Contents lists available at ScienceDirect

Journal of Econometrics

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

Factor models for matrix-valued high-dimensional time series[☆]

Dong Wang [a], Xialu Liu [b], Rong Chen [c,*]

[a] Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, United States
[b] Management Information Systems Department, San Diego State University, San Diego, CA 92182, United States
[c] Department of Statistics, Rutgers University, Piscataway, NJ 08854, United States

ARTICLE INFO

Article history:
Available online xxxx

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. Engle, 1982; Bollerslev, 1986; Tong, 1990), and nonparametric models (e.g. Fan and Yao, 2003). Multivariate time series and panel time series, 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; Stock and Watson, 2004; Lütkepohl, 2005; Tsay, 2014, and others). Such analysis not only reveals the temporal dynamics of the time series, but also explores the relationship among a group of 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, countries routinely 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 model. 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.

☆ Chen's research was supported in part by US National Science Foundation grants IIS-1741390, DMS-1503409 and DMS-1209085.
* Corresponding author.
E-mail address: rongchen@stat.rutgers.edu (R. Chen).

https://doi.org/10.1016/j.jeconom.2018.09.013

Please cite this article in press as: Wang D., et al., Factor models for matrix-valued high-dimensional time series. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.09.013.

Table 1
Illustration of a matrix-valued time series.

	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	$X_{t,31}$	$X_{t,32}$	$\cdots$	$X_{t,3p}$
Payout Ratio	$X_{t,41}$	$X_{t,42}$	$\cdots$	$X_{t,4p}$

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 entire group of variables while fully preserve and utilize its matrix structure.

There are many other examples. Investors may be interested in the time series of a group of financials (e.g. asset/equity ratio, dividend per share, and revenue) for a group of companies, the 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 study such a matrix-valued time series.

Matrix-valued data 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.

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 aim 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 two real data examples are given in Sections 5 and 6. Section 7 provides a brief summary. All proofs are in the 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 $p_1 \times p_2$,

$$\boldsymbol{X}_t = \begin{pmatrix} X_{t,11} & \cdots & X_{t,1p_2} \\ \vdots & \ddots & \vdots \\ X_{t,p_11} & \cdots & X_{t,p_1p_2} \end{pmatrix}.$$

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 $\boldsymbol{F}_t$'s 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 a 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. We first introduce some notation. For a matrix $\boldsymbol{A}$, we use $\boldsymbol{a}_{i\cdot}$ and $\boldsymbol{a}_j$ to represent the ith row and the jth 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 shown in Table 1 and consider the first column of $\boldsymbol{X}_t$ (the US economic indicators),

$$\overset{\text{US}}{\begin{pmatrix} \text{GDP} \\ \text{Unem} \\ \text{Inf} \\ \text{PayR} \end{pmatrix}_t} = C_{11} \overset{\boldsymbol{f}_{t,1}}{\begin{pmatrix} \text{F-GDP} \\ \text{F-Unem} \\ \text{F-Inf} \\ \text{F-PayR} \end{pmatrix}_t} + \overset{\cdots}{\cdots} + C_{1k_2} \overset{\boldsymbol{f}_{t,k_2}}{\begin{pmatrix} \text{F-GDP} \\ \text{F-Unem} \\ \text{F-Inf} \\ \text{F-PayR} \end{pmatrix}_t} + \mathbf{e}_{t,US}.$$

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 $\boldsymbol{X}_t$,

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

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.

Obviously column and row interaction would be of interests and of importance. One way to introduce interaction is to assume an additive structure, by combining the column and row factor models

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

However, the number of factors in this model is large ($k_1 \times p_2 + p_1 \times k_2$). A more parsimonious model would be a direct interaction as in model (1). In this case the number of factors is only $k_1 \times k_2$.

Interpretation II: We can view the model (1) as a two-step hierarchical model.

Step 1: For each fixed row $i = 1, 2, \ldots, p_1$, 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 $\{\boldsymbol{g}_{t,i\cdot} = (G_{t,i1}, \ldots, G_{t,ik_2}),\ t = 1, 2, \ldots, T\}$ under a standard vector factor model setting. That is,

$$(X_{t,i1}, \ldots, X_{t,ip_2}) = (G_{t,i1}, \ldots, G_{t,ik_2})\boldsymbol{C}^{(i)'} + (H_{t,i1}, \ldots, H_{t,ip_2}), \quad t = 1, 2, \ldots, T.$$

Let $\boldsymbol{G}_t$ be the $p_1 \times k_2$ matrix formed with p_1 rows of $\boldsymbol{g}_{t,i\cdot}$. Also denote $\boldsymbol{H}_t$ as the $p_1 \times p_2$ error matrix formed with the rows of $\{\boldsymbol{h}_{t,i\cdot} = (H_{t,i1} \ldots, H_{t,ip_2})\}$.

Step 2: Suppose each column $j = 1, 2, \ldots, k_2$ of the assembled factor matrix $\boldsymbol{G}_t$ obtained in Step 1 also assumes the factor structure, with a $p_1 \times k_1$ loading matrix $\boldsymbol{R}^{(j)}$ and a k_1 dimensional factor $\boldsymbol{f}_{t,\cdot j}$. That is,

$$\begin{pmatrix} G_{t,1j} \\ \vdots \\ G_{t,p_1j} \end{pmatrix} = \boldsymbol{R}^{(j)} \begin{pmatrix} F_{t,1j} \\ \vdots \\ F_{t,k_1j} \end{pmatrix} + \begin{pmatrix} H^*_{t,1j} \\ \vdots \\ H^*_{t,p_1j} \end{pmatrix}, \quad t = 1, 2, \ldots, T.$$

This step reveals the common factors that drive the co-moments in $\boldsymbol{G}_t$. Let $\boldsymbol{F}_t$ be the $k_1 \times k_2$ matrix formed with the columns $\boldsymbol{f}_{t,\cdot j}$. And let $\boldsymbol{H}^*_t$ be the $p_1 \times k_2$ error matrix formed with columns $\{\boldsymbol{h}^*_{t,\cdot j} = (H^*_{t,1j}, \ldots, H^*_{t,p_1j})'\}$.

Step 3: Assembly: With the above two-step factor analysis and notation, assume $\boldsymbol{R}^{(1)} = \cdots = \boldsymbol{R}^{(k_2)} = \boldsymbol{R}$ and $\boldsymbol{C}^{(1)} = \cdots = \boldsymbol{C}^{(p_1)} = \boldsymbol{C}$, we have

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

Hence

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

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

Here we provide some additional remarks of 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

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

Please cite this article in press as: Wang D., et al., Factor models for matrix-valued high-dimensional time series. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.09.013.

where $\boldsymbol{\Phi}$ is a $p_1p_2 \times k$ 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 $k = k_1k_2$. Then model (3) is a special case of model (2), with a Kronecker product structured loading matrix. Hence model (1) is a restricted version of model (2), assuming a special structure for the loading spaces. The number of parameters for the loading matrix $\boldsymbol{\Phi}$ in model (2) is $(p_1k_1) \times (p_2k_2)$ whereas it is $p_1k_1 + p_2k_2$ for the loading matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ in model (1). Therefore, model (1) significantly reduces the dimension of the problem.

Remark 2. Interpretation II also reveals the reduction in the number of factors comparing to using factor models for each column panel or row panel. Note that, if one ignores the interconnection between the rows and obtain individual factor models for each row, as in Step 1, the total number of factors is $p_1 \times k_2$. These factors may have connections across rows. Step 2 exploits such correlations and uses another factor model to reduce the number of factors from $p_1 \times k_2$ to $k_1 \times k_2$.

Remark 3. We also observed that in practice, the total number of factors used in our model may be larger than the number of factors needed in the vectorized factor model (2). This is possible since the vectorized model simultaneously exploits common driving features in all series, while the matrix factor model does it by working on the row vectors separately first (Step 1), then condensing them by the columns (Step 2). Such a two-step approach may result in redundancy (highly correlated factors) which may be further simplified. Because we are forcing the factors to assume a neat matrix structure, it is difficult to have simplifications such as having one or several elements in the factor matrix $\boldsymbol{F}_t$ be constant zero. Since k_1 and k_2 are usually small, we will tolerate such redundancy. One extension is to assume that the factor matrix $\boldsymbol{F}_t$, after a certain rotation, has a block diagonal structure, resulting in a multi-term factor model

$$\boldsymbol{X}_t = \sum_{i=1}^{s} \boldsymbol{R}_i \boldsymbol{F}_{it} \boldsymbol{C}_i' + \boldsymbol{E}_t, \quad t = 1, 2, \ldots, T, \tag{4}$$

where $\boldsymbol{F}_{it}$ is a $k_{i1} \times k_{i2}$ factor matrix, and $\sum_{i=1}^{s} k_{i1} = k_1$ and $\sum_{i=1}^{s} k_{i2} = k_2$. This will reduce the number of factors from $k_1 \times k_2$ to $\sum_{i=1}^{s} k_{i1}k_{i2}$, with corresponding dimension reduction in the loading matrices as well. We are currently investigating the properties and estimation procedures of such a multi-term factor model.

Remark 4. As in all factor model setting, the properties or assumptions on the observed process $\boldsymbol{X}_t$ are inferred from the assumptions on the factors and the noise processes, since the observed series are assumed to be linear combinations of the factor processes plus the noise process. Indirectly, we assume that all autocovariance matrices of lag $h \geq 1$ of all series lie in a structured $k_1k_2 \times k_1k_2$ space, but no assumption on the contemporary covariance matrix, as we do not assume any contemporary covariance structure on the error $\boldsymbol{E}_t$.

Remark 5. 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); 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, and establish their convergence rates. 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 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 the 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 $k_1 \times k_1$ and $k_2 \times k_2$. Then the triplets $(\boldsymbol{R}, \boldsymbol{F}_t, \boldsymbol{C})$ 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); 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 $\mathcal{M}(\boldsymbol{C})$, and referred to as row factor loading space and column factor loading space, respectively.

