Factor Models for Matrix-Valued High-Dimensional Time Series *

Dong Wang
Princeton University
Xialu Liu
San Diego State University
Rong Chen
Rutgers University

October, 2016

Abstract

In finance, economics and many other fields, observations in a matrix form are often observed over time. For example, many economic indicators are obtained in different countries over time. Various financial characteristics of many companies are reported over time. Although it is natural to turn a matrix observation into a long vector then use standard vector time series models or factor analysis, it is often the case that the columns and rows of a matrix represent different sets of information that are closely interrelated in a very structural way. We propose a novel factor model that maintains and utilizes the matrix structure to achieve greater dimensional reduction as well as finding clearer and more interpretable factor structures. Estimation procedure and its theoretical properties are investigated and demonstrated with simulated and real examples.

1 Introduction

Time series analysis is widely used in many applications. Univariate time series, when one observes one variable through time, is well studied, with linear models (e.g. Box and Jenkins, 1976; Brockwell and Davis, 1991; Tsay, 2005), nonlinear models (e.g. Tong, 1990; Engle, 1982; Bollerslev, 1986) and nonparametric models (e.g. Fan and Yao, 2003). Multivariate time series and panel time, when one observes a vector or a panel of variables through time, is also a long studied but still active field (e.g. Tiao and Box, 1981; Tiao and Tsay, 1989; Engle and Kroner, 1995; Lütkepohl, 2005; Tsay, 2014; Stock and Watson, 2004, and others). Such analysis not only reveals the temporal dynamics of the time series, but also explores the relationship among a group of

*Corresponding author: Rong Chen, Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA. Email: rongchen@stat.rutgers.edu.

	US	Japan	$\cdots$	China
GDP	$X_{t,11}$	$X_{t,12}$	$\cdots$	$X_{t,1p}$
Unemployment	$X_{t,21}$	$X_{t,22}$	$\cdots$	$X_{t,2p}$
Inflation	$X_{t,31}$	$X_{t,32}$	$\cdots$	$X_{t,3p}$
Industry Production	$X_{t,41}$	$X_{t,42}$	$\cdots$	$X_{t,4p}$

Table 1: Illustration of a matrix-valued time series

time series, using the available information more fully. Often, the investigation of the relationship among the time series is the objective of the study.

Matrix-valued time series, when one observes a group of variables structured in a well defined matrix form over time, has not been studied. Such a time series is encountered in many applications. For example, in economics, a group of countries will report a set of economic indicators (e.g. GDP growth, unemployment rate, inflation index and others) every quarter. Table 1 depicts such a matrix-valued time series. One can concentrate on one cell in Table 1, say US Unemployment rate series $\{X_{t,21},\ t = 1, 2 \ldots\}$ and build a univariate time series. Or one can concentrate on one column in Table 1, say, all economic indicators of US $\{(X_{t,11}, \ldots, X_{t,41})'\}$ and study it as a vector time series. Similarly, if one is interested in modeling GDP growth of the group of countries, a panel time series model can be built for the first row $\{(X_{t,11}, \ldots, X_{t,1p})\}$ in Table 1. However, there are certainly relationships among all variables in the table and the matrix structure is extremely important. For example, the variables in the same column (same country) would have stronger inter-relationship. Same for the variables in the same row (same indicator). Hence it is important to analyze the whole group of variables while fully preserve and utilize its matrix structure.

There are many other examples. Investors may be interested in a group of financials (e.g. asset/equity ratio, dividend per share, and revenue) for a group of companies. Other examples include import-export volume among a group of countries, pollution and environmental variables (e.g. PM2.5, ozone level, temperature, moisture, wind speed, etc) observed at a group of stations. In this article we will study such a matrix valued time series.

Matrix-valued data analysis has been studied (e.g. Gupta and Nagar, 2000; Kollo and von Rosen, 2006; Werner et al., 2008; Leng and Tang, 2012; Yin and Li, 2012; Zhao and Leng, 2014; Zhou, 2014; Zhou and Li, 2014). Their study mainly focuses on independent observations. The concept of matrix-valued time series was introduced by Walden and Serroukh (2002), applied in signal and image processing. Still, the temporal dependence of the time series was not fully exploited for model building and model validation.

In this article, we focus on high-dimensional matrix-valued time series data. In cases, we may allow the dimensions of the matrix to be as large as, or even larger than the length of the observations. A well-known issue often accompanying with high-dimensional data is the curse of dimensionality. We adopt a factor model approach. Factor analysis can effectively reduce the number of parameters involved, and is a powerful statistical approach to extracting hidden driving processes, or latent factor processes, from an observed stochastic process. In the past decades, factor models for high-dimensional time series data have drawn great attention from both econometricians and statisticians (e.g. Chamberlain and Rothschild, 1983; Forni et al., 2000; Bai and Ng, 2002; Hallin and Liška, 2007; Pan and Yao, 2008; Lam et al., 2011; Fan et al., 2011; Lam and Yao, 2012; Fan et al., 2013; Chang et al., 2015; Liu and Chen, 2016).

With the above observations and motivations, in this article, we are aiming to develop factor models for matrix-valued time series, which fully explore the matrix structure. The rest of this article is organized as follows. In Section 2, detailed model settings are introduced and interpretations are discussed in detail. Section 3 presents an estimation procedure. The theoretical properties of the estimators are also studied. Simulation results are shown in Section 4 and a real data example is given in Section 5. All proofs are in Appendix.

2 Matrix Factor Models

Let $\boldsymbol{X}_t$ $(t = 1, \ldots, T)$ be a matrix-valued time series, where each $\boldsymbol{X}_t$ is a matrix of size $p_1 \times p_2$,

$$\boldsymbol{X}_t = \begin{pmatrix} X_{t,11} & \cdots & X_{t,1p_2} \\ \vdots & \ddots & \vdots \\ X_{t,p_1 1} & \cdots & X_{t,p_1 p_2} \end{pmatrix}.$$

We propose the following factor model for matrix-valued time series,

$$\boldsymbol{X}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}' + \boldsymbol{E}_t, \quad t = 1, 2, \ldots, T. \tag{1}$$

Here, $\boldsymbol{F}_t$ is a $k_1 \times k_2$ unobserved matrix-valued time series of common **fundamental factors**, $\boldsymbol{R}$ is a $p_1 \times k_1$ front loading matrix, $\boldsymbol{C}$ is a $p_2 \times k_2$ back loading matrix, and $\boldsymbol{E}_t$ is a $p_1 \times p_2$ error matrix. In model (1), the common fundamental factors $\boldsymbol{F}_t$ drive all dynamics and co-movement of $\boldsymbol{X}_t$. $\boldsymbol{R}$ and $\boldsymbol{C}$ reflect the importance of common factors and their interactions.

Similar to multivariate factor models, we assume that the matrix-valued time series is driven by a few latent factors. Unlike the classical factor model, the factors $\boldsymbol{F}_t$'s in Model (1) are assumed to be organized in matrix form. Correspondingly, we adopt two loading matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ to capture the dependency between each individual time series in the matrix observations and the matrix factors. In the following we provide two interpretations of the loading matrices. Here, we first introduce some notations. For a matrix $\boldsymbol{A}$, we use $\boldsymbol{A}_{i\cdot}$ and $\boldsymbol{A}_{\cdot j}$ to represent the i-th row and the j-th column of $\boldsymbol{A}$ respectively, and A_{ij} to denote the ij-th element of $\boldsymbol{A}$.

Interpretation I: To isolate effects, assume $k_1 = p_1$ and $\boldsymbol{R} = \boldsymbol{I}_{p_1}$, then $\boldsymbol{X}_t = \boldsymbol{F}_t\boldsymbol{C}' + \boldsymbol{E}_t$. In this case, each column of $\boldsymbol{X}_t$ is a linear combination of the columns of $\boldsymbol{F}_t$. Take the example of Table 1 and consider the first column of $\boldsymbol{X}_t$ (the US economic indicators),

$$
\begin{array}{ccccc}
\text{US} & & \boldsymbol{F}_{t,\cdot 1} & \dots & \boldsymbol{F}_{t,\cdot k_2} \\
\begin{pmatrix} \text{GDP} \\ \text{Unem} \\ \text{Inf} \\ \text{PayR} \end{pmatrix}_t & = C_{11} & \begin{pmatrix} \text{F-GDP} \\ \text{F-Unem} \\ \text{F-Inf} \\ \text{F-PayR} \end{pmatrix}_t & + \ldots + C_{1k_2} & \begin{pmatrix} \text{F-GDP} \\ \text{F-Unem} \\ \text{F-Inf} \\ \text{F-PayR} \end{pmatrix}_t + \boldsymbol{E}_{t,\cdot US}.
\end{array}
$$

It is seen that the US GDP only depends on the first row of $\boldsymbol{F}_t$. Similarly, other countries' GDP also only depends on the first row of $\boldsymbol{F}_t$. Hence we can view the first row of $\boldsymbol{F}_t$ as the GDP factors. Similarly, the second row of $\boldsymbol{F}_t$ can be considered as the unemployment factors. There is no interaction between the indicators in this setting (when $\boldsymbol{R} = \boldsymbol{I}$). The loading matrix $\boldsymbol{C}$ reflects how each country (column of $\boldsymbol{X}_t$) depends on the columns of $\boldsymbol{F}_t$, hence reflects column interactions, or the interactions between the countries. Because of this, we will call $\boldsymbol{C}$ the column loading matrix.

Similarly, the rows of $\boldsymbol{F}_t$ can be viewed as common factors of all rows of $\boldsymbol{X}_t$, and the front loading matrix $\boldsymbol{R}$ as row loading matrix. Again, assume $k_2 = p_2$ and $\boldsymbol{C} = \boldsymbol{I}_{p_2}$, it follows that $\boldsymbol{X}_t = \boldsymbol{R}\boldsymbol{F}_t + \boldsymbol{E}_t$. Then each row of $\boldsymbol{X}_t$ is a linear combination of the rows of $\boldsymbol{F}_t$. Consider the first row of $\boldsymbol{X}_t$,

$$
\begin{array}{rll}
\text{US} \quad \text{Japan} \ \ldots \ \text{China} & \quad \text{US} \qquad \text{Japan} \quad \ldots \quad \text{China} & \\
(\text{GDP}, \text{GDP}, \ldots, \text{GDP})_t = & R_{11}(\text{F-US}, \text{F-Japan}, \ldots, \text{F-China})_t & \boldsymbol{F}_{t,1\cdot} \\
& +R_{12}(\text{F-US}, \text{F-Japan}, \ldots, \text{F-China})_t & \boldsymbol{F}_{t,2\cdot} \\
& + \ldots & \vdots \\
& +R_{1k_1}(\text{F-US}, \text{F-Japan}, \ldots, \text{F-China})_t & \boldsymbol{F}_{t,k_1\cdot} \\
& +\boldsymbol{E}_{t,GDP\cdot}. &
\end{array}
$$

It is seen that all economic movements (of each country) are driven by k_1 (row) common factors. For example, every US's indicator depends on only the first column of $\boldsymbol{F}_t$. Hence the first column of $\boldsymbol{F}_t$ can be viewed as the US factor. And the second column of $\boldsymbol{F}_t$ can be viewed as Japan factor. The loading matrix $\boldsymbol{R}$ reflects how each indicator depends on the rows of $\boldsymbol{F}_t$. It reflects row interactions, the interactions between the indicators within each country. Because of this, we will call $\boldsymbol{R}$ the row loading matrix.

Interpretation II: We can view the model as a two-step hierarchical model.
Step 1: For each fixed $i = 1, 2, \ldots, p_1$, using data $\{\boldsymbol{X}_{t,i\cdot},\ t = 1, 2, \ldots, T\}$, we can find a $p_2 \times k_2$ dimensional loading matrix $\boldsymbol{C}_i$ and k_2 dimensional factors $\{\boldsymbol{G}_{t,i\cdot},\ t = 1, 2, \ldots, T\}$ through factor model. That is,

$$
[X_{t,i1}, \ldots, X_{t,ip_2}] = [G_{t,i1}, \ldots, G_{t,ik_2}]\boldsymbol{C}_i' + [Z_{t,i1}, \ldots, Z_{t,ip_2}], \quad t = 1, 2, \ldots, T.
$$

Let the i-th row of the $p_1 \times k_2$ matrix $\boldsymbol{G}_t$ and the $p_1 \times p_2$ matrix $\boldsymbol{Z}_t$ be the factor $\boldsymbol{G}_{t,i\cdot}$ and the error term $\boldsymbol{Z}_{t,i\cdot}$ respectively.

