{f,t-q})]_{j\cdot}\Big)\Big\}\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\Big\{\Big(\sum_{q\geq 0}a_{f,q}[\mathrm{mat}_k(\mathcal{X}_{f,t-q})]'_{j\cdot}\Big)\Big(\sum_{q\geq 0}a_{f,q}[\mathrm{mat}_k(\mathcal{X}_{f,t-q})]_{j\cdot}\Big),\\
&\qquad\qquad\Big(\sum_{q\geq 0}a_{f,q}[\mathrm{mat}_k(\mathcal{X}_{f,s-q})]'_{j\cdot}\Big)\Big(\sum_{q\geq 0}a_{f,q}[\mathrm{mat}_k(\mathcal{X}_{f,s-q})]_{j\cdot}\Big)\Big\}\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\sum_{q\geq 0}a_{f,q}^4\mathrm{Var}\Big([\mathrm{mat}_k(\mathcal{X}_{f,t-q})]'_{j\cdot}[\mathrm{mat}_k(\mathcal{X}_{f,t-q})]_{j\cdot}\Big)\\
&\quad+\frac{1}{T^2}\sum_{t=1}^{T}\sum_{q\geq 0}\sum_{p\neq q}a_{f,q}^2a_{f,p}^2\mathrm{Var}\Big([\mathrm{mat}_k(\mathcal{X}_{f,t-q})]'_{j\cdot}[\mathrm{mat}_k(\mathcal{X}_{f,t-p})]_{j\cdot}\Big)\\
&=\frac{r_{\text{-}k}}{T^2}\sum_{t=1}^{T}\sum_{q\geq 0}a_{f,q}^4\mathrm{Var}\Big([\mathrm{mat}_k(\mathcal{X}_{f,t-q})]_{j1}^2\Big)+\frac{r_{\text{-}k}}{T^2}\sum_{t=1}^{T}\sum_{q\geq 0}\sum_{p\neq q}a_{f,q}^2a_{f,p}^2\\
&=\frac{1}{T^2}O\Big(\sum_{t=1}^{T}\sum_{q\geq 0}a_{f,q}^4+\sum_{t=1}^{T}\sum_{q\geq 0}\sum_{p\neq q}a_{f,q}^2a_{f,p}^2\Big)=O\Big\{\frac{1}{T^2}\sum_{t=1}^{T}\Big(\sum_{q\geq 0}a_{f,q}^2\Big)^2\Big\}=O\Big(\frac{1}{T}\Big)=o(1),
\end{aligned}
$$

where the third equality uses the independence in Assumption (E2). This completes the proof of result 3, and hence the Proposition. □

To prove Theorem 3.1, we first present some lemmas and prove them. From (3.18),

$$
\widehat{\mathbf{Q}}_{j\cdot}=\widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\widehat{S}_{ij}=\widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}Y_{t,ih}Y_{t,jh}. \tag{3.24}
$$

With the notations in (3.19), (3.1) can be written as $\mathbf{Y}_t=\mathbf{Q}\mathbf{F}_{Z,t}\boldsymbol{\Lambda}'+\mathbf{E}_t$, and hence for $i,j\in[d_k],h\in[d_{\text{-}k}]$,

$$
\begin{aligned}
Y_{t,ih}&=\Big(\sum_{n=1}^{r_k}\sum_{m=1}^{r_{\text{-}k}}Q_{in}\Lambda_{hm}F_{Z,t,nm}\Big)+E_{t,ih}\\
&=\mathbf{Q}'_{i\cdot}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)+E_{t,ih}=\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}+E_{t,ih}.
\end{aligned}
$$

Hence the product $Y_{t,ih}Y_{t,jh}$ in (3.24) can be written as

$$\begin{aligned} Y_{t,ih}Y_{t,jh} = &\Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}\Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{j\cdot} + E_{t,ih}E_{t,jh} \\ &+ E_{t,jh}\Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot} + E_{t,ih}\Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{j\cdot}. \end{aligned} \tag{3.25}$$

We then have, from (3.24) and (3.25) that

$$\begin{aligned} \widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}_j\mathbf{Q}_{j\cdot} = \widehat{\mathbf{D}}^{-1}\Big\{ &\sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih}E_{t,jh} \\ &+ \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,jh}\Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot} \\ &+ \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih}\Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{j\cdot}\Big\} \\ =: \widehat{\mathbf{D}}^{-1}\Big(&\mathcal{I}_j + \mathcal{II}_j + \mathcal{III}_j\Big), \end{aligned} \tag{3.26}$$

where

$$\begin{aligned} \mathcal{I}_j &:= \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih}E_{t,jh}, \\ \mathcal{II}_j &:= \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,jh}\Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}, \\ \mathcal{III}_j &:= \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih}\Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{j\cdot}. \end{aligned}$$

The following lemma bounds the terms $\mathcal{I}_j, \mathcal{II}_j$, and $\mathcal{III}_j$.

**Lemma 3.1** *Under Assumptions (O1), (F1), (L1), (E1) and (E2), we have*

$$\frac{1}{d_k}\sum_{j=1}^{d_k} \|\mathcal{I}_j\|_F^2 = O_P\Big(\frac{d}{T} + d_{\text{-}k}^2\Big), \tag{3.27}$$

$$\frac{1}{d_k}\sum_{j=1}^{d_k} \|\mathcal{II}_j\|_F^2 = O_P\Big(\frac{d_{\text{-}k}d_k^{\alpha_{k,1}}}{T}\Big) = \frac{1}{d_k}\sum_{j=1}^{d_k} \|\mathcal{III}_j\|_F^2. \tag{3.28}$$

***Proof of Lemma 3.1.*** To prove (3.27), we decompose

$$\begin{aligned}
\mathcal{I}_j &= \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{-k}} \frac{1}{|\psi_{ij,h}|} \sum_{t \in \psi_{ij,h}} E_{t,ih} E_{t,jh} \\
&= \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{-k}} \left\{ \frac{\sum_{t\in\psi_{ij,h}} \big(E_{t,ih}E_{t,jh} - \mathbb{E}[E_{t,ih}E_{t,jh}]\big)}{|\psi_{ij,h}|} + \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}E_{t,jh}]}{|\psi_{ij,h}|} \right\} \qquad (3.29) \\
&=: \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \xi_{ij} + \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \eta_{ij}, \quad \text{where}
\end{aligned}$$

$$\xi_{ij} := \sum_{h=1}^{d_{-k}} \frac{\sum_{t\in\psi_{ij,h}} \big(E_{t,ih}E_{t,jh} - \mathbb{E}[E_{t,ih}E_{t,jh}]\big)}{|\psi_{ij,h}|}, \quad \eta_{ij} := \sum_{h=1}^{d_{-k}} \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}E_{t,jh}]}{|\psi_{ij,h}|}.$$

We want to show the following:

$$\sum_{j=1}^{d_k} \Big\| \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \xi_{ij} \Big\|_F^2 = O_P\Big(\frac{dd_k}{T}\Big), \qquad (3.30)$$

$$\sum_{j=1}^{d_k} \Big\| \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \eta_{ij} \Big\|_F^2 = O_P\big(dd_{-k}\big). \qquad (3.31)$$

To show (3.30), first note that $\mathbb{E}\xi_{ij} = 0$, and also by Assumption (O1),

$$\begin{aligned}
\mathbb{E}|\xi_{ij}|^2 &= \text{Var}(\xi_{ij}) \le \frac{1}{\psi_0^2 T^2} \text{Var}\Big\{ \sum_{h=1}^{d_{-k}} \sum_{t\in\psi_{ij,h}} \Big( E_{t,ih}E_{t,jh} - \mathbb{E}\big[E_{t,ih}E_{t,jh}\big]\Big)\Big\} \\
&\le \frac{1}{\psi_0^2 T^2} \sum_{h=1}^{d_{-k}} \sum_{l=1}^{d_{-k}} \sum_{t\in\psi_{ij,h}} \sum_{s\in\psi_{ij,l}} \Big|\text{Cov}\Big(E_{t,ih}E_{t,jh} - \mathbb{E}\big[E_{t,ih}E_{t,jh}\big], E_{s,il}E_{s,jl} - \mathbb{E}\big[E_{s,il}E_{s,jl}\big]\Big)\Big| \\
&= \frac{1}{\psi_0^2 T^2} \sum_{h=1}^{d_{-k}} \sum_{t\in\psi_{ij,h}} \sum_{l=1}^{d_{-k}} \sum_{s\in\psi_{ij,l}} \Big|\text{Cov}\Big(E_{t,ih}E_{t,jh}, E_{s,il}E_{s,jl}\Big)\Big| \le \frac{cd_{-k}}{\psi_0^2 T}, \qquad (3.32)
\end{aligned}$$

where the last inequality and the constant $c$ are from result 1 of Proposition 3.2 (hereafter Proposition 3.2.1, etc.). Then by the Cauchy–Schwarz inequality,

$$\sum_{j=1}^{d_k} \Big\| \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \xi_{ij} \Big\|_F^2 \le \sum_{j=1}^{d_k} \Big( \sum_{i=1}^{d_k} \|\widehat{\mathbf{Q}}_{i\cdot}\|_F^2 \Big) \Big( \sum_{i=1}^{d_k} \xi_{ij}^2 \Big) = O_P\Big(\frac{dd_k}{T}\Big),$$

which is (3.30). To show (3.31), note that if we define

$$\rho_{ij,h} := \frac{\frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}E_{t,jh}]}{\big(\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}^2]\big)^{1/2} \big(\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,jh}^2]\big)^{1/2}},$$

then $|\rho_{ij,h}| < 1$ and hence $\rho_{ij,h}^2 \leq |\rho_{ij,h}|$. It is then easy to prove also that

$$\rho_{ij} := \frac{\sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}E_{t,jh}]}{\big(\sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}^2]\big)^{1/2} \big(\sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,jh}^2]\big)^{1/2}},$$

also satisfy $|\rho_{ij}| \leq 1$ and $\rho_{ij}^2 \leq |\rho_{ij}|$. By Proposition 3.2.1,

$$\Big| \sum_{h=1}^{d_{\text{-}k}} \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}^2]}{|\psi_{ij,h}|} \Big| = O_P(d_{\text{-}k}),$$

and hence

$$\begin{aligned}
\eta_{ij}^2 &= \Big(\sum_{h=1}^{d_{\text{-}k}} \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}E_{t,jh}]}{|\psi_{ij,h}|}\Big)^2 = \rho_{ij}^2 \Big(\sum_{h=1}^{d_{\text{-}k}} \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}^2]}{|\psi_{ij,h}|}\Big)\Big(\sum_{h=1}^{d_{\text{-}k}} \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,jh}^2]}{|\psi_{ij,h}|}\Big) \\
&= |\rho_{ij}| \Big(\sum_{h=1}^{d_{\text{-}k}} \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}^2]}{|\psi_{ij,h}|}\Big)^{1/2} \Big(\sum_{h=1}^{d_{\text{-}k}} \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,jh}^2]}{|\psi_{ij,h}|}\Big)^{1/2} \cdot O_P(d_{\text{-}k}) \\
&= \Big| \sum_{h=1}^{d_{\text{-}k}} \frac{\sum_{t\in\psi_{ij,h}} \mathbb{E}[E_{t,ih}E_{t,jh}]}{|\psi_{ij,h}|} \Big| \cdot O_P(d_{\text{-}k}) = \big|\eta_{ij}\big| O_P(d_{\text{-}k}).
\end{aligned}$$

Using the above, we then have

$$\begin{aligned}
&\sum_{j=1}^{d_k} \Big\| \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot}\eta_{ij} \Big\|_F^2 \leq \sum_{j=1}^{d_k} \Big( \sum_{i=1}^{d_k} \|\widehat{\mathbf{Q}}_{i\cdot}\|_F^2 \Big) \Big( \sum_{i=1}^{d_k} \eta_{ij}^2 \Big) \leq r_k \sum_{j=1}^{d_k}\sum_{i=1}^{d_k} \eta_{ij}^2 \\
&= O_P(d_{\text{-}k}) \cdot \sum_{j=1}^{d_k}\sum_{i=1}^{d_k} |\eta_{ij}| = O_P(d_{\text{-}k}) \cdot \frac{1}{|\psi_0 T|} \sum_{t\in\psi_{ij,h}} \sum_{h=1}^{d_{\text{-}k}} \sum_{j=1}^{d_k}\sum_{i=1}^{d_k} \big|\mathbb{E}[E_{t,ih}E_{t,jh}]\big| = O_P(dd_{\text{-}k}),
\end{aligned} \tag{3.33}$$

where the second last equality used Assumption (O1), and the last equality used Proposition 3.2.1. This proves (3.31). Using (3.30) and (3.31), from (3.29) we have

$$\begin{aligned}
\frac{1}{d_k}\sum_{j=1}^{d_k} \|\mathcal{I}_j\|_F^2 &= \frac{1}{d_k}\sum_{j=1}^{d_k} \Big\| \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \xi_{ij,h} + \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \eta_{ij,h} \Big\|_F^2 \\
&\leq \frac{2}{d_k}\sum_{j=1}^{d_k} \Big\| \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot}\xi_{ij} \Big\|_F^2 + \frac{2}{d_k}\sum_{j=1}^{d_k} \Big\| \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{i\cdot}\eta_{ij} \Big\|_F^2 = O_P\Big(\frac{d}{T} + d_{\text{-}k}^2\Big).
\end{aligned}$$

This completes the proof of (3.27). To prove (3.28), consider

$$
\begin{aligned}
\frac{1}{d_k}\sum_{j=1}^{d_k}\|\mathcal{II}_j\|_F^2 &= \frac{1}{d_k}\sum_{j=1}^{d_k}\Big\|\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}\Big\|_F^2 \\
&\le \frac{1}{d_k}\sum_{j=1}^{d_k}\Big(\sum_{i=1}^{d_k}\|\widehat{\mathbf{Q}}_{i\cdot}\|_F^2\Big)\cdot\sum_{i=1}^{d_k}\Big\{\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}\Big\}^2 \\
&= \frac{r_k}{d_k}\sum_{j=1}^{d_k}\sum_{i=1}^{d_k}\Big\{\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}\Big\}^2 \\
&= \frac{r_k d_{\text{-}k}^2}{d_k}\sum_{j=1}^{d_k}\sum_{i=1}^{d_k}\Bigg(\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\frac{1}{d_{\text{-}k}\cdot|\psi_{ij,h}|}E_{t,jh}[\otimes_{l\in[K]\setminus\{k\}}\mathbf{A}_l]'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\Bigg)^2 \\
&= \frac{r_k d_{\text{-}k}^2}{d_k}\sum_{j=1}^{d_k}\sum_{i=1}^{d_k}\|\mathbf{u}_i\|_F^2\Bigg(\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\frac{1}{d_{\text{-}k}\cdot|\psi_{ij,h}|}E_{t,jh}\mathbf{v}'_h\mathbf{F}'_t\frac{1}{\|\mathbf{u}_i\|_F}\mathbf{u}_i\Bigg)^2,
\end{aligned}
\tag{3.34}
$$

where $\mathbf{A}=\mathbf{A}_k$ and $\mathbf{F}_t=\text{mat}_k(\mathcal{F}_t)$ above, and we define $\mathbf{v}_h := [\otimes_{l\in[K]\setminus\{k\}}\mathbf{A}_l]_{h\cdot}$, $\mathbf{u}_i := \mathbf{A}_{i\cdot}$. By Proposition 3.2.2, the last bracket in the last line of (3.34) is $O_P(d_{\text{-}k}^{-1}T^{-1})$, and hence

$$
\frac{1}{d_k}\sum_{j=1}^{d_k}\|\mathcal{II}_j\|_F^2 = O_P\Big(\frac{d_{\text{-}k}}{d_kT}\Big)\sum_{j=1}^{d_k}\sum_{i=1}^{d_k}\|\mathbf{u}_i\|_F^2 = O_P\Big(\frac{d_{\text{-}k}}{d_kT}\Big)\sum_{j=1}^{d_k}\big\|\mathbf{A}\big\|_F^2 = O_P\Big(\frac{d_{\text{-}k}d_k^{\alpha_{k,1}}}{T}\Big),
$$

where the last equality follows since for any $l\in[K]$, $\|\mathbf{A}_l\|_F^2 = O_P\big(\text{tr}(\mathbf{Z}_l)\big) = O_P\big(d_l^{\alpha_{l,1}}\big)$ by Assumption (L1). The bound corresponding to $\mathcal{III}_j$ can be proved similarly (omitted), and hence (3.28) is established. This concludes the proof of Lemma 3.1. □

**Lemma 3.2** *Under Assumptions (O1), (M1), (F1), (L1), (E1), (E2) and (R1), with* $\mathbf{H}_j$ *and* $\widehat{\mathbf{D}}$ *from* (3.19)*, we have*

$$
\big\|\widehat{\mathbf{D}}^{-1}\big\|_F = O_P\Big(d_k^{\alpha_{k,1}-\alpha_{k,r_k}}\prod_{j=1}^{K}d_j^{-\alpha_{j,1}}\Big), \tag{3.35}
$$

$$
\frac{1}{d_k}\sum_{j=1}^{d_k}\Big\|\widehat{\mathbf{Q}}_{j\cdot}-\mathbf{H}_j\mathbf{Q}_{j\cdot}\Big\|_F^2 = O_P\Big\{d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})-1}\Big(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\Big)\prod_{j=1}^{K}d_j^{2(1-\alpha_{j,1})}\Big\}. \tag{3.36}
$$

***Proof of Lemma 3.2.*** First, we bound the term $\big\|\widehat{\mathbf{D}}^{-1}\big\|_F^2$ by finding the lower bound of $\lambda_{r_k}(\widehat{\mathbf{D}})$. To do this, define $\omega_k := d_k^{\alpha_{k,r_k}-\alpha_{k,1}}\prod_{j=1}^{K}d_j^{\alpha_{j,1}}$, and consider the decomposition

$$
\widehat{\mathbf{S}} = \mathbf{R}^* + (\widetilde{\mathbf{R}}-\mathbf{R}^*) + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3, \tag{3.37}
$$

where for a unit vector $\boldsymbol{\gamma}$,

$$\begin{aligned}
R(\boldsymbol{\gamma}) &:= \frac{1}{\omega_k}\boldsymbol{\gamma}'\widehat{\mathbf{S}}\boldsymbol{\gamma} = \frac{1}{\omega_k}\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\gamma_i\gamma_j\widehat{S}_{ij} = \frac{1}{\omega_k}\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\gamma_i\gamma_j\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}Y_{t,ih}Y_{t,jh} \\
&=: R^*(\boldsymbol{\gamma}) + (\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma})) + R_1 + R_2 + R_3, \quad \text{with} \\
\widetilde{R}(\boldsymbol{\gamma}) &:= \frac{1}{\omega_k}\boldsymbol{\gamma}'\widetilde{\mathbf{R}}\boldsymbol{\gamma} := \frac{1}{\omega_k}\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\frac{\gamma_i\gamma_j}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{j\cdot}, \\
R^*(\boldsymbol{\gamma}) &:= \frac{1}{\omega_k}\boldsymbol{\gamma}'\mathbf{R}^*\boldsymbol{\gamma} := \frac{1}{\omega_k}\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\frac{\gamma_i\gamma_j}{T}\sum_{t=1}^{T}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{j\cdot}, \\
R_1 &:= \frac{1}{\omega_k}\boldsymbol{\gamma}'\mathbf{R}_1\boldsymbol{\gamma} := \frac{1}{\omega_k}\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\frac{\gamma_i\gamma_j}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,ih}E_{t,jh}, \\
R_2 &:= \frac{1}{\omega_k}\boldsymbol{\gamma}'\mathbf{R}_2\boldsymbol{\gamma} := \frac{1}{\omega_k}\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\frac{\gamma_i\gamma_j}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}, \\
R_3 &:= \frac{1}{\omega_k}\boldsymbol{\gamma}'\mathbf{R}_3\boldsymbol{\gamma} := \frac{1}{\omega_k}\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\frac{\gamma_i\gamma_j}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,ih}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{j\cdot},
\end{aligned} \tag{3.38}$$

and we used (3.25) for the expansion above. Then we have the decomposition

$$R(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma}) = R(\boldsymbol{\gamma}) - \widetilde{R}(\boldsymbol{\gamma}) + \widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma}). \tag{3.39}$$

Similar to the treatment of the term $\mathcal{I}_j$ in the proof of Lemma 3.1, since $\|\boldsymbol{\gamma}\| = 1$,

$$\begin{aligned}
|R_1| &\le \frac{1}{\omega_k}\Big|\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\gamma_i\gamma_j\xi_{ij}\Big| + \frac{1}{\omega_k}\Big|\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\gamma_i\gamma_j\eta_{ij}\Big| \le \frac{1}{\omega_k}\Big(\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\xi_{ij}^2\Big)^{\frac{1}{2}} + \frac{1}{\omega_k}\Big(\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\eta_{ij}^2\Big)^{\frac{1}{2}} \\
&= O_P\Bigg\{d_k^{\alpha_{k,1}-\alpha_{k,r_k}}\left(\frac{1}{T^{1/2}d_{\text{-}k}^{1/2}} + \frac{1}{d_k^{1/2}}\right)\prod_{j=1}^{K}d_j^{1-\alpha_{j,1}}\Bigg\} = O_P\Big(d\,[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}]/\omega_k\Big),
\end{aligned} \tag{3.40}$$

where the second last equality is from (3.32) and part of (3.33). Together with Assumption (R1), (3.40) implies that as $T, d_k, d_{\text{-}k} \to \infty$, we have $R_1 \xrightarrow{p} 0$.

From (3.34) and the arguments for $\mathcal{II}_j$ immediately afterwards, we see that

$$|R_2| \le \frac{1}{\omega_k}\Bigg\{\sum_{j=1}^{d_k}\sum_{i=1}^{d_k}\Bigg(\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}\Big(\sum_{m=1}^{r_{\text{-}k}}\Lambda_{hm}\mathbf{F}_{Z,t,\cdot m}\Big)'\mathbf{Q}_{i\cdot}\Bigg)^2\Bigg\}^{1/2}$$

$$= O_P(\omega_k^{-1}) \cdot O_P\Big(T^{-1/2} d^{1/2} \prod_{j=1}^{K} d_j^{\alpha_{j,1}/2}\Big) = O_P\Big((dg_s)^{1/2} T^{-1/2}/\omega_k\Big) = o_P(1), \quad (3.41)$$

where the last equality is from Assumption (R1). The term $R_3$ can be proved to have the same rate with same lines of proof as for $R_2$. Hence we have

$$\sup_{\|\boldsymbol{\gamma}\|=1} |R(\boldsymbol{\gamma}) - \widetilde{R}(\boldsymbol{\gamma})| = \sup_{\|\boldsymbol{\gamma}\|=1} |R_1 + R_2 + R_3| \xrightarrow{p} 0. \quad (3.42)$$

Similar to the proof of (R6) in Lemma 4 in Xiong and Pelger (2023), using the definition $\mathbf{v}_h := [\otimes_{l\in[K]\setminus\{k\}} \mathbf{A}_l]_{h\cdot}$ as before,

$$\begin{aligned}
&\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma}) \\
&= \frac{1}{\omega_k} \sum_{i=1}^{d_k}\sum_{j=1}^{d_k} \gamma_i\gamma_j \sum_{h=1}^{d_{\text{-}k}} \Big( \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \mathbf{Q}_{i\cdot}' \mathbf{F}_{Z,t} \boldsymbol{\Lambda}_{h\cdot} \boldsymbol{\Lambda}_{h\cdot}' \mathbf{F}_{Z,t}' \mathbf{Q}_{j\cdot} - \frac{1}{T}\sum_{t=1}^{T} \mathbf{Q}_{i\cdot}' \mathbf{F}_{Z,t} \boldsymbol{\Lambda}_{h\cdot} \boldsymbol{\Lambda}_{h\cdot}' \mathbf{F}_{Z,t}' \mathbf{Q}_{j\cdot}\Big) \\
&= \frac{1}{\omega_k} \sum_{i=1}^{d_k}\sum_{j=1}^{d_k} \gamma_i\gamma_j \sum_{h=1}^{d_{\text{-}k}} \mathbf{A}_{i\cdot}' \Big( \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \mathbf{F}_t \mathbf{v}_h \mathbf{v}_h' \mathbf{F}_t' - \frac{1}{T}\sum_{t=1}^{T} \mathbf{F}_t \mathbf{v}_h \mathbf{v}_h' \mathbf{F}_t' \Big) \mathbf{A}_{j\cdot} \\
&=: \frac{1}{\omega_k} \sum_{i=1}^{d_k}\sum_{j=1}^{d_k} \gamma_i\gamma_j \sum_{h=1}^{d_{\text{-}k}} \mathbf{A}_{i\cdot}' \Delta_{F,k,ij,h} \mathbf{A}_{j\cdot}, \quad \text{where} \qquad (3.43)\\
&\Delta_{F,k,ij,h} := \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \mathbf{F}_t \mathbf{v}_h \mathbf{v}_h' \mathbf{F}_t' - \frac{1}{T}\sum_{t=1}^{T} \mathbf{F}_t \mathbf{v}_h \mathbf{v}_h' \mathbf{F}_t'.
\end{aligned}$$

By the Cauchy–Schwarz inequality, we then have

$$\big|\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma})\big| \le \frac{1}{\omega_k} \Big\{ \sum_{i=1}^{d_k}\sum_{j=1}^{d_k} \Big[ \mathbf{A}_{i\cdot}' \Big( \sum_{h=1}^{d_{\text{-}k}} \Delta_{F,k,ij,h} \Big) \mathbf{A}_{j\cdot} \Big]^2 \Big\}^{1/2}. \quad (3.44)$$

With Assumption (M1), using the standard rate of convergence in the weak law of large number for $\alpha$-mixing sequence and the fact that the elements in $\mathcal{F}_t$ are independent from Assumption (F1), since $\Delta_{F,k,ij,h}$ has fixed dimension, we have for each $k \in [K]$, $i, j \in [d_k]$ and $h \in [d_{\text{-}k}]$,

$$\|\Delta_{F,k,ij,h}\|_F \le \Big\| \sum_{t\in\psi_{ij,h}} \frac{\mathbf{F}_t \mathbf{v}_h \mathbf{v}_h' \mathbf{F}_t'}{|\psi_{ij,h}|} - \mathbf{v}_h' \mathbf{v}_h \boldsymbol{\Sigma}_k \Big\|_F + \Big\| \sum_{t=1}^{T} \frac{\mathbf{F}_t \mathbf{v}_h \mathbf{v}_h' \mathbf{F}_t'}{T} - \mathbf{v}_h' \mathbf{v}_h \boldsymbol{\Sigma}_k \Big\|_F = O_P\Big( \frac{\|\mathbf{v}_h\|^2}{T^{1/2}} \Big). \quad (3.45)$$

From Assumption (L1), we then have

$$\sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\Big\{\mathbf{A}_{i\cdot}'\Big(\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\Big)\mathbf{A}_{j\cdot}\Big\}^2 \le \sum_{i=1}^{d_k}\sum_{j=1}^{d_k}\|\mathbf{A}_{i\cdot}\|_F^2\|\mathbf{A}_{j\cdot}\|_F^2\Big(\sum_{h=1}^{d_{\text{-}k}}\|\Delta_{F,k,ij,h}\|_F\Big)^2$$
$$\le \|\mathbf{A}_k\|_F^4\cdot O_P\Big\{\frac{1}{T}\Big(\sum_{h=1}^{d_{\text{-}k}}\|\mathbf{v}_h\|^2\Big)^2\Big\} = \|\mathbf{A}_k\|_F^4\cdot O_P\Big(\frac{1}{T}\|\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j\|_F^4\Big)$$
$$= O_P\Big(\frac{1}{T}\prod_{j=1}^{K}\|\mathbf{A}_j\|_F^4\Big) = O_P\Big(\frac{1}{T}\prod_{j=1}^{K}d_j^{2\alpha_{j,1}}\Big), \tag{3.46}$$

and hence from (3.44), we have by Assumption (R1) that

$$\big|\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma})\big| = O_P\Big(\frac{1}{\sqrt{T}}d_k^{\alpha_{k,1}-\alpha_{k,r_k}}\Big) = o_P(d^{-1/2}g_s^{1/2}) = o_P(1), \tag{3.47}$$

where the second last equality is from Assumption (R1). Next, with Proposition 3.2.3, consider

$$\lambda_{r_k}(\mathbf{R}^*) = \lambda_{r_k}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}\mathbf{F}_{Z,t}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}_{Z,t}'\mathbf{Q}'\Big)$$
$$= \lambda_{r_k}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}_k\mathbf{F}_t\big[\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j\big]'\big[\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j\big]\mathbf{F}_t'\mathbf{A}_k'\Big)$$
$$\ge \lambda_{r_k}(\mathbf{A}_k'\mathbf{A}_k)\cdot\lambda_{r_k}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\big[\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j\big]'\big[\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j\big]\mathbf{F}_t'\Big)$$
$$\asymp_P d_k^{\alpha_{k,r_k}}\cdot\lambda_{r_k}(\text{tr}(\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j'\mathbf{A}_j)\boldsymbol{\Sigma}_k) \asymp_P d_k^{\alpha_{k,r_k}}\prod_{j\in[K]\setminus\{k\}}d_j^{\alpha_{j,1}} = \omega_k.$$

With this, going back to the decomposition (3.37),

$$\omega_k^{-1}\lambda_{r_k}(\widehat{\mathbf{D}}) = \frac{\lambda_{r_k}(\widehat{\mathbf{S}})}{\omega_k} \ge \frac{\lambda_{r_k}(\mathbf{R}^*)}{\omega_k} - \sup_{\|\boldsymbol{\gamma}\|=1}|\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma})| - \sup_{\|\boldsymbol{\gamma}\|=1}|R_1+R_2+R_3| \asymp_P 1,$$

where we used (3.42) and (3.47). Hence finally,

$$\big\|\widehat{\mathbf{D}}^{-1}\big\|_F = O_P\big(\lambda_{r_k}^{-1}(\widehat{\mathbf{D}})\big) = O_P(\omega_k^{-1}) = O_P\Big(d_k^{\alpha_{k,1}-\alpha_{k,r_k}}\prod_{j=1}^{K}d_j^{-\alpha_{j,1}}\Big),$$

which completes the proof of (3.35).

To prove (3.36), from (3.26) we obtain

$$\frac{1}{d_k}\sum_{j=1}^{d_k}\Big\|\widehat{\mathbf{Q}}_{j\cdot}-\mathbf{H}_j\mathbf{Q}_{j\cdot}\Big\|_F^2=\frac{1}{d_k}\sum_{j=1}^{d_k}\Big\|\widehat{\mathbf{D}}^{-1}\Big(\mathcal{I}_j+\mathcal{II}_j+\mathcal{III}_j\Big)\Big\|_F^2$$
$$\leq \|\widehat{\mathbf{D}}^{-1}\|_F^2\Big(\frac{1}{d_k}\sum_{j=1}^{d_k}\Big\|\mathcal{I}_j+\mathcal{II}_j+\mathcal{III}_j\Big\|_F^2\Big)$$
$$\leq \|\widehat{\mathbf{D}}^{-1}\|_F^2\Big\{\Big(\frac{2}{d_k}\sum_{j=1}^{d_k}\|\mathcal{I}_j\|_F^2\Big)+\Big(\frac{4}{d_k}\sum_{j=1}^{d_k}\|\mathcal{II}_j\|_F^2\Big)+\Big(\frac{4}{d_k}\sum_{j=1}^{d_k}\|\mathcal{III}_j\|_F^2\Big)\Big\}$$
$$=O_P\Big\{d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})-1}\Big(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\Big)\prod_{j=1}^{K}d_j^{2(1-\alpha_{j,1})}\Big\},$$

where the last line used (3.35) and Lemma 3.1. This concludes the proof of Lemma 3.2. □

**Lemma 3.3** *Let all the assumptions in Lemma 3.2 hold. For any $j\in[d_k]$, with $\mathbf{H}_j$ and $\mathbf{H}^a$ from* (3.19) *and the notation $\eta=1-\psi_0$,*

$$\big\|\mathbf{H}_j-\mathbf{H}^a\big\|_F^2=O_P\Big\{\min\Big(\frac{1}{T},\frac{\eta^2}{(1-\eta)^2}\Big)d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}\Big\}=o_P(1).$$

***Proof of Lemma 3.3.*** Firstly, consider $\Delta_{F,k,ij,h}$ from (3.43), where

$$\|\Delta_{F,k,ij,h}\|_F=\Big\|\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}\mathbf{F}_t\mathbf{v}_h\mathbf{v}_h'\mathbf{F}_t'-\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{v}_h\mathbf{v}_h'\mathbf{F}_t'\Big\|_F$$
$$=\Big\|\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}^c}\mathbf{F}_t\mathbf{v}_h\mathbf{v}_h'\mathbf{F}_t'\Big\|_F+\Big\|\Big(\frac{1}{|\psi_{ij,h}|}-\frac{1}{T}\Big)\sum_{t=1}^{T}\mathbf{F}_t\mathbf{v}_h\mathbf{v}_h'\mathbf{F}_t'\Big\|_F$$
$$\leq O_P\Big(\frac{T\eta}{T-T\eta}\big\|\mathbf{v}_h\big\|^2\Big)+\frac{T\eta}{T(T-T\eta)}\cdot O_P\Big(T\big\|\mathbf{v}_h\big\|^2\Big)=O_P\Big(\frac{\eta}{1-\eta}\big\|\mathbf{v}_h\big\|^2\Big).$$

Combining this with (3.45), we have

$$\big\|\Delta_{F,k,ij,h}\big\|_F=O_P\Big\{\min\Big(\frac{1}{\sqrt{T}},\frac{\eta}{1-\eta}\Big)\big\|\mathbf{v}_h\big\|^2\Big\}. \tag{3.48}$$

Note also the following two results that

$$\mathbf{H}_j = \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}\mathbf{\Lambda}'_{h\cdot}\mathbf{F}'_{Z,t}\mathbf{Q}_{i\cdot}\mathbf{\Lambda}'_{h\cdot}\mathbf{F}'_{Z,t}$$
$$= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}\mathbf{Q}'_{i\cdot}\mathbf{F}_{Z,t}\mathbf{\Lambda}_{h\cdot}\mathbf{\Lambda}'_{h\cdot}\mathbf{F}'_{Z,t};$$
$$\mathbf{H}^a = \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}'_{i\cdot}\mathbf{F}_{Z,t}\mathbf{\Lambda}_{h\cdot}\mathbf{\Lambda}'_{h\cdot}\mathbf{F}'_{Z,t}.$$

We then have, for $\widehat{\mathbf{A}}_{i\cdot} = \mathbf{Z}_k^{1/2}\widehat{\mathbf{Q}}_{i\cdot}$ and using (3.48),

$$\begin{aligned}
&\left\|\mathbf{H}_j - \mathbf{H}^a\right\|_F^2 \\
&= \left\|\widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\Big(\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}\mathbf{Q}'_{i\cdot}\mathbf{F}_{Z,t}\mathbf{\Lambda}_{h\cdot}\mathbf{\Lambda}'_{h\cdot}\mathbf{F}'_{Z,t} - \frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}'_{i\cdot}\mathbf{F}_{Z,t}\mathbf{\Lambda}_{h\cdot}\mathbf{\Lambda}'_{h\cdot}\mathbf{F}'_{Z,t}\Big)\right\|_F^2 \\
&= \left\|\widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\mathbf{A}'_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\mathbf{Z}_k^{1/2}\right\|_F^2 \le \left\|\widehat{\mathbf{D}}^{-1}\right\|_F^2\cdot\left\|\sum_{i=1}^{d_k}\widehat{\mathbf{A}}_{i\cdot}\mathbf{A}'_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\right\|_F^2 \\
&= O_P(\omega_k^{-2})\cdot\max_{i\in[d_k]}\left\|\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\right\|_F^2\cdot\Big(\sum_{i=1}^{d_k}\|\widehat{\mathbf{A}}_{i\cdot}\|\cdot\|\mathbf{A}_{i\cdot}\|\Big)^2 \\
&= O_P(\omega_k^{-2})\cdot O_P\Big\{\min\Big(\frac{1}{T},\frac{\eta^2}{(1-\eta)^2}\Big)\cdot\big\|\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j\big\|_F^4\Big\}\cdot\big\|\widehat{\mathbf{A}}_k\big\|_F^2\cdot\|\mathbf{A}_k\|_F^2 \\
&= O_P\Big\{\min\Big(\frac{1}{T},\frac{\eta^2}{(1-\eta)^2}\Big)d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}\Big\} = o_P(1),
\end{aligned}$$

where we used (3.48) in the second last line, Assumption (L1) in the second last equality, and Assumption (R1) in the last equality. This completes the proof of Lemma 3.3. □

***Proof of Theorem 3.1.*** The first result is shown in Lemma 3.2. Together with Lemma 3.3,

$$
\begin{aligned}
&\frac{1}{d_k}\sum_{j=1}^{d_k}\big\|\widehat{\mathbf{Q}}_{j\cdot}-\mathbf{H}^a\mathbf{Q}_{j\cdot}\big\|^2 \le \frac{1}{d_k}\sum_{j=1}^{d_k}\big\|\widehat{\mathbf{Q}}_{j\cdot}-\mathbf{H}_j\mathbf{Q}_{j\cdot}\big\|_F^2+\frac{1}{d_k}\sum_{j=1}^{d_k}\Big\|\Big(\mathbf{H}_j-\mathbf{H}^a\Big)\mathbf{Q}_{j\cdot}\Big\|_F^2\\
&\le \frac{1}{d_k}\sum_{j=1}^{d_k}\big\|\widehat{\mathbf{Q}}_{j\cdot}-\mathbf{H}_j\mathbf{Q}_{j\cdot}\big\|_F^2+\frac{1}{d_k}\sum_{j=1}^{d_k}\big\|\mathbf{H}_j-\mathbf{H}^a\big\|_F^2\big\|\mathbf{Q}_{j\cdot}\big\|_F^2\\
&= O_P\Big\{d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})-1}\Big(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\Big)\prod_{j=1}^{K}d_j^{2(1-\alpha_{j,1})}\Big\}\\
&\quad+O_P\Big\{\min\Big(\frac{1}{T},\frac{\eta^2}{(1-\eta)^2}\Big)d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})-1}\Big\}\\
&= O_P\left(d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})-1}\left[\Big(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\Big)\frac{d^2}{g_s^2}+\min\Big\{\frac{1}{T},\frac{\eta^2}{(1-\eta)^2}\Big\}\right]\right),
\end{aligned}
$$

where we used $d_k^{-1}\sum_{j=1}^{d_k}\|\mathbf{Q}_{j\cdot}\|_F^2 = d_k^{-1}\|\mathbf{Q}_k\|_F^2 = O(d_k^{-1})$ by Assumption (L1).

Thus together with Assumption (R1), we may note that

$$
\begin{aligned}
\mathbf{I}_{r_k} &= \widehat{\mathbf{Q}}'\widehat{\mathbf{Q}} = \widehat{\mathbf{Q}}'\big[\widehat{\mathbf{Q}}-\mathbf{Q}\mathbf{H}^{a'}\big]+\widehat{\mathbf{Q}}'\mathbf{Q}\mathbf{H}^{a'} = \mathbf{Q}'\widehat{\mathbf{Q}}\mathbf{H}^{a'}+o_P(1)\\
&= \mathbf{H}^a\mathbf{Q}'\mathbf{Q}\mathbf{H}^{a'}+o_P(1) = \mathbf{H}^a\mathbf{\Sigma}_{A,k}\mathbf{H}^{a'}+o_P(1),
\end{aligned}
$$

where the last equality used Assumption (L1) and it is immediate that $\mathbf{H}^a$ has full rank asymptotically. We also have

$$
\begin{aligned}
&\sigma_1(\mathbf{H}^a)\cdot\sigma_{r_k}(\mathbf{\Sigma}_{A,k})\cdot\sigma_{r_k}(\mathbf{H}^{a'}) \le \sigma_1(\mathbf{H}^a)\cdot\sigma_{r_k}(\mathbf{\Sigma}_{A,k}\mathbf{H}^{a'})\\
&\le \sigma_1(\mathbf{H}^a\mathbf{\Sigma}_{A,k}\mathbf{H}^{a'}) = O_P(\sigma_1(\mathbf{I}_{r_k})) = O_P(1),
\end{aligned}
$$

which implies $\|\mathbf{H}^a\|_F = O_P(1)$ by (L1). This completes the proof of the theorem. □

Before we prove the consistency results for our imputations, we want to prove asymptotic normality for our factor loading estimators first. Consistency for the imputations will then use the rate obtained from asymptotic normality of the estimated factor loading matrices. We present a lemma first before proving Theorem 3.3.

**Lemma 3.4** *Let Assumption (O1), (M1), (F1), (L1), (L2), (E1), (E2) and (R1) hold. For a given $k \in [K]$, let $\mathbf{R}^*$ be from* (3.37) *and $\omega_k := d_k^{\alpha_{k,r_k}-\alpha_{k,1}}g_s$. Then*

$$
\begin{aligned}
&\omega_k^{-1}\mathbf{R}^* \xrightarrow{p} \text{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k})\cdot\omega_k^{-1}\mathbf{A}_k\mathbf{A}_k',\\
&\omega_k^{-1}\widehat{\mathbf{D}}_k \xrightarrow{p} \omega_k^{-1}\mathbf{D}_k := \omega_k^{-1}\text{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k})\cdot\text{diag}\{\lambda_j(\mathbf{A}_k'\mathbf{A}_k)\mid j\in[r_k]\},\\
&\quad\mathbf{H}_k^a \xrightarrow{p} \mathbf{H}_k^{a,*} := (\text{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}))^{1/2}\cdot\mathbf{D}_k^{-1/2}\Upsilon_k'\mathbf{Z}_k^{1/2},
\end{aligned}
$$

*where* $\Upsilon_k$ *is the eigenvector matrix of* $\mathrm{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}) \cdot \omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{\Sigma}_{A,k}\mathbf{Z}_k^{1/2}$.

***Proof of Lemma 3.4.*** First, let $\widehat{\mathbf{S}}, \widehat{\mathbf{Q}}, \widehat{\mathbf{D}}, \mathbf{H}^a$ be from (3.19), and $\mathbf{R}^*, \widetilde{\mathbf{R}}, \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ from (3.37). Define also $\mathbf{H}^{a,*} := \mathbf{H}_k^{a,*}$, then we have

$$\begin{aligned} &\frac{1}{\omega_k}(\widehat{\mathbf{D}} - \widehat{\mathbf{Q}}'\mathbf{R}^*\widehat{\mathbf{Q}}) = \frac{1}{\omega_k}\widehat{\mathbf{Q}}'(\widehat{\mathbf{S}} - \mathbf{R}^*)\widehat{\mathbf{Q}} \\ &= \frac{1}{\omega_k}\widehat{\mathbf{Q}}'(\widetilde{\mathbf{R}} - \mathbf{R}^*)\widehat{\mathbf{Q}} + \frac{1}{\omega_k}\widehat{\mathbf{Q}}'\mathbf{R}_1\widehat{\mathbf{Q}} + \frac{1}{\omega_k}\widehat{\mathbf{Q}}'\mathbf{R}_2\widehat{\mathbf{Q}} + \frac{1}{\omega_k}\widehat{\mathbf{Q}}'\mathbf{R}_3\widehat{\mathbf{Q}} = o_P(1), \end{aligned} \tag{3.49}$$

where the last equality follows from the proof of Lemma 3.2.

Using the structure in Assumption (F1), we have

$$\begin{aligned} &\mathbb{E}\Big(\frac{1}{\omega_k}\mathbf{R}^*\Big) = \frac{1}{\omega_k T}\sum_{t=1}^{T}\mathbb{E}\Big\{\mathbf{A}_k\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}_{f,t-q}\Big)\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}_{f,t-q}'\Big)\mathbf{A}_k'\Big\} \\ &= \mathrm{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}) \cdot \frac{1}{\omega_k T}\sum_{t=1}^{T}\sum_{q\geq 0} a_{f,q}^2\mathbf{A}_k\mathbf{A}_k' = \mathrm{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}) \cdot \frac{1}{\omega_k}\mathbf{A}_k\mathbf{A}_k'. \end{aligned}$$

Meanwhile, we have

$$\begin{aligned} \mathrm{Var}\Big(\frac{\mathbf{R}_{ij}^*}{\omega_k}\Big) &= \frac{1}{\omega_k^2 T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\Big\{\mathbf{A}_{k,i\cdot}'\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}_{f,t-q}\Big)\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}_{f,t-q}'\Big)\mathbf{A}_{k,j\cdot}, \\ &\qquad\qquad \mathbf{A}_{k,i\cdot}'\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}_{f,s-q}\Big)\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}_{f,s-q}'\Big)\mathbf{A}_{k,j\cdot}\Big\} \\ &= \frac{1}{\omega_k^2 T^2}\sum_{t=1}^{T}\sum_{q\geq 0}\sum_{p\geq 0} a_{f,q}^2 a_{f,p}^2 \cdot \mathrm{Var}\Big(\mathbf{A}_{k,i\cdot}'\mathbf{X}_{f,t-q}\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}\mathbf{X}_{f,t-p}'\mathbf{A}_{k,j\cdot}\Big) \\ &= \frac{1}{\omega_k^2 T^2}\sum_{t=1}^{T}\sum_{q,p\geq 0} a_{f,q}^2 a_{f,p}^2\, O_P\big(\|\mathbf{A}_{\text{-}k}\|_F^2\big) = O_P\Big(T^{-1}d_k^{-2\alpha_{k,r_k}}\prod_{j\in[K]\setminus\{k\}} d_j^{-\alpha_{j,1}}\Big) = o_P(1), \end{aligned}$$

where we used Assumption (E2) in the third last equality, both (L1) and (F1) in the second last, and (R1) in the last. We can then conclude $\omega_k^{-1}\mathbf{R}^* \xrightarrow{p} \mathrm{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}) \cdot \omega_k^{-1}\mathbf{A}_k\mathbf{A}_k'$. Together with (3.49) and Assumption (L2), we obtain the limit of $\omega_k^{-1}\widehat{\mathbf{D}}$ as

$$\omega_k^{-1}\widehat{\mathbf{D}} \xrightarrow{p} \omega_k^{-1}\mathbf{D} = \omega_k^{-1}\mathrm{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}) \cdot \mathrm{diag}\{\lambda_j(\mathbf{A}_k'\mathbf{A}_k) \mid j \in [r_k]\}. \tag{3.50}$$

Define further

$$\mathbf{R}_{\text{res}} := \omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'((\widetilde{\mathbf{R}} - \mathbf{R}^*) + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3)\widehat{\mathbf{Q}}.$$

With similar arguments in the proof of Lemma 3.2, we have $\|\mathbf{R}_{\text{res}}\|_F = o_P(\|\mathbf{Z}_k^{1/2}\|_F)$. Left-multiply both sides of $\widehat{\mathbf{S}}\widehat{\mathbf{Q}} = \widehat{\mathbf{Q}}\widehat{\mathbf{D}}$ by $\omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'$, we can write

$$\begin{aligned}(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})(\omega_k^{-1}\widehat{\mathbf{D}}) &= \omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{S}}\widehat{\mathbf{Q}} = \omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'\mathbf{R}^*\widehat{\mathbf{Q}} + \mathbf{R}_{\text{res}} \\ &= \Big[\omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'\mathbf{R}^*\widehat{\mathbf{Q}}(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})^{-1} + \mathbf{R}_{\text{res}}(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})^{-1}\Big](\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}}).\end{aligned}$$

Hence, each column of $\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}}$ is an eigenvector of the matrix

$$\omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'\mathbf{R}^*\widehat{\mathbf{Q}}(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})^{-1} + \mathbf{R}_{\text{res}}(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})^{-1}.$$

We have $(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})'(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}}) \xrightarrow{p} (\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}))^{-1}\cdot\mathbf{D}$, since

$$\begin{aligned}&\omega_k^{-1}\Big\{(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})'(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}}) - (\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}))^{-1}\cdot\mathbf{D}\Big\} \\ &= \Big\{\frac{1}{\omega_k}\widehat{\mathbf{Q}}'\mathbf{A}'_k\mathbf{A}_k\widehat{\mathbf{Q}} - \frac{1}{\omega_k}(\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}))^{-1}\widehat{\mathbf{Q}}'\mathbf{R}^*\widehat{\mathbf{Q}}\Big\} \\ &\quad + \Big\{\frac{1}{\omega_k}(\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}))^{-1}\widehat{\mathbf{Q}}'\mathbf{R}^*\widehat{\mathbf{Q}} - \frac{1}{\omega_k}(\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}))^{-1}\,\mathbf{D}\Big\},\end{aligned}$$

which is $o_P(1)$ from the limit of $\omega_k^{-1}\mathbf{R}^*$ (for the first square bracket) and from (3.49) and (3.50) (for the second square bracket). Hence the eigenvalues of $(\mathbf{Q}'\widehat{\mathbf{Q}})'(\mathbf{Q}'\widehat{\mathbf{Q}})$ are asymptotically bounded away from zero and infinity by Assumption (L1), and also $\|(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})^{-1}\|_F = O_P(\|\mathbf{Z}_k^{-1/2}\|_F)$. Let

$$\Upsilon_k^* := (\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}))^{1/2}\cdot(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})\mathbf{D}^{-1/2}.$$

Using the limit of $\omega_k^{-1}\mathbf{R}^*$, we have

$$\begin{aligned}\omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'\mathbf{R}^*\widehat{\mathbf{Q}}(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})^{-1} &\xrightarrow{p} \text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k})\cdot\omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'\mathbf{Q}\mathbf{Z}_k\mathbf{Q}'\widehat{\mathbf{Q}}(\mathbf{Q}'\widehat{\mathbf{Q}})^{-1}\mathbf{Z}_k^{-1/2} \\ &= \text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k})\cdot\omega_k^{-1}\mathbf{Z}_k^{1/2}\mathbf{Q}'\mathbf{Q}\mathbf{Z}_k^{1/2} \\ &= \text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k})\cdot\omega_k^{-1}\mathbf{Z}_k^{1/2}\boldsymbol{\Sigma}_{A,k}\mathbf{Z}_k^{1/2},\end{aligned}$$

and $\|\mathbf{R}_{\text{res}}(\mathbf{Z}_k^{1/2}\mathbf{Q}'\widehat{\mathbf{Q}})^{-1}\|_F = o_P(1)$ from the above. By Assumption (L2) and eigenvector perturbation theories, there exists a unique eigenvector matrix $\Upsilon_k$ of $\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k})\cdot\omega_k^{-1}\mathbf{Z}_k^{1/2}\boldsymbol{\Sigma}_{A,k}\mathbf{Z}_k^{1/2}$ such that $\|\Upsilon_k - \Upsilon_k^*\| = o_P(1)$. Therefore, we have

$$\mathbf{Q}'\widehat{\mathbf{Q}} = (\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}))^{-1/2}\cdot\mathbf{Z}_k^{-1/2}\Upsilon_k^*\mathbf{D}^{1/2} \xrightarrow{p} (\text{tr}(\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}))^{-1/2}\cdot\mathbf{Z}_k^{-1/2}\Upsilon_k\mathbf{D}^{1/2}.$$

Thus, we have

$$\begin{aligned}\mathbf{H}^a &= \widehat{\mathbf{D}}^{-1}\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{Q}}'\mathbf{A}_k\mathbf{F}_t\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}\mathbf{F}_t'\mathbf{Z}_k^{1/2} = \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Q}}'\mathbf{R}^*\mathbf{A}_k(\mathbf{A}_k'\mathbf{A}_k)^{-1}\mathbf{Z}_k^{1/2} = \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Q}}'\mathbf{R}^*\mathbf{Q}\mathbf{\Sigma}_{A,k}^{-1} \\ &= \text{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k})\cdot\mathbf{D}^{-1}\widehat{\mathbf{Q}}'\mathbf{Q}\mathbf{Z}_k\mathbf{Q}'\mathbf{Q}\mathbf{\Sigma}_{A,k}^{-1} + o_P(1) \xrightarrow{p} (\text{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}))^{1/2}\cdot\mathbf{D}^{-1/2}\Upsilon_k'\mathbf{Z}_k^{1/2}.\end{aligned} \tag{3.51}$$

This completes the proof of Lemma 3.4. □

***Proof of Theorem 3.3.*** Suppose we focus on the $k$-th mode, and hence we adapt all notations by omitting the subscript $k$ for the ease of notational simplicity; see (3.19) for example. Moreover, we set $\mathbf{X}_{e,t} := \text{mat}_k(\mathcal{X}_{e,t}), \mathbf{X}_{\epsilon,t} := \text{mat}_k(\mathcal{X}_{\epsilon,t})$ and $\mathbf{X}_{f,t} := \text{mat}_k(\mathcal{X}_{f,t})$.

To proceed, we first decompose

$$\widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}^a\mathbf{Q}_{j\cdot} = (\widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}_j\mathbf{Q}_{j\cdot}) + (\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{j\cdot}. \tag{3.52}$$

Consider the first term $(\widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}_j\mathbf{Q}_{j\cdot})$. Using the decomposition in (3.26),

$$\begin{aligned}\widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}_j\mathbf{Q}_{j\cdot} &= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}(\widehat{\mathbf{Q}}_{i\cdot} - \mathbf{H}_j\mathbf{Q}_{i\cdot})\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})_{h\cdot}'\mathbf{F}_t'\mathbf{A}_{i\cdot} \\ &+ \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}(\widehat{\mathbf{Q}}_{i\cdot} - \mathbf{H}_j\mathbf{Q}_{i\cdot})\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,ih}(\mathbf{A}_{\text{-}k})_{h\cdot}'\mathbf{F}_t'\mathbf{A}_{j\cdot} \\ &+ \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}(\widehat{\mathbf{Q}}_{i\cdot} - \mathbf{H}_j\mathbf{Q}_{i\cdot})\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,ih}E_{t,jh} \\ &+ \mathcal{I}_{H,j} + \mathcal{I}_{H,j}^* + \mathcal{II}_{H,j} + \mathcal{III}_{H,j}, \quad \text{where}\end{aligned} \tag{3.53}$$

$$\begin{aligned}\mathcal{I}_{H,j} &:= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\mathbf{H}^a\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})_{h\cdot}'\mathbf{F}_t'\mathbf{A}_{i\cdot}, \\ \mathcal{I}_{H,j}^* &:= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}(\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})_{h\cdot}'\mathbf{F}_t'\mathbf{A}_{i\cdot}, \\ \mathcal{II}_{H,j} &:= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\mathbf{H}_j\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,ih}(\mathbf{A}_{\text{-}k})_{h\cdot}'\mathbf{F}_t'\mathbf{A}_{j\cdot}, \\ \mathcal{III}_{H,j} &:= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\mathbf{H}_j\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,ih}E_{t,jh}, \quad \text{with } \mathbf{A}_{\text{-}k} := \otimes_{l\in[K]\setminus\{k\}}\mathbf{A}_l.\end{aligned}$$

We want to show that $\mathcal{I}_{H,j}$ is the leading term among those in (3.53). To this end, we will show $\sqrt{T\omega_B} \cdot \mathcal{I}_{H,j}$ converges to a normal distribution with mean zero and variance of constant order (see (3.63) later), so that $\mathcal{I}_{H,j}$ is of order $(T\omega_B)^{-1/2}$ exactly. Then it suffices to show that the rate $(T\omega_B)^{-1}$ is dominating the following rates multiplied by the rate of $\|\widehat{\mathbf{D}}^{-1}\|_F^2 = O_P\big\{d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})} g_s^{-2}\big\}$ from Lemma 3.2:

$$\Big\| \sum_{i=1}^{d_k} (\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot} \Big\|^2, \tag{3.54}$$

$$\Big\| \sum_{i=1}^{d_k} \mathbf{H}_j\mathbf{Q}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{j\cdot} \Big\|^2, \tag{3.55}$$

$$\Big\| \sum_{i=1}^{d_k} \mathbf{H}_j\mathbf{Q}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih}E_{t,jh} \Big\|^2, \tag{3.56}$$

$$\Big\| \sum_{i=1}^{d_k} (\widehat{\mathbf{Q}}_{i\cdot} - \mathbf{H}_j\mathbf{Q}_{i\cdot}) \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot} \Big\|^2, \tag{3.57}$$

$$\Big\| \sum_{i=1}^{d_k} (\widehat{\mathbf{Q}}_{i\cdot} - \mathbf{H}_j\mathbf{Q}_{i\cdot}) \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{j\cdot} \Big\|^2, \tag{3.58}$$

$$\Big\| \sum_{i=1}^{d_k} (\widehat{\mathbf{Q}}_{i\cdot} - \mathbf{H}_j\mathbf{Q}_{i\cdot}) \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih}E_{t,jh} \Big\|^2. \tag{3.59}$$

Note that we can easily see the rates of (3.55) and (3.56) are greater than those of (3.58) and (3.59) respectively, using Lemma 3.2 and the Cauchy–Schwarz inequality.

Consider (3.54) first. We have

$$\begin{aligned}
&\Big\| \sum_{i=1}^{d_k} (\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot} \Big\|^2 \\
&\le \Big( \sum_{i=1}^{d_k} \|\mathbf{H}_j - \mathbf{H}^a\|_F^2 \cdot \|\mathbf{Q}_{i\cdot}\|^2 \Big) \cdot \sum_{i=1}^{d_k} \Big( \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot} \Big)^2 \\
&= O(d_{\text{-}k}^2) \cdot \|\mathbf{H}_j - \mathbf{H}^a\|_F^2 \cdot \sum_{i=1}^{d_k} \Big( \sum_{h=1}^{d_{\text{-}k}} \sum_{t\in\psi_{ij,h}} \frac{1}{d_{\text{-}k} \cdot |\psi_{ij,h}|} E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot} \Big)^2 \\
&= O(d_{\text{-}k}^2) \cdot \|\mathbf{H}_j - \mathbf{H}^a\|_F^2 \cdot \sum_{i=1}^{d_k} \|\mathbf{u}_i\|^2 \Big( \sum_{h=1}^{d_{\text{-}k}} \sum_{t\in\psi_{ij,h}} \frac{1}{d_{\text{-}k} \cdot |\psi_{ij,h}|} E_{t,jh}\mathbf{v}'_h\mathbf{F}'_t \frac{1}{\|\mathbf{u}_i\|}\mathbf{u}_i \Big)^2 \\
&= O_P\Big\{ d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})} \Big(\frac{d_{\text{-}k}}{T^2}\Big) d_k^{\alpha_{k,1}} \Big\}, \quad \text{with } \mathbf{v}_h := (\mathbf{A}_{\text{-}k})_{h\cdot},\ \mathbf{u}_i := \mathbf{A}_{i\cdot},
\end{aligned}$$

where we used Lemma 3.3, Proposition 3.2.2 and (L1) in the last equality.

To bound (3.55), note from Assumption (E1) and (E2) that we can write

$$E_{t,ih} = \sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-q}\mathbf{A}_{e,\text{-}k,h\cdot} + [\text{mat}_k(\mathbf{\Sigma}_\epsilon)]_{ih}\sum_{q\geq 0} a_{\epsilon,q}(\mathbf{X}_{\epsilon,t-q})_{ih}.$$

Consider first $\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\big(\sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-q}\mathbf{A}_{e,\text{-}k,h\cdot}\big)(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{j\cdot}$. By Assumption (O1), (E1), (E2) and (F1), we have

$$\begin{aligned}
&\mathbb{E}\Big\{\Big[\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\Big(\sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-q}\mathbf{A}_{e,\text{-}k,h\cdot}\Big)(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{j\cdot}\Big]^2\Big\}\\
&= \text{Cov}\Big\{\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}(\mathbf{A}_{\text{-}k})'_{h\cdot}\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}'_{f,t-q}\Big)\mathbf{A}_{j\cdot}\Big(\sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-q}\mathbf{A}_{e,\text{-}k,h\cdot}\Big),\\
&\qquad\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}(\mathbf{A}_{\text{-}k})'_{h\cdot}\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}'_{f,t-q}\Big)\mathbf{A}_{j\cdot}\Big(\sum_{q\geq 0} a_{e,q}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-q}\mathbf{A}_{e,\text{-}k,h\cdot}\Big)\Big\}\\
&= \sum_{h=1}^{d_{\text{-}k}}\sum_{l=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}\cap\psi_{ij,l}}\sum_{q\geq 0} a_{f,q}^2 a_{e,q}^2\cdot\|\mathbf{A}_{j\cdot}\|^2\,\|(\mathbf{A}_{\text{-}k})_{h\cdot}\|\,\|(\mathbf{A}_{\text{-}k})_{l\cdot}\|\,\|\mathbf{A}_{e,\text{-}k,h\cdot}\|\,\|\mathbf{A}_{e,\text{-}k,l\cdot}\|\,\|\mathbf{A}_{e,k,i\cdot}\|^2\\
&= O(T)\cdot\|\mathbf{A}_{j\cdot}\|^2\cdot\|\mathbf{A}_{e,k,i\cdot}\|^2.
\end{aligned}\tag{3.60}$$

Consider also $\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\mathbf{Q}_{i\cdot}\big([\text{mat}_k(\mathbf{\Sigma}_\epsilon)]_{ih}\sum_{q\geq 0} a_{\epsilon,q}(\mathbf{X}_{\epsilon,t-q})_{ih}\big)(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{j\cdot}$. Similarly, by Assumption (O1), (E1), (E2) and (F1), we have

$$\begin{aligned}
&\mathbb{E}\Big\{\Big\|\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\mathbf{Q}_{i\cdot}\Big([\text{mat}_k(\mathbf{\Sigma}_\epsilon)]_{ih}\sum_{q\geq 0} a_{\epsilon,q}(\mathbf{X}_{\epsilon,t-q})_{ih}\Big)(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{j\cdot}\Big\|^2\Big\}\\
&= \text{Cov}\Big\{\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\mathbf{Q}_{i\cdot}(\mathbf{A}_{\text{-}k})'_{h\cdot}\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}'_{f,t-q}\Big)\mathbf{A}_{j\cdot}\Big([\text{mat}_k(\mathbf{\Sigma}_\epsilon)]_{ih}\sum_{q\geq 0} a_{\epsilon,q}(\mathbf{X}_{\epsilon,t-q})_{ih}\Big),\\
&\qquad\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\mathbf{Q}_{i\cdot}(\mathbf{A}_{\text{-}k})'_{h\cdot}\Big(\sum_{q\geq 0} a_{f,q}\mathbf{X}'_{f,t-q}\Big)\mathbf{A}_{j\cdot}\Big([\text{mat}_k(\mathbf{\Sigma}_\epsilon)]_{ih}\sum_{q\geq 0} a_{\epsilon,q}(\mathbf{X}_{\epsilon,t-q})_{ih}\Big)\Big\}\\
&=\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\sum_{q\geq 0} a_{f,q}^2 a_{\epsilon,q}^2\Sigma_{\epsilon,k,h,ii}\cdot\|\mathbf{A}_{j\cdot}\|^2\,\|(\mathbf{A}_{\text{-}k})_{h\cdot}\|^2\,\|\mathbf{Q}_{i\cdot}\|^2 = O(T)\cdot\|\mathbf{A}_{j\cdot}\|^2\,\|\mathbf{A}_{\text{-}k}\|^2\,\|\mathbf{Q}\|^2.
\end{aligned}\tag{3.61}$$

Hence it holds that

$$
\begin{aligned}
&\Big\| \sum_{i=1}^{d_k} \mathbf{H}_j \mathbf{Q}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih} (\mathbf{A}_{\text{-}k})'_{h\cdot} \mathbf{F}'_t \mathbf{A}_{j\cdot} \Big\|^2 \\
&\le \|\mathbf{H}_j\|_F^2 \cdot \bigg\{ \Big\| \sum_{i=1}^{d_k} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \mathbf{Q}_{i\cdot} \Big\{ [\text{mat}_k(\mathbf{\Sigma}_\epsilon)]_{ih} \sum_{q\ge 0} a_{\epsilon,q} (\mathbf{X}_{\epsilon,t-q})_{ih} \Big\} (\mathbf{A}_{\text{-}k})'_{h\cdot} \mathbf{F}'_t \mathbf{A}_{j\cdot} \Big\|^2 \\
&\quad + \Big( \sum_{i=1}^{d_k} \|\mathbf{Q}_{i\cdot}\|^2 \Big) \sum_{i=1}^{d_k} \Big\| \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} \Big\{ \sum_{q\ge 0} a_{e,q} \mathbf{A}'_{e,k,i\cdot} \mathbf{X}_{e,t-q} \mathbf{A}_{e,\text{-}k,h\cdot} \Big\} (\mathbf{A}_{\text{-}k})'_{h\cdot} \mathbf{F}'_t \mathbf{A}_{j\cdot} \Big\|^2 \bigg\} \\
&= O_P\Big\{ d_k^{-\alpha_{k,1}} \Big(\frac{1}{T}\Big) \prod_{j=1}^K d_j^{\alpha_{j,1}} \Big\},
\end{aligned}
$$

where we used Assumption (L1), (3.60) and (3.61) in the last equality.

For (3.56), by Assumption (O1), (E1), (E2), and the proof of Proposition 3.2 we have

$$
\begin{aligned}
&\text{Var}\Big( \sum_{i=1}^{d_k} \sum_{h=1}^{d_{\text{-}k}} \sum_{t\in\psi_{ij,h}} \mathbf{Q}_{i\cdot} E_{t,ih} E_{t,jh} \Big) \\
&= O\Big( \sum_{i,w=1}^{d_k} \sum_{h,l=1}^{d_{\text{-}k}} \sum_{t\in\psi_{ij,h}\cap\psi_{wj,l}} \sum_{n=1}^{r_{e,k}} \sum_{m=1}^{r_{e,\text{-}k}} \sum_{q\ge 0} a_{e,q}^4 A_{e,k,in} A_{e,k,wn} A_{e,k,jn}^2 A_{e,\text{-}k,hm}^2 A_{e,\text{-}k,lm}^2 \\
&\quad \cdot \|\mathbf{Q}_{i\cdot}\| \cdot \|\mathbf{Q}_{w\cdot}\| \cdot \text{Var}((\mathbf{X}_{e,t-q})_{nm}^2) \Big) + O\Big( \sum_{i=1}^{d_k} \sum_{h=1}^{d_{\text{-}k}} \sum_{t\in\psi_{ij,h}\cap\psi_{wj,l}} \sum_{q\ge 0} a_{\epsilon,q}^4 \Sigma_{\epsilon,k,h,ii} \Sigma_{\epsilon,k,h,jj} \\
&\quad \cdot \|\mathbf{Q}_{i\cdot}\|^2 \cdot \text{Var}((\mathbf{X}_{\epsilon,t-q})_{ih} (\mathbf{X}_{\epsilon,t-q})_{jh}) \Big) = O(T + T d_{\text{-}k}) = O(T d_{\text{-}k}).
\end{aligned}
$$

Moreover, it holds that

$$
\mathbb{E}\Big( \sum_{i=1}^{d_k} \sum_{h=1}^{d_{\text{-}k}} \sum_{t\in\psi_{ij,h}} E_{t,ih} E_{t,jh} \Big) = \sum_{i=1}^{d_k} \sum_{h=1}^{d_{\text{-}k}} \sum_{t\in\psi_{ij,h}} \Big( \|\mathbf{A}_{e,\text{-}k,h\cdot}\|^2 \|\mathbf{A}_{e,k,i\cdot}\| \|\mathbf{A}_{e,k,j\cdot}\| + \Sigma_{\epsilon,k,h,ij} \Big) = O(T d_{\text{-}k}),
$$

and with $\max_i \|\mathbf{Q}_{i\cdot}\|^2 \le \|\mathbf{A}_{j\cdot}\|^2 \cdot \|\mathbf{Z}_k^{-1/2}\|^2 = O_P\big(d_k^{-\alpha_{k,r_k}}\big)$, we thus have

$$
\begin{aligned}
\Big\| \sum_{i=1}^{d_k} \mathbf{H}_j \mathbf{Q}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih} E_{t,jh} \Big\|^2 &\le \|\mathbf{H}_j\|_F^2 \cdot \Big\| \sum_{i=1}^{d_k} \mathbf{Q}_{i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{ij,h}|} \sum_{t\in\psi_{ij,h}} E_{t,ih} E_{t,jh} \Big\|^2 \\
&= O\Big( \frac{d_{\text{-}k}}{T} + d_{\text{-}k}^2 d_k^{-\alpha_{k,r_k}} \Big).
\end{aligned}
$$

Now consider (3.57). Similar to (3.54), we have

$$\begin{aligned}
&\Big\|\sum_{i=1}^{d_k}(\widehat{\mathbf{Q}}_{i\cdot}-\mathbf{H}_j\mathbf{Q}_{i\cdot})\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\Big\|^2\\
&\le\Big(\sum_{i=1}^{d_k}\|\widehat{\mathbf{Q}}_{i\cdot}-\mathbf{H}_j\mathbf{Q}_{i\cdot}\|_F^2\Big)\cdot\sum_{i=1}^{d_k}\Big(\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\Big)^2\\
&=O(d_{\text{-}k}^2)\cdot\Big(\sum_{i=1}^{d_k}\|\widehat{\mathbf{Q}}_{i\cdot}-\mathbf{H}_j\mathbf{Q}_{i\cdot}\|_F^2\Big)\cdot\sum_{i=1}^{d_k}\|\mathbf{u}_i\|^2\Big(\sum_{h=1}^{d_{\text{-}k}}\sum_{t\in\psi_{ij,h}}\frac{1}{d_{\text{-}k}\cdot|\psi_{ij,h}|}E_{t,jh}\mathbf{v}'_h\mathbf{F}'_t\frac{1}{\|\mathbf{u}_i\|}\mathbf{u}_i\Big)^2\\
&=O_P\Big\{d_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}\Big(\frac{d_{\text{-}k}}{T}\Big)\Big(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\Big)\prod_{j=1}^{K}d_j^{2(1-\alpha_{j,1})}\Big\},
\end{aligned}$$

where we used Lemma 3.2, Proposition 3.2.2 and (L1) in the last equality.

Finally, we consider the following ratios with $d_1,\dots,d_K,T\to\infty$:

$$\begin{aligned}
&\Big\|\sum_{i=1}^{d_k}(\mathbf{H}_j-\mathbf{H}^a)\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\Big\|^2 d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}g_s^{-2}\Big/(T\omega_B)^{-1}\\
&=O_P\Big(d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}\cdot\frac{1}{T}\Big)=o_P(1),\\
&\Big\|\sum_{i=1}^{d_k}\mathbf{H}_j\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,ih}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{j\cdot}\Big\|^2 d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}g_s^{-2}\Big/(T\omega_B)^{-1}\\
&=O_P\Big(d_k^{-\alpha_{k,1}}\prod_{j\in[K]\setminus\{k\}}d_j^{\alpha_{j,1}-1}\Big)=o_P(1),\\
&\Big\|\sum_{i=1}^{d_k}\mathbf{H}_j\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,ih}E_{t,jh}\Big\|^2 d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}g_s^{-2}\Big/(T\omega_B)^{-1}\\
&=O_P\Big(Td_{\text{-}k}d_k^{-\alpha_{k,r_k}-\alpha_{k,1}}\Big)=o_P(1),\\
&\Big\|\sum_{i=1}^{d_k}(\widehat{\mathbf{Q}}_{i\cdot}-\mathbf{H}_j\mathbf{Q}_{i\cdot})\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\Big\|^2 d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}g_s^{-2}\Big/(T\omega_B)^{-1}\\
&=O_P\Big\{d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}\Big(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\Big)\prod_{j=1}^{K}d_j^{2(1-\alpha_{j,1})}\Big\}=o_P(1),
\end{aligned}$$

by Assumptions (R1) and the rate assumptions $d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}}=o(d_{\text{-}k})$ (from the statement of Theorem 3.2) and $Td_{\text{-}k}=o(d_k^{\alpha_{k,1}+\alpha_{k,r_k}})$. Hence $\mathcal{I}_{H,j}$ is indeed the dominating term in (3.53). In other words, we have

$$\widehat{\mathbf{Q}}_{j\cdot}-\mathbf{H}_j\mathbf{Q}_{j\cdot}=\mathcal{I}_{H,j}+o_P(1). \tag{3.62}$$

Then we want to show that

$$\begin{aligned}\sqrt{T\omega_B}\cdot\mathcal{I}_{H,j} &= \sqrt{T\omega_B}\cdot\widehat{\mathbf{D}}^{-1}\mathbf{H}^a\sum_{i=1}^{d_k}\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\\ &\xrightarrow{p}\sqrt{T\omega_B}\cdot\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\sum_{i=1}^{d_k}\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\\ &\xrightarrow{\mathcal{D}}\mathcal{N}(\mathbf{0},T\omega_B\cdot\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\boldsymbol{\Xi}_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}),\end{aligned}\tag{3.63}$$

where $\mathbf{D}$ and $\mathbf{H}_k^{a,*}$ are from Lemma 3.4, and we require Assumption (AD1) for the covariance matrix to have constant rate. In fact, using Lemma 3.4 and Proposition 3.2, the upper bound is of constant order by

$$\begin{aligned}&\Big\|T\omega_B\cdot\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\boldsymbol{\Xi}_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}\Big\|_F\\ &=O\big(T\omega_B\big)\cdot\big\|\mathbf{D}^{-1}\big\|_F^2\cdot\Bigg\|\sum_{i=1}^{d_k}\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\Bigg\|^2\\ &=O\Big\{\frac{T}{d_{\text{-}k}d_k^{\alpha_{k,1}}}\cdot\Big(\sum_{i=1}^{d_k}\|\mathbf{Q}_{i\cdot}\|^2\Big)\cdot\sum_{i=1}^{d_k}\Big(\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\Big)^2\Big\}=O_P(1).\end{aligned}$$

We will adapt the central limit theorem for $\alpha$-mixing processes (Fan and Yao (2003), Theorem 2.21). Due to the existence of missing data and the general missing patterns that we allow, we construct an auxiliary time series to facilitate the proof. Formally, define $\{\mathbf{B}_{j,t}\}_{t\in[T]}$ as

$$\mathbf{B}_{j,t}:=\sqrt{\omega_B}\cdot\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\mathbf{Q}_{i\cdot}\frac{T}{|\psi_{ij,h}|}\cdot E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\cdot\mathbb{1}\{t\in\psi_{ij,h}\}.$$

Hence we have the following,

$$\sqrt{T\omega_B}\cdot\mathcal{I}_{H,j}\xrightarrow{p}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{B}_{j,t}.$$

It is easy to see that $\mathbb{E}[\mathbf{B}_{j,t}]=\mathbf{0}$ by Assumption (E1), (E2) and (F1). Moreover, $\mathbf{B}_{j,t}$ is also $\alpha$-mixing over $t$. To see this, consider

$$\begin{aligned}&E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\\ &=\Big(\sum_{q\geq 0}a_{e,q}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-q}\mathbf{A}_{e,\text{-}k,h\cdot}+\Sigma_{\epsilon,k,h,ii}^{1/2}\sum_{q\geq 0}a_{\epsilon,q}(\mathbf{X}_{\epsilon,t-q})_{ih}\Big)\mathbf{A}'_{\text{-}k,h\cdot}\Big(\sum_{q\geq 0}a_{f,q}\mathbf{X}'_{f,t-q}\Big)\mathbf{A}_{i\cdot}.\end{aligned}$$

Define $\mathbf{b}_{e,t}:=\sum_{q\geq 0}a_{e,q}\mathbf{X}_{e,t-q}$, $\mathbf{b}_{\epsilon,ih,t}:=\sum_{q\geq 0}a_{\epsilon,q}(\mathbf{X}_{\epsilon,t-q})_{ih}$ and $\mathbf{b}_{f,t}:=\sum_{q\geq 0}a_{f,q}\mathbf{X}_{f,t-q}$

which are independent of each other by Assumption (E2), we can then rewrite

$$\mathbf{B}_{j,t} = h\Big(\mathbf{b}_{e,t}, (\mathbf{b}_{\epsilon,ih,t})_{i\in[d_k],h\in[d_{\text{-}k}]}, \mathbf{b}_{f,t}\Big),$$

for some function $h$, and hence Theorem 5.2 in Bradley (2005) implies the $\alpha$-mixing property. Then similar to Chen and Fan (2023), it is left to show that there exists an $m > 2$ such that $\mathbb{E}[\|\mathbf{B}_{j,t}\|^m] \le C$ for some constant $C$. With Assumption (E1), (E2) and (F1) and similar to the proof of Proposition 3.2, we have

$$\mathbb{E}\Big(\sum_{h=1}^{d_{\text{-}k}} \frac{T}{|\psi_{ij,h}|} \cdot E_{t,jh}\mathbf{u}'\mathbf{F}_t\mathbf{v} \cdot \mathbb{1}\{t\in\psi_{ij,h}\}\Big)^2 = O(d_{\text{-}k}),$$

where $\mathbf{u} \in \mathbb{R}^{r_k}$ and $\mathbf{v} \in \mathbb{R}^{r_{\text{-}k}}$ are any deterministic vectors of constant order. Hence

$$\begin{aligned}
&\mathbb{E}\big(\|\mathbf{B}_{j,t}\|^m\big)\\
&\le \omega_B^{m/2}\,\|\mathbf{D}^{-1}\|_F^m\,\|\mathbf{H}_k^{a,*}\|_F^m\,\Big(\sum_{i=1}^{d_k}\|\mathbf{Q}_{i\cdot}\|^2\Big)^{m/2}\\
&\cdot\,\mathbb{E}\left(\Big\{\sum_{i=1}^{d_k}\Big(\sum_{h=1}^{d_{\text{-}k}}\frac{T}{|\psi_{ij,h}|}\,E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\cdot\mathbb{1}\{t\in\psi_{ij,h}\}\Big)^2\Big\}^{m/2}\right)\\
&= O\Big\{\Big(\omega_B d_{\text{-}k} d_k^{\alpha_{k,1}}\Big)^{m/2}\Big\}\cdot\|\mathbf{D}^{-1}\|_F^m = O_P\Big\{\Big(\omega_B d_{\text{-}k} d_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}\prod_{j=1}^{K} d_j^{-2\alpha_{j,1}}\Big)^{m/2}\Big\} = O_P(1),
\end{aligned}$$

where we used Lemma 3.2 and the definition of $\omega_B$ in the last line. Theorem 2.21 in Fan and Yao (2003) then applies. With (3.62), (3.63) and Lemma 3.4, we can directly establish that

$$\sqrt{T\omega_B}\cdot(\widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}_j\mathbf{Q}_{j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, T\omega_B\cdot\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\boldsymbol{\Xi}_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}\big). \tag{3.64}$$

Consider now the second term in (3.52). By Lemma 3.3 and 3.4, we have

$$\big\|(\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{j\cdot}\big\|^2 \le \|\mathbf{H}_j - \mathbf{H}^a\|_F^2\cdot\|\mathbf{A}_{j\cdot}\|^2\cdot\|\mathbf{Z}_k^{-1/2}\|^2 = O_P\Big\{\min\Big(\frac{1}{T}, \frac{\eta^2}{(1-\eta)^2}\Big)d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}}\Big\},$$

implying

$$\begin{aligned}
&\big\|(\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{j\cdot}\big\|^2\Big/\Big\|\widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\mathbf{H}^a\mathbf{Q}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}E_{t,jh}(\mathbf{A}_{\text{-}k})'_{h\cdot}\mathbf{F}'_t\mathbf{A}_{i\cdot}\Big\|^2\\
&= O_P\Big\{\min\Big(1, \frac{T\eta^2}{(1-\eta)^2}\Big)d_k^{3(\alpha_{k,1}-\alpha_{k,r_k})}\prod_{j\in[K]\setminus\{k\}} d_j^{2\alpha_{j,1}-1}\Big\},
\end{aligned}$$

which is unrealistic to be $o_P(1)$ in the presence of missing data in general. Thus $(\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{j\cdot}$ contributes to the asymptotic distribution of $(\widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}^a\mathbf{Q}_{j\cdot})$. Rewrite

$$
\begin{aligned}
&(\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{j\cdot} \\
&= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\sum_{h=1}^{d_{\text{-}k}}\Big(\frac{1}{|\psi_{ij,h}|}\sum_{t\in\psi_{ij,h}}\mathbf{Q}_{i\cdot}'\mathbf{F}_{Z,t}\mathbf{\Lambda}_{h\cdot}\mathbf{\Lambda}_{h\cdot}'\mathbf{F}_{Z,t}' - \frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}_{i\cdot}'\mathbf{F}_{Z,t}\mathbf{\Lambda}_{h\cdot}\mathbf{\Lambda}_{h\cdot}'\mathbf{F}_{Z,t}'\Big)\mathbf{Q}_{j\cdot} \\
&= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{i\cdot}\mathbf{A}_{i\cdot}'\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\mathbf{Z}_k^{1/2}\mathbf{Q}_{j\cdot} \\
&= \widehat{\mathbf{D}}^{-1}\sum_{i=1}^{d_k}(\widehat{\mathbf{Q}}_{i\cdot} - \mathbf{H}^a\mathbf{Q}_{i\cdot})\mathbf{A}_{i\cdot}'\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\mathbf{Z}_k^{1/2}\mathbf{Q}_{j\cdot} + \widehat{\mathbf{D}}^{-1}\mathbf{H}^a\sum_{i=1}^{d_k}\mathbf{Q}_{i\cdot}\mathbf{A}_{i\cdot}'\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\mathbf{Z}_k^{1/2}\mathbf{Q}_{j\cdot}.
\end{aligned}
$$

Note the first term is dominated by the second term due to Theorem 3.1. Using Assumption (AD2) and the Slutsky's theorem, we have

$$
\begin{aligned}
&\sqrt{Td_k^{\alpha_{k,r_k}}}\cdot\widehat{\mathbf{D}}^{-1}\mathbf{H}^a\sum_{i=1}^{d_k}\mathbf{Q}_{i\cdot}\mathbf{A}_{i\cdot}'\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\mathbf{Z}_k^{1/2}\mathbf{Q}_{j\cdot} \\
&\to\mathcal{N}\big(\mathbf{0},\mathbf{D}^{-1}\mathbf{H}_k^{a,*}h_{k,j}(\mathbf{A}_{j\cdot})(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}\big)\quad\mathcal{G}^T\text{-stably.}
\end{aligned}
\tag{3.65}
$$

Furthermore, $\mathcal{I}_{H,j}$ and $(\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{j\cdot}$ are asymptotically independent since the randomness of $\mathcal{I}_{H,j}$ comes from $E_{t,jh}(\mathbf{A}_{\text{-}k})_{h\cdot}'\mathbf{F}_t'$ while that of $(\mathbf{H}_j - \mathbf{H}^a)\mathbf{Q}_{j\cdot}$ comes from $\Delta_{F,k,ij,h}$. From (3.64) and (3.65), we conclude that

$$
\sqrt{Td_k^{\alpha_{k,r_k}}}\cdot(\widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}^a\mathbf{Q}_{j\cdot})\xrightarrow{\mathcal{D}}\mathcal{N}\big(\mathbf{0},\mathbf{D}^{-1}\mathbf{H}_k^{a,*}(Td_k^{\alpha_{k,r_k}}\cdot\mathbf{\Xi}_{k,j} + h_{k,j}(\mathbf{A}_{j\cdot}))(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}\big).
$$

On the other hand, if we have finite missingness or asymptotically vanishing missingness such that

$$
\min\Big\{1,\frac{T\eta^2}{(1-\eta)^2}\Big\}\cdot d_k^{3(\alpha_{k,1}-\alpha_{k,r_k})}\prod_{j\in[K]\setminus\{k\}}d_j^{2\alpha_{j,1}-1} = Td^{-1}g_s^2g_\eta d_k^{1+\alpha_{k,1}-3\alpha_{k,r_k}} = o(1),
$$

then (3.65) is dominated by (3.64), and hence it holds at the absence of (AD2) that

$$
\sqrt{T\omega_B}\cdot(\widehat{\mathbf{Q}}_{j\cdot} - \mathbf{H}^a\mathbf{Q}_{j\cdot})\xrightarrow{\mathcal{D}}\mathcal{N}(\mathbf{0},T\omega_B\cdot\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\mathbf{\Xi}_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}).
$$

This completes the proof of Theorem 3.3. □

***Proof of Theorem 3.4.*** By Lemma 3.4, $\widehat{\mathbf{D}}_k$ is consistent for $\mathbf{D}_k$, and $\mathbf{H}_k^a$ is consistent for $\mathbf{H}_k^{a,*}$. Similar to the proof of Theorem 5 in Chen and Fan (2023), it suffice to prove that the HAC estimator $\widehat{\mathbf{\Sigma}}_{HAC}$ based on $\{\widehat{\mathbf{Q}}_k,\text{mat}_k(\widehat{\mathcal{C}}_t),\text{mat}_k(\widehat{\mathcal{E}}_t)\}_{t\in[T]}$ is a consistent estimator for

$\mathbf{H}_k^a \Xi_{k,j} (\mathbf{H}_k^a)'$. Recall that

$$\begin{aligned}
&\mathbf{H}_k^a \Xi_{k,j} (\mathbf{H}_k^a)' \\
&= \mathrm{Var}\Big( \sum_{i=1}^{d_k} \mathbf{H}_k^a \mathbf{Q}_{k,i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{k,ij,h}|} \sum_{t \in \psi_{k,ij,h}} E_{t,jh} (\mathbf{A}_{\text{-}k})'_{h\cdot} \mathbf{F}'_t \mathbf{A}_{k,i\cdot} \Big) \\
&= \mathrm{Var}\Big\{ \sum_{i=1}^{d_k} \Big( \widehat{\mathbf{D}}_k^{-1} \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{Q}}_k' \mathbf{Q}_k \mathbf{F}_{Z,t} \mathbf{\Lambda}' \mathbf{\Lambda} \mathbf{F}'_{Z,t} \Big) \mathbf{Q}_{k,i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{k,ij,h}|} \sum_{t \in \psi_{k,ij,h}} E_{t,jh} (\mathbf{A}_{\text{-}k})'_{h\cdot} \mathbf{F}'_t \mathbf{A}_{k,i\cdot} \Big\} \\
&= \mathrm{Var}\Big\{ \sum_{i=1}^{d_k} \Big( \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{D}}_k^{-1} \widehat{\mathbf{Q}}_k' \cdot \mathrm{mat}_k(\mathcal{C}_t)\, \mathrm{mat}_k(\mathcal{C}_t)_{i\cdot} \Big) \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{k,ij,h}|} \sum_{t \in \psi_{k,ij,h}} \mathrm{mat}_k(\mathcal{E}_t)_{jh}\, \mathrm{mat}_k(\mathcal{C}_t)_{ih} \Big\}.
\end{aligned}$$

By Theorem 3.1, and the rate assumption in the statement of Theorem 3.4, we have $\widehat{\mathbf{Q}}_k$ being consistent for a version of $\mathbf{Q}_k$ (in Frobenius norm) for any $k \in [K]$. By Theorem 3.2 and the assumption that the rate for individual common component imputation error is going to 0, $\widehat{\mathcal{C}}_{t,i_1,\ldots,i_K}$ is consistent for $\mathcal{C}_{t,i_1,\ldots,i_K}$ for any $k \in [K], i_k \in [d_k], t \in [T]$. Hence, it also holds that $\widehat{\mathcal{E}}_{t,i_1,\ldots,i_K}$ is consistent for $\mathcal{E}_{t,i_1,\ldots,i_K}$ for any $k \in [K], i_k \in [d_k], t \in [T]$. We can finally conclude that $\widehat{\mathbf{\Sigma}}_{HAC}$ is estimating $\mathbf{H}_k^a \Xi_{k,j} (\mathbf{H}_k^a)'$ consistently (Newey and West, 1987), which is result 1. We can also show a similar result for $\widehat{\mathbf{\Sigma}}_{HAC}^{\Delta}$, which is result 2 (details omitted). Combining both results, and consider the general statement of Theorem 3.3, we can easily conclude result 3. This completes the proof of the theorem. □

We will present two other lemmas before proving Theorem 3.2. While we stick with the notations in (3.19), we use the following also hereafter:

$$\begin{aligned}
&\mathbf{y}_t := \mathbf{vec}\big(\mathcal{Y}_t\big),\ \mathbf{m}_t := \mathbf{vec}\big(\mathcal{M}_t\big),\ \mathbf{f}_{Z,t} := \mathbf{vec}\big(\mathcal{F}_{Z,t}\big),\ \boldsymbol{\varepsilon}_t := \mathbf{vec}\big(\mathcal{E}_t\big), \mathbf{c}_t := \mathbf{vec}\big(\mathcal{C}_t\big), \\
&\mathbf{f}_t := \mathbf{vec}\big(\mathcal{F}_t\big),\ \mathbf{H}_{\otimes} := \mathbf{H}_K^a \otimes \cdots \otimes \mathbf{H}_1^a,\ \mathbf{A}_{\otimes} := \mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_1,\ \mathbf{Z}_{\otimes} := \mathbf{Z}_K \otimes \cdots \otimes \mathbf{Z}_1,
\end{aligned} \tag{3.66}$$

where the hat versions (if any) of the above are defined similarly.

**Lemma 3.5** *Under the assumptions in Theorem 3.2, for any $k \in [K]$ and $j \in [d_k]$,*

$$\big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{k,j\cdot}\big\|_F^2 = O_P\Big( T^{-1} d_{\text{-}k} d_k^{3\alpha_{k,1} - 2\alpha_{k,r_k}} g_s^{-2} + d^2 g_s^{-2} d_k^{2\alpha_{k,1} - 3\alpha_{k,r_k} - 2} + g_\eta d_k^{2\alpha_{k,1} - 3\alpha_{k,r_k}} \Big). \tag{3.67}$$

***Proof of Lemma 3.5.*** First, consider the case when $Td_{\text{-}k} = o\big(d_k^{\alpha_{k,r_k} + \alpha_{k,1}}\big)$. From (3.64) in the proof of Theorem 3.3, we have $\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_{k,j} \mathbf{Q}_{k,j\cdot}\|_F^2 = O_P(T^{-1} \omega_B^{-1})$. Thus,

$$\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{k,j\cdot}\|_F^2 = O_P\Big( \|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_{k,j} \mathbf{Q}_{k,j\cdot}\|_F^2 + \|(\mathbf{H}_{k,j} - \mathbf{H}_k^a) \mathbf{Q}_{k,j\cdot}\|_F^2 \Big)$$

$$= O_P\Big((T\omega_B)^{-1} + g_\eta d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}}\Big) = O_P\Big(T^{-1}d_{\text{-}k}d_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-2} + g_\eta d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}}\Big), \tag{3.68}$$

where we used Lemma 3.3 in the second equality, and

$$\big\|\mathbf{Q}_{k,j\cdot}\big\|^2 = \|\mathbf{Z}_k^{-1/2}\mathbf{A}_{k,j\cdot}\|^2 = O_P(d_k^{-\alpha_{k,r_k}}).$$

Now suppose $Td_{\text{-}k} = o\big(d_k^{\alpha_{k,r_k}+\alpha_{k,1}}\big)$ fails to hold. From the decomposition of $\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_{k,j}\mathbf{Q}_{k,j\cdot}$ in (3.53), $\mathcal{I}_{H,j}$ is not the leading term anymore, and the leading term among the expressions from (3.54) to (3.59) will be (3.56). It has rate

$$O_P\Big(\frac{d_{\text{-}k}}{T} + d_{\text{-}k}^2 d_k^{-\alpha_{k,r_k}}\Big) = O_P\big(d_{\text{-}k}^2 d_k^{-\alpha_{k,r_k}}\big),$$

where the above equality used the fact that $Td_{\text{-}k} = o\big(d_k^{\alpha_{k,r_k}+\alpha_{k,1}}\big)$ does not hold. Together with the bound on $\|\widehat{\mathbf{D}}_k^{-1}\|_F$ from Lemma 3.2 , we have

$$\big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_{k,j}\mathbf{Q}_{k,j\cdot}\big\|_F^2 = O_P\Big(d^2 g_s^{-2} d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-2}\Big). \tag{3.69}$$

Combining (3.68) and (3.69), we arrive at the statement of the lemma. □

**Lemma 3.6** *Under the Assumptions in Theorem 3.2, with the notations in* (3.19) *and* (3.66)*, we have the following for any* $j \in [d]$*:*

$$\big\|\mathbf{Q}_\otimes \mathbf{H}_\otimes'\big\|_F^2 = O_P(1), \tag{3.70}$$

$$\big\|\widehat{\mathbf{Q}}_{\otimes,j\cdot} - \mathbf{H}_\otimes \mathbf{Q}_{\otimes,j\cdot}\big\|^2 = O_P\Big\{\max_{k\in[K]}\Big(T^{-1}d_{\text{-}k}d_k^{3\alpha_{k,1}-\alpha_{k,r_k}}g_s^{-2}g_w^{-1} + d^2 g_s^{-2}g_w^{-1}d_k^{2\alpha_{k,1}-2\alpha_{k,r_k}-2} + g_\eta g_w^{-1}d_k^{2\alpha_{k,1}-2\alpha_{k,r_k}}\Big)\Big\}, \tag{3.71}$$

$$\big\|\widehat{\mathbf{Q}}_\otimes - \mathbf{Q}_\otimes \mathbf{H}_\otimes'\big\|_F^2 = O_P\Big\{\max_{k\in[K]}\Big(T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-2} + d^2 g_s^{-2}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1} + g_\eta d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big)\Big\}. \tag{3.72}$$

***Proof of Lemma 3.6.*** For (3.70), with Assumption (L1) we have

$$\big\|\mathbf{Q}_\otimes \mathbf{H}_\otimes'\big\|_F^2 \le \big\|\mathbf{H}_\otimes\big\|_F^2 \cdot \prod_{k=1}^K \big\|\mathbf{Q}_k\big\|_F^2 = O_P(1).$$

To show (3.71), for any $j \in [d_k]$, by a simple induction argument (omitted),

$$\begin{aligned}
&\big\|\widehat{\mathbf{Q}}_{\otimes,j\cdot} - \mathbf{H}_\otimes \mathbf{Q}_{\otimes,j\cdot}\big\|^2 = \big\|(\widehat{\mathbf{Q}}_\otimes - \mathbf{Q}_\otimes \mathbf{H}_\otimes')_{j\cdot}\big\|^2 \\
&= \Big\|\Big\{(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1) - (\mathbf{Q}_K \mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1 \mathbf{H}_1')\Big\}_{j\cdot}\Big\|^2 \\
&\leq \sum_{k=1}^K \Big(\max_{j\in[d_k]} \big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{k,j\cdot}\big\|^2 \prod_{\ell\in[K]\setminus\{k\}} \max_{j\in[d_\ell]} \big\|\widehat{\mathbf{Q}}_{\ell,j\cdot}\big\|^2\Big) \\
&= O_P\Big(\max_{k\in[K]}\Big\{\Big(T^{-1}d_{\text{-}k}d_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-2} + d^2 g_s^{-2} d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-2} + g_\eta d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}}\Big) \prod_{\ell\in[K]\setminus\{k\}} d_\ell^{-\alpha_{\ell,r_\ell}}\Big\}\Big) \\
&= O_P\Big\{\max_{k\in[K]}\Big(T^{-1}d_{\text{-}k}d_k^{3\alpha_{k,1}-\alpha_{k,r_k}} g_s^{-2} g_w^{-1} + d^2 g_s^{-2} g_w^{-1} d_k^{2\alpha_{k,1}-2\alpha_{k,r_k}-2} + g_\eta g_w^{-1} d_k^{2\alpha_{k,1}-2\alpha_{k,r_k}}\Big)\Big\},
\end{aligned}$$

where the second last equality used (3.67) and

$$\begin{aligned}
&\big\|\widehat{\mathbf{Q}}_{\ell,j\cdot}\big\|^2 \leq 2\Big(\big\|\widehat{\mathbf{Q}}_{\ell,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{\ell,j\cdot}\big\|^2 + \big\|\mathbf{H}_\ell^a \mathbf{Q}_{\ell,j\cdot}\big\|^2\Big) \\
&= O_P\Big(\big\|\widehat{\mathbf{Q}}_{\ell,j\cdot} - \mathbf{H}_\ell^a \mathbf{Q}_{\ell,j\cdot}\big\|^2 + \big\|\mathbf{H}_\ell^a \mathbf{Z}_\ell^{-1/2} \mathbf{A}_{\ell,j\cdot}\big\|^2\Big) \\
&= O_P\Big\{\big\|\widehat{\mathbf{Q}}_{\ell,j\cdot} - \mathbf{H}_\ell^a \mathbf{Q}_{\ell,j\cdot}\big\|^2 + \big\|\mathbf{H}_\ell^a\big\|_F^2 \cdot \Big(d_\ell^{-\alpha_{\ell,r_\ell}/2}\Big)^2 \cdot 1\Big\} = O_P\big(d_\ell^{-\alpha_{\ell,r_\ell}}\big).
\end{aligned}$$

Finally it also holds that

$$\begin{aligned}
&\big\|\widehat{\mathbf{Q}}_\otimes - \mathbf{Q}_\otimes \mathbf{H}_\otimes'\big\|_F^2 = \big\|(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1) - (\mathbf{Q}_K \mathbf{H}_K^{a'} \otimes \cdots \otimes \mathbf{Q}_1 \mathbf{H}_1^{a'})\big\|_F^2 \\
&= O(1) \cdot \sum_{k=1}^K \big\|\widehat{\mathbf{Q}}_k - \mathbf{Q}_k \mathbf{H}_k^{a'}\big\|_F^2 = O\Big(\max_{k\in[K]} \sum_{j=1}^{d_k} \big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{k,j\cdot}\big\|^2\Big) \\
&= O_P\Big\{\max_{k\in[K]}\Big(T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}} g_s^{-2} + d^2 g_s^{-2} d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1} + g_\eta d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big)\Big\},
\end{aligned}$$

where the second equality could be shown by a simple induction argument using $\|\mathbf{Q}_k\| = O(1)$ (omitted), and the last equality is from (3.67). □

***Proof of Theorem 3.2.*** The equation (3.7) is essentially

$$\begin{aligned}
\widehat{\mathbf{f}}_{Z,t} &= \Big(\sum_{j=1}^d m_{t,j} \widehat{\mathbf{Q}}_{\otimes,j\cdot} \widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1} \Big(\sum_{j=1}^d m_{t,j} \widehat{\mathbf{Q}}_{\otimes,j\cdot} y_{t,j}\Big) \\
&= \Big(\sum_{j=1}^d m_{t,j} \widehat{\mathbf{Q}}_{\otimes,j\cdot} \widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1} \Big(\sum_{j=1}^d m_{t,j} \widehat{\mathbf{Q}}_{\otimes,j\cdot} (\mathbf{Q}_{\otimes,j\cdot}' \mathbf{f}_{Z,t} + \varepsilon_{t,j})\Big) \\
&= \Big(\sum_{j=1}^d m_{t,j} \widehat{\mathbf{Q}}_{\otimes,j\cdot} \widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1} \Big(\sum_{j=1}^d m_{t,j} \widehat{\mathbf{Q}}_{\otimes,j\cdot} \mathbf{Q}_{\otimes,j\cdot}' \mathbf{f}_{Z,t}\Big)
\end{aligned}$$

$$
\begin{aligned}
&+\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\varepsilon_{t,j}\Big)\\
&=\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)(\mathbf{H}_{\otimes}')^{-1}\mathbf{f}_{Z,t}\\
&+\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\varepsilon_{t,j}\Big)\\
&+\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}(\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}-\widehat{\mathbf{Q}}_{\otimes,j\cdot})'\Big)(\mathbf{H}_{\otimes}')^{-1}\mathbf{f}_{Z,t}\\
&=: (\mathbf{H}_{\otimes}')^{-1}\mathbf{f}_{Z,t}+\widetilde{\boldsymbol{\varepsilon}}_{H,t}+\widetilde{\boldsymbol{\varepsilon}}_{t}+\widetilde{\mathbf{f}}_{Z,t},\quad\text{where} \qquad (3.73)
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{\boldsymbol{\varepsilon}}_{H,t} &:= \Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d} m_{t,j}\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}\varepsilon_{t,j}\Big),\\
\widetilde{\boldsymbol{\varepsilon}}_{t} &:= \Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d} m_{t,j}(\widehat{\mathbf{Q}}_{\otimes,j\cdot}-\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot})\varepsilon_{t,j}\Big),\\
\widetilde{\mathbf{f}}_{Z,t} &:= \Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\Big)^{-1}\Big(\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}(\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}-\widehat{\mathbf{Q}}_{\otimes,j\cdot})'\Big)(\mathbf{H}_{\otimes}')^{-1}\mathbf{f}_{Z,t}.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\Big\|\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}' - \sum_{j=1}^{d} m_{t,j}\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}\mathbf{Q}_{\otimes,j\cdot}'\mathbf{H}_{\otimes}'\Big\|_F\\
&\le \sum_{j=1}^{d}\big\|\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}'-\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}\mathbf{Q}_{\otimes,j\cdot}'\mathbf{H}_{\otimes}'\big\|_F\\
&\le \sum_{j=1}^{d}\big\|\widehat{\mathbf{Q}}_{\otimes,j\cdot}-\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}\big\|^2 + 2\sum_{j=1}^{d}\big\|\widehat{\mathbf{Q}}_{\otimes,j\cdot}-\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}\big\|\cdot\big\|\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}\big\|\\
&= O_P\Big(\big\|\widehat{\mathbf{Q}}_{\otimes}-\mathbf{Q}_{\otimes}\mathbf{H}_{\otimes}'\big\|_F^2+\big\|\widehat{\mathbf{Q}}_{\otimes}-\mathbf{Q}_{\otimes}\mathbf{H}_{\otimes}'\big\|_F\Big)\\
&= O_P\Big\{\max_{k\in[K]}\Big(T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-2}+d^2g_s^{-2}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1}+g_\eta d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big)^{\frac{1}{2}}\Big\}=o_P(1),
\end{aligned}
$$

where we used the Cauchy–Schwarz inequality, (3.70) and (3.72) in the last equality, and Assumption (R1). Hence we have

$$
\sum_{j=1}^{d} m_{t,j}\widehat{\mathbf{Q}}_{\otimes,j\cdot}\widehat{\mathbf{Q}}_{\otimes,j\cdot}' \xrightarrow{p} \sum_{j=1}^{d} m_{t,j}\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}\mathbf{Q}_{\otimes,j\cdot}'\mathbf{H}_{\otimes}'. \qquad (3.74)
$$

Note that by (3.51) we have $\|\mathbf{H}_\otimes^{-1}\|_F = O_P(1)\cdot\|\mathbf{Q}_K'\widehat{\mathbf{Q}}_K\otimes\cdots\otimes\mathbf{Q}_1'\widehat{\mathbf{Q}}_1\| = O_P(1)$, which will be used later in the proof. To bound $\mathbf{f}_{Z,t}$, from Assumption (F1), we have

$$\mathbb{E}\|\mathbf{f}_t\|^2 = r = O(1),$$

and hence with Assumption (L1),

$$\|\mathbf{f}_{Z,t}\|^2 \le \|\mathbf{Z}_\otimes^{1/2}\|_F^2\cdot\|\mathbf{f}_t\|^2 = O_P\Bigg(\prod_{j=1}^K d_j^{\alpha_{j,1}}\Bigg) = O_P(g_s).$$

With (3.72), (3.74) and the above result,

$$\begin{aligned}\|\widetilde{\mathbf{f}}_{Z,t}\|^2 &= O_P(1)\cdot\Big(\sum_{j=1}^d \big\|\widehat{\mathbf{Q}}_{\otimes,j\cdot}\big\|^2\Big)\Big(\sum_{j=1}^d \big\|\mathbf{H}_\otimes\mathbf{Q}_{\otimes,j\cdot}-\widehat{\mathbf{Q}}_{\otimes,j\cdot}\big\|^2\Big)\\ &\quad\cdot\big\|(\mathbf{\Sigma}_{A,K}\otimes\cdots\otimes\mathbf{\Sigma}_{A,1})^{-1}\big\|_F^2\cdot\big\|\mathbf{H}_\otimes^{-1}\big\|_F^6\cdot\|\mathbf{f}_{Z,t}\|^2\\ &= O_P\Big\{\max_{k\in[K]}\Big(T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1}+d^2g_s^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1}+g_\eta g_s d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big)\Big\},\end{aligned} \tag{3.75}$$

where we also used Assumption (L1) in the last equality. Similarly, by (3.71),

$$\begin{aligned}\|\widetilde{\boldsymbol{\varepsilon}}_t\|^2 &= O_P\Big(\big\|(\mathbf{\Sigma}_{A,K}\otimes\ldots\otimes\mathbf{\Sigma}_{A,1})^{-1}\big\|_F^2\,\|\mathbf{H}_\otimes^{-1}\|_F^4\max_{j\in[d]}\big\|\mathbf{H}_\otimes\mathbf{Q}_{\otimes,j\cdot}-\widehat{\mathbf{Q}}_{\otimes,j\cdot}\big\|^2\sum_{j,\ell=1}^d|\mathbb{E}\varepsilon_{t,j}\varepsilon_{t,\ell}|\Big)\\ &= O_P\Big\{\max_{k\in[K]}\Big(T^{-1}dd_{\text{-}k}d_k^{3\alpha_{k,1}-\alpha_{k,r_k}}g_s^{-2}g_w^{-1}+d^3g_s^{-2}g_w^{-1}d_k^{2\alpha_{k,1}-2\alpha_{k,r_k}-2}+dg_\eta g_w^{-1}d_k^{2\alpha_{k,1}-2\alpha_{k,r_k}}\Big)\Big\},\end{aligned} \tag{3.76}$$

since $\sum_{j,\ell=1}^d|\mathbb{E}\varepsilon_{t,j}\varepsilon_{t,\ell}| = O(d)$ by Assumption (E1). By the same token,

$$\begin{aligned}\|\widetilde{\boldsymbol{\varepsilon}}_{H,t}\|^2 &= O_P(1)\cdot\big\|(\mathbf{\Sigma}_{A,K}\otimes\cdots\otimes\mathbf{\Sigma}_{A,1})^{-1}\big\|_F^2\cdot\|\mathbf{H}_\otimes^{-1}\|_F^2\\ &\quad\cdot\big\|\mathbf{Z}_\otimes^{-1/2}\big\|_F^2\cdot\Big\|\sum_{j=1}^d m_{t,j}\mathbf{A}_{\otimes,j\cdot}\varepsilon_{t,j}\Big\|^2 = O_P(d/g_w),\end{aligned} \tag{3.77}$$

where we used

$$\mathbb{E}\Big\|\sum_{j=1}^d m_{t,j}\mathbf{A}_{\otimes,j\cdot}\epsilon_{t,j}\Big\|^2 \le \max_{j\in[d]}\big\|\mathbf{A}_{\otimes,j\cdot}\big\|_F^2\cdot\sum_{j,\ell=1}^d\big|\mathbb{E}\epsilon_{t,j}\epsilon_{t,\ell}\big| = O(d).$$

Therefore, we have from (3.75), (3.77) and (3.76),

$$\begin{aligned}&\big\|\widehat{\mathbf{f}}_{Z,t}-(\mathbf{H}_{\otimes}')^{-1}\mathbf{f}_{Z,t}\big\|^2\leq\|\widetilde{\boldsymbol{\varepsilon}}_{H,t}\|^2+\|\widetilde{\boldsymbol{\varepsilon}}_t\|^2+\|\widetilde{\mathbf{f}}_{Z,t}\|^2\\&=O_P\Big\{\max_{k\in[K]}\Big(T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1}+d^2g_s^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1}+g_\eta g_s d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big)+\frac{d}{g_w}\Big\},\end{aligned}$$

where we also used $1/2<\alpha_{k,r_k}\leq\alpha_{k,1}\leq 1$ from Assumption (L1) to conclude that $dg_s^{-1}g_w^{-1}=o(1)$, so that in fact $\|\widetilde{\boldsymbol{\varepsilon}}_t\|^2=o_P\big(\|\widetilde{\mathbf{f}}_{Z,t}\|^2\big)$.

Now from (3.73) and using the notations in (3.66), we can obtain the vectorised imputed values, which are the vectorised estimated common components, as $\widehat{\mathbf{c}}_t=\widehat{\mathbf{Q}}_{\otimes}\widehat{\mathbf{f}}_{Z,t}$ for any $t\in[T]$. Then for $j\in[d]$, we have the squared individual imputation error as

$$\begin{aligned}&(\widehat{\mathcal{C}}_{t,i_1,\ldots,i_K}-\mathcal{C}_{t,i_1,\ldots,i_K})^2=(\widehat{\mathbf{c}}_t-\mathbf{c}_t)_j^2=\big(\widehat{\mathbf{Q}}_{\otimes,j\cdot}'\widehat{\mathbf{f}}_{Z,t}-\mathbf{Q}_{\otimes,j\cdot}'\mathbf{f}_{Z,t}\big)^2\\&=\big\{\big(\widehat{\mathbf{Q}}_{\otimes,j\cdot}-\mathbf{H}_{\otimes}\mathbf{Q}_{\otimes,j\cdot}\big)'\big((\mathbf{H}_{\otimes}')^{-1}\mathbf{f}_{Z,t}+\widetilde{\boldsymbol{\varepsilon}}_{H,t}+\widetilde{\boldsymbol{\varepsilon}}_t+\widetilde{\mathbf{f}}_{Z,t}\big)+\mathbf{A}_{\otimes,j\cdot}'\mathbf{Z}_{\otimes}^{-1/2}\mathbf{H}_{\otimes}'\big(\widetilde{\boldsymbol{\varepsilon}}_{H,t}+\widetilde{\boldsymbol{\varepsilon}}_t+\widetilde{\mathbf{f}}_{Z,t}\big)\big\}^2\\&=O_P\Big\{\max_{k\in[K]}\Big(T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1}g_w^{-1}\\&\qquad+d^2g_s^{-1}g_w^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1}+g_\eta g_s g_w^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big)+\frac{d}{g_w^2}\Big\},\end{aligned}$$

where we used (3.71), (3.72) and Assumption (R1) in the last equality.

Lastly, we have the average imputation error as the following,

$$\begin{aligned}&\frac{1}{Td}\sum_{t=1}^T\sum_{i_1,\ldots,i_K=1}^{d_1,\ldots,d_K}(\widehat{\mathcal{C}}_{t,i_1,\ldots,i_K}-\mathcal{C}_{t,i_1,\ldots,i_K})^2\\&=\frac{1}{Td}\sum_{t=1}^T\|\widehat{\mathbf{c}}_t-\mathbf{c}_t\|^2=\frac{1}{Td}\sum_{t=1}^T\big\|\widehat{\mathbf{Q}}_{\otimes}\widehat{\mathbf{f}}_{Z,t}-\mathbf{Q}_{\otimes}\mathbf{f}_{Z,t}\big\|^2\\&=\frac{1}{Td}\sum_{t=1}^T\Big\|\big(\widehat{\mathbf{Q}}_{\otimes}-\mathbf{Q}_{\otimes}\mathbf{H}_{\otimes}'\big)(\mathbf{H}_{\otimes}')^{-1}\mathbf{f}_{Z,t}+\widehat{\mathbf{Q}}_{\otimes}\big(\widetilde{\boldsymbol{\varepsilon}}_{H,t}+\widetilde{\boldsymbol{\varepsilon}}_t+\widetilde{\mathbf{f}}_{Z,t}\big)\Big\|^2\\&=O_P\Big\{\max_{k\in[K]}\Big(T^{-1}d_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1}+dg_s^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1}+d^{-1}g_\eta g_s d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\Big)+\frac{1}{g_w}\Big\},\end{aligned}\tag{3.78}$$

where we used (3.75), (3.77), (3.76) and Lemma 3.6 in the last equality, and the fact that $\|\widehat{\mathbf{Q}}_{\otimes}\|_F^2=r=O(1)$. This completes the proof of Theorem 3.2. □

***Proof of Corollary 3.1.*** It is direct from Theorem 3.1 and Theorem 3.2. □

***Proof of Theorem 3.5.*** Firstly, we use the notations in (3.19), and define also $\mathbf{Z}:=\mathbf{Z}_k$ and

$\mathbf{R}^* := \mathbf{R}_k^*$, which coincides with the $\mathbf{R}^*$ defined in (3.37). Then for $j \in [r_k]$,

$$\begin{aligned}
&\lambda_j(\mathbf{R}^*) = \lambda_j\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}\mathbf{F}_{Z,t}\mathbf{\Lambda}'\mathbf{\Lambda}\mathbf{F}_{Z,t}'\mathbf{Q}'\Big)\\
&= \lambda_j\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}_k\mathbf{F}_t\big[\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j\big]'\big[\otimes_{j\in[K]\setminus\{k\}}\mathbf{A}_j\big]\mathbf{F}_t'\mathbf{A}_k'\Big)\\
&= \lambda_j\Big(\mathbf{A}_k'\mathbf{A}_k\cdot\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t[\otimes_{\ell\in[K]\setminus\{k\}}\mathbf{A}_\ell]'[\otimes_{\ell\in[K]\setminus\{k\}}\mathbf{A}_\ell]\mathbf{F}_t'\Big)\\
&\asymp_P \lambda_j\big(\mathbf{A}_k'\mathbf{A}_k\cdot\text{tr}(\otimes_{\ell\in[K]\setminus\{k\}}\mathbf{A}_\ell'\mathbf{A}_\ell)\big) = \lambda_j(\mathbf{A}_k'\mathbf{A}_k)\prod_{\ell\in[K]\setminus\{k\}}\text{tr}(\mathbf{A}_\ell'\mathbf{A}_\ell)\\
&\asymp \lambda_j(\mathbf{Z}\mathbf{Q}'\mathbf{Q})\prod_{\ell\in[K]\setminus\{k\}}\sum_{i=1}^{r_k}d_\ell^{\alpha_{\ell,i}} \asymp \lambda_j\big(\mathbf{\Sigma}_{A,k}^{1/2}\mathbf{Z}\mathbf{\Sigma}_{A,k}^{1/2}\big)\cdot\prod_{\ell\in[K]\setminus\{k\}}d_\ell^{\alpha_{\ell,1}}\\
&\asymp \lambda_j(\mathbf{Z})d_k^{-\alpha_{k,1}}g_s \asymp g_s d_k^{\alpha_{k,j}-\alpha_{k,1}},
\end{aligned} \tag{3.79}$$

where the third line uses Assumption (F1), and Assumption (L1) in the second last line. The last line uses Theorem 1 of Ostrowski (1959) on the eigenvalues of a congruent transformation $\mathbf{\Sigma}_{A,k}^{1/2}\mathbf{Z}\mathbf{\Sigma}_{A,k}^{1/2}$ of $\mathbf{Z}$, and from Assumption (L1) that $\mathbf{\Sigma}_{A,k}$ has eigenvalues uniformly bounded away from 0 and infinity.