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 $k_i \times k_i$ non-singular 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 $\boldsymbol{C}$ is equivalent to the estimation of column spaces of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$.

Write

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

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

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

Eq. (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); 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 noise 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 estimated in a similar way using the transposes of $\boldsymbol{X}_t$'s.

Let the jth column of $\boldsymbol{X}_t, \boldsymbol{R}, \boldsymbol{C}, \boldsymbol{Q}_i$ and $\boldsymbol{E}_t$ be $\boldsymbol{x}_{t,j}, \boldsymbol{r}_j, \boldsymbol{c}_j, \boldsymbol{q}_{i,j}$ 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 kth 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) and by the assumption that $\boldsymbol{E}_t$ is white, it follows that

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

for $h \geq 1$. For a pre-determined integer h_0, define

$$\boldsymbol{M}_1 = \sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2}\boldsymbol{\Omega}_{x,ij}(h)\boldsymbol{\Omega}_{x,ij}'(h). \tag{10}$$

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

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

Suppose the matrix $\boldsymbol{M}_1$ has rank k_1 (Condition 5 in Section 3.2). From (11), we can see that each column of $\boldsymbol{M}_1$ is a linear combination of columns of $\boldsymbol{Q}_1$, and thus the matrices $\boldsymbol{M}_1$ and $\boldsymbol{Q}_1$ have the same column spaces, that is, $\mathcal{M}(\boldsymbol{M}_1) = \mathcal{M}(\boldsymbol{Q}_1)$. It follows that the eigen-space of $\boldsymbol{M}_1$ 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 $\boldsymbol{M}_1$. Assume that $\boldsymbol{M}_1$ has k_1 distinct nonzero eigenvalues, and let $\boldsymbol{q}_{1,j}$ be the unit eigenvector corresponding to the jth largest eigenvalue. As there are two unit eigenvectors corresponding to each eigenvalue, we use the one with positive $\boldsymbol{1}'\boldsymbol{q}_{1,j}$. We can now uniquely define $\boldsymbol{Q}_1$ by

$$\boldsymbol{Q}_1 = (\boldsymbol{q}_{1,1}, \boldsymbol{q}_{1,2}, \ldots, \boldsymbol{q}_{1,k_1}).$$

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

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

$$\widehat{\boldsymbol{M}}_1 = \sum_{h=1}^{h_0}\sum_{i=1}^{p_2}\sum_{j=1}^{p_2} \widehat{\boldsymbol{\Omega}}_{x,ij}(h)\widehat{\boldsymbol{\Omega}}'_{x,ij}(h). \tag{13}$$

Then, $\mathcal{M}(\boldsymbol{Q}_1)$ can be estimated by $\mathcal{M}(\widehat{\boldsymbol{Q}}_1)$, where $\widehat{\boldsymbol{Q}}_1 = \{\widehat{\boldsymbol{q}}_{1,1},\ldots,\widehat{\boldsymbol{q}}_{1,k_1}\}$, and $\widehat{\boldsymbol{q}}_{1,1},\ldots\widehat{\boldsymbol{q}}_{1,k_1}$ are the eigenvectors of $\widehat{\boldsymbol{M}}_1$ 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 $\widehat{\lambda}_{1,1} \geq \widehat{\lambda}_{1,2} \geq \cdots \geq \widehat{\lambda}_{1,p_1} \geq 0$ be the ordered eigenvalues of $\widehat{\boldsymbol{M}}_1$. Then

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

For $\boldsymbol{Q}_2$ and k_2, they can be estimated by performing the same procedure on the transposes of $\boldsymbol{X}_t$'s to construct $\boldsymbol{M}_2$ and $\widehat{\boldsymbol{M}}_2$. 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 the properties of Kronecker product, it follows that

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

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,

$$\widehat{\boldsymbol{S}}_t = \widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}'_1\boldsymbol{X}_t\widehat{\boldsymbol{Q}}_2\widehat{\boldsymbol{Q}}'_2. \tag{14}$$

Remark 6. Theoretically any h_0 can be used to estimate the loading spaces, as long as one of the $\boldsymbol{\Omega}_{x,ij}(h)$ is of full rank for $i,j = 1,\ldots,p_2$ and $h = 1,\ldots,h_0$. Although they converge at the same rate, the estimate from the lag where the autocorrelation maximizes is most efficient. We demonstrate the impact of h_0 in the matrix factor model setting in Section 4. As the autocorrelation is often at its strongest at small time lags, a relatively small h_0 is usually adopted (Lam et al., 2011; Chang et al., 2015; Liu and Chen, 2016). Larger h_0 strengthens the signal, but also adds more noises in the estimation of $\boldsymbol{M}_i$.

Remark 7. K-fold cross-validation procedures can be adopted for model selection between matrix-valued factor models in (1) and vector-valued factor models in (2), and among the models with different number of factors. Specifically, we first partition the data D into k subsets $D_1,\ldots,D_k$, and fit a factor model with each of the $D\backslash D_k$ sets. Then we use the estimated loading spaces, together with the data in D_k to obtain the dynamic signal process $\boldsymbol{S}_t$ for D_k, and obtain out-of-sample residuals. Residual sum of squares (RSS) of the K folds is then adopted for model comparison. Rolling-validation which uses only the data before the block for estimation can be used as well.

3.2. *Theoretical properties of the estimators*

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 rank($\boldsymbol{Y}$), $\|\boldsymbol{Y}\|_2$, $\|\boldsymbol{Y}\|_F$, $\|\boldsymbol{Y}\|_{\min}$, and $\sigma_j(\boldsymbol{Y})$ to denote the rank, the spectral norm, the Frobenius norm, the smallest nonzero singular value and the jth largest singular value of $\boldsymbol{Y}$. When $\boldsymbol{Y}$ is a square matrix, we denote by 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

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

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

Condition 1. The vector-valued process vec($\boldsymbol{F}_t$) is α-mixing. Specifically, 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 $\{\text{vec}(\boldsymbol{F}_t) : i \leq t \leq j\}$.

Condition 2. Let $F_{t,ij}$ be the ij-th entry of $\boldsymbol{F}_t$. For any $i = 1, \dots, k_1$, $j = 1, \dots, k_2$, and $t = 1, \dots, T$, we assume that $\mathrm{E}(|F_{t,ij}|^{2\gamma}) \le C$, where C is a positive constant, and γ is given in Condition 1. In addition, there exists a $1 \le h \le h_0$ such that $\mathrm{rank}(\boldsymbol{\Sigma}_f(h)) \ge k$, and $\|\boldsymbol{\Sigma}_f(h)\|_2 \asymp O(1) \asymp \sigma_k(\boldsymbol{\Sigma}_f(h))$, where $k = \max\{k_1, k_2\}$, as p_1 and p_2 go to infinity and k_1 and k_2 are fixed. For $i = 1, \dots, k_1$ and $j = 1, \dots, k_2$, $\frac{1}{T-h}\sum_{t=1}^{T-h}\mathrm{Cov}(\boldsymbol{f}_{t,i\cdot}, \boldsymbol{f}_{t+h,i\cdot}) \ne \boldsymbol{0}$, $\frac{1}{T-h}\sum_{t=1}^{T-h}\mathrm{Cov}(\boldsymbol{f}_{t,\cdot j}, \boldsymbol{f}_{t+h,\cdot j}) \ne \boldsymbol{0}$.

The latent process does not have to be stationary, but needs to satisfy the mixing condition (Condition 1) and boundedness condition (Condition 2). They are weaker than stationarity. For example, when a process has a deterministic seasonal variance component or with a deterministic regime switching mechanism, it is not stationary but mixing. 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.

Under Condition 2, $\boldsymbol{\Sigma}_f(h)$ may not be of full rank, which indicates that it is allowed to involve some extent of redundancy in the factors. Condition 2 also guarantees that there is no redundant row or column in $\boldsymbol{F}_t$, and in each row or column there is at least one factor which has serial dependence at lag h. The greater dimension reduction can be achieved by a multi-term factor model in (4). We are currently investigating the properties and estimation procedures of such a multi-term factor model.

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 $\|\boldsymbol{C}\|_2^2 \asymp p_2^{1-\delta_2} \asymp \|\boldsymbol{C}\|_{\min}^2$, as p_1 and p_2 go to infinity and k_1 and k_2 are fixed.

The rates δ_1 and δ_2 are called the strength for row factors and the strength for column factors, respectively. They measure the relative growth rate of the amount of information carried by the observed process $\boldsymbol{X}_t$ on the 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 $\boldsymbol{X}_t$ on the factors grows more slowly than the noises introduced as p_i increases. For detailed discussion of factor strength, see Lam and Yao (2012).

Condition 5. $\boldsymbol{M}_i$ has k_i distinct positive eigenvalues for $i = 1,2$.

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 $\boldsymbol{M}_i$ are distinct, we can uniquely define $\boldsymbol{Q}_i$ as $\boldsymbol{Q}_i = \{\boldsymbol{q}_{i,1}, \dots, \boldsymbol{q}_{i,k_i}\}$, where $\boldsymbol{q}_{i,1}, \dots, \boldsymbol{q}_{i,k_i}$ are the unit eigenvectors of $\boldsymbol{M}_i$ corresponding to its k_i largest eigenvalues $\{\lambda_{i,1} > \lambda_{i,2} \dots > \lambda_{i,k_i}\}$ which make $\boldsymbol{1}'\boldsymbol{q}_{i,1}, \boldsymbol{1}'\boldsymbol{q}_{i,2}, \dots,$ and $\boldsymbol{1}'\boldsymbol{q}_{i,k_i}$ all positive, for $i = 1,2$.

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*

$$\|\widehat{\boldsymbol{Q}}_i - \boldsymbol{Q}_i\|_2 = O_p(p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}), \ \text{for } i = 1, 2.$$

Concerning 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.

When p_1 and p_2 are fixed, the convergence rate for estimating the loading matrices are $\sqrt{T}$. If the loadings are strong ($\delta_i = 0$), the rate is also $\sqrt{T}$, since the signal is as strong as the noise, and the increase in dimensions will not affect the estimation of the loading spaces. When δ_i's are not 0, the noise increases faster than useful information. In this case, increases in dimension will dilute the information, resulting in less efficient estimators.

