Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Factor Models for High-Dimensional Tensor Time Series

Rong Chen, Dan Yang & Cun-Hui Zhang

To cite this article: Rong Chen, Dan Yang & Cun-Hui Zhang (2021): Factor Models for High-Dimensional Tensor Time Series, Journal of the American Statistical Association, DOI: 10.1080/01621459.2021.1912757

To link to this article: https://doi.org/10.1080/01621459.2021.1912757

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 00, NO. 0, 1–23
https://doi.org/10.1080/01621459.2021.1912757

Taylor & Francis
Taylor & Francis Group

Factor Models for High-Dimensional Tensor Time Series

Rong Chen[a], Dan Yang[b], and Cun-Hui Zhang[a]

[a]Department of Statistics, Rutgers University, Piscataway, NJ; [b]Faculty of Business and Economics, The University of Hong Kong, Hong Kong

ABSTRACT

Large tensor (multi-dimensional array) data routinely appear nowadays in a wide range of applications, due to modern data collection capabilities. Often such observations are taken over time, forming tensor time series. In this article we present a factor model approach to the analysis of high-dimensional dynamic tensor time series and multi-category dynamic transport networks. This article presents two estimation procedures along with their theoretical properties and simulation results. We present two applications to illustrate the model and its interpretations.

ARTICLE HISTORY
Received May 2020
Accepted Mar 2021

KEYWORDS
Autocovariance matrices; Cross-covariance matrices; Dimension reduction; Dynamic transport network; Eigen-analysis; Factor models; Import–export; Tensor time series; Traffic; Unfolding

1. Introduction

As modern data collection capability has led to massive accumulation of data over time, high-dimensional time series observed in tensor form are becoming more and more commonly seen in various fields such as economics, finance, engineering, environmental sciences, medical research, and others. For example, Figure 1 shows the monthly import–export volume time series of four categories of products (Chemical, Food, Machinery and Electronic, and Footwear and Headwear) among six countries (the United States, Canada, Mexico, Germany, the United Kingdom and France) from January 2001 to December 2016. At each time point, the observations can be arranged into a three-dimensional tensor, with the diagonal elements for each product category unavailable. This is a part of a larger dataset with 15 product categories and 22 countries which we will study in detail in Section 7.1. Univariate time series deals with one item in the tensor (e.g., Food export series of the United States to Canada). Panel time series analysis focuses on the co-movement of one row (fiber) in the tensor (e.g., Food export of the United States to all other countries). Vector time series analysis also focuses on the co-movement of one fiber in the tensor (e.g., Export of the United States to Canada in all product categories). Wang, Liu, and Chen (2019), Chen and Chen (2019), and Chen, Tsay, and Chen (2019) studied matrix time series. Their analysis deals with a matrix slice of the tensor (e.g., the import–export activities between all the countries in one product category). In this article, we develop a factor model for the analysis of the entire tensor time series simultaneously.

The import-export network belongs to the general class of dynamic transport (traffic) network. The focus of such a network is the volume of traffic on the links between the nodes on the network. The availability of complex and diverse network data, recorded over periods of time and in very large scales, brings new opportunities of interesting applications as well as challenging problems in models, methods and theory (Aggarwal and Subbian 2014). For example, weekly activities in different forms (e.g., text messages, email, phone conversations, and personal interactions) and on different topics (politics, food, travel, photo, emotions, etc.) among friends on a social network form a transport network similar to the import–export network, but as a four-dimensional tensor time series. The number of passengers flying between a group of cities with a group of airlines in different seat classes on different days of the week can be represented as a five-dimensional tensor time series. In Section 7.2 we will present a second example on taxi traffic patterns in New York city. With the city being divided into 69 zones, we study the volume of passenger pickups and drop-offs by taxis among the zones, at different hours during the day as a daily time series of a $69 \times 69 \times 24$ tensor.

In the existing literature, most statistical inference methods in network analysis are confined to static network data such as social network (Snijders 2006; Hanneke, Fu, and Xing 2010; Goldenberg et al. 2010; Zhao, Levina, and Zhu 2012; Kolaczyk and Csárdi 2014; Phan and Airoldi 2015; Ji and Jin 2016). However, most networks are dynamic in nature. Thus, an important challenge is to develop stochastic models/processes that capture the dynamic dependence and dynamic changes of a network.

Besides dynamic traffic networks, tensor time series are observed in many other applications. For example, many economic indicators such as GDP, unemployment rate and inflation index are reported quarterly by many countries, forming a matrix-valued time series. Functional MRI produces a sequence of three-dimensional brain images (forming three-dimensional tensors) that changes with different stimulants. Temperature and salinity levels observed at a regular grid of locations and a set

CONTACT Rong Chen rongchen@stat.rutgers.edu Department of Statistics, Rutgers University, Piscataway, NJ 08854
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

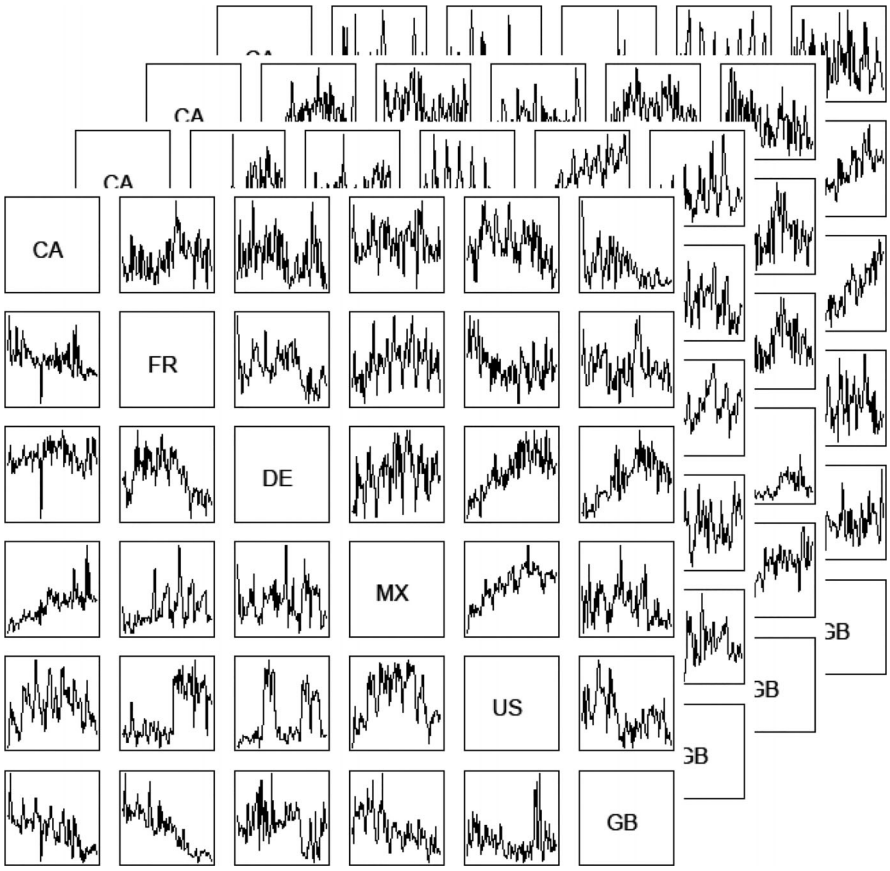

Figure 1. Monthly import-export volume time series of four categories of products (Chemical, Food, Machinery and Electronic, and Footwear and Headwear) among six countries (United States, Canada, Mexico, Germany, UK and France) from January 2001 to December 2016.

of different depth in the ocean form three-dimensional tensors and are observed over time.

Such tensor systems are often very large. Thirty economic indicators from 30 countries yield total 900 individual time series. Import–export volume of 15 product categories among 20 countries makes up almost 6000 individual time series. FMRI images often consist of hundreds of thousands of voxels observed over time.

The aim of this article is to develop a factor model to systematically study the dynamics of tensor systems by jointly modeling the entire tensor simultaneously, while preserving both the tensor structure and the time series structure. This is different from the more conventional time series analysis which deals with scalar or vector observations (Box and Jenkins 1976; Tong 1990; Brockwell and Davis 1991; Härdle, Chen, and Luetkepohl 1997; Shumway and Stoffer 2002; Fan and Yao 2003; Tsay 2005; Tsay and Chen 2018) and multivariate time series analysis (Hannan 1970; Lütkepohl 1993), panel time series analysis (Geweke 1977; Sargent and Sims 1977; Hsiao 2003; Baltagi 2005) and spatial–temporal modelling (Bennett 1979; Cressie 1993; Handcock and Wallis 1994; Wikle, Berliner, and Cressie 1998; Mardia et al. 1998; Stein 1999; Wikle and Cressie 1999; Irwin, Cressie, and Johannesson 2000; Stroud, Muller, and Sanso 2001; Woolrich et al. 2004).

We mainly focus on the cases where the tensor dimension is large. When dealing with many time series simultaneously, dimension reduction is one of the main approaches to extracting common information from the data without being overwhelmed by the idiosyncratic variations. One of the most powerful tools for dimension reduction in time series analysis is the dynamic factor model in which "common" information is summarized into a small number of factors and the co-movement of the time series is assumed to be driven by these factors and their inherited dynamic structures (Chamberlain 1983; Peña and Box 1987; Forni et al. 2000; Bai 2003; Bai and Ng 2008; Pan and Yao 2008; Connor, Hagmann, and Linton 2012; Stock and Watson 2012). We will follow this approach in our development.

It is noted that any tensor can be stacked into a long vector. Hence, it is tempting to simply use the standard factor analysis designed for vector or panel time series. Although such an approach is simple, straightforward and relatively well understood, it has two drawbacks in dealing with tensor time series data. First, stacking a tensor into a long vector loses the structural information, creating difficulties in interpreting results. When rows, columns and depths have special meanings and close relationships within classification, models directly using such structures typically provide better interpretation and meaningful results. Second, dimension reduction under the stacked vector structure is more difficult. Although the tensor factor model we study here can be viewed as a special case of the factor model with the stacked vector time series, the tensor structure used in the tensor factor model induces naturally a special form of the loading matrix in the corresponding vectorized factor model with a much smaller number of parameters. Using the tensor structure allows more effective

dimension reduction, a crucial component in the analysis of high-dimensional time series under consideration here. In simulation studies on matrix factor models, Wang, Liu, and Chen (2019) showed that, if the data-generating process is indeed in a matrix factor model form, it is better to use the structure than stacking the tensor into a long vector and using vector factor models. We expect similar difficulties to directly extend low-dimensional time series models and methods to our setting. For example, tensor time series can be modeled as autoregression with multilinear coefficients structures as in Hoff (2015) and Chen, Xiao, and Yang (2020), but the number of unknown parameters would typically be too large in the high-dimensional case.

The tensor factor model in this article is similar to the matrix factor model studied in Wang, Liu, and Chen (2019). Specifically, we use a Tucker-type decomposition to relate the high-dimensional tensor observations to a low-dimensional latent tensor factor that is assumed to vary over time. Two estimation approaches, named TOPUP and TIPUP, are studied. Asymptotic properties of these estimators are investigated and compared to each other. The estimation procedure used in Wang, Liu, and Chen (2019) in the matrix setting is essentially the TOPUP. We show that the convergence rate they obtained for the TOPUP can be improved. On the other hand, the TIPUP has a faster rate than the TOPUP, under a mildly more restrictive condition on the level of signal omission and cancellation. The theoretical study here also covers the case where the dimensions of the tensor factor increase with the dimensions of the observed tensor time series.

The article is organized as follows. Section 2 contains some preliminary information about the factor models that we will adopt and the basic notations of tensor analysis. Section 3 introduces a general framework of factor models for large tensor time series, which is assumed to be the sum of a signal part and a noise part. The signal part has a multi-linear factor form, consisting of a low-dimensional tensor that varies over time, and a set of fixed loading matrices in a Tucker-type decomposition. Section 4 introduces TOPUP and TIPUP as two general estimation procedures. Section 5 proves their theoretical properties. Section 6 presents some simulation studies to demonstrate the performance of the estimation procedures. Section 7 illustrates the model and its interpretations in two real data applications. The appendix contains proof of the theorems, some discussion, some additional simulation results and additional results of the examples.

2. Preliminary: Dynamic Factor Models and Tensor Operations

In this section, we briefly review the approach of the linear factor model to panel time series data and tensor data analysis. Both serve as foundations of our approach to tensor time series.

Let $\{(x_{i,t})_{d\times T}\}$ be a panel time series. The dynamic factor model assumes

$$\begin{aligned} &\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \text{ or equivelently} \\ &x_{it} = a_{i1}f_{1t} + \ldots + a_{ir}f_{rt} + \varepsilon_{it} \text{ for } i = 1, \ldots, d, \end{aligned} \tag{1}$$

where $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{rt})^\top$ is a a vector-valued unobserved latent factor time series with dimension $r \ll d$; The row vector $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{ir})$, treated as unknown and deterministic, is called factor loading of the ith series. The collection of all $\boldsymbol{a}_i$ forms the loading matrix $\boldsymbol{A}$. The idiosyncratic noise $\boldsymbol{\varepsilon}_t$ is assumed to be uncorrelated with the factors $\boldsymbol{f}_t$ in all leads and lags. Both $\boldsymbol{A}$ and $\boldsymbol{f}_t$ are unobserved hence some further model assumptions are needed. Two different types of model assumptions are adopted in the literature. One type of models assumes that a common factor must have impact on "most" (defined asymptotically) of the time series, but allows the idiosyncratic noise to have weak cross-correlations and weak autocorrelations (Geweke 1977; Sargent and Sims 1977; Forni et al. 2000; Stock and Watson 2012; Bai and Ng 2008; Stock and Watson 2006; Bai and Ng 2002; Hallin and Liška 2007; Chamberlain 1983; Chamberlain and Rothschild 1983; Connor, Hagmann, and Linton 2012; Connor and Linton 2007; Fan, Liao, and Wang 2016; Fan, Wang, and Zhong 2019; Peña and Poncela 2006; Bai and Li 2012). Under such sets of assumptions, principle component analysis (PCA) of the sample covariance matrix is typically used to estimate the space spanned by the columns of the loading matrix, with various extensions. Another type of models assumes that the factors accommodate all dynamics, making the idiosyncratic noise "white" with no autocorrelation but allowing substantial contemporary cross-correlation among the error process (Peña and Box 1987; Pan and Yao 2008; Lam, Yao, and Bathia 2011a; Lam and Yao 2012; Chang, Guo, and Yao 2018). The estimation of the loading space is done by an eigen analysis based on the nonzero lag autocovariance matrices. In this article we adopt the second approach in our model development.

The key feature of the factor model is that all co-movements of the data are driven by the common factor $\boldsymbol{f}_t$ and the factor loading $\boldsymbol{a}_i$ provides a link between the underlying factors and the ith series x_{it}. This approach has three major benefits: (i) It achieves great reduction in model complexity (i.e., the number of parameters) as the autocovariance matrices are now determined by the loading matrix $\boldsymbol{A}$ and the much smaller autocovariance matrix of the factor process $\boldsymbol{f}_t$; (ii) The hidden dynamics (the co-movements) become transparent, leading to clearer and more insightful understanding. This is especially important when the co-movement of the time series is complex and difficult to discover without proper modeling of the full panel; (iii) The estimated factors can be used as input and instrumental variables in models in downstream data analyses, providing summarized and parsimonious information of the whole series.

In the following we briefly review tensor operations mainly for the purpose of fixing the notation in our later discussion. For more detailed information, see Kolda and Bader (2009).

A tensor is a multidimensional array. The order of a tensor is the number of dimensions, also known as the number of modes. Fibers of a tensor are the higher order analogue of matrix rows and columns, which can be obtained by fixing all but one of the modes. For example, a matrix is a tensor of order 2, and a matrix column is a mode-1 fiber and a matrix row is a mode-2 fiber.

Consider an order-K tensor $\mathcal{X} \in \mathbb{R}^{d_1\times\cdots\times d_K}$. Following Kolda and Bader (2009), the mode-k product of $\mathcal{X}$ with a matrix $\boldsymbol{A} \in \mathbb{R}^{\tilde{d}_k\times d_k}$ is an order-K tensor of size $d_1 \times \cdots \times d_{k-1} \times \tilde{d}_k \times d_{k+1} \times \cdots \times d_K$ and will be denoted by $\mathcal{X} \times_k \boldsymbol{A}$. Elementwise, $(\mathcal{X} \times_k \boldsymbol{A})_{i_1\cdots i_{k-1}ji_{k+1}\cdots i_K} = \sum_{i_k=1}^{d_k} x_{i_1\cdots i_k\cdots i_K} a_{ji_k}$. Similarly, the

mode-k product of an order-K tensor with a vector $\boldsymbol{a} \in \mathbb{R}^{d_k}$ is an order-$(K-1)$ tensor of size $d_1 \times \cdots \times d_{k-1} \times d_{k+1} \times \cdots \times d_K$ and denoted by $\mathcal{X} \times_k \boldsymbol{a}$. Elementwise, $(\mathcal{X} \times_k \boldsymbol{a})_{i_1 \cdots i_{k-1} i_{k+1} \cdots i_K} = \sum_{i_k=1}^{d_k} x_{i_1 \cdots i_k \cdots i_K} a_{i_k}$. Let $d = d_1 \ldots d_K$ and $d_{-k} = d/d_k$. The mode-k unfolding matrix $\mathrm{mat}_k(\mathcal{X})$ is a $d_k \times d_{-k}$ matrix by assembling all d_{-k} mode-k fibers as columns of the matrix. One may also stack a tensor into a vector. Specifically, $\mathrm{vec}(\mathcal{X})$ is a vector in $\mathbb{R}^d$ formed by stacking mode-1 fibers of $\mathcal{X}$ in the order of modes $2, \ldots, K$.

The CP decomposition (Carroll and Chang 1970; Harshman 1970) and Tucker decomposition (Tucker 1963, 1964, 1966) are two major extensions of the matrix singular value decomposition (SVD) to tensors of higher order. Recall that the SVD of a matrix $\boldsymbol{X} \in \mathbb{R}^{d_1 \times d_2}$ of rank r has two equivalent forms: $\boldsymbol{X} = \sum_{l=1}^{r} \lambda_l \boldsymbol{u}_l^{(1)} \boldsymbol{u}_l^{(2)\top}$, which decomposes a matrix into a sum of r rank-one matrices, and $\boldsymbol{X} = \boldsymbol{U}_1 \boldsymbol{\Lambda}_r \boldsymbol{U}_2^\top$, where $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are orthonormal matrices of size $d_1 \times r$ and $d_2 \times r$ spanning the column and row spaces of $\boldsymbol{X}$, respectively, and $\boldsymbol{\Lambda}_r$ is an $r \times r$ diagonal matrix with r positive singular values on its diagonal. In parallel, CP decomposes an order-K tensor $\mathcal{X}$ into a sum of rank one tensors, $\mathcal{X} = \sum_{l=1}^{r} \lambda_l \boldsymbol{u}_l^{(1)} \otimes \boldsymbol{u}_l^{(2)} \otimes \cdots \otimes \boldsymbol{u}_l^{(K)} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$, where "$\otimes$" represents the tensor product. The vectors $\boldsymbol{u}_l^{(k)} \in \mathbb{R}^{d_k}, l = 1, 2, \ldots, r$, are not necessarily orthogonal to each other, which differs from the matrix SVD. The Tucker decomposition boils down to K orthonormal matrices $\boldsymbol{U}_k \in \mathbb{R}^{d_k \times r_k}$ containing basis vectors spanning mode-k fibers of the tensor, a potentially much smaller "core" tensor $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_K}$ and the relationship

$$\mathcal{X} = \mathcal{G} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \cdots \times_K \boldsymbol{U}_K = \mathcal{G} \times_{k=1}^{K} \boldsymbol{U}_k. \quad (2)$$

Here the "loading" $\boldsymbol{U}_k$ is the left singular matrix of the mode-k unfolding $\mathrm{mat}_k(\mathcal{X})$ of the tensor, and the core tensor $\mathcal{G}$ is similar to the $\boldsymbol{\Lambda}_r$ in the middle of matrix SVD. While $\mathcal{G}$ is typically not diagonal, its matrix unfolding $\mathrm{mat}_k(\mathcal{G})$ has orthogonal rows for every mode k, that is, the slabs of $\mathcal{G}$ in the same mode are orthogonal to each other for each mode.

3. A Tensor Factor Model

In tensor times series, the observed tensors would depend on $t = 1, \ldots, T$ and be denoted by $\mathcal{X}_t \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ as a series of order-K tensors. By absorbing time, we may stack $\mathcal{X}_t$ into an order-$(K+1)$ tensor $\mathcal{Y} \in \mathbb{R}^{d_1 \times \cdots \times d_K \times T}$, with time t as the $(K+1)$-th mode, referred to as the time-mode. We assume the following decomposition

$$\mathcal{Y} = \mathcal{S} + \mathcal{R}, \text{ or equivalently } \mathcal{X}_t = \mathcal{M}_t + \mathcal{E}_t, \quad (3)$$

where $\mathcal{S}$ is the dynamic signal component and $\mathcal{R}$ is a white-noise part. In the second expression (3), $\mathcal{M}_t$ and $\mathcal{E}_t$ are the corresponding signal and noise components of $\mathcal{X}_t$, respectively. As discussed in Section 2, there are two different approaches to factor analysis, based on the assumption on the dynamic structure of noise $\mathcal{E}_t$. In this article we assume that the noise $\mathcal{E}_t$ are uncorrelated (white) across time, but with arbitrary contemporary covariance structure, following Lam and Yao (2012).

In this model, all dynamics are contained in the signal component $\mathcal{M}_t$. We assume that $\mathcal{M}_t$ is in a lower-dimensional space and has certain multilinear decomposition. We further assume that any component in this multilinear decomposition that involves the time-mode is random and dynamic, and will be called a factor component (depending on its order, it will be called a scalar factor f_t, a vector factor $\boldsymbol{f}_t$, a matrix factor $\boldsymbol{F}_t$, or a tensor factor $\mathcal{F}_t$), which when concatenated along the time-mode forms a higher order object (respectively as $\boldsymbol{g}, \boldsymbol{G}, \mathcal{G}$). Any components of $\mathcal{M}_t$ other than $\mathcal{F}_t$ are assumed to be deterministic and will be called the loading components.

Although it is tempting to directly model $\mathcal{S}$ with standard tensor decomposition approaches to find its lower dimensional structure, the dynamics and dependency in the time direction (auto-dependency) are important and should be treated differently. Traditional tensor decomposition using tensor SVD/PCA on $\mathcal{S}$ ignores the special role of the time-mode and the covariance structure in the time direction, and treats the signal $\mathcal{S}$ as deterministic (Richard and Montanari 2014; Anandkumar, Ge, and Janzamin 2014; Hopkins, Shi, and Steurer 2015; Sun et al. 2016). Such a direct approach often leads to inferior inference results as demonstrated in Wang, Liu, and Chen (2019). In our approach, the component in the time direction is considered as latent and random. As a result, our model assumptions and interpretations, and their corresponding estimation procedures and theoretical properties are significantly different.