Since $\widehat{\mathbf{S}} = \mathbf{R}^* + (\widehat{\mathbf{S}} - \mathbf{R}^*)$, for $j \in [r_k]$, we have by Weyl's inequality that

$$\begin{aligned}
&|\lambda_j(\widehat{\mathbf{S}}) - \lambda_j(\mathbf{R}^*)| \le \big\|\widehat{\mathbf{S}} - \mathbf{R}^*\big\| \le \big\|\widetilde{\mathbf{R}} - \mathbf{R}^*\big\| + \big\|\mathbf{R}_1\big\| + \big\|\mathbf{R}_2\big\| + \big\|\mathbf{R}_3\big\|\\
&\le \omega_k\Big(\sup_{\|\boldsymbol{\gamma}\|=1}|\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma})| + \sup_{\|\boldsymbol{\gamma}\|=1}R_1 + \sup_{\|\boldsymbol{\gamma}\|=1}R_2 + \sup_{\|\boldsymbol{\gamma}\|=1}R_3\Big) = o_P(\omega_k),
\end{aligned} \tag{3.80}$$

where we use the decomposition in (3.37) in the first line, and $\omega_k := g_s d_k^{\alpha_{k,r_k}-\alpha_{k,1}}$ is defined at the beginning of the proof of Lemma 3.2. The second line uses $\widetilde{R}(\boldsymbol{\gamma}), R^*(\boldsymbol{\gamma}), R_1, R_2$ and $R_3$ defined in (3.38), and the convergence in probability in (3.42) and (3.47).

Secondly, with Assumption (R1) and our choice of $\xi$ (see also (3.40)),

$$\xi/\omega_k \asymp d g_s^{-1} d_k^{\alpha_{k,1}-\alpha_{k,r_k}}[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}] = o(1). \tag{3.81}$$

For $r_k > 1$, if $j \in [r_k - 1]$, using (3.80) and (3.81), consider

$$\begin{aligned}
\frac{\lambda_{j+1}(\widehat{\mathbf{S}}) + \xi}{\lambda_j(\widehat{\mathbf{S}}) + \xi} &\le \frac{\lambda_{j+1}(\mathbf{R}^*) + \xi + |\lambda_{j+1}(\widehat{\mathbf{S}}) - \lambda_{j+1}(\mathbf{R}^*)|}{\lambda_j(\mathbf{R}^*) + \xi - |\lambda_j(\widehat{\mathbf{S}}) - \lambda_j(\mathbf{R}^*)|} = \frac{\lambda_{j+1}(\mathbf{R}^*) + o_P(\omega_k)}{\lambda_j(\mathbf{R}^*) + o_P(\omega_k)}\\
&= \frac{\lambda_{j+1}(\mathbf{R}^*)}{\lambda_j(\mathbf{R}^*)}(1 + o_P(1)) \asymp_P d_k^{\alpha_{k,j+1}-\alpha_{k,j}},
\end{aligned} \tag{3.82}$$

where the last line uses (3.79). Also, for $j \in [r_k - 1]$,

$$\frac{\lambda_{r_k+1}(\widehat{\mathbf{S}}) + \xi}{\lambda_{r_k}(\widehat{\mathbf{S}}) + \xi} = \frac{\lambda_{r_k+1}(\widehat{\mathbf{S}}) + \xi}{\omega_k(1 + o_P(1))} = O_P\Big(\frac{\lambda_{r_k+1}(\widehat{\mathbf{S}})}{\omega_k} + \frac{\xi}{\omega_k}\Big) \tag{3.83}$$

$$= O_P\Big(\sup_{\|\boldsymbol{\gamma}\|=1}(\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma}) + R_1 + R_2 + R_3) + \xi/\omega_k\Big)$$

$$= O_P(\xi/\omega_k) = o_P(d_k^{\alpha_{k,j+1}-\alpha_{k,j}}), \tag{3.84}$$

where the second last equality uses (3.40), (3.41) and (3.47) together with our choice of $\xi$, and the last equality uses the extra rate assumption in the statement of the theorem. In the third equality, we assume the following is true (to be shown at the end of this proof):

$$\lambda_j(\widehat{\mathbf{S}}) = \lambda_j((\widetilde{\mathbf{R}} - \mathbf{R}^*) + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3), \quad j = r_k + 1, \ldots, d_k, \tag{3.85}$$

so that

$$\frac{\lambda_j(\widehat{\mathbf{S}})}{\omega_k} = \lambda_j\Big(\frac{1}{\omega_k}((\widetilde{\mathbf{R}} - \mathbf{R}^*) + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3)\Big) \leq \sup_{\|\boldsymbol{\gamma}\|=1}((\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma})) + R_1 + R_2 + R_3).$$

Hence for $j = r_k + 1, \ldots, \lfloor d_k/2 \rfloor$ (true also for $r_k = 1$),

$$\frac{\lambda_{j+1}(\widehat{\mathbf{S}}) + \xi}{\lambda_j(\widehat{\mathbf{S}}) + \xi} \geq \frac{\xi/\omega_k}{\sup_{\|\boldsymbol{\gamma}\|=1}((\widetilde{R}(\boldsymbol{\gamma}) - R^*(\boldsymbol{\gamma})) + R_1 + R_2 + R_3) + \xi/\omega_k} \geq \frac{1}{C} \tag{3.86}$$

in probability for some generic constant $C > 0$, where the last inequality uses (3.40), (3.41) and (3.47) together with our choice of $\xi$. Combining (3.82), (3.84) and (3.86), we can easily see that our proposed $\widehat{r}_k$ is a consistent estimator for $r_k$.

If $r_k = 1$, then (3.84) becomes

$$\frac{\lambda_{r_k+1}(\widehat{\mathbf{S}}) + \xi}{\lambda_{r_k}(\widehat{\mathbf{S}}) + \xi} = O_P(\xi/\omega_k) = o_P(1).$$

When combined with (3.86) which is true also for $r_k = 1$, we can see that $\widehat{r}_k = 1$ in probability, showing that $\widehat{r}_k$ is a consistent estimator of $r_k$.

It remains to show (3.85). To this end, from (3.79) and (3.80), the first $r_k$ eigenvalues of $\widehat{\mathbf{S}}$ coincides with those of $\mathbf{R}^*$ asymptotically, so that the first $r_k$ eigenvectors corresponding to $\widehat{\mathbf{S}}$ coincides with those for $\mathbf{R}^*$ asymptotically as $T, d_k \to \infty$, which are necessarily in $\mathcal{N}^\perp :=$ Span$(\mathbf{Q})$, the linear span of the columns of $\mathbf{Q}$ (see (3.13), where $\mathbf{R}^*$ is sandwiched by $\mathbf{Q}$ and $\mathbf{Q}'$). This means that the $(r_k + 1)$-th largest eigenvalue of $\widehat{\mathbf{S}}$ and beyond will asymptotically have eigenvectors in $\mathcal{N}$, the orthogonal complement of $\mathcal{N}^\perp$. Then for any unit vectors $\boldsymbol{\gamma} \in \mathcal{N}$, we have from the definitions of $\mathbf{R}^*$, $\widetilde{\mathbf{R}}$. $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ in (3.38) that

$$\boldsymbol{\gamma}'\widehat{\mathbf{S}}\boldsymbol{\gamma} = \boldsymbol{\gamma}'\big(\mathbf{R}^* + (\widetilde{\mathbf{R}} - \mathbf{R}^*) + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3\big)\boldsymbol{\gamma} = \boldsymbol{\gamma}'\big((\widetilde{\mathbf{R}} - \mathbf{R}^*) + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3\big)\boldsymbol{\gamma},$$

which is equivalent to (3.85). This completes the proof of the theorem. □

***Proof of Proposition 3.1.*** For simplicity, first consider the scenario with conditions 1 and 2 satisfied. We can show stable convergence in law similar to Proposition 3.1 in Xiong and Pelger (2023). First, using Assumption (F1) we can write

$$\begin{aligned}
&\sqrt{T d_k^{\alpha_{k,r_k}}} \cdot \mathbf{D}^{-1}\mathbf{H}_k^{a,*} \sum_{i=1}^{d_k} \mathbf{Q}_{k,i\cdot}\mathbf{A}'_{k,i\cdot} \sum_{h=1}^{d_{\text{-}k}} \Delta_{F,k,ij,h}\mathbf{A}_{k,j\cdot} \\
&= \sum_{t=1}^{T} \sqrt{\frac{d_k^{\alpha_{k,r_k}}}{T}} \cdot \sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}} \left(\frac{T\cdot \mathbb{1}\{t\in\psi_{k,ij,h}\}}{|\psi_{k,ij,h}|} - 1\right) \\
&\quad \cdot \mathbf{D}^{-1}\mathbf{H}_k^{a,*}\mathbf{Q}_{k,i\cdot}\mathbf{A}'_{k,i\cdot}\big(\text{mat}_k(\mathcal{F}_t)\mathbf{v}_{k,h}\mathbf{v}'_{k,h}\text{mat}_k(\mathcal{F}_t)' - \mathbf{v}'_{k,h}\mathbf{v}_{k,h}\boldsymbol{\Sigma}_k\big)\mathbf{A}_{k,j\cdot}.
\end{aligned}$$

Define the filtration $\mathcal{G}^T := \sigma(\cup_{s=1}^T \mathcal{G}_s)$ where the sigma-algebra $\mathcal{G}_s := \sigma(\{\mathcal{M}_{t,i_1,\ldots,i_K} \mid t \le s\}, \mathbf{A}_1,\ldots,\mathbf{A}_K)$. Let $\mathbf{u} \in \mathbb{R}^{r_k}$ be a non-random unit vector. For a given $k\in[K], j\in[d_k]$, define also the random variable

$$\begin{aligned}
g_{k,j,t} &:= \mathbf{u}'\sqrt{\frac{d_k^{\alpha_{k,r_k}}}{T}} \cdot \sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}} \Big(\frac{T\cdot \mathbb{1}\{t\in\psi_{k,ij,h}\}}{|\psi_{k,ij,h}|} - 1\Big) \\
&\quad \cdot \mathbf{D}^{-1}\mathbf{H}_k^{a,*}\mathbf{Q}_{k,i\cdot}\mathbf{A}'_{k,i\cdot}(\text{mat}_k(\mathcal{F}_t)\mathbf{v}_{k,h}\mathbf{v}'_{k,h}\text{mat}_k(\mathcal{F}_t)' - \mathbf{v}'_{k,h}\mathbf{v}_{k,h}\boldsymbol{\Sigma}_k)\mathbf{A}_{k,j\cdot}.
\end{aligned}$$

Since each entry in $\mathcal{F}_t$ is i.i.d. by Assumption (F1) and is independent of $(\mathcal{M}_t, \mathbf{A}_1,\ldots,\mathbf{A}_K)$ by Assumptions (O1) and (L1), we have $\mathbb{E}[g_{k,j,t} \mid \mathcal{G}_{t-1}] = 0$. Define

$$\begin{aligned}
\boldsymbol{\Xi}_{F,k} &:= \text{Var}\Big\{\textbf{vec}\big(\text{mat}_k(\mathcal{F}_t)\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}\text{mat}_k(\mathcal{F}_t)' - \text{tr}(\mathbf{A}_{\text{-}k}\mathbf{A}'_{\text{-}k})\boldsymbol{\Sigma}_k\big)\Big\}, \\
\mathbf{x}_{F,k,j,il} &:= \textbf{vec}\big([\mathbf{A}'_{k,j\cdot}\otimes(\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\mathbf{Q}_{k,i\cdot}\mathbf{A}'_{k,i\cdot})]\boldsymbol{\Xi}_{F,k}[\mathbf{A}'_{k,j\cdot}\otimes(\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\mathbf{Q}_{k,l\cdot}\mathbf{A}'_{k,l\cdot})]'\big),
\end{aligned}$$

so that we have

$$\begin{aligned}
&\big\|\mathbf{x}_{F,k,j,il}\big\|^2 \le \big\|\boldsymbol{\Xi}_{F,k}\big\|_F^2 \cdot \big\|\mathbf{D}^{-1}\big\|_F^4 \cdot \big\|\mathbf{Z}_k^{-1/2}\big\|_F^4 \\
&= O_P\Big(d_k^{4\alpha_{k,1}-6\alpha_{k,r_k}} \prod_{j=1}^{K} d_j^{-4\alpha_{j,1}} \prod_{j\in[K]\setminus\{k\}} d_j^{4\alpha_{j,1}}\Big) = O_P\big(d_k^{-6\alpha_{k,r_k}}\big),
\end{aligned}$$

leading to

$$d_k^{\alpha_{k,r_k}}\,\mathbb{E}\Big\|\sum_{i=1}^{d_k}\sum_{l=1}^{d_k}\Big(\frac{T\cdot\mathbb{1}\{t\in\psi_{k,ij}\}}{|\psi_{k,ij}|}-1\Big)\Big(\frac{T\cdot\mathbb{1}\{t\in\psi_{k,lj}\}}{|\psi_{k,lj}|}-1\Big)\big(\mathbf{x}_{F,k,j,il}-\mathbb{E}[\mathbf{x}_{F,k,j,il}]\big)\Big\|^2$$
$$=O_P\big(d_k^{4-5\alpha_{k,r_k}}\big)=o_P(1).$$

Hence, it holds that

$$\begin{aligned}
&\sum_{t=1}^{T}\mathbb{E}\big[g_{k,j,t}^2\mid\mathcal{G}_{t-1}\big]\\
&=\frac{d_k^{\alpha_{k,r_k}}}{T}\cdot\sum_{t=1}^{T}\mathbb{E}\Big\{\mathbf{u}'\sum_{i=1}^{d_k}\sum_{l=1}^{d_k}\Big(\frac{T\cdot\mathbb{1}\{t\in\psi_{k,ij}\}}{|\psi_{k,ij}|}-1\Big)\Big(\frac{T\cdot\mathbb{1}\{t\in\psi_{k,lj}\}}{|\psi_{k,lj}|}-1\Big)\\
&\quad\cdot\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\big[\mathbf{A}_{k,j\cdot}'\otimes(\mathbf{Q}_{k,i\cdot}\mathbf{A}_{k,i\cdot}')\big]\mathbf{vec}\big(\text{mat}_k(\mathcal{F}_t)\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}\text{mat}_k(\mathcal{F}_t)'-\text{tr}(\mathbf{A}_{\text{-}k}\mathbf{A}_{\text{-}k}')\mathbf{\Sigma}_k\big)\\
&\quad\cdot\mathbf{vec}\big(\text{mat}_k(\mathcal{F}_t)\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}\text{mat}_k(\mathcal{F}_t)'-\text{tr}(\mathbf{A}_{\text{-}k}\mathbf{A}_{\text{-}k}')\mathbf{\Sigma}_k\big)'\\
&\quad\cdot\big[\mathbf{A}_{k,j\cdot}'\otimes(\mathbf{Q}_{k,l\cdot}\mathbf{A}_{k,l\cdot}')\big]'(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}\mathbf{u}\mid\mathcal{G}_{t-1}\Big\}\\
&\xrightarrow{p}d_k^{2+\alpha_{k,r_k}}\cdot\lim_{d_k\to\infty}\frac{1}{d_k^2}\sum_{i=1}^{d_k}\sum_{l=1}^{d_k}\Big(\frac{T\cdot\mathbb{1}\{t\in\psi_{k,ij}\}}{|\psi_{k,ij}|}-1\Big)\Big(\frac{T\cdot\mathbb{1}\{t\in\psi_{k,lj}\}}{|\psi_{k,lj}|}-1\Big)\\
&\quad\cdot\mathbf{u}'\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\big[\mathbf{A}_{k,j\cdot}'\otimes(\mathbf{Q}_k'\mathbf{A}_k)\big]\mathbf{\Xi}_{F,k}\big[\mathbf{A}_{k,j\cdot}'\otimes(\mathbf{Q}_k'\mathbf{A}_k)\big]'(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}\mathbf{u}\\
&\xrightarrow{p}d_k^{2+\alpha_{k,r_k}}\omega_{\psi,k,j}\cdot\mathbf{u}'\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\big[\mathbf{A}_{k,j\cdot}'\otimes(\mathbf{Q}_k'\mathbf{A}_k)\big]\mathbf{\Xi}_{F,k}\big[\mathbf{A}_{k,j\cdot}'\otimes(\mathbf{Q}_k'\mathbf{A}_k)\big]'(\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}\mathbf{u},
\end{aligned}$$

which satisfies the nesting condition of Theorem 6.1 in Häusler and Luschgy (2015). From Assumption (O1), we have $|(T\cdot\mathbb{1}\{t\in\psi_{k,ij,h}\}/|\psi_{k,ij,h}|)-1|\le\max(\psi_0^{-1}-1,1)$. Hence with $\epsilon$ from Proposition 3.1, we have

$$\begin{aligned}
&\sum_{t=1}^{T}\mathbb{E}\big[g_{k,j,t}^{2+\epsilon}\mid\mathcal{G}_{t-1}\big]\\
&\le\|\mathbf{A}_{k,j\cdot}\|^{2+\epsilon}\cdot d_k^{\alpha_{k,r_k}(1+\epsilon/2)}T^{-(1+\epsilon/2)}\cdot\big\|\mathbf{D}^{-1}\mathbf{H}_k^{a,*}\big\|^{2+\epsilon}\sum_{t=1}^{T}\Big\|\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\Big(\frac{T\cdot\mathbb{1}\{t\in\psi_{k,ij,h}\}}{|\psi_{k,ij,h}|}-1\Big)\\
&\quad\cdot\mathbf{Q}_{k,i\cdot}\mathbf{A}_{k,i\cdot}'\Big\{\text{mat}_k(\mathcal{F}_t)\mathbf{v}_{k,h}\mathbf{v}_{k,h}'\text{mat}_k(\mathcal{F}_t)'-\mathbf{v}_{k,h}'\mathbf{v}_{k,h}\mathbf{\Sigma}_k\Big\}\Big\|^{2+\epsilon}\\
&=O_P\Big\{(d_k^{\alpha_{k,1}-\alpha_{k,r_k}}g_s^{-1})^{2+\epsilon}\cdot\frac{d_k^{\alpha_{k,r_k}(1+\epsilon/2)}}{T^{\epsilon/2}}\cdot d_k^{-\alpha_{k,r_k}(1+\epsilon/2)}g_s^{2+\epsilon}\Big\}=O_P\Big(\frac{d_k^{\alpha_{k,1}-\alpha_{k,r_k}}}{T^{\epsilon/2}}\Big)=o_P(1),
\end{aligned}$$

which is sufficient for the conditional Lindeberg condition in Häusler and Luschgy (2015) to hold. Then by the stable martingale central limit theorem (Theorem 6.1 in Häusler and Luschgy

(2015)), we have

$$\sum_{t=1}^{T} g_{k,j,t} \to \mathcal{N}\Big(0, \mathbf{D}^{-1}\mathbf{H}_k^{a,*}\big[h_{k,j}(\mathbf{A}_{k,j\cdot})\big](\mathbf{H}_k^{a,*})'\mathbf{D}^{-1}\Big) \quad \mathcal{G}^T\text{-stably as } T \to \infty,$$

where $h_{k,j}(\mathbf{A}_{k,j\cdot}) = d_k^{2+\alpha_{k,r_k}} \omega_{\psi,k,j} \cdot \big[\mathbf{A}'_{k,j\cdot} \otimes (\mathbf{Q}'_k \mathbf{A}_k)\big] \boldsymbol{\Xi}_{F,k} \big[\mathbf{A}'_{k,j\cdot} \otimes (\mathbf{Q}'_k \mathbf{A}_k)\big]'$.

When condition 2 is relaxed, all the previous steps can be repeated by noticing that we now have $|\psi_{k,ij,h}| \xrightarrow{p} c^*_{k,ij}$ for some constant $c^*_{k,ij}$, and there exists some constant $p^*_{k,ij}$ such that as $T, d_1, \ldots, d_K \to \infty$,

$$\sum_{h=1}^{d_{\text{-}k}} \Big(\frac{T \cdot \mathbb{1}\{t \in \psi_{k,ij,h}\}}{|\psi_{k,ij,h}|} - 1\Big)(\text{mat}_k(\mathcal{F}_t)\mathbf{v}_{k,h}\mathbf{v}'_{k,h}\text{mat}_k(\mathcal{F}_t)' - \mathbf{v}'_{k,h}\mathbf{v}_{k,h}\boldsymbol{\Sigma}_k)$$
$$\xrightarrow{p} \Big(\frac{Tp^*_{k,ij}}{c^*_{k,ij}} - 1\Big)\big\{\text{mat}_k(\mathcal{F}_t)\mathbf{A}'_{\text{-}k}\mathbf{A}_{\text{-}k}\text{mat}_k(\mathcal{F}_t)' - \text{tr}(\mathbf{A}_{\text{-}k}\mathbf{A}'_{\text{-}k})\boldsymbol{\Sigma}_k\big\},$$

with $\omega^*_{\psi,k,j} := d_k^{-2} \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} (Tp^*_{k,ij}/c^*_{k,ij} - 1)(Tp^*_{k,lj}/c^*_{k,lj} - 1)$. Then the convergence of $\sum_{t=1}^{T} \mathbb{E}[g^2_{k,j,t} \mid \mathcal{G}_{t-1}]$ can be similarly constructed, etc. This completes the proof of Proposition 3.1. □

# Chapter 4

# On Testing Kronecker Product Structure in Tensor Factor Models

## 4.1 Introduction

With rapid advance in information technology, high-dimensional time series data observed in tensor form are becoming more readily available for analysis in fields such as finance, economics, bioinformatics or computer science, to name but a few areas. In many cases, low-rank structures in the tensor time series observed can be exploited, facilitating analysis and interpretations. The most commonly used devices are the CP-decomposition and the multilinear/Tucker decomposition of a tensor, leading to CP-decomposition tensor factor models and Tucker-decomposition tensor factor models, respectively. See Section 2.3 for a review on factor models. While tensor time series can be transformed back to vector time series through vectorisation and be analysed using traditional factor models for vector time series, the tensor structure of the data is lost and hence any corresponding interpretations from it. Moreover, vectorisation increases the dimension of the factor loading matrix significantly relative to the sample size, leading potentially to less accurate estimation and inferences (Chen and Lam, 2024b).

However, a tensor factor model comes with its assumptions. For using the Tucker decomposition in particular, a tensor factor model assumes that the factor loading matrix for the vectorised data is the Kronecker product of smaller dimensional factor loading matrices. For instance, suppose at each $t \in [T]$, a mean-zero matrix $\mathbf{Y}_t \in \mathbb{R}^{d_1 \times d_2}$ is observed. Consider a matrix factor model of the form

$$\mathbf{Y}_t = \mathbf{A}_1 \mathbf{F}_t \mathbf{A}_2' + \mathbf{E}_t, \tag{4.1}$$

where $\mathbf{F}_t \in \mathbb{R}^{r_1 \times r_2}$ is the core factor, $\mathbf{A}_k \in \mathbb{R}^{d_k \times r_k}$ is the mode-$k$ factor loading matrix, i.e., $\mathbf{A}_1$ and $\mathbf{A}_2$ are respectively the row and column loading matrices, and $\mathbf{E}_t$ is the noise matrix.

The vectorisation of (4.1) is

$$\mathbf{vec}(\mathbf{Y}_t) = (\mathbf{A}_2 \otimes \mathbf{A}_1)\,\mathbf{vec}(\mathbf{F}_t) + \mathbf{vec}(\mathbf{E}_t) \equiv \mathbf{A}_V\,\mathbf{vec}(\mathbf{F}_t) + \mathbf{vec}(\mathbf{E}_t), \tag{4.2}$$

where $\mathbf{A}_V := \mathbf{A}_2 \otimes \mathbf{A}_1$, which is a vector factor model for the time series data $\{\mathbf{vec}(\mathbf{Y}_t)\}$ with factor loading matrix $\mathbf{A}_V$. Clearly, the implicit assumption of a Kronecker product structure for $\mathbf{A}_V$ when using a matrix factor model for matrix-valued time series data should be the first thing to check before such a factor model is applied.

Motivated by this simple example, we propose a test in this chapter to test the Kronecker product structure of the factor loading matrix implied in the vectorised data when using a Tucker-decomposition tensor factor model (TFM), and extend it to higher order tensors. He et al. (2023a) has also noted this implicit assumption in a Tucker-decomposition matrix factor model, and proposes to test the "boundary" cases of each column (resp. row) of the data following a factor model with a common factor loading matrix, but with possibly distinct factors, or the whole matrix is just pure noise. Model (4.2) with a general $\mathbf{A}_V$ also implies a vector factor model with potentially different factor loading matrices for each column (resp. row) of the data, but they share the same factors. To explore the data as a matrix, connectedness through having a set of shared common factors rather than having the same factor loading matrix with all distinct factors is more meaningful. Practically, (4.2) is an alternative model easier to be satisfied by data than the "boundary" cases in He et al. (2023a), since the data still follows a more general factor model, just the implied Kronecker product structure in the factor loading matrix $\mathbf{A}_V$ is lost. This comes as no surprise then, that in all of the tests in He et al. (2023a) for their real data analyses, they cannot reject the null hypothesis of a matrix factor model. An easier alternative such as (4.2) with just a general $\mathbf{A}_V$ can provide a more critical test for the null hypothesis of a matrix factor model. See our portfolio return example in Section 4.5.2 for cases where our test can reject the null hypothesis of a matrix factor model, when He et al. (2023a) cannot.

We also stress that our model is fundamentally different from those used in testing for Kronecker product structure in the covariance matrix of the data. For example, Yu et al. (2022b) and Guggenberger et al. (2023) both propose tests for the Kronecker product structure of the covariance matrix of a vectorised matrix data. For model (4.1), even in the simplest hypothetical case of $\mathbf{E}_t$ and $\mathbf{F}_t$ being independent and $\mathbf{F}_t$ contains independent standard normal random variables, we have

$$\text{Cov}\{\mathbf{vec}(\mathbf{Y}_t)\} = \mathbf{A}_2\mathbf{A}_2' \otimes \mathbf{A}_1\mathbf{A}_1' + \text{Cov}\{\mathbf{vec}(\mathbf{E}_t)\},$$

so that the covariance matrix is never exactly of Kronecker product structure because of $\mathbf{E}_t$. Moreover, even with $\mathbf{E}_t = \mathbf{0}$, both $\mathbf{A}_1\mathbf{A}_1'$ and $\mathbf{A}_2\mathbf{A}_2'$ are of low rank, which is different from the full rank component matrices in the two papers mentioned above.

Our contributions in this chapter are threefold. Firstly, as a first in the literature, we propose a test to test directly a Tucker-decomposition TFM against the alternative of a (tensor) factor model with Kronecker product structure lost in some of its factor loading matrices. As shown in Section 4.3, for higher order tensors, testing against a tensor-decomposition TFM can be against a tensor factor model for the reshaped data, but not necessarily the vectorised data. This gives rise to flexibility and in fact statistical power in practical situations. Secondly, our analysis allows for weak factors, with our theoretical results developed to spell out rates of convergence explicitly. Last but not least, as a useful by-product, we developed tensor reshape theorems which can be useful in their own rights.

The rest of this chapter is organised as follows. Section 4.2 defines the tensor reshape operation used for our tests. Section 4.3 introduces the Kronecker product structure set and pinpoints exactly through a theorem when a tensor time series $\{\mathcal{Y}_t\}$ follows a Tucker-decomposition TFM. This becomes the basis for the construction of our test statistics. Section 4.4 lays down all the assumptions for this chapter, and presents the main theoretical results for our test statistics to be valid. Section 4.5 presents our simulation results and two sets of real data analyses. Finally, Section 4.6 provides details for model identification. Both our test and the tensor reshape operator can be implemented by the R package `KOFM`, available on R CRAN. Section 4.7 includes all the proofs. Hereafter in this chapter, we use the following definition

$$d := \prod_{k=1}^{K} d_k, \quad d_{\text{-}k} := d/d_k, \quad r := \prod_{k=1}^{K} r_k, \quad r_{\text{-}k} := r/r_k.$$

## 4.2 Introduction to Tensor Reshape

In this section, we introduce tensor reshape. Given an order-$K$ tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_K}$ and a set with ordered, strictly ascending elements $\{a_1, \ldots, a_\ell\} \subseteq [K]$, the $\textsc{reshape}(\cdot, \cdot)$ operator is defined as follows:

$$\text{If } \ell = 1, \quad \textsc{reshape}(\mathcal{X}, \{a_1\}) = \textsc{fold}_K\big\{\text{mat}_{a_1}(\mathcal{X}), \{I_1, \ldots, I_{a_1-1}, I_{a_1+1}, \ldots, I_K, I_{a_1}\}\big\};$$

$$\text{if } \ell = 2, \quad \textsc{reshape}(\mathcal{X}, \{a_1, a_2\})$$

$$= \textsc{fold}_{K-1}(\mathcal{X}_{a_1 \sim a_2}, \{I_1, \ldots, I_{a_1-1}, I_{a_1+1}, \ldots, I_{a_2-1}, I_{a_2+1}, \ldots, I_K, I_{a_1} I_{a_2}\}),$$

$$\text{where } \mathcal{X}_{a_1 \sim a_2} = \begin{pmatrix} \text{mat}_{a_1}\big[\textsc{fold}\{\text{mat}_{a_2}(\mathcal{X})_{1\cdot}, \{I_1, \ldots, I_{a_2-1}, I_{a_2+1}, \ldots, I_K\}\}\big] \\ \cdots \\ \text{mat}_{a_1}\big[\textsc{fold}\big\{\text{mat}_{a_2}(\mathcal{X})_{I_{a_2}\cdot}, \{I_1, \ldots, I_{a_2-1}, I_{a_2+1}, \ldots, I_K\}\big\}\big] \end{pmatrix};$$

$$\text{if } \ell \geq 3, \quad \textsc{reshape}(\mathcal{X}, \{a_1, \ldots, a_\ell\})$$

$$= \textsc{reshape}\big\{\textsc{reshape}(\mathcal{X}, \{a_{\ell-1}, a_\ell\}), \{a_1, \ldots, a_{\ell-2}, K-1\}\big\}.$$

Hence, reshaping an order-$K$ tensor along $\{a_1, \dots, a_\ell\}$ results in an order-$(K-\ell+1)$ tensor. A heuristic view of $\textsc{reshape}(\mathcal{X}, \{a_1, \dots, a_\ell\})$ is that all modes of $\mathcal{X}$ with indices $\{a_1, \dots, a_\ell\}$ are "merged" into a single mode acting as the last mode as a result. Note that one may recover $\mathcal{X}$ from $\textsc{reshape}(\mathcal{X}, \{a_1, \dots, a_\ell\})$ given the original dimension of $\mathcal{X}$ and $\{a_1, \dots, a_\ell\}$. To help readers to understand the reshape operator, we also present Figure 4.1 as an visualization.

As a simple example on tensor reshape, consider a matrix $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$. Trivially, we have

$$\textsc{reshape}(\mathbf{X}, \{2\}) = \mathbf{X}, \quad \textsc{reshape}(\mathbf{X}, \{1\}) = \mathbf{X}'.$$

Moreover, $\textsc{reshape}(\mathbf{X}, \{1,2\}) = \textsc{fold}_1(\mathbf{X}_{1\sim 2}, \{I_1 I_2\}) = \mathbf{X}_{1\sim 2} = \mathbf{vec}(\mathbf{X})$ since

$$\mathbf{X}_{1\sim 2} = \begin{pmatrix} \mathrm{mat}_1\left[\textsc{fold}\left\{\mathrm{mat}_2(\mathbf{X})_{1\cdot}, \{I_1\}\right\}\right] \\ \cdots \\ \mathrm{mat}_1\left[\textsc{fold}\left\{\mathrm{mat}_2(\mathbf{X})_{I_2\cdot}, \{I_1\}\right\}\right] \end{pmatrix} = \begin{pmatrix} (\mathbf{X}')_{1\cdot} \\ \cdots \\ (\mathbf{X}')_{I_2\cdot} \end{pmatrix} = \mathbf{vec}(\mathbf{X}).$$

In fact, it holds for any order-$K$ tensor $\mathcal{X}$ that $\textsc{reshape}(\mathcal{X}, [K]) = \mathbf{vec}(\mathcal{X})$.

We discuss some useful algebra of tensor reshape in the following. First, the reshape operator is linear in the first argument, i.e.,

$$\begin{aligned} &\textsc{reshape}(b_1\mathcal{X}_1 + b_2\mathcal{X}_2, \{a_1, \dots, a_\ell\}) \\ &= b_1 \cdot \textsc{reshape}(\mathcal{X}_1, \{a_1, \dots, a_\ell\}) + b_2 \cdot \textsc{reshape}(\mathcal{X}_2, \{a_1, \dots, a_\ell\}). \end{aligned}$$

Moreover, for two sets $\{a_1, \dots, a_\ell\}, \{b_1, \dots, b_g\}$ such that $a_\ell < b_1$ (i.e., all elements in the first set are less than those in the second), it holds that

$$\begin{aligned} &\textsc{reshape}(\mathcal{X}, \{a_1, \dots, a_\ell, b_1, \dots, b_g\}) \\ &= \textsc{reshape}\left\{\textsc{reshape}(\mathcal{X}, \{b_1, \dots, b_g\}), \{a_1, \dots, a_\ell, K-g+1\}\right\}, \end{aligned}$$

where $\{a_1, \dots, a_\ell, K-g+1\}$ is indeed strictly ascending since

$$a_\ell \le b_1 - 1 \le b_g - (g-1) - 1 = b_g - g \le K - g.$$

## 4.3 A Factor Model and Kronecker Product Structure Test

### 4.3.1 Factor models and Kronecker product structure

This subsection introduces the concept of factor models with Kronecker product structure and lays down the technical details for the testing problem. For an integral reading experience, readers can go straight to Section 4.3.2 where equations and terms can be referred back to

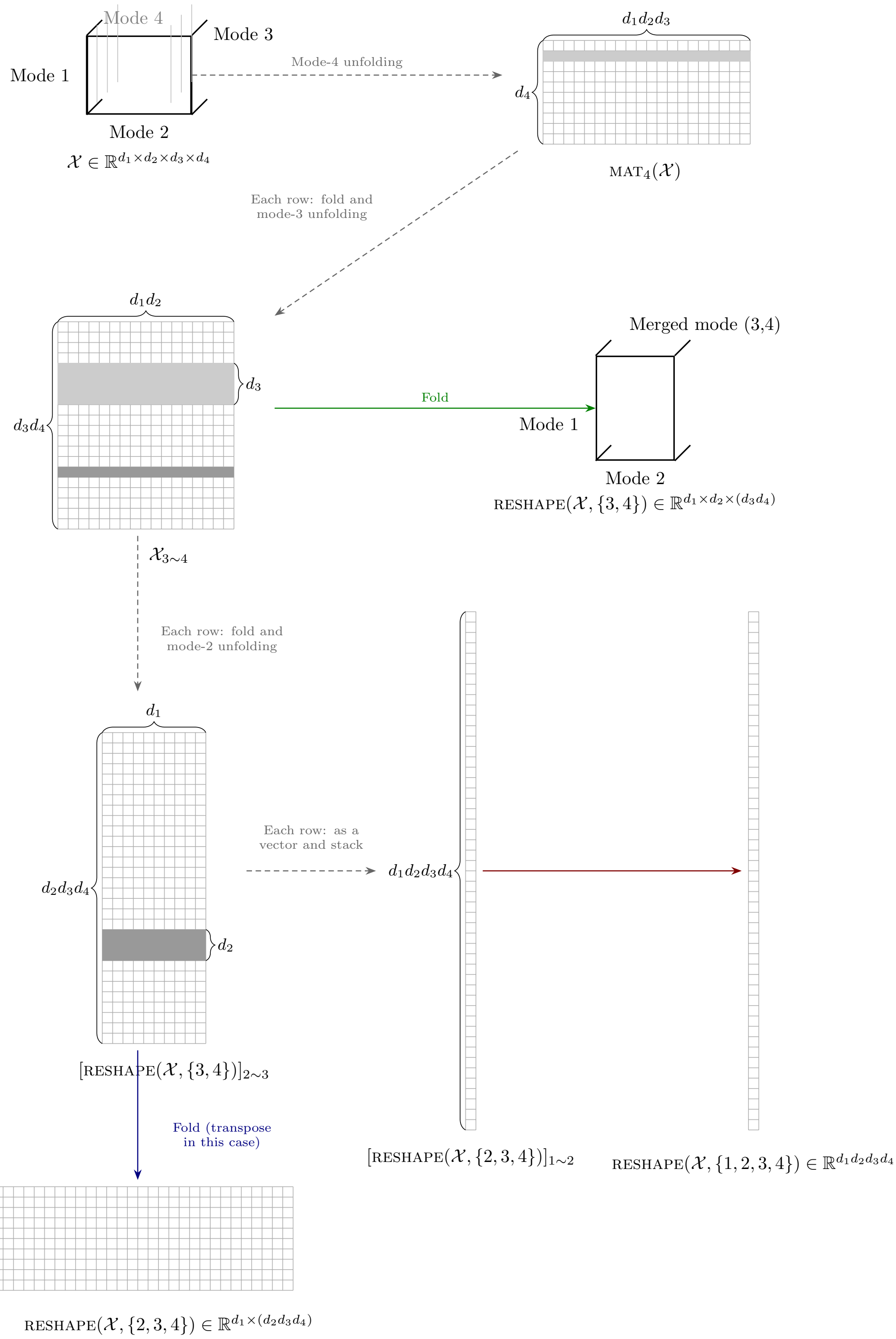


Figure 4.1: Illustration of the reshape operator for an order-4 tensor $\mathcal{X}$ along $\mathcal{A}$. The last step in the reshape with $\mathcal{A} = \{3,4\}$, $\{2,3,4\}$, and $\{1,2,3,4\}$, are respectively denoted by green, blue, and red arrows.

Section 4.3.1 whenever necessary. We begin by introducing the Kronecker product structure set which facilitates description of our models.

**Definition 4.1** (Kronecker product structure set). *Given an ordered set of positive integers* $\{b_1, \ldots, b_\kappa\}$*, the Kronecker product structure set is defined as*

$$\mathcal{K}_{b_1 \times \cdots \times b_\kappa} := \big\{ \mathbf{A} \mid \mathbf{A} = \mathbf{A}_\kappa \otimes \cdots \otimes \mathbf{A}_1 \\ \text{with } \mathbf{A}_j \in \mathbb{R}^{b_j \times u_j} \text{ of finite rank } u_j \le b_j,\ \|\mathbf{A}_{j,\cdot i}\|^2 \asymp b_j^{\delta_{j,i}},\ \delta_{j,i} \in (0,1] \big\}.$$

The Kronecker product structure set defined by Definition 4.1 characterises the factor loading matrix, and requiring $\delta_{j,i} > 0$ is to ensure certain factor strength in each loading matrix. See Assumptions (L1) and (L2) in Section 4.4.1 for the technical details. The form of factor models is depicted below, with the feature of Kronecker product structure.

**Definition 4.2** (Factor models and Kronecker product structure). *Given a series of mean-zero order-$K$ tensors $\mathcal{Y}_t \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ for $t \in [T]$ and a set with ordered, ascending elements $\mathcal{A} = \{a_1, \ldots, a_\ell\} \subseteq [K]$, we say $\{\mathcal{Y}_t\}$ follows a factor model along $\mathcal{A}$ if for $t \in [T]$,*

$$\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}) = \mathcal{C}_{\text{reshape},t} + \mathcal{E}_{\text{reshape},t} = \mathcal{F}_{\text{reshape},t} \times_{j=1}^{K-\ell+1} \mathbf{A}_{\text{reshape},j} + \mathcal{E}_{\text{reshape},t}, \tag{4.3}$$

*where $\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}) \in \mathbb{R}^{p_1 \times \cdots \times p_{K-\ell+1}}$ (for some $p_1, \ldots, p_{K-\ell+1}$) is the order-$(K-\ell+1)$ tensor by reshaping $\mathcal{Y}_t$ along $\mathcal{A}$, the common component $\mathcal{C}_{\text{reshape},t}$ consists of the core factor $\mathcal{F}_{\text{reshape},t} \in \mathbb{R}^{\pi_1 \times \cdots \times \pi_{K-\ell+1}}$ and loading matrices $\mathbf{A}_{\text{reshape},j} \in \mathbb{R}^{p_j \times \pi_j}$ with finite rank $\pi_j \le p_j$ for $j \in [K-\ell+1]$, and $\mathcal{E}_{\text{reshape},t}$ is the noise. We further make the following classifications.*

1. *$\{\mathcal{Y}_t\}$ has a Kronecker product structure if $\mathbf{A}_{\text{reshape},K-\ell+1} \in \mathcal{K}_{d_{a_1} \times \cdots \times d_{a_\ell}}$;*

2. *$\{\mathcal{Y}_t\}$ has no Kronecker product structure along $\mathcal{A}$ if $\mathbf{A}_{\text{reshape},K-\ell+1} \notin \mathcal{K}_{d_{a_1} \times \cdots \times d_{a_\ell}}$.*

Definition 4.2 formally defines the form of factor models considered in this chapter. A key information lying in Definition 4.2.1 is that if the Kronecker product structure holds along some $\mathcal{A}$, the structure holds along any $\mathcal{A}$; see the discussion below Theorem 4.1 for details. Note that if $\ell = 1$ in Definition 4.2, i.e., $\mathcal{A}$ contains only one element (representing the mode index), for each order-$K$ tensor $\mathcal{Y}_t$, $\textsc{reshape}(\mathcal{Y}_t, \{a_1\})$ is the order-$K$ tensor constructed from $\mathcal{Y}_t$ by treating mode-$a_1$ as mode-$K$. Hence, the factor model of $\mathcal{Y}_t$ along $\{a_1\}$ returns to a Tucker-decomposition TFM (Chen et al., 2022a; Barigozzi et al., 2023b) of $\mathcal{Y}_t$ with mode indices changed. For instance, we may read (4.3) along $\mathcal{A} = \{K\}$ as

$$\mathcal{Y}_t = \mathcal{C}_{\text{reshape},t} + \mathcal{E}_{\text{reshape},t} = \mathcal{F}_{\text{reshape},t} \times_1 \mathbf{A}_{\text{reshape},1} \times_2 \cdots \times_K \mathbf{A}_{\text{reshape},K} + \mathcal{E}_{\text{reshape},t}.$$

Hence Definition 4.2.1 automatically describes $\{\mathcal{Y}_t\}$ if $\ell = 1$, implying that Kronecker product structure is only non-trivial for $\ell \geq 2$ (hence $K \geq 2$). To demystify Definition 4.2.1, we next present Theorem 4.1 which, as a first in the literature, spells out the equivalence of Tucker-decomposition TFM under tensor reshape.

**Theorem 4.1** (Tensor Reshape Theorem I). *With the notations in Definition 4.2, $\{\mathcal{Y}_t\}$ following* (4.3) *along any given $\mathcal{A} = \{a_1, \dots, a_\ell\} \subseteq [K]$ with a Kronecker product structure is equivalent to $\{\mathcal{Y}_t\}$ following a Tucker-decomposition factor model such that*

$$\mathcal{Y}_t = \mathcal{C}_t + \mathcal{E}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \cdots \times_K \mathbf{A}_K + \mathcal{E}_t, \tag{4.4}$$

*where $\mathcal{C}_t$ is the common component, $\mathcal{F}_t \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is the core factor, each $\mathbf{A}_k \in \mathbb{R}^{d_k \times r_k}$ is the mode-$k$ loading matrix, and $\mathcal{E}_t$ is the noise. More importantly, with $\mathcal{A}^* := [K] \setminus \mathcal{A}$,*

$$\mathcal{F}_{\text{reshape},t} = \text{RESHAPE}(\mathcal{F}_t, \mathcal{A}), \quad \mathcal{E}_{\text{reshape},t} = \text{RESHAPE}(\mathcal{E}_t, \mathcal{A}),$$
$$\mathbf{A}_{\text{reshape},K-\ell+1} = \otimes_{i \in \mathcal{A}} \mathbf{A}_i, \quad \mathbf{A}_{\text{reshape},j} = \mathbf{A}_{\mathcal{A}^*_j} \quad \textit{for } j \in [K - \ell].$$

*Moreover, the model* (4.4) *uniquely determines parameters in* (4.3)*, and* (4.3) *determines those in* (4.4) *up to an arbitrary set $\{\mathbf{A}_i\}_{i \in \mathcal{A}}$.*

Theorem 4.1 reveals that a factor model on $\{\mathcal{Y}_t\}$ with Kronecker product structure in Definition 4.2 is in fact a Tucker-decomposition TFM on $\{\mathcal{Y}_t\}$. This forms the foundation for the hypothesis test design later. The identification of (4.3) and (4.4) are relegated to Section 4.6.

**Remark 4.1** *Both* (4.3) *and* (4.4) *are based on a Tucker decomposition for the observed tensor. Other tensor decompositions are possible, such as the CP decomposition (Kolda and Bader, 2009) and the Low Separation Rank (LSR) decomposition (Taki et al., 2024), etc. As CP decomposition is a special Tucker decomposition, our defined factor model is more general. The LSR decomposition is generalised further from Tucker decomposition, but the structure is less helpful here and brings in unnecessary complication due to the arbitrary separation rank.*

### 4.3.2 A test on Kronecker product structure

The testing problem on Kronecker product structure is formally defined in this subsection, with an example on an order-2 tensor (i.e., a matrix) time series given at the end. For each $t \in [T]$, we observe a mean-zero order-$K$ tensor $\mathcal{Y}_t \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with $K \geq 2$ (otherwise the test is trivial as explained in Section 4.3.1). Without loss of generality, let $v < K$ be a given positive integer and denote $\mathcal{A} = \{v, v+1, \dots, K-1, K\}$ which contains the mode indices along which the Kronecker product structure might be lost; see the alternative hypothesis $H_1$ below.

Suppose $\{\mathcal{Y}_t\}$ follows a factor model along $\mathcal{A}$ as in Definition 4.2, with notations therein except that we now read (4.3) as

$$\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}) = \mathcal{C}_{\text{reshape},t} + \mathcal{E}_{\text{reshape},t} = \mathcal{F}_{\text{reshape},t} \times_{j=1}^{v-1} \mathbf{A}_j \times_v \mathbf{A}_V + \mathcal{E}_{\text{reshape},t}, \tag{4.5}$$

where $\mathbf{A}_j \in \mathbb{R}^{d_j \times r_j}$ for $j \in [v-1]$ (if $v > 1$) and $\mathbf{A}_V \in \mathbb{R}^{d_V \times r_V}$ with $d_V := \prod_{i=v}^K d_i$. Essentially, the order-$v$ tensor $\textsc{reshape}(\mathcal{Y}_t, \mathcal{A})$ follows a Tucker-decomposition TFM. The set $\{r_1, \ldots, r_{v-1}, r_V\}$ is assumed known and any consistent estimators (e.g. Han et al., 2022; Chen and Lam, 2024b) can be used in practice. With $\mathcal{K}_{d_v \times \cdots \times d_K}$ defined in Definition 4.1, we consider a hypothesis test as follows:

$$\begin{aligned} &H_0 : \{\mathcal{Y}_t\} \text{ has a Kronecker product structure, i.e., } \mathbf{A}_V \in \mathcal{K}_{d_v \times \cdots \times d_K}; \\ &H_1 : \{\mathcal{Y}_t\} \text{ has no Kronecker product structure along } \mathcal{A}\text{, i.e., } \mathbf{A}_V \notin \mathcal{K}_{d_v \times \cdots \times d_K}. \end{aligned} \tag{4.6}$$

Besides the complexity of being a composite testing problem, the difficulty of (4.6) is elevated by the fact that $\mathcal{Y}_t$ under the alternative has no explicit form without reshaping along $\mathcal{A}$. Fortunately, the factor structure in (4.5) is stable under both hypotheses. That is, the estimation of $\{\mathcal{F}_{\text{reshape},t}, \mathbf{A}_1, \ldots, \mathbf{A}_{v-1}, \mathbf{A}_V\}$ is always feasible. In particular, thanks to Theorem 4.1, we have the following under $H_0$:

$$\mathcal{Y}_t = \mathcal{C}_t + \mathcal{E}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \cdots \times_{v-1} \mathbf{A}_{v-1} \times_v \mathbf{A}_v \times_{v+1} \cdots \times_K \mathbf{A}_K + \mathcal{E}_t, \tag{4.7}$$

where $\mathbf{A}_k \in \mathbb{R}^{d_k \times r_k}$ for $k \in [K]$ (hence the first $v-1$ loading matrices are exactly those in (4.5)), and that

$$\textsc{reshape}(\mathcal{F}_t, \mathcal{A}) = \mathcal{F}_{\text{reshape},t}, \quad \textsc{reshape}(\mathcal{E}_t, \mathcal{A}) = \mathcal{E}_{\text{reshape},t}, \quad \mathbf{A}_K \otimes \mathbf{A}_{K-1} \otimes \cdots \otimes \mathbf{A}_v = \mathbf{A}_V.$$

**Example 4.1** *Let $\mathbf{Y}_t \in \mathbb{R}^{d_1 \times d_2}$ ($t \in [T]$) be matrix-valued observations. For the setup, we can only specify $\mathcal{A} = \{1, 2\}$ (which is the only non-trivial case here as discussed in Section 4.3.1). The hypothesis test* (4.6) *is simplified as follows, with $\mathcal{A}$ reflected by the vectorisation:*

$$\begin{aligned} &H_0 : \mathbf{Y}_t = \mathbf{A}_1 \mathbf{F}_t \mathbf{A}_2' + \mathbf{E}_t; \\ &H_1 : \mathbf{vec}(\mathbf{Y}_t) = \mathbf{A}_V \mathbf{vec}(\mathbf{F}_t) + \mathbf{vec}(\mathbf{E}_t), \quad \textit{with } \mathbf{A}_V \notin \mathcal{K}_{d_1 \times d_2}. \end{aligned}$$

### 4.3.3 Constructing the test statistic

Despite the obscure $\mathcal{K}_{d_v \times \cdots \times d_K}$ in (4.6), we may resort to the Tucker-decomposition TFM in (4.7) under $H_0$. To construct the test, we first obtain estimators for the (standardised) loading matrices in (4.5). For $j \in [v-1]$, $\widetilde{\mathbf{Q}}_j$ is defined as the eigenvector matrix corresponding to the

$r_j$ largest eigenvalues of

$$\frac{1}{T}\sum_{t=1}^{T} \text{RESHAPE}(\mathcal{Y}_t, \mathcal{A})_{(j)} \, \text{RESHAPE}(\mathcal{Y}_t, \mathcal{A})'_{(j)},$$

where $\text{RESHAPE}(\mathcal{Y}_t, \mathcal{A})_{(j)}$ represents the mode-$j$ unfolding matrix of $\text{RESHAPE}(\mathcal{Y}_t, \mathcal{A})$. Similarly, $\widetilde{\mathbf{Q}}_V$ is the eigenvector matrix corresponding to the $r_V$ largest eigenvalues of

$$\frac{1}{T}\sum_{t=1}^{T} \text{RESHAPE}(\mathcal{Y}_t, \mathcal{A})_{(v)} \, \text{RESHAPE}(\mathcal{Y}_t, \mathcal{A})'_{(v)}.$$

Then $\mathcal{C}_{\text{reshape},t}$ and $\mathcal{E}_{\text{reshape},t}$ are respectively estimated by

$$\widetilde{\mathcal{C}}_{\text{reshape},t} := \text{RESHAPE}(\mathcal{Y}_t, \mathcal{A}) \times_{j=1}^{v-1} (\widetilde{\mathbf{Q}}_j \widetilde{\mathbf{Q}}_j') \times_v (\widetilde{\mathbf{Q}}_V \widetilde{\mathbf{Q}}_V'), \tag{4.8}$$

$$\widetilde{\mathcal{E}}_{\text{reshape},t} := \text{RESHAPE}(\mathcal{Y}_t, \mathcal{A}) - \widetilde{\mathcal{C}}_{\text{reshape},t}. \tag{4.9}$$

For (4.7), $\widehat{\mathbf{Q}}_j$ for $j \in [v-1]$ is defined as the eigenvector matrix corresponding to the $r_j$ largest eigenvalues of $T^{-1}\sum_{t=1}^{T} \mathbf{Y}_{t,(j)}\mathbf{Y}'_{t,(j)}$. Next, denote $\mathcal{R}$ as the set of all divisor combinations of $r_V$, i.e.,

$$\mathcal{R} := \Big\{(\pi_1, \pi_2, \ldots, \pi_{K-v+1}) \,\Big|\, \prod_{j=1}^{K-v+1} \pi_j = r_V \text{ with each } \pi_j \in \mathbb{Z}^+, \pi_j \le d_{j+v-1}\Big\}. \tag{4.10}$$

Let the $m$-th element of $\mathcal{R}$ be $(\pi_{m,1}, \ldots, \pi_{m,K-v+1})$. Then for $i \in \{v, v+1, \ldots, K\}$, we obtain $\widehat{\mathbf{Q}}_{m,i}$ as the eigenvector matrix corresponding to the $\pi_{m,i-v+1}$ largest eigenvalues of $T^{-1}\sum_{t=1}^{T} \mathbf{Y}_{t,(i)}\mathbf{Y}'_{t,(i)}$. The common component and residual estimators are hence obtained as

$$\widehat{\mathcal{C}}_{m,t} := \mathcal{Y}_t \times_{j=1}^{v-1} (\widehat{\mathbf{Q}}_j \widehat{\mathbf{Q}}_j') \times_{i=v}^{K} (\widehat{\mathbf{Q}}_{m,i} \widehat{\mathbf{Q}}_{m,i}'), \tag{4.11}$$

$$\widehat{\mathcal{E}}_{m,t} := \mathcal{Y}_t - \widehat{\mathcal{C}}_{m,t}. \tag{4.12}$$

Let $\widetilde{\mathcal{E}}_t$ be the order-$K$ tensor with the same dimension as $\mathcal{Y}_t$ such that $\text{RESHAPE}(\widetilde{\mathcal{E}}_t, \mathcal{A}) = \widetilde{\mathcal{E}}_{\text{reshape},t}$. Define $k^* := \arg\min_{k\in[K]}\{d_k\}$ and denote the mode-$k^*$ unfolding of $\widetilde{\mathcal{E}}_t$ and $\widehat{\mathcal{E}}_{m,t}$ as $\widetilde{\mathbf{E}}_{t,(k^*)}$ and $\widehat{\mathbf{E}}_{m,t,(k^*)}$, respectively. Theorem 4.2 (in Section 4.4.2) tells us that there exists $m \in [|\mathcal{R}|]$ such that for each $t \in [T]$, $j \in [d/d_{k^*}]$, both

$$x_{j,t} := \frac{1}{d_{k^*}}\sum_{i=1}^{d_{k^*}} \widetilde{E}^2_{t,(k^*),ij}, \quad y_{m,j,t} := \frac{1}{d_{k^*}}\sum_{i=1}^{d_{k^*}} \widehat{E}^2_{m,t,(k^*),ij},$$

are asymptotically distributed the same under $H_0$, and $x_{j,t}$ in particular is distributed the same under either $H_0$ or $H_1$. Let $\mathbb{P}_{x,j}$ and $\mathbb{P}_{y,m,j}$ respectively denote the empirical probability measures induced by the empirical cumulative distribution functions for $\{x_{j,t}\}_{t\in[T]}$ and $\{y_{m,j,t}\}_{t\in[T]}$:

$$\mathbb{F}_{x,j}(c) := \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{x_{j,t} \le c\}, \quad \mathbb{F}_{y,m,j}(c) := \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{y_{m,j,t} \le c\}. \tag{4.13}$$

Let $\widehat{q}_{x,j}(\alpha) := \inf\big\{c \mid \mathbb{F}_{x,j}(c) \ge 1-\alpha\big\}$. The intuition here is that if $H_0$ is satisfied, then over different $j \in [d/d_{k^*}]$, the cumulative distribution functions $\mathbb{F}_{x,j}(\cdot)$ and $\mathbb{F}_{y,m,j}(\cdot)$ should be similar. However, if $H_1$ is true, then we expect the residuals in $\widehat{\mathbf{E}}_{m,t,(k^*)}$ to be inflated, so that $\mathbb{P}_{y,m,j}\{y_{m,j,t} \ge \widehat{q}_{x,j}(\alpha)\}$ is expected to be larger than $\alpha$; see the theoretical statement in Theorem 4.3. To incorporate it across different $j \in [d/d_{k^*}]$, we compare the 5% quantile of $T^{-1}\sum_{t=1}^{T}\mathbb{1}\{y_{m,j,t} \ge \widehat{q}_{x,j}(\alpha)\}$ over $j \in [d/d_{k^*}]$ to $\alpha$, and expect it to be larger than $\alpha$ under $H_1$.

Since with the wrong number of factors, a particular $m \in [|\mathcal{R}|]$ will in general inflate the residuals $y_{m,j,t}$ further, in practice, to be on the conservative side, we reject $H_0$ if

$$\min_{m\in[|\mathcal{R}|]} \Big\{ 5\% \text{ quantile of } \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{y_{m,j,t} \ge \widehat{q}_{x,j}(\alpha)\} \text{ over } j \in [d/d_{k^*}] \Big\} > \alpha, \tag{4.14}$$

noting that exactly one element in $\mathcal{R}$ represents the true number of factors on the modes with indices in $\mathcal{A}$. We also point out that there are other possible ways to aggregate the information from each $j$, but (4.14) empirically works well and circumvents possible issues such as heavy-tailed noise, under- or over-estimation on the number of factors, and insufficient data dimensions; see Section 4.5.1.

**Remark 4.2** *(Explanation of $\mathcal{R}$ in* (4.10)*). It is possible to perform the test directly using the number of factors for modes in $\mathcal{A}$, i.e., $\mathcal{R}$ only contains the number of factors $r_j$, $j = v, \ldots, K$, in* (4.7)*. This is guaranteed by the Tucker-decomposition TFM under $H_0$ in* (4.6)*. However, usually in practice we need to estimate the number of factors which are invalid under $H_1$ in* (4.6)*. This leads to unstable estimated number of factors and hence unstable test statistic, which we address by introducing $\mathcal{R}$ in* (4.10).

## 4.4 Assumptions and Theoretical Results

### 4.4.1 Assumptions

This subsection presents all the assumptions for testing $H_0$ against $H_1$ in (4.6). Another version (with only different notations) of Assumptions (L1) and (L2) for the identification of (4.3) and (4.4) is included in Section 4.6, with identification theorem presented and proved there.

(L1) *For each $j \in [v-1]$, we assume that $\mathbf{A}_j$ in* (4.5) *is of full rank and as $d_j \to \infty$,*

$$\mathbf{Z}_j^{-1/2}\mathbf{A}_j'\mathbf{A}_j\mathbf{Z}_j^{-1/2} \to \mathbf{\Sigma}_{A,j}, \tag{4.15}$$

*where $\mathbf{\Sigma}_{A,j}$ is positive definite with all eigenvalues bounded away from 0 and infinity, and $\mathbf{Z}_j$ is a diagonal matrix with $(\mathbf{Z}_j)_{hh} \asymp d_j^{\delta_{j,h}}$ for $h \in [r_j]$ and the ordered factor strengths $1/2 < \delta_{j,r_j} \le \cdots \le \delta_{j,1} \le 1$.*

*We assume that $\mathbf{A}_V$ also has the above form with $\mathbf{Z}_V$ and $\mathbf{\Sigma}_{A,V}$, except that only the maximum and minimum factor strengths are ordered, i.e., $1/2 < \delta_{V,r_V} \le \delta_{V,h} \le \delta_{V,1} \le 1$ for any $h \in [r_V]$.*

(L2) *With $\mathcal{A} = \{v, v+1, \dots, K\}$, we assume that for each $i \in \mathcal{A}$, $\mathbf{A}_i$ in* (4.7) *is of full rank and as $d_i \to \infty$,*

$$\mathbf{Z}_i^{-1/2}\mathbf{A}_i'\mathbf{A}_i\mathbf{Z}_i^{-1/2} \to \mathbf{\Sigma}_{A,i}, \tag{4.16}$$

*where $\mathbf{\Sigma}_{A,i}$ is positive definite with all eigenvalues bounded away from 0 and infinity, and $\mathbf{Z}_i$ is a diagonal matrix with $(\mathbf{Z}_i)_{hh} \asymp d_i^{\delta_{i,h}}$ for $h \in [r_i]$ and the ordered factor strengths $1/2 < \delta_{i,r_i} \le \cdots \le \delta_{i,1} \le 1$.*

(F1) (Time series in $\mathcal{F}_{\text{reshape},t}$). *There is $\mathcal{X}_{\text{reshape},f,t}$ the same dimension as $\mathcal{F}_{\text{reshape},t}$ such that $\mathcal{F}_{\text{reshape},t} = \sum_{w\ge 0} a_{f,w}\mathcal{X}_{\text{reshape},f,t-w}$. The time series $\{\mathcal{X}_{\text{reshape},f,t}\}$ has i.i.d. elements with mean* 0*, variance* 1 *and uniformly bounded fourth order moments. The coefficients $a_{f,w}$ satisfy $\sum_{w\ge 0} a_{f,w}^2 = 1$ and $\sum_{w\ge 0} |a_{f,w}| \le c$ for some constant c.*

(E1) (Decomposition of $\mathcal{E}_t$). *The noise $\mathcal{E}_t$ (such that $\mathcal{E}_{\text{reshape},t} = \text{RESHAPE}(\mathcal{E}_t, \mathcal{A})$) can be decomposed as*

$$\mathcal{E}_t = \mathcal{F}_{e,t} \times_1 \mathbf{A}_{e,1} \times_2 \cdots \times_K \mathbf{A}_{e,K} + \mathbf{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t, \tag{4.17}$$

*where order-$K$ tensors $\mathcal{F}_{e,t} \in \mathbb{R}^{r_{e,1}\times\cdots\times r_{e,K}}$ and $\boldsymbol{\epsilon}_t \in \mathbb{R}^{d_1\times\cdots\times d_K}$ contain independent mean zero elements with unit variance, with the two time series $\{\boldsymbol{\epsilon}_t\}$ and $\{\mathcal{F}_{e,t}\}$ being independent. The order-$K$ tensor $\mathbf{\Sigma}_\epsilon$ contains the standard deviations of the corresponding elements in $\boldsymbol{\epsilon}_t$, and has elements uniformly bounded.*

*Moreover, $\mathbf{A}_{e,k} \in \mathbb{R}^{d_k\times r_{e,k}}$ ($k \in [K]$) is approximately sparse such that $\|\mathbf{A}_{e,k}\|_1 = O(1)$.*

(E2) (Time series in $\mathcal{E}_t$). *There is $\mathcal{X}_{e,t}$ the same dimension as $\mathcal{F}_{e,t}$, and $\mathcal{X}_{\epsilon,t}$ the same dimension as $\boldsymbol{\epsilon}_t$, such that $\mathcal{F}_{e,t} = \sum_{q\ge 0} a_{e,q}\mathcal{X}_{e,t-q}$ and $\boldsymbol{\epsilon}_t = \sum_{q\ge 0} a_{\epsilon,q}\mathcal{X}_{\epsilon,t-q}$, with $\{\mathcal{X}_{e,t}\}$ and $\{\mathcal{X}_{\epsilon,t}\}$ independent of each other. $\{\mathcal{X}_{e,t}\}$ has independent elements while $\{\mathcal{X}_{\epsilon,t}\}$ has i.i.d. elements, and all elements have mean zero with unit variance and uniformly bounded fourth order moments. Both $\{\mathcal{X}_{e,t}\}$ and $\{\mathcal{X}_{\epsilon,t}\}$ are independent of $\{\mathcal{X}_{\text{reshape},f,t}\}$ from (F1).*

*The coefficients* $a_{e,q}$ *and* $a_{\epsilon,t}$ *are such that for some constant* $c$*,*

$$\sum_{q\geq 0} a_{e,q}^2 = \sum_{q\geq 0} a_{\epsilon,q}^2 = 1, \quad \sum_{q\geq 0} |a_{e,q}|, \sum_{q\geq 0} |a_{\epsilon,q}| \leq c.$$

(R1) (Rate assumptions). *With* $g_s := \prod_{k=1}^{K} d_k^{\delta_{k,1}}$ *and* $\gamma_s := d_V^{\delta_{V,1}} \prod_{j=1}^{v-1} d_j^{\delta_{j,1}}$*, we assume that*

$$dg_s^{-2}T^{-1}d_k^{2(\delta_{k,1}-\delta_{k,r_k})+1}, \quad dg_s^{-1}d_k^{\delta_{k,1}-\delta_{k,r_k}-1/2},$$
$$d\gamma_s^{-2}T^{-1}d_V^{2(\delta_{V,1}-\delta_{V,r_V})+1}, \quad d\gamma_s^{-1}d_V^{\delta_{V,1}-\delta_{V,r_V}-1/2} = o(1).$$

(R2) (Further rate assumptions). *With* $g_w := \prod_{k=1}^{K} d_k^{\delta_{k,r_k}}$ *and* $\gamma_w := d_V^{\delta_{V,r_V}} \prod_{j=1}^{v-1} d_j^{\delta_{j,r_j}}$*, we assume that*

$$\max_{k\in[K]} \left\{ d_k^{2(\delta_{k,1}-\delta_{k,r_k})} \left( \frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}} \right) \frac{d^2}{g_s g_w} \right\},$$
$$\max_{j\in[v-1]} \left\{ d_j^{2(\delta_{j,1}-\delta_{j,r_j})} \left( \frac{1}{Td_{\text{-}k}d_j^{1-\delta_{j,1}}} + \frac{1}{d_j^{1+\delta_{j,r_j}}} \right) \frac{d^2}{\gamma_s \gamma_w} \right\},$$
$$d_V^{2(\delta_{V,1}-\delta_{V,r_V})} \left( \frac{1}{Tdd_V^{-\delta_{V,1}}} + \frac{1}{d_V^{1+\delta_{V,r_V}}} \right) \frac{d^2}{\gamma_s\gamma_w}, \quad \frac{d}{\gamma_w^2}, \quad \frac{d}{g_w^2} = o\Big( \max_{k\in[K]} \{d_k^{-1}\} \Big).$$

With (L1), the standardised loading matrix $\mathbf{Q}_j := \mathbf{A}_j\mathbf{Z}_j^{-1/2}$ satisfies $\mathbf{Q}_j'\mathbf{Q}_j \to \mathbf{\Sigma}_{A,j}$ for $j \in [v-1]$, and $\mathbf{Q}_V := \mathbf{A}_V\mathbf{Z}_V^{-1/2}$ satisfies $\mathbf{Q}_V'\mathbf{Q}_V \to \mathbf{\Sigma}_{A,V}$. Similar implication holds for (L2), except that (L2) is only valid under $H_0$. Hence with (L2), $\mathbf{Z}_V$ and $\mathbf{\Sigma}_{A,V}$ in (L1) satisfy

$$\mathbf{Z}_V = \mathbf{Z}_K \otimes \cdots \otimes \mathbf{Z}_v, \quad \mathbf{\Sigma}_{A,V} = \mathbf{\Sigma}_{A,K} \otimes \cdots \otimes \mathbf{\Sigma}_{A,v}. \tag{4.18}$$

Note that the factor strength requirement for $\mathbf{A}_V$ in (L1) is satisfied by (L2), since from (4.18) $(\mathbf{Z}_V)_{r_V r_V} \asymp \prod_{i=v}^{K} d_i^{\delta_{i,r_i}} \geq d_V^{\min_{i=v}^{K}\delta_{i,r_i}} > d_V^{1/2}$. Assumption (L1) characterises the loading matrix behaviour generally for (4.6), and the additional (L2) is specific for the null. Both assumptions allow for weak factors which are common feature in the literature (Lam and Yao, 2012; Onatski, 2012; Cen and Lam, 2025b). When all factors are pervasive, for instance, (4.15) can be interpreted as $d_j^{-1}\mathbf{A}_j'\mathbf{A}_j \to \mathbf{\Sigma}_{A,j}$ if all factors are pervasive, which coincides with Assumption 3 of Chen and Fan (2023) for matrix time series.

Assumption (F1) assumes that $\mathcal{F}_{\text{reshape},t}$ is a general linear process with weakly serial dependence. Theorem 4.1 ensures that the core factor in (4.7) (under $H_0$) reserves its structure of

(F1) such that

$$\mathcal{F}_t = \sum_{w \geq 0} a_{f,w} \mathcal{X}_{f,t-w}, \text{ with } \textsc{reshape}(\mathcal{X}_{f,t}, \mathcal{A}) = \mathcal{X}_{\text{reshape},f,t}. \tag{4.19}$$

Note that it holds for each $k \in [K]$, as $T \to \infty$,

$$\frac{1}{T} \sum_{t=1}^{T} \text{mat}_k(\mathcal{F}_t) \, \text{mat}_k(\mathcal{F}_t)' \xrightarrow{p} r_{\text{-}k} \, \mathbf{I}_{r_k}, \tag{4.20}$$

which is direct from Proposition 1.3 in the supplement of Cen and Lam (2025b). In comparison, Barigozzi et al. (2023b) assumes the form of (4.20) with $r_{\text{-}k} \, \mathbf{I}_{r_k}$ replaced by a positive definite matrix. This does not imply (F1) is particularly stronger as our factor loading matrices already incorporate some positive definite matrices by (L1) and (L2).

Assumptions (E1) and (E2) depict a general noise time series on the factor models (4.5) and (4.7). It is worth noting that the noise tensor $\mathcal{E}_t$ is allowed to be (weakly) dependent across modes and time, regardless of the existence of Kronecker product structure. From (4.17),

$$\begin{aligned} \textsc{reshape}(\mathcal{E}_t, \mathcal{A}) = \textsc{reshape}(\mathcal{F}_{e,t}, \mathcal{A}) \times_1 \mathbf{A}_{e,1} \times_2 \cdots \times_{v-1} \mathbf{A}_{e,v-1} \times_v (\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_v) \\ + \textsc{reshape}(\boldsymbol{\Sigma}_\epsilon, \mathcal{A}) * \textsc{reshape}(\boldsymbol{\epsilon}_t, \mathcal{A}), \end{aligned} \tag{4.21}$$

so that the structure of (E1) and (E2) are preserved by $\textsc{reshape}(\mathcal{E}_t, \mathcal{A})$. Assumption (R1) details the rate assumptions on factor strengths and is hence satisfied automatically when all factors are pervasive. Assumption (R2) also concerns factor strength and would hold for all strong factors if $v > 1$; for $v = 1$, (R2) holds when $\min_{k \in [K]} d_k = o(T)$ in addition to strong factors.

**Remark 4.3** *When $v = 1$ and all factors are strong, (R2) requires $d_{k*} = \min_{k \in [K]} d_k = o(T)$ which seems restricted. This is to ensure the asymptotic normality when we aggregate $d_{k*}$ number of estimated residuals in $x_{j,t}$ and $y_{m,j,t}$ in Section 4.3.3. However, from the proof of Theorem 4.2 in Section 4.7, it is feasible to aggregate $d_{k*}^{\beta}$ for any $0 < \beta < 2$ such that $d_{k*}^{\beta} = o(T)$. Therefore, (R2) is arguably as mild as Assumption B5 in He et al. (2023a). We only briefly discuss how to construct the test statistic differently in the following, and choose not to pursue such an aggregation scheme to keep the practical procedure as simple as possible.*

*Suppose we follow the same procedure in Section 4.3.3 with $\widetilde{\mathcal{E}}_t$ and $\widehat{\mathcal{E}}_{m,t}$ obtained. Next, we need to specify some $d^\dagger \to \infty$ that divides $d$ and is small enough (such that $\rho = o(1)$ in Theorem 4.3 with $d_{k^*}$ replaced by $d^\dagger$). With this, we simply re-arrange the two residual tensors and construct $\widetilde{\mathbf{E}}_t^\dagger = \textsc{fold}\{\textbf{vec}(\widetilde{\mathcal{E}}_t), \{d^\dagger, d/d^\dagger\}\}$ and $\widehat{\mathbf{E}}_{m,t}^\dagger = \textsc{fold}\{\textbf{vec}(\widehat{\mathcal{E}}_{m,t}), \{d^\dagger, d/d^\dagger\}\}$. Then the remaining procedure is the same as in Section 4.3.3 with $\widetilde{\mathbf{E}}_{t,(k^*)}$ and $\widehat{\mathbf{E}}_{m,t,(k^*)}$ replaced by $\widetilde{\mathbf{E}}_t^\dagger$ and $\widehat{\mathbf{E}}_{m,t}^\dagger$, respectively.*

### 4.4.2 Main results and practical test design

We first present below the results for our residual estimators in (4.9) and (4.12), which inspire the testing procedure in Section 4.3.3. Following Theorem 4.2, the theoretical guarantee of the test is also provided.

**Theorem 4.2** *Let Assumptions (F1), (L1), (L2), (E1), (E2), (R1) and (R2) hold. With the notations in Section 4.3.3, under $H_0$, there exists $m \in [|\mathcal{R}|]$ such that for each $t \in [T]$, $j \in [d/d_k^*]$,*

$$\frac{\sum_{i=1}^{d_{k^*}} \left(\widehat{E}^2_{m,t,(k^*),ij} - \Sigma^2_{\epsilon,(k^*),ij}\right)}{\sqrt{\sum_{i=1}^{d_{k^*}} \mathrm{Var}(\epsilon^2_{t,(k^*),ij})\Sigma^4_{\epsilon,(k^*),ij}}}, \quad \frac{\sum_{i=1}^{d_{k^*}} \left(\widetilde{E}^2_{t,(k^*),ij} - \Sigma^2_{\epsilon,(k^*),ij}\right)}{\sqrt{\sum_{i=1}^{d_{k^*}} \mathrm{Var}(\epsilon^2_{t,(k^*),ij})\Sigma^4_{\epsilon,(k^*),ij}}} \xrightarrow{p} Z_{j,t},$$

*where $Z_{j,t} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ and $Z_{h,t}$ is independent of $Z_{\ell,t}$ for $h \neq \ell$. Under $H_1$, the asymptotic result for $\widetilde{E}_{t,(k^*),ij}$ above still holds true.*

**Theorem 4.3** *Let all the assumptions in Theorem 4.2 hold. In addition, each element in the time series $\{\mathcal{X}_{\text{reshape},f,t}\}$, $\{\mathcal{X}_{e,t}\}$ and $\{\mathcal{X}_{\epsilon,t}\}$ has sub-Gaussian tail. With the notations in Section 4.3.3, we have the following for any $j \in [d/d_{k^*}]$ under $H_0$. There exists $m \in [|\mathcal{R}|]$ such that, as $T, d_1, \dots, d_K \to \infty$,*

$$\mathbb{P}_{y,m,j}\{y_{m,j,t} > \widehat{q}_{x,j}(\alpha)\} \le \alpha + O_P(\rho), \quad \textit{where}$$

$$\begin{aligned}\rho = &\left[\max_{k\in[K]} \left\{d_k^{\delta_{k,1}-\delta_{k,r_k}}\left(\frac{1}{(Td_{\text{-}k}d_k^{1-\delta_{k,1}})^{1/2}} + \frac{1}{d_k^{(1+\delta_{k,r_k})/2}}\right)\frac{d}{(g_s g_w)^{1/2}}\right\} + \frac{d^{1/2}}{g_w}\right] \\ &\cdot \log^2(T)\Big(\prod_{k=1}^K \log^2(d_k)\Big) d_{k^*}^{1/2} + \log^2(T)\log(d_V)\Big(\prod_{k=1}^{v-1}\log(d_k)\Big)\Big(\prod_{k=1}^K \log^2(d_k)\Big) d_{k^*}^{1/2} \\ &\cdot \left[\max_{j\in[v-1]} \left\{d_j^{\delta_{j,1}-\delta_{j,r_j}}\left(\frac{1}{(Td_{\text{-}k}d_j^{1-\delta_{j,1}})^{1/2}} + \frac{1}{d_j^{(1+\delta_{j,r_j})/2}}\right)\frac{d}{(\gamma_s\gamma_w)^{1/2}}\right\}\right. \\ &\left. + d_V^{\delta_{V,1}-\delta_{V,r_V}}\left(\frac{1}{(Tdd_V^{-\delta_{V,1}})^{1/2}} + \frac{1}{d_V^{(1+\delta_{V,r_V})/2}}\right)\frac{d}{(\gamma_s\gamma_w)^{1/2}} + \frac{d^{1/2}}{\gamma_w}\right].\end{aligned}$$

Theorem 4.3 suggests that if some factors are weaker, then the rate in the probability statement above will be inflated. When all factors are pervasive, define $d_{\max} = \max_{k\in[K]}\{d_k\}$, and we may simplify $\rho$ as

$$\rho = \left\{\frac{1}{d_{k^*}^{1/2}} + \Big(\frac{d_{k^*}d_{\max}}{Td}\Big)^{1/2} + \Big(\frac{d_{k^*}d_V}{Td}\Big)^{1/2}\log(d_V)\Big(\prod_{k=1}^{v-1}\log(d_k)\Big)\right\}\log^2(T)\Big(\prod_{k=1}^K \log^2(d_k)\Big).$$

Hence $\rho = o(1)$ as long as $T, d_1, \dots, d_K$ are of the same order, but it appears that when $d_V = d$, i.e., $\mathcal{A} = [K]$, the current test requires $d_{k^*} \log(d) \log^2(T) \prod_{k=1}^{K} \log^2(d_k) = o(T)$. However, this can be circumvented as explained in Remark 4.3. Theorem 4.3 presents the grounds for our construction of the test statistic in (4.14). For related explanations, see the discussions immediately after (4.13), and before Remark 4.2.

We also point out that the result in Theorem 4.3 holds exactly the same with all quantities constructed from $y_{m,j,t}$ and $x_{j,t}$ replaced by quantities from $\max_{j\in[d_{\text{-}k}]} y_{m,j,t}$ and $\max_{j\in[d_{\text{-}k}]} x_{j,t}$, respectively; see also Theorem 5.8 which shows such a result but on testing nested model structures in a matrix-valued time series.

The setup of the problem (4.6) specifies the set $\mathcal{A}$ which is only needed in $H_1$ due to (4.7) under $H_0$. It is direct to specify $\mathcal{A}$ for a series of matrix-valued observations (i.e., order-2 tensor), see Example 4.1. However, for a general order-$K$ tensor with $K \geq 3$, $\mathcal{A}$ might be misspecified without any prior knowledge. To resolve this, we present the second theorem on tensor reshape as follows.

**Theorem 4.4** (Tensor Reshape Theorem II). *Consider a tensor time series $\{\mathcal{Y}_t\}$ and a set of mode indices $\mathcal{A}$. With Definition 4.2, the time series $\{\text{RESHAPE}(\mathcal{Y}_t, \mathcal{A})\}$ has a Kronecker product structure if and only if $\{\mathcal{Y}_t\}$ either has a Kronecker product structure or has no Kronecker product structure along a subset of $\mathcal{A}$.*

Suppose now $\{\mathcal{Y}_t\}$ has no Kronecker product structure along some $\mathcal{A}^*$. Theorem 4.4 tells us that testing the Kronecker product structure of the reshaped series $\{\text{RESHAPE}(\mathcal{Y}_t, \mathcal{A})\}$ effectively tests if $\mathcal{A}^* \subseteq \mathcal{A}$. In light of this, a testing design is feasible when $\mathcal{A}$ is unspecified, with a minimal assumption that $\text{RESHAPE}(\mathcal{Y}_t, [K]) = \mathbf{vec}(\mathcal{Y}_t)$ has a factor structure, i.e., the vectorised $\mathcal{Y}_t$ follows a vector factor model. For illustration, consider $\text{RESHAPE}(\mathcal{Y}_t, [K] \setminus \{1\}) = \text{RESHAPE}(\mathcal{Y}_t, \{2, \ldots, K\})$ which is a matrix. Using the property of Reshape$(\cdot, \cdot)$, we have

$$\text{RESHAPE}\big(\text{RESHAPE}(\mathcal{Y}_t, \{2, \ldots, K\}), \{1, 2\}\big) = \text{RESHAPE}(\mathcal{Y}_t, \{1, 2, \ldots, K\}) = \mathbf{vec}(\mathcal{Y}_t).$$

According to Definition 4.2, $\{\text{RESHAPE}(\mathcal{Y}_t, \{2, \ldots, K\})\}$ follows a factor model along $\{1, 2\}$. This is always correctly specified since $\{\mathbf{vec}(\mathcal{Y}_t)\}$ follows a factor model (which also implies $\mathcal{A}^* \subseteq [K]$). By Theorem 4.4, $\text{RESHAPE}(\mathcal{Y}_t, \{2, \ldots, K\})$ has no Kronecker product structure if and only if $1 \in \mathcal{A}^*$. Hence on testing (4.6) with $\mathcal{Y}_t$ replaced by $\{\text{RESHAPE}(\mathcal{Y}_t, \{2, \ldots, K\})\}$ and $\mathcal{A} = \{1, 2\}$, rejection of the null implies $1 \in \mathcal{A}^*$.

By the fact that $\{\mathbf{vec}(\mathcal{Y}_t)\}$ with any permutation on $\mathbf{vec}(\mathcal{Y}_t)$ also follows a factor model, the above scheme is in fact valid on $\text{RESHAPE}(\mathcal{Y}_t, [K] \setminus \{k\})$ for any $k \in [K]$. Eventually, $\mathcal{A}^*$ can be identified, and the above procedure is summarised into the following algorithm.

Practical testing algorithm

1. Given an order-$K$ tensor time series $\{\mathcal{Y}_t\}$ with $K \geq 2$ and $\mathbf{vec}(\mathcal{Y}_t)$ following a factor model with $r_{\text{vec}}$ number of factors, initialise $\widehat{\mathcal{A}}^* = \phi$, the empty set.

2. Initialise $k = 1$. Define a test as (4.6) with $\{\mathcal{Y}_t\}$ replaced by $\{\textsc{reshape}(\mathcal{Y}_t, [K] \setminus \{k\})\}$ and $\mathcal{A}$ by $\{1, 2\}$.

3. Follow the steps in Section 4.3.3 to test the problem in step 2, with $r_V$ replaced by $r_{\text{vec}}$. If the null is rejected, include $k$ in the set $\widehat{\mathcal{A}}^*$.

4. Repeat from step 2 to step 3 with $k = 2, 3, \ldots, K$. Output $\widehat{\mathcal{A}}^*$.

With the algorithm output, we interpret that $\{\mathcal{Y}_t\}$ has no Kronecker product structure along $\widehat{\mathcal{A}}^*$. In practice, $\widehat{\mathcal{A}}^*$ being an empty set implies $\{\mathcal{Y}_t\}$ has a Kronecker product structure.

**Remark 4.4** *Definition 4.2 considers the absence of Kronecker product structure over a single set $\mathcal{A}$ only, which does not fully characterise all scenarios for $\mathcal{Y}_t$ with order at least 4. However, we do not pursue this complication here, albeit our practical design can be readily adapted.*

# 4.5 Numerical Results

## 4.5.1 Simulations

In this subsection, we demonstrate the empirical performance of our test by Monte Carlo simulations. As discussed in Section 4.3.1, the test is only non-trivial when the data order $K$ is at least 2. We hence consider from $K = 2$ to $K = 4$.

The data generating processes adapt Assumptions (F1), (E1) and (E2). Specifically, we set the number of factors as $r_k = 2$ for any $k \in [K]$, and first generate $\mathcal{F}_t$ in (4.7) with each element being independent standardised AR(2) with AR coefficients 0.7 and -0.3. The elements in $\mathcal{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are generated similarly, but their AR coefficients are (-0.5, 0.5) and (0.4, 0.4) respectively. The standard deviation of each element in $\boldsymbol{\epsilon}_t$ is generated by i.i.d. $|\mathcal{N}(0, 1)|$. Unless otherwise specified, all innovation processes in constructing $\mathcal{F}_t$, $\mathcal{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are i.i.d. standard normal. For each $j \in [v-1]$, each factor loading matrix $\mathbf{A}_j$ is generated independently with $\mathbf{A}_j = \mathbf{U}_j\mathbf{B}_j$, where each entry of $\mathbf{U}_j \in \mathbb{R}^{d_j \times r_j}$ is i.i.d. $\mathcal{N}(0, 1)$, and $\mathbf{B}_j \in \mathbb{R}^{r_j \times r_j}$ is diagonal with the $h$-th diagonal entry being $d_j^{-\zeta_{j,h}}$, $0 \leq \zeta_{j,h} \leq 0.5$. Pervasive factors have $\zeta_{j,h} = 0$, while weak factors have $0 < \zeta_{j,h} \leq 0.5$. Each entry of $\mathbf{A}_{e,j} \in \mathbb{R}^{d_j \times r_{e,j}}$ is i.i.d. $\mathcal{N}(0, 1)$, but has independent probability of 0.95 being set exactly to 0. We set $r_{e,k} = 2$ for all $j \in [v-1]$ throughout all experiments. For any $\mathcal{A}$ (specified later), we obtain

$$\textsc{reshape}(\mathcal{F}_t, \mathcal{A}) = \mathcal{F}_{\text{reshape},t}, \quad \textsc{reshape}(\mathcal{E}_t, \mathcal{A}) = \mathcal{E}_{\text{reshape},t}.$$

Lastly, similar to $\{\mathbf{A}_j\}_{j\in[v-1]}$, we generate $\{\mathbf{A}_v, \ldots, \mathbf{A}_K\}$ and let $\mathbf{A}_V = \mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_v$ under $H_0$, or generate $\mathbf{A}_V$ directly under $H_1$. Whenever $r_V$ is required, it is computed as $\prod_{j\in\mathcal{A}} r_j$. According to (4.5) and (4.7), we then respectively construct $\textsc{reshape}(\mathcal{Y}_t, \mathcal{A})$ (and hence the corresponding $\mathcal{Y}_t$) and $\mathcal{Y}_t$ directly.

We consider a series of performance indicators and each simulation setting is repeated 500 times. With notations in Section 4.3.3, we calculate the following with $\alpha \in \{0.01, 0.05\}$:

$$
\begin{aligned}
&\widehat{\alpha} := \min_{m\in[|\mathcal{R}|]} \Big\{ \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{y_{m,1,t} \geq \widehat{q}_{x,1}(\alpha)\} \Big\}, \\
&\widehat{p} := \mathbb{1}\{\widehat{q}_\alpha \leq \alpha\}, \text{ where} \qquad (4.22)\\
&\widehat{q}_\alpha := \min_{m\in[|\mathcal{R}|]} \Big\{ 5\% \text{ quantile of } \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{y_{m,j,t} \geq \widehat{q}_{x,j}(\alpha)\} \text{ over } j \in [d/d_{k^*}] \Big\},
\end{aligned}
$$

where $\widehat{\alpha}$ is the significance level under the measure $\mathbb{P}_{y,m,1}$ taken minimum over $m \in [|\mathcal{R}|]$, and $\widehat{p}$ is an indicator function of the decision rule (4.14) leading to retaining $H_0$. Under $H_0$, we expect $\widehat{\alpha}$ to be close to $\alpha$ and $\widehat{p}$ to be 1 according to Theorem 4.3.

### Test size and power

Consider first $H_0$ with $\mathcal{A}$ containing the last two modes of $\mathcal{Y}_t$, i.e., $\mathcal{A} = \{1, 2\}$ for $K = 2$, $\mathcal{A} = \{2, 3\}$ for $K = 3$ and $\mathcal{A} = \{3, 4\}$ for $K = 4$. We experiment on all pervasive factors. Table 4.1 presents the simulation results under various settings for $K = 2, 3, 4$, and all of them well align with Theorem 4.3. Note that for $K = 3, 4$, all $\widehat{p}$'s are 1, and for $K = 2$, the proportion of repetitions with $\widehat{p} = 1$ is increasing with dimensions and time in general. The results under $H_1$ are presented in Table 4.2 which confirms the power of our test. While larger dimensions generally improve the test performance, it is unsurprising from Table 4.2 that under the same $(T, d_k)$ setting, testing the Kronecker product structure along two modes on $\mathcal{Y}_t$ is harder for higher-order $\mathcal{Y}_t$. This is reasonable since the testing problem (4.6) is genuinely harder when $\mathbf{A}_V$ plays a less significant role in a higher-order data. To demonstrate this, suppose $K = 3$, $(T, d_1, d_2, d_3) = (360, 10, 15, 20)$, and all factors are pervasive. We experiment through $\mathcal{A} = \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$. The results reported in Table 4.3 indeed shows that when the tested loading matrix $\mathbf{A}_V$ has a larger size, the test has larger power in general. The setting with $\mathcal{A} = \{2, 3\}$ is an exception, suggesting a potential issue of unbalanced spatial dimensions.

### Robustness for weak factor, heavy-tailed noise and misspecified number of factors

In the following, we fix $K = 3$ and $\mathcal{A} = \{2, 3\}$ to investigate the robustness of our test. Consider Setting I and II, each with four sub-settings:

| | $K=2$ | | | | $K=3$ | | | | $K=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T=120$ | $d_k=15$ | | $d_k=30$ | | $d_k=15$ | | $d_k=30$ | | $d_k=10$ | | $d_k=15$ | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .020 | .071 | .020 | .078 | .013 | .055 | .013 | .055 | .012 | .054 | .012 | .053 |
| $\widehat{p}$ | .974 | .836 | .996 | .860 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $T=360$ | $d_k=15$ | | $d_k=30$ | | $d_k=15$ | | $d_k=30$ | | $d_k=10$ | | $d_k=15$ | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .011 | .057 | .012 | .059 | .010 | .051 | .010 | .052 | .010 | .051 | .010 | .051 |
| $\widehat{p}$ | .988 | .862 | 1 | .842 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $T=720$ | $d_k=15$ | | $d_k=30$ | | $d_k=15$ | | $d_k=30$ | | $d_k=10$ | | $d_k=15$ | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .011 | .053 | .012 | .054 | .010 | .051 | .010 | .051 | .010 | .050 | .010 | .051 |
| $\widehat{p}$ | .994 | .916 | 1 | .920 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Table 4.1: Results of $\widehat{\alpha}$ and $\widehat{p}$ under $H_0$ in (4.6) for $K = 2, 3, 4$. For each setting, $d_k$ is the same for all $k \in [K]$. Each cell is the average of $\widehat{\alpha}$ or $\widehat{p}$ computed under the corresponding setting over 500 runs.

| | $K=2$ | | | | $K=3$ | | | | $K=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T=120$ | $d_k=15$ | | $d_k=30$ | | $d_k=15$ | | $d_k=30$ | | $d_k=10$ | | $d_k=15$ | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .839 | .898 | .928 | .956 | .674 | .742 | .790 | .834 | .583 | .655 | .649 | .712 |
| $\widehat{p}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .012 | .002 | 0 | 0 |
| $T=360$ | $d_k=15$ | | $d_k=30$ | | $d_k=15$ | | $d_k=30$ | | $d_k=10$ | | $d_k=15$ | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .818 | .888 | .917 | .951 | .659 | .738 | .776 | .832 | .571 | .653 | .636 | .709 |
| $\widehat{p}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .002 | 0 | 0 | 0 |
| $T=720$ | $d_k=15$ | | $d_k=30$ | | $d_k=15$ | | $d_k=30$ | | $d_k=10$ | | $d_k=15$ | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .817 | .885 | .918 | .951 | .652 | .731 | .787 | .837 | .559 | .640 | .629 | .701 |
| $\widehat{p}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .002 | 0 | 0 | 0 |

Table 4.2: Results of $\widehat{\alpha}$ and $\widehat{p}$ under $H_1$ in (4.6) for $K = 2, 3, 4$. Refer to Table 4.1 for the explanation of each cell.

| $H_0$ | $\mathcal{A} = \{1,2\}$ | | $\mathcal{A} = \{1,3\}$ | | $\mathcal{A} = \{2,3\}$ | | $\mathcal{A} = \{1,2,3\}$ | |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .010 | .051 | .010 | .052 | .010 | .052 | .014 | .063 |
| $\widehat{p}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | .956 |
| $H_1$ | $\mathcal{A} = \{1,2\}$ | | $\mathcal{A} = \{1,3\}$ | | $\mathcal{A} = \{2,3\}$ | | $\mathcal{A} = \{1,2,3\}$ | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .702 | .775 | .705 | .779 | .673 | .748 | .927 | .959 |
| $\widehat{p}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 4.3: Results of $\widehat{\alpha}$ and $\widehat{p}$ over different $\mathcal{A}$'s in (4.6) for $(T, d_1, d_2, d_3) = (360, 15, 20, 25)$. Refer to Table 4.1 for the explanation of each cell. For each $\mathcal{A} = \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}$, the number of rows of $\mathbf{A}_V$ in (4.6) is respectively $300, 375, 500, 7500$.

(Ia) $T = 180,\ d_1 = d_2 = d_3 = 15$. All factors are pervasive with $\zeta_{j,h} = 0$.

(Ib) Same as (Ia), but one factor is weak with $\zeta_{j,1} = 0.1$.

(Ic) Same as (Ia), but both factors are weak with $\zeta_{j,1} = \zeta_{j,2} = 0.1$.

(Id) Same as (Ia), but all innovation processes in constructing $\mathcal{F}_t$, $\mathcal{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are i.i.d. $t_3$.

(IIa–d) Same as (Ia) to (Id) respectively, except that $r_V$ is randomly specified from $\{2, 3, 4, 5, 6\}$ with equal probability.

Setting (Ia) is our benchmark and all other settings feature some defects from weak factors, heavy-tailed noise, or misspecified number of factors. Table 4.4 reports the results for both $H_0$ and $H_1$. In contrast to (Ia), all other settings have lower test power to various extents. However, the size of the test is hardly influenced by weak factors or heavy-tailed noise from the results of (Ib), (Ic) and (Id). Although number-of-factor misspecification is detrimental, our decision rule $\widehat{p}$ still has satisfying performance.

**Numerical performance of the practical testing algorithm**

On the practical testing algorithm which does not require $\mathcal{A}$ to be specified, we consider Setting III and IV with $K = 3$, and each has three sub-settings:

(IIIa) $T = 360,\ d_1 = d_2 = d_3 = 10$. All factors are strong and the data has a Kronecker product structure.

(IIIb) Same as (IIIa), but the data has no Kronecker product structure along $\{2, 3\}$.

(IIIc) Same as (IIIa), but the data has no Kronecker product structure along $\{1, 2, 3\}$.

(IVa–c) Same as (IIIa) to (IIIc) respectively, except that $T = 720$.

| Setting I | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $H_0$ | (Ia) | | (Ib) | | (Ic) | | (Id) | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .008 | .054 | .008 | .053 | .008 | .053 | .008 | .053 |
| $\widehat{p}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $H_1$ | (Ia) | | (Ib) | | (Ic) | | (Id) | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .691 | .765 | .593 | .684 | .441 | .553 | .519 | .693 |
| $\widehat{p}$ | 0 | 0 | .014 | 0 | .034 | .004 | .070 | .002 |
| Setting II | | | | | | | | |
| $H_0$ | (IIa) | | (IIb) | | (IIc) | | (IId) | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .054 | .113 | .035 | .091 | .021 | .074 | .030 | .092 |
| $\widehat{p}$ | .972 | .932 | .988 | .966 | .998 | .996 | .996 | .964 |
| $H_1$ | (IIa) | | (IIb) | | (IIc) | | (IId) | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| $\widehat{\alpha}$ | .553 | .652 | .504 | .620 | .378 | .509 | .424 | .596 |
| $\widehat{p}$ | .034 | 0 | .018 | .002 | .036 | .006 | .128 | .004 |

Table 4.4: Results of $\widehat{\alpha}$ and $\widehat{p}$ under $H_0$ and $H_1$ in (4.6) over sub-settings of Setting I and II. Refer to Table 4.1 for the explanation of each cell.

| | Setting III | | | | | | Setting IV | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (IIIa) | | (IIIb) | | (IIIc) | | (IVa) | | (IVb) | | (IVc) | |
| $\alpha$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| Mode 1 | 0 | .024 | .030 | .202 | 1 | 1 | 0 | 0 | .004 | .228 | 1 | 1 |
| Mode 2 | 0 | .036 | 1 | 1 | 1 | 1 | 0 | .004 | 1 | 1 | 1 | 1 |
| Mode 3 | 0 | .034 | .998 | 1 | 1 | 1 | 0 | .002 | 1 | 1 | 1 | 1 |

Table 4.5: Results of Setting III and IV for the practical testing algorithm. Each cell is the fraction of the corresponding mode identified over 500 runs for the corresponding sub-settings.

Table 4.5 verifies that our algorithm is able to test the Kronecker product structure of a given data without pre-specifying $\mathcal{A}$. The performance is improved with more observations, and the level of $\alpha = 0.01$ works particularly well.

### 4.5.2 Real data analysis

We apply our test on two real data examples described as follows.

1. New York City taxi traffic. The data considered includes all individual taxi rides operated by Yellow Taxi within Manhattan Island of New York City, published at

   https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

   The dataset contains trip records within the period of January 1, 2018 to December 31, 2022. We focus on the pick-up and drop-off dates/times, and pick-up and drop-off locations which are coded according to 69 predefined zones in the dataset. Moreover, each day is divided into 24 hourly periods to represent the pick-up and drop-off times, with the first hourly period from 0 a.m. to 1 a.m. Hence each day we have $\mathcal{Y}_t \in \mathbb{R}^{69\times 69\times 24}$, where $y_{i_1,i_2,i_3,t}$ is the number of trips from zone $i_1$ to zone $i_2$ and the pick-up time is within the $i_3$-th hourly period on day $t$. We consider business days and non-business days separately, so that we will analyse two tensor time series. The business-day series and the non-business-day series are 1,260 and 566 days long, respectively.

2. Fama–French portfolio returns. This is a set of portfolio returns data, where stocks are respectively categorised into ten levels of market equity and book-to-equity ratio which is the book equity for the last fiscal year divided by the end-of-year market equity; both criteria use NYSE deciles as breakpoints at the end of June each year. See details in https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_100_port_sz.html.

   The stocks in each of the $10 \times 10$ categories form exactly two portfolios, one being value-weighted, and the other of equal-weight. That is, we will study two sets of 10 by 10 portfolios with their time series. We use monthly data from January 2010 to June

2021, and hence for both value-weighted and equal-weighted portfolios we have each of our data set as an order-2 tensor $\mathcal{X}_t \in \mathbb{R}^{10\times 10}$ for $t \in [138]$.

The two taxi series are order-3 tensor time series, and we only test their Kronecker product structure along $\mathcal{A} = \{1, 2\}$, i.e., we speculate that there is a merged "location" factor instead of "pick-up" and "drop-off" factors along mode-1 and -2 respectively. On the other hand, the two portfolio series are order-2 tensor time series, hence naturally we test along $\mathcal{A} = \{1, 2\}$. Furthermore, we remove the market effect via the capital asset pricing model (CAPM) as

$$\mathbf{vec}\big(\mathcal{X}_t\big) = \mathbf{vec}\big(\bar{\mathcal{X}}\big) + (r_t - \bar{r})\boldsymbol{\beta} + \mathbf{vec}\big(\mathcal{Y}_t\big),$$

where $\mathbf{vec}\big(\mathcal{X}_t\big) \in \mathbb{R}^{100}$ is the vectorised returns at time $t$, $\mathbf{vec}\big(\bar{\mathcal{X}}\big)$ is the sample mean of $\mathbf{vec}\big(\mathcal{X}_t\big)$, $\boldsymbol{\beta}$ is the coefficient vector, $r_t$ is the return of the NYSE composite index at time $t$, $\bar{r}$ is the sample mean of $r_t$, and $\mathbf{vec}\big(\mathcal{Y}_t\big)$ is the CAPM residual. The least squares solution is

$$\widehat{\boldsymbol{\beta}} = \frac{\sum_{t=1}^{138}(r_t - \bar{r})\{\mathbf{vec}\big(\mathcal{X}_t\big) - \mathbf{vec}\big(\bar{\mathcal{X}}\big)\}}{\sum_{t=1}^{138}(r_t - \bar{r})^2},$$

so that the estimated residual series $\{\widehat{\mathcal{Y}}_t\}_{t\in[138]}$ with $\widehat{\mathcal{Y}}_t \in \mathbb{R}^{10\times 10}$ is constructed as $\{\mathbf{vec}\big(\mathcal{X}_t\big) - \mathbf{vec}\big(\bar{\mathcal{X}}\big) - (r_t - \bar{r})\widehat{\boldsymbol{\beta}}\}_{t\in[138]}$.

Hence, we study six time series in total: business-day taxi series, non-business-day taxi series, value-weighted portfolio series, equal-weighted portfolio series, value-weighted residual series and equal-weighted residual series. For each series, we perform the test described in Section 4.3.3. To estimate the rank, we use BCorTh by Chen and Lam (2024b), iTIP-ER by Han et al. (2022) and RTFA-ER by He et al. (2022b) directly on each time series due to their large dimensions. Each mode of the six series has one or two estimated number of factors. Since the test results are similar for those rank settings, we present the results with two factors each mode and hence $\widehat{r}_V = 4$.

In addition, we also conduct the hypotheses tests in He et al. (2023a) on our matrix time series data sets. To explain their hypotheses, for a matrix time series $\{\mathbf{Y}_t\}$ with $\mathbf{Y}_t \in \mathbb{R}^{d_1\times d_2}$, under the null we have (4.7) for $K = 2$:

$$H_0 : \mathbf{Y}_t = \mathbf{A}_1\mathbf{F}_t\mathbf{A}_2' + \mathbf{E}_t,$$

where $\mathbf{F}_t \in \mathbb{R}^{r_1\times r_2}$. However, under their two alternatives we test

$$H_{1,\text{row}} : r_2 = 0, \quad H_{1,\text{col}} : r_1 = 0,$$

where according to He et al. (2023a), $r_1 > 0, r_2 = 0$ (resp. $r_2 > 0, r_1 = 0$) denotes a one-way factor model along the row dimension, so that $\mathbf{Y}_t = \mathbf{A}_1\mathbf{F}_{1,t} + \mathbf{E}_t$ with $\mathbf{F}_{1,t} \in \mathbb{R}^{r_1\times d_2}$ (resp. the

| | $\widehat{\alpha}$ | | $\widehat{q}_\alpha$ | | Tests in He et al. (2023a) | |
|---|---|---|---|---|---|---|
| | 1% | 5% | 1% | 5% | $H_0$ versus $H_{1,\text{row}}$ | $H_0$ versus $H_{1,\text{col}}$ |
| Business-day taxi | .020 | .093 | .002 | .003 | - | - |
| Non-business-day taxi | .018 | .095 | .004 | .011 | - | - |
| Value-weighted portfolio | .058 | .087 | **.011** | **.053** | Not reject | Not reject |
| Equal-weighted portfolio | .036 | .051 | **.018** | .039 | Not reject | Not reject |
| Value-weighted residual | .022 | .065 | **.011** | .047 | Not reject | Not reject |
| Equal-weighted residual | .014 | .051 | **.011** | .047 | Not reject | Not reject |

Table 4.6: Test results for the studied series. The first two columns report the results for our hypothesis of interest (4.6) with $\mathcal{A} = \{1, 2\}$; $\widehat{q}_\alpha$ larger than the corresponding $\alpha$ level is in bold. The last two columns report the results according to He et al. (2023a).

column dimension, so that $\mathbf{Y}_t = \mathbf{F}_{2,t}\mathbf{A}_2' + \mathbf{E}_t$ with $\mathbf{F}_{2,t} \in \mathbb{R}^{d_1 \times r_2}$), and $r_1 = r_2 = 0$ denotes the absence of any factor structure, so that $\mathbf{Y}_t = \mathbf{E}_t$. All hyperparameter setups in Table 8 and 9 in He et al. (2023a) are experimented and all conclusions are the same.

Table 4.6 reports $\widehat{\alpha}$ and $\widehat{q}_\alpha$ defined in (4.22), with $\alpha = 0.01,\ 0.05$, together with the corresponding tests by He et al. (2023a). For our hypothesis of interest, there is no evidence to reject the null for the two taxi series, but there is mild evidence (especially at $1\%$ level, with $\widehat{\alpha}$ observed to be mildly larger than 1%) to conclude that for the Fama–French time series, there is no Kronecker product structure along $\{1, 2\}$. In other words, there is evidence to suggest that the portfolio return series has structures deviating from the low-rank structure along its respective categorisations by market equity and book-to-equity ratio, meaning the vectorised data may have a more distinct factor structure. The comparisons between the portfolio and residual series justifies the removal of the market effect, which is intuitive as the market effect should be pervasive in financial returns and is irrelevant of our categorisations. In contrast, we cannot reject the null by considering those alternative hypotheses considered in He et al. (2023a).

## 4.6 Details on Identification

This section concerns the identification of the model in Definition 4.2, following a discussion on Definition 4.1. First, consider Definition 4.2.1 so that in (4.3), we have $\mathbf{A}_{\text{reshape},K-\ell+1} \in \mathcal{K}_{d_{a_1} \times \cdots \times d_{a_\ell}}$. Given a general $\mathcal{A} = \{a_1, \ldots, a_\ell\}$, the ordered set of matrices $\{\mathbf{A}_j\}_{j \in [\kappa]}$ decomposing (as in Definition 4.1) $\mathbf{A}_{\text{reshape},K-\ell+1}$ might not be unique. For instance, suppose $K = 2$, $d_1 = d_2$, $\mathcal{A} = \{1, 2\}$ (hence $\mathbf{A}_{\text{reshape},K-\ell+1} \equiv \mathbf{A}_{\text{reshape},1}$ has $d_1^2$ rows) and let $\mathbf{A}_{\text{reshape},1} = d_1^{-1/4}\mathbf{1}_{d_1^2}$, then we have

$$\mathbf{A}_{\text{reshape},1} = \underbrace{\big(d_1^{-1/4}\mathbf{1}_{d_1}\big)}_{\mathbf{A}_1} \otimes \underbrace{\mathbf{1}_{d_1}}_{\mathbf{A}_2} = \underbrace{\mathbf{1}_{d_1}}_{\ddot{\mathbf{A}}_1} \otimes \underbrace{\big(d_1^{-1/4}\mathbf{1}_{d_1}\big)}_{\ddot{\mathbf{A}}_2}, \tag{4.23}$$

where it is clear that $\|\mathbf{A}_{1,\cdot 1}\|^2 \asymp d_1^{1/2}$ and $\|\ddot{\mathbf{A}}_{1,\cdot 1}\|^2 \asymp d_1$. Such defeat can be rectified by allocating the "factor strength" in $\mathbf{A}_{\text{reshape},K-\ell+1}$ to each mode in $\mathcal{A}$; see the following Assumption (S1) as one example.

(S1) *For* (4.4) *such that* $\mathbf{A}_{\text{reshape},K-\ell+1} = \otimes_{i\in\mathcal{A}}\mathbf{A}_i$ *for a given* $\mathcal{A}$*, we assume for any* $i \in \mathcal{A}$*,*

$$\frac{\|\mathbf{A}_i\|_F^2}{r_i d_i} = \left(\frac{\|\mathbf{A}_{\text{reshape},K-\ell+1}\|_F^2}{\prod_{i\in\mathcal{A}} r_i d_i}\right)^{1/|\mathcal{A}|}.$$

The issue of indeterminacy in the factor strength is fixed by (S1) which has same spirit of Assumption (IC) in Chen and Lam (2024a). Its heuristic is to allocate the factor strength in $\mathbf{A}_{\text{reshape},K-\ell+1}$ to each mode according to their number of factors and dimensions. Note that Assumption (S1) holds automatically if all factors are pervasive. Now recall the example (4.23), $\mathbf{A}_1 = \mathbf{A}_2 = d_1^{-1/8}\mathbf{1}_{d_1}$ are identified by (S1). Note that the discussion above is only for completeness and would not influence our testing problem.

Next, we identify models in Definition 4.2 and Theorem 4.1. For (4.3) and (4.4), we state below the two assumptions (L1') and (L2'), which are (notation-wise) general versions of (L1) and (L2), respectively.

(L1') (Factor strength in $\mathbf{A}_{\text{reshape},j}$). *For each* $j \in [K-\ell]$*, we assume that* $\mathbf{A}_{\text{reshape},j}$ *in* (4.3) *is of full rank and as* $I_j \to \infty$*,*

$$\mathbf{Z}_{\text{reshape},j}^{-1/2}\mathbf{A}_{\text{reshape},j}'\mathbf{A}_{\text{reshape},j}\mathbf{Z}_{\text{reshape},j}^{-1/2} \to \mathbf{\Sigma}_{\text{reshape},A,j}, \tag{4.24}$$

*where* $\mathbf{\Sigma}_{\text{reshape},A,j}$ *is positive definite with all eigenvalues bounded away from 0 and infinity, and* $\mathbf{Z}_{\text{reshape},j}$ *is a diagonal matrix with* $(\mathbf{Z}_{\text{reshape},j})_{hh} \asymp I_j^{\delta_{\text{reshape},j,h}}$ *for* $h \in [\pi_j]$ *and the ordered factor strengths* $1/2 < \delta_{\text{reshape},j,\pi_j} \le \cdots \le \delta_{\text{reshape},j,1} \le 1$.

*We assume that* $\mathbf{A}_{\text{reshape},j}$ *for* $j = K-\ell+1$ *also has the above form, except that only the maximum and minimum factor strengths are ordered, i.e.,* $1/2 < \delta_{\text{reshape},j,\pi_j} \le \delta_{\text{reshape},j,h} \le \delta_{\text{reshape},j,1} \le 1$ *for any* $h \in [\pi_j]$.

(L2') (Factor strength in $\mathbf{A}_i$). *For* (4.4) *with a given* $\mathcal{A}$*, we assume that for each* $i \in \mathcal{A}$*,* $\mathbf{A}_i$ *is of full rank and as* $d_i \to \infty$*,*

$$\mathbf{Z}_i^{-1/2}\mathbf{A}_i'\mathbf{A}_i\mathbf{Z}_i^{-1/2} \to \mathbf{\Sigma}_{A,i}, \tag{4.25}$$

*where* $\mathbf{\Sigma}_{A,i}$ *is positive definite with all eigenvalues bounded away from 0 and infinity, and* $\mathbf{Z}_i$ *is a diagonal matrix with* $(\mathbf{Z}_i)_{hh} \asymp d_i^{\delta_{i,h}}$ *for* $h \in [r_i]$ *and the ordered factor strengths* $1/2 < \delta_{i,r_i} \le \cdots \le \delta_{i,1} \le 1$.

Theorem 4.5 presents the identification of the model (4.3) both in general and with Kronecker product structure (equivalently (4.4) according to Theorem 4.1). Its proof is given directly after the statement.

**Theorem 4.5** (Identification). *Let Assumption (F1) and (L1') hold, and $\mathcal{A}$ is given. Then the factor structure in* (4.3) *is asymptotically identified up to some invertible matrix $\mathbf{M}_j \in \mathbb{R}^{\pi_j \times \pi_j}$ such that the following sets of factor structure are equivalent,*

$$\left(\mathcal{F}_{\text{reshape},t}, \left\{\mathbf{A}_{\text{reshape},j}\right\}_{j\in[K-\ell+1]}\right) = \left(\mathcal{F}_{\text{reshape},t} \times_{j=1}^{K-\ell+1} \mathbf{M}_j^{-1}, \left\{\mathbf{A}_{\text{reshape},j}\mathbf{M}_j\right\}_{j\in[K-\ell+1]}\right).$$

*Let Assumption (L2') further holds. With a Kronecker product structure on* (4.3)*, we have* (4.4) *where for each $k \in [K]$, $\mathbf{A}_k$ has unique rank and the factor structure in* (4.4) *is asymptotically identified up to some invertible matrices.*

***Proof of Theorem 4.5.*** Consider first (4.3). Let $\left(\ddot{\mathcal{F}}_{\text{reshape},t}, \left\{\ddot{\mathbf{A}}_{\text{reshape},j}\right\}_{j\in[K-\ell+1]}\right)$ be another set of parameters such that $\ddot{\mathcal{F}}_{\text{reshape},t} \times_{j=1}^{K-\ell+1} \ddot{\mathbf{A}}_{\text{reshape},j} = \mathcal{F}_{\text{reshape},t} \times_{j=1}^{K-\ell+1} \mathbf{A}_{\text{reshape},j}$. Define

$$\mathbf{A}_{\text{reshape,-}j} := \mathbf{A}_{\text{reshape},K-\ell+1} \otimes \cdots \otimes \mathbf{A}_{\text{reshape},j+1} \otimes \mathbf{A}_{\text{reshape},j-1} \otimes \cdots \otimes \mathbf{A}_{\text{reshape},1}.$$

Define $\ddot{\mathbf{A}}_{\text{reshape,-}j}$ similarly. Without loss of generality, for any $j \in [K-\ell+1]$ we write

$$\ddot{\mathbf{A}}_{\text{reshape},j} = \mathbf{A}_{\text{reshape},j}\mathbf{M}_{\text{reshape},j} + \mathbf{\Gamma}_{\text{reshape},j}, \text{ where } \mathbf{\Gamma}'_{\text{reshape},j}\mathbf{A}_{\text{reshape},j} = \mathbf{0}, \tag{4.26}$$

with $\mathbf{M}_{\text{reshape},j} \in \mathbb{R}^{\pi_j \times \pi_j}$ and $\mathbf{\Gamma}_{\text{reshape},j} \in \mathbb{R}^{I_j \times \pi_j}$, but can have zero columns. Then

$$\begin{aligned}
\mathbf{0} &= \mathbf{\Gamma}'_{\text{reshape},j}\mathbf{A}_{\text{reshape},j}\mathbf{F}_{\text{reshape},t,(j)}\mathbf{A}'_{\text{reshape,-}j}\ddot{\mathbf{A}}_{\text{reshape,-}j} \\
&= \mathbf{\Gamma}'_{\text{reshape},j}\ddot{\mathbf{A}}_{\text{reshape},j}\ddot{\mathbf{F}}_{\text{reshape},t,(j)}\ddot{\mathbf{A}}'_{\text{reshape,-}j}\ddot{\mathbf{A}}_{\text{reshape,-}j} \\
&= \mathbf{\Gamma}'_{\text{reshape},j}\mathbf{\Gamma}_{\text{reshape},j}\ddot{\mathbf{F}}_{\text{reshape},t,(j)}\ddot{\mathbf{A}}'_{\text{reshape,-}j}\ddot{\mathbf{A}}_{\text{reshape,-}j},
\end{aligned}$$

which can only be true in general if $\mathbf{\Gamma}_{\text{reshape},j} = \mathbf{0}$ since $\ddot{\mathbf{F}}_{\text{reshape},t,(j)}$ is random by Assumption (F1) and $\ddot{\mathbf{A}}'_{\text{reshape,-}j}\ddot{\mathbf{A}}_{\text{reshape,-}j}$ converges to some full rank matrix by (L1'). Hence $\mathbf{M}_{\text{reshape},j}$ has full rank, and $\ddot{\mathbf{A}}_{\text{reshape},j}$ and $\mathbf{A}_{\text{reshape},j}$ share the same column space. $\mathcal{F}_{\text{reshape},t}$ is identified once $\left\{\mathbf{A}_{\text{reshape},j}\right\}_{j\in[K-\ell+1]}$ is given correspondingly.