Theorem 2. *With Conditions 1–5 and $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2} = o(1)$, the eigenvalues $\{\widehat{\lambda}_{i,1}, \dots, \widehat{\lambda}_{i,p_i}\}$ of $\widehat{\boldsymbol{M}}_i$ which are sorted in descending order satisfy*

$$|\widehat{\lambda}_{i,j} - \lambda_{i,j}| = O_p(p_1^{2-\delta_1}p_2^{2-\delta_2}T^{-1/2}), \quad \text{for } j = 1, 2, \dots, k_i,$$

and

$$|\widehat{\lambda}_{i,j}| = O_p(p_1^2p_2^2T^{-1}), \quad \text{for } j = k_i + 1, \dots, p_i,$$

where $\lambda_{i,1} > \lambda_{i,2} \dots > \lambda_{i,k_i}$ are eigenvalues of $\boldsymbol{M}_i$, for $i = 1,2$.

Theorem 2 shows that the estimators for nonzero eigenvalues of $\boldsymbol{M}_i$ converge more slowly than those for the zero eigenvalues. It provides the theoretical support for the ratio estimator proposed in Section 3.1.

The following theorem demonstrates the theoretical properties of the estimator $\widehat{\boldsymbol{S}}_t$ in (14).

Theorem 3. *If Conditions 1–5 hold, $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$, and $\|\boldsymbol{\Sigma}_e\|_2$ is bounded, we have*

$$p_1^{-1/2} p_2^{-1/2} \|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2 = O_p(a_1 \|\widehat{\boldsymbol{Q}}_1 - \boldsymbol{Q}_1\|_2) + O_p(a_2 \|\widehat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2\|_2) + O_p(p_1^{-1/2} p_2^{-1/2})$$
$$= O_p(p_1^{\delta_1/2} p_2^{\delta_2/2} T^{-1/2} + p_1^{-1/2} p_2^{-1/2}),$$

where $a_1 \asymp a_2 \asymp O(p_1^{-\delta_1/2} p_2^{-\delta_2/2} T^{-1/2})$.

The theorem shows that, in order to estimate the signal $\boldsymbol{S}_t$ consistently, dimensions p_1 and p_2 must go to infinity, in order to have sufficient information on $\boldsymbol{S}_t$ at each time point t.

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.

4. Simulation

In this section, we study the numerical performance of the proposed matrix-valued 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 $k_1 k_2$ independent processes with noise $N(0, 1)$ where the types and coefficients of the processes will be specified later. 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 $\boldsymbol{0}$ and a Kronecker product covariance structure, that is, $\mathrm{Cov}(\mathrm{vec}(\boldsymbol{E}_t)) = \boldsymbol{\Gamma}_2 \otimes \boldsymbol{\Gamma}_1$, 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. For all simulations, the reported results are based on 200 simulation runs.

We first study the performance of our proposed approach on estimating the loading spaces. In this part, the $k_1 k_2 = 6$ latent factors are independent AR(1) processes with the AR coefficients $[-0.5\ 0.6;\ 0.8\ -0.4;\ 0.7\ 0.3]$. We consider three pairs of (δ_1, δ_2) combinations: $(0.5, 0.5)$, $(0.5, 0)$ and $(0, 0)$. For each pair of δ_1 and δ_2, the two dimensions (p_1, p_2) are chosen to be $(20, 20)$, $(20, 50)$ and $(50, 50)$. The sample size T is selected as $0.5 p_1 p_2$, $p_1 p_2$, and $2 p_1 p_2$. We take $h_0 = 1$ since it is sufficient for AR(1) model as will be shown later.

Table 2 shows the results for estimating the loading spaces $\mathcal{M}(\boldsymbol{Q}_1)$ and $\mathcal{M}(\boldsymbol{Q}_2)$. The accuracies are measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$ and $\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$ using the correct k_1 and k_2, respectively. The results show that with stronger signals and more data sample points, the approach increases the estimation accuracies. Moreover, increasing the strength of one loading matrix can improve the estimation accuracies for both loading spaces.

With the same simulated data, we compare the proposed matrix-valued approach and the vector-valued approach in Lam and Yao (2012) through the estimation accuracy of the total loading matrix $\boldsymbol{Q} = \boldsymbol{Q}_2 \otimes \boldsymbol{Q}_1$. In what follows, the subscripts mat and vec denote our approach and Lam and Yao (2012)'s method, respectively. The loading space $\widehat{\boldsymbol{Q}}_{\mathrm{mat}}$ is computed as $\widehat{\boldsymbol{Q}}_{\mathrm{mat}} = \widehat{\boldsymbol{Q}}_2 \otimes \widehat{\boldsymbol{Q}}_1$ once we obtain estimates of $\widehat{\boldsymbol{Q}}_1$ and $\widehat{\boldsymbol{Q}}_2$ through our approach. For the vector-valued approach, we apply Lam and Yao (2012)'s method to the observations $\{\mathrm{vec}(\boldsymbol{X}_t), t = 1, 2, \ldots, T\}$ to obtain $\widehat{\boldsymbol{Q}}_{\mathrm{vec}}$. Table 3 presents the results for the estimation accuracies of $\boldsymbol{Q}$ measured by $\mathcal{D}_{\mathrm{vec}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ and $\mathcal{D}_{\mathrm{mat}}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$. It shows that the matrix approach efficiently improves the estimation accuracy over the vector-valued approach.

We next demonstrate the performance of the matrix-valued approach on estimating the number of factors, k_1 and k_2. The data are the same as the data in Table 2 with $\delta_1 = \delta_2 = 0$ and hence the true rank pair is $(3, 2)$. Table 4 shows the relative frequencies of estimated rank pairs over 200 simulation runs. The four pairs $(2, 1)$, $(2, 2)$, $(3, 1)$ and $(3, 2)$ have high appearances in all of the combinations of p_1, p_2 and T. The row for the true rank pair $(3, 2)$ is highlighted. It shows that the relative frequency of correctly estimating the true rank pair improves with increasing sample size T. Table 5 shows a comparison between the matrix and vector-valued approaches on estimating the total number of latent factors $k = k_1 k_2$. The column with the true rank $k = 6$ is highlighted. The results show that the two approaches have similar performance when the sample size T is large. For smaller T, the probability of the matrix-valued approach to select the rank pair $(3, 1)$ is high and hence the frequency of estimating the true rank k decreases.

We now study the effects of the lag parameter h_0. The $k_1 k_2$ factors are assumed to be independent and follow the same model which is either an AR(1) or an MA(2) model. For the AR(1) model, the coefficients of all the factors are 0.9, 0.6 or

Table 2
Means and standard deviations (in parentheses) of $\mathcal{D}(\widehat{\mathbf{Q}}_i, \mathbf{Q}_i)$, $i = 1, 2$, over 200 simulation runs. For ease of presentation, all numbers in this table are the true numbers multiplied by 10.

				$T = .5 * p_1 * p_2$		$T = p_1 * p_2$		$T = 2 * p_1 * p_2$	
δ_1	δ_2	p_1	p_2	$\mathcal{D}(\widehat{\mathbf{Q}}_1, \mathbf{Q}_1)$	$\mathcal{D}(\widehat{\mathbf{Q}}_2, \mathbf{Q}_2)$	$\mathcal{D}(\widehat{\mathbf{Q}}_1, \mathbf{Q}_1)$	$\mathcal{D}(\widehat{\mathbf{Q}}_2, \mathbf{Q}_2)$	$\mathcal{D}(\widehat{\mathbf{Q}}_1, \mathbf{Q}_1)$	$\mathcal{D}(\widehat{\mathbf{Q}}_2, \mathbf{Q}_2)$
0.5	0.5	20	20	5.96(0.19)	7.12(0.03)	5.80(0.07)	7.09(0.01)	5.73(0.04)	7.08(0.01)
		20	50	5.87(0.15)	7.07(0.02)	5.77(0.04)	7.05(0.01)	5.74(0.02)	7.04(0.01)
		50	50	6.26(0.56)	7.05(0.01)	5.73(0.13)	7.04(0.00)	5.61(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)	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)	0.54(0.18)	0.53(0.17)
		50	50	3.68(0.48)	3.44(1.23)	1.31(0.20)	0.65(0.19)	0.51(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)	0.24(0.04)	0.22(0.04)
		20	50	0.25(0.06)	0.36(0.07)	0.16(0.03)	0.26(0.05)	0.10(0.02)	0.18(0.03)
		50	50	0.13(0.02)	0.12(0.02)	0.09(0.01)	0.08(0.01)	0.06(0.01)	0.06(0.01)

Table 3
Means and standard deviations (in parentheses) of $\mathcal{D}(\widehat{\mathbf{Q}}, \mathbf{Q})$ over 200 replicates. For ease of presentation, all numbers are the true numbers multiplied by 10.

				$T = .5 * p_1 * p_2$		$T = p_1 * p_2$		$T = 2 * p_1 * p_2$	
δ_1	δ_2	p_1	p_2	$\mathcal{D}_{\text{vec}}(\widehat{\mathbf{Q}}, \mathbf{Q})$	$\mathcal{D}_{\text{mat}}(\widehat{\mathbf{Q}}, \mathbf{Q})$	$\mathcal{D}_{\text{vec}}(\widehat{\mathbf{Q}}, \mathbf{Q})$	$\mathcal{D}_{\text{mat}}(\widehat{\mathbf{Q}}, \mathbf{Q})$	$\mathcal{D}_{\text{vec}}(\widehat{\mathbf{Q}}, \mathbf{Q})$	$\mathcal{D}_{\text{mat}}(\widehat{\mathbf{Q}}, \mathbf{Q})$
0.5	0.5	20	20	8.75(0.17)	8.26(0.07)	8.24(0.18)	8.19(0.03)	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)	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)	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)	4.37(0.45)	1.64(0.72)
		20	50	5.64(0.24)	6.75(1.13)	4.75(0.35)	2.30(0.80)	3.37(0.45)	0.78(0.20)
		50	50	5.07(0.10)	4.92(0.94)	4.46(0.29)	1.47(0.23)	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)	2.07(0.13)	0.33(0.04)
		20	50	2.84(0.18)	0.44(0.07)	2.13(0.10)	0.30(0.05)	1.56(0.07)	0.21(0.03)
		50	50	1.85(0.10)	0.18(0.02)	1.34(0.06)	0.12(0.01)	0.97(0.04)	0.09(0.01)

Table 4
Relative frequency of estimated rank pair $(\hat{k}_1, \hat{k}_2)$ over 200 runs. The row with the true rank pair (3, 2) is highlighted. Here $p = p_1 p_2$.