Step 2: For each $j = 1, 2, \ldots, k_2$, conduct a factor analysis on $\{\boldsymbol{G}_{t,\cdot j},\ t = 1, 2, \ldots, T\}$, and obtain a $p_1 \times k_1$ loading matrix $\boldsymbol{R}_j$ and k_1 dimensional factors $\boldsymbol{F}_{t,\cdot j}$. That is,

$$\begin{bmatrix} G_{t,1j} \\ \vdots \\ G_{t,p_1 j} \end{bmatrix} = \boldsymbol{R}_j \begin{bmatrix} F_{t,1j} \\ \vdots \\ F_{t,k_1 j} \end{bmatrix} + \begin{bmatrix} Z^*_{t,1j} \\ \vdots \\ Z^*_{t,p_1 j} \end{bmatrix}, \quad t = 1, 2, \ldots, T.$$

This step tries to find common factors that drives the co-moments in $\boldsymbol{G}_t$. Now let the j-th column of the $k_1 \times k_2$ matrix $\boldsymbol{F}_t$ and the $p_1 \times k_2$ matrix $\boldsymbol{Z}^*_t$ be the factor $\boldsymbol{F}_{t,\cdot j}$ and the error term $\boldsymbol{Z}^*_{t,\cdot j}$ respectively.

Step 3: Assembly: With the above two-step factor analysis and notations, assume $\boldsymbol{R}_1 = \ldots = \boldsymbol{R}_{k_2} = \boldsymbol{R}$ and $\boldsymbol{C}_1 = \ldots = \boldsymbol{C}_{p_1} = \boldsymbol{C}$, we have

$$\boldsymbol{X}_t = \boldsymbol{G}_t \boldsymbol{C}^{'} + \boldsymbol{Z}_t \quad \text{and} \quad \boldsymbol{G}_t = \boldsymbol{R}\boldsymbol{F}_t + \boldsymbol{Z}^*_t.$$

Hence

$$\boldsymbol{X}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}^{'} + \boldsymbol{Z}^*_t\boldsymbol{C}^{'} + \boldsymbol{Z}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}^{'} + \boldsymbol{E}_t,$$

where $\boldsymbol{E}_t = \boldsymbol{Z}^*_t\boldsymbol{C}^{'} + \boldsymbol{Z}_t$. It is identical to (1).

Here we provide some additional remarks of the Model (1).

Remark 1: Let vec$(\cdot)$ be the vectorization operator, i.e., vec$(\cdot)$ converts a matrix to a vector by stacking columns of the matrix on top of each other. The classical factor analysis treats vec$(\boldsymbol{X}_t)$ as the observations, and a factor model is in the form of

$$\text{vec}(\boldsymbol{X}_t) = \boldsymbol{\Phi}\mathbf{f}_t + \mathbf{e}_t, \quad t = 1, 2, \ldots, T, \tag{2}$$

where $\boldsymbol{\Phi}$ is a $p_1 p_2 \times k$ loading matrix, $\mathbf{f}_t$ of length k is the latent factor, $\mathbf{e}_t$ is the error term, and k is the total number of factors. On the other hand, note that Model (1) can be re-written as

$$\text{vec}(\boldsymbol{X}_t) = (\boldsymbol{C} \otimes \boldsymbol{R})\text{vec}(\boldsymbol{F}_t) + \text{vec}(\boldsymbol{E}_t). \tag{3}$$

Assume $k = k_1 k_2$. Then Equation (3) is equivalent to Equation (2) with a Kronecker product structured loading matrix. Hence Model (1) is a restricted version of Model (2), with a special structure for loading spaces. The number of parameters for the loading matrix $\boldsymbol{\Phi}$ in Model (2) is $(p_1 k_1) \times (p_2 k_2)$ whereas it is $p_1 k_1 + p_2 k_2$ for loading matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ in Model (1). Therefore, Model (1) significantly reduces the dimension of the problem.

Remark 2: Similar models as Model (1) have been proposed and studied when conducting principal component analysis on matrix-valued data (e.g. Paatero and Tapper, 1994; Yang et al., 2004; Ye, 2005; Ding and Ye, 2005; Zhang and Zhou, 2005; Crainiceanu et al., 2011; Wang et al., 2016). In those studies, the matrix-valued observations $\boldsymbol{X}_t$ are assumed to be independent and they primarily focused on principal component analysis. To the best of our knowledge, our paper is the first one considering factor models for matrix-valued time series data.

In this article, we extend the methods described in Lam et al. (2011) and Lam and Yao (2012) for vector-valued factor model (2) to matrix-valued factor model (1). We propose estimators for the loading spaces and the numbers of row and column factors, investigate their theoretical properties, establish their converge rates, and construct a test for the Kronecker product structure of the loading matrix. Simulated and real examples are presented to illustrate the performance of the proposed estimators, to compare the asymptotics under different conditions with different factor strengths, and to explore the interactions between row and column factors.

3 Estimation and Modeling Procedures

Because of the latent nature of the factors, various assumptions are imposed to 'define' a factor. Two common assumptions are used. One assumes that factors must have impact on most of the series, and weak serial dependence is allowed for the idiosyncratic noise process, see Chamberlain and Rothschild (1983); Forni et al. (2000); Bai and Ng (2002); Hallin and Liška (2007), among others. Another assumes that the factors should capture all dynamics of the observed process, hence the idiosyncratic noise process has no serial dependence (but may have strong cross-sectional dependence), see Pan and Yao (2008); Lam et al. (2011); Lam and Yao (2012); Chang et al. (2015); Liu and Chen (2016). Here we adopt the second assumption and assume that the vectorized error $\text{vec}(\boldsymbol{E}_t)$ is a white noise process with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_e$, and is independent of the factor process $\text{vec}(\boldsymbol{F}_t)$. For ease of presentation, we will assume that the process $\boldsymbol{F}_t$ has mean $\mathbf{0}$, and the observations $\boldsymbol{X}_t$'s are centered and standardized through out this paper.

For the vector-valued factor model (2), it is well-known that there exists an identifiable issue among the factors $\mathbf{f}_t$ and the loading matrix $\boldsymbol{\Phi}$. Similar problem also arises in the proposed matrix-valued factor model (1). Let $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ be two invertible matrices of sizes $p_1 \times p_1$ and $p_2 \times p_2$. Then the triples $(\boldsymbol{R}, \boldsymbol{F}_t, \boldsymbol{C})$ and $(\boldsymbol{R}\boldsymbol{U}_1, \boldsymbol{U}_1^{-1}\boldsymbol{F}_t\boldsymbol{U}_2^{-1}, \boldsymbol{C}\boldsymbol{U}_2')$ are equivalent under Model (1), and hence Model (1) is not identifiable. However, with a similar argument as in Lam et al. (2011) and Lam and Yao (2012), the column spaces of the loading matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ are uniquely determined. Hence, in the following, we will focus on the estimation of the column spaces of $\boldsymbol{R}$ and $\boldsymbol{C}$, denoted by $\mathcal{M}(\boldsymbol{R})$ and $\mathcal{M}(\boldsymbol{C})$, called row factor loading space and column factor loading space respectively.

By the QR decomposition, we can further decompose $\boldsymbol{R}$ and $\boldsymbol{C}$ as follows,

$$\boldsymbol{R} = \boldsymbol{Q}_1\boldsymbol{W}_1, \text{ and } \boldsymbol{C} = \boldsymbol{Q}_2\boldsymbol{W}_2,$$

where $\boldsymbol{Q}_i$ is a $p_i \times k_i$ matrix with orthonormal columns and $\boldsymbol{W}_i$ is a $k_i \times k_i$ upper triangular matrix, for $i = 1, 2$. Let $\mathcal{M}(\boldsymbol{Q}_i)$ denote the column space of $\boldsymbol{Q}_i$. Then we have $\mathcal{M}(\boldsymbol{Q}_1) = \mathcal{M}(\boldsymbol{R})$ and $\mathcal{M}(\boldsymbol{Q}_2) = \mathcal{M}(\boldsymbol{C})$. Hence, the estimation of column spaces of $\boldsymbol{R}$ and $\boldsymbol{C}$ are equivalent to the estimation of column spaces of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$.

Write

$$\boldsymbol{Z}_t = \boldsymbol{W}_1\boldsymbol{F}_t\boldsymbol{W}_2', \quad t = 1, 2, \ldots, T, \tag{4}$$

as a transformed latent factor process. Then, the Model (1) can be re-expressed as

$$\boldsymbol{X}_t = \boldsymbol{Q}_1 \boldsymbol{Z}_t \boldsymbol{Q}_2' + \boldsymbol{E}_t, \quad t = 1, 2, \ldots, T. \tag{5}$$

Equation (5) can be viewed as another formulation of the matrix-valued factor model with orthonormal loading matrices. Since $\mathcal{M}(\boldsymbol{R}) = \mathcal{M}(\boldsymbol{Q}_1)$ and $\mathcal{M}(\boldsymbol{C}) = \mathcal{M}(\boldsymbol{Q}_2)$, we will perform analysis on Model (1) and (5) interchangeably whenever one is more convenient than the other.

3.1 Estimation

To estimate the matrix-valued factor model (1), we follow closely the idea of Lam et al. (2011) and Lam and Yao (2012) in estimating vector-valued factor models. The key idea is to calculate auto-cross-covariances of the time series then construct a Box-Ljung type of statistics in matrix. Under the matrix factor model and white idiosyncratic noises assumption, the space spanned by such a matrix is directly linked with the loading matrices. In what follows, we will illustrate the method to obtain an estimate of $\mathcal{M}(\boldsymbol{R})$. The column space of $\boldsymbol{C}$ can be obtained in a similar way on the transposes of $\boldsymbol{X}_t$'s.

Let the j-th columns of $\boldsymbol{X}_t$, $\boldsymbol{R}$, $\boldsymbol{C}$, $\boldsymbol{Q}_i$ and $\boldsymbol{E}_t$ be $\boldsymbol{x}_{t,j}$, $\boldsymbol{R}_j$, $\boldsymbol{C}_i$, $\boldsymbol{q}_{i,j}$ and $\boldsymbol{\epsilon}_{t,j}$ respectively. Let $\boldsymbol{R}_{k\cdot}'$, $\boldsymbol{C}_{k\cdot}'$ and $\boldsymbol{Q}_{i,k\cdot}'$ be the row vectors that denote the k-th row of $\boldsymbol{R}$, $\boldsymbol{C}$ and $\boldsymbol{Q}_i$ respectively. Then it follows from (1) and (5) that

$$\boldsymbol{x}_{t,j} = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}_{j\cdot} + \boldsymbol{\epsilon}_{t,j} = \boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_{2,j\cdot} + \boldsymbol{\epsilon}_{t,j}, \quad j = 1, 2, \ldots, p_2. \tag{6}$$

From the zero mean assumptions of both $\boldsymbol{F}_t$ and $\boldsymbol{E}_t$, we have $\mathrm{E}(\boldsymbol{x}_{t,j}) = \boldsymbol{0}$.

Let h be a positive integer. Define

$$\boldsymbol{\Omega}_{zq,ij}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} \mathrm{Cov}(\boldsymbol{Z}_t\boldsymbol{Q}_{2,i\cdot}, \boldsymbol{Z}_{t+h}\boldsymbol{Q}_{2,j\cdot}), \tag{7}$$

$$\boldsymbol{\Omega}_{x,ij}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} \mathrm{Cov}(\boldsymbol{x}_{t,i}, \boldsymbol{x}_{t+h,j}), \tag{8}$$

for $i, j = 1, 2, \ldots, p_2$. By plugging (6) into (8), it follows that

$$\boldsymbol{\Omega}_{x,ij}(h) = \boldsymbol{Q}_1\boldsymbol{\Omega}_{zq,ij}\boldsymbol{Q}_1'. \tag{9}$$

For a pre-determined integer h_0, define

$$\boldsymbol{M} = \sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2} \boldsymbol{\Omega}_{x,ij}(h)\boldsymbol{\Omega}_{x,ij}'(h). \tag{10}$$

By Equations (9) and (10), it follows that

$$\boldsymbol{M} = \boldsymbol{Q}_1\left(\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2} \boldsymbol{\Omega}_{zq,ij}(h)\boldsymbol{\Omega}_{zq,ij}'(h)\right)\boldsymbol{Q}_1'. \tag{11}$$

Suppose Condition 5 in Section 3.2 holds, i.e., the matrix $\boldsymbol{M}$ has rank k_1. From (11), we can see that each column of $\boldsymbol{M}$ is a linear combination of columns of $\boldsymbol{Q}_1$, and thus the matrices $\boldsymbol{M}$ and $\boldsymbol{Q}_1$ have the same column spaces, that is, $\mathcal{M}(\boldsymbol{M}) = \mathcal{M}(\boldsymbol{Q}_1)$. It follows that the eigen-space of $\boldsymbol{M}$ is the same as $\mathcal{M}(\boldsymbol{Q}_1)$. Hence, $\mathcal{M}(\boldsymbol{Q}_1)$ can be estimated by the space spanned by the eigenvectors of the sample version of $\boldsymbol{M}$.

