

Threshold Matrix-Variate Factor Models

Xialu Liu ^{*1} and Elynn Y. Chen ^{† 2}

¹San Diego State University

²Stern School of Business, New York University

Abstract

Motivated by the growing availability of complex time series observed in real applications, we propose a threshold matrix-variate factor model, which simultaneously addresses the sample-wise and time-wise complexities of a time series. The sample-wise complexity is characterized by modeling matrix-variate observations directly, while the time-wise complexity is modeled by a threshold variable to describe the non-linearity in time series. The estimators for loading spaces and threshold values are introduced and their asymptotic properties are investigated. Our matrix-variate models compress data more efficiently than traditional vectorization-based models. Furthermore, we greatly extend the scope of current research on threshold factor models by removing several restrictive assumptions, including existence of only one threshold, fixed factor dimensions across different regimes, and stationarity within regime. Under the relaxed assumptions, the proposed estimators are consistent even when the numbers of factors are overestimated. Simulated and real examples are presented to illustrate the proposed methods.

KEYWORDS: Curse of dimensionality; Factor models; High-dimensional time series; Matrix-variate time series; Thresholding effect.

^{*}Xialu Liu is an Associate Professor, Management Information Systems Department, San Diego State University, San Diego, CA 92182. Email: xialu.liu@sdsu.edu. Xialu Liu is the corresponding author.

[†]Elynn Y. Chen is an Assistant Professor at the Stern School of Business, New York University. Email: elynn.chen@stern.nyu.edu. The authors thank the editor, the associate editor, and the referee for the insightful comments and suggestions which significantly improve the paper.

1 Introduction

Advances in information technology have brought a growing number of high-dimensional matrix-variate time series in forms of dynamic panels (Chen et al., 2020a), dynamic networks (Chen and Chen, 2019), and multi-variate spatial temporal observations (Yu et al., 2015). To reduce dimensions and to reveal underlying correlation structures, Wang et al. (2019) proposed a matrix-variate factor model which is defined, for time series of $\mathbf{X}_t \in \mathbb{R}^{p_1 \times p_2}$, as

$$\mathbf{X}_t = \mathbf{R}\mathbf{F}_t\mathbf{C}' + \mathbf{E}_t, \quad t = 1, 2, \dots, T. \quad (1)$$

Here, the observed matrix $\mathbf{X}_t$ is driven by a latent factor matrix $\mathbf{F}_t \in \mathbb{R}^{k_1 \times k_2}$ of a smaller dimension (i.e. $k_1 \ll p_1$ and $k_2 \ll p_2$), plus a noise matrix $\mathbf{E}_t$. Unobserved $\mathbf{R} \in \mathbb{R}^{p_1 \times k_1}$ and $\mathbf{C} \in \mathbb{R}^{p_2 \times k_2}$ are row and column loading matrices, respectively. The temporal dynamics of $\mathbf{X}_t$ is entirely driven by the latent factor $\mathbf{F}_t$ in the sense that $\mathbf{E}_t$ is assumed to serially independent. Model (1) extends the classic factor analysis to directly modeling matrix-variate observations. Compared with those estimated from the vectorized factor model, estimators obtained from (1) have been shown to achieve faster convergence rate (Wang et al., 2019) and provide better interpretation in real applications such as international import-export trading (Chen and Chen, 2019) and multi-variate spatial temporal environment data (Chen et al., 2020b).

Factor models for high-dimensional time series have been extensively investigated over the past decades. The most widely-studied model in econometrics, called approximate factor model, searches for common factors that affect the dynamics of most of time series in the cross-section dimension, and allows limited time-series and cross-section dependence in the idiosyncratic component; See Assumption B and Assumption C in Bai (2003). Many existing theoretical results on factor models were derived under this setting; See examples Chen et al. (2014), Barigozzi et al. (2018), Barigozzi and Cho (2020), Massacci (2017), Ma and Su (2018), Wu (2021) and references therein. Following the assumptions in the approximate factor model, Chen and Fan (2021) extended the model for matrix-variate time series analysis. However, Lam et al. (2011) pointed out that the rigorous definition of factors and idiosyncratic noise in the approximate factor model can only be established when the dimension goes to infinity, and they proposed another approach to decompose the observed data. Lam et al. (2011) separates the time series into two parts: a dynamic part driven by factors which captures all the temporal dependence of the observed process and a static noise which is time-independent. Since the noise has no serial dependence, the decomposition is unique when the cross-section dimension is finite. The model proposed by Lam

et al. (2011) became popular recently and were discussed in Chang et al. (2015), Liu and Chen (2016), Chen et al. (2020a), Liu and Chen (2020), and Liu and Zhang (2022). Wang et al. (2019) extended the factor models for matrix time series analysis based on Lam et al. (2011). Two types of factor models are compared in Remark 2. Throughout this paper we follow the factor model settings proposed by Lam et al. (2011).