	$p_1 = 20,\ p_2 = 20$			$p_1 = 20,\ p_2 = 50$			$p_1 = 50,\ p_2 = 50$		
$(\hat{k}_1, \hat{k}_2)$	$T = .5p$	$T = p$	$T = 2p$	$T = .5p$	$T = p$	$T = 2p$	$T = .5p$	$T = p$	$T = 2p$
(2,1)	0.2	0.055	0	0.32	0.005	0	0	0	0
(2,2)	0.055	0.04	0	0.025	0.005	0	0	0	0
(3,1)	0.19	0.215	0.01	0.47	0.325	0.005	0.005	0	0
(3,2)	0.365	0.66	0.985	0.17	0.665	0.995	0.995	1	1
Others	0.19	0.03	0.005	0.015	0	0	0	0	0

Table 5
Relative frequency of estimated total rank $\hat{k}$ over 200 replicates for both the vector and matrix-valued approaches. The column with the true rank $k = 6$ is highlighted. Here $p = p_1 p_2$.

			$\hat{k} = 1$		$\hat{k} = 2$		$\hat{k} = 3$		$\hat{k} = 4$		$\hat{k} = 6$		Others	
p_1	p_2	T	vec	mat	vec	mat	vec	mat	vec	mat	vec	mat	vec	mat
20	20	$.5p$	0.25	0.125	0.33	0.22	0.035	0.19	0.015	0.07	0.345	0.365	0.025	0.03
		p	0.055	0.02	0.105	0.06	0	0.215	0	0.045	0.83	0.66	0.01	0
		$2p$	0	0	0.005	0	0	0.01	0	0.005	0.995	0.985	0	0
20	50	$.5p$	0.03	0.015	0.62	0.32	0	0.47	0	0.025	0.34	0.17	0.01	0
		p	0	0	0.14	0.005	0	0.325	0	0.005	0.86	0.665	0	0
		$2p$	0	0	0	0	0	0.005	0	0	1	0.995	0	0
50	50	$.5p$	0.07	0	0	0	0	0.005	0	0	0.93	0.995	0	0
		p	0	0	0	0	0	0	0	0	1	1	0	0
		$2p$	0	0	0	0	0	0	0	0	1	1	0	0

0.3. For the MA(2) model, we consider the case $f_t = e_t + 0.9e_{t-2}$. We take $\delta_1 = \delta_2 = 0$, $T = p_1 p_2$ and compare the estimation accuracies of the two loading spaces for four lag choices, $h_0 = 1, 2, 3, 4$. Table 6 shows the results of $\mathcal{D}(\widehat{\mathbf{Q}}_1, \mathbf{Q}_1)$

Table 6
Means and standard deviations (in parentheses) of $\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$ and $\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$ for different lag parameter h_0. All numbers are the true numbers multiplied by 10.

			$h_0 = 1$		$h_0 = 2$		$h_0 = 3$		$h_0 = 4$	
AR(1)	p_1	p_2	$\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$	$\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$
0.9	20	20	0.13(0.02)	0.10(0.02)	0.13(0.02)	0.10(0.02)	0.14(0.02)	0.10(0.02)	0.14(0.02)	0.11(0.02)
	20	50	0.05(0.01)	0.07(0.01)	0.05(0.01)	0.07(0.01)	0.05(0.01)	0.07(0.01)	0.06(0.01)	0.08(0.01)
	50	50	0.03(0.00)	0.03(0.00)	0.03(0.00)	0.03(0.00)	0.03(0.00)	0.03(0.00)	0.03(0.00)	0.03(0.00)
0.6	20	20	0.36(0.07)	0.26(0.04)	0.41(0.08)	0.27(0.05)	0.47(0.10)	0.28(0.05)	0.53(0.12)	0.29(0.05)
	20	50	0.15(0.03)	0.19(0.02)	0.18(0.04)	0.20(0.03)	0.21(0.05)	0.21(0.03)	0.24(0.06)	0.21(0.03)
	50	50	0.09(0.01)	0.07(0.01)	0.10(0.01)	0.08(0.01)	0.11(0.02)	0.08(0.01)	0.12(0.02)	0.08(0.01)
0.3	20	20	1.56(0.72)	0.57(0.13)	2.31(0.99)	0.60(0.16)	2.82(1.04)	0.63(0.18)	3.12(1.05)	0.67(0.18)
	20	50	0.64(0.21)	0.45(0.10)	1.14(0.49)	0.49(0.13)	1.66(0.68)	0.55(0.19)	2.13(0.82)	0.63(0.24)
	50	50	0.26(0.05)	0.17(0.03)	0.39(0.08)	0.18(0.03)	0.56(0.11)	0.20(0.04)	0.74(0.14)	0.21(0.04)
MA(2)	20	20	2.60(1.11)	0.88(0.28)	0.48(0.12)	0.27(0.05)	0.59(0.15)	0.28(0.05)	0.68(0.17)	0.28(0.06)
	20	50	2.76(1.16)	1.13(0.56)	0.21(0.04)	0.21(0.03)	0.27(0.06)	0.22(0.04)	0.32(0.07)	0.22(0.04)
	50	50	2.85(1.15)	0.68(0.23)	0.11(0.02)	0.08(0.01)	0.13(0.02)	0.08(0.01)	0.15(0.02)	0.08(0.01)

Table 7
Means and standard deviations (in parentheses) of estimation accuracies of $\boldsymbol{Q}$ and $\boldsymbol{S}$. All numbers are the true numbers multiplied by 10.

$p_1 = p_2$	T	$\mathcal{D}_{vec}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{mat}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	$\mathcal{D}_{vec}(\widehat{\boldsymbol{S}}, \boldsymbol{S})$	$\mathcal{D}_{mat}(\widehat{\boldsymbol{S}}, \boldsymbol{S})$
10	50	6.26(0.38)	3.66(0.92)	4.05(0.28)	3.41(0.39)
	200	4.11(0.33)	1.42(0.42)	3.02(0.19)	2.62(0.15)
	1000	2.12(0.13)	0.50(0.09)	2.48(0.05)	2.40(0.04)
	5000	0.99(0.05)	0.21(0.03)	2.38(0.02)	2.36(0.01)
20	50	5.65(0.39)	2.36(1.26)	3.07(0.27)	1.86(0.61)
	200	3.64(0.23)	0.71(0.16)	1.96(0.16)	1.11(0.05)
	1000	1.88(0.11)	0.29(0.04)	1.25(0.05)	1.02(0.01)
	5000	0.87(0.03)	0.13(0.01)	1.04(0.01)	1.00(0.00)
50	50	5.81(0.35)	3.17(1.47)	2.64(0.26)	1.95(0.79)
	200	3.78(0.24)	0.62(0.19)	1.57(0.16)	0.56(0.09)
	1000	2.04(0.10)	0.21(0.03)	0.82(0.06)	0.42(0.01)
	5000	0.97(0.04)	0.09(0.01)	0.49(0.02)	0.40(0.00)

and $\mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$. It is seen that, for the AR(1) processes, taking $h_0 = 1$ is sufficient. Larger h_0 in fact decreases the performance, especially for small AR coefficient cases. Note that larger h_0 increases the signal strength in the matrix $\boldsymbol{M}$, but also increases the noise level in its sample version $\hat{\boldsymbol{M}}$. For an AR(1) model with small AR coefficient, the autocorrelation in higher lags is relatively small hence the additional signal strength is limited. For the MA(2) process, one must use $h_0 \geq 2$ since lag 1 autocovariance matrix is zero and does not provide any information. $h_0 = 2$ performs the best, since all higher lags carry no additional information, but add significant amount of noise.

Next we study the performance of recovering the signal $\boldsymbol{S}_t$. The latent factors are simulated in the same way as the data in Table 2. We take $\delta_1 = \delta_2 = 0$, $p_1 = p_2 = 10, 20, 50$, and $T = 50, 200, 1000, 5000$. The recovery accuracy of $\widehat{\boldsymbol{S}}_t$, denoted by $\mathcal{D}(\widehat{\boldsymbol{S}}, \boldsymbol{S})$, is estimated by the average of $\|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2$ for $t = 1, 2, \ldots, T$ further normalized by $\sqrt{p_1 p_2}$, that is, $\mathcal{D}(\widehat{\boldsymbol{S}}, \boldsymbol{S}) = p_1^{-1/2} p_2^{-1/2} \left(\sum_{t=1}^{T} \|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2 / T\right)$. Table 7 presents the results of $\mathcal{D}(\widehat{\boldsymbol{S}}, \boldsymbol{S})$ and $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ for the two approaches. It shows that, when T is relatively large (hence the $\boldsymbol{Q}$ is estimated relatively accurately), increasing p improves the estimation of $\boldsymbol{S}$. For the same p and relatively large T, further increasing T has a limited benefit in improving the estimation of $\boldsymbol{S}$. The estimation accuracies of $\boldsymbol{S}$ of the proposed matrix-valued approach are better than that of the vector-valued approach, though the relative improvement decreases as T increases, even the improvement of estimating $\boldsymbol{Q}$ is significant.