Now we construct the sample versions of these quantities and introduce the estimation procedure as follows. For any positive integer h and a pre-scribed positive integer h_0, let

$$\widehat{\boldsymbol{\Omega}}_{x,ij}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{x}_{t,i}\boldsymbol{x}'_{t+h,j}, \tag{12}$$

$$\widehat{\boldsymbol{M}} = \sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2} \widehat{\boldsymbol{\Omega}}_{x,ij}(h)\widehat{\boldsymbol{\Omega}}'_{x,ij}(h). \tag{13}$$

Then, $\mathcal{M}(\boldsymbol{Q}_1)$ can be estimated by $\mathcal{M}(\widehat{\boldsymbol{Q}}_1)$, where $\widehat{\boldsymbol{Q}}_1 = \{\widehat{\boldsymbol{q}}_1, \ldots, \widehat{\boldsymbol{q}}_{k_1}\}$, and $\widehat{\boldsymbol{q}}_1, \ldots \widehat{\boldsymbol{q}}_{k_1}$ are the eigenvectors of $\widehat{\boldsymbol{M}}$ corresponding to its k_1 largest eigenvalues.

In practice, the number of row factors k_1 is usually unknown. This quantity can be estimated through a similar eigenvalue ratio estimator as described in Lam and Yao (2012). Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_{p_1} \geq 0$ be the ordered eigenvalues of $\widehat{\boldsymbol{M}}$. Then

$$\hat{k}_1 = \arg\min_{1\leq i\leq p_1/2} \frac{\hat{\lambda}_{i+1}}{\hat{\lambda}_i}.$$

For $\boldsymbol{Q}_2$ and k_2, they can be found by performing the same procedure on the transposes of $\boldsymbol{X}_t$'s. Once $\widehat{\boldsymbol{Q}}_1$ and $\widehat{\boldsymbol{Q}}_2$ are obtained, the estimate of $\boldsymbol{Z}_t$ can be found via a general linear regression analysis, since

$$\text{vec}(\boldsymbol{X}_t) = (\boldsymbol{Q}_2 \otimes \boldsymbol{Q}_1)\text{vec}(\boldsymbol{Z}_t) + \text{vec}(\boldsymbol{E}_t).$$

Together with the orthonormal properties of both $\widehat{\boldsymbol{Q}}_1$ and $\widehat{\boldsymbol{Q}}_2$ and properties of Kronecker product, it follows that

$$\widehat{\boldsymbol{Z}}_t = \widehat{\boldsymbol{Q}}'_1\boldsymbol{X}_t\widehat{\boldsymbol{Q}}_2.$$

3.2 Theoretical properties of the estimator

In this section, we study the asymptotic properties of the estimators under the setting that all T, p_1 and p_2 grow to infinity while k_1 and k_2 are being fixed. In the following, for any matrix $\boldsymbol{Y}$, we use $\|\boldsymbol{Y}\|_2$, $\|\boldsymbol{Y}\|_F$, and $\|\boldsymbol{Y}\|_{\min}$ to denote the spectral norm, the Frobenius norm, and the smallest nonzero singular value of $\boldsymbol{Y}$. When $\boldsymbol{Y}$ is a square matrix, we denote by $\text{tr}(\boldsymbol{Y})$, $\lambda_{\max}(\boldsymbol{Y})$ and $\lambda_{\min}(\boldsymbol{Y})$ the trace, maximum and minimum eigenvalues of $\boldsymbol{Y}$ respectively. We write $a \asymp b$ when $a = O(b)$ and $b = O(a)$. Define

$$\boldsymbol{\Sigma}_f(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} \text{Cov}\left(\text{vec}(\boldsymbol{F}_t), \text{vec}(\boldsymbol{F}_{t+h})\right), \quad \text{and} \quad \boldsymbol{\Sigma}_e = \text{Cov}(\text{vec}(\boldsymbol{E}_t)).$$

The following regularity conditions are imposed before we derive the asymptotics of our estimators.

Condition 1. The vector-valued process $\mathrm{vec}(\boldsymbol{F}_t)$ is α-mixing. For some $\gamma > 2$, the mixing coefficients satisfy the condition $\sum_{h=1}^{\infty} \alpha(h)^{1-2/\gamma} < \infty$, where

$$\alpha(h) = \sup_{i} \sup_{A \in \mathcal{F}_{-\infty}^{i}, B \in \mathcal{F}_{i+h}^{\infty}} |P(A \cap B) - P(A)P(B)|,$$

and $\mathcal{F}_i^j$ is the σ-field generated by $\{\mathrm{vec}(\boldsymbol{F}_t) : i \le t \le j\}$.

Condition 2. Let $F_{t,ij}$ be the ij-th entry of $\boldsymbol{F}_t$ for any $i = 1, \ldots, k_1$, $j = 1, \ldots, k_2$, and $t = 1, \ldots, T$. Then, $\mathrm{E}(|F_{t,ij}|^{2\gamma}) \le C$, where C is a positive constant, and γ is given in Condition 1. There exists an $1 \le h \le h_0$ such that $\boldsymbol{\Sigma}_f(h)$ is of full rank and $\|\boldsymbol{\Sigma}_f(h)\|_2 \asymp O(1) \asymp \|\boldsymbol{\Sigma}_f(h)\|_{\min}$.

The latent process does not have to be stationary, and only needs to satisfy the mixing condition. We do not need to assume any specific model for the latent process $\{\boldsymbol{F}_t\}$ since we only use the eigen-analysis based on autocovariances of the observed process at nonzero lags. These two features make the estimation procedure we adopt more attractive and general than the standard principal component analysis approach.

Condition 3. Each element of $\boldsymbol{\Sigma}_e$ remains bounded as p_1 and p_2 increase to infinity.

In model (1), $\boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}'$ can be viewed as the signal part of the observation $\boldsymbol{X}_t$, and $\boldsymbol{E}_t$ as the noise. The signal strength, or the strength of the factors, can be measured by the L_2-norm of the loading matrices which are assumed to grow with the dimensions.

Condition 4. There exist constants δ_1 and $\delta_2 \in [0, 1]$ such that $\|\boldsymbol{R}\|_2^2 \asymp p_1^{1-\delta_1} \asymp \|\boldsymbol{R}\|_{\min}^2$ and $\|\boldsymbol{C}\|_2^2 \asymp p_2^{1-\delta_2} \asymp \|\boldsymbol{C}\|_{\min}^2$.

The rate δ_1 and δ_2 are called the strength for row factors and the strength for column factors respectively. They measure relative growth rate of the amount of information which the observed process $\boldsymbol{X}_t$ carries about common factors as the dimensions increase, with respect to the growth rate of the amount of noise. When $\delta_i = 0$, the factors are strong; when $\delta_i > 0$, the factors are weak, which means the information contained in $\boldsymbol{X}_t$ about the factors grows more slowly than the noises introduced as p_i increases.

Condition 5. $\boldsymbol{M}$ has k_1 distinct positive eigenvalues.

As stated in Section 3, only $\mathcal{M}(\boldsymbol{Q}_1)$ and $\mathcal{M}(\boldsymbol{Q}_2)$ are uniquely determined, while $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are not. However, when the eigenvalues of $\boldsymbol{M}$ are distinct, we could uniquely define $\boldsymbol{Q}_1$ as $\boldsymbol{Q}_1 = \{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_{k_1}\}$, where $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_{k_1}$ are eigenvectors of $\boldsymbol{M}$ corresponding to its k_1 largest eigenvalues $\{\lambda_1 > \lambda_2 \ldots > \lambda_{k_1}\}$.

The following theorems show the rate of convergence for estimators of loading spaces and the eigenvalues.

Theorem 1. *Under conditions 1-5 and* $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$*, it holds that*

$$\|\widehat{\boldsymbol{Q}}_1 - \boldsymbol{Q}_1\|_2 = O_p(p_1^{\delta_1} p_2^{\delta_2} T^{-1/2}), \quad \textit{and} \quad \|\widehat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2\|_2 = O_p(p_1^{\delta_1} p_2^{\delta_2} T^{-1/2}).$$

For the impact of δ_i's, it is not surprising that the stronger the factors are, the more useful information the observed process carries and the faster the estimators converge. More interestingly, the strengths of row factors and column factors δ_1 and δ_2 determine the rates together. An increase in the strength of row factors is able to improve the estimation of the column factors loading space and vice versa.

For the impact of p_1 and p_2, we first consider the case that δ_i's are not 0, which means that noise increases faster than useful information. Dimension increases will dilute the information of the latent factor process. Less efficient estimators result from "curse of dimension". Assuming that δ_i's are 0, which means that the signal is as strong as noise, and dimension increases will not affect the estimation of loading spaces.

Theorem 2. *With conditions 1-5 and* $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$*, the eigenvalues* $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_{p_1}\}$ *of* $\widehat{\boldsymbol{M}}$ *which is sorted in descending order satisfy*

$$\begin{aligned} |\hat{\lambda}_i - \lambda_i| &= O_p(p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}), \quad \textit{for } i = 1, 2, \ldots, k_1, \\ \textit{and} \qquad |\hat{\lambda}_i| &= O_p(p_1^2 p_2^2 T^{-1/2}), \quad \textit{for } i = k_1 + 1, \ldots, p_1, \end{aligned}$$

where $\lambda_1 > \lambda_2 \ldots > \lambda_{k_1}$ *are eigenvalues of* $\boldsymbol{M}$*.*

Theorem 2 shows that the estimators for nonzero eigenvalues of $\boldsymbol{M}$ converge more slowly than those for the zero eigenvalues. It provides the theoretical support for the ratio estimator proposed in Section 3.1. If we use the transpose of $\boldsymbol{X}_t$ to construct $\boldsymbol{M}$ with the same procedure, we will obtain similar results for estimation of its eigenvalues.

Let $\boldsymbol{S}_t$ be the dynamic signal part of $\boldsymbol{X}_t$, that is, $\boldsymbol{S}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}' = \boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'$. Then a natural estimator of $\boldsymbol{S}_t$ is given by,

$$\widehat{\boldsymbol{S}}_t = \hat{\boldsymbol{Q}}_1\hat{\boldsymbol{Q}}_1'\boldsymbol{X}_t\hat{\boldsymbol{Q}}_2\hat{\boldsymbol{Q}}_2'.$$

The following theorem demonstrates the theoretical properties of estimator of $\boldsymbol{S}_t$.

Theorem 3. *If Conditions 1-5 hold,* $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$*, and* $\|\mathrm{E}(\boldsymbol{E}_t\boldsymbol{E}_t')\|_2/p_{\max}$ *is bounded, we have*

$$p_1^{-1/2} p_2^{-1/2} \|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2 = O_p(p_1^{\delta_1/2} p_2^{\delta_2/2} T^{-1/2} + p_{\min}^{-1/2}),$$

where $p_{\max} = \max\{p_1, p_2\}$*, and* $p_{\min} = \min\{p_1, p_2\}$*.*

Since $\boldsymbol{Q}_i$ is not identifiable in Model (1), another measure to quantify the accuracy of factor loading matrices estimation is the distance between $\mathcal{M}(\boldsymbol{Q}_i)$ and $\mathcal{M}(\widehat{\boldsymbol{Q}}_i)$. For two orthogonal matrices $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ of sizes $p \times q_1$ and $p \times q_2$, define

$$\mathcal{D}(\boldsymbol{O}_1, \boldsymbol{O}_2) = \left(1 - \frac{1}{\max(q_1, q_2)} \mathrm{tr}(\boldsymbol{O}_1\boldsymbol{O}_1'\boldsymbol{O}_2\boldsymbol{O}_2')\right)^{1/2}.$$

Then $\mathcal{D}(\boldsymbol{O}_1, \boldsymbol{O}_2)$ is a quantity between 0 and 1. It is equal to 0 if the column spaces of $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ are the same and 1 if they are orthogonal.