Here, we propose a specific model for tensor time series, based on a decomposition similar to Tucker decomposition. Specifically, we assume that

$$\mathcal{S} = \mathcal{G} \times_1 \boldsymbol{A}_1 \times_2 \cdots \times_K \boldsymbol{A}_K \text{ or equivalently}$$
$$\mathcal{M}_t = \mathcal{F}_t \times_1 \boldsymbol{A}_1 \times_2 \cdots \times_K \boldsymbol{A}_K \quad (4)$$

where $\mathcal{F}_t$ is itself a tensor times series of dimension $r_1 \times \cdots \times r_K$ with $r_k \ll d_k$ and $\boldsymbol{A}_k$ are $d_k \times r_k$ loading matrices. We assume without loss of generality in the sequel that $\boldsymbol{A}_k$ is of rank r_k. In this article we consider the case that the order of the tensor K is fixed but the dimensions $d_1, \ldots, d_K \to \infty$ and ranks $r_1, \ldots, r_K$ can be fixed or diverge.

Model (4) resembles a Tucker-type decomposition similar to Equation (2) where the core tensor $\mathcal{G} \in \mathbb{R}^{r_1 \times \cdots \times r_K \times T}$ is the factor term and the loading matrices $\boldsymbol{A}_k \in \mathbb{R}^{d_k \times r_k}$ are constant matrices, whose column spaces are identifiable. The core tensor $\mathcal{F}_t$ is usually much smaller than $\mathcal{X}_t$ in dimension. It drives all the comovements of individual time series in $\mathcal{X}_t$. For matrix time series, model (4) becomes $\boldsymbol{M}_t = \boldsymbol{F}_t \times_1 \boldsymbol{A}_1 \times_2 \boldsymbol{A}_2 = \boldsymbol{A}_1 \boldsymbol{F}_t \boldsymbol{A}_2^\top$. A notable difference between the decompositions (4) and (2) is that the slabs of $\mathcal{F}_t$ are not guaranteed to be orthogonal. For example, in the matrix case, Equation (2) becomes the SVD but $\boldsymbol{F}_t$ is not required to have orthogonal rows or columns in Equation (4). The matrix version of Model (4) was considered in Wang, Liu, and Chen (2019), which also provided several model interpretations. Most of their interpretations can be extended to the tensor factor model. In this article we consider more general model settings and more powerful estimation procedures.

As $\mathcal{F}_t \to \mathcal{X}_t = \mathcal{F}_t \times_{k=1}^{K} \boldsymbol{A}_k + \mathcal{E}_t$ is a linear mapping from $\mathbb{R}^{r_1 \times \cdots \times r_K}$ to $\mathbb{R}^{d_1 \times \cdots \times d_K}$. It can be written as a matrix acting on vectors as in

$$\mathrm{vec}(\mathcal{X}_t) = \big(\odot_{k=1}^{K} \boldsymbol{A}_k \big) \mathrm{vec}(\mathcal{F}_t) + \mathrm{vec}(\mathcal{E}_t), \quad (5)$$

where $\odot_{k=1}^{K} \boldsymbol{A}_k = \boldsymbol{A}_K \odot \cdots \odot \boldsymbol{A}_1$ is the Kronecker product as a $d \times r$ matrix, $d = \prod_{k=1}^{K} d_k$, $r = \prod_{k=1}^{K} r_k$, and $\mathrm{vec}(\cdot)$ is

the tensor stacking operator as described in Section 2. While $\otimes$ is often used to denote the Kronecker product, we shall avoid this usage as $\otimes$ is preserved to denote the tensor product in this article. For example, in the case of $K = 2$ with observation $\boldsymbol{X}_t \in \mathbb{R}^{d_1 \times d_2}$, $\boldsymbol{X}_{t-h} \otimes \boldsymbol{X}_t$ is a $d_1 \times d_2 \times d_1 \times d_2$ tensor of order four, not a matrix of dimension $d_1^2 \times d_2^2$, as we would need to consider the mode-2 unfolding of $\boldsymbol{X}_{t-h} \otimes \boldsymbol{X}_t$ as a $d_2 \times (d_1^2 d_2)$ matrix. The Kronecker expression in Equation (5) exhibits the same form as in the factor model for panel time series except that the loading matrix of size $d \times r$ in the vector factor model is assumed to have a Kronecker product structure of K matrices of much smaller sizes $d_i \times r_i$ $(i = 1, \ldots, K)$. Hence, the tensor factor model reduces the number of parameters in the loading matrices from $dr = d_1 r_1 \ldots d_K r_K$ in the stacked vector version to $d_1 r_1 + \ldots + d_K r_K$, a very significant dimension reduction. The dimension reduction comes from the assumption imposed on the loading matrices.

Remark 1. (Model identification) It is well known that factor models have severe identification problems. To ensure identification, various normalization and constraints can be imposed (Bekker 1986; Neudecker 1990; Bai and Wang 2014, 2015). In order to obtain the most general solutions, in this article we work under the assumption that there is a natural data-generating process that produces the observed data. The generating process consists of a *natural* (unobserved) dynamic factor process $\mathcal{F}_t$ and its corresponding loading matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_K$, whose combination leads to the signal process $\mathcal{M}_t$ and the observed process $\mathcal{X}_t$. For example, the model would accommodate a stationary $\mathcal{F}_t$ of fixed dimension even when the dimension d_k of the observed $\mathcal{X}_t$ and the singular values of $\boldsymbol{A}_k$ all diverge to infinity. As the factor process and the loading matrices cannot be identified even with the noiseless signal process $\mathcal{M}_t$, we do not attempt to impose constraints and normalization in order to estimate them. Instead, we focus on the projection matrix $\boldsymbol{P}_k = \boldsymbol{A}_k(\boldsymbol{A}_k^\top \boldsymbol{A}_k)^{-1}\boldsymbol{A}_k^\top$, which is uniquely defined, as well as other uniquely defined functionals of $\mathcal{M}_t$. The general theory we developed are based on conditions specified on the signal process $\mathcal{M}_t$ and the noise process $\mathcal{E}_t$, even though we often refer to $\mathcal{F}_t$ and $\boldsymbol{A}_k$ in the discussions. In specific examples, we may specify structures of $\mathcal{F}_t$ and $\boldsymbol{A}_k$ and derive specific results based on the general theory.

While the consistent estimation of the orthogonal projection $\boldsymbol{P}_k$ does not require the identifiability of the natural factor series $\mathcal{F}_t$, we will use a canonical form of the factor model (4) to facilitate more explicit explanation of our methodology and theory.

Canonical Factor Series: Let $\boldsymbol{A}_k = \|\boldsymbol{A}_k\|_{\mathrm{S}} \boldsymbol{U}_k \boldsymbol{D}_k \boldsymbol{V}_k^\top$ be the scaled SVD of $\boldsymbol{A}_k$ with diagonal $\boldsymbol{D}_k$ satisfying $\|\boldsymbol{D}_k\|_{\mathrm{S}} = 1$ and $\mathrm{rank}(\boldsymbol{D}_k) = r_k$. The canonical form of the model in (3) and (4) is

$$\mathcal{X}_t = \left(\lambda \mathcal{F}_t^{(\mathrm{cano})}\right) \times_{k=1}^K \boldsymbol{U}_k + \mathcal{E}_t, \tag{6}$$

where $\lambda = \prod_{k=1}^K \|\boldsymbol{A}_k\|_{\mathrm{S}}$ and $\mathcal{F}_t^{(\mathrm{cano})} = \mathcal{F}_t \times_{k=1}^K (\boldsymbol{D}_k \boldsymbol{V}_k^\top) = \lambda^{-1}\left(\mathcal{M}_t \times_{k=1}^K \boldsymbol{U}_k^\top\right) \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is the canonical factor series. In this canonical form, $\boldsymbol{P}_k = \boldsymbol{U}_k \boldsymbol{U}_k^\top$. In the matrix case, we observe $\boldsymbol{X}_t = \boldsymbol{A}_1 \boldsymbol{F}_t \boldsymbol{A}_2^\top + \boldsymbol{E}_t = \lambda \boldsymbol{U}_1 \boldsymbol{F}_t^{(\mathrm{cano})} \boldsymbol{U}_2^\top + \boldsymbol{E}_t$ respectively in the natural and canonical forms.

One-Factor Model: We refer to the simplest case of $r_k = 1\ \forall k \le K$ as the One-Factor Model,

$$\mathcal{X}_t = \lambda f_t(\boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_K) + \mathcal{E}_t, \tag{7}$$

where $f_t \in \mathbb{R}$ is a weakly stationary process and $\boldsymbol{u}_k \in \mathbb{R}^{d_k}$ are deterministic with $\|\boldsymbol{u}_k\|_2 = 1\ \forall k \le K$. As there is no rotation in $\mathbb{R}$, the natural and canonical factor series are the same, $f_t^{(\mathrm{cano})} = f_t$. In the matrix case ($K = 2$), (7) becomes $\boldsymbol{X}_t = \lambda f_t \boldsymbol{u}_1 \boldsymbol{u}_2^\top$.

Remark 2. In our theoretical development, we do not impose any specific structure for the dynamics of the relatively low-dimensional factor process $\mathcal{F}_t \in \mathbb{R}^{r_1 \times \cdots \times r_K}$, except conditions on the spectrum norm and singular values of certain matrices in the unfolding of the average of the cross-product $(T-h)^{-1}(\sum_{t=h+1}^T \mathcal{M}_{t-h} \otimes \mathcal{M}_t)$. As $\mathcal{M}_t = \mathcal{F}_t \times_{k=1}^K \boldsymbol{A}_k$, these conditions on $\mathcal{M}_t$ would hold when we impose certain structures on $\mathcal{F}_t$ and $\boldsymbol{A}_k$. Specifically, as $\mathcal{M}_t = \left(\lambda \mathcal{F}_t^{(\mathrm{cano})}\right) \times_{k=1}^K \boldsymbol{U}_k$ with orthonormal $\boldsymbol{U}_k$, these spectrum norm and singular values can be scaled to the corresponding ones with respect to $(T-h)^{-1}\sum_{t=h+1}^T \mathcal{F}_{t-h}^{(\mathrm{cano})} \otimes \mathcal{F}_t^{(\mathrm{cano})}$ in the canonical form Equation (6), or to those with respect to the natural $(T-h)^{-1}\sum_{t=h+1}^T \mathcal{F}_{t-h} \otimes \mathcal{F}_t$ under proper conditions on $\boldsymbol{D}_k$ or $\boldsymbol{A}_k$. For relatively low-dimensional $\mathcal{F}_t$, the consistency of such spectrum norms and singular values to their population version in the matrix unfolding of $\mathbb{E}\left[(T-h)^{-1}(\sum_{t=h+1}^T \mathcal{M}_{t-h} \otimes \mathcal{M}_t)\right]$ has been extensively studied in the literature with many options such as various mixing conditions, especially for fixed $r_1, \ldots, r_K$.

Remark 3. The above tensor factor model does not assume any structure on the noises except that the noise process is white as explained below (1) and below (3). In particular, it allows structures for the contemporary cross-correlation of the elements of $\mathcal{E}_t$. For example, one may assume that $\mathcal{E}_t$ follows the array Normal distribution (Hoff 2011) in the form $\mathcal{E}_t = \mathcal{Z}_t \times_1 \boldsymbol{\Sigma}_1^{1/2} \times_2 \boldsymbol{\Sigma}_2^{1/2} \times_3 \cdots \times_K \boldsymbol{\Sigma}_K^{1/2}$ where all elements in $\mathcal{Z}_t$ are iid $N(0,1)$. Hence, each of the $\boldsymbol{\Sigma}_i$ can be viewed as the common covariance matrix of mode-i fiber in the tensor $\mathcal{E}_t$. More efficient estimators may be constructed to use such a structure but is out of the scope of this article.

Remark 4. The core factor tensor $\mathcal{F}_t$ may collapse to a lower dimensional tensor. When $r_k = 1$, the factor tensor collapses its mode-k fiber to a scalar. The corresponding loading matrix $\boldsymbol{A}_k$ is then a $d_k \times 1$ vector, and every mode-k fiber of the signal tensor $\mathcal{M}_t$ is proportional to the vector $\boldsymbol{A}_k$. It seems possible to test the hypothesis $r_k = 1$ by studying the proportionality of the fibers but the problem is out of scope of this article.

Remark 5. We note that the array Normal distribution of Hoff (2011) corresponds to model (4) with iid normal entries in $\mathcal{F}_t$ and no observational noise $\mathcal{E}_t$. Loh and Tao-Kai (2000) used similar structure for spatial data. Hafner, Linton, and Tang (2020) and Linton and Tang (2020) considered decomposing the covariance matrix of a vector time series into Kronecker products. Their approach is in some way related to our model, as they essentially arrange the vector time series into a tensor form. However, their objective is quite different from ours.

4. Estimation Procedures

Low-rank tensor approximation is a delicate task. To begin with, the best rank-r approximation to a tensor may not exist (de Silva and Lim 2008) or is NP hard to compute (Hillar and Lim 2013). On the other hand, despite such inherent difficulties, many heuristic techniques are widely used and often enjoy great successes in practice. Richard and Montanari (2014) and Hopkins, Shi, and Steurer (2015), among others, have considered a rank-one spiked tensor model $\mathcal{S} + \mathcal{R}$ as a vehicle to investigate the requirement of signal-to-noise ratio (SNR) for consistent estimation under different constraints of computational resources, where $\mathcal{S} = \lambda \boldsymbol{u}_1 \otimes \boldsymbol{u}_2 \otimes \boldsymbol{u}_3$ for some deterministic unit vectors $\boldsymbol{u}_k \in \mathbb{R}^{d_k}$ and all entries of $\mathcal{R}$ are iid standard normal. As shown by Richard and Montanari (2014), in the symmetric case where $d_1 = d_2 = d_3 = d$, $\mathcal{S}$ can be estimated consistently by the MLE when $\sqrt{d}/\lambda = o(1)$. Similar to the case of spiked PCA (Koltchinskii, Lounici, and Tsybakov 2011; Negahban and Wainwright 2011), it can be shown that the rate achieved by the MLE is minimax among all estimators when $\mathcal{S}$ is treated as deterministic. However, at the same time it is also unsatisfactory as the MLE of $\mathcal{S}$ is NP hard to compute even in this simplest rank one case. Additional discussion of this and some other key differences between matrix and tensor estimations can be found in recent studies of related tensor completion problems (Barak and Moitra 2016; Yuan and Zhang 2016, 2017; Xia and Yuan 2017; Zhang et al. 2019).

A commonly used heuristic to overcome this computational difficulty is tensor unfolding. In the following we proposed two estimation methods that are based on the tensor unfolding of lagged cross-product, the tensor version of the auto-co-moments. Both methods are motivated by the dynamic and random nature of the latent factor process and the whiteness assumption on the error process. As mentioned earlier, due to identification issues, we only attempt to estimate the column space of $\boldsymbol{A}_k$, which can be viewed as the kth principle space of the signal tensor time series $\mathcal{M}_t = \mathcal{F}_t \times_{k=1}^{K} \boldsymbol{A}_k$ in Equation (4). Equivalently, our estimation target is the orthogonal projection

$$\boldsymbol{P}_k = \boldsymbol{A}_k \big(\boldsymbol{A}_k^\top \boldsymbol{A}_k\big)^{-1} \boldsymbol{A}_k^\top. \tag{8}$$

The lagged cross-product operator, which we denote by $\boldsymbol{\Sigma}_h$, can be viewed as the $(2K)$-tensor

$$\begin{aligned}\boldsymbol{\Sigma}_h &= \mathbb{E}\Big[\sum_{t=h+1}^{T} \frac{\mathcal{X}_{t-h} \otimes \mathcal{X}_t}{T-h}\Big] \\ &= \mathbb{E}\Big[\sum_{t=h+1}^{T} \frac{\mathcal{M}_{t-h} \otimes \mathcal{M}_t}{T-h}\Big] \in \mathbb{R}^{d_1 \times \cdots \times d_K \times d_1 \times \cdots \times d_K},\end{aligned}$$

$h = 1, \ldots, h_0$. We consider two estimation methods based on the sample version of $\boldsymbol{\Sigma}_h$,

$$\overline{\boldsymbol{\Sigma}}_h = \sum_{t=h+1}^{T} \frac{\mathcal{X}_{t-h} \otimes \mathcal{X}_t}{T-h}, \quad h = 1, \ldots, h_0. \tag{9}$$

As $\mathcal{M}_t = \mathcal{M}_t \times_{k=1}^{K} \boldsymbol{P}_k$ for all t, we have

$$\boldsymbol{\Sigma}_h = \boldsymbol{\Sigma}_h \times_{k=1}^{2K} \boldsymbol{P}_k = \mathbb{E}\Big[\sum_{t=h+1}^{T} \frac{\mathcal{F}_{t-h} \otimes \mathcal{F}_t}{T-h}\Big] \times_{k=1}^{2K} \boldsymbol{P}_k \boldsymbol{A}_k.$$

with the notation $\boldsymbol{A}_k = \boldsymbol{A}_{k-K}$ and $\boldsymbol{P}_k = \boldsymbol{P}_{k-K}$ for $k > K$. Once consistent estimates $\widehat{\boldsymbol{P}}_k$ are obtained for $\boldsymbol{P}_k$, the estimation of other aspects of $\boldsymbol{\Sigma}_h$ can be carried out based on the low-rank projection of Equation (9),

$$\overline{\boldsymbol{\Sigma}}_h \times_{k=1}^{2K} \widehat{\boldsymbol{P}}_k = \sum_{t=h+1}^{T} \frac{(\mathcal{X}_{t-h} \times_{k=1}^{K} \widehat{\boldsymbol{P}}_k) \otimes (\mathcal{X}_t \times_{k=1}^{K} \widehat{\boldsymbol{P}}_k)}{T-h},$$

as if the low-rank tensor time series $\mathcal{X}_t \times_{k=1}^{K} \widehat{\boldsymbol{P}}_k$ is observed. For the estimation of $\boldsymbol{P}_k$, we propose two methods, and both methods can be written in terms of the mode-k matrix unfolding $\mathrm{mat}_k(\mathcal{X}_t)$ of $\mathcal{X}_t$ as follows.

(i) TOPUP method: Let $d = \prod_{k=1}^{K} d_k$ and $d_{-k} = d/d_k$. Define

$$\boldsymbol{V}_{k,h} = \sum_{t=h+1}^{T} \frac{\mathrm{mat}_k(\mathcal{X}_{t-h}) \otimes \mathrm{mat}_k(\mathcal{X}_t)}{T-h}, \tag{10}$$

which organize all lag-h cross-products of fiber time series in a $d_k \times d_{-k} \times d_k \times d_{-k}$ tensor. Define

$$\mathrm{TOPUP}_k = \big(\mathrm{mat}_1(\boldsymbol{V}_{k,h}), h = 1, \ldots, h_0\big) \tag{11}$$

as a $d_k \times (d_{-k} d h_0)$ matrix. We estimate the left singular matrices of $\mathbb{E}[\mathrm{TOPUP}_k]$ by

$$\widehat{\boldsymbol{P}_{k,m}} = \mathrm{PLSVD}_m\big(\mathrm{TOPUP}_k\big), \tag{12}$$

where PLSVD_m stands for the orthogonal projection to the span of the first m left singular vectors of a matrix. In particular, we estimate the projection $\boldsymbol{P}_k$ by $\widehat{\boldsymbol{P}}_k = \widehat{\boldsymbol{P}_{k,r_k}}$.

The above method is expected to yield consistent estimates of $\boldsymbol{P}_k$ under proper conditions on the dimensionality, signal strength and noise level since Equations (11) and (4) imply

$$\begin{aligned}&\mathbb{E}\big[\mathrm{TOPUP}_k\big] \\ &= \big(\mathrm{mat}_1\big(\textstyle\sum_{t=h+1}^{T} \mathbb{E}\big(\mathrm{mat}_k(\mathcal{M}_{t-h}) \otimes \mathrm{mat}_k(\mathcal{M}_t)\big)/(T-h)\big), h = 1, \ldots, h_0\big) \\ &= \mathrm{mat}_k\big(\big\{\textstyle\sum_{t=h+1}^{T} \mathbb{E}\big(\mathcal{F}_{t-h} \otimes \mathcal{F}_t\big)/(T-h)\big\} \times_{k=1}^{2K} \boldsymbol{A}_k, h = 1, \ldots, h_0\big) \\ &= \boldsymbol{A}_k \mathrm{mat}_k\big(\textstyle\sum_{t=h+1}^{T} \mathbb{E}\big(\mathcal{F}_{t-h} \otimes \mathcal{F}_t\big)/(T-h)\big\} \times_{\ell=1}^{k-1} \boldsymbol{A}_\ell \\ &\qquad \times_{\ell=k+1}^{2K} \boldsymbol{A}_\ell, h = 1, \ldots, h_0\big).\end{aligned} \tag{13}$$

This is a product of two matrices, with $\boldsymbol{A}_k = \boldsymbol{P}_k \boldsymbol{A}_k$ on the left.

We note that the left singular vectors of TOPUP_k are the same as the eigenvectors in the PCA of the $d_k \times d_k$ nonnegative-definite matrix

$$\begin{aligned}\widehat{\boldsymbol{W}}_k &= (\mathrm{TOPUP}_k)(\mathrm{TOPUP}_k)^\top \\ &= \sum_{h=1}^{h_0} \mathrm{mat}_1\big(\boldsymbol{V}_{k,h}\big) \mathrm{mat}_1^\top\big(\boldsymbol{V}_{k,h}\big),\end{aligned} \tag{14}$$

which can be viewed as the sample version of

$$\boldsymbol{W}_k = \mathbb{E}\big[\mathrm{TOPUP}_k\big] \mathbb{E}\big[\mathrm{TOPUP}_k^\top\big]. \tag{15}$$

It follows from Equation (13) that $\boldsymbol{W}_k$ has a sandwich formula with $\boldsymbol{A}_k$ on the left and $\boldsymbol{A}_k^\top$ on the right.