Next, we conduct a 10-fold cross-validation study. The data are generated in the same way as the data in Table 2 with $\delta_1 = \delta_2 = 0$, $p_1 = p_2 = 20$ and $T = 1000$. We vary the estimated number of factors k_1 and k_2 from all combinations of $k_1 = 1, 2, 3, 4$ and $k_2 = 1, 2, 3$. The means of the out-of-sample RSS/SST are reported in Table 8. For the matrix-valued approach, the RSS/SST decreases rapidly when k_1 and k_2 increase, before they reach the true rank pair $(3, 2)$ (highlighted in the table). Then the RSS/SST value remain roughly the same with increasing estimated ranks when $k_1 > 3$ and $k_2 > 2$. For the vector-valued approach, $k = k_1 k_2$, hence the values in the table are the same for the same $k_1 k_2$ value (e.g. $(2, 3)$ and $(3, 2)$ are equivalent). Its performance improves quickly as k increases until $k = 6$, the true number of factors. Then the performance remains relatively the same for $\hat{k} > 6$.

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 is grouped into 100 portfolios, according to ten levels of market capital (size) and ten levels of book to equity ratio (BE). Their

Table 8
Means of out-of-sample RSS/SST for 10-fold cross-validation over 200 simulation runs. The cell corresponding to the true order (3, 2) is highlighted.

	$\hat{k}_2 = 1$		$\hat{k}_2 = 2$		$\hat{k}_2 = 3$	
$\hat{k}_1$	vec	mat	vec	mat	vec	mat
1	0.83	0.83	0.72	0.75	0.63	0.75
2	0.72	0.71	0.58	0.56	0.49	0.55
3	0.63	0.67	0.49	0.47	0.46	0.46
4	0.58	0.66	0.46	0.47	0.44	0.43

Table 9
Fama–French series: Size loading matrix after rotation and scaling.

Factor	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1	-13	-14	-13	-13	-10	-5	-2	1	6	7
2	0	0	-2	3	5	12	12	18	15	5

Table 10
Fama–French series: BtoE loading matrix after rotation and scaling.

Factor	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8	BE9	BE10
1	-21	-14	-11	-9	-4	-1	-1	-4	1	3
2	-9	2	3	7	9	10	10	10	13	14

monthly returns from January 1964 to December, 2015 for total 624 months and overall 62,400 observations are used in this analysis. For more detailed information, see http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

All the 100 series are clearly related to the overall market condition. In this analysis we simply subtract the corresponding monthly excess market return from each of the series, resulting in 100 market-adjusted return series. We chose not to fit a standard CAPM model to each of the series to remove the market effect, as it will involve estimating 100 different betas. The market return data are obtained from the same website above.

Fig. 1 shows the time series plot of the 100 series (standardized), and Figs. 2 and 3 show the logarithms and ratios of eigenvalues of $\widehat{\boldsymbol{M}}_1$ and $\widehat{\boldsymbol{M}}_2$ for the row (size) and column (BE) loading matrices. Since the series shows very small autocorrelation beyond $h = 1$, in this example we use $h_0 = 1$. The results by using $h_0 = 2$ are similar. Although the eigenvalue ratio estimate presented in Section 3.1 indicates $k_1 = k_2 = 1$, we use $k_1 = k_2 = 2$ here for illustration. Tables 9 and 10 show the estimated loading matrices after a varimax rotation that maximizes the variance of the squared factor loadings, scaled by 30 for a cleaner view. For size, it is seen that there are possibly two or three groups. The 1-st to 5-th smallest size portfolios load heavily (with roughly equal weights) on the first row of the factor matrix, while the 6-th to 9-th smallest size portfolios load heavily (with roughly equal weights) on the second row of the factor matrix. The largest (10-th) size portfolio behaves similar to the other larger size portfolios, but with some differences. 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 the 8-th to 10-th size portfolio) and the smallest 30% of the companies (combining our 1st to 3rd size portfolio).

Turning to the book to equity ratio, Table 10 shows a different pattern in the column loading matrix. There seem to have three groups. The smallest 2-nd to 4-th BE portfolios load heavily on the first column of the factor matrix; the 5th to 10th BE portfolios load heavily on the second columns of the factor matrix. The smaller (1st) BE portfolios load heavily on both columns of the factor matrix, with different loading coefficients.

Fig. 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, factor trading and other usage, though further analysis is needed to assess their effectiveness. Cross-correlation study shows that there are not many significant cross-correlation of lag larger than 0 among the factors, though the factors show some strong contemporary correlation. Note that the factor matrices are subject to rotation — in our case we performed rotation to reveal the group structure in the loading matrices. A principle component analysis of the four factor series reveals that three principle components can explain 98% of the variation in the four factors, hence there may still be some redundancy in the factors and the model may be further simplified.

Fig. 5 shows the logarithms and ratios of eigenvalues of $\widehat{\boldsymbol{M}}$ in Lam et al. (2011) for a vectorized factor model (2). Models with various number of factors were estimated and a comparison is shown in Table 11 using a version of rolling-validation. Specifically, for each year between 1996 to 2015, we use all data available before the year to fit a matrix (or vector) factor model and estimate the corresponding loading matrices. Using these estimated loading matrices and the observed 12 months of the data in the year, we estimate the factors and the corresponding residuals. Total sum of squares of the 12 residuals of

Please cite this article in press as: Wang D., et al., Factor models for matrix-valued high-dimensional time series. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.09.013.

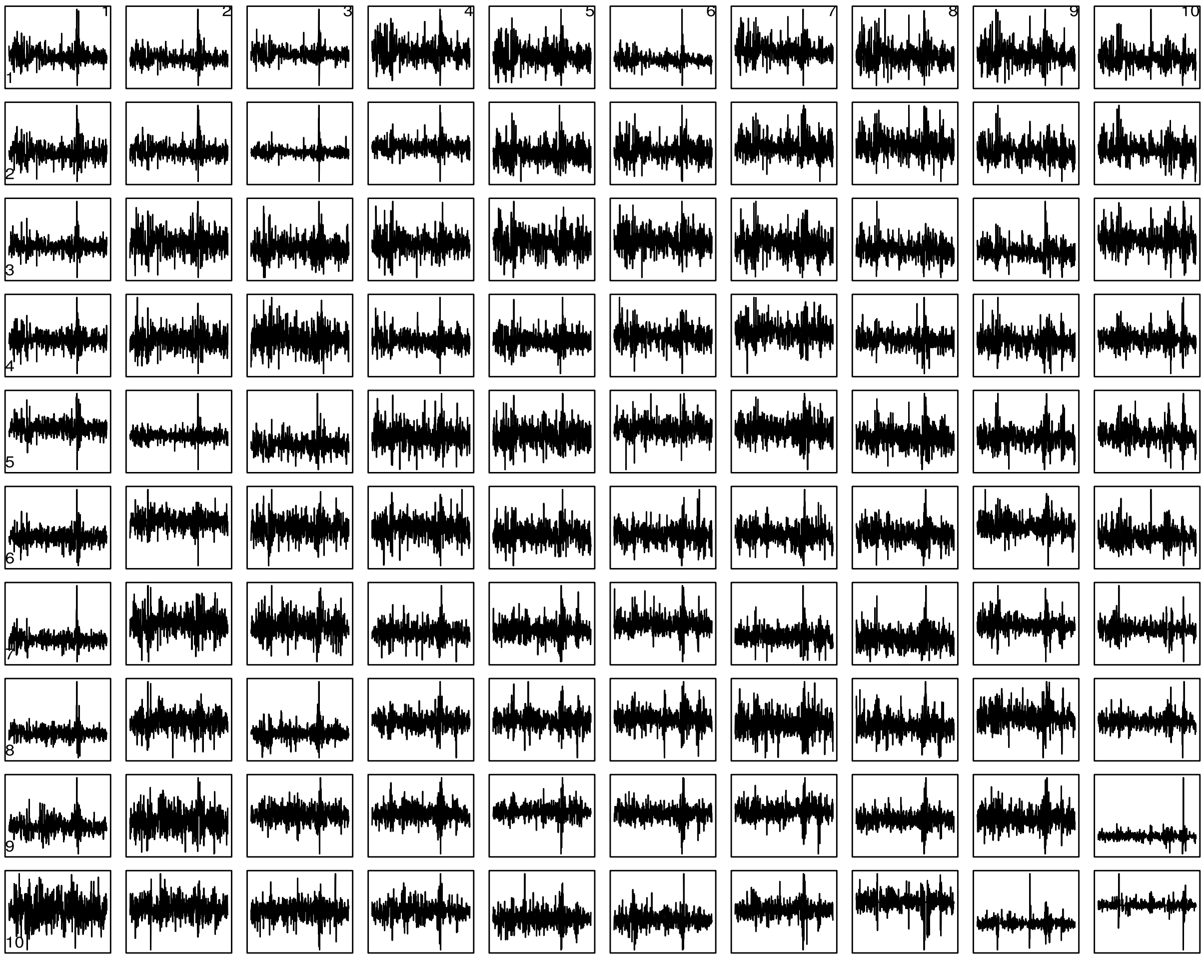

Fig. 1. Time series plot of Fama–French 10 by 10 series.