Suppose (4.3) has a Kronecker product structure and consider (4.4). By an argument similar to (4.26) but over $k \in [K]$ (omitted), each matrix $\mathbf{A}_k$ for $k \in [K]$ has a unique rank and is identified up to some invertible matrix using (L2'). Hence $\mathcal{F}_t$ is also identified, which completes the proof of the theorem. □

## 4.7 Proof of Theorems and Auxiliary Results

**A high-level summary of proofs**: The design of (4.11) and (4.12) over all divisor combinations of $r_V$ is due to pseudo-ranks from mode-$v$ to mode-$K$ of $\mathcal{Y}_t$ under $H_1$ in (4.6). On the other hand, under $H_0$, there must be one $m \in [|\mathcal{R}|]$ such that $(\pi_{m,1}, \dots, \pi_{m,K-v+1}) = (r_v, \dots, r_K)$. Hence we consider throughout the proof that $\{r_k\}_{k\in[K]}$ in (4.7) (hence under $H_0$) is correctly specified. Hence, we simplify the notations $\widehat{\mathbf{Q}}_{m,i}$, $\widehat{\mathcal{C}}_{m,t}$ and $\widehat{\mathcal{E}}_{m,t}$ as $\widehat{\mathbf{Q}}_i$, $\widehat{\mathcal{C}}_t$ and $\widehat{\mathcal{E}}_t$, respectively. Whenever (L2) is assumed, we are implicitly considering $H_0$ in (4.6), and

$$
\begin{aligned}
&\mathbf{Q}_{\text{-}k} := \mathbf{Q}_K \otimes \cdots \otimes \mathbf{Q}_{k+1} \otimes \mathbf{Q}_{k-1} \otimes \cdots \otimes \mathbf{Q}_1 \quad \text{for } k \in [K],\\
&\mathcal{F}_{Z,t} := \mathcal{F}_t \times_{k=1}^K \mathbf{Z}_k^{1/2} \quad \text{with } \mathcal{F}_t \text{ from (4.19)}.
\end{aligned}
$$

Theorem 4.1 reveals the importance of the reshape operator and is the key to formalise the testing problem. Lemma 4.1 to Lemma 4.4 serve as technical steps. The steps of all other proofs are summarised as follows.

1. Under $H_0$, consider (4.7). We first derive the rate of convergence for $\widehat{\mathbf{Q}}_k$ as an estimator of $\mathbf{Q}_k$ for each $k \in [K]$ (Lemma 4.5). Then the rates for the corresponding core factor and hence the common component can also be obtained (Lemma 4.6).

2. Under $H_0$, consider (4.5). We derive the rates of convergence for $\{\widetilde{\mathbf{Q}}_j\}_{j\in[v-1]}$ and $\widetilde{\mathbf{Q}}_V$ as estimators of $\{\mathbf{Q}_j\}_{j\in[v-1]}$ and $\mathbf{Q}_V$, respectively. Similar to step 1, the rate for the common component is obtained. See Lemma 4.7.

3. With Steps 1 and 2, we then show that $\sum_{i=1}^{d_{k^*}} \widehat{E}^2_{m,t,(k^*),ij}$ and $\sum_{i=1}^{d_{k^*}} \widetilde{E}^2_{t,(k^*),ij}$ under $H_0$ have the same distribution asymptotically (Theorem 4.2).

4. With Step 3 and the uniform rates for the common components in Lemma 4.10, the test statistic can be constructed with theoretical support (Theorem 4.3).

By the interplay between the core factor and the noise in Assumptions (F1), (E1) and (E2), we state below Lemma 4.1 which is direct from Proposition 1.1 and 1.2 of Cen and Lam (2025b).

**Lemma 4.1** *Let Assumptions (F1), (E1) and (E2) hold. Then*

1. *(Weak correlation of noise $\mathcal{E}_t$ across different modes and time). There exists some positive constant $C < \infty$ so that for any $t \in [T]$, $k \in [K]$, $i_k, j \in [d_k]$, $h \in [d_{\text{-}k}]$, we have*

$\mathbb{E}\mathcal{E}_{t,i_1,\ldots,i_K} = 0$, $\mathbb{E}\mathcal{E}^4_{t,i_1,\ldots,i_K} \le C$, *and*

$$\sum_{j=1}^{d_k}\sum_{l=1}^{d_{\text{-}k}} \Big|\mathbb{E}[E_{t,(k),i_k h}E_{t,(k),jl}]\Big| \le C,$$
$$\sum_{l=1}^{d_{\text{-}k}}\sum_{s=1}^{T} \Big|\mathrm{Cov}\big(E_{t,(k),i_k h}E_{t,(k),jh}, E_{s,(k),i_k l}E_{s,(k),jl}\big)\Big| \le C.$$

2. *(Weak dependence between factor $\mathcal{F}_t$ and noise $\mathcal{E}_t$). There exists some positive constant $C < \infty$ so that for any $k \in [K]$, $j \in [d_k]$, and any deterministic vectors $\mathbf{u} \in \mathbb{R}^{r_k}$ and $\mathbf{v} \in \mathbb{R}^{r_{\text{-}k}}$ with constant magnitudes, it holds for $\mathcal{F}_t$ in* (4.7) *that*

$$\mathbb{E}\left(\frac{1}{(d_{\text{-}k}T)^{1/2}}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} E_{t,(k),jh}\mathbf{u}'\mathbf{F}_{t,(k)}\mathbf{v}\right)^2 \le C.$$

3. *Statement 2 holds similarly for* RESHAPE$(\mathcal{F}_t, \mathcal{A})$ *and* RESHAPE$(\mathcal{E}_t, \mathcal{A})$.

**Lemma 4.2** *Under Assumption (F1), with $\gamma_v := \prod_{j=1}^{v-1} r_j$, we have as $T \to \infty$,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_{t,(k)}\mathbf{M}\mathbf{F}'_{t,(k)} \xrightarrow{p} \mathrm{tr}(\mathbf{M})\cdot\mathbf{I}_{r_k}, \tag{4.27}$$
$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_{\text{reshape},t,(j)}\mathbf{N}\mathbf{F}'_{\text{reshape},t,(j)} \xrightarrow{p} \mathrm{tr}(\mathbf{N})\cdot\mathbf{I}_{r_j}, \tag{4.28}$$
$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_{\text{reshape},t,(v)}\mathbf{W}\mathbf{F}'_{\text{reshape},t,(v)} \xrightarrow{p} \mathrm{tr}(\mathbf{W})\cdot\mathbf{I}_{r_V}, \tag{4.29}$$

*where $\mathbf{F}_{t,(k)} \in \mathbb{R}^{r_k\times r_{\text{-}k}}$ for $k \in [K]$ is the mode-$k$ unfolding of $\mathcal{F}_t$ in* (4.7)*, and $\mathbf{M}$ is any $r_{\text{-}k}\times r_{\text{-}k}$ matrix independent of $\{\mathcal{F}_t\}_{t\in[T]}$ with $\|\mathbf{M}\|_F$ bounded in probability; similarly for $\mathbf{F}_{\text{reshape},t,(j)} \in \mathbb{R}^{r_k\times(r_V\gamma_v/r_j)}$ for $j \in [v-1]$ in* (4.5)*, and for $\mathbf{F}_{\text{reshape},t,(v)} \in \mathbb{R}^{r_V\times\gamma_v}$ in* (4.5).

***Proof of Lemma 4.2.*** We show (4.27) and the other two follow similarly. With (F1),

$$\sum_{s=1}^{T}\Big(\sum_{w\ge0} a_{f,w}a_{f,w-(t-s)}\Big)\Big(\sum_{q\ge0} a_{f,q}a_{f,q-(t-s)}\Big)$$
$$\le \Big(\sum_{s=1}^{T}\sum_{w\ge0}|a_{f,w}||a_{f,w-(t-s)}|\Big)\Big(\sum_{q\ge0}|a_{f,q}|\Big)\cdot\max_q|a_{f,q}| = O(1)\cdot\Big(\sum_{w\ge0}|a_{f,w}|\Big)^2 = O(1), \tag{4.30}$$

where the last two equality used Assumption (F1). Similarly, it holds that

$$\sum_{s=1}^{T}\sum_{w\geq 0} a_{f,w}^2 a_{f,w-(t-s)}^2 \leq \Big(\sum_{w\geq 0} a_{f,w}^2\Big)^2 = O(1). \tag{4.31}$$

Now for $i \neq j \in [r_k]$, by Assumption (F1) we have

$$\mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}'_{t,(k),i\cdot}\mathbf{M}\mathbf{F}_{t,(k),j\cdot}\Big) = \mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}'_{t,(k),i\cdot}\Big)\mathbb{E}\Big(\mathbf{M}\mathbf{F}_{t,(k),j\cdot}\Big) = 0.$$

For $i = j \in [r_k]$, with $\mathcal{X}_{f,t}$ from (4.19), we have

$$\mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}'_{t,(k),i\cdot}\mathbf{M}\mathbf{F}_{t,(k),i\cdot}\Big) = \frac{1}{T}\sum_{t=1}^{T}\sum_{w\geq 0}\sum_{q\geq 0} a_{f,w}a_{f,q}\mathbb{E}\Big(\mathbf{X}'_{f,t-w,(k),i\cdot}\mathbf{M}\mathbf{X}_{f,t-q,(k),i\cdot}\Big)$$
$$= \frac{1}{T}\sum_{t=1}^{T}\sum_{w\geq 0}\sum_{q\geq 0} a_{f,w}a_{f,q}\text{tr}(\mathbf{M})\mathbb{1}\{w=q\} = \text{tr}(\mathbf{M})\sum_{w\geq 0} a_{f,w}^2 = \text{tr}(\mathbf{M}),$$

so that $\mathbb{E}\big(T^{-1}\sum_{t=1}^{T}\mathbf{F}_{t,(k)}\mathbf{M}\mathbf{F}'_{t,(k)}\big) = \text{tr}(\mathbf{M})\cdot\mathbf{I}_{r_k}$. Finally, consider any $i, j \in [r_k]$,

$$\text{Var}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}'_{t,(k),i\cdot}\mathbf{M}\mathbf{F}_{t,(k),j\cdot}\Big) = \text{Var}\Big(\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{r_{\text{-}k}}\sum_{v=1}^{r_{\text{-}k}} F_{t,(k),il}M_{lv}F_{t,(k),jv}\Big)$$
$$= \frac{1}{T^2}\text{Cov}\Big(\sum_{t=1}^{T}\sum_{l=1}^{r_{\text{-}k}}\sum_{v=1}^{r_{\text{-}k}}\sum_{w\geq 0}\sum_{q\geq 0} a_{f,w}a_{f,q}X_{f,t-w,(k),il}M_{lv}X_{f,t-q,(k),jv},$$
$$\sum_{s=1}^{T}\sum_{g=1}^{r_{\text{-}k}}\sum_{u=1}^{r_{\text{-}k}}\sum_{h\geq 0}\sum_{m\geq 0} a_{f,h}a_{f,m}X_{f,s-h,(k),ig}M_{gu}X_{f,s-m,(k),ju}\Big)$$
$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big(\sum_{w\geq 0} a_{f,w}a_{f,w-(t-s)}\Big)\Big(\sum_{q\geq 0} a_{f,q}a_{f,q-(t-s)}\Big)\Big(\sum_{l=1}^{r_{\text{-}k}}\sum_{v=1}^{r_{\text{-}k}} M_{lv}^2\Big)$$
$$+ \frac{1}{T^2}\sum_{t=1}^{T}\sum_{w\geq 0}\sum_{q\geq 0}\sum_{s=1}^{T}\sum_{h\geq 0}\sum_{m\geq 0} a_{f,w}a_{f,q}a_{f,h}a_{f,m}\Big(\sum_{l=1}^{r_{\text{-}k}} M_{ll}^2\Big)$$
$$\cdot\,\text{Cov}\Big(X_{f,t-w,(k),il}X_{f,t-q,(k),il}, X_{f,s-h,(k),il}X_{f,s-m,(k),il}\Big)$$
$$= O\Big(\frac{1}{T}\Big)\cdot\Big(\sum_{l=1}^{r_{\text{-}k}}\sum_{v=1}^{r_{\text{-}k}} M_{lv}^2\Big) + O\Big(\frac{1}{T}\Big)\cdot\Big(\sum_{l=1}^{r_{\text{-}k}} M_{ll}^2\Big) = O\Big(\frac{1}{T}\Big)\cdot\|\mathbf{M}\|_F^2 = o(1),$$

where the third equality considered $i = j$ and $i \neq j$ separately, and the fourth used (4.30) and (4.31). This completes the proof of (4.27) and hence Lemma 4.2. □

**Lemma 4.3** *(Bounding $\sum_{t=1}^{T}\mathbf{R}_{k,t}$). Under Assumptions (F1), (L1), (L2), (E1) and (E2), it* *holds that*

$$\Big\| \sum_{t=1}^{T} \mathbf{Q}_k \mathbf{F}_{Z,t,(k)} \mathbf{Q}_{\text{-}k}' \mathbf{E}_{t,(k)}' \Big\|_F^2 = O_P\Big( T d_k^{1+\delta_{k,1}} d_{\text{-}k} \Big), \tag{4.32}$$

$$\Big\| \sum_{t=1}^{T} \mathbf{E}_{t,(k)} \mathbf{E}_{t,(k)}' \Big\|_F^2 = O_P\Big( T d_k^2 d_{\text{-}k} + T^2 d_k d_{\text{-}k}^2 \Big). \tag{4.33}$$

*Hence, with* $\mathbf{R}_{k,t}$ *defined in* (4.36)*, we have*

$$\Big\| \sum_{t=1}^{T} \mathbf{R}_{k,t} \Big\|_F^2 = O_P\Big( T d_k^2 d_{\text{-}k} + T^2 d_k d_{\text{-}k}^2 \Big).$$

***Proof of Lemma 4.3.*** It is not hard to see (4.32) holds as follows,

$$\Big\| \sum_{t=1}^{T} \mathbf{Q}_k \mathbf{F}_{Z,t,(k)} \mathbf{Q}_{\text{-}k}' \mathbf{E}_{t,(k)}' \Big\|_F^2 = \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} \Big( \sum_{t=1}^{T} \mathbf{A}_{k,i\cdot}' \mathbf{F}_{t,(k)} \mathbf{A}_{\text{-}k}' \mathbf{E}_{t,(k),l\cdot} \Big)^2$$
$$= \sum_{i=1}^{d_k} \|\mathbf{A}_{k,i\cdot}\|^2 \cdot \sum_{l=1}^{d_k} \Big( \sum_{h=1}^{d_{\text{-}k}} \sum_{t=1}^{T} E_{t,(k),lh} \frac{1}{\|\mathbf{A}_{k,i\cdot}\|} \mathbf{A}_{k,i\cdot}' \mathbf{F}_{t,(k)} \mathbf{A}_{\text{-}k,h\cdot} \Big)^2 = O_P\big( T d_k^{1+\delta_{k,1}} d_{\text{-}k} \big),$$

where the last equality is from Assumptions (L1), (L2) and Lemma 4.1.

Consider now (4.33). First, from Assumption (E1), for any $k \in [K]$, $i \in [d_k]$, $j \in [d_{\text{-}k}]$,

$$E_{t,(k),ij} = \mathbf{A}_{e,k,i\cdot}' \mathbf{F}_{e,t,(k)} \mathbf{A}_{e,\text{-}k,j\cdot} + \Sigma_{\epsilon,(k),ij} \epsilon_{t,(k),ij},$$

where $\mathbf{A}_{e,\text{-}k} := \mathbf{A}_{e,K} \otimes \cdots \otimes \mathbf{A}_{e,k+1} \otimes \mathbf{A}_{e,k-1} \otimes \cdots \otimes \mathbf{A}_{e,1}$. Then with Assumption (E2),

$$\mathrm{Cov}(E_{t,(k),ij}, E_{t,(k),lj}) = \mathbf{A}_{e,k,i\cdot}' \mathbf{A}_{e,k,l\cdot} \|\mathbf{A}_{e,\text{-}k,j\cdot}\|^2 + \Sigma_{\epsilon,(k),ij}^2 \mathbb{1}_{\{i=l\}},$$

and together with Lemma 4.1,

$$\mathbb{E}\Big( \Big\| \sum_{t=1}^{T} \mathbf{E}_{t,(k)} \mathbf{E}_{t,(k)}' \Big\|_F^2 \Big) = \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} \mathbb{E}\Big\{ \Big( \sum_{t=1}^{T} \sum_{j=1}^{d_{\text{-}k}} E_{t,(k),ij} E_{t,(k),lj} \Big)^2 \Big\}$$
$$= \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} \Big\{ \sum_{t=1}^{T} \sum_{j=1}^{d_{\text{-}k}} \sum_{s=1}^{T} \sum_{h=1}^{d_{\text{-}k}} \mathrm{Cov}(E_{t,(k),ij} E_{t,(k),lj}, E_{s,(k),ih} E_{s,(k),lh})$$
$$+ \Big( \sum_{t=1}^{T} \sum_{j=1}^{d_{\text{-}k}} \mathbb{E}[E_{t,(k),ij} E_{t,(k),lj}] \Big)^2 \Big\}$$
$$= O\big( T d_k^2 d_{\text{-}k} \big) + \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} O\Big( T \mathbf{A}_{e,k,i\cdot}' \mathbf{A}_{e,k,l\cdot} \|\mathbf{A}_{e,\text{-}k}\|_F^2 + T d_{\text{-}k} \mathbb{1}_{\{i=l\}} \Big)^2 = O\big( T d_k^2 d_{\text{-}k} + T^2 d_k d_{\text{-}k}^2 \big).$$

Recall from (4.36) it is defined that

$$\mathbf{R}_{k,t} = \mathbf{Q}_k \mathbf{F}_{Z,t,(k)} \mathbf{Q}_{\text{-}k}' \mathbf{E}_{t,(k)}' + \mathbf{E}_{t,(k)} \mathbf{Q}_{\text{-}k} \mathbf{F}_{Z,t,(k)}' \mathbf{Q}_k' + \mathbf{E}_{t,(k)} \mathbf{E}_{t,(k)}',$$

the rate on $\| \sum_{t=1}^T \mathbf{R}_{k,t} \|_F^2$ is direct from (4.32) and (4.33). This completes the proof. □

**Lemma 4.4** *Let Assumptions (F1), (L1), (L2), (E1), (E2) and (R1) hold. For* $k \in [K]$*, define* $\widehat{\mathbf{D}}_k$ *as the diagonal matrix with the first largest* $r_k$ *eigenvalues of* $T^{-1} \sum_{t=1}^T \mathbf{Y}_{t,(k)} \mathbf{Y}_{t,(k)}'$ *on the main diagonal, such that* $\widehat{\mathbf{D}}_k = \widehat{\mathbf{Q}}_k' \big( T^{-1} \sum_{t=1}^T \mathbf{Y}_{t,(k)} \mathbf{Y}_{t,(k)}' \big) \widehat{\mathbf{Q}}_k$*. Define* $\omega_k := d_k^{\delta_{k,r_k} - \delta_{k,1}} g_s$*, then*

$$\big\| \widehat{\mathbf{D}}_k^{-1} \big\|_F = O_P\big( \omega_k^{-1} \big).$$

***Proof of Lemma 4.4.*** Observe that $\widehat{\mathbf{D}}_k$ has size $r_k \times r_k$, it suffices to find the lower bound of $\lambda_{r_k}(\widehat{\mathbf{D}}_k)$. To do this, consider the decomposition

$$\frac{1}{T} \sum_{t=1}^T \mathbf{Y}_{t,(k)} \mathbf{Y}_{t,(k)}' = \frac{1}{T} \sum_{t=1}^T \mathbf{Q}_k \mathbf{F}_{Z,t,(k)} \mathbf{Q}_{\text{-}k}' \mathbf{Q}_{\text{-}k} \mathbf{F}_{Z,t,(k)}' \mathbf{Q}_k' + \frac{1}{T} \sum_{t=1}^T \mathbf{R}_{k,t}, \tag{4.34}$$

which is direct from (4.36). Then for a unit vector $\boldsymbol{\gamma} \in \mathbb{R}^{d_k}$, we can define

$$S_k(\boldsymbol{\gamma}) := \frac{1}{\omega_k} \boldsymbol{\gamma}' \Big( \frac{1}{T} \sum_{t=1}^T \mathbf{Y}_{t,(k)} \mathbf{Y}_{t,(k)}' \Big) \boldsymbol{\gamma} =: S_k^*(\boldsymbol{\gamma}) + \widetilde{S}_k(\boldsymbol{\gamma}), \quad \text{with}$$

$$S_k^*(\boldsymbol{\gamma}) := \frac{1}{\omega_k} \boldsymbol{\gamma}' \Big( \frac{1}{T} \sum_{t=1}^T \mathbf{Q}_k \mathbf{F}_{Z,t,(k)} \mathbf{Q}_{\text{-}k}' \mathbf{Q}_{\text{-}k} \mathbf{F}_{Z,t,(k)}' \mathbf{Q}_k' \Big) \boldsymbol{\gamma}, \quad \widetilde{S}_k(\boldsymbol{\gamma}) := \frac{1}{\omega_k} \boldsymbol{\gamma}' \Big( \frac{1}{T} \sum_{t=1}^T \mathbf{R}_{k,t} \Big) \boldsymbol{\gamma}.$$

Since $\|\boldsymbol{\gamma}\| = 1$, we have by Lemma 4.3,

$$\begin{aligned} |\widetilde{S}_k(\boldsymbol{\gamma})|^2 &\leq \frac{1}{\omega_k^2 T^2} \Big\| \sum_{t=1}^T \mathbf{R}_{k,t} \Big\|_F^2 \\ &= O_P\Big( T^{-1} d_k^{1+2(\delta_{k,1} - \delta_{k,r_k})} d g_s^{-2} + d_k^{2(\delta_{k,1} - \delta_{k,r_k} - 1/2)} d^2 g_s^{-2} \Big) = o_P(1), \end{aligned}$$

where the last equality used Assumption (R1). Next, with Assumption (F1) and Lemma 4.2,

$$\begin{aligned} &\lambda_{r_k}\Big( \frac{1}{T} \sum_{t=1}^T \mathbf{Q}_k \mathbf{F}_{Z,t,(k)} \mathbf{Q}_{\text{-}k}' \mathbf{Q}_{\text{-}k} \mathbf{F}_{Z,t,(k)}' \mathbf{Q}_k' \Big) = \lambda_{r_k}\Big( \frac{1}{T} \sum_{t=1}^T \mathbf{A}_k \mathbf{F}_{t,(k)} \mathbf{A}_{\text{-}k}' \mathbf{A}_{\text{-}k} \mathbf{F}_{t,(k)}' \mathbf{A}_k' \Big) \\ &\geq \lambda_{r_k}(\mathbf{A}_k' \mathbf{A}_k) \cdot \lambda_{r_k}\Big( \frac{1}{T} \sum_{t=1}^T \mathbf{F}_{t,(k)} \mathbf{A}_{\text{-}k}' \mathbf{A}_{\text{-}k} \mathbf{F}_{t,(k)}' \Big) \asymp_P d_k^{\delta_{k,r_k}} \cdot \lambda_{r_k}(\mathrm{tr}(\mathbf{A}_{\text{-}k}' \mathbf{A}_{\text{-}k}) \mathbf{I}_{r_k}) \asymp_P \omega_k. \end{aligned}$$

With this, going back to the decomposition (4.34),

$$\begin{aligned}\omega_k^{-1}\lambda_{r_k}(\widehat{\mathbf{D}}_k) &= \omega_k^{-1}\lambda_{r_k}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_{t,(k)}\mathbf{Y}'_{t,(k)}\Big)\\ &\geq \omega_k^{-1}\lambda_{r_k}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}_k\mathbf{F}_{Z,t,(k)}\mathbf{Q}'_{\text{-}k}\mathbf{Q}_{\text{-}k}\mathbf{F}'_{Z,t,(k)}\mathbf{Q}'_k\Big) - \sup_{\|\gamma\|=1}|\widetilde{S}_k(\gamma)| \asymp_P 1,\end{aligned}$$

implying $\big\|\widehat{\mathbf{D}}_k^{-1}\big\|_F = O_P\big(\lambda_{r_k}^{-1}(\widehat{\mathbf{D}}_k)\big) = O_P\big(\omega_k^{-1}\big)$. This completes the proof of Lemma 4.4. □

**Lemma 4.5** *(Consistency of $\{\widehat{\mathbf{Q}}_k\}_{k\in[K]}$). Let Assumptions (F1), (L1), (L2), (E1), (E2) and (R1) hold. For any $k \in [K]$, define an $r_k \times r_k$ matrix*

$$\mathbf{H}_k := T^{-1}\widehat{\mathbf{D}}_k^{-1}\widehat{\mathbf{Q}}'_k\mathbf{Q}_k\sum_{t=1}^{T}\big(\mathbf{F}_{Z,t,(k)}\mathbf{Q}'_{\text{-}k}\mathbf{Q}_{\text{-}k}\mathbf{F}'_{Z,t,(k)}\big).$$

*As $T, d_1, \ldots, d_K \to \infty$ we have $\mathbf{H}_k$ invertible with $\|\mathbf{H}_k\|_F = O_P(1)$ and*

$$\big\|\widehat{\mathbf{Q}}_k - \mathbf{Q}_k\mathbf{H}'_k\big\|_F^2 = O_P\left\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\left(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\right)\frac{d^2}{g_s^2}\right\}.$$

***Proof of Lemma 4.5.*** First, we may write (4.7) as

$$\mathcal{Y}_t = \mathcal{F}_{Z,t} \times_1 \mathbf{Q}_1 \times_2 \cdots \times_K \mathbf{Q}_K + \mathcal{E}_t. \tag{4.35}$$

For any $k \in [K]$, taking the mode-$k$ unfolding on (4.35), we have

$$\mathbf{Y}_{t,(k)} = \mathbf{Q}_k\mathbf{F}_{Z,t,(k)}\mathbf{Q}'_{\text{-}k} + \mathbf{E}_{t,(k)},$$

where $\mathbf{F}_{Z,t,(k)}$ denotes the mode-$k$ unfolding of $\mathcal{F}_{Z,t}$. Hence,

$$\begin{aligned}\mathbf{Y}_{t,(k)}\mathbf{Y}'_{t,(k)} &= \mathbf{Q}_k\mathbf{F}_{Z,t,(k)}\mathbf{Q}'_{\text{-}k}\mathbf{Q}_{\text{-}k}\mathbf{F}'_{Z,t,(k)}\mathbf{Q}'_k + \mathbf{R}_{k,t}, \text{ where}\\ \mathbf{R}_{k,t} &:= \mathbf{Q}_k\mathbf{F}_{Z,t,(k)}\mathbf{Q}'_{\text{-}k}\mathbf{E}'_{t,(k)} + \mathbf{E}_{t,(k)}\mathbf{Q}_{\text{-}k}\mathbf{F}'_{Z,t,(k)}\mathbf{Q}'_k + \mathbf{E}_{t,(k)}\mathbf{E}'_{t,(k)}.\end{aligned} \tag{4.36}$$

Recall from Lemma 4.4 that $\widehat{\mathbf{D}}_k$ is the $r_k \times r_k$ diagonal matrix with the first largest $r_k$ eigenvalues of $T^{-1}\sum_{t=1}^{T}\mathbf{Y}_{t,(k)}\mathbf{Y}'_{t,(k)}$ on the main diagonal, and since $\widehat{\mathbf{Q}}_k$ consists of the corresponding eigenvectors, we have

$$\widehat{\mathbf{Q}}_k\widehat{\mathbf{D}}_k = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_{t,(k)}\mathbf{Y}'_{t,(k)}\widehat{\mathbf{Q}}_k.$$

With (4.36), we can write the $j$-th row of estimated mode-$k$ factor loading as

$$\widehat{\mathbf{Q}}_{k,j\cdot} = \frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{k,i\cdot}\sum_{t=1}^{T}\left(\mathbf{Y}_{t,(k)}\mathbf{Y}'_{t,(k)}\right)_{ij}$$
$$= \frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{k,i\cdot}\mathbf{Q}'_{k,i\cdot}\sum_{t=1}^{T}\left(\mathbf{F}_{Z,t,(k)}\mathbf{Q}'_{\text{-}k}\mathbf{Q}_{\text{-}k}\mathbf{F}'_{Z,t,(k)}\right)\mathbf{Q}_{k,j\cdot} + \frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{k,i\cdot}\sum_{t=1}^{T}(\mathbf{R}_{k,t})_{ij}.$$

Hence with the definition $\mathbf{H}_k = T^{-1}\widehat{\mathbf{D}}_k^{-1}\widehat{\mathbf{Q}}'_k\mathbf{Q}_k\sum_{t=1}^{T}\left(\mathbf{F}_{Z,t,(k)}\mathbf{Q}'_{\text{-}k}\mathbf{Q}_{\text{-}k}\mathbf{F}'_{Z,t,(k)}\right)$, decompose

$$\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k\mathbf{Q}_{k,j\cdot} = \frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{k,i\cdot}\sum_{t=1}^{T}(\mathbf{R}_{k,t})_{ij} \tag{4.37}$$
$$= \frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\left(\widehat{\mathbf{Q}}_{k,i\cdot} - \mathbf{H}_k\mathbf{Q}_{k,i\cdot}\right)\sum_{t=1}^{T}(\mathbf{R}_{k,t})_{ij} + \frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\mathbf{H}_k\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}(\mathbf{R}_{k,t})_{ij}. \tag{4.38}$$

With the decomposition (4.37), it holds that

$$\left\|\widehat{\mathbf{Q}}_k - \mathbf{Q}_k\mathbf{H}'_k\right\|_F^2 = \sum_{j=1}^{d_k}\left\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k\mathbf{Q}_{k,j\cdot}\right\|^2 = \sum_{j=1}^{d_k}\left\|\frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\widehat{\mathbf{Q}}_{k,i\cdot}\sum_{t=1}^{T}(\mathbf{R}_{k,t})_{ij}\right\|^2$$
$$= \sum_{j=1}^{d_k}\left\|\frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\widehat{\mathbf{Q}}'_k\Big(\sum_{t=1}^{T}\mathbf{R}_{k,t}\Big)_{j\cdot}\right\|^2 \le \frac{1}{T^2}\cdot\|\widehat{\mathbf{D}}_k^{-1}\|_F^2\cdot\|\widehat{\mathbf{Q}}_k\|_F^2\cdot\left\|\sum_{t=1}^{T}\mathbf{R}_{k,t}\right\|_F^2$$
$$= O_P\left(d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\left(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k}\right)\frac{d^2}{g_s^2}\right) = o_P(1), \tag{4.39}$$

where the second last equality used Lemma 4.3 and 4.4, and the last used Assumption (R1).

Before improving the rate of (4.39), we now use it to show $\mathbf{H}_k$ has full rank and $\|\mathbf{H}_k\|_F = O_P(1)$ asymptotically. To this end, it is sufficient to observe

$$\mathbf{I}_{r_k} = \widehat{\mathbf{Q}}'_k\widehat{\mathbf{Q}}_k = \widehat{\mathbf{Q}}'_k\left(\widehat{\mathbf{Q}}_k - \mathbf{Q}_k\mathbf{H}'_k\right) + \widehat{\mathbf{Q}}'_k\mathbf{Q}_k\mathbf{H}'_k = \mathbf{Q}'_k\widehat{\mathbf{Q}}_k\mathbf{H}'_k + o_P(1)$$
$$= \mathbf{H}_k\mathbf{Q}'_k\mathbf{Q}_k\mathbf{H}'_k + o_P(1) = \mathbf{H}_k\mathbf{\Sigma}_{A,k}\mathbf{H}'_k + o_P(1),$$

where the last equality used Assumption (L1) or (L2) and it is immediate that $\mathbf{H}_k$ has full rank asymptotically. Let $\sigma_i(\mathbf{X})$ denote the $i$-th largest singular value for any give matrix $\mathbf{X}$, we have

$$\sigma_1(\mathbf{H}_k)\cdot\sigma_{r_k}(\mathbf{\Sigma}_{A,k})\cdot\sigma_{r_k}(\mathbf{H}_k^T) \le \sigma_1(\mathbf{H}_k)\cdot\sigma_{r_k}(\mathbf{\Sigma}_{A,k}\mathbf{H}_k^T)$$
$$\le \sigma_1(\mathbf{H}_k\mathbf{\Sigma}_{A,k}\mathbf{H}'_k) = O_P(\sigma_1(\mathbf{I}_{r_k})) = O_P(1),$$

which implies $\|\mathbf{H}_k\|_F = O_P(1)$ by Assumption (L1) or (L2).

Consider the decomposition (4.38), we have

$$
\begin{aligned}
&\left\|\widehat{\mathbf{Q}}_k-\mathbf{Q}_k\mathbf{H}_k'\right\|_F^2=\sum_{j=1}^{d_k}\left\|\widehat{\mathbf{Q}}_{k,j\cdot}-\mathbf{H}_k\mathbf{Q}_{k,j\cdot}\right\|^2\\
&=O_P\Bigg(\sum_{j=1}^{d_k}\left\|\frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\mathbf{H}_k\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}(\mathbf{R}_{k,t})_{ij}\right\|^2\\
&\quad+\sum_{j=1}^{d_k}\left\|\frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\sum_{i=1}^{d_k}\left(\widehat{\mathbf{Q}}_{k,i\cdot}-\mathbf{H}_k\mathbf{Q}_{k,i\cdot}\right)\sum_{t=1}^{T}(\mathbf{R}_{k,t})_{ij}\right\|^2\Bigg)\\
&=O_P\Bigg(\sum_{j=1}^{d_k}\left\|\frac{1}{T}\widehat{\mathbf{D}}_k^{-1}\mathbf{H}_k\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}(\mathbf{R}_{k,t})_{ij}\right\|^2\Bigg)\\
&\quad+O_P\Bigg(\frac{1}{T^2}\left\|\widehat{\mathbf{D}}_k^{-1}\right\|_F^2\cdot\left\|\widehat{\mathbf{Q}}_k-\mathbf{Q}_k\mathbf{H}_k'\right\|_F^2\cdot\left\|\sum_{t=1}^{T}\mathbf{R}_{k,t}\right\|_F^2\Bigg)\\
&=O_P\Big(T^{-2}d_k^{2(\delta_{k,1}-\delta_{k,r_k})}g_s^{-2}\Big)\cdot\sum_{j=1}^{d_k}\left(\mathcal{I}_{1,j}+\mathcal{I}_{2,j}+\mathcal{I}_{3,j}\right)+o_P\Big(\left\|\widehat{\mathbf{Q}}_k-\mathbf{Q}_k\mathbf{H}_k'\right\|_F^2\Big),
\end{aligned}
\tag{4.40}
$$

where the last equality used $\|\mathbf{H}_k\|_F=O_P(1)$, Lemma 4.3 and 4.4, and the definitions

$$
\begin{aligned}
\mathcal{I}_{1,j}&:=\left\|\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}\mathbf{Q}_{k,i\cdot}'\mathbf{F}_{Z,t,(k)}\mathbf{Q}_{\text{-}k}'\mathbf{E}_{t,(k),j\cdot}\right\|^2,\\
\mathcal{I}_{2,j}&:=\left\|\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}\mathbf{E}_{t,(k),i\cdot}'\mathbf{Q}_{\text{-}k}\mathbf{F}_{Z,t,(k)}'\mathbf{Q}_{k,j\cdot}\right\|^2,\\
\mathcal{I}_{3,j}&:=\left\|\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}\mathbf{E}_{t,(k),i\cdot}'\mathbf{E}_{t,(k),j\cdot}\right\|^2.
\end{aligned}
$$

By the Cauchy–Schwarz inequality,

$$
\begin{aligned}
\mathcal{I}_{1,j}&=\left\|\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}\mathbf{A}_{k,i\cdot}'\mathbf{F}_{t,(k)}\mathbf{A}_{\text{-}k}'\mathbf{E}_{t,(k),j\cdot}\right\|^2\\
&\le\Big(\sum_{i=1}^{d_k}\left\|\mathbf{Q}_{k,i\cdot}\right\|^2\Big)\Big\{\sum_{i=1}^{d_k}\Big(\sum_{t=1}^{T}\mathbf{A}_{k,i\cdot}'\mathbf{F}_{t,(k)}\mathbf{A}_{\text{-}k}'\mathbf{E}_{t,(k),j\cdot}\Big)^2\Big\}\\
&=\left\|\mathbf{Q}_k\right\|_F^2\cdot\Big\{\sum_{i=1}^{d_k}\|\mathbf{A}_{k,i\cdot}\|^2\Big(\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\frac{\mathbf{A}_{k,i\cdot}'}{\|\mathbf{A}_{k,i\cdot}\|}\mathbf{F}_{t,(k)}\mathbf{A}_{\text{-}k,h\cdot}\mathbf{E}_{t,(k),jh}\Big)^2\Big\}=O_P\Big(Td_k^{\delta_{k,1}}d_{\text{-}k}\Big),
\end{aligned}
$$

where the last equality used Assumptions (L1), (L2) and Lemma 4.1.

Consider $\mathcal{I}_{2,j}$. By Assumptions (E1) and (E2), for any $t \in [T], k \in [K], i \in [j_k], h \in [d_{\text{-}k}]$,

$$E_{t,(k),ih} = \sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-w,(k)}\mathbf{A}_{e,\text{-}k,h\cdot} + \Sigma_{\epsilon,(k),ih}\sum_{w\geq 0} a_{\epsilon,w}X_{\epsilon,t-w,(k),ih}.$$

By Assumptions (F1), (E1) and (E2), we first have

$$\begin{aligned}
&\mathbb{E}\Big\{\Big(\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}(\sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-w,(k)}\mathbf{A}_{e,\text{-}k,h\cdot})\mathbf{A}'_{\text{-}k,h\cdot}\mathbf{F}'_{t,(k)}\mathbf{A}_{k,j\cdot}\Big)^2\Big\}\\
&= \text{Cov}\Big(\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\mathbf{A}'_{\text{-}k,h\cdot}(\sum_{w\geq 0} a_{f,w}\mathbf{X}'_{f,t-w,(k)})\mathbf{A}_{k,j\cdot}(\sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-w,(k)}\mathbf{A}_{e,\text{-}k,h\cdot}),\\
&\qquad \sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\mathbf{A}'_{\text{-}k,h\cdot}(\sum_{w\geq 0} a_{f,w}\mathbf{X}'_{f,t-w,(k)})\mathbf{A}_{k,j\cdot}(\sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-w,(k)}\mathbf{A}_{e,\text{-}k,h\cdot})\Big)\\
&= \sum_{h=1}^{d_{\text{-}k}}\sum_{l=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\sum_{w\geq 0} a_{f,w}^2 a_{e,w}^2\cdot\|\mathbf{A}_{k,j\cdot}\|^2\cdot\|\mathbf{A}_{\text{-}k,h\cdot}\|\cdot\|\mathbf{A}_{\text{-}k,l\cdot}\|\cdot\|\mathbf{A}_{e,\text{-}k,h\cdot}\|\cdot\|\mathbf{A}_{e,\text{-}k,l\cdot}\|\cdot\|\mathbf{A}_{e,k,i\cdot}\|^2\\
&= O(T)\cdot\|\mathbf{A}_{k,j\cdot}\|^2\|\mathbf{A}_{e,k,i\cdot}\|^2.
\end{aligned} \tag{4.41}$$

Similarly, it holds that

$$\begin{aligned}
&\mathbb{E}\Big\{\Big\|\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\mathbf{Q}_{k,i\cdot}\Big(\Sigma_{\epsilon,(k),ih}\sum_{w\geq 0} a_{\epsilon,w}X_{\epsilon,t-w,(k),ih}\Big)\mathbf{A}'_{\text{-}k,h\cdot}\mathbf{F}'_{t,(k)}\mathbf{A}_{k,j\cdot}\Big\|^2\Big\}\\
&= \text{Cov}\Big(\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\mathbf{Q}_{k,i\cdot}\Big(\Sigma_{\epsilon,(k),ih}\sum_{w\geq 0} a_{\epsilon,w}X_{\epsilon,t-w,(k),ih}\Big)\mathbf{A}'_{\text{-}k,h\cdot}\mathbf{F}'_{t,(k)}\mathbf{A}_{k,j\cdot},\\
&\qquad \sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\mathbf{Q}_{k,i\cdot}\Big(\Sigma_{\epsilon,(k),ih}\sum_{w\geq 0} a_{\epsilon,w}X_{\epsilon,t-w,(k),ih}\Big)\mathbf{A}'_{\text{-}k,h\cdot}\mathbf{F}'_{t,(k)}\mathbf{A}_{k,j\cdot}\Big)\\
&= \sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\sum_{w\geq 0} a_{f,w}^2 a_{\epsilon,w}^2\cdot\|\mathbf{A}_{k,j\cdot}\|^2\|\mathbf{A}_{\text{-}k,h\cdot}\|^2\Sigma_{\epsilon,(k),ih}^2\|\mathbf{Q}_{k,i\cdot}\|^2 = O(T)\cdot\|\mathbf{A}_{k,j\cdot}\|^2\|\mathbf{A}_{\text{-}k}\|^2.
\end{aligned} \tag{4.42}$$

Hence using Lemma 4.4, it holds that

$$\begin{aligned}
\mathcal{I}_{2,j} &= \Big\|\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}\mathbf{E}'_{t,(k),i\cdot}\mathbf{A}_{\text{-}k}\mathbf{F}'_{t,(k)}\mathbf{A}_{k,j\cdot}\Big\|^2\\
&= \Big\|\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}E_{t,(k),ih}\mathbf{A}'_{\text{-}k,h\cdot}\mathbf{F}'_{t,(k)}\mathbf{A}_{k,j\cdot}\Big\|^2
\end{aligned}$$

$$
\begin{aligned}
&\leq 2\Big\{\Big\|\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\mathbf{Q}_{k,i\cdot}\Big(\Sigma_{\epsilon,(k),ih}\sum_{w\geq 0}a_{\epsilon,w}X_{\epsilon,t-w,(k),ih}\Big)\mathbf{A}'_{\text{-}k,h\cdot}\mathbf{F}'_{t,(k)}\mathbf{A}_{k,j\cdot}\Big\|^2\Big\}\\
&\quad+2\Big\{\|\mathbf{Q}_k\|^2\cdot\sum_{i=1}^{d_k}\Big(\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}(\sum_{w\geq 0}a_{e,w}\mathbf{A}'_{e,k,i\cdot}\mathbf{X}_{e,t-w,(k)}\mathbf{A}_{e,\text{-}k,h\cdot})\mathbf{A}'_{\text{-}k,h\cdot}\mathbf{F}'_{t,(k)}\mathbf{A}_{k,j\cdot}\Big)^2\Big\}\\
&=O_P\Big(\|\mathbf{A}_{k,j\cdot}\|^2\cdot Td_k^{-\delta_{k,1}}g_s\Big)
\end{aligned}
$$

where we used (4.41) and (4.42) in the last equality.

For $\mathcal{I}_{3,j}$, let $r_{e,\text{-}k}:=\prod_{p\neq k}r_{e,p}$. By the noise structure in Assumptions (E1) and (E2),

$$
\begin{aligned}
&\operatorname{Var}\Big(\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\mathbf{Q}_{k,i\cdot}E_{t,(k),ih}E_{t,(k),jh}\Big)\\
&=O\Big(\sum_{i=1}^{d_k}\sum_{u=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{l=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\sum_{n=1}^{r_{e,k}}\sum_{m=1}^{r_{e,\text{-}k}}\sum_{w\geq 0}a_{e,w}^4A_{e,k,in}A_{e,k,un}A_{e,k,jn}^2A_{e,\text{-}k,hm}^2A_{e,\text{-}k,lm}^2\\
&\quad\cdot\|\mathbf{Q}_{k,i\cdot}\|\cdot\|\mathbf{Q}_{k,u\cdot}\|\cdot\operatorname{Var}(X_{e,t-w,(k),nm}^2)\Big)\\
&\quad+O\Big(\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\sum_{w\geq 0}a_{\epsilon,w}^4\Sigma_{\epsilon,(k),ih}^2\Sigma_{\epsilon,(k),jh}^2\cdot\|\mathbf{Q}_{k,i\cdot}\|^2\cdot\operatorname{Var}(X_{e,t-w,(k),ih}X_{e,t-w,(k),jh})\\
&=O(T+Td_{\text{-}k})=O(Td_{\text{-}k}).
\end{aligned}
$$

Moreover, it also holds that

$$
\begin{aligned}
&\mathbb{E}\Big(\sum_{i=1}^{d_k}\Big|\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}E_{t,(k),ih}E_{t,(k),jh}\Big|\Big)\\
&=\sum_{i=1}^{d_k}\Big|\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}\Big(\|\mathbf{A}_{e,\text{-}k,h\cdot}\|^2\cdot\|\mathbf{A}_{e,k,i\cdot}\|\cdot\|\mathbf{A}_{e,k,j\cdot}\|+\Sigma_{\epsilon,(k),ih}\mathbb{1}_{\{i=j\}}\Big)\Big|=O(Td_{\text{-}k}),
\end{aligned}
$$

together with $\max_i\|\mathbf{Q}_{k,i\cdot}\|^2\leq\|\mathbf{A}_{k,j\cdot}\|^2\cdot\|\mathbf{Z}_k^{-1/2}\|^2=O_P\big(d_k^{-\delta_{k,r_k}}\big)$, we arrive at

$$
\mathcal{I}_{3,j}=\Big\|\sum_{i=1}^{d_k}\sum_{h=1}^{d_{\text{-}k}}\mathbf{Q}_{k,i\cdot}\sum_{t=1}^{T}E_{t,(k),ih}E_{t,(k),jh}\Big\|^2=O_P\Big(Td_{\text{-}k}+T^2d_k^{-\delta_{k,r_k}}d_{\text{-}k}^2\Big).
$$

Finally, for (4.40) we have

$$
\begin{aligned}
&\big\|\widehat{\mathbf{Q}}_k-\mathbf{Q}_k\mathbf{H}'_k\big\|_F^2\\
&=O_P\Big\{T^{-2}d_k^{2(\delta_{k,1}-\delta_{k,r_k})}g_s^{-2}\cdot\Big(Tdd_k^{\delta_{k,1}}+T^2d_k^{1-\delta_{k,r_k}}d_{\text{-}k}^2\Big)\Big\}+o_P\Big(\big\|\widehat{\mathbf{Q}}_k-\mathbf{Q}_k\mathbf{H}'_k\big\|_F^2\Big)\\
&=O_P\bigg\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\bigg(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}}+\frac{1}{d_k^{1+\delta_{k,r_k}}}\bigg)\frac{d^2}{g_s^2}\bigg\}.
\end{aligned}
$$

This completes the proof of Lemma 4.5. □

**Lemma 4.6** *(Consistency of $\widehat{\mathcal{C}}_t$). Let all the assumptions in Lemma 4.5 hold. With $\{\mathbf{Z}_k\}_{k\in[K]}$ from Assumptions (L1) and (L2) and $\{\mathbf{H}_k\}_{k\in[K]}$ from the statement of Lemma 4.5, define*

$$\mathbf{Z}_\otimes := \mathbf{Z}_K \otimes \cdots \otimes \mathbf{Z}_1, \quad \mathbf{H}_\otimes := \mathbf{H}_K \otimes \cdots \otimes \mathbf{H}_1, \quad \mathbf{vec}\big(\widehat{\mathcal{F}}_t\big) := \big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big)'\mathbf{vec}\big(\mathcal{Y}_t\big).$$

*Then the estimators of the vectorised (renormalised) core factor and the $i$-th entry of the vectorised common component for* (4.7) *are consistent such that*

$$\begin{aligned}
&\big\|\mathbf{vec}\big(\widehat{\mathcal{F}}_t\big) - (\mathbf{H}_\otimes')^{-1}\mathbf{vec}\big(\mathcal{F}_{Z,t}\big)\big\|^2 = \big\|\mathbf{vec}\big(\widehat{\mathcal{F}}_t\big) - (\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big)\big\|^2 \\
&= O_P\left(\max_{k\in[K]}\left\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\left(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\right)\frac{d^2}{g_s}\right\} + \frac{d}{g_w}\right), \qquad (4.43)\\
&\Big\{\big(\mathbf{vec}\big(\widehat{\mathcal{C}}_t\big)\big)_i - \big(\mathbf{vec}\big(\mathcal{C}_t\big)\big)_i\Big\}^2 \\
&= O_P\left(\max_{k\in[K]}\left\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\left(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\right)\frac{d^2}{g_s g_w}\right\} + \frac{d}{g_w^2}\right). \qquad (4.44)
\end{aligned}$$

***Proof of Lemma 4.6.*** By (4.7), we have

$$\begin{aligned}
&\mathbf{vec}\big(\widehat{\mathcal{F}}_t\big) - (\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big) = \big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big)'\mathbf{vec}\big(\mathcal{Y}_t\big) - (\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big) \\
&= \big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big)'\mathbf{vec}\big(\mathcal{F}_t \times_{k=1}^K \mathbf{A}_k + \mathcal{E}_t\big) - (\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big) \\
&= \big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big)'\big\{\big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big) - \big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big)\big\}(\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big) \\
&\quad + \big\{\big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big\}'\mathbf{vec}\big(\mathcal{E}_t\big) \\
&\quad + \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)'\mathbf{vec}\big(\mathcal{E}_t\big) =: \mathcal{I}_{f,1} + \mathcal{I}_{f,2} + \mathcal{I}_{f,3}.
\end{aligned} \qquad (4.45)$$

We first show by an induction argument that for any positive integer $K$,

$$\big\|\big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big\|_F = O_P\big(\max_{k\in[K]}\big\|\widehat{\mathbf{Q}}_k - \mathbf{Q}_k\mathbf{H}_k'\big\|_F\big). \qquad (4.46)$$

The initial case for $K = 1$ is trivial. Suppose (4.46) holds for $K - 1$, then

$$\begin{aligned}
&\big\|\big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big\|_F \\
&= \big\|\big(\widehat{\mathbf{Q}}_K - \mathbf{Q}_K\mathbf{H}_K'\big) \otimes \big(\widehat{\mathbf{Q}}_{K-1} \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) \\
&\quad + \mathbf{Q}_K\mathbf{H}_K' \otimes \big\{\big(\widehat{\mathbf{Q}}_{K-1} \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_{K-1}\mathbf{H}_{K-1}' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big\}\big\|_F \\
&= O_P\Big(\big\|\widehat{\mathbf{Q}}_K - \mathbf{Q}_K\mathbf{H}_K'\big\|_F + \big\|\big(\widehat{\mathbf{Q}}_{K-1} \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_{K-1}\mathbf{H}_{K-1}' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big\|_F\Big),
\end{aligned}$$

which concludes (4.46). Hence for $\mathcal{I}_{f,1}$, with Lemma 4.5 we immediately have

$$\begin{aligned}\|\mathcal{I}_{f,1}\|^2 &= \Big\|\big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big)'\big\{\big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big) - \big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big)\big\} \\ &\quad \cdot (\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big)\Big\|^2 \\ &= O_P\Big(\big\|\big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big\|_F^2 \cdot \big\|\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big)\big\|^2\Big) \\ &= O_P\left(\max_{k\in[K]}\left\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\left(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\right)\frac{d^2}{g_s}\right\}\right),\end{aligned} \tag{4.47}$$

where the last equality used (4.46), Assumptions (F1), (L1) and (L2).

For $\mathcal{I}_{f,2}$, observe first throughout the proof of Lemma 4.5, the consistency of $\widehat{\mathbf{Q}}_{k,j\cdot}$ for $k \in [K], j \in [d_k]$ can be shown with the same argument (omitted), before eventually being aggregated over all $d_k$ rows. That is,

$$\big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k\mathbf{Q}_{k,j\cdot}\big\|^2 = O_P\left\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})-1}\left(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\right)\frac{d^2}{g_s^2}\right\}. \tag{4.48}$$

Then we have $\big\|\widehat{\mathbf{Q}}_{k,j\cdot}\big\|^2 = O_P\big(\big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k\mathbf{Q}_{k,j\cdot}\big\|^2 + \big\|\mathbf{H}_k\mathbf{Z}_k^{-1/2}\mathbf{A}_{k,j\cdot}\big\|^2\big) = O_P\big(d_k^{-\delta_{k,r_k}}\big)$ which used (4.48), Assumptions (L1) (or (L2)) and (R1). Note that this rate $d_k^{-\delta_{k,r_k}}$ is the same as the one for $\big\|\mathbf{Q}_{k,j\cdot}\big\|^2$, shown in the proof of Lemma 4.5. Moreover, it holds that for any positive integer $k$ that

$$\begin{aligned}&\max_{\ell\in[d]}\Big\|\big[\big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big]_{\ell\cdot}\Big\|^2 \\ &= O\Big(\max_{k\in[K]}\Big\{\max_{j\in[d_k]}\big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k\mathbf{Q}_{k,j\cdot}\big\|^2 \prod_{j\in[K]\setminus\{k\}}\max_{i\in[d_j]}\big\|\widehat{\mathbf{Q}}_{j,i\cdot}\big\|^2\Big\}\Big),\end{aligned} \tag{4.49}$$

which can be shown by an induction argument for which the initial case for $K = 1$ is trivial, and the induction step is seen by

$$\begin{aligned}&\max_{\ell\in[d]}\Big\|\big[\big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big]_{\ell\cdot}\Big\|^2 \\ &= O\Bigg(\max_{j\in[d_K]}\big\|\widehat{\mathbf{Q}}_{K,j\cdot} - \mathbf{H}_K\mathbf{Q}_{K,j\cdot}\big\|^2 \cdot \prod_{k=1}^{K-1}\max_{i\in[d_k]}\big\|\widehat{\mathbf{Q}}_{k,i\cdot}\big\|^2 \\ &\quad + \max_{i\in[d_K]}\|\mathbf{Q}_{K,i\cdot}\|^2 \max_{i\in\left[\prod_{k=1}^{K-1} d_k\right]}\Big\|\big[\big(\widehat{\mathbf{Q}}_{K-1} \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_{K-1}\mathbf{H}_{K-1}' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)\big]_{i\cdot}\Big\|^2\Bigg).\end{aligned}$$

Hence, for $\mathcal{I}_{f,2}$ we have

$$\begin{aligned}
\|\mathcal{I}_{f,2}\|^2 &= \Big\| \big\{ \big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big) \big\}' \mathbf{vec}\big(\mathcal{E}_t\big) \Big\|^2 \\
&= \max_{\ell \in [d]} \Big\| \big[ \big(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1\big) - \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big) \big]_{\ell \cdot} \Big\|^2 \\
&\quad \cdot \sum_{j,l=1}^{d} \Big| \mathbb{E}\big(\mathbf{vec}\big(\mathcal{E}_t\big)\big)_j \big(\mathbf{vec}\big(\mathcal{E}_t\big)\big)_l \Big| \\
&= O_P\Bigg( \max_{k \in [K]} \Bigg\{ d_k^{2\delta_{k,1} - \delta_{k,r_k}} \Bigg( \frac{1}{T d_{\text{-}k} d_k^{2-\delta_{k,1}}} + \frac{1}{d_k^{2+\delta_{k,r_k}}} \Bigg) \frac{d^3}{g_s^2 g_w} \Bigg\} \Bigg),
\end{aligned} \tag{4.50}$$

where the last equality used (4.48), (4.49) and the first result on Lemma 4.1.1.

Lastly, consider $\mathcal{I}_{f,3}$. With Assumptions (L1) and (L2), we have

$$\begin{aligned}
\|\mathcal{I}_{f,3}\|^2 &= \Big\| \big(\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1'\big)' \mathbf{vec}\big(\mathcal{E}_t\big) \Big\|^2 = O_P\Big( \big\| \big(\mathbf{Q}_K' \otimes \cdots \otimes \mathbf{Q}_1'\big) \mathbf{vec}\big(\mathcal{E}_t\big) \big\|^2 \Big) \\
&= O_P\Big( \big\| \mathbf{Z}_\otimes^{-1/2} \big\|_F^2 \cdot \big\| \big(\mathbf{A}_K' \otimes \cdots \otimes \mathbf{A}_1'\big) \mathbf{vec}\big(\mathcal{E}_t\big) \big\|^2 \Big) \\
&= O_P\Big\{ g_w^{-1} \cdot \sum_{j=1}^{d} \big\| \big(\mathbf{vec}\big(\mathcal{E}_t\big)\big)_j \big(\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_1\big)_{j\cdot} \big\|^2 \Big\} = O_P(d/g_w),
\end{aligned} \tag{4.51}$$

where the last equality used (L1) and (L2) and the first result on Lemma 4.1.1 that

$$\begin{aligned}
&\mathbb{E}\Big\{ \sum_{j=1}^{d} \big\| \big(\mathbf{vec}\big(\mathcal{E}_t\big)\big)_j \big(\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_1\big)_{j\cdot} \big\|^2 \Big\} \\
&\le \max_{j \in [d]} \big\| \big(\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_1\big)_{j\cdot} \big\|^2 \sum_{j,l=1}^{d} \Big| \mathbb{E}\big(\mathbf{vec}\big(\mathcal{E}_t\big)\big)_j \big(\mathbf{vec}\big(\mathcal{E}_t\big)\big)_l \Big| = O(d).
\end{aligned}$$

Combining (4.45), (4.47), (4.50) and (4.51), we obtain (4.43).

It remains to show (4.44). To this end, from (4.7) and (4.11) (where $\widehat{\mathcal{C}}_{m,t}$ is simplified as $\widehat{\mathcal{C}}_t$, explained in the summary of proofs), we have

$$\begin{aligned}
&\mathbf{vec}\big(\widehat{\mathcal{C}}_t\big) - \mathbf{vec}\big(\mathcal{C}_t\big) = (\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1)\mathbf{vec}\big(\widehat{\mathcal{F}}_t\big) - \mathbf{vec}\big(\mathcal{F}_t \times_{k=1}^{K} \mathbf{A}_k\big) \\
&= (\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1)\mathbf{vec}\big(\widehat{\mathcal{F}}_t\big) - (\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1')(\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big) \\
&= (\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1)\Big\{ \mathbf{vec}\big(\widehat{\mathcal{F}}_t\big) - (\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big) \Big\} \\
&\quad + \Big\{ (\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_1) - (\mathbf{Q}_K\mathbf{H}_K' \otimes \cdots \otimes \mathbf{Q}_1\mathbf{H}_1') \Big\} (\mathbf{H}_\otimes')^{-1}\mathbf{Z}_\otimes^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big),
\end{aligned}$$

which implies that

$$\begin{aligned}
&\Big\{\big(\mathbf{vec}(\widehat{\mathcal{C}}_t)\big)_i - \big(\mathbf{vec}(\mathcal{C}_t)\big)_i\Big\}^2 \\
&= O_P\Bigg(\prod_{k=1}^{K}\max_{j\in[d_k]}\big\|\widehat{\mathbf{Q}}_{k,j\cdot}\big\|^2\cdot\big\|\mathbf{vec}\big(\widehat{\mathcal{F}}_t\big)-(\mathbf{H}'_{\otimes})^{-1}\mathbf{Z}_{\otimes}^{1/2}\mathbf{vec}\big(\mathcal{F}_t\big)\big\|^2 \\
&\qquad + \max_{\ell\in[d]}\Big\|\big[\big(\widehat{\mathbf{Q}}_K\otimes\cdots\otimes\widehat{\mathbf{Q}}_1\big)-\big(\mathbf{Q}_K\mathbf{H}'_K\otimes\cdots\otimes\mathbf{Q}_1\mathbf{H}'_1\big)\big]_{\ell\cdot}\Big\|^2\cdot\big\|\mathbf{Z}_{\otimes}^{1/2}\big\|_F^2\Bigg) \\
&= O_P\left(\max_{k\in[K]}\left\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\left(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}}+\frac{1}{d_k^{1+\delta_{k,r_k}}}\right)\frac{d^2}{g_s g_w}\right\}+\frac{d}{g_w^2}\right),
\end{aligned}$$

where the last equality used (4.43), (4.48), (4.49), Assumptions (L1) and (L2). This completes the proof of Lemma 4.6. □

**Lemma 4.7** *(Consistency of $\{\widetilde{\mathbf{Q}}_j\}_{j\in[v-1]}$, $\widetilde{\mathbf{Q}}_V$ and $\widetilde{\mathcal{C}}_{\text{reshape},t}$). Let Assumptions (F1), (L1), (E1), (E2) and (R1) hold, and consider the model* (4.5)*. For $j\in[v-1]$, define $\widetilde{\mathbf{D}}_j$ as the $r_j\times r_j$ diagonal matrix with the first largest $r_j$ eigenvalues of*

$$\frac{1}{T}\sum_{t=1}^{T}\text{RESHAPE}(\mathcal{Y}_t,\mathcal{A})_{(j)}\text{RESHAPE}(\mathcal{Y}_t,\mathcal{A})'_{(j)}$$

*on the main diagonal such that $\widetilde{\mathbf{D}}_j=\widetilde{\mathbf{Q}}'_j\big(T^{-1}\sum_{t=1}^{T}\text{RESHAPE}(\mathcal{Y}_t,\mathcal{A})_{(j)}\text{RESHAPE}(\mathcal{Y}_t,\mathcal{A})'_{(j)}\big)\widetilde{\mathbf{Q}}_j$. Similarly, $\widetilde{\mathbf{D}}_V$ denotes the $r_V\times r_V$ diagonal matrix with the first largest $r_V$ eigenvalues of*

$$\frac{1}{T}\sum_{t=1}^{T}\text{RESHAPE}(\mathcal{Y}_t,\mathcal{A})_{(v)}\text{RESHAPE}(\mathcal{Y}_t,\mathcal{A})'_{(v)}$$

*on the main diagonal. Correspondingly, define an $r_j\times r_j$ matrix and an $r_V\times r_V$ matrix that*

$$\begin{aligned}
\widetilde{\mathbf{H}}_j &:= T^{-1}\widetilde{\mathbf{D}}_j^{-1}\widetilde{\mathbf{Q}}'_j\mathbf{Q}_j\sum_{t=1}^{T}\mathbf{F}_{\text{reshape},Z,t,(j)}\mathbf{F}'_{\text{reshape},Z,t,(j)}, \\
\widetilde{\mathbf{H}}_V &:= T^{-1}\widetilde{\mathbf{D}}_V^{-1}\widetilde{\mathbf{Q}}'_V\mathbf{Q}_V\sum_{t=1}^{T}\mathbf{F}_{\text{reshape},Z,t,(v)}\mathbf{F}'_{\text{reshape},Z,t,(v)},
\end{aligned}$$

*where $\mathcal{F}_{\text{reshape},Z,t}:=\mathcal{F}_{\text{reshape},t}\times_{j=1}^{v-1}\mathbf{Z}_j^{1/2}\times_v\mathbf{Z}_V^{1/2}$. As $T,d_1,\ldots,d_{v-1},d_V\to\infty$ we have $\{\widetilde{\mathbf{H}}_j\}_{j\in[v-1]}$, $\widetilde{\mathbf{H}}_V$ are invertible, and for $j\in[v-1]$ that $\|\widetilde{\mathbf{H}}_j\|_F=O_P(1)$ and $\|\widetilde{\mathbf{H}}_V\|_F=O_P(1)$. For each $j\in[v-1]$,*

$$\big\|\widetilde{\mathbf{Q}}_j-\mathbf{Q}_j\widetilde{\mathbf{H}}'_j\big\|_F^2=O_P\left\{d_j^{2(\delta_{j,1}-\delta_{j,r_j})}\left(\frac{1}{Td_{\text{-}k}d_j^{1-\delta_{j,1}}}+\frac{1}{d_j^{1+\delta_{j,r_j}}}\right)\frac{d^2}{\gamma_s^2}\right\}, \tag{4.52}$$

$$\big\|\widetilde{\mathbf{Q}}_V - \mathbf{Q}_V\widetilde{\mathbf{H}}_V'\big\|_F^2 = O_P\bigg\{d_V^{2(\delta_{V,1}-\delta_{V,r_V})}\bigg(\frac{1}{Tdd_V^{-\delta_{V,1}}} + \frac{1}{d_V^{1+\delta_{V,r_V}}}\bigg)\frac{d^2}{\gamma_s^2}\bigg\}, \tag{4.53}$$

*where $\gamma_s$ is defined in Assumption (R1). Lastly, the $i$-th entry of the vectorised common component in* (4.5) *is also consistent such that*

$$\begin{aligned}&\Big\{\big(\mathbf{vec}(\widetilde{\mathcal{C}}_{\text{reshape},t})\big)_i - \big(\mathbf{vec}(\mathcal{C}_{\text{reshape},t})\big)_i\Big\}^2\\ &= O_P\Bigg(\max_{j\in[v-1]}\bigg\{d_j^{2(\delta_{j,1}-\delta_{j,r_j})}\bigg(\frac{1}{Td_{\text{-}k}d_j^{1-\delta_{j,1}}} + \frac{1}{d_j^{1+\delta_{j,r_j}}}\bigg)\frac{d^2}{\gamma_s\gamma_w}\bigg\}\\ &\quad + d_V^{2(\delta_{V,1}-\delta_{V,r_V})}\bigg(\frac{1}{Tdd_V^{-\delta_{V,1}}} + \frac{1}{d_V^{1+\delta_{V,r_V}}}\bigg)\frac{d^2}{\gamma_s\gamma_w} + \frac{d}{\gamma_w^2}\Bigg),\end{aligned} \tag{4.54}$$

*where $\gamma_w$ is defined in Assumption (R2).*

***Proof of Lemma 4.7.*** Consider (4.21), we have by Assumptions (E1) and (E2) that

$$\begin{aligned}\textsc{reshape}(\mathcal{F}_{e,t}, \mathcal{A}) &= \sum_{q\geq 0} a_{e,q}\textsc{reshape}(\mathcal{X}_{e,t-q}, \mathcal{A}),\\ \textsc{reshape}(\boldsymbol{\epsilon}_t, \mathcal{A}) &= \sum_{q\geq 0} a_{\epsilon,q}\textsc{reshape}(\mathcal{X}_{\epsilon,t-q}, \mathcal{A}),\end{aligned}$$

which implies that the structure depicted in (E1) and (E2) for the noise $\mathcal{E}_t$ in (4.7) holds for $\mathcal{E}_{\text{reshape},t}$ in (4.5). Read $\textsc{reshape}(\mathcal{Y}_t, \mathcal{A})$ as an order-$v$ tensor and consider the factor model (4.5). Statements (4.52) and (4.53) can be shown in exactly the same way as Lemma 4.5 (without (L2) now; details omitted here), given that all rate conditions used in the proof of Lemma 4.5 are fulfilled for (4.5). In other words, it remains to show the rate conditions similar to the last equality of (4.39) are satisfied. For $j \in [v-1]$, (4.52) requires the same rate condition as Lemma 4.5. Hence we are left with the rate conditions for (4.53), i.e.,

$$d\gamma_s^{-2}T^{-1}d_V^{2(\delta_{V,1}-\delta_{V,r_V})+1} = o(1), \quad d\gamma_s^{-1}d_V^{\delta_{V,1}-\delta_{V,r_V}-1/2} = o(1),$$

which are included in Assumption (R1). With (4.52) and (4.53) shown, (4.54) follows similarly as Lemma 4.6 (omitted). The proof of Lemma 4.7 is then complete. □

**Lemma 4.8** *Let Assumptions (F1), (E1) and (E2) hold. Then with the sub-Gaussian tail assumption in the statement of Theorem 4.3, for any $k \in [K]$ and any deterministic vectors $\mathbf{u} \in \mathbb{R}^{r_k}$ and $\mathbf{v} \in \mathbb{R}^{r_{\text{-}k}}$ with constant magnitudes, for $\mathcal{F}_t$ in* (4.7) *we have*

$$\max_{j\in[p]}\frac{1}{(d_{\text{-}k}T)^{1/2}}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} E_{t,(k),jh}\mathbf{u}'\mathbf{F}_{t,(k)}\mathbf{v} = O_P\{\log(d_k)\}.$$

*The result holds similarly for* $\textsc{reshape}(\mathcal{F}_t, \mathcal{A})$ *and* $\textsc{reshape}(\mathcal{E}_t, \mathcal{A})$.

***Proof of Lemma 4.8.*** From Assumption (F1), (E1) and (E2), we may rewrite $E_{t,(k),jh}$ and $\mathbf{u}'\mathbf{F}_{t,(k)}\mathbf{v}$ as

$$E_{t,(k),jh} = \mathbf{A}'_{e,r,j\cdot}\Big(\sum_{w\geq 0} a_{e,w}\mathbf{X}_{e,t-w,(k)}\Big)\mathbf{A}_{e,c,h\cdot} + \Sigma_{\epsilon,(k),jh}\Big(\sum_{g\geq 0} a_{\epsilon,g}X_{\epsilon,t-g,(k),jh}\Big),$$
$$\mathbf{u}'\mathbf{F}_{t,(k)}\mathbf{v} = \sum_{m=1}^{r_k}\sum_{n=1}^{r_{\text{-}k}}\sum_{l\geq 0} a_{f,l}X_{f,t-l,(k),mn}u_m v_n,$$

so that for any $j \in [d_k]$,

$$\begin{aligned}
&\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} E_{t,(k),jh}\mathbf{u}'\mathbf{F}_{t,(k)}\mathbf{v} \\
&= \sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} \mathbf{A}'_{e,r,j\cdot}\Big(\sum_{w\geq 0} a_{e,w}\mathbf{X}_{e,t-w,(k)}\Big)\mathbf{A}_{e,c,h\cdot}\sum_{m=1}^{r_k}\sum_{n=1}^{r_{\text{-}k}}\sum_{l\geq 0} a_{f,l}X_{f,t-l,(k),mn}u_m v_n \\
&\quad + \sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} \Sigma_{\epsilon,(k),jh}\Big(\sum_{g\geq 0} a_{\epsilon,g}X_{\epsilon,t-g,(k),jh}\Big)\sum_{m=1}^{r_k}\sum_{n=1}^{r_{\text{-}k}}\sum_{l\geq 0} a_{f,l}X_{f,t-l,(k),mn}u_m v_n.
\end{aligned} \tag{4.55}$$

Consider first the second term above, i.e.,

$$\sum_{m=1}^{r_k}\sum_{n=1}^{r_{\text{-}k}} u_m v_n \sum_{l\geq 0}\sum_{g\geq 0} a_{f,l}a_{\epsilon,g}\sum_{t=1}^{T}\Big(\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}X_{\epsilon,t-g,(k),jh}\Big)X_{f,t-l,(k),mn}.$$

Fix $l \geq 0, g \geq 0$. By the sub-Gaussian tail assumption in the statement of Theorem 4.3, for each $t \in [T]$, we have $\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}X_{\epsilon,t-g,(k),jh} \sim \text{subG}(C_1 d_{\text{-}k})$, with arbitrary constant $C_1 > 0$ such that $C_1 q = \sum_{h=1}^{d_{\text{-}k}}\Sigma^2_{\epsilon,(k),jh}$, which is independent over $g$. Notice that $X_{f,t-l,(k),mn} \sim \text{subG}(1)$ by the sub-Gaussian tail, then

$$(\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}X_{\epsilon,t-g,(k),jh})X_{f,t-l,(k),mn} \sim \text{subE}(\sqrt{C_1 d_{\text{-}k}})$$

which is independent over $t$, and hence

$$\sum_{t=1}^{T}(\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}X_{\epsilon,t-g,(k),jh})X_{f,t-l,(k),mn} \sim \text{subE}(\sqrt{C_1 d_{\text{-}k}T}).$$

Then sum those sub-exponential random variables over $l \geq 0$, $g \geq 0$, we have by (E2),

$$\sum_{l\geq 0}\sum_{g\geq 0} a_{f,l}a_{\epsilon,g}\sum_{t=1}^{T}\Big(\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}X_{\epsilon,t-g,(k),jh}\Big)X_{f,t-l,(k),mn} \sim \text{subE}\big(\sqrt{C_2 d_{\text{-}k}T}\big),$$

with some arbitrary constant $C_2 > 0$. As $r_k$, $r_{\text{-}k}$, $r_{e,k}$ and $r_{e,\text{-}k}$ are all constants, we conclude that the entire second term in (4.55), together with the first term therein, are also sub-exponential with parameter of the rate $\sqrt{d_{\text{-}k}T}$. Therefore, for each $j \in [d_k]$, it holds that

$$(d_{\text{-}k}T)^{-1/2}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} E_{t,(k),jh}\mathbf{u}'\mathbf{F}_{t,(k)}\mathbf{v}$$

is sub-exponential with parameter of constant rate. Using the union bound, with some arbitrary constant $C_3 > 0$, we have

$$\mathbb{P}\Big(\max_{j\in[d_k]}\frac{1}{(d_{\text{-}k}T)^{1/2}}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} E_{t,(k),jh}\mathbf{u}'\mathbf{F}_{t,(k)}\mathbf{v} \geq \varepsilon\Big) \leq \exp\big\{\log(d_k) - C_3\varepsilon\big\}, \tag{4.56}$$

implying that $\max_{j\in[d_k]}(d_{\text{-}k}T)^{-1/2}\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} E_{t,(k),jh}\mathbf{u}'\mathbf{F}_{t,(k)}\mathbf{v} = O_P\{\log(d_k)\}$. This concludes the proof for the display in the lemma, and such a result for $\text{RESHAPE}(\mathcal{F}_t, \mathcal{A})$ and $\text{RESHAPE}(\mathcal{E}_t, \mathcal{A})$ follows trivially by treating $\text{RESHAPE}(\mathcal{F}_t, \mathcal{A})$ and $\text{RESHAPE}(\mathcal{E}_t, \mathcal{A})$ as another tensor core factor and noise which have the same structure as $\mathcal{F}_t$ and $\mathcal{E}_t$, respectively. This concludes the proof for the lemma. □

**Lemma 4.9** *Let all assumptions in Lemma 4.5 hold, and let the sub-Gaussian tail assumption in the statement of Theorem 4.3 also hold. Then with $\mathbf{R}_{k,t}$ defined in* (4.36) *for any $k \in [K]$, we have*

$$\max_{j\in[d_k]}\Big\|\Big(\sum_{t=1}^{T}\mathbf{R}_{k,t}\Big)_{\cdot j}\Big\|^2 = O_P\big\{(Td + T^2 d_{\text{-}k}^2)\log^2(d_k)\big\}.$$

***Proof of Lemma 4.9.*** Essentially, we need to show similar results in Lemma 4.3. To this end, we show the corresponding versions of (4.32) and (4.33). To start with, using Lemma 4.8, we have for any $k \in [K]$,

$$\max_{j\in[d_k]}\Big\|\Big(\sum_{t=1}^{T}\mathbf{Q}_k F_{Z,t,(k)}\mathbf{Q}'_{\text{-}k}\mathbf{E}'_{t,(k)}\Big)_{\cdot j}\Big\|_F^2$$
$$= \sum_{i=1}^{d_k}\|\mathbf{A}_{k,i\cdot}\|^2\max_{j\in[d_k]}\Big(\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T} E_{t,(k),jh}\frac{1}{\|\mathbf{A}_{k,i\cdot}\|}\mathbf{A}'_{k,i\cdot}\mathbf{F}_{t,(k)}\mathbf{A}_{\text{-}k,h\cdot}\Big)^2 = O_P\big\{Td_k^{\delta_{k,1}}d_{\text{-}k}\log^2(d_k)\big\}.$$

Similarly,

$$\begin{aligned}
&\max_{j\in[d_k]}\Big\|\big(\sum_{t=1}^{T}\mathbf{E}_{t,(k)}\mathbf{Q}_{\text{-}k}F'_{Z,t,(k)}\mathbf{Q}'_k\big)_{\cdot j}\Big\|_F^2\\
&=\max_{j\in[d_k]}\Big\{\|\mathbf{A}_{k,j\cdot}\|^2\cdot\sum_{l=1}^{d_k}\Big(\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}E_{t,(k),lh}\frac{1}{\|\mathbf{A}_{k,j\cdot}\|}\mathbf{A}'_{k,j\cdot}\mathbf{F}_{t,(k)}\mathbf{A}_{\text{-}k,h\cdot}\Big)^2\Big\}\\
&\le\max_{j\in[d_k]}\|\mathbf{A}_{k,j\cdot}\|^2\max_{j\in[d_k]}\sum_{l=1}^{d_k}\Big(\sum_{h=1}^{d_{\text{-}k}}\sum_{t=1}^{T}E_{t,(k),lh}\frac{1}{\|\mathbf{A}_{k,j\cdot}\|}\mathbf{A}'_{k,j\cdot}\mathbf{F}_{t,(k)}\mathbf{A}_{\text{-}k,h\cdot}\Big)^2=O_P\big\{Td\log^2(d_k)\big\}.
\end{aligned}$$

Next, consider

$$\max_{j\in[d_k]}\Big\|\sum_{t=1}^{T}\big(\mathbf{E}_{t,(k)}\mathbf{E}'_{t,(k)}\big)_{\cdot j}\Big\|_F^2=\max_{j\in[d_k]}\sum_{i=1}^{d_k}\Big(\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}E_{t,(k),ih}E_{t,(k),jh}\Big)^2. \tag{4.57}$$

Given $j\in[d_k]$, first consider first $i\neq j$. By Assumption (E1) and (E2), we can write

$$\begin{aligned}
E_{t,(k),jh}&=\sum_{m=1}^{r_{e,k}}\sum_{n=1}^{r_{e,\text{-}k}}A_{e,k,jm}A_{e,\text{-}k,hn}\Big(\sum_{w\ge 0}a_{e,w}X_{e,t-w,(k),mn}\Big)+\Sigma_{\epsilon,(k),jh}\Big(\sum_{g\ge 0}a_{\epsilon,g}X_{\epsilon,t-g,(k),jh}\Big),\\
E_{t,(k),ih}&=\sum_{\tau=1}^{r_{e,k}}\sum_{\gamma=1}^{r_{e,\text{-}k}}A_{e,k,i\tau}A_{e,\text{-}k,h\gamma}\Big(\sum_{l\ge 0}a_{e,l}X_{e,t-l,(k),\tau\gamma}\Big)+\Sigma_{\epsilon,(k),ih}\Big(\sum_{\xi\ge 0}a_{\epsilon,\xi}X_{\epsilon,t-\xi,(k),ih}\Big).
\end{aligned}$$

Then among all terms in the expansion of $\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}E_{t,(k),ih}E_{t,(k),jh}$, consider

$$\sum_{g\ge 0}\sum_{\xi\ge 0}a_{\epsilon,g}a_{\epsilon,\xi}\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}\Sigma_{\epsilon,(k),ih}X_{\epsilon,t-g,(k),jh}X_{\epsilon,t-\xi,(k),ih}.$$

Fix $g\ge 0$ and $\xi\ge 0$, then it is direct from the sub-Gaussian tail that

$$\Sigma_{\epsilon,(k),jh}\Sigma_{\epsilon,(k),ih}X_{\epsilon,t-g,(k),jh}X_{\epsilon,t-\xi,(k),ih}$$

is sub-exponential with parameter of constant order and independent over $h\in[d_{\text{-}k}]$ and $t\in[T]$. This implies $\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}\Sigma_{\epsilon,(k),ih}X_{\epsilon,t-g,(k),jh}X_{\epsilon,t-\xi,(k),ih}$ is sub-exponential with parameter of order $(Td_{\text{-}k})^{1/2}$, which hence also holds true for

$$\sum_{g\ge 0}\sum_{\xi\ge 0}a_{\epsilon,g}a_{\epsilon,\xi}\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}\Sigma_{\epsilon,(k),ih}X_{\epsilon,t-g,(k),jh}X_{\epsilon,t-\xi,(k),ih}$$

by Assumption (E2). Thus,

$$\max_{j\in[d_k]}\sum_{i\neq j}^{d_k}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}\Big(\sum_{g\geq 0}a_{\epsilon,g}X_{\epsilon,t-g,(k),jh}\Big)\Sigma_{\epsilon,(k),ih}\Big(\sum_{\xi\geq 0}a_{\epsilon,\xi}X_{\epsilon,t-\xi,(k),ih}\Big)\Big\}^2 = O_P\big\{Td\log^2(d_k)\big\}. \tag{4.58}$$

Using the same argument above, with the independence between $\{\mathcal{X}_{e,t}\}$ and $\{\mathcal{X}_{\epsilon,t}\}$ from (E2), we have

$$\max_{j\in[d_k]}\sum_{i=1}^{d_k}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\sum_{m=1}^{r_{e,k}}\sum_{n=1}^{r_{e,\text{-}k}}A_{e,k,jm}A_{e,\text{-}k,hn} \cdot\Big(\sum_{w\geq 0}a_{e,w}X_{e,t-w,(k),mn}\Big)\Sigma_{\epsilon,(k),ih}\Big(\sum_{\xi\geq 0}a_{\epsilon,\xi}X_{\epsilon,t-\xi,(k),ih}\Big)\Big\}^2 = O_P\big(Td\log^2(d_k)\big),$$
$$\max_{j\in[d_k]}\sum_{i=1}^{d_k}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\Sigma_{\epsilon,(k),jh}\Big(\sum_{g\geq 0}a_{\epsilon,g}X_{\epsilon,t-g,(k),jh}\Big) \cdot\sum_{\tau=1}^{r_{e,k}}\sum_{\gamma=1}^{r_{e,\text{-}k}}A_{e,k,i\tau}A_{e,\text{-}k,h\gamma}\Big(\sum_{l\geq 0}a_{e,l}X_{e,t-l,(k),\tau\gamma}\Big)\Big\}^2 = O_P\big\{Td\log^2(d_k)\big\}. \tag{4.59}$$

In the expansion of $\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}E_{t,(k),ih}E_{t,(k),jh}$ for any $i\in[d_k]$, consider now

$$\sum_{w\geq 0}\sum_{l\geq 0}a_{e,w}a_{e,l}\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\sum_{m,\tau=1}^{r_{e,k}}\sum_{n,\gamma=1}^{r_{e,\text{-}k}}A_{e,k,i\tau}A_{e,\text{-}k,h\gamma}A_{e,k,jm}A_{e,\text{-}k,hn}X_{e,t-w,(k),mn}X_{e,t-l,(k),\tau\gamma},$$

which is sub-exponential with mean of order $T$ and parameter of order $\|\mathbf{A}_{e,k}\|_\infty\cdot(T)^{1/2}$ by Assumption (E1), (E2) and the sub-Gaussian tail. Hence by the sparsity of $\mathbf{A}_{e,k}$ according to (E1) again,

$$\max_{j\in[d_k]}\sum_{i=1}^{d_k}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\sum_{m=1}^{r_{e,k}}\sum_{n=1}^{r_{e,\text{-}k}}A_{e,k,jm}A_{e,\text{-}k,hn}\Big(\sum_{w\geq 0}a_{e,w}X_{e,t-w,(k),mn}\Big) \sum_{\tau=1}^{r_{e,k}}\sum_{\gamma=1}^{r_{e,\text{-}k}}A_{e,k,i\tau}A_{e,\text{-}k,h\gamma}\Big(\sum_{l\geq 0}a_{e,l}X_{e,t-l,(k),\tau\gamma}\Big)\Big\}^2 = O_P\big\{T^2\log^2(d_k)\big\}. \tag{4.60}$$

To bound (4.57), it remains to consider

$$\sum_{g\geq 0}\sum_{\xi\geq 0}a_{\epsilon,g}a_{\epsilon,\xi}\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}}\Sigma^2_{\epsilon,(k),jh}X^2_{\epsilon,t-g,(k),jh},$$

which is sub-exponential with parameter of order $(Td_{\text{-}k})^{1/2}$, similar to the case as in (4.58), except that the mean is of order $Td_{\text{-}k}$. Therefore,

$$\max_{j\in[d_k]} \mathbb{1}\{i=j\}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{d_{\text{-}k}} \Sigma_{\epsilon,(k),jh}\Big(\sum_{g\geq 0} a_{\epsilon,g}X_{\epsilon,t-g,(k),jh}\Big)\Sigma_{\epsilon,(k),ih}\Big(\sum_{\xi\geq 0} a_{\epsilon,\xi}X_{\epsilon,t-\xi,(k),ih}\Big)\Big\}^2$$
$$= O_P\big\{T^2 d_{\text{-}k}^2 \log^2(d_k)\big\}. \tag{4.61}$$

Finally for (4.57), combining (4.58), (4.59), (4.60) and (4.61), we have

$$\max_{j\in[d_k]} \Big\|\sum_{t=1}^{T}\big(\mathbf{E}_{t,(k)}\mathbf{E}_{t,(k)}'\big)_{\cdot j}\Big\|_F^2 = O_P\big\{(Td + T^2 d_{\text{-}k}^2)\log^2(d_k)\big\}.$$

This concludes the proof of the lemma. □

**Lemma 4.10** *Let all assumptions in Lemma 4.9 hold. Then we have*

$$\max_{j\in[d_k]} \big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k\mathbf{Q}_{k,j\cdot}\big\|^2 = O_P\Big\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\Big(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\Big)\frac{d^2}{g_s^2}\log^2(d_k)\Big\}, \tag{4.62}$$

$$\max_{t\in[T]} \big\|\textsc{vec}(\widehat{\mathcal{F}}_t) - (\mathbf{H}_\otimes')^{-1}\textsc{vec}(\mathcal{F}_{Z,t})\big\|^2$$
$$= O_P\Big(\Big[\max_{k\in[K]}\Big\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\Big(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\Big)\frac{d^2}{g_s}\Big\} + \frac{d}{g_w}\Big]\log^2(T)\Big). \tag{4.63}$$

*Thus, we have*

$$\max_{i\in[d],t\in[T]} \Big\{\big(\textsc{vec}(\widehat{\mathcal{C}}_t)\big)_i - \big(\textsc{vec}(\mathcal{C}_t)\big)_i\Big\}^2$$
$$= O_P\Big(\Big[\max_{k\in[K]}\Big\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\Big(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\Big)\frac{d^2}{g_s g_w}\Big\} + \frac{d}{g_w^2}\Big]\log^2(T)\prod_{k=1}^{K}\log^2(d_k)\Big).$$

*Similarly,*

$$\max_{i\in[d],t\in[T]} \Big\{\big(\textsc{vec}(\widetilde{\mathcal{C}}_{\text{reshape},t})\big)_i - \big(\textsc{vec}(\mathcal{C}_{\text{reshape},t})\big)_i\Big\}^2$$
$$= O_P\Big(\Big[\max_{j\in[v-1]}\Big\{d_j^{2(\delta_{j,1}-\delta_{j,r_j})}\Big(\frac{1}{Td_{\text{-}k}d_j^{1-\delta_{j,1}}} + \frac{1}{d_j^{1+\delta_{j,r_j}}}\Big)\frac{d^2}{\gamma_s\gamma_w}\Big\}$$
$$+ d_V^{2(\delta_{V,1}-\delta_{V,r_V})}\Big(\frac{1}{Tdd_V^{-\delta_{V,1}}} + \frac{1}{d_V^{1+\delta_{V,r_V}}}\Big)\frac{d^2}{\gamma_s\gamma_w} + \frac{d}{\gamma_w^2}\Big]\log^2(T)\log^2(d_V)\prod_{k=1}^{v-1}\log^2(d_k)\Big).$$

***Proof of Lemma 4.10.*** To see (4.62), from the proof of (4.39) in Lemma 4.5, we have

$$\max_{j\in[d_k]} \big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k\mathbf{Q}_{k,j\cdot}\big\|^2 = \max_{j\in[d_k]} \Big\|T^{-1}\widehat{\mathbf{D}}_k^{-1}\widehat{\mathbf{Q}}_k'\Big(\sum_{t=1}^{T}\mathbf{R}_{k,t}\Big)_{\cdot j}\Big\|^2$$
$$\leq T^{-2}\cdot\|\widehat{\mathbf{D}}_k^{-1}\|_F^2\cdot\|\widehat{\mathbf{Q}}_k\|_F^2\cdot\max_{j\in[d_k]}\Big\|\Big(\sum_{t=1}^{T}\mathbf{R}_{k,t}\Big)_{\cdot j}\Big\|_F^2$$
$$= O_P\big\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})} g_s^{-2}(T^{-1}d + d_{\text{-}k}^2)\log^2(d_k)\big\},$$

where the last equality used Lemma 4.4 and Lemma 4.9. With the rate further refined as in and (4.40) and repeat the procedures in the proof of Lemma 4.9, we conclude $\max_{j\in[d_k]}\big\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k\mathbf{Q}_{k,j\cdot}\big\|^2$ is inflated by $\log^2(d_k)$ compared to the rate in Lemma 4.5.

For (4.63), by inspecting (4.45), it suffices to characterize the change from the rate of $\|\mathcal{F}_t\|_F^2$ to the rate of $\max_{t\in[T]}\|\mathcal{F}_t\|_F^2$, while all other rates follow the similar arguments in the proof of Lemma 4.9 by using sub-exponential distributions. With the sub-Gaussian tail, $\|\mathcal{F}_t\|_F^2$ is sub-exponential with both mean and parameter of constant order, so that $\max_{t\in[T]}\|\mathcal{F}_t\|_F^2 = O_P\{\log^2(T)\}$. By checking all the rates in the expansion (4.45) are inflated by $\log^2(T)$, (4.63) is hence concluded.

Finally, with all previous results and recall the expansion of $\textsc{vec}(\widehat{\mathcal{C}}_t) - \textsc{vec}(\mathcal{C}_t)$ in the proof of Lemma 4.6, the rate of $\max_{i\in[d],t\in[T]}\{(\textsc{vec}(\widehat{\mathcal{C}}_t))_i - (\textsc{vec}(\mathcal{C}_t))_i\}^2$ is inflated by

$$\log^2(T)\prod_{k=1}^{K}\log^2(d_k)$$

compared to the individual rate of $\{(\textsc{vec}(\widehat{\mathcal{C}}_t))_i - (\textsc{vec}(\mathcal{C}_t))_i\}^2$. The result for $\widetilde{\mathcal{C}}_{\text{reshape},t}$ holds similarly by all previous arguments and the proof of Lemma 4.7, except that the dimension is different. This ends the proof of the lemma. □

***Proof of Theorem 4.1.*** Let $\mathcal{A} = \{a_1,\ldots,a_\ell\}$ be given. We first show that the Tucker-decomposition tensor factor model (4.4) implies (4.3) with $\mathbf{A}_{\text{reshape},K-\ell+1} \in \mathcal{K}_{d_{a_1}\times\cdots\times d_{a_\ell}}$. Suppose $\mathcal{Y}_t = \mathcal{F}_t \times_{k=1}^{K}\mathbf{A}_k + \mathcal{E}_t$.

Consider first $\ell = 1$. With any $\mathcal{A} = \{a_1\}$ (hence the corresponding $\mathcal{A}^* = [K]\setminus\mathcal{A}$ as defined in Theorem 4.1), it is direct that

$$\textsc{reshape}\big(\mathcal{Y}_t,\mathcal{A}\big) = \textsc{reshape}\big(\mathcal{F}_t\times_{k=1}^{K}\mathbf{A}_k + \mathcal{E}_t,\mathcal{A}\big)$$
$$= \textsc{reshape}\big(\mathcal{F}_t\times_{k=1}^{K}\mathbf{A}_k,\{a_1\}\big) + \textsc{reshape}\big(\mathcal{E}_t,\mathcal{A}\big)$$
$$= \textsc{fold}_K\big(\text{mat}_{a_1}(\mathcal{F}_t\times_{k=1}^{K}\mathbf{A}_k),\{d_1,\ldots,d_{a_1-1},d_{a_1+1},\ldots,d_K,d_{a_1}\}\big) + \textsc{reshape}\big(\mathcal{E}_t,\mathcal{A}\big)$$
$$= \textsc{fold}_K\big(\mathbf{A}_{a_1}\text{mat}_{a_1}(\mathcal{F}_t)\mathbf{A}_{\text{-}a_1}',\{d_1,\ldots,d_{a_1-1},d_{a_1+1},\ldots,d_K,d_{a_1}\}\big) + \textsc{reshape}\big(\mathcal{E}_t,\mathcal{A}\big),$$
$$= \textsc{reshape}\big(\mathcal{F}_t,\mathcal{A}\big)\times_{i=1}^{K-1}\mathbf{A}_{\mathcal{A}_i^*}\times_K\mathbf{A}_{a_1} + \textsc{reshape}\big(\mathcal{E}_t,\mathcal{A}\big),$$

where the last equality used the fact that

$$\begin{aligned}
&\mathrm{mat}_K\big\{\textsc{reshape}\big(\mathcal{F}_t,\mathcal{A}\big)\times_{i=1}^{K-1}\mathbf{A}_{\mathcal{A}_i^*}\times_K\mathbf{A}_{a_1}\big\}=\mathbf{A}_{a_1}\mathrm{mat}_K\big\{\textsc{reshape}\big(\mathcal{F}_t,\{a_1\}\big)\big\}\mathbf{A}'_{\text{-}a_1}\\
&=\mathbf{A}_{a_1}\mathrm{mat}_K\big\{\textsc{fold}_K\big(\mathrm{mat}_{a_1}(\mathcal{F}_t),\{d_1,\ldots,d_{a_1-1},d_{a_1+1},\ldots,d_K,d_{a_1}\}\big)\big\}\mathbf{A}'_{\text{-}a_1}\\
&=\mathbf{A}_{a_1}\mathrm{mat}_{a_1}(\mathcal{F}_t)\mathbf{A}'_{\text{-}a_1}.
\end{aligned} \tag{4.64}$$

Hence, $\mathcal{Y}_t$ follows (4.3) with variables defined according to Theorem 4.1, and the model has a Kronecker structure product since $\mathbf{A}_{a_1}\in\mathcal{K}_{d_{a_1}}$.

Consider $\ell=2$ (implying at least $K=2$). In the following, we use the neater notation that mode-$k$ unfolding of some tensor $\mathcal{X}$ is $\mathbf{X}_{(k)}$. Without loss of generality, let $\mathcal{A}=\{a,b\}$ (with $a<b$) and the corresponding $\mathcal{A}^*=[K]\setminus\mathcal{A}$. Take the mode-$b$ unfolding on each $\mathcal{Y}_t$, we obtain

$$\mathbf{Y}_{t,(b)}=\mathbf{A}_b\mathbf{F}_{t,(b)}\big(\mathbf{A}_K\otimes\cdots\otimes\mathbf{A}_{b+1}\otimes\mathbf{A}_{b-1}\otimes\cdots\otimes\mathbf{A}_a\otimes\cdots\otimes\mathbf{A}_1\big)'+\mathbf{E}_{t,(b)},$$

then for each row (as a column vector) of $\mathbf{Y}_{t,(b)}$, we fold it back to an order-$(K-1)$ tensor along the remaining dimensions, i.e., for any $i$-th row $\mathbf{Y}_{t,(b),i\cdot}$ with $i\in[d_b]$,

$$\begin{aligned}
&\mathcal{Y}_{t,(b),i}:=\textsc{fold}\big(\mathbf{Y}_{t,(b),i\cdot},\{d_1,\ldots,d_{b-1},d_{b+1},\ldots,d_K\}\big)\\
&=\textsc{fold}\big\{\big(\mathbf{A}_K\otimes\cdots\otimes\mathbf{A}_{b+1}\otimes\mathbf{A}_{b-1}\otimes\cdots\otimes\mathbf{A}_a\otimes\cdots\otimes\mathbf{A}_1\big)\\
&\quad\cdot\mathbf{F}'_{t,(b)}\mathbf{A}_{b,i\cdot},\{d_1,\ldots,d_{b-1},d_{b+1},\ldots,d_K\}\big\}+\textsc{fold}\big(\mathbf{E}_{t,(b),i\cdot},\{d_1,\ldots,d_{b-1},d_{b+1},\ldots,d_K\}\big)\\
&=\textsc{fold}\big(\mathbf{F}'_{t,(b)}\mathbf{A}_{b,i\cdot},\{r_1,\ldots,r_{b-1},r_{b+1},\ldots,r_K\}\big)\times_{k=1}^{b-1}\mathbf{A}_k\times_{h=b}^{K-1}\mathbf{A}_{h+1}\\
&\quad+\textsc{fold}\big(\mathbf{E}_{t,(b),i\cdot},\{d_1,\ldots,d_{b-1},d_{b+1},\ldots,d_K\}\big).
\end{aligned}$$

Define $\mathbf{A}_{\text{-}b,\text{-}a}:=\mathbf{A}_K\otimes\cdots\otimes\mathbf{A}_{b+1}\otimes\mathbf{A}_{b-1}\otimes\cdots\otimes\mathbf{A}_{a+1}\otimes\mathbf{A}_{a-1}\otimes\cdots\otimes\mathbf{A}_1$, where by convention $\mathbf{A}_{\text{-}b,\text{-}a}=1$ if $K=2$. Take the mode-$a$ unfolding on $\mathcal{Y}_{t,(b),i}$, we have

$$\begin{aligned}
&\big(\mathcal{Y}_{t,(b),i}\big)_{(a)}\\
&=\mathbf{A}_a\big\{\textsc{fold}\big(\mathbf{F}'_{t,(b)}\mathbf{A}_{b,i\cdot},\{r_1,\ldots,r_{a-1},r_{a+1},\ldots,r_K\}\big)\big\}_{(a)}\mathbf{A}'_{\text{-}b,\text{-}a}\\
&\quad+\big\{\textsc{fold}\big(\mathbf{E}_{t,(b),i\cdot},\{d_1,\ldots,d_{b-1},d_{b+1},\ldots,d_K\}\big)\big\}_{(a)}\\
&=\sum_{j=1}^{r_b}\mathbf{A}_a\big\{\textsc{fold}\big(A_{b,ij}\mathbf{F}_{t,(b),j\cdot},\{r_1,\ldots,r_{b-1},r_{b+1},\ldots,r_K\}\big)\big\}_{(a)}\mathbf{A}'_{\text{-}b,\text{-}a}\\
&\quad+\big\{\textsc{fold}\big(\mathbf{E}_{t,(b),i\cdot},\{d_1,\ldots,d_{b-1},d_{b+1},\ldots,d_K\}\big)\big\}_{(a)}\\
&=\big(\mathbf{A}'_{b,i\cdot}\otimes\mathbf{I}_{d_a}\big)\big(\mathbf{I}_{r_b}\otimes\mathbf{A}_a\big)\begin{pmatrix}\big\{\textsc{fold}\big(\mathbf{F}_{t,(b),1\cdot},\{r_1,\ldots,r_{b-1},r_{b+1},\ldots,r_K\}\big)\big\}_{(a)}\mathbf{A}'_{\text{-}b,\text{-}a}\\ \cdots\\ \big\{\textsc{fold}\big(\mathbf{F}_{t,(b),r_b\cdot},\{r_1,\ldots,r_{b-1},r_{b+1},\ldots,r_K\}\big)\big\}_{(a)}\mathbf{A}'_{\text{-}b,\text{-}a}\end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&+ \left\{\text{FOLD}\big(\mathbf{E}_{t,(b),i\cdot}, \{d_1, \ldots, d_{b-1}, d_{b+1}, \ldots, d_K\}\big)\right\}_{(a)} \\
&= \left(\mathbf{A}'_{b,i\cdot} \otimes \mathbf{A}_a\right) \begin{pmatrix} \left\{\text{FOLD}\big(\mathbf{F}_{t,(b),1\cdot}, \{r_1, \ldots, r_{b-1}, r_{b+1}, \ldots, r_K\}\big)\right\}_{(a)} \\ \cdots \\ \left\{\text{FOLD}\big(\mathbf{F}_{t,(b),r_b\cdot}, \{r_1, \ldots, r_{b-1}, r_{b+1}, \ldots, r_K\}\big)\right\}_{(a)} \end{pmatrix} \mathbf{A}'_{\text{-}b,\text{-}a} \\
&+ \left\{\text{FOLD}\big(\mathbf{E}_{t,(b),i\cdot}, \{d_1, \ldots, d_{b-1}, d_{b+1}, \ldots, d_K\}\big)\right\}_{(a)}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathcal{Y}_{t,a\sim b} &= \begin{pmatrix} \big(\mathcal{Y}_{t,(b),1}\big)_{(a)} \\ \cdots \\ \big(\mathcal{Y}_{t,(b),d_b}\big)_{(a)} \end{pmatrix} \\
&= \left(\mathbf{A}_b \otimes \mathbf{A}_a\right) \begin{pmatrix} \left\{\text{FOLD}\big(\mathbf{F}_{t,(b),1\cdot}, \{r_1, \ldots, r_{b-1}, r_{b+1}, \ldots, r_K\}\big)\right\}_{(a)} \\ \cdots \\ \left\{\text{FOLD}\big(\mathbf{F}_{t,(b),r_b\cdot}, \{r_1, \ldots, r_{b-1}, r_{b+1}, \ldots, r_K\}\big)\right\}_{(a)} \end{pmatrix} \mathbf{A}'_{\text{-}b,\text{-}a} \\
&\quad + \begin{pmatrix} \left\{\text{FOLD}\big(\mathbf{E}_{t,(b),1\cdot}, \{d_1, \ldots, d_{b-1}, d_{b+1}, \ldots, d_K\}\big)\right\}_{(a)} \\ \cdots \\ \left\{\text{FOLD}\big(\mathbf{E}_{t,(b),d_b\cdot}, \{d_1, \ldots, d_{b-1}, d_{b+1}, \ldots, d_K\}\big)\right\}_{(a)} \end{pmatrix},
\end{aligned}
$$

so that by definition of the reshape operator with $\ell = 2$,

$$
\begin{aligned}
&\text{RESHAPE}(\mathcal{Y}_t, \{a, b\}) = \text{FOLD}_{K-1}(\mathcal{Y}_{t,a\sim b}, \{d_1, \ldots, d_{a-1}, d_{a+1}, \ldots, d_{b-1}, d_{b+1}, \ldots, d_K, d_a d_b\}) \\
&= \text{RESHAPE}(\mathcal{F}_t, \{a, b\}) \times_{i=1}^{K-2} \mathbf{A}_{\mathcal{A}_i^*} \times_{K-1} (\mathbf{A}_b \otimes \mathbf{A}_a) + \text{RESHAPE}(\mathcal{E}_t, \{a, b\}),
\end{aligned}
$$

where the last line used similar arguments (omitted) as (4.64). This implies $\mathcal{Y}_t$ follows (4.3) with a Kronecker structure product as $(\mathbf{A}_b \otimes \mathbf{A}_a) \in \mathcal{K}_{d_a \times d_b}$.

Finally, consider any $\ell \geq 3$. We use an induction argument. With the definition of tensor reshape in Section 4.2 and the above for $\ell = 2$, the initial case $\ell = 3$ can be shown by

$$
\begin{aligned}
&\text{RESHAPE}(\mathcal{Y}_t, \{a_1, a_2, a_3\}) = \text{RESHAPE}\big\{\text{RESHAPE}(\mathcal{Y}_t, \{a_2, a_3\}), \{a_1, K-1\}\big\} \\
&= \text{RESHAPE}\Big\{\text{RESHAPE}(\mathcal{F}_t, \{a_2, a_3\}) \times_{i=1}^{K-2} \mathbf{A}_{[[K]\backslash\{a_2,a_3\}]_i} \times_{K-1} (\mathbf{A}_{a_3} \otimes \mathbf{A}_{a_2}) \\
&\quad + \text{RESHAPE}(\mathcal{E}_t, \{a_2, a_3\}), \{a_1, K-1\}\Big\} \\
&= \text{RESHAPE}\big\{\text{RESHAPE}(\mathcal{F}_t, \{a_2, a_3\}), \{a_1, K-1\}\big\} \times_{i=1}^{K-3} \mathbf{A}_{[([K]\backslash\{a_2,a_3\})\backslash\{a_1\}]_i} \\
&\quad \times_{K-2} (\mathbf{A}_{a_3} \otimes \mathbf{A}_{a_2} \otimes \mathbf{A}_{a_1}) + \text{RESHAPE}\big\{\text{RESHAPE}(\mathcal{E}_t, \{a_2, a_3\}), \{a_1, K-1\}\big\} \\
&= \text{RESHAPE}(\mathcal{F}_t, \{a_1, a_2, a_3\}) \times_{i=1}^{K-3} \mathbf{A}_{[[K]\backslash\{a_1,a_2,a_3\}]_i} \times_{K-2} (\mathbf{A}_{a_3} \otimes \mathbf{A}_{a_2} \otimes \mathbf{A}_{a_1}) \\
&\quad + \text{RESHAPE}(\mathcal{E}_t, \{a_1, a_2, a_3\}),
\end{aligned}
\tag{4.65}
$$

where the last equality used again the definition of tensor shape and that $(\mathbf{A}_{a_3} \otimes \mathbf{A}_{a_2} \otimes \mathbf{A}_{a_1}) \in \mathcal{K}_{d_{a_1} \times d_{a_2} \times d_{a_3}}$. Now if for all $\ell \in [L]$ with $L \geq 3$, (4.4) implies (4.3) with $\mathbf{A}_{\text{reshape},K-\ell+1} \in \mathcal{K}_{d_{a_1} \times \cdots \times d_{a_\ell}}$, which is then also true for $\ell = L+1$ in a similar argument (omitted) as (4.65). This completes the induction.

Given any $\mathcal{A}$, note also that if Assumption (F1) holds with $\mathcal{X}_{\text{reshape},t}$ and $\mathcal{F}_{\text{reshape},t}$ replaced by $\mathcal{X}_t$ and $\mathcal{F}_t$ respectively (with $\mathcal{X}_t$ and $\mathcal{F}_t$ from (4.19)), then it is immediate from the linearity of the reshape operator that

$$\mathcal{F}_{\text{reshape},t} = \text{RESHAPE}(\mathcal{F}_t, \mathcal{A}) = \text{RESHAPE}\Big(\sum_{w \geq 0} a_{f,w} \mathcal{X}_{f,t-w}, \mathcal{A}\Big) = \sum_{w \geq 0} a_{f,w} \text{RESHAPE}(\mathcal{X}_{f,t-w}, \mathcal{A}),$$

which implies $\mathcal{F}_{\text{reshape},t}$ follows Assumption (F1) by $\mathcal{X}_{\text{reshape},t} = \text{RESHAPE}(\mathcal{X}_{f,t}, \mathcal{A})$.

We have now proved that (4.4) uniquely implies (4.3) with $\mathbf{A}_{\text{reshape},K-\ell+1} \in \mathcal{K}_{d_{a_1} \times \cdots \times d_{a_\ell}}$, and a version of Assumption (F1) on $\{\mathcal{F}_t\}$ (from (4.19)) implies Assumption (F1) on $\mathcal{F}_{\text{reshape},t}$. It remains to show the other way around (for some $\mathcal{A}$), but all the previous steps are reversible (note that in particular, the reshape operator is reversible as long as the dimension of the original tensor is known) and $\mathbf{A}_{\text{reshape},K-\ell+1} \in \mathcal{K}_{d_{a_1} \times \cdots \times d_{a_\ell}}$ ensures the existence of an appropriate set of low-rank matrices. Therefore, the proof for the theorem is completed. □

***Proof of Theorem 4.2.*** Under $H_0$, consider (4.7). Since there exists $m \in [|\mathcal{R}|]$ such that

$$(\pi_{m,1}, \ldots, \pi_{m,K-v+1}) = (r_v, \ldots, r_K),$$

we only consider such $m$ and simplify $\widehat{\mathcal{C}}_{m,t}$ as $\widehat{\mathcal{C}}_t$ and $\widehat{\mathcal{E}}_{m,t}$ as $\widehat{\mathcal{E}}_t$ (see the explanation at the beginning of Section 4.7). By (4.12) and Assumption (E1),

$$\begin{aligned}\widehat{\mathbf{E}}_{t,(k^*)} &= \text{mat}_{k^*}(\widehat{\mathcal{E}}_t) = \text{mat}_{k^*}((\mathcal{C}_t - \widehat{\mathcal{C}}_t) + \mathcal{F}_{e,t} \times_1 \mathbf{A}_{e,1} \times_2 \cdots \times_K \mathbf{A}_{e,K} + \mathbf{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t) \\ &= (\mathcal{C}_t - \widehat{\mathcal{C}}_t)_{(k^*)} + \mathbf{A}_{e,k^*} \mathbf{F}_{e,t,(k^*)} \mathbf{A}'_{e,\text{-}k^*} + \mathbf{\Sigma}_{\epsilon,(k^*)} * \boldsymbol{\epsilon}_{t,(k^*)},\end{aligned}$$

where $\mathbf{A}_{e,\text{-}k^*} := \mathbf{A}_{e,K} \otimes \cdots \otimes \mathbf{A}_{e,k^*+1} \otimes \mathbf{A}_{e,k^*-1} \otimes \cdots \otimes \mathbf{A}_{e,1}$. Hence, for any $t \in [T]$ and

$j \in [d/d_k^*]$, we have

$$\frac{1}{d_{k^*}}\sum_{i=1}^{d_{k^*}}\big(\widehat{E}^2_{t,(k^*),ij} - \Sigma^2_{\epsilon,(k^*),ij}\big) = \frac{1}{d_{k^*}}\big(\widehat{\mathbf{E}}'_{t,(k^*)}\widehat{\mathbf{E}}_{t,(k^*)}\big)_{jj} - \frac{1}{d_{k^*}}\sum_{i=1}^{d_{k^*}}\Sigma^2_{\epsilon,(k^*),ij} = \sum_{h=1}^{6}\mathcal{I}_{e,h},$$

$$\begin{aligned}
\text{where } \mathcal{I}_{e,1} &:= \frac{1}{d_{k^*}}\big\{\big(\mathbf{\Sigma}_{\epsilon,(k^*)} * \boldsymbol{\epsilon}_{t,(k^*)}\big)'\big(\mathbf{\Sigma}_{\epsilon,(k^*)} * \boldsymbol{\epsilon}_{t,(k^*)}\big)\big\}_{jj} - \frac{1}{d_{k^*}}\sum_{i=1}^{d_{k^*}}\Sigma^2_{\epsilon,(k^*),ij},\\
\mathcal{I}_{e,2} &:= \frac{1}{d_{k^*}}\big\{(\mathcal{C}_t - \widehat{\mathcal{C}}_t)'_{(k^*)}(\mathcal{C}_t - \widehat{\mathcal{C}}_t)_{(k^*)}\big\}_{jj},\\
\mathcal{I}_{e,3} &:= \frac{1}{d_{k^*}}\big\{\mathbf{A}_{e,\text{-}k^*}\mathbf{F}'_{e,t,(k^*)}\mathbf{A}'_{e,k^*}\mathbf{A}_{e,k^*}\mathbf{F}_{e,t,(k^*)}\mathbf{A}'_{e,\text{-}k^*}\big\}_{jj},\\
\mathcal{I}_{e,4} &:= O_P\Big(d_{k^*}^{-1}\big\{(\mathcal{C}_t - \widehat{\mathcal{C}}_t)'_{(k^*)}\mathbf{A}_{e,k^*}\mathbf{F}_{e,t,(k^*)}\mathbf{A}'_{e,\text{-}k^*}\big\}_{jj}\Big),\\
\mathcal{I}_{e,5} &:= O_P\Big(d_{k^*}^{-1}\big\{(\mathcal{C}_t - \widehat{\mathcal{C}}_t)'_{(k^*)}\big(\mathbf{\Sigma}_{\epsilon,(k^*)} * \boldsymbol{\epsilon}_{t,(k^*)}\big)\big\}_{jj}\Big),\\
\mathcal{I}_{e,6} &:= O_P\Big(d_{k^*}^{-1}\big\{\mathbf{A}_{e,\text{-}k^*}\mathbf{F}'_{e,t,(k^*)}\mathbf{A}'_{e,k^*}\big(\mathbf{\Sigma}_{\epsilon,(k^*)} * \boldsymbol{\epsilon}_{t,(k^*)}\big)\big\}_{jj}\Big).
\end{aligned} \tag{4.66}$$

Consider $\mathcal{I}_{e,2}$. From Lemma 4.6, recall that

$$\begin{aligned}
&\big\{\big(\mathbf{vec}(\widehat{\mathcal{C}}_t)\big)_i - \big(\mathbf{vec}(\mathcal{C}_t)\big)_i\big\}^2\\
&= O_P\left(\max_{k\in[K]}\left\{d_k^{2(\delta_{k,1}-\delta_{k,r_k})}\left(\frac{1}{Td_{\text{-}k}d_k^{1-\delta_{k,1}}} + \frac{1}{d_k^{1+\delta_{k,r_k}}}\right)\frac{d^2}{g_s g_w}\right\} + \frac{d}{g_w^2}\right),
\end{aligned}$$

which is the squared error for each entry of $\widehat{\mathcal{C}}_t$. With Assumption (R2), the above squared error rate is $o\big(1/\min_{k\in[K]}\{d_k\}\big) = o\big(d_{k^*}^{-1}\big)$. Hence $\mathcal{I}_{e,2} = o_P\big(d_{k^*}^{-1}\big)$.

With Assumption (E1), $\|\mathbf{A}_{e,k^*}\|_F = O(1)$, $\|\mathbf{A}_{e,\text{-}k^*}\|_F = O(1)$ and $r_{e,k}$ for $k \in [K]$ are finite, so that $\mathcal{I}_{e,3}$, $\mathcal{I}_{e,6} = O_P\big(d_{k^*}^{-1}\big)$. By the Cauchy–Schwarz inequality, immediately $\mathcal{I}_{e,4} = O_P\big(\mathcal{I}_{e,2}^{1/2}\cdot\mathcal{I}_{e,3}^{1/2}\big) = o_P\big(d_{k^*}^{-1}\big)$.