Theorem 4. *If Conditions 1-5 hold and* $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$*, we have*

$$\mathcal{D}(\widehat{\boldsymbol{Q}}_i, \boldsymbol{Q}_i) = O_p(p_1^{\delta_1} p_2^{\delta_2} T^{-1/2}), \quad \textit{for } i = 1, 2.$$

Theorem 4 shows that the error to estimate loading spaces is on the same order as that for the estimated $\boldsymbol{Q}_i$'s.

3.3 An F-test for model comparison

As the matrix factor model (1) is a special case of the vectorized model (2), it is of interests to test whether the significant dimension reduction of (1) is valid. Here we develop an F-test for comparison.

Define $\mathcal{M}(\boldsymbol{B}_i)$ as the orthogonal complement of $\mathcal{M}(\boldsymbol{Q}_i)$, where $\boldsymbol{B}_i = (\boldsymbol{b}_{i,1}, \ldots, \boldsymbol{b}_{i,p_i-k_i})$ is a $p_i \times (p_i - k_i)$ matrix and $(\boldsymbol{Q}_i, \boldsymbol{B}_i)$ forms a $p_i \times p_i$ orthonormal matrix, i.e., $\boldsymbol{B}_i'\boldsymbol{Q}_i = \boldsymbol{0}$, and $\boldsymbol{B}_i'\boldsymbol{B}_i = \boldsymbol{I}_{p_i-k_i}$, for $i = 1, 2$. The $p_1 - k_1$ eigenvectors of $\boldsymbol{M}$ corresponding to its zero eigenvalues forms $\mathcal{M}(\boldsymbol{B}_1)$. The columns of the estimated matrix $\widehat{\boldsymbol{B}}_1$ are the k_1 orthonormal eigenvectors of $\widehat{\boldsymbol{M}}$ corresponding to its $p_1 - k_1$ minimal eigenvalues. $\boldsymbol{B}_2$ is defined and $\widehat{\boldsymbol{B}}_2$ is obtained in a similar way.

If model (1) is true and $\text{vec}(\boldsymbol{E}_\text{t})$ is a Gaussian white noise process with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}_e$, then $\text{vec}(\boldsymbol{B}_1'\boldsymbol{X}_\text{t}\boldsymbol{B}_2)$ is a white-noise process as well, and $\text{vec}(\boldsymbol{B}_1'\boldsymbol{X}_\text{t}\boldsymbol{B}_2) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, (\boldsymbol{B}_2' \otimes \boldsymbol{B}_1')\boldsymbol{\Sigma}_\text{e}(\boldsymbol{B}_2 \otimes \boldsymbol{B}_1))$.

The estimation of (2) can be done following Lam et al. (2011). Define $\mathcal{M}(\boldsymbol{\Psi})$ as the orthogonal complement of $\mathcal{M}(\boldsymbol{\Phi})$, where $\boldsymbol{\Psi}$ is a $(p_1p_2) \times (p_1p_2 - k_1k_2)$ matrix, i.e., $\boldsymbol{\Psi}'\boldsymbol{\Phi} = \boldsymbol{0}$, $\boldsymbol{\Psi}'\boldsymbol{\Psi} = \boldsymbol{I}_{p_1p_2-k_1k_2}$. If model (2) is true, $\boldsymbol{\Psi}'\text{vec}(\boldsymbol{X}_t) \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi}'\boldsymbol{\Sigma}_e\boldsymbol{\Psi})$.

Then we construct an F-test for the matrix-valued factor model as follows,

$$H_0 : \text{Model (1) is true}, \quad H_a : \text{Model (2) is true}.$$

We reject H_0 if

$$F = \frac{(\text{SSE}_r - \text{SSE}_f)/[(p_1k_2 + p_2k_1 - 2k_1k_2)T]}{\text{SSE}_f/[(p_1p_2 - k_1k_2)T]} > F_{\alpha,(p_1k_2+p_2k_1-2k_1k_2)T,(p_1p_2-k_1k_2)T},$$

where

$$\text{SSE}_f = \sum_{t=1}^{T} \text{vec}(\boldsymbol{X}_t)'\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Sigma}}_f^{-1}\widehat{\boldsymbol{\Psi}}'\text{vec}(\boldsymbol{X}_t), \quad \widehat{\boldsymbol{\Sigma}}_f = \frac{\sum_{t=1}^{T} \widehat{\boldsymbol{\Psi}}'\text{vec}(\boldsymbol{X}_t)\text{vec}(\boldsymbol{X}_t)'\widehat{\boldsymbol{\Psi}}}{T-1},$$

$$\text{SSE}_r = \sum_{t=1}^{T} \text{vec}(\boldsymbol{X}_t)'(\widehat{\boldsymbol{B}}_2 \otimes \widehat{\boldsymbol{B}}_1)\widehat{\boldsymbol{\Sigma}}_r^{-1}(\widehat{\boldsymbol{B}}_2' \otimes \widehat{\boldsymbol{B}}_1')\text{vec}(\boldsymbol{X}_t),$$

$$\text{and } \widehat{\boldsymbol{\Sigma}}_r = \frac{\sum_{t=1}^{T}(\widehat{\boldsymbol{B}}_2' \otimes \widehat{\boldsymbol{B}}_1')\text{vec}(\boldsymbol{X}_t)\text{vec}(\boldsymbol{X}_t)'(\widehat{\boldsymbol{B}}_2 \otimes \widehat{\boldsymbol{B}}_1)}{T-1}.$$

4 Simulation

In this section, we study the numerical performance of our proposed approach. In all simulations, the observed data $\boldsymbol{X}_t$'s are simulated according to Model (1),

$$\boldsymbol{X}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}' + \boldsymbol{E}_t, \quad t = 1, 2, \ldots, T.$$

We choose the dimensions of the latent factor process $\boldsymbol{F}_t$ to be $k_1 = 3$ and $k_2 = 2$. The entries of $\boldsymbol{F}_t$ are simulated as k_1k_2 independent AR(1) processes with noise $N(0,1)$, and the AR coefficients are $[-0.5 \ 0.6; \ 0.8 \ -0.4; \ 0.7 \ 0.3]$. The entries of $\boldsymbol{R}$ and $\boldsymbol{C}$ are independently sampled from the uniform distribution $U(-p_i^{-\delta_i/2}, p_i^{-\delta_i/2})$ for $i = 1, 2$ respectively. The error process $\boldsymbol{E}_t$ is a white noise process with mean $\mathbf{0}$ and a Kronecker product covariance structure, that is,

$$\text{Cov}(\text{vec}(\boldsymbol{E}_t)) = \boldsymbol{\Gamma}_2 \otimes \boldsymbol{\Gamma}_1,$$

where $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are of sizes $p_1 \times p_1$ and $p_2 \times p_2$ respectively. Both $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ have values 1 on the diagonal entries and 0.2 on the off-diagonal entries.

We consider three pairs of (δ_1, δ_2) to be $(0.5, 0.5)$, $(0.5, 0)$ and $(0, 0)$. For each pair of δ_i's, the dimensions (p_1, p_2) are chosen to be $(20, 20)$, $(20, 50)$ and $(50, 50)$. The sample size T are selected as $0.5p_1p_2$, p_1p_2, $1.5p_1p_2$ and $2p_1p_2$. For a given combination of these parameters, we will run the simulation 200 times. In all simulations, we take $h_0 = 1$.

We first study and compare the matrix-valued approach developed in this paper for Model (1) and the vector-valued approach in Lam and Yao (2012) for Model (2). In what follows, we use subscripts mat and vec to denote our approach and Lam and Yao (2012)'s method respectively when estimating a parameter. First, we apply the vector-valued approach to the observations $\{\text{vec}(\boldsymbol{X}_t)\}$ $(t = 1, 2, \ldots, T)$ and obtain the estimates $\hat{k}_{\text{vec}}$ and $\widehat{\boldsymbol{Q}}_{\text{vec}}$. In this case, the true values of k and $\boldsymbol{Q}$ are $k = k_1k_2 = 6$ and $\boldsymbol{Q} = \boldsymbol{Q}_2 \otimes \boldsymbol{Q}_1$. Then we employ our matrix-valued approach to $\boldsymbol{X}_t$ and obtain estimates of $\hat{k}_1$, $\hat{k}_2$, $\widehat{\boldsymbol{Q}}_1$ and $\widehat{\boldsymbol{Q}}_2$. The total number of factors is estimated by $\hat{k}_{\text{mat}} = \hat{k}_1\hat{k}_2$, and the loading matrix $\boldsymbol{Q}$ is estimated by $\widehat{\boldsymbol{Q}}_{\text{mat}} = \widehat{\boldsymbol{Q}}_2 \otimes \widehat{\boldsymbol{Q}}_1$.

Table 2 and Table 3 summarize the simulation results of the comparison. f_{vec} denotes the relative frequency of correctly estimating the true number of factors k over 200 simulation runs for the vector approach. $\mathcal{D}_{\text{vec}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ represents the estimation accuracy when estimating $\boldsymbol{Q}$. The quantities f_{mat} and $\mathcal{D}_{\text{mat}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ are similarly defined for the matrix approach. We can see that the matrix approach efficiently improve the estimation accuracy. Moreover, with stronger signals and more data samples, both approaches increase their estimation accuracy.

Table 4 and Table 5 show the results for the matrix approach when estimating the left and right components. Here, f_i, $i = 1, 2$, indicates the relative frequency of correctly estimating the true rank k_i over 200 simulations, and $\mathcal{D}(\widehat{\boldsymbol{Q}}_i, \boldsymbol{Q}_i)$ denotes the estimation accuracy when estimating the loading space. We can see that increasing the strength of one loading matrix can improve estimation accuracy for both loading spaces.

				$T = .5p_1p_2$		$T = p_1p_2$		$T = 1.5p_1p_2$		$T = 2p_1p_2$	
δ_1	δ_2	p_1	p_2	f_{vec}	f_{mat}	f_{vec}	f_{mat}	f_{vec}	f_{mat}	f_{vec}	f_{mat}
0.5	0.5	20	20	0	0	0	0.02	0	0.06	0	0.245
		20	50	0	0	0	0	0	0.04	0	0.275
		50	50	0	0	0	0	0	0	0	0
0.5	0	20	20	0	0.045	0	0.02	0	0.01	0	0
		20	50	0	0.04	0	0.01	0	0.005	0	0
		50	50	0	0	0	0	0	0	0	0
0	0	20	20	0	0.365	0.83	0.66	0.965	0.89	0.995	0.985
		20	50	0	0.17	0.86	0.665	0.995	0.98	1	0.995
		50	50	0	0.995	1	1	1	1	1	1

Table 2: Relative frequency of correctly estimating k.

				$T = .5 * p_1 * p_2$		$T = p_1 * p_2$		$T = 1.5 * p_1 * p_2$		$T = 2 * p_1 * p_2$	
δ_1	δ_2	p_1	p_2	$\mathcal{D}_{\text{vec}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{\text{mat}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{\text{vec}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{\text{mat}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{\text{vec}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{\text{mat}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{\text{vec}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{\text{mat}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$
0.5	0.5	20	20	8.75(0.17)	8.26(0.07)	8.24(0.18)	8.19(0.03)	7.88(0.18)	8.17(0.02)	7.62(0.17)	8.16(0.02)
		20	50	8.72(0.10)	8.20(0.06)	8.40(0.09)	8.15(0.01)	8.15(0.14)	8.14(0.01)	7.92(0.16)	8.13(0.01)
		50	50	8.51(0.14)	8.34(0.22)	7.62(0.14)	8.13(0.05)	7.10(0.09)	8.10(0.01)	6.81(0.06)	8.09(0.01)
0.5	0	20	20	6.40(0.29)	7.19(1.13)	5.50(0.31)	4.66(1.45)	4.82(0.39)	2.51(0.99)	4.37(0.45)	1.64(0.72)
		20	50	5.64(0.24)	6.75(1.13)	4.75(0.35)	2.30(0.80)	3.98(0.50)	1.14(0.30)	3.37(0.45)	0.78(0.20)
		50	50	5.07(0.10)	4.92(0.94)	4.46(0.29)	1.47(0.23)	3.59(0.60)	0.83(0.12)	2.73(0.46)	0.59(0.08)
0	0	20	20	3.64(0.23)	0.71(0.16)	2.77(0.16)	0.48(0.08)	2.33(0.14)	0.38(0.05)	2.07(0.13)	0.33(0.04)
		20	50	2.84(0.18)	0.44(0.07)	2.13(0.10)	0.30(0.05)	1.77(0.08)	0.25(0.04)	1.56(0.07)	0.21(0.03)
		50	50	1.85(0.10)	0.18(0.02)	1.34(0.06)	0.12(0.01)	1.11(0.04)	0.10(0.01)	0.97(0.04)	0.09(0.01)

Table 3: Means and standard deviations (in parentheses) of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$. For ease of presentation, all numbers in this table are the true numbers multiplied by 10.