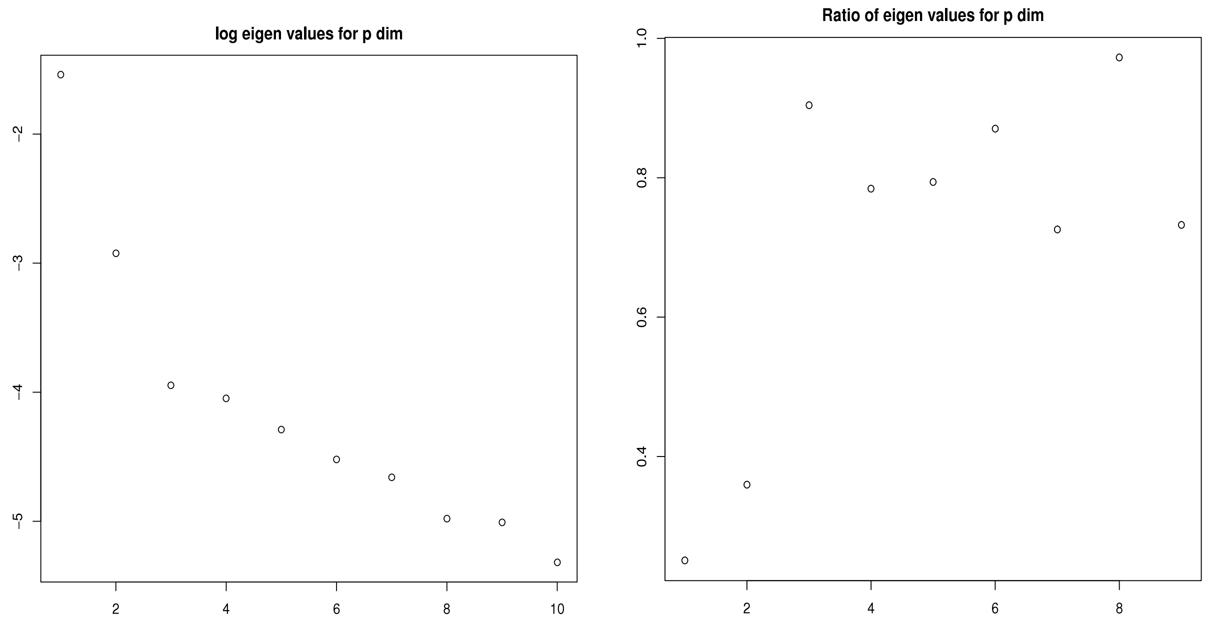

Fig. 2. Fama–French series: Logarithms and ratios of eigenvalues of $\widehat{\boldsymbol{M}}_1$ for the row (size) loading matrix.

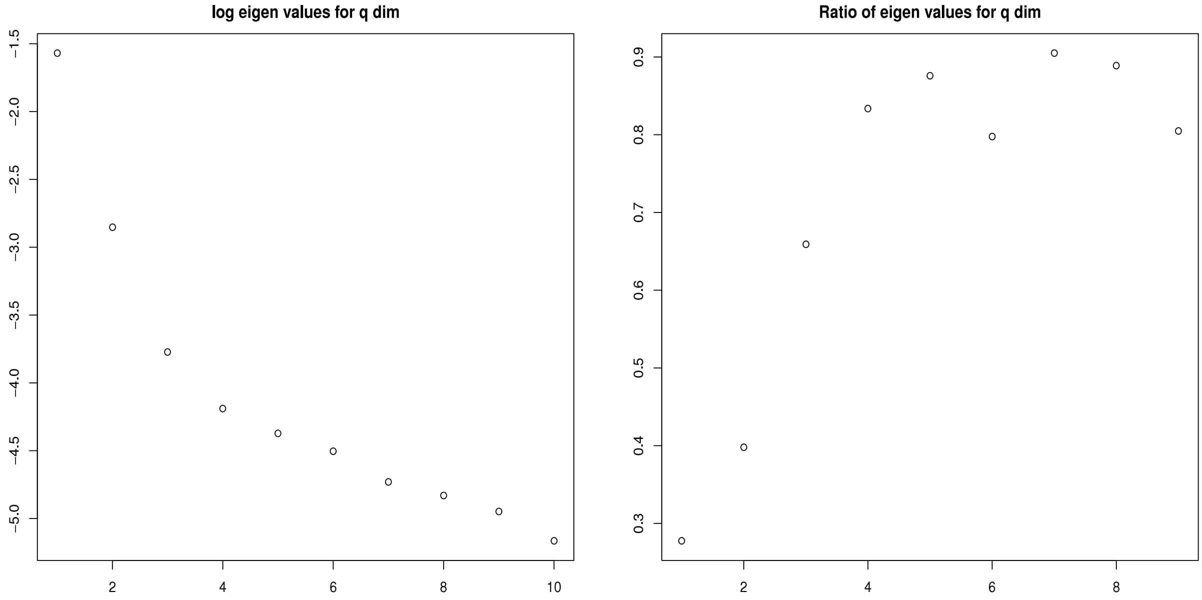

Fig. 3. Fama–French series: Logarithms and ratios of eigenvalues of $\widehat{\boldsymbol{M}}_2$ for the column (BE) loading matrix.

Table 11
Comparison of different models for Fama–French series.

	Factor	RSS	# factors	# parameters
Matrix model	(0,0)	29,193	0	0
Matrix model	(2,2)	14,973	4	40
Matrix model	(2,3)	14,514	6	50
Matrix model	(3,2)	14,166	6	50
Matrix model	(3,3)	13,530	9	60
Vector model	3	16,262	3	300
Vector model	4	15,365	4	400
Vector model	5	14,565	5	500
Vector model	6	14,149	6	600

the 100 series of the 20 years are reported. The RSS corresponding to model $(0, 0)$ is the total sum of squares of the observed 100 series of the 20 years being studied. It is seen that the matrix factor model with $(2, 2)$ factor matrices performs better than the vectorized factor model with equal number of factors and many more parameters in the loading matrices. The $(3, 2)$ matrix factor model performs similarly as the 6-factor vectorized factor model, but the number of parameters used is much smaller.

6. Real example: series of company financials

In this example we analyze the series of financial data reported by a group of 200 companies. We constructed 16 financial characteristics based on company quarterly financial reports. The list of variables and their definitions is given in Appendix 2. The period is from the first quarter of 2006 to the fourth quarter of 2015 for 10 years with total 40 observations. The total number of time series is 3,200.

Figs. 6 and 7 show the eigenvalues and their ratios of $\widehat{\boldsymbol{M}}_1$ and $\widehat{\boldsymbol{M}}_2$ for row factors and column factors. The estimated dimensions k_1 and k_2 are both 3, though we use $k_1 = 5$ and $k_2 = 20$ for this illustration, with interesting results. Estimation is done using $h_0 = 2$.

The estimated row loading matrix is rotated to maximize its variance, with potential grouping shown by the shaded areas in Table 12. It shows the loading of each financial on the five rows of the factor matrix, after proper scaling (30 times) and reordering for easy visualization. The two financials in Group 1 load almost exclusively on Row 1 of the factor matrix, with almost the same weights. The six financials in Group 2 load heavily on Row 2, again with almost the same weights. The three financials in Group 3 load on Rows 3 and 4, with somewhat different weights. Finally, the five financials in Group 4 mainly load on Row 5 of the factor matrix, with Payout. Ratio having opposite weights from the others.

The detailed grouping is shown in Table 13. Group 1 consists of asset to equity ratio and liability to equity ratio. They are two very closely related measures. Group 2 consists of six measures on earnings and returns. Group 3 consists of cash

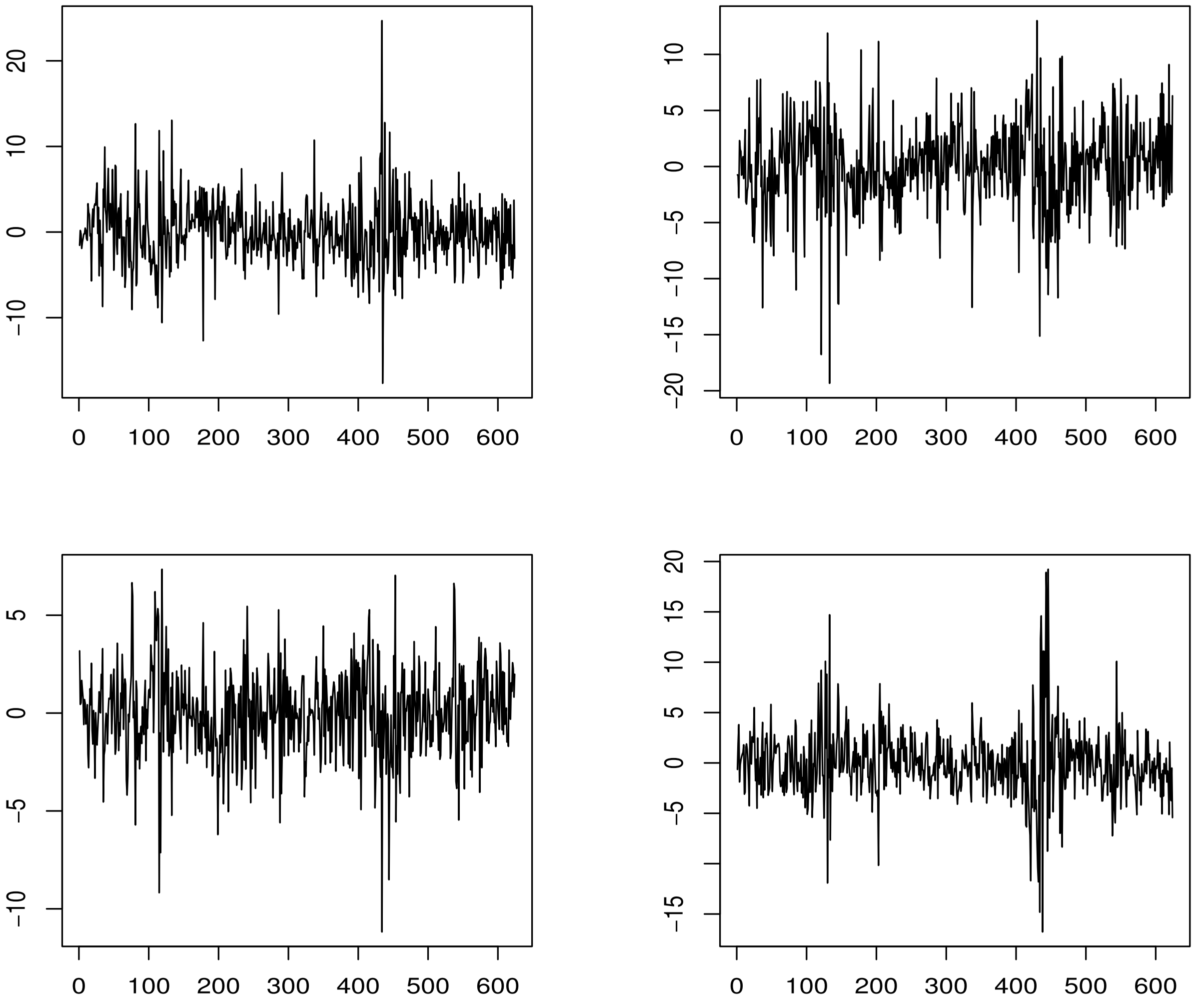

Fig. 4. Fama–French series: Estimated factors.