Consider $\mathcal{I}_{e,1}$, noting that

$$\mathcal{I}_{e,1} = \frac{1}{d_{k^*}}\sum_{i=1}^{d_{k^*}}\Sigma^2_{\epsilon,(k^*),ij}\big(\epsilon^2_{t,(k^*),ij} - 1\big),$$

so that with Theorem 1 in Ayvazyan and Ulyanov (2023), we have

$$Z_{j,t} := \frac{d_{k^*}^{-1}\sum_{i=1}^{d_{k^*}}\Sigma^2_{\epsilon,(k^*),ij}\big(\epsilon^2_{t,(k^*),ij} - 1\big)}{\sqrt{d_{k^*}^{-2}\sum_{i=1}^{d_{k^*}}\text{Var}\big(\epsilon^2_{t,(k^*),ij}\big)\Sigma^4_{\epsilon,(k^*),ij}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

implying $\mathcal{I}_{e,1}$ is of the rate $d_{k^*}^{-1/2}$ exactly. Note that $Z_{j,t}$'s are independent of each other by

Assumption (E1). It also follows that $\mathcal{I}_{e,5} = O_P\big(\mathcal{I}_{e,2}^{1/2}\mathcal{I}_{e,1}^{1/2}\big) = o_P\big(d_{k^*}^{-3/4}\big)$. Finally, with (4.66),

$$\frac{\sum_{i=1}^{d_{k^*}}\big(\widehat{E}_{t,(k^*),ij}^2 - \Sigma_{\epsilon,(k^*),ij}^2\big)}{\sqrt{\sum_{i=1}^{d_{k^*}}\text{Var}(\epsilon_{t,(k^*),ij}^2)\Sigma_{\epsilon,(k^*),ij}^4}} = \frac{d_{k^*}^{-1}\sum_{i=1}^{d_{k^*}}\big(\widehat{E}_{t,(k^*),ij}^2 - \Sigma_{\epsilon,(k^*),ij}^2\big)}{\sqrt{d_{k^*}^{-2}\sum_{i=1}^{d_{k^*}}\text{Var}(\epsilon_{t,(k^*),ij}^2)\Sigma_{\epsilon,(k^*),ij}^4}}$$
$$= Z_{j,t}(1+o_P(1)) \xrightarrow{p} Z_{j,t} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

This shows the asymptotic result for $\widehat{\mathcal{E}}_t$ in Theorem 4.2 (i.e., the first asymptotic result). The result for $\widetilde{\mathcal{E}}_t$ can be shown in the same manner, except that for $\mathcal{I}_{e,2}$, we use Assumption (R2) on the squared error for each entry of $\widetilde{\mathcal{C}}_{\text{reshape},t}$ which is the following from Lemma 4.7,

$$\Big\{\big(\textbf{vec}\big(\widetilde{\mathcal{C}}_{\text{reshape},t}\big)\big)_i - \big(\textbf{vec}\big(\mathcal{C}_{\text{reshape},t}\big)\big)_i\Big\}^2$$
$$= O_P\Bigg(\max_{j\in[v-1]}\Bigg\{d_j^{2(\delta_{j,1}-\delta_{j,r_j})}\Bigg(\frac{1}{Td_{\text{-}k}d_j^{1-\delta_{j,1}}} + \frac{1}{d_j^{1+\delta_{j,r_j}}}\Bigg)\frac{d^2}{\gamma_s\gamma_w}\Bigg\}$$
$$+ d_V^{2(\delta_{V,1}-\delta_{V,r_V})}\Bigg(\frac{1}{Tdd_V^{-\delta_{V,1}}} + \frac{1}{d_V^{1+\delta_{V,r_V}}}\Bigg)\frac{d^2}{\gamma_s\gamma_w} + \frac{d}{\gamma_w^2}\Bigg).$$

Under $H_0$, we arrive at the same conclusion with the same $Z_{j,t}$'s. Moreover, the asymptotic result for $\widehat{\mathcal{E}}_t$ holds true under $H_1$. This concludes the proof of the theorem. □

***Proof of Theorem 4.3.*** We work with the exact $m \in [|\mathcal{R}|]$ satisfying the statement of Theorem 4.2. Firstly, using triangle inequality, for each $t \in [T]$, $j \in [d/d_k^*]$, consider

$$\max_{t\in[T]}|x_{j,t} - y_{m,j,t}| \le \max_{t\in[T]}\Bigg|\frac{1}{d_{k^*}}\sum_{i=1}^{d_{k^*}}\big(\widehat{E}_{t,(k^*),ij}^2 - \widetilde{E}_{t,(k^*),ij}^2\big)\Bigg| \le \sum_{\ell=2}^{6}\max_{t\in[T]}(|\mathcal{I}_{e,\ell}| + |\widetilde{\mathcal{I}}_{e,\ell}|),$$

where $\mathcal{I}_{e,\ell}$, $\ell = 2,\ldots,6$ is defined as in (4.66), and $\widetilde{\mathcal{I}}_{e,\ell}$ for $\ell = 2,\ldots,6$ is defined exactly the same as $\mathcal{I}_{e,\ell}$, except that $\widehat{\mathcal{C}}_t$ is replaced by $\widetilde{\mathcal{C}}_{\text{reshape},t}$.

From Lemma 4.10, the uniform error rates for $\widehat{\mathcal{C}}_t$ and $\widetilde{\mathcal{C}}_{\text{reshape},t}$ is

$$\max_{i\in[d],t\in[T]}\Big\{\big|\big(\textbf{vec}\big(\widehat{\mathcal{C}}_t\big)\big)_i - \big(\textbf{vec}\big(\mathcal{C}_t\big)\big)_i\big|,\ \big|\big(\textbf{vec}\big(\widetilde{\mathcal{C}}_{\text{reshape},t}\big)\big)_i - \big(\textbf{vec}\big(\mathcal{C}_{\text{reshape},t}\big)\big)_i\big|\Big\}$$
$$= O_P\Bigg\{\Bigg[\max_{k\in[K]}\Bigg\{d_k^{\delta_{k,1}-\delta_{k,r_k}}\Bigg(\frac{1}{(Td_{\text{-}k}d_k^{1-\delta_{k,1}})^{1/2}} + \frac{1}{d_k^{(1+\delta_{k,r_k})/2}}\Bigg)\frac{d}{(g_sg_w)^{1/2}}\Bigg\} + \frac{d^{1/2}}{g_w}\Bigg]$$
$$\cdot\log(T)\prod_{k=1}^{K}\log(d_k) + \Bigg[\max_{j\in[v-1]}\Bigg\{d_j^{\delta_{j,1}-\delta_{j,r_j}}\Bigg(\frac{1}{(Td_{\text{-}k}d_j^{1-\delta_{j,1}})^{1/2}} + \frac{1}{d_j^{(1+\delta_{j,r_j})/2}}\Bigg)\frac{d}{(\gamma_s\gamma_w)^{1/2}}\Bigg\}$$
$$+ d_V^{\delta_{V,1}-\delta_{V,r_V}}\Bigg(\frac{1}{(Tdd_V^{-\delta_{V,1}})^{1/2}} + \frac{1}{d_V^{(1+\delta_{V,r_V})/2}}\Bigg)\frac{d}{(\gamma_s\gamma_w)^{1/2}} + \frac{d^{1/2}}{\gamma_w}\Bigg]\log(T)\log(d_V)\prod_{k=1}^{v-1}\log(d_k)\Bigg\}, \tag{4.67}$$

which is denoted by $O_P\{\varrho(T, d_1, \ldots, d_K)\}$ for simplicity. From the proof of Theorem 4.2, we can see that $\mathcal{I}_{e,5}$ and also $\widetilde{\mathcal{I}}_{e,5}$ have the slowest rate of convergence for a fixed indices $t \in [T]$ and $i \in [d]$. Taking maximum over all possible indices, using the sub-Gaussian tail and (4.67), we thus have

$$
\begin{aligned}
&\max_{t\in[T]} |x_{j,t} - y_{m,j,t}| = O_P(\mathcal{I}_{e,5} + \widetilde{\mathcal{I}}_{e,5}) \\
&= \max_{i\in[d],t\in[T]} \Big\{ \big|\big(\mathbf{vec}(\widehat{\mathcal{C}}_t)\big)_i - \big(\mathbf{vec}(\mathcal{C}_t)\big)_i\big|,\ \big|\big(\mathbf{vec}(\widetilde{\mathcal{C}}_{\text{reshape},t})\big)_i - \big(\mathbf{vec}(\mathcal{C}_{\text{reshape},t})\big)_i\big| \Big\} \\
&\quad \cdot O_P\big(\max_{i\in[d],t\in[T]} |(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} * \boldsymbol{\epsilon}_t)_i|\big) \\
&= O_P\Big\{\varrho(T, d_1, \ldots, d_K)\log(T)\prod_{k=1}^K \log(d_k)\Big\}.
\end{aligned}
\tag{4.68}
$$

Next, we assess the approximate "gap" size of the $x_{j,t}$'s over $t \in [T]$. To this end, using Theorem 4.2, and the fact that $\sum_{i=1}^{d_{k^*}} \text{var}(\epsilon^2_{t,(k^*),ij})\Sigma^4_{\epsilon,(k^*),ij}$ has order $d_{k^*}$, we have

$$
x_{j,t} \asymp_P d_{k^*}^{-1}\sum_{i=1}^{d_{k^*}} \Sigma^2_{\epsilon,(k^*),ij} + \frac{1}{d_{k^*}^{1/2}} Z_{j,t} \asymp_P d_{k^*}^{-1}\sum_{i=1}^{d_{k^*}} \Sigma^2_{\epsilon,(k^*),ij} + d_{k^*}^{-1/2},
\tag{4.69}
$$

showing that the "gap" between two ordered $x_{j,t}$'s is $O_P(T^{-1}d_{k^*}^{-1/2})$.

With the "gap" size and uniform error in (4.68), consider

$$
\begin{aligned}
&\sup_{c\in\mathbb{R}} |\mathbb{F}_{y,m,j}(c) - \mathbb{F}_{x,j}(c)| = \sup_{c\in\mathbb{R}} \Big|\frac{1}{T}\sum_{t=1}^T \big[\mathbb{1}\{y_{m,j,t} \le c\} - \mathbb{1}\{x_{j,t} \le c\}\big]\Big| \\
&\le \sup_{c\in\mathbb{R}} \Big|\frac{1}{T}\sum_{t=1}^T \big[\mathbb{1}\{x_{j,t} \pm \max_{t\in[T]} |x_{j,t} - y_{m,j,t}| \le c\} - \mathbb{1}\{x_{\alpha,t} \le c\}\big]\Big| \\
&= O_P\Big\{\frac{1}{T}\max_{t\in[T]} |x_{\alpha,t} - y_{\alpha,t}| \Big/ \big(T^{-1}d_{k^*}^{-1/2}\big)\Big\} = O_P\Big\{\varrho(T, d_1, \ldots, d_K) d_{k^*}^{1/2}\log(T)\prod_{k=1}^K \log(d_k)\Big\},
\end{aligned}
$$

where the last line used (4.68). Hence in particular,

$$
\begin{aligned}
&\mathbb{P}_{y,m,j}\{y_{m,j,t} > \widehat{q}_{x,j}(\alpha)\} = 1 - \mathbb{P}_{y,m,j}\{y_{m,j,t} \le \widehat{q}_{x,j}(\alpha)\} = 1 - \mathbb{F}_{y,m,j}(\widehat{q}_{x,j}(\alpha)) \\
&\le 1 - \mathbb{F}_{x,j}(\widehat{q}_{x,j}(\alpha)) + \sup_{c\in\mathbb{R}} |\mathbb{F}_{y,m,j}(c) - \mathbb{F}_{x,j}(c)| \\
&\le \alpha + O_P\Big\{\varrho(T, d_1, \ldots, d_K) d_{k^*}^{1/2}\log(T)\prod_{k=1}^K \log(d_k)\Big\},
\end{aligned}
$$

which is the desired result we want, and this completes the proof of the theorem. □

***Proof of Theorem 4.4.*** Let $\mathcal{Y}_t \in \mathbb{R}^{d_1\times\cdots\times d_K}$ be an order-$K$ tensor and $\mathcal{A} = \{a_1, \ldots, a_\ell\}$.

Then each $\textsc{reshape}(\mathcal{Y}_t, \mathcal{A})$ is an order-$(K-\ell+1)$ tensor. If $\{\textsc{reshape}(\mathcal{Y}_t, \mathcal{A})\}$ has a Kronecker product structure, then Theorem 4.1 allows us to write for $t \in [T]$,

$$\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}) = \mathcal{F}_{\text{reshape},t} \times_{j=1}^{K-\ell} \mathbf{A}_{\text{reshape},j} \times_{K-\ell+1} \mathbf{A}_{\text{reshape},K-\ell+1} + \mathcal{E}_{\text{reshape},t}, \tag{4.70}$$

for some core factor $\{\mathcal{F}_{\text{reshape},t}\}$, loading matrices $\{\mathbf{A}_{\text{reshape},j}\}_{j\in[K-\ell+1]}$, and noise $\mathcal{E}_{\text{reshape},t}$. Immediately by Definition 4.2, if $\mathbf{A}_{\text{reshape},K-\ell+1} \in \mathcal{K}_{d_{a_1}\times\cdots\times d_{a_\ell}}$, then $\mathcal{Y}_t$ has a Kronecker product structure; otherwise, $\{\mathcal{Y}_t\}$ has no Kronecker product structure along $\mathcal{A}$.

It remains to show that if $\{\mathcal{Y}_t\}$ either has a Kronecker product structure or has no Kronecker product structure along some set $\mathcal{A}^*$, then $\{\textsc{reshape}(\mathcal{Y}_t, \mathcal{A})\}$ with $\mathcal{A}^* \subseteq \mathcal{A}$ has a Kronecker product structure. The case where $\{\mathcal{Y}_t\}$ has a Kronecker product structure is trivial. We then only need to consider that $\{\mathcal{Y}_t\}$ has no Kronecker product structure along $\mathcal{A}^*$, which implicitly assume a factor model of $\{\mathcal{Y}_t\}$ along $\mathcal{A}^*$ by Definition 4.2. Without loss of generality, let $\mathcal{A}^* := \{K-g+1, \ldots, K\}$, otherwise redefine the mode indices of $\mathcal{Y}_t$. For the set $\mathcal{A}$ with $\mathcal{A}^* \subseteq \mathcal{A}$, we now read $\mathcal{A} = \{a_1, \ldots, a_{\ell-g}, K-g+1, \ldots, K\}$. Using the last property of tensor reshape in Section 4.2 (which can be easily seen by induction), we have

$$\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}) = \textsc{reshape}\big\{\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}^*), \{a_1, \ldots, a_{\ell-g}, K-g+1\}\big\}. \tag{4.71}$$

Now that $\{\mathcal{Y}_t\}$ has no Kronecker product structure along $\mathcal{A}^*$, similar to the form (4.70), we have for $t \in [T]$,

$$\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}^*) = \mathcal{F}^*_{\text{reshape},t} \times_{j=1}^{K-g} \mathbf{A}^*_{\text{reshape},j} \times_{K-g+1} \mathbf{A}^*_{\text{reshape},K-g+1} + \mathcal{E}^*_{\text{reshape},t},$$

which implies the time series $\{\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}^*)\}$ follows a Tucker-decomposition TFM. According to Theorem 4.1, $\{\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}^*)\}$ follows a factor model along any index set of $\{\textsc{reshape}(\mathcal{Y}_t, \mathcal{A}^*)\}$ (with Kronecker product structure; but the factor model form is sufficient for our claim). In particular, with the index set $\{a_1, \ldots, a_{\ell-g}, K-g+1\}$, we conclude from (4.71) that $\{\textsc{reshape}(\mathcal{Y}_t, \mathcal{A})\}$ has a Kronecker product structure. This also completes the proof of the theorem. □

# Chapter 5

# Matrix-Valued Factor Model with Time-Varying Main Effects

## 5.1 Introduction

Matrix-valued time series factor models, a generalisation of vector time series factor models, have been utilised a lot for dimension reduction and prediction in recent years in fields such as finance, economics, medical science and meteorology. Important earlier theoretical and methodological developments include Wang et al. (2019), Chen et al. (2020), Chen and Fan (2023) and He et al. (2024b), which are all on matrix-valued factor models using the Tucker decomposition for the common component. Chang et al. (2023) uses the CP decomposition for the common component, while Guan (2024) considers Tucker decomposition of the common component but taking in covariates in the loadings. See a detailed survey of matrix factor models in Section 2.3.4. With Tucker decomposition, a matrix-valued time series factor model (FM) can be written as

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{R}\mathbf{F}_t\mathbf{C}' + \mathbf{E}_t, \tag{5.1}$$

where $\mathbf{Y}_t \in \mathbb{R}^{p\times q}$ is the observed matrix at time $t$, $\boldsymbol{\mu} \in \mathbb{R}^{p\times q}$ is the mean matrix, $\mathbf{R} \in \mathbb{R}^{p\times k_r}$ and $\mathbf{C} \in \mathbb{R}^{q\times k_c}$ are the row and column factor loading matrices respectively, $\mathbf{F}_t \in \mathbb{R}^{k_r\times k_c}$ is the core factor matrix at time $t$, and finally $\mathbf{E}_t \in \mathbb{R}^{p\times q}$ is the noise matrix at time $t$. If we set $\mathbf{R} := (\boldsymbol{\alpha}_{p\times r}, \widetilde{\mathbf{R}}_{p\times(k_r-r-\ell)}, \mathbf{1}_{p\times\ell})$, $\mathbf{C} := (\mathbf{1}_{q\times r}, \widetilde{\mathbf{C}}_{q\times(k_c-r-\ell)}, \boldsymbol{\gamma}_{q\times\ell})$ and $\mathbf{F}_t :=$ diag$((\mathbf{g}_t)_{r\times r}, (\widetilde{\mathbf{F}}_t)_{(k_r-r-\ell)\times(k_c-r-\ell)}, (\mathbf{h}_t)_{\ell\times\ell})$ (He et al., 2023a), where $\mathbf{1}_{m\times n}$ is a matrix of ones of size $m \times n$, then (5.1) becomes

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_{r\times q} + \mathbf{1}_{p\times\ell}\mathbf{h}_t\boldsymbol{\gamma}' + \widetilde{\mathbf{R}}\widetilde{\mathbf{F}}_t\widetilde{\mathbf{C}}' + \mathbf{E}_t. \tag{5.2}$$

If the rows of $\mathbf{Y}_t$ represent different countries and the columns represent different economic indicators, then since the $j$-th row of $\boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_{r\times q}$ is $\boldsymbol{\alpha}_{j\cdot}'\mathbf{g}_t\mathbf{1}_{r\times q}$, where $\boldsymbol{\alpha}_{j\cdot}$ is the $j$-th row of $\boldsymbol{\alpha}$ as a column vector, it means that each element in the $j$-th row is the same, with value $\boldsymbol{\alpha}_{j\cdot}'\mathbf{g}_t\mathbf{1}_r$. Hence we can argue that $\mathbf{g}_t$ represents common global factors affecting all countries, although each country loads differently on $\mathbf{g}_t$. Similarly, $\mathbf{h}_t$ represents latent economic states across different economic indicators. The term $\widetilde{\mathbf{R}}\widetilde{\mathbf{F}}_t\widetilde{\mathbf{C}}'$ can be viewed as an interaction term, while $\boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_{r\times q}$ the countries' main effects, and $\mathbf{1}_{p\times\ell}\mathbf{h}_t\boldsymbol{\gamma}'$ the economic states' main effects.

Three problems arise upon inspecting (5.1) and (5.2) however. Firstly, for (5.1) to transform to (5.2), $\mathbf{R}$ and $\mathbf{C}$ are both of reduced rank. In the literature for model (5.1), we always need $\mathbf{R}$ and $\mathbf{C}$ to be of full rank at least asymptotically (see for example, Assumption (B2) in He et al. (2023a) or Equation (8) in Chen and Fan (2023)) for estimation purpose.

Secondly, model (5.2) is not general enough, unless $r$ and $\ell$ can be large. For example, if $r$ is small, each country is driven only by few global common factors affecting all countries, on top of the factors in $\widetilde{\mathbf{F}}_t$. This will not be a problem, if not for the fact that there can be latent common factors only driving a small group of countries/economic indicators. For instance, there can be a few small European countries which do not share global common factors with the majority of European countries, but with other middle-Eastern countries. Such "grouping" of countries usually comes with their corresponding groups of unique factors. These factors become "weak" country effects, shared only among "small" number of countries, essentially inflating $r$ while inducing a sparse $\boldsymbol{\alpha}$. Constraint factor modelling in Chen et al. (2020) certainly helps, but we do not always know the exact group of countries which share latent common factors.

The final problem is related to the second one. The inability of (5.2) to accommodate "weak" country/economic states effects originates from the fact that $\widetilde{\mathbf{R}}\widetilde{\mathbf{F}}_t\widetilde{\mathbf{C}}'$ in (5.1) contains only pervasive factors, which is essentially assumed across all past works in factor models for matrix-valued time series. In a general order tensor setting, Cen and Lam (2025b) and Chen and Lam (2024b) have both allowed weak factors in the common component of the factor model.

One way to generalise (5.2) to address all aforementioned problems is to note that

$$\boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_{r\times q} = (\boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_r)\mathbf{1}_q' =: \boldsymbol{\alpha}_t\mathbf{1}_q', \quad \mathbf{1}_{p\times\ell}\mathbf{h}_t\boldsymbol{\gamma}' = \mathbf{1}_p(\boldsymbol{\gamma}\mathbf{h}_t'\mathbf{1}_\ell)' =: \mathbf{1}_p\boldsymbol{\beta}_t',$$

where $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are the time-varying row and column main effects respectively. If we manage to estimate the two vectors $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ without any low-rank constraints as in the equation above, then the second problem is naturally solved. Formally allowing for weak factors in the loading matrices, like those in Lam and Yao (2012) for a vector factor model, solves the third problem. Finally, with these problems solved, we can go back to assuming full rank row and column factor loading matrices (see Assumption (L1) in Section 5.3.1) to solve the first problem.

In this chapter, we contribute to the literature in several important ways. Firstly, we generalise model (5.2) to (5.3) which is the time-varying main effects factor model (MEFM), incorporating all relaxations described in the previous paragraph. Secondly, we provide estimation and inference methods and the corresponding theoretical guarantees, on top of a separate ratio-based method for identifying the core rank of $\mathbf{F}_t$, with consistency proved. Third and perhaps the most important of all, we provide a statistical test on the null of FM in (5.1), with $\mathbf{R}$ and $\mathbf{C}$ both of full rank, is sufficient against the more general MEFM in (5.3). A rejected null hypothesis then implies there are row and/or column main effects that is not of a low rank structure like those in (5.2), essentially pointing to the existence of "weak" main effects.

The rest of this chapter is organised as follows. Section 5.2 introduces MEFM formally, laying down important identification conditions and estimation methodologies for all the components in the model. Section 5.3 presents the assumptions for MEFM and the consistency and asymptotic normality results for its estimators. In particular, the test for FM versus MEFM is detailed in Section 5.3.6, while the core rank estimator for $\mathbf{F}_t$ is presented in Section 5.3.7. Finally, Section 5.4 presents our extensive simulation results and details the NYC Taxi traffic data analysis, pinpointing the presence of weak hourly main effects in the data. Our method is available in the R package `MEFM`, with instruction in its reference manual on R CRAN. All proofs of the theorems are relegated to Section 5.5.

## 5.2 Model and Estimation

### 5.2.1 Main effect matrix factor model

We propose the time-varying **M**ain **E**ffect matrix **F**actor **M**odel (MEFM) such that for $t \in [T]$,

$$\mathbf{Y}_t = \mu_t \mathbf{1}_p \mathbf{1}_q' + \boldsymbol{\alpha}_t \mathbf{1}_q' + \mathbf{1}_p \boldsymbol{\beta}_t' + \mathbf{C}_t + \mathbf{E}_t, \tag{5.3}$$

where $\mathbf{Y}_t$ is a $p \times q$ observed matrix at time $t$, $\mu_t$ is a scalar representing the grand mean of $\mathbf{Y}_t$, $\boldsymbol{\alpha}_t \in \mathbb{R}^p$ and $\boldsymbol{\beta}_t \in \mathbb{R}^q$ are the row and column main effects at time $t$, respectively. The common component $\mathbf{C}_t := \mathbf{A}_r \mathbf{F}_t \mathbf{A}_c'$ is latent, where $\mathbf{F}_t \in \mathbb{R}^{k_r \times k_c}$ is the core factor series with unknown number of factors $k_r$ and $k_c$, and $\mathbf{A}_r$ and $\mathbf{A}_c$ are the row and column factor loading matrices, with size $p \times k_r$ and $q \times k_c$, respectively. Lastly, $\mathbf{E}_t$ is the idiosyncratic noise series with the same dimension as $\mathbf{Y}_t$.

Unlike FM in (5.2), the main effects $\alpha_t$ and $\boldsymbol{\beta}_t$ in MEFM are not restricted to be of low rank, which significantly improves the flexibility of FM, and allows for a test of FM in (5.1) in the end. In fact, setting concatenated matrices $\ddot{\mathbf{A}}_r = (\mathbf{1}_p, \mathbf{I}_p, \mathbf{A}_r, \mathbf{1}_p)$ and $\ddot{\mathbf{A}}_c = (\mathbf{1}_q, \mathbf{1}_q, \mathbf{A}_c, \mathbf{I}_q)$, block diagonal matrix $\ddot{\mathbf{F}}_t = \text{diag}\{\mu_t, \boldsymbol{\alpha}_t, \mathbf{F}_t, \boldsymbol{\beta}_t'\}$, then we can read (5.3) as

$$\mathbf{Y}_t = \ddot{\mathbf{A}}_r \ddot{\mathbf{F}}_t \ddot{\mathbf{A}}_c' + \mathbf{E}_t.$$

However, we observe that the dimension of the factor series is now $(2+p+k_r) \times (2+q+k_c)$, and hence there is not much dimension reduction for $\mathbf{Y}_t$, and both $\ddot{\mathbf{A}}_r$ and $\ddot{\mathbf{A}}_c$ have no full column ranks. This observation suggests again that MEFM is more general than FM, and numerical results in Section 5.4 actually show that even an approximate estimation by FM in general comes at a cost of using very large number of factors.

Given the above motivation of MEFM, we point out that the form of MEFM can be obtained by FM in general, see Remark 5.2 for details. For generality purpose, $\mathbf{Y}_t$ can have nonzero mean but we can always demean the data as the sample mean is not our main parameter of interest. The right hand side of (5.3) is entirely latent and hence we propose Assumption (IC1) below to identify the grand mean and the row and column effects.

(IC1) (Identification). *For any $t \in [T]$, we assume that*

$$\mathbf{1}_p'\boldsymbol{\alpha}_t = \mathbf{1}_q'\boldsymbol{\beta}_t = 0, \quad \mathbf{1}_p'\mathbf{A}_r = \mathbf{0}, \quad \mathbf{1}_q'\mathbf{A}_c = \mathbf{0}.$$

However, we require further identification between the factors and the factor loading matrices. To do this, we normalise the loading matrices to $\mathbf{Q}_r = \mathbf{A}_r\mathbf{Z}_r^{-1/2}$ and $\mathbf{Q}_c = \mathbf{A}_c\mathbf{Z}_c^{-1/2}$, where $\mathbf{Z}_r = \text{diag}(\mathbf{A}_r'\mathbf{A}_r)$ and $\mathbf{Z}_c = \text{diag}(\mathbf{A}_c'\mathbf{A}_c)$, measuring the sparsity of each column of loading matrices and hence the factor strength. For example, $\mathbf{F}_t$ pervasive in the $j$-th row will have the $j$-th column of $\mathbf{A}_r$ dense and hence the $j$-th diagonal entry of $\mathbf{Z}_r$ will be of order $p$. For technical details, see Assumption (L1). We leave the identification to Section 5.3.1. Assumption (IC1) also facilitates the estimation of $\mu_t$, $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$, and we discuss in the next subsection how to estimate the grand mean, the row and column effects, and the row and column factor loading matrices in (5.3).

**Remark 5.1** *One advantage of MEFM lies in Assumption (IC1) that potentially allows the grand mean and main effects to be nonstationary. However, we should also warn that since the model identification relies heavily on the identification assumption contemporaneously involving all entries in the row effect or column effect, any form of nonstationarity in the main effects might be unnatural and restricted. For example, $\{(\boldsymbol{\alpha}_t)_1\}_{t\in[T]}$ can potentially have seasonality or even unit roots, but requires that each $\boldsymbol{\alpha}_t$ satisfies Assumption (IC1). One way to circumvent this arguably unrealistic (IC1) is to use alternative identifications. For instance, in one of our ongoing projects, we are exploring the direction of requiring $\min_{i\in[p]}(\boldsymbol{\alpha}_t)_i = \min_{j\in[q]}(\boldsymbol{\beta}_t)_j = 0$ which is a valid identification condition and meanwhile provides a straightforward setup of sparse main effects. More general nonstationarity structures merit further investigation, but we do not expect the common component series to be nonstationary.*

### 5.2.2 Estimation of the main effects and factor components

The factor structure is hidden in $\mathbf{Y}_t$ and we need to estimate the time-varying grand mean and main effects first. For the grand mean, right-multiplying by $\mathbf{1}_q$ and left-multiplying by $\mathbf{1}_p'$ on both sides of (5.3) results in $\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q = pq\mu_t + \mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q$ by Assumption (IC1). Hence for each $t \in [T]$, we obtain the moment estimator for the time-varying grand mean as

$$\widehat{\mu}_t := \mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q/pq.$$

Also, right-multiplying by $\mathbf{1}_q$ and left-multiplying by $\mathbf{1}_p'$ lead respectively to $\mathbf{Y}_t\mathbf{1}_q = q\mu_t\mathbf{1}_p + q\boldsymbol{\alpha}_t + \mathbf{E}_t\mathbf{1}_q$ and $\mathbf{1}_p'\mathbf{Y}_t = p\mu_t\mathbf{1}_q' + p\boldsymbol{\beta}_t' + \mathbf{1}_p'\mathbf{E}_t$. Therefore, we obtain the time-varying row and column effect estimators as

$$\widehat{\boldsymbol{\alpha}}_t := q^{-1}\mathbf{Y}_t\mathbf{1}_q - \widehat{\mu}_t\mathbf{1}_p, \quad \widehat{\boldsymbol{\beta}}_t' := p^{-1}\mathbf{1}_p'\mathbf{Y}_t - \widehat{\mu}_t\mathbf{1}_q'.$$

Finally, we introduce the following to estimate the factor structure,

$$\begin{aligned}\widehat{\mathbf{L}}_t &:= \mathbf{Y}_t - \widehat{\mu}_t\mathbf{1}_p\mathbf{1}_q' - \widehat{\boldsymbol{\alpha}}_t\mathbf{1}_q' - \mathbf{1}_p\widehat{\boldsymbol{\beta}}_t' \\ &= \mathbf{Y}_t + (pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' - q^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t = \mathbf{M}_p\mathbf{Y}_t\mathbf{M}_q, \end{aligned} \tag{5.4}$$

where $\mathbf{M}_m := \mathbf{I}_m - m^{-1}\mathbf{1}_m\mathbf{1}_m'$ for any positive integer $m$. From the above, $\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$ admits $\mathbf{1}_p$ in its null space, and $\widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t$ admits $\mathbf{1}_q$ instead. The factor structure can hence be estimated, with $\widehat{\mathbf{Q}}_r$ constructed as the eigenvectors of $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$ corresponding to the first $k_r$ largest eigenvalues, and $\widehat{\mathbf{Q}}_c$ the eigenvectors of $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t$ corresponding to the first $k_c$ largest eigenvalues.

We can then estimate the factor time series $\mathbf{F}_{Z,t} = \mathbf{Z}_r^{1/2}\mathbf{F}_t\mathbf{Z}_c^{1/2}$, and the common component $\mathbf{C}_t$, respectively as

$$\widehat{\mathbf{F}}_{Z,t} := \widehat{\mathbf{Q}}_r'\widehat{\mathbf{L}}_t\widehat{\mathbf{Q}}_c = \widehat{\mathbf{Q}}_r'\mathbf{Y}_t\widehat{\mathbf{Q}}_c, \quad \widehat{\mathbf{C}}_t := \widehat{\mathbf{Q}}_r\widehat{\mathbf{F}}_{Z,t}\widehat{\mathbf{Q}}_c' = \widehat{\mathbf{Q}}_r\widehat{\mathbf{Q}}_r'\mathbf{Y}_t\widehat{\mathbf{Q}}_c\widehat{\mathbf{Q}}_c'. \tag{5.5}$$

Finally, the residuals $\mathbf{E}_t$ is estimated by

$$\widehat{\mathbf{E}}_t := \widehat{\mathbf{L}}_t - \widehat{\mathbf{C}}_t. \tag{5.6}$$

**Remark 5.2** *Suppose we have a traditional matrix-valued factor model such that* $\acute{\mathbf{Y}}_t = \acute{\mathbf{C}}_t + \acute{\mathbf{E}}_t$ *where* $\acute{\mathbf{Y}}_t$, $\acute{\mathbf{C}}_t$, *and* $\acute{\mathbf{E}}_t$ *are* $p \times q$ *matrices representing the observation, common component, and noise, respectively. Suppose also* $\acute{\mathbf{C}}_t = \mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'$. *Then we can construct*

$$\begin{aligned}&\acute{\mu}_t := (pq)^{-1}\mathbf{1}_p'\acute{\mathbf{C}}_t\mathbf{1}_q, \quad \acute{\boldsymbol{\alpha}}_t := q^{-1}\acute{\mathbf{C}}_t\mathbf{1}_q - \acute{\mu}_t\mathbf{1}_p = q^{-1}\mathbf{M}_p\acute{\mathbf{C}}_t\mathbf{1}_q, \\ &\acute{\boldsymbol{\beta}}_t := p^{-1}\acute{\mathbf{C}}_t'\mathbf{1}_p - \acute{\mu}_t\mathbf{1}_q = p^{-1}\mathbf{M}_q\acute{\mathbf{C}}_t'\mathbf{1}_p.\end{aligned}$$

*Hence we can express FM in the following MEFM form satisfying (IC1):*

$$\acute{\mathbf{Y}}_t = \acute{\mu}_t \mathbf{1}_p \mathbf{1}_q' + \acute{\boldsymbol{\alpha}}_t \mathbf{1}_q' + \mathbf{1}_p \acute{\boldsymbol{\beta}}_t' + (\acute{\mathbf{C}}_t - \acute{\mu}_t \mathbf{1}_p \mathbf{1}_q' - \acute{\boldsymbol{\alpha}}_t \mathbf{1}_q' - \mathbf{1}_p \acute{\boldsymbol{\beta}}_t') + \acute{\mathbf{E}}_t,$$

*where*

$$\acute{\mathbf{C}}_t - \acute{\mu}_t \mathbf{1}_p \mathbf{1}_q' - \acute{\boldsymbol{\alpha}}_t \mathbf{1}_q' - \mathbf{1}_p \acute{\boldsymbol{\beta}}_t' = (\mathbf{M}_p \mathbf{A}_r) \mathbf{F}_t (\mathbf{M}_q \mathbf{A}_c)',$$

*is the common component. Since* $\mathbf{M}_m \mathbf{1}_m = \mathbf{0}$*, it is easy to see that*

$$\mathbf{1}_p'(\mathbf{M}_p \mathbf{A}_r) = \mathbf{0}, \quad \mathbf{1}_q'(\mathbf{M}_q \mathbf{A}_c) = \mathbf{0}.$$

*It is also easy to verify that* $\mathbf{1}_p' \acute{\boldsymbol{\alpha}}_t = \mathbf{1}_q' \acute{\boldsymbol{\beta}}_t = 0$*. Hence a traditional matrix-valued factor model can be expressed as MEFM in* (5.3) *to satisfy (IC1).*

## 5.3 Assumptions and Theoretical Results

### 5.3.1 Assumptions

A set of assumptions on the factor structure is imposed below, and in particular, we allow factors to have different strengths, as in Lam and Yao (2012) and Chen and Lam (2024b).

(M1) (Alpha mixing). *The vector processes* $\{\mathbf{vec}(\mathbf{F}_t)\}$ *and* $\{\mathbf{vec}(\mathbf{E}_t)\}$ *are* $\alpha$*-mixing, respectively. A vector process* $\{\mathbf{x}_t : t = 0, \pm 1, \pm 2, \dots\}$ *is* $\alpha$*-mixing if, for some* $\gamma > 2$*, the mixing coefficients satisfy the condition that*

$$\sum_{h=1}^{\infty} \alpha(h)^{1-2/\gamma} < \infty,$$

*where* $\alpha(h) = \sup_\tau \sup_{A \in \mathcal{H}_{-\infty}^{\tau}, B \in \mathcal{H}_{\tau+h}^{\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ *and* $\mathcal{H}_\tau^s$ *is the* $\sigma$*-field generated by* $\{\mathbf{x}_t : \tau \le t \le s\}$*.*

(F1) (Time series in $\mathbf{F}_t$). *There is* $\mathbf{X}_{f,t}$ *with the same dimension as* $\mathbf{F}_t$*, such that* $\mathbf{F}_t = \sum_{w \ge 0} a_{f,w} \mathbf{X}_{f,t-w}$*. The time series* $\{\mathbf{X}_{f,t}\}$ *has i.i.d. elements with mean* 0 *and variance* 1*, with uniformly bounded fourth order moments. The coefficients* $a_{f,w}$ *are such that* $\sum_{w \ge 0} a_{f,w}^2 = 1$ *and* $\sum_{w \ge 0} |a_{f,w}| \le c$ *for some constant* $c$*.*

(L1) (Factor strength). *We assume that* $\mathbf{A}_r$ *and* $\mathbf{A}_c$ *are of full rank and independent of factors and errors series. Furthermore, as* $p, q \to \infty$*,*

$$\mathbf{Z}_r^{-1/2} \mathbf{A}_r' \mathbf{A}_r \mathbf{Z}_r^{-1/2} \to \boldsymbol{\Sigma}_{A,r}, \quad \mathbf{Z}_c^{-1/2} \mathbf{A}_c' \mathbf{A}_c \mathbf{Z}_c^{-1/2} \to \boldsymbol{\Sigma}_{A,c},$$

*where* $\mathbf{Z}_r = \text{diag}(\mathbf{A}_r'\mathbf{A}_r)$, $\mathbf{Z}_c = \text{diag}(\mathbf{A}_c'\mathbf{A}_c)$, *and both* $\mathbf{\Sigma}_{A,r}$ *and* $\mathbf{\Sigma}_{A,c}$ *are positive definite with all eigenvalues bounded away from 0 and infinity. We assume* $(\mathbf{Z}_r)_{jj} \asymp p^{\delta_{r,j}}$ *for* $j \in [k_r]$ *and* $1/2 < \delta_{r,k_r} \leq \cdots \leq \delta_{r,2} \leq \delta_{r,1} \leq 1$. *Similarly, we assume* $(\mathbf{Z}_c)_{jj} \asymp p^{\delta_{c,j}}$ *for* $j \in [k_c]$, *with* $1/2 < \delta_{c,k_c} \leq \cdots \leq \delta_{c,2} \leq \delta_{c,1} \leq 1$.

With Assumption (L1), we can denote $\mathbf{Q}_r := \mathbf{A}_r\mathbf{Z}_r^{-1/2}$ and $\mathbf{Q}_c := \mathbf{A}_c\mathbf{Z}_c^{-1/2}$. Hence $\mathbf{Q}_r'\mathbf{Q}_r \to \mathbf{\Sigma}_{A,r}$ and $\mathbf{Q}_c'\mathbf{Q}_c \to \mathbf{\Sigma}_{A,c}$.

(E1) (Decomposition of $\mathbf{E}_t$). *We assume that*

$$\mathbf{E}_t = \mathbf{A}_{e,r}\mathbf{F}_{e,t}\mathbf{A}_{e,c}' + \mathbf{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t,$$

*where* $\mathbf{F}_{e,t}$ *is a matrix of size* $k_{e,r} \times k_{e,c}$, *containing independent elements with mean* 0 *and variance* 1. *The matrix* $\boldsymbol{\epsilon}_t \in \mathbb{R}^{p\times q}$ *contains independent elements with mean 0 and variance 1, with* $\{\boldsymbol{\epsilon}_t\}$ *independent of* $\{\mathbf{F}_{e,t}\}$. *The matrix* $\mathbf{\Sigma}_\epsilon$ *contains the standard deviations of the corresponding elements in* $\boldsymbol{\epsilon}_t$, *and has elements uniformly bounded away from 0 and infinity.*

*Moreover,* $\mathbf{A}_{e,r}$ *and* $\mathbf{A}_{e,c}$ *are (approximately) sparse matrices with sizes* $p \times k_{e,r}$ *and* $q \times k_{e,c}$ *respectively, such that* $\|\mathbf{A}_{e,r}\|_1, \|\mathbf{A}_{e,c}\|_1 = O(1)$, *with* $k_{e,r}, k_{e,c} = O(1)$.

(E2) (Time Series in $\mathbf{E}_t$). *There is* $\mathbf{X}_{e,t}$ *the same dimension as* $\mathbf{F}_{e,t}$, *and* $\mathbf{X}_{\epsilon,t}$ *the same dimension as* $\boldsymbol{\epsilon}_t$, *such that* $\mathbf{F}_{e,t} = \sum_{w\geq 0} a_{e,w}\mathbf{X}_{e,t-w}$ *and* $\boldsymbol{\epsilon}_t = \sum_{w\geq 0} a_{\epsilon,w}\mathbf{X}_{\epsilon,t-w}$, *with* $\{\mathbf{X}_{e,t}\}$ *and* $\{\mathbf{X}_{\epsilon,t}\}$ *independent of each other.* $\{\mathbf{X}_{e,t}\}$ *has independent elements while* $\{\mathbf{X}_{\epsilon,t}\}$ *has i.i.d. elements, and all elements have mean zero with unit variance and uniformly bounded fourth order moments. Both* $\{\mathbf{X}_{e,t}\}$ *and* $\{\mathbf{X}_{\epsilon,t}\}$ *are independent of* $\{\mathbf{X}_{f,t}\}$ *from (F1).*

*The coefficients* $a_{e,w}$ *and* $a_{\epsilon,w}$ *are such that* $\sum_{w\geq 0} a_{e,w}^2 = \sum_{w\geq 0} a_{\epsilon,w}^2 = 1$ *and for some constant* $c$ *that* $\sum_{w\geq 0} |a_{e,w}|, \sum_{w\geq 0} |a_{\epsilon,w}| \leq c$.

(R1) (Rate assumptions). *We assume that,*

$$T^{-1}p^{2(1-\delta_{r,k_r})}q^{1-2\delta_{c,1}} = o(1), \quad p^{1-2\delta_{r,k_r}}q^{2(1-\delta_{c,1})} = o(1),$$
$$T^{-1}q^{2(1-\delta_{c,k_c})}p^{1-2\delta_{r,1}} = o(1), \quad q^{1-2\delta_{c,k_c}}p^{2(1-\delta_{r,1})} = o(1).$$

Assumption (F1) introduces serial dependence into the factors, and (E1) and (E2) introduce both cross-sectional and temporal dependence in the noise. The factor structure depicted by (F1), (E1) and (E2) is the same as the one in Cen and Lam (2025b). Note that although Assumption (M1) also features in serial dependence, it is mainly used to construct asymptotic normality of estimators. We refer to Remark 3.1 for further explanations.

By (L1), we have $\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c' = \mathbf{Q}_r\mathbf{Z}_r^{1/2}\mathbf{F}_t\mathbf{Z}_c^{1/2}\mathbf{Q}_c'$, so we aim to estimate $(\mathbf{Q}_r, \mathbf{Q}_c, \mathbf{F}_{Z,t})$ where $\mathbf{F}_{Z,t} := \mathbf{Z}_r^{1/2}\mathbf{F}_t\mathbf{Z}_c^{1/2}$. Unlike the traditional approximate factor model which assumes all factors are pervasive, we allow factors to have different strength similar to Lam and Yao (2012) and Chen and Lam (2024b). To be precise, a column of $\mathbf{A}_r$ (resp. $\mathbf{A}_c$) is dense (i.e., a pervasive factor) if the corresponding $\delta_{r,j} = 1$ (resp. $\delta_{c,j} = 1$), otherwise the column represents a weak factor as it is sparse to certain extent.

Due to the presence of potentially weak factors, we require rate conditions in Assumption (R1) for consistency to hold. If all factors are pervasive, then (R1) holds trivially. We point out that the first and second (or the third and fourth) conditions in (R1) are exactly the same as the first and third conditions of Assumption (R1) in Cen and Lam (2025b) for matrix time series.

### 5.3.2 Identification of the model

With Assumptions (IC1) and (L1), the model (5.3) is identified according to Theorem 5.1 below.

**Theorem 5.1** *(Identification). With Assumption (IC1), each $\mu_t$, $\boldsymbol{\alpha}_t$, and $\boldsymbol{\beta}_t$ can be identified. The common component is hence identified, and if (L1) is also satisfied, the factor structure is identified up to some invertible matrices $\mathbf{M}_r \in \mathbb{R}^{k_r\times k_r}$ and $\mathbf{M}_c \in \mathbb{R}^{k_c\times k_c}$ such that* $(\mathbf{Q}_r, \mathbf{Q}_c, \mathbf{F}_{Z,t}) = (\mathbf{Q}_r\mathbf{M}_r, \mathbf{Q}_c\mathbf{M}_c, \mathbf{M}_r^{-1}\mathbf{F}_{Z,t}\mathbf{M}_c^{-1})$.

### 5.3.3 Rate of convergence for various estimators

To present the consistency of the loading estimators, define

$$\mathbf{H}_r := T^{-1}\widehat{\mathbf{D}}_r^{-1}\widehat{\mathbf{Q}}_r'\mathbf{Q}_r\sum_{t=1}^T(\mathbf{F}_{Z,t}\mathbf{Q}_c'\mathbf{Q}_c\mathbf{F}_{Z,t}'),$$
$$\mathbf{H}_c := T^{-1}\widehat{\mathbf{D}}_c^{-1}\widehat{\mathbf{Q}}_c'\mathbf{Q}_c\sum_{t=1}^T(\mathbf{F}_{Z,t}'\mathbf{Q}_r'\mathbf{Q}_r\mathbf{F}_{Z,t}),$$

where $\widehat{\mathbf{D}}_r := \widehat{\mathbf{Q}}_r'(T^{-1}\sum_{t=1}^T\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t')\widehat{\mathbf{Q}}_r$ is the $k_r\times k_r$ diagonal matrix consisting of eigenvalues of $T^{-1}\sum_{t=1}^T\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$, and similarly $\widehat{\mathbf{D}}_c := \widehat{\mathbf{Q}}_c'(T^{-1}\sum_{t=1}^T\widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t)\widehat{\mathbf{Q}}_c$ is the $k_c\times k_c$ diagonal matrix of eigenvalues of $T^{-1}\sum_{t=1}^T\widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t$.

**Theorem 5.2** *Under Assumptions (IC1), (M1), (F1), (L1), (E1), (E2) and (R1), we have*

$$
\begin{aligned}
(\widehat{\mu}_t - \mu_t)^2 &= O_P(p^{-1}q^{-1}),\\
p^{-1}\|\widehat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t\|^2 &= O_P(q^{-1}),\\
q^{-1}\|\widehat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t\|^2 &= O_P(p^{-1}),\\
p^{-1}\|\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r'\|_F^2 &= O_P\Big\{T^{-1}p^{1-2\delta_{r,k_r}}q^{1-2\delta_{c,1}} + p^{-2\delta_{r,k_r}}q^{2(1-\delta_{c,1})}\Big\},\\
q^{-1}\|\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c'\|_F^2 &= O_P\Big\{T^{-1}q^{1-2\delta_{c,k_c}}p^{1-2\delta_{r,1}} + q^{-2\delta_{c,k_c}}p^{2(1-\delta_{r,1})}\Big\}.
\end{aligned}
$$

From Theorem 5.2, the consistency for the loading matrix estimators requires Assumption (R1). If all factors are pervasive, the (squared) convergence rates for the row (resp. column) loading matrix will be $\max(1/(Tpq), 1/p^2)$ (resp. $\max(1/(Tpq), 1/q^2)$), which are consistent with those in Chen and Fan (2023) after the same normalization of the loading matrices.

**Theorem 5.3** *Under the assumptions in Theorem 5.2, we have the following:*

1. *The error of the estimated factor series has rate*

$$
\begin{aligned}
&\|\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}\|_F^2\\
&= O_P\big(p^{1-\delta_{r,k_r}}q^{1-\delta_{c,k_c}} + T^{-1}p^{1+2\delta_{r,1}-2\delta_{r,k_r}}q^{1-\delta_{c,1}} + p^{1+\delta_{r,1}-3\delta_{r,k_r}}q^{2-\delta_{c,1}}\\
&\qquad + T^{-1}q^{1+2\delta_{c,1}-2\delta_{c,k_c}}p^{1-\delta_{r,1}} + q^{1+\delta_{c,1}-3\delta_{c,k_c}}p^{2-\delta_{r,1}}\big).
\end{aligned}
$$

2. *For any $t \in [T], i \in [p], j \in [q]$, the squared error of the estimated individual common component is*

$$
\begin{aligned}
&(\widehat{C}_{t,ij} - C_{t,ij})^2\\
&= O_P\big(p^{1-2\delta_{r,k_r}}q^{1-2\delta_{c,k_c}} + T^{-1}p^{1+2\delta_{r,1}-3\delta_{r,k_r}}q^{1-\delta_{c,1}-\delta_{c,k_c}} + p^{1+\delta_{r,1}-4\delta_{r,k_r}}q^{2-\delta_{c,1}-\delta_{c,k_c}}\\
&\qquad + T^{-1}q^{1+2\delta_{c,1}-3\delta_{c,k_c}}p^{1-\delta_{r,1}-\delta_{r,k_r}} + q^{1+\delta_{c,1}-4\delta_{c,k_c}}p^{2-\delta_{r,1}-\delta_{r,k_r}}\big).
\end{aligned}
$$

We state the above results separating from Theorem 5.2 since they have used some arguments from the proof of Theorem 5.5. If all factors are pervasive, it is clear that individual common components are consistent with rate $(pq)^{-1/2} + T^{-1/2}(q^{-1/2} + p^{-1/2}) + p^{-1} + q^{-1} = \max(1/(Tq)^{1/2}, 1/(Tp)^{1/2}, 1/p, 1/q)$. This rate coincides with Theorem 4 of Chen and Fan (2023) for instance.

### 5.3.4 Asymptotic normality of estimators

We present the asymptotic normality of various estimators in this subsection, together with the estimation of the corresponding covariance matrices for practical inferences. Before that, we need three more assumptions.

(L2) (Eigenvalues). *The eigenvalues of the* $k_r \times k_r$ *matrix* $\mathbf{\Sigma}_{A,r}\mathbf{Z}_r$ *from Assumption (L1) are distinct, and so are those of the* $k_c \times k_c$ *matrix* $\mathbf{\Sigma}_{A,c}\mathbf{Z}_c$.

(AD1) *Define* $\mathbf{\Gamma}_r^*$ *as the eigenvector matrix of* $\text{tr}(\mathbf{A}_c'\mathbf{A}_c) \cdot p^{-\delta_{r,k_r}} q^{-\delta_{c,1}} \mathbf{Z}_r^{1/2}\mathbf{\Sigma}_{A,r}\mathbf{Z}_r^{1/2}$, $\mathbf{\Gamma}_c^*$ *as the eigenvector matrix of* $\text{tr}(\mathbf{A}_r'\mathbf{A}_r) \cdot q^{-\delta_{c,k_c}} p^{-\delta_{r,1}} \mathbf{Z}_c^{1/2}\mathbf{\Sigma}_{A,c}\mathbf{Z}_c^{1/2}$, *and*

$$\begin{aligned}
\mathbf{H}_r^* &:= \text{tr}(\mathbf{A}_c'\mathbf{A}_c)^{1/2} \cdot \mathbf{D}_r^{-1/2}(\mathbf{\Gamma}_r^*)'\mathbf{Z}_r^{1/2}, \\
\mathbf{H}_c^* &:= \text{tr}(\mathbf{A}_r'\mathbf{A}_r))^{1/2} \cdot \mathbf{D}_c^{-1/2}(\mathbf{\Gamma}_c^*)'\mathbf{Z}_c^{1/2}, \\
\mathbf{D}_r &:= \text{tr}(\mathbf{A}_c'\mathbf{A}_c) \cdot \text{diag}\{\lambda_1(\mathbf{A}_r'\mathbf{A}_r), \ldots, \lambda_{k_r}(\mathbf{A}_r'\mathbf{A}_r)\}, \\
\mathbf{D}_c &:= \text{tr}(\mathbf{A}_r'\mathbf{A}_r) \cdot \text{diag}\{\lambda_1(\mathbf{A}_c'\mathbf{A}_c), \ldots, \lambda_{k_c}(\mathbf{A}_c'\mathbf{A}_c)\}, \\
\mathbf{\Xi}_{r,j} &:= \operatorname*{plim}_{p,q,T\to\infty} \text{Var}\Big\{ \sum_{i=1}^{p} \mathbf{Q}_{r,i\cdot} \sum_{t=1}^{T} (\mathbf{C}_t\mathbf{E}_t')_{ij} \Big\}, \\
\mathbf{\Xi}_{c,j} &:= \operatorname*{plim}_{p,q,T\to\infty} \text{Var}\Big\{ \sum_{i=1}^{q} \mathbf{Q}_{c,i\cdot} \sum_{t=1}^{T} (\mathbf{C}_t'\mathbf{E}_t)_{ij} \Big\}.
\end{aligned}$$

*We assume both* $Tp^{2\delta_{r,k_r}-\delta_{r,1}} q^{2\delta_{c,1}-1} \cdot \big\|\mathbf{D}_r^{-1}\mathbf{H}_r^*\mathbf{\Xi}_{r,j}(\mathbf{H}_r^*)'\mathbf{D}_r^{-1}\big\|_F$ *and* $Tq^{2\delta_{c,k_c}-\delta_{c,1}} p^{2\delta_{r,1}-1} \cdot \big\|\mathbf{D}_c^{-1}\mathbf{H}_c^*\mathbf{\Xi}_{c,j}(\mathbf{H}_c^*)'\mathbf{D}_c^{-1}\big\|_F$ *are of constant order.*

(R2) (Further rate assumptions). We have

$$\begin{aligned}
&T^{-1}p^{1+2\delta_{r,1}-3\delta_{r,k_r}} q^{1-\delta_{c,1}-\delta_{c,k_c}}, \quad p^{1+\delta_{r,1}-4\delta_{r,k_r}} q^{2-\delta_{c,1}-\delta_{c,k_c}}, \\
&T^{-1}q^{1+2\delta_{c,1}-3\delta_{c,k_c}} p^{1-\delta_{r,1}-\delta_{r,k_r}}, \quad q^{1+\delta_{c,1}-4\delta_{c,k_c}} p^{2-\delta_{r,1}-\delta_{r,k_r}} = o(1).
\end{aligned}$$

Assumption (AD1) appears in Cen and Lam (2025b) as well, and similar to the discussion therein, it is a lower bound condition as we can show the upper bound in the theoretical proof. Essentially, this assumption facilitates the proof of the asymptotic normality of each row of $\widehat{\mathbf{Q}}_r$ and $\widehat{\mathbf{Q}}_c$, by asserting that in the decomposition of $\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r$ (resp. $\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c$), certain terms are dominating others even in the lower bound, and hence is truly dominating rather than just having the upper bounds dominating other upper bounds as in the proofs of similar theorems in the broader literature of factor models. Assumption (R2) is needed to make sure that the estimated common component $\widehat{\mathbf{C}}_t$ is consistent element-wise (see Theorem 5.3). This is satisfied automatically when all factors are pervasive, for instance.

**Theorem 5.4** *Let all assumptions in Theorem 5.2 hold, and let* $\Sigma_{\epsilon,ij}$ *be the* $(i, j)$ *entry of* $\mathbf{\Sigma}_\epsilon$

*in Assumption (E1). Assume also for $i \in [p]$ and $j \in [q]$,*

$$\gamma_{\alpha,i}^2 := \lim_{q\to\infty} \frac{1}{q}\sum_{j=1}^{q} \Sigma_{\epsilon,ij}^2, \quad \gamma_{\beta,j}^2 := \lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p} \Sigma_{\epsilon,ij}^2, \quad \gamma_{\mu}^2 := \lim_{p,q\to\infty} \frac{1}{pq}\sum_{i\in[p],j\in[q]} \Sigma_{\epsilon,ij}^2.$$

*Then for each $t \in [T]$,*

$$\sqrt{pq}(\widehat{\mu}_t - \mu_t) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma_\mu^2).$$

*Take a finite integer $m$ and integers $i_1 < \cdots < i_m$ ($i_\ell \in [p]$). Define $\boldsymbol{\theta}_{\alpha,t} := (\alpha_{t,i_1}, \ldots, \alpha_{t,i_m})'$ and similarly for $\widehat{\boldsymbol{\theta}}_{\alpha,t}$, where $\alpha_{t,i}$ is the $i$-th element of $\boldsymbol{\alpha}_t$. For a fixed $t \in [T]$,*

$$\sqrt{q}\,(\widehat{\boldsymbol{\theta}}_{\alpha,t} - \boldsymbol{\theta}_{\alpha,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \text{diag}(\gamma_{\alpha,i_1}^2, \ldots, \gamma_{\alpha,i_m}^2)).$$

*Similarly, take integers $j_1 < \cdots < j_m$ where $j_\ell \in [q]$. Define $\boldsymbol{\theta}_{\beta,t} := (\beta_{t,j_1}, \ldots, \beta_{t,j_m})'$ and similarly for $\widehat{\boldsymbol{\theta}}_{\beta,t}$, where $\beta_{t,j}$ is the $j$-th element of $\boldsymbol{\beta}_t$. Then for a fixed $t \in [T]$,*

$$\sqrt{p}(\widehat{\boldsymbol{\theta}}_{\beta,t} - \boldsymbol{\theta}_{\beta,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \text{diag}(\gamma_{\beta,j_1}^2, \ldots, \gamma_{\beta,j_m}^2)).$$

*Moreover, for $i \in [p]$ and $j \in [q]$, if the rate for $\widehat{C}_{t,ij} - C_{t,ij}$ in Theorem 5.3 is $o(1)$, then*

$$\widehat{\gamma}_{\alpha,i}^2 := q^{-1}(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii}, \quad \widehat{\gamma}_{\beta,j}^2 := p^{-1}(\widehat{\mathbf{E}}_t'\widehat{\mathbf{E}}_t)_{jj}, \quad \widehat{\gamma}_{\mu}^2 := p^{-1}\sum_{i=1}^{p}\widehat{\gamma}_{\alpha,i}^2 = q^{-1}\sum_{j=1}^{q}\widehat{\gamma}_{\beta,j}^2$$

*are consistent estimators for $\gamma_{\alpha,i}^2$, $\gamma_{\beta,j}^2$ and $\gamma_\mu^2$ respectively under Assumption (R2), so that*

$$\begin{aligned} \sqrt{pq}\,\widehat{\gamma}_\mu^{-1}(\widehat{\mu}_t - \mu_t) &\xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \\ \sqrt{q}\,\text{diag}(\widehat{\gamma}_{\alpha,i_1}^{-1}, \ldots, \widehat{\gamma}_{\alpha,i_m}^{-1})(\widehat{\boldsymbol{\theta}}_{\alpha,t} - \boldsymbol{\theta}_{\alpha,t}) &\xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_m), \\ \sqrt{p}\,\text{diag}(\widehat{\gamma}_{\beta,j_1}^{-1}, \ldots, \widehat{\gamma}_{\beta,j_m}^{-1})(\widehat{\boldsymbol{\theta}}_{\beta,t} - \boldsymbol{\theta}_{\beta,t}) &\xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_m). \end{aligned}$$

Recall from Remark 5.2 that FM can be expressed in MEFM, and hence the ability to make inferences on the elements of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ does not facilitate a test for the necessity of MEFM over FM. For such a test, please see Section 5.3.6. Theorem 5.4 gives us the ability to infer on the level of row and column main effects at each time point, which is important if we have target comparisons we want to make for these effects. For instance, if each row represents a country, we can easily compare the main effects at time $t$ for the first country against the average of the second and third simply by considering $\mathbf{g} := (1, -1/2, -1/2)'$, $\boldsymbol{\theta}_{\alpha,t} := (\alpha_{t,1}, \alpha_{t,2}, \alpha_{t,3})'$ and using Theorem 5.4 to arrive at

$$\sqrt{q}\,(\mathbf{g}'\text{diag}(\widehat{\gamma}_{\alpha,1}^2, \widehat{\gamma}_{\alpha,2}^2, \widehat{\gamma}_{\alpha,3}^2)\mathbf{g})^{-1/2}\mathbf{g}'(\widehat{\boldsymbol{\theta}}_{\alpha,t} - \boldsymbol{\theta}_{\alpha,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

**Theorem 5.5** *Let all the assumptions under Theorem 5.2 hold, in addition to (AD1) and (L2). Suppose $k_r$ and $k_c$ are fixed and $p, q, T \to \infty$. If $Tq = o(p^{\delta_{r,1}+\delta_{r,k_r}})$, we have*

$$(Tp^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1})^{1/2} \cdot (\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, T^{-1}p^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1} \cdot \mathbf{D}_r^{-1}\mathbf{H}_r^*\mathbf{\Xi}_{r,j}(\mathbf{H}_r^*)'\mathbf{D}_r^{-1}\big).$$

*On the other hand, if $Tp = o(q^{\delta_{c,1}+\delta_{c,k_c}})$, we have*

$$(Tq^{2\delta_{c,k_c}-\delta_{c,1}}p^{2\delta_{r,1}-1})^{1/2} \cdot (\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, T^{-1}q^{2\delta_{c,k_c}-\delta_{c,1}}p^{2\delta_{r,1}-1} \cdot \mathbf{D}_c^{-1}\mathbf{H}_c^*\mathbf{\Xi}_{c,j}(\mathbf{H}_c^*)'\mathbf{D}_c^{-1}\big).$$

Theorem 5.5 is essentially Theorem 3 of Cen and Lam (2025b) when $K = 2$ and $\eta = 0$ (full observations), having the same rate of convergence under potentially weak factors. Hence our MEFM estimation has successfully estimated and removed all time-varying main effects and grand mean, leaving the estimation of the common component exactly the same as in FM.

### 5.3.5 Estimation of the asymptotic covariance matrix

To practically use Theorem 5.5 for inference, we need to estimate the covariance matrices for $\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}$ and $\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}$. Based on $\{\widehat{\mathbf{D}}_r, \widehat{\mathbf{Q}}_r, \widehat{\mathbf{C}}_t, \widehat{\mathbf{E}}_t\}_{t\in[T]}$ and $\{\widehat{\mathbf{D}}_c, \widehat{\mathbf{Q}}_c, \widehat{\mathbf{C}}_t, \widehat{\mathbf{E}}_t\}_{t\in[T]}$ respectively, we use the heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West, 1987).

For $\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}$, with $\eta_r$ such that $\eta_r \to \infty$, $\eta_r/(Tp^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1})^{1/4} \to 0$, define an HAC estimator

$$\widehat{\mathbf{\Sigma}}_{r,j}^{HAC} := \mathbf{D}_{r,0,j} + \sum_{\nu=1}^{\eta_r}\Big(1 - \frac{\nu}{1+\eta_r}\Big)\Big(\mathbf{D}_{r,\nu,j} + \mathbf{D}_{r,\nu,j}'\Big), \text{ where}$$
$$\mathbf{D}_{r,\nu,j} := \sum_{t=1+\nu}^{T}\Big\{\sum_{i=1}^{p}\Big(T^{-1}\widehat{\mathbf{D}}_r^{-1}\widehat{\mathbf{Q}}_r'\sum_{s=1}^{T}\widehat{\mathbf{C}}_s\widehat{\mathbf{C}}_{s,i\cdot}\Big)(\widehat{\mathbf{C}}_t\widehat{\mathbf{E}}_t')_{ij}\Big\} \cdot \Big\{\sum_{i=1}^{p}\Big(T^{-1}\widehat{\mathbf{D}}_r^{-1}\widehat{\mathbf{Q}}_r'\sum_{s=1}^{T}\widehat{\mathbf{C}}_s\widehat{\mathbf{C}}_{s,i\cdot}\Big)(\widehat{\mathbf{C}}_{t-\nu}\widehat{\mathbf{E}}_{t-\nu}')_{ij}\Big\}'.$$

For $\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}$, with $\eta_c$ such that $\eta_c \to \infty$, $\eta_c/(Tq^{2\delta_{c,k_c}-\delta_{c,1}}p^{2\delta_{r,1}-1})^{1/4} \to 0$, define

$$\widehat{\mathbf{\Sigma}}_{c,j}^{HAC} := \mathbf{D}_{c,0,j} + \sum_{\nu=1}^{\eta_c}\Big(1 - \frac{\nu}{1+\eta_c}\Big)\Big(\mathbf{D}_{c,\nu,j} + \mathbf{D}_{c,\nu,j}'\Big), \text{ where}$$
$$\mathbf{D}_{c,\nu,j} := \sum_{t=1+\nu}^{T}\Big\{\sum_{i=1}^{q}\Big(T^{-1}\widehat{\mathbf{D}}_c^{-1}\widehat{\mathbf{Q}}_c'\sum_{s=1}^{T}\widehat{\mathbf{C}}_s'\widehat{\mathbf{C}}_{s,\cdot i}\Big)(\widehat{\mathbf{C}}_t'\widehat{\mathbf{E}}_t)_{ij}\Big\}$$

$$\cdot \Big\{ \sum_{i=1}^{q} \Big( T^{-1} \widehat{\mathbf{D}}_c^{-1} \widehat{\mathbf{Q}}_c' \sum_{s=1}^{T} \widehat{\mathbf{C}}_s' \widehat{\mathbf{C}}_{s,\cdot i} \Big) (\widehat{\mathbf{C}}_{t-\nu}' \widehat{\mathbf{E}}_{t-\nu})_{ij} \Big\}'.$$

**Theorem 5.6** *Let all the assumptions under Theorem 5.2 hold, in addition to (L2), (AD1) and (R2). Suppose $k_r$ and $k_c$ are fixed and $p, q, T \to \infty$. If $Tq = o(p^{\delta_{r,1}+\delta_{r,k_r}})$, then*

1. $\widehat{\mathbf{D}}_r^{-1} \widehat{\boldsymbol{\Sigma}}_{r,j}^{HAC} \widehat{\mathbf{D}}_r^{-1}$ *is consistent for* $\mathbf{D}_r^{-1} \mathbf{H}_r^* \boldsymbol{\Xi}_{r,j} (\mathbf{H}_r^*)' \mathbf{D}_r^{-1}$;
2. $T \cdot \big(\widehat{\boldsymbol{\Sigma}}_{r,j}^{HAC}\big)^{-1/2} \widehat{\mathbf{D}}_r (\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r \mathbf{Q}_{r,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_r})$.

*On the other hand, if $Tp = o(q^{\delta_{c,1}+\delta_{c,k_c}})$, then*

3. $\widehat{\mathbf{D}}_c^{-1} \widehat{\boldsymbol{\Sigma}}_{c,j}^{HAC} \widehat{\mathbf{D}}_c^{-1}$ *is consistent for* $\mathbf{D}_c^{-1} \mathbf{H}_c^* \boldsymbol{\Xi}_{c,j} (\mathbf{H}_c^*)' \mathbf{D}_c^{-1}$;
4. $T \cdot \big(\widehat{\boldsymbol{\Sigma}}_{c,j}^{HAC}\big)^{-1/2} \widehat{\mathbf{D}}_c (\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c \mathbf{Q}_{c,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_c})$.

### 5.3.6 Testing the sufficiency of FM versus MEFM

In the last subsection, we introduce how to make inferences on various parameters of MEFM. However, to test if FM is sufficient against our proposed MEFM, simple inferences on the model parameters are not enough in the face of Remark 5.2. Formally, we want to test, for the time horizon $t \in [T]$,

$$H_0 : \text{FM is sufficient over } \ t \in [T] \quad \longleftrightarrow \quad H_1 : \text{MEFM is needed over } \ t \in [T].$$

The above problem is complicated by the fact that, in Section 5.2.1, we have seen that MEFM can always be expressed as FM if we are willing to potentially consider a large number of factors. So, how "large" an increase in the number of factors do we consider unacceptable?

Remark 5.2 tells us that a special form of MEFM can be expressed back in FM:

$$\mathbf{Y}_t = \mu_t \mathbf{1}_p \mathbf{1}_q' + \boldsymbol{\alpha}_t \mathbf{1}_q' + \mathbf{1}_p \boldsymbol{\beta}_t' + \mathbf{M}_p \acute{\mathbf{C}}_t \mathbf{M}_q + \mathbf{E}_t = \mathbf{A}_r \mathbf{F}_t \mathbf{A}_c' + \mathbf{E}_t, \quad t \in [T],$$

where $\acute{\mathbf{C}}_t := \mathbf{A}_r \mathbf{F}_t \mathbf{A}_c'$ and

$$\mu_t := (pq)^{-1} \mathbf{1}_p' \acute{\mathbf{C}}_t \mathbf{1}_q, \quad \boldsymbol{\alpha}_t := q^{-1} \mathbf{M}_p \acute{\mathbf{C}}_t \mathbf{1}_q, \quad \boldsymbol{\beta}_t := p^{-1} \mathbf{M}_q \acute{\mathbf{C}}_t' \mathbf{1}_p.$$

If $\mathbf{A}_r$ has rank $k_r$ satisfying Assumption (L1) and $\mathbf{A}_c$ has rank $k_c$, the potential rank of $\mathbf{M}_p \mathbf{A}_r$ is $k_r - 1$ (when a column in $\mathbf{A}_r$ is parallel to $\mathbf{1}_p$), and that of $\mathbf{M}_q \mathbf{A}_c$ is $k_c - 1$ (when a column in $\mathbf{A}_c$ is parallel to $\mathbf{1}_q$), demonstrating that FM can have an increase in the number of factors, albeit still finite.

Another special example is when both $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are zero, but $\mu_t \neq 0$. Then we can write MEFM as

$$\mathbf{Y}_t = \mu_t \mathbf{1}_p \mathbf{1}_q' + \mathbf{A}_r \mathbf{F}_t \mathbf{A}_c' + \mathbf{E}_t = (\mathbf{A}_r, \mathbf{1}_p) \begin{pmatrix} \mathbf{F}_t & \mathbf{0} \\ \mathbf{0}' & \mu_t/(pq) \end{pmatrix} \begin{pmatrix} \mathbf{A}_c' \\ \mathbf{1}_q' \end{pmatrix} + \mathbf{E}_t,$$

which is FM with loading matrices $(\mathbf{A}_r, \mathbf{1}_p)$ and $(\mathbf{A}_c, \mathbf{1}_q)$ respectively, and an increase by 1 for both the number of row and column factors.

In light of the above examples, we deem FM sufficient if and only if the number of factors in the FM is still finite and any model variables satisfy the Assumptions in Section 5.3.1.

To be able to test $H_0$ against $H_1$, define $\check{\mathbf{E}}_t$ to be the residual matrix after a fitting of FM (a similar procedure to fitting MEFM but treating $\mu_t$, $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ as zero), with

$$\check{\mathbf{E}}_t := \mathbf{Y}_t - \check{\mathbf{C}}_t, \quad \text{where } \check{\mathbf{C}}_t := \check{\mathbf{A}}_r \check{\mathbf{A}}_r' \mathbf{Y}_t \check{\mathbf{A}}_c \check{\mathbf{A}}_c',$$

with $\check{\mathbf{A}}_r$ and $\check{\mathbf{A}}_c$ the $p \times \ell_r$ and $q \times \ell_c$ eigenmatrices of $\sum_{t=1}^T \mathbf{Y}_t \mathbf{Y}_t'$ and $\sum_{t=1}^T \mathbf{Y}_t' \mathbf{Y}_t$ respectively.

**Theorem 5.7** *Let all the assumptions in Theorem 5.2 hold, on top of (R2). Also assume that* $\widehat{C}_{t,ij} - C_{t,ij} = o_P(\min(p^{-1/2}, q^{-1/2}))$ *in Theorem 5.3. Suppose* $k_r, k_c, \ell_r$ *and* $\ell_c$ *are all fixed and known. Then under* $H_0$*, for each* $t \in [T]$*, we have*

$$\frac{(\widehat{\mathbf{E}}_t \widehat{\mathbf{E}}_t')_{ii} - \sum_{j=1}^q \Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{j=1}^q \mathrm{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}}, \; \frac{(\check{\mathbf{E}}_t \check{\mathbf{E}}_t')_{ii} - \sum_{j=1}^q \Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{j=1}^q \mathrm{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}} \xrightarrow{\mathcal{D}} Z_{i,t} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \;\; \text{for each } \; i \in [p];$$
$$\frac{(\widehat{\mathbf{E}}_t' \widehat{\mathbf{E}}_t)_{jj} - \sum_{i=1}^p \Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{i=1}^p \mathrm{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}}, \; \frac{(\check{\mathbf{E}}_t' \check{\mathbf{E}}_t)_{jj} - \sum_{i=1}^p \Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{i=1}^p \mathrm{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}} \xrightarrow{\mathcal{D}} W_{j,t} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \;\; \text{for each } \; j \in [q],$$

*where* $Z_{h,t}$ *is independent of* $Z_{\ell,t}$ *and* $W_{h,t}$ *is independent of* $W_{\ell,t}$ *for* $h \neq \ell$*. The same asymptotic results hold true under* $H_1$ *for* $(\widehat{\mathbf{E}}_t \widehat{\mathbf{E}}_t')_{ii}$ *and* $(\widehat{\mathbf{E}}_t' \widehat{\mathbf{E}}_t)_{jj}$ *respectively for* $i \in [p], j \in [q]$*.*

The assumption $\widehat{C}_{t,ij} - C_{t,ij} = o_P(\min(p^{-1/2}, q^{-1/2}))$ is satisfied, for instance, when all factors are pervasive and $T, p, q$ are of the same order. Theorem 5.7 tells us that for each $t \in [T]$, both

$$x_{\alpha,t} := \max_{i \in [p]} \widehat{\gamma}_{\alpha,i}^2 = \max_{i \in [p]} \{q^{-1}(\widehat{\mathbf{E}}_t \widehat{\mathbf{E}}_t')_{ii}\}, \quad y_{\alpha,t} := \max_{i \in [p]} \check{\gamma}_{\alpha,i}^2 := \max_{i \in [p]} \{q^{-1}(\check{\mathbf{E}}_t \check{\mathbf{E}}_t')_{ii}\}$$

are distributed approximately the same for large $q$ under $H_0$, and $x_{\alpha,t}$ in particular is distributed the same no matter under $H_0$ or $H_1$. Similarly, define

$$x_{\beta,t} := \max_{j \in [q]} \widehat{\gamma}_{\beta,j}^2 = \max_{j \in [q]} \{p^{-1}(\widehat{\mathbf{E}}_t' \widehat{\mathbf{E}}_t)_{jj}\}, \quad y_{\beta,t} := \max_{j \in [q]} \check{\gamma}_{\beta,j}^2 := \max_{j \in [q]} \{p^{-1}(\check{\mathbf{E}}_t' \check{\mathbf{E}}_t)_{jj}\},$$

which are distributed approximately the same for large $p$ under $H_0$ from Theorem 5.7, and $x_{\beta,t}$

in particular is distributed the same no matter under $H_0$ or $H_1$. To utilize Theorem 5.7 in testing $H_0$, we impose an additional assumption on the core factor and idiosyncratic noise as follows.

(E3) (Tail condition in $\mathbf{F}_t$ and $\mathbf{E}_t$). *Each element in the time series* $\{\mathbf{X}_{f,t}\}$, $\{\mathbf{X}_{e,t}\}$ *and* $\{\mathbf{X}_{\epsilon,t}\}$ *has sub-Gaussian tail.*

This assumption allows us to make convergence statements in quantiles to be defined in Theorem 5.8 below. Define $\mathbb{F}_{x,\alpha}$, $\mathbb{F}_{y,\alpha}$, $\mathbb{F}_{x,\beta}$ and $\mathbb{F}_{y,\beta}$ the empirical cumulative distribution functions for the series $\{x_{\alpha,t}\}_{t\in[T]}$, $\{y_{\alpha,t}\}_{t\in[T]}$, $\{x_{\beta,t}\}_{t\in[T]}$ and $\{y_{\beta,t}\}_{t\in[T]}$ respectively:

$$\mathbb{F}_{x,\alpha}(c) := \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\{x_{\alpha,t}\leq c\},\quad \mathbb{F}_{y,\alpha}(c) := \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\{y_{\alpha,t}\leq c\},$$
$$\mathbb{F}_{x,\beta}(c) := \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\{x_{\beta,t}\leq c\},\quad \mathbb{F}_{y,\beta}(c) := \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\{y_{\beta,t}\leq c\}.$$

**Theorem 5.8** *Let Assumption (E3) and all the assumptions in Theorem 5.7 hold. Moreover, we assume for simplicity of presentation that all factors are pervasive. Define for* $0<\theta<1$,

$$\widehat{q}_{x,\alpha}(\theta) := \inf\{c \mid \mathbb{F}_{x,\alpha}(c)\geq\theta\},\quad \widehat{q}_{x,\beta}(\theta) := \inf\{c \mid \mathbb{F}_{x,\beta}(c)\geq\theta\},$$

*Then under* $H_0$*, as* $T,p,q\to\infty$*, we have for each* $t\in[T]$*,*

$$\mathbb{P}_{y,\alpha}[y_{\alpha,t}>\widehat{q}_{x,\alpha}(\theta)]\leq 1-\theta+O_P\Big\{\Big(\frac{1}{\sqrt{p}}+\frac{1}{\sqrt{q}}+\frac{1}{\sqrt{T}}+\frac{\sqrt{q}}{p}+\sqrt{\frac{q}{Tp}}\Big)\log^2(T)\log(p)\log^2(q)\Big\},$$
$$\mathbb{P}_{y,\beta}[y_{\beta,t}>\widehat{q}_{x,\beta}(\theta)]\leq 1-\theta+O_P\Big\{\Big(\frac{1}{\sqrt{p}}+\frac{1}{\sqrt{q}}+\frac{1}{\sqrt{T}}+\frac{\sqrt{p}}{q}+\sqrt{\frac{p}{Tq}}\Big)\log^2(T)\log^2(p)\log(q)\Big\},$$

*where* $\mathbb{P}_{y,\alpha}$ *and* $\mathbb{P}_{y,\beta}$ *are empirical probability measures induced by* $\mathbb{F}_{y,\alpha}$ *and* $\mathbb{F}_{y,\beta}$ *respectively.*

The assumption of pervasive factors is for the ease of presentation of the rate added to the two probability statements above. But if some factors are weaker, then the convergence rate of the common components will be adversely affected, and the rate in the probability statements above will be inflated.

With Theorem 5.8, we can test $H_0$ at significance level $1-\theta$ asymptotically using the test statistics $y_{\alpha,t}$ and $y_{\beta,t}$, and rejection rules $y_{\alpha,t}\geq\widehat{q}_{x,\alpha}(\theta)$ and $y_{\beta,t}\geq\widehat{q}_{x,\beta}(\theta)$ respectively. Since we have $y_{\alpha,t}$ and $y_{\beta,t}$ for $t\in[T]$, we can assess the significance level under $H_0$ by calculating

$$\text{Significance levels} = T^{-1}\sum_{t=1}^{T}\mathbb{1}\{y_{\alpha,t}\geq\widehat{q}_{x,\alpha}(\theta)\},\quad T^{-1}\sum_{t=1}^{T}\mathbb{1}\{y_{\beta,t}\geq\widehat{q}_{x,\beta}(\theta)\},$$

and see if they are close to $1-\theta$. If $H_0$ is not true, then if any $\alpha_{t,i}$ is large, we expect $y_{\alpha,t}$ to be large. Or, if any $\beta_{t,j}$ is large, we expect $y_{\beta,t}$ to be large.

In practice for testing $H_0$ against $H_1$, we estimate $k_r$ and $k_c$, and set $\ell_r = k_r + 1$ and $\ell_c = k_c + 1$ in light of the previous argument on how a special form of MEFM can be expressed back in FM. For the estimation of $k_r$ and $k_c$, see Section 5.3.7.

**Remark 5.3** *The size of our test is theoretically guaranteed by Theorem 5.8, while the test power is shown by numerical results later in Section 5.4.1. To appreciate the difficulty in deriving the test power, recall that any FM can be rewritten as a MEFM according to Remark 5.2. Hence, any attempt in studying the power in terms of the magnitude of variables in MEFM, e.g. L2-norm of $\boldsymbol{\alpha}_t$, would fail, since a large $\boldsymbol{\alpha}_t$ could be present in an FM. A possible direction to tackle this problem is to formalise some local alternative hypotheses that are relatively easier to work on, which we defer to our future endeavours.*

### 5.3.7 Estimation of the number of factors

From (5.4), we have $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$ essentially being the row sample covariance matrix and $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t$ the column sample covariance matrix. We then propose the eigenvalue-ratio estimators for the number of factors as

$$\widehat{k}_r := \arg\min_j \left\{ \frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r}{\lambda_j\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r},\ j \in [\lfloor p/2 \rfloor] \right\}, \quad \xi_r \asymp pq\big[(Tq)^{-1/2} + p^{-1/2}\big], \tag{5.7}$$

$$\widehat{k}_c := \arg\min_j \left\{ \frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t\big) + \xi_c}{\lambda_j\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t\big) + \xi_c},\ j \in [\lfloor q/2 \rfloor] \right\}, \quad \xi_c \asymp pq\big[(Tp)^{-1/2} + q^{-1/2}\big]. \tag{5.8}$$

Ratio-based estimators are widely studied by researchers. For example, an eigenvalue-ratio estimator is considered in Lam and Yao (2012) and Ahn and Horenstein (2013), while a cumulative eigenvalue ratio estimator is proposed by Zhang et al. (2024a). Our proposed estimator is similar to the perturbed eigenvalue-ratio estimators as in Pelger (2019). Technically, we can minimise (5.7) (resp. (5.8)) over any $j \in [p]$ (resp. $j \in [q]$), but it is very reasonable to assume $k_r \le p/2$ and $k_c \le q/2$ in all applications of factor models. The correction terms $\xi_r$ and $\xi_c$ are added to stabilise the ratio so that consistency follows from the theorem below.

**Theorem 5.9** *Under Assumptions (IC1), (M1), (F1), (L1), (E1), (E2) and (R1), we have:*

1. *$\widehat{k}_r$ is a consistent estimator of $k_r$ if*

$$\begin{cases} p^{1-\delta_{r,k_r}} q^{1-\delta_{c,1}}\big[(Tq)^{-1/2} + p^{-1/2}\big] = o(p^{\delta_{r,j+1}-\delta_{r,j}}), & j \in [k_r - 1] \text{ with } k_r \ge 2; \\ p^{1-\delta_{r,1}} q^{1-\delta_{c,1}}\big[(Tq)^{-1/2} + p^{-1/2}\big] = o(1), & k_r = 1. \end{cases}$$

2. $\widehat{k}_c$ *is a consistent estimator of* $k_c$ *if*

$$\begin{cases} q^{1-\delta_{c,k_c}}p^{1-\delta_{r,1}}[(Tp)^{-1/2}+q^{-1/2}] = o(q^{\delta_{c,j+1}-\delta_{c,j}}), & j \in [k_c-1] \textit{ with } k_c \geq 2; \\ q^{1-\delta_{c,1}}p^{1-\delta_{r,1}}[(Tp)^{-1/2}+q^{-1/2}] = o(1), & k_c = 1. \end{cases}$$

The extra rate conditions in the theorem are due to existence of potential weak factors and are trivially satisfied for pervasive factors. The theorem is similar to the consistency result in Cen and Lam (2025b) for matrix-valued factor models, and this implies that the number of factors in MEFM can be well estimated just as in the case of FM.

# 5.4 Numerical Results

## 5.4.1 Simulations

We demonstrate the performance of our estimators in this subsection. We will experiment different settings to assess consistency results as described in Theorem 5.2 and 5.3, followed by the asymptotic normality of our estimators in Theorem 5.4 and 5.5, where the covariance matrices can be constructed by their consistent estimators by Theorem 5.4 and Theorem 5.6, respectively. We then showcase the results for the rank estimators described in Theorem 5.9. As it is a first to consider matrix factor model with time-varying grand mean and main effects, we unveil the differences between MEFM and FM using numerical results that will illustrate Theorem 5.7.

For the data generating process, we use Assumptions (E1), (E2), and (F1) to generate general linear processes for the noise and factor series in model (5.3). To be precise, the elements in $\mathbf{F}_t$ are independent standardised AR(5) with AR coefficients 0.7, 0.3, -0.4, 0.2, and -0.1. The elements in $\mathbf{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are generated similarly, but their AR coefficients are (-0.7, -0.3, -0.4, 0.2, 0.1) and (0.8, 0.4, -0.4, 0.2, -0.1) respectively. The standard deviation of each element in $\boldsymbol{\epsilon}_t$ is generated by i.i.d. $|\mathcal{N}(0,1)|$. To test how robust our method is under heavy-tailed distribution, we consider two distributions for the innovation process in generating $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$: 1) i.i.d. $\mathcal{N}(0,1)$; 2) i.i.d. $t_3$.

The row factor loading matrix $\mathbf{A}_r$ is generated with $\mathbf{A}_r = \mathbf{M}_p\mathbf{U}_r\mathbf{B}_r$, where each entry of $\mathbf{U}_r \in \mathbb{R}^{p\times k_r}$ is i.i.d. $\mathcal{N}(0,1)$, and $\mathbf{B}_r \in \mathbb{R}^{k_r\times k_r}$ is diagonal with the $j$-th diagonal entry being $p^{-\zeta_{r,j}}$, $0 \leq \zeta_{r,j} \leq 0.5$. Pervasive (strong) factors have $\zeta_{r,j} = 0$, while weak factors have $0 < \zeta_{r,j} \leq 0.5$. Note that $\mathbf{M}_p$ is defined in (5.4) so that (IC1) is satisfied. In a similar way, the column factor loading matrix $\mathbf{A}_c$ is generated independently. Each entry of $\mathbf{A}_{e,r} \in \mathbb{R}^{p\times k_{e,r}}$ is i.i.d. $\mathcal{N}(0,1)$ and has independent probability of 0.95 being set exactly to 0, and $\mathbf{A}_{e,c}$ is generated similarly. We fix $k_{e,r} = k_{e,c} = 2$ throughout the subsection.

For any $t \in [T]$, we generate $\mu_t = v_{\mu,t}$, $\boldsymbol{\alpha}_t = \mathbf{M}_p\mathbf{v}_{\alpha,t}$ and $\boldsymbol{\beta}_t = \mathbf{M}_q\mathbf{v}_{\beta,t}$, where $v_{\mu,t}$ is

$\mathcal{N}(m_\mu, \sigma_\mu^2)$, each element of $\mathbf{v}_{\alpha,t}$ is i.i.d. $\mathcal{N}(m_\alpha, \sigma_\alpha^2)$ and that of $\mathbf{v}_{\beta,t}$ is i.i.d. $\mathcal{N}(m_\beta, \sigma_\beta^2)$. We set $m_\mu = m_\alpha = m_\beta = 0$ and $\sigma_\mu = \sigma_\alpha = \sigma_\beta = 1$, and every experiment in this subsection is repeated 1000 times unless specified otherwise.

**Accuracy of various estimators**

To assess the accuracy of our estimators, we define the relative mean squared errors (MSE) for $\mu_t$, $\boldsymbol{\alpha}_t$, $\boldsymbol{\beta}_t$ and $\mathbf{C}_t$ as the following, respectively,

$$\text{relative MSE}_\mu = \frac{\sum_{t=1}^T (\mu_t - \widehat{\mu}_t)^2}{\sum_{t=1}^T \mu_t^2}, \quad \text{relative MSE}_{\boldsymbol{\alpha}} = \frac{\sum_{t=1}^T \|\boldsymbol{\alpha}_t - \widehat{\boldsymbol{\alpha}}_t\|^2}{\sum_{t=1}^T \|\boldsymbol{\alpha}_t\|^2},$$
$$\text{relative MSE}_{\boldsymbol{\beta}} = \frac{\sum_{t=1}^T \|\boldsymbol{\beta}_t - \widehat{\boldsymbol{\beta}}_t\|^2}{\sum_{t=1}^T \|\boldsymbol{\beta}_t\|^2}, \quad \text{relative MSE}_{\mathbf{C}} = \frac{\sum_{t=1}^T \|\mathbf{C}_t - \widehat{\mathbf{C}}_t\|_F^2}{\sum_{t=1}^T \|\mathbf{C}_t\|_F^2}.$$

For measuring the accuracy of our factor loading matrix estimators, we use the column space distance,

$$\mathcal{D}(\mathbf{Q}, \widehat{\mathbf{Q}}) = \left\|\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}' - \widehat{\mathbf{Q}}(\widehat{\mathbf{Q}}'\widehat{\mathbf{Q}})^{-1}\widehat{\mathbf{Q}}'\right\|,$$

for any given $\mathbf{Q}$ and $\widehat{\mathbf{Q}}$, which is a common measure in the literature such as Chen et al. (2022a) and Chen and Fan (2023).

We consider the following settings:

(Ia) $T = 100,\ p = q = 40,\ k_r = 1,\ k_c = 2$. All factors are pervasive with $\zeta_{r,j} = \zeta_{c,j} = 0$. All innovation processes in constructing $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are i.i.d. standard normal.

(Ib) Same as (Ia), but one factor is weak with $\zeta_{r,1} = 0.2$ and $\zeta_{c,1} = 0.2$. Set also $m_\alpha = -2$.

(Ic) Same as (Ia), but all innovation processes are i.i.d. $t_3$.

(Id) Same as (Ib), but $T = 100,\ p = q = 80$ and $\sigma_\alpha = 2$.

(Ie) Same as (Id), but $T = 200$.

(IIa-e) Same as (Ia) to (Ie) respectively, except that we generate $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ using white noise rather than AR(5).

Setting (IIa)–(IIe) are to investigate how temporal dependence in the noise affects our results.

We report the boxplots of accuracy measures for our estimators from Figure 5.1–5.6. Note first that stronger temporal dependence leads to larger variance of our estimators in general. The serial dependence mainly undermines the performance of our loading matrix estimators as shown in Figures 5.5 and 5.6, which in turn affects our common component estimator.

Considering the comparisons among (Ia) to (Ie), we see that relative MSE$_\mu$ can be improved by increasing the spatial dimensions, but is not affected by weak factors. Similar results can be

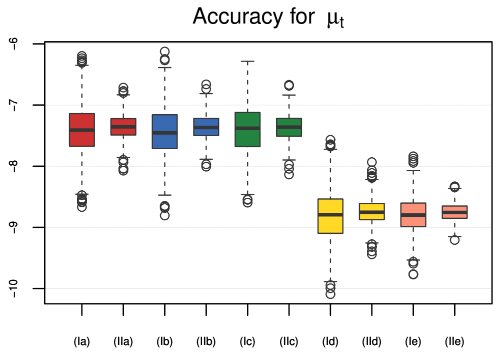


Figure 5.1: Plot of the relative MSE for $\mu_t$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

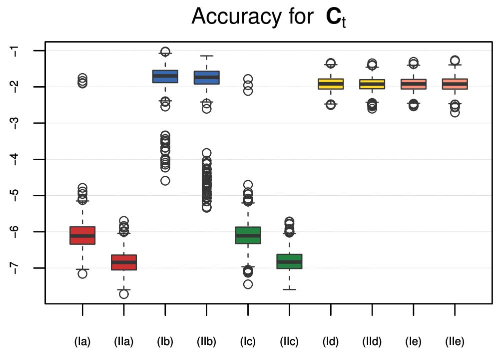


Figure 5.2: Plot of the relative MSE for $\mathbf{C}_t$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

seen from Figure 5.3 and Figure 5.4 for relative MSE$_\alpha$ and relative MSE$_\beta$. The detrimental effects of heavy-tailed innovation processes in Setting (Ic) are most reflected in the corresponding boxplots in Figure 5.4.

Weak factors can be detrimental to the accuracy of the factor loading matrix estimators, as can be seen by the significant rise in the factor loading space errors from Setting (Ia) to (Ib) in Figure 5.5 and 5.6. In fact, $\widehat{k}_c$ barely captures the second factor under Setting (Ib) and (IIb). Comparing Setting (Ib) with (Id), Figure 5.5 and 5.6 show that increase in data dimensions slightly improves our factor loading matrix estimators, which is consistent to the simulation results in Wang et al. (2019) for instance.

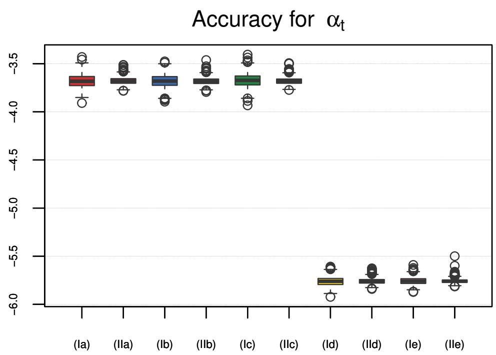


Figure 5.3: Plot of the relative MSE for $\boldsymbol{\alpha}_t$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

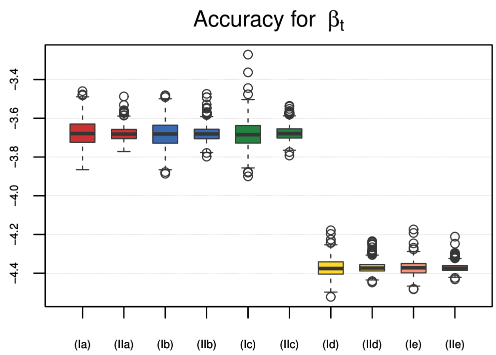


Figure 5.4: Plot of the relative MSE for $\boldsymbol{\beta}_t$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

### Performance for the estimation of the number of factors

We demonstrate the performance of our estimators for the number of factors, as described in Theorem 5.9. First, we set $\xi_r = pq[(Tq)^{-1/2} + p^{-1/2}]/5$ and $\xi_c = pq[(Tp)^{-1/2} + q^{-1/2}]/5$,

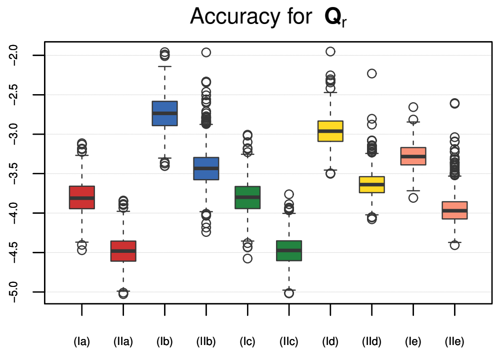


Figure 5.5: Plot of the row space distance $\mathcal{D}(\mathbf{Q}_r, \widehat{\mathbf{Q}}_r)$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

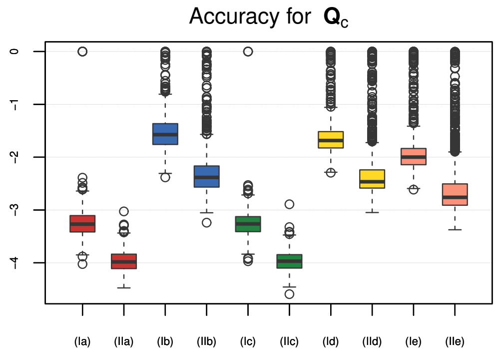


Figure 5.6: Plot of the column space distance $\mathcal{D}(\mathbf{Q}_c, \widehat{\mathbf{Q}}_c)$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

so that the conditions for $\xi_r$ and $\xi_c$ in (5.7) and (5.8) are respectively satisfied. A wide range of values other than $1/5$ for $\xi_r$ and $\xi_c$ are experimented, but $1/5$ is working the best in vast majority of settings, and hence we do not recommend treating this as a tuning parameter.

We present the results for each of the following settings:

(IIIa) $k_r = k_c = 3$. All factors are pervasive with $\zeta_{r,j} = \zeta_{c,j} = 0$ for all $j \in [3]$. All innovation processes involved are i.i.d. standard normal.

(IIIb) Same as (IIIa), but some factors are weak with $\zeta_{r,1} = \zeta_{c,1} = \zeta_{c,2} = 0.2$.

(IIIc) Same as (IIIa), but all factors are weak with $\zeta_{r,j} = \zeta_{c,j} = 0.2$ for all $j \in [3]$.

We experiment the above settings with $(p, q)$ pairs among $(10, 10)$, $(10, 20)$ and $(20, 20)$, with the choice $T = 0.5 \cdot pq$ or $T = pq$. The setup is similar to Wang et al. (2019) and Chen and Fan (2023), but we use smaller sets of dimensions since the accuracy of our estimators are approaching 1 with larger dimensions, which reveal little intricacies among different settings.

From the results in Table 5.1, our eigenvalue-ratio estimators is working well with MEFM. The accuracy of $\widehat{k}_r$ and $\widehat{k}_c$ suffers from the existence of weak factors, which is also seen in traditional FM (see for instance Chen and Lam (2024b) and Cen and Lam (2025b)). In particular, the accuracy of our estimators drops significantly as we move from Setting (IIIa) to (IIIc), and in general large dimensions are beneficial to our estimation. Lastly, note that although we have two weak factors in the column loading matrix while there is only one weak factor in the row loading matrix, the correct proportion of $\widehat{k}_c$ is much larger than that of $\widehat{k}_r$ for $(p, q) = (10, 20)$. This hints at the importance of data dimensions over factor strength, which can also be seen from the fact that the results for $(p, q) = (20, 20)$ under Setting (IIIc) are comparable with those for $(p, q) = (10, 10)$ under Setting (IIIa).

| $(\widehat{k}_r, \widehat{k}_c)$ | $p, q = 10, 10$ | | $p, q = 10, 20$ | | $p, q = 20, 20$ | |
|---|---|---|---|---|---|---|
| | $T = .5pq$ | $T = pq$ | $T = .5pq$ | $T = pq$ | $T = .5pq$ | $T = pq$ |
| | Setting (IIIa) | | | | | |
| $(2, 3)$ | 0.121 | 0.112 | 0.128 | 0.11 | 0 | 0.004 |
| $(3, 2)$ | 0.124 | 0.111 | 0.004 | 0.003 | 0.001 | 0.001 |
| $(3, 3)$ | **0.583** | **0.659** | **0.833** | **0.855** | **0.999** | **0.995** |
| other | 0.172 | 0.118 | 0.035 | 0.032 | 0 | 0 |
| | Setting (IIIb) | | | | | |
| $(2, 3)$ | 0.135 | 0.13 | 0.23 | 0.257 | 0.228 | 0.149 |
| $(3, 2)$ | 0.079 | 0.096 | 0.024 | 0.017 | 0.022 | 0.02 |
| $(3, 3)$ | **0.136** | **0.17** | **0.289** | **0.347** | **0.556** | **0.637** |
| other | 0.65 | 0.604 | 0.457 | 0.379 | 0.194 | 0.194 |
| | Setting (IIIc) | | | | | |
| $(2, 3)$ | 0.082 | 0.085 | 0.218 | 0.254 | 0.089 | 0.096 |
| $(3, 2)$ | 0.075 | 0.124 | 0.04 | 0.035 | 0.088 | 0.089 |
| $(3, 3)$ | **0.073** | **0.096** | **0.209** | **0.257** | **0.614** | **0.646** |
| other | 0.77 | 0.695 | 0.533 | 0.454 | 0.209 | 0.169 |

Table 5.1: Results for Setting (IIIa) to (IIIc). Each cell reports the frequency of $(\widehat{k}_r, \widehat{k}_c)$ under the setting in the corresponding column. The true number of factors is $(k_r, k_c) = (3, 3)$, and the cells corresponding to correct estimations are bolded.

**Asymptotic normality**

We numerically demonstrate the asymptotic normality results in Theorems 5.4 and 5.5 in the following. For the ease of demonstration, we consider $t = 10$ only for the asymptotic distribution of $\widehat{\mu}_t$, $\widehat{\boldsymbol{\theta}}_{\alpha,t} = (\widehat{\alpha}_{t,1}, \widehat{\alpha}_{t,2}, \widehat{\alpha}_{t,3})'$ and $\widehat{\boldsymbol{\theta}}_{\beta,t} = (\widehat{\beta}_{t,1}, \widehat{\beta}_{t,2}, \widehat{\beta}_{t,3})'$, and for $\widehat{\boldsymbol{\theta}}_{\alpha,t}$ and $\widehat{\boldsymbol{\theta}}_{\beta,t}$ we will only report results for the third component. We will also demonstrate the asymptotic normality for $(\widehat{\mathbf{Q}}_c)_{1\cdot}$ and present the results for $(\widehat{\mathbf{Q}}_c)_{11}$, i.e., the first entry of the first row in the column loading matrix estimator. To consistently estimate its covariance matrix, we use Theorem 5.6 with $\eta_c = \lfloor (Tpq)^{1/4}/5 \rfloor$.

We use heavy-tailed innovations to investigate the robustness of our results, hence Setting (Ic) is adapted except that we generate $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ using AR(1) with coefficient $-0.2$. Due to the different rates of convergence in Theorems 5.4 and 5.5, we specify different dimensions $(T, p, q)$ in the following settings:

$$\widehat{\mu}_t : (80, 100, 100), \quad \widehat{\boldsymbol{\theta}}_{\alpha,t} : (60, 60, 300), \quad \widehat{\boldsymbol{\theta}}_{\beta,t} : (60, 300, 60), \quad (\widehat{\mathbf{Q}}_c)_{1\cdot} : (60, 60, 300),$$

where the dimension setting for $(\widehat{\mathbf{Q}}_c)_{1\cdot}$ is to align with the rate conditions in Theorem 5.5 that

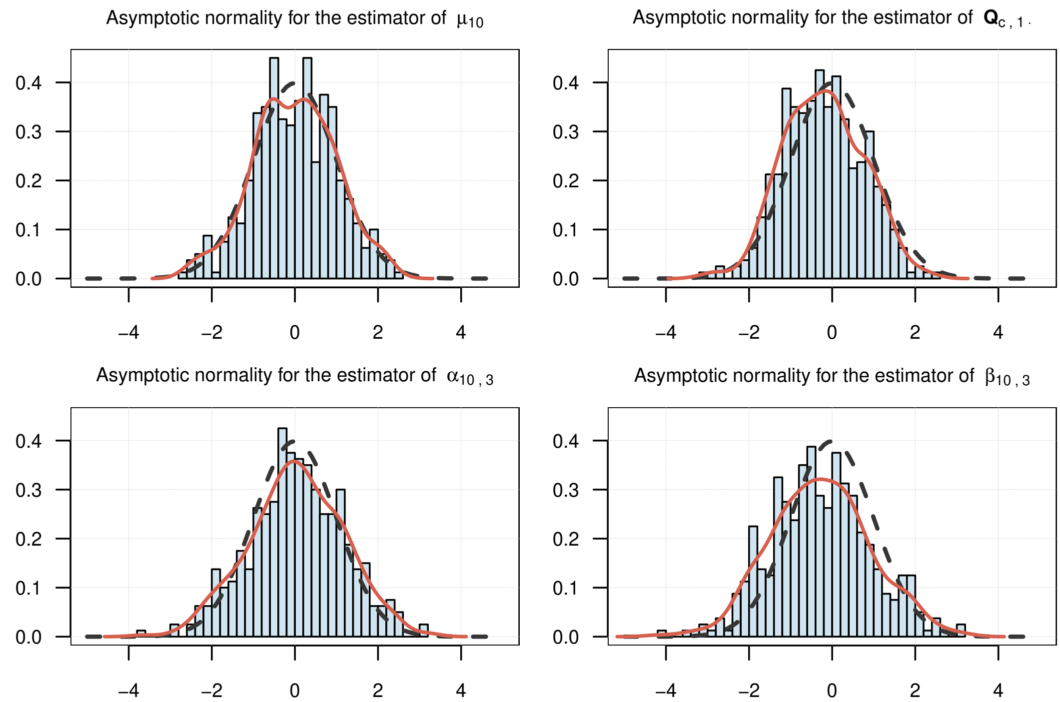


Figure 5.7: Histograms showing the asymptotic normality of $\sqrt{pq}\,\widehat{\gamma}_\mu^{-1}(\widehat{\mu}_{10} - \mu_{10})$ (top-left), $[T\,(\widehat{\mathbf{\Sigma}}_{c,1}^{HAC})^{-1/2}\widehat{\mathbf{D}}_c(\widehat{\mathbf{Q}}_{c,1\cdot} - \mathbf{H}_1^a\mathbf{Q}_{c,1\cdot})]_1$ (top-right), $\sqrt{q}\,[\mathrm{diag}(\widehat{\gamma}_{\alpha,1}^{-1}, \widehat{\gamma}_{\alpha,2}^{-1}, \widehat{\gamma}_{\alpha,3}^{-1})(\widehat{\boldsymbol{\theta}}_{\alpha,10} - \boldsymbol{\theta}_{\alpha,10})]_3$ (bottom-left), and $\sqrt{p}\,[\mathrm{diag}(\widehat{\gamma}_{\beta,1}^{-1}, \widehat{\gamma}_{\beta,2}^{-1}, \widehat{\gamma}_{\beta,3}^{-1})(\widehat{\boldsymbol{\theta}}_{\beta,10} - \boldsymbol{\theta}_{\beta,10})]_3$ (bottom-right). Each panel plots the empirical density (red), with the density curve for $\mathcal{N}(0,1)$ (black, dotted) also shown.

$Tp/q^2 \to 0$ under pervasive factors. Each setting is repeated 400 times, and we present the histograms of our four estimators in Figure 5.7.

Our plots stand as empirical evidence of Theorem 5.4, 5.5, and 5.6. It might worth noting that the spread of the normalised empirical density for $\widehat{\beta}_{10,3}$ is slightly larger than expected by comparing with the superimposed standard normal. The same problem is not seen in the histogram for $\widehat{\alpha}_{10,3}$. With true $(k_r, k_c) = (1,2)$, the common component estimation using $(p,q) = (300,60)$ is worse than that using $(p,q) = (60,300)$ due to insufficient column dimension relative to $k_c$. Hence it leads to worse estimators for errors and $(\widehat{\gamma}_{\beta,1}^{-1}, \widehat{\gamma}_{\beta,2}^{-1}, \widehat{\gamma}_{\beta,3}^{-1})$ under $(p,q) = (300,60)$. Hence inference performances on the time-varying row and column effect estimators are affected by the latent number of factors.

### Testing MEFM versus FM

We now demonstrate numerical results for Corollary 5.8. Consider the two scenarios:

1. *(Global effect.)* The entries of at least one of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are in general nonzero for each $t$.

| | Size | Setting (IVa) | | | Setting (IVb) | | | Setting (IVc) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | 0 | 0.1 | 0.5 | 1 | 0.1 | 0.5 | 1 | 2 | 5 | 10 |
| **reject**$_\alpha$ | $5_{(4)}$ | $11_{(7)}$ | $63_{(31)}$ | $96_{(15)}$ | $13_{(8)}$ | $53_{(30)}$ | $86_{(23)}$ | $37_{(17)}$ | $77_{(24)}$ | $85_{(27)}$ |
| **reject**$_\beta$ | $5_{(4)}$ | $11_{(7)}$ | $52_{(28)}$ | $87_{(22)}$ | $13_{(8)}$ | $62_{(32)}$ | $96_{(16)}$ | $14_{(8)}$ | $28_{(16)}$ | $48_{(26)}$ |

Table 5.2: Results for Setting (IVa) to (IVc). Each cell reports the mean and SD (subscripted, in bracket), both multiplied by 100. The parameters for Settings (IVa), (IVb) and (IVc) are $u_\alpha$, $u_\beta$ and $u_{local}$, respectively. Setting (IVa) with $u_\alpha = 0$ is reported in the first column, representing the size of the test.

2. *(Local effect.)* The entries of at least one of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are sparse at each $t$, i.e., there are some nonzero entries in at least one of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ with all other entries zero.

Throughout this subsection, we generate the time-varying grand mean and main effects using Rademacher random variables such that $v_{\mu,t}$ is i.i.d. Rademacher multiplied by some $u_\mu$ and each entry of $\mathbf{v}_{\alpha,t}$, $\mathbf{v}_{\beta,t}$ is i.i.d. Rademacher multiplied by some $u_\alpha$, $u_\beta$ respectively, recalling that $\mu_t = v_{\mu,t}$, $\boldsymbol{\alpha}_t = \mathbf{M}_p\mathbf{v}_{\alpha,t}$ and $\boldsymbol{\beta}_t = \mathbf{M}_q\mathbf{v}_{\beta,t}$. Hence, setting $u_\mu = u_\alpha = u_\beta = 0$ corresponds to generating a traditional FM. We set $k_r = k_c = 2$, and consider the settings:

(IVa) $T = p = q = 40$. All factors are pervasive with $\zeta_{r,j} = \zeta_{c,j} = 0$. All innovation processes in constructing $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and $\boldsymbol{\epsilon}_t$ are i.i.d. standard normal. Set $u_\mu = u_\beta = 0$, and we select $u_\alpha$ from $0.1,\ 0.5,\ 1$.

(IVb) Same as (IVa), but fix $u_\alpha = 0.1$ and select $u_\beta$ from $0.1,\ 0.5,\ 1$.

(IVc) Same as (IVa), except that $u_\alpha = 1$, and when generating $\boldsymbol{\alpha}_t = \mathbf{M}_p\mathbf{v}_{\alpha,t}$ as specified previously, we only keep the first $u_{local}$ entries of $\mathbf{v}_{\alpha,t}$ as nonzero where $u_{local}$ is selected from $2,\ 5,\ 10$.

Setting (IVa) and (IVb) are designed for testing global effects, and Setting (IVc) for local effects. For each setting, we construct $y_{\alpha,t}$, $y_{\beta,t}$ and use $\theta = 0.95$ in Corollary 5.8. Each experiment is repeated 400 times and we report both **reject**$_\alpha := T^{-1}\sum_{t=1}^T \mathbb{1}\{y_{\alpha,t} \geq \widehat{q}_{x,\alpha}(0.95)\}$ and **reject**$_\beta := T^{-1}\sum_{t=1}^T \mathbb{1}\{y_{\beta,t} \geq \widehat{q}_{x,\beta}(0.95)\}$.

As explained under Corollary 5.8, we expect **reject**$_\alpha$ and **reject**$_\beta$ to be close to $1 - \theta = 0.05$ if FM is sufficient. From Table 5.2, our proposed test works well since it suggests FM is insufficient as we strengthen $\boldsymbol{\alpha}_t$ or $\boldsymbol{\beta}_t$. In particular, even if the signal of $\boldsymbol{\alpha}_t$ is not strong enough such as $u_\alpha = 0.1$, Setting (IVb) shows that additional signals from $\boldsymbol{\beta}_t$ allows us to reject the use of FM. The comparison between **reject**$_\alpha$ and **reject**$_\beta$ is indicative of which effect is stronger. According to the results for (IVc) in the table, our test is capable of detecting local effect such that **reject**$_\alpha$ is far from $0.05$ even when only two entries in $\boldsymbol{\alpha}_t$ are nonzero.

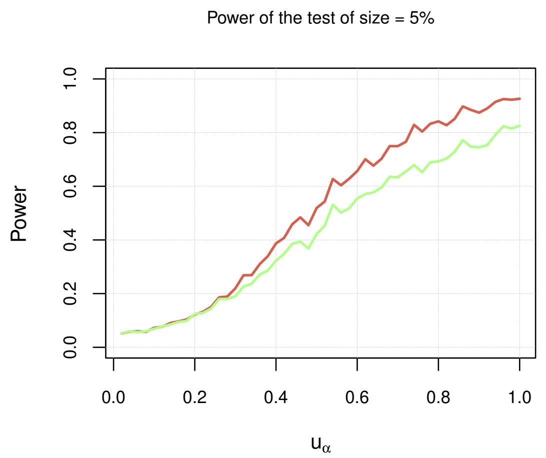


Figure 5.8: Statistical power curve of testing the null hypothesis that FM is sufficient for the given series, against the alternative that MEFM is necessary. Each power value is computed as the average over 400 runs of $\mathbf{reject}_\alpha$ (in red) and $\mathbf{reject}_\beta$ (in green) under Setting (IVa) except that $(T,\ p,\ q) = (60,\ 80,\ 80)$.

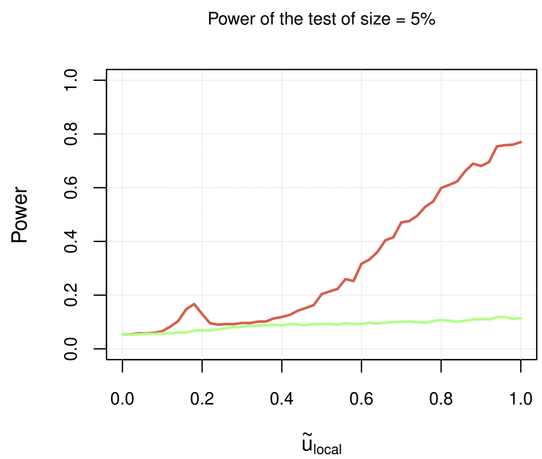


Figure 5.9: Statistical power curve of testing. Refer to the caption of Figure 5.8 for details on the hypothesis test and power computation. The data is generated under Setting (IVc) except that $(T, p, q) = (60, 80, 80)$ and $\boldsymbol{\alpha}_t$ is generated such that $\boldsymbol{\alpha}_1 = \widetilde{u}_{local}(1, 1, \text{-}2, 0, \ldots, 0)'$, $\boldsymbol{\alpha}_2 = \widetilde{u}_{local}(1, 2, \text{-}3, 0, \ldots, 0)'$, and $\boldsymbol{\alpha}_3 = \widetilde{u}_{local}(2, \text{-}5, 3, 0, \ldots, 0)'$, followed by $\boldsymbol{\alpha}_{3\ell+i} = \boldsymbol{\alpha}_i$ for positive integer $\ell$ and $i = 1, 2, 3$, so that each $\boldsymbol{\alpha}_t$ has nonzero entries only in the first three indices.

Extensive experiments on different dimensions, factor strengths or grand mean magnitudes are performed. All indicate similar interpretation as the above settings and hence the results are not shown here. The power curve for Setting (IVa) is also presented in Figure 5.8 to support the use of our test, with $(T,\ p,\ q) = (60,\ 80,\ 80)$ and $u_\alpha$ ranging from $0.02$ to $1$. Besides, we also show the power curve for local effect in Figure 5.9, for Setting (IVc) except that $(T,\ p,\ q) = (60,\ 80,\ 80)$ and we generate $\boldsymbol{\alpha}_t$ as described in the caption. Both power curves show that the test is able to reject the use of FM if signals from the time-varying main effects are large, either globally or locally. In both figures, when $u_\alpha$ is close to $0.02$ or $\widetilde{u}_{local}$ close to $0$, the value of the power curves are all very close to $0.05$, which is exactly what we want for the size of the tests.

### 5.4.2 Real data analysis: NYC taxi traffic

We analyse a set of taxi traffic data in New York City in this example. The data includes all individual taxi rides operated by Yellow Taxi in New York City, published at https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

For simplicity, we only consider the rides within Manhattan Island, which comprises most of the data. The dataset contains 842 million trip records from January 1, 2013 to December 31, 2022. Each trip record includes features such as pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. Our example here focuses on the drop-off dates/times and locations.

To classify the drop-off locations in Manhattan, they are coded according to 69 predefined zones in the dataset. Moreover, each day is divided into 24 hourly periods to represent the drop-off times each day, with the first hourly period from 0 a.m. to 1 a.m. The total number of rides moving among the zones within each hour are recorded, yielding data $\mathbf{Y}_t \in \mathbb{R}^{69\times 24}$ each day, where $y_{i_1,i_2,t}$ is the number of trips to zone $i_1$ and the pick-up time is within the $i_2$-th hourly period on day $t$.