As $\boldsymbol{A}_k$ is assumed to have rank r_k, its column space is identical to that of $\mathbb{E}[\mathrm{TOPUP}_k]$ in Equation (13) or that of $\boldsymbol{W}_k$ in Equation (15) as long as they are also of rank r_k. Thus, $\boldsymbol{P}_k$ is identifiable from the population version of TOPUP_k. However, further identification of the lagged cross-product operator by the TOPUP would involve parameters specific to the TOPUP approach. For example, the TOPUP (12) is designed to estimate

$\boldsymbol{P}_{k,m} = \sum_{j=1}^{m} \boldsymbol{u}_{k,j}\boldsymbol{u}_{k,j}^\top$ where $\boldsymbol{u}_{k,j}$ is the jth left singular vector of the matrix in Equation (13). While $\boldsymbol{P}_{k,r_k} = \boldsymbol{P}_k$ as given in Equation (8), $\boldsymbol{P}_{k,m}$ is specific to the TOPUP in general. Even then, the singular vector $\boldsymbol{u}_{k,m}$ is identifiable only up to the sign through the projections $\boldsymbol{P}_{k,m}$ and $\boldsymbol{P}_{k,m-1}$ provided a sufficiently large gap between the $(m-1)$th, the mth and the $(m+1)$th singular values of the matrix in Equation (13).

For the ease of discussion, we consider for example the case of $k = 1$ and $K = 2$ with stationary factor $\mathcal{F}_t$ where $\mathcal{X}_t \in \mathbb{R}^{d_1 \times d_2}$ is a matrix. The TOPUP estimates the column space of the loading matrix $\boldsymbol{A}_1$ by the span of the first r_k eigenvectors of

$$\widehat{\boldsymbol{W}}_1 = \sum_{h=1}^{h_0} \sum_{j_1=1}^{d_2} \sum_{j_2=1}^{d_2} \boldsymbol{V}_{1,h,j_1j_2} \boldsymbol{V}_{1,h,j_1j_2}^\top \tag{16}$$

as in Equation (14), where $\boldsymbol{V}_{1,h,j_1j_2} = (T-h)^{-1} \sum_{t=h+1}^{T} \boldsymbol{x}_{\cdot,j_1,t-h}\boldsymbol{x}_{\cdot,j_2,t}^\top \in \mathbb{R}^{d_1 \times d_1}$ are the mode-$(1,3)$ matrices of $\boldsymbol{V}_{1,h}$ in Equation (10). By Equation (15), the population version of $\widehat{\boldsymbol{W}}_1$ is

$$\begin{aligned} \boldsymbol{W}_1 &= \sum_{h=1}^{h_0} \sum_{j_1,j_2} \big(\mathbb{E}\big[\boldsymbol{V}_{1,h,j_1j_2}\big]\big)\big(\mathbb{E}\big[\boldsymbol{V}_{1,h,j_1j_2}\big]\big)^\top \\ &= \boldsymbol{A}_1 \Bigg(\sum_{h=1}^{h_0} \sum_{j_1,j_2} \boldsymbol{\Gamma}_{1,h,j_1j_2} \boldsymbol{\Gamma}_{1,h,j_1j_2}^\top \Bigg) \boldsymbol{A}_1^\top, \end{aligned} \tag{17}$$

where $\boldsymbol{\Gamma}_{1,h,j_1j_2} = \mathbb{E}[\mathcal{F}_{t-h} \otimes \mathcal{F}_t] \times_2 \boldsymbol{A}_{2,j_1\cdot} \times_3 \boldsymbol{A}_1 \times_4 \boldsymbol{A}_{2,j_2\cdot} \in \mathbb{R}^{r_1 \times d_1}$. Because $\boldsymbol{W}_1$ is a nonnegative-definite matrix sandwiched by $\boldsymbol{A}_1$ and $\boldsymbol{A}_1^\top$, the column space of $\boldsymbol{A}_1$ and the column space of $\boldsymbol{W}_1$ are the same when the matrix between $\boldsymbol{A}_1$ and $\boldsymbol{A}_1^\top$ in Equation (17) is of full rank.

This procedure uses the time-lagged outer-product of all mode-1 fibers of the observed tensor to extract information about $\boldsymbol{P}_1$ and aggregate the information by applying the PCA to the sum over their Hermitian squares. By considering positive lags $h > 0$, we explicitly use the assumption that the noise process is white, hence avoiding having to deal with the contemporaneous covariance structure of the noise $\mathcal{E}_t$, as $\mathcal{E}_t$ disappears in $\mathbb{E}\big[\boldsymbol{V}_{1,h,j_1j_2}\big] = \boldsymbol{A}_1\boldsymbol{\Gamma}_{1,h,j_1j_2}$ for all $h > 0$. We also note that while the PCA of $\widehat{\boldsymbol{W}}_k$ in Equation 14) is equivalent to the SVD in Equation (12) for the estimation of $\boldsymbol{P}_k$, it can be computationally more efficient to perform the SVD directly in many cases.

We call this TOPUP (Time series Outer-Product Unfolding Procedure) as the tensor product in the matrix unfolding in Equation (10) is a direct extension of the vector outer product. The TOPUP reduces to the algorithm in Wang, Liu, and Chen (2019) for matrix time series.

(ii) TIPUP method: The TIPUP (Time series Inner-Product Unfolding Procedure) can be simply described as the replacement of the tensor product in (10) with the inner product:

$$\boldsymbol{V}_{k,h}^* = \sum_{t=h+1}^{T} \frac{\mathrm{mat}_k(\mathcal{X}_{t-h})\mathrm{mat}_k^\top(\mathcal{X}_t)}{T-h} \in \mathbb{R}^{d_k \times d_k}. \tag{18}$$

Consequently, the TOPUP_k in (11) is replaced by

$$\mathrm{TIPUP}_k = \big(\boldsymbol{V}_{k,h}^*, h = 1, \ldots, h_0\big) \in \mathbb{R}^{d_k \times (d_k h_0)}. \tag{19}$$

The estimator $\widehat{\boldsymbol{P}_{k,m}}$ is then defined as

$$\widehat{\boldsymbol{P}_{k,m}} = \mathrm{PLSVD}_m\big(\mathrm{TIPUP}_k\big). \tag{20}$$

Again TIPUP is expected to yield consistent estimates of $\boldsymbol{P}_k$ in (8) as

$$\begin{aligned} &\mathbb{E}\big[\mathrm{TIPUP}_k\big] \\ &= \big(\big\langle \boldsymbol{\Sigma}_h, \boldsymbol{I}_{k,k+K} \big\rangle_{\{k,k+K\}^c}, h = 1, \ldots, h_0\big) \\ &= \boldsymbol{A}_k\Big(\big\langle \mathbb{E}\big[\textstyle\sum_{t=h+1}^{T} (\mathcal{F}_{t-h} \otimes \mathcal{F}_t)/(T-h)\big] \\ &\qquad \times_{\ell \neq k, 1 \le \ell \le 2K} \boldsymbol{A}_\ell, \boldsymbol{I}_{k,k+K} \big\rangle_{\{k,k+K\}^c}, h \le h_0\Big), \end{aligned} \tag{21}$$

where $\boldsymbol{I}_{k,k+K}$ is the $(2K)$-tensor with elements $(\boldsymbol{I}_{k,k+K})_{\boldsymbol{i},\boldsymbol{j}} = I\{\boldsymbol{i}_{-k} = \boldsymbol{j}_{-k}\}$ at $\boldsymbol{i} = (i_1, \ldots, i_K)$ and $\boldsymbol{j} = \{j_1, \ldots, j_K\}$, and $\langle \cdot, \cdot \rangle_{\{k,k+K\}^c}$ is the inner product summing over indices other than $\{k, k+K\}$.

We use the superscript * to indicate the TIPUP counterpart of TOPUP quantities, for example,

$$\widehat{\boldsymbol{W}}_k^* = \big(\mathrm{TIPUP}_k\big)\big(\mathrm{TIPUP}_k\big)^\top = \sum_{h=1}^{h_0} \boldsymbol{V}_{k,h}^* \boldsymbol{V}_{k,h}^{*\top}, \tag{22}$$

is the sample version of

$$\boldsymbol{W}_k^* = \mathbb{E}\big[\mathrm{TIPUP}_k\big]\mathbb{E}\big[\mathrm{TIPUP}_k^\top\big]. \tag{23}$$

We note that by Equation (21) $\boldsymbol{W}_k^*$ is again sandwiched between $\boldsymbol{A}_k$ and $\boldsymbol{A}_k^\top$. For $k = 1$ and $K = 2$, the TIPUP estimates $\boldsymbol{P}_1$ by applying the PCA to

$$\widehat{\boldsymbol{W}}_1^* = \sum_{h=1}^{h_0} \Bigg(\sum_{j=1}^{d_2} \boldsymbol{V}_{1,h,jj} \Bigg) \Bigg(\sum_{j=1}^{d_2} \boldsymbol{V}_{1,h,jj} \Bigg)^\top \tag{24}$$

as in Equation (22), where $\boldsymbol{V}_{1,h,j_1j_2} \in \mathbb{R}^{d_1 \times d_1}$ is as in Equation (16). The noiseless version of the above $\widehat{\boldsymbol{W}}_1^*$ is

$$\boldsymbol{W}_1^* = \boldsymbol{A}_1 \Bigg[\sum_{h=1}^{h_0} \Bigg(\sum_{j=1}^{d_2} \boldsymbol{\Gamma}_{1,h,jj} \Bigg) \Bigg(\sum_{j=1}^{d_2} \boldsymbol{\Gamma}_{1,h,jj} \Bigg)^\top \Bigg] \boldsymbol{A}_1^\top \tag{25}$$

as in Equation (23), where $\boldsymbol{\Gamma}_{1,h,j_1j_2}$ is as in Equation (17). If the middle term in Equation (25) is of full rank, then the column space of $\boldsymbol{W}_1^*$ is the same as that of $\boldsymbol{A}_1$.

As in the case of the TOPUP, for the estimation of the auto-covariance operator beyond $\boldsymbol{P}_k$, the TIPUP would typically only identify parameters specific to the approach. For example, similar to the TOPUP (12) estimation of the projection $\boldsymbol{P}_{k,m}$ as discussed below Equation (15), the TIPUP (20) aims to estimate $\boldsymbol{P}_{k,m}^* = \sum_{j=1}^{m} \boldsymbol{u}_{k,j}^* \boldsymbol{u}_{k,j}^{*\top}$ with $\boldsymbol{u}_{k,j}^*$ being the jth left singular vectors of (21). However, while $\boldsymbol{P}_{k,r_k}^* = \boldsymbol{P}_k = \boldsymbol{P}_{k,r_k}$ in the full-rank case as both (17) and (25) are sandwiched between $\boldsymbol{A}_k$ and $\boldsymbol{A}_k^\top$, it is evident that $\boldsymbol{P}_{k,m}^* \neq \boldsymbol{P}_{k,m}$ for $m \in [1, r_k)$ in general as the "fillings" in the two sandwiches are not guaranteed to be the same.

Remark 6. The differences between the TOPUP and TIPUP, and the pros and cons: First, taking $K = 2$, $k = 1$ as an example, the TOPUP estimation of $\boldsymbol{P}_1$ uses the auto-cross-product matrices $\boldsymbol{V}_{1,h,j_1j_2} \in \mathbb{R}^{d_1 \times d_1}$ in Equation (16) between *all* the mode-1 fibers in $\mathcal{X}_{t-h}$ and *all* the mode-1 fibers in $\mathcal{X}_t$,

with all possible combinations of j_1 and j_2, while the TIPUP only uses $\boldsymbol{Y}_{1,h,jj}$ with $j = j_1 = j_2$ in Equation (24). Hence, the TIPUP may suffer a loss of efficiency from *signal omission* but may also benefit from this feature when the SNR is low in $\boldsymbol{Y}_{1,h,j_1j_2}$ with $j_1 \neq j_2$. Second, the TOPUP takes the Hermitian squares of all the $\boldsymbol{Y}_{1,h,j_1j_2}$ in Equation (16) before the summation over j_1 and j_2, while the TIPUP takes the summation of $\boldsymbol{Y}_{1,h,jj}$ in Equation (24) before the Hermitian square. As a result, the order of summation and Hermitian square is also switched in the noiseless $\boldsymbol{W}_1$ and $\boldsymbol{W}_1^*$ in Equations (17) and (25), respectively. Hence, the TIPUP may suffer a loss of efficiency, or even the complete loss of identifiability, from *signal cancellation* by taking the summation first. This can be seen more clearly from Equation (25) as the matrix in between $\boldsymbol{A}_1$ and $\boldsymbol{A}_1^\top$ has to be of full rank for consistent estimation of $\boldsymbol{P}_1$ and the sum $\sum_j \boldsymbol{\Gamma}_{1,h,jj}$ is not guaranteed to be of full rank even when all $\boldsymbol{\Gamma}_{1,h,jj}$ are. For the TOPUP, $\boldsymbol{W}_1$ is of full rank when at least one of $\boldsymbol{\Gamma}_{1,h,j_1j_2}$ is of full rank. Meanwhile, compared with the TOPUP, the TIPUP may benefit from *noise omission and noise cancellation* by omitting terms and taking the summation first, which result in faster convergence rate. Here the noise $\boldsymbol{Y}_{1,h,j_1j_2} - \mathbb{E}[\boldsymbol{Y}_{1,h,j_1j_2}]$ comes from both the systematic noise $\mathcal{E}_t$ and the randomness of the factor series. The above discussion is valid for the general $K \geq 2$ as both the TOPUP and TIPUP begin by unfolding the tensors $\mathcal{X}_t$ to matrices $\mathrm{mat}_k(\mathcal{X}_t)$, respectively, in Equations (10) and (18), given the mode k, because both the estimation procedures deal with each mode separately by estimating each loading space individually. For the vector time series $\boldsymbol{x}_t$ with $K = 1$, the TOPUP and TIPUP are identical with $\boldsymbol{Y}_{1,h} = \boldsymbol{Y}_{1,h}^* = \sum_{t=h+1}^{T} \boldsymbol{x}_{t-h}\boldsymbol{x}_t^\top/(T-h)$ in Equations (11) and (19). In Section 5, we will show that in the One-Factor Model (7), $\boldsymbol{W}_k^* = \boldsymbol{W}_k$ so that the TIPUP benefits from noise omission and cancellation without suffering from signal omission or signal cancellation.

The detailed asymptotic convergence rates of both methods presented in Section 5 reflect the differences. In Section 6 we show an example in which the auto-covariance matrices cancel each other for the TIPUP. We note that such complete cancellation does not occur often and can often be avoided by using a larger h_0 in estimation, although partial cancellation can still have a significant impact on the performance of TIPUP in finite samples.

Remark 7. The problem of determining the rank r_k in practice and its associated testing procedure is out of the scope of this article and remains an important and challenging problem to investigate. Many developed procedures for factor model, including the information criteria approach (Bai and Ng 2002, 2007; Hallin and Liška 2007; Amengual and Watson 2007) and ratio of eigenvalues approach (Lam and Yao 2012; Pan and Yao 2008; Lam, Yao, and Bathia 2011b; Ahn and Horenstein 2013), can be extended.

Remark 8. There are other possible estimators using $\boldsymbol{Y}_{k,h}$. For example, in the case of $K = 2, k = 1$, similar to TOPUP in Equation (16) and TIPUP in Equation (24), one may use $\widehat{\boldsymbol{W}}_1 = \sum_{h=1}^{h_0}\sum_{j=1}^{d_2} \boldsymbol{Y}_{1,h,jj}\boldsymbol{Y}_{1,h,jj}^\top$ or $\widehat{\boldsymbol{W}}_1 = \sum_{h=1}^{h_0}\sum_{j_1=1}^{d_2}\big(\sum_{j_2=1}^{d_2}\boldsymbol{Y}_{1,h,j_1j_2}\big)\big(\sum_{j_2=1}^{d_2}\boldsymbol{Y}_{1,h,j_1j_2}\big)^\top$. They may have certain advantages in different cases and can be analyzed with the tools we developed. However, for the sake of space we refrain from further discussion of such variations of our approach.

Remark 9. *i*TOPUP and *i*TIPUP: One can construct iterative procedures based on the TOPUP and TIPUP respectively. Again, consider the case of $K = 2$ and $k = 1$. If a version of $\boldsymbol{U}_2 \in \mathbb{R}^{d_2\times r_2}$ is given with $\boldsymbol{P}_2 = \boldsymbol{U}_2\boldsymbol{U}_2^\top$, $\boldsymbol{P}_1$ can be estimated via the TOPUP or TIPUP using $\tilde{\mathcal{X}}_t^{(1)} = \times_2\boldsymbol{U}_2^\top \in \mathbb{R}^{d_1\times r_2}$. Intuitively the performance improves since $\tilde{\mathcal{X}}_t^{(1)}$ is of much lower dimension than $\mathcal{X}_t$ as $r_2 \ll d_2$. With the results of the TOPUP and TIPUP as the starting points, one can alternate the estimation of $\boldsymbol{P}_k$ given other estimated loading matrices until convergence. They have similar flavor as tensor power iteration methods. Numerical experiments show that the iterative procedures do indeed outperform the simple implementation of the TOPUP and TIPUP. However, their asymptotic properties require more detailed analysis and are out of the scope of this article. The benefit of such iteration has been shown in tensor completion (Xia and Yuan 2017) and tensor de-noising (Zhang and Xia 2018) among other examples.

5. Theoretical Results

Here, we present some theoretical properties of the proposed estimators. Recall that our aim is to estimate the projection $\boldsymbol{P}_k$ in Equation (8) to the column space of $\boldsymbol{A}_k$. We focus on the spectrum norm loss $\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\|_{\mathrm{S}}$, which is connected to the angular error $\theta(\widehat{\boldsymbol{P}}_k, \boldsymbol{P}_k)$ (or the largest canonical angle between the column spaces of $\widehat{\boldsymbol{P}}_k$ and $\boldsymbol{P}_k$) via

$$\begin{aligned}\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\|_{\mathrm{S}} &= \sin\big(\theta(\widehat{\boldsymbol{P}}_k, \boldsymbol{P}_k)\big)\\ &= \max_{1\leq m\leq r_k}\big|\sin\big(\mathrm{angle}(\widehat{\boldsymbol{b}}_{k,m}, \boldsymbol{b}_{k,m})\big)\big|, \qquad (26)\end{aligned}$$

where $\widehat{\boldsymbol{b}}_{k,m}$ and $\boldsymbol{b}_{k,m}$ are, respectively, the mth left- and right-singular vectors of $\widehat{\boldsymbol{P}}_k\boldsymbol{P}_k$.

5.1. Main Condition and Notation

We shall consider the theoretical properties of the estimators under the following main condition:

Condition A: $\mathcal{E}_t$ are independent Gaussian tensors conditionally on the entire process of $\{\mathcal{F}_t\}$. In addition, we assume that for some constant $\sigma > 0$, we have

$$\overline{\mathbb{E}}(\boldsymbol{u}^\top\mathrm{vec}(\mathcal{E}_t))^2 \leq \sigma^2\|\boldsymbol{u}\|_2^2, \quad \boldsymbol{u}\in\mathbb{R}^d, \qquad (27)$$

where $\overline{\mathbb{E}}$ is the conditional expectation given $\{\mathcal{F}_t, 1 \leq t \leq T\}$ and $d = \prod_{k=1}^K d_k$.

We note that the normality assumption, which ensures fast convergence rates in our analysis, is imposed for technical convenience. In fact we only need to impose the sub-Gaussian condition for the results below. For distributions with heavier tails, the convergence rate may be slower.

Condition A, which holds with equality when $\mathcal{E}_t$ has iid $N(0,\sigma^2)$ entries, allows the entries of $\mathcal{E}_t$ to have a range of

dependency structures and different covariance structures for different t. Under Condition A, we develop a general theory to describe the ability of the TOPUP and TIPUP estimators to average out the noise $\mathcal{E}_t$. It guarantees consistency and provides convergence rates in the estimation of the principle spaces of the signal $\mathcal{M}_t$, or equivalently the projections $\boldsymbol{P}_k$, under proper conditions on the magnitude and certain singular value of the lagged cross-product of $\mathcal{M}_t$, allowing the ranks r_k to grow as well as d_k in a sequence of experiments with $T \to \infty$.

We then apply our general theory in two scenarios in the general tensor factor model which we refer to as Fixed-Rank Factor Model and Factor Strength Model, The first one, which is also the simpler, imposes assumptions on the factor series $\mathcal{F}_t$ and the loading matrices directly as follows.

Fixed-Rank Factor Model: Assume Condition A holds in Equations (3) and (4), the ranks $r_k = \text{rank}(\boldsymbol{A}_k)$ are fixed, the canonical factor series $\mathcal{F}_t^{(\text{cano})} = \mathcal{F}_t \times_{k=1}^K (\boldsymbol{D}_k \boldsymbol{V}_k^\top)$ in Equation (6) is weakly stationary, and

$$\mathbb{P}\Big\{ \sum_{t=h+1}^{T} \frac{\mathcal{F}_{t-h}^{(\text{cano})} \otimes \mathcal{F}_t^{(\text{cano})}}{T-h} \to \mathbb{E}\big[\mathcal{F}_{T-h}^{(\text{cano})} \otimes \mathcal{F}_T^{(\text{cano})}\big], \text{ as } T \to \infty \Big\} = 1. \quad (28)$$

Here the diagonal $\boldsymbol{D}_k$ and orthonormal $\boldsymbol{V}_k$ are deterministic $r_k \times r_k$ matrices from the scaled SVD $\boldsymbol{A}_k = \|\boldsymbol{A}_k\|_S \boldsymbol{U}_k \boldsymbol{D}_k \boldsymbol{V}_k^\top$.

Similarly, we define the *natural* Fixed-Rank Factor Model as the one in which the above assumptions hold with $\mathcal{F}_t^{(\text{cano})}$ replaced by $\mathcal{F}_t$. As $\text{rank}(\boldsymbol{A}_k) = r_k$, $\boldsymbol{D}_k$ is positive-definite, so that the two versions of the Fixed-Rank Factor Model are equivalent when the $r_k \times r_k$ matrices $\boldsymbol{D}_k$ and $\boldsymbol{V}_k$ can be viewed as fixed. As a special case of the Fixed-Rank Factor Model, the One-Factor Model (7) has identical canonical and natural factor series. We leave to the existing literature for the verification of (28) as many options are available.

In the Factor Strength Model, to be introduced and discussed in Section 5.4. we will consider cases where the ranks $r_1, \ldots, r_K$ are allowed to diverge with the dimensions $d_1, \ldots, d_K$, and the conditions on the process $\mathcal{M}_t$ are expressed in terms of certain factor strength and related quantities.