				$T = .5 * p_1 * p_2$		$T = p_1 * p_2$		$T = 1.5 * p_1 * p_2$		$T = 2 * p_1 * p_2$	
δ_1	δ_2	p_1	p_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
0.5	0.5	20	20	0	0.255	0.02	0.87	0.06	0.985	0.245	1
		20	50	0	0.545	0	1	0.04	1	0.275	1
		50	50	0	0.16	0	1	0	1	0	1
0.5	0	20	20	0.015	0	0	0	0	0	0	0
		20	50	0	0	0	0	0	0	0	0
		50	50	0	0	0	0	0	0	0	0
0	0	20	20	0.56	0.465	0.875	0.705	0.99	0.895	0.995	0.985
		20	50	0.64	0.195	0.99	0.67	1	0.98	1	0.995
		50	50	1	0.995	1	1	1	1	1	1

Table 4: Relative frequency of correctly estimating k_1 and k_2.

				$T = .5 * p_1 * p_2$		$T = p_1 * p_2$		$T = 1.5 * p_1 * p_2$		$T = 2 * p_1 * p_2$	
δ_1	δ_2	p_1	p_2	$\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$
0.5	0.5	20	20	5.96(0.19)	7.12(0.03)	5.80(0.07)	7.09(0.01)	5.75(0.04)	7.09(0.01)	5.73(0.04)	7.08(0.01)
		20	50	5.87(0.15)	7.07(0.02)	5.77(0.04)	7.05(0.01)	5.75(0.02)	7.04(0.01)	5.74(0.02)	7.04(0.01)
		50	50	6.26(0.56)	7.05(0.01)	5.73(0.13)	7.04(0.00)	5.64(0.04)	7.04(0.00)	5.61(0.03)	7.03(0.00)
0.5	0	20	20	5.36(0.41)	5.42(2.22)	4.27(1.13)	1.66(1.70)	2.39(1.02)	0.68(0.30)	1.52(0.75)	0.54(0.17)
		20	50	5.02(0.67)	5.15(1.61)	1.82(0.77)	1.32(0.60)	0.82(0.27)	0.74(0.30)	0.54(0.18)	0.53(0.17)
		50	50	3.68(0.48)	3.44(1.23)	1.31(0.20)	0.65(0.19)	0.73(0.11)	0.37(0.11)	0.51(0.07)	0.28(0.08)
0	0	20	20	0.55(0.16)	0.44(0.10)	0.36(0.08)	0.31(0.06)	0.28(0.05)	0.25(0.05)	0.24(0.04)	0.22(0.04)
		20	50	0.25(0.06)	0.36(0.07)	0.16(0.03)	0.26(0.05)	0.12(0.02)	0.21(0.04)	0.10(0.02)	0.18(0.03)
		50	50	0.13(0.02)	0.12(0.02)	0.09(0.01)	0.08(0.01)	0.07(0.01)	0.07(0.01)	0.06(0.01)	0.06(0.01)

Table 5: Means and standard deviations (in parentheses) of $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$'s over 200 simulation runs. For ease of presentation, all numbers in this table are the true numbers multiplied by 10.

Factor	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1	1	2	0	-3	-5	-14	-11	-11	-20	-3
2	-3	3	2	-1	-5	6	-1	4	-8	27
3	-15	-15	-12	-11	-11	1	-4	-2	5	0

Table 6: Size loading matrix after rotation and scaling

5 Real Example: Fama-French 10 by 10 series

In this section we illustrate the matrix factor model using the Fama-French 10 by 10 return series. A universe of stocks are grouped into 100 portfolios, according to ten levels of market capital (size) and ten levels of book to equity ratio (BE). Their monthly returns from January 1964 to December, 2015 for total 624 months and overall 62,400 observations. For more detailed information, see *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*.

Figure 1 shows the time series plot of the 100 series (standardized) and Figure 2 shows the estimated log eigenvalues for the row (size) and column (BE) loading matrices. Although the eigenvalue ratio estimate presented in Section 3.1 indicates $k_1 = k_2 = 1$, we use $k_1 = k_2 = 3$ here for illustration. Tables 6 and 7 shows the estimated loading matrices after a rotation that maximizes element variation, and a proper scaling for a cleaner view. For size, it is clearly seen that there are three groups. The 6-th to 9-th largest size portfolios load heavily (with roughly equal weights) on the first row of the factor matrix, while the 1-st to 5-th smallest size portfolios load heavily (with roughly equal weights) on the third row of the factor matrix. The portfolio consisting of the largest companies loads only on the second row of the factor matrix. We note that the Fama-French size factor proposed in Fama and French (1993) is constructed using the return differences of the largest 30% of the companies (combining our 8-th to 10-th size portfolio) and the smallest 30% of the companies (combining our 1st to 3rd size portfolio). Figure 3(a) shows the grouping.

Turning to the Book to Equity ratio, Table 7 shows a different pattern in the column loading matrix. There seem to have four groups. The smallest 3 BE portfolios load heavily on the first column of the factor matrix; the 4th to 6th BE portfolios load heavily on the first two columns of the factor matrix, with opposite signs; the 7th to 9th BE portfolios load heavily on the third columns of the factor matrix; and the largest BE portfolio loads only on the second column. It can also be seen that the portfolios load on the first column of the factor matrix with increasing weights. Figure 3(b) shows the grouping.

Figure 4 shows the estimated factor matrices over time. It can be potentially used to replace the Fama-French size factor (SMB) and book to equity factor (HML) in a Fama-French factor model for asset pricing.

Figure 5 shows the log eigenvalues of the loading matrix for a vectorized factor model (2). Models with various number of factors were estimated and a comparison is shown in Table 8, where RSS denotes residual sum of squares and SST denotes the total sum of squares. Clearly

Factor	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8	BE9	BE10
1	-59	-45	-41	*-32*	*-26*	*-23*	-15	-4	14	7
2	-2	-20	5	*10*	*14*	*28*	3	3	-3	92
3	-20	13	-2	12	14	7	35	51	72	-3

Table 7: BtoE loading matrix after rotation and scaling

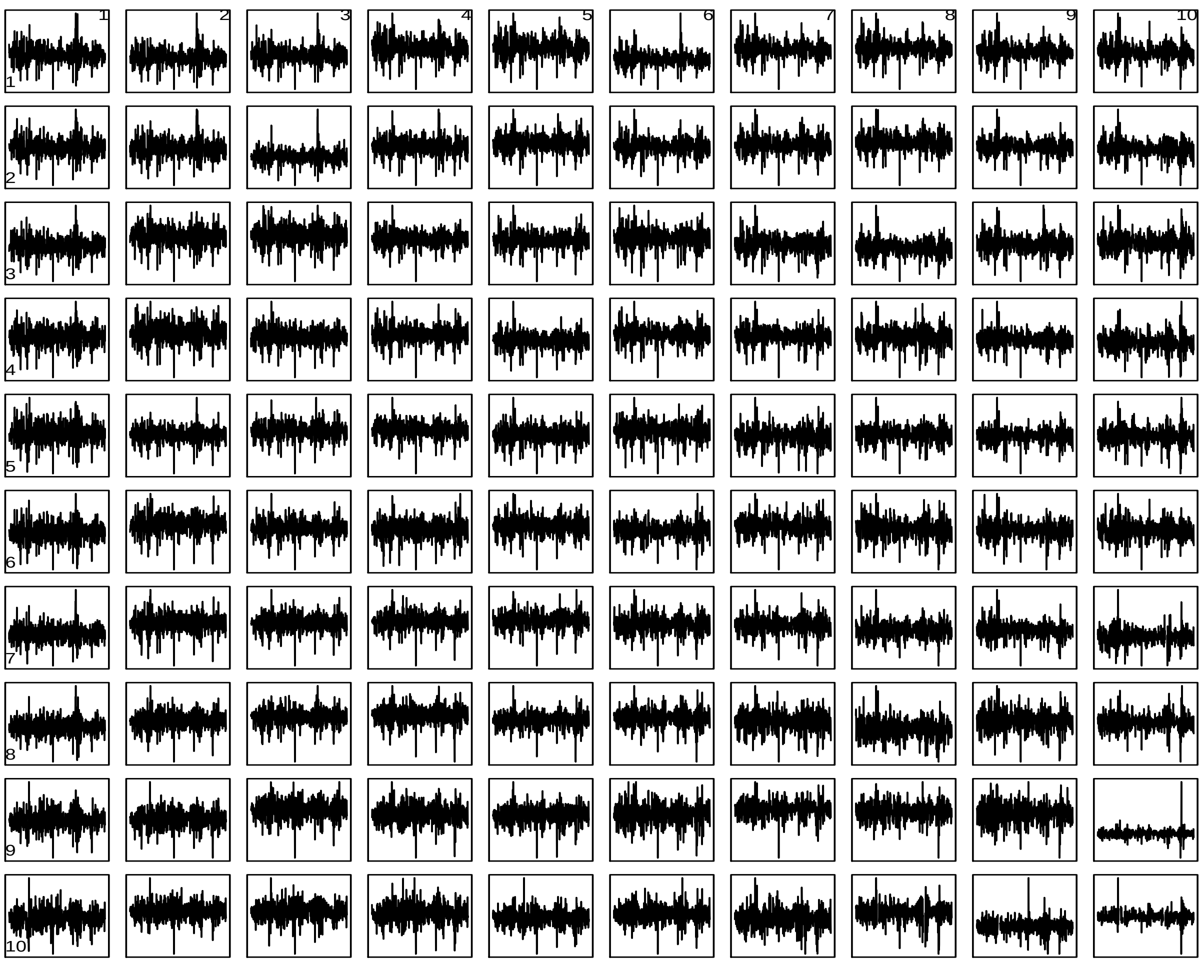


Figure 1: Time Series plot of Fama-French 10 by 10 series

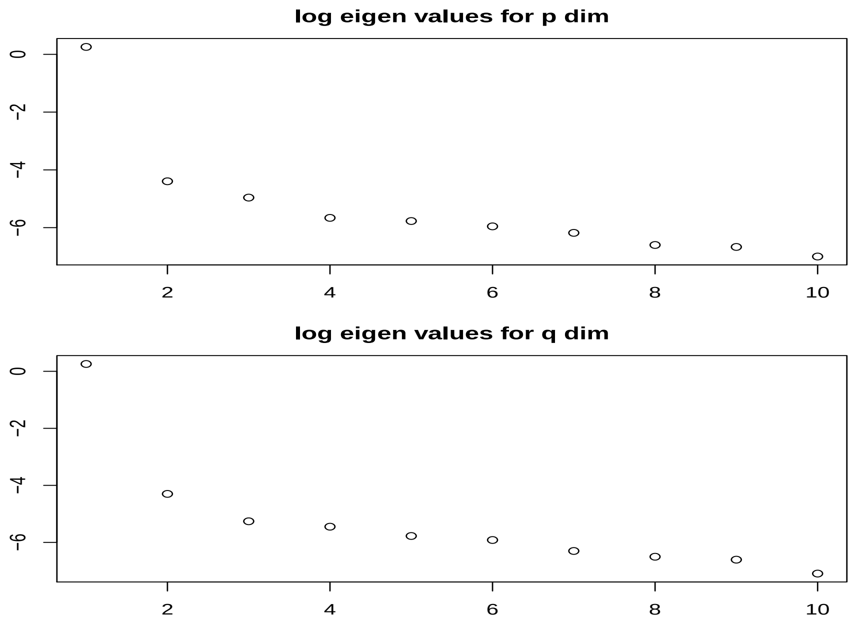
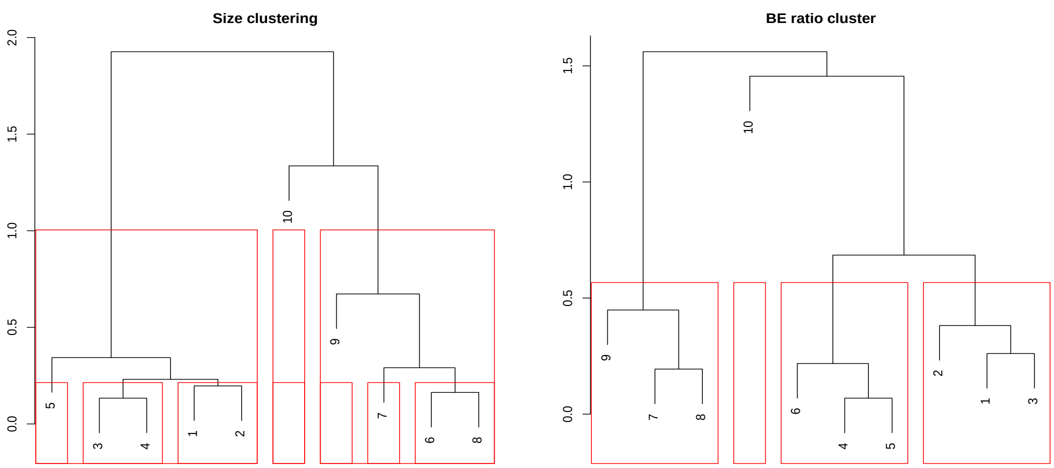