We consider the non-business-day series which is 1,133 days long, within the period of January 1, 2013 to December 31, 2022. Using MEFM, the estimated rank of the core factors is $(2, 2)$ according to our proposed eigenvalue ratio estimator. As mentioned in Section 5.3.6, we therefore use $(3, 3)$ as the number of factors to estimate FM and test if FM is sufficient. We compute $\textbf{reject}_\alpha = 0.064$ and $\textbf{reject}_\beta = 0.133$ which are defined in Section 5.4.1. They should be close to $1 - \theta = 0.05$ according to Corollary 5.8 if FM is sufficient. Hence we reject the use of traditional FM due to the signals in $\widehat{\boldsymbol{\beta}}_t$.

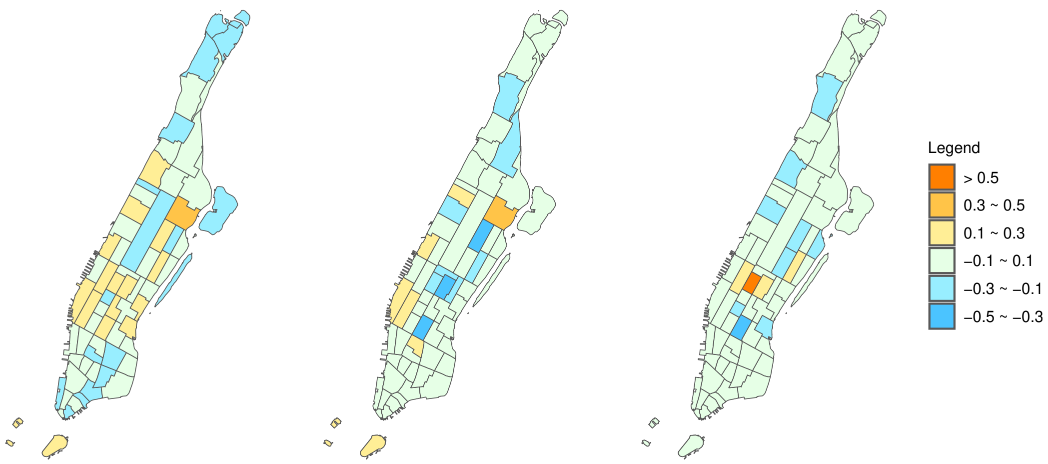


Figure 5.10: Estimated loading on three dropoff factors using MEFM, i.e., $\widehat{\mathbf{Q}}_{1,\cdot 1}$ (left), $\widehat{\mathbf{Q}}_{1,\cdot 2}$ (middle) and $\widehat{\mathbf{Q}}_{1,\cdot 3}$ (right).

To compare MEFM with FM, we use core rank $(3, 3)$ to estimate MEFM for the rest of this example. Figure 5.10 and 5.11 illustrate the heatmaps of the estimated loading columns on the three dropoff factors using MEFM and FM, respectively. From both heatmaps, we can

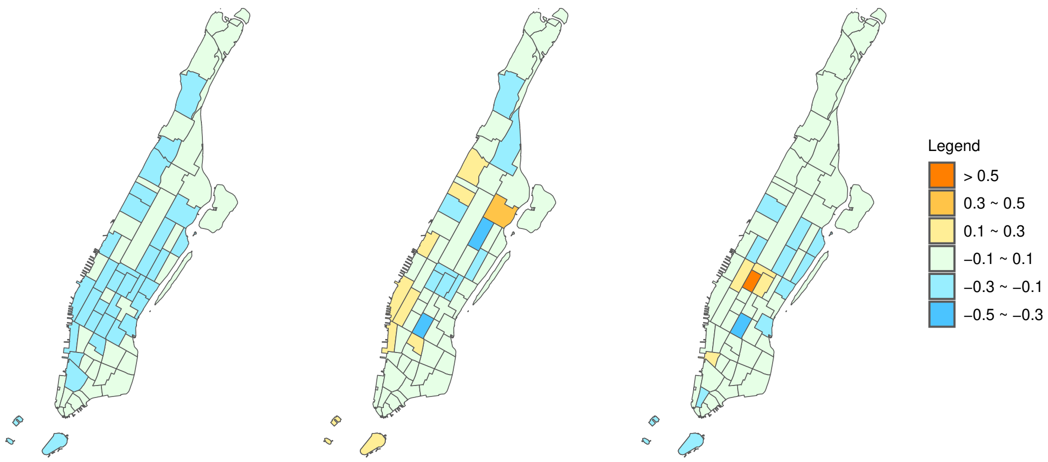


Figure 5.11: Estimated loading on three dropoff factors using FM, similar to Figure 5.10.

identify the first factor as active areas, the second as dining and sports areas and the third as downtown areas. The three factors are similar to their corresponding counterparts, except that the first factor estimated using MEFM is more indicative on the active areas to taxi traffic in Manhattan by its emphasised orange zone which corresponds to East Harlem.

To gain further understanding on the taxi traffic, we show the scaled $\widehat{\mathbf{Q}}_2$ by MEFM and FM in Tables 5.3 and 5.4, respectively. We can see that for the rush hours between 6 p.m. to 11 p.m., the estimated loadings almost vanish for MEFM, which is consistent with the fact that $\widehat{\boldsymbol{\beta}}_t$ captures the common hour effect on Manhattan life style. This also provides an intuition why the time-varying column/hour effect is strong, since in non-business days, the way that daily hours affecting the taxi traffic can change drastically over time as compared to the same when Manhattan zones are considered. For demonstration, we plot both $\widehat{\beta}_{t,2}$ and $\widehat{\beta}_{t,18}$ in Figure 5.12, where the former series features the mid-night effects and the latter features the night-life effects. Both series demonstrate seasonality before COVID-19 as shown on the plot.

The business-day series is also analysed, but since both $\textbf{reject}_\alpha$ and $\textbf{reject}_\beta$ are not significant, the estimated model is not shown here. The fact that the time-varying hour effect is not strong for business days is probably due to a rather routine working hours. Thus the hour effect is hardly changing and can be absorbed into a fixed mean, so that FM would be sufficient.

## 5.5 Proof of Theorems and Auxiliary Results

***Proof of Theorem 5.1.*** Suppose we have another set of parameters, $(\widetilde{\mu}_t, \widetilde{\boldsymbol{\alpha}}_t, \widetilde{\boldsymbol{\beta}}_t, \widetilde{\mathbf{Q}}_r, \widetilde{\mathbf{Q}}_c, \widetilde{\mathbf{F}}_{Z,t})$ for $t \in [T]$, also satisfying (5.3). For each $t \in [T]$, left-multiplying by $\mathbf{1}_p'$ and right-multiplying

| $0_{am}$ | | 2 | | 4 | | 6 | | 8 | | 10 | | $12_{pm}$ | | 2 | | 4 | | 6 | | 8 | | 10 | | $12_{am}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -2 | -5 | -6 | -7 | -7 | -7 | -6 | -5 | -3 | 0 | 3 | 5 | 6 | 6 | 5 | 5 | 4 | 4 | 5 | 5 | 2 | 0 | 0 | -1 |
| 2 | 6 | 5 | 3 | 1 | -1 | -4 | -5 | -6 | -7 | -7 | -6 | -5 | -3 | -2 | -1 | -2 | -1 | -1 | 2 | 5 | 8 | 6 | 6 | 7 |
| 3 | -1 | -13 | -9 | -6 | -2 | 2 | 4 | 5 | 6 | 4 | 2 | -2 | -4 | -5 | -4 | -3 | -2 | -1 | 0 | 2 | 5 | 4 | 7 | 9 |

Table 5.3: Estimated loading matrix $\widehat{\mathbf{Q}}_2$ using MEFM, after scaling. Magnitudes larger than 6 are highlighted in red.

| $0_{am}$ | | 2 | | 4 | | 6 | | 8 | | 10 | | $12_{pm}$ | | 2 | | 4 | | 6 | | 8 | | 10 | | $12_{am}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -5 | -5 | -4 | -3 | -2 | -1 | -1 | -1 | -2 | -3 | -4 | -5 | -5 | -6 | -5 | -5 | -5 | -5 | -6 | -6 | -6 | -5 | -5 | -5 |
| 2 | 5 | 7 | 7 | 5 | 4 | 2 | 0 | -2 | -4 | -6 | -6 | -6 | -6 | -5 | -4 | -4 | -3 | -3 | -2 | 1 | 4 | 4 | 5 | 6 |
| 3 | 1 | -13 | -10 | -9 | -6 | -3 | -1 | 0 | 1 | 0 | -1 | -3 | -3 | -3 | -3 | -2 | -1 | 0 | 3 | 6 | 6 | 6 | 8 | 11 |

Table 5.4: Estimated loading matrix $\widehat{\mathbf{Q}}_2$ using FM, after scaling. Magnitudes larger than 5 are highlighted in red.

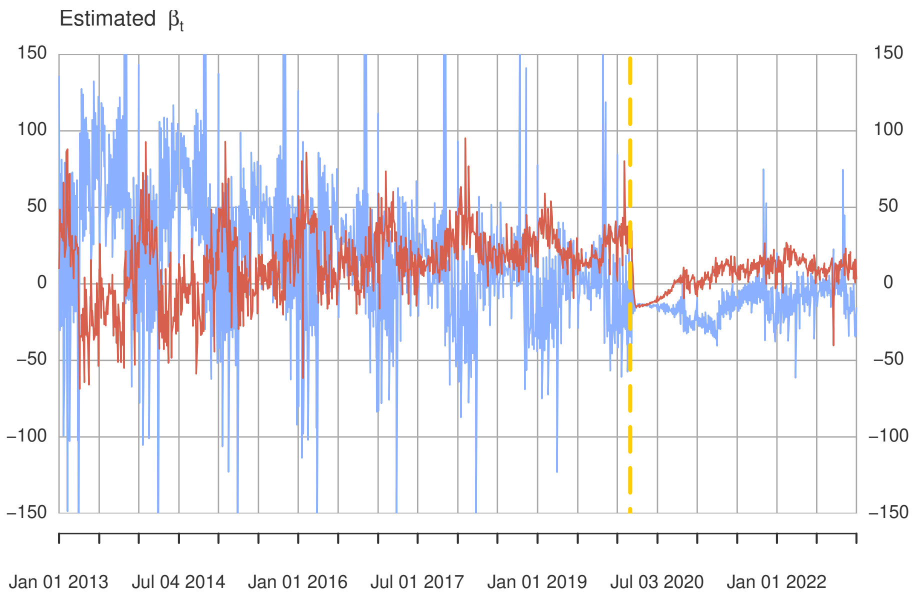


Figure 5.12: Plot of the estimated hour effects for periods from 1 a.m. to 2 a.m. (in blue) and from 5 p.m. to 6.p.m. (in red). The date for the first confirmed case of COVID-19 in New York is also shown (dotted yellow vertical line).

by $\mathbf{1}_q$ on (5.3), we arrive at $pq\widetilde{\mu}_t = pq\mu_t$ from (IC1), so $\mu_t$ is identified. Similarly, $\widetilde{\boldsymbol{\alpha}}_t = \boldsymbol{\alpha}_t$ and $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_t$, by separately left-multiplying $\mathbf{1}_p'$ and right-multiplying $\mathbf{1}_q$ on $\mathbf{Y}_t$.

We hence have $\widetilde{\mathbf{Q}}_r\widetilde{\mathbf{F}}_{Z,t}\widetilde{\mathbf{Q}}_c' = \mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c'$, but the factor loading matrices and factor series require further identification due to the multiplicative form. Without loss of generality, write

$$\widetilde{\mathbf{Q}}_r = \mathbf{Q}_r\mathbf{M}_r + \boldsymbol{\Gamma}_r, \quad \text{where } \boldsymbol{\Gamma}_r'\mathbf{Q}_r = \mathbf{0},$$

with $\mathbf{M}_r \in \mathbb{R}^{k_r\times k_r}$ and $\boldsymbol{\Gamma}_r \in \mathbb{R}^{p\times k_r}$, but can have zero columns. Then we have

$$\mathbf{0} = \boldsymbol{\Gamma}_r'\mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c'\widetilde{\mathbf{Q}}_c = \boldsymbol{\Gamma}_r'\widetilde{\mathbf{Q}}_r\widetilde{\mathbf{F}}_{Z,t}\widetilde{\mathbf{Q}}_c'\widetilde{\mathbf{Q}}_c = \boldsymbol{\Gamma}_r'\boldsymbol{\Gamma}_r\widetilde{\mathbf{F}}_{Z,t}\widetilde{\mathbf{Q}}_c'\widetilde{\mathbf{Q}}_c,$$

which can only be true in general if $\boldsymbol{\Gamma}_r = \mathbf{0}$ since $\widetilde{\mathbf{F}}_t$ is random and $\widetilde{\mathbf{Q}}_c'\widetilde{\mathbf{Q}}_c \to \boldsymbol{\Sigma}_{A,c}$ due to (L1). Using (L1), $\mathbf{M}_r$ is of full rank and hence $\widetilde{\mathbf{Q}}_r$ and $\mathbf{Q}_r$ share the same column space. Similarly, the factor loading space of $\mathbf{Q}_c$ is identified, and $\mathbf{F}_{Z,t}$ is hence identified once $\mathbf{Q}_r$ and $\mathbf{Q}_c$ are given correspondingly. □

***Proof of Theorem 5.2.*** By Assumption (IC1), we have $\mu_t = \mathbf{1}_p'(\mathbf{Y}_t - \mathbf{E}_t)\mathbf{1}_q/pq$ and hence

$$(\widehat{\mu}_t - \mu_t)^2 = \frac{1}{p^2q^2}\Big(\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\Big)^2 = \frac{1}{p^2q^2}\Big(\sum_{i=1}^p\sum_{j=1}^q E_{t,ij}\Big)^2. \tag{5.9}$$

Assumption (E1) implies each entry of $\mathbf{E}_t$ has zero mean and bounded fourth moment, and

$$\mathbb{E}\Big\{\Big(\sum_{i=1}^p\sum_{j=1}^q E_{t,ij}\Big)^2\Big\} = \text{Var}\Big(\sum_{i=1}^p\sum_{j=1}^q E_{t,ij}\Big) = \sum_{i=1}^p\sum_{l=1}^p\sum_{j=1}^q\sum_{h=1}^q \text{Cov}(E_{t,ij}, E_{t,lh}) = O(pq), \tag{5.10}$$

where we used Lemma 5.1 in the last equality. Thus with (5.9), $(\widehat{\mu}_t - \mu_t)^2 = O_P(p^{-1}q^{-1})$.

Similar to the rate for $\widehat{\mu}_t$, by again (IC1) we have $\boldsymbol{\alpha}_t = q^{-1}\mathbf{Y}_t\mathbf{1}_q - \mu_t\mathbf{1}_p - q^{-1}\mathbf{E}_t\mathbf{1}_q$ and $\boldsymbol{\beta}_t = p^{-1}\mathbf{Y}_t'\mathbf{1}_p - \mu_t\mathbf{1}_q - p^{-1}\mathbf{E}_t'\mathbf{1}_p$. Then we have

$$\frac{1}{p}\cdot\|\widehat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t\|^2 = \frac{1}{p}\cdot\Big\|(\mu_t - \widehat{\mu}_t)\mathbf{1}_p + q^{-1}\mathbf{E}_t\mathbf{1}_q\Big\|^2, \tag{5.11}$$

$$\frac{1}{q}\cdot\|\widehat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t\|^2 = \frac{1}{q}\cdot\Big\|(\mu_t - \widehat{\mu}_t)\mathbf{1}_q + p^{-1}\mathbf{E}_t'\mathbf{1}_p\Big\|^2. \tag{5.12}$$

From (5.9) we have $\|(\mu_t - \widehat{\mu}_t)\mathbf{1}_p\|^2 = O_P(q^{-1})$ and $\|(\mu_t - \widehat{\mu}_t)\mathbf{1}_q\|^2 = O_P(p^{-1})$. Furthermore, by Lemma 5.1,

$$\mathbb{E}\Big(\|q^{-1}\mathbf{E}_t\mathbf{1}_q\|^2\Big) = q^{-2}\cdot\sum_{i=1}^p\text{Var}\Big(\sum_{j=1}^q E_{t,ij}\Big) = O(pq^{-1}).$$

Similarly, $\|p^{-1}\mathbf{E}_t'\mathbf{1}_p\|^2 = O_P(qp^{-1})$. Then we have (5.11) and (5.12) as

$$\frac{1}{p}\cdot\|\widehat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t\|^2 = O_P(q^{-1}), \quad \frac{1}{q}\cdot\|\widehat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t\|^2 = O_P(p^{-1}).$$

In the rest of the proof, we show consistency of the factor loading estimators. From (5.4),

$$\begin{aligned}
\widehat{\mathbf{L}}_t &= \mathbf{Y}_t + (pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' - q^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t,\\
\widehat{\mathbf{L}}_t' &= \mathbf{Y}_t' + (pq)^{-1}\mathbf{1}_q\mathbf{1}_p'\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p - q^{-1}\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t' - p^{-1}\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p',
\end{aligned}$$

and hence the following decomposition

$$\begin{aligned}
\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t' = \;& \mathbf{Y}_t\mathbf{Y}_t' + (pq)^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p'\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p - q^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t' - p^{-1}\mathbf{Y}_t\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p'\\
&+ (pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\mathbf{Y}_t' + (pq)^{-2}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_p'\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p\\
&- q^{-1}(pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t' - p^{-1}(pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p'\\
&- q^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t' - q^{-1}(pq)^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_p'\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p + q^{-2}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t'\\
&+ q^{-1}p^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t\mathbf{Y}_t' - (pq)^{-1}p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p'\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p\\
&+ q^{-1}p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t' + p^{-2}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p'\\
=\;& \mathbf{Y}_t\mathbf{Y}_t' + (pq)^{-1}\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p' - p^{-1}\mathbf{Y}_t\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p' - q^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t\mathbf{Y}_t'\\
&- (pq)^{-1}p^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + (pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\mathbf{Y}_t' + p^{-2}\mathbf{1}_p'\mathbf{Y}_t\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p'\\
=\;& \mathbf{Y}_t\mathbf{Y}_t' + \mathcal{Q}_1 - \mathcal{Q}_2 - \mathcal{Q}_3 - \mathcal{Q}_4 - \mathcal{Q}_5 + \mathcal{Q}_6 + \mathcal{Q}_7, \quad \text{where}\\
\mathcal{Q}_1 :=\;& (pq)^{-1}\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p', \quad \mathcal{Q}_2 := p^{-1}\mathbf{Y}_t\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p', \quad \mathcal{Q}_3 := q^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t',\\
\mathcal{Q}_4 :=\;& p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t\mathbf{Y}_t', \quad \mathcal{Q}_5 := (pq)^{-1}p^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p',\\
\mathcal{Q}_6 :=\;& (pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\mathbf{Y}_t', \quad \mathcal{Q}_7 := p^{-2}\mathbf{1}_p'\mathbf{Y}_t\mathbf{Y}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p'.
\end{aligned}$$

From (5.3), $\mathbf{1}_q'\mathbf{Y}_t'\mathbf{1}_p = pq\mu_t + \mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p$ and $\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p' = q\mu_t\mathbf{1}_p\mathbf{1}_p' + q\boldsymbol{\alpha}_t\mathbf{1}_p' + \mathbf{E}_t\mathbf{1}_q\mathbf{1}_p'$. Thus,

$$\begin{aligned}
\mathcal{Q}_1 &= (pq)^{-1}(pq^2\mu_t^2\mathbf{1}_p\mathbf{1}_p' + pq^2\mu_t\boldsymbol{\alpha}_t\mathbf{1}_p' + pq\mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p'\\
&\quad + q\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + q\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\boldsymbol{\alpha}_t\mathbf{1}_p' + \mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p')\\
&= q\mu_t^2\mathbf{1}_p\mathbf{1}_p' + q\mu_t\boldsymbol{\alpha}_t\mathbf{1}_p' + \mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + p^{-1}\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p'\\
&\quad + p^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\boldsymbol{\alpha}_t\mathbf{1}_p' + (pq)^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p'.
\end{aligned} \tag{5.13}$$

Similarly, we have $\mathbf{Y}_t\mathbf{Y}_t' = (\mu_t\mathbf{1}_p\mathbf{1}_q' + \boldsymbol{\alpha}_t\mathbf{1}_q' + \mathbf{1}_p\boldsymbol{\beta}_t' + \mathbf{C}_t + \mathbf{E}_t)(\mu_t\mathbf{1}_q\mathbf{1}_p' + \mathbf{1}_q\boldsymbol{\alpha}_t' + \boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{C}_t' + \mathbf{E}_t')$.

Further with Assumption (IC1),

$$\begin{aligned}\mathcal{Q}_2 = p^{-1}(&\mu_t^2\mathbf{1}_p\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_p' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{1}_q\boldsymbol{\alpha}_t' + \mu_t\mathbf{1}_p\mathbf{1}_q'\boldsymbol{\beta}_t\mathbf{1}_p' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{C}_t' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' \\ &+ \mu_t\boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_p' + \boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{1}_q\boldsymbol{\alpha}_t' + \boldsymbol{\alpha}_t\mathbf{1}_q'\boldsymbol{\beta}_t\mathbf{1}_p' + \boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{C}_t' + \boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{E}_t' \\ &+ \mu_t\mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{1}_q\mathbf{1}_p' + \mathbf{1}_p\boldsymbol{\beta}'\mathbf{1}_q\boldsymbol{\alpha}_t' + \mathbf{1}_p\boldsymbol{\beta}_t'\boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{C}_t' + \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{E}_t' \\ &+ \mu_t\mathbf{C}_t\mathbf{1}_q\mathbf{1}_p' + \mathbf{C}_t\mathbf{1}_q\boldsymbol{\alpha}_t' + \mathbf{C}_t\boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{C}_t\mathbf{C}_t' + \mathbf{C}_t\mathbf{E}_t' \\ &+ \mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + \mathbf{E}_t\mathbf{1}_q\boldsymbol{\alpha}_t' + \mathbf{E}_t\boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{E}_t\mathbf{C}_t' + \mathbf{E}_t\mathbf{E}_t')\mathbf{1}_p\mathbf{1}_p' \\ = q&\mu_t^2\mathbf{1}_p\mathbf{1}_p' + p^{-1}\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + q\mu_t\boldsymbol{\alpha}_t\mathbf{1}_p' + p^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\boldsymbol{\alpha}_t\mathbf{1}_p' + \mathbf{1}_p\boldsymbol{\beta}_t'\boldsymbol{\beta}_t\mathbf{1}_p' \\ &+ p^{-1}\boldsymbol{\beta}_t'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + \mathbf{C}_t\boldsymbol{\beta}_t\mathbf{1}_p' + p^{-1}\mathbf{C}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p' + \mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + \mathbf{E}_t\boldsymbol{\beta}_t\mathbf{1}_p' + p^{-1}\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p'.\end{aligned} \tag{5.14}$$

Since $\mathbf{Y}_t\mathbf{1}_q = q\mu_t\mathbf{1}_p + q\boldsymbol{\alpha}_t + \mathbf{E}_t\mathbf{1}_q$, we have

$$\begin{aligned}\mathcal{Q}_3 &= q^{-1}(q\mu_t\mathbf{1}_p + q\boldsymbol{\alpha}_t + \mathbf{E}_t\mathbf{1}_q)(q\mu_t\mathbf{1}_p' + q\boldsymbol{\alpha}_t' + \mathbf{1}_q'\mathbf{E}_t') \\ &= q\mu_t^2\mathbf{1}_p\mathbf{1}_p' + q\mu_t\mathbf{1}_p\boldsymbol{\alpha}_t' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' + q\mu_t\boldsymbol{\alpha}_t\mathbf{1}_p' + q\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t' + \boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{E}_t' \\ &\quad + \mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + \mathbf{E}_t\mathbf{1}_q\boldsymbol{\alpha}_t' + q^{-1}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'.\end{aligned} \tag{5.15}$$

Similar to (5.14), we have

$$\begin{aligned}\mathcal{Q}_4 = q&\mu_t^2\mathbf{1}_p\mathbf{1}_p' + q\mu_t\mathbf{1}_p\boldsymbol{\alpha}_t' + \mu_t\mathbf{1}_p\mathbf{1}_q'\boldsymbol{\beta}_t\mathbf{1}_p' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{C}_t' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' + \mu_t\mathbf{1}_p\mathbf{1}_p'\boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_p' \\ &+ p^{-1}\mathbf{1}_p\mathbf{1}_p'\boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{1}_q\boldsymbol{\alpha}_t' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\boldsymbol{\alpha}_t\mathbf{1}_q'\boldsymbol{\beta}_t\mathbf{1}_p' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{C}_t' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{E}_t' \\ &+ \mu_t\mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{1}_q\mathbf{1}_p' + \mathbf{1}_p\boldsymbol{\beta}'\mathbf{1}_q\boldsymbol{\alpha}_t' + \mathbf{1}_p\boldsymbol{\beta}_t'\boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{C}_t' + \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{E}_t' + p^{-1}\mu_t\mathbf{1}_p\mathbf{1}_p'\mathbf{C}_t\mathbf{1}_q\mathbf{1}_p' \\ &+ p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{C}_t\mathbf{1}_q\boldsymbol{\alpha}_t' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{C}_t\boldsymbol{\beta}_t\mathbf{1}_p' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{C}_t\mathbf{C}_t' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{C}_t\mathbf{E}_t' \\ &+ p^{-1}\mu_t\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\boldsymbol{\alpha}_t' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\boldsymbol{\beta}_t\mathbf{1}_p' \\ &+ p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{C}_t' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t' \\ = q&\mu_t^2\mathbf{1}_p\mathbf{1}_p' + q\mu_t\mathbf{1}_p\boldsymbol{\alpha}_t' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' + \boldsymbol{\beta}_t'\boldsymbol{\beta}_t\mathbf{1}_p\mathbf{1}_p' + \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{C}_t' + \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{E}_t' \\ &+ p^{-1}\mu_t\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_p' + p^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\boldsymbol{\alpha}_t' + p^{-1}\mathbf{1}_p'\mathbf{E}_t\boldsymbol{\beta}_t\mathbf{1}_p\mathbf{1}_p' \\ &+ p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{C}_t' + p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t',\end{aligned} \tag{5.16}$$

where the last equality used Assumption (IC1). For $\mathcal{Q}_5$ and $\mathcal{Q}_6$, we have

$$\begin{aligned}\mathcal{Q}_5 &= (pq)^{-1}p^{-1}(pq\mu_t + \mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q)(pq\mu_t + \mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)\mathbf{1}_p\mathbf{1}_p' \\ &= q\mu_t^2\mathbf{1}_p\mathbf{1}_p' + p^{-1}\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + p^{-1}\mu_t\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_p' + (pq)^{-1}p^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p',\end{aligned} \tag{5.17}$$

$$\mathcal{Q}_6 = (\mu_t + (pq)^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)(q\mu_t\mathbf{1}_p\mathbf{1}_p' + q\mathbf{1}_p\boldsymbol{\alpha}_t' + \mathbf{1}_p\mathbf{1}_q'\boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{1}_p\mathbf{1}_q'\mathbf{C}_t' + \mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t')$$

$$
\begin{aligned}
&= q\mu_t^2 \mathbf{1}_p \mathbf{1}_p' + q\mu_t \mathbf{1}_p \boldsymbol{\alpha}_t' + \mu_t \mathbf{1}_p \mathbf{1}_q' \mathbf{E}_t' + p^{-1}\mu_t \mathbf{1}_q' \mathbf{E}_t' \mathbf{1}_p \mathbf{1}_p \mathbf{1}_p' \\
&\quad + p^{-1}\mathbf{1}_q' \mathbf{E}_t' \mathbf{1}_p \mathbf{1}_p \boldsymbol{\alpha}_t' + (pq)^{-1} \mathbf{1}_q' \mathbf{E}_t' \mathbf{1}_p \mathbf{1}_p \mathbf{1}_q' \mathbf{E}_t'. \qquad (5.18)
\end{aligned}
$$

Lastly for $\mathcal{Q}_7$, we have similar to (5.14) that

$$
\begin{aligned}
\mathcal{Q}_7 &= p^{-2}\mathbf{1}_p'(\mu_t^2 \mathbf{1}_p \mathbf{1}_q' \mathbf{1}_q \mathbf{1}_p' + \mu_t \mathbf{1}_p \mathbf{1}_q' \mathbf{1}_q \boldsymbol{\alpha}_t' + \mu_t \mathbf{1}_p \mathbf{1}_q' \boldsymbol{\beta}_t \mathbf{1}_p' + \mu_t \mathbf{1}_p \mathbf{1}_q' \mathbf{C}_t' + \mu_t \mathbf{1}_p \mathbf{1}_q' \mathbf{E}_t' \\
&\quad + \mu_t \boldsymbol{\alpha}_t \mathbf{1}_q' \mathbf{1}_q \mathbf{1}_p' + \boldsymbol{\alpha}_t \mathbf{1}_q' \mathbf{1}_q \boldsymbol{\alpha}_t' + \boldsymbol{\alpha}_t \mathbf{1}_q' \boldsymbol{\beta}_t \mathbf{1}_p' + \boldsymbol{\alpha}_t \mathbf{1}_q' \mathbf{C}_t' + \boldsymbol{\alpha}_t \mathbf{1}_q' \mathbf{E}_t' \\
&\quad + \mu_t \mathbf{1}_p \boldsymbol{\beta}_t' \mathbf{1}_q \mathbf{1}_p' + \mathbf{1}_p \boldsymbol{\beta}' \mathbf{1}_q \boldsymbol{\alpha}_t' + \mathbf{1}_p \boldsymbol{\beta}_t' \boldsymbol{\beta}_t \mathbf{1}_p' + \mathbf{1}_p \boldsymbol{\beta}_t' \mathbf{C}_t' + \mathbf{1}_p \boldsymbol{\beta}_t' \mathbf{E}_t' \\
&\quad + \mu_t \mathbf{C}_t \mathbf{1}_q \mathbf{1}_p' + \mathbf{C}_t \mathbf{1}_q \boldsymbol{\alpha}_t' + \mathbf{C}_t \boldsymbol{\beta}_t \mathbf{1}_p' + \mathbf{C}_t \mathbf{C}_t' + \mathbf{C}_t \mathbf{E}_t' \\
&\quad + \mu_t \mathbf{E}_t \mathbf{1}_q \mathbf{1}_p' + \mathbf{E}_t \mathbf{1}_q \boldsymbol{\alpha}_t' + \mathbf{E}_t \boldsymbol{\beta}_t \mathbf{1}_p' + \mathbf{E}_t \mathbf{C}_t' + \mathbf{E}_t \mathbf{E}_t') \mathbf{1}_p \mathbf{1}_p \mathbf{1}_p' \qquad (5.19) \\
&= p^{-2}(pq\mu_t^2 \mathbf{1}_p' + pq\mu_t \boldsymbol{\alpha}_t' + p\mu_t \mathbf{1}_q' \mathbf{E}_t' + p\boldsymbol{\beta}_t' \boldsymbol{\beta}_t \mathbf{1}_p' + p\boldsymbol{\beta}_t' \mathbf{C}_t' + p\boldsymbol{\beta}_t' \mathbf{E}_t' \\
&\quad + \mu_t \mathbf{1}_p' \mathbf{E}_t \mathbf{1}_q \mathbf{1}_p' + \mathbf{1}_p' \mathbf{E}_t \mathbf{1}_q \boldsymbol{\alpha}_t' + \mathbf{1}_p' \mathbf{E}_t \boldsymbol{\beta}_t \mathbf{1}_p' + \mathbf{1}_p' \mathbf{E}_t \mathbf{C}_t' + \mathbf{1}_p' \mathbf{E}_t \mathbf{E}_t') \mathbf{1}_p \mathbf{1}_p \mathbf{1}_p' \\
&= q\mu_t^2 \mathbf{1}_p \mathbf{1}_p' + p^{-1}\mu_t \mathbf{1}_q' \mathbf{E}_t' \mathbf{1}_p \mathbf{1}_p \mathbf{1}_p' + \boldsymbol{\beta}_t' \boldsymbol{\beta}_t \mathbf{1}_p \mathbf{1}_p' + p^{-1} \boldsymbol{\beta}_t' \mathbf{E}_t' \mathbf{1}_p \mathbf{1}_p \mathbf{1}_p' \\
&\quad + p^{-1}\mu_t \mathbf{1}_p' \mathbf{E}_t \mathbf{1}_q \mathbf{1}_p \mathbf{1}_p' + p^{-1} \mathbf{1}_p' \mathbf{E}_t \boldsymbol{\beta}_t \mathbf{1}_p \mathbf{1}_p' + p^{-2} \mathbf{1}_p' \mathbf{E}_t \mathbf{E}_t' \mathbf{1}_p \mathbf{1}_p \mathbf{1}_p'.
\end{aligned}
$$

With (5.13), (5.14), (5.15), (5.16), (5.17), (5.18) and (5.19), we have

$$
\begin{aligned}
&\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t' \\
&= \mathbf{Y}_t\mathbf{Y}_t' + \mathcal{Q}_1 - \mathcal{Q}_2 - \mathcal{Q}_3 - \mathcal{Q}_4 - \mathcal{Q}_5 + \mathcal{Q}_6 + \mathcal{Q}_7 \\
&= q\mu_t^2\mathbf{1}_p\mathbf{1}_p' + q\mu_t\mathbf{1}_p\boldsymbol{\alpha}_t' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' + q\mu_t\boldsymbol{\alpha}_t\mathbf{1}_p' + q\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t' + \boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{E}_t' + \mathbf{1}_p\boldsymbol{\beta}_t'\boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{C}_t' \\
&\quad + \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{E}_t' + \mathbf{C}_t\boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{C}_t\mathbf{C}_t' + \mathbf{C}_t\mathbf{E}_t' + \mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + \mathbf{E}_t\mathbf{1}_q\boldsymbol{\alpha}_t' + \mathbf{E}_t\boldsymbol{\beta}_t\mathbf{1}_p' + \mathbf{E}_t\mathbf{C}_t' + \mathbf{E}_t\mathbf{E}_t' \\
&\quad + q\mu_t^2\mathbf{1}_p\mathbf{1}_p' + q\mu_t\boldsymbol{\alpha}_t\mathbf{1}_p' + \mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + p^{-1}\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + p^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\boldsymbol{\alpha}_t\mathbf{1}_p' \\
&\quad + (pq)^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' - q\mu_t^2\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' - q\mu_t\boldsymbol{\alpha}_t\mathbf{1}_p' - p^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\boldsymbol{\alpha}_t\mathbf{1}_p' \\
&\quad - \mathbf{1}_p\boldsymbol{\beta}_t'\boldsymbol{\beta}_t\mathbf{1}_p' - p^{-1}\boldsymbol{\beta}_t'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' - \mathbf{C}_t\boldsymbol{\beta}_t\mathbf{1}_p' - p^{-1}\mathbf{C}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p' - \mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' - \mathbf{E}_t\boldsymbol{\beta}_t\mathbf{1}_p' \\
&\quad - p^{-1}\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p' - q\mu_t^2\mathbf{1}_p\mathbf{1}_p' - q\mu_t\mathbf{1}_p\boldsymbol{\alpha}_t' - \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' - q\mu_t\boldsymbol{\alpha}_t\mathbf{1}_p' - q\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t' - \boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{E}_t' \\
&\quad - \mu_t\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' - \mathbf{E}_t\mathbf{1}_q\boldsymbol{\alpha}_t' - q^{-1}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t' - q\mu_t^2\mathbf{1}_p\mathbf{1}_p' - q\mu_t\mathbf{1}_p\boldsymbol{\alpha}_t' - \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' - \boldsymbol{\beta}_t'\boldsymbol{\beta}_t\mathbf{1}_p\mathbf{1}_p' \\
&\quad - \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{C}_t' - \mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{E}_t' - p^{-1}\mu_t\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\boldsymbol{\alpha}_t' - p^{-1}\mathbf{1}_p'\mathbf{E}_t\boldsymbol{\beta}_t\mathbf{1}_p\mathbf{1}_p' \\
&\quad - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{C}_t' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t' - q\mu_t^2\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mu_t\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_p' \\
&\quad - (pq)^{-1}p^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + q\mu_t^2\mathbf{1}_p\mathbf{1}_p' + q\mu_t\mathbf{1}_p\boldsymbol{\alpha}_t' + \mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' + p^{-1}\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' \\
&\quad + p^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\boldsymbol{\alpha}_t' + (pq)^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' + q\mu_t^2\mathbf{1}_p\mathbf{1}_p' + p^{-1}\mu_t\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + \boldsymbol{\beta}_t'\boldsymbol{\beta}_t\mathbf{1}_p\mathbf{1}_p' \\
&\quad + p^{-1}\boldsymbol{\beta}_t'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' + p^{-1}\mu_t\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_p' + p^{-1}\mathbf{1}_p'\mathbf{E}_t\boldsymbol{\beta}_t\mathbf{1}_p\mathbf{1}_p' + p^{-2}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p' \\
&= \mathbf{C}_t\mathbf{C}_t' + \mathbf{C}_t\mathbf{E}_t' + \mathbf{E}_t\mathbf{C}_t' + \mathbf{E}_t\mathbf{E}_t' + (pq)^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + (pq)^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' \\
&\quad - p^{-1}\mathbf{C}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{C}_t' - p^{-1}\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t' - q^{-1}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t' \\
&\quad - (pq)^{-1}p^{-1}(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\mathbf{1}_p\mathbf{1}_p' + p^{-2}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p'.
\end{aligned} \tag{5.20}
$$

Swapping the roles of row and column factor loadings, we can arrive at similarly

$$
\begin{aligned}
\widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t = {}& \mathbf{C}_t'\mathbf{C}_t + \mathbf{C}_t'\mathbf{E}_t + \mathbf{E}_t'\mathbf{C}_t + \mathbf{E}_t'\mathbf{E}_t + (pq)^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_q' + (pq)^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q\mathbf{1}_p'\mathbf{E}_t \\
& - q^{-1}\mathbf{C}_t'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q' - q^{-1}\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\mathbf{C}_t - q^{-1}\mathbf{E}_t'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q' - q^{-1}\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\mathbf{E}_t - p^{-1}\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t \\
& - (pq)^{-1}q^{-1}(\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q)^2\mathbf{1}_q\mathbf{1}_q' + q^{-2}\mathbf{1}_q'\mathbf{E}_t'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q\mathbf{1}_q'.
\end{aligned} \tag{5.21}
$$

For ease of notation, we define

$$
\begin{aligned}
\mathbf{R}_{r,t} := {}& \mathbf{C}_t\mathbf{E}_t' + \mathbf{E}_t\mathbf{C}_t' + \mathbf{E}_t\mathbf{E}_t' + (pq)^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p' + (pq)^{-1}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t' \\
& - p^{-1}\mathbf{C}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{C}_t' - p^{-1}\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t' - q^{-1}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t' \\
& - (pq)^{-1}p^{-1}(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\mathbf{1}_p\mathbf{1}_p' + p^{-2}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p', \qquad (5.22) \\
\mathbf{R}_{c,t} := {}& \mathbf{C}_t'\mathbf{E}_t + \mathbf{E}_t'\mathbf{C}_t + \mathbf{E}_t'\mathbf{E}_t + (pq)^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_q' + (pq)^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q\mathbf{1}_p'\mathbf{E}_t \\
& - q^{-1}\mathbf{C}_t'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q' - q^{-1}\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\mathbf{C}_t - q^{-1}\mathbf{E}_t'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q' - q^{-1}\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\mathbf{E}_t - p^{-1}\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t
\end{aligned}
$$

$$- (pq)^{-1} q^{-1} (\mathbf{1}_p' \mathbf{E}_t \mathbf{1}_q)^2 \mathbf{1}_q \mathbf{1}_q' + q^{-2} \mathbf{1}_q' \mathbf{E}_t' \mathbf{E}_t \mathbf{1}_q \mathbf{1}_q \mathbf{1}_q', \tag{5.23}$$

so that from (5.20) and (5.21), we can write

$$\widehat{\mathbf{L}}_t \widehat{\mathbf{L}}_t' = \mathbf{C}_t \mathbf{C}_t' + \mathbf{R}_{r,t}, \quad \widehat{\mathbf{L}}_t' \widehat{\mathbf{L}}_t = \mathbf{C}_t' \mathbf{C}_t + \mathbf{R}_{c,t}.$$

Recall that we denote by $\widehat{\mathbf{D}}_r$ the $k_r \times k_r$ diagonal matrix with the first largest $k_r$ eigenvalues of $T^{-1} \sum_{t=1}^T \widehat{\mathbf{L}}_t \widehat{\mathbf{L}}_t'$ on the main diagonal, and since $\widehat{\mathbf{Q}}_r$ consists of the corresponding eigenvectors,

$$\widehat{\mathbf{Q}}_r \widehat{\mathbf{D}}_r = T^{-1} \sum_{t=1}^T \widehat{\mathbf{L}}_t \widehat{\mathbf{L}}_t' \widehat{\mathbf{Q}}_r. \tag{5.24}$$

With (5.20) and $\mathbf{C}_t \mathbf{C}_t' = \mathbf{Q}_r \mathbf{F}_{Z,t} \mathbf{Q}_c' \mathbf{Q}_c \mathbf{F}_{Z,t}' \mathbf{Q}_r'$, we can write the $j$-th row of estimated row factor loading as

$$\begin{aligned}
\widehat{\mathbf{Q}}_{r,j\cdot} &= T^{-1} \widehat{\mathbf{D}}_r^{-1} \sum_{i=1}^p \widehat{\mathbf{Q}}_{r,i\cdot} \sum_{t=1}^T (\widehat{\mathbf{L}}_t \widehat{\mathbf{L}}_t')_{ij} \\
&= T^{-1} \widehat{\mathbf{D}}_r^{-1} \sum_{i=1}^p \widehat{\mathbf{Q}}_{r,i\cdot} \mathbf{Q}_{r,i\cdot}' \sum_{t=1}^T (\mathbf{F}_{Z,t} \mathbf{Q}_c' \mathbf{Q}_c \mathbf{F}_{Z,t}') \mathbf{Q}_{r,j\cdot} + T^{-1} \widehat{\mathbf{D}}_r^{-1} \sum_{i=1}^p \widehat{\mathbf{Q}}_{r,i\cdot} \sum_{t=1}^T (\mathbf{R}_{r,t})_{ij}.
\end{aligned}$$

Thus with the definition $\mathbf{H}_r = T^{-1} \widehat{\mathbf{D}}_r^{-1} \widehat{\mathbf{Q}}_r' \mathbf{Q}_r \sum_{t=1}^T (\mathbf{F}_{Z,t} \mathbf{Q}_c' \mathbf{Q}_c \mathbf{F}_{Z,t}')$, we have

$$\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r \mathbf{Q}_{r,j\cdot} = T^{-1} \widehat{\mathbf{D}}_r^{-1} \sum_{i=1}^p \widehat{\mathbf{Q}}_{r,i\cdot} \sum_{t=1}^T (\mathbf{R}_{r,t})_{ij},$$

and hence we have

$$\begin{aligned}
&\|\widehat{\mathbf{Q}}_r - \mathbf{Q}_r \mathbf{H}_r'\|_F^2 = \sum_{j=1}^p \|\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r \mathbf{Q}_{r,j\cdot}\|^2 = \sum_{j=1}^p \left\| T^{-1} \widehat{\mathbf{D}}_r^{-1} \widehat{\mathbf{Q}}_r' \Big( \sum_{t=1}^T \mathbf{R}_{r,t} \Big)_{\cdot j} \right\|^2 \\
&\le T^{-2} \cdot \|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\widehat{\mathbf{Q}}_r\|_F^2 \cdot \left\| \sum_{t=1}^T \mathbf{R}_{r,t} \right\|_F^2 = O_P\Big( T^{-1} p^{2(1-\delta_{r,k_r})} q^{1-2\delta_{c,1}} + p^{1-2\delta_{r,k_r}} q^{2(1-\delta_{c,1})} \Big),
\end{aligned}$$

where the last equality used Lemma 5.2 and Lemma 5.3. The consistency of $\widehat{\mathbf{Q}}_c$ can be similarly shown (omitted here). This completes the proof of Theorem 5.2. □

***Proof of Theorem 5.3.*** From (5.4), we can first write

$$\begin{aligned}
\widehat{\mathbf{L}}_t &= \mathbf{Y}_t + (pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' - q^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t \\
&= \mu_t\mathbf{1}_p\mathbf{1}_q' + \boldsymbol{\alpha}_t\mathbf{1}_q' + \mathbf{1}_p\boldsymbol{\beta}_t' + \mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c' + \mathbf{E}_t \\
&\quad + (pq)^{-1}\mathbf{1}_p'(\mu_t\mathbf{1}_p\mathbf{1}_q' + \boldsymbol{\alpha}_t\mathbf{1}_q' + \mathbf{1}_p\boldsymbol{\beta}_t' + \mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c' + \mathbf{E}_t)\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' \\
&\quad - q^{-1}(\mu_t\mathbf{1}_p\mathbf{1}_q' + \boldsymbol{\alpha}_t\mathbf{1}_q' + \mathbf{1}_p\boldsymbol{\beta}_t' + \mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c' + \mathbf{E}_t)\mathbf{1}_q\mathbf{1}_q' \\
&\quad - p^{-1}\mathbf{1}_p\mathbf{1}_p'(\mu_t\mathbf{1}_p\mathbf{1}_q' + \boldsymbol{\alpha}_t\mathbf{1}_q' + \mathbf{1}_p\boldsymbol{\beta}_t' + \mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c' + \mathbf{E}_t) \\
&= \mu_t\mathbf{1}_p\mathbf{1}_q' + \boldsymbol{\alpha}_t\mathbf{1}_q' + \mathbf{1}_p\boldsymbol{\beta}_t' + \mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c' + \mathbf{E}_t + (pq)^{-1}\mathbf{1}_p'\mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' \\
&\quad + (pq)^{-1}\mathbf{1}_p'\boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' + (pq)^{-1}\mathbf{1}_p'\mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' + (pq)^{-1}\mathbf{1}_p'\mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c'\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' \\
&\quad + (pq)^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' - q^{-1}\mu_t\mathbf{1}_p\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_q' - q^{-1}\boldsymbol{\alpha}_t\mathbf{1}_q'\mathbf{1}_q\mathbf{1}_q' - q^{-1}\mathbf{1}_p\boldsymbol{\beta}_t'\mathbf{1}_q\mathbf{1}_q' \\
&\quad - q^{-1}\mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c'\mathbf{1}_q\mathbf{1}_q' - q^{-1}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mu_t\mathbf{1}_p\mathbf{1}_q' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\boldsymbol{\alpha}_t\mathbf{1}_q' \\
&\quad - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{1}_p\boldsymbol{\beta}_t' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t \\
&= \mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c' + \mathbf{E}_t + (pq)^{-1}\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' - q^{-1}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t,
\end{aligned}$$

where the last equality used Assumption (IC1). Thus, we have

$$\begin{aligned}
&\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1} = \widehat{\mathbf{Q}}_r'\widehat{\mathbf{L}}_t\widehat{\mathbf{Q}}_c - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1} \\
&= \widehat{\mathbf{Q}}_r'(\mathbf{Q}_r\mathbf{H}_r')(\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}(\mathbf{Q}_c\mathbf{H}_c')'\widehat{\mathbf{Q}}_c - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1} + \widehat{\mathbf{Q}}_r'\mathbf{E}_t\widehat{\mathbf{Q}}_c \\
&\quad + (pq)^{-1}\widehat{\mathbf{Q}}_r'\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\widehat{\mathbf{Q}}_c - q^{-1}\widehat{\mathbf{Q}}_r'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\widehat{\mathbf{Q}}_c - p^{-1}\widehat{\mathbf{Q}}_r'\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\widehat{\mathbf{Q}}_c \\
&= \widehat{\mathbf{Q}}_r'(\mathbf{Q}_r\mathbf{H}_r' - \widehat{\mathbf{Q}}_r)(\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}(\mathbf{Q}_c\mathbf{H}_c' - \widehat{\mathbf{Q}}_c)'\widehat{\mathbf{Q}}_c + \widehat{\mathbf{Q}}_r'(\mathbf{Q}_r\mathbf{H}_r' - \widehat{\mathbf{Q}}_r)(\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1} \\
&\quad + (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}(\mathbf{Q}_c\mathbf{H}_c' - \widehat{\mathbf{Q}}_c)'\widehat{\mathbf{Q}}_c + (\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r)'\mathbf{E}_t(\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c') \\
&\quad + (\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r)'\mathbf{E}_t\mathbf{Q}_c\mathbf{H}_c' + \mathbf{H}_r'\mathbf{Q}_r'\mathbf{E}_t(\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c') + \mathbf{H}_r'\mathbf{Q}_r'\mathbf{E}_t\mathbf{Q}_c\mathbf{H}_c' \\
&\quad + (pq)^{-1}\widehat{\mathbf{Q}}_r'\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\widehat{\mathbf{Q}}_c - q^{-1}(\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r')'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\widehat{\mathbf{Q}}_c - q^{-1}\mathbf{H}_r\mathbf{Q}_r'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\widehat{\mathbf{Q}}_c \\
&\quad - p^{-1}\widehat{\mathbf{Q}}_r'\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t(\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c') - p^{-1}\widehat{\mathbf{Q}}_r'\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{Q}_c\mathbf{H}_c' \\
&=: \mathcal{I}_{F,1} + \mathcal{I}_{F,2} + \mathcal{I}_{F,3} + \mathcal{I}_{F,4} + \mathcal{I}_{F,5} + \mathcal{I}_{F,6} + \mathcal{I}_{F,7} + \mathcal{I}_{F,8} \\
&\quad - \mathcal{I}_{F,9} - \mathcal{I}_{F,10} - \mathcal{I}_{F,11} - \mathcal{I}_{F,12}, \quad \text{where}
\end{aligned} \tag{5.25}$$

$$\begin{aligned}
&\mathcal{I}_{F,1} := \widehat{\mathbf{Q}}_r'(\mathbf{Q}_r\mathbf{H}_r' - \widehat{\mathbf{Q}}_r)(\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}(\mathbf{Q}_c\mathbf{H}_c' - \widehat{\mathbf{Q}}_c)'\widehat{\mathbf{Q}}_c, \\
&\mathcal{I}_{F,2} := \widehat{\mathbf{Q}}_r'(\mathbf{Q}_r\mathbf{H}_r' - \widehat{\mathbf{Q}}_r)(\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}, \quad \mathcal{I}_{F,3} := (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}(\mathbf{Q}_c\mathbf{H}_c' - \widehat{\mathbf{Q}}_c)'\widehat{\mathbf{Q}}_c, \\
&\mathcal{I}_{F,4} := (\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r)'\mathbf{E}_t(\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c'), \quad \mathcal{I}_{F,5} := (\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r)'\mathbf{E}_t\mathbf{Q}_c\mathbf{H}_c', \\
&\mathcal{I}_{F,6} := \mathbf{H}_r'\mathbf{Q}_r'\mathbf{E}_t(\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c'), \quad \mathcal{I}_{F,7} := \mathbf{H}_r'\mathbf{Q}_r'\mathbf{E}_t\mathbf{Q}_c\mathbf{H}_c', \quad \mathcal{I}_{F,8} := (pq)^{-1}\widehat{\mathbf{Q}}_r'\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q'\widehat{\mathbf{Q}}_c, \\
&\mathcal{I}_{F,9} := q^{-1}(\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r')'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\widehat{\mathbf{Q}}_c, \quad \mathcal{I}_{F,10} := q^{-1}\mathbf{H}_r\mathbf{Q}_r'\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\widehat{\mathbf{Q}}_c, \\
&\mathcal{I}_{F,11} := p^{-1}\widehat{\mathbf{Q}}_r'\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t(\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c'), \quad \mathcal{I}_{F,12} := p^{-1}\widehat{\mathbf{Q}}_r'\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{Q}_c\mathbf{H}_c'.
\end{aligned}$$

First consider $\mathbf{F}_{Z,t}$, by its definition and Assumption (L1) we have

$$\|\mathbf{F}_{Z,t}\|_F^2 \le \|\mathbf{F}_t\|_F^2 \cdot \|\mathbf{Z}_r^{1/2}\|_F^2 \cdot \|\mathbf{Z}_c^{1/2}\|_F^2 = O_P(p^{\delta_{r,1}} q^{\delta_{c,1}}).$$

Then for $\mathcal{I}_{F,1}$, we have

$$\begin{aligned}
\|\mathcal{I}_{F,1}\|_F^2 &= O_P(p^{\delta_{r,1}} q^{\delta_{c,1}}) \cdot \|\mathbf{Q}_r\mathbf{H}_r' - \widehat{\mathbf{Q}}_r\|_F^2 \cdot \|\mathbf{Q}_c\mathbf{H}_c' - \widehat{\mathbf{Q}}_c\|_F^2 \\
&= O_P\big(T^{-2} p^{2-2\delta_{r,k_r}} q^{2-2\delta_{c,k_c}} + T^{-1} p^{3-2\delta_{r,k_r}} q^{2-\delta_{c,1}-3\delta_{c,k_c}} \\
&\quad + T^{-1} q^{3-2\delta_{c,k_c}} p^{2-\delta_{r,1}-3\delta_{r,k_r}} + p^{3-\delta_{r,1}-3\delta_{r,k_r}} q^{3-\delta_{c,1}-3\delta_{c,k_c}}\big),
\end{aligned}$$

where we used Lemma 5.6 in the last equality. Similarly for $\mathcal{I}_{F,2}$ and $\mathcal{I}_{F,3}$,

$$\begin{aligned}
\|\mathcal{I}_{F,2}\|_F^2 &= O_P\big(T^{-1} p^{1+2\delta_{r,1}-2\delta_{r,k_r}} q^{1-\delta_{c,1}} + p^{1+\delta_{r,1}-3\delta_{r,k_r}} q^{2-\delta_{c,1}}\big), \\
\|\mathcal{I}_{F,3}\|_F^2 &= O_P\big(T^{-1} q^{1+2\delta_{c,1}-2\delta_{c,k_c}} p^{1-\delta_{r,1}} + q^{1+\delta_{c,1}-3\delta_{c,k_c}} p^{2-\delta_{r,1}}\big).
\end{aligned}$$

For $\mathcal{I}_{F,4}$, from Assumptions (E1) and (E2) we easily have $\|\mathbf{E}_t\|_F^2 = O(pq)$, so we have

$$\begin{aligned}
\|\mathcal{I}_{F,4}\|_F^2 &= O_P(pq) \cdot \|\mathbf{Q}_r\mathbf{H}_r' - \widehat{\mathbf{Q}}_r\|_F^2 \cdot \|\mathbf{Q}_c\mathbf{H}_c' - \widehat{\mathbf{Q}}_c\|_F^2 \\
&= O_P\big(T^{-2} p^{3-\delta_{r,1}-2\delta_{r,k_r}} q^{3-\delta_{c,1}-2\delta_{c,k_c}} + T^{-1} p^{4-\delta_{r,1}-2\delta_{r,k_r}} q^{3-2\delta_{c,1}-3\delta_{c,k_c}} \\
&\quad + T^{-1} q^{4-\delta_{c,1}-2\delta_{c,k_c}} p^{3-2\delta_{r,1}-3\delta_{r,k_r}} + p^{4-2\delta_{r,1}-3\delta_{r,k_r}} q^{4-2\delta_{c,1}-3\delta_{c,k_c}}\big),
\end{aligned}$$

For $\mathcal{I}_{F,5}$, consider first

$$\begin{aligned}
&\mathbb{E}\Big\{\|\mathbf{E}_t\mathbf{A}_c\|_F^2\Big\} \le pk_c \max_{i\in[p],j\in[k_c]} \mathbb{E}\Big\{(\mathbf{E}_{t,i\cdot}'\mathbf{A}_{c,\cdot j})^2\Big\} \\
&= pk_c \max_{i\in[p],j\in[k_c]} \sum_{n=1}^{q}\sum_{l=1}^{q} \mathrm{Cov}(E_{t,in}, E_{t,il}) \cdot A_{c,nj}A_{c,lj} = O(pq),
\end{aligned} \tag{5.26}$$

where the last equality used Assumptions (E1) and (E2). Thus,

$$\begin{aligned}
\|\mathcal{I}_{F,5}\|_F^2 &= O_P(pq) \cdot \|\mathbf{Q}_r\mathbf{H}_r' - \widehat{\mathbf{Q}}_r\|_F^2 \cdot \|\mathbf{Z}_c^{-1/2}\|_F^2 \\
&= O_P\big(T^{-1} p^{2+\delta_{r,1}-2\delta_{r,k_r}} q^{2-2\delta_{c,1}-\delta_{c,k_c}} + p^{2-3\delta_{r,k_r}} q^{3-2\delta_{c,1}-\delta_{c,k_c}}\big),
\end{aligned}$$

where we used Lemma 5.6 and Assumption (L1). Similarly for $\mathcal{I}_{F,6}$,

$$\|\mathcal{I}_{F,6}\|_F^2 = O_P\big(T^{-1} q^{2+\delta_{c,1}-2\delta_{c,k_c}} p^{2-2\delta_{r,1}-\delta_{r,k_r}} + q^{2-3\delta_{c,k_c}} p^{3-2\delta_{r,1}-\delta_{r,k_r}}\big).$$

By Assumptions (E1) and (E2) again, we have

$$\mathbb{E}\Big\{\|\mathbf{A}_r'\mathbf{E}_t\mathbf{A}_c\|_F^2\Big\} \leq k_r k_c \max_{i\in[k_r],j\in[k_c]} \mathbb{E}\Big\{(\mathbf{A}_{r,\cdot i}'\mathbf{E}_t\mathbf{A}_{c,\cdot j})^2\Big\}$$
$$= k_r k_c \max_{i\in[k_r],j\in[k_c]} \sum_{m=1}^{p}\sum_{n=1}^{q}\sum_{h=1}^{p}\sum_{l=1}^{q} \mathrm{Cov}(E_{t,mn}, E_{t,hl}) \cdot A_{r,mi}A_{c,nj}A_{r,hi}A_{c,lj} = O(pq),$$

hence for $\mathcal{I}_{F,7}$, it holds that

$$\|\mathcal{I}_{F,7}\|_F^2 = O_P(pq) \cdot \|\mathbf{Z}_r^{-1/2}\|_F^2 \cdot \|\mathbf{Z}_c^{-1/2}\|_F^2 = O_P\big(p^{1-\delta_{r,k_r}} q^{1-\delta_{c,k_c}}\big).$$

Consider $\mathcal{I}_{F,8}$, recall that $(\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q)^2 = O_P(pq)$ from (5.10) and hence

$$\|\mathcal{I}_{F,8}\|_F^2 = O_P(p^{-2}q^{-2}) \cdot (\mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q)^2 \cdot \|\mathbf{1}_p\mathbf{1}_q'\|_F^2 = O_P(1).$$

For $\mathcal{I}_{F,9}$, note that $\mathbb{E}[\|\mathbf{E}_t\mathbf{1}_q\|_F^2]$ has the same rate as (5.26) since $k_c$ is fixed, and hence

$$\begin{aligned}\|\mathcal{I}_{F,9}\|_F^2 &= O_P(q^{-2}) \cdot \|\mathbf{Q}_r\mathbf{H}_r' - \widehat{\mathbf{Q}}_r\|_F^2 \cdot \|\mathbf{E}_t\mathbf{1}_q\|_F^2 \cdot \|\mathbf{1}_q\|^2 \\ &= O_P\big(T^{-1}p^{2+\delta_{r,1}-2\delta_{r,k_r}} q^{1-2\delta_{c,1}} + p^{2-3\delta_{r,k_r}} q^{2-2\delta_{c,1}}\big).\end{aligned}$$

From (5.10), we also have $\mathbb{E}[\|\mathbf{A}_r'\mathbf{E}_t\mathbf{1}_q\|_F^2] = O(pq)$, so for $\mathcal{I}_{F,10}$ we have

$$\|\mathcal{I}_{F,10}\|_F^2 = O_P(q^{-2}) \cdot \|\mathbf{Z}_r^{-1/2}\|_F^2 \cdot \|\mathbf{A}_r'\mathbf{E}_t\mathbf{1}_q\|_F^2 \cdot \|\mathbf{1}_q\|^2 = O_P\big(p^{1-\delta_{r,k_r}}\big).$$

Lastly, the rates for $\mathcal{I}_{F,11}$ and $\mathcal{I}_{F,12}$ can be obtained similarly as $\mathcal{I}_{F,9}$ and $\mathcal{I}_{F,10}$,

$$\begin{aligned}&\|\mathcal{I}_{F,11}\|_F^2 = O_P\big(T^{-1}q^{2+\delta_{c,1}-2\delta_{c,k_c}} p^{1-2\delta_{r,1}} + q^{2-3\delta_{c,k_c}} p^{2-2\delta_{r,1}}\big), \\ &\|\mathcal{I}_{F,12}\|_F^2 = O_P\big(q^{1-\delta_{c,k_c}}\big).\end{aligned}$$

Therefore, with all the rates from $\mathcal{I}_{F,1}$ to $\mathcal{I}_{F,12}$ in (5.25), by using Assumption (R1) we have

$$\begin{aligned}&\|\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}\|_F^2 \\ &= O_P\big(p^{1-\delta_{r,k_r}} q^{1-\delta_{c,k_c}} + T^{-1}p^{1+2\delta_{r,1}-2\delta_{r,k_r}} q^{1-\delta_{c,1}} + p^{1+\delta_{r,1}-3\delta_{r,k_r}} q^{2-\delta_{c,1}} \\ &\qquad + T^{-1}q^{1+2\delta_{c,1}-2\delta_{c,k_c}} p^{1-\delta_{r,1}} + q^{1+\delta_{c,1}-3\delta_{c,k_c}} p^{2-\delta_{r,1}}\big).\end{aligned}$$

This shows the first statement of Theorem 5.3.

For the remaining proof, consider $\widehat{C}_{t,ij} - C_{t,ij}$ for any $t \in [T], i \in [p], j \in [q]$. First,

$$\begin{aligned}
&\widehat{C}_{t,ij} - C_{t,ij} = \widehat{\mathbf{Q}}_{r,i\cdot}'\widehat{\mathbf{F}}_{Z,t}\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{Q}_{r,i\cdot}'\mathbf{F}_{Z,t}\mathbf{Q}_{c,j\cdot} \\
&= (\widehat{\mathbf{Q}}_{r,i\cdot} - \mathbf{H}_r\mathbf{Q}_{r,i\cdot})'(\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1})(\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}) \\
&\quad + (\widehat{\mathbf{Q}}_{r,i\cdot} - \mathbf{H}_r\mathbf{Q}_{r,i\cdot})'(\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1})\mathbf{H}_c\mathbf{Q}_{c,j\cdot} \\
&\quad + (\widehat{\mathbf{Q}}_{r,i\cdot} - \mathbf{H}_r\mathbf{Q}_{r,i\cdot})'(\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}(\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}) \\
&\quad + (\widehat{\mathbf{Q}}_{r,i\cdot} - \mathbf{H}_r\mathbf{Q}_{r,i\cdot})'(\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{Q}_{c,j\cdot} + \mathbf{Q}_{r,i\cdot}'\mathbf{H}_r'(\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1})(\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}) \\
&\quad + \mathbf{Q}_{r,i\cdot}'\mathbf{H}_r'(\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1})\mathbf{H}_c\mathbf{Q}_{c,j\cdot} + \mathbf{Q}_{r,i\cdot}'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}(\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}).
\end{aligned} \tag{5.27}$$

Notice that from Assumption (L1),

$$\begin{aligned}
\|\mathbf{Q}_{r,i\cdot}\|^2 &= O(\|\mathbf{A}_{r,i\cdot}\|^2) \cdot \|\mathbf{Z}_r^{-1/2}\|_F^2 = O(p^{-\delta_{r,k_r}}), \\
\|\mathbf{Q}_{c,j\cdot}\|^2 &= O(\|\mathbf{A}_{c,j\cdot}\|^2) \cdot \|\mathbf{Z}_c^{-1/2}\|_F^2 = O(q^{-\delta_{c,k_c}}).
\end{aligned}$$

Together with Lemma 5.6, the first statement of Theorem 5.3 and Assumption (R1), we have

$$\begin{aligned}
&(\widehat{C}_{t,ij} - C_{t,ij})^2 \\
&= O_P\big(p^{1-2\delta_{r,k_r}}q^{1-2\delta_{c,k_c}} + T^{-1}p^{1+2\delta_{r,1}-3\delta_{r,k_r}}q^{1-\delta_{c,1}-\delta_{c,k_c}} + p^{1+\delta_{r,1}-4\delta_{r,k_r}}q^{2-\delta_{c,1}-\delta_{c,k_c}} \\
&\qquad + T^{-1}q^{1+2\delta_{c,1}-3\delta_{c,k_c}}p^{1-\delta_{r,1}-\delta_{r,k_r}} + q^{1+\delta_{c,1}-4\delta_{c,k_c}}p^{2-\delta_{r,1}-\delta_{r,k_r}}\big).
\end{aligned}$$

This completes the proof of Theorem 5.3. □

***Proof of Theorem 5.4.*** We first consider $\widehat{\mu}_t$, which is given by

$$\begin{aligned}
\widehat{\mu}_t &= \mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q/(pq) = \mu_t + \mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q/(pq) = \mu_t + \frac{1}{pq}\sum_{i,j} E_{t,ij}, \text{ so that} \\
\widehat{\mu}_t - \mu_t &= \frac{1}{pq}\mathbf{1}_p'\mathbf{A}_{e,r}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'\mathbf{1}_q + \frac{1}{pq}\sum_{i,j}\Sigma_{\epsilon,ij}\epsilon_{t,ij} =: I_{\mu,1} + I_{\mu,2}.
\end{aligned}$$

By Assumption (E1), since $\|\mathbf{A}_{e,r}\|_1, \|\mathbf{A}_{e,c}\|_1 = O(1)$, we have $I_{\mu,1} = O_P(1/(pq))$. Also, $I_{\mu,2} = O((pq)^{-1/2})$ since $\mathbb{E}(I_{\mu,2}) = 0$ and $\text{Var}(I_{\mu,2}) = (pq)^{-2}\sum_{i,j}\Sigma_{\epsilon,ij}^2 = O(1/(pq))$. Hence $I_{\mu,1}$ is dominated by $I_{\mu,2}$, and

$$\sqrt{pq}(\widehat{\mu}_t - \mu_t) = \sqrt{pq}I_{\mu,2}(1 + o_P(1)) = \frac{1}{\sqrt{pq}}\sum_{i,j}\Sigma_{\epsilon,ij}\epsilon_{t,ij}(1 + o_P(1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma_\mu^2),$$

where we use Theorem 1 in Ayvazyan and Ulyanov (2023) for the convergence in distribution.

For $\widehat{\boldsymbol{\alpha}}_t$, consider the decomposition

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t &= (\mu_t - \widehat{\mu}_t)\mathbf{1}_p + q^{-1}\mathbf{E}_t\mathbf{1}_q \\ &= (\mu_t - \widehat{\mu}_t)\mathbf{1}_p + q^{-1}\mathbf{A}_{e,r}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'\mathbf{1}_q + q^{-1}(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)\mathbf{1}_q,\end{aligned}$$

so that

$$\widehat{\alpha}_{t,i} - \alpha_{t,i} = (\mu_t - \widehat{\mu}_t) + q^{-1}(\mathbf{A}_{e,r})_{i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'\mathbf{1}_q + q^{-1}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}\epsilon_{t,ij}.$$

From what we have proved above we have $|\widehat{\mu}_t - \mu_t| = O_P(1/\sqrt{pq})$. Since $\|\mathbf{A}_{e,r}\|_1, \|\mathbf{A}_{e,c}\|_1 = O(1)$, we also have $q^{-1}(\mathbf{A}_{e,r})_{i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'\mathbf{1}_q = O_P(1/q)$. Finally, $\mathbb{E}(q^{-1}(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)\mathbf{1}_q) = \mathbf{0}$ and $\text{Var}(q^{-1}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}\epsilon_{t,ij}) = q^{-2}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2 = O(q^{-1})$, implying that $q^{-1}(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)\mathbf{1}_q = O_P(q^{-1/2})$ element-wise, thus dominating other terms. Hence for $i \in [p]$, using Theorem 1 in Ayvazyan and Ulyanov (2023),

$$\sqrt{q}(\widehat{\alpha}_{t,i} - \alpha_{t,i}) = q^{-1/2}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}\epsilon_{t,ij}(1 + o_P(1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma_{\alpha,i}^2).$$

By Assumption (E1), each element of $q^{-1}(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)\mathbf{1}_q$ is independent of each other. Hence with integers $i_1 < \cdots < i_m$, $m$ being finite and $\boldsymbol{\theta}_{\alpha,t} := (\alpha_{t,i_1}, \ldots, \alpha_{t,i_m})'$, by Theorem 1 in Ayvazyan and Ulyanov (2023),

$$\sqrt{q}(\widehat{\boldsymbol{\theta}}_{\alpha,t} - \boldsymbol{\theta}_{\alpha,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \text{diag}(\gamma_{\alpha,i_1}^2, \ldots, \gamma_{\alpha,i_m}^2)).$$

We omit the proof of asymptotic normality for $\widehat{\boldsymbol{\beta}}_t$ since the arguments used are in parallel to those used for $\widehat{\boldsymbol{\alpha}}_t$, using the independence of the columns in $\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t$ by Assumption (E1).

The rest of the proof is done if we can prove that $\widehat{\gamma}_{\alpha,i}$, $\widehat{\gamma}_{\beta,j}$ and $\widehat{\gamma}_\mu$ are consistent estimators for $\gamma_{\alpha,i}$, $\gamma_{\beta,j}$ and $\gamma_\mu$ respectively. From (5.6), since we assume $\widehat{C}_{t,ij} - C_{t,ij} = o_P(1)$ from Theorem 5.3, then element-wise we have

$$\begin{aligned}\widehat{\mathbf{E}}_t &= \widehat{\mathbf{L}}_t - \widehat{\mathbf{C}}_t \\ &= (\mu_t - \widehat{\mu}_t)\mathbf{1}_p\mathbf{1}_q' + (\boldsymbol{\alpha}_t - \widehat{\boldsymbol{\alpha}}_t)\mathbf{1}_q' + \mathbf{1}_p(\boldsymbol{\beta}_t - \widehat{\boldsymbol{\beta}}_t)' + (\mathbf{C}_t - \widehat{\mathbf{C}}_t) + \mathbf{E}_t = \mathbf{E}_t(1 + o_P(1)).\end{aligned}$$

Hence we have

$$\begin{aligned}q^{-1}(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii} &= \{q^{-1}(\mathbf{A}_{e,r})_{i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'\mathbf{A}_{e,c}\mathbf{F}_{e,t}'(\mathbf{A}_{e,r})_{i\cdot}' + q^{-1}(\mathbf{A}_{e,r})_{i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)_{i\cdot}' \\ &\quad + q^{-1}(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)_{i\cdot}(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)_{i\cdot}'\}(1 + o_P(1)) \\ &= O_P(q^{-1}) + O_P(q^{-1}) + \frac{1}{q}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2(\epsilon_{t,ij}^2 - 1)(1 + o_P(1)) + \frac{1}{q}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2\end{aligned}$$

$$\xrightarrow{\mathcal{P}} \gamma_{\alpha,i}^2,$$

where we used the Markov inequality to arrive at $q^{-1}\sum_{j=1}^q \Sigma_{\epsilon,ij}^2(\epsilon_{t,ij}^2 - 1) = O_P(q^{-1/2})$, knowing that each $\Sigma_{\epsilon,ij}$ is bounded away from infinity by Assumption (E1). A parallel argument (omitted) can show that $\widehat{\gamma}_{\beta,j}^2$ is consistent for $\gamma_{\beta,j}^2$. Finally,

$$\widehat{\gamma}_\mu^2 = p^{-1}\sum_{i=1}^p q^{-1}(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii} \xrightarrow{\mathcal{P}} p^{-1}\sum_{i=1}^p \gamma_{\alpha,i}^2 = \gamma_\mu^2.$$

This completes the proof of the theorem. □

***Proof of Theorem 5.5.*** We construct the asymptotic normality for rows of our factor loading estimators. We only prove the result for the row loading estimator, and the proof for the column loading estimator would be similar (omitted). For any $j \in [p]$, consider the decomposition

$$\begin{aligned}
&\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot} = T^{-1}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \widehat{\mathbf{Q}}_{r,i\cdot}\sum_{t=1}^T(\mathbf{R}_{r,t})_{ij} \\
&= T^{-1}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p(\widehat{\mathbf{Q}}_{r,i\cdot} - \mathbf{H}_r\mathbf{Q}_{r,i\cdot})\sum_{t=1}^T(\mathbf{R}_{r,t})_{ij} + T^{-1}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{R}_{r,t})_{ij} \qquad (5.28)\\
&= \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4 + \mathcal{I}_5 - \mathcal{I}_6 - \mathcal{I}_7 - \mathcal{I}_8 - \mathcal{I}_9 - \mathcal{I}_{10} - \mathcal{I}_{11} + \mathcal{I}_{12} + \mathcal{I}_{13}, \quad \text{where}
\end{aligned}$$

$$\mathcal{I}_1 := \frac{1}{T}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{C}_t\mathbf{E}_t')_{ij}, \quad \mathcal{I}_2 := \frac{1}{T}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{E}_t\mathbf{C}_t')_{ij},$$

$$\mathcal{I}_3 := \frac{1}{T}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{E}_t\mathbf{E}_t')_{ij}, \quad \mathcal{I}_4 := \frac{1}{Tpq}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p')_{ij},$$

$$\mathcal{I}_5 := \frac{1}{Tpq}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t')_{ij},$$

$$\mathcal{I}_6 := \frac{1}{Tp}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{C}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p')_{ij}, \quad \mathcal{I}_7 := \frac{1}{Tp}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{C}_t')_{ij},$$

$$\mathcal{I}_8 := \frac{1}{Tp}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p')_{ij}, \quad \mathcal{I}_9 := \frac{1}{Tp}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t')_{ij},$$

$$\mathcal{I}_{10} := \frac{1}{Tp}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t')_{ij},$$

$$\mathcal{I}_{11} := \frac{1}{Tp^2q}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T((\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\mathbf{1}_p\mathbf{1}_p')_{ij},$$

$$\mathcal{I}_{12} := \frac{1}{Tp^2}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p')_{ij},$$

$$\mathcal{I}_{13} := \frac{1}{T}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^{p}(\widehat{\mathbf{Q}}_{r,i\cdot} - \mathbf{H}_r\mathbf{Q}_{r,i\cdot})\sum_{t=1}^{T}(\mathbf{R}_{r,t})_{ij}.$$

We shall show that $\mathcal{I}_1$ is the leading term among the decomposition in (5.28). To obtain the rate for $\mathcal{I}_2$, from Assumptions (E1) and (E2) we have for any $i \in [p], h \in [q]$,

$$E_{t,ih} = \sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot} + (\mathbf{\Sigma}_\epsilon)_{ih}\sum_{w\geq 0} a_{\epsilon,w}(\mathbf{X}_{\epsilon,t-w})_{ih}.$$

Consider first $\sum_{t=1}^{T}\sum_{h=1}^{q}\big(\sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\big)\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot}$. We have from Assumptions (F1), (E1) and (E2) that

$$\begin{aligned}
&\mathbb{E}\Big\{\Big[\sum_{t=1}^{T}\sum_{h=1}^{q}\Big(\sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\Big)\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot}\Big]^2\Big\}\\
&= \mathrm{Cov}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{q}\mathbf{A}'_{c,h\cdot}\Big(\sum_{w\geq 0} a_{f,w}\mathbf{X}'_{f,t-w}\Big)\mathbf{A}_{r,j\cdot}\Big(\sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\Big),\\
&\qquad \sum_{t=1}^{T}\sum_{h=1}^{q}\mathbf{A}'_{c,h\cdot}\Big(\sum_{w\geq 0} a_{f,w}\mathbf{X}'_{f,t-w}\Big)\mathbf{A}_{r,j\cdot}\Big(\sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\Big)\Big\}\\
&= \sum_{h=1}^{q}\sum_{l=1}^{q}\sum_{t=1}^{T}\sum_{w\geq 0} a_{f,w}^2 a_{e,w}^2 \cdot \|\mathbf{A}_{r,j\cdot}\|^2 \cdot \|\mathbf{A}_{c,h\cdot}\| \cdot \|\mathbf{A}_{c,l\cdot}\| \cdot \|\mathbf{A}_{e,c,h\cdot}\| \cdot \|\mathbf{A}_{e,c,l\cdot}\| \cdot \|\mathbf{A}_{e,r,i\cdot}\|^2\\
&= O(T) \cdot \|\mathbf{A}_{r,j\cdot}\|^2 \cdot \|\mathbf{A}_{e,r,i\cdot}\|^2.
\end{aligned} \tag{5.29}$$

Consider also $\sum_{i=1}^{p}\sum_{h=1}^{q}\sum_{t=1}^{T}\mathbf{Q}_{r,i\cdot}((\mathbf{\Sigma}_\epsilon)_{ih}\sum_{w\geq 0} a_{\epsilon,w}(\mathbf{X}_{\epsilon,t-w})_{ih})\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot}$. Similarly, by Assumptions (E1), (E2) and (F1), we have

$$\begin{aligned}
&\mathbb{E}\Big\{\Big\|\sum_{i=1}^{p}\sum_{h=1}^{q}\sum_{t=1}^{T}\mathbf{Q}_{r,i\cdot}\Big((\mathbf{\Sigma}_\epsilon)_{ih}\sum_{w\geq 0} a_{\epsilon,w}(\mathbf{X}_{\epsilon,t-w})_{ih}\Big)\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot}\Big\|^2\Big\}\\
&= \mathrm{Cov}\Big\{\sum_{i=1}^{p}\sum_{h=1}^{q}\sum_{t=1}^{T}\mathbf{Q}_{r,i\cdot}\Big((\mathbf{\Sigma}_\epsilon)_{ih}\sum_{w\geq 0} a_{\epsilon,w}(\mathbf{X}_{\epsilon,t-w})_{ih}\Big)\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot},\\
&\qquad \sum_{i=1}^{p}\sum_{h=1}^{q}\sum_{t=1}^{T}\mathbf{Q}_{r,i\cdot}\Big((\mathbf{\Sigma}_\epsilon)_{ih}\sum_{w\geq 0} a_{\epsilon,w}(\mathbf{X}_{\epsilon,t-w})_{ih}\Big)\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot}\Big\}\\
&= \sum_{i=1}^{p}\sum_{h=1}^{q}\sum_{t=1}^{T}\sum_{w\geq 0} a_{f,w}^2 a_{\epsilon,w}^2 \cdot \|\mathbf{A}_{r,j\cdot}\|^2 \cdot \|\mathbf{A}_{c,h\cdot}\|^2 \cdot (\mathbf{\Sigma}_\epsilon)_{ih}^2 \cdot \|\mathbf{Q}_{r,i\cdot}\|^2\\
&= O(T) \cdot \|\mathbf{A}_{r,j\cdot}\|^2 \cdot \|\mathbf{A}_c\|^2 \cdot \|\mathbf{Q}_r\|^2.
\end{aligned} \tag{5.30}$$

Hence using Lemma 5.3, it holds that

$$
\begin{aligned}
&\|\mathcal{I}_2\|^2 \leq \frac{1}{T^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^{p} \mathbf{Q}_{r,i\cdot} \sum_{t=1}^{T}\sum_{h=1}^{q} E_{t,ih}\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot}\Big\|^2 \\
&= O_P\Big(\frac{1}{T^2}p^{-2\delta_{r,k_r}}q^{-2\delta_{c,1}}\Big)\Big\{\Big\| \sum_{i=1}^{p}\sum_{h=1}^{q}\sum_{t=1}^{T} \mathbf{Q}_{r,i\cdot}\Big((\boldsymbol{\Sigma}_\epsilon)_{ih}\sum_{w\geq 0} a_{\epsilon,w}(\mathbf{X}_{\epsilon,t-w})_{ih}\Big)\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot}\Big\|^2 \\
&+ \Big(\sum_{i=1}^{p}\|\mathbf{Q}_{r,i\cdot}\|^2\Big)\sum_{i=1}^{p}\Big[\sum_{t=1}^{T}\sum_{h=1}^{q}\Big(\sum_{w\geq 0} a_{e,w}\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\Big)\mathbf{A}'_{c,h\cdot}\mathbf{F}'_t\mathbf{A}_{r,j\cdot}\Big]^2\Big\} \\
&= O_P\Big(\frac{1}{T^2}p^{-2\delta_{r,k_r}}q^{-2\delta_{c,1}}\cdot Tq^{\delta_{c,1}}\Big) = O_P\big(T^{-1}p^{-2\delta_{r,k_r}}q^{-\delta_{c,1}}\big),
\end{aligned}
$$

where we used Assumption (L1), (5.29) and (5.30) in the last equality.

For $\mathcal{I}_3$, first notice from the noise structure in Assumptions (E1) and (E2), we have

$$
\begin{aligned}
&\mathrm{Var}\Big(\sum_{i=1}^{p}\sum_{h=1}^{q}\sum_{t=1}^{T}\mathbf{Q}_{r,i\cdot}E_{t,ih}E_{t,jh}\Big) \\
&= O(1)\cdot\sum_{i=1}^{p}\sum_{u=1}^{p}\sum_{h=1}^{q}\sum_{l=1}^{q}\sum_{t=1}^{T}\sum_{n=1}^{k_{e,r}}\sum_{m=1}^{k_{e,c}}\sum_{w\geq 0} a_{e,w}^4 A_{e,r,in}A_{e,r,un}A_{e,r,jn}^2A_{e,c,hm}^2A_{e,c,lm}^2 \\
&\cdot\|\mathbf{Q}_{r,i\cdot}\|\cdot\|\mathbf{Q}_{r,u\cdot}\|\cdot\mathrm{Var}((\mathbf{X}_{e,t-w})_{nm}^2) \\
&+ O(1)\cdot\sum_{i=1}^{p}\sum_{h=1}^{q}\sum_{t=1}^{T}\sum_{w\geq 0} a_{\epsilon,w}^4(\boldsymbol{\Sigma}_\epsilon)_{ih}^2(\boldsymbol{\Sigma}_\epsilon)_{jh}^2\cdot\|\mathbf{Q}_{r,i\cdot}\|^2\cdot\mathrm{Var}((\mathbf{X}_{\epsilon,t-w})_{ih}(\mathbf{X}_{\epsilon,t-w})_{jh}) \\
&= O(T+Tq) = O(Tq).
\end{aligned}
$$

Moreover, it holds that

$$
\begin{aligned}
&\mathbb{E}\Big(\sum_{i=1}^{p}\Big|\sum_{h=1}^{q}\sum_{t=1}^{T}E_{t,ih}E_{t,jh}\Big|\Big) \\
&= \sum_{i=1}^{p}\Big|\sum_{h=1}^{q}\sum_{t=1}^{T}\big(\|\mathbf{A}_{e,c,h\cdot}\|^2\cdot\|\mathbf{A}_{e,r,i\cdot}\|\cdot\|\mathbf{A}_{e,r,j\cdot}\| + (\boldsymbol{\Sigma}_\epsilon)_{ih}\mathbb{1}_{\{i=j\}}\big)\Big| = O(Tq),
\end{aligned}
$$

and with $\max_i\|\mathbf{Q}_{r,i\cdot}\|^2 \leq \|\mathbf{A}_{r,j\cdot}\|^2\cdot\|\mathbf{Z}_r^{-1/2}\|^2 = O_P\big(p^{-\delta_{r,k_r}}\big)$, we thus have

$$
\begin{aligned}
&\|\mathcal{I}_3\|^2 \leq \frac{1}{T^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2\cdot\|\mathbf{H}_r\|_F^2\cdot\Big\|\sum_{i=1}^{p}\mathbf{Q}_{r,i\cdot}\sum_{t=1}^{T}\sum_{h=1}^{q}E_{t,ih}E_{t,jh}\Big\|^2 \\
&= O_P\Big(T^{-2}p^{-2\delta_{r,k_r}}q^{-2\delta_{c,1}}(Tq + T^2q^2p^{-\delta_{r,k_r}})\Big) = O_P\big(T^{-1}p^{-2\delta_{r,k_r}}q^{1-2\delta_{c,1}} + p^{-3\delta_{r,k_r}}q^{2-2\delta_{c,1}}\big).
\end{aligned}
\tag{5.31}
$$

Consider now $\mathcal{I}_4$ and $\mathcal{I}_5$. From the proof of (5.49) in Lemma 5.2,

$$\|\mathcal{I}_4\|^2 \le \frac{1}{T^2p^2q^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^p \mathbf{Q}_{r,i\cdot} \sum_{t=1}^T (\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p')_{ij} \Big\|^2$$

$$\le \frac{1}{T^2p^2q^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big(\sum_{i=1}^p \|\mathbf{Q}_{r,i\cdot}\|^2\Big) \sum_{i=1}^p \Big(\sum_{t=1}^T \mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p'\Big)_{ij}^2$$

$$= O_P\big(T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2-2\delta_{c,1}}\big) \cdot \Big\| \Big(\sum_{t=1}^T \mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p'\Big)_{\cdot j} \Big\|^2$$

$$= O_P\big(T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2-2\delta_{c,1}}\big) \cdot \sum_{k=1}^p \mathbb{E}\Big\{\Big[\sum_{t=1}^T \Big(\sum_{i=1}^p\sum_{u=1}^q E_{t,iu}\Big)\sum_{h=1}^q E_{t,kh}\Big]^2\Big\}$$

$$= O_P\Big(T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2-2\delta_{c,1}}(Tp^2q^2 + T^2pq^2)\Big)$$

$$= O_P\big(T^{-1}p^{-2\delta_{r,k_r}}q^{-2\delta_{c,1}} + p^{-1-2\delta_{r,k_r}}q^{-2\delta_{c,1}}\big),$$

where we used Lemma 5.3 in the third line and (5.63) in the second last equality. Similarly,

$$\|\mathcal{I}_5\|^2 \le \frac{1}{T^2p^2q^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^p \mathbf{Q}_{r,i\cdot} \sum_{t=1}^T (\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_q'\mathbf{E}_t')_{ij} \Big\|^2$$

$$= O_P\big(T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2-2\delta_{c,1}}\big) \cdot \Big\| \Big(\sum_{t=1}^T \mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p'\Big)_{j\cdot} \Big\|^2$$

$$= O_P\big(T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2-2\delta_{c,1}}\big) \cdot p \cdot \mathbb{E}\Big[\Big\{\sum_{t=1}^T \Big(\sum_{i=1}^p\sum_{u=1}^q E_{t,iu}\Big)\sum_{h=1}^q E_{t,jh}\Big\}^2\Big]$$

$$= O_P\big\{T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2-2\delta_{c,1}}(Tp^2q^2 + T^2pq^2)\big\}$$

$$= O_P\big(T^{-1}p^{-2\delta_{r,k_r}}q^{-2\delta_{c,1}} + p^{-1-2\delta_{r,k_r}}q^{-2\delta_{c,1}}\big),$$

where we used again Lemma 5.3 in the third line and (5.63) in the second last equality.

Consider now $\mathcal{I}_6$, note first from Assumptions (E1) and (E2),

$$E_{t,hu} = \sum_{w\ge 0} a_{e,w}\mathbf{A}_{e,r,h\cdot}'\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,u\cdot} + (\mathbf{\Sigma}_\epsilon)_{hu}\sum_{w\ge 0} a_{\epsilon,w}(\mathbf{X}_{\epsilon,t-w})_{hu}.$$

Thus, we have

$$\mathrm{Cov}(E_{t,hu}, E_{s,vl}) = \mathbb{E}[\mathbf{A}_{e,r,h\cdot}'\mathbf{F}_{e,t}\mathbf{A}_{e,c,u\cdot}\mathbf{A}_{e,c,l\cdot}'\mathbf{F}_{e,s}'\mathbf{A}_{e,r,v\cdot}] + \mathbb{1}_{\{h=v\}}\mathbb{1}_{\{u=l\}}\,(\mathbf{\Sigma}_\epsilon)_{hu}^2\,\mathbb{E}[\boldsymbol{\epsilon}_{t,\cdot u}'\boldsymbol{\epsilon}_{s,\cdot l}]$$

$$= \mathbf{A}_{e,c,l\cdot}'\mathbf{A}_{e,c,u\cdot}\mathbf{A}_{e,r,h\cdot}'\mathbf{A}_{e,r,v\cdot} \cdot \sum_{w\ge 0} a_{e,w}a_{e,w+|t-s|} + \mathbb{1}_{\{h=v\}}\mathbb{1}_{\{u=l\}} \cdot (\mathbf{\Sigma}_\epsilon)_{hu}^2 \sum_{w\ge 0} a_{\epsilon,w}a_{\epsilon,w+|t-s|}.$$

Hence if we fix $t \in [T], h \in [p], u \in [q]$, then for any deterministic vectors $\mathbf{n} \in \mathbb{R}^{k_r}$ and

$\mathbf{g}, \mathbf{j} \in \mathbb{R}^{k_c}$, we have

$$\sum_{s=1}^{T}\sum_{v=1}^{p}\sum_{l=1}^{q}\mathbb{E}\big(E_{t,hu}\,\mathbf{n}'\mathbf{F}_t\mathbf{g}\cdot E_{s,vl}\,\mathbf{j}'\mathbf{F}_s'\mathbf{n}\big) = \sum_{s=1}^{T}\sum_{v=1}^{p}\sum_{l=1}^{q}\mathrm{Cov}(E_{t,hu}, E_{s,vl})\cdot\mathbb{E}\Big(\mathbf{n}'\mathbf{F}_t\mathbf{g}\mathbf{j}'\mathbf{F}_s'\mathbf{n}\Big)$$
$$= \sum_{v=1}^{p}\sum_{l=1}^{q}\Big\{O\big(\mathbf{A}_{e,c,l\cdot}'\mathbf{A}_{e,c,u\cdot}\mathbf{A}_{e,r,h\cdot}'\mathbf{A}_{e,r,v\cdot}\big)\cdot\sum_{w\geq 0}\sum_{m\geq 0}\sum_{s=1}^{T}a_{e,w}a_{e,w+|t-s|}a_{f,m}a_{f,m+|t-s|}$$
$$+ O\big(\mathbb{1}_{\{h=v\}}\mathbb{1}_{\{u=l\}}\cdot(\mathbf{\Sigma}_\epsilon)_{hu}^2\big)\cdot\sum_{w\geq 0}\sum_{m\geq 0}\sum_{s=1}^{T}a_{\epsilon,w}a_{\epsilon,w+|t-s|}a_{f,m}a_{f,m+|t-s|}\Big\}$$
$$= \sum_{v=1}^{p}\sum_{l=1}^{q}O\big(\mathbf{A}_{e,c,l\cdot}'\mathbf{A}_{e,c,u\cdot}\mathbf{A}_{e,r,h\cdot}'\mathbf{A}_{e,r,v\cdot} + \mathbb{1}_{\{h=v\}}\mathbb{1}_{\{u=l\}}\cdot(\mathbf{\Sigma}_\epsilon)_{hu}^2\big) = O(1),$$

where for the second last equality, we argue for the first term in the second last line only, as the second term could be shown similarly:

$$\sum_{w\geq 0}\sum_{m\geq 0}\sum_{s=1}^{T}a_{e,w}a_{e,w+|t-s|}a_{f,m}a_{f,m+|t-s|} = \sum_{w\geq 0}\sum_{m\geq 0}a_{e,w}a_{f,m}\sum_{s=1}^{T}a_{e,w+|t-s|}a_{f,m+|t-s|}$$
$$\leq \sum_{w\geq 0}\sum_{m\geq 0}|a_{e,w}|\,|a_{f,m}|\cdot\Big(\sum_{s=1}^{T}a_{e,w+|t-s|}^2\Big)^{\frac{1}{2}}\Big(\sum_{s=1}^{T}a_{f,m+|t-s|}^2\Big)^{\frac{1}{2}} \leq \sum_{w\geq 0}\sum_{m\geq 0}|a_{e,w}|\,|a_{f,m}| \leq c^2,$$

where the constant $c$ is from Assumptions (F1) and (E2). Finally,

$$\begin{aligned}\mathbb{E}\Big\{\Big(\sum_{h=1}^{p}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t')_{ih}\Big)^2\Big\} &= \mathbb{E}\Big\{\Big(\sum_{u=1}^{q}\sum_{h=1}^{p}\sum_{t=1}^{T}E_{t,hu}\mathbf{A}_{c,u\cdot}'\mathbf{F}_t'\mathbf{A}_{r,i\cdot}\Big)^2\Big\}\\ = \sum_{t=1}^{T}\sum_{h=1}^{p}\sum_{u=1}^{q}\sum_{s=1}^{T}\sum_{v=1}^{p}\sum_{l=1}^{q}&\mathbb{E}\Big(E_{t,hu}\mathbf{A}_{r,i\cdot}'\mathbf{F}_t\mathbf{A}_{c,u\cdot}\cdot E_{s,vl}\mathbf{A}_{c,l\cdot}'\mathbf{F}_s'\mathbf{A}_{r,i\cdot}\Big) = O(Tpq).\end{aligned} \tag{5.32}$$

Thus, we have

$$\|\mathcal{I}_6\|^2 \leq \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2\cdot\|\mathbf{H}_r\|_F^2\cdot\Big\|\sum_{i=1}^{p}\mathbf{Q}_{r,i\cdot}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p')_{ij}\Big\|^2$$
$$= \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2\cdot\|\mathbf{H}_r\|_F^2\cdot\Big\|\sum_{h=1}^{p}\sum_{i=1}^{p}\mathbf{Q}_{r,i\cdot}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t')_{ih}(\mathbf{1}_p\mathbf{1}_p')_{hj}\Big\|^2$$
$$\leq \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2\cdot\|\mathbf{H}_r\|_F^2\cdot\|\mathbf{Q}_r\|_F^2\cdot\sum_{i=1}^{p}\Big(\sum_{h=1}^{p}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t')_{ih}\Big)^2$$
$$= O_P\big(T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2\delta_{c,1}}\cdot Tp^2q\big) = O_P\big(T^{-1}p^{-2\delta_{r,k_r}}q^{1-2\delta_{c,1}}\big),$$

where we used Cauchy–Schwarz inequality in the third line, both Lemma 5.3 and (5.32) in the second last equality. Similarly, we have the following for $\mathcal{I}_7$,

$$\begin{aligned}
&\|\mathcal{I}_7\|^2 \leq \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^p \mathbf{Q}_{r,i\cdot} \sum_{t=1}^T (\mathbf{1}_p\mathbf{1}_p'\mathbf{E}_t\mathbf{C}_t')_{ij} \Big\|^2 \\
&= \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{h=1}^p \sum_{i=1}^p \mathbf{Q}_{r,i\cdot} \sum_{t=1}^T (\mathbf{1}_p\mathbf{1}_p')_{ih}(\mathbf{E}_t\mathbf{C}_t')_{hj} \Big\|^2 \\
&\leq \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \|\mathbf{Q}_r\|_F^2 \cdot p\Big\{ \sum_{h=1}^p \sum_{t=1}^T (\mathbf{C}_t\mathbf{E}_t')_{jh} \Big\}^2 \\
&= O_P\big(T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2\delta_{c,1}} \cdot Tp^2q\big) = O_P\big(T^{-1}p^{-2\delta_{r,k_r}}q^{1-2\delta_{c,1}}\big).
\end{aligned}$$

For $\mathcal{I}_8$ and $\mathcal{I}_9$, their rates can be shown to be the same as that for $\mathcal{I}_3$ by the following,

$$\begin{aligned}
&\|\mathcal{I}_8\|^2 \leq \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^p \mathbf{Q}_{r,i\cdot} \sum_{t=1}^T (\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p')_{ij} \Big\|^2 \\
&= \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{h=1}^p \sum_{i=1}^p \sum_{l=1}^q \sum_{t=1}^T \mathbf{Q}_{r,i\cdot} E_{t,il}E_{t,hl} \Big\|^2 \\
&\leq \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot p\sum_{h=1}^p \Big\| \sum_{i=1}^p \sum_{l=1}^q \sum_{t=1}^T \mathbf{Q}_{r,i\cdot} E_{t,il}E_{t,hl} \Big\|^2 \\
&= O_P\big(T^{-2}p^{-2-2\delta_{r,k_r}}q^{-2\delta_{c,1}} \cdot p^2(Tq + T^2q^2p^{-\delta_{r,k_r}})\big) \\
&= O_P\big(T^{-1}p^{-2\delta_{r,k_r}}q^{1-2\delta_{c,1}} + p^{-3\delta_{r,k_r}}q^{2-2\delta_{c,1}}\big),
\end{aligned}$$

where we used Lemma 5.3 and (5.31) in the second last equality. The proof for $\|\mathcal{I}_9\|^2$ is similar to the above by using the proof of (5.31) previously and omitted here.

For $\mathcal{I}_{10}$, first observe from the proof of (5.63), we also have for any $j \in [p]$,

$$\sum_{i=1}^p \mathbb{E}\Big\{ \Big( \sum_{t=1}^T \sum_{l=1}^q E_{t,il} \sum_{h=1}^q E_{t,jh} \Big)^2 \Big\} = O\big(Tpq^2 + T^2q^2\big),$$

then together with Lemma 5.3, it holds that

$$\begin{aligned}
&\|\mathcal{I}_{10}\|^2 \leq \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^p \mathbf{Q}_{r,i\cdot} \sum_{t=1}^T (\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t')_{ij} \Big\|^2 \\
&= \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^p \mathbf{Q}_{r,i\cdot} \sum_{t=1}^T \sum_{l=1}^q E_{t,il} \sum_{h=1}^q E_{t,jh} \Big\|^2 \\
&\leq \frac{1}{T^2p^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \|\mathbf{Q}_r\|_F^2 \cdot \sum_{i=1}^p \Big( \sum_{t=1}^T \sum_{l=1}^q E_{t,il} \sum_{h=1}^q E_{t,jh} \Big)^2 \\
&= O_P\big(T^{-1}p^{-1-2\delta_{r,k_r}}q^{2-2\delta_{c,1}} + p^{-2-2\delta_{r,k_r}}q^{2-2\delta_{c,1}}\big).
\end{aligned}$$

Consider now $\mathcal{I}_{11}$, we have

$$\begin{aligned}\|\mathcal{I}_{11}\|^2 &\leq \frac{1}{T^2p^4q^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^{p} \mathbf{Q}_{r,i\cdot} \sum_{t=1}^{T}(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\Big\|^2 \\ &\leq \frac{1}{T^2p^4q^2}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \|\mathbf{Q}_r\|_F^2 \cdot p\Big\{\sum_{t=1}^{T}(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\Big\}^2 \\ &= O_P\big(T^{-2}p^{-4-2\delta_{r,k_r}}q^{-2-2\delta_{c,1}} \cdot T^2p^3q^2\big) = O_P\big(p^{-1-2\delta_{r,k_r}}q^{-2\delta_{c,1}}\big),\end{aligned}$$

where we used Lemma 5.3 and the rate from (5.10) in the second last equality.

For $\mathcal{I}_{12}$, we have

$$\begin{aligned}\|\mathcal{I}_{12}\|^2 &\leq \frac{1}{T^2p^4}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \Big\| \sum_{i=1}^{p} \mathbf{Q}_{r,i\cdot} \sum_{t=1}^{T}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\Big\|^2 \\ &\leq \frac{1}{T^2p^4}\|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\mathbf{H}_r\|_F^2 \cdot \|\mathbf{Q}_r\|_F^2 \cdot p\Big(\sum_{t=1}^{T}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\Big)^2 \\ &= O_P\big(T^{-2}p^{-4-2\delta_{r,k_r}}q^{-2\delta_{c,1}} \cdot T^2p^3q^2\big) = O_P\big(p^{-1-2\delta_{r,k_r}}q^{2-2\delta_{c,1}}\big),\end{aligned}$$

where the last line used the following result which can be shown similar to (5.56),

$$\begin{aligned}&\mathbb{E}\Big\{\Big(\sum_{t=1}^{T}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\Big)^2\Big\} = \mathbb{E}\Big\{\Big(\sum_{t=1}^{T}\sum_{h=1}^{q}\sum_{i=1}^{p}\sum_{j=1}^{p}E_{t,ih}E_{t,jh}\Big)^2\Big\} \\ &= \sum_{t=1}^{T}\sum_{h=1}^{q}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{s=1}^{T}\sum_{l=1}^{q}\sum_{m=1}^{p}\sum_{n=1}^{p}\mathrm{Cov}(E_{t,ih}E_{t,jh}, E_{s,ml}E_{s,nl}) \\ &\quad + \Big(\sum_{t=1}^{T}\sum_{h=1}^{q}\sum_{i=1}^{p}\sum_{j=1}^{p}\mathbb{E}[E_{t,ih}E_{t,jh}]\Big)^2 = O\big(Tp^2q^2 + T^2p^2q^2\big) = O\big(T^2p^2q^2\big).\end{aligned}$$

Lastly, $\|\mathcal{I}_{13}\|^2$ is dominated by the terms from $\mathcal{I}_1$ to $\mathcal{I}_{12}$ using Cauchy–Schwarz inequality and Theorem 5.2. We require the term $\mathcal{I}_1$ to be truly dominating by using Assumption (AD1) and we equivalently compare the rates without the term $\widehat{\mathbf{D}}_r^{-1}$. Notice the rates for $\|\mathcal{I}_2\|^2$, $\|\mathcal{I}_4\|^2$, $\|\mathcal{I}_5\|^2$, $\|\mathcal{I}_6\|^2$, $\|\mathcal{I}_7\|^2$, $\|\mathcal{I}_{10}\|^2$, $\|\mathcal{I}_{11}\|^2$ and $\|\mathcal{I}_{12}\|^2$ are bounded above by the rate for $\|\mathcal{I}_3\|^2$ which is the same as $\|\mathcal{I}_8\|^2$ and $\|\mathcal{I}_9\|^2$. Thus, it suffices to consider the following ratio as $p, q, T \to \infty$,

$$\|\mathcal{I}_3\|^2/\|\mathcal{I}_1\|^2 = O_P\big(1/p^{\delta_{r,1}} + Tq/p^{\delta_{r,1}+\delta_{r,k_r}}\big) = o_P(1),$$

by the rate assumption $Tq = o(p^{\delta_{r,1}+\delta_{r,k_r}})$. Therefore, $\mathcal{I}_1$ is dominating over other terms in

(5.28) and hence we have

$$
\begin{aligned}
\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot} &= \mathcal{I}_1 + o_P(1) = \frac{1}{T}\widehat{\mathbf{D}}_r^{-1}\sum_{i=1}^{p}\mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t')_{ij} + o_P(1) \\
&\xrightarrow{p} \frac{1}{T}\mathbf{D}_r^{-1}\mathbf{H}_r^*\sum_{i=1}^{p}\mathbf{Q}_{r,i\cdot}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t')_{ij},
\end{aligned}
\tag{5.33}
$$

where the last line used Lemma 5.4 and Lemma 5.5 in which $\mathbf{D}_r$ and $\mathbf{H}_r^*$ are defined, respectively. In the rest of the proof, we show

$$
\sqrt{T\omega_B}\cdot\frac{1}{T}\mathbf{D}_r^{-1}\mathbf{H}_r^*\sum_{i=1}^{p}\mathbf{Q}_{r,i\cdot}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t')_{ij} \xrightarrow{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, T^{-1}\omega_B\cdot\mathbf{D}_r^{-1}\mathbf{H}_r^*\mathbf{\Xi}_{r,j}(\mathbf{H}_r^*)'\mathbf{D}_r^{-1}\big),
$$

with $\omega_B := p^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1}$ and $\mathbf{\Xi}_{r,j} := \operatorname{plim}_{p,q,T\to\infty}\operatorname{Var}\big(\sum_{i=1}^{p}\mathbf{Q}_{r,i\cdot}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t')_{ij}\big)$. We will adapt the central limit theorem for $\alpha$-mixing processes as depicted in Theorem 2.21 in Fan and Yao (2003). First, define $\mathbf{B}_{j,t} := \sqrt{\omega_B}\cdot\mathbf{D}_r^{-1}\mathbf{H}_r^*\sum_{i=1}^{p}\mathbf{Q}_{r,i\cdot}(\mathbf{C}_t\mathbf{E}_t')_{ij}$, and also let $\mathbf{b}_{e,t} := \sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}$, $\mathbf{b}_{\epsilon,il,t} := \sum_{w\geq 0}a_{\epsilon,w}(\mathbf{X}_{\epsilon,t-w})_{il}$ and $\mathbf{b}_{f,t} := \sum_{w\geq 0}a_{f,w}\mathbf{X}_{f,t-w}$ which are independent of each other by Assumption (E2).

Since we may write $\mathbf{B}_{j,t} = h(\mathbf{b}_{e,t}, (\mathbf{b}_{\epsilon,il,t})_{i\in[p],l\in[q]}, \mathbf{b}_{f,t})$ for some function $h$, we conclude $\mathbf{B}_{j,t}$ is $\alpha$-mixing using Theorem 5.2 in Bradley (2005). Observe that $\mathbb{E}[\mathbf{B}_{j,t}] = \mathbf{0}$, and we show in the following that there exists an $m > 2$ such that $\mathbb{E}[\|\mathbf{B}_{j,t}\|^m] \leq C$ for some constant $C$,

$$
\begin{aligned}
\mathbb{E}[\|\mathbf{B}_{j,t}\|^m] &\leq \omega_B^{m/2}\cdot\|\mathbf{D}_r^{-1}\|_F^m\cdot\|\mathbf{H}_r^*\|_F^m\cdot\|\mathbf{Q}_r\|_F^m\cdot\mathbb{E}\Big[\Big\{\sum_{i=1}^{p}\Big(\sum_{l=1}^{q}E_{t,jl}\mathbf{A}_{c,h\cdot}'\mathbf{F}_t'\mathbf{A}_{r,i\cdot}\Big)^2\Big\}^{m/2}\Big] \\
&= O\Big((p^{2\delta_{r,k_r}}q^{2\delta_{c,1}})^{m/2}\Big)\cdot\|\mathbf{D}_r^{-1}\|_F^m = O(1),
\end{aligned}
$$

where we used Lemma 5.1.2 and the definition of $\omega_B$ in the second last equality, and Lemma 5.4 in the last equality. Theorem 2.21 in Fan and Yao (2003) then applies, and hence

$$
\begin{aligned}
\sqrt{T\omega_B}\cdot\frac{1}{T}\mathbf{D}_r^{-1}\mathbf{H}_r^*\sum_{i=1}^{p}\mathbf{Q}_{r,i\cdot}\sum_{t=1}^{T}(\mathbf{C}_t\mathbf{E}_t')_{ij} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{B}_{j,t} \\
&\xrightarrow{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, T^{-1}\omega_B\cdot\mathbf{D}_r^{-1}\mathbf{H}_r^*\mathbf{\Xi}_{r,j}(\mathbf{H}_r^*)'\mathbf{D}_r^{-1}\big).
\end{aligned}
$$

Together with (5.33), we arrive at

$$
\begin{aligned}
&(Tp^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1})^{1/2}\cdot(\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}) \\
&\xrightarrow{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, T^{-1}p^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1}\cdot\mathbf{D}_r^{-1}\mathbf{H}_r^*\mathbf{\Xi}_{r,j}(\mathbf{H}_r^*)'\mathbf{D}_r^{-1}\big).
\end{aligned}
\tag{5.34}
$$

This completes the proof of Theorem 5.5. □

***Proof of Theorem 5.6.*** We only prove the scenario for $Tq = o(p^{\delta_{r,1}+\delta_{r,k_r}})$, which is on showing the constructed estimator for the row loading estimator is consistent. Note $\widehat{\mathbf{D}}_r$ consistently estimates $\mathbf{D}_r$ by Lemma 5.4, and $\mathbf{H}_r$ consistently estimates $\mathbf{H}_r^*$ by Lemma 5.5. Then

$$\mathbf{H}_r \mathrm{Var}\Big\{\sum_{i=1}^p \mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{C}_t\mathbf{E}_t')_{ij}\Big\}\mathbf{H}_a' = \mathrm{Var}\Big\{\sum_{i=1}^p \mathbf{H}_r\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{C}_t\mathbf{E}_t')_{ij}\Big\}$$
$$= \mathrm{Var}\Big\{\sum_{i=1}^p\Big(T^{-1}\widehat{\mathbf{D}}_r^{-1}\widehat{\mathbf{Q}}_r'\mathbf{Q}_r\sum_{s=1}^T\mathbf{F}_{Z,s}\mathbf{Q}_c'\mathbf{Q}_c\mathbf{F}_{Z,s}'\Big)\mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{C}_t\mathbf{E}_t')_{ij}\Big\}$$
$$= \mathrm{Var}\Big\{\sum_{t=1}^T\sum_{i=1}^p\Big(T^{-1}\widehat{\mathbf{D}}_r^{-1}\widehat{\mathbf{Q}}_r'\sum_{s=1}^T\mathbf{C}_s\mathbf{C}_{s,i\cdot}\Big)(\mathbf{C}_t\mathbf{E}_t')_{ij}\Big\}.$$

By Theorem 5.2, $(\widehat{\mu}_t, \widehat{\boldsymbol{\alpha}}_t, \widehat{\boldsymbol{\beta}}_t)$ is consistent for $(\mu_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t)$. By Theorem 5.3 and the rate assumption in the statement of Theorem 5.6, $\widehat{\mathbf{C}}_t$ is consistent for $\mathbf{C}_t$ and hence $\widehat{\mathbf{E}}_t$ is consistent for $\mathbf{E}_t$. Thus, we conclude that $\widehat{\boldsymbol{\Sigma}}_{r,j}^{HAC}$ is estimating $\mathbf{H}_r\boldsymbol{\Xi}_{r,j}\mathbf{H}_r'$ consistently (Newey and West (1987)) and hence result 1 is implied. Result 2 then follows, and results 3 and 4 can be shown similarly (details omitted). This completes the proof of the Theorem 5.6. □

***Proof of Theorem 5.7.*** Combining (5.4) and (5.6), we have

$$\begin{aligned}\widehat{\mathbf{E}}_t &= \widehat{\mathbf{L}}_t - \widehat{\mathbf{C}}_t = \mathbf{M}_p\mathbf{Y}_t\mathbf{M}_q - \widehat{\mathbf{C}}_t = \mathbf{M}_p\mathbf{C}_t\mathbf{M}_q + \mathbf{M}_p\mathbf{E}_t\mathbf{M}_q - \widehat{\mathbf{C}}_t \\ &= (\mathbf{C}_t - \widehat{\mathbf{C}}_t) + \mathbf{M}_p\mathbf{A}_{e,r}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'\mathbf{M}_q + \mathbf{M}_p(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)\mathbf{M}_q,\end{aligned}$$

where the second line used (IC1) being satisfied, so that $\mathbf{M}_p\mathbf{A}_r = \mathbf{A}_r$ and $\mathbf{M}_q\mathbf{A}_c = \mathbf{A}_c$. Hence

$$q^{-1}(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii} - q^{-1}\sum_{j=1}^q\Sigma_{\epsilon,ij}^2 = \sum_{i=1}^6 I_i, \text{ where}$$

$$\begin{aligned}I_1 &:= q^{-1}\{\mathbf{M}_p(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)\mathbf{M}_q(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)'\mathbf{M}_p\}_{ii} - q^{-1}\sum_{j=1}^q\Sigma_{\epsilon,ij}^2, \\ I_2 &:= q^{-1}\{(\widehat{\mathbf{C}}_t - \mathbf{C}_t)(\widehat{\mathbf{C}}_t - \mathbf{C}_t)'\}_{ii}, \\ I_3 &:= q^{-1}\{\mathbf{M}_p\mathbf{A}_{e,r}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'\mathbf{M}_q\mathbf{A}_{e,c}\mathbf{F}_{e,t}'\mathbf{A}_{e,r}'\mathbf{M}_p\}_{ii}, \\ I_4 &:= O_P(q^{-1}\{(\mathbf{C}_t - \widehat{\mathbf{C}}_t)\mathbf{M}_q\mathbf{A}_{e,c}\mathbf{F}_{e,t}'\mathbf{A}_{e,r}'\mathbf{M}_p\}_{ii}), \\ I_5 &:= O_P(q^{-1}\{(\mathbf{C}_t - \widehat{\mathbf{C}}_t)\mathbf{M}_q(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)'\mathbf{M}_p\}_{ii}), \\ I_6 &:= O_P(q^{-1}\{\mathbf{M}_p\mathbf{A}_{e,r}\mathbf{F}_{e,t}\mathbf{A}_{e,c}'\mathbf{M}_q(\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)'\mathbf{M}_p\}_{ii}).\end{aligned} \tag{5.35}$$

By an assumption in the statement of the theorem, we have $I_2 = o_P(q^{-1})$. Since $\|\mathbf{M}_p\mathbf{A}_{e,r}\|_1 \leq \|\mathbf{M}_p\|_1\|\mathbf{A}_{e,r}\|_1 = O(1)$, with the finiteness of $k_r$ and $k_c$, we have $I_3, I_6 = O_P(q^{-1})$, and by the Cauchy–Schwarz inequality, $I_4 = O_P(I_2^{1/2}\, I_6^{1/2}) = o_P(q^{-1})$. Writing $\eta_{t,ij} := (\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t)_{ij}$, we can expand

$$
\begin{aligned}
I_1 &= \frac{1}{q}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2(\epsilon_{t,ij}^2-1) - \frac{2}{q}\sum_{j=1}^{q}\Big(\frac{1}{p}\sum_{\ell=1}^{p}\eta_{t,\ell j}\Big)\eta_{t,ij} \\
&\quad + \frac{2}{q}\Big(\frac{\mathbf{1}_p'\boldsymbol{\eta}_t\mathbf{1}_q}{pq}\Big)\sum_{j=1}^{q}\eta_{t,ij} + \frac{1}{q}\sum_{j=1}^{q}\Big(\frac{1}{p}\sum_{i=1}^{p}\eta_{t,ij}\Big)^2 - \Big(\frac{\mathbf{1}_p'\boldsymbol{\eta}_t\mathbf{1}_q}{pq}\Big)^2 - \Big(\frac{1}{q}\sum_{j=1}^{q}\eta_{t,ij}\Big)^2, \\
&= \frac{1}{q}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2(\epsilon_{t,ij}^2-1) - \frac{2}{pq}\sum_{j=1}^{q}\eta_{t,ij}^2 - \frac{2}{pq}\sum_{j=1}^{q}\sum_{\ell\neq i}\eta_{t,\ell j}\eta_{t,ij} \\
&\quad + O_P(q^{-1}p^{-1/2}) + O_P(p^{-1}) + O_P(p^{-1}q^{-1}) + O_P(q^{-1}) \\
&= \frac{1}{q}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2(\epsilon_{t,ij}^2-1)(1+o_P(1)),
\end{aligned}
\tag{5.36}
$$

where all rates of convergence above are obtained from applying the Markov inequality. Hence

$$
\begin{aligned}
\frac{(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii} - \sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^4\text{Var}(\epsilon_{t,ij}^2)}} &= \frac{q^{-1}(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii} - q^{-1}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2}{\sqrt{q^{-2}\sum_{j=1}^{q}\text{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}} \\
&= \frac{q^{-1}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2(\epsilon_{t,ij}^2-1)(1+o_P(1))}{\sqrt{q^{-2}\sum_{j=1}^{q}\text{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}} \\
&\xrightarrow{\mathcal{D}} Z_{i,t} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),
\end{aligned}
$$

where the last line follows from Theorem 1 in Ayvazyan and Ulyanov (2023), with

$$
Z_{i,t} := \frac{q^{-1}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2(\epsilon_{t,ij}^2-1)}{\sqrt{q^{-2}\sum_{j=1}^{q}\text{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}}, \quad i \in [p],
$$

so that we can easily see that the $Z_{i,t}$'s are independent of each other by Assumption (E1). The proof for $(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii}$ completes since from the calculations for $I_1$, we see that

$$
I_5 = O_P(I_2^{1/2}\cdot 1) = o_P(q^{-1/2}).
$$

For $(\check{\mathbf{E}}_t\check{\mathbf{E}}_t')_{ii}$ under $H_0$, note that we have

$$
\check{\mathbf{E}}_t = (\mathbf{C}_t - \check{\mathbf{C}}_t) + \mathbf{A}_{e,r}\mathbf{F}_{e,t}\mathbf{A}_{e,c}' + (\boldsymbol{\Sigma}_\epsilon * \boldsymbol{\epsilon}_t),
$$

with the rate for $\check{C}_{t,ij} - C_{t,ij}$ the same as that for $\widehat{C}_{t,ij} - C_{t,ij}$ since the estimation procedure for FM is essentially the same with the same assumptions on the factor loadings (apart from (IC1) which is not important for FM), the factors and the noise. Hence the proof we employed so far can be replicated with $\mathbf{M}_p$ and $\mathbf{M}_q$ replaced by the corresponding sized identity matrices, and we arrive at the same conclusion with the same $Z_{i,t}$'s.