Table 12
Financial series: Loading matrix after a varimax rotation and scaling.

Row Factor	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16
1	21	21	-1	-1	-1	-1	-1	4	-2	2	1	0	1	0	0	-1
2	1	1	-13	-9	-11	-11	-13	-12	7	-6	-1	3	-1	1	0	1
3	0	0	0	-10	0	-1	0	-1	-11	9	-24	-4	-1	2	-2	0
4	0	0	-3	2	5	4	-2	-2	19	21	-2	4	1	3	0	-1
5	0	0	-1	1	2	0	-2	-2	-4	2	1	9	-8	-13	-14	-18

Table 13
Financial series: Grouping of company financials.

Group 1	AssetE.R	LiabilityE.R				
Group 2	Earnings.R	EPS	Oper.M	Profit.Margin	ROA	ROE
Group 3	Cash.PS	Gross.Margin	Revenue.PS			
Group 4	Payout.R	Profit.G.Q	Profit.G.Y	Revenue.G.Q	Revenue.G.Y	

and revenue per share, and gross margin. Group 4 consists of profit growth and revenue growth comparing to the previous quarter and the same quarter last year. The Payout Ratio variable is also included in the group. Such groupings are relatively expected.

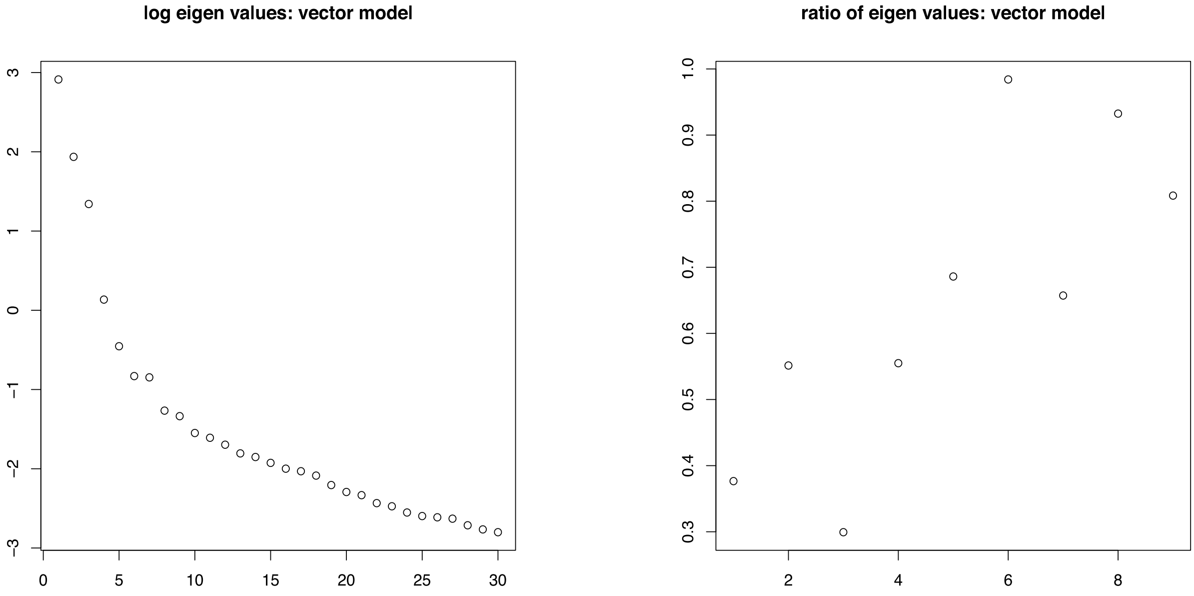

Fig. 5. Fama–French series: Logarithms and ratios of eigenvalues of $\hat{\boldsymbol{M}}$ for the vectorized factor model.

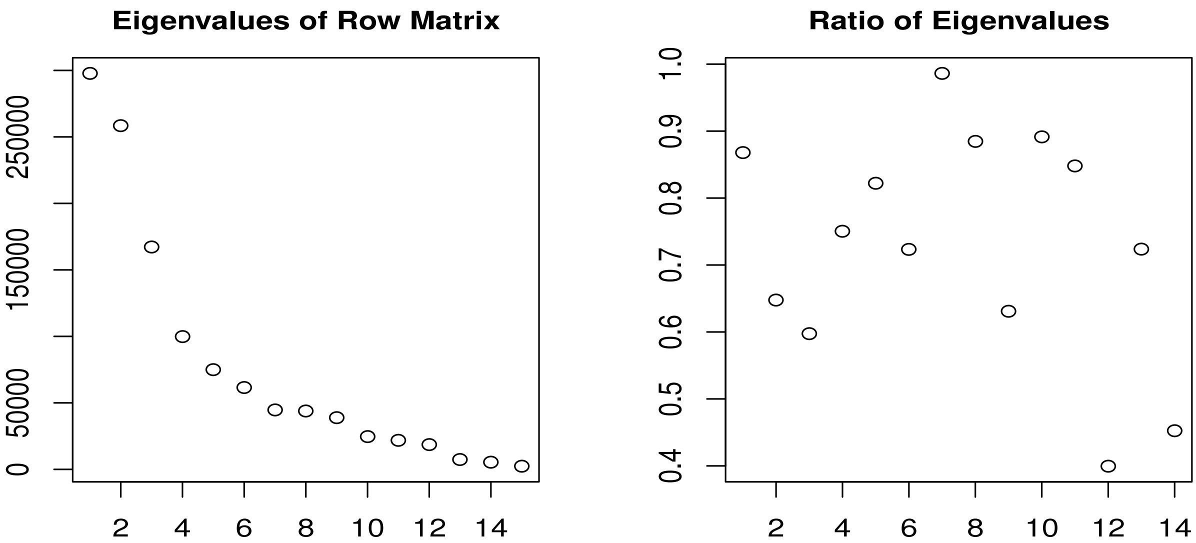

Fig. 6. Financial series: Eigenvalues and their ratios of $\widehat{\boldsymbol{M}}_1$.

Based on the 200 rows of the estimated columns loading matrix (corresponding to the companies), after rotation to maximize the variance, the companies are grouped into 6 groups. Table 14 shows the grouping corresponding to the industry classification index. The pattern is not as clear as the clustering of the row loading matrix but we still make some interesting discoveries. Industrial companies are mainly clustered in Groups 1 to 3; Health Care companies in Groups 2 and 3; Information Technology companies in Groups 1, 3 and 5; and Materials companies in Group 3. Looking from the other angle, we find that Group 4 mainly contains Energy companies; Group 5 mainly contains Consumer Discretionary, Financials and Information Technology companies; Group 6 mainly contains Utility companies.

Fig. 8 shows the total 100 factor series in the 5 by 20 factor matrix. Interpretation of the factors is difficult. There are significant redundancy and correlation among the factors, since we have 100 factors but the time series length is only 40. Clearly the model tends to overfit. This example is for illustration purpose only, though we do find interesting features.

Table 15 shows a simple comparison between the matrix factor models and vectorized factor models of various size and number of factors. Since the time series is short, the table shows in-sample residual sum of squares. Again, it is seen that

Please cite this article in press as: Wang D., et al., Factor models for matrix-valued high-dimensional time series. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.09.013.

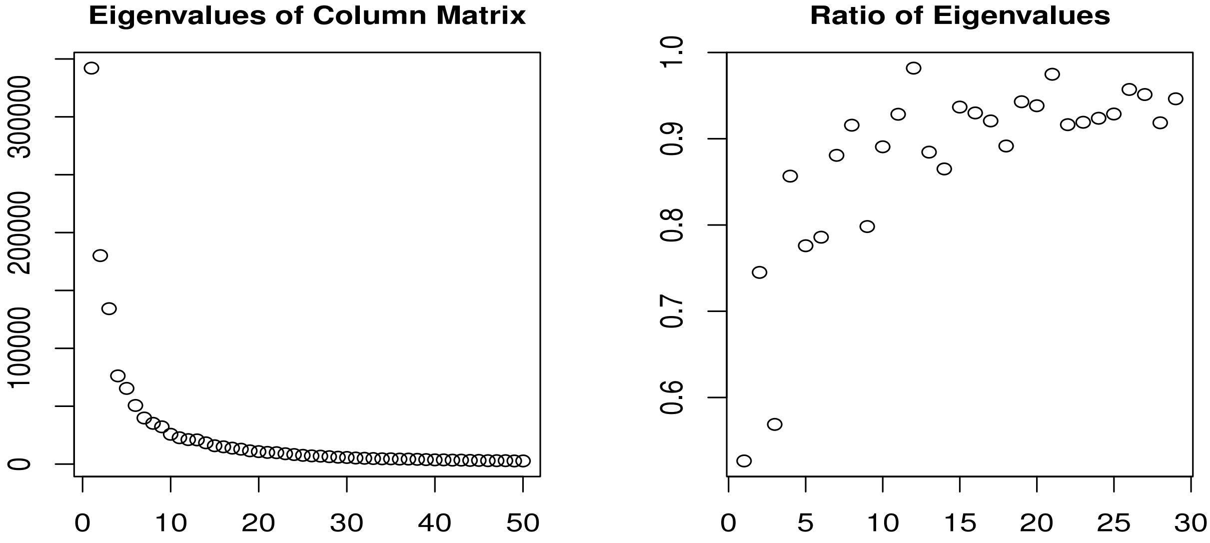

Fig. 7. Financial series: Eigenvalues and their ratios of $\widehat{\boldsymbol{M}}_2$.