Figure 2: Fama-French series: estimate log eigenvalues for row and column loading matrices



,

Figure 3: Fama-French series: clustering loading matrices after rotation

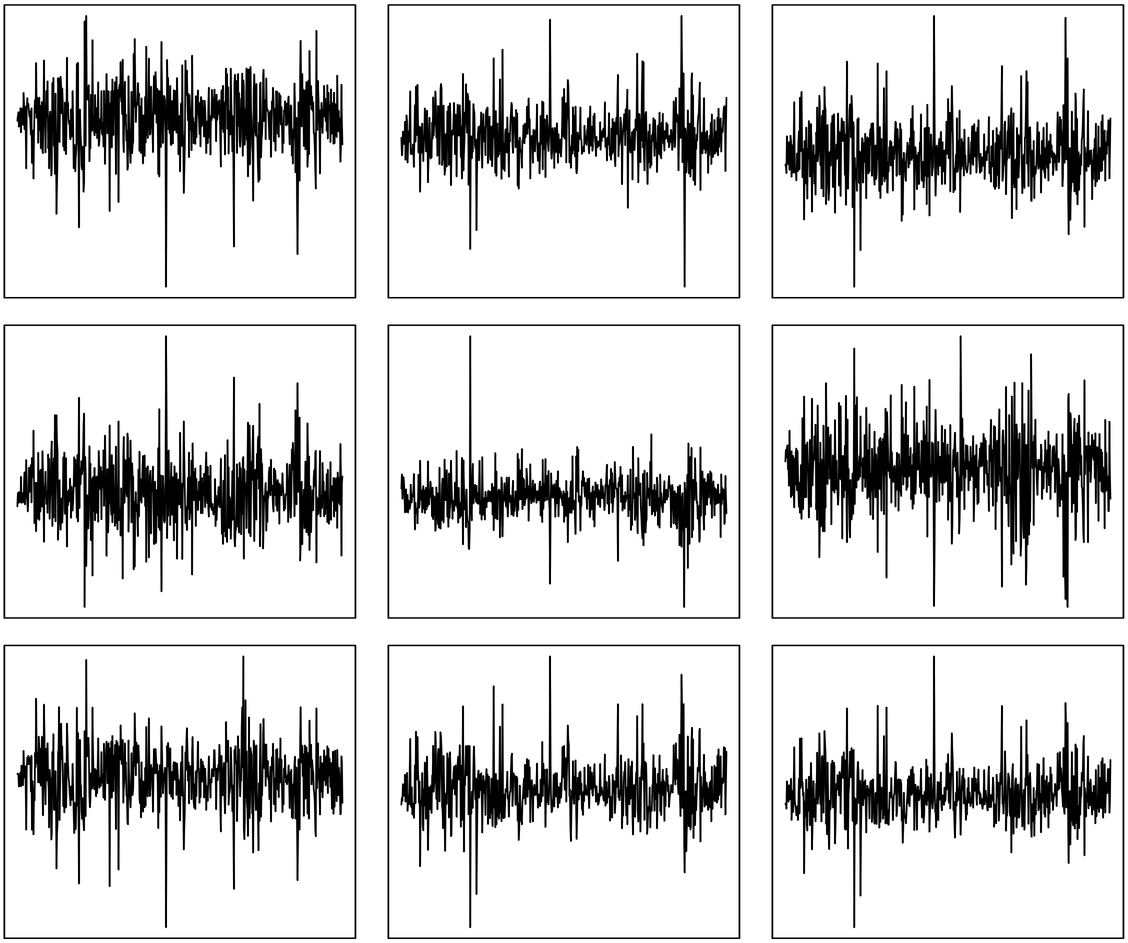


Figure 4: Fama-French series: estimated factors

	factor	RSS	RSS/SST	# factors	# parameters
Matrix model	(2,3)	12,285	0.197	6	50
Matrix model	(3,2)	12,109	0.195	6	50
Matrix model	(3,3)	10,672	0.171	9	60
Matrix model	(4,3)	9,989	0.161	12	70
Matrix model	(4,4)	9,519	0.153	16	80
Vector model	7	10,670	0.172	7	700
Vector model	8	10,196	0.164	8	800
Vector model	9	9,860	0.158	9	900

Table 8: Comparison of different models for Fama-French series

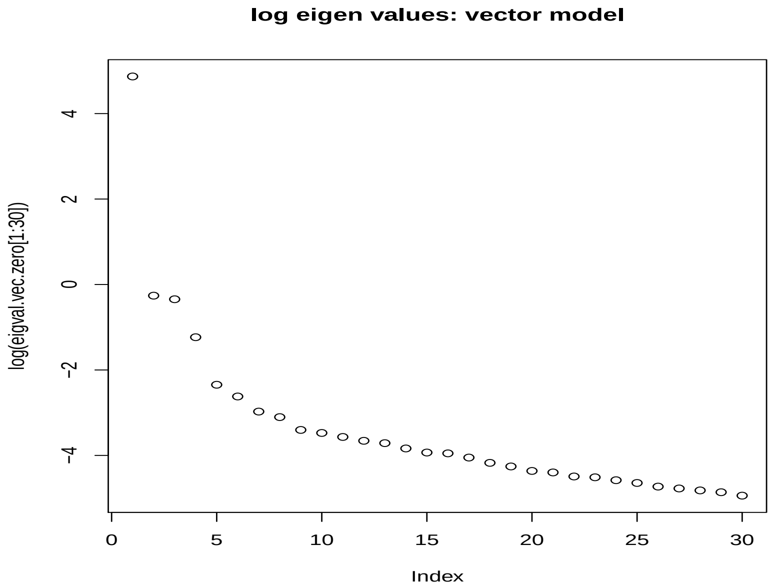


Figure 5: Fama-French series: log eigenvalues of the vectorized model

the matrix factor uses much less number of parameters in loading matrices to achieve similar estimation performance, though the number of factors may be slightly larger.

References

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

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

Brockwell, P. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and meanvariance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

Chang, J., Guo, B., and Yao, Q. (2015). High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. *Journal of Econometrics*, 189(2):297–312.

Crainiceanu, C. M., Caffo, B. S., Luo, S., Zipunnikov, V. M., and Punjabi, N. M. (2011). Population Value Decomposition, a Framework for the Analysis of Image Populations. *Journal of the American Statistical Association*, 106(495):775–790.

Ding, C. and Ye, J. (2005). 2-Dimensional Singular Value Decomposition for 2D Maps and Images. In *Proc. SIAM Int'l Conf. Data Mining (SDM'05)*, pages 32–43.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflactions. *Econometrika*, 59:987–1007.

Engle, R. and Kroner, K. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1):122–150.

Fama, E. F. and French, K. R. (1993). The cross-section of expected stock returns. *Journal of Finance*, 47:427–465.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, 39(6):3320.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: identification and estimation. *Review of Economics and Statistics*, 82(4):540–554.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions*. Chapman & Hall/CRC, Boca Raton, FL.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Kollo, T. and von Rosen, D. (2006). *Advanced multivariate statistics with matrices*, volume 579. Springer.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Annals of Statistics*, 40(2):694–726.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *Journal of the American Statistical Association*, 107(499):1187–1200.

Liu, X. and Chen, R. (2016). Regime-switching factor models for high-dimensional time series. *Statistica Sinica*.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer, Berlin.

Paatero, P. and Tapper, U. (1994). Positive matrix factorization: a non-negative factor model wiht optimal utilization of errorestimates of data vaelus. *Biometrika*, 5(1):111–126.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95(2):365–379.

Stock, J. and Watson, M. (2004). An empirical comparison of methods for forecasting using many predictors. *Technical Report*, Department of Economics, Havard University.

Tiao, G. and Box, G. (1981). Modelling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802–816.

Tiao, G. and Tsay, R. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 51(2):157–213.

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

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

Tsay, R. (2014). *Multivariate Time Series Analysis.* New York: Wiley.

Walden, A. and Serroukh, A. (2002). Wavelet analysis of matrix-valued time series. *Proceedings: Mathematical, Physical and engineering Sciences*, 458(2017):157–179.

Wang, D., Shen, H., and Truong, Y. (2016). Efficient dimension reduction for high-dimensional matrix-valued data. *Neurocomputing*, 190:25–34.

Werner, K., Jansson, M., and Stoica, P. (2008). On estimation of covariance matrices with Kronecker product structure. 56(2):478–491.

Yang, J., Zhang, D., Frangi, A. F., and Yang, J. (2004). Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. 26(1):131–137.

Ye, J. (2005). Generalized Low Rank Approximations of Matrices. *Machine Learning*, 61(1-3):167–191.

Yin, J. and Li, H. (2012). Model selection and estimation in the matrix normal graphical model. *Journal of Multivariate Analysis*, 107(0):119 – 140.

Zhang, D. and Zhou, Z. (2005). $(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing*, 69(1-3):224–231.

Zhao, J. and Leng, C. (2014). Structured lasso for regression with matrix covariates. *Statistica Sinica*, 24:799–814.

Zhou, H. and Li, L. (2014). Regularized matrix regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):463–483.

Zhou, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *Annals of Statistics*, 42(2):532–562.

Appendix: Proofs

We start by defining some quantities used in the proofs. Write

$$\begin{aligned}
\boldsymbol{\Omega}_{s,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \mathrm{Cov}(\boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}_{i\cdot}, \boldsymbol{R}\boldsymbol{F}_{t+h}\boldsymbol{C}_{j\cdot}),\\
\boldsymbol{\Omega}_{fa,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \mathrm{Cov}(\boldsymbol{F}_t\boldsymbol{C}_{i\cdot}, \boldsymbol{F}_{t+h}\boldsymbol{C}_{j\cdot})\\
\widehat{\boldsymbol{\Omega}}_{s,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}_{i\cdot}\boldsymbol{C}'_{j\cdot}\boldsymbol{F}'_{t+h}\boldsymbol{R}',\\
\widehat{\boldsymbol{\Omega}}_{s\epsilon,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}_{i\cdot}\boldsymbol{\epsilon}'_{t+h,j},\\
\widehat{\boldsymbol{\Omega}}_{\epsilon s,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{\epsilon}_{t,i}\boldsymbol{C}'_{j\cdot}\boldsymbol{F}'_{t+h}\boldsymbol{R}',\\
\widehat{\boldsymbol{\Omega}}_{\epsilon,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{\epsilon}_{t,i}\boldsymbol{\epsilon}'_{t+h,j},\\
\widehat{\boldsymbol{\Omega}}_{fa,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{F}_t\boldsymbol{C}_{i\cdot}\boldsymbol{C}'_{j\cdot}\boldsymbol{F}'_{t+h}.
\end{aligned}$$

The following lemma establishes the entry-wise convergence rate of the covariance matrix estimation of the vectorized latent factor process $\mathrm{vec}(\boldsymbol{F}_t)$.