For $(\widehat{\mathbf{E}}_t'\widehat{\mathbf{E}}_t)_{jj}$ under both $H_0$ and $H_1$ and $(\check{\mathbf{E}}_t'\check{\mathbf{E}}_t)_{jj}$ under $H_0$, the proofs are parallel to that for $(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii}$, and we omit them here. □

***Proof of Theorem 5.8.*** We provide the details of proof for the inequality concerning $\mathbb{P}_{y,\alpha}(y_{\alpha,t} \geq \widehat{q}_{y,\alpha}(\theta))$. The other one involves similar details and its proof is omitted.

Firstly, using triangle inequality, consider

$$\begin{aligned}\max_{t\in[T]}|x_{\alpha,t} - y_{\alpha,t}| &\leq \max_{i\in[p],t\in[T]}|q^{-1}(\widehat{E}_t\widehat{E}_t')_{ii} - q^{-1}(\check{E}_t\check{E}_t')_{ii}| \\ &\leq \max_{i\in[p],t\in[T]}|I_1^*| + \sum_{\ell=2}^{6}\max_{i\in[p],t\in[T]}(|I_\ell| + |\widetilde{I}_\ell|),\end{aligned}$$

where $I_\ell$, $\ell = 2,\ldots,6$ is defined as in (5.35), and $\widetilde{I}_\ell$ for $\ell = 2,\ldots,6$ is defined exactly the same as $I_\ell$, except that $\mathbf{M}_p$ and $\mathbf{M}_q$ in the definitions of the $I_\ell$'s are replaced by identity matrices of appropriate sizes, and $\widehat{\mathbf{C}}_t$ is replaced by $\check{\mathbf{C}}_t$. The definition of $I_1^*$ is the same as $I_1$ in (5.35), except that the term $q^{-1}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2(\epsilon_{t,ij}^2 - 1)$ is absent in (5.36).

With pervasive factors, the uniform error rates for $\widehat{\mathbf{C}}_t$ (and $\check{\mathbf{C}}_t$, since the form of $\check{\mathbf{C}}_t$ is the same as that in (5.5), with rates for $\check{\mathbf{Q}}_r$ and $\check{\mathbf{Q}}_c$ similar to $\widehat{\mathbf{Q}}_r$ and $\widehat{\mathbf{Q}}_c$ respectively) from Lemma 5.9 is

$$\begin{aligned}&\max_{i\in[p],j\in[q],t\in[T]}\{|C_{t,ij} - \widehat{C}_{t,ij}|,\ |C_{t,ij} - \check{C}_{t,ij}|\} \\ &= O_P\{((pq)^{-1/2} + (Tq)^{-1/2} + (Tp)^{-1/2} + p^{-1} + q^{-1})\log(T)\log(p)\log(q)\}. \qquad (5.37)\end{aligned}$$

From the proof of Theorem 5.7, we can see that $I_1^*$ has faster convergence rate than other terms, and in fact $I_5$ (and hence $\widetilde{I}_5$) has the slowest rate of convergence for a fixed indices $t \in [T]$ and $i \in [p]$. Taking maximum over all possible indices, using Assumption (E3) and (5.37), we have

$$\begin{aligned}&\max_{t\in[T]}|x_{\alpha,t} - y_{\alpha,t}| = O_P(I_5) = O_P(\max_{t\in[T],i\in[p],j\in[q]}|C_{t,ij} - \widehat{C}_{t,ij}| \cdot \max_{t\in[T],i\in[p],j\in[q]}|(\mathbf{\Sigma}_{\boldsymbol{\epsilon}} \circ \boldsymbol{\epsilon}_t)_{ij}|) \\ &= O_P\{((pq)^{-1/2} + (Tq)^{-1/2} + (Tp)^{-1/2} + p^{-1} + q^{-1})\log^2(T)\log^2(p)\log^2(q)\}. \qquad (5.38)\end{aligned}$$

Next, we assess the approximate "gap" size of the $x_{\alpha,t}$'s over $t \in [T]$. To this end, using Theorem 5.7, and the fact that $\sum_{j=1}^{q}\text{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4$ has order $q$ uniformly over $i \in [p]$, we have

$$x_{\alpha,t} \asymp_P \max_{i\in[p]}\Big\{q^{-1}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2 + \frac{1}{\sqrt{q}}\max_{i\in[p]}Z_{i,t}\Big\} \asymp_P \max_{i\in[p]}\Big\{q^{-1}\sum_{j=1}^{q}\Sigma_{\epsilon,ij}^2\Big\} + \frac{\log(p)}{\sqrt{q}}, \qquad (5.39)$$

showing that the "gap" between two ordered $x_{\alpha,t}$'s is $O_P\{\log(p)/(Tq^{1/2})\}$.

With the "gap" size and uniform error in (5.38), consider

$$\begin{aligned}
\sup_{c\in\mathbb{R}} |\mathbb{F}_{y,\alpha}(c) - \mathbb{F}_{x,\alpha}(c)| &= \sup_{c\in\mathbb{R}} \Big| \frac{1}{T}\sum_{t=1}^T \big[\mathbf{1}\{y_{\alpha,t} \le c\} - \mathbf{1}\{x_{\alpha,t} \le c\}\big]\Big| \\
&\le \sup_{c\in\mathbb{R}} \Big| \frac{1}{T}\sum_{t=1}^T \big[\mathbf{1}\{x_{\alpha,t} \pm \max_{t\in[T]} |x_{\alpha,t} - y_{\alpha,t}| \le c\} - \mathbf{1}\{x_{\alpha,t} \le c\}\big]\Big| \\
&= O_P\Big(\frac{1}{T}\max_{t\in[T]} |x_{\alpha,t} - y_{\alpha,t}| \Big/ \Big(\frac{\log(p)}{T\sqrt{q}}\Big)\Big) \\
&= O_P\Big\{\Big(\frac{1}{\sqrt{p}} + \frac{1}{\sqrt{q}} + \frac{1}{\sqrt{T}} + \frac{\sqrt{q}}{p} + \sqrt{\frac{q}{Tp}}\Big)\log^2(T)\log(p)\log^2(q)\Big\},
\end{aligned}$$

where the last line used (5.38). Hence in particular,

$$\begin{aligned}
&\mathbb{P}_{y,\alpha}(y_{\alpha,t} \le \widehat{q}_{x,\alpha}(\theta)) = \mathbb{F}_{y,\alpha}(\widehat{q}_{x,\alpha}(\theta)) \ge \mathbb{F}_{x,\alpha}(\widehat{q}_{x,\alpha}(\theta)) - \sup_{c\in\mathbb{R}} |\mathbb{F}_{y,\alpha}(c) - \mathbb{F}_{x,\alpha}(c)| \\
&\ge \theta + O_P\Big\{\Big(\frac{1}{\sqrt{p}} + \frac{1}{\sqrt{q}} + \frac{1}{\sqrt{T}} + \frac{\sqrt{q}}{p} + \sqrt{\frac{q}{Tp}}\Big)\log^2(T)\log(p)\log^2(q)\Big\},
\end{aligned}$$

which is the result we want. □

***Proof of Theorem 5.9.*** First consider $\widehat{k}_r$, i.e., result 1 in Theorem 5.9. For $j \in [k_r]$,

$$\begin{aligned}
&\lambda_j\Big(\frac{1}{T}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\Big) = \lambda_j\Big(\frac{1}{T}\sum_{t=1}^T \mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'\mathbf{A}_c\mathbf{F}_t'\mathbf{A}_r'\Big) = \lambda_j\Big(\mathbf{A}_r'\mathbf{A}_r \cdot \frac{1}{T}\sum_{t=1}^T \mathbf{F}_t\mathbf{A}_c'\mathbf{A}_c\mathbf{F}_t'\Big) \\
&\asymp_P \operatorname{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\lambda_j(\mathbf{A}_r'\mathbf{A}_r) = \|\mathbf{A}_c\|_F^2 \cdot \lambda_j(\mathbf{\Sigma}_{A,r}^{1/2}\mathbf{Z}_r\mathbf{\Sigma}_{A,r}^{1/2}) \asymp q^{\delta_{c,1}}\cdot\lambda_j(\mathbf{Z}_r) = p^{\delta_{r,j}}q^{\delta_{c,1}},
\end{aligned} \tag{5.40}$$

where in the second line, we used Assumption (F1) in the first step and Assumption (L1) in the second. For the second last step, we used Theorem 1 of Ostrowski (1959) on the eigenvalues of a congruent transformation $\mathbf{\Sigma}_{A,r}^{1/2}\mathbf{Z}_r\mathbf{\Sigma}_{A,r}^{1/2}$ of $\mathbf{Z}_r$, where we further used Assumption (L1) that all eigenvalues of $\mathbf{\Sigma}_{A,r}$ are bounded away from 0 and infinity.

Since we have $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t' = T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t' + T^{-1}\sum_{t=1}^T \mathbf{R}_{r,t}$ from (5.22), it holds by Weyl's inequality that for $j \in [k_r]$,

$$\Big|\lambda_j\Big(\frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\Big) - \lambda_j\Big(\frac{1}{T}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\Big)\Big| \le \Big\|\frac{1}{T}\sum_{t=1}^T \mathbf{R}_{r,t}\Big\| = o_P(\omega_r), \tag{5.41}$$

where $\omega_r = p^{\delta_{r,k_r}}q^{\delta_{c,1}}$ is defined in Lemma 5.3, and the last equality used from the proof of Lemma 5.3 that $\boldsymbol{\gamma}'(T^{-1}\sum_{t=1}^T \mathbf{R}_{r,t})\boldsymbol{\gamma} = o_P(\omega_r)$ for any unit vector $\boldsymbol{\gamma} \in \mathbb{R}^p$. With our choice of $\xi_r$, we also have

$$\xi_r/\omega_r \asymp p^{1-\delta_{r,k_r}} q^{1-\delta_{c,1}} \big[(Tq)^{-1/2} + p^{-1/2}\big] = o(1), \tag{5.42}$$

where we used Assumption (R1) in the last equality. Hence for $k_r > 1$, if $j \in [k_r - 1]$, using (5.41) and (5.42) we have

$$\begin{aligned}
&\frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r}{\lambda_j\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r} \\
&\le \frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\big) + \xi_r + \big|\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) - \lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\big)\big|}{\lambda_j\big(T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\big) + \xi_r - \big|\lambda_j\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) - \lambda_j\big(T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\big)\big|} \\
&= \frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\big) + o_P(\omega_r)}{\lambda_j\big(T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\big) + o_P(\omega_r)} = \frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\big)}{\lambda_j\big(T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\big)}\big(1 + o_P(1)\big) \asymp_P p^{\delta_{r,j+1}-\delta_{r,j}},
\end{aligned} \tag{5.43}$$

where the last line used (5.40). Moreover, for any $j \in [k_r - 1]$,

$$\begin{aligned}
&\frac{\lambda_{k_r+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r}{\lambda_{k_r}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r} = \frac{\lambda_{k_r+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r}{\omega_r(1 + o_P(1))} \\
&= O_P\Big\{\frac{\lambda_{k_r+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big)}{\omega_r} + \frac{\xi_r}{\omega_r}\Big\} \\
&= O_P\Big\{O_P\Big(pq[(Tq)^{-1/2} + p^{-1/2}]/\omega_r\Big) + \xi_r/\omega_r\Big\} = O_P(\xi_r/\omega_r) = o_P(p^{\delta_{r,j+1}-\delta_{r,j}}),
\end{aligned} \tag{5.44}$$

where we used (5.42) and the proof of Lemma 5.3 in the first equality, our choice of $\xi_r$ in the second last, and the extra rate assumption in the statement of the theorem in the last. In the third equality, we used the following (which will be shown at the end of the this proof),

$$\lambda_j\Big(\frac{1}{T}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\Big) = O_P(T^{-1/2}pq^{1/2} + p^{1/2}q), \quad j = k_r + 1, \dots, p. \tag{5.45}$$

Hence for $j = k_r + 1, \dots, \lfloor p/2 \rfloor$ (true also for $k_r = 1$),

$$\frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r}{\lambda_j\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r} \ge \frac{\xi_r/\omega_r}{O_P(\xi_r/\omega_r) + \xi_r/\omega_r} \ge \frac{1}{C} \tag{5.46}$$

in probability for some generic positive constant $C$, where we used again our choice of $\xi_r$ and (5.45) in the first inequality. Combining (5.43), (5.44) and (5.46), we may conclude our proposed $\widehat{k}_r$ is consistent for $k_r$.

If $k_r = 1$, then from our choice of $\xi_r$, (5.44) becomes

$$\frac{\lambda_{k_r+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r}{\lambda_{k_r}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r} = O_P(\xi_r/\omega_r) = o_P(1).$$

Together with (5.46) which holds true for $k_r = 1$, we also conclude $\widehat{k}_r$ is consistent for $k_r$.

It remains to show (5.45). To this end, from (5.40) and (5.41), the first $k_r$ eigenvalues of $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$ coincides with those of $T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'$ asymptotically, so that the first $k_r$ eigenvectors corresponding to $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$ coincides with those for $T^{-1}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'$ asymptotically as $T, p \to \infty$, which are necessarily in $\mathcal{N}^\perp := \text{Span}(\mathbf{Q}_r)$, the linear span of the columns of $\mathbf{Q}_r$. This means that the $(k_r + 1)$-th largest eigenvalue of $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$ and beyond will asymptotically have eigenvectors in $\mathcal{N}$, the orthogonal complement of $\mathcal{N}^\perp$. Then for any unit vectors $\boldsymbol{\gamma} \in \mathcal{N}$, we have from (5.64) and Lemma 5.2 that

$$\begin{aligned}
\boldsymbol{\gamma}'\Big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\Big)\boldsymbol{\gamma} &= \boldsymbol{\gamma}'\Big(\frac{1}{T}\sum_{t=1}^T \mathbf{R}_{r,t}\Big)\boldsymbol{\gamma} \\
&= O_P\Big(T^{-1}\Big\|\sum_{t=1}^T \mathbf{R}_{r,t}\Big\|_F\Big) = O_P(T^{-1/2}pq^{1/2} + p^{1/2}q),
\end{aligned}$$

which is equivalent to (5.45). This completes the proof of Theorem 5.9. □

As we have the same factor structure as Cen and Lam (2025b), we state Lemma 5.1 below for further use and refer readers to Cen and Lam (2025b) for the proof in detail.

**Lemma 5.1** *Let Assumptions (F1), (E1) and (E2) hold. Then*

1. *(Weak correlation of noise $\mathbf{E}_t$ across different rows, columns and times). there exists some positive constant $C < \infty$ so that for any $t \in [T], i, j \in [p], h \in [q]$,*

$$\sum_{k=1}^p \sum_{l=1}^q \Big|\mathbb{E}[E_{t,ih}E_{t,kl}]\Big| \le C,$$
$$\sum_{l=1}^q \sum_{s=1}^T \Big|\text{cov}(E_{t,ih}E_{t,jh}, E_{s,il}E_{s,jl})\Big| \le C.$$

2. *(Weak dependence between factor $\mathbf{F}_t$ and noise $\mathbf{E}_t$). there exists some positive constant $C < \infty$ so that for any $j \in [p], i \in [q]$, and any deterministic vectors $\mathbf{u} \in \mathbb{R}^{k_r}$ and $\mathbf{v} \in \mathbb{R}^{k_c}$ with constant magnitudes,*

$$\mathbb{E}\Big\{\frac{1}{(qT)^{1/2}}\sum_{h=1}^q\sum_{t=1}^T E_{t,jh}\mathbf{u}'\mathbf{F}_t\mathbf{v}\Big\}^2 \le C, \quad \mathbb{E}\Big\{\frac{1}{(pT)^{1/2}}\sum_{h=1}^p\sum_{t=1}^T E_{t,hi}\mathbf{v}'\mathbf{F}_t'\mathbf{u}\Big\}^2 \le C.$$

3. *(Further results on factor $\mathbf{F}_t$). for any $t \in [T]$, all elements in $\mathbf{F}_t$ are independent of each other, with mean 0 and unit variance. Moreover,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t' \xrightarrow{p} \mathbf{\Sigma}_r := k_c\mathbf{I}_{k_r}, \quad \frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t'\mathbf{F}_t \xrightarrow{p} \mathbf{\Sigma}_c := k_r\mathbf{I}_{k_c},$$

*with the number of factors $k_r$ and $k_c$ fixed as $\min\{T, p, q\} \to \infty$.*

**Lemma 5.2** *(Bounding $\sum_{t=1}^{T}\mathbf{R}_{r,t}$). Under Assumptions (F1), (L1), (E1) and (E2), we have*

$$\Big\|\sum_{t=1}^{T}\mathbf{C}_t\mathbf{E}_t'\Big\|_F^2 = O_P(Tp^{1+\delta_{r,1}}q), \tag{5.47}$$

$$\Big\|\sum_{t=1}^{T}\mathbf{E}_t\mathbf{E}_t'\Big\|_F^2 = O_P(Tp^2q + T^2pq^2), \tag{5.48}$$

$$\Big\|\sum_{t=1}^{T}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p'\Big\|_F^2 = O_P(Tp^3q^2 + T^2p^2q^2), \tag{5.49}$$

$$\Big\|\sum_{t=1}^{T}\mathbf{C}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 = O_P(Tp^{3+\delta_{r,1}}q), \tag{5.50}$$

$$\Big\|\sum_{t=1}^{T}\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 = O_P(Tp^4q + T^2p^3q^2), \tag{5.51}$$

$$\Big\|\sum_{t=1}^{T}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\Big\|_F^2 = O_P(Tp^2q^2 + T^2pq^2), \tag{5.52}$$

$$\Big\|\sum_{t=1}^{T}(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 = O_P(T^2p^4q^2), \tag{5.53}$$

$$\Big\|\sum_{t=1}^{T}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 = O_P(Tp^6q + T^2p^5q^2). \tag{5.54}$$

*Thus, with $\mathbf{R}_{r,t}$ defined in* (5.22)*, we have*

$$\Big\|\sum_{t=1}^{T}\mathbf{R}_{r,t}\Big\|_F^2 = O_P(Tp^2q + T^2pq^2).$$

***Proof of Lemma 5.2.*** Using $\mathbf{C}_t = \mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'$, we have (5.47) holds as follows,

$$\Big\|\sum_{t=1}^{T}\mathbf{C}_t\mathbf{E}_t'\Big\|_F^2 = \Big\|\sum_{t=1}^{T}\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'\mathbf{E}_t'\Big\|_F^2 = \sum_{i=1}^{p}\sum_{l=1}^{p}\Big(\sum_{t=1}^{T}\mathbf{A}_{r,i\cdot}'\mathbf{F}_t\mathbf{A}_c'\mathbf{E}_{t,l\cdot}\Big)^2$$
$$= \sum_{i=1}^{p}\|\mathbf{A}_{r,i\cdot}\|^2\cdot\sum_{l=1}^{p}\Big(\sum_{h=1}^{q}\sum_{t=1}^{T}E_{t,lh}\frac{1}{\|\mathbf{A}_{r,i\cdot}\|}\mathbf{A}_{r,i\cdot}'\mathbf{F}_t\mathbf{A}_{c,h\cdot}\Big)^2 = O_P\Big(Tp^{1+\delta_{r,1}}q\Big),$$

where the last equality is from Assumption (L1) and Lemma 5.1.

To show (5.48), first notice from Assumption (E1),

$$E_{t,ij} = \mathbf{A}'_{e,r,i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,j\cdot} + \Sigma_{\epsilon,ij}\epsilon_{t,ij}.$$

With Assumption (E2), we have

$$\text{Cov}(E_{t,ij}, E_{t,kj}) = \mathbf{A}'_{e,r,i\cdot}\mathbf{A}_{e,r,k\cdot}\|\mathbf{A}_{e,c,j\cdot}\|^2 + \Sigma^2_{\epsilon,ij}\mathbb{1}_{\{i=k\}},$$

and hence using Lemma 5.1,

$$\begin{aligned}
&\mathbb{E}\Big(\Big\|\sum_{t=1}^T \mathbf{E}_t\mathbf{E}_t'\Big\|_F^2\Big) = \sum_{i=1}^p\sum_{k=1}^p \mathbb{E}\Big\{\Big(\sum_{t=1}^T\sum_{j=1}^q E_{t,ij}E_{t,kj}\Big)^2\Big\} \\
&= \sum_{i=1}^p\sum_{k=1}^p\Big\{\sum_{t=1}^T\sum_{j=1}^q\sum_{s=1}^T\sum_{l=1}^q \text{Cov}(E_{t,ij}E_{t,kj}, E_{s,il}E_{s,kl}) + \Big(\sum_{t=1}^T\sum_{j=1}^q \mathbb{E}[E_{t,ij}E_{t,kj}]\Big)^2\Big\} \\
&= O(Tp^2q) + \sum_{i=1}^p\sum_{k=1}^p O\Big(T\cdot\mathbf{A}'_{e,r,i\cdot}\mathbf{A}_{e,r,k\cdot}\|\mathbf{A}_{e,c}\|_F^2 + Tq\cdot\mathbb{1}_{\{i=k\}}\Big)^2 = O(Tp^2q + T^2pq^2).
\end{aligned}$$

For (5.49), consider first

$$\begin{aligned}
&\sum_{t=1}^T\sum_{s=1}^T \text{Cov}(E_{t,ij}E_{t,kh}, E_{s,lm}E_{s,kn}) \\
&= \sum_{t=1}^T\sum_{s=1}^T \text{Cov}\Big\{(\mathbf{A}'_{e,r,i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,j\cdot} + \Sigma_{\epsilon,ij}\epsilon_{t,ij})(\mathbf{A}'_{e,r,k\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,h\cdot} + \Sigma_{\epsilon,kh}\epsilon_{t,kh}), \qquad (5.55) \\
&\qquad (\mathbf{A}'_{e,r,l\cdot}\mathbf{F}_{e,s}\mathbf{A}_{e,c,m\cdot} + \Sigma_{\epsilon,lm}\epsilon_{s,lm})(\mathbf{A}'_{e,r,k\cdot}\mathbf{F}_{e,s}\mathbf{A}_{e,c,n\cdot} + \Sigma_{\epsilon,kn}\epsilon_{s,kn})\Big\} \\
&= \sum_{t=1}^T\sum_{s=1}^T \text{Cov}\Big(\mathbf{A}'_{e,r,i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,h\cdot}, \mathbf{A}'_{e,r,l\cdot}\mathbf{F}_{e,s}\mathbf{A}_{e,c,m\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{F}_{e,s}\mathbf{A}_{e,c,n\cdot}\Big) \\
&\quad + \sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\Big(\mathbf{A}'_{e,r,i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,l\cdot}\mathbf{F}_{e,s}\mathbf{A}_{e,c,m\cdot}\Big)\cdot\mathbb{E}\Big(\Sigma_{\epsilon,kh}\epsilon_{t,kh}\Sigma_{\epsilon,kn}\epsilon_{s,kn}\Big) \\
&\quad + \sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\Big(\mathbf{A}'_{e,r,i\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{F}_{e,s}\mathbf{A}_{e,c,n\cdot}\Big)\cdot\mathbb{E}\Big(\Sigma_{\epsilon,kh}\epsilon_{t,kh}\Sigma_{\epsilon,lm}\epsilon_{s,lm}\Big) \\
&\quad + \sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\Big(\mathbf{A}'_{e,r,k\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,l\cdot}\mathbf{F}_{e,s}\mathbf{A}_{e,c,m\cdot}\Big)\cdot\mathbb{E}\Big(\Sigma_{\epsilon,ij}\epsilon_{t,ij}\Sigma_{\epsilon,kn}\epsilon_{s,kn}\Big) \\
&\quad + \sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\Big(\mathbf{A}'_{e,r,k\cdot}\mathbf{F}_{e,t}\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{F}_{e,s}\mathbf{A}_{e,c,n\cdot}\Big)\cdot\mathbb{E}\Big(\Sigma_{\epsilon,ij}\epsilon_{t,ij}\Sigma_{\epsilon,lm}\epsilon_{s,lm}\Big)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\Big(\Sigma_{\epsilon,ij}\epsilon_{t,ij}\Sigma_{\epsilon,kh}\epsilon_{t,kh},\Sigma_{\epsilon,lm}\epsilon_{s,lm}\Sigma_{\epsilon,kn}\epsilon_{s,kn}\Big)\\
=&\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\Big\{\mathbf{A}'_{e,r,i\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,h\cdot},\\
&\qquad\mathbf{A}'_{e,r,l\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,m\cdot}\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,n\cdot}\Big\}\\
&+\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\Big\{\mathbf{A}'_{e,r,i\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,l\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,m\cdot}\Big\}\\
&\qquad\cdot\mathbb{E}\Big(\Sigma_{\epsilon,kh}\epsilon_{t,kh}\Sigma_{\epsilon,kn}\epsilon_{s,kn}\Big)\\
&+\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\Big\{\mathbf{A}'_{e,r,i\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,n\cdot}\Big\}\\
&\qquad\cdot\mathbb{E}\Big(\Sigma_{\epsilon,kh}\epsilon_{t,kh}\Sigma_{\epsilon,lm}\epsilon_{s,lm}\Big)\\
&+\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\Big\{\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,l\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,m\cdot}\Big\}\\
&\qquad\cdot\mathbb{E}\Big(\Sigma_{\epsilon,ij}\epsilon_{t,ij}\Sigma_{\epsilon,kn}\epsilon_{s,kn}\Big)\\
&+\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\Big\{\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,n\cdot}\Big\}\\
&\qquad\cdot\mathbb{E}\Big(\Sigma_{\epsilon,ij}\epsilon_{t,ij}\Sigma_{\epsilon,lm}\epsilon_{s,lm}\Big)\\
&+\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\Big(\Sigma_{\epsilon,ij}\epsilon_{t,ij}\Sigma_{\epsilon,kh}\epsilon_{t,kh},\Sigma_{\epsilon,lm}\epsilon_{s,lm}\Sigma_{\epsilon,kn}\epsilon_{s,kn}\Big).
\end{aligned}
\tag{5.56}
$$

Consider the six terms in the last equality above, we have the first term as

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\Big\{\mathbf{A}'_{e,r,i\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,h\cdot},\\
&\mathbf{A}'_{e,r,l\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,m\cdot}\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\geq 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,n\cdot}\Big\}\\
=&\sum_{t=1}^{T}\sum_{w\geq 0}a_{e,w}^2\cdot\mathbb{E}\Big(\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,n\cdot}\Big)\\
&\qquad\cdot\mathbb{E}\Big(\mathbf{A}'_{e,r,k\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,l\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,m\cdot}\Big)\\
&+\sum_{t=1}^{T}\sum_{w\geq 0}a_{e,w}^2\cdot\mathbb{E}\Big(\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,l\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,m\cdot}\Big)\\
&\qquad\cdot\mathbb{E}\Big(\mathbf{A}'_{e,r,k\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,n\cdot}\Big)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{t=1}^{T}\sum_{w\ge 0} a_{e,w}^4\cdot \mathbb{E}\Big(\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\\
&\qquad\qquad\cdot \mathbf{A}'_{e,r,l\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,m\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,n\cdot}\Big)\\
&-\sum_{t=1}^{T}\sum_{w\ge 0} a_{e,w}^4\cdot \mathbb{E}\Big(\mathbf{A}'_{e,r,i\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,h\cdot}\Big)\\
&\qquad\qquad\cdot \mathbb{E}\Big(\mathbf{A}'_{e,r,l\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,m\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{X}_{e,t-w}\mathbf{A}_{e,c,n\cdot}\Big)\\
&= O\Big(\sum_{t=1}^{T}\mathbf{A}'_{e,r,k\cdot}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,i\cdot}\mathbf{A}_{e,c,n\cdot}\mathbf{A}'_{e,r,l\cdot}\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{A}_{e,c,m\cdot}\\
&\quad+\sum_{t=1}^{T}\mathbf{A}'_{e,r,l\cdot}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,i\cdot}\mathbf{A}_{e,c,m\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{A}_{e,c,n\cdot}\\
&\quad+\sum_{t=1}^{T}\sum_{w\ge 0} a_{e,w}^4\,\|\mathbf{A}_{e,r,k\cdot}\|^2\,\|\mathbf{A}_{e,r,i\cdot}\|\,\|\mathbf{A}_{e,c,j\cdot}\|\,\|\mathbf{A}_{e,c,h\cdot}\|\,\|\mathbf{A}_{e,r,l\cdot}\|\,\|\mathbf{A}_{e,c,m\cdot}\|\,\|\mathbf{A}_{e,c,n\cdot}\|\\
&\quad-\sum_{t=1}^{T}\sum_{w\ge 0} a_{e,w}^4\cdot \mathbf{A}'_{e,r,k\cdot}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,i\cdot}\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,k\cdot}\mathbf{A}_{e,c,m\cdot}\mathbf{A}'_{e,r,l\cdot}\mathbf{A}_{e,c,n\cdot}\Big),
\end{aligned}
\tag{5.57}
$$

where we used (E2) in the last equality that each entry in $\{\mathbf{X}_{e,t}\}$ is independent with uniformly bounded fourth moment. Similarly, the remaining terms in last equality of (5.56) are

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\Big\{\mathbf{A}'_{e,r,i\cdot}\Big(\sum_{w\ge 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,l\cdot}\Big(\sum_{w\ge 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,m\cdot}\Big\}\\
&\qquad\cdot \mathbb{E}\Big(\Sigma_{\epsilon,kh}\epsilon_{t,kh}\Sigma_{\epsilon,kn}\epsilon_{s,kn}\Big)\\
&=\sum_{t=1}^{T}\mathbf{A}'_{e,r,l\cdot}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,i\cdot}\mathbf{A}_{e,c,m\cdot}\cdot\Sigma_{\epsilon,kh}\Sigma_{\epsilon,kn}\cdot \mathbb{1}_{\{h=n\}},
\end{aligned}
\tag{5.58}
$$

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\Big\{\mathbf{A}'_{e,r,i\cdot}\Big(\sum_{w\ge 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\ge 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,n\cdot}\Big\}\\
&\qquad\cdot \mathbb{E}\Big(\Sigma_{\epsilon,kh}\epsilon_{t,kh}\Sigma_{\epsilon,lm}\epsilon_{s,lm}\Big)\\
&=\sum_{t=1}^{T}\mathbf{A}'_{e,r,k\cdot}\mathbf{A}_{e,c,j\cdot}\mathbf{A}'_{e,r,i\cdot}\mathbf{A}_{e,c,n\cdot}\cdot\Sigma_{\epsilon,kh}\Sigma_{\epsilon,lm}\cdot \mathbb{1}_{\{k=l\}}\mathbb{1}_{\{h=m\}},
\end{aligned}
\tag{5.59}
$$

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\Big\{\mathbf{A}'_{e,r,k\cdot}\Big(\sum_{w\ge 0}a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,h\cdot}\mathbf{A}'_{e,r,l\cdot}\Big(\sum_{w\ge 0}a_{e,w}\mathbf{X}_{e,s-w}\Big)\mathbf{A}_{e,c,m\cdot}\Big\}\\
&\qquad\cdot \mathbb{E}\Big(\Sigma_{\epsilon,ij}\epsilon_{t,ij}\Sigma_{\epsilon,kn}\epsilon_{s,kn}\Big)
\end{aligned}
$$

$$= \sum_{t=1}^{T} \mathbf{A}'_{e,r,l\cdot} \mathbf{A}_{e,c,h\cdot} \mathbf{A}'_{e,r,k\cdot} \mathbf{A}_{e,c,m\cdot} \cdot \Sigma_{\epsilon,ij} \Sigma_{\epsilon,kn} \cdot \mathbb{1}_{\{i=k\}} \mathbb{1}_{\{j=n\}}, \tag{5.60}$$

$$\sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}\Big\{ \mathbf{A}'_{e,r,k\cdot} \Big( \sum_{w \geq 0} a_{e,w} \mathbf{X}_{e,t-w} \Big) \mathbf{A}_{e,c,h\cdot} \mathbf{A}'_{e,r,k\cdot} \Big( \sum_{w \geq 0} a_{e,w} \mathbf{X}_{e,s-w} \Big) \mathbf{A}_{e,c,n\cdot} \Big\}$$
$$\cdot \mathbb{E}\Big( \Sigma_{\epsilon,ij} \epsilon_{t,ij} \Sigma_{\epsilon,lm} \epsilon_{s,lm} \Big)$$
$$= \sum_{t=1}^{T} \mathbf{A}'_{e,r,k\cdot} \mathbf{A}_{e,c,h\cdot} \mathbf{A}'_{e,r,k\cdot} \mathbf{A}_{e,c,n\cdot} \cdot \Sigma_{\epsilon,ij} \Sigma_{\epsilon,lm} \cdot \mathbb{1}_{\{i=l\}} \mathbb{1}_{\{j=m\}}, \tag{5.61}$$

$$\sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{Cov}\Big( \Sigma_{\epsilon,ij} \epsilon_{t,ij} \Sigma_{\epsilon,kh} \epsilon_{t,kh}, \Sigma_{\epsilon,lm} \epsilon_{s,lm} \Sigma_{\epsilon,kn} \epsilon_{s,kn} \Big)$$
$$= O(T) \cdot \Big( \mathbb{1}_{\{i=l=k\}} \mathbb{1}_{\{j=h=m=n\}} + \mathbb{1}_{\{i=l\}} \mathbb{1}_{\{j=m\}} \mathbb{1}_{\{h=n\}} + \mathbb{1}_{\{i=l=k\}} \mathbb{1}_{\{j=n\}} \mathbb{1}_{\{h=m\}} \Big). \tag{5.62}$$

Using (5.57), (5.58), (5.59), (5.60), (5.61) and (5.62), we arrive at an expression for (5.56). Thus, (5.49) can be obtained as

$$\mathbb{E}\Big( \Big\| \sum_{t=1}^{T} \mathbf{1}'_q \mathbf{E}'_t \mathbf{1}_p \mathbf{E}_t \mathbf{1}_q \mathbf{1}'_p \Big\|_F^2 \Big) = p \sum_{k=1}^{p} \mathbb{E}\Big\{ \Big[ \sum_{t=1}^{T} \Big( \sum_{i=1}^{p} \sum_{j=1}^{q} E_{t,ij} \Big) \sum_{h=1}^{q} E_{t,kh} \Big]^2 \Big\}$$
$$= p \sum_{k=1}^{p} \Big\{ \sum_{t=1}^{T} \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{h=1}^{q} \sum_{s=1}^{T} \sum_{l=1}^{p} \sum_{m=1}^{q} \sum_{n=1}^{q} \mathrm{Cov}(E_{t,ij} E_{t,kh}, E_{s,lm} E_{s,kn})$$
$$+ \Big( \sum_{t=1}^{T} \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{h=1}^{q} \mathbb{E}[E_{t,ij} E_{t,kh}] \Big)^2 \Big\}$$
$$= O(Tp^3q^2) + p \sum_{k=1}^{p} \Big\{ \sum_{t=1}^{T} \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{h=1}^{q} \Big( \mathbf{A}'_{e,r,i\cdot} \mathbf{A}_{e,r,k\cdot} \mathbf{A}'_{e,c,j\cdot} \mathbf{A}_{e,c,h\cdot} + \Sigma^2_{\epsilon,ij} \mathbb{1}_{\{i=k\}} \mathbb{1}_{\{j=h\}} \Big) \Big\}^2$$
$$= O(Tp^3q^2 + T^2p^2q^2). \tag{5.63}$$

By (5.47) and (5.48), we can obtain (5.50) and (5.51), respectively as follows,

$$\Big\| \sum_{t=1}^{T} \mathbf{C}_t \mathbf{E}'_t \mathbf{1}_p \mathbf{1}'_p \Big\|_F^2 \leq \Big\| \sum_{t=1}^{T} \mathbf{C}_t \mathbf{E}'_t \Big\|_F^2 \cdot \|\mathbf{1}_p \mathbf{1}'_p\|_F^2 = O_P(Tp^{3+\delta_{r,1}} q),$$
$$\Big\| \sum_{t=1}^{T} \mathbf{E}_t \mathbf{E}'_t \mathbf{1}_p \mathbf{1}'_p \Big\|_F^2 \leq \Big\| \sum_{t=1}^{T} \mathbf{E}_t \mathbf{E}'_t \Big\|_F^2 \cdot \|\mathbf{1}_p \mathbf{1}'_p\|_F^2 = O_P(Tp^4 q + T^2 p^3 q^2).$$

Similar to the proof of (5.49), we can show (5.52) by

$$\begin{aligned}
&\mathbb{E}\Big(\Big\|\sum_{t=1}^{T}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\Big\|_F^2\Big) = \sum_{i=1}^{p}\sum_{k=1}^{p}\mathbb{E}\Big\{\Big[\sum_{t=1}^{T}\Big(\sum_{j=1}^{q}E_{t,ij}\Big)\Big(\sum_{h=1}^{q}E_{t,kh}\Big)\Big]^2\Big\}\\
&= \sum_{i=1}^{p}\sum_{k=1}^{p}\Big\{\sum_{t=1}^{T}\sum_{j=1}^{q}\sum_{h=1}^{q}\sum_{s=1}^{T}\sum_{m=1}^{q}\sum_{n=1}^{q}\mathrm{Cov}(E_{t,ij}E_{t,kh}, E_{s,im}E_{s,kn})\\
&\qquad + \Big(\sum_{t=1}^{T}\sum_{j=1}^{q}\sum_{h=1}^{q}\mathbb{E}[E_{t,ij}E_{t,kh}]\Big)^2\Big\}\\
&= O(Tp^2q^2) + \sum_{i=1}^{p}\sum_{k=1}^{p}\Big\{\sum_{t=1}^{T}\sum_{j=1}^{q}\sum_{h=1}^{q}\big(\mathbf{A}_{e,r,i\cdot}'\mathbf{A}_{e,r,k\cdot}\mathbf{A}_{e,c,j\cdot}'\mathbf{A}_{e,c,h\cdot} + \Sigma_{\epsilon,ij}^2\mathbb{1}_{\{i=k\}}\mathbb{1}_{\{j=h\}}\big)\Big\}^2\\
&= O(Tp^2q^2 + T^2pq^2).
\end{aligned}$$

From (5.10), we can obtain (5.53) such that

$$\Big\|\sum_{t=1}^{T}(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 = \Big\{\sum_{t=1}^{T}(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\Big\}^2\cdot\|\mathbf{1}_p\mathbf{1}_p'\|_F^2 = O_P(T^2p^4q^2).$$

Lastly, from (5.48) we have

$$\Big\|\sum_{t=1}^{T}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 \le \|\mathbf{1}_p\|^2\cdot\|\mathbf{1}_p\|^2\cdot\Big\|\sum_{t=1}^{T}\mathbf{E}_t\mathbf{E}_t'\Big\|_F^2\cdot\|\mathbf{1}_p\mathbf{1}_p'\|_F^2 = O_P(Tp^6q + T^2p^5q^2).$$

From (5.22), we have

$$\begin{aligned}
\Big\|\sum_{t=1}^{T}\mathbf{R}_{r,t}\Big\|_F^2 &= O_P\Bigg(\Big\|\sum_{t=1}^{T}\mathbf{C}_t\mathbf{E}_t'\Big\|_F^2 + \Big\|\sum_{t=1}^{T}\mathbf{E}_t\mathbf{E}_t'\Big\|_F^2 + (pq)^{-2}\Big\|\sum_{t=1}^{T}\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p\mathbf{E}_t\mathbf{1}_q\mathbf{1}_p'\Big\|_F^2\\
&\quad + p^{-2}\Big\|\sum_{t=1}^{T}\mathbf{C}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 + p^{-2}\Big\|\sum_{t=1}^{T}\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 + p^{-2}\Big\|\sum_{t=1}^{T}\mathbf{E}_t\mathbf{1}_q\mathbf{1}_q'\mathbf{E}_t'\Big\|_F^2\\
&\quad + (pq)^{-2}p^{-2}\Big\|\sum_{t=1}^{T}(\mathbf{1}_q'\mathbf{E}_t'\mathbf{1}_p)^2\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2 + p^{-4}\Big\|\sum_{t=1}^{T}\mathbf{1}_p'\mathbf{E}_t\mathbf{E}_t'\mathbf{1}_p\mathbf{1}_p\mathbf{1}_p'\Big\|_F^2\Bigg)\\
&= O_P(Tp^2q + T^2pq^2),
\end{aligned}$$

which completes the proof of Lemma 5.2. □

**Lemma 5.3** *Let Assumptions (M1), (F1), (L1), (E1), (E2) and (R1) hold. Then define* $\omega_r := p^{\delta_{r,k_r}}q^{\delta_{c,1}}$ *and* $\omega_c := q^{\delta_{c,k_c}}p^{\delta_{r,1}}$*. We have*

$$\|\widehat{\mathbf{D}}_r^{-1}\|_F = O_P(\omega_r^{-1}), \quad \|\widehat{\mathbf{D}}_c^{-1}\|_F = O_P(\omega_c^{-1}).$$

***Proof of Lemma 5.3.*** It suffices to show the bound of $\|\widehat{\mathbf{D}}_r^{-1}\|_F$, since that of $\|\widehat{\mathbf{D}}_c^{-1}\|_F$ would be similar. First, we bound the term $\|\widehat{\mathbf{D}}_r^{-1}\|_F^2$ by finding the lower bound of $\lambda_{k_r}(\widehat{\mathbf{D}}_r)$. To do this, consider the decomposition

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t' = \frac{1}{T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t' + \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_{r,t}, \tag{5.64}$$

so that for a unit vector $\boldsymbol{\gamma}\in\mathbb{R}^p$, we can define

$$S_r(\boldsymbol{\gamma}) := \frac{1}{\omega_r}\boldsymbol{\gamma}'\Big(\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\Big)\boldsymbol{\gamma} =: S_r^*(\boldsymbol{\gamma}) + \widetilde{S}_r(\boldsymbol{\gamma}), \quad \text{with}$$
$$S_r^*(\boldsymbol{\gamma}) := \frac{1}{\omega_r}\boldsymbol{\gamma}'\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t'\Big)\boldsymbol{\gamma}, \quad \widetilde{S}_r(\boldsymbol{\gamma}) := \frac{1}{\omega_r}\boldsymbol{\gamma}'\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_{r,t}\Big)\boldsymbol{\gamma}.$$

Since $\|\boldsymbol{\gamma}\| = 1$, we have by Lemma 5.2,

$$|\widetilde{S}_r(\boldsymbol{\gamma})|^2 \le \frac{1}{\omega_r^2T^2}\Big\|\sum_{t=1}^{T}\mathbf{R}_{r,t}\Big\|_F^2 = O_P\Big(T^{-1}p^{2(1-\delta_{r,k_r})}q^{1-2\delta_{c,1}} + p^{1-2\delta_{r,k_r}}q^{2(1-\delta_{c,1})}\Big) = o_P(1),$$

where the last equality used Assumption (R1). Next, with Assumption (F1), consider

$$\lambda_{k_r}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t'\Big) = \lambda_{k_r}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'\mathbf{A}_c\mathbf{F}_t'\mathbf{A}_r'\Big) \ge \lambda_{k_r}(\mathbf{A}_r'\mathbf{A}_r)\cdot\lambda_{k_r}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{A}_c'\mathbf{A}_c\mathbf{F}_t'\Big)$$
$$\asymp_P p^{\delta_{r,k_r}}\cdot\lambda_{k_r}(\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)\boldsymbol{\Sigma}_r) \asymp_P p^{\delta_{r,k_r}}q^{\delta_{c,1}} = \omega_r.$$

With this, going back to the decomposition (5.64),

$$\omega_r^{-1}\lambda_{k_r}(\widehat{\mathbf{D}}_r) = \omega_r^{-1}\lambda_{k_r}\Big(\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\Big) \ge \omega_r^{-1}\lambda_{k_r}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t'\Big) - \sup_{\|\boldsymbol{\gamma}\|=1}|\widetilde{S}_r(\boldsymbol{\gamma})| \asymp_P 1,$$

and hence finally $\big\|\widehat{\mathbf{D}}_r^{-1}\big\|_F = O_P\big(\lambda_{k_r}^{-1}(\widehat{\mathbf{D}}_r)\big) = O_P(\omega_r^{-1})$, which completes the proof of Lemma 5.3. □

**Lemma 5.4** *(Limit of $\widehat{\mathbf{D}}_r$ and $\widehat{\mathbf{D}}_c$). Let Assumptions (F1), (L1), (E1), (E2) and (R1) hold. With $\widehat{\mathbf{D}}_r, \widehat{\mathbf{D}}_c$ and $\omega_r, \omega_c$ defined in Lemma 5.3, we have*

$$\omega_r^{-1}\widehat{\mathbf{D}}_r \xrightarrow{p} \omega_r^{-1}\mathbf{D}_r := \omega_r^{-1}\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\mathrm{diag}\{\lambda_j(\mathbf{A}_r'\mathbf{A}_r) \mid j\in[k_r]\},$$
$$\omega_c^{-1}\widehat{\mathbf{D}}_c \xrightarrow{p} \omega_c^{-1}\mathbf{D}_c := \omega_c^{-1}\mathrm{tr}(\mathbf{A}_r'\mathbf{A}_r)\cdot\mathrm{diag}\{\lambda_j(\mathbf{A}_c'\mathbf{A}_c) \mid j\in[k_c]\}.$$

***Proof of Lemma 5.4.*** It suffices to show the limit of $\widehat{\mathbf{D}}_r$, as the proof for $\widehat{\mathbf{D}}_c$ will be similar.

We first show

$$\frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t' \xrightarrow{p} \text{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}\mathbf{A}_r\mathbf{A}_r', \tag{5.65}$$

where $\lambda_{k_r}(\text{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}\mathbf{A}_r\mathbf{A}_r')\asymp 1$. By Assumption (F1), we have

$$\mathbb{E}\Big(\frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t'\Big) = \frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbb{E}(\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'\mathbf{A}_c\mathbf{F}_t'\mathbf{A}_r') = \text{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}\mathbf{A}_r\mathbf{A}_r',$$

which used the independence structure among elements in $\mathbf{F}_t$. Furthermore, for any $i,j\in[p]$,

$$\begin{aligned}
&\text{Var}\Big\{\Big(\frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t'\Big)_{ij}\Big\}\\
&= \frac{1}{\omega_r^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\text{Cov}\Big\{\mathbf{A}_{r,i\cdot}'\Big(\sum_{w\geq 0}a_{f,w}\mathbf{X}_{f,t-w}\Big)\mathbf{A}_c'\mathbf{A}_c\Big(\sum_{w\geq 0}a_{f,w}\mathbf{X}_{f,t-w}'\Big)\mathbf{A}_{r,j\cdot},\\
&\qquad\qquad \mathbf{A}_{r,i\cdot}'\Big(\sum_{w\geq 0}a_{f,w}\mathbf{X}_{f,s-w}\Big)\mathbf{A}_c'\mathbf{A}_c\Big(\sum_{w\geq 0}a_{f,w}\mathbf{X}_{f,s-w}'\Big)\mathbf{A}_{r,j\cdot}\Big\}\\
&= \frac{1}{\omega_r^2T^2}\sum_{t=1}^{T}\sum_{w\geq 0}\sum_{l\geq 0}a_{f,w}^2a_{f,l}^2\cdot\text{Var}\Big(\mathbf{A}_{r,i\cdot}'\mathbf{X}_{f,t-w}\mathbf{A}_c'\mathbf{A}_c\mathbf{X}_{f,t-l}'\mathbf{A}_{r,j\cdot}\Big)\\
&= \frac{1}{\omega_r^2T^2}\sum_{t=1}^{T}\sum_{w\geq 0}\sum_{l\geq 0}a_{f,w}^2a_{f,l}^2\cdot O(\|\mathbf{A}_c\|_F^4) = O(T^{-1}p^{-2\delta_{r,k_r}}) = o(1),
\end{aligned}$$

where we used Assumption (F1) in the third last equality, and both (L1) and (F1) in the second last, which concludes (5.65). Then it holds that

$$\begin{aligned}
&\Big\|\frac{1}{\omega_r T}\sum_{t=1}^{T}\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t' - \text{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}\mathbf{A}_r\mathbf{A}_r'\Big\|_F^2\\
&\leq 2\cdot\Big\|\frac{1}{\omega_r T}\sum_{t=1}^{T}\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t' - \frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t'\Big\|_F^2 + 2\cdot\Big\|\frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t' - \text{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}\mathbf{A}_r\mathbf{A}_r'\Big\|_F^2\\
&= 2\cdot\Big\|\frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbf{R}_{r,t}\Big\|_F^2 + 2\cdot\Big\|\frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbf{C}_t\mathbf{C}_t' - \text{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}\mathbf{A}_r\mathbf{A}_r'\Big\|_F^2\\
&= O_P\Big(T^{-1}p^{2(1-\delta_{r,k_r})}q^{1-2\delta_{c,1}} + p^{1-2\delta_{r,k_r}}q^{2(1-\delta_{c,1})}\Big) + o_P(1) = o_P(1),
\end{aligned}$$

where the second last equality used Lemma 5.2 and (5.65), and the last used Assumption (R1). Using the inequality that for the $i$-th eigenvalue of matrices $\widehat{\mathbf{A}}$ and $\mathbf{A}$, $|\lambda_i(\widehat{\mathbf{A}}) - \lambda_i(\mathbf{A})| \leq \|\widehat{\mathbf{A}} - \mathbf{A}\| \leq \|\widehat{\mathbf{A}} - \mathbf{A}\|_F$, we have for any $i\in[k_r]$,

$$|(\omega_r^{-1}\widehat{\mathbf{D}}_r)_{ii} - (\omega_r^{-1}\mathbf{D}_r)_{ii}| = o_P(1).$$

Thus, $\omega_r^{-1}\widehat{\mathbf{D}}_r \xrightarrow{p} \omega_r^{-1}\mathbf{D}_r$. This completes the proof of Lemma 5.4. □

**Lemma 5.5** *(Limit of $\mathbf{H}_r$ and $\mathbf{H}_c$). Under Assumptions (F1), (L1), (E1), (E2), (R1) and (L2),*

$$\mathbf{H}_r \xrightarrow{p} \mathbf{H}_r^* := (\text{tr}(\mathbf{A}_c'\mathbf{A}_c))^{1/2} \cdot \mathbf{D}_r^{-1/2}(\mathbf{\Gamma}_r^*)'\mathbf{Z}_r^{1/2},$$
$$\mathbf{H}_c \xrightarrow{p} \mathbf{H}_c^* := (\text{tr}(\mathbf{A}_r'\mathbf{A}_r))^{1/2} \cdot \mathbf{D}_c^{-1/2}(\mathbf{\Gamma}_c^*)'\mathbf{Z}_c^{1/2},$$

*where $\mathbf{D}_r, \mathbf{D}_c$ are defined in Lemma 5.4, and $\mathbf{\Gamma}_r^*, \mathbf{\Gamma}_c^*$ are the eigenvector matrices of* $\text{tr}(\mathbf{A}_c'\mathbf{A}_c) \cdot \omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{\Sigma}_{A,r}\mathbf{Z}_r^{1/2}$ *and* $\text{tr}(\mathbf{A}_r'\mathbf{A}_r) \cdot \omega_c^{-1}\mathbf{Z}_c^{1/2}\mathbf{\Sigma}_{A,c}\mathbf{Z}_c^{1/2}$*, respectively.*

***Proof of Lemma 5.5.*** The proof of the two limits are similar, and hence we only show the probability limit of $\mathbf{H}_r$ is $\mathbf{H}_r^*$. First, left-multiply $\omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{Q}_r'$ on (5.24), we have

$$\begin{aligned}(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)(\omega_r^{-1}\widehat{\mathbf{D}}_r) &= \omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\Big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\Big)\widehat{\mathbf{Q}}_r \\ &= \Big(\frac{1}{T}\sum_{t=1}^T \omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c'\mathbf{Q}_c\mathbf{F}_{Z,t}'\mathbf{Z}_r^{-1/2} + \mathbf{R}_{r,\text{res}}\Big)(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r),\end{aligned}$$

where $\mathbf{R}_{r,\text{res}} := T^{-1}\sum_{t=1}^T \omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\mathbf{R}_{r,t}\widehat{\mathbf{Q}}_r(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)^{-1}$. This implies each column of $(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)$ is an eigenvector of the matrix $\big(T^{-1}\sum_{t=1}^T \omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c'\mathbf{Q}_c\mathbf{F}_{Z,t}'\mathbf{Z}_r^{-1/2} + \mathbf{R}_{r,\text{res}}\big)$. Note that

$$\begin{aligned}&\omega_r^{-1}\Big\{\big(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r\big)'\big(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r\big) - \text{tr}(\mathbf{A}_c'\mathbf{A}_c)^{-1}\cdot\mathbf{D}_r\Big\} \\ &= \omega_r^{-1}\widehat{\mathbf{Q}}'\mathbf{A}_r'\mathbf{A}_r\widehat{\mathbf{Q}} - \frac{1}{\text{tr}(\mathbf{A}_c'\mathbf{A}_c)}\widehat{\mathbf{Q}}'\Big(\frac{1}{\omega_r T}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\Big)\widehat{\mathbf{Q}} \\ &\quad + \frac{1}{\text{tr}(\mathbf{A}_c'\mathbf{A}_c)}\widehat{\mathbf{Q}}'\Big(\frac{1}{\omega_r T}\sum_{t=1}^T \mathbf{C}_t\mathbf{C}_t'\Big)\widehat{\mathbf{Q}} - \text{tr}(\mathbf{A}_c'\mathbf{A}_c)^{-1}\cdot\mathbf{D}_r,\end{aligned}$$

whose Frobenius norm is $o_P(1)$ by (5.65) and Lemma 5.4. We arrive at

$$(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)'(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r) \xrightarrow{p} \text{tr}(\mathbf{A}_c'\mathbf{A}_c)^{-1}\cdot\mathbf{D}_r.$$

Thus, the eigenvalues of $(\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)'(\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)$ are asymptotically bounded away from zero and infinity by Assumption (L1), and hence $\|(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)^{-1}\|_F = O_P(\|\mathbf{Z}_r^{-1/2}\|_F)$. Therefore,

$$\begin{aligned}\|\mathbf{R}_{r,\text{res}}\|_F^2 &= O_P(1)\cdot\Big\|\frac{1}{\omega_r T}\sum_{t=1}^T \mathbf{R}_{r,t}\Big\|_F^2 \cdot \|\mathbf{Z}_r^{1/2}\|_F^2 \cdot \|(\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)^{-1}\|_F^2 \\ &= O_P\Big(T^{-1}p^{2(1-\delta_{r,k_r})}q^{1-2\delta_{c,1}} + p^{1-2\delta_{r,k_r}}q^{2(1-\delta_{c,1})}\Big) = o_P(1),\end{aligned} \tag{5.66}$$

where we used Lemma 5.2 in the second last equality and Assumption (R1) in the last.

Denote the following as the normalisation of $\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r$,

$$\mathbf{\Gamma}_r := (\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c))^{1/2} \cdot (\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\widehat{\mathbf{Q}}_r)\mathbf{D}_r^{-1/2}.$$

Using the limit in (5.65), we have

$$\frac{1}{T}\sum_{t=1}^{T}\omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\mathbf{Q}_r\mathbf{F}_{Z,t}\mathbf{Q}_c'\mathbf{Q}_c\mathbf{F}_{Z,t}'\mathbf{Z}_r^{-1/2} = \frac{1}{\omega_r T}\sum_{t=1}^{T}\mathbf{A}_r'\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'\mathbf{A}_c\mathbf{F}_t'(\mathbf{A}_r'\mathbf{A}_r)(\mathbf{A}_r'\mathbf{A}_r)^{-1}$$
$$\xrightarrow{p} \mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}(\mathbf{A}_r'\mathbf{A}_r)(\mathbf{A}_r'\mathbf{A}_r)(\mathbf{A}_r'\mathbf{A}_r)^{-1} = \mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{\Sigma}_{A,r}\mathbf{Z}_r^{1/2}.$$

By (5.66), Assumption (L2) and eigenvector perturbation theory, there exists a unique eigenvector matrix $\mathbf{\Gamma}_r^*$ of $\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\omega_r^{-1}\mathbf{Z}_r^{1/2}\mathbf{\Sigma}_{A,r}\mathbf{Z}_r^{1/2}$ such that $\|\mathbf{\Gamma}_r^* - \mathbf{\Gamma}_r\| = o_P(1)$. Thus,

$$\widehat{\mathbf{Q}}_r'\mathbf{Q}_r = (\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c))^{-1/2}\mathbf{D}_r^{1/2}\mathbf{\Gamma}_r'\mathbf{Z}_r^{-1/2} \xrightarrow{p} (\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c))^{-1/2}\mathbf{D}_r^{1/2}(\mathbf{\Gamma}_r^*)'\mathbf{Z}_r^{-1/2},$$

and hence using (5.65) again, we obtain

$$\mathbf{H}_r = \widehat{\mathbf{D}}_r^{-1}\widehat{\mathbf{Q}}_r'\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'\mathbf{A}_c\mathbf{F}_t'\mathbf{A}_r'\Big)\mathbf{Q}_r\mathbf{\Sigma}_{A,r}^{-1} \xrightarrow{p} \mathbf{D}_r^{-1}\widehat{\mathbf{Q}}_r'\Big(\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\mathbf{A}_r\mathbf{A}_r'\Big)\mathbf{Q}_r\mathbf{\Sigma}_{A,r}^{-1}$$
$$= \mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)\cdot\mathbf{D}_r^{-1}\widehat{\mathbf{Q}}_r'\mathbf{Q}_r\mathbf{Z}_r^{1/2}\mathbf{Z}_r^{1/2}\mathbf{Q}_r'\mathbf{Q}_r\mathbf{\Sigma}_{A,r}^{-1} \xrightarrow{p} (\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c))^{1/2}\cdot\mathbf{D}_r^{-1/2}(\mathbf{\Gamma}_r^*)'\mathbf{Z}_r^{1/2},$$

which completes the proof of Lemma 5.5. □

**Lemma 5.6** *Under the assumptions in Theorem 5.3, for any $j\in[p], l\in[q]$,*

$$\|\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}\|^2 = O_P\big(T^{-1}p^{\delta_{r,1}-2\delta_{r,k_r}}q^{1-2\delta_{c,1}} + p^{-3\delta_{r,k_r}}q^{2-2\delta_{c,1}}\big), \tag{5.67}$$
$$\|\widehat{\mathbf{Q}}_{c,l\cdot} - \mathbf{H}_c\mathbf{Q}_{c,l\cdot}\|^2 = O_P\big(T^{-1}q^{\delta_{c,1}-2\delta_{c,k_c}}p^{1-2\delta_{r,1}} + q^{-3\delta_{c,k_c}}p^{2-2\delta_{r,1}}\big). \tag{5.68}$$

***Proof of Lemma 5.6.*** For $\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}$, first consider the case when $Tq = o(p^{\delta_{r,k_r}+\delta_{r,1}})$. From (5.34) in the proof of Theorem 5.5, we have

$$\|\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}\|^2 = O_P\big((Tp^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1})^{-1}\big). \tag{5.69}$$

Now suppose $Tq = o(p^{\delta_{r,k_r}+\delta_{r,1}})$ fails to hold. From the decomposition of $\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}$ in (5.28), the leading term among the expressions will be $\mathcal{I}_3$. It has rate

$$\|\mathcal{I}_3\|^2 = O_P\big(T^{-1}p^{-2\delta_{r,k_r}}q^{1-2\delta_{c,1}} + p^{-3\delta_{r,k_r}}q^{2-2\delta_{c,1}}\big) = O_P\big(p^{-3\delta_{r,k_r}}q^{2-2\delta_{c,1}}\big),$$

where the last equality used the fact that $Tq = o\big(p^{\delta_{r,k_r}+\delta_{r,1}}\big)$ does not hold. Thus we have

$$\|\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}\|^2 = O_P\big(p^{-3\delta_{r,k_r}}q^{2-2\delta_{c,1}}\big). \tag{5.70}$$

Combining (5.69) and (5.70), we arrive at the statement of (5.67). The proof for (5.68) is similar and omitted here, which ends of the proof of Lemma 5.6. □

**Lemma 5.7** *Let Assumptions (F1), (E1), (E2) and (E3) hold. Then for any deterministic vectors* $\mathbf{u} \in \mathbb{R}^{k_r}$ *and* $\mathbf{v} \in \mathbb{R}^{k_c}$ *with constant magnitudes,*

$$\max_{j\in[p]} \frac{1}{(qT)^{1/2}} \sum_{h=1}^{q}\sum_{t=1}^{T} E_{t,jh}\mathbf{u}'\mathbf{F}_t\mathbf{v} = O_P(\log(p)),$$
$$\max_{i\in[q]} \frac{1}{(pT)^{1/2}} \sum_{h=1}^{p}\sum_{t=1}^{T} E_{t,hi}\mathbf{v}'\mathbf{F}_t'\mathbf{u} = O_P(\log(q)).$$

***Proof of Lemma 5.7.*** We only show the first result as the proof for the second is similar. From Assumption (F1), (E1) and (E2), we may rewrite $E_{t,jh}$ and $\mathbf{u}'\mathbf{F}_t\mathbf{v}$ as

$$E_{t,jh} = \mathbf{A}'_{e,r,j\cdot}\Big(\sum_{w\geq 0} a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,h\cdot} + \Sigma_{\epsilon,jh}\Big(\sum_{g\geq 0} a_{\epsilon,g}X_{\epsilon,t-g,jh}\Big),$$
$$\mathbf{u}'\mathbf{F}_t\mathbf{v} = \sum_{m=1}^{k_r}\sum_{n=1}^{k_c}\sum_{l\geq 0} a_{f,l}X_{f,t-l,mn}u_m v_n,$$

so that for any $j \in [p]$,

$$\begin{aligned}\sum_{h=1}^{q}\sum_{t=1}^{T} E_{t,jh}\mathbf{u}'\mathbf{F}_t\mathbf{v} &= \sum_{h=1}^{q}\sum_{t=1}^{T}\mathbf{A}'_{e,r,j\cdot}\Big(\sum_{w\geq 0} a_{e,w}\mathbf{X}_{e,t-w}\Big)\mathbf{A}_{e,c,h\cdot}\sum_{m=1}^{k_r}\sum_{n=1}^{k_c}\sum_{l\geq 0} a_{f,l}X_{f,t-l,mn}u_m v_n \\ &+ \sum_{h=1}^{q}\sum_{t=1}^{T}\Sigma_{\epsilon,jh}\Big(\sum_{g\geq 0} a_{\epsilon,g}X_{\epsilon,t-g,jh}\Big)\sum_{m=1}^{k_r}\sum_{n=1}^{k_c}\sum_{l\geq 0} a_{f,l}X_{f,t-l,mn}u_m v_n.\end{aligned} \tag{5.71}$$

Consider first the second term above, i.e.,

$$\sum_{m=1}^{k_r}\sum_{n=1}^{k_c} u_m v_n \sum_{l\geq 0}\sum_{g\geq 0} a_{f,l}a_{\epsilon,g}\sum_{t=1}^{T}\Big(\sum_{h=1}^{q}\Sigma_{\epsilon,jh}X_{\epsilon,t-g,jh}\Big)X_{f,t-l,mn}.$$

Fix $l \geq 0$, $g \geq 0$. By Assumption (E3), for each $t \in [T]$, we have $\sum_{h=1}^{q}\Sigma_{\epsilon,jh}X_{\epsilon,t-g,jh} \sim$ subG$(C_1q)$, with arbitrary constant $C_1 > 0$ such that $C_1q = \sum_{h=1}^{q}\Sigma^2_{\epsilon,jh}$, which is independent over $g$. Notice that $X_{f,t-l,mn} \sim$ subG$(1)$ by (E3), then $(\sum_{h=1}^{q}\Sigma_{\epsilon,jh}X_{\epsilon,t-g,jh})X_{f,t-l,mn} \sim$ subE$(\sqrt{C_1q})$ which is independent over $t$, and hence $\sum_{t=1}^{T}(\sum_{h=1}^{q}\Sigma_{\epsilon,jh}X_{\epsilon,t-g,jh})X_{f,t-l,mn} \sim$

subE$(\sqrt{C_1 qT})$. Then sum those sub-exponential random variables over $l \geq 0$, $g \geq 0$, we have by (E2),

$$\sum_{l\geq 0}\sum_{g\geq 0} a_{f,l}a_{\epsilon,g}\sum_{t=1}^{T}\Big(\sum_{h=1}^{q}\Sigma_{\epsilon,jh}X_{\epsilon,t-g,jh}\Big)X_{f,t-l,mn} \sim \text{subE}\big(\sqrt{C_2 qT}\big),$$

with some arbitrary constant $C_2 > 0$. As $k_r$, $k_c$, $k_{e,r}$ and $k_{e,c}$ are all constants, we conclude that the entire second term in (5.71), together with the first term therein, are also sub-exponential with parameter of the rate $\sqrt{qT}$. Therefore, for each $j \in [p]$, it holds that

$$(qT)^{-1/2}\sum_{h=1}^{q}\sum_{t=1}^{T} E_{t,jh}\mathbf{u}'\mathbf{F}_t\mathbf{v}$$

is sub-exponential with parameter of constant rate. Using the union bound, with some arbitrary constant $C_3 > 0$, we have

$$\mathbb{P}\Big(\max_{j\in[p]}\frac{1}{(qT)^{1/2}}\sum_{h=1}^{q}\sum_{t=1}^{T}E_{t,jh}\mathbf{u}'\mathbf{F}_t\mathbf{v} \geq \varepsilon\Big) \leq \exp\big\{\log(p) - C_3\varepsilon\big\}, \tag{5.72}$$

implying that $\max_{j\in[p]}(qT)^{-1/2}\sum_{h=1}^{q}\sum_{t=1}^{T}E_{t,jh}\mathbf{u}'\mathbf{F}_t\mathbf{v} = O_P(\log(p))$. This concludes the proof for the lemma. □

**Lemma 5.8** *Let all assumptions in Theorem 5.3 hold, and let Assumption (E3) also hold. Then with $\mathbf{R}_{r,t}$ defined in* (5.22) *and $\mathbf{R}_{c,t}$ in* (5.23)*, we have*

$$\max_{j\in[p]}\Big\|\big(\sum_{t=1}^{T}\mathbf{R}_{r,t}\big)_{\cdot j}\Big\|^2 = O_P\big\{(Tpq + T^2q^2)\log^2(p)\big\}, \tag{5.73}$$

$$\max_{j\in[q]}\Big\|\big(\sum_{t=1}^{T}\mathbf{R}_{c,t}\big)_{\cdot j}\Big\|^2 = O_P\big\{(Tpq + T^2p^2)\log^2(q)\big\}. \tag{5.74}$$

***Proof of Lemma 5.8.*** Consider (5.73). Essentially, we need to show similar results in Lemma 5.2. To this end, we show the corresponding versions of (5.47) and (5.48), and the remaining terms are based on the derived results and can be shown (omitted here) using the same machinery of sub-exponential distribution and the union bound as in (5.72). To start with, using the first result in Lemma 5.7, we have

$$\begin{aligned}\max_{j\in[p]}\Big\|\big(\sum_{t=1}^{T}\mathbf{C}_t\mathbf{E}_t'\big)_{\cdot j}\Big\|_F^2 &= \sum_{i=1}^{p}\|\mathbf{A}_{r,i\cdot}\|^2\cdot\max_{l\in[p]}\Big(\sum_{h=1}^{q}\sum_{t=1}^{T}E_{t,lh}\frac{1}{\|\mathbf{A}_{r,i\cdot}\|}\mathbf{A}_{r,i\cdot}'\mathbf{F}_t\mathbf{A}_{c,h\cdot}\Big)^2\\ &= O_P\big(Tp^{\delta_{r,1}}q\log^2(p)\big).\end{aligned}$$

Similarly,

$$\max_{j\in[p]}\Big\|\big(\sum_{t=1}^{T}\mathbf{E}_t\mathbf{C}_t'\big)_{\cdot j}\Big\|_F^2=\max_{i\in[p]}\Big\{\|\mathbf{A}_{r,i\cdot}\|^2\cdot\sum_{l=1}^{p}\Big(\sum_{h=1}^{q}\sum_{t=1}^{T}E_{t,lh}\frac{1}{\|\mathbf{A}_{r,i\cdot}\|}\mathbf{A}_{r,i\cdot}'\mathbf{F}_t\mathbf{A}_{c,h\cdot}\Big)^2\Big\}$$
$$\le\max_{i\in[p]}\|\mathbf{A}_{r,i\cdot}\|^2\cdot\max_{i\in[p]}\sum_{l=1}^{p}\Big(\sum_{h=1}^{q}\sum_{t=1}^{T}E_{t,lh}\frac{1}{\|\mathbf{A}_{r,i\cdot}\|}\mathbf{A}_{r,i\cdot}'\mathbf{F}_t\mathbf{A}_{c,h\cdot}\Big)^2=O_P\big(Tpq\log^2(p)\big).$$

Next, consider

$$\max_{j\in[p]}\Big\|\sum_{t=1}^{T}\big(\mathbf{E}_t\mathbf{E}_t'\big)_{\cdot j}\Big\|_F^2=\max_{j\in[p]}\sum_{i=1}^{p}\Big(\sum_{t=1}^{T}\sum_{h=1}^{q}E_{t,ih}E_{t,jh}\Big)^2. \tag{5.75}$$

Given $j\in[p]$, first consider first $i\neq j$. By Assumption (E1) and (E2), we can write

$$E_{t,jh}=\sum_{m=1}^{k_{e,r}}\sum_{n=1}^{k_{e,c}}A_{e,r,jm}A_{e,c,hn}\Big(\sum_{w\ge 0}a_{e,w}X_{e,t-w,mn}\Big)+\Sigma_{\epsilon,jh}\Big(\sum_{g\ge 0}a_{\epsilon,g}X_{\epsilon,t-g,jh}\Big),$$
$$E_{t,ih}=\sum_{\tau=1}^{k_{e,r}}\sum_{\gamma=1}^{k_{e,c}}A_{e,r,i\tau}A_{e,c,h\gamma}\Big(\sum_{l\ge 0}a_{e,l}X_{e,t-l,\tau\gamma}\Big)+\Sigma_{\epsilon,ih}\Big(\sum_{\xi\ge 0}a_{\epsilon,\xi}X_{\epsilon,t-\xi,ih}\Big).$$

Then among all terms in the expansion of $\sum_{t=1}^{T}\sum_{h=1}^{q}E_{t,ih}E_{t,jh}$, consider

$$\sum_{g\ge 0}\sum_{\xi\ge 0}a_{\epsilon,g}a_{\epsilon,\xi}\sum_{t=1}^{T}\sum_{h=1}^{q}\Sigma_{\epsilon,jh}\Sigma_{\epsilon,ih}X_{\epsilon,t-g,jh}X_{\epsilon,t-\xi,ih}.$$

Fix $g\ge 0$ and $\xi\ge 0$, then it is direct from Assumption (E3) that $\Sigma_{\epsilon,jh}\Sigma_{\epsilon,ih}X_{\epsilon,t-g,jh}X_{\epsilon,t-\xi,ih}$ is sub-exponential with parameter of constant order and independent over $h\in[q]$ and $t\in[T]$. This implies $\sum_{t=1}^{T}\sum_{h=1}^{q}\Sigma_{\epsilon,jh}\Sigma_{\epsilon,ih}X_{\epsilon,t-g,jh}X_{\epsilon,t-\xi,ih}$ is sub-exponential with parameter of order $(Tq)^{1/2}$, which also holds true for $\sum_{g\ge 0}\sum_{\xi\ge 0}a_{\epsilon,g}a_{\epsilon,\xi}\sum_{t=1}^{T}\sum_{h=1}^{q}\Sigma_{\epsilon,jh}\Sigma_{\epsilon,ih}X_{\epsilon,t-g,jh}X_{\epsilon,t-\xi,ih}$ by Assumption (E2). Thus,

$$\max_{j\in[p]}\sum_{i\neq j}^{p}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{q}\Sigma_{\epsilon,jh}\Big(\sum_{g\ge 0}a_{\epsilon,g}X_{\epsilon,t-g,jh}\Big)\Sigma_{\epsilon,ih}\Big(\sum_{\xi\ge 0}a_{\epsilon,\xi}X_{\epsilon,t-\xi,ih}\Big)\Big\}^2=O_P\big(Tpq\log^2(p)\big). \tag{5.76}$$

Using the same argument above, with the independence between $\{\mathbf{X}_{e,t}\}$ and $\{\mathbf{X}_{\epsilon,t}\}$ from

(E2), we have

$$\begin{aligned}
&\max_{j\in[p]}\sum_{i=1}^{p}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{q}\sum_{m=1}^{k_{e,r}}\sum_{n=1}^{k_{e,c}}A_{e,r,jm}A_{e,c,hn}\Big(\sum_{w\geq 0}a_{e,w}X_{e,t-w,mn}\Big)\Sigma_{\epsilon,ih}\Big(\sum_{\xi\geq 0}a_{\epsilon,\xi}X_{\epsilon,t-\xi,ih}\Big)\Big\}^2\\
&=O_P\big(Tpq\log^2(p)\big),\\
&\max_{j\in[p]}\sum_{i=1}^{p}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{q}\Sigma_{\epsilon,jh}\Big(\sum_{g\geq 0}a_{\epsilon,g}X_{\epsilon,t-g,jh}\Big)\sum_{\tau=1}^{k_{e,r}}\sum_{\gamma=1}^{k_{e,c}}A_{e,r,i\tau}A_{e,c,h\gamma}\Big(\sum_{l\geq 0}a_{e,l}X_{e,t-l,\tau\gamma}\Big)\Big\}^2\\
&=O_P\big(Tpq\log^2(p)\big).
\end{aligned} \tag{5.77}$$

In the expansion of $\sum_{t=1}^{T}\sum_{h=1}^{q}E_{t,ih}E_{t,jh}$ for any $i\in[p]$, consider now

$$\sum_{w\geq 0}\sum_{l\geq 0}a_{e,w}a_{e,l}\sum_{t=1}^{T}\sum_{h=1}^{q}\sum_{m=1}^{k_{e,r}}\sum_{n=1}^{k_{e,c}}\sum_{\tau=1}^{k_{e,r}}\sum_{\gamma=1}^{k_{e,c}}A_{e,r,i\tau}A_{e,c,h\gamma}A_{e,r,jm}A_{e,c,hn}X_{e,t-w,mn}X_{e,t-l,\tau\gamma},$$

which is sub-exponential with mean of order $T$ and parameter of order $\|\mathbf{A}_{e,r}\|_\infty\cdot(T)^{1/2}$ by Assumption (E1), (E2) and (E3). Hence by the sparsity of $\mathbf{A}_{e,r}$ according to (E1) again,

$$\begin{aligned}
&\max_{j\in[p]}\sum_{i=1}^{p}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{q}\sum_{m=1}^{k_{e,r}}\sum_{n=1}^{k_{e,c}}A_{e,r,jm}A_{e,c,hn}\Big(\sum_{w\geq 0}a_{e,w}X_{e,t-w,mn}\Big)\\
&\sum_{\tau=1}^{k_{e,r}}\sum_{\gamma=1}^{k_{e,c}}A_{e,r,i\tau}A_{e,c,h\gamma}\Big(\sum_{l\geq 0}a_{e,l}X_{e,t-l,\tau\gamma}\Big)\Big\}^2=O_P\big(T^2\log^2(p)\big).
\end{aligned} \tag{5.78}$$

To bound (5.75), it remains to consider

$$\sum_{g\geq 0}\sum_{\xi\geq 0}a_{\epsilon,g}a_{\epsilon,\xi}\sum_{t=1}^{T}\sum_{h=1}^{q}\Sigma_{\epsilon,jh}^2X_{\epsilon,t-g,jh}^2,$$

which is sub-exponential with parameter of order $(Tq)^{1/2}$, similar to the case as in (5.76), except that the mean is of order $Tq$. Therefore,

$$\begin{aligned}
&\max_{j\in[p]}\mathbb{1}\{i=j\}\Big\{\sum_{t=1}^{T}\sum_{h=1}^{q}\Sigma_{\epsilon,jh}\Big(\sum_{g\geq 0}a_{\epsilon,g}X_{\epsilon,t-g,jh}\Big)\Sigma_{\epsilon,ih}\Big(\sum_{\xi\geq 0}a_{\epsilon,\xi}X_{\epsilon,t-\xi,ih}\Big)\Big\}^2\\
&=O_P\big(T^2q^2\log^2(p)\big).
\end{aligned} \tag{5.79}$$

Finally for (5.75), combining (5.76), (5.77), (5.78) and (5.79), we have

$$\max_{j\in[p]} \Big\| \sum_{t=1}^{T} \big(\mathbf{E}_t \mathbf{E}_t'\big)_{\cdot j} \Big\|_F^2 = O_P\big((Tpq + T^2 q^2)\log^2(p)\big).$$

This ends the proof for (5.73). The result (5.74) follows similarly to (5.73) and this concludes the proof of the lemma. □

**Lemma 5.9** *Let all assumptions in Theorem 5.3 hold, and let Assumption (E3) also hold. Then we have*

$$\begin{aligned}
&\max_{j\in[p]} \big\|\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r \mathbf{Q}_{r,j\cdot}\big\|^2 \\
&= O_P\Big\{\big(T^{-1} p^{1-2\delta_{r,k_r}} q^{1-2\delta_{c,1}} + p^{-2\delta_{r,k_r}} q^{2(1-\delta_{c,1})}\big)\log^2(p)\Big\}, \qquad (5.80)\\
&\max_{j\in[q]} \big\|\widehat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c \mathbf{Q}_{c,j\cdot}\big\|^2 \\
&= O_P\Big\{\big(T^{-1} q^{1-2\delta_{c,k_c}} p^{1-2\delta_{r,1}} + q^{-2\delta_{c,k_c}} p^{2(1-\delta_{r,1})}\big)\log^2(q)\Big\}, \qquad (5.81)\\
&\max_{t\in[T]} \|\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}\|_F^2 \\
&= O_P\Big\{\big(p^{1-\delta_{r,k_r}} q^{1-\delta_{c,k_c}} + T^{-1} p^{1+2\delta_{r,1}-2\delta_{r,k_r}} q^{1-\delta_{c,1}} + p^{1+\delta_{r,1}-3\delta_{r,k_r}} q^{2-\delta_{c,1}} \\
&\qquad + T^{-1} q^{1+2\delta_{c,1}-2\delta_{c,k_c}} p^{1-\delta_{r,1}} + q^{1+\delta_{c,1}-3\delta_{c,k_c}} p^{2-\delta_{r,1}}\big)\log^2(T)\Big\}. \qquad (5.82)
\end{aligned}$$

*Thus, we have*

$$\begin{aligned}
&\max_{i\in[p],j\in[q],t\in[T]} (\widehat{C}_{t,ij} - C_{t,ij})^2 \\
&= O_P\Big\{\big(p^{1-2\delta_{r,k_r}} q^{1-2\delta_{c,k_c}} + T^{-1} p^{1+2\delta_{r,1}-3\delta_{r,k_r}} q^{1-\delta_{c,1}-\delta_{c,k_c}} + p^{1+\delta_{r,1}-4\delta_{r,k_r}} q^{2-\delta_{c,1}-\delta_{c,k_c}} \\
&\qquad + T^{-1} q^{1+2\delta_{c,1}-3\delta_{c,k_c}} p^{1-\delta_{r,1}-\delta_{r,k_r}} + q^{1+\delta_{c,1}-4\delta_{c,k_c}} p^{2-\delta_{r,1}-\delta_{r,k_r}}\big)\log^2(T)\log^2(p)\log^2(q)\Big\}.
\end{aligned}$$

***Proof of Lemma 5.9.*** To see (5.80), from the proof of the consistency for the row loading matrix in Theorem 5.2, we have

$$\begin{aligned}
&\max_{j\in[p]} \big\|\widehat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r \mathbf{Q}_{r,j\cdot}\big\|^2 = \max_{j\in[p]} \Big\| T^{-1}\widehat{\mathbf{D}}_r^{-1}\widehat{\mathbf{Q}}_r' \Big(\sum_{t=1}^{T} \mathbf{R}_{r,t}\Big)_{\cdot j}\Big\|^2 \\
&\le T^{-2}\cdot \|\widehat{\mathbf{D}}_r^{-1}\|_F^2 \cdot \|\widehat{\mathbf{Q}}_r\|_F^2 \cdot \max_{j\in[p]} \Big\|\Big(\sum_{t=1}^{T} \mathbf{R}_{r,t}\Big)_{\cdot j}\Big\|_F^2 \\
&= O_P\Big(T^{-1} p^{1-2\delta_{r,k_r}} q^{1-2\delta_{c,1}} \log^2(p) + p^{-2\delta_{r,k_r}} q^{2(1-\delta_{c,1})}\log^2(p)\Big),
\end{aligned}$$

where the last equality used Lemma 5.3 and Lemma 5.8. In a similar way, (5.81) holds also by

Lemma 5.3 and Lemma 5.8.

For (5.82), by inspecting (5.25), it suffices to characterize the change from the rate of $\|\mathbf{F}_t\|_F^2$ to the rate of $\max_{t\in[T]}\|\mathbf{F}_t\|_F^2$, while all other rates follow the similar arguments in the proof of Lemma 5.8 by using sub-exponential distributions. With Assumption (E3), $\|\mathbf{F}_t\|_F^2$ is sub-exponential with both mean and parameter of constant order, so that $\max_{t\in[T]}\|\mathbf{F}_t\|_F^2 = O_P(\log^2(T))$. By checking all the rates in the expansion (5.25) are inflated by $\log^2(T)$, (5.82) is hence concluded. Finally, with all previous results and recall the expansion in (5.27), the rate of $\max_{i\in[p],j\in[q],t\in[T]}(\widehat{C}_{t,ij} - C_{t,ij})^2$ is inflated by $\log^2(T)\log^2(p)\log^2(q)$ compared to the individual rate of $(\widehat{C}_{t,ij} - C_{t,ij})^2$. This ends the proof of the lemma. □

# Chapter 6

# Spatial Autoregressive Models with Change Point Detection

## 6.1 Introduction

The study of spatial dependence in regional science gives rise to the techniques in spatial econometrics that we commonly use nowadays. Restricting to cross-sectional data only, a very general form of a model describing spatial dependence can be $\mathbf{y} = f(\mathbf{y}) + \boldsymbol{\epsilon}$ (Anselin, 1988), where $\mathbf{y}$ denotes a vector of $d$ observed units, and $\boldsymbol{\epsilon}$ denotes an error term. A prominent and widely used candidate model is the spatial autoregressive model (see for example LeSage and Pace (2009)), which assumes a known *spatial weight matrix* $\mathbf{W}$ with zero diagonal and $f(\mathbf{y})$ of the form $f(\mathbf{y}) = \rho\mathbf{W}\mathbf{y}$ (or $f(\mathbf{y}) = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta}$ for a model with matrix of covariates $\mathbf{X}$), where $\rho$ is called the spatial correlation coefficient.

Users of these models need to specify the $d \times d$ spatial weight matrix $\mathbf{W}$, which can be a contiguity matrix of 0 and 1, a matrix of inverse distances between two cities/regions, relative amount of import export, etc. An obvious shortcoming for practitioners is to specify an "accurate" spatial weight matrix for use, often in the face of too many potential choices. This leads to a series of attempts to estimate the spatial weight matrix itself from data. For instance, see Pinkse et al. (2002) and Sun (2016) for models dealing with cross-sectional data only, both allowing for nonlinear spatial weight matrix estimation. Beenstock and Felsenstein (2012), Bhattacharjee and Jensen-Butler (2013), Lam and Souza (2020) and Higgins and Martellosio (2023) use spatial panel data for spatial weight matrix estimation, with Lam and Souza (2020) and Higgins and Martellosio (2023) allowing for multiple specified spatial weight matrices through a linear combination of them with constant coefficients.

Recent advances in spatial econometrics allow researchers to specify more complex models with an observed panel $\{\mathbf{y}_t\}$. Zhang and Shen (2015) considers partially linear covariate effects and constant spatial interactions using a sieve method to estimate a nonlinear function, while

Sun and Malikov (2018) considers varying coefficients in both the spatial correlation coefficient (with underlying variables differ over observed units) and the covariate effects, assuming the nonlinear functions are smooth for kernel estimations. Liang et al. (2022) uses kernel estimation on a model with constant spatial interactions but deterministic time-varying coefficient functions for the covariates, while Chang et al. (2025) generalises the model to include an unknown random time trend and deterministic time-varying spatial correlation coefficient, still using kernel estimation. Hong et al. (2024) investigates a model similar to Sun and Malikov (2018), but adds dynamic terms involving $\mathbf{y}_{t-1}$.

However, all the above allow for one specified spatial weight matrix only. As mentioned before, practitioners often face with too many potential choices for a spatial weight matrix. Combining the flexibility of allowing for multiple specified spatial weight matrices as input in Lam and Souza (2020) and varying effects in spatial interactions over observed variables or time directly, we propose a model similar to that in Lam and Souza (2020), but with varying coefficients in the linear combination of the spatial weight matrices. The varying coefficients can be varying over some observed variables (stochastic) or time directly (non-stochastic).

Our contribution in this chapter are three-folds. Firstly, using basis representations, we allow for the varying coefficients to be either stochastic or directly time-varying, without the need for any smoothness conditions. Hence the final estimated spatial weight matrix can be either stochastic or deterministic, e.g., directly time dependent. Secondly, our adaptive LASSO estimators are proved to have the oracle properties, so that ill-specified spatial weight matrices which are irrelevant in the end will be dropped with probability going to 1 as the dimension $d$ and the sample size $T$ go to infinity. At the same time, the effects of relevant spatial weight matrices can be seen to be truly varying or not, again with probability going to 1 as $d, T \to \infty$. This greatly facilitates the interpretability of the spillover effects over time. Last but not least, our framework includes special cases such as spatial autoregressive models with structural changes (Li, 2018) or threshold variables (Deng, 2018; Li and Lin, 2024). Section 6.5 explores the applications to multiple change points detection in both spatial autoregressive models with structural changes or threshold variables, suggesting an applicable algorithm for consistent change points detection in both cases.

The rest of this chapter is organised as follows. Section 6.2 introduces the spatial autoregressive model and presents a procedure using adaptive LASSO to estimate the spatial fixed effect, spatial autoregressive parameters in a basis expansion, and the regression coefficients. Section 6.3 includes the required assumptions and the theoretical guarantees on the parameter estimators. Section 6.4 covers the algorithm for practical implementations including model selection and covariance matrix estimation for our estimators. Section 6.5 focuses on change points detection for a spatial autoregressive model with threshold variables or structural changes. Finally, numerical results are presented in Section 6.6, with a case study of enterprise profits in

China. Section 6.7 provides additional details and simulations, whereas all technical proofs and additional lemmas are deferred to Section 6.8.

# 6.2 Model and Estimation

## 6.2.1 Spatial autoregressive model

We propose a framework of spatial autoregressive models with fixed effects such that for each time $t \in [T]$,

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{p} \Big(\phi_{j,0}^* + \sum_{k=1}^{l_j} \phi_{j,k}^* z_{j,k,t}\Big) \mathbf{W}_j \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \tag{6.1}$$

where $\mathbf{y}_t \in \mathbb{R}^d$ is the observed vector at time $t$, and $\boldsymbol{\mu}^*$ is a constant vector of spatial fixed effects. Each $\mathbf{W}_j \in \mathbb{R}^{d\times d}$ is a pre-specified spatial weight matrix provided by researchers to feature the spillover effects of cross-sectional units from their neighbours. Each $\mathbf{W}_j$ has zero entries on its main diagonal with no restrictions on the signs of off-diagonal entries, and can be asymmetric. Each term $(\phi_{j,0}^* + \sum_{k=1}^{l_j} \phi_{j,k}^* z_{j,k,t})$ is essentially a spatial correlation coefficient for the spatial weight matrix $\mathbf{W}_j$ (see also Lam and Souza (2020)), which can be time-varying by being presented as either a basis expansion using some non-random pre-specified set of basis $\{z_{j,k,t}\}$, or an affine combination of random variables $\{z_{j,k,t}\}$. In either case, we call the $\{z_{j,k,t}\}$'s the dynamic variables hereafter. For $j \in [p]$, $k \in [l_j]$, the parameters $\phi_{j,0}^*, \phi_{j,k}^*$ are unknown and need to be estimated. The covariate matrix $\mathbf{X}_t$ has size $d \times r$, with $\boldsymbol{\beta}^*$ the corresponding unknown regression coefficients of length $r$. Finally, $\boldsymbol{\epsilon}_t$ is the idiosyncratic noise with zero mean.

Without loss of generality, we assume $\mathbf{X}_t$ to have zero mean. Otherwise, we read

$$\boldsymbol{\mu}^* + \mathbf{X}_t \boldsymbol{\beta}^* = \big[\boldsymbol{\mu}^* + \mathbb{E}(\mathbf{X}_t)\boldsymbol{\beta}^*\big] + \big[\mathbf{X}_t - \mathbb{E}(\mathbf{X}_t)\big]\boldsymbol{\beta}^*,$$

which leads to estimating $[\boldsymbol{\mu}^* + \mathbb{E}(\mathbf{X}_t)\boldsymbol{\beta}^*]$ as the spatial fixed effects instead. We can rewrite (6.1) as a traditional spatial autoregressive model $\mathbf{y}_t = \boldsymbol{\mu}^* + \mathbf{W}_t^* \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t$ by defining the true spatial weight matrix at time $t$ as

$$\mathbf{W}_t^* := \sum_{j=1}^{p} \Big(\phi_{j,0}^* + \sum_{k=1}^{l_j} \phi_{j,k}^* z_{j,k,t}\Big) \mathbf{W}_j, \text{ with } -1 < \rho_t^* := \sum_{j=1}^{p} \Big(\phi_{j,0}^* + \sum_{k=1}^{l_j} \phi_{j,k}^* z_{j,k,t}\Big) < 1. \tag{6.2}$$

The restrictions on $\rho_t^*$ for all $t \in [T]$ ensure that the model is stationary. See Assumptions (M2) and (M2') for technical details. We define $L := p + \sum_{j=1}^{p} l_j$ and further reformulate (6.1) as

$$\mathbf{y}_t = \boldsymbol{\mu}^* + (\boldsymbol{\Lambda}_t \boldsymbol{\Phi}^*)\mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \text{ where} \tag{6.3}$$

$$\boldsymbol{\Lambda}_t := (\boldsymbol{\Lambda}_{1,t}, \boldsymbol{\Lambda}_{2,t}, \ldots, \boldsymbol{\Lambda}_{p,t}) \in \mathbb{R}^{d\times dL}, \text{ with } \boldsymbol{\Lambda}_{j,t} := (\mathbf{W}_j, z_{j,1,t}\mathbf{W}_j, \ldots, z_{j,l_j,t}\mathbf{W}_j) \in \mathbb{R}^{d\times(d+dl_j)},$$
$$\boldsymbol{\Phi}^* := (\boldsymbol{\Phi}_1^{*\prime}, \boldsymbol{\Phi}_2^{*\prime}, \ldots, \boldsymbol{\Phi}_p^{*\prime})' \in \mathbb{R}^{dL\times d}, \text{ with } \boldsymbol{\Phi}_j^* := (\phi_{j,0}^*\mathbf{I}_d, \phi_{j,1}^*\mathbf{I}_d, \ldots, \phi_{j,l_j}^*\mathbf{I}_d)'.$$

Due to the endogeneity in $\mathbf{y}_t$ and potentially $\mathbf{X}_t$, we assume that a set of valid instrumental variables $\mathbf{U}_t$ are available for $t \in [T]$. More specifically, each $\mathbf{U}_t$ is independent of $\boldsymbol{\epsilon}_t$ but is correlated with $\mathbf{y}_t$ and the endogenous $\mathbf{X}_t$. Note that if $\mathbf{X}_t$ is exogenous, we may simply have $\mathbf{U}_t = \mathbf{X}_t$. Following Kelejian and Prucha (1998), we can construct instruments $\mathbf{B}_t$ as a $d \times v$ matrix with $v \geq r$ by interacting each given spatial weight matrix with $\mathbf{U}_t$ such that $\mathbf{B}_t$ is composed of at least a subset of linearly independent columns in[1]

$$\Big\{\mathbf{U}_t, \mathbf{W}_1\mathbf{U}_t, \mathbf{W}_1^2\mathbf{U}_t, \ldots, \mathbf{W}_p\mathbf{U}_t, \mathbf{W}_p^2\mathbf{U}_t, \ldots\Big\}.$$

To enhance interpretability of the true spatial weight matrix $\mathbf{W}_t^*$, we assume the dynamic feature of model (6.1) is driven only by a few $\{z_{j,k,t}\}$. That is, the vector of coefficients $\boldsymbol{\phi}^* := (\boldsymbol{\phi}_1^{*\prime}, \boldsymbol{\phi}_2^{*\prime}, \ldots, \boldsymbol{\phi}_p^{*\prime})'$ (with $\boldsymbol{\phi}_j^* := (\phi_{j,0}^*, \phi_{j,1}^*, \ldots, \phi_{j,l_j}^*)'$) is sparse. Using the LASSO (Tibshirani, 1996), an $L_1$ penalty $\lambda\|\boldsymbol{\phi}\|_1$ can be included in a regression problem to shrink the estimators toward zero and some of them to exactly zero, where $\lambda > 0$ is a tuning parameter. However, this form of regularisation penalises uniformly on each entry, which may lead to over- or under-penalisation. The former induces bias while the latter fails sign-consistency, i.e., zeros are estimated exactly as zeros and nonzeros are estimated with the correct signs.

To ensure the zero-consistency in variable selection, a necessary "irrepresentable condition" is often imposed (Zhao and Yu, 2006). Subsequently, Zou (2006) reweighs the regularization to be $\lambda\mathbf{u}'|\boldsymbol{\phi}|$ where $|\cdot|$ is applied entrywise and $\mathbf{u}$ contains the inverse of the initial estimators of $\boldsymbol{\phi}^*$. Now the sign-consistency can be ensured even without the irrepresentable condition if the estimators in $\mathbf{u}$ are $\sqrt{T}$-consistent. Such a framework adaptively penalizes the magnitude of the estimators and is hence called "adaptive LASSO". To this end, we start by profiling out $\boldsymbol{\beta}$. To make use of the instruments, define $\bar{\mathbf{B}} := T^{-1}\sum_{t=1}^T \mathbf{B}_t$. If $\boldsymbol{\phi}$ (and hence $\boldsymbol{\Phi}$) is given, by multiplying $(\mathbf{B}_t - \bar{\mathbf{B}})'$ and summing over all $t \in [T]$ on both sides of (6.3), we then have

$$\sum_{t=1}^T (\mathbf{B}_t - \bar{\mathbf{B}})'(\mathbf{I}_d - \boldsymbol{\Lambda}_t\boldsymbol{\Phi})\mathbf{y}_t = \sum_{t=1}^T (\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{X}_t\boldsymbol{\beta} + \sum_{t=1}^T (\mathbf{B}_t - \bar{\mathbf{B}})'\boldsymbol{\epsilon}_t,$$

where the true values $\boldsymbol{\Phi}^*$ and $\boldsymbol{\beta}^*$ are replaced by $\boldsymbol{\Phi}$ and $\boldsymbol{\beta}$, respectively. Note that the spatial fixed effect $\boldsymbol{\mu}^*$ vanishes since $\sum_{t=1}^T (\mathbf{B}_t - \bar{\mathbf{B}})'\boldsymbol{\mu}^* = \mathbf{0}$. Therefore, the least squares estimator

[1] Ideally, we should have $\mathbf{B}_t$ of the form $[\sum_{j=1}^p (\phi_{j,0} + \sum_{k=1}^{l_j} \phi_{j,k} z_{j,k,t})\mathbf{W}_j]^m \mathbf{U}_t$ for $m = 0, 1, 2, \ldots$. However, each $\phi_{j,0}$ and $\phi_{j,k}$ is unknown and hence we exclude any cross-terms with more than one $\mathbf{W}_j$.

of $\boldsymbol{\beta}^*$ given $\boldsymbol{\phi}$ can be denoted as

$$\boldsymbol{\beta}(\boldsymbol{\phi}) = \Big\{ \sum_{s=1}^{T} \mathbf{X}_s'(\mathbf{B}_s - \bar{\mathbf{B}}) \sum_{t=1}^{T} (\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{X}_t \Big\}^{-1} \sum_{s=1}^{T} \mathbf{X}_s'(\mathbf{B}_s - \bar{\mathbf{B}}) \sum_{t=1}^{T} (\mathbf{B}_t - \bar{\mathbf{B}})'(\mathbf{I}_d - \boldsymbol{\Lambda}_t \boldsymbol{\Phi})\mathbf{y}_t. \tag{6.4}$$

To facilitate formulating the adaptive LASSO problem by accommodating the instrumental variables, we write $\boldsymbol{\gamma} := v^{-1}\mathbf{1}_v$ and denote the $i$-th row of $\mathbf{B}_t$ and $\bar{\mathbf{B}}$ by $\mathbf{B}_{t,i\cdot}$ and $\bar{\mathbf{B}}_{i\cdot}$, respectively. For $i \in [d]$, $t \in [T]$, define the outcome and covariates filtered through instrumental variables as

$$\mathbf{y}_{B,i,t} := (\mathbf{B}_{t,i\cdot} - \bar{\mathbf{B}}_{i\cdot})'\boldsymbol{\gamma}\mathbf{y}_t, \quad \mathbf{X}_{B,i} := \sum_{t=1}^{T} (\mathbf{B}_{t,i\cdot} - \bar{\mathbf{B}}_{i\cdot})'\boldsymbol{\gamma}\mathbf{X}_t.$$

The least squares problem is then

$$\widetilde{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \frac{1}{2T} \sum_{i=1}^{d} \Big\| \sum_{t=1}^{T} (\mathbf{I}_d - \boldsymbol{\Lambda}_t \boldsymbol{\Phi})\mathbf{y}_{B,i,t} - \mathbf{X}_{B,i}\boldsymbol{\beta}(\boldsymbol{\phi}) \Big\|^2. \tag{6.5}$$

Using this solution as an initial estimator, the adaptive LASSO problem becomes solving for

$$\widehat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \frac{1}{2T} \sum_{i=1}^{d} \Big\| \sum_{t=1}^{T} (\mathbf{I}_d - \boldsymbol{\Lambda}_t \boldsymbol{\Phi})\mathbf{y}_{B,i,t} - \mathbf{X}_{B,i}\boldsymbol{\beta}(\boldsymbol{\phi}) \Big\|^2 + \lambda \mathbf{u}'|\boldsymbol{\phi}|, \tag{6.6}$$
$$\text{subj. to} \quad \|\boldsymbol{\Lambda}_t \boldsymbol{\Phi}\|_\infty < 1, \quad \text{with} \quad |\mathbf{z}_t'\boldsymbol{\phi}| < 1 \quad \text{for any } t \in [T],$$

where $\mathbf{z}_t := (\mathbf{z}_{1,t}', \mathbf{z}_{2,t}', \ldots, \mathbf{z}_{p,t}')'$, $\mathbf{z}_{j,t} := (1, z_{j,1,t}, \ldots, z_{j,l_j,t})'$, $\mathbf{u} := (|\widetilde{\phi}_{1,0}|^{-1}, \ldots, |\widetilde{\phi}_{p,l_p}|^{-1})'$, $|\boldsymbol{\phi}| := (|\phi_{1,0}|, \ldots, |\phi_{p,l_p}|)'$ and $\lambda$ is a tuning parameter. With $\widehat{\boldsymbol{\phi}}$ (and hence $\widehat{\boldsymbol{\Phi}}$), the adaptive LASSO estimators for $\boldsymbol{\beta}^*$ can be obtained by $\widehat{\boldsymbol{\beta}} := \boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})$ and the fixed effect estimator by

$$\widehat{\boldsymbol{\mu}} := \frac{1}{T} \sum_{t=1}^{T} \Big\{ (\mathbf{I}_d - \boldsymbol{\Lambda}_t \widehat{\boldsymbol{\Phi}})\mathbf{y}_t - \mathbf{X}_t \widehat{\boldsymbol{\beta}} \Big\}. \tag{6.7}$$

### 6.2.2 Full matrix notations

To facilitate both the theoretical results and practical implementation, the least squares and the adaptive LASSO problems are presented in matrix notations in this subsection. Define first

$$\mathbf{B} := T^{-1/2} d^{-a/2} \, (\mathbf{B}_\gamma - \bar{\mathbf{B}}_\gamma) := T^{-1/2} d^{-a/2} \, \mathbf{I}_d \otimes \big\{ (\mathbf{I}_T \otimes \boldsymbol{\gamma}')(\mathbf{B}_1 - \bar{\mathbf{B}}, \ldots, \mathbf{B}_T - \bar{\mathbf{B}})' \big\}, \tag{6.8}$$

where $a$ is a constant gauging the correlation between $\mathbf{B}_t$ and $\mathbf{X}_t$ so that a larger $a$ generally means $\mathbf{B}_t$ is correlated with more covariates in $\mathbf{X}_t$. See Assumption (R4) for technical details. As in Lam and Souza (2020), in practice we can set $a = 1$ to compute $\mathbf{B}$, without changing the optimal values of any tuning parameters or estimators in the adaptive LASSO problem below.

For ease of notation, denote $z_{j,0,t} = 1$ for all $j \in [p], t \in [T]$. We now rewrite (6.1) as

$$\mathbf{y} = \boldsymbol{\mu}^* \otimes \mathbf{1}_T + \mathbf{V}\boldsymbol{\phi}^* + \mathbf{X}_{\boldsymbol{\beta}^*}\mathbf{vec}\big(\mathbf{I}_d\big) + \boldsymbol{\epsilon}, \quad \text{where} \tag{6.9}$$

$$\begin{aligned}
\mathbf{y} &:= \mathbf{vec}\big((\mathbf{y}_1, \ldots, \mathbf{y}_T)'\big), \quad \boldsymbol{\epsilon} := \mathbf{vec}\big((\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T)'\big),\\
\mathbf{V} &:= (\mathbf{V}_1, \ldots, \mathbf{V}_p), \quad \mathbf{V}_j := \big[\boldsymbol{\Gamma}_{j,0}\mathbf{vec}\big(\mathbf{W}_j'\big), \boldsymbol{\Gamma}_{j,1}\mathbf{vec}\big(\mathbf{W}_j'\big), \ldots, \boldsymbol{\Gamma}_{j,l_j}\mathbf{vec}\big(\mathbf{W}_j'\big)\big],\\
\boldsymbol{\Gamma}_{j,k} &:= \mathbf{I}_d \otimes \big(z_{j,k,1}\mathbf{y}_1, \ldots, z_{j,k,T}\mathbf{y}_T\big)', \quad \mathbf{X}_{\boldsymbol{\beta}^*} := \mathbf{I}_d \otimes \big\{\big(\mathbf{I}_T \otimes \boldsymbol{\beta}^{*\prime}\big)\big(\mathbf{X}_1, \ldots, \mathbf{X}_T\big)'\big\}.
\end{aligned}$$

In this form, the model now has design matrix $\mathbf{V}$ in a classical linear regression setting, except that the endogenous variables $\mathbf{y}_t$ are present in $\mathbf{V}$. We thus obtain the augmented model by left-multiplying both sides of (6.9) by $\mathbf{B}'$:

$$\mathbf{B}'\mathbf{y} = \mathbf{B}'\mathbf{V}\boldsymbol{\phi}^* + \mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}^*}\mathbf{vec}\big(\mathbf{I}_d\big) + \mathbf{B}'\boldsymbol{\epsilon}, \tag{6.10}$$

where the augmented spatial fixed effect vanishes since $\mathbf{B}'(\boldsymbol{\mu}^* \otimes \mathbf{1}_T) = T^{-1/2}d^{-a/2}\boldsymbol{\mu}' \otimes \big\{(\mathbf{B}_1 - \bar{\mathbf{B}}, \ldots, \mathbf{B}_T - \bar{\mathbf{B}})(\mathbf{I}_T \otimes \boldsymbol{\gamma})\mathbf{1}_T\big\} = \mathbf{0}$. For any matrix $\mathbf{C}$, denote $\mathbf{C}^{\otimes} := \mathbf{I}_T \otimes \mathbf{C}$ throughout this chapter. We can also read (6.1) as

$$\mathbf{y}^{\nu} = \mathbf{1}_T \otimes \boldsymbol{\mu}^* + \sum_{j=1}^{p}\sum_{k=0}^{l_j} \phi^*_{j,k}\mathbf{W}_j^{\otimes}\mathbf{y}^{\nu}_{j,k} + \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}^{\nu}, \quad \text{where} \tag{6.11}$$

$$\begin{aligned}
\mathbf{y}^{\nu} &:= (\mathbf{y}_1', \ldots, \mathbf{y}_T')', \quad \mathbf{y}^{\nu}_{j,k} := (z_{j,k,1}\mathbf{y}_1', \ldots, z_{j,k,T}\mathbf{y}_T')',\\
\boldsymbol{\epsilon}^{\nu} &:= (\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_T')', \quad \mathbf{X} := (\mathbf{X}_1', \ldots, \mathbf{X}_T')'.
\end{aligned}$$

Thus with $\mathbf{B}^{\nu} := (\mathbf{B}_1' - \bar{\mathbf{B}}', \ldots, \mathbf{B}_T' - \bar{\mathbf{B}}')'$, we may write (6.4) in matrix form as

$$\boldsymbol{\beta}(\boldsymbol{\phi}) = \big(\mathbf{X}'\,\mathbf{B}^{\nu}\,\mathbf{B}^{\nu\prime}\,\mathbf{X}\big)^{-1}\mathbf{X}'\,\mathbf{B}^{\nu}\,\mathbf{B}^{\nu\prime}\,\Big(\mathbf{y}^{\nu} - \sum_{j=1}^{p}\sum_{k=0}^{l_j}\phi_{j,k}\mathbf{W}_j^{\otimes}\mathbf{y}^{\nu}_{j,k}\Big). \tag{6.12}$$

Together with (6.10), the least squares problem in (6.5) can be described as

$$\widetilde{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \frac{1}{2T}\Big\|\mathbf{B}'\mathbf{y} - \mathbf{B}'\mathbf{V}\boldsymbol{\phi} - \mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\phi})}\mathbf{vec}\big(\mathbf{I}_d\big)\Big\|^2. \tag{6.13}$$

With the least squares estimator $\widetilde{\boldsymbol{\phi}}$, the problem in (6.6) in matrix notation is

$$\begin{aligned}
\widehat{\boldsymbol{\phi}} &= \arg\min_{\boldsymbol{\phi}} \frac{1}{2T}\Big\|\mathbf{B}'\mathbf{y} - \mathbf{B}'\mathbf{V}\boldsymbol{\phi} - \mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\phi})}\mathbf{vec}\big(\mathbf{I}_d\big)\Big\|^2 + \lambda\mathbf{u}'|\boldsymbol{\phi}|, \\
&\quad \text{subj. to} \quad \|\boldsymbol{\Lambda}_t\boldsymbol{\Phi}\|_{\infty} < 1, \quad \text{with} \quad |\mathbf{z}_t'\boldsymbol{\phi}| < 1 \quad \text{for any } t \in [T].
\end{aligned} \tag{6.14}$$

Note that the squared error in both (6.13) and (6.14) are still implicit in $\boldsymbol{\phi}$ due to the term $\boldsymbol{\beta}(\boldsymbol{\phi})$. To this end, define

$$\mathbf{Y}_W := \big(\mathbf{W}_1^{\otimes}\mathbf{y}_{1,0}^{\nu}, \ldots, \mathbf{W}_1^{\otimes}\mathbf{y}_{1,l_1}^{\nu}, \ldots, \mathbf{W}_p^{\otimes}\mathbf{y}_{p,0}^{\nu}, \ldots, \mathbf{W}_p^{\otimes}\mathbf{y}_{p,l_p}^{\nu}\big),$$
$$\boldsymbol{\Xi} := T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\big(\mathbf{X}'\,\mathbf{B}^{\nu}\,\mathbf{B}^{\nu\prime}\,\mathbf{X}\big)^{-1}\mathbf{X}'\,\mathbf{B}^{\nu}\,\mathbf{B}^{\nu\prime}.$$

With all lengthy simplification steps relegated to Section 6.8, (6.13) can be rewritten as

$$\begin{aligned}\widetilde{\boldsymbol{\phi}} &= \arg\min_{\boldsymbol{\phi}} \frac{1}{2T}\Big\|\mathbf{B}'\mathbf{y} - \boldsymbol{\Xi}\mathbf{y}^{\nu} - \big(\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W\big)\boldsymbol{\phi}\Big\|^2 \\ &= \big\{\big(\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W\big)'\big(\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W\big)\big\}^{-1}\big(\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W\big)'\big(\mathbf{B}'\mathbf{y} - \boldsymbol{\Xi}\mathbf{y}^{\nu}\big).\end{aligned} \tag{6.15}$$

Moreover, the adaptive LASSO problem in (6.14) can be written as

$$\widehat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \frac{1}{2T}\Big\|\mathbf{B}'\mathbf{y} - \boldsymbol{\Xi}\mathbf{y}^{\nu} - \big(\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W\big)\boldsymbol{\phi}\Big\|^2 + \lambda\mathbf{u}'|\boldsymbol{\phi}|, \tag{6.16}$$
$$\text{subj. to} \quad \|\boldsymbol{\Lambda}_t\boldsymbol{\Phi}\|_{\infty} < 1, \quad \text{with} \quad |\mathbf{z}_t'\boldsymbol{\phi}| < 1 \quad \text{for any } t \in [T].$$

## 6.3 Assumptions and Theoretical Results

We first present some notations involving the measure of serial dependence of all time series variables, which is gauged by the functional dependence measure introduced by Wu (2005). We state all the assumptions used in this chapter in Section 6.3.1. Denote $\{\mathbf{x}_t\} = \{\mathbf{vec}(\mathbf{X}_t)\}$ and $\{\mathbf{b}_t\} = \{\mathbf{vec}(\mathbf{B}_t)\}$ to be the vectorised processes for $\{\mathbf{X}_t\}$ and $\{\mathbf{B}_t\}$ with length $dr$ and $dv$, respectively. For $t \in [T]$, assume that

$$\mathbf{x}_t = \big[f_i(\mathcal{F}_t)\big]_{i\in[dr]}, \quad \mathbf{b}_t = \big[g_i(\mathcal{G}_t)\big]_{i\in[dv]}, \quad \boldsymbol{\epsilon}_t = \big[h_i(\mathcal{H}_t)\big]_{i\in[d]}, \tag{6.17}$$

where $f_i(\cdot)$'s, $g_i(\cdot)$'s, $h_i(\cdot)$'s are measurable functions defined on the real line, and $\mathcal{F}_t = (\ldots, \mathbf{e}_{x,t-1}, \mathbf{e}_{x,t})$, $\mathcal{G}_t = (\ldots, \mathbf{e}_{b,t-1}, \mathbf{e}_{b,t})$, $\mathcal{H}_t = (\ldots, \mathbf{e}_{\epsilon,t-1}, \mathbf{e}_{\epsilon,t})$ are defined by i.i.d. processes $\{\mathbf{e}_{x,t}\}$, $\{\mathbf{e}_{b,t}\}$ and $\{\mathbf{e}_{\epsilon,t}\}$ respectively, with $\{\mathbf{e}_{b,t}\}$ independent of $\{\mathbf{e}_{\epsilon,t}\}$ but correlated with $\{\mathbf{e}_{x,t}\}$. For $q > 0$, we define

$$\begin{aligned}\theta^x_{t,q,i} &:= \big\|x_{t,i} - \ddot{x}_{t,i}\big\|_q = \big(\mathbb{E}|x_{t,i} - \ddot{x}_{t,i}|^q\big)^{1/q}, \;\; i \in [dr], \\ \theta^b_{t,q,i} &:= \big\|b_{t,i} - \ddot{b}_{t,i}\big\|_q = \big(\mathbb{E}|b_{t,i} - \ddot{b}_{t,i}|^q\big)^{1/q}, \;\; i \in [dv], \\ \theta^{\epsilon}_{t,q,i} &:= \big\|\epsilon_{t,i} - \ddot{\epsilon}_{t,i}\big\|_q = \big(\mathbb{E}|\epsilon_{t,i} - \ddot{\epsilon}_{t,i}|^q\big)^{1/q}, \;\; i \in [d],\end{aligned} \tag{6.18}$$

where $\ddot{x}_{t,i} = f_i(\ddot{\mathcal{F}}_t)$, $\ddot{\mathcal{F}}_t = (\ldots, \mathbf{e}_{x,\text{-}1}, \ddot{\mathbf{e}}_{x,0}, \mathbf{e}_{x,1}, \ldots, \mathbf{e}_{x,t})$, with $\ddot{\mathbf{e}}_{x,0}$ independent of all other $\mathbf{e}_{x,j}$'s. Hence $\ddot{x}_{t,i}$ is a coupled version of $x_{t,i}$ with $\mathbf{e}_{x,0}$ replaced by an i.i.d. copy $\ddot{\mathbf{e}}_{x,0}$. We define $\ddot{b}_{t,i}$ and $\ddot{\epsilon}'_{t,i}$ similarly. Intuitively, a large $\theta^x_{t,q,i}$ implies strong serial correlation in $\mathbf{x}_t$ and incorporates some tail conditions of $f_i(\cdot)$'s, i.e., how $f_i(\cdot)$ frames $\mathbf{e}_{x,0}$ at time $t$ and how exaggerated the functional $f_i(\cdot)$ is.