We will express general error bounds for the TOPUP and TIPUP in terms of norms and singular values of certain low-dimensional matrices quadratic in the canonical factor series $\mathcal{F}_t^{(\text{cano})}$ in (6). Let $r = \prod_{k=1}^K r_k$ and $r_{-k} = r/r_k$. Parallel to (10) define

$$\boldsymbol{\Phi}_{k,h}^{(\text{cano})} = \text{mat}_k\Big(\sum_{t=h+1}^{T} \frac{\mathcal{F}_{t-h}^{(\text{cano})} \otimes \mathcal{F}_t^{(\text{cano})}}{T-h} \Big) \in \mathbb{R}^{r_k \times (r_{-k} r)}. \quad (29)$$

By (3), (4) and (6), the $\boldsymbol{V}_{k,h}$ in Equation(10) is related to Equation (29) by

$$\overline{\mathbb{E}}[\text{mat}_1(\boldsymbol{V}_{k,h})] = \lambda^2 \boldsymbol{U}_k \boldsymbol{\Phi}_{k,h}^{(\text{cano})} \boldsymbol{U}_{[2k]\setminus\{k\}}^\top$$

where $\boldsymbol{U}_k$ is as in (6), $\boldsymbol{U}_{[2k]\setminus\{k\}} = \odot_{j \in [2K]\setminus\{k\}} \boldsymbol{U}_j \in \mathbb{R}^{(d_{-k}d) \times (r_{-k} r)}$ with $[2K] = \{1, \ldots, 2K\}$ and $\boldsymbol{U}_{k+K} = \boldsymbol{U}_k$, and $\overline{\mathbb{E}}$ is the conditional expectation in Condition A. It follows that by Equation (11)

$$\overline{\mathbb{E}}\big[\text{TOPUP}_k\big] = \lambda^2 \boldsymbol{U}_k \boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})} \big(\text{diag}\big(\boldsymbol{U}_{[2k]\setminus\{k\}}\big)\big)^\top, \quad (30)$$

where $\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})} = (\boldsymbol{\Phi}_{k,1}^{(\text{cano})}, \ldots, \boldsymbol{\Phi}_{k,h_0}^{(\text{cano})}) \in \mathbb{R}^{r_k \times (r_{-k} r h_0)}$ and $\text{diag}(\boldsymbol{U}_{[2k]\setminus\{k\}}) \in \mathbb{R}^{(d_{-k} d h_0) \times (r_{-k} r h_0)}$ with h_0 blocks in the diagonal. As $\boldsymbol{U}_k$ are orthonormal, $\boldsymbol{U}_{[2k]\setminus\{k\}}$ and $\text{diag}(\boldsymbol{U}_{[2k]\setminus\{k\}})$ are orthonormal matrices of the respective dimensions. Thus, in connection to the PCA of Equations (14) and (15),

$$\boldsymbol{W}_k = \lambda^4 \boldsymbol{U}_k \mathbb{E}\Big[\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})}\Big] \mathbb{E}\Big[\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})\top}\Big] \boldsymbol{U}_k^\top.$$

Moreover, as in Equation (28), the analysis of the TOPUP also involves the matrix

$$\boldsymbol{\Phi}_h^{(\text{cano})} = \sum_{t=h+1}^{T} \frac{\text{vec}(\mathcal{F}_{t-h}^{(\text{cano})}) \otimes \text{vec}(\mathcal{F}_t^{(\text{cano})})}{T-h} \in \mathbb{R}^{r \times r}. \quad (31)$$

For the TIPUP, the matrix parallel to Equation (18) is

$$\boldsymbol{\Phi}_{k,h}^{(\text{cano})*} = \sum_{t=h+1}^{T} \frac{\text{mat}_k(\mathcal{F}_{t-h}^{(\text{cano})}) \text{mat}_k^\top(\mathcal{F}_t^{(\text{cano})})}{T-h} \in \mathbb{R}^{r_k \times r_k}. \quad (32)$$

Similar to the derivation of Equation (30), $\overline{\mathbb{E}}\big[\boldsymbol{V}_{k,h}^*\big] = \lambda^2 \boldsymbol{U}_k \boldsymbol{\Phi}_{k,h}^{(\text{cano})*} \boldsymbol{U}_k^\top$ by Equations (3), (4), (6), and (18), so that

$$\overline{\mathbb{E}}\big[\text{TIPUP}_k\big] = \lambda^2 \boldsymbol{U}_k \boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*} \big(\text{diag}\big(\boldsymbol{U}_k\big)\big)^\top \quad (33)$$

by (19), where $\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*} = (\boldsymbol{\Phi}_{k,1}^{(\text{cano})*}, \ldots, \boldsymbol{\Phi}_{k,h_0}^{(\text{cano})*}) \in \mathbb{R}^{r_k \times (r_k h_0)}$ and $\text{diag}\big(\boldsymbol{U}_k\big) \in \mathbb{R}^{(d_k h_0) \times (r_k h_0)}$ has h_0 blocks in the diagonal. Again, as $\text{diag}\big(\boldsymbol{U}_k\big)$ is orthonormal,

$$\boldsymbol{W}_k^* = \lambda^4 \boldsymbol{U}_k \mathbb{E}\Big[\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*}\Big] \mathbb{E}\Big[\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*\top}\Big] \boldsymbol{U}_k^\top. \quad (34)$$

in connection to the PCA of Equations (22) and (23).

5.2. Theoretical Properties of TOPUP:

Based on Condition A and the notation in the above subsection, we present some error bounds for the TOPUP in the following theorem.

Theorem 1. Let λ be as in (6), $d = \prod_{k=1}^K d_k$, $d_{-k} = d/d_k$, $r = \prod_{k=1}^K r_k$, $r_{-k} = r/r_k$ and

$$\Delta_{k,h} = \frac{\sigma(2Td)^{1/2}}{\lambda(T-h)} \Big\{ \Big(d_{-k}^{-1/2} + r_{-k}^{1/2}\Big) \big\| \boldsymbol{\Phi}_{k,0}^{(\text{cano})*} \big\|_S^{1/2} + \Big(d_{-k}^{-1/2} + \sqrt{r/d_k}\Big) \big\| \boldsymbol{\Phi}_0^{(\text{cano})} \big\|_S^{1/2} \Big\} + \frac{\sigma^2 d_{-k}^{1/2}}{\lambda^2} \Big\{ \frac{(d_{-k}^{-1}+1)\sqrt{2d}}{\sqrt{T-h}} + \frac{2d_k}{T-h} \Big\} \quad (35)$$

with the matrices $\boldsymbol{\Phi}_0^{(\text{cano})}$ and $\boldsymbol{\Phi}_{k,0}^{(\text{cano})*}$ in Equations (31) and (32), respectively. Suppose Condition A holds with the conditional expectation $\overline{\mathbb{E}}$. Then, $\overline{\mathbb{E}}[\text{mat}_1(\boldsymbol{V}_{k,h})] = \lambda^2 \boldsymbol{U}_k \boldsymbol{\Phi}_{k,h}^{(\text{cano})} \boldsymbol{U}_{[2k]\setminus\{k\}}^\top$ and

$$\overline{\mathbb{E}} \big\| \lambda^{-2} \text{mat}_1\big(\boldsymbol{V}_{k,h}\big) - \boldsymbol{U}_k \boldsymbol{\Phi}_{k,h}^{(\text{cano})} \boldsymbol{U}_{[2k]\setminus\{k\}}^\top \big\|_S \le \Delta_{k,h}, \quad (36)$$

$$\overline{\mathbb{E}} \big\| \lambda^{-2} \text{TOPUP}_k - \boldsymbol{U}_k \boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})} \big(\text{diag}\big(\boldsymbol{U}_{[2k]\setminus\{k\}}\big)\big)^\top \big\|_S \le \sqrt{h_0} \Delta_{k,h_0},$$

for all k and $h_0 \le T/4$, with the matrices $\boldsymbol{U}_k$ in Equation (6), $\boldsymbol{\Phi}_{k,h}^{(\text{cano})}$ in Equation (29) and $\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})}$ and $\text{diag}\big(\boldsymbol{U}_{[2k]\setminus\{k\}}\big)$ in

(30). Moreover, for the estimator $\widehat{\boldsymbol{P}}_k = \widehat{\boldsymbol{P}_{k,r_k}}$ with $m = r_k$ in Equation (12),

$$\overline{\mathbb{E}}\big[\big\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\big\|_{\mathrm{S}}\big] \le 2\sqrt{h_0}\Delta_{k,h_0}\big/\sigma_{r_k}\big(\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0}\big) \tag{37}$$

for the spectrum loss in Equation (26), where $\sigma_m(\cdot)$ is the m-th largest singular value.

Remark 10. As the removal of the effect of the systematic noise $\mathcal{E}_t$ and dimension reduction are the most important tasks in the analysis of SVD-based estimators of the projections $\boldsymbol{P}_k$, the main theorems in this article are stated in the general form in terms of the conditional expectation $\overline{\mathbb{E}}$ with error bounds involving only norms and singular values of auto-cross-product matrices of the canonical factor series. As discussed in Remark 2 of Section 3, we intentionally leave open many options to study such error bounds to maintain generality and simplicity. Still, deterministic error bounds are provided in the corollaries below for more specific models. In the noiseless case of $\sigma = 0$, $\Delta_{k,h} = 0$ in Equations (36), and (37) has the interpretation

$$\mathbb{P}\big\{\widehat{\boldsymbol{P}}_k = \boldsymbol{P}_k\big\} = \mathbb{P}\big\{\mathrm{rank}\big(\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0}\big) = r_k\big\},$$

so that the recovery of $\boldsymbol{P}_k$ is feasible as long as the factor series is almost surely in general position, without requiring the consistency of the random matrix $\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0} \in \mathbb{R}^{r_k\times(r^2h_0/r_k)}$.

Theorem 1 asserts that the observable TOPUP_k can be viewed as a noisy version of a low-rank matrix sandwiched between $\boldsymbol{U}_k$ and another orthonormal matrix. The spectrum bound (36) for this noise leads to the error bound for the estimation of $\boldsymbol{P}_k = \boldsymbol{U}_k\boldsymbol{U}_k^\top$ in Equation (37) and additional results in Corollaries 1 and 3 below. The proof of the theorem is shown in Appendix A. The following corollary gives explicit error bounds for the Fixed-Rank Factor Model.

Corollary 1. Let h_0 be fixed. Suppose $\mathbb{E}\big[\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0}\big]$ is of rank r_k for the matrix $\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0} \in \mathbb{R}^{r_k\times(r^2h_0/r_k)}$ in (30). Then, in the Fixed-Rank Factor Model, Equations (36) and (37) hold with

$$\begin{aligned}\Delta_{k,h} \asymp \sqrt{d/d_k}\big\{&(\sigma/\lambda)\sqrt{d_k/T}\\ &+(\sigma/\lambda)^2\big(\sqrt{d/T} + d_k/T\big)\big\},\end{aligned} \tag{38}$$

where $\lambda = \prod_{k=1}^K \|\boldsymbol{A}_k\|_{\mathrm{S}}$. Consequently, the TOPUP (12) satisfies

$$\big\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\big\|_{\mathrm{S}} \lesssim \sqrt{d/d_k}\big\{(\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T}\big\}. \tag{39}$$

If in addition $\mathbb{E}\big[\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0}\big]$ has distinct nonzero singular values, then Equation (12) satisfies

$$\begin{aligned}&\big\|\widehat{\boldsymbol{P}_{k,m}} - \boldsymbol{P}_{k,m}\big\|_{\mathrm{S}}\\ &\quad\lesssim \sqrt{d/d_k}\big\{(\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T}\big\}\end{aligned} \tag{40}$$

where $\boldsymbol{P}_{k,m} = \sum_{j=1}^m \boldsymbol{u}_{k,j}\boldsymbol{u}_{k,j}^\top$ with $\boldsymbol{u}_{k,j}$ being the jth left singular vector of $\mathbb{E}\big[\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0}\big]$.

In the One-Factor Model (7) with $\hat\rho_h = \sum_{t=h+1}^T f_{t-h}f_t/(T-h)$, we have

$$\begin{aligned}\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,h} &= \boldsymbol{\Phi}^{(\mathrm{cano})*}_{k,h} = \hat\rho_h,\\ \sigma_1\big(\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0}\big) &= \big\|\boldsymbol{\Phi}^{(\mathrm{cano})}_{k,1:h_0}\big\|_2 = \big(\textstyle\sum_{h=1}^{h_0}\hat\rho_h^2\big)^{1/2},\end{aligned} \tag{41}$$

in view of Equations (31), (32) and (30). Thus, when $\sum_{h=1}^{h_0}\hat\rho_h^2/h_0 \asymp 1 \asymp \hat\rho_0$, (39) of Corollary 1 becomes

$$\begin{aligned}\big|\sin\big(\mathrm{angle}(\widehat{\boldsymbol{u}}_k, \boldsymbol{u}_k)\big)\big| &= \big\|\widehat{\boldsymbol{u}}_k\widehat{\boldsymbol{u}}_k^\top - \boldsymbol{u}_k\boldsymbol{u}_k^\top\big\|_{\mathrm{S}}\\ &\lesssim \sqrt{d/d_k}\big\{(\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T}\big\}\end{aligned} \tag{42}$$

for the spectrum loss in Equation (26) and the TOPUP estimates $\widehat{\boldsymbol{u}}_k$ of $\boldsymbol{u}_k$.

In the case of $K = 2$ where the matrix time series $\boldsymbol{X}_t = \boldsymbol{A}_1\boldsymbol{F}_t\boldsymbol{A}_2^\top + \boldsymbol{E}_t$ is observed, properties of the TOPUP were studied in Wang, Liu, and Chen (2019) under the conditions of Corollary 1 assuming $(\lambda/\sigma)^2 \asymp d_1^{1-\delta_1'}d_2^{1-\delta_2'} = d^{1-\delta_0'}$ for some $\delta_0' \in [0,1]$. Their error bounds yield the convergence rate

$$\big\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\big\|_{\mathrm{S}} \lesssim d^{\delta_0'}/T^{1/2} \asymp d^{\delta_0'/2}(\sigma/\lambda)\sqrt{d/T} \asymp (\sigma/\lambda)^2 d/T^{1/2}$$

The result of Wang, Liu, and Chen (2019) can be viewed as an extension of the results of Lam, Yao, and Bathia (2011a) from the vector factor model (1) with $K = 1$ to the matrix factor model with $K = 2$. In comparison, (39) is sharper and also applies to the tensor factor model with $K \ge 3$.

5.3. Theoretical Properties of TIPUP

We summarize our analysis of the TIPUP procedure in the following theorem.

Theorem 2. Let λ be as in Equation (6), $\boldsymbol{\Phi}^{(\mathrm{cano})*}_{k,h}$ as in Equation (32), $\boldsymbol{\Phi}^{(\mathrm{cano})*}_{k,1:h_0}$ as in Equation (33), $d = \prod_{k=1}^K d_k$ and

$$\begin{aligned}\Delta^*_{k,h} = &\frac{2\sigma(8Td_k)^{1/2}}{\lambda(T-h)}\big\|\boldsymbol{\Phi}^{(\mathrm{cano})*}_{k,0}\big\|_{\mathrm{S}}^{1/2}\\ &+\frac{\sigma^2}{\lambda^2}\bigg(\frac{\sqrt{8d}}{\sqrt{T-h}} + \frac{2d_k}{T-h}\bigg).\end{aligned} \tag{43}$$

Suppose Condition A holds. Then, $\overline{\mathbb{E}}\big[\boldsymbol{V}^*_{k,h}\big] = \lambda^2\boldsymbol{U}_k\boldsymbol{\Phi}^{(\mathrm{cano})*}_{k,h}\boldsymbol{U}_k^\top$ and

$$\begin{aligned}\overline{\mathbb{E}}\big\|\boldsymbol{V}^*_{k,h}/\lambda^2 - \boldsymbol{U}_k\boldsymbol{\Phi}^{(\mathrm{cano})*}_{k,h}\boldsymbol{U}_k^\top\big\|_{\mathrm{S}} &\le \Delta^*_{k,h},\\ \overline{\mathbb{E}}\big\|\mathrm{TIPUP}_k/\lambda^2 - \boldsymbol{U}_k\boldsymbol{\Phi}^{(\mathrm{cano})*}_{k,1:h_0}\boldsymbol{U}_k^\top\big\|_{\mathrm{S}} &\le h_0^{1/2}\Delta^*_{k,h_0},\end{aligned} \tag{44}$$

for all k and $h_0 \le T/4$. Moreover, for the estimator $\widehat{\boldsymbol{P}}_k = \widehat{\boldsymbol{P}_{k,r_k}}$ with $m = r_k$ in (20),

$$\overline{\mathbb{E}}\big\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\big\|_{\mathrm{S}} \le 2\sqrt{h_0}\Delta^*_{k,h_0}\big/\sigma_{r_k}\big(\boldsymbol{\Phi}^{(\mathrm{cano})*}_{k,1:h_0}\big). \tag{45}$$

Parallel to Theorem 1, Theorem 2 asserts that the observable TIPUP_k can be viewed as a noisy low-rank matrix sandwiched between $\boldsymbol{U}_k$ and $\boldsymbol{U}_k^\top$. The spectrum bound (44) for this noise leads to the error bound for the estimation of $\boldsymbol{P}_k = \boldsymbol{U}_k\boldsymbol{U}_k^\top$ in (45) and additional results in Corollaries 2 and 3. The proof of the theorem is shown in Appendix A. Again we consider the Fixed-Rank Factor Model in a corollary.

Corollary 2. Let h_0 be fixed. Suppose $\mathbb{E}[\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0}]$ is of rank r_k for the matrix $\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0} \in \mathbb{R}^{r_k\times(r_kh_0)}$ in Equation (33). Then, in the Fixed-Rank Factor Model, Equations (44) and (45) hold with

$$\Delta^*_{k,h} \asymp (\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\big(\sqrt{d/T} + d_k/T\big), \quad (46)$$

where $\lambda = \prod_{k=1}^K \|\boldsymbol{A}_k\|_{\mathrm{S}}$. Consequently, the TIPUP (20) satisfies

$$\big\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\big\|_{\mathrm{S}} \lesssim (\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T}. \quad (47)$$

If in addition $\mathbb{E}[\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0}]$ has distinct nonzero singular values, then Equation (20) satisfies

$$\big\|\widehat{\boldsymbol{P}_{k,m}} - \boldsymbol{P}^*_{k,m}\big\|_{\mathrm{S}} \lesssim (\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T}, \quad 1 \le m \le r_k, \quad (48)$$

where $\boldsymbol{P}^*_{k,m} = \sum_{j=1}^m \boldsymbol{u}^*_{k,j}\boldsymbol{u}^{*\top}_{k,j}$ with $\boldsymbol{u}^*_{k,j}$ being the j-th left singular vector of $\mathbb{E}[\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0}]$.

For the One-Factor Model (7), (41) applies to the TIPUP and Equation (47) of Corollary 2 becomes

$$\big|\sin\big(\text{angle}(\widehat{\boldsymbol{u}}_k, \boldsymbol{u}_k)\big)\big| = \big\|\widehat{\boldsymbol{u}}_k\widehat{\boldsymbol{u}}_k^\top - \boldsymbol{u}_k\boldsymbol{u}_k^\top\big\|_{\mathrm{S}} \lesssim (\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T} \quad (49)$$

for the TIPUP estimate $\widehat{\boldsymbol{u}}_k$ of $\boldsymbol{u}_k$ under the condition $\sum_{h=1}^{h_0}\hat{\rho}_h^2/h_0 \asymp 1 \asymp \hat{\rho}_0$.

We note that the convergence rate in Equations (47), (48) and (49) is faster by a factor of $\sqrt{d/d_k} = \prod_{j\neq k} d_j^{1/2}$ than the corresponding rate for the TOPUP in Equations (39), (40) and (42) under the respective conditions. However, it is important to recognize that the full rank condition on the population matrix $\mathbb{E}[\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0}]$ in Corollary 2, which guarantees that potential signal omission and cancellation do not change the rate, is significantly stronger than the corresponding full rank condition on the matrix $\mathbb{E}[\mathbf{\Phi}^{(\text{cano})}_{k,1:h_0}]$ in Corollary 1. As we explained in Equations (17), (25) and Remark 6 below Equation (25), the matrix $\mathbb{E}[\mathbf{\Phi}^{(\text{cano})}_{k,1:h_0}]$ is of full rank if and only if the following sum

$$\sum_{h=1}^{h_0}\sum_{j_1=1}^{r/r_k}\sum_{j_2=1}^{r/r_k}\Big(\mathbf{\Gamma}^{(\text{cano})}_{k,h,j_1j_2}\Big)\Big(\mathbf{\Gamma}^{(\text{cano})}_{k,h,j_1j_2}\Big)^\top \in \mathbb{R}^{r_k\times r_k} \quad (50)$$

is positive-definite, where $\mathbf{\Gamma}^{(\text{cano})}_{k,h,j_1j_2} = \mathbb{E}\big[\big(\text{mat}_k(\mathcal{F}^{(\text{cano})}_{t-h})\boldsymbol{e}_{j_1}\big)\big(\text{mat}_k(\mathcal{F}^{(\text{cano})}_{t-h})\boldsymbol{e}_{j_2}\big)^\top\big]$ with the canonic unit vectors $\boldsymbol{e}_j$ in $\mathbb{R}^{r/r_k}$. Meanwhile, $\mathbb{E}[\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0}]$ is of full rank if and only if

$$\sum_{h=1}^{h_0}\Big(\sum_{j=1}^{r/r_k}\mathbf{\Gamma}^{(\text{cano})}_{k,h,jj}\Big)\Big(\sum_{j=1}^{r/r_k}\mathbf{\Gamma}^{(\text{cano})}_{k,h,jj}\Big)^\top \in \mathbb{R}^{r_k\times r_k} \quad (51)$$

is positive-definite. Compared with Equation (50), (51) omits the signals in $\mathbf{\Gamma}^{(\text{cano})}_{k,h,j_1j_2}$ with $j_1 \neq j_2$ and may cancel the signals in $\mathbf{\Gamma}^{(\text{cano})}_{k,h,jj}$. While the signal omission and cancellation rarely cause the rank deficiency of the matrix in Equation (51), the resulting loss of efficiency may still have an impact on the finite sample performance as our simulation results demonstrate. In any cases, there is no signal omission or cancellation in TIPUP in the One-Factor Model as shown in Equations (49) and (41), and in general we may alleviate the signal cancellation in TIPUP by taking a larger h_0 in Equation (51).

5.4. Models With Diverging Ranks

In Theorems 1 and 2, the risk bounds for the TOPUP and TIPUP are expressed in the scale $\lambda = \prod_{k=1}^K \|\boldsymbol{A}_k\|_{\mathrm{S}}$ in (6) and in terms of spectrum norms of the matrices $\mathbf{\Phi}^{(\text{cano})}_0 \in \mathbb{R}^{r\times r}$ in Equation (31) and $\mathbf{\Phi}^{(\text{cano})*}_{k,0} \in \mathbb{R}^{r_k\times r_k}$ in Equation (32) and the singular values of the matrices $\mathbf{\Phi}^{(\text{cano})}_{k,1:h_0} \in \mathbb{R}^{r_k\times(r_{-k}rh_0)}$ in Equation (30) and $\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0} \in \mathbb{R}^{r_k\times(r_kh_0)}$ in Equation (33). When the ranks r_k and the maximum lag h_0 are fixed, it would be quite reasonable to treat these norms and singular values as constants in the Fixed-Rank Factor Model as in Corollaries 1 and 2. However, the situation is different when the ranks r_k are allowed to diverge. First, as the loading of the factor series $\mathcal{F}_t$ may not be uniform in different directions, we may desire to allow the condition numbers $\|\boldsymbol{D}_k^{-1}\|_{\mathrm{S}}^2$ of $\boldsymbol{A}_k^\top\boldsymbol{A}_k \in \mathbb{R}^{r_k\times r_k}$ to diverge. Second, we may need to take into account the impact of the ranks and heterogeneity of the singular values of these matrices on the estimation error. To this end, we consider below the Factor Strength Model in which the scaling constant λ and the combined impacts of the above discussed and other features of the canonical factor series are summarized in terms of certain signal strength parameters, following the tradition of Lam, Yao, and Bathia (2011a) among others.