Lemma 1. *Let $F_{t,ij}$ denote the ij-th entry of $\boldsymbol{F}_t$. Under Conditions 1 and 2, for any $i,k = 1,2,\ldots,k_1$, and $j,l = 1,2,\ldots,k_2$, it follows that*

$$\left|\frac{1}{T-h}\sum_{t=1}^{T-h}\Big(F_{t,ij}F_{t+h,kl} - \mathrm{Cov}(F_{t,ij}, F_{t+h,kl})\Big)\right| = O_p(T^{-1/2}). \tag{14}$$

Proof. Under Conditions 1 and 2, by Davydov's inequality, it follows that

$$\begin{aligned}
&\mathrm{E}\left(\frac{1}{T-h}\sum_{t=1}^{T-h}\Big(F_{t,ij}F_{t+h,kl} - \mathrm{Cov}(F_{t,ij}, F_{t+h,kl})\Big)\right)^2\\
&= \frac{1}{(T-h)^2}\sum_{|t_1-t_2|\le h} \mathrm{E}[F_{t_1,ij}F_{t_1+h,kl} - \mathrm{E}(F_{t_1,ij}F_{t_1+h,kl})][F_{t_2,ij}F_{t_2+h,kl} - \mathrm{E}(F_{t_2,ij}F_{t_2+h,kl})]\}\\
&\quad + \frac{1}{(T-h)^2}\sum_{|t_1-t_2|> h} \mathrm{E}[F_{t_1,ij}F_{t_1+h,kl} - \mathrm{E}(F_{t_1,ij}F_{t_1+h,kl})][F_{t_2,ij}F_{t_2+h,kl} - \mathrm{E}(F_{t_2,ij}F_{t_2+h,kl})]\}\\
&\le \frac{C}{T-h} + \frac{C}{T-h}\sum_{u=1}^{T-2h-1}\alpha(u)^{1-2/\gamma} = O(T^{-1}).
\end{aligned}$$

Here C denotes a constant. □

Under the matrix-valued factor Model (1), the $\boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}'$ can be view as the signal part and $\boldsymbol{E}_t$ as noise. The following lemma concerns the rates of convergence for estimation of the signal, the noise, and the interaction between the two.

Lemma 2. *Under Conditions 1-4, it holds that*

$$\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{s,ij}(h)-\boldsymbol{\Omega}_{s,ij}(h)\|_2^2 = O_p(p_1^{2-2\delta_1}p_2^{2-2\delta_2}T^{-1}), \tag{15}$$

$$\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{s\epsilon,ij}(h)\|_2^2 = O_p(p_1^{2-\delta_1}p_2^{2-\delta_2}T^{-1}), \tag{16}$$

$$\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{\epsilon s,ij}(h)\|_2^2 = O_p(p_1^{2-\delta_1}p_2^{2-\delta_2}T^{-1}), \tag{17}$$

$$\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{\epsilon,ij}(h)\|_2^2 = O_p(p_1^{2}p_2^{2}T^{-1}). \tag{18}$$

Proof. Firstly, we have

$$\begin{aligned}
&\|\widehat{\boldsymbol{\Omega}}_{fa,ij}(h)-\boldsymbol{\Omega}_{fa,ij}(h)\|_2^2 \le \|\widehat{\boldsymbol{\Omega}}_{fa,ij}(h)-\boldsymbol{\Omega}_{fa,ij}(h)\|_F^2\\
&= \|\mathrm{vec}(\widehat{\boldsymbol{\Omega}}_{fa,ij}(h)-\boldsymbol{\Omega}_{fa,ij}(h))\|_2^2\\
&= \Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t-\mathrm{E}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t))\mathrm{vec}(\boldsymbol{C}_{i\cdot}\boldsymbol{C}_{j\cdot}')\Big\|_2^2\\
&\le \Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t-\mathrm{E}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t))\Big\|_2^2\|\mathrm{vec}(\boldsymbol{C}_{i\cdot}\boldsymbol{C}_{j\cdot}')\|_2^2\\
&\le \Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t-\mathrm{E}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t))\Big\|_F^2\|\boldsymbol{C}_{i\cdot}\boldsymbol{C}_{j\cdot}'\|_F^2\\
&= \Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t-\mathrm{E}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t))\Big\|_F^2\|\boldsymbol{C}_{i\cdot}\|_2^2\cdot\|\boldsymbol{C}_{j\cdot}\|_2^2.
\end{aligned}$$

Hence, by Condition 4 and Lemma 1, it follows that

$$\begin{aligned}
&\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{s,ij}(h)-\boldsymbol{\Omega}_{s,ij}(h)\|_2^2 = \sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\boldsymbol{R}(\widehat{\boldsymbol{\Omega}}_{fa,ij}(h)-\boldsymbol{\Omega}_{fa,ij}(h))\boldsymbol{R}'\|_2^2\\
&\le \|\boldsymbol{R}\|_2^4\Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t-\mathrm{E}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t))\Big\|_F^2\left(\sum_{i=1}^{p_2}\|\boldsymbol{C}_{i\cdot}\|_2^2\right)^2\\
&= \|\boldsymbol{R}\|_2^4\Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t-\mathrm{E}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t))\Big\|_F^2\|\boldsymbol{C}\|_F^4\\
&\le k_2^2\|\boldsymbol{R}\|_2^4\Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t-\mathrm{E}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t))\Big\|_F^2\|\boldsymbol{C}\|_2^4 = O_p(p_1^{2-2\delta_1}p_2^{2-2\delta_2}T^{-1}),
\end{aligned}$$

where the last inequality follows from Condition 4 and Lemma 1.

Similarly, for the interaction component between signal and noise, we have

$$\begin{aligned}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{s\epsilon,ij}(h)\|_2^2 \leq \sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\boldsymbol{R}\|_2^2\Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}\boldsymbol{F}_t\boldsymbol{C}_{i\cdot}\boldsymbol{\epsilon}'_{t+h,j}\Big\|_2^2 \\ \leq \quad \|\boldsymbol{R}\|_2^2\left(\sum_{j=1}^{p_2}\Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}\boldsymbol{\epsilon}_{t+h,j}\otimes\boldsymbol{F}_t\Big\|_2^2\right)\left(\sum_{i=1}^{p_2}\|\boldsymbol{C}_{i\cdot}\|_2^2\right) \\ = \quad O_p(p_1^{2-\delta_1}p_2^{2-\delta_2}T^{-1}),\end{aligned}$$

and

$$\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{\epsilon s,ij}(h)\|_2^2 \quad = \quad O_p(p_1^{2-\delta_1}p_2^{2-\delta_2}T^{-1}).$$

Lastly, for the noise term, we have

$$\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{\epsilon,ij}(h)\|_2^2 = \sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\frac{1}{T-h}\sum_{t=1}^{T-h}\boldsymbol{\epsilon}_{t,i}\boldsymbol{\epsilon}'_{t+h,j}\|_2^2 = O_p(p_1^2p_2^2T^{-1}).$$

□

With the four rates established in Lemma 2, we can now study the rate of convergence for the observed covariance matrix $\widehat{\boldsymbol{\Omega}}_{x,ij}(h)$.

Lemma 3. *Under Conditions 1-4, it holds that*

$$\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\|_2^2 = O_p(p_1^2p_2^2T^{-1}).$$

Proof. From the definition of $\widehat{\boldsymbol{\Omega}}_{x,ij}(h)$ in (12), we can decompose $\widehat{\boldsymbol{\Omega}}_{x,ij}(h)$ into four parts as follows,

$$\begin{aligned}\widehat{\boldsymbol{\Omega}}_{x,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h}\boldsymbol{x}_{t,i}\boldsymbol{x}'_{t+h,j} \\ &= \frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}_{i\cdot}+\boldsymbol{\epsilon}_{t,i})(\boldsymbol{R}\boldsymbol{F}_{t+h}\boldsymbol{C}_{j\cdot}+\boldsymbol{\epsilon}_{t+h,j})' \\ &= \widehat{\boldsymbol{\Omega}}_{s,ij}(h)+\widehat{\boldsymbol{\Omega}}_{s\epsilon,ij}(h)+\widehat{\boldsymbol{\Omega}}_{\epsilon s,ij}(h)+\widehat{\boldsymbol{\Omega}}_{\epsilon,ij}(h).\end{aligned}$$

Then by Lemma 2, it follows that

$$\begin{aligned}&\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\big\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\big\|_2^2 \\ \leq\ & 4\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\Big(\|\widehat{\boldsymbol{\Omega}}_{s,ij}(h)-\boldsymbol{\Omega}_{s,ij}(h)\|_2^2+\|\widehat{\boldsymbol{\Omega}}_{s\epsilon,ij}(h)\|_2^2+\|\widehat{\boldsymbol{\Omega}}_{\epsilon s,ij}(h)\|_2^2+\|\widehat{\boldsymbol{\Omega}}_{\epsilon,ij}(h)\|_2^2\Big) \\ =\ & O_p(p_1^2p_2^2T^{-1}). \qquad (19)\end{aligned}$$

□

Lemma 4. *Under Conditions 1-4, and $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}=o(1)$, it holds that*

$$\|\widehat{\boldsymbol{M}}-\boldsymbol{M}\|_2=O_p(p_1^{2-\delta_1}p_2^{2-\delta_2}T^{-1/2}).$$

Proof. From the definitions of $\widehat{\boldsymbol{M}}$ and $\boldsymbol{M}$ in (13) and (11), it follows that

$$\begin{aligned}
\|\widehat{\boldsymbol{M}}-\boldsymbol{M}\|_2 &= \Big\|\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\Big(\widehat{\boldsymbol{\Omega}}_{x,ij}(h)\widehat{\boldsymbol{\Omega}}'_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\boldsymbol{\Omega}'_{x,ij}(h)\Big)\Big\|_2\\
&\le \sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\Big(\|(\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h))(\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h))'\|_2+2\|\boldsymbol{\Omega}_{x,ij}(h)\|_2\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\|_2\Big)\\
&\le \sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\|_2^2+2\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\boldsymbol{\Omega}_{x,ij}(h)\|_2\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\|_2.
\end{aligned}$$

We have

$$\begin{aligned}
\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\boldsymbol{\Omega}_{x,ij}(h)\|_2^2 &= \sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\boldsymbol{R}\boldsymbol{\Omega}_{fa,ij}(h)\boldsymbol{R}'\|_2^2\le\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\boldsymbol{R}\|_2^4\|\boldsymbol{\Omega}_{fa,ij}(h)\|_2^2\\
&\le \|\boldsymbol{R}\|_2^4\cdot\Big\|\frac{1}{T-h}\sum_{t=1}^{T-h}\mathrm{E}(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t)\Big\|_2^2\cdot\left(\sum_{i=1}^{p_2}\|\boldsymbol{C}_{i\cdot}\|_2^2\right)^2\\
&= O(p_1^{2-2\delta_1}p_2^{2-2\delta_2}). \qquad (20)
\end{aligned}$$

By (20) and Lemma 3,

$$\begin{aligned}
&\left(\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\boldsymbol{\Omega}_{x,ij}(h)\|_2\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\|_2\right)^2\\
&\le \sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\boldsymbol{\Omega}_{x,ij}(h)\|_2^2\cdot\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\|_2^2\\
&\le O_p(p_1^{2-2\delta_1}p_2^{2-2\delta_2}p_1^2p_2^2T^{-1})=O_p(p_1^{4-2\delta_1}p_2^{4-2\delta_2}T^{-1}), \qquad (21)
\end{aligned}$$

where the second inequality follows from Cauchy-Schwarz inequality.

From (21), Lemma 3, and the condition $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}=o(1)$, it follows that

$$\|\widehat{\boldsymbol{M}}-\boldsymbol{M}\|_2=O_p(p_1^{2-\delta_1}p_2^{2-\delta_2}T^{-1/2}).$$

□

Lemma 5. *Under Conditions 2 and 3, we have*

$$\lambda_i(\boldsymbol{M})\asymp p_1^{2-2\delta_1}p_2^{2-2\delta_2},\quad i=1,2,...k_1,$$

where $\lambda_i(\boldsymbol{M})$ denotes the i-th largest singular value of $\boldsymbol{M}$.