### 6.3.1 Assumptions

We present here the assumptions for our model. In summary, (I1) helps to identify the model; assumptions prefixed "M" renders the model framework; those prefixed "R" are more technical.

(I1) (Identification). *All the eigenvalues of* $\mathbf{Q}'\mathbf{Q}$ *are uniformly bounded away from* $0$*, where*

$$\mathbf{Q} = \big[\mathbb{E}(\mathbf{B}'\mathbf{V}), \mathbb{E}(\mathbf{B}'\widetilde{\mathbf{X}})\big], \quad \widetilde{\mathbf{X}} = (\mathbf{x}_{1,1\cdot}, \ldots, \mathbf{x}_{T,1\cdot}, \ldots, \mathbf{x}_{1,d\cdot}, \ldots, \mathbf{x}_{T,d\cdot})'.$$

(M1) (Time series in $\mathbf{X}_t$, $\mathbf{B}_t$ and $\boldsymbol{\epsilon}_t$). *The processes* $\{\mathbf{X}_t\}$, $\{\mathbf{B}_t\}$ *and* $\{\boldsymbol{\epsilon}_t\}$ *are second-order stationary and and satisfy* (6.17)*, with* $\{\mathbf{X}_t\}$ *and* $\{\boldsymbol{\epsilon}_t\}$ *having mean zero. The tail condition* $\mathbb{P}(|Z| > z) \le D_1 \exp\big(-D_2 z^{\ell}\big)$ *is satisfied for the variables* $B_{t,ij}$, $X_{t,ij}$, $\epsilon_{t,i}$ *by the same constants* $D_1$, $D_2$ *and* $\ell$. *With* (6.18)*, define the tail sum*

$$\Theta^x_{m,q} = \sum_{t=m}^{\infty} \max_{i\in[dr]} \theta^x_{t,q,i}, \quad \Theta^b_{m,q} = \sum_{t=m}^{\infty} \max_{i\in[dv]} \theta^b_{t,q,i}, \quad \Theta^\epsilon_{m,q} = \sum_{t=m}^{\infty} \max_{i\in[d]} \theta^\epsilon_{t,q,i}.$$

*We assume that for some* $w > 2$, $\Theta^x_{m,2w}$, $\Theta^b_{m,2w}$, $\Theta^\epsilon_{m,2w} \le Cm^{-\alpha}$ *with* $\alpha$, $C > 0$ *being constants that can depend on* $w$.

(M2) (True spatial weight matrix $\mathbf{W}^*_t$ with non-random basis $z_{j,k,t}$). $\mathbf{W}^*_t$ *defined in* (6.2) *uses a uniformly bounded non-stochastic basis* $\{z_{j,k,t}\}$ *for* $j \in [p]$, $k \in [l_j]$. *There exists a constant* $\eta > 0$ *such that for all* $t \in [T]$, $\|\mathbf{W}^*_t\|_\infty < \eta < 1$ *uniformly as* $d \to \infty$. *The elements in* $\mathbf{W}^*_t$ *can be negative, and* $\mathbf{W}^*_t$ *can be asymmetric. Furthermore,* $\rho^*_t$ *defined in* (6.2) *satisfies* $|\rho^*_t| < 1$.

(M2') (True spatial weight matrix $\mathbf{W}^*_t$ with random $z_{j,k,t}$). *Same as Assumption (M2), except that* $\{z_{j,k,t}\}$ *is a zero mean stochastic process with support* $[-1, 1]$*, such that* $z_{j,k,t} = u_{j,k}(\mathcal{U}_t)$ *similar to* (6.17)*, with* $\mathbb{E}(z_{j,k,t}\mathbf{X}_t) = \mathbf{0}$, $\mathbb{E}(z_{j,k,t}\boldsymbol{\epsilon}_t) = \mathbf{0}$*, and* $\Theta^z_{m,2w} \le Cm^{-\alpha}$ *as in Assumption (M1). Furthermore:*

1. *there exists* $\eta > 0$ *such that* $\sum_{j=1}^{p}\sum_{k=0}^{l_j} \|\phi^*_{j,k}\mathbf{W}_j\|_\infty < \eta < 1$ *uniformly as* $d \to \infty$.
2. $\sum_{j=1}^{p}\sum_{k=0}^{l_j} |\phi^*_{j,k}| < 1$.

(R1) *Denote the* $d^2L \times L$ *block diagonal matrix* $\mathbf{D}_W := \operatorname{diag}\big\{\mathbf{I}_{1+l_1} \otimes \mathbf{vec}\big(\mathbf{W}_1'\big), \dots, \mathbf{I}_{1+l_p} \otimes \mathbf{vec}\big(\mathbf{W}_p'\big)\big\}$*. Then there exists a constant* $u > 0$ *such that the* $L$*-th largest singular value of* $\mathbf{D}_W$ *satisfies* $\sigma_L^2(\mathbf{D}_W) \geq du > 0$ *uniformly as* $d \to \infty$*.*

*Moreover, there exists a constant* $c > 0$ *such that* $\max_j \left\{\|\mathbf{W}_j\|_1, \|\mathbf{W}_j\|_\infty\right\} \leq c < \infty$ *uniformly as* $d \to \infty$*.*

(R2) *Write* $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_\epsilon^{1/2}\boldsymbol{\epsilon}_t^*$ *with* $\boldsymbol{\Sigma}_\epsilon$ *being the covariance matrix of* $\boldsymbol{\epsilon}_t$*. Assume* $\big\|\boldsymbol{\Sigma}_\epsilon\big\|_{\max} \leq \sigma_{\max}^2 < \infty$ *uniformly as* $d \to \infty$*. The same applies to the variance of the elements in* $\mathbf{B}_t$*.*

*Assume also* $\big\|\boldsymbol{\Sigma}_\epsilon^{1/2}\big\|_\infty \leq S_\epsilon < \infty$ *uniformly as* $d \to \infty$*, with* $\{\epsilon_{t,i}^*\}_{i\in[d]}$ *being a martingale difference with respect to the filtration generated by* $\sigma(\epsilon_{t,1}, \dots, \epsilon_{t,i})$*. Furthermore,* $\{\boldsymbol{\epsilon}_t^*\}_{t\in[T]}$ *satisfies the tail condition and the functional dependence in Assumption (M1).*

(R3) *All singular values of* $\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)$ *are uniformly larger than* $du$ *for some constant* $u > 0$*, while the maximum singular value is of order* $d$*. Individual entries in the matrix* $\mathbb{E}(\mathbf{b}_t\mathbf{x}_t')$ *are uniformly bounded away from infinity, with* $\mathbf{x}_t$ *and* $\mathbf{b}_t$ *defined in* (6.17).

(R4) *With the same* $a \in [0, 1]$ *introduced in* (6.8)*, we define*

$$\mathbf{G} := d^{-a}\,\mathbf{I}_d \otimes \{\mathbb{E}(\check{\mathbf{G}})\,\mathbb{E}(\check{\mathbf{G}})'\}, \quad \check{\mathbf{G}} := (\check{\mathbf{G}}_{1,0}, \dots, \check{\mathbf{G}}_{1,l_1}, \dots, \check{\mathbf{G}}_{p,0}, \dots, \check{\mathbf{G}}_{p,l_p}),$$
$$\check{\mathbf{G}}_{j,k} := \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\boldsymbol{\beta}^{*\prime}\,\mathbf{X}_t'\,\boldsymbol{\Pi}_t^{*\prime}, \quad \boldsymbol{\Pi}_t^* := (\mathbf{I}_d - \mathbf{W}_t^*)^{-1}.$$

*We assume that* $\mathbf{G}$ *has full rank and there exists a constant* $u > 0$ *such that* $\lambda_{\min}(\mathbf{G}) \geq u > 0$ *and* $\lambda_{\max}(\mathbf{G}) < \infty$ *uniformly as* $d \to \infty$*.*

(R5) *For the same constant* $a$ *as in Assumption (R4), we have for each* $d$*,*

$$\max_{i\in[d]}\sum_{j=1}^{d}\big\|\mathbb{E}(\mathbf{B}_{t,i\cdot}\mathbf{X}_{t,j\cdot}')\big\|_{\max}, \;\; \max_{j\in[d]}\sum_{i=1}^{d}\big\|\mathbb{E}(\mathbf{B}_{t,i\cdot}\mathbf{X}_{t,j\cdot}')\big\|_{\max} = O(d^a).$$

*At the same time, assume* $\mathbb{E}(\mathbf{X}_t \otimes \mathbf{B}_t\boldsymbol{\gamma})$ *has all singular values of order* $d^{1+a}$*.*

(R6) *With* $\boldsymbol{\Pi}_t^*$ *in Assumption (R4), define*

$$\ddot{\mathbf{G}} := (\ddot{\mathbf{G}}_{1,0}, \dots, \ddot{\mathbf{G}}_{1,l_1}, \dots, \ddot{\mathbf{G}}_{p,0}, \dots, \ddot{\mathbf{G}}_{p,l_p}),$$
$$\ddot{\mathbf{G}}_{j,k} := \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\big(\mathbf{I}_d \otimes \boldsymbol{\beta}^{*\prime}\,\mathbf{X}_t'\,\boldsymbol{\Pi}_t^{*\prime}\big).$$

*Assume that* $\mathbb{E}(\ddot{\mathbf{G}})\mathbb{E}(\ddot{\mathbf{G}})'$ *has full rank and that there exists a constant* $u > 0$ *such that* $\lambda_v\big(\mathbb{E}(\ddot{\mathbf{G}})\mathbb{E}(\ddot{\mathbf{G}})'\big) \geq u > 0$ *and* $\lambda_{\max}\big(\mathbb{E}(\ddot{\mathbf{G}})\mathbb{E}(\ddot{\mathbf{G}})'\big) < \infty$ *uniformly as* $d \to \infty$*.*

(R7) *Define the predictive dependence measures*

$$P_0^b(B_{t,jk}) := \mathbb{E}(B_{t,jk} \mid \mathcal{G}_0) - \mathbb{E}(B_{t,jk} \mid \mathcal{G}_{\text{-}1}), \quad P_0^\epsilon(\epsilon_{t,j}) := \mathbb{E}(\epsilon_{t,j} \mid \mathcal{H}_0) - \mathbb{E}(\epsilon_{t,j} \mid \mathcal{H}_{\text{-}1}),$$

*with* $\mathcal{G}_t$ *and* $\mathcal{H}_t$ *specified after* (6.17). *Assume*

$$\sum_{t\geq 0} \max_{j\in[d]} \max_{k\in[v]} \left\|P_0^b(B_{t,jk})\right\|_2 < \infty, \quad \sum_{t\geq 0} \max_{j\in[d]} \left\|P_0^\epsilon(\epsilon_{t,j})\right\|_2 < \infty.$$

(R8) *For* $b \in [0,1]$, *the eigenvalues of* $\text{Var}(\boldsymbol{\epsilon}_t)$ *and* $\text{Var}(d^{-b/2}\mathbf{B}_{t,\cdot k})$ *are uniformly bounded away from zero and infinity, and respectively dominate the singular values of* $\mathbb{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}')$ *and* $d^{-b}\mathbb{E}\big\{[\mathbf{B}_{t,\cdot k} - \mathbb{E}(\mathbf{B}_{t,\cdot k})][\mathbf{B}_{t+\tau,\cdot k} - \mathbb{E}(\mathbf{B}_{t,\cdot k})]'\big\}$ *for any* $\tau \neq 0$. *The sum of the* $i$*-th largest singular values over all lags* $\tau \in \mathbb{Z}$ *for each* $i \in [d]$ *is assumed to be finite for both autocovariance matrices of* $\{\boldsymbol{\epsilon}_t\}$ *and* $\{d^{-b/2}\mathbf{B}_t\}$.

(R9) *Define* $c_T = gT^{-1/2}\log^{1/2}(T \vee d)$ *for some constant* $g > 0$. *The tuning parameter for the adaptive LASSO problem* (6.6) *is* $\lambda = Cc_T$ *for some constant* $C > 0$.

(R10) (Rate assumptions). *We assume that as* $L, d, T \to \infty$,

$$c_T L^{3/2} d^{1-a},\ \ Ld^{-1},\ \ L^2 d^3 T^{2-w},\ \ d^{b+2a+1/w}T^{-1} = o(1),$$
$$d^{-1/w}\log(T \vee d),\ \ d^{b-a-1/w}\log^{-1}(T \vee d) = O(1).$$

In the sequel, we discuss in detail the identification of the model, the structure of the true spatial weight matrix $\mathbf{W}_t^*$, and some technical assumptions made above. To show the coefficients $\boldsymbol{\phi}^*$ and $\boldsymbol{\beta}^*$ in (6.10) are identified under Assumption (I1), suppose that two sets of parameters $(\check{\boldsymbol{\phi}}, \check{\boldsymbol{\beta}})$ and $(\acute{\boldsymbol{\phi}}, \acute{\boldsymbol{\beta}})$ both satisfy model (6.10). Then we have

$$\mathbf{B}'\,\mathbf{V}\check{\boldsymbol{\phi}} + \mathbf{B}'\,\mathbf{X}_{\check{\boldsymbol{\beta}}}\mathbf{vec}\big(\mathbf{I}_d\big) = \mathbf{B}'\,\mathbf{V}\acute{\boldsymbol{\phi}} + \mathbf{B}'\,\mathbf{X}_{\acute{\boldsymbol{\beta}}}\mathbf{vec}\big(\mathbf{I}_d\big).$$

By noticing that $\mathbf{B}'\,\mathbf{X}_{\boldsymbol{\beta}}\mathbf{vec}\big(\mathbf{I}_d\big) = \mathbf{B}'\,\widetilde{\mathbf{X}}\boldsymbol{\beta}$, we may rearrange the above and arrive at

$$\mathbf{0} = \mathbf{B}'\,\mathbf{V}(\check{\boldsymbol{\phi}} - \acute{\boldsymbol{\phi}}) + \mathbf{B}'\,\widetilde{\mathbf{X}}(\check{\boldsymbol{\beta}} - \acute{\boldsymbol{\beta}}) = \begin{pmatrix}\mathbf{B}'\,\mathbf{V} & \mathbf{B}'\,\widetilde{\mathbf{X}}\end{pmatrix}\begin{pmatrix}\check{\boldsymbol{\phi}} - \acute{\boldsymbol{\phi}} \\ \check{\boldsymbol{\beta}} - \acute{\boldsymbol{\beta}}\end{pmatrix}.$$

Taking expectation and left-multiplying by $(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$ on both sides, we obtain $\check{\boldsymbol{\phi}} = \acute{\boldsymbol{\phi}}$ and $\check{\boldsymbol{\beta}} = \acute{\boldsymbol{\beta}}$. Now with $\boldsymbol{\phi}^*$ and $\boldsymbol{\beta}^*$ identified, we also have $\boldsymbol{\mu}^*$ uniquely identified.

The time series components in the model is depicted in Assumption (M1) such that weak serial dependence is allowed. Such a definition of functional or physical dependence of time series is used by various previous work such as Shao (2010). We assume exponential tails so that a Nagaev-type inequality for functional dependent data can be used (Liu et al., 2013).

Assumptions (M2) and (M2') describe the structure of $\mathbf{W}_t^*$ under two different settings for $\{z_{j,k,t}\}$. The row sum condition for $\mathbf{W}_t^*$ ensures the model (6.1) to be uniformly stationary. (M2) treats $\{z_{j,k,t}\}$ as a non-random series while (M2') allows $\{z_{j,k,t}\}$ to be stochastic. Note that in (M2'), the stationarity is guaranteed with $\|\mathbf{W}_t^*\|_\infty < \eta < 1$ and $|\rho_t^*| < 1$ with probability 1. The assumption of zero mean and support $[-1, 1]$ simplifies our presentation. For a general finite mean and bounded support, we may rewrite each $\phi_{j,k}^* z_{j,k,t}$ as

$$\begin{aligned}\phi_{j,k}^* z_{j,k,t} &= \phi_{j,k}^* \mathbb{E}(z_{j,k,t}) + \phi_{j,k}^* \big[z_{j,k,t} - \mathbb{E}(z_{j,k,t})\big] \\ &= \phi_{j,k}^* \mathbb{E}(z_{j,k,t}) + \big(\phi_{j,k}^* z_{j,k}^*\big)\big\{\big[z_{j,k,t} - \mathbb{E}(z_{j,k,t})\big]/z_{j,k}^*\big\}.\end{aligned}$$

Note that we set $z_{j,0,t} = 1$ for $j \in [p], t \in [T]$. This is justified by the fact that we may allow some $\{z_{j,k,t}\}$ for $j \in \mathcal{P} \subseteq [p], k \in \mathcal{L}_j \subseteq [l_j]$ satisfying (M2) with all other $\{z_{j,k,t}\}$ satisfying (M2'), since all theoretical results hold with either assumption.

Note that although Assumptions (M2) and (M2') constrain the magnitude of the true spatial weight matrix $\mathbf{W}_t^*$ which depends on the spatial weight matrix candidates $\mathbf{W}_j$, the dynamic variables $z_{j,k,t}$, and the coefficients $\phi_{j,k}^*$, our estimator would remain effective even if $\phi_{j,k}^*$ are very small. In detail, $\phi_{j,k}^*$ being potentially very small lies in the scenario where the signals from $\mathbf{W}_j$ and $z_{j,k,t}$ are potentially very strong, so our least squares and adaptive LASSO problems would not be affected. Furthermore, the set of technical assumptions, e.g. Assumption (R1), ensures the performance of our estimators, which we explain below.

(R1) describes how sparse each spatial weight matrix candidate is. It is worthwhile pointing out that although each $\mathbf{W}_j$ is not necessarily linearly independent with each other by (R1), we actually implicitly impose such linear independence condition from (I1) through the combination of $\{z_{j,k,t}\}$ and $\mathbf{W}_j$ in $\mathbf{B}'\mathbf{V}$. (R2) is included as a technical addition to (M2).

Assumptions (R3) to (R6) all draw on the relation between $\mathbf{B}_t$ and $\mathbf{X}_t$. Their dependence structure is non-trivial due to the extra complication from spatial weight matrix and the time-varying components here. Naturally, $\mathbf{B}_t$ needs to be correlated with $\mathbf{X}_t$ to a certain extent, captured by an unknown constant $a$ which facilitates the presentation of theoretical results. For instance, as an immediate consequence of (R5), we may derive $\big\|\mathbb{E}(\mathbf{X}_t \otimes \mathbf{B}_t\boldsymbol{\gamma})\big\|_1 = O(d^{1+a})$, which is a key ingredient in obtaining the rates for $\big\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}^*\big\|_1$ in Theorem 6.1.

The predictive dependence measure defined in (R7) allows us to apply the central limit theorem for data with functional dependence and the Assumption (R7) can be satisfied by, for example, causal AR processes. Assumption (R8) further restricts the serial correlation in the noise. It also introduces another constant $b$ that characterises how elements in $\mathbf{B}_t$ are contemporarily and temporally correlated with each other. From the comparison of rates derived in the proofs, we conclude that $b$ is actually bounded above by $1/w$, which is intuitive since a large $w$ in (M1) generally implies light-tails and hence a small $b$. Lastly, (R9) sets the rate for $\lambda$ and (R10) characterises the relation between $T$, $d$ and $L$. As an example, (R10) is satisfied when $w = 6$, $a = 1/2$, $L = O(d^{1/3})$ and $T \asymp d^2$.

### 6.3.2 Main results

In this subsection, we formally present the main results for our model estimators.

**Theorem 6.1** *Let all assumptions in Section 6.3.1 hold ((M2) or (M2')). Given any $\boldsymbol{\phi}$ as an estimator of $\boldsymbol{\phi}^*$, with $c_T$ is defined in Assumption (R9), $\boldsymbol{\beta}(\boldsymbol{\phi})$ according to* (6.12) *satisfies*

$$\left\|\boldsymbol{\beta}(\boldsymbol{\phi}) - \boldsymbol{\beta}^*\right\|_1 = O_P\Big(\left\|\boldsymbol{\phi} - \boldsymbol{\phi}^*\right\|_1 + c_T d^{-\frac{1}{2}+\frac{1}{2w}}\Big).$$

*In particular, the least squares estimator $\widetilde{\boldsymbol{\phi}}$ in* (6.15) *and $\widetilde{\boldsymbol{\beta}} := \boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})$ under $L = O(1)$ satisfy*

$$\left\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}^*\right\|_1 = O_P\big(c_T d^{-\frac{1}{2}+\frac{1}{2w}}\big) = \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1.$$

Theorem 6.1 serves as a foundational step for the results hereafter. From the theorem, the error of our least squares estimator $\boldsymbol{\beta}(\boldsymbol{\phi})$ might be inflated by the plugged-in estimator for $\boldsymbol{\phi}^*$. With a dense estimator $\widetilde{\boldsymbol{\phi}}$, we arrive at the rate $c_T d^{-\frac{1}{2}+\frac{1}{2w}}$. The dependence of the rate on $w$ confirms that weaker temporal dependence in the data results in better estimation, as expected. We now present the sign-consistency of our adaptive LASSO estimator.

**Theorem 6.2** *(Oracle property for $\widehat{\boldsymbol{\phi}}$). Let all assumptions in Section 6.3.1 hold (either (M2) or (M2')), except that (R4) and (R6) are satisfied with $\check{\mathbf{G}}$ and $\ddot{\mathbf{G}}$ respectively replaced by*

$$\check{\mathbf{G}} = \check{\mathbf{G}}_H, \quad \ddot{\mathbf{G}} = \ddot{\mathbf{G}}_H, \quad \textit{with } H := \big\{i : (\boldsymbol{\phi}^*)_i \neq 0\big\},$$

*where for any matrix with L columns, $(\cdot)_H$ denotes the matrix with its columns restricted on $H$. Then as $T, d \to \infty$, with probability approaching 1,* $\mathrm{sign}(\widehat{\boldsymbol{\phi}}_H) = \mathrm{sign}(\boldsymbol{\phi}^*_H)$ *and $\widehat{\boldsymbol{\phi}}_{H^c} = \mathbf{0}$.*

*If we further assume the smallest eigenvalue of $\mathbf{R}_H \mathbf{S}_\gamma \mathbf{S}_\gamma' \mathbf{R}_H'$ is of constant order, with $\mathbf{R}_H$ and $\mathbf{S}_\gamma$ defined below, then $\widehat{\boldsymbol{\phi}}_H$ is asymptotically normal with rate $T^{-1/2} d^{-(1-b)/2}$ such that*

$$T^{1/2}\big(\mathbf{R}_H \mathbf{S}_\gamma \mathbf{R}_\beta \boldsymbol{\Sigma}_\beta \mathbf{R}_\beta' \mathbf{S}_\gamma' \mathbf{R}_H'\big)^{-1/2}(\widehat{\boldsymbol{\phi}}_H - \boldsymbol{\phi}^*_H) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, \mathbf{I}_{|H|}\big),$$

*where*

$$
\begin{aligned}
\mathbf{R}_\beta &= \big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t),\\
\boldsymbol{\Sigma}_\beta &= \sum_\tau \mathbb{E}\big\{[\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]'\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}'[\mathbf{B}_{t+\tau} - \mathbb{E}(\mathbf{B}_t)]\big\},\\
\mathbf{R}_H &= \big[(\mathbf{H}_{20} - \mathbf{H}_{10})_H'(\mathbf{H}_{20} - \mathbf{H}_{10})_H\big]^{-1}(\mathbf{H}_{20} - \mathbf{H}_{10})_H',\\
\mathbf{S}_\gamma &= \Big\{\mathbb{E}\big[\mathbf{X}_{t,1\cdot}(\mathbf{B}_{t,1\cdot} - \mathbb{E}(\mathbf{B}_{t,1\cdot}))'\big]\boldsymbol{\gamma}, \ldots, \mathbb{E}\big[\mathbf{X}_{t,1\cdot}(\mathbf{B}_{t,d\cdot} - \mathbb{E}(\mathbf{B}_{t,d\cdot}))'\big]\boldsymbol{\gamma},\\
&\qquad \ldots,\\
&\qquad \mathbb{E}\big[\mathbf{X}_{t,d\cdot}(\mathbf{B}_{t,1\cdot} - \mathbb{E}(\mathbf{B}_{t,1\cdot}))'\big]\boldsymbol{\gamma}, \ldots, \mathbb{E}\big[\mathbf{X}_{t,d\cdot}(\mathbf{B}_{t,d\cdot} - \mathbb{E}(\mathbf{B}_{t,d\cdot}))'\big]\boldsymbol{\gamma}\Big\}',\\
\mathbf{H}_{10} &= \Big\{\big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbb{E}\big(\mathbf{U}_{\mathbf{x},1,0}(\boldsymbol{\beta}^* \otimes \mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime}\big)\big]\mathbf{vec}\big(\mathbf{W}_1'\big)\\
&\qquad + \big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbb{E}\big(\mathbf{U}_{\boldsymbol{\mu},1,0}\boldsymbol{\Pi}_t^{*\prime}\big)\big]\mathbf{vec}\big(\mathbf{W}_1'\big),\\
&\qquad \ldots,\\
&\qquad \big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbb{E}\big(\mathbf{U}_{\mathbf{x},p,l_p}(\boldsymbol{\beta}^* \otimes \mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime}\big)\big]\mathbf{vec}\big(\mathbf{W}_p'\big)\\
&\qquad + \big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbb{E}\big(\mathbf{U}_{\boldsymbol{\mu},p,l_p}\boldsymbol{\Pi}_t^{*\prime}\big)\big]\mathbf{vec}\big(\mathbf{W}_p'\big)\Big\},\\
\mathbf{H}_{20} &= \mathbb{E}(\mathbf{X}_t \otimes \mathbf{B}_t\boldsymbol{\gamma})\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\\
&\quad \cdot \Big\{\mathbf{V}_{\mathbf{W}_1',v}'\mathbb{E}\big[(\mathbf{I}_d \otimes \mathbf{U}_{\mathbf{x},1,0})\mathbf{V}_{\boldsymbol{\Pi}_t^*,r}\big]\boldsymbol{\beta}^* + \mathbf{V}_{\mathbf{W}_1',v}'\mathbb{E}\big[(\mathbf{I}_d \otimes \mathbf{U}_{\boldsymbol{\mu},1,0})\mathbf{vec}\big(\boldsymbol{\Pi}_t^{*\prime}\big)\big],\\
&\qquad \ldots,\\
&\qquad \mathbf{V}_{\mathbf{W}_p',v}'\mathbb{E}\big[(\mathbf{I}_d \otimes \mathbf{U}_{\mathbf{x},p,l_p})\mathbf{V}_{\boldsymbol{\Pi}_t^*,r}\big]\boldsymbol{\beta}^* + \mathbf{V}_{\mathbf{W}_p',v}'\mathbb{E}\big[(\mathbf{I}_d \otimes \mathbf{U}_{\boldsymbol{\mu},p,l_p})\mathbf{vec}\big(\boldsymbol{\Pi}_t^{*\prime}\big)\big]\Big\},\\
\mathbf{U}_{\mathbf{x},j,k} &= \frac{1}{T}\sum_{t=1}^T z_{j,k,t}\mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\mathbf{x}_t', \quad \mathbf{U}_{\boldsymbol{\mu},j,k} = \frac{1}{T}\sum_{t=1}^T z_{j,k,t}\mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\boldsymbol{\mu}^{*\prime},
\end{aligned}
$$

*with the notation* $\mathbf{V}_{\mathbf{H},K} = \big(\mathbf{I}_K \otimes \mathbf{h}_1', \ldots, \mathbf{I}_K \otimes \mathbf{h}_n'\big)'$ *for a given* $n \times d$ *matrix* $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_n)'$.

Regarding the new assumptions in Theorem 6.2, the term $\check{\mathbf{G}}$ (resp. $\ddot{\mathbf{G}}$) in (R4) (resp. (R6)) is simply replaced by its $H$-restricted version. Moreover, we show in the proof that $\mathbf{R}_H\mathbf{S}_\gamma\mathbf{S}_\gamma'\mathbf{R}_H'$ has its largest eigenvalue of constant order, and hence the requirement on its smallest eigenvalue is not particularly strong. Nonetheless, two key results are obtained: $\widehat{\boldsymbol{\phi}}$ consistently estimate the zeros in $\boldsymbol{\phi}^*$ as exact zeros, and are asymptotically normal on the nonzero entries in $\boldsymbol{\phi}^*$. The convergence rate is $T^{-1/2}d^{-(1-b)/2}$, which is worse off if more variables in $\mathbf{B}_t$ are correlated.

Theorem 6.2 enables us to perform inference on $\widehat{\boldsymbol{\phi}}_H$ in practice, with the covariance matrix replaced by the plug-in estimator (see Section 6.4 for more details). If $\{z_{j,k,t}\}$'s are non-stochastic, inference on $\rho_t^*$ by $\widehat{\rho}_t := (\mathbf{z}_t)_H'\widehat{\boldsymbol{\phi}}_H$ is also feasible, since

$$
T^{1/2}\big((\mathbf{z}_t)_H'\mathbf{R}_H\mathbf{S}_\gamma\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta'\mathbf{S}_\gamma'\mathbf{R}_H'(\mathbf{z}_t)_H\big)^{-1/2}(\widehat{\rho}_t - \rho_t^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).
$$

Lastly, we present the consistency of the spatial weight matrix estimator and the spatial fixed effect estimator. Note that Theorem 6.3 implies the spectral norm error of the spatial weight matrix estimator $\widehat{\mathbf{W}}_t$ also satisfies $\|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\| = O_P\big(T^{-1/2}d^{-(1-b)/2}\big)$.

**Theorem 6.3** *Let assumptions in Theorems 6.1 and 6.2 hold. Then, for* $\widehat{\mathbf{W}}_t := \sum_{j=1}^{p}\Big(\widehat{\phi}_{j,0} + \sum_{k=1}^{l_j}\widehat{\phi}_{j,k}z_{j,k,t}\Big)\mathbf{W}_j$ *and the spatial fixed effect estimator* $\widehat{\boldsymbol{\mu}}$ *defined in* (6.7)*, we have*

$$\|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\|_\infty = O_P\big(T^{-1/2}d^{-(1-b)/2}\big) = \|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\|_1, \quad \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_{\max} = O_P(c_T).$$

## 6.4 Practical Implementation

In Section 6.2.2, we estimate $(\boldsymbol{\mu}^*, \boldsymbol{\phi}^*, \boldsymbol{\beta}^*)$ by first obtaining a penalised estimator for $\boldsymbol{\phi}^*$, followed by the least squares estimator for $\boldsymbol{\beta}^*$ and $\boldsymbol{\mu}^*$. The step-by-step algorithm is now presented.

Algorithm for $(\boldsymbol{\mu}^*, \boldsymbol{\phi}^*, \boldsymbol{\beta}^*)$ Estimation

1. Compute the least squares estimator stated in (6.15) and denote it as $\widetilde{\boldsymbol{\phi}}$.
2. Construct $\mathbf{u}$ using $\widetilde{\boldsymbol{\phi}}$. Using the Least Angle Regressions (LARS) (Efron et al., 2004), solve the adaptive LASSO problem stated in (6.16), and denote the solution by $\widehat{\boldsymbol{\phi}}$.
3. Using (6.12), obtain the least squares estimator for $\boldsymbol{\beta}^*$ as $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})$.
4. According to (6.3), construct $\widehat{\boldsymbol{\Phi}}$ using $\widehat{\boldsymbol{\phi}}$ and obtain the least squares estimator for $\boldsymbol{\mu}^*$ as $\widehat{\boldsymbol{\mu}} = T^{-1}\sum_{t=1}^{T}\big\{(\mathbf{I}_d - \boldsymbol{\Lambda}_t\widehat{\boldsymbol{\Phi}})\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}\big\}$.

The tuning parameter $\lambda$ in step 2 can be determined via minimising the following BIC:

$$\mathrm{BIC}(\lambda) = \log\Big(\frac{1}{T}\Big\|\mathbf{B}'\mathbf{y} - \mathbf{B}'\mathbf{V}\widehat{\boldsymbol{\phi}} - \mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})}\mathbf{vec}\big(\mathbf{I}_d\big)\Big\|^2\Big) + |\widehat{H}|\,\frac{\log(T)}{T}\,\log(\log(L)), \quad (6.19)$$

which is inspired by Wang et al. (2009), where $\widehat{\boldsymbol{\phi}}$ is the adaptive LASSO solution with parameter $\lambda$ and $\widehat{H}$ is the set of indices on which $\widehat{\boldsymbol{\phi}}$ is nonzero. Note that although $\mathbf{B}$ contains the unknown constant $a$, the optimal $\lambda$ is independent of it. A procedure for assessing the goodness of fit of a given set of dynamic variables $\{z_{j,k,t}\}$ can also be facilitated by (6.19). In detail, we can compare (6.1) with the null model $\mathcal{H}_0 : \mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{p}\phi_{j,0}^*\mathbf{W}_j\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t$ using BIC. For $\mathcal{H}_0$, we compute the following BIC, inspired by Wang and Leng (2007):

$$\mathrm{BIC}(\mathcal{H}_0) = \log\Big(\frac{1}{T}\Big\|\mathbf{B}'\mathbf{y} - \mathbf{B}'\mathbf{V}\widehat{\boldsymbol{\phi}} - \mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})}\mathbf{vec}\big(\mathbf{I}_d\big)\Big\|^2\Big) + |\widehat{H}|\,\frac{\log(T)}{T}, \quad (6.20)$$

with the variables constructed under $\mathcal{H}_0$ and the above algorithm implemented accordingly.

Finally, to utilise the asymptotic normality of $\widehat{\boldsymbol{\phi}}_H$ for feasible inference, we require estimators for $\mathbf{R}_H$, $\mathbf{R}_\beta$, $\boldsymbol{\Sigma}_\beta$ and $\mathbf{S}_\gamma$ in Theorem 6.2. By replacing the expected values by their sample estimates, using $\widehat{H} = \{i : (\widehat{\boldsymbol{\phi}})_i \neq 0\}$ to estimate the set $H$ and leveraging all the consistency results for $\boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})$, $\widehat{\boldsymbol{\mu}}$ and $\widehat{H}$, we obtain estimators $\widehat{\mathbf{R}}_{\widehat{H}}$, $\widehat{\mathbf{R}}_\beta$ and $\widehat{\mathbf{S}}_\gamma$. For $\boldsymbol{\Sigma}_\beta$, we use a consistent estimator of $\boldsymbol{\epsilon}_t$ denoted by $\widehat{\boldsymbol{\epsilon}}_t := \mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\mathbf{W}}_t\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})$. As $\boldsymbol{\Sigma}_\beta$ involves an infinite sum, we can sum up to a cut-off $\tau^*$ after which the sum changes little, and denote the constructed estimator $\widehat{\boldsymbol{\Sigma}}_\beta$. Putting everything together, the covariance matrix of $\widehat{\boldsymbol{\phi}}_H$ can be estimated by $T^{-1}\widehat{\mathbf{R}}_{\widehat{H}}\widehat{\mathbf{S}}_\gamma\widehat{\mathbf{R}}_\beta\widehat{\boldsymbol{\Sigma}}_\beta\widehat{\mathbf{R}}_\beta'\widehat{\mathbf{S}}_\gamma'\widehat{\mathbf{R}}_{\widehat{H}}'$.

## 6.5 Change Point Detection and Estimation in Spatial Autoregressive Models

### 6.5.1 Threshold spatial autoregressive models

The early work by Tong (1978) proposes a regime switching mechanism via the threshold autoregressive model. Since then, it has been studied extensively for panel data in the past few decades (Hansen, 1999). More recently, such threshold structure is used by researchers in spatial econometrics; see, for example, threshold spatial autoregressive models for cross-sectional data by Deng (2018) and Li and Lin (2024), and spatial panel data models with threshold effects also on regression coefficient by Meng and Yang (2023). One benefit of the framework introduced in this chapter is that threshold variables can be directly adapted into (6.1). As a simple example, we consider

$$\mathbf{y}_t = \begin{cases} \boldsymbol{\mu}^* + \phi_1^*\mathbf{W}_1\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & q_t \leq \gamma^*, \\ \boldsymbol{\mu}^* + \phi_2^*\mathbf{W}_2\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & q_t > \gamma^*. \end{cases} \tag{6.21}$$

This is a spatial autoregressive model with regime switching on the spatial weight matrix, where $q_t$ is some observed threshold variable with an unknown threshold value $\gamma^*$. By rewriting (6.21) in the form of (6.1), with $z_{1,1,t} := \mathbb{1}_{\{q_t \leq \gamma^*\}}$ and $z_{2,1,t} := \mathbb{1}_{\{q_t > \gamma^*\}}$, we have

$$\mathbf{y}_t = \boldsymbol{\mu}^* + z_{1,1,t}\phi_1^*\mathbf{W}_1\mathbf{y}_t + z_{2,1,t}\phi_2^*\mathbf{W}_2\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t. \tag{6.22}$$

We consider the estimation of the threshold value $\gamma^*$. Suppose there is a domain of possible threshold values $\Gamma = [\gamma_{\text{min}}, \gamma_{\text{max}}]$, a standard approach in threshold models is to search the minimum regression error over the intersection $\Gamma \cap \{q_1, \ldots, q_T\}$; see, for example, Deng (2018). Our framework provides an alternative approach. Denote the elements in the intersection by

$\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_L$, with $L \equiv |\Gamma \cap \{q_1, \ldots, q_T\}|$, and let $\gamma^*$ be identified as one of them. Let $z_{1,l,t} := \mathbb{1}_{\{q_t \leq \gamma_l\}}$ and $z_{2,l,t} := \mathbb{1}_{\{q_t > \gamma_l\}}$ for all $l \in [L]$. Then, we can consider a spatial autoregressive model such that

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{l=1}^{L} z_{1,l,t}\phi_{1,l}^* \mathbf{W}_1 \mathbf{y}_t + \sum_{l=1}^{L} z_{2,l,t}\phi_{2,l}^* \mathbf{W}_2 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \tag{6.23}$$

where $\phi_{1,l}^*$ and $\phi_{2,l}^*$ would be nonzero[2] only for $\gamma_l = \gamma^*$. The threshold value can be selected consistently in one step by the oracle property of our adaptive LASSO estimator in Theorem 6.2. We present this result in Corollary 6.1. Given the sparse solution $(\widehat{\phi}_{1,l}, \widehat{\phi}_{2,l})_{l \in [L]}$ of (6.23), we can re-estimate all model parameters. Note that such an approach remains applicable for $L$ growing with $T$. In practice, the order of $L$ may not fulfil (R10), but we can circumvent this issue by a sequential procedure. See Remark 6.1 in Section 6.5.2 for more details.

Our framework also allows us to consider a spatial autoregressive model with regimes switching on the spatial correlation coefficients, similar to Li (2022):

$$\mathbf{y}_t = \begin{cases} \boldsymbol{\mu}^* + \phi_1^* \mathbf{W}_1 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & q_t \leq \gamma^*, \\ \boldsymbol{\mu}^* + \phi_2^* \mathbf{W}_1 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & q_t > \gamma^*. \end{cases} \tag{6.24}$$

To estimate the parameters in (6.24), Li (2022) uses quasi maximum likelihood (QML) estimators and traverses over a finite parameter space for the threshold value $\gamma^*$. In contrast, a one-step estimation is again feasible by our framework. To this end, we read (6.24) in the form,

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \Big(\phi_{1,0}^* + \sum_{l=1}^{L} z_{1,l,t}\phi_{1,l}^*\Big) \mathbf{W}_1 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \tag{6.25}$$

where $z_{1,l,t}$ for $l \in [L]$ is as previously defined. With our adaptive LASSO estimators, only $\widehat{\phi}_{1,0}$ and one $\widehat{\phi}_{1,l}$ such that $\gamma_l = \gamma^*$ are expected to be nonzero. The consistency of such estimator for $\gamma^*$ is included in Corollary 6.1.

**Corollary 6.1** *(Threshold value estimation consistency). Given all the assumptions in Theorem 6.2:*
*(a) For model* (6.21)*, $\gamma^*$ can be consistently estimated by the set of estimators $\{\gamma_l : \widehat{\phi}_{1,l} \neq 0, \widehat{\phi}_{2,l} \neq 0\}$, where $\{\widehat{\phi}_{1,l}, \widehat{\phi}_{2,l}\}_{l \in [L]}$ is the adaptive LASSO solution for $\{\phi_{1,l}^*, \phi_{2,l}^*\}_{l \in [L]}$ in* (6.23).
*(b) For model* (6.24)*, $\gamma^*$ can be consistently estimated by the set of estimators $\{\gamma_l : \widehat{\phi}_{1,l} \neq 0\}$, where $\{\widehat{\phi}_{1,l}\}_{l \in [L]}$ is the adaptive LASSO solution for $\{\phi_{1,l}^*\}_{l \in [L]}$ in* (6.25).

[2] In practice, we often have no prior information on which spatial weight matrix corresponds to the regime $q_t \leq \gamma^*$. This can be resolved in (6.23) by writing $\big(\sum_{l=1}^{L} z_{1,l,t}\phi_{1,l}^*\big)$ and $\big(\sum_{l=1}^{L} z_{2,l,t}\phi_{2,l}^*\big)$ as $\big(\phi_{1,0}^* + \sum_{l=1}^{L} z_{1,l,t}\phi_{1,l}^*\big)$ and $\big(\phi_{2,0}^* + \sum_{l=1}^{L} z_{2,l,t}\phi_{2,l}^*\big)$, respectively.

In fact, our framework (6.1) can be applied to spatial autoregressive models with more complicated threshold structures in the spatial weight matrix $\mathbf{W}_t^*$, i.e., regimes from multiple threshold variables with multiple threshold values. For illustrations, consider the following spatial autoregressive model with $(k+1)$ regimes, where $k$ can be unknown:

$$\mathbf{y}_t = \begin{cases} \boldsymbol{\mu}^* + \phi_1^* \mathbf{W}_1 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & q_t \leq \gamma_1^*, \\ \boldsymbol{\mu}^* + \phi_2^* \mathbf{W}_1 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & \gamma_1^* < q_t \leq \gamma_2^*, \\ \cdots \\ \boldsymbol{\mu}^* + \phi_k^* \mathbf{W}_1 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & q_t > \gamma_k^*. \end{cases} \tag{6.26}$$

Model (6.26) can be written in the form of (6.25), with the consistency of $\{\phi_{1,l}^*\}_{l\in[L]}$ guaranteed by Corollary 6.2. This also implies that the estimation on $k$ is consistent.

**Corollary 6.2** *(Consistency on the number of threshold regimes and multiple threshold values estimation). Let all the assumptions in Theorem 6.2 hold. For model* (6.26)*, let* $\{\widehat{\phi}_{1,l}\}_{l\in[L]}$ *denote the adaptive LASSO solution for* $\{\phi_{1,l}^*\}_{l\in[L]}$ *in* (6.25)*. Then,* $\widehat{k} := |\{\gamma_l : \widehat{\phi}_{1,l} \neq 0\}|$ *estimates k consistently. Moreover, for every* $i \in [\widehat{k}]$*, the* $i$*-th smallest element in* $\{\gamma_l : \widehat{\phi}_{1,l} \neq 0\}$ *estimates* $\gamma_i^*$ *consistently.*

As we often have limited prior knowledge on the parameter space in practice, we recommend using our framework in an exploratory way. This should help researchers discover more reasonable threshold structures in the data.

### 6.5.2 Spatial autoregressive models with structural change points

Structural changes in the relationship of variables in econometric models have been studied extensively in the literature; see, for example, Sengupta (2017) and Barigozzi and Trapani (2020). As a second example demonstrating the applicability of our framework, we consider the spatial autoregressive model with a structural change:

$$\mathbf{y}_t = \begin{cases} \boldsymbol{\mu}^* + \phi_1^* \mathbf{W}_1 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & t \leq t^*, \\ \boldsymbol{\mu}^* + \phi_2^* \mathbf{W}_2 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & t > t^*, \end{cases} \tag{6.27}$$

where $t^*$ is some unknown change location. Similar to the threshold model example, this can also be expressed in the form of (6.22), now with $z_{1,1,t} := \mathbb{1}_{\{t\leq t^*\}}$ and $z_{2,1,t} := \mathbb{1}_{\{t>t^*\}}$. It worths to notice, despite the models taking the same form, the dynamic variables $z_{1,1,t}$ and $z_{2,1,t}$ are random in the threshold model but non-random in the change point model here. Recall that our main results hold for both types of $\{z_{j,k,t}\}$; see Section 6.3.1 for a more detailed discussion.

To estimate the change location $t^*$, Li (2018) calculates the quasi maximum likelihood for each possible change location and sets the maximiser as the estimator. When the set of possible change locations is large, this approach requires a significant number of model fittings. Using our framework, it is again possible to estimate $t^*$ in one go. Let $\mathcal{T}$ denote the set of all candidate change point locations such that $\mathcal{T} = \{t_1, \dots, t_{|\mathcal{T}|}\}$. Then, we rewrite model (6.27) as

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{l=1}^{|\mathcal{T}|} z_{1,l,t}\phi^*_{1,l}\mathbf{W}_1\mathbf{y}_t + \sum_{l=1}^{|\mathcal{T}|} z_{2,l,t}\phi^*_{2,l}\mathbf{W}_2\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \tag{6.28}$$

where $z_{1,l,t} := \mathbb{1}_{\{t \le t_l\}}$ and $z_{2,l,t} := \mathbb{1}_{\{t > t_l\}}$. We then follow the same argument used in the threshold model below (6.23) and the consistency of our change location estimate is again guaranteed.

**Corollary 6.3** *(Consistency on change location estimation). Let all the assumptions in Theorem 6.2 hold, with L replaced by $|\mathcal{T}|$. Consider* (6.27) *and assume that $t^* \in \mathcal{T}$. The change location $t^*$ can be consistently estimated by the set of estimators $\{l \in [|\mathcal{T}|] : \widehat{\phi}_{1,l} \neq 0,\, \widehat{\phi}_{2,l} \neq 0\}$, where $\{\widehat{\phi}_{1,l}, \widehat{\phi}_{2,l}\}_{l\in[|\mathcal{T}|]}$ is the adaptive LASSO solution for $\{\phi^*_{1,l}, \phi^*_{2,l}\}_{l\in[|\mathcal{T}|]}$ in* (6.28).

Similar to (6.26), we can also consider a multiple change model:

$$\mathbf{y}_t = \begin{cases} \boldsymbol{\mu}^* + \sum_{j=1}^{p} \phi^*_{j,1}\mathbf{W}_j\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & t \le t_1^*, \\ \boldsymbol{\mu}^* + \sum_{j=1}^{p} \phi^*_{j,2}\mathbf{W}_j\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & t_1^* < t \le t_2^*, \\ \cdots \\ \boldsymbol{\mu}^* + \sum_{j=1}^{p} \phi^*_{j,k}\mathbf{W}_j\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, & t > t_k^*. \end{cases} \tag{6.29}$$

This model allows for $k$ change points in $\mathbf{W}_t^*$ consisting of $p$ spatial weight candidates, with the number of change points $k$ unknown. The result below confirms the consistency of the estimations in both change point numbers and locations.

**Corollary 6.4** *(Consistency on the estimations for the number of changes and the change locations). Given a set of all candidate change point locations $\mathcal{T}$ and assume that $t_i^* \in \mathcal{T}$ for all $i \in [k]$. Let all the assumptions in Theorem 6.2 hold, with L replaced by $|\mathcal{T}|$. For model* (6.29), *let $\{\widehat{\phi}_{j,l}\}_{j\in[p],l\in[|\mathcal{T}|]}$ denote the adaptive LASSO solution for $\{\phi^*_{j,l}\}_{j\in[p],l\in[|\mathcal{T}|]}$ in*

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{p}\Big(\phi^*_{j,0} + \sum_{l=1}^{|\mathcal{T}|} z_{j,l,t}\phi^*_{j,l}\Big)\mathbf{W}_j\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t,$$

*where $z_{j,l,t} := \mathbb{1}_{\{t\le t_l\}}$ with $t_l$ being the $l$-th element of $\mathcal{T}$ for $j \in [p], l \in \big[|\mathcal{T}|\big]$. Write $\widehat{\mathcal{T}} := \{l \neq 0 : \widehat{\phi}_{j,l} \neq 0$ for some $j \in [p]\}$. Then, $\widehat{k} := |\widehat{\mathcal{T}}|$ estimates $k$ consistently and, for every $i \in [\widehat{k}]$, the $i$-th smallest element in $\widehat{\mathcal{T}}$ estimates $t_i^*$ consistently.*

**Remark 6.1** *Throughout Section 6.5.2, the size of the set $\mathcal{T}$ is restricted by Assumption (R10). Specifying appropriate $\mathcal{T}$ requires prior information, which might be infeasible in practice. Without this, a set $\mathcal{T}$ with a large size may violate Assumption (R10). The order of $L$ in Section 6.5.1 raises a similar concern.*

*For practical implementation, we may resort to a divide-and-conquer scheme in the following manner. Consider model* (6.27) *for instance. We first partition $\mathcal{T}$ into subsets $\mathcal{T} = \cup_j \mathcal{T}_j$ such that each $\mathcal{T}_j$ satisfies (R10) (with $L$ replaced by $|\mathcal{T}_j|$). On each subset, we run the estimation algorithm and obtain all identified potential change locations within the subset. Then, we aggregate all those locations into a set $\widetilde{\mathcal{T}}$, which can be shown, using Corollary 6.3 on each $\mathcal{T}_j$, to satisfy (R10) (with $L$ replaced by $|\widetilde{\mathcal{T}}|$). Finally, we estimate* (6.28) *with $\mathcal{T}$ replaced by $\widetilde{\mathcal{T}}$ to determine the change point location. Simulations in Section 6.6.1 confirms the effectiveness of this scheme.*

## 6.6 Numerical Studies

### 6.6.1 Simulations

In this subsection, we conduct Monte Carlo simulations to demonstrate the performance of our estimators. For the general setting, we consider

$$\mathbf{y}_t = \Big\{\mathbf{I}_d - \big(0.2 + 0.2z_{1,1,t} + 0z_{1,2,t}\big)\mathbf{W}_1 - \big(0 + 0z_{2,1,t} + 0.3z_{2,2,t}\big)\mathbf{W}_2\Big\}^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big), \tag{6.30}$$

where $\boldsymbol{\mu}^*$ and $\boldsymbol{\beta}^*$ are vectors of 1's, $\mathbf{W}_1$ is generated to have two neighbours ahead and two behind as in Kelejian and Prucha (1998), and $\mathbf{W}_2$ is a contiguity matrix with off-diagonal entries being i.i.d. Bernoulli $(0.2)$. The true parameter is $\boldsymbol{\phi}^* = (0.2,\ 0.2,\ 0,\ 0,\ 0,\ 0.3)'$. The disturbance $\boldsymbol{\epsilon}_t$ is jointly Gaussian with its variance-covariance matrix having 1 on the diagonal and each upper triangular entries 0.1 with probability 0.2 and 0 otherwise. For any row of $\mathbf{W}_1$ or $\mathbf{W}_2$ with row sum exceeding one, we divide each entry by the $L_1$ norm of the row. We use independent standard normal random variables for the dynamic variables $\{z_{1,1,t}\}$, $\{z_{1,2,t}\}$, $\{z_{2,1,t}\}$ and $\{z_{2,2,t}\}$. The covariate matrix $\mathbf{X}_t$ has three columns, with each entry generated as independent standard normal, except that the third column is endogenous by adding $0.5\,\boldsymbol{\epsilon}_t$. Let $\mathbf{X}_{\text{exo},t}$ be $\mathbf{X}_t$ with the disturbance part removed, then the instruments can be set as $\mathbf{B}_t = \big[\mathbf{X}_{\text{exo},t}, \mathbf{W}_1\mathbf{X}_{\text{exo},t}, \mathbf{W}_2\mathbf{X}_{\text{exo},t}\big]$. The tuning parameter for the adaptive LASSO is selected by minimising the BIC in (6.19).

We experiment $d = 25, 50, 75$ and $T = 50, 100, 150$, with each setting repeated 1000 times. Results are presented in Table 6.1. In there, MSE is the mean squared error; specificity is the proportion of true zeros estimated as zeros; sensitivity is the proportion of nonzeros estimated as nonzeros. The MSE results corroborate the consistency of parameter estimation in Theorem 6.1 and 6.3, while the specificity and sensitivity results corroborate the sparsity consistency in Theorem 6.2. Zeros in $\phi^*$ can be selected with high accuracy, yet the sensitivity results suggest a mild over-identification of zeros. Both increasing the spatial dimension $d$ and time span $T$ improve the performance of our estimators in general, except that when $T$ increases from 100 to 150 and $d = 25$, all measures get a bit worse, which is similarly seen in Table 1 of Lam and Souza (2020). This might suggest the issue of a data set with unbalanced dimensions in practice.

| | $T = 50$ | | | $T = 100$ | | | $T = 150$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $d = 25$ | $d = 50$ | $d = 75$ | $d = 25$ | $d = 50$ | $d = 75$ | $d = 25$ | $d = 50$ | $d = 75$ |
| $\widehat{\phi}$ MSE | .002 | .001 | .002 | .000 | .000 | .000 | .003 | .001 | .000 |
| | (.008) | (.001) | (.001) | (.000) | (.000) | (.000) | (.007) | (.000) | (.000) |
| $\widehat{\beta}$ MSE | .087 | .029 | .080 | .005 | .008 | .001 | .088 | .013 | .005 |
| | (.434) | (.011) | (.025) | (.004) | (.002) | (.001) | (.278) | (.005) | (.002) |
| $\widehat{\mu}$ MSE | .039 | .038 | .044 | .013 | .010 | .011 | .036 | .010 | .008 |
| | (.124) | (.010) | (.010) | (.006) | (.002) | (.002) | (.083) | (.003) | (.002) |
| $\widehat{\phi}$ Specificity | .992 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | .994 | 1.000 | 1.000 |
| | (.064) | (.000) | (.000) | (.000) | (.000) | (.000) | (.048) | (.000) | (.000) |
| $\widehat{\phi}$ Sensitivity | .921 | .983 | .992 | .910 | .961 | .991 | .873 | .983 | .956 |
| | (.133) | (.069) | (.048) | (.132) | (.099) | (.051) | (.148) | (.070) | (.103) |

Table 6.1: Simulation results for the general setting (6.30). Mean and standard deviation (in brackets) of the corresponding error measures over 1000 repetitions are presented.

To better illustrate the asymptotic normality for $\widehat{\phi}$ in Theorem 6.2, we use the same data generating mechanism as above with $(T, d) = (200, 50)$, except that $\phi^* = (0, \text{-}0.5, 0.5, 0, 0, 0)'$, that $\mathbf{X}_t$ is exogenous and that $\epsilon_t$ has a diagonal covariance matrix. For ease of presentation, we fix $\widehat{H} = \{2, 3\}$ which is the index set of true nonzero parameters. The remaining components of the covariance matrix are estimated according to the last part of Section 6.4. Figure 6.1 displays the histogram of $T^{1/2}(\widehat{\mathbf{R}}_{\widehat{H}}\widehat{\mathbf{S}}_\gamma\widehat{\mathbf{R}}_\beta\widehat{\mathbf{\Sigma}}_\beta\widehat{\mathbf{R}}'_\beta\widehat{\mathbf{S}}'_\gamma\widehat{\mathbf{R}}'_{\widehat{H}})^{-1/2}(\widehat{\phi}_{\widehat{H}} - \phi^*_{\widehat{H}})$. The plots show good normal approximation to the distribution of this quantity and confirm the result in Theorem 6.2. Some discrepancies are present on the tails, potentially due to insufficient dimensions $T$ and $d$. We leave potential improvements to future studies.

**Change point analysis**

We now demonstrate the performance of our dynamic framework with structural changes as described in Section 6.5.2, with simulation results for the threshold model in Section 6.5.1 included in Section 6.7. We first consider a spatial autoregressive model with a single change

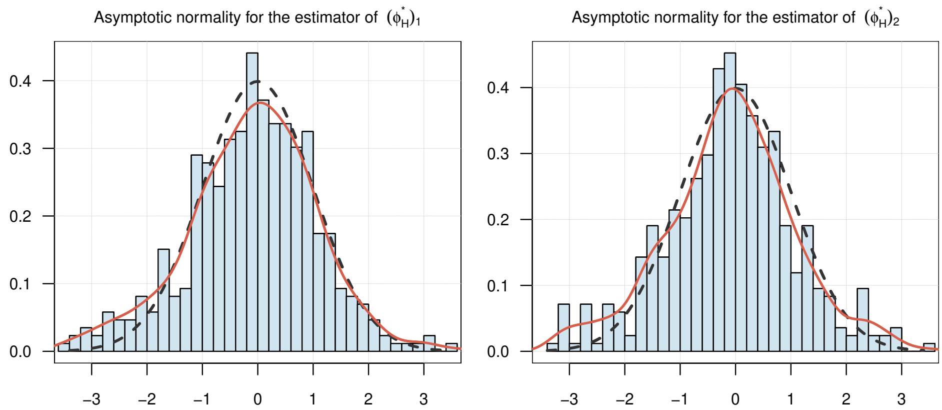


Figure 6.1: Histogram of $T^{1/2}(\widehat{\mathbf{R}}_{\widehat{H}}\widehat{\mathbf{S}}_{\gamma}\widehat{\mathbf{R}}_{\beta}\widehat{\mathbf{\Sigma}}_{\beta}\widehat{\mathbf{R}}'_{\beta}\widehat{\mathbf{S}}'_{\gamma}\widehat{\mathbf{R}}'_{\widehat{H}})^{-1/2}(\widehat{\boldsymbol{\phi}}_{\widehat{H}} - \boldsymbol{\phi}^*_{\widehat{H}})$ for $(T, d) = (200, 50)$, shown for the first coordinate (left panel) and the second coordinate (right panel). The red curves are the empirical density, and the black dotted curves are the density for $\mathcal{N}(0, 1)$.

such that

$$\mathbf{y}_t = \big(\mathbf{I}_d - 0.3 \cdot \mathbb{1}_{\{t\leq 30\}}\mathbf{W}_1 - 0.3 \cdot \mathbb{1}_{\{t>30\}}\mathbf{W}_2\big)^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big), \tag{6.31}$$

where $\{\mathbf{W}_1, \mathbf{W}_2, \boldsymbol{\mu}^*, \mathbf{X}_t, \boldsymbol{\beta}^*\}$ are constructed in the same way as in (6.30). To showcase the robustness of our estimators under heavy-tailed noise, we generate $\boldsymbol{\epsilon}_t$ by i.i.d. $\mathcal{N}(0, 1)$ and $t_6$, respectively. The model (6.31) represents a change on the true spatial weight matrix at $t = 30$ from $0.3\,\mathbf{W}_1$ to $0.3\,\mathbf{W}_2$. We then fit a model

$$\mathbf{y}_t = \Big(\mathbf{I}_d - \sum_{l=1}^{\lfloor T/\Delta\rfloor - 1} z_{1,l,t}\phi^*_{1,l}\mathbf{W}_1 - \sum_{l=1}^{\lfloor T/\Delta\rfloor - 1} z_{2,l,t}\phi^*_{2,l}\mathbf{W}_2\Big)^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big), \quad \text{where} \tag{6.32}$$

$$z_{1,l,t} = \mathbb{1}_{\{t\leq t_l\}}, \quad z_{2,l,t} = \mathbb{1}_{\{t>t_l\}}, \quad t_l = \Delta \cdot l.$$

We consider a grid of candidate change locations, spaced at intervals of $\Delta = 5$. From (6.31), every $(\phi^*_{1,l}, \phi^*_{2,l})$ equals $(0, 0)$ except the one $l$ such that $t_l = 30$. Table 6.2 displays the results with each $(T, d)$ setting specified, and Table 6.3 shows the results under a stronger change signal such that the true spatial weight matrix changes from $0.5\,\mathbf{W}_1$ to $0.5\,\mathbf{W}_2$.

Results for both weak and strong change signals display similar patterns. Unsurprisingly, the change detection slightly suffers from fat-tailed noise and weaker signals. The accuracy of detection benefits from the increasing spatial dimension. A larger $T$ seems to undermine the $\widehat{\boldsymbol{\phi}}$ True-Unique measure, but this is essentially due to more dynamic variables ($z_{1,l,t}$ and $z_{2,l,t}$) used in the setting.

| $\epsilon_t$ | i.i.d. $\mathcal{N}(0,1)$ | | | | i.i.d. $t_6$ | | | |
|---|---|---|---|---|---|---|---|---|
| $(T,d)$ | $(50,25)$ | $(50,50)$ | $(100,50)$ | $(100,75)$ | $(50,25)$ | $(50,50)$ | $(100,50)$ | $(100,75)$ |
| $\widehat{\phi}$ MSE | .008 | .004 | .004 | .003 | .010 | .004 | .004 | .003 |
| | (.007) | (.003) | (.002) | (.002) | (.009) | (.003) | (.003) | (.002) |
| $\widehat{\phi}$ True-Unique | .611 | .965 | .668 | .938 | .559 | .944 | .647 | .898 |
| | (.488) | (.184) | (.472) | (.242) | (.497) | (.231) | (.479) | (.303) |

Table 6.2: Simulation results for the model (6.32) with a weak change signal. $\widehat{\phi}$ *True-Unique* is defined to be 1 if the only nonzero pair $(\phi^*_{1,l}, \phi^*_{2,l})$ corresponds to $t_l = 30$. Mean and standard deviation (in brackets) of the corresponding error measures over 500 repetitions are presented.

| $\epsilon_t$ | i.i.d. $\mathcal{N}(0,1)$ | | | | i.i.d. $t_6$ | | | |
|---|---|---|---|---|---|---|---|---|
| $(T,d)$ | $(50,25)$ | $(50,50)$ | $(100,50)$ | $(100,75)$ | $(50,25)$ | $(50,50)$ | $(100,50)$ | $(100,75)$ |
| $\widehat{\phi}$ MSE | .006 | .001 | .003 | .001 | .007 | .001 | .004 | .001 |
| | (.011) | (.001) | (.006) | (.001) | (.012) | (.003) | (.007) | (.002) |
| $\widehat{\phi}$ True-Unique | .826 | .994 | .846 | .990 | .759 | .984 | .811 | .972 |
| | (.379) | (.077) | (.361) | (.100) | (.428) | (.126) | (.392) | (.166) |

Table 6.3: Simulation results for the model (6.32) with a strong change signal. Refer to Table 6.2 for the definition on $\widehat{\phi}$ *True-Unique*. Mean and standard deviation (in brackets) of the corresponding error measures over 500 repetitions are presented.

**Experiments on the divide-and-conquer scheme in Remark 6.1**

We now demonstrate the numerical performance of the divide-and-conquer scheme in Remark 6.1. Under $(T,d) = (100,75)$, consider an extension of (6.31) with two change points:

$$\mathbf{y}_t = \big(\mathbf{I}_d - 0.8\,\mathbb{1}_{\{t\le 30\}}\mathbf{W}_1 + 0.9\,\mathbb{1}_{\{t\le 60\}}\mathbf{W}_1 + 0.9\,\mathbb{1}_{\{t>60\}}\mathbf{W}_2\big)^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big). \quad (6.33)$$

That is, the spatial weight matrix changes from -$0.1\mathbf{W}_1$ to -$0.9\mathbf{W}_1$ at $t = 30$, followed by a change from -$0.9\mathbf{W}_1$ to -$0.9\mathbf{W}_2$ at $t = 60$. Suppose it is only known a priori that on $\mathcal{T} = \{2,4,\ldots,98,100\}$[3], the spatial weight matrix might change from $\mathbf{W}_1$ to $\mathbf{W}_2$ and the spatial correlation coefficients might change as well. We wish to estimate the number of changes and the change locations. Following Remark 6.1, we construct the ordered sets $\mathcal{T}_1 = \{2,4,\ldots,20\}$, $\mathcal{T}_2 = \{20,22,\ldots,40\}$, $\mathcal{T}_3 = \{40,42,\ldots,60\}$, $\mathcal{T}_4 = \{60,62,\ldots,80\}$ and $\mathcal{T}_5 = \{80,82,\ldots,100\}$. Note that we add "overlaps" between adjacent sets to circumvent falsely identifying change candidates at the margin. Now, the size of each set is $11$, which is

[3]Change point identified at the last observed time point $T = 100$ represents no change in the structure.

reasonable according to the numerical results in Tables 6.2 and 6.3. For each $j \in [5]$, consider[4]

$$\mathbf{y}_t = \Big(\mathbf{I}_d - \sum_{l=1}^{|\mathcal{T}_j|} z_{j,1,l,t}\phi^*_{j,1,l}\mathbf{W}_1 - \sum_{l=1}^{|\mathcal{T}_j|} z_{j,2,l,t}\phi^*_{j,2,l}\mathbf{W}_2\Big)^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big), \quad \text{where}$$

$$z_{j,1,l,t} := \mathbb{1}_{\{t \le (\mathcal{T}_j)_l\}}, \quad z_{j,2,l,t} := \mathbb{1}_{\{t > (\mathcal{T}_j)_l\}}, \quad \text{with } (\mathcal{T}_j)_l \text{ being the } l\text{-th element in } \mathcal{T}_j.$$

As in Remark 6.1, all time points corresponding to nonzero estimates of $\phi^*_{j,1,l}$ or $\phi^*_{j,2,l}$ for $j \in [5], l \in [|\mathcal{T}_j|]$ are identified and collected to form a refined candidate set $\widetilde{\mathcal{T}}$, where marginal time points $\{20, 40, 60, 80\}$ are discarded if they are not identified in all $\mathcal{T}_j$ containing them. Finally, consider

$$\mathbf{y}_t = \Big(\mathbf{I}_d - \sum_{l=1}^{|\widetilde{\mathcal{T}}|} z_{1,l,t}\phi^*_{1,l}\mathbf{W}_1 - \sum_{l=1}^{|\widetilde{\mathcal{T}}|} z_{2,l,t}\phi^*_{2,l}\mathbf{W}_2\Big)^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big), \quad \text{where}$$

$$z_{1,l,t} = \mathbb{1}_{\{t \le (\widetilde{\mathcal{T}})_l\}}, \quad z_{2,l,t} = \mathbb{1}_{\{t > (\widetilde{\mathcal{T}})_l\}}, \quad \text{with } (\widetilde{\mathcal{T}})_l \text{ being the } l\text{-th element in } \widetilde{\mathcal{T}}.$$

Then, change points are estimated as the timestamps corresponding to nonzero estimates for $\phi^*_{1,l}$ or $\phi^*_{2,l}$. The histogram for the estimated change locations over 500 repetitions is shown in the left panel of Figure 6.2 and is encouraging. To further quantify the performance of our scheme, we use the Adjusted Rand index (ARI) of the estimated time segmentation against the truth[5] (Rand, 1971; Hubert and Arabie, 1985), a measure frequently used by change point researchers (Wang and Samworth, 2017). The average ARI across all runs is $0.901$, again suggesting that our scheme is performing very well.

We also consider (6.33) under no change or, equivalently, one change at $t = 100$:

$$\mathbf{y}_t = \big(\mathbf{I}_d + 0.9 \cdot \mathbb{1}_{\{t \le T\}}\mathbf{W}_1\big)^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big). \tag{6.34}$$

We follow the same exact procedure to estimate (6.33), and the histogram for the estimated change points over 500 runs is shown in the right panel of Figure 6.2. In $98\%$ of the experiments, exactly $T = 100$ is identified, meaning no change is detected, which corresponds to a $2\%$ false change discovery rate. Furthermore, the average ARI[6] is $0.980$.

[4]Note that the last dynamic variable in $\mathcal{T}_5$ for $\mathbf{W}_2$, $z_{5,2,11,t}$, is 0 for all $t$, so we directly specify $\phi^*_{5,2,11}$ as 0.

[5]The estimated time segmentation assigns the same labels to time points between the estimated change points, with different labels assigned after each change point. For true time partitioning with changes at $\{30, 60\}$, the intervals $\{1, 2, \ldots, 30\}$, $\{31, 32, \ldots, 60\}$, and $\{61, 62, \ldots, 100\}$ are labelled as 1, 2, and 3, respectively.

[6]The true time partition, when there are no changes, labels all time points as 1.

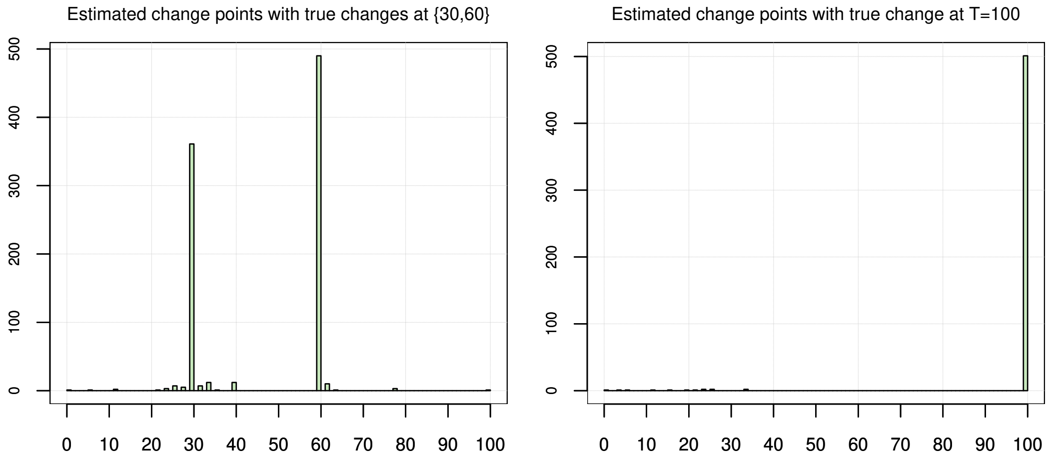


Figure 6.2: Histograms of estimated change locations under true models (6.33) (left panel) and (6.34) (right panel). Both experiments are repeated 500 times.

### 6.6.2 Real data analysis: enterprise monthly profits

In this case study, we use our proposed model to analyse the total profits of enterprises for a selection of provincial regions in China. Our panel data covers $T = 86$ monthly periods from March 2016 to August 2024 and 25 provinces and 4 direct-administered municipalities (i.e., $d = 29$); see Section 6.7 for more details. The set of covariates (all standardised) consists of Consumer Price Index (*CPI*), Purchasing Price Index for industrial producers (*PPI*) and output of electricity (*elec*). The data is available at the National Bureau of Statistics of China: https://data.stats.gov.cn/english/.

We consider three spatial weight matrix candidates, with each row standardised by its $L_1$ norm if the row sum exceeds one: inverse distance matrix using inverse of geographical distances between locations computed by the Geodesic WGS-84 System ($\mathbf{W}_1$), contiguity matrix ($\mathbf{W}_2$), municipality matrix such that all direct-administered municipalities are neighbours ($\mathbf{W}_3$).

We treat the covariates as exogenous for two reasons: *CPI* and *PPI* are largely independent of the internal economic activities specific to enterprises within each province or municipality, and electricity supply as a public utility is often price inelastic. Using the aforementioned covariates and spatial weight matrices, we first specify a time-invariant spatial autoregressive model as our null model:

$$\textit{profit}_t = \boldsymbol{\mu}^* + (\phi^*_{1,0}\mathbf{W}_1 + \phi^*_{2,0}\mathbf{W}_2 + \phi^*_{3,0}\mathbf{W}_3)\,\textit{profit}_t + (\textit{CPI}_t, \textit{PPI}_t, \textit{elec}_t)\,\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t. \quad (6.35)$$

We estimate the parameters in (6.35) using our adaptive LASSO estimators. The estimated coefficients $\{\widehat{\phi}_{1,0}, \widehat{\phi}_{2,0}, \widehat{\phi}_{3,0}, \widehat{\boldsymbol{\beta}}\}$ are presented in Table 6.4, together with the BIC computed according to (6.20). The table also shows the standard errors of $\widehat{\phi}_{1,0}$ and $\widehat{\phi}_{3,0}$ based on Theorem 6.2. The respective p-values for testing $\widehat{\phi}_{1,0} = 0$ and $\widehat{\phi}_{3,0} = 0$ are both less than $0.0001$, revealing some spillovers among the neighbours of provinces and municipalities. Interestingly, $\widehat{\phi}_{3,0}$ suggests a negative spillover effect among the four direct-administered municipalities, which could be explained by that the enterprises within municipalities are main competitors in the market.

| | $\widehat{\phi}_{1,0}$ | $\widehat{\phi}_{2,0}$ | $\widehat{\phi}_{3,0}$ | $\widehat{\beta}_{CPI}$ | $\widehat{\beta}_{PPI}$ | $\widehat{\beta}_{elec}$ | BIC |
|---|---|---|---|---|---|---|---|
| Null model | 15.184 | .000 | -.285 | .021 | .053 | .394 | 2.790 |
| | (3.725) | | (.066) | | | | |

Table 6.4: Estimated coefficients for model (6.35), with standard errors (in brackets) computed according to the last part of Section 6.4. $\widehat{\beta}_{CPI}$, $\widehat{\beta}_{PPI}$ and $\widehat{\beta}_{elec}$ denote the estimates of $\boldsymbol{\beta}^*$ corresponding to $CPI_t$, $PPI_t$ and $elec_t$, respectively.

Hereafter, we refer to (6.35) as the null model. The rest of the analysis is performed in an exploratory fashion such that we consider spatial autoregressive models of the form (6.1) with some $l_j$ and dynamic variables $\{z_{j,k,t}\}$. We consider the following models:

$$\begin{aligned}
&\text{Model 1}: \mathit{profit}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{3}\Big(\phi^*_{j,0} + \sum_{k=1}^{15}\phi^*_{j,k}\mathbb{1}\{t \le 5+5k\}\Big)\mathbf{W}_j\,\mathit{profit}_t + (\mathit{CPI}_t, \mathit{PPI}_t, \mathit{elec}_t)\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t;\\
&\text{Model 2}: \mathit{profit}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{3}\Big(\phi^*_{j,0} + \sum_{k=1}^{9}\phi^*_{j,k}\mathbb{1}\{\mathrm{sd}(\mathit{profit}_{t-5}) \le \gamma_k\}\Big)\mathbf{W}_j\,\mathit{profit}_t\\
&\qquad\qquad + (\mathit{CPI}_t, \mathit{PPI}_t, \mathit{elec}_t)\,\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t,\\
&\qquad \text{where } (\gamma_1, \ldots, \gamma_9) = (.177, .193, .202, .207, .217, .219, .231, .250, .302);\\
&\text{Model 3}: \mathit{profit}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{3}\Big(\phi^*_{j,0} + \sum_{k=1}^{5}\phi^*_{j,k}\mathbb{1}\{t \text{ divides } 2k\}\Big)\mathbf{W}_j\,\mathit{profit}_t + (\mathit{CPI}_t, \mathit{PPI}_t, \mathit{elec}_t)\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t;\\
&\text{Model 4}: \mathit{profit}_t = \boldsymbol{\mu}^* + \Big(\phi^*_{1,0}\alpha_{3,1}(\mathbf{W}_1) + \phi^*_{2,0}\alpha_{3,2}(\mathbf{W}_1) + \phi^*_{3,0}\alpha_{3,3}(\mathbf{W}_1)\Big)\,\mathit{profit}_t\\
&\qquad\qquad + (\mathit{CPI}_t, \mathit{PPI}_t, \mathit{elec}_t)\,\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t.
\end{aligned} \tag{6.36}$$

Model 1 represents a spatial autoregressive model with the spatial weight matrix potentially changing at $(10, 15, 20, \ldots, 75, 80)$. Model 2 is a self-exciting threshold spatial autoregressive model with the standard deviation of $\mathit{profit}_{t-5}$ as the threshold variable. The sequence of threshold value is in fact the empirical quantile, from $10\%$ to $90\%$, of $\mathrm{sd}(\mathit{profit}_{t-5})$. Model 3 is similar to the null model but accounts for monthly spillovers for lags of two, four, six, eight and ten months. Model 4 adapts our framework to time-invariant and nonlinear spatial weight matrices, where $\alpha_{3,1}(\mathbf{W}_1)$, $\alpha_{3,2}(\mathbf{W}_1)$ and $\alpha_{3,3}(\mathbf{W}_1)$ denote the matrices formed by series expansion using the order-3 normalised Laguerre functions (inspired by Sun (2016)) based on

$\{(\mathbf{W}_1)_{ij}^{-1}\}_{i,j\in[29]}$.

Table 6.5 reports the estimated parameters and BIC for each model. The nonzero $\widehat{\phi}_{1,9}$ for Model 1 corresponds to a change in the spillovers featured by $\mathbf{W}_1$ in March 2020, potentially suggesting inactive economic activities due to COVID-19 starting at the beginning of 2020.

| | nonzero $\widehat{\phi}_{j,k}$ | | | | $\widehat{\beta}_{CPI}$ | $\widehat{\beta}_{PPI}$ | $\widehat{\beta}_{elec}$ | BIC |
|---|---|---|---|---|---|---|---|---|
| Model 1 | $\widehat{\phi}_{1,9} = 18.030$ | | | | .044 | .038 | .309 | 2.978 |
| | (13.601) | | | | | | | |
| Model 2 | $\widehat{\phi}_{1,3} = 9.721$ | $\widehat{\phi}_{1,9} = 18.997$ | $\widehat{\phi}_{3,7} = -.522$ | | .030 | .038 | .368 | 2.744 |
| | (3.487) | (.717) | (.001) | | | | | |
| Model 3 | $\widehat{\phi}_{1,0} = 15.424$ | $\widehat{\phi}_{2,1} = .335$ | $\widehat{\phi}_{2,3} = -.331$ | $\widehat{\phi}_{3,5} = -.401$ | .023 | .052 | .398 | 2.747 |
| | (.000) | (.000) | (.000) | (.000) | | | | |
| Model 4 | $\widehat{\phi}_{1,0} = .048$ | $\widehat{\phi}_{2,0} = -.034$ | $\widehat{\phi}_{3,0} = .114$ | | .043 | .056 | .512 | 2.848 |
| | (.001) | (.003) | (.001) | | | | | |

Table 6.5: Estimated coefficients for different models specified in (6.36), with standard errors (in brackets) computed according to the last part of Section 6.4. Refer to Table 6.4 for the definitions of $\widehat{\beta}_{CPI}$, $\widehat{\beta}_{PPI}$ and $\widehat{\beta}_{elec}$.

Model 2 has the best BIC among all models shown here, including the null model. From Table 6.5, four threshold regions are identified as

$$\widehat{\mathbf{W}}_t = \begin{cases} 28.718\,\mathbf{W}_1 - 0.522\,\mathbf{W}_3, & \mathrm{sd}(\mathit{profit}_{t-5}) \le 0.202; \\ 18.997\,\mathbf{W}_1 - 0.522\,\mathbf{W}_3, & 0.202 < \mathrm{sd}(\mathit{profit}_{t-5}) \le 0.231; \\ 18.997\,\mathbf{W}_1, & 0.231 < \mathrm{sd}(\mathit{profit}_{t-5}) \le 0.302; \\ 0, & \mathrm{sd}(\mathit{profit}_{t-5}) > 0.302. \end{cases}$$

For better illustration, the series of $\widehat{\mathbf{W}}_t$ among Beijing, Shanghai and Guangdong are plotted in Figure 6.3. We see that the spillovers between Beijing and Shanghai is more significant than their respective spillovers with Guangdong.

On Model 3, Table 6.5 suggests that the effect of $\mathbf{W}_1$ (representing domestic spillovers) remains constant, that of $\mathbf{W}_2$ (representing more local spillovers) persists every two months but roughly cancels out every half year, and the "municipality spillover" by $\mathbf{W}_3$ occurs every December. The various spillover patterns featured by the expert spatial weight matrices are intriguing and warrant further investigation.

Model 4 considers a time-invariant spillover effect. An example of the estimated spatial weight matrix is displayed in Figure 6.4. It depicts how the spillovers diminish with the geographical distance. Lastly, the analysis on the total profits data serves to demonstrate our proposed spatial autoregressive framework. More comprehensive investigations are required to further understand the spatial relations among industrial enterprises in Chinese provinces

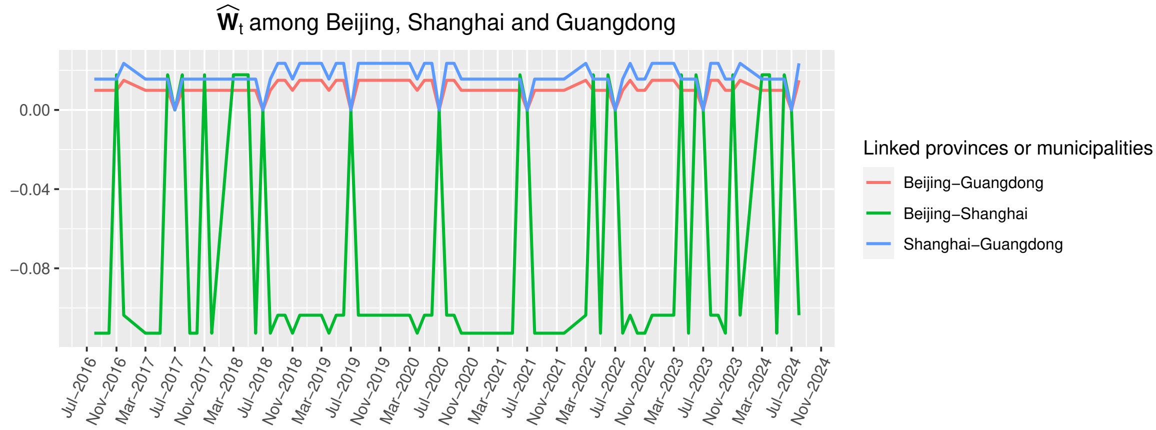


Figure 6.3: Illustration of $\widehat{\mathbf{W}}_t$ of Model 2 in (6.36) among Beijing, Shanghai and Guangdong, from August 2016 to August 2024.

and direct-administered municipalities.

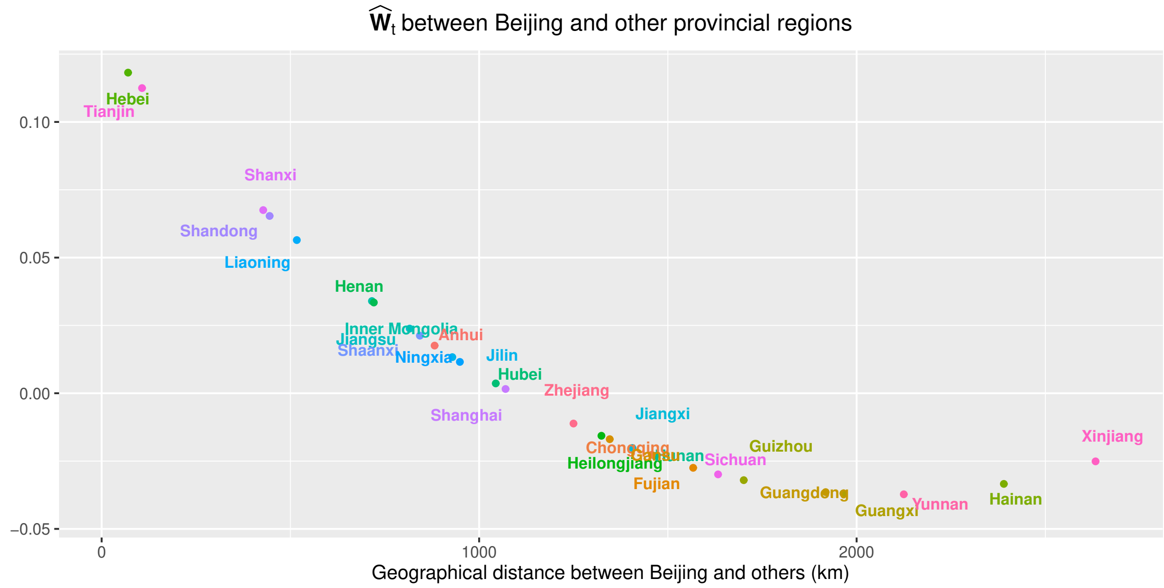


Figure 6.4: Illustration of (time-invariant) $\widehat{\mathbf{W}}_t$ of Model 4 in (6.36) between Beijing and other provincial regions, against their geographical distances.

## 6.7 Additional Details and Simulations

### Additional explanations for Section 6.2.2

Note from (6.12) and the definition of $\mathbb{B}$ in (6.8), we have

$$
\begin{aligned}
&\mathbb{B}'\Big(\mathbf{I}_d \otimes \Big\{\Big(\mathbf{I}_T \otimes \Big\{(\mathbf{y}^\nu)'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\Big\}\Big)(\mathbf{X}_1,\ldots,\mathbf{X}_T)'\Big\}\Big)\mathbf{vec}\big(\mathbf{I}_d\big) \\
&= T^{-1/2}d^{-a/2}\Big(\mathbf{I}_d \otimes \Big\{(\mathbf{B}_1-\bar{\mathbf{B}},\ldots,\mathbf{B}_T-\bar{\mathbf{B}})(\mathbf{I}_T\otimes\boldsymbol{\gamma}) \\
&\quad \cdot \Big(\mathbf{I}_T \otimes \Big\{(\mathbf{y}^\nu)'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\Big\}\Big)(\mathbf{X}_1,\ldots,\mathbf{X}_T)'\Big\}\Big)\mathbf{vec}\big(\mathbf{I}_d\big) \\
&= T^{-1/2}d^{-a/2}\Big(\mathbf{I}_d \otimes \Big\{\sum_{t=1}^{T}(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}(\mathbf{y}^\nu)'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\mathbf{X}_t'\Big\}\Big)\mathbf{vec}\big(\mathbf{I}_d\big) \\
&= T^{-1/2}d^{-a/2}\mathbf{vec}\big(\sum_{t=1}^{T}(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}(\mathbf{y}^\nu)'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\mathbf{X}_t'\big) \\
&= T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t\otimes(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\mathbf{vec}\big((\mathbf{y}^\nu)'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\big) \\
&= T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t\otimes(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{y}^\nu.
\end{aligned}
$$

Similarly, by the definition of $\mathbf{Y}_W$,

$$
\begin{aligned}
&\mathbb{B}'\Big(\mathbf{I}_d \otimes \Big\{\Big(\mathbf{I}_T \otimes \Big\{\Big(\sum_{j=1}^{p}\sum_{k=0}^{l_j}\phi_{j,k}(\mathbf{y}^\nu_{j,k})'(\mathbf{W}_j^\otimes)'\Big) \\
&\qquad \cdot \mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\Big\}\Big)(\mathbf{X}_1,\ldots,\mathbf{X}_T)'\Big\}\Big)\mathbf{vec}\big(\mathbf{I}_d\big) \\
&= T^{-1/2}d^{-a/2}\Big(\mathbf{I}_d \otimes \Big\{(\mathbf{B}_1-\bar{\mathbf{B}},\ldots,\mathbf{B}_T-\bar{\mathbf{B}})(\mathbf{I}_T\otimes\boldsymbol{\gamma}) \\
&\quad \cdot \Big(\mathbf{I}_T \otimes \Big\{\Big(\sum_{j=1}^{p}\sum_{k=0}^{l_j}\phi_{j,k}(\mathbf{y}^\nu_{j,k})'(\mathbf{W}_j^\otimes)'\Big) \\
&\quad \cdot \mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\Big\}\Big)(\mathbf{X}_1,\ldots,\mathbf{X}_T)'\Big\}\Big)\mathbf{vec}\big(\mathbf{I}_d\big) \\
&= T^{-1/2}d^{-a/2}\Big(\mathbf{I}_d \otimes \Big\{\sum_{t=1}^{T}(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}\Big(\sum_{j=1}^{p}\sum_{k=0}^{l_j}\phi_{j,k}(\mathbf{y}^\nu_{j,k})'(\mathbf{W}_j^\otimes)'\Big) \\
&\quad \cdot \mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\mathbf{X}_t'\Big\}\Big)\mathbf{vec}\big(\mathbf{I}_d\big) \\
&= T^{-1/2}d^{-a/2}\mathbf{vec}\Big\{\sum_{t=1}^{T}(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}(\boldsymbol{\phi}'\mathbf{Y}_W')\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big]^{-1}\mathbf{X}_t'\Big\}
\end{aligned}
$$

$$= T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\big[\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big]^{-1}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{Y}_W\boldsymbol{\phi}.$$

With (6.12), the term inside the squared Euclidean norm of (6.13) can be further written:

$$\begin{aligned}
&\mathbf{B}'\mathbf{y} - \mathbf{B}'\mathbf{V}\boldsymbol{\phi} - \mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\phi})}\mathbf{vec}\big(\mathbf{I}_d\big)\\
&= \mathbf{B}'\mathbf{y} - \mathbf{B}'\mathbf{V}\boldsymbol{\phi} - \mathbf{B}'\Big(\mathbf{I}_d \otimes \Big\{\Big(\mathbf{I}_T \otimes \Big\{(\mathbf{y}^{\nu})'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big]^{-1}\Big\}\Big)\\
&\quad \cdot \big(\mathbf{X}_1, \dots, \mathbf{X}_T\big)'\Big\}\Big)\mathbf{vec}\big(\mathbf{I}_d\big)\\
&\quad + \mathbf{B}'\Big(\mathbf{I}_d \otimes \Big\{\Big(\mathbf{I}_T \otimes \Big\{\Big(\sum_{j=1}^{p}\sum_{k=0}^{l_j}\phi_{j,k}(\mathbf{y}^{\nu}_{j,k})'(\mathbf{W}_j^{\otimes})'\Big)\\
&\quad \cdot \mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big[\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big]^{-1}\Big\}\Big)(\mathbf{X}_1, \dots, \mathbf{X}_T)'\Big\}\Big)\mathbf{vec}\big(\mathbf{I}_d\big)\\
&= \mathbf{B}'\mathbf{y} - T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\big[\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big]^{-1}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{y}^{\nu}\\
&\quad - \Big\{\mathbf{B}'\mathbf{V} - T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\big[\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big]^{-1}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{Y}_W\Big\}\boldsymbol{\phi}\\
&= \mathbf{B}'\mathbf{y} - \boldsymbol{\Xi}\mathbf{y}^{\nu} - (\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W)\boldsymbol{\phi}.
\end{aligned} \tag{6.37}$$

## Experiments on the threshold spatial autoregressive models

In the following, we demonstrate numerical results for the threshold spatial autoregressive model in Section 6.5.1. Consider

$$\mathbf{y}_t = \big(\mathbf{I}_d - 0.3\, z_{1,t}\mathbf{W}_1 - 0.8\, z_{2,t}\mathbf{W}_2\big)^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big), \tag{6.38}$$

where $\mathbf{W}_1$, $\mathbf{W}_2$, $\boldsymbol{\mu}^*$, $\mathbf{X}_t$, $\boldsymbol{\beta}^*$ and $\boldsymbol{\epsilon}_t$ are constructed in the same way as those in (6.30), $z_{1,t} = \mathbb{1}_{\{q_t \le \gamma^*\}}$ and $z_{2,t} = \mathbb{1}_{\{q_t > \gamma^*\}}$ for some threshold variables and values $q_t$ and $\gamma^*$. Hence, (6.38) represents a threshold spatial autoregressive model with changes in both coefficients and spatial weight matrices. To estimate the parameters and the threshold value simultaneously, we fit a model of the form

$$\mathbf{y}_t = \Big(\mathbf{I}_d - \sum_{l=1}^{19} z_{1,l,t}\phi^*_{1,l}\mathbf{W}_1 - \sum_{l=1}^{19} z_{2,l,t}\phi^*_{2,l}\mathbf{W}_2\Big)^{-1}\big(\boldsymbol{\mu}^* + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t\big), \quad \text{where} \tag{6.39}$$

$$z_{1,l,t} = \mathbb{1}_{\{q_t \le \widehat{\gamma}_l\}}, \quad z_{2,l,t} = \mathbb{1}_{\{q_t > \widehat{\gamma}_l\}}, \quad \widehat{\gamma}_l = (5\% \cdot l)\text{-th empirical quantile of } \{q_t\}_{t \in [T]}.$$

As discussed in Section 6.5.1, we expect $(\phi^*_{1,1}, \dots, \phi^*_{1,19})$ are all zero except for the one corresponding to $z_{1,l,t}$ such that $\widehat{\gamma}_l$ is the nearest to $\gamma^*$. Denote the index of such $\widehat{\gamma}_l$ as $l^*$. Similarly,

$(\phi_{2,1}^*, \ldots, \phi_{2,19}^*)$ are all zero except for the one corresponding to $z_{1,l^*,t}$. It should also hold true that $(\phi_{1,l^*}^*, \phi_{2,l^*}^*) \approx (0.3, 0.8)$. We experiment two types of threshold variables:

1) *(AR(5))* $q_t$ is AR(5) with i.i.d. $\mathcal{N}(0, 1)$ innovations, with $\gamma^* = 0.3$;

2) *(Self-exciting on mean)* $q_t = \mathbf{1}'\mathbf{y}_{t-1}/d$, i.e., the regime changes in a self-exciting manner on the mean of the previous data point, with $\gamma^* = 1.5$.

Results for $d = 50, 75$ and $T = 100, 150$ are presented in Table 6.6 with $\widehat{\boldsymbol{\phi}}$ and $\widehat{\gamma}_l$ estimated from (6.39), where

$$\widehat{\boldsymbol{\phi}} \text{ MSE} := \text{MSE of } \widehat{\boldsymbol{\phi}} \text{ with } \boldsymbol{\phi} \text{ all zero except } (\phi_{1,l^*}^*, \phi_{2,l^*}^*) \text{ set as } (0.3, 0.8),$$
$$\widehat{\boldsymbol{\phi}} \text{ True-Unique} := \mathbb{1}\{\widehat{\phi}_{1,l^*} \text{ and } \widehat{\phi}_{2,l^*} \text{ are both nonzero in } \widehat{\boldsymbol{\phi}} \text{ uniquely}\},$$
$$\widehat{\gamma}_l \text{ MSE} := \text{MSE of } \widehat{\gamma}_l \text{ with true threshold value } \gamma^*.$$

$\widehat{\boldsymbol{\phi}}$ True-Unique is the key to demonstrating the validity of our algorithm as it relates to both specificity and sensitivity. More importantly, it measures if the estimated threshold value is unique. On computing $\widehat{\gamma}_l$-MSE with multiple $\widehat{\gamma}_l$ values, i.e., the intersecting index set is not a singleton, we choose the $l$ corresponding to the largest $\widehat{\phi}_{1,l^*}$. Table 6.6 confirms that our procedure is capable of estimating the threshold value and other model parameters in one go. Although the estimator of $\gamma^*$ is coarse up to the $5\%$ empirical quantile of the threshold variable, Table 6.6 shows that increasing the data dimensions improves the performance of $\widehat{\gamma}_l$. In practice, re-estimation using a finer grid based on such initial threshold estimator could be performed.

| $q_t$ setting | AR(5) | | | | Self-exciting on mean | | | |
|---|---|---|---|---|---|---|---|---|
| $(T, d)$ | $(100, 50)$ | $(100, 75)$ | $(150, 50)$ | $(150, 75)$ | $(100, 50)$ | $(100, 75)$ | $(150, 50)$ | $(150, 75)$ |
| $\widehat{\boldsymbol{\phi}}$ MSE | .002 | .003 | .017 | .020 | .009 | .007 | .003 | .001 |
| | (.002) | (.003) | (.013) | (.011) | (.011) | (.010) | (.006) | (.003) |
| $\widehat{\boldsymbol{\phi}}$ True-Unique | .562 | .439 | .706 | .844 | .537 | .621 | .797 | .938 |
| | (.497) | (.500) | (.456) | (.363) | (.499) | (.485) | (.403) | (.242) |
| $\widehat{\gamma}_l$ MSE | .343 | .066 | .207 | .025 | .128 | .031 | .080 | .005 |
| | (.870) | (.309) | (.707) | (.193) | (.881) | (.126) | (.802) | (.023) |

Table 6.6: Simulation results for the threshold model (6.39). Mean and standard deviation (in brackets) of the corresponding error measures over 500 repetitions are presented.

## Additional details for the enterprise profits data in Section 6.6.2

Due to missingness, we exclude January and February data. The 25 provinces included in the data are Hebei, Shanxi, Inner Mongolia, Liaoning, Jilin, Heilongjiang, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Hainan, Sichuan, Guizhou, Yunnan, Shaanxi, Gansu, Ningxia, Xinjiang, and the 4 direct-administered municipalities are Beijing, Tianjin, Shanghai and Chongqing.

A snippet of the total profits for August 2024 is shown in Figure 6.5, where the map is produced using the R package `hchinamap`. From the estimation of our null model, it is revealed from $\widehat{\boldsymbol{\mu}}$ that Guangdong, Beijing, Jiangsu and Shanghai have significantly larger spatial fixed effects than other provinces or municipalities.

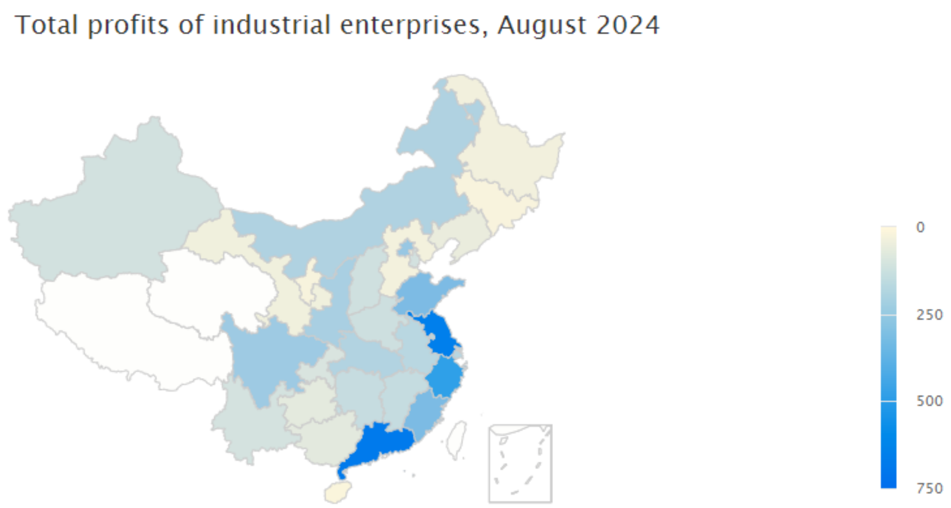


Figure 6.5: Illustration of the total profits of industrial enterprises within considered Chinese provinces and direct-administered municipalities, in 100 million yuans.

## 6.8 Proof of Theorems and Auxiliary Results

To prove our main theorems, recall first $B_{t,ij}$ and $X_{t,ij}$ represent the $(i,j)$ entry of $\mathbf{B}_t$ and $\mathbf{X}_t$ respectively. Define $\mathcal{M} = \bigcap_{i=1}^{13} \mathcal{A}_i$, where

$$
\begin{aligned}
\mathcal{A}_1 &= \left\{ \max_{i,q\in[d]} \max_{j,l\in[r]} \left| \frac{1}{T} \sum_{t=1}^{T} [B_{t,ij} X_{t,ql} - \mathbb{E}(B_{t,ij} X_{t,ql})] \right| < c_T \right\}, \\
\mathcal{A}_2 &= \left\{ \max_{i,q\in[d]} \max_{j\in[r]} \left| \frac{1}{T} \sum_{t=1}^{T} B_{t,ij} \epsilon_{t,q} \right| < c_T \right\}, \\
\mathcal{A}_3 &= \left\{ \max_{j\in[r]} \left| \frac{1}{T} \sum_{t=1}^{T} \sum_{q=1}^{d} B_{t,qj} \epsilon_{t,q} \right| < c_T d^{\frac{1}{2}+\frac{1}{2w}} \right\}, \\
\mathcal{A}_4 &= \left\{ \max_{i\in[d]} \max_{j\in[r]} \left| \bar{B}_{\cdot,ij} - \mathbb{E}[B_{t,ij}] \right| < c_T \right\}, \\
\mathcal{A}_5 &= \left\{ \max_{q\in[d]} \left| \bar{\epsilon}_{\cdot,q} \right| < c_T \right\}, \\
\mathcal{A}_6 &= \left\{ \max_{i\in[d]} \max_{j\in[r]} \left| \bar{X}_{\cdot,ij} \right| < c_T \right\}, \\
\mathcal{A}_7 &= \left\{ \max_{j\in[r]} \left| \sum_{q=1}^{d} \bar{B}_{\cdot,qj} \bar{\epsilon}_{\cdot,q} \right| < 2^{1/2} c_T d^{1/2} \log^{1/2}(T \vee d) S_\epsilon (\mu_{b,\max} + c_T) \right\}, \\
\mathcal{A}_8 &= \left\{ \max_{m\in[p]} \max_{n\in[l_m]\cup\{0\}} \max_{i,q\in[d]} \max_{j\in[v]} \max_{l\in[r]} \left| \frac{1}{T} \sum_{t=1}^{T} [z_{m,n,t} B_{t,ij} X_{t,ql} - \mathbb{E}(z_{m,n,t} B_{t,ij} X_{t,ql})] \right| < c_T \right\}, \\
\mathcal{A}_9 &= \left\{ \max_{m\in[p]} \max_{n\in[l_m]\cup\{0\}} \max_{i\in[d]} \max_{j\in[r]} \left| \frac{1}{T} \sum_{t=1}^{T} z_{m,n,t} X_{t,ij} \right| < c_T \right\}, \\
\mathcal{A}_{10} &= \left\{ \max_{m\in[p]} \max_{n\in[l_m]\cup\{0\}} \max_{i,q\in[d]} \max_{j\in[v]} \left| \frac{1}{T} \sum_{t=1}^{T} z_{m,n,t} B_{t,ij} \epsilon_{t,q} \right| < c_T \right\}, \\
\mathcal{A}_{11} &= \left\{ \max_{m\in[p]} \max_{n\in[l_m]\cup\{0\}} \max_{q\in[d]} \left| \frac{1}{T} \sum_{t=1}^{T} z_{m,n,t} \epsilon_{t,q} \right| < c_T \right\}, \\
\mathcal{A}_{12} &= \left\{ \max_{m\in[p]} \max_{n\in[l_m]\cup\{0\}} \max_{i\in[d]} \max_{j\in[v]} \left| \frac{1}{T} \sum_{t=1}^{T} [z_{m,n,t} B_{t,ij} - \mathbb{E}(z_{m,n,t} B_{t,ij})] \right| < c_T \right\}, \\
\mathcal{A}_{13} &= \left\{ \max_{m\in[p]} \max_{n\in[l_m]\cup\{0\}} \left| \frac{1}{T} \sum_{t=1}^{T} z_{m,n,t} \right| < c_T \vee z_{\max} \right\},
\end{aligned}
\tag{6.40}
$$

where $\bar{B}_{\cdot,ij} := T^{-1} \sum_{t=1}^{T} B_{t,ij}$, $\bar{X}_{\cdot,ij} := T^{-1} \sum_{t=1}^{T} X_{t,ij}$, $\bar{\epsilon}_{\cdot,q} := T^{-1} \sum_{t=1}^{T} \epsilon_{t,q}$, $\mu_{b,\max} := \max_{i,j} \left| \mathbb{E}[B_{t,ij}] \right|$ being a constant implied by Assumption (M1), and $z_{\max}$ is the upper bound for $\{z_{j,k,t}\}$ implied in (M2) with $z_{\max} = 1$ for $\{z_{j,0,t}\}$ by default. Our main theoretical results depict the properties of estimators on the set $\mathcal{M}$ which holds with probability approaching 1 as $T, d \to \infty$ by (R10), as shown in Lemma 6.2 which is similar to Theorem S.1 of Lam and Souza (2020).

To prove Lemma 6.2, we first quote a Nagaev-type inequality for functional dependent data from Theorem 2(ii), 2(iii) and Section 4 of Liu et al. (2013), presented as the following lemma.

**Lemma 6.1** *For a zero mean time series process* $\mathbf{x}_t = \mathbf{f}(\mathcal{F}_t)$ *defined in* (6.17) *with dependence measure* $\theta^x_{t,q,i}$ *defined in* (6.18)*, assume* $\Theta^x_{m,2w} \leq Cm^{-\alpha}$ *as in Assumption (M1). Then there exists constants* $C_1$*,* $C_2$ *and* $C_3$ *independent of* $n, T$ *and the index* $i$ *such that*

$$\mathbb{P}\Big(\Big|\frac{1}{T}\sum_{t=1}^{T} x_{t,i}\Big| > n\Big) \leq \frac{C_1 T^{w(1/2-\widetilde{\alpha})}}{(Tn)^w} + C_2 \exp(-C_3 T^{\widetilde{\beta}} n^2),$$

*where* $\widetilde{\alpha} = \alpha \wedge (1/2 - 1/w)$ *and* $\widetilde{\beta} = (3 + 2\widetilde{\alpha}w)/(1 + w)$*.*

*Furthermore, assume another zero mean time series process* $\{\mathbf{e}_t\}$ *(can be the same process* $\{\mathbf{x}_t\}$*) with* $\Theta^e_{m,2w}$ *as in Assumption (M1). Then provided* $\max_i \|x_{t,i}\|_{2w}, \max_j \|e_{t,j}\|_{2w} \leq c_0 < \infty$ *where* $c_0$ *is a constant, the above Nagaev-type inequality holds for the product process* $\{x_{t,i}e_{t,j} - \mathbb{E}(x_{t,i}e_{t,j})\}$*.*

*The above results also hold for any zero mean non-stationary process* $\mathbf{x}_t = \mathbf{f}_t(\mathcal{F}_t)$ *provided that* $\max_i \|x_{t,i}\|_w < \infty$ *and* $\Theta^{x,*}_{m,2w} \leq Cm^{-\alpha}$*, where* $\Theta^{x,*}_{m,q}$ *is uniform tail sum defined in the following with* $\ddot{x}_{t,i}$ *being the coupled version of* $x_{t,i}$ *as in* (6.18)*:*

$$\Theta^{x,*}_{m,q} := \sum_{t=m}^{\infty} \max_i \theta^{x,*}_{t,q,i} := \sum_{t=m}^{\infty} \max_i \sup_t \big\|x_{t,i} - \ddot{x}_{t,i}\big\|_q.$$

We present Lemma 6.2 below. Note that we assume $\alpha > 1/2 - 1/w$ which can be relaxed at the cost of more complicated rates and longer proofs presented here, and it simplifies the form of Lemma 6.1 as $w(1/2 - \widetilde{\alpha}) = \widetilde{\beta} = 1$.

**Lemma 6.2** *Let Assumptions (M1), (M2) (or (M2')), and (R2) hold and* $\alpha > 1/2 - 1/w$ *in Assumption (M1). Suppose for the application of the Nagaev-type inequality in Lemma 6.1 for the processes in* $\mathcal{M} = \bigcap_{i=1}^{13} \mathcal{A}_i$ *where* $\mathcal{A}_i$ *is defined in* (6.40)*, the constants* $C_1$*,* $C_2$ *and* $C_3$ *are the same. Then with* $g \geq \sqrt{3/C_3}$ *where* $g$ *is the constant defined in* $c_T = gT^{-1/2}\log^{1/2}(T \vee d)$*, we have*

$$\mathbb{P}(\mathcal{M}) \geq 1 - 14C_1 r^2 vL\Big(\frac{C_3}{3}\Big)^{w/2} \frac{d^2}{T^{w/2-1}\log^{w/2}(T \vee d)} - \frac{14C_2 r^2 vLd^2}{T^3 \vee d^3} - \frac{2r}{T \vee d}.$$

***Proof of Lemma 6.2.*** As shown in Theorem S.1 of Lam and Souza (2020), the tail condition in Assumption (M1) implies $\|\cdot\|_{2w}$ is bounded for the processes $\{B_{t,ij}X_{t,ql} - \mathbb{E}(B_{t,ij}X_{t,ql})\}$,

$\{B_{t,ij}\epsilon_{t,q}\}$, $\{X_{t,ij}\}$ and $\{\epsilon_{t,q}\}$, and further with Assumption (R2) we have

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_1^c) &\leq C_1 r^2 \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d^2}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 r^2 d^2}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_2^c) &\leq C_1 r \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d^2}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 r d^2}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_3^c) &\leq C_1 \Big(\frac{C_3}{3}\Big)^{w/2} \frac{r}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 r}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_4^c) &\leq C_1 r \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 r d}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_5^c) &\leq C_1 \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 d}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_6^c) &\leq C_1 r \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 r d}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_7^c) &\leq \frac{2r}{T\vee d} + \mathbb{P}(\mathcal{A}_4^c) + \mathbb{P}(\mathcal{A}_5^c).
\end{aligned}$$

Consider the remaining sets. First let Assumption (M2) hold and notice $\|\cdot\|_{2w}$ is bounded for the processes $\{z_{m,n,t}B_{t,ij}X_{t,ql} - \mathbb{E}(z_{m,n,t}B_{t,ij}X_{t,ql})\}$, $\{z_{m,n,t}B_{t,ij}\epsilon_{t,q}\}$, $\{z_{m,n,t}X_{t,ij}\}$ and $\{z_{m,n,t}\epsilon_{t,q}\}$, and their uniform tail sums satisfy the condition in Lemma 6.1. Thus, apply Lemma 6.1 first on $\mathcal{A}_8^c$ and we have by the union bound,

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_8^c) &\leq \sum_{m\in[p]}\sum_{n\in[l_m]}\sum_{i,q\in[d]}\sum_{j\in[v]}\sum_{l\in[r]} \mathbb{P}\left(\Big|\frac{1}{T}\sum_{t=1}^{T}[z_{m,n,t}B_{t,ij}X_{t,ql} - \mathbb{E}(z_{m,n,t}B_{t,ij}X_{t,ql})]\Big| \geq c_T\right)\\
&\leq d^2 r v L\Big(\frac{C_1 T}{(Tc_T)^w} + C_2\exp(-C_3 T c_T^2)\Big)\\
&\leq C_1 r v L\Big(\frac{C_3}{3}\Big)^{w/2} \frac{d^2}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 r v L d^2}{T^3\vee d^3}.
\end{aligned}$$

Similarly using Lemma 6.1, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_9^c) &\leq C_1 r L \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 r L d}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_{10}^c) &\leq C_1 v L \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d^2}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 v L d^2}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_{11}^c) &\leq C_1 L \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 L d}{T^3 \vee d^3},\\
\mathbb{P}(\mathcal{A}_{12}^c) &\leq C_1 v L \Big(\frac{C_3}{3}\Big)^{w/2} \frac{d}{T^{w/2-1}\log^{w/2}(T\vee d)} + \frac{C_2 v L d}{T^3 \vee d^3},
\end{aligned}$$

while $\mathbb{P}(\mathcal{A}_{13}^c) = 0$ by Assumption (M2). On the other hand, if we have Assumption (M2'), the above results remain valid, except that

$$\mathbb{P}(\mathcal{A}_{13}^c) \le C_1 L\Big(\frac{C_3}{3}\Big)^{w/2} \frac{1}{T^{w/2-1}\log^{w/2}(T \vee d)} + \frac{C_2 L}{T^3 \vee d^3},$$

by applying Lemma 6.1 given the bounded support and tail sum assumption in (M2'). For any $\{z_{j,0,t}\}$, we may treat it as a non-stochastic basis as in (M2) and the result follows. Lastly, by $\mathbb{P}(\mathcal{M}) \ge 1 - \sum_{i=1}^{13} \mathbb{P}(\mathcal{A}_i^c)$ we complete the proof of Lemma 6.2. □

We present the following lemma with a short proof as well, and we will utilise the defined notation $\mathbf{V}_{\mathbf{H},K}$ (which is the same definition in Theorem 6.2) in the proof of main theorems.