For the vector factor model (1) with bounded r_1 and $\|\boldsymbol{D}_1^{-1}\|_{\mathrm{S}}$ and $\lambda = \|\boldsymbol{A}_1\|_{\mathrm{S}} \asymp \sigma d^{(1-\delta')/2}$ for a $\delta' \in [0,1]$, Lam, Yao, and Bathia (2011a) proved the convergence rate of $(\sigma/\lambda)^2 d/T^{1/2}$ for their estimator of $\boldsymbol{P}_1$, which can be viewed as the TOPUP for $K = 1$. The quantity δ', which reflects the signal-to-noise ratio in the factor model, can be referred to as the factor strength index (Bai and Ng 2002; Doz, Giannone, and Reichlin 2011; Lam, Yao, and Bathia 2011a; Lam and Yao 2012; Wang, Liu, and Chen 2019). When $\delta' = 0$, the information contained in the signal $\boldsymbol{A}\boldsymbol{f}_t$ grows linearly with the dimension d. In this case the factor is said to be "strong" and the convergence rate is $T^{-1/2}$ in Lam, Yao, and Bathia (2011a) for $K = 1$ and Wang, Liu, and Chen (2019) for $K = 2$. When $0 < \delta' \le 1$, the information in the signal increases more slowly than the dimension and the factor is said to be "weak." Compared with Wang, Liu, and Chen (2019), Corollary 1 provides the same rate for strong factors and faster rate for weak factors. It is still true that one would need longer time series to compensate the weakness in signal to achieve consistent estimation of the loading spaces. Below we provide a concise description of the relationship between the signal strength and convergence rates of TOPUP and TIPUP in the Factor Strength Model.

Factor Strength Model: Let the canonical factor series in Equation (6) be scaled such that

$$\mathbb{E}\big[\text{trace}\big(\mathbf{\Phi}^{(\text{cano})}_0\big)\big] = r, \quad \lambda^2 = \sigma^2 d^{1-\delta_0}/r, \quad (52)$$

for the $\mathbf{\Phi}^{(\text{cano})}_0$ in (31) and some $\delta_0 \in \mathbb{R}$. Suppose Condition A holds, that with probability $1 + o(1)$

$$\text{trace}\big(\mathbf{\Phi}^{(\text{cano})}_0\big) \asymp r, \quad (53)$$

$$\sigma_{r_k}\big(\mathbf{\Phi}^{(\text{cano})}_{k,1:h_0}\big) \ge c_1 d^{\delta_0-\delta_1} h_0^{1/2} r/(r_k r_0)^{1/2}, \quad (54)$$

$$\sigma_m\big(\mathbf{\Phi}^{(\text{cano})}_{k,1:h_0}\big) - \sigma_{m+1}\big(\mathbf{\Phi}^{(\text{cano})}_{k,1:h_0}\big) \ge c_1 d^{\delta_0-\delta_1} h_0^{1/2} r/(r_k^3 r_0)^{1/2}, \quad (55)$$

$$\sigma_{r_k}\big(\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0}\big) \ge c_1 d^{\delta_0-\delta_1^*} h_0^{1/2} r/(r_k r^*_{k,0})^{1/2}, \quad (56)$$

$$\sigma_m\big(\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0}\big) - \sigma_{m+1}\big(\mathbf{\Phi}^{(\text{cano})*}_{k,1:h_0}\big) \ge c_1 d^{\delta_0-\delta_1^*} h_0^{1/2} r/(r_k^3 r^*_{k,0})^{1/2}, \quad (57)$$

for $1 \le m < r_k$ and the matrices in (31), (32), (30) and (33), and that with probability $1 + o(1)$

$$\left\|\overline{\mathbb{E}}\left[\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})}\right] - \mathbb{E}\left[\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})}\right]\right\|_{\text{S}} \ll d^{\delta_0-\delta_1} h_0^{1/2} r/(r_k^3 r_0)^{1/2}, \quad (58)$$

$$\left\|\overline{\mathbb{E}}\left[\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})*}\right] - \mathbb{E}\left[\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})*}\right]\right\|_{\text{S}} \ll d^{\delta_0-\delta_1^*} h_0^{1/2} r/(r_k^3 r_{k,0}^*)^{1/2} \quad (59)$$

where $\delta_1^* \ge \delta_0$, $\delta_1 \ge \delta_0$ and $c_1 > 0$ are constants, $r_0 = \text{trace}\left(\mathbf{\Phi}_0^{(\text{cano})}\right)/\left\|\mathbf{\Phi}_0^{(\text{cano})}\right\|_{\text{S}} \in [1, r]$ is the effective rank of $\mathbf{\Phi}_0^{(\text{cano})}$, and $r_{k,0}^* = \text{trace}\left(\mathbf{\Phi}_{k,0}^{(\text{cano})*}\right)/\left\|\mathbf{\Phi}_{k,0}^{(\text{cano})*}\right\|_{\text{S}} \in [1, r_k]$ is the effective rank of $\mathbf{\Phi}_{k,0}^{(\text{cano})*}$. In each statement in the sequel, only the referenced conditions among Equations (53)–(59) are imposed when the Factor Strength Model is invoked.

Some explanations of the model are in order: Condition (52) can be always achieved by scaling $\|\boldsymbol{A}_k\|_{\text{S}}$ properly and then picking δ_0 accordingly. By Equations (6) and (31), Equation (52) is equivalent to

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{\|\text{vec}(\mathcal{M}_t)\|_2^2}{\sigma^2 d\, T}\right] = d^{-\delta_0},$$

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{\|\text{vec}(\mathcal{F}_t^{(\text{cano})})\|_2^2}{r\, T}\right] = 1. \quad (60)$$

Thus, $d^{-\delta_0} = r\lambda^2/(\sigma^2 d)$ has the interpretation as the SNR, and the mean squared value of the canonical factor series $\{\mathcal{F}_t^{(\text{cano})}, 1 \le t \le T\}$ is normalized to 1. It follows from Equations (30) and (52) that Equation (54) is equivalent to $(r_k r_0)^{1/2}\sigma_{r_k}\left(\overline{\mathbb{E}}\left[\text{TOPUP}_k/h_0^{1/2}\right]\right) \ge c_1\sigma^2 d^{1-\delta_1}$, so that $d^{-\delta_1}$ can be viewed as the order of the SNR relevant to the TOPUP with proper adjustments for the ranks and the total number of lags involved. See Equations (61) and (62) in Lemma 1 and the discussion below the lemma for the rational of the adjustments. Similarly, by Equations (33) and (52) $d^{-\delta_1^*}$ can be viewed as the order of the SNR relevant to the TIPUP. For $K = 1$ and $\delta_0 = \delta_1$, the δ_0 in Equation (53) is comparable with the δ' in Lam, Yao, and Bathia (2011a). For $K = 2$ and $\delta_0 = \delta_1$, the d^{δ_0} in Equation (53) is comparable with the $d_1^{\delta_1'} d_2^{\delta_2'}$ in Wang, Liu, and Chen (2019). The Factor Strength Model is more general than the Fixed-Rank Factor Model as the factor series $\mathcal{F}_t$ is not required to be weakly stationary and $r_1, \ldots, r_K, h_0$ are allowed to diverge. In the Fixed-Rank Factor Model, Equations (53)–(57) all hold as their left-hand sides are all of the constant order, provided that the respective canonical population matrices $\mathbb{E}\left[\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})}\right]$ and $\mathbb{E}\left[\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})*}\right]$ are of full rank and have distinct singular values. The Factor Strength Model also allow weaker signals with $\delta_1 > \delta_0$ or $\delta_1^* > \delta_0$. The following lemma provides some relationships among the norms and singular values in Equations (53)–(59).

Lemma 1. Let λ and $\mathcal{F}_t^{(\text{cano})}$ be as in Equation (6) and $\mathbf{\Phi}_0^{(\text{cano})} \in \mathbb{R}^{r\times r}$ be as in Equation (31). Then,

$$r_0\left\|\mathbf{\Phi}_0^{(\text{cano})}\right\|_{\text{S}} = r_{k,0}^*\left\|\mathbf{\Phi}_{k,0}^{(\text{cano})*}\right\|_{\text{S}} = \text{trace}\left(\mathbf{\Phi}_0^{(\text{cano})}\right) = \sum_{t=1}^{T} \frac{\|\text{vec}(\mathcal{M}_t)\|_2^2}{\lambda^2 T}, \quad (61)$$

where $r_0 \in [1, r]$ and $r_{k,0}^* \in [1, r_k]$ are the effective ranks in Equations (54)–(59). Moreover,

$$\sigma_{r_k}\left(\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})}\right) \le \sum_{m=1}^{r_k} \frac{\sigma_m\left(\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})}\right)}{r_k} \le \frac{h_0^{1/2}\text{trace}\left(\mathbf{\Phi}_0^{(\text{cano})}\right)}{(r_k r_0)^{1/2}(1 - h_0/T)} \quad (62)$$

and

$$\sigma_{r_k}\left(\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})*}\right) \le \frac{\sum_{m=1}^{r_k}\sigma_m\left(\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})*}\right)}{r_k} \le \frac{h_0^{1/2}\text{trace}\left(\mathbf{\Phi}_0^{(\text{cano})}\right)}{(r_k r_{k,0}^*)^{1/2}(1 - h_0/T)}. \quad (63)$$

It follows from Equation (61) that Equation (53) holds when $\sum_{t=1}^{T}\|\text{vec}(\mathcal{F}_t^{(\text{cano})})\|_2^2/(rT) = 1 + o_{\mathbb{P}}(1)$, so that (53) is expected to follow from Equation (52). By Equation (53), the right-hand side of Equation (62) is of the order $h_0^{1/2} r/(r_k r_0)^{1/2}$. Thus, Equation (54) with $\delta_1 = \delta_0$ means that the two sides of Equation (62) are of the same order but Equation (54) also allows larger δ_1 for weaker signals. For the interpretation of Equation (55), we note that with $\delta_1 = \delta_0$ it follows from Equation (54) when all the singular values $\mathbf{\Phi}_{k,1:h_0}^{(\text{cano})}$ are of the same order and all the singular-value gaps are of the same order for each k, but Equation (55) also allows relaxation with $\delta_1 > \delta_0$. The interpretation of Equations (56) and (57) is parallel as the right-hand side of (63) is of the order $h_0^{1/2} r/(r_k r_{k,0}^*)^{1/2}$ by Equation (53). We allow $\delta_1^* \ge \delta_1$ to take into account the impact of potential signal omission and cancellation by the TIPUP. Finally, Equations (58) and (59) simply guarantee the population version of Equations (55) and (57), respectively. Again, we leave to the existing literature for the analysis of the low-dimensional matrices in Equations (53)–(59), as many options are available.

Corollary 3. Let $h_0 \le T/4$. In the Factor Strength Model, the TOPUP yields

$$\left\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\right\|_{\text{S}} \lesssim d^{\delta_1-\delta_0}\sqrt{d_{-k}}\sqrt{(r_0/r_{k,0}^*) \vee 1} \times\left\{(\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T}/r_{-k}\right\} \quad (64)$$

under conditions (53) and (54), and for $1 \le m < r_k$ and the $\boldsymbol{P}_{k,m}$ in (40)

$$\left\|\widehat{\boldsymbol{P}_{k,m}} - \boldsymbol{P}_{k,m}\right\|_{\text{S}} \lesssim d^{\delta_1-\delta_0}\sqrt{d_{-k}}r_k\sqrt{(r_0/r_{k,0}^*) \vee 1} \times\left\{(\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T}/r_{-k}\right\} \quad (65)$$

under conditions (53), (55) and (58), where $d_{-k} = d/d_k$ and $r_{-k} = r/r_k$; The TIPUP yields

$$\left\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\right\|_{\text{S}} \lesssim d^{\delta_1^*-\delta_0}\left\{(\sigma/\lambda)\sqrt{d_k/(r_{-k}T)} + (\sigma/\lambda)^2\sqrt{d/T}/r_{-k}\right\} \quad (66)$$

under conditions (53) and (56), and for $1 \le m < r_k$ and the $\boldsymbol{P}_{k,m}^*$ in (48)

$$\left\|\widehat{\boldsymbol{P}_{k,m}^*} - \boldsymbol{P}_{k,m}^*\right\|_{\text{S}} \lesssim d^{\delta_1^*-\delta_0} r_k\left\{(\sigma/\lambda)\sqrt{d_k/(r_{-k}T)} + (\sigma/\lambda)^2\sqrt{d/T}/r_{-k}\right\} \quad (67)$$

under conditions (53), (57) and (59).

Compared with the results in Lam, Yao, and Bathia (2011a) and Wang, Liu, and Chen (2019) and Corollaries 1 and 2 which involve $\{d_1, \ldots, d_K, \lambda, T\}$, the bounds in Corollary 3 also involve the ranks $\{r_0, r^*_{k,0}, r_1, \ldots, r_K\}$ as they are allowed to diverge in the Factor Strength Model. We note that while $\boldsymbol{P}_{k,m} = \boldsymbol{P}^*_{k,m}$ for $m = r_k$, the estimation targets in Equations (65) and (67) are not the same in general for $1 \le m < r_k$ as discussed in the paragraph below (25).

For the vector series $\boldsymbol{x}_t = \lambda \boldsymbol{U}_1 \boldsymbol{f}_t^{(\text{cano})} + \boldsymbol{e}_t$ in the canonical form with $\boldsymbol{U}_1$ being the left singular matrix of $\boldsymbol{A}_1$ and $\lambda = \|\boldsymbol{A}_1\|_{\text{S}}$, the TOPUP and TIPUP are identical, and $\boldsymbol{\Phi}^{(\text{cano})}_{1,h} = \boldsymbol{\Phi}^{(\text{cano})*}_{1,h} = \sum_{t=h+1}^{T} \boldsymbol{f}^{(\text{cano})}_{t-h} \boldsymbol{f}^{(\text{cano})\top}_{t} / (T-h)$, $\boldsymbol{\Phi}^{(\text{cano})}_0 = \boldsymbol{\Phi}^{(\text{cano})}_{1,0} = \boldsymbol{\Phi}^{(\text{cano})*}_{1,0}$ in Equations (29)–(32) and (52)–(57). and the rates in Corollary 3 are simplified to

$$\big\|\widehat{\boldsymbol{P}}_1 - \boldsymbol{P}_1\big\|_{\text{S}} \lesssim d_1^{\delta_1 - \delta_0}(\sigma/\lambda)\sqrt{d_1/T}(1 + \sigma/\lambda).$$

While the conditions in the Factor Strength Model are all expressed in terms of the canonical factor series, Equation (64) of Corollary 3 for the TOPUP estimation of $\boldsymbol{P}_k$ also leads to the oracle inequality under parallel conditions on the natural factor series $\mathcal{F}_t$, as $\mathcal{F}^{(\text{cano})}_t = \mathcal{F}_t \otimes_{k=1}^{K} (\boldsymbol{D}_k \boldsymbol{V}_k^{\top})$ in view of Equations (4) and (6). Similar to Equations (29), (32), (30) and (31), define

$$\boldsymbol{\Phi}_{k,h} = \sum_{t=h+1}^{T} \frac{\text{mat}_k(\mathcal{F}_{t-h} \otimes \mathcal{F}_t)}{T-h},$$

$$\boldsymbol{\Phi}^*_{k,h} = \sum_{t=h+1}^{T} \frac{\text{mat}_k(\mathcal{F}_{t-h})\text{mat}_k^{\top}(\mathcal{F}_t)}{T-h},$$

$\boldsymbol{\Phi}_{k,1:h_0} = \big(\boldsymbol{\Phi}_{k,1}, \ldots, \boldsymbol{\Phi}_{k,h_0}\big)$ and $\boldsymbol{\Phi}_0 = \sum_{t=1}^{T} \text{vec}(\mathcal{F}_t)\text{vec}^{\top}(\mathcal{F}_t)/T$. Corollary 3 yields the following.

Corollary 4. Let $d^{-\delta_0}$ be the SNR in the sense of $\sum_{t=1}^{T} \|\text{vec}(\mathcal{M}_t)\|_2^2/(\sigma^2 dT) \asymp d^{-\delta_0}$. Let $\lambda = \prod_{k=1}^{K} \|\boldsymbol{A}_k\|_{\text{S}}$. Suppose $\boldsymbol{A}_k$ are scaled such that $\lambda^2 = \sigma^2 d^{1-\delta_0}/r$. Let $\boldsymbol{U}_k \boldsymbol{D}_k \boldsymbol{V}_k^{\top}$ be the SVD of $\boldsymbol{A}_k/\|\boldsymbol{A}_k\|_{\text{S}}$ and $\kappa_0 = \prod_{k=1}^{K} \sigma_{r_k}(\boldsymbol{D}_k)$. If $\kappa_0^2 \sigma_{r_k}\big(\boldsymbol{\Phi}_{k,1:h_0}\big) \ge c_1 d^{\delta_0 - \delta_1} h_0^{1/2} r/(r_k r_0)^{1/2}$, then

$$\big\|\widehat{\boldsymbol{P}}_k - \boldsymbol{P}_k\big\|_{\text{S}} \lesssim d^{\delta_1 - \delta_0}\sqrt{d_{-k}}\sqrt{(r_0/r^*_{k,0}) \vee 1}\big\{(\sigma/\lambda)\sqrt{d_k/T} + (\sigma/\lambda)^2\sqrt{d/T}/r_{-k}\big\}, \tag{68}$$

where $\widehat{\boldsymbol{P}}_k$ is the TOPUP estimate of $\boldsymbol{P}_k$, r_0 and $r^*_{k,0}$ are the effective ranks in Lemma 1. Moreover, $\kappa_0^2 \le r_0 \|\boldsymbol{\Phi}_0\|_{\text{S}}/\text{trace}(\boldsymbol{\Phi}_0) \le \kappa_0^{-2}$ and $\kappa_0^2 \le r^*_{k,0}\|\boldsymbol{\Phi}^*_{0,k}\|_{\text{S}}/\text{trace}(\boldsymbol{\Phi}_0) \le \kappa_0^{-2}$.

It follows from Corollary 4 that for the TOPUP estimation of the projection $\boldsymbol{P}_k$ to the column space of $\boldsymbol{A}_k$, conditions for Equation (64) can be equivalently stated in terms of the natural factor series when the condition numbers $\|\boldsymbol{D}_k^{-1}\|_{\text{S}}^2$ of $\boldsymbol{A}_k^{\top}\boldsymbol{A}_k$ are uniformly bounded in the Factor Strength Model. However, parallel variation of Equation (66) in the TIPUP theory is unclear as the impact of $\boldsymbol{D}_k$ inside the inner-product in Equations (18) and (32) cannot be directly deciphered for $h \ne 0$. Similarly, neither parallel variations of Equations (65) and (67) are available as the impact of $\boldsymbol{D}_k$ on the gaps of the singular values is unclear. Below we consider a special case with more explicit results.

Bi-Orthogonal CP Factor Model: Let $r_1 = \cdots = r_k$ and $\boldsymbol{A}_k = (\boldsymbol{a}_{k,1}, \ldots, \boldsymbol{a}_{k,r_1})$. Suppose

$$\mathcal{M}_t = \mathcal{F}_t \otimes_{k=1}^{K} \boldsymbol{A}_k = \lambda \sum_{i=1}^{r_1} \kappa_i f_{it}\Big(\otimes_{k=1}^{K} \boldsymbol{b}_{k,i}\Big), \tag{69}$$

where $\lambda = \prod_{k=1}^{K} \|\boldsymbol{A}_k\|_{\text{S}}$, $\boldsymbol{b}_{k,i} = \boldsymbol{a}_{k,i}/\|\boldsymbol{a}_{k,i}\|_2$, $\kappa_i = \big(\prod_{k=1}^{K} \|\boldsymbol{a}_{k,i}\|_2\big)/\lambda$ and the factor series $\mathcal{F}_t$ is diagonal with elements $f_{i_1 i_2 \cdots i_K t} = f_{it} I\{i_1 = \ldots = i_K = i\}$. We assume that the loading vectors $\boldsymbol{a}_{k,i}$ are bi-orthogonal: For each $i_1 \ne i_2$, $\boldsymbol{a}_{k,i_1}^{\top}\boldsymbol{a}_{k,i_2} = 0$ for at least two $k \le K$, or equivalently

$$\min_{1 \le i_1 < i_2 \le r_1} \#\big\{k \le K : \boldsymbol{b}_{k,i_1}^{\top}\boldsymbol{b}_{k,i_2} = 0\big\} \ge 2. \tag{70}$$

Moreover, without loss of generality by scaling, assume always that $\mathbb{E}\big[\sum_{t=1}^{T} f_{it}^2/T\big] = 1$ and that Equations (52) and (60) hold with the SNR $d^{-\delta_0}$ and $\lambda^2 = \sigma^2 d^{1-\delta_0}/r_1^K$ in agreement with the λ in Equation (69). We shall call (69) the Fully Orthogonal CP Factor Model when $\mathbb{E}\big[\sum_{t=h+1}^{T} f_{i_1,t-h} f_{i_2,t}/T\big] = 0$ and $\boldsymbol{b}_{k,i_1}^{\top}\boldsymbol{b}_{k,i_2} = 0$ for all $1 \le i_1 < i_2 \le r_1$ and $1 \le h \le h_0$.