One important assumption of the matrix variate factor model (1) is that the loading matrices $\mathbf{R}$ and $\mathbf{C}$ affects the observation in a time-invariant linear fashion. However, literature has shown the presence of nonlinearity in many time series data (Tong and Lim, 1980; Tsay, 1989, 1998; Fan and Yao, 2008). Particularly, the threshold effect is observed in many cases of factor models (Massacci, 2017; Liu and Chen, 2020; Wu, 2021). For example, CAPM theory indicates that the expected market return is an important factor for the expected return of various assets, and Liu and Chen (2016) found that its impact (loadings) on any individual asset may be different depending on whether a stock market is volatile or stable. In economics, risk-free rate, unemployment, and economic growth are crucial factors of all economic activities and serve as their performance indicators. Indeed, the loadings of these factors may vary under different fiscal policies or in different stages of the economic cycle (Kim and Nelson, 1998).

In this paper, we introduce a threshold matrix-variate factor model, which relaxes the time-invariant assumption on $\mathbf{R}$ and $\mathbf{C}$ and extends model (1) to allow for the threshold effect (Tong and Lim, 1980; Tsay, 1989, 1998). The theoretical research on vector-variate threshold factor models has emerged in literature in the past few years. Massacci (2017) used the least squares method to estimated the threshold and proposed model selection criteria. Liu and Chen (2020) developed a projection-based objective function to estimate the threshold value. Wu (2021) discussed the case when the threshold variable divides the data into multiple regimes and proposed an algorithm to identify regime shifts. Our threshold matrix-variate factor model generalizes its vector-variate counterpart to directly handle the intrinsic matrix structure. We achieve greater dimension reduction and obtain better estimation than the vectorization-based procedure via a carefully-designed estimation procedure that exploits intrinsic matrix and regime-switching structure.

Besides that, we also make the following contributions to the current literature on threshold factor models. Firstly, we consider the case where dimensions of the latent factor matrix may vary in different regimes. This removes the limitation of the methods proposed in Massacci (2017), Liu and Chen (2020), and Wu (2021), all of which require the number of factors to remain the

same across regimes. Secondly, our algorithm is able to identify the thresholding mechanism when the number of thresholds is unknown. In contrast, current literature in threshold factor models (Massacci, 2017; Liu and Chen, 2020) only allows one threshold to divide data in two regimes. Thirdly, the proposed estimation procedure can be applied to data that neither the factor or the noise process is stationary, while the method in Wu (2021) is limited to data in which latent factor and noise process are both strictly stationary. We only require the factor process to satisfy the mixing condition and the moments of factor and noise processes to be bounded. Our methods can successfully identify regime switchings and estimate thresholds and loadings when the means and variances of factor and noise processes are time-varying; See simulation results in Section 3.3. In summary, the proposed methods greatly relax these restrictive assumptions and thus extend the scope of potential applications of threshold factor models.

The rest of the paper is organized as follows. Section 2.1 introduces the threshold factor model with two regimes for high-dimensional matrix-variate time series. Section 2.2 discusses the estimators for loading spaces and presents their theoretical properties. Section 2.3 investigates the threshold estimation and establishes the asymptotics of the proposed threshold estimator. The estimation for numbers of factors is studied in Section 2.4, and we show that if the numbers of factors are overestimated the proposed estimators are still consistent. In Section 2.5 we consider the threshold matrix factor models with multiple thresholds and introduce the procedure for model identification. We apply our methods to simulated and real data, and present the results in Sections 3 and 4, respectively. Section 5 provides the conclusions. The lemmas and detailed mathematical proofs are in the Appendix 1. Appendix 2 presents more simulation results.

Here is some notation we will use throughout the paper. Let $\text{vec}(\cdot)$ be the vectorization operator, which converts a matrix to a vector by stacking columns of the matrix on top of each other. For any matrix $\mathbf{H}$, let $\|\mathbf{H}\|_2$ and $\|\mathbf{H}\|_F$ be the L-2 norm and Frobenius norm of $\mathbf{H}$, respectively, $\sigma_i(\mathbf{H})$ be its i -th largest singular value, $\text{rank}(\mathbf{H})$ be its rank, and $\|\mathbf{H}\|_{\min}$ be the square root of the minimum nonzero eigenvalue of $\mathbf{H}'\mathbf{H}$. For a squarer matrix $\mathbf{H}$, $\text{tr}(\mathbf{H})$ denotes its trace. We use $\mathbf{h}_u$ and $\mathbf{h}_v$ to represent the vectors with the entries in the u -th row and the v -th column of $\mathbf{H}$ respectively, and h_{uv} to represent the (u, v) -th entry of $\mathbf{H}$. We write $a \asymp b$, if $a = O(b)$ and $b = O(a)$. For a constant a , $[a]$ is the largest integer less than or equal to a . We use C 's and C_i 's to denote generic uniformly positive constants which only depend on the parameters.