**Lemma 6.3** *For any $n \times d$ matrix $\mathbf{H} = (\mathbf{h}_1, \dots, \mathbf{h}_n)'$ and any $d \times K$ matrix $\mathbf{M}$, define*

$$\mathbf{V}_{\mathbf{H},K} = \begin{pmatrix} \mathbf{I}_K \otimes \mathbf{h}_1 \\ \vdots \\ \mathbf{I}_K \otimes \mathbf{h}_n \end{pmatrix}.$$

*We then have $\mathbf{H}\mathbf{M} = \big\{\mathbf{I}_n \otimes \mathbf{vec}\big(\mathbf{M}\big)'\big\}\mathbf{V}_{\mathbf{H},K}$.*

***Proof of Lemma 6.3.*** Notice that

$$\big\{\mathbf{I}_n \otimes \mathbf{vec}\big(\mathbf{M}\big)'\big\}\mathbf{V}_{\mathbf{H},K} = \begin{pmatrix} \mathbf{vec}\big(\mathbf{M}\big)'(\mathbf{I}_K \otimes \mathbf{h}_1) \\ \vdots \\ \mathbf{vec}\big(\mathbf{M}\big)'(\mathbf{I}_K \otimes \mathbf{h}_n) \end{pmatrix},$$

whose $j$-th row (as a column vector) is hence $(\mathbf{I}_K \otimes \mathbf{h}_j')\mathbf{vec}\big(\mathbf{M}\big) = \mathbf{vec}\big(\mathbf{h}_j'\mathbf{M}\big) = \mathbf{vec}\big(\mathbf{M}'\mathbf{h}_j\big) = \mathbf{M}'\mathbf{h}_j$ which is the $j$-th row of $\mathbf{H}\mathbf{M}$ indeed. □

**Remark 6.2** *With the notation of $\mathbf{V}_{\mathbf{H},K}$, we may write any $d \times K$ matrix $\mathbf{M}$ as*

$$\mathbf{M} = \mathbf{I}_d\mathbf{M} = \big\{\mathbf{I}_d \otimes \mathbf{vec}\big(\mathbf{M}\big)'\big\}\mathbf{V}_{\mathbf{I}_d,K},$$

*which will be useful if we are interested in the interaction only between $\mathbf{A}$, $\mathbf{M}$ in $\mathbf{ABM}$, with $\mathbf{A} \in \mathbb{R}^{l\times r}$ and $\mathbf{B} \in \mathbb{R}^{r\times d}$, since we have*

$$\mathbf{ABM} = (\mathbf{B}'\mathbf{A}')'\mathbf{M} = \mathbf{V}_{\mathbf{B}',l}'\big\{\mathbf{I}_d \otimes \mathbf{vec}\big(\mathbf{A}'\big)\big\}\mathbf{M} = \mathbf{V}_{\mathbf{B}',l}'\big\{\mathbf{I}_d \otimes \mathbf{vec}\big(\mathbf{A}'\big)\mathbf{vec}\big(\mathbf{M}\big)'\big\}\mathbf{V}_{\mathbf{I}_d,K}.$$

*Moreover, notice that in Lemma 6.3, if $K = 1$, i.e. $\mathbf{M}$ is a vector, then $\mathbf{V}_{\mathbf{H},1} = \mathbf{vec}\big(\mathbf{H}'\big)$ and Lemma 6.3 simply coincides with the fact that*

$$\mathbf{BM} = \mathbf{vec}\big(\mathbf{M}'\mathbf{B}'\big) = \mathbf{vec}\big(\mathbf{M}'\mathbf{B}'\mathbf{I}_r\big) = \big(\mathbf{I}_r \otimes \mathbf{M}'\big)\mathbf{vec}\big(\mathbf{B}'\big) = \Big(\mathbf{I}_r \otimes \mathbf{vec}\big(\mathbf{M}\big)'\Big)\mathbf{V}_{\mathbf{B},1}.$$

*Thus,* $\mathbf{V}_{\mathbf{H},K}$ *can be seen as the "$K$-block vectorisation" of* $\mathbf{H}$*, as a generalised vectorisation.*

**Lemma 6.4** *Let the assumptions in Theorem 6.2 hold. Let* $\mathbf{R}_\beta$ *and* $\mathbf{\Sigma}_\beta$ *be defined in Theorem 6.2. For* $I_2 = \big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}T^{-2}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu$*, we have* $I_2$ *asymptotically normal with rate* $T^{-1/2}d^{-(1-b)/2}$ *such that*

$$T^{1/2}(\mathbf{R}_\beta\mathbf{\Sigma}_\beta\mathbf{R}_\beta')^{-1/2}I_2 \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\mathbf{I}_r).$$

***Proof of Lemma 6.4.*** Given any nonzero $\boldsymbol{\alpha} \in \mathbb{R}^r$ with $\|\boldsymbol{\alpha}\|_1 \leq c < \infty$, we construct below the asymptotic normality of $\boldsymbol{\alpha}'I_2$ which is $T^{1/2}d^{(1-b)/2}$-convergent. First, we have the following decomposition, with the second term dominating the first by (6.44):

$$\begin{aligned}\boldsymbol{\alpha}'I_2 &= \big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}\big(T^{-1}\mathbf{X}'\mathbf{B}^\nu - \mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\big)T^{-1}(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu \\ &\quad + \big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)T^{-1}(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu.\end{aligned}$$

Then recall $\mathbf{R}_\beta = \big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)$, we have

$$\boldsymbol{\alpha}'I_2 = \frac{1}{T}\sum_{t=1}^T \boldsymbol{\alpha}'\mathbf{R}_\beta(\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))'\boldsymbol{\epsilon}_t(1+o_P(1)).$$

To construct the asymptotic normality of $\boldsymbol{\alpha}'I_2$, we want to show the as in (6.68) that

$$\sum_{t\geq 0}\Big\|P_0(\boldsymbol{\alpha}'\mathbf{R}_\beta(\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))'\boldsymbol{\epsilon}_t)\Big\|_2 < \infty, \tag{6.41}$$

so that Theorem 3 (ii) of Wu (2011) can be applied. With the definition $s_2 := \boldsymbol{\alpha}'\mathbf{R}_\beta\mathbf{\Sigma}_\beta\mathbf{R}_\beta'\boldsymbol{\alpha}$, we have $T^{1/2}s_2^{-1/2}\boldsymbol{\alpha}'I_2 \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ and hence

$$T^{1/2}(\mathbf{R}_\beta\mathbf{\Sigma}_\beta\mathbf{R}_\beta')^{-1/2}I_2 \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\mathbf{I}_r).$$

Similar to the proof of (6.68), we have (6.41) hold by Assumption (R7) and the following,

$$\begin{aligned}&\Big\|P_0(\boldsymbol{\alpha}'\mathbf{R}_\beta(\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))'\boldsymbol{\epsilon}_t)\Big\|_2 \\ &= \Big\|\boldsymbol{\alpha}'\mathbf{R}_\beta\big\{P_0((\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))')\mathbb{E}_0(\boldsymbol{\epsilon}_t)\big\} + \boldsymbol{\alpha}'\mathbf{R}_\beta\big\{\mathbb{E}_{-1}((\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))')P_0(\boldsymbol{\epsilon}_t)\big\}\Big\|_2 \\ &\leq \Big\{2\boldsymbol{\alpha}'\mathbf{R}_\beta\mathbb{E}\Big\{P_0((\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))')\,\mathbb{E}_0(\boldsymbol{\epsilon}_t)\,\mathbb{E}_0(\boldsymbol{\epsilon}_t')\,P_0(\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))\Big\}\mathbf{R}_\beta'\boldsymbol{\alpha}\Big\}^{1/2} \\ &\quad + \Big\{2\boldsymbol{\alpha}'\mathbf{R}_\beta\mathbb{E}\Big\{\mathbb{E}_{-1}((\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))')\,P_0(\boldsymbol{\epsilon}_t)\,P_0(\boldsymbol{\epsilon}_t')\,\mathbb{E}_{-1}(\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))\Big\}\mathbf{R}_\beta'\boldsymbol{\alpha}\Big\}^{1/2} \\ &= O\Big(\big\|\boldsymbol{\alpha}\big\|_1\big\|\mathbf{R}_\beta\big\|_\infty\Big) \\ &\quad \cdot \Big(d\cdot\max_{j\in[d]}\mathbb{E}^{1/2}(\mathbb{E}_0^2(\epsilon_{t,j}))\cdot\max_{j\in[d]}\max_{k\in[v]}\big\|P_0(B_{t,jk})\big\|_2 + d\cdot\sigma_{\max}\max_{j\in[d]}\big\|P_0(\epsilon_{t,j})\big\|_2\Big)\end{aligned}$$

$$= O\Big( \max_{j\in[d]} \big\|P_0^{\epsilon}(\epsilon_{t,j})\big\|_2 + \max_{j\in[d]} \max_{k\in[v]} \big\|P_0^{b}(B_{t,jk})\big\|_2 \Big),$$

where the second last equality used Assumption (R2), and the last used $\big\|\mathbf{R}_\beta\big\|_\infty = O(d^{-2}\cdot d) = O(d^{-1})$ by (6.42) and Assumption (R3).

It remains to show $\boldsymbol{\alpha}' I_2$ is of order $T^{-1/2} d^{-(1-b)/2}$. To this end, we only need to show $s_2$ is of order $d^{-(1-b)}$. First, $\mathbf{R}_\beta \mathbf{R}_\beta' = \big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}$ which has all eigenvalues of order $d^{-2}$ from (6.42) and Assumption (R3). Consider any $j$-th diagonal element of $\boldsymbol{\Sigma}_\beta$, we have

$$\begin{aligned}(\boldsymbol{\Sigma}_\beta)_{jj} &= \sum_\tau \mathbb{E}\big\{(\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))_{\cdot j}' \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t+\tau}' (\mathbf{B}_{t+\tau} - \mathbb{E}(\mathbf{B}_t))_{\cdot j}\big\} \\ &= \sum_\tau \operatorname{tr}\big\{\mathbb{E}((\mathbf{B}_{t+\tau} - \mathbb{E}(\mathbf{B}_t))_{\cdot j} (\mathbf{B}_{t+\tau} - \mathbb{E}(\mathbf{B}_t))_{\cdot j}')\, \mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t+\tau}')\big\},\end{aligned}$$

which is finite and has order exactly $d^{1+b}$ by Assumption (R8). Notice the dimension of $\boldsymbol{\Sigma}_\beta$ is $r \times r$, the order of eigenvalues of $\boldsymbol{\Sigma}_\beta$ is hence exactly $d^{1+b}$. The order of $s_2$ is $d^{b-1}$ by

$$\|\boldsymbol{\alpha}\|_1^2\, \lambda_{\min}(\mathbf{R}_\beta \mathbf{R}_\beta')\, \lambda_{\min}(\boldsymbol{\Sigma}_\beta) \le s_2 \le \|\boldsymbol{\alpha}\|_1^2\, \lambda_{\max}(\mathbf{R}_\beta \mathbf{R}_\beta')\, \lambda_{\max}(\boldsymbol{\Sigma}_\beta).$$

This completes the proof of Lemma 6.4. □

***Proof of Corollary 6.1, 6.2, 6.3 and 6.4.*** All are direct from Theorem 6.2. □

***Proof of Theorem 6.1.*** From (6.11) and (6.12), we have

$$\begin{aligned}\boldsymbol{\beta}(\boldsymbol{\phi}^*) &= \Big(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\Big)^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\Big(\mathbf{y}^\nu - \sum_{j=1}^{p}\sum_{k=0}^{l_j} \phi_{j,k}^* \mathbf{W}_j^{\otimes}\mathbf{y}_{j,k}^\nu\Big) \\ &= \Big(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\Big)^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\Big(\mathbf{1}_T \otimes \boldsymbol{\mu}^* + \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}^\nu\Big) \\ &= \boldsymbol{\beta}^* + \Big(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\Big)^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu.\end{aligned}$$

We now define a diagonal matrix $\mathbf{D}_{z_{j,k}} := \operatorname{diag}(z_{j,k,1}\mathbf{I}_d, \ldots, z_{j,k,T}\mathbf{I}_d) \in \mathbb{R}^{dT\times dT}$, with diagonal blocks $z_{j,k,1}\mathbf{I}_d, \ldots, z_{j,k,T}\mathbf{I}_d$, and $\boldsymbol{\Pi}^{*\otimes} := \big(\mathbf{I}_{Td} - \sum_{j=1}^p \sum_{k=0}^{l_j} \phi_{j,k}^* \mathbf{W}_j^{\otimes} \mathbf{D}_{z_{j,k}}\big)^{-1}$. We then have $\mathbf{y}^\nu = \boldsymbol{\Pi}^{*\otimes}\big(\mathbf{1}_T \otimes \boldsymbol{\mu}^* + \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}^\nu\big)$. Thus,

$$\begin{aligned}\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}}) &= \Big(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\Big)^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\Big(\mathbf{y}^\nu - \sum_{j=1}^{p}\sum_{k=0}^{l_j} \widetilde{\phi}_{j,k} \mathbf{W}_j^{\otimes}\mathbf{y}_{j,k}^\nu\Big) \\ &= \boldsymbol{\beta}(\boldsymbol{\phi}^*) + \Big(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\Big)^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\Big(\sum_{j=1}^{p}\sum_{k=0}^{l_j} (\phi_{j,k}^* - \widetilde{\phi}_{j,k}) \mathbf{W}_j^{\otimes}\mathbf{y}_{j,k}^\nu\Big) \\ &= \boldsymbol{\beta}(\boldsymbol{\phi}^*) + \Big(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\Big)^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\end{aligned}$$

$$\cdot\Big\{\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\mathbf{W}_j^{\otimes}\mathbf{D}_{z_{j,k}}\mathbf{\Pi}^{*\otimes}\big(\mathbf{1}_T\otimes\boldsymbol{\mu}^*+\mathbf{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}^{\nu}\big)\Big\}.$$

We can hence decompose $\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})-\boldsymbol{\beta}^*=\sum_{j=1}^{5}I_j$ where

$$I_1:=\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)-T^{-2}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big](\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})-\boldsymbol{\beta}^*),$$
$$I_2:=\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}T^{-2}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\boldsymbol{\epsilon}^{\nu},$$
$$I_3:=\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}T^{-2}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\mathbf{W}_j^{\otimes}\mathbf{D}_{z_{j,k}}\mathbf{\Pi}^{*\otimes}\mathbf{X}\boldsymbol{\beta}^*,$$
$$I_4:=\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}T^{-2}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\mathbf{W}_j^{\otimes}\mathbf{D}_{z_{j,k}}\mathbf{\Pi}^{*\otimes}\boldsymbol{\epsilon}^{\nu},$$
$$I_5:=\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}T^{-2}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\mathbf{W}_j^{\otimes}\mathbf{D}_{z_{j,k}}\mathbf{\Pi}^{*\otimes}(\mathbf{1}_T\otimes\boldsymbol{\mu}^*).$$

Notice we can take any $t\in[T]$ for $\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)$ and $\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)$ due to Assumption (M1). To bound $I_1$ to $I_4$, we first have

$$\big\|\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}\big\|_1\leq\frac{r^{1/2}}{\lambda_r\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]}\leq\frac{r^{1/2}}{d^2u^2},\tag{6.42}$$

where $\lambda_r\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]=\sigma_r^2(\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t))\geq d^2u^2$ with $u>0$ being a constant by Assumption (R3). Next, define

$$\mathbf{U}=\mathbf{I}_d\otimes T^{-1}\sum_{t=1}^{T}\mathbf{vec}\big(\mathbf{B}_t-\bar{\mathbf{B}}\big)\mathbf{x}_t',\qquad\mathbf{U}_0=\mathbf{I}_d\otimes\mathbb{E}(\mathbf{b}_t\mathbf{x}_t').$$

By Lemma 6.3, we then have

$$\frac{1}{T}\mathbf{X}'\mathbf{B}^{v}=\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'(\mathbf{B}_t-\bar{\mathbf{B}})=\frac{1}{T}\sum_{t=1}^{T}\Big\{\big(\mathbf{I}_d\otimes\mathbf{x}_t'\big)\mathbf{V}_{\mathbf{I}_d,r}\Big\}'\Big\{\big(\mathbf{I}_d\otimes\mathbf{vec}\big(\mathbf{B}_t-\bar{\mathbf{B}}\big)'\big)\mathbf{V}_{\mathbf{I}_d,v}\Big\}$$
$$=\frac{1}{T}\sum_{t=1}^{T}\mathbf{V}_{\mathbf{I}_d,r}'\big(\mathbf{I}_d\otimes\mathbf{x}_t\big)\big(\mathbf{I}_d\otimes\mathbf{vec}\big(\mathbf{B}_t-\bar{\mathbf{B}}\big)'\big)\mathbf{V}_{\mathbf{I}_d,v}=\mathbf{V}_{\mathbf{I}_d,r}'\mathbf{U}'\mathbf{V}_{\mathbf{I}_d,v}.$$

Similarly, $\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t) = \mathbf{V}_{\mathbf{I}_d,r}'\mathbf{U}_0'\mathbf{V}_{\mathbf{I}_d,v}$. Thus on the set $\mathcal{M}$ in Lemma 6.2, with (6.42) we have

$$\begin{aligned}
\|I_1\|_1 &\le \frac{r^{1/2}}{d^2u^2}\cdot\big\|\mathbf{V}_{\mathbf{I}_d,r}'\mathbf{U}_0'\mathbf{V}_{\mathbf{I}_d,v}\mathbf{V}_{\mathbf{I}_d,v}'\mathbf{U}_0 - \mathbf{V}_{\mathbf{I}_d,r}'\mathbf{U}'\mathbf{V}_{\mathbf{I}_d,v}\mathbf{V}_{\mathbf{I}_d,v}'\mathbf{U}\big\|_1\cdot\big\|\mathbf{V}_{\mathbf{I}_d,r}(\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})-\boldsymbol{\beta}^*)\big\|_1\\
&= O\Big(\frac{1}{d}\Big)\Big(\big\|\mathbf{V}_{\mathbf{I}_d,r}'(\mathbf{U}_0-\mathbf{U})'\mathbf{V}_{\mathbf{I}_d,v}\mathbf{V}_{\mathbf{I}_d,v}'\mathbf{U}_0\big\|_1+\big\|\mathbf{V}_{\mathbf{I}_d,r}'\mathbf{U}'\mathbf{V}_{\mathbf{I}_d,v}\mathbf{V}_{\mathbf{I}_d,v}'(\mathbf{U}_0-\mathbf{U})\big\|_1\Big)\big\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})-\boldsymbol{\beta}^*\big\|_1\\
&= O\Big(\frac{1}{d}\Big)\Big\{d\,\|\mathbf{U}_0-\mathbf{U}\|_{\max}\|\mathbf{U}_0\|_{\max}+d\,\|\mathbf{U}_0-\mathbf{U}\|_{\max}\Big(\|\mathbf{U}_0-\mathbf{U}\|_{\max}+\|\mathbf{U}_0\|_{\max}\Big)\Big\}\\
&\qquad\cdot\big\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})-\boldsymbol{\beta}^*\big\|_1 = O\Big(c_T\big\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})-\boldsymbol{\beta}^*\big\|_1\Big),
\end{aligned} \tag{6.43}$$

where the last equality used Assumption (R3) which implies $\|\mathbf{U}_0\|_{\max}$ is bounded by some constant, and Lemma 6.2 (using $\mathcal{A}_1,\ \mathcal{A}_4,\ \mathcal{A}_6$) that

$$\begin{aligned}
&\|\mathbf{U}-\mathbf{U}_0\|_{\max} = \Big\|\frac{1}{T}\sum_{t=1}^T\big[\mathbf{vec}\big(\mathbf{B}_t-\bar{\mathbf{B}}\big)\mathbf{x}_t'\big]-\mathbb{E}(\mathbf{b}_t\mathbf{x}_t')\Big\|_{\max}\\
&= \Big\|\frac{1}{T}\sum_{t=1}^T\mathbf{b}_t\mathbf{x}_t'-\mathbb{E}(\mathbf{b}_t\mathbf{x}_t')-\mathbf{vec}\big(\bar{\mathbf{B}}\big)\frac{1}{T}\sum_{t=1}^T\mathbf{x}_t'\Big\|_{\max}\le c_T+\Big\|\mathbf{vec}\big(\bar{\mathbf{B}}\big)\frac{1}{T}\sum_{t=1}^T\mathbf{x}_t'\Big\|_{\max}\\
&\le c_T+\Big(\big\|\mathbf{vec}\big(\bar{\mathbf{B}}\big)-\mathbb{E}[\mathbf{B}_t]\big\|_{\max}+\|\mathbb{E}[\mathbf{B}_t]\|_{\max}\Big)\Big\|\frac{1}{T}\sum_{t=1}^T\mathbf{x}_t'\Big\|_{\max}\le c_T+c_T(c_T+\mu_{b,\max}).
\end{aligned} \tag{6.44}$$

Similarly for $I_2$, we have on the set $\mathcal{M}$ that

$$\begin{aligned}
&\|I_2\|_1 \le \frac{r^{1/2}}{d^2u^2}\cdot\big\|\mathbf{V}_{\mathbf{I}_d,r}'\mathbf{U}'\mathbf{V}_{\mathbf{I}_d,v}\big\|_1\cdot\big\|T^{-1}(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu\big\|_1\\
&= O\Big(\frac{1}{d}\Big)\big\|T^{-1}(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu\big\|_1 = O\Big(\frac{1}{d}\Big)\Big\|\frac{1}{T}\sum_{t=1}^T(\mathbf{B}_t'-\bar{\mathbf{B}}')\boldsymbol{\epsilon}_t\Big\|_1\\
&= O_P\Big(\frac{1}{d}\Big)\Big\{c_Td^{\frac{1}{2}+\frac{1}{2w}}+c_Td^{1/2}\log^{1/2}(T\vee d)S_\epsilon(\mu_{b,\max}+c_T)\Big\} = O_P\big(c_Td^{-\frac{1}{2}+\frac{1}{2w}}\big),
\end{aligned} \tag{6.45}$$

where the first equality used $\mathcal{A}_1,\ \mathcal{A}_4,\ \mathcal{A}_6$ in $\mathcal{M}$, the third used $\mathcal{A}_3,\ \mathcal{A}_7$ in $\mathcal{M}$, and the last used Assumption (R10).

For $I_3$, first recall that

$$\boldsymbol{\Pi}_t^* = (\mathbf{I}_d-\mathbf{W}_t^*)^{-1} = \Big(\mathbf{I}_d-\sum_{j=1}^p\sum_{k=0}^{l_j}\phi_{j,k}^*z_{j,k,t}\mathbf{W}_j\Big)^{-1}, \tag{6.46}$$

and hence from Assumption (M2) (resp. (M2')) we have $\|\boldsymbol{\Pi}_t^*\|_\infty \le 1/(1-\eta) = O(1)$ (resp. $\|\boldsymbol{\Pi}_t^*\|_\infty = O_P(1)$) using that $(\mathbf{I}_d-\mathbf{W}_t^*)$ is strictly diagonally dominant. Then by (6.42), we have on the set $\mathcal{M}$ that

$$\begin{aligned}
\big\|I_3\big\|_1 &\le \frac{r^{1/2}}{d^2u^2}\,\big\|\mathbf{V}'_{\mathbf{I}_d,r}\mathbf{U}'\mathbf{V}_{\mathbf{I}_d,v}\big\|_1 \Big\|T^{-1}(\mathbf{B}^{\nu})'\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\mathbf{W}_j^{\otimes}\mathbf{D}_{z_{j,k}}\mathbf{\Pi}^{*\otimes}\mathbf{X}\Big\|_1 \big\|\boldsymbol{\beta}^*\big\|_1\\
&= O\Big(\frac{1}{d}\Big)\Big\|\frac{1}{T}\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})(\mathbf{B}^{\nu})'\mathbf{W}_j^{\otimes}\mathbf{D}_{z_{j,k}}\mathbf{\Pi}^{*\otimes}\mathbf{X}\Big\|_1\\
&= O\Big(\frac{1}{d}\Big)\Big\|\frac{1}{T}\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\sum_{t=1}^{T}z_{j,k,t}(\mathbf{B}'_t-\bar{\mathbf{B}}')\mathbf{W}_j\mathbf{\Pi}^*_t\mathbf{X}_t\Big\|_1\\
&= O\Big(\frac{1}{d}\Big)\Big\{\max_{q\in[r]}\max_{s\in[v]}\Big|\frac{1}{T}\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\sum_{t=1}^{T}\sum_{i=1}^{d}z_{j,k,t}\mathbf{W}'_{j,\cdot i}(\mathbf{B}_{t,\cdot s}-\bar{\mathbf{B}}_{\cdot s})\mathbf{X}'_{t,\cdot q}\mathbf{\Pi}^*_{t,i\cdot}\Big|\Big\}\\
&= O\Big(\frac{1}{d}\Big)\Big(\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1\\
&\qquad\cdot\max_{q\in[r],s\in[v],j\in[p]}\ \max_{k\in[l_j]\cup\{0\}}\ \max_{m,n\in[d]}\Big|\frac{1}{T}\sum_{t=1}^{T}z_{j,k,t}(B_{t,ms}-\bar{B}_{ms})X_{t,nq}\Big|\sum_{i=1}^{d}\|\mathbf{W}_{j,\cdot i}\|_1\|\mathbf{\Pi}^*_{t,i\cdot}\|_1\Big)\\
&= O_P\Big(\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1\big[c_T+1+c_T(c_T+\mu_{b,\max})\big]\Big) = O_P\Big(\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1\Big),
\end{aligned} \tag{6.47}$$

where the first equality used $\mathcal{A}_1$, $\mathcal{A}_4$, $\mathcal{A}_6$ in $\mathcal{M}$, the second last used $\mathcal{A}_4$, $\mathcal{A}_8$, $\mathcal{A}_9$ in $\mathcal{M}$ and Assumptions (R1) and (R3). Similarly, for $I_4$ on the set $\mathcal{M}$,

$$\begin{aligned}
\big\|I_4\big\|_1 &\le \frac{r^{1/2}}{d^2u^2}\cdot\big\|\mathbf{V}'_{\mathbf{I}_d,r}\mathbf{U}'\mathbf{V}_{\mathbf{I}_d,v}\big\|_1\cdot\big\|T^{-1}(\mathbf{B}^{\nu})'\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\mathbf{W}_j^{\otimes}\mathbf{D}_{z_{j,k}}\mathbf{\Pi}^{*\otimes}\boldsymbol{\epsilon}^{\nu}\big\|_1\\
&= O\Big(\frac{1}{d}\Big)\Big\{\max_{s\in[v]}\Big|\frac{1}{T}\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\sum_{t=1}^{T}\sum_{i=1}^{d}z_{j,k,t}\mathbf{W}'_{j,\cdot i}(\mathbf{B}_{t,\cdot s}-\bar{\mathbf{B}}_{\cdot s})\boldsymbol{\epsilon}'_t\mathbf{\Pi}^*_{t,i\cdot}\Big|\Big\}\\
&= O\Big(\frac{1}{d}\Big)\Big\{\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1\\
&\qquad\cdot\max_{s\in[v]}\max_{j\in[p]}\ \max_{k\in[l_j]\cup\{0\}}\ \max_{m,q\in[d]}\Big|\frac{1}{T}\sum_{t=1}^{T}z_{j,k,t}(B_{t,ms}-\bar{B}_{ms})\epsilon_{t,q}\Big|\sum_{i=1}^{d}\|\mathbf{W}_{j,\cdot i}\|_1\|\mathbf{\Pi}^*_{t,i\cdot}\|_1\Big\}\\
&= O_P\Big(\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1\big[c_T+c_T(c_T+\mu_{b,\max})\big]\Big) = O_P\Big(c_T\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1\Big),
\end{aligned} \tag{6.48}$$

where the first equality used $\mathcal{A}_1$, $\mathcal{A}_4$, $\mathcal{A}_6$ in $\mathcal{M}$, the second last used $\mathcal{A}_4$, $\mathcal{A}_{10}$, $\mathcal{A}_{11}$ in $\mathcal{M}$ and

Assumptions (R1) and (R3). For $I_5$ we also have on the set $\mathcal{M}$,

$$
\begin{aligned}
&\big\|I_5\big\|_1 \leq \frac{r^{1/2}}{d^2u^2}\,\big\|\mathbf{V}'_{\mathbf{I}_{d,r}}\mathbf{U}'\mathbf{V}_{\mathbf{I}_{d,v}}\big\|_1 \cdot \big\|\frac{1}{T}(\mathbf{B}^\nu)'\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\mathbf{W}_j^{\otimes}\mathbf{D}_{z_{j,k}}\mathbf{\Pi}^{*\otimes}(\mathbf{1}_T\otimes\boldsymbol{\mu}^*)\big\|_1 \\
&= O\Big(\frac{1}{d}\Big)\Big\{\max_{s\in[v]}\Big|\frac{1}{T}\sum_{j=1}^{p}\sum_{k=0}^{l_j}(\phi^*_{j,k}-\widetilde{\phi}_{j,k})\sum_{t=1}^{T}\sum_{i=1}^{d}z_{j,k,t}\mathbf{W}'_{j,\cdot i}(\mathbf{B}_{t,\cdot s}-\bar{\mathbf{B}}_{\cdot s})\boldsymbol{\mu}^{*\prime}\mathbf{\Pi}^*_{t,i\cdot}\Big|\Big\} \\
&= O\Big(\frac{1}{d}\Big)\Big\{\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1 \\
&\qquad \cdot \max_{s\in[v]}\max_{j\in[p]}\max_{k\in[l_j]\cup\{0\}}\max_{m\in[d]}\Big|\frac{1}{T}\sum_{t=1}^{T}z_{j,k,t}(B_{t,ms}-\bar{B}_{ms})\Big|\,\|\boldsymbol{\mu}^*\|_{\max}\sum_{i=1}^{d}\|\mathbf{W}_{j,\cdot i}\|_1\|\mathbf{\Pi}^*_{t,i\cdot}\|_1\Big\} \\
&= O_p\Big(\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1\big[c_T+(z_{\max}c_T\vee c_T)\big]\Big) = O_p\Big(c_T\big\|\boldsymbol{\phi}^*-\widetilde{\boldsymbol{\phi}}\big\|_1\Big),
\end{aligned}
$$

where the first equality used $\mathcal{A}_1$, $\mathcal{A}_4$, $\mathcal{A}_6$ in $\mathcal{M}$, the second last used $\mathcal{A}_4$, $\mathcal{A}_{12}$, $\mathcal{A}_{13}$ in $\mathcal{M}$ and Assumptions (R1), (R3), and $\mathbb{E}(z_{j,k,t}\mathbf{B}_t)=\mathbf{0}$ from (M2') if (M2) is not satisfied.

From (6.43), (6.45), (6.47) and (6.48), combining with Lemma 6.2, we have

$$
\big\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})-\boldsymbol{\beta}^*\big\|_1 \leq \sum_{j=1}^{5}\big\|I_j\big\|_1 = O_p\Big(\big\|\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^*\big\|_1 + c_T d^{-\frac{1}{2}+\frac{1}{2w}}\Big). \tag{6.49}
$$

For the remaining proof of Theorem 6.1, we work on the rate for $\big\|\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^*\big\|_1$. From (6.15),

$$
\begin{aligned}
\widetilde{\boldsymbol{\phi}} &= \big[(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)\big]^{-1}(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{y}-\Xi\mathbf{y}^\nu) \\
&= \big[(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)\big]^{-1}(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}^*}\mathbf{vec}(\mathbf{I}_d)+\mathbf{B}'\boldsymbol{\epsilon}-\Xi\mathbf{y}^\nu) \\
&\quad + \big[(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)\big]^{-1}(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\Xi\mathbf{Y}_W)\boldsymbol{\phi}^* + \boldsymbol{\phi}^* \\
&= \boldsymbol{\phi}^* + \big[(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)\big]^{-1}(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'\mathbf{B}'\boldsymbol{\epsilon} \\
&\quad - T^{-1/2}d^{-a/2}\cdot\big[(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)\big]^{-1}(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)' \\
&\qquad\qquad \cdot\mathbf{vec}\Big\{\sum_{t=1}^{T}(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}(\boldsymbol{\epsilon}^\nu)'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X})^{-1}\mathbf{X}'_t\Big\},
\end{aligned}
$$

where the second equality used (6.10), and the third used the fact that $\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\phi})}\mathbf{vec}(\mathbf{I}_d) = \Xi\mathbf{y}^\nu - \Xi\mathbf{Y}_W\boldsymbol{\phi}$ from (6.37) and $\boldsymbol{\beta}(\boldsymbol{\phi}^*) = \boldsymbol{\beta}^* + \big(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big)^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu$. Thus, we may decompose

$$
\begin{aligned}
&\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^* = D_1 - D_2, \text{ where} \\
&\quad D_1 = \big[(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)\big]^{-1}(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'\mathbf{B}'\boldsymbol{\epsilon}, \\
&\quad D_2 = T^{-1/2}d^{-a/2}\cdot\big[(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)\big]^{-1}(\mathbf{B}'\mathbf{V}-\Xi\mathbf{Y}_W)'
\end{aligned}
$$

$$\cdot \Big\{ \sum_{t=1}^{T} \Big( \mathbf{I}_d \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma} \Big) \mathbf{X}_t \Big\} (\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\boldsymbol{\epsilon}^{\nu}.$$

To bound the above, recall first the following definitions (from the statement of Theorem 6.2) for $j \in [p], k \in [l_j] \cup \{0\}$,

$$\mathbf{U}_{\mathbf{x},j,k} := \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}\mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\mathbf{x}_t', \;\; \mathbf{U}_{\boldsymbol{\mu},j,k} := \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}\mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\boldsymbol{\mu}^{*\prime},$$
$$\mathbf{U}_{\boldsymbol{\epsilon},j,k} := \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}\mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\boldsymbol{\epsilon}_t'.$$

Consider $\Xi\mathbf{Y}_W$, from its definition we have,

$$\begin{aligned}
\Xi\mathbf{Y}_W &= T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\big[\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big]^{-1}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})' \\
&\quad \cdot (\mathbf{W}_1^{\otimes}\mathbf{y}_{1,0}^{\nu}, \ldots, \mathbf{W}_1^{\otimes}\mathbf{y}_{1,l_1}^{\nu}, \ldots, \mathbf{W}_p^{\otimes}\mathbf{y}_{p,0}^{\nu}, \ldots, \mathbf{W}_p^{\otimes}\mathbf{y}_{p,l_p}^{\nu}) \\
&= T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\big[\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X}\big]^{-1}\mathbf{X}'\mathbf{B}^{\nu} \\
&\quad \cdot \Big\{\sum_{t=1}^{T} z_{1,0,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{W}_1\mathbf{y}_t, \cdots, \sum_{t=1}^{T} z_{p,l_p,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{W}_p\mathbf{y}_t\Big\}
\end{aligned}$$

From (6.46), we have $\mathbf{y}_t = \boldsymbol{\Pi}_t^*\boldsymbol{\mu}^* + \boldsymbol{\Pi}_t^*\mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\Pi}_t^*\boldsymbol{\epsilon}_t$. It hence holds for any $j \in [p], k \in [l_j] \cup \{0\}$ by Lemma 6.3 that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{W}_j\mathbf{y}_t \\
&= \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}\mathbf{V}_{\mathbf{W}_j',v}'\big(\mathbf{I}_d \otimes \mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\big)\big(\boldsymbol{\Pi}_t^*\boldsymbol{\mu}^* + \boldsymbol{\Pi}_t^*\mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\Pi}_t^*\boldsymbol{\epsilon}_t\big) \\
&= \mathbf{V}_{\mathbf{W}_j',v}'(\mathbf{I}_d \otimes \mathbf{U}_{\mathbf{x},j,k})\mathbf{V}_{\boldsymbol{\Pi}_t^*,r}\boldsymbol{\beta}^* + \mathbf{V}_{\mathbf{W}_j',v}'(\mathbf{I}_d \otimes \mathbf{U}_{\boldsymbol{\mu},j,k})\mathbf{vec}\big(\boldsymbol{\Pi}_t^{*\prime}\big) + \mathbf{V}_{\mathbf{W}_j',v}'(\mathbf{I}_d \otimes \mathbf{U}_{\boldsymbol{\epsilon},j,k})\mathbf{vec}\big(\boldsymbol{\Pi}_t^{*\prime}\big).
\end{aligned} \tag{6.50}$$

Consider also $\mathbf{B}'\mathbf{V}$, we have

$$\begin{aligned}
&\mathbf{B}'\mathbf{V} \\
&= T^{-1/2}d^{-a/2}\Big(\mathbf{I}_d \otimes \big\{(\mathbf{B}_1 - \bar{\mathbf{B}}, \ldots, \mathbf{B}_T - \bar{\mathbf{B}})(\mathbf{I}_T \otimes \boldsymbol{\gamma})\big\}\Big) \\
&\quad \cdot \Big\{\big[\mathbf{I}_d \otimes (z_{1,0,1}\mathbf{y}_1, \ldots, z_{1,0,T}\mathbf{y}_T)'\big]\mathbf{vec}\big(\mathbf{W}_1'\big), \ldots, \big[\mathbf{I}_d \otimes (z_{1,l_1,1}\mathbf{y}_1, \ldots, z_{1,l_1,T}\mathbf{y}_T)'\big]\mathbf{vec}\big(\mathbf{W}_1'\big), \\
&\quad \ldots,
\end{aligned}$$

$$
\begin{aligned}
&\quad \big[\mathbf{I}_d \otimes (z_{p,0,1}\mathbf{y}_1, \ldots, z_{p,0,T}\mathbf{y}_T)'\big]\,\mathbf{vec}\big(\mathbf{W}_p'\big), \ldots, \big[\mathbf{I}_d \otimes (z_{p,l_p,1}\mathbf{y}_1, \ldots, z_{p,l_p,T}\mathbf{y}_T)'\big]\,\mathbf{vec}\big(\mathbf{W}_p'\big)\Big\} \\
&= T^{-1/2}d^{-a/2}\Big\{\big[\mathbf{I}_d \otimes (\mathbf{B}_1 - \bar{\mathbf{B}}, \ldots, \mathbf{B}_T - \bar{\mathbf{B}})(\mathbf{I}_T \otimes \boldsymbol{\gamma})(z_{1,0,1}\mathbf{y}_1, \ldots, z_{1,0,T}\mathbf{y}_T)'\big]\,\mathbf{vec}\big(\mathbf{W}_1'\big), \\
&\qquad \ldots, \big[\mathbf{I}_d \otimes (\mathbf{B}_1 - \bar{\mathbf{B}}, \ldots, \mathbf{B}_T - \bar{\mathbf{B}})(\mathbf{I}_T \otimes \boldsymbol{\gamma})(z_{p,l_p,1}\mathbf{y}_1, \ldots, z_{p,l_p,T}\mathbf{y}_T)'\big]\,\mathbf{vec}\big(\mathbf{W}_p'\big)\Big\} \\
&= T^{-1/2}d^{-a/2}\Big\{\Big[\mathbf{I}_d \otimes \sum_{t=1}^{T} z_{1,0,t}(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\mathbf{y}_t'\Big]\,\mathbf{vec}\big(\mathbf{W}_1'\big), \\
&\qquad \ldots, \Big[\mathbf{I}_d \otimes \sum_{t=1}^{T} z_{p,l_p,t}(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\mathbf{y}_t'\Big]\,\mathbf{vec}\big(\mathbf{W}_p'\big)\Big\}.
\end{aligned}
$$

Similar to $\boldsymbol{\Xi}\mathbf{Y}_W$, for $j \in [p]$, $k \in [l_p] \cup \{0\}$ we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\mathbf{y}_t' = \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}(\boldsymbol{\mu}^{*\prime}\boldsymbol{\Pi}_t^{*\prime} + \boldsymbol{\beta}^{*\prime}\mathbf{X}_t'\boldsymbol{\Pi}_t^{*\prime} + \boldsymbol{\epsilon}_t'\boldsymbol{\Pi}_t^{*\prime}) \\
&= \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\,\mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\boldsymbol{\mu}^{*\prime}\boldsymbol{\Pi}_t^{*\prime} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\,\mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\mathbf{x}_t'(\boldsymbol{\beta}^* \otimes \mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\,\mathbf{vec}\big(\mathbf{B}_t - \bar{\mathbf{B}}\big)\boldsymbol{\epsilon}_t'\boldsymbol{\Pi}_t^{*\prime} \\
&= (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbf{U}_{\boldsymbol{\mu},j,k}\boldsymbol{\Pi}_t^{*\prime} + (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbf{U}_{\mathbf{x},j,k}(\boldsymbol{\beta}^* \otimes \mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime} + (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbf{U}_{\boldsymbol{\epsilon},j,k}\boldsymbol{\Pi}_t^{*\prime}.
\end{aligned}
\tag{6.51}
$$

With (6.50) and (6.51), recall from the statement of Theorem 6.2 the definitions for $\mathbf{H}_{10}$ and $\mathbf{H}_{20}$. As a heuristic for $\mathbf{H}_{10}$ and $\mathbf{H}_{20}$, they are essentially $T^{-1/2}d^{a/2}\mathbf{B}'\mathbf{V}$ and $T^{-1/2}d^{a/2}\boldsymbol{\Xi}\mathbf{Y}_W$ at the population level, respectively. For the rest of the proof for Theorem 6.1, we find the rate of $D_2$, followed by constructing the asymptotic normality of the dominating term in the expansion of $D_1$. For $D_2$, we further decompose $D_2 = F_1 + F_2 - F_3$ where

$$
\begin{aligned}
F_1 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}\Big\{(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10}) \\
&\quad - T^{-1}d^{a}(\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W)\Big\}D_2, \\
F_2 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}\big[(T^{-1/2}d^{a/2}\mathbf{B}'\mathbf{V} - \mathbf{H}_{10}) - (T^{-1/2}d^{a/2}\boldsymbol{\Xi}\mathbf{Y}_W - \mathbf{H}_{20})\big]' \\
&\quad \cdot \Big\{\frac{1}{T}\sum_{t=1}^{T}\Big(\mathbf{I}_d \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\mathbf{X}_t\Big\}(\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\boldsymbol{\epsilon}^{\nu}, \\
F_3 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}(\mathbf{H}_{20} - \mathbf{H}_{10})' \\
&\quad \cdot \Big\{\frac{1}{T}\sum_{t=1}^{T}\Big(\mathbf{I}_d \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\mathbf{X}_t\Big\}(\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}^{\nu}(\mathbf{B}^{\nu})'\boldsymbol{\epsilon}^{\nu}.
\end{aligned}
$$

To bound the L1 norm of $F_1$ to $F_3$, first observe that by Assumptions (R1) and (R4) we have

$$\begin{aligned}
\sigma_L^2(\mathbf{H}_{10}) &\geq \sigma_L^2\Bigg(\Big\{\big[\mathbf{I}_d\otimes(\boldsymbol{\gamma}'\otimes\mathbf{I}_d)\,\mathbb{E}\big(\mathbf{U}_{\mathbf{x},1,0}(\boldsymbol{\beta}^*\otimes\mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime}\big)\big]\mathbf{vec}\big(\mathbf{W}_1'\big),\\
&\qquad \ldots,\big[\mathbf{I}_d\otimes(\boldsymbol{\gamma}'\otimes\mathbf{I}_d)\,\mathbb{E}\big(\mathbf{U}_{\mathbf{x},p,l_p}(\boldsymbol{\beta}^*\otimes\mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime}\big)\big]\mathbf{vec}\big(\mathbf{W}_p'\big)\Big\}\Bigg)\\
&\geq \sigma_L^2(\mathbf{D}_W)\,\sigma_{d^2}^2\Bigg(\Big\{\big[\mathbf{I}_d\otimes(\boldsymbol{\gamma}'\otimes\mathbf{I}_d)\,\mathbb{E}\big(\mathbf{U}_{\mathbf{x},1,0}(\boldsymbol{\beta}^*\otimes\mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime}\big)\big],\\
&\qquad \ldots,\big[\mathbf{I}_d\otimes(\boldsymbol{\gamma}'\otimes\mathbf{I}_d)\,\mathbb{E}\big(\mathbf{U}_{\mathbf{x},p,l_p}(\boldsymbol{\beta}^*\otimes\mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime}\big)\big]\Big\}\Bigg)\geq Cd\cdot d^a = Cd^{1+a},
\end{aligned}$$

where $C>0$ is a generic constant. Similarly, by Assumptions (R1), (R3), (R5) and (R6),

$$\begin{aligned}
\sigma_L(\mathbf{H}_{20}) &\geq \sigma_r\Big(\mathbb{E}(\mathbf{X}_t\otimes\mathbf{B}_t\boldsymbol{\gamma})\Big)\,\sigma_r\Big(\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}\Big)\,\sigma_r\Big(\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\Big)\\
&\quad\cdot\sigma_{\min}\Big(\Big\{\mathbf{V}_{\mathbf{W}_1',r}'\mathbb{E}\big[(\mathbf{I}_d\otimes\mathbf{U}_{\mathbf{x},1,0})\mathbf{V}_{\boldsymbol{\Pi}_t^*,r}\big]\boldsymbol{\beta}^*,\ldots,\mathbf{V}_{\mathbf{W}_p',r}'\mathbb{E}\big[(\mathbf{I}_d\otimes\mathbf{U}_{\mathbf{x},p,l_p})\mathbf{V}_{\boldsymbol{\Pi}_t^*,r}\big]\boldsymbol{\beta}^*\Big\}\Big)\\
&\geq \frac{Cd^{1+a}\cdot d}{\lambda_{\max}[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)]}\cdot\sigma_v[\mathbb{E}(\ddot{\mathbf{G}})]\cdot\sigma_L(\mathbf{D}_W)\\
&\geq \frac{Cd^{1+a}\cdot d\cdot d^{1/2}}{\lambda_{\max}[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)]}\geq Cd^{1/2+a},
\end{aligned}$$

with some arbitrary constant $C>0$. Notice $\mathbf{H}_{20}$ has the smallest singular value of order larger then that for $\mathbf{H}_{10}$, so $\sigma_L^2(\mathbf{H}_{20}-\mathbf{H}_{10})\geq Cd^{1+a}$ for some $C>0$. Thus,

$$\Big\|[(\mathbf{H}_{20}-\mathbf{H}_{10})'(\mathbf{H}_{20}-\mathbf{H}_{10})]^{-1}\Big\|_1\leq\frac{L^{1/2}}{\lambda_{\min}[(\mathbf{H}_{20}-\mathbf{H}_{10})'(\mathbf{H}_{20}-\mathbf{H}_{10})]}\leq\frac{L^{1/2}}{Cd^{1+a}}. \tag{6.52}$$

Consider $F_1$ first and we hence have on $\mathcal{M}$,

$$\begin{aligned}
\|F_1\|_1 &\leq \frac{L^{1/2}\cdot L}{Cd^{1+a}}\Big\{\big\|(\mathbf{H}_{20}-T^{-1/2}d^{a/2}\boldsymbol{\Xi}\mathbf{Y}_W)+(T^{-1/2}d^{a/2}\mathbf{B}'\mathbf{V}-\mathbf{H}_{10})\big\|_{\max}\big\|\mathbf{H}_{20}-\mathbf{H}_{10}\big\|_1\\
&\qquad+\big\|T^{-1/2}d^{a/2}(\mathbf{B}'\mathbf{V}-\boldsymbol{\Xi}\mathbf{Y}_W)\big\|_{\max}\\
&\qquad\cdot\Big\|(\mathbf{H}_{20}-T^{-1/2}d^{a/2}\boldsymbol{\Xi}\mathbf{Y}_W)+(T^{-1/2}d^{a/2}\mathbf{B}'\mathbf{V}-\mathbf{H}_{10})\Big\|_1\Big\}\|D_2\|_1\\
&= O\Big\{\frac{L^{3/2}}{d^{1+a}}\big[c_T\cdot d^2+1\cdot(c_Td^2+c_Td^2)\big]\Big\}\|D_2\|_1 = O\Big(c_TL^{3/2}d^{1-a}\,\|D_2\|_1\Big),
\end{aligned} \tag{6.53}$$

where the last line used the following rates to be shown later,

$$\big\|\mathbf{H}_{20}-T^{-1/2}d^{a/2}\boldsymbol{\Xi}\mathbf{Y}_W\big\|_{\max}=O_P(c_T), \tag{6.54}$$

$$\left\|\mathbf{H}_{10} - T^{-1/2} d^{a/2} \mathbf{B}'\mathbf{V}\right\|_{\max} = O_P(c_T). \tag{6.55}$$

For neat presentation, we define the following terms whose norms will be bounded and involved in (6.54) and later,

$$\begin{aligned}
&\mathbf{A}_1 := \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}, && \mathbf{A}_1^0 := \mathbb{E}(\mathbf{X}_t \otimes \mathbf{B}_t\boldsymbol{\gamma}),\\
&\mathbf{A}_2 := \Big(\frac{1}{T}\mathbf{X}'\mathbf{B}^{\nu}\frac{1}{T}(\mathbf{B}^{\nu})'\mathbf{X}\Big)^{-1}, && \mathbf{A}_2^0 := \big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1},\\
&\mathbf{A}_3 := \frac{1}{T}\mathbf{X}'\mathbf{B}^{\nu}, && \mathbf{A}_3^0 := \mathbb{E}(\mathbf{X}_t'\mathbf{B}_t),\\
&\mathbf{A}_{4,j,k} := \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{W}_j(\boldsymbol{\Pi}_t^*\boldsymbol{\mu}^* + \boldsymbol{\Pi}_t^*\mathbf{X}_t\boldsymbol{\beta}^*), &&\\
&\mathbf{A}_{4,j,k}^0 := \mathbb{E}(\mathbf{A}_{4,j,k}), && \mathbf{A}_{5,j,k} := \frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{W}_j\boldsymbol{\Pi}_t^*\boldsymbol{\epsilon}_t.
\end{aligned}$$

On $\mathcal{M}$, we immediately have from Lemma 6.2 (using $\mathcal{A}_1, \mathcal{A}_4, \mathcal{A}_6$),

$$\left\|\mathbf{A}_1 - \mathbf{A}_1^0\right\|_{\max} = O\big(c_T + c_T(c_T + \mu_{b,\max})\big) = O(c_T), \tag{6.56}$$

which also gives $\left\|\mathbf{A}_1\right\|_{\max} \leq \left\|\mathbf{A}_1^0\right\|_{\max} + \left\|\mathbf{A}_1 - \mathbf{A}_1^0\right\|_{\max} = O(1 + c_T) = O(1)$. Hence with Assumptions (R5) and (R10), we also have on $\mathcal{M}$ that

$$\left\|\mathbf{A}_1\right\|_1 \leq \left\|\mathbf{A}_1^0\right\|_1 + \left\|\mathbf{A}_1 - \mathbf{A}_1^0\right\|_1 = O\big(d^{1+a} + c_T d^2\big) = O(d^{1+a}). \tag{6.57}$$

Similarly, with Lemma 6.2 (using $\mathcal{A}_1$) and Assumption (R3), we have on $\mathcal{M}$,

$$\left\|\mathbf{A}_3 - \mathbf{A}_3^0\right\|_1 = O(c_T d), \;\; \left\|\mathbf{A}_3^0\right\|_1 = O(d), \;\; \left\|\mathbf{A}_3\right\|_1 \leq \left\|\mathbf{A}_3 - \mathbf{A}_3^0\right\|_1 + \left\|\mathbf{A}_3^0\right\|_1 = O(d). \tag{6.58}$$

Rewrite $\mathbf{A}_2^0 = (\mathbf{A}_3^0\mathbf{A}_3^{0\prime})^{-1}$ and $\mathbf{A}_2 = (\mathbf{A}_3\mathbf{A}_3')^{-1}$, by Assumption (R3),

$$\left\|\mathbf{A}_2^0\right\|_1 \leq \frac{r^{1/2}}{\lambda_{\min}(\mathbf{A}_3^0\mathbf{A}_3^{0\prime})} = O(d^{-2}). \tag{6.59}$$

Moreover, from (6.58) we have on $\mathcal{M}$

$$\begin{aligned}
\left\|(\mathbf{A}_2^0)^{-1} - \mathbf{A}_2^{-1}\right\|_1 &\leq \left\|\mathbf{A}_3^0\mathbf{A}_3^{0\prime} - \mathbf{A}_3\mathbf{A}_3'\right\|_1 \leq \left\|\mathbf{A}_3^0 - \mathbf{A}_3\right\|_1\left\|\mathbf{A}_3^{0\prime}\right\|_1 + \left\|\mathbf{A}_3\right\|_1\left\|\mathbf{A}_3^{0\prime} - \mathbf{A}_3'\right\|_1\\
&= O(c_T d \cdot d + d \cdot c_T d) = O(c_T d^2).
\end{aligned}$$

Thus, rewrite $\mathbf{A}_2 - \mathbf{A}_2^0 = (\mathbf{A}_2 - \mathbf{A}_2^0)[(\mathbf{A}_2^0)^{-1} - \mathbf{A}_2^{-1}]\mathbf{A}_2^0 + \mathbf{A}_2^0[(\mathbf{A}_2^0)^{-1} - \mathbf{A}_2^{-1}]\mathbf{A}_2^0$, then on $\mathcal{M}$,

$$\big\|\mathbf{A}_2 - \mathbf{A}_2^0\big\|_1 = o\Big(\big\|\mathbf{A}_2 - \mathbf{A}_2^0\big\|_1\Big) + O\big(c_T d^2 \cdot d^{-4}\big) = O(c_T d^{-2}).$$

Consider now $\mathbf{A}_{4,j,k}^0$ for any $j \in [p], k \in [l_j] \cup \{0\}$. First, we have on $\mathcal{M}$,

$$\begin{aligned}
&\Big\|\mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{W}_j\mathbf{\Pi}_t^*\boldsymbol{\mu}^*\Big)\Big\|_1 \\
&\le \Big\|\mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\Big)\Big\|_{\max}\big\|\mathbf{W}_j\big\|_1\big\|\mathbf{\Pi}_t^*\boldsymbol{\mu}^*\big\|_1 = O\Big(d \cdot \big\|\mathbf{\Pi}_t^*\big\|_\infty\big\|\boldsymbol{\mu}^*\big\|_{\max}\Big) = O(d),
\end{aligned}$$

where the last equality used Assumption (R1). Thus on $\mathcal{M}$,

$$\begin{aligned}
&\big\|\mathbf{A}_{4,j,k}^0\big\|_1 \\
&\le \big\|\mathbf{V}'_{\mathbf{W}'_j,v}\big[\mathbf{I}_d \otimes \mathbb{E}(\mathbf{U}_{\mathbf{x},j,k})\big]\mathbf{V}_{\mathbf{\Pi}_t^*,r}\boldsymbol{\beta}^*\big\|_1 + \Big\|\mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\mathbf{W}_j\mathbf{\Pi}_t^*\boldsymbol{\mu}^*\Big)\Big\|_1 \\
&\le \big\|\mathbf{V}'_{\mathbf{W}'_j,v}\big[\mathbf{I}_d \otimes \mathbb{E}(\mathbf{U}_{\mathbf{x},j,k})\big]\big\|_{\max}\big\|\mathbf{V}_{\mathbf{\Pi}_t^*,r}\big\|_1\big\|\boldsymbol{\beta}^*\big\|_1 + O(d) \\
&\le \max_{i,n\in[d]}\max_{m\in[v]}\max_{q\in[r]} \mathbf{W}'_{j,\cdot i}\mathbb{E}\Big\{\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})_{\cdot m}X_{t,nq}\Big\}\big\|\mathbf{V}_{\mathbf{\Pi}_t^*,r}\big\|_1\big\|\boldsymbol{\beta}^*\big\|_1 + O(d) = O(d),
\end{aligned}$$

where the last equality used Assumption (R1). Furthermore, we have on $\mathcal{M}$ that

$$\begin{aligned}
&\big\|\mathbf{A}_{4,j,k} - \mathbf{A}_{4,j,k}^0\big\|_1 \\
&\le \Big\|\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})' - \mathbb{E}\Big\{\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\Big\}\Big\|_{\max}\big\|\mathbf{W}_j\big\|_1\big\|\mathbf{\Pi}_t^*\boldsymbol{\mu}^*\big\|_1 \\
&\quad + \big\|\mathbf{V}'_{\mathbf{W}'_j,v}\big\{\mathbf{I}_d \otimes [\mathbf{U}_{\mathbf{x},j,k} - \mathbb{E}(\mathbf{U}_{\mathbf{x},j,k})]\big\}\mathbf{V}_{\mathbf{\Pi}_t^*,r}\boldsymbol{\beta}^*\big\|_1 \\
&= O\Big(\big[c_T + 1 \cdot c_T + 1 \cdot c_T + 1 \cdot (c_T \vee 0)\big]\, d\Big) + O\Big(\big\|\mathbf{V}_{\mathbf{\Pi}_t^*,r}\big\|_1\big\|\boldsymbol{\beta}^*\big\|_1\Big) \cdot \big\|\mathbf{W}'_j\big\|_1 \\
&\quad \cdot \max_{n,m\in[d]}\max_{s\in[v]}\max_{q\in[r]}\Big|\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})_{ms}X_{t,nq} - \mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})_{ms}X_{t,nq}\Big)\Big| \\
&= O(c_T d),
\end{aligned} \tag{6.60}$$

where the second last equality used Assumption (R1) and $\mathcal{A}_4$, $\mathcal{A}_{12}$, $\mathcal{A}_{13}$, while the last used $\mathcal{A}_4$, $\mathcal{A}_8$, $\mathcal{A}_9$. In particular, we used the following as an immediate result of $\mathcal{A}_{13}$,

$$\max_{m\in[p]}\max_{n\in[l_m]\cup\{0\}}\Big|\frac{1}{T}\sum_{t=1}^{T} z_{m,n,t} - \mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^{T} z_{m,n,t}\Big)\Big| \le c_T \vee 0.$$

Lastly for any $j \in [p], k \in [l_j] \cup \{0\}$, similar to (6.48), we have on $\mathcal{M}$ that

$$\begin{aligned}\big\|\mathbf{A}_{5,j,k}\big\|_1 &= \big\|\mathbf{V}'_{\mathbf{W}'_j,v}\big(\mathbf{I}_d \otimes \mathbf{U}_{\boldsymbol{\epsilon},j,k}\big)\mathbf{V}_{\mathbf{\Pi}^*_t,r}\big\|_1 \leq \big\|\mathbf{V}'_{\mathbf{W}'_j,v}\big(\mathbf{I}_d \otimes \mathbf{U}_{\boldsymbol{\epsilon},j,k}\big)\big\|_{\max}\big\|\mathbf{V}_{\mathbf{\Pi}^*_t,r}\big\|_1 \\ &\leq \big\|\mathbf{W}'_j\big\|_1 \max_{m,n\in[d]}\max_{s\in[v]}\Big|\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})_{ms}\epsilon_{t,n}\Big| \cdot \big\|\mathbf{V}_{\mathbf{\Pi}^*_t,r}\big\|_1 = O(c_T d).\end{aligned} \tag{6.61}$$

Consider now (6.54), we have

$$\begin{aligned}&\Big\|\mathbf{H}_{20} - T^{-1/2}d^{a/2}\mathbf{\Xi}\mathbf{Y}_W\Big\|_{\max} = \max_{j,k}\Big\|\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3(\mathbf{A}_{4,j,k} + \mathbf{A}_{5,j,k}) - \mathbf{A}^0_1\mathbf{A}^0_2\mathbf{A}^0_3\mathbf{A}^0_{4,j,k}\Big\|_{\max} \\ &\leq \max_{j,k}\big\|\mathbf{A}_1\big\|_{\max}\big\|\mathbf{A}_2\big\|_1\big\|\mathbf{A}_3\big\|_1\big\|\mathbf{A}_{5,j,k}\big\|_1 \\ &\quad + \max_{j,k}\Big\{\big\|\mathbf{A}_1\big\|_{\max}\big\|\mathbf{A}_2\mathbf{A}_3\mathbf{A}_{4,j,k} - \mathbf{A}^0_2\mathbf{A}^0_3\mathbf{A}^0_{4,j,k}\big\|_1 + \big\|\mathbf{A}_1 - \mathbf{A}^0_1\big\|_{\max}\big\|\mathbf{A}^0_2\mathbf{A}^0_3\mathbf{A}^0_{4,j,k}\big\|_1\Big\},\end{aligned}$$

$$\begin{aligned}\text{with} \qquad &\max_{j,k}\big\|\mathbf{A}_2\mathbf{A}_3\mathbf{A}_{4,j,k} - \mathbf{A}^0_2\mathbf{A}^0_3\mathbf{A}^0_{4,j,k}\big\|_1 \\ &\leq \max_{j,k}\Big\{\big\|\mathbf{A}_2\big\|_1\big\|\mathbf{A}_3 - \mathbf{A}^0_3\big\|_1\big\|\mathbf{A}_{4,j,k}\big\|_1 \\ &\qquad + \big\|\mathbf{A}_2\big\|_1\big\|\mathbf{A}^0_3\big\|_1\big\|\mathbf{A}_{4,j,k} - \mathbf{A}^0_{4,j,k}\big\|_1 + \big\|\mathbf{A}_2 - \mathbf{A}^0_2\big\|_1\big\|\mathbf{A}^0_3\big\|_1\big\|\mathbf{A}^0_{4,j,k}\big\|_1\Big\}.\end{aligned}$$

Together with all the rates from (6.56) to (6.61), we have (6.54) true on $\mathcal{M}$.

For (6.55), consider for any $j \in [p], k \in [l_j] \cup \{0\}$, we have on $\mathcal{M}$ that

$$\begin{aligned}&\Big\|\Big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbf{U}_{\boldsymbol{\epsilon},j,k}\mathbf{\Pi}^{*\prime}_t\Big]\mathbf{vec}\big(\mathbf{W}'_j\big)\Big\|_{\max} = \max_{i\in[d]}\big\|(\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbf{U}_{\boldsymbol{\epsilon},j,k}\mathbf{\Pi}^{*\prime}_t\mathbf{W}_{j,i\cdot}\big\|_{\max} \\ &\leq \max_{i\in[d]}\big\|(\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbf{U}_{\boldsymbol{\epsilon},j,k}\big\|_{\max}\big\|\mathbf{\Pi}^*_t\big\|_\infty\big\|\mathbf{W}_{j,i\cdot}\big\|_1 \\ &= O(1) \cdot \|\boldsymbol{\gamma}\|_1 \cdot \max_{m,n\in[d]}\max_{s\in[v]}\Big|\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})_{ms}\epsilon_{t,n}\Big| = O(c_T),\end{aligned} \tag{6.62}$$

where the second last equality used Assumption (R1) and the result below (6.46), and the last is similar to (6.48). In a similar way on $\mathcal{M}$,

$$\begin{aligned}&\Big\|\Big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\big\{\mathbf{U}_{\boldsymbol{\mu},j,k} - \mathbb{E}(\mathbf{U}_{\boldsymbol{\mu},j,k})\big\}\mathbf{\Pi}^{*\prime}_t\Big]\mathbf{vec}\big(\mathbf{W}'_j\big)\Big\|_{\max} \\ &\leq \max_{i\in[d]}\big\|(\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\big\{\mathbf{U}_{\boldsymbol{\mu},j,k} - \mathbb{E}(\mathbf{U}_{\boldsymbol{\mu},j,k})\big\}\big\|_{\max}\big\|\mathbf{\Pi}^*_t\big\|_\infty\big\|\mathbf{W}_{j,i\cdot}\big\|_1 \\ &= O\Big(\|\boldsymbol{\gamma}\|_1\|\boldsymbol{\mu}^*\|_{\max}\Big) \cdot \Big\|\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})' - \mathbb{E}\Big\{\frac{1}{T}\sum_{t=1}^{T} z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})'\Big\}\Big\|_{\max} = O(c_T),\end{aligned} \tag{6.63}$$

with the last line similar to (6.60) which is also involved in the last line of the following,

$$\begin{aligned}
&\Big\|\Big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\big\{\mathbf{U}_{\mathbf{x},j,k} - \mathbb{E}(\mathbf{U}_{\mathbf{x},j,k})\big\}(\boldsymbol{\beta}^* \otimes \mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime}\Big]\mathbf{vec}\big(\mathbf{W}_j'\big)\Big\|_{\max} \\
&\le \max_{i\in[d]} \big\|(\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\big\{\mathbf{U}_{\mathbf{x},j,k} - \mathbb{E}(\mathbf{U}_{\mathbf{x},j,k})\big\}\big\|_{\max}\big\|\boldsymbol{\beta}^* \otimes \mathbf{I}_d\big\|_1\big\|\boldsymbol{\Pi}_t^*\big\|_\infty\big\|\mathbf{W}_{j,i\cdot}\big\|_1 \\
&= O\big(\|\boldsymbol{\gamma}\|_1\big) \cdot \max_{m,n\in[d]} \max_{s\in[v]} \max_{q\in[r]} \Big| \frac{1}{T}\sum_{t=1}^T z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})_{ms} X_{t,nq} \\
&\qquad\qquad - \mathbb{E}\Big(\frac{1}{T}\sum_{t=1}^T z_{j,k,t}(\mathbf{B}_t - \bar{\mathbf{B}})_{ms} X_{t,nq}\Big)\Big| = O(c_T).
\end{aligned} \tag{6.64}$$

Combining (6.62), (6.63) and (6.64), we have (6.55) true by the following,

$$\begin{aligned}
&\Big\|\mathbf{H}_{10} - T^{-1/2}d^{a/2}\mathbf{B}'\mathbf{V}\Big\|_{\max} \\
&\le \max_{j,k} \Big\|\Big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\mathbf{U}_{\boldsymbol{\epsilon},j,k}\boldsymbol{\Pi}_t^{*\prime}\Big]\mathbf{vec}\big(\mathbf{W}_j'\big)\Big\|_{\max} \\
&\quad + \max_{j,k} \Big\|\Big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\big\{\mathbf{U}_{\boldsymbol{\mu},j,k} - \mathbb{E}(\mathbf{U}_{\boldsymbol{\mu},j,k})\big\}\boldsymbol{\Pi}_t^{*\prime}\Big]\mathbf{vec}\big(\mathbf{W}_j'\big)\Big\|_{\max} \\
&\quad + \max_{j,k} \Big\|\Big[\mathbf{I}_d \otimes (\boldsymbol{\gamma}' \otimes \mathbf{I}_d)\big\{\mathbf{U}_{\mathbf{x},j,k} - \mathbb{E}(\mathbf{U}_{\mathbf{x},j,k})\big\}(\boldsymbol{\beta}^* \otimes \mathbf{I}_d)\boldsymbol{\Pi}_t^{*\prime}\Big]\mathbf{vec}\big(\mathbf{W}_j'\big)\Big\|_{\max}.
\end{aligned}$$

Next for $F_2$ and $F_3$, we consider first on $\mathcal{M}$,

$$\begin{aligned}
&\Big\|\Big\{\frac{1}{T}\sum_{t=1}^T \Big(\mathbf{I}_d \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\mathbf{X}_t\Big\}(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu\Big\|_1 \\
&= \Big\|\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3\,\frac{1}{T}\sum_{t=1}^T(\mathbf{B}_t - \bar{\mathbf{B}})'\boldsymbol{\epsilon}_t\Big\|_1 \le \big\|\mathbf{A}_1\big\|_1\big\|\mathbf{A}_2\big\|_1\big\|\mathbf{A}_3\big\|_1 \cdot v\max_{j\in[r]}\Big|\frac{1}{T}\sum_{t=1}^T\sum_{q=1}^d B_{t,qj}\epsilon_{t,q}\Big| \\
&= O\big(d^{1+a}\cdot d^{-2}\cdot d\cdot c_T d^{\frac{1}{2}+\frac{1}{2w}}\big) = O\big(c_T d^{\frac{1}{2}+\frac{1}{2w}+a}\big),
\end{aligned} \tag{6.65}$$

where the first equality used the fact that $\mathbf{X}_t = \mathbf{X}_t \otimes 1$, the second last used (6.57), (6.58), (6.59) and $\mathcal{A}_3$ in Lemma 6.2. Then for $F_2$, we have on $\mathcal{M}$ that

$$\begin{aligned}
\|F_2\|_1 &\le \frac{L^{1/2}\cdot L}{Cd^{1+a}}\Big(\Big\|(\mathbf{H}_{20} - T^{-1/2}d^{a/2}\boldsymbol{\Xi}\mathbf{Y}_W)\Big\|_{\max} + \Big\|(T^{-1/2}d^{a/2}\mathbf{B}'\mathbf{V} - \mathbf{H}_{10})\Big\|_{\max}\Big) \\
&\quad \cdot \Big\|\Big\{\frac{1}{T}\sum_{t=1}^T \Big(\mathbf{I}_d \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\mathbf{X}_t\Big\}(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu\Big\|_1 \\
&= O\big(L^{3/2}\cdot d^{-1-a}\cdot c_T\cdot c_T d^{\frac{1}{2}+\frac{1}{2w}+a}\big) = O\big(c_T^2 L^{3/2} d^{-\frac{1}{2}+\frac{1}{2w}}\big),
\end{aligned} \tag{6.66}$$

where the last line used (6.52), (6.54), (6.55) and (6.65). Similarly, on $\mathcal{M}$ we have $\|F_3\|_1 =$

$O\big(c_T L^{3/2} d^{-\frac{1}{2}+\frac{1}{2w}}\big)$, and hence together with $L = O(1)$, (6.53) and (6.66), it holds on $\mathcal{M}$ that

$$\big\|D_2\big\|_1 \le \big\|F_1\big\|_1 + \big\|F_2\big\|_1 + \big\|F_3\big\|_1 = O\big(c_T d^{-\frac{1}{2}+\frac{1}{2w}}\big). \tag{6.67}$$

Similar to the way we decompose $D_2$, we can rewrite $D_1 = F_4 + F_5 - F_6$ where

$$\begin{aligned}
F_4 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}\Big\{(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10}) \\
&\quad - T^{-1} d^a (\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W)'(\mathbf{B}'\mathbf{V} - \boldsymbol{\Xi}\mathbf{Y}_W)\Big\} D_1, \\
F_5 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}\Big\{(T^{-1/2} d^{a/2}\mathbf{B}'\mathbf{V} - \mathbf{H}_{10}) - (T^{-1/2} d^{a/2}\boldsymbol{\Xi}\mathbf{Y}_W - \mathbf{H}_{20})\Big\}' \\
&\quad \cdot T^{-1/2} d^{a/2} \mathbf{B}'\boldsymbol{\epsilon}, \\
F_6 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}(\mathbf{H}_{20} - \mathbf{H}_{10})' \cdot T^{-1/2} d^{a/2} \mathbf{B}'\boldsymbol{\epsilon}.
\end{aligned}$$

From (6.53), it is direct that $F_4 = O\big(c_T L^{3/2} d^{1-a} \|D_1\|_1\big)$ on $\mathcal{M}$. Moreover, (6.54) and (6.55) imply that $F_5$ has a smaller rate than that of $F_6$. Given $L = O(1)$, we next construct the asymptotic normality of $\boldsymbol{\alpha}' F_6$ for any given nonzero $\boldsymbol{\alpha} \in \mathbb{R}^L$ with $\|\boldsymbol{\alpha}\|_1 \le c < \infty$.

Denote by $\mathbf{R}_1 := [(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}(\mathbf{H}_{20} - \mathbf{H}_{10})'$, we have

$$\begin{aligned}
\boldsymbol{\alpha}' F_6 &= \frac{1}{T}\boldsymbol{\alpha}'\mathbf{R}_1\Big(\mathbf{I}_d \otimes \big\{(\mathbf{B}_1 - \bar{\mathbf{B}}, \ldots, \mathbf{B}_T - \bar{\mathbf{B}})(\mathbf{I}_T \otimes \boldsymbol{\gamma})\big\}\Big)\mathbf{vec}\big((\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T)'\big) \\
&= \frac{1}{T}\boldsymbol{\alpha}'\mathbf{R}_1\Big(\mathbf{I}_d \otimes \big\{(\mathbf{B}_1 - \bar{\mathbf{B}})\boldsymbol{\gamma}, \ldots, (\mathbf{B}_T - \bar{\mathbf{B}})\boldsymbol{\gamma}\big\}\Big)\sum_{t=1}^T \Big\{\boldsymbol{\epsilon}_t \otimes \big(\mathbb{1}_{\{j=t\}}\big)_{j\in[T]}\Big\} \\
&= \frac{1}{T}\sum_{t=1}^T \boldsymbol{\alpha}'\mathbf{R}_1\Big(\boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\Big) = \frac{1}{T}\sum_{t=1}^T \boldsymbol{\alpha}'\mathbf{R}_1\Big(\boldsymbol{\epsilon}_t \otimes [\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]\boldsymbol{\gamma}(1 + o_P(1))\Big),
\end{aligned}$$

where the vector $\big(\mathbb{1}_{\{j=t\}}\big)_{j\in[T]} \in \mathbb{R}^T$ has value $\mathbb{1}_{\{j=t\}}$ at each $j$-th entry, and the last equality used $\mathcal{A}_4$ in Lemma 6.2. Hence to apply Theorem 3 (ii) of Wu (2011), we need to show

$$\sum_{t\ge 0}\Big\|P_0\Big\{\boldsymbol{\alpha}'\mathbf{R}_1\big(\boldsymbol{\epsilon}_t \otimes [\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]\boldsymbol{\gamma}\big)\Big\}\Big\|_2 < \infty, \tag{6.68}$$

where $P_0(\cdot) := \mathbb{E}_0(\cdot) - \mathbb{E}_{-1}(\cdot)$ and $\mathbb{E}_i(\cdot) := \mathbb{E}(\cdot \mid \sigma(\mathcal{G}_i, \mathcal{H}_i))$. Notice that

$$\begin{aligned}
&\Big\|P_0\Big\{\boldsymbol{\alpha}'\mathbf{R}_1\big(\boldsymbol{\epsilon}_t \otimes [\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]\boldsymbol{\gamma}\big)\Big\}\Big\|_2 \\
&= \Big\|\boldsymbol{\alpha}'\mathbf{R}_1\Big\{P_0(\boldsymbol{\epsilon}_t) \otimes \mathbb{E}_0([\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]\boldsymbol{\gamma})\Big\} + \boldsymbol{\alpha}'\mathbf{R}_1\Big\{\mathbb{E}_{-1}(\boldsymbol{\epsilon}_t) \otimes P_0([\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]\boldsymbol{\gamma})\Big\}\Big\|_2 \\
&\le \Big(2\boldsymbol{\alpha}'\mathbf{R}_1\Big\{\mathbb{E}\big(P_0(\boldsymbol{\epsilon}_t)\,P_0(\boldsymbol{\epsilon}_t)'\big) \otimes \mathbb{E}\big(\mathbb{E}_0([\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]\boldsymbol{\gamma})\,\mathbb{E}_0(\boldsymbol{\gamma}'[\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]')\big)\Big\}\mathbf{R}_1'\boldsymbol{\alpha}\Big)^{\frac{1}{2}} \\
&\quad + \Big(2\boldsymbol{\alpha}'\mathbf{R}_1\Big\{\mathbb{E}\big(\mathbb{E}_{-1}(\boldsymbol{\epsilon}_t)\mathbb{E}_{-1}(\boldsymbol{\epsilon}_t)'\big) \otimes \mathbb{E}\big(P_0([\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]\boldsymbol{\gamma})P_0(\boldsymbol{\gamma}'[\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t)]')\big)\Big\}\mathbf{R}_1'\boldsymbol{\alpha}\Big)^{\frac{1}{2}}
\end{aligned}$$

$$= O\Big(\big\|\boldsymbol{\alpha}\big\|_1\big\|\mathbf{R}_1\big\|_\infty\Big) \cdot \Big(\max_{j\in[d]}\big\|P_0(\epsilon_{t,j})\big\|_2 \cdot \max_{j\in[d]}\text{Var}^{1/2}(\mathbf{B}'_{t,j\cdot}\boldsymbol{\gamma}) + \sigma_{\max}\max_{j\in[d]}\big\|P_0(\mathbf{B}'_{t,j\cdot}\boldsymbol{\gamma})\big\|_2\Big)$$
$$= O\Big(\max_{j\in[d]}\big\|P_0^\epsilon(\epsilon_{t,j})\big\|_2 + \max_{j\in[d]}\max_{k\in[v]}\big\|P_0^b(B_{t,jk})\big\|_2\Big),$$

where the second last equality used $\text{Var}(\cdot) = \text{Var}(\mathbb{E}_i(\cdot)) + \mathbb{E}(\text{Var}_i(\cdot)) \geq \text{Var}(\mathbb{E}_i(\cdot))$, and the last used Assumption (R2) and $\big\|\mathbf{R}_1\big\|_\infty = O(1)$ which is implied from (6.52), $\big\|\mathbf{H}_{10}\big\|_1 = O(d)$ and $\big\|\mathbf{H}_{20}\big\|_1 = O(d^{1+a})$. With Assumption (R7), (6.68) is true. Therefore, with definition

$$s_1 := \boldsymbol{\alpha}'\mathbf{R}_1\boldsymbol{\Sigma}\mathbf{R}_1'\boldsymbol{\alpha}, \;\; \boldsymbol{\Sigma} := \sum_\tau \mathbb{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}') \otimes \mathbb{E}\big[(\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))\boldsymbol{\gamma}\boldsymbol{\gamma}'(\mathbf{B}_{t+\tau} - \mathbb{E}(\mathbf{B}_t))'\big],$$

we have by Theorem 3 (ii) of Wu (2011) that

$$T^{1/2}s_1^{-1/2}\boldsymbol{\alpha}'F_6 \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Then equivalently we have

$$T^{1/2}(\mathbf{R}_1\boldsymbol{\Sigma}\mathbf{R}_1')^{-1/2}F_6 \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_L), \tag{6.69}$$

so that $F_6$ is at least $T^{1/2}d^{(1+a-b)/2}$-convergent which used $\lambda_{\max}(\mathbf{R}_1\mathbf{R}_1') = O(d^{-1-a})$ from (6.52), and all eigenvalues of $d^{-b}\boldsymbol{\Sigma}$ uniformly bounded from 0 and infinity by Assumption (R8). Hence, $\big\|D_1\big\|_1 = O\big(\|F_6\|_1\big) = O\big(T^{-1/2}d^{-(1+a-b)/2}\big)$ on $\mathcal{M}$, and by (6.67) we have

$$\big\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}^*\big\|_1 = O_P\Big(\big\|D_1\big\|_1 + \big\|D_2\big\|_1\Big) = O_P\big(T^{-1/2}d^{-(1+a-b)/2} + c_T d^{-\frac{1}{2}+\frac{1}{2w}}\big) = O_P\big(c_T d^{-\frac{1}{2}+\frac{1}{2w}}\big),$$

where the last equality used Assumption (R10). With the above plugged into (6.49), the proof of Theorem 6.1 is complete. □

***Proof of Theorem 6.2.*** By the KKT condition, $\widehat{\boldsymbol{\phi}}$ is a solution to the adaptive LASSO problem in (6.16) if and only if there exists a subgradient

$$\mathbf{h} = \partial(\mathbf{u}'|\widehat{\boldsymbol{\phi}}|) = \left\{\mathbf{h} \in \mathbb{R}^L : \begin{cases} h_i = u_i \,\text{sign}(\widehat{\phi}_i), & \widehat{\phi}_i \neq 0; \\ |h_i| \leq u_i, & \text{otherwise.} \end{cases}\right\},$$

such that differentiating the expression in (6.16) with respect to $\boldsymbol{\phi}$, we have

$$T^{-1}(\boldsymbol{\Xi}\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'(\boldsymbol{\Xi}\mathbf{Y}_W - \mathbf{B}'\mathbf{V})\boldsymbol{\phi} + T^{-1}(\boldsymbol{\Xi}\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'(\mathbf{B}'\mathbf{y} - \boldsymbol{\Xi}\mathbf{y}^\nu) = -\lambda\mathbf{h}.$$

Substituting (6.10) in the above, we arrive at

$$-\lambda\mathbf{h} = T^{-1}(\boldsymbol{\Xi}\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'(\boldsymbol{\Xi}\mathbf{Y}_W - \mathbf{B}'\mathbf{V})\boldsymbol{\phi}$$

$$
\begin{aligned}
&+ T^{-1}(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'(\mathbf{B}'\mathbf{V}\boldsymbol{\phi}^* + \mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}^*}\mathbf{vec}(\mathbf{I}_d) + \mathbf{B}'\boldsymbol{\epsilon} - \Xi\mathbf{y}^\nu) \\
&= T^{-1}(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'\mathbf{B}'\mathbf{V}(\boldsymbol{\phi}^* - \boldsymbol{\phi}) \\
&\quad + T^{-1}(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'\mathbf{B}'\boldsymbol{\epsilon} + T^{-1}(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}^*}\mathbf{vec}(\mathbf{I}_d) \\
&\quad + T^{-1}(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'(\Xi\mathbf{Y}_W\boldsymbol{\phi}^* - \Xi\mathbf{y}^\nu + \Xi\mathbf{Y}_W(\boldsymbol{\phi} - \boldsymbol{\phi}^*)) \\
&= T^{-1}(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})(\boldsymbol{\phi} - \boldsymbol{\phi}^*) + T^{-1}(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'\mathbf{B}'\boldsymbol{\epsilon} \\
&\quad + T^{-1}(\Xi\mathbf{Y}_W - \mathbf{B}'\mathbf{V})'\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}^* - \boldsymbol{\beta}(\boldsymbol{\phi}^*)}\mathbf{vec}(\mathbf{I}_d),
\end{aligned}
$$

where the last equality used the fact that $\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\phi}^*)}\mathbf{vec}(\mathbf{I}_d) = \Xi\mathbf{y}^\nu - \Xi\mathbf{Y}_W\boldsymbol{\phi}^*$ from (6.37). Then we may conclude that there exists a sign-consistent solution $\widehat{\boldsymbol{\phi}}$ if and only if

$$
\begin{cases}
-\lambda\mathbf{h}_H = T^{-1}(\Xi\mathbf{Y}_{W,H} - \mathbf{B}'\mathbf{V}_H)'(\Xi\mathbf{Y}_{W,H} - \mathbf{B}'\mathbf{V}_H)(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*) \\
\qquad\qquad + T^{-1}(\Xi\mathbf{Y}_{W,H} - \mathbf{B}'\mathbf{V}_H)'\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}^* - \boldsymbol{\beta}(\boldsymbol{\phi}^*)}\mathbf{vec}(\mathbf{I}_d) + T^{-1}(\Xi\mathbf{Y}_{W,H} - \mathbf{B}'\mathbf{V}_H)'\mathbf{B}'\boldsymbol{\epsilon}, \\
\lambda\mathbf{u}_{H^c} \geq \left|T^{-1}(\Xi\mathbf{Y}_{W,H^c} - \mathbf{B}'\mathbf{V}_{H^c})'\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}^* - \boldsymbol{\beta}(\boldsymbol{\phi}^*)}\mathbf{vec}(\mathbf{I}_d) + T^{-1}(\Xi\mathbf{Y}_{W,H^c} - \mathbf{B}'\mathbf{V}_{H^c})'\mathbf{B}'\boldsymbol{\epsilon}\right|,
\end{cases}
\tag{6.70}
$$

where $\mathbf{A}_H$ and $\mathbf{a}_H$ denote the corresponding submatrix $\mathbf{A}$ with columns restricted on the set $H$ and subvector $\mathbf{a}$ with entries restricted on the set $H$, respectively. Similarly $(\cdot)_{H^c}$ is defined. Consider the first equation in (6.70), similar to how $D_2$ is decomposed in the proof of Theorem 6.1, we write $\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^* = \sum_{j=1}^4 I_{\phi,j}$ where

$$
\begin{aligned}
I_{\phi,1} &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'_H(\mathbf{H}_{20} - \mathbf{H}_{10})_H]^{-1}\Big\{(\mathbf{H}_{20} - \mathbf{H}_{10})'_H(\mathbf{H}_{20} - \mathbf{H}_{10})_H \\
&\quad - T^{-1}d^a(\mathbf{B}'\mathbf{V}_H - \Xi\mathbf{Y}_{W,H})'(\mathbf{B}'\mathbf{V}_H - \Xi\mathbf{Y}_{W,H})\Big\}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*), \\
I_{\phi,2} &= -[(\mathbf{H}_{20} - \mathbf{H}_{10})'_H(\mathbf{H}_{20} - \mathbf{H}_{10})_H]^{-1}d^a\lambda\mathbf{h}_H, \\
I_{\phi,3} &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'_H(\mathbf{H}_{20} - \mathbf{H}_{10})_H]^{-1}T^{-1}d^a(\Xi\mathbf{Y}_{W,H} - \mathbf{B}'\mathbf{V}_H)'\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\phi}^*) - \boldsymbol{\beta}^*}\mathbf{vec}(\mathbf{I}_d), \\
I_{\phi,4} &= [(\mathbf{H}_{20} - \mathbf{H}_{10})'_H(\mathbf{H}_{20} - \mathbf{H}_{10})_H]^{-1}T^{-1}d^a(\mathbf{B}'\mathbf{V}_H - \Xi\mathbf{Y}_{W,H})'\mathbf{B}'\boldsymbol{\epsilon}.
\end{aligned}
$$

Similar to $F_1$ in the proof of Theorem 6.1, we may derive that

$$
\|I_{\phi,1}\|_{\max} = O_P\Big(c_T d^{1-a}\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*\|_{\max}\Big) = o_P\Big(\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*\|_{\max}\Big),
$$

where the first equality used the fact that Assumption (R1) implies for a positive constant $u$ that $\sigma^2_{|H|}\{(\mathbf{D}_W)_H\} \geq du > 0$ uniformly as $d \to \infty$, and the conditions in the statement of Theorem 6.2, and the second used (R10). Similarly, with Assumption (R9) we have

$$
\|I_{\phi,2}\|_{\max} = O_P\big(d^{-1-a} \cdot d^a \cdot \lambda\big) = O_P\big(c_T d^{-1}\big).
$$

For $I_{\phi,4}$, we may decompose it as the following with the second term dominating the first term similarly to $F_5$ and $F_6$ in the proof of Theorem 6.1,

$$
\begin{aligned}
I_{\phi,4} = &\Big([(\mathbf{H}_{20}-\mathbf{H}_{10})'_H(\mathbf{H}_{20}-\mathbf{H}_{10})_H]^{-1}\Big[(T^{-1/2}d^{a/2}\mathbf{B}'\mathbf{V}_H-\mathbf{H}_{10,H}) \\
&-(T^{-1/2}d^{a/2}\boldsymbol{\Xi}\mathbf{Y}_{W,H}-\mathbf{H}_{20,H})\Big]'\cdot T^{-1/2}d^{a/2}\mathbf{B}'\boldsymbol{\epsilon}\Big) \\
&-\Big([(\mathbf{H}_{20}-\mathbf{H}_{10})'_H(\mathbf{H}_{20}-\mathbf{H}_{10})_H]^{-1}(\mathbf{H}_{20}-\mathbf{H}_{10})'_H\cdot T^{-1/2}d^{a/2}\mathbf{B}'\boldsymbol{\epsilon}\Big).
\end{aligned}
$$

The second term in the above has rate $T^{-1/2}d^{-(1+a-b)/2}$ by exactly the same way to construct asymptotic normality of $F_6$ in (6.69), except for the restriction to the set $H$ here (proof omitted). Thus,

$$
\|I_{\phi,4}\|_{\max} = O_P\big(T^{-1/2}d^{-(1+a-b)/2}\big).
$$

We next construct the asymptotic normality for $I_{\phi,3}$ and show that its convergence rate is of order $T^{-1/2}d^{-(1-b)/2}$ which is dominating over those of $I_{\phi,1}$, $I_{\phi,2}$ and $I_{\phi,4}$ by Assumption (R10). Recall $\mathbf{R}_H=[(\mathbf{H}_{20}-\mathbf{H}_{10})'_H(\mathbf{H}_{20}-\mathbf{H}_{10})_H]^{-1}(\mathbf{H}_{20}-\mathbf{H}_{10})'_H$, and let nonzero $\boldsymbol{\alpha}\in\mathbb{R}^{|H|}$ such that $\|\boldsymbol{\alpha}\|_1\le c<\infty$. Then we have

$$
\begin{aligned}
\boldsymbol{\alpha}'I_{\phi,3} &= \boldsymbol{\alpha}'\mathbf{R}_H T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}}-\bar{\mathbf{B}}_{\boldsymbol{\gamma}})'\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\phi}^*)-\boldsymbol{\beta}^*}\mathbf{vec}\big(\mathbf{I}_d\big)(1+o_P(1)) \\
&= \boldsymbol{\alpha}'\mathbf{R}_H T^{-1}\Big(\mathbf{I}_d\otimes\big\{(\mathbf{B}_1-\bar{\mathbf{B}},\ldots,\mathbf{B}_T-\bar{\mathbf{B}})(\mathbf{I}_T\otimes\boldsymbol{\gamma}(\boldsymbol{\beta}(\boldsymbol{\phi}^*)-\boldsymbol{\beta}^*)')(\mathbf{X}_1,\ldots,\mathbf{X}_T)'\big\}\Big) \\
&\qquad\cdot\mathbf{vec}\big(\mathbf{I}_d\big)(1+o_P(1)) \\
&= \boldsymbol{\alpha}'\mathbf{R}_H\frac{1}{T}\sum_{t=1}^{T}\mathbf{vec}\big((\mathbf{B}_t-\mathbb{E}(\mathbf{B}_t))\,\boldsymbol{\gamma}\,(\boldsymbol{\beta}(\boldsymbol{\phi}^*)-\boldsymbol{\beta}^*)'\mathbf{X}_t'\big)(1+o_P(1)) \\
&= \boldsymbol{\alpha}'\mathbf{R}_H\,\mathbf{vec}\Big\{\frac{1}{T}\sum_{t=1}^{T}\big[\boldsymbol{\gamma}'(\mathbf{B}_{t,i\cdot}-\mathbb{E}(\mathbf{B}_{t,i\cdot}))\mathbf{X}'_{t,j\cdot}(\boldsymbol{\beta}(\boldsymbol{\phi}^*)-\boldsymbol{\beta}^*)\big]_{i,j\in[d]}\Big\}(1+o_P(1)) \\
&= \boldsymbol{\alpha}'\mathbf{R}_H\mathbf{S}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}(\boldsymbol{\phi}^*)-\boldsymbol{\beta}^*)(1+o_P(1)),
\end{aligned}
$$

where the third last equality used $\mathcal{A}_4$ in Lemma 6.2 and the last used $\mathcal{A}_1$. From (6.44) and Lemma 6.4, we have

$$
\begin{aligned}
&T^{\frac{1}{2}}(\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta')^{-\frac{1}{2}}\big(\boldsymbol{\beta}(\boldsymbol{\phi}^*)-\boldsymbol{\beta}^*\big) = T^{\frac{1}{2}}(\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta')^{-\frac{1}{2}}\Big\{\big(\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\mathbf{X}\big)^{-1}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu\Big\} \\
&= T^{\frac{1}{2}}(\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta')^{-\frac{1}{2}}\Big\{\big[\mathbb{E}(\mathbf{X}_t'\mathbf{B}_t)\mathbb{E}(\mathbf{B}_t'\mathbf{X}_t)\big]^{-1}T^{-2}\mathbf{X}'\mathbf{B}^\nu(\mathbf{B}^\nu)'\boldsymbol{\epsilon}^\nu(1+o(1))\Big\}\xrightarrow{\mathcal{D}}\mathcal{N}(\mathbf{0},\mathbf{I}_r).
\end{aligned}
$$

Define $s_3:=\boldsymbol{\alpha}'\mathbf{R}_H\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta'\mathbf{S}_{\boldsymbol{\gamma}}'\mathbf{R}_H'\boldsymbol{\alpha}$, then $T^{1/2}s_3^{-1/2}\boldsymbol{\alpha}'I_{\phi,3}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)$ and equivalently

$$
T^{1/2}(\mathbf{R}_H\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta'\mathbf{S}_{\boldsymbol{\gamma}}'\mathbf{R}_H')^{-1/2}I_{\phi,3}\xrightarrow{\mathcal{D}}\mathcal{N}(\mathbf{0},\mathbf{I}_{|H|}).
$$

As shown in the proof of Lemma 6.4, the eigenvalues of $\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta'$ are of order $d^{b-1}$. Similar to $\mathbf{R}_1$ in the proof of Theorem 6.1, $\lambda_{\max}(\mathbf{R}_H\mathbf{R}_H') = O(d^{-1-a})$. We also have $\lambda_{\max}(\mathbf{S}_\gamma'\mathbf{S}_\gamma) = O(d^{1+a})$ by Assumption (R5). Combining them, we have

$$\begin{aligned}\|\boldsymbol{\alpha}\|_1^2\,\lambda_{\min}(\mathbf{R}_H\mathbf{S}_\gamma\mathbf{S}_\gamma'\mathbf{R}_H')\,\lambda_{\min}(\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta') \le s_3 \\ \le \|\boldsymbol{\alpha}\|_1^2\,\lambda_{\max}(\mathbf{R}_H\mathbf{R}_H')\,\lambda_{\max}(\mathbf{S}_\gamma'\mathbf{S}_\gamma)\,\lambda_{\max}(\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta'),\end{aligned}$$

with the right hand side (of $s_3$) of order $d^{b-1}$. The left hand side (of $s_3$) is of the same order by the assumption in the statement of Theorem 6.2 that $\mathbf{R}_H\mathbf{S}_\gamma\mathbf{S}_\gamma'\mathbf{R}_H'$ has the smallest eigenvalue of constant order. Thus, $s_3$ is of order exactly $d^{b-1}$ and hence $\boldsymbol{\alpha}' I_{\phi,3}$ has order $T^{-1/2}d^{-(1-b)/2}$. It implies $I_{\phi,3}$ is the leading term in $\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*$ whose asymptotic normality therefore holds.

As $I_{\phi,1}$ to $I_{\phi,4}$ are all $o_P(1)$, we conclude $\text{sign}(\widehat{\boldsymbol{\phi}}_H) = \text{sign}(\boldsymbol{\phi}_H^*)$. It remains to show the second part in (6.70) for the zero consistency of $\widehat{\boldsymbol{\phi}}_{H^c}$.

To this end, notice similar to $I_{\phi,3}$ but with restriction on the set $H^c$, we have

$$\begin{aligned}&\big\|T^{-1}(\boldsymbol{\Xi}\mathbf{Y}_{W,H^c} - \mathbf{B}'\mathbf{V}_{H^c})'\mathbf{B}'\mathbf{X}_{\boldsymbol{\beta}^*-\boldsymbol{\beta}(\boldsymbol{\phi}^*)}\mathbf{vec}\big(\mathbf{I}_d\big)\big\|_{\max} \\ &= O_P\big(T^{-1/2}d^{-(1-b)/2}\cdot d^{-a}\cdot d^{1+2a}\big) = O_P\big(T^{-1/2}d^{\frac{1}{2}+\frac{b}{2}+a}\big),\end{aligned}$$

which used $\big\|(\mathbf{H}_{20} - \mathbf{H}_{10})'(\mathbf{H}_{20} - \mathbf{H}_{10})\big\|_{\max} \le \sigma_1^2(\mathbf{H}_{20} - \mathbf{H}_{10}) = O(d^{1+2a})$ similarly to the steps above (6.52). In the same manner, we also have from $I_{\phi,4}$ that

$$\big\|T^{-1}(\boldsymbol{\Xi}\mathbf{Y}_{W,H^c} - \mathbf{B}'\mathbf{V}_{H^c})'\mathbf{B}'\boldsymbol{\epsilon}\big\|_{\max} = O_P\big(T^{-1/2}d^{\frac{1}{2}+\frac{b}{2}+\frac{a}{2}}\big).$$

The left hand side of the second inequality in (6.70) has minimum value of

$$\frac{\lambda}{\big\|\widetilde{\boldsymbol{\phi}}_{H^c}\big\|_{\max}} \ge \frac{\lambda}{\big\|\widetilde{\boldsymbol{\phi}}_{H^c} - \boldsymbol{\phi}_{H^c}^*\big\|_{\max}} \ge \frac{\lambda}{\big\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}^*\big\|_{\max}},$$

so it suffices to show

$$\big(T^{-1/2}d^{\frac{1}{2}+\frac{b}{2}+a}\big)\cdot\big\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}^*\big\|_{\max} = o_P(c_T),$$

which is true by Assumption (R10) and Theorem 6.1 in which each entry of $F_6$ can be shown to be asymptotically normal. This completes the proof of Theorem 6.2. □

***Proof of Theorem 6.3.*** We have

$$\begin{aligned}\big\|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\big\|_\infty &= \Big\|\sum_{j=1}^{p}\Big\{(\widehat{\phi}_{j,0} - \phi_{j,0}^*) + \sum_{k=1}^{l_j}(\widehat{\phi}_{j,k} - \phi_{j,k}^*)z_{j,k,t}\Big\}\mathbf{W}_j\Big\|_\infty \\ &= O_P\Big(\big\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*\big\|_1\cdot\max_j\|\mathbf{W}_j\|_\infty\Big) = O_P\big(T^{-1/2}d^{-(1-b)/2}\big),\end{aligned} \tag{6.71}$$

where the last equality used Theorem 6.2, Assumptions (M2) (or (M2')) and (R1). Observe that we have similarly $\big\|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\big\|_1 = O_P\big(T^{-1/2}d^{-(1-b)/2}\big)$ by Assumption (R1), and hence

$$\big\|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\big\| \leq \Big(\big\|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\big\|_1 \big\|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\big\|_\infty\Big)^{1/2} = O_P\big(T^{-1/2}d^{-(1-b)/2}\big).$$

With $\mathbf{\Pi}_t^*$ defined in (6.46), we can decompose

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}^* &= \frac{1}{T}\sum_{t=1}^T \Big\{(\mathbf{I}_d - \mathbf{\Lambda}_t\widehat{\mathbf{\Phi}})\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}\Big\} - \frac{1}{T}\sum_{t=1}^T \Big\{(\mathbf{I}_d - \mathbf{\Lambda}_t\mathbf{\Phi}^*)\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}^* - \boldsymbol{\epsilon}_t\Big\} \\
&= \frac{1}{T}\sum_{t=1}^T \Big\{\big(\mathbf{W}_t^* - \widehat{\mathbf{W}}_t\big)\big(\mathbf{\Pi}_t^*\boldsymbol{\mu}^* + \mathbf{\Pi}_t^*\mathbf{X}_t\boldsymbol{\beta}^* + \mathbf{\Pi}_t^*\boldsymbol{\epsilon}_t\big)\Big\} + \bar{\mathbf{X}}(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}) + \bar{\boldsymbol{\epsilon}},
\end{aligned}$$

so that combining (6.71), Lemma 6.2 and Theorem 6.1, we have

$$\begin{aligned}
\big\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\big\|_{\max} = O_P\Big\{ & \max_t \big\|\widehat{\mathbf{W}}_t - \mathbf{W}_t^*\big\|_\infty \\
& \cdot \big(\big\|\boldsymbol{\mu}^*\big\|_{\max} + c_T\big\|\boldsymbol{\beta}^*\big\|_{\max} + c_T\big) + c_T\big\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\big\|_1 + c_T\Big\} = O_P(c_T).
\end{aligned}$$

This completes the proof of Theorem 6.3. □

# Bibliography

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Anselin, L. (1988). Model validation in spatial econometrics: A review and evaluation of alternative approaches. *International Regional Science Review*, 11:279–316.

Ayvazyan, S. A. and Ulyanov, V. V. (2023). A multivariate clt for weighted sums with rate of convergence of order o(1/n). In Belomestny, D., Butucea, C., Mammen, E., Moulines, E., Reiß, M., and Ulyanov, V. V., editors, *Foundations of Modern Statistics*, pages 225–257, Cham. Springer International Publishing.

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

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2019). Rank regularized estimation of approximate factor models. *Journal of Econometrics*, 212(1):78–96. Big Data in Dynamic Predictive Econometric Modeling.

Bai, J. and Ng, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association*, 116(536):1746–1763.

Bai, J. and Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2):1893–1916.

Barigozzi, M. and Cho, H. (2020). Consistent estimation of high-dimensional factor models when the factor number is over-estimated. *Electronic Journal of Statistics*, 14(2):2892 – 2921.

Barigozzi, M., Cho, H., and Maeng, H. (2024). Tail-robust factor modelling of vector and tensor time series in high dimensions. arXiv:2407.09390.

Barigozzi, M. and Hallin, M. (2024). *Dynamic Factor Models: A Genealogy*, pages 3–24. Springer Nature Switzerland, Cham.

Barigozzi, M., He, Y., Li, L., and Trapani, L. (2023a). Robust tensor factor analysis. arXiv:2303.18163v2.

Barigozzi, M., He, Y., Li, L., and Trapani, L. (2023b). Statistical inference for large-dimensional tensor factor model by iterative projections. arXiv:2206.09800v2.

Barigozzi, M. and Trapani, L. (2020). Sequential testing for structural stability in approximate factor models. *Stochastic Processes and their Applications*, 130(8):5149–5187.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, 3:77–85.

Beenstock, M. and Felsenstein, D. (2012). Nonparametric estimation of the spatial connectivity matrix using spatial panel data. *Geographical Analysis*, 44(4):386–397.

Bhattacharjee, A. and Jensen-Butler, C. (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics*, 43(4):617–634.

Bradley, R. C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, 2:107 – 144.

Cahan, E., Bai, J., and Ng, S. (2023). Factor-based imputation of missing values and covariances in panel data of large dimensions. *Journal of Econometrics*, 233(1):113–131.

Carroll, J. D. and Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35:283–319.

Cen, Z., Chen, Y., and Lam, C. (2024). Inference on dynamic spatial autoregressive models with change point detection. arXiv:2411.18773.

Cen, Z. and Lam, C. (2025a). On testing kronecker product structure in tensor factor models. arXiv:2501.11208.

Cen, Z. and Lam, C. (2025b). Tensor time series imputation through tensor factor modelling. *Journal of Econometrics*, 249:105974.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

Chang, H.-Y., Song, X., and Yu, J. (2025). Trending time-varying coefficient spatial panel data models. *Journal of Business & Economic Statistics*, 43(1):191–203.

Chang, J., Du, Y., Huang, G., and Yao, Q. (2024). Identification and estimation for matrix time series cp-factor models. arXiv:2410.05634.

Chang, J., He, J., Yang, L., and Yao, Q. (2023). Modelling matrix time series via a tensor CP-decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(1):127–148.

Chapon, A., Ouarda, T. B., and Hamdi, Y. (2023). Imputation of missing values in environmental time series by d-vine copulas. *Weather and Climate Extremes*, 41:100591.

Chen, B., Chen, E. Y., Bolivar, S., and Chen, R. (2024a). Time-varying matrix factor models. arXiv:2404.01546.

Chen, B., Han, Y., and Yu, Q. (2024b). Cp tensor factor models. arXiv:2406.17278.

Chen, E. Y. and Chen, R. (2022). Modeling dynamic transport network with matrix factor models: an application to international trade flow. *Journal of Data Science*, 21(3):490–507.

Chen, E. Y. and Fan, J. (2023). Statistical inference for high-dimensional matrix-variate factor models. *Journal of the American Statistical Association*, 118(542):1038–1055.

Chen, E. Y., Tsay, R. S., and Chen, R. (2020). Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association*, 115(530):775–793.

Chen, E. Y., Xia, D., Cai, C., and Fan, J. (2024c). Semi-parametric tensor factor analysis by iteratively projected singular value decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(3):793–823.

Chen, R., Xiao, H., and Yang, D. (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics*, 222(1, Part B):539–560. Annals Issue: Financial Econometrics in the Age of the Digital Economy.

Chen, R., Yang, D., and Zhang, C.-H. (2022a). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association*, 117(537):94–116.

Chen, W. and Lam, C. (2024a). Factor strength estimation in vector and matrix time series factor models. arXiv:2405.07294v1.

Chen, W. and Lam, C. (2024b). Rank and factor loadings estimation in time series tensor factor model by pre-averaging. *Annals of Statistics*, 52(1):364–391.

Chen, X., Lei, M., Saunier, N., and Sun, L. (2022b). Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):12301–12310.

Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15(3):373–394.

Davidson, J. (2021). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278.

Deng, Y. (2018). Estimation for the spatial autoregressive threshold model. *Economics Letters*, 171:172–175.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics*, 32(2):407 – 499.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fan, J., Guo, J., and Zheng, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *Journal of the American Statistical Association*, 117(538):852–861.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Fan, J. and Yao, Q. (2003). Nonlinear time series: Nonparametric and parametric methods. *New York: Springer-Verlag*.

Favier, G. (2019). *Historical Elements of Matrices and Tensors*, chapter 1, pages 1–7. John Wiley & Sons, Ltd.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–554.

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory*, 17(6):1113–1141.

Freyaldenhoven, S. (2019). A Generalized Factor Model with Local Factors. Working Papers 19-23, Federal Reserve Bank of Philadelphia.

Freyaldenhoven, S. (2022). Factor models with local factors — determining the number of relevant factors. *Journal of Econometrics*, 229(1):80–102.

Gao, Z. and Tsay, R. S. (2023). A two-way transformed factor model for matrix-variate time series. *Econometrics and Statistics*, 27:83–101.

Geweke, J. (1977). *The Dynamic Factor Analysis of Economic Time Series Models*. Workshop series / Social Systems Research Institute, University of Wisconsin. University of Wisconsin.

Guan, L. (2024). Smooth and probabilistic parafac model with auxiliary covariates. *Journal of Computational and Graphical Statistics*, 33(2):538–550.

Guggenberger, P., Kleibergen, F., and Mavroeidis, S. (2023). A test for kronecker product structure covariance matrix. *Journal of Econometrics*, 233(1):88–112.

Hallin, M. and Lippi, M. (2013). Factor models in high-dimensional time series—a time-domain approach. *Stochastic Processes and their Applications*, 123:2678—-2695.

Han, Y., Chen, R., Yang, D., and Zhang, C.-H. (2024a). Tensor factor model estimation by iterative projection. *The Annals of Statistics*, 52(6):2641 – 2667.

Han, Y., Chen, R., Zhang, C.-H., and Yao, Q. (2024b). Simultaneous decorrelation of matrix time series. *Journal of the American Statistical Association*, 119(546):957–969.

Han, Y., Yang, D., Zhang, C.-H., and Chen, R. (2024c). Cp factor model for dynamic tensors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5):1383–1413.

Han, Y., Zhang, C. H., and Chen, R. (2022). Rank determination in tensor factor model. *Electronic Journal of Statistic*, 16:1726–1803.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, 93(2):345–368.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84.

Häusler, E. and Luschgy, H. (2015). *Stable Martingale Central Limit Theorems*, pages 67–122. Springer International Publishing, Cham.

He, Y., Kong, X., Trapani, L., and Yu, L. (2023a). One-way or two-way factor model for matrix sequences? *Journal of Econometrics*, 235(2):1981–2004.

He, Y., Kong, X., Trapani, L., and Yu, L. (2024a). Online change-point detection for matrix-valued time series with latent two-way factor structure. *The Annals of Statistics*, 52(4):1646 – 1670.

He, Y., Kong, X., Yu, L., and Zhang, X. (2022a). Large-dimensional factor analysis without moment constraints. *Journal of Business & Economic Statistics*, 40(1):302–312.

He, Y., Kong, X., Yu, L., Zhang, X., and Zhao, C. (2024b). Matrix factor analysis: From least squares to iterative projection. *Journal of Business & Economic Statistics*, 42(1):322–334.

He, Y., Kong, X.-B., Liu, D., and Zhao, R. (2023b). Robust statistical inference for large-dimensional matrix-valued time series via iterative huber regression. *arXiv:2306.03317*.

He, Y., Wang, Y., Yu, L., Zhou, W., and Zhou, W.-X. (2022b). Matrix kendall's tau in high-dimensions: A robust statistic for matrix factor model. *arXiv:2207.09633*.

Higgins, A. and Martellosio, F. (2023). Shrinkage estimation of network spillovers with factor structured errors. *Journal of Econometrics*, 233(1):66–87.

Hirzel, A. H., Hausser, J., Chessel, D., and Perrin, N. (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, 83(7):2027–2036.

Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189.

Hochreiter, S., Clevert, D.-A., and Obermayer, K. (2006). A new summarization method for affymetrix probe level data. *Bioinformatics*, 22(8):943–949.

Hong, H., Ju, G., Li, Q., and Yan, K. X. (2024). Varying-coefficient spatial dynamic panel data models with fixed effects: Theory and application. *Journal of Econometrics*, 245(1):105883.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2:193–218.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics*, 28(3):397–409.

Kazijevs, M. and Samad, M. D. (2023). Deep imputation of missing values in time series health data: A review with benchmarking. *Journal of Biomedical Informatics*, 144:104440.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17:99–121.

Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, 14(3):105–122.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3):455–500.

Kong, X.-B., Liu, Y.-X., Yu, L., and Zhao, P. (2024). Matrix quantile factor model. arXiv:2208.08693v3.

Lam, C. (2021). Rank determination for time series tensor factor model using correlation thresholding. *Manuscript*.

Lam, C. and Cen, Z. (2024). Matrix-valued factor model with time-varying main effects. arXiv:2406.00128.

Lam, C. and Souza, P. C. (2020). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, 38(3):693–710.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*. Chapman-Hall, Boca Raton (USA).

Lettau, M. (2022). High-dimensional factor models with an application to mutual fund characteristics. NBER Working Papers 29833, National Bureau of Economic Research, Inc.

Li, H., Li, Q., and Shi, Y. (2017a). Determining the number of factors when the number of factors can increase with sample size. *Journal of Econometrics*, 197:76–86.

Li, K. (2018). Spatial Panel Data Models with Structural Change. MPRA Paper 85388, University Library of Munich, Germany.

Li, K. (2022). Threshold Spatial Autoregressive Model. MPRA Paper 113568, University Library of Munich, Germany.

Li, K. and Lin, W. (2024). Threshold spatial autoregressive model. *Journal of Econometrics*, 244(1):105841.

Li, Z., Wang, Q., and Yao, J. (2017b). Identifying the number of factors from singular values of a large sample auto-covariance matrix. *The Annals of Statistics*, 45(1):257–288.

Li, Z. and Xiao, H. (2021). Multi-linear tensor autoregressive models. arXiv:2110.00928.

Liang, X., Gao, J., and Gong, X. (2022). Semiparametric spatial autoregressive panel data model with fixed effects and time-varying coefficients. *Journal of Business & Economic Statistics*, 40(4):1784–1802.

Liu, T., Yuan, M., and Zhao, H. (2022). Characterizing spatiotemporal transcriptome of the human brain via low-rank tensor decomposition. *Statistics in Biosciences*.

Liu, W., Xiao, H., and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statistica Sinica*, 23(3):1257–1272.

Liu, X. and Chen, E. Y. (2022). Identification and estimation of threshold matrix-variate factor models. *Scandinavian Journal of Statistics*, 49(3):1383–1417.

McCrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215.

Meng, X. and Yang, Z. (2023). Threshold spatial panel data models with fixed effects. Technical report, Working Paper. Singapore Singapore Management University.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92(4):1004–1016.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258.

Ostrowski, A. M. (1959). A quantitative formulation of sylvester's law of inertia. *Proceedings of the National Academy of Sciences of the United States of America*, 45(5):740–744.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95(2):365–379.

Pan, X., Li, Z., Qin, S., Yu, M., and Hu, H. (2021). Sclrtc: imputation for single-cell rna-seq data via low-rank tensor completion. *BMC Genomics*, 22(860).

Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, 208(1):23–42. Special Issue on Financial Engineering and Risk Management.

Peña, D. and Box, G. E. P. (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82(399):836–843.

Pinkse, J., Slade, M. E., and Brett, C. (2002). Spatial price competition: A semiparametric appraoch. *Econometrica*, 70(3):1111–1153.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850.

Ross, S. A. (1972). Portfolio and capital market theory with arbitrary preferences and distributions: The general validity of the mean-variance approach in large markets. Rodney l. white center for financial research working papers, Wharton School Rodney L. White Center for Financial Research.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341 – 360.

Sengupta, A. (2017). Testing for a structural break in a spatial panel model. *Econometrics*, 5(1).

Shao, X. (2010). Nonstationarity-extended whittle estimation. *Econometric Theory*, 26(4):1060–1087.

Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Sidorov, D. I. (2010). On mixing conditions for sequences of moving averages. *Theory of Probability & Its Applications*, 54(2):339–347.

Spearman, C. (1904). “General intelligence”, objectively determined and measured. *The American Journal of Psychology*, 15(2):201–293.

Spearman, C. (1927). *The abilities of man*. Macmillan.

Stock, J. and Watson, M. (2005). Implications of dynamic factor models for var analysis. *NBER Working Papers*. No. 11467.

Stock, J. H. and Watson, M. W. (1998). Diffusion indexes. NBER Working Papers 6702, National Bureau of Economic Research, Inc.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Su, L. and Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101.

Sun, Y. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics*, 195(1):134–153.

Sun, Y. and Malikov, E. (2018). Estimation and inference in functional-coefficient spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 203(2):359–378.

Taki, B., Sarwate, A. D., and Bajwa, W. U. (2024). Low separation rank in tensor generalized linear models: An asymptotic analysis. In *2024 58th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6.

Tao, M., Su, J., and Wang, L. (2019). Land cover classification of PolSAR image using tensor representation and learning. *Journal of Applied Remote Sensing*, 13(1):016516.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Tong, H. (1978). On a threshold model in pattern recognition and signal processing. *Pattern recognition and signal processing*, pages 575–586.

Trzcinka, C. (1986). On the number of factors in the arbitrage pricing model. *Journal of Finance*, 41:347–368.

Tsay, R. S. (2024). Matrix-variate time series analysis: A brief review and some new developments. *International Statistical Review*, 92(2):246–262.

Tucker, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In Harris, C. W., editor, *Problems in measuring change.*, pages 122–137. University of Wisconsin Press, Madison WI.

Wang, D., Liu, X., and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, 208(1):231–248. Special Issue on Financial Engineering and Risk Management.

Wang, D., Zheng, Y., and Li, G. (2024). High-dimensional low-rank tensor autoregressive time series modeling. *Journal of Econometrics*, 238(1):105544.

Wang, H. and Leng, C. (2007). Unified lasso estimation by least squares approximation. *Journal of the American Statistical Association*, 102(479):1039–1048.

Wang, H., Li, B., and Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):671–683.

Wang, T. and Samworth, R. J. (2017). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(1):57–83.

Wansbeek, T., Neudecker, H., and Kapteyn, A. (1986). An approach to n-mode components analysis. *Psychometrika*, 51:269–275.

Wu, S. and Bi, P. (2023). Autoregressive moving average model for matrix time series. *Statistical Theory and Related Fields*, 7(4):318–335.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and Its Interface*, 4:207–226.

Xiong, R. and Pelger, M. (2023). Large dimensional latent factor modeling with missing observations and applications to causal inference. *Journal of Econometrics*, 233(1):271–301.

Xu, S., Yuan, C., and Guo, J. (2024). Quasi maximum likelihood estimation for large-dimensional matrix factor models. *Journal of Business & Economic Statistics*, 0(0):1–15.

Yokota, T., Lee, N., and Cichocki, A. (2017). Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria. *IEEE Transactions on Signal Processing*, 65(5):1196–1206.

Yu, C., Li, D., Jiang, F., and Zhu, K. (2024a). Matrix garch model: Inference and application. *Journal of the American Statistical Association*, 0(0):1–31.

Yu, L., He, Y., Kong, X., and Zhang, X. (2022a). Projected estimation for large-dimensional matrix factor models. *Journal of Econometrics*, 229(1):201–217.

Yu, L., Xie, J., and Zhou, W. (2022b). Testing kronecker product covariance matrices for high-dimensional matrix-variate data. *Biometrika*, 110(3):799–814.

Yu, L., Zhao, P., and Zhou, W. (2024b). Testing the number of common factors by bootstrapped sample covariance matrix in high-dimensional factor models. *Journal of the American Statistical Association*, 0(0):1–12.

Yuan, C., Gao, Z., He, X., Huang, W., and Guo, J. (2023). Two-way dynamic factor models for high-dimensional matrix-valued time series. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(5):1517–1537.

Zhang, A. (2019). Cross: Efficient low-rank tensor completion. *The Annals of Statistics*, 47(2):936 – 964.

Zhang, A. R. and Xia, D. (2018). Tensor svd: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64:7311–7338.

Zhang, B., Pan, G., Yao, Q., and Zhou, W. (2024a). Factor modeling for clustering high-dimensional time series. *Journal of the American Statistical Association*, 119(546):1252–1263.

Zhang, H.-F. (2024). Additive autoregressive models for matrix valued time series. *Journal of Time Series Analysis*, 45(3):398–420.

Zhang, X., Li, G., Liu, C. C., and Guo, J. (2024b). Tucker tensor factor models: matricization and mode-wise pca estimation. arXiv:2206.02508v4.

Zhang, X., Liu, C. C., Guo, J., Yuen, K. C., and Welsh, A. H. (2024c). Modeling and learning on high-dimensional matrix-variate sequences. *Journal of the American Statistical Association*, 0(0):1–16.

Zhang, Y. and Shen, D. (2015). Estimation of semi-parametric varying-coefficient spatial panel data models with random-effects. *Journal of Statistical Planning and Inference*, 159:64–80.

Zhang, Y., Zhou, B., Cai, X., Guo, W., Ding, X., and Yuan, X. (2021). Missing value imputation in multivariate time series with end-to-end generative adversarial networks. *Information Sciences*, 551:67–82.

Zhao, J., Rong, C., Lin, C., and Dang, X. (2023). Multivariate time series data imputation using attention-based mechanism. *Neurocomputing*, 542:126238.

Zhao, P. and Yu, B. (2006). On model selection consistency of lasso. *Journal of Machine Learning Research*, 7(90):2541–2563.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.