Proposition 1. In the above Bi-Orthogonal CP Factor Model, let $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{r_1 t})^{\top}$, $\overline{\boldsymbol{\Sigma}}_{f,h} = (T-h)^{-1}\sum_{t=h+1}^{T} \boldsymbol{f}_{t-h}\boldsymbol{f}_t^{\top}$, $\boldsymbol{D}_0 = \text{diag}(\kappa_1, \ldots, \kappa_{r_1}) \in \mathbb{R}^{r_1 \times r_1}$, $\overline{\boldsymbol{\Sigma}}_{\kappa f,h} = \boldsymbol{D}_0 \overline{\boldsymbol{\Sigma}}_{f,h} \boldsymbol{D}_0$ and $\boldsymbol{B}_k = (\boldsymbol{b}_{k,1}, \ldots, \boldsymbol{b}_{k,r_1})$. Then, $r_1^K = r = \mathbb{E}\big[\text{trace}\big(\boldsymbol{\Phi}_0^{(\text{cano})}\big)\big] = \mathbb{E}\big[\text{trace}\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,0}\big)\big] = \sum_{i=1}^{r_1} \kappa_i^2$ and

$$\sigma_m\big(\boldsymbol{\Phi}_0^{(\text{cano})}\big) = \sigma_m\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,0}\big), \tag{71}$$

$$\sigma_m\big(\boldsymbol{\Phi}^{(\text{cano})}_{k,1:h_0}\big) = \sigma_m^{1/2}\big(\textstyle\sum_{h=1}^{h_0} \boldsymbol{B}_k \text{diag}\big(\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,h}\big)\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,h}\big)^{\top}\big)\boldsymbol{B}_k^{\top}\big), \tag{72}$$

$$\sigma_m\big(\boldsymbol{\Phi}^{(\text{cano})*}_{k,0}\big) = \sigma_m\big(\boldsymbol{B}_k \text{diag}\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,0}\big)\boldsymbol{B}_k^{\top}\big), \tag{73}$$

$$\sigma_m\big(\boldsymbol{\Phi}^{(\text{cano})*}_{k,1:h_0}\big) = \sigma_m^{1/2}\big(\textstyle\sum_{h=1}^{h_0} \big(\boldsymbol{B}_k \text{diag}\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,h}\big)\boldsymbol{B}_k^{\top}\big)^2\big). \tag{74}$$

Thus, in Equations (52)–(59), $\boldsymbol{\Phi}_0^{(\text{cano})}$, $\boldsymbol{\Phi}^{(\text{cano})*}_{k,h}$ and $\boldsymbol{\Phi}^{(\text{cano})}_{k,h}$ can be replaced by $\overline{\boldsymbol{\Sigma}}_{\kappa f,0}$, $\boldsymbol{B}_k \text{diag}\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,h}\big)\boldsymbol{B}_k^{\top}$ and $\big\{\boldsymbol{B}_k \text{diag}\big(\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,h}\big)\big(\overline{\boldsymbol{\Sigma}}_{\kappa f,h}\big)^{\top}\big)\boldsymbol{B}_k^{\top}\big\}^{1/2}$ respectively.

In the Bi-Orthogonal CP Factor Model with orthonormal $\boldsymbol{B}_k$, conditions (54) and (56) for the estimation of $\boldsymbol{P}_k$ become

$$\min_{1 \le i \le r_1} \Big\{\sum_{h=1}^{h_0}\sum_{j=1}^{r_1} \big(\overline{\boldsymbol{\Sigma}}_{\kappa f,h}\big)_{i,j}^2\Big\}^{1/2} \ge c_1 d^{\delta_0 - \delta_1} r_1^K \sqrt{h_0}/(r_1 r_0)^{1/2},$$

$$\min_{1 \le i \le r_1} \Big\{\sum_{h=1}^{h_0} \big(\overline{\boldsymbol{\Sigma}}_{\kappa f,h}\big)_{i,i}^2\Big\}^{1/2} \ge c_1 d^{\delta_0 - \delta_1^*} r_1^K \sqrt{h_0}/(r_1 r^*_{k,0})^{1/2},$$

respectively by Proposition 1 with $(\overline{\boldsymbol{\Sigma}}_{\kappa f,h})_{i,j} = \kappa_i \kappa_j \sum_{t=h+1}^{T} f_{i,t-h} f_{j,t}/(T-h)$. Thus, while the TOPUP uses the information in all elements of the matrix $\overline{\boldsymbol{\Sigma}}_{\kappa f,h}$, the TIPUP would fail due to the omission of signals in the off-diagonal elements when $\mathbb{E}[(\overline{\boldsymbol{\Sigma}}_{\kappa f,h})_{ii}] = 0$ for all $1 \le h \le h_0$ for at least one $i \le r_1$. When $\boldsymbol{B}_k$ is not orthonormal in the Bi-Orthogonal CP Factor Model, the TIPUP is also subject to mild signal cancellation in the diagonal elements of the

matrix. However, the TIPUP enjoys noise omission and noise cancellation as exhibited in the faster rates in Equations (66) and (67) under the respective conditions. In the Fully Orthogonal CP Factor Model where $\boldsymbol{B}_k$ is orthonormal and $\mathbb{E}[\overline{\boldsymbol{\Sigma}}_{\kappa f,h}]$ is diagonal, $\sigma_m\big(\mathbb{E}[\boldsymbol{\Phi}^{(\text{cano})}_{k,1:h_0}]\big) = \sigma_m\big(\mathbb{E}[\boldsymbol{\Phi}^{(\text{cano})*}_{k,1:h_0}]\big)$ and $\sigma_m\big(\mathbb{E}[\boldsymbol{\Phi}^{(\text{cano})}_{0}]\big) = \sigma_m\big(\mathbb{E}[\boldsymbol{\Phi}^{(\text{cano})*}_{k,0}]\big)$ so that the TIPUP yields smaller error bounds in Corollary 3 with $\delta_1^* = \delta_1$.

6. Simulation Results

Here we present some empirical study on the performance of the estimation procedures, with various experimental configurations. We also check the performance of a standard tensor decomposition procedure which incorporates time as an additional tensor dimension, and treats the factor as deterministic without temporal structure. The loading matrices are then estimated using SVD of the mode-1 (or 2, 3) matricization of the expanded tensor $\mathcal{Y}$ in (3) to estimate the column space of $\boldsymbol{A}_1$ (or $\boldsymbol{A}_2$, $\boldsymbol{A}_3$). We will call it the unfolding procedure (UP). The main difference between UP and the estimators TIPUP and TOPUP is that UP does not incorporate the assumption that the noise is white, while the TIPUP and TOPUP take full advantage of that assumption.

To compare the performance of different methods in finite samples, we generated observations from the following two-dimensional model:

$$\boldsymbol{X}_t = 2\boldsymbol{A}_1\boldsymbol{F}_t\boldsymbol{A}_2^\top + \boldsymbol{E}_t \tag{75}$$

where $\boldsymbol{F}_t = [f_{1t}, f_{2t}]$ is a 1×2 factor, with two independent AR(1) processes $f_{it} = \phi_i f_{it-1} + e_{it}$. The noise $\boldsymbol{E}_t$ is generated the same way as the simulation in the rank one case. The elements of the loadings $\boldsymbol{A}_1$ (a $d_1 \times 1$ matrix) and $\boldsymbol{A}_2$ (a $d_2 \times 2$ matrix) are generated from iid N(0,1), then normalized so that $\|\boldsymbol{A}_1\|_2 = 1$ ($\boldsymbol{A}_1$ is vector) and $\boldsymbol{A}_2$ is orthonormal through QR decomposition. We use dimension $d_1 = d_2 = 16$ here. Figures 2 and 3 show the comparison of the estimation methods, using boxplots of the logarithm of the estimation error in 100 simulation runs. The estimation error of $\boldsymbol{A}_1$ is calculated the same way as that in the rank one case (since $\boldsymbol{A}_1$ is a vector). The estimation error of $\boldsymbol{A}_2$ is the spectral norm of the difference between $\widehat{\boldsymbol{P}}_2 = \hat{\boldsymbol{A}}_2(\hat{\boldsymbol{A}}_2^\top\hat{\boldsymbol{A}}_2)^{-1}\hat{\boldsymbol{A}}_2^\top$ and $\boldsymbol{P}_2 = \boldsymbol{A}_2(\boldsymbol{A}_2^\top\boldsymbol{A}_2)^{-1}\boldsymbol{A}_2^\top$, that is, the sine of the largest canonical angle between the column spaces of $\hat{\boldsymbol{A}}_2$ and $\boldsymbol{A}_2$. We consider TOPUP and TIPUP with $h_0 = 1$ and $h_0 = 2$. UP denotes the results using simple tensor decomposition via unfolding. The top panel is for estimating $\boldsymbol{A}_1$ and the bottom panel for $\boldsymbol{A}_2$.

Figure 2 shows the results of using $\phi_1 = 0.8$ and $\phi_2 = -0.8$ in the AR processes of the factors and sample sizes $T = 256$ and 1024. Note that with $\phi_1 = 0.8$ and $\phi_2 = -0.8$, we have

$$\mathbb{E}\big[\boldsymbol{\Phi}^{(\text{cano})*}_{k=1,h}\big] = \mathbb{E}[\boldsymbol{F}_t\boldsymbol{F}_{t-h}^\top] = (\phi_1^h + \phi_2^h)\sigma_f^2,\ h = 0, 1, 2,$$

so that $\mathbb{E}\big[\boldsymbol{\Phi}^{(\text{cano})*}_{k=1,1:(h_0=1)}\big] = 0$. It violates the full-rank condition for the TIPUP in estimating $\boldsymbol{A}_1$ with $h_0 = 1$. Essentially the signal in $\sum_{t=h+1}^{T}\boldsymbol{X}_t\boldsymbol{X}_{t-1}^\top$ completely cancelled out so that the results of the TIPUP with $h_0 = 1$ in the top two panels in Figure 2 are significantly worse than the respective results of the TOPUP with $\mathbb{E}\big[\boldsymbol{\Phi}^{(\text{cano})}_{k=1,1:(h_0=1)}\big] = (\phi_1\sigma_f^2, 0, 0, \phi_2\sigma_f^2)$. On the other hand, for $h_0 = 2$, the signal cancellation does not happen to the TIPUP for the $h = 2$ term, $\mathbb{E}\big[\boldsymbol{\Phi}^{(\text{cano})*}_{k=1,1:(h_0=2)}\big] = (0, (\phi_1^2 + \phi_2^2)\sigma_f^2)$, so that the TIPUP is comparable with the TOPUP with $\mathbb{E}\big[\boldsymbol{\Phi}^{(\text{cano})}_{k=1,1:(h_0=2)}\big] = (\phi_1\sigma_f^2, 0, 0, \phi_2\sigma_f^2, \phi_1^2\sigma_f^2, 0, 0, \phi_2^2\sigma_f^2)$ as well as the TOPUP with $h_0 = 1$. Still, for the larger T, the performance of the TIPUP with $h_0 = 2$ is slightly worse compared with the TOPUP with either $h_0 = 2$ or $h_0 = 1$, due to partial signal cancellation and weaker signal for lag 2. The TOPUP does not have such a cancellation problem since it is based on the sum of the squares of column-wise autocovariance. The cancellation problem should not be very common in practice. For example, there is no cancellation for estimating $\boldsymbol{A}_2$ when using the TIPUP in this setting, since $\mathbb{E}[\boldsymbol{F}_t^\top\boldsymbol{F}_{t-1}]$ is a full rank matrix. Since the TIPUP in general has a faster convergence rate, its performance

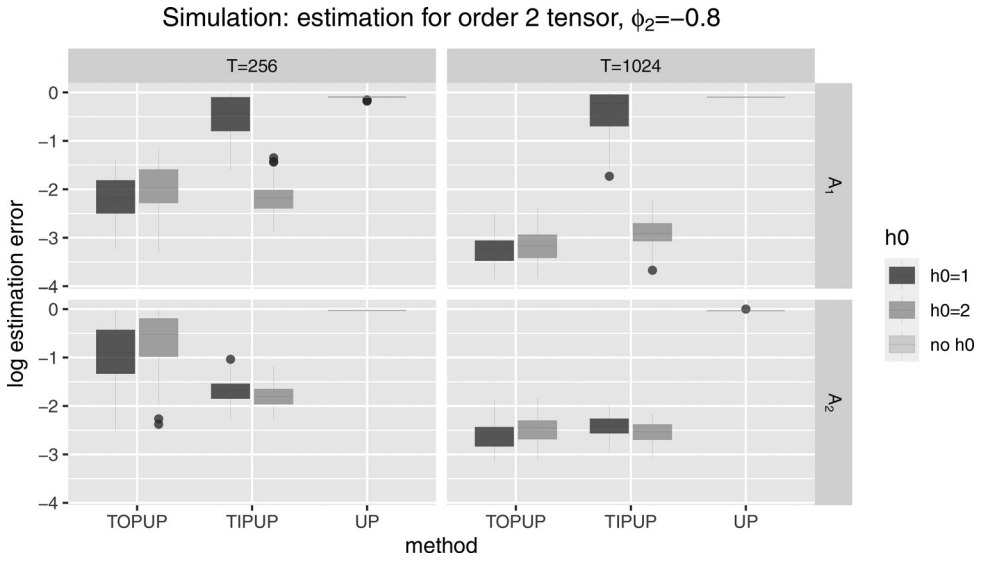

Figure 2. Finite sample comparison between TIPUP, TOPUP with different h_0 and UP under model (75) where $\phi_1 = 0.8$, $\phi_2 = -0.8$. The boxplots show the logarithms of the estimation errors. The top row is for estimation of the column space of the mode 1 loading matrix $\boldsymbol{A}_1$ and bottom for $\boldsymbol{A}_2$. The left column is for $T = 256$ and right for $T = 1024$.

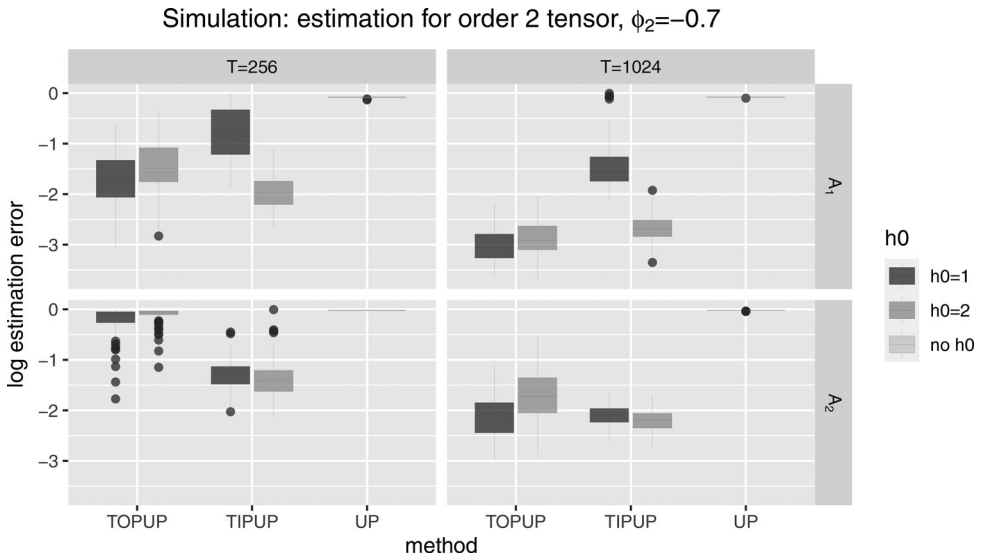

Figure 3. Finite sample comparison between TIPUP, TOPUP with different h_0 and UP under model (75) where $\phi_1 = 0.8$, $\phi_2 = -0.7$. The boxplots show the logarithms of the estimation errors. The top row is for estimation of the column space of the mode 1 loading matrix $\boldsymbol{A}_1$ and bottom for $\boldsymbol{A}_2$. The left column is for $T = 256$ and right for $T = 1024$.

is better than that of the TOPUP, especially for small sample sizes, as shown in the bottom panel of Figure 2.

Figure 3 shows the results of using $\phi_1 = 0.8$ and $\phi_2 = -0.7$ in the AR processes of the factors. Here although $\mathbb{E}[\boldsymbol{F}_t \boldsymbol{F}_{t-1}^\top]$ is not zero, TIPUP with $h_0 = 1$ for estimating $\boldsymbol{A}_1$ is still worse than TOPUP due to the partial cancellation, though not as severe as that in the complete cancellation $\phi_2 = -0.8$ case.

The UP procedure is always the worst performer in our experiments, most probably due to the presence of contemporary correlation in the noise $\boldsymbol{E}_t$ and the fact that TOPUP and TIPUP utilize the whiteness assumption on $\boldsymbol{E}_t$. Additional simulation results, including the relative estimation error comparing TTPUP or TOPUP with UP for this example, are given in Appendix C.1. An additional simulation example on the analysis of order 3 tensor time series is given in Appendix C.2, which shows similar results to the matrix times series example here. The simulation study to verify the theoretical results is given in Appendix C.4.

7. Applications

7.1. Tensor Factor Models for Import–Export Transport Networks

Here we analyze the multi-category import-export network data as illustrated in Figure 1. The dataset contains the monthly total export among 22 countries in North American and Europe in 15 product categories from January 2010 to December 2016 (length 84), so that the original dataset can be viewed as a four-way tensor of dimension $22 \times 22 \times 15 \times 84$, with missing values for the export from any country to itself. For simplicity, we treated the missing diagonal values as zero in the analysis. More sophisticated imputation can be implemented. The details of the data, countries and product categories are given in Appendix B. Following Linnemann (1966), to reduce the effect of incidental transactions of large trades or unusual shipping delays, a three-month moving average of the series is used, so that $\mathcal{X}_t \in \mathbb{R}^{22\times 22\times 15}$ with $t = 1, \ldots, 82$. Each element $x_{i,j,k,t}$ is the three-month moving average of total export from country i to country j in category k in the tth month.

Figure 4 shows the total volume from year 2010 to 2017 in two categories of products (Machinery and Electronic, and Footwear and Headwear) among 22 countries in North American and Europe. The arrows show the trade direction and the width of an arrow reflects the volume of the trade. Clearly the networks are quite different for different product categories. For example, Mexico is a large importer and exporter of Machinery and Electronic as it serves as one of the major part suppliers in the product chain of machinery and electronics. On the other hand, Italy is the largest exporter of Footwear and Headwear.

Under our general framework presented in Section 3, we use the following model for the dynamic transport networks. Let $\mathcal{X}_t$ be the observed tensor at time t. The element $x_{i_1 i_2 i_3, t}$ is the trading volume from country i_1 (the exporter) to country i_2 (the importer) of product type i_3. Let

$$\mathcal{X}_t = \mathcal{F}_t \times_1 \boldsymbol{A}_1 \times_2 \boldsymbol{A}_2 \times_3 \boldsymbol{A}_3 + \mathcal{E}_t \tag{76}$$

where $\mathcal{X}_t \in \mathbb{R}^{d_1\times d_2\times d_3}$ ($d_1 = d_2$), $\mathcal{F}_t \in \mathbb{R}^{r_1\times r_2\times r_3}$ ($r \ll d$), and $\boldsymbol{A}_i \in \mathbb{R}^{d_i\times r_i}$. This is similar to the DEDICOM model (Harshman 1978; Kolda, Bader, and Kenny 2005; Kolda and Bader 2006). Chen and Chen (2019) provided some interpretations of the factors in a uni-category import-export network under the matrix factor model setting of Wang, Liu, and Chen (2019).

In the following, we provide some interpretation of the model. Consider the loading matrix $\boldsymbol{A}_3$. It can be viewed as the loading matrix of a standard factor model

$$\boldsymbol{x}_{i_1 i_2 \cdot, t} = \boldsymbol{A}_3 \boldsymbol{f}^{(3)}_{i_1 i_2 \cdot, t} + \boldsymbol{\varepsilon}^{(3)}_{i_1 i_2 \cdot, t}$$

of the mode-3 fiber $\boldsymbol{x}_{i_1 i_2 \cdot, t}$ for all (t, i_1, i_2). This is essentially unfolding the order 4 tensor $\mathcal{Y}$ with dimensions $d_1 \times d_2 \times d_3 \times T$ into a $d_3 \times (d_1 d_2 T)$ matrix and fit a standard factor model with $d_1 d_2 T$ factors, each a vector of dimension r_3. These factors drive the co-moment of all mode-3 fibers $\mathcal{X}_{i_1, i_2, \cdot, t}$ at time t. The loading matrix reflects how each element of the mode-3

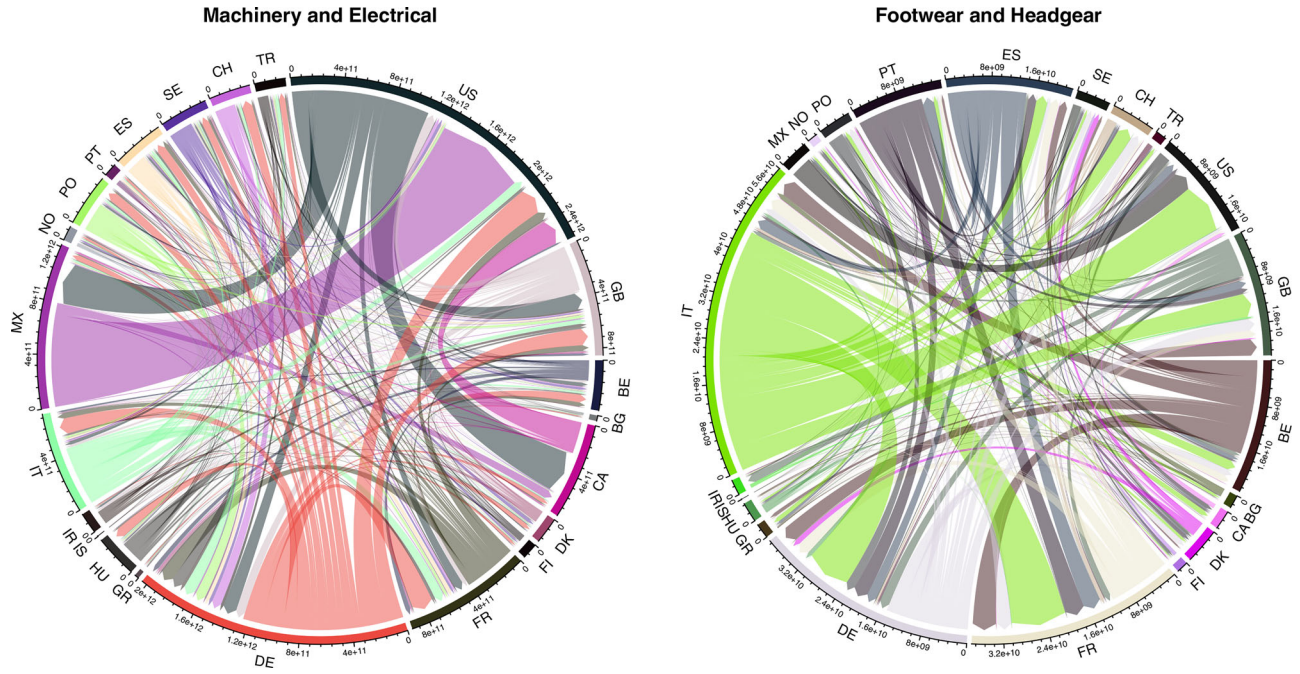

Figure 4. Time aggregated import-export volume of Machinery and Electronic products and Footwear and Headwear products among 22 countries in North American and Europe.

Table 1. Estimated loading matrix $\boldsymbol{A}_3$ for category fiber.