Proof. Define

$$\boldsymbol{\Omega}_{\mathrm{f,v}}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} \mathrm{E}\big[\mathrm{vec}(\boldsymbol{F}_t)\mathrm{vec}(\boldsymbol{F}_{t+h})'\big]. \tag{22}$$

$$\begin{aligned}
\boldsymbol{\Omega}_{xa,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \mathrm{E}\big[(\boldsymbol{C}_{i\cdot}\otimes \boldsymbol{I}_{k_1})\mathrm{vec}(\boldsymbol{F}_t)\mathrm{vec}(\boldsymbol{F}_{t+h})'(\boldsymbol{C}_{j\cdot}\otimes \boldsymbol{I}_{k_1})\big] \\
&= (\boldsymbol{C}_i \otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{\mathrm{f,v}}(h)(\boldsymbol{C}_j\otimes \boldsymbol{I}_{k_1}).
\end{aligned}$$

Under Conditions 2-3 and by properties of Kronecker product, we have

$$\begin{aligned}
\lambda_{k_1}(\boldsymbol{M}) &= \lambda_{k_1}\left(\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\boldsymbol{R}\boldsymbol{\Omega}_{xa,ij}(h)\boldsymbol{R}'\boldsymbol{R}\boldsymbol{\Omega}_{xa,ij}(h)\boldsymbol{R}'\right) \\
&\geq \|\boldsymbol{R}\|_{\min}^4\cdot\lambda_{k_1}\left(\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\boldsymbol{\Omega}_{xa,ij}(h)\boldsymbol{\Omega}_{xa,ij}(h)'\right) \\
&= \|\boldsymbol{R}\|_{\min}^4\cdot\lambda_{k_1}\left(\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}(\boldsymbol{C}_{i\cdot}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}(h)(\boldsymbol{a}_{2,j}\otimes \boldsymbol{I}_{k_1})(\boldsymbol{C}_{j\cdot}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}'(h)(\boldsymbol{C}_{i\cdot}\otimes \boldsymbol{I}_{k_2})\right) \\
&= \|\boldsymbol{R}\|_{\min}^4\cdot\lambda_{k_1}\left(\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}(\boldsymbol{C}_{i\cdot}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}(h)(\boldsymbol{C}_{j\cdot}\boldsymbol{C}_{j\cdot}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}'(h)(\boldsymbol{C}_{i\cdot}\otimes \boldsymbol{I}_{k_1})\right) \\
&= \|\boldsymbol{R}\|_{\min}^4\cdot\lambda_{k_1}\left(\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}(\boldsymbol{C}_{i\cdot}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}(h)(\boldsymbol{C}\boldsymbol{C}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}'(h)(\boldsymbol{C}_{i\cdot}\otimes \boldsymbol{I}_{k_2})\right) \\
&= \|\boldsymbol{R}\|_{\min}^4\cdot\lambda_{k_1}\left(\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}(\boldsymbol{C}_{i\cdot}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}(h)(\boldsymbol{C}\otimes \boldsymbol{I}_{k_1})(\boldsymbol{C}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}'(h)(\boldsymbol{C}_{i\cdot}\otimes \boldsymbol{I}_{k_1})\right) \\
&= \|\boldsymbol{R}\|_{\min}^4\cdot\lambda_{k_1}\left(\sum_{h=1}^{h_0}\sum_{i=1}^{p_2}(\boldsymbol{C}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}'(h)(\boldsymbol{C}_{i\cdot}\otimes \boldsymbol{I}_{k_1})(\boldsymbol{C}_{i\cdot}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}(h)(\boldsymbol{C}\otimes \boldsymbol{I}_{k_1})\right) \\
&= \|\boldsymbol{R}\|_{\min}^4\cdot\lambda_{k_1}\left(\sum_{h=1}^{h_0}(\boldsymbol{C}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}'(h)(\boldsymbol{C}\boldsymbol{C}'\otimes \boldsymbol{I}_{k_1})\boldsymbol{\Omega}_{f,v}(h)(\boldsymbol{C}\otimes \boldsymbol{I}_{k_1})\right) \\
&= O_p(p_1^{2-2\delta_1}p_2^{2-2\delta_2}).
\end{aligned}$$

□

Proof of Theorem 1

Proof. By Lemmas 1-4, and Lemma 3 in Lam et al. (2011), Theorem 1 follows. □

Proof of Theorem 2

Proof. The proof is quite similar to that of Theorem 1 of Lam and Yao (2012). We denote $\hat{\lambda}_j$ and $\widehat{\boldsymbol{q}}_j$ for the j-th largest eigenvalue of $\widehat{\boldsymbol{M}}$ and its corresponding eigenvector, respectively.

The corresponding population eigenvalues are denoted by λ_j and $\boldsymbol{q}_j$ for the matrix $\boldsymbol{M}$. Let $\widehat{\boldsymbol{Q}}_1 = (\widehat{\boldsymbol{q}}_1, \ldots, \widehat{\boldsymbol{q}}_{k_1})$ and $\boldsymbol{Q}_1 = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k_1})$. We have

$$\lambda_j = \boldsymbol{q}_j' \boldsymbol{M} \boldsymbol{q}_j, \quad \text{and} \qquad \widehat{\lambda}_j = \widehat{\boldsymbol{q}}_j' \widehat{\boldsymbol{M}} \widehat{\boldsymbol{q}}_j, \qquad j = 1, \ldots, p_1.$$

We can decompose $\widehat{\lambda}_j - \lambda_j$ by

$$\widehat{\lambda}_j - \lambda_j = \widehat{\boldsymbol{q}}_j' \widehat{\boldsymbol{M}} \, \widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j' \boldsymbol{M} \boldsymbol{q}_j = I_1 + I_2 + I_3 + I_4 + I_5,$$

where

$$I_1 = (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)'(\widehat{\boldsymbol{M}} - \boldsymbol{M})\widehat{\boldsymbol{q}}_j, \quad I_2 = (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)' \boldsymbol{M} (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j),$$
$$I_3 = (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)' \boldsymbol{M} \boldsymbol{q}_j, \quad I_4 = \boldsymbol{q}_j'(\widehat{\boldsymbol{M}} - \boldsymbol{M})\widehat{\boldsymbol{q}}_j, \quad I_5 = \boldsymbol{q}_j' \boldsymbol{M} (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j).$$

For $j = 1, \ldots, k_1$, $\|\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j\|_2 \le \|\widehat{\boldsymbol{Q}}_1 - \boldsymbol{Q}_1\|_2 = O_p(h_T)$, where $h_T = p_1^{\delta_1} p_2^{\delta_2} T^{-1/2}$ by Theorem 1, and $\|\boldsymbol{M}\|_2 = O_p(p_1^{2-\delta_1} p_2^{2-\delta_2})$. By Lemma 4, we have $\|I_1\|_2$ and $\|I_2\|_2$ are of order $O_p(p_1^{2-2\delta_1} p_2^{2-2\delta_2} h_T^2)$ and $\|I_3\|_2$, $\|I_4\|_2$ and $\|I_5\|_2$ are of order $O_p(p_1^{2-2\delta_1} p_2^{2-2\delta_2} h_T)$. So $|\widehat{\lambda}_j - \lambda_j| = O_p(p_1^{2-2\delta_1} p_2^{2-2\delta_2} h_T) = O_p(p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2})$.
For $j = k_1 + 1, \ldots, p_1$, define,

$$\widetilde{\boldsymbol{M}} = \sum_{h=1}^{h_0} \sum_{i=1}^{p_2} \sum_{j=1}^{p_2} \widehat{\boldsymbol{\Omega}}_{i,j}(h) \boldsymbol{\Omega}_{i,j}'(h), \quad \widehat{\boldsymbol{B}}_1 = (\widehat{\boldsymbol{q}}_{k_1+1}, \ldots, \widehat{\boldsymbol{q}}_{p_1}), \quad \text{and} \quad \boldsymbol{B}_1 = (\boldsymbol{q}_{k_1+1}, \ldots, \boldsymbol{q}_{p_1}).$$

It can be shown that $\|\widehat{\boldsymbol{B}}_1 - \boldsymbol{B}_1\|_2 = O_p(h_T)$, similar to proof of Theorem 1 with Lemma 3 in Lam et al. (2011). Hence, $\|\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j\|_2 \le \|\widehat{\boldsymbol{B}}_1 - \boldsymbol{B}_1\|_2 = O_p(h_T)$.
Since $\lambda_j = 0$, for $j = k_1 + 1, \ldots, p_1$, consider the decomposition

$$\widehat{\lambda}_j = \widehat{\boldsymbol{q}}_j' \widehat{\boldsymbol{M}} \widehat{\boldsymbol{q}}_j = K_1 + K_2 + K_3,$$

where

$$K_1 = \widehat{\boldsymbol{q}}_j'(\widehat{\boldsymbol{M}} - \widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{M}}' + \boldsymbol{M})' \widehat{\boldsymbol{q}}_j, \quad K_2 = 2\widehat{\boldsymbol{q}}_j'(\widetilde{\boldsymbol{M}} - \boldsymbol{M})(\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j),$$
$$K_3 = (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)' \boldsymbol{M} (\widehat{\boldsymbol{q}}_j' - \boldsymbol{q}_j).$$

By Lemma 2 and Lemma 4,

$$\begin{aligned}
K_1 &= \sum_{h=1}^{h_0} \| \sum_{i=1}^{p_2} \sum_{j=1}^{p_2} (\widehat{\boldsymbol{\Omega}}_{i,j}(h) - \boldsymbol{\Omega}_{i,j}(h)) \widehat{\boldsymbol{q}}_j \|_2^2 \le \sum_{h=1}^{h_0} \sum_{i=1}^{p_2} \sum_{j=1}^{p_2} \|\widehat{\boldsymbol{\Omega}}_{i,j}(h) - \boldsymbol{\Omega}_{i,j}(h)\|_2^2 = O_p(p_1^2 p_2^2 T^{-1}), \\
|K_2| &= O_p(\|\widetilde{\boldsymbol{M}} - \boldsymbol{M}\|_2 \cdot \|\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j\|_2) = O_p(\|\widetilde{\boldsymbol{M}} - \boldsymbol{M}\|_2 \cdot \|\widehat{\boldsymbol{B}}_1 - \boldsymbol{B}_1\|_2) = O_p(p_1^2 p_2^2 T^{-1}), \\
|K_3| &= O_p(\|\widehat{\boldsymbol{B}}_k - \boldsymbol{B}_k\|_2^2 \cdot \|\boldsymbol{M}_k\|_2) = O_p(p_1^{2-2\delta_1} p_2^{2-2\delta_2} h_T^2) = O_p(p_1^2 p_2^2 n^{-1}).
\end{aligned}$$

Hence $\lambda_j = O_p(p_1^2 p_2^2 T^{-1})$. □

Proof of Theorem 3

Proof.

$$\begin{aligned}\hat{\boldsymbol{S}}_t - \boldsymbol{S}_t =& \hat{\boldsymbol{Q}}_1\hat{\boldsymbol{Q}}_1'\boldsymbol{X}_t\hat{\boldsymbol{Q}}_2\hat{\boldsymbol{Q}}_2' - \boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2' = \hat{\boldsymbol{Q}}_1\hat{\boldsymbol{Q}}_1'(\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2' + \boldsymbol{E}_t)\hat{\boldsymbol{Q}}_2\hat{\boldsymbol{Q}}_2' - \boldsymbol{Q}_1\boldsymbol{Q}_1'\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'\boldsymbol{Q}_2\boldsymbol{Q}_2' \\ =& \hat{\boldsymbol{Q}}_1\hat{\boldsymbol{Q}}_1'\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'(\hat{\boldsymbol{Q}}_2\hat{\boldsymbol{Q}}_2' - \boldsymbol{Q}_2\boldsymbol{Q}_2') + (\hat{\boldsymbol{Q}}_1\hat{\boldsymbol{Q}}_1' - \boldsymbol{Q}_1\boldsymbol{Q}_1')\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2' + \hat{\boldsymbol{Q}}_1\hat{\boldsymbol{Q}}_1'\boldsymbol{E}_t\hat{\boldsymbol{Q}}_2\hat{\boldsymbol{Q}}_2' \\ =& I_1 + I_2 + I_3\end{aligned}$$

By Theorem 1, it follows

$$\begin{aligned}& I_1 \le 2\|\boldsymbol{Z}_t\|_2\|\hat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2\|_2 = O_p(p_1^{1/2-\delta_1/2}p_2^{1/2-\delta_2/2}\|\hat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2\|_2) = O_p(p_1^{1/2+\delta_1/2}p_2^{1/2+\delta_2/2}T^{-1/2}), \\ & I_2 \le 2\|\hat{\boldsymbol{Q}}_1 - \boldsymbol{Q}_1\|_2\|\boldsymbol{Z}_t\|_2 = O_p(p_1^{1/2+\delta_1/2}p_2^{1/2+\delta_2/2}T^{-1/2}) \\ & I_3 = O(p_{\max}^{1/2}).\end{aligned}$$

We have $p_1^{-1/2}p_2^{-1/2}\|\hat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2 = O_p(p_1^{\delta_1/2}p_2^{\delta_2/2}T^{-1/2} + p_{\min}^{-1/2})$. □

Proof of Theorem 4

Proof. We assume that $\boldsymbol{Q}_1$ is uniquely defined as $\boldsymbol{Q}_1 = (\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_{k_1})$, where $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k_1}$ are eigenvectors of $\boldsymbol{M}$ corresponding to the largest k_1 eigenvalues $\lambda_1, \ldots, \lambda_{k_1}$, and $\lambda_1 > \lambda_2 > \ldots > \lambda_{k_1}$. Then similar to proof of Theorem 3 in Liu and Chen (2016), we can obtain the results. □