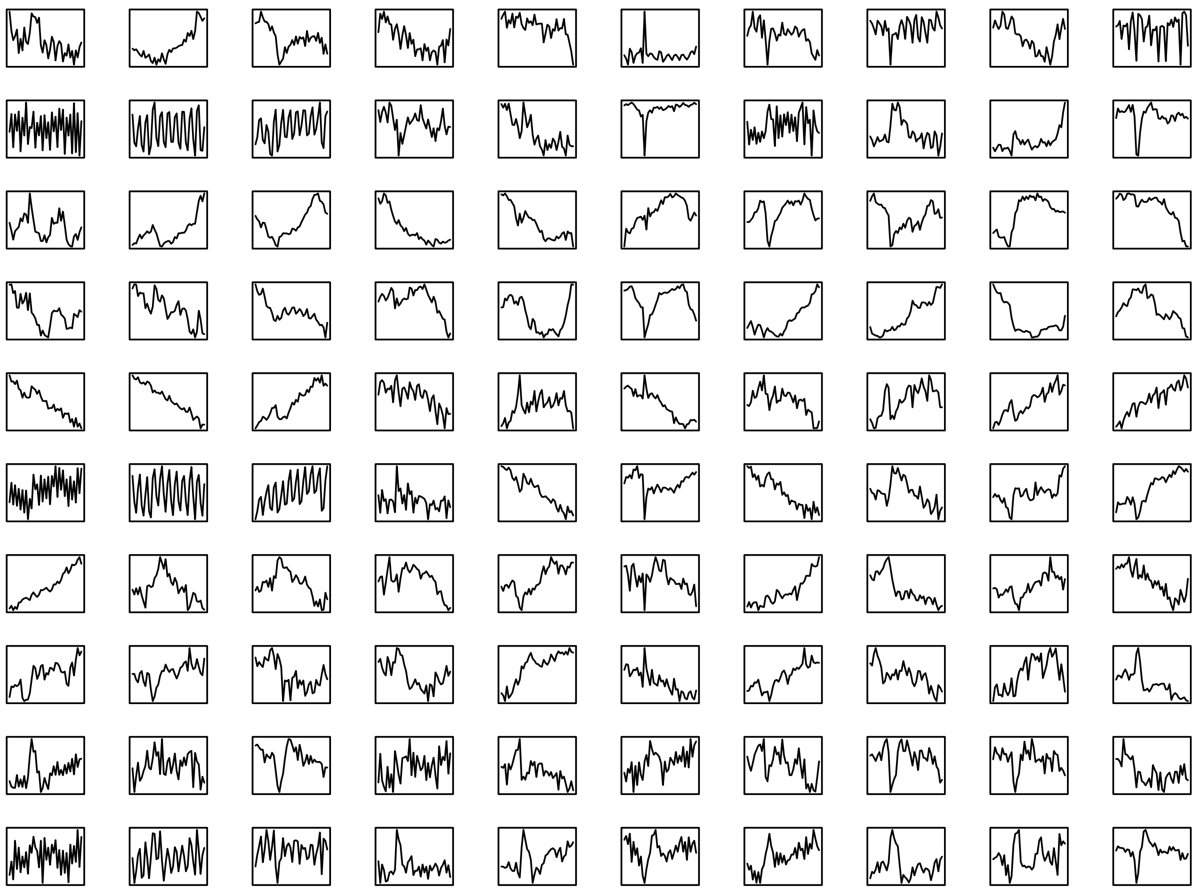

Fig. 8. Financial series: Plot of the 100 series in the factor matrix.

the matrix factor models use much fewer parameters in loading matrices to achieve similar estimation performance. The number of parameters involved is large as we are jointly modeling 3,200 time series.

Table 14
Financial series: Matching the companies and the industry.

Group	1	2	3	4	5	6
Consumer discretionary	2	3	4	1	5	4
Consumer staples	3	4	6	1	0	1
Energy	2	3	4	9	0	4
Financials	0	5	2	0	4	0
Health care	0	5	17	0	1	2
Industrials	12	7	17	0	1	2
Information technology	4	0	12	0	5	0
Materials	2	5	8	1	1	1
Telecommunications services	0	2	1	0	0	0
Utilities	0	6	5	1	0	13

Table 15
Financial series: Comparison of different models for company financials series.

	Factor	RSS	RSS/SST	# factors	# parameters
Matrix model	(4,10)	86,739	0.701	40	2,064
Matrix model	(4,20)	74,848	0.610	80	4,064
Matrix model	(5,10)	84.517	0.688	50	2,080
Matrix model	(5,20)	71,535	0.582	100	4,080
Matrix model	(5,30)	65,037	0.530	150	6,080
Vector model	3	79,704	0.650	3	9,600
Vector model	4	73,457	0.598	4	12,800
Vector model	5	68,428	0.557	5	16,000
Vector model	6	63,031	0.514	6	19,200

7. Summary

In this paper we propose a matrix factor model for high-dimensional matrix-valued time series, along with an estimation procedure. Theoretical analysis shows the asymptotic properties of the estimators. Simulation and real examples are used to illustrate the model and finite sample properties of the estimators. The real examples show the usefulness of the model and its ability to reveal interesting features of high-dimensional time series. Significant amount of effort is needed to investigate model validation and model comparison procedures for the proposed model. Extensions to multi-term model and approaches to simply reducing factor redundancy are important research topics. Extending the model to dynamic factor model with an imposed dynamic structure on the factor matrix will be useful in terms of prediction and better understanding the dynamic nature of the matrix-valued time series.

Acknowledgments

We thank the Editors and two anonymous referees for their helpful insightful comments and suggestions, which lead to significant improvement of the paper in motivation and justification, design of simulation study and the analysis of real examples.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2018.09.013.

References

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31 (3), 307–327.
Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis, Forecasting and Control. Holden Day, San Francisco.
Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer.
Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and meanvariance analysis on large asset markets. Econometrica 51 (5), 1281–1304.
Chang, J., Guo, B., Yao, Q., 2015. High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. J. Econometrics 189 (2), 297–312.
Crainiceanu, C.M., Caffo, B.S., Luo, S., Zipunnikov, V.M., Punjabi, N.M., 2011. Population value decomposition, a framework for the analysis of image Populations. J. Amer. Statist. Assoc. 106 (495), 775–790.
Ding, C., Ye, J., 2005. 2-Dimensional singular value decomposition for 2D maps and images. In: Proc. SIAM Int'L Conf. Data Mining, SDM'05, pp. 32–43.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united Kingdom inflactions. Econometrika 59, 987–1007.
Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized arch. Econometric Theory 11 (1), 122–150.
Fama, E.F., French, K.R., 1993. The cross-section of expected stock returns. J. Finance 47, 427–465.
Fan, J., Liao, Y., Mincheva, M., 2011. High dimensional covariance matrix estimation in approximate factor models. Ann. Statist. 39 (6), 3320.
Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (4), 603–680.

Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: identification and estimation. Rev. Econ. Stat. 82 (4), 540–554.
Gupta, A.K., Nagar, D.K., 2000. Matrix Variate Distributions. Chapman & Hall/CRC, Boca Raton, FL.
Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. J. Amer. Statist. Assoc. 102 (478), 603–617.
Kollo, T., von Rosen, D., 2006. Advanced Multivariate Statistics with Matrices. 579, Springer.
Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: inference for the number of factors. Ann. Statist. 40 (2), 694–726.
Lam, C., Yao, Q., Bathia, N., 2011. Estimation of latent factors for high-dimensional time series. Biometrika 98 (4), 901–918.
Leng, C., Tang, C.Y., 2012. Sparse matrix graphical models. J. Amer. Statist. Assoc. 107 (499), 1187–1200.
Liu, X., Chen, R., 2016. Regime-Switching factor models for high-dimensional time series. Statist. Sinica 26, 1427–1451.
Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer, Berlin.
Paatero, P., 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., Yao, Q., 2008. Modelling multiple time series via common factors. Biometrika 95 (2), 365–379.
Stock, J.H., Watson, M.W., 2004, An empirical comparison of methods for forecasting using many predictors. In: Technical Report, Department of Economics, Havard University.
Tiao, G.C., Box, G.E.P., 1981. Modelling multiple time series with applications. J. Amer. Statist. Assoc. 76 (376), 802–816.
Tiao, G.C., Tsay, R.S., 1989. Model specification in multivariate time series. J. R. Stat. Soc. Ser. B Stat. Methodol. 51 (2), 157–213.
Tong, H., 1990. Nonlinear Time Series Analysis: A Dynamical System Approach. Oxford University Press, London.
Tsay, R.S., 2005. Analysis of Financial Time Series. Wiley, New York.
Tsay, R.S., 2014. Multivariate Time Series Analysis. Wiley, New York.
Walden, A.T., Serroukh, A., 2002. Wavelet analysis of matrix-valued time series. Proc. Math. Phys. Eng. Sci. 458 (2017), 157–179.
Wang, D., Shen, H., Truong, Y., 2016. Efficient dimension reduction for high-dimensional matrix-valued data. Neurocomputing 190, 25–34.
Werner, K., Jansson, M., Stoica, P., 2008. On estimation of covariance matrices with Kronecker product structure. IEEE Trans. Signal Process. 56 (2), 478–491.
Yang, J., Zhang, D., Frangi, A.F., Yang, J., 2004. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 26 (1), 131–137.
Ye, J., 2005. Generalized Low Rank Approximations of Matrices. Mach. Learn. 61 (1–3), 167–191.
Yin, J., Li, H., 2012. Model selection and estimation in the matrix normal graphical model. J. Multivariate Anal. 107, 119–140.
Zhang, D., Zhou, Z., 2005. $(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing 69 (1), 224–231.
Zhao, J., Leng, C., 2014. Structured lasso for regression with matrix covariates. Statist. Sinica 24, 799–814.
Zhou, S., 2014. Gemini: graph estimation with matrix variate normal instances. Ann. Statist. 42 (2), 532–562.
Zhou, H., Li, L., 2014. Regularized matrix regression. J. R. Stat. Soc. Ser. B Stat. Methodol. 76 (2), 463–483.

Please cite this article in press as: Wang D., et al., Factor models for matrix-valued high-dimensional time series. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.09.013.