	Animal	*Vegetable*	Food	*Mineral*	*Chemical*	*Plastic*	Leather	*Wood*	*Textile*	Footwear	*Stone*	*Metal*	*Machinery*	*Transportation*	*Misc*
1	0	2	0	0	−1	0	0	−2	1	0	−1	−1	**29**	0	7
2	0	1	0	**30**	−1	0	0	2	−1	0	0	1	0	0	3
3	0	−1	1	0	**29**	1	0	0	0	0	0	0	−1	0	8
4	0	0	2	0	−1	0	0	2	−1	0	0	1	0	**30**	3
5	6	5	6	0	2	16	1	7	6	1	4	19	3	0	−9
6	−1	0	1	0	−1	−3	0	1	0	0	**29**	−1	−1	0	6

NOTE: Matrix is rotated via varimax. Elements are multiplied by 30 and truncated to integer.

fiber is related to the factors. Note that this scheme is only for interpretation. Th estimation procedure is based on a different set-up.

Table 1 shows an estimate of $\boldsymbol{A}_3$ of the import-export data under the tensor factor model, using $r_3 = 6$ factors. The estimation is based on the TIPUP procedure with $h_0 = 2$. The loading matrix is rotated using the varimax procedure for better interpretation. All numbers are multiplied by 30 then truncated to integers for clearer viewing.

It can be seen that there is a group structure. For example, Factors 1, 2, 3, 4, and 6 can be interpreted as the Machinery and Electrical factor, Mineral factor, Chemicals factor, Transportation factor, and Stone and Glass factor, respectively, since the corresponding product categories load heavily and almost exclusively on them. On the other hand, Factor 5 is mixed, with large loadings by Metals and Plastics/Rubbers, and medium loadings by Animal, Vegetable, and Food products. We will view each factor as a "condensed product group." Figure 5 shows the clustering of the product categories according to their loading vectors.

The factor matrix $\mathcal{F}_{\cdot,\cdot,i_3,t}$ (for a fixed i_3) can be viewed as the trading pattern among several *trading hubs* for the i_3-th condensed product groups (product factor). One can imagine that the export of a product by a country would first go through a virtual "export hub," then to a virtual "import hub," before

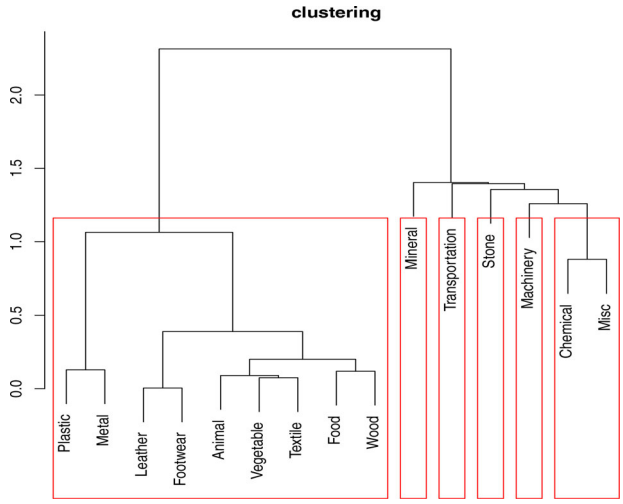

Figure 5. Clustering of product categories by their loading coefficients

arriving at the country that imports the product. Each row of the matrix $\boldsymbol{F}_{\cdot,\cdot,i_3,t}$ represents an export hub and each column represents an import hub. The elements $\mathcal{F}_{i_1,i_2,i_3,t}$ can be viewed as the volume of the condensed product group i_3 moved from export hub i_1 to import hub i_2 at time t. The corresponding loading matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ reflects the trading activities of

Table 2. Estimated loading matrix A_1 for the export fiber (hub).

	BE	BG	*CA*	DK	FI	*FR*	*DE*	GR	HU	IS	*IR*	IT	*MX*	*NO*	*PO*	*PT*	*ES*	SE	*CH*	*TR*	*US*	GB
1	4	0	**80**	0	1	2	−4	0	0	0	5	0	−3	3	0	1	2	0	3	0	3	5
2	−1	0	−4	0	0	1	−6	0	1	0	−2	2	1	1	1	0	0	0	1	0	**102**	1
3	9	0	−1	2	1	12	29	0	2	0	7	9	−3	1	4	1	7	3	6	2	1	8
4	−8	0	5	0	0	0	15	0	0	0	−7	2	**104**	−2	−2	−1	−4	1	−3	−1	0	2

NOTE: Matrix is rotated via varimax and column normalized. Values are in percentage.

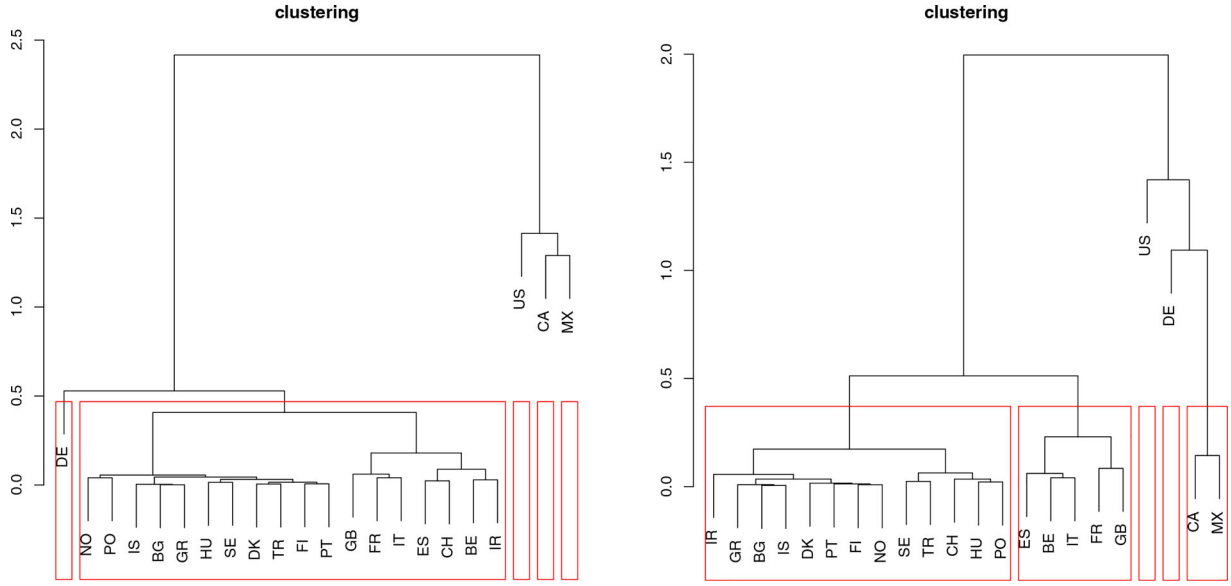

Figure 6. Clustering of countries by their export (left) and import (right) loading coefficients

Table 3. Estimated loading matrix A_2 for the import fiber (hub). Matrix is rotated via varimax and column normalized. Values are in percentage.

	BE	BG	*CA*	DK	FI	*FR*	*DE*	GR	*HU*	IS	*IR*	*IT*	*MX*	NO	*PO*	*PT*	*ES*	*SE*	*CH*	TR	*US*	*GB*
1	1	0	2	0	0	−1	0	0	0	0	0	0	−2	0	0	0	1	0	0	0	**100**	−1
2	0	0	57	0	0	2	−5	0	0	0	1	−2	44	0	−1	−1	−2	−1	0	1	0	7
3	10	1	−2	3	2	22	−3	1	4	0	1	11	0	2	6	2	6	5	8	4	0	18
4	7	0	4	0	0	0	**68**	1	−2	0	4	4	1	1	−2	1	8	0	−2	2	0	5

each country through each of the export and import hubs, respectively. We normalize each column of the loading matrices to sum up to one, so the value can be viewed as the proportion of activities of each country contributes to the hubs. Tables 2 and 3 show the estimated loading matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ after varimax rotation and column normalization, using four export hubs (E1 to E4) and four import hubs (I1 to I4). All values are in percentage. There are a few negative values since we do not constrain the loadings to be positive. The interpretation of the negative values is tricky. Fortunately, there are not many and the values are small. From Table 2, it is seen that Canada, the United States and Mexico heavily load on export hubs E1, E2, and E4, respectively, while European countries mainly load on export hub E3. The clustering based on loading coefficients of $\boldsymbol{A}_1$ of each country is shown in the left panel of Figure 6. The three countries in North America are very different from the European countries. In Europe, Germany behaves differently from the others as an exporter. For imports, seen from Table 3, the United States and Germany load heavily on hubs I1 and I4, respectively, while Canada and Mexico share hub I2. The European countries other than Germany mainly load on hub I3. The clustering based on loading coefficients of $\boldsymbol{A}_2$ of each country is shown in the right panel of Figure 6. It seems that the European countries (other than Germany) can be divided into two groups of similar import behavior, mainly based on the size of their economies.

The left panel of Figure 7 shows the trade transport network for the condensed product group 1 (mainly Machinery and Electrical). Several interesting features emerge. Export hub E3 (European hub) has the largest trading volume, and the goods mainly go to import hub I3 (European hub) and hub I1 (U.S. hub). This is understandable as trades among the many countries in Europe accumulate, and the United States is one of the largest importers. Mexico dominates export hub E4 and it mainly exports to import hub I1, used by the United States, confirming what is shown in the left panel of Figure 4. The United States is also a large exporter of machinery and electrical, occupying export hub E2, which mainly exports to import hub I2 used by Mexico and Canada.

On the other hand, for the network of condensed product group 2 (mainly mineral products) shown by the right panel of Figure 7, the dynamic is quite different. Export hub E1, mainly used by Canada, is the largest hub for mineral products. The import hub I1 is the largest import hub, mainly used by the

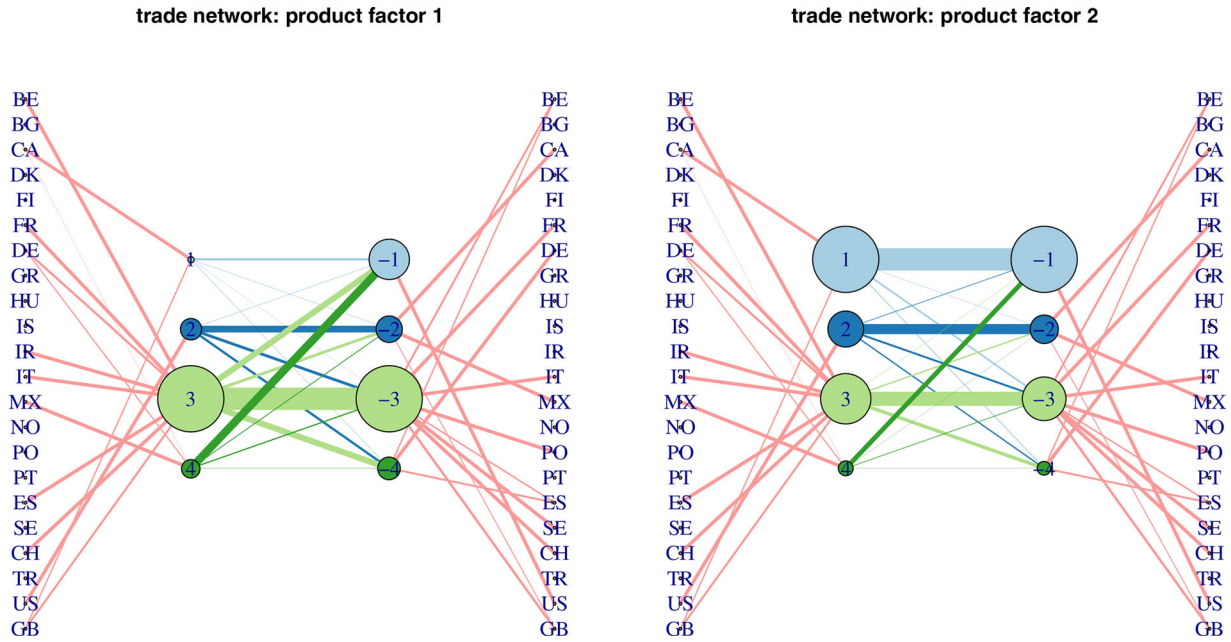

Figure 7. Trade network for condensed product group 1 (left) and group 2 (right). Export and import hubs are on the left and right of the center network respectively. Line width is proportional to the total volume of trade between the hubs for the last three years (2015 to 2017). Vertex size is proportional to total volume of trades through the hub. The line width between the countries and the hubs is proportional to the corresponding loading coefficients, for coefficients larger than 0.05 only.

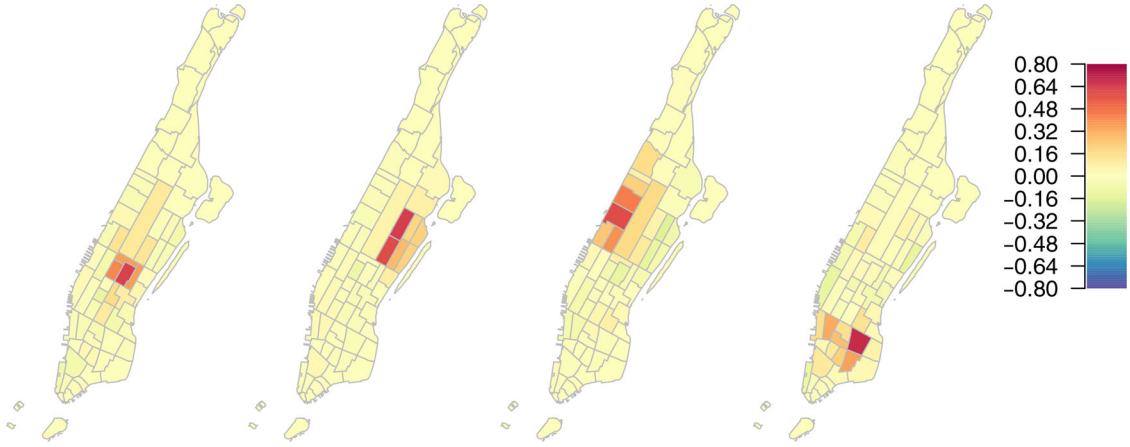

Figure 8. Loadings on four pickup factors for business day series

United States. Most of its volume come through export hubs E1 (used mainly by Canada) and E4 (used mainly by Mexico). The network plots of other product groups are shown in Appendix B.

We remark that this analysis is just for illustration and showcasing the interpretation of the model. A more formal analysis would include the determination of the number of factors and model comparison procedures. Results of a parametric bootstrap study for this example are given in Appendix C.3.

Remark 11. The above analysis is based on the original observation. It can be seen that the volume (or the scale) of the individual time series tends to dominate the factor structure. Factor and principle component analysis always depend on the choice of using covariance structure or correlation structure. In this example, if the individual time series are standardized for the analysis, the results lack economic interpretation. Some small countries and small product categories become dominating factors as their variations can be (proportionally) much larger than the large countries and large product categories, after standardization. From the view of economics, the global trade network is indeed dominated by large countries and large product categories. Our analysis shows a pattern that is intuitively true, but also revealed hidden structures that are difficult to study by looking at trading patterns individually.

7.2. Taxi Traffic in New York City

In this example we analyze taxi traffic pattern in New York city. The data include all individual taxi rides operated by Yellow Taxi within New York City, maintained by the Taxi & Limousine Commission of New York City and published at

https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

The dataset contains 1.4 billion trip records within the period of January 1, 2009 to December 31, 2017, among these 1.2 billion

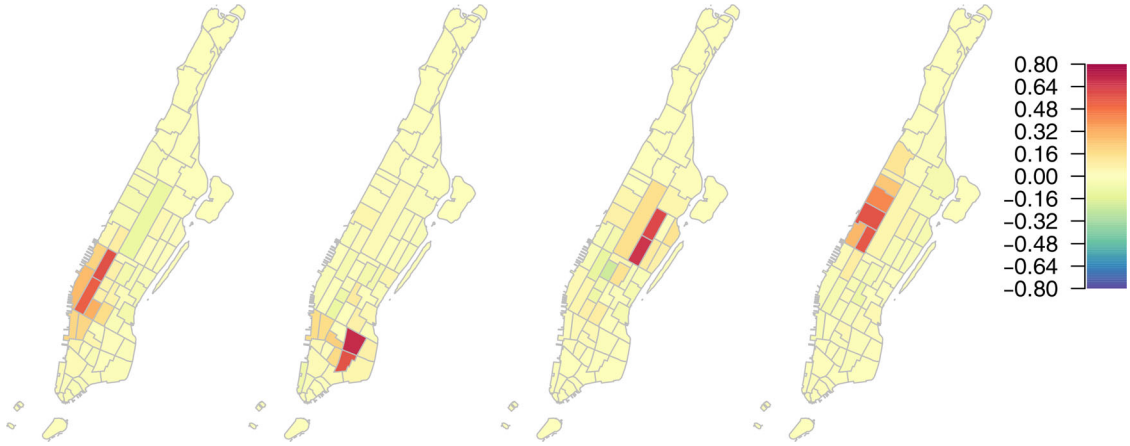

Figure 9. Loadings on four pickup factors for nonbusiness day series

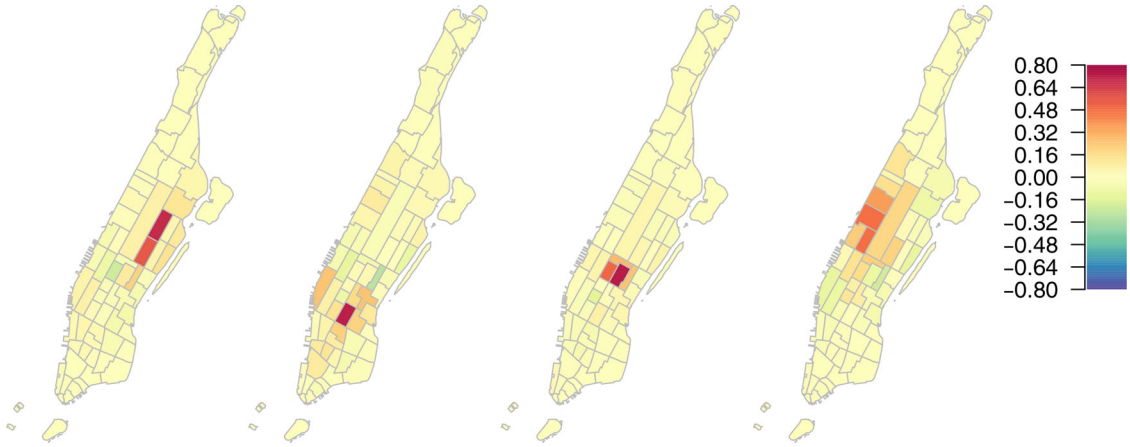

Figure 10. Loadings on four dropoff factors for business day series

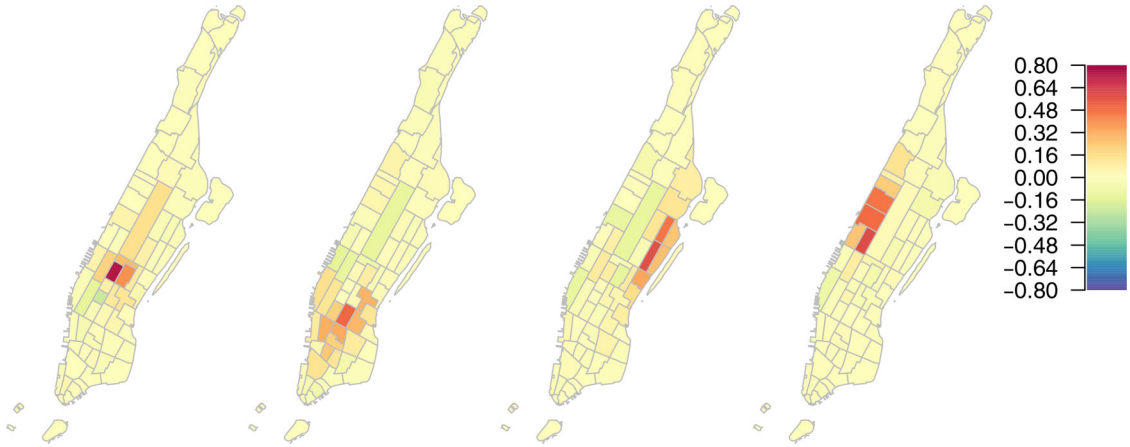

Figure 11. Loadings on four dropoff factors for nonbusiness day series

are for rides within Manhattan Island. Each trip record includes fields capturing 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. As we are interested in the movements of passengers using the taxi service, our study focuses on the pick-up and drop-off dates/times, and pick-up and drop-off locations of each ride. To simplify the discussion, we only consider rides within Manhattan Island.

The pick-up and drop-off location in Manhattan are coded according to 69 predefined zones in the dataset after 2016 and we will use them to classify the pick-up and drop-off locations. To account for time variation during the day, we divide each day into 24 hourly periods. The first hourly period is from 0 a.m. to 1 a.m. The total number of rides moving among the zones within each hour is recorded, yielding a $\mathcal{X}_t \in \mathbb{R}^{69\times 69\times 24}$ tensor for each day. Here $x_{i_1,i_2,i_3,t}$ is the number of trips from zone i_1 (the pick-up zone) to zone i_2 (the drop-off zone) and the

pickup time within the i_3-th hourly period in day t. We consider business day and non-business day separately and ignore the gaps created by the separation. Hence, we will analyze two tensor time series. The business-day series is 2,262 days long, and the non-business-day series is 1025 day long, within the period of January 1, 2009 to December 31, 2017.

After some exploratory analysis, we decide to use the tensor factor model with a $4 \times 4 \times 4$ core factor tensor and estimate the model using the TIPUP estimator with $h_0 = 1$. The TOPUP produces similar results.

Figure 8 shows the heatmap of the loading matrix $\boldsymbol{A}_1$ (related to pick-up locations) of the 69 zones in Manhattan. It is seen that during business days, the midtown/Times square area is heavily loaded on Factor 1, upper east side on Factor 2, upper west side on Factor 3 and lower east side on Factor 4. For non-business days, the loading matrix is significantly different, as shown in Figure 9. The area on the lower west side near Chelsea (with many restaurants and bars) that heavily loads on the first factor is not active for pickups during the business day.

Table 4. Label of representing areas identified under the tensor factor model with area description.

	Source factor				
	Business		Non-Bus		
Area	p	d	p	d	description
1 Upper east	2	1	3		Affluent neighborhoods and museums
2 Midtown/Times square	1	3		1	Tourism and office buildings
3 Upper west/Lincoln square	3	4	4	4	Affluent neighborhoods and performing arts
4 East village/Lower east	4		2		Historic district with art
5 Union square		2		2	Transportation hub with shops and restaurants
6 Clinton east/Chelsea			1		Lots of restaurants and bars
7 Yorkvill/Lenox hill				3	A few universities

NOTE: "p" stands for pickup and "d" for dropoff.

Figures 10 and 11 show the loading matrices $\boldsymbol{A}_2$ (related to dropoff locations) for business days and non-business days, respectively. For dropoff during business days, the areas that load heavily on the factors are quite similar to that for pick-up, except the area that loads heavily on Factor 2. This area is around Union Square which is a big transportation hub servicing the surrounding tri-state area (New York, Connecticut and New Jersey), and a heavy shopping/restaurant area. For non-business days, the dropoff area that heavily loads on Factor 3 (Yorkvill/Lenox hill) is different from all the areas used for both pickup and dropoff and for both business days and non-business days. To simplify our presentation and to show comparable results in different settings, we will roughly match the pickup and dropoff factors by their corresponding heavily loaded areas, shown in Table 4 with brief area descriptions.

Tables 5 and 6 show the loading matrix $\boldsymbol{A}_3$ (on the time of day dimension) for business day and non-business day, respectively, after varimax rotation. The shaded cells roughly show the dominating periods of each of the factors, though the change is more continuously and smooth. It is seen that, for business days, the morning rush-hours between 6 a.m. and 9 a.m. are heavy and almost exclusively loaded on factor 1 and we will name this factor the *morning rush-hour* factor. The business hours from 8am to 3 p.m. heavily load on Factor 2 (the *business hour* factor), the evening rush-hours from 3 p.m. to 8 p.m. load heavily on Factor 3 (the *evening rush-hour* factor) and the night life hours from 8 p.m. to 1 a.m. load on Factor 4 (the *night life* factor). On the other hand, for nonbusiness days, we have morning activities between 8 a.m. and 1 p.m. (the *morning* factor), afternoon/evening activities between 12 p.m. to 9 p.m. (the *afternoon/evening* factor), and night activities between 9pm to 12am (the *early night* factor) and 12 a.m. to 4 a.m. (the *late night* factor).

Figures 12 and 13 show the traffic network plots between the areas defined in Table 4 during different time factor periods. The width of the lines reflects total traffic volume between the major areas over the entire time series (the sum of the factors $f_{k_1k_2k_3,t}$ over time t.) The size of the vertices reflects total number of pickups (left vertices) and dropoffs (right vertices) in the area during the time factor period.

The figures reveal many interesting patterns. For example, during the morning rush-hours of business days, traffic mainly

Table 5. Estimated loading matrix $\boldsymbol{A}_3$ for hour of day fiber. Business day.

	0am		2		4		6		8		10		12pm		2		4		6		8		10		12am
1	−2	−1	−1	−1	1	10	47	72	42	14	2	−6	−11	−9	−8	−5	−1	5	5	4	1	2	0	−2
2	0	0	0	0	−1	−4	−13	−5	32	46	36	35	38	33	29	19	9	1	−2	−5	−3	−3	−1	1
3	−5	−4	−3	−2	−1	1	4	6	−15	−25	−6	4	7	9	19	31	32	43	47	39	22	14	4	−6
4	28	18	11	7	4	1	0	−8	2	14	4	−2	−3	−2	−7	−15	−13	−11	1	19	35	41	46	47

NOTE: Matrix is rotated via varimax. Values are in percentage.

Table 6. Estimated loading matrix $\boldsymbol{A}_3$ for hour of day fiber. Nonbusiness day.

	0am		2		4		6		8		10		12pm		2		4		6		8		10		12am
1	−20	−3	11	10	5	3	9	19	34	47	47	35	23	14	10	12	3	−4	−13	−16	−14	−13	−5	10
2	19	0	−13	−11	−3	0	0	−2	−3	−2	6	17	25	29	30	27	29	34	39	33	22	17	5	−17
3	−11	3	14	7	−2	−4	−4	−3	−2	2	−1	−5	−3	1	0	4	−3	−3	2	17	20	24	45	78
4	53	52	45	37	21	8	6	5	4	2	1	2	−2	−4	−4	−6	−2	0	0	−1	4	6	0	−10

NOTE: Matrix is rotated via varimax. Values are in percentage.

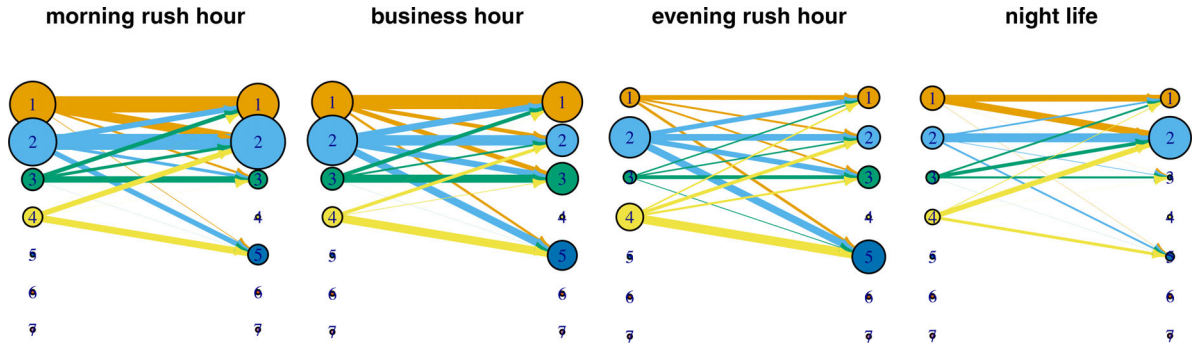

Figure 12. Network Plots during the four time factor periods for business day series

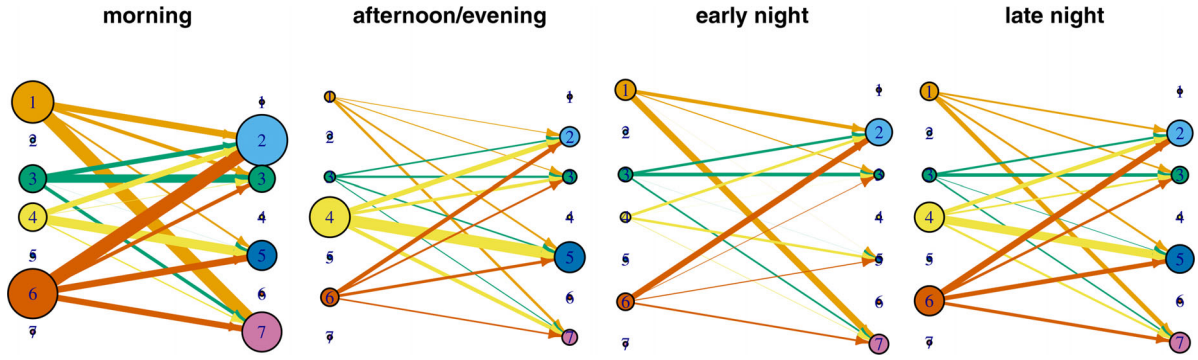

Figure 13. Network Plots during the four time factor periods for nonbusiness day series

goes from Areas 1 and 2 (upper east and midtown) to Area 2 (midtown). There is only a small amount of traffic to Area 5. During the business hours and early evening hours, traffic is mainly within Areas 1 and 2. During the evening rush-hour, the main pickup area is midtown and the main dropoff area is the Union square where many people take public transportation to the surrounding tri-state area. During the night life hours, main traffic is toward Area 2 (midtown), since Times square is popular among tourists and night-life goers.

For nonbusiness days, the pattern is very different. During morning time from 8 a.m. to 12 p.m., most traffic takes place from Area 6 (Chelsea) to Area 2 (midtown) and from Area 1 (upper east side) to Area 7 (Yorkvill/Lenox hill); during afternoon/evening from 12 p.m. to 9 p.m., many riders take taxi from Area 4 (lower east) to Area 5 (Union square); during early night (from 8 p.m. to 12 a.m.), the traffic volume is much smaller, mainly from Areas 1 (upper east) and 6 (Chelsea) to Areas 7 (Yorkvill/Lenox hill) and 2 (midtown); during late night from 12 a.m. to 5 a.m., the traffic is heavier than early night, mainly dominated by pickups from Areas 4 (lower east) and 6 (Chelsea) and dropoffs in Areas 5 (Union square) and 2 (midtown). The late night dropoff to Union square is very plausible since people need to go to transportation hub to go back home after a long night in New York city after midnight.

Again, this analysis is for demonstration of the tensor factor model only. More thorough and sophisticated analysis may be needed to fully understand the traffic pattern.

Funding

Chen's research is supported in part by National Science Foundation grants DMS-1503409, DMS-1737857, IIS-1741390, CCF-1934924 and DMS-2027855. Yang's research is supported in part by NSF grant IIS-1741390, Hong Kong GRF 17301620 and CRF C7162-20GF. Zhang's research is supported in part by NSF grants DMS-1721495, IIS-1741390 and CCF-1934924. The authors thank the editor, an associate editor, and three anonymous referees for their helpful comments.

ORCID

Rong Chen http://orcid.org/0000-0003-0793-4546

References

Aggarwal, C., and Subbian, K. (2014), "Evolutionary Network Analysis: A Survey," *ACM Computing Surveys (CSUR)*, 47, Article no. 10. [1]

Ahn, S. C., and Horenstein, A. R. (2013), "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227. [8]

Amengual, D., and Watson, M. W. (2007), "Consistent Estimation of the Number of Dynamic Factors in a Large *n* and *t* Panel," *Journal of Business & Economic Statistics*, 25, 91–96. [8]

Anandkumar, A., Ge, R., and Janzamin, M. (2014), "Guaranteed Non-orthogonal Tensor Decomposition Via Alternating Rank-1 Updates," [4]

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171. [2]

Bai, J., and Li, K. (2012), "Statistical Analysis of Factor Models of High Dimension," *Annals of Statistics*, 40, 436–465. [3]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [3,8,11]

——— (2007), "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business and Economic Statistics*, 25, 52–60. [8]

——— (2008), "Large Dimensional Factor Analysis," *Foundations and Trends in Econometrics*, 3, 89–163. [2,3]

Bai, J., and Wang, P. (2014), "Identification Theory for High Dimensional Static and Dynamic Factor Models," *Journal of Econometrics*, 178, 794–804. [5]

——— (2015), "Identification and Bayesian Estimation of Dynamic Factor Models," *Journal of Business & Economic Statistics*, 33, 221–240. [5]

Baltagi, B. (2005), *Econometric Analysis of Panel Data*. 3rd ed., Chichester: Wiley. [2]

Barak, B., and Moitra, A. (2016), "Noisy Tensor Completion Via the Sum-of-squares Hierarchy," In *Conference on Learning Theory*, PLMR, vol. 49, 417–445. [6]

Bekker, P. (1986), "A Note on the Identification of Restricted Factor Loading Matrices," *Psychometrika*, 51, 607–611. [5]

Bennett, R. (1979), *Spatial Time Series*, London: Pion. [2]

Box, G., and Jenkins, G. (1976), *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden Day. [2]

Brockwell, P., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer. [2]

Carroll, J., 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. [4]

Chamberlain, G. (1983), "Funds, Factors, and Diversification in Arbitrage Pricing Models," *Econometrica*, 51, 1305–23. [2,3]

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean–Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–304. [3]

Chang, J., Guo, B., and Yao, Q. (2018), "Principal Component Analysis for Second-order Stationary Vector Time Series," *The Annals of Statistics*, 46, 2094–2124. [3]

Chen, R., Xiao, H., and Yang, D. (2021), "Autoregressive Models for Matrix-valued Time Series," *Journal of Econometrics*, 222, 539–560. [3]

Chen, Y., and Chen, R. (2019), "Modeling Dynamic Transport Network With Matrix Factor Models: With an Application to International Trade Flow," *arxiv*: 1901.00769. [1,15]

Chen, Y., Tsay, R., and Chen, R. (2020), "Constrained Factor Models for High-dimensional Matrix-variate Time Series," *Journal of the American Statistical Association*, 115, 775–793. [1]

Connor, G., Hagmann, M., and Linton, O. (2012), "Efficient Semiparametric Estimation of the Fama-French Model and Extensions," *Econometrica*, 80, 713–754. [2,3]

Connor, G., and Linton, O. (2007), "Semiparametric Estimation of a Characteristic-based Factor Model of Stock Returns," *Journal of Empirical Finance*, 14, 694–717. [3]

Cressie, N. (1993), *Statistics for Spatial Data*. 2nd ed., New York: Wiley [2]

de Silva, V., and Lim, L. (2008), "Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem," *SIAM Journal on Matrix Analysis and Applications*, 30, 1084–1127. [6]

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, 188–205. [11]

Fan, J., Liao, Y., and Wang, W. (2016), "Projected Principal Component Analysis in Factor Models," *Annals of Statistics*, 44, 219–254. [3]

Fan, J., Wang, W., and Zhong, Y. (2019), "Robust Covariance Estimation for Approximate Factor Models," *Journal of Econometrics*, 208, 5–22. [3]

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer. [2]

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Dynamic-Factor Model: Identification And Estimation," *The Review of Economics and Statistics*, 82, 540–554. [2,3]

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, eds. Aigner, D. J., and Goldberger, A. S. Amsterdam: North-Holland. [2,3]

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010), "A Survey of Statistical Network Models," *Foundations and Trends in Machine Learning*, 2, 129–233. [1]

Hafner, C. M., Linton, O. B., and Tang, H. (2020+), "Estimation of a Multiplicative Correlation Structure in the Large Dimensional Case," *Journal Econometrics*, page in press. [5]

Hallin, M., and Liška, R. (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102, 603–617. [3,8]

Handcock, M., and Wallis, J. (1994), "An Approach to Statistical Spatial-temporal Modeling of Meteorological Fields (With Discussion)," *Journal of the American Statistical Association*, 89, 368–390. [2]

Hannan, E. (1970). *Multiple Time Series*, New York: Wiley. [2]

Hanneke, S., Fu, W., and Xing, E. P. (2010), "Discrete Temporal Models of Social Networks," *Electronic Journal of Statistics*, 4, 585–605. [1]

Härdle, W., Chen, R., and Luetkepohl, H. (1997), "A Review of Nonparametric Time Series Analysis," *International Statistical Review*, 65, 49–72. [2]

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, 84. [4]

——— (1978), "Models for Analysis of Asymmetrical Relationships Among n Objects or Stimuli," in *First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology*, vol. 5. McMaster University, Hamilton, Ontario. [15]

Hillar, C. J., and Lim, L. (2013), "Most Tensor Problems Are np-hard," *Journal of ACM*, 60, 45:1–45:39. [6]

Hoff, P. D. (2011), "Separable Covariance Arrays Via the Tucker Product, With Applications to Multivariate Relational Data," *Bayesian Analysis*, 6, 179–196. [5]

——— (2015), "Multilinear Tensor Regression for Longitudinal Relational Data," *Annals of Applied Statistics*, 9, 1169–1193. [3]

Hopkins, S. B., Shi, J., and Steurer, D. (2015), "Tensor Principal Component Analysis Via Sum-of-squares Proofs," *JMLR*, 40. [4,6]

Hsiao, C. (2003), *Analysis of Panel Data*, 2nd ed., Cambridge, UK: Cambridge University Press. [2]

Irwin, M., Cressie, N., and Johannesson, G. (2000), "Spatial–temporal Nonlinear Filtering Based on Hierarchical Statistical Models," *Test*, 11, 249–302. [2]

Ji, P., and Jin, J. (2016), "Coauthorship and Citation Networks for Statisticians," *Annals of Applied Statistics*, 10, 1779–1812. [1]

Kolaczyk, E. D., and Csárdi, G. (2014), *Statistical Analysis of Network Data With R*, vol. 65. New York: Springer. [1]

Kolda, T., and Bader, B. (2006), "The Tophits Model for Higher-order Web Link Analysis," in *Workshop on Link Analysis, Counterterrorism and Security*, vol. 7, 26–29. [15]

Kolda, T. G., and Bader, B. W. (2009), "Tensor Decompositions and Applications," *SIAM Review*, 51, 455–500. [3]

Kolda, T. G., Bader, B. W., and Kenny, J. P. (2005), "Higher-order Web Link Analysis Using Multilinear Algebra," in *Fifth IEEE International Conference on Data Mining*, 8pp. IEEE. [15]

Koltchinskii, V., Lounici, K., and Tsybakov, A. B. (2011), "Nuclear-norm Penalization and Optimal Rates for Noisy Low-rank Matrix Completion," *Annals of Statistics*, 39, 2302–2329. [6]

Lam, C., and Yao, Q. (2012), "Factor Modeling for High-dimensional Time Series: Inference for the Number of Factors," *The Annals of Statistics*, 40, 694–726. [3,4,8,11]

Lam, C., Yao, Q., and Bathia, N. (2011a), "Estimation of Latent Factors for High-dimensional Time Series," *Biometrika*, 98, 901–918. [3,10,11,12,13]

——— (2011b), "Factor Modeling for High Dimensional Time Series," in *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, ed. F. Ferraty, pp. 203–207. Physica-Verlag HD. [8]

Linnemann, H. (1966), *An Econometric Study of International Trade Flows*, vol. 234. Amsterdam: North-Holland Publishing Company. [15]

Linton, O. B., and Tang, H. (2020), "Estimation of the Kronecker Covariance Model by Quadratic Form," arxiv:1906.08908. [5]

Loh, W.-L., and Tao-Kai (2000), "Estimating Structured Correlation Matrices in Smooth Gaussian Random Field Models," *The Annals of Statistics*, 28, 880–904. [5]

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag. [2]

Mardia, K., Goodall, C., Redfern, E., and Alonso, F. (1998), "The Kriged Kalman Filter" (with discussion), *Test*, 7, 217–285. [2]

Negahban, S., and Wainwright, M. J. (2011), "Estimation of (Near) Low-rank Matrices With Noise and High-dimensional Scaling," *Annals of Statistics*, 39, 1069–1097. [6]

Neudecker, H. (1990), "On the Identification of Restricted Factor Loading Matrices: An Alternative Condition," *Journal of Mathematical Psychology*, 34, 237–241. [5]

Pan, J., and Yao, Q. (2008), "Modelling Multiple Time Series Via Common Factors," *Biometrika*, 95, 365–379. [2,3,8]

Peña, D., and Poncela, P. (2006), "Nonstationary Dynamic Factor Analysis," *Journal of Statistical Planning and Inference*, 136, 1237–1257. [3]

Peña, D., and Box, G. E. P. (1987), "Identifying a Simplifying Structure in Time Series," *Journal of the American statistical Association*, 82, 836–843. [2,3]

Phan, T. Q., and Airoldi, E. M. (2015), "A Natural Experiment of Social Network Formation and Dynamics," *Proceedings of the National Academy of Sciences*, 112, 6595–6600. [1]

Richard, E., and Montanari, A. (2014), "A Statistical Model for Tensor PCA," in *Advances in Neural Information Processing Systems*, pp. 2897–2905. [4,6]

Sargent, T. J., and Sims, C. A. (1977), "Business Cycle Modeling Without Pretending to Have Too Much a Priori Economic Theory," *New Methods in Business Cycle Research*, 1, 145–168. [2,3]

Shumway, R., and Stoffer, D. (2002), *Time Series Analysis and Its Applications*. New York, NY: Springer. [2]

Snijders, T. A. (2006), "Statistical Methods for Network Dynamics," in *Proceedings of the XLIII Scientific Meeting, Italian Statistical Society*, number 1994, 281–296. Padova: CLEUP, Italy. [1]

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer. [2]

Stock, J. H., and Watson, M. W. (2006), "Forecasting With Many Predictors," in *Handbook of Economic Forecasting*, eds. Elliott, G., Granger, C., and Timmermann, A. pp. 515–554. Amsterdam: Elsevier. [3]

Stock, J. H., and Watson, M. W. (2012), "Dynamic Factor Models," in *The Oxford Handbook of Economic Forecasting*, eds. Clements, M. P., and Hendry, D. F. Oxford: Oxford University Press. [2,3]

Stroud, J. R., Muller, P., and Sanso, B. (2001), "Dynamic Models for Spatiotemporal Data," *Journal of the Royal Statistical Society*, Series B, 63, 673–689. [2]

Sun, W. W., Lu, J., Liu, H., and Cheng, G. (2016), "Provable Sparse Tensor Decomposition," *Journal of the Royal Statistical Society*, Series B, [4]

Tong, H. (1990), *Nonlinear Time Series Analysis: A Dynamical System Approach*, London: Oxford University Press. [2]

Tsay, R. (2005), *Analysis of Financial Time Series*. New York: Wiley. [2]

Tsay, R., and Chen, R. (2018), *Nonlinear Time Series Analysis*, Hoboken, NJ: Wiley. [2]

Tucker, L. R. (1963), "Implications of Factor Analysis of Three-way Matrices for Measurement of Change," in *Problems in Measuring Change*, eds. Chester W. Harris, pp. 122–137. Madison: University of Wisconsin Press. [4]

——— (1964), "The Extension of Factor Analysis to Three-dimensional Matrices," in *Contributions to Mathematical Psychology*, pp. 110–127. New York: Holt, Rinehart and Winston. [4]

——— (1966), "Some Mathematical Notes on Three-mode Factor Analysis," *Psychometrika*, 31, 279–311. [4]

Wang, D., Liu, X., and Chen, R. (2019), "Factor Models for Matrix-valued High-dimensional Time Series," *Journal of Econometrics*, 208, 231–248. [1,3,4,7,10,11,12,13,15]

Wikle, C., Berliner, L., and Cressie, N. (1998), "Hierarchical Bayesian Space-time Models," *Environmental and Ecological Statistics*, 5, 117–154. [2]

Wikle, C., and Cressie, N. (1999), "A Dimension-reduced Approach to Space-time Kalman Filtering," *Biometrika*, 86, 815–829. [2]

Woolrich, M., Jenkinson, M., Brady, J., and Smith, S. (2004), "Fully Bayesian Spatio-temporal Modeling of fMRI Data," *IEEE Transactions on Medical Imaging*, 23, 213–231. [2]

Xia, D., and Yuan, M. (2017), "On Polynomial Time Methods for Exact Low-rank Tensor Completion," *Foundations of Computational Mathematics*, 19(6), 1265–1313. [6,8]

Yuan, M., and Zhang, C.-H. (2016), "On Tensor Completion Via Nuclear Norm Minimization," *Foundations of Computational Mathematics*, 16, 1031–1068. [6]

——— (2017), "Incoherent Tensor Norms and Their Applications in Higher Order Tensor Completion," *IEEE Transactions on Information Theory*, 63, 6753–6766. [6]

Zhang, A. (2019), "Cross: Efficient Low-rank Tensor Completion," *The Annals of Statistics*, 47, 936–964. [6]

Zhang, A., and Xia, D. (2018), "Tensor svd: Statistical and Computational Limits," *IEEE Transactions on Information Theory*, 64, 7311–7338. [8]

Zhao, Y., Levina, E., and Zhu, J. (2012), "Consistency of Community Detection in Networks Under Degree-corrected Stochastic Block Models," *Annals of Applied Statistics*, 40, 2266–2290. [1]