2 Estimation and Identification of Threshold Matrix Factor Models

In this Section, we will introduce the threshold matrix factor models, propose the estimation methods, and develop an algorithm for model identification. The asymptotic properties of the proposed estimators will also be investigated. We first consider a special case for threshold matrix factor models with two regimes in Section 2.1, and propose methods to estimate the loading spaces and threshold value when the numbers of factors are known in Section 2.2 and Section 2.3, respectively. In practice, the numbers of factors are also unobserved, and Section 2.4 considers the estimation for the numbers of factors. Section 2.5 discusses the case when the number of regimes is unknown and presents the algorithm to identify the number of regimes and estimate multiple threshold values.

2.1 Two-regime threshold matrix factor model

Let $\mathbf{X}_t$ be a $p_1 \times p_2$ observed matrix-variate time series, where

$$\mathbf{X}_t = \begin{cases} \mathbf{R}_1 \mathbf{F}_{t,1} \mathbf{C}'_1 + \mathbf{E}_t & z_t < r_0, \\ \mathbf{R}_2 \mathbf{F}_{t,2} \mathbf{C}'_2 + \mathbf{E}_t & z_t \geq r_0, \end{cases} \quad t = 1, \dots, T. \quad (2)$$

$\mathbf{F}_{t,i}$ is a $k_{1,i} \times k_{2,i}$ matrix-variate time series which consists of fundamental latent factors for $i = 1, 2$. $\mathbf{R}_i$ is an unobserved $p_1 \times k_{1,i}$ row loading matrix in regime i , and $\mathbf{C}_i$ is an unobserved $p_2 \times k_{2,i}$ column loading matrix in regime i , for $i = 1, 2$. $\mathbf{E}_t$ is a $p_1 \times p_2$ matrix which is the noise process and has no serial dependence. z_t is an observed threshold variable, controlling the switchings between two regimes. Loading matrices $\{\mathbf{R}_i, \mathbf{C}_i\}$ are different across regimes.

Threshold models were first proposed by Tong and Lim (1980) to interpret nonlinear dynamics in time series data, and were widely applied in economics (Wu and Chen, 2007), finance (Tsay, 1989; Massacci, 2017), biology (Tong and Lim, 1980), and environmental science (Tsay, 1998). The threshold variable is always assumed to be observable, which is one of the limitations of this class of models. There are various choices of threshold variables used in literature. Specifically, the threshold variable can be a lag variable of the observed time series (Tong and Lim, 1980; Tong, 1990), an exogenous variable (Tsay, 1998), or a linear or nonlinear combination of observed data and/or exogenous variables (Chen, 1995).

Two regimes in model (2) are only distinguished by loading matrices, which is a common assumption in factor models for high dimensional vector time series with multiple regimes; See

examples in [Chen et al. \(2014\)](#); [Baltagi et al. \(2017\)](#); [Massacci \(2017\)](#). One advantage to focus exclusively on the regime switchings in loadings is that the factor and noise processes can accommodate non-stationarity within regimes, including but not limited to time-varying mean and time-varying variance; See Section 3.3. However, as a result, our methods cannot be used to make inferences about the changes in the moments of $\mathbf{F}_{t,1}$, $\mathbf{F}_{t,2}$, or $\mathbf{E}_t$ as [Barigozzi et al. \(2018\)](#) does.

Remark 1. Model (2) with $(k_{1,1}k_{2,1} + k_{1,2}k_{2,2})$ factors is a special case of matrix factor model ([Wang et al., 2019](#)), since it can be re-written as

$$\mathbf{X}_t = \tilde{\mathbf{R}}\tilde{\mathbf{F}}_t\tilde{\mathbf{C}}' + \mathbf{E}_t, \quad (3)$$

where

$$\tilde{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_2 \end{pmatrix}, \quad \tilde{\mathbf{F}}_t = \begin{pmatrix} \mathbf{F}_{t,1}I_{t,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{t,2}I_{t,2} \end{pmatrix}, \quad \tilde{\mathbf{C}} = \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{pmatrix},$$

where $I_{t,1} = I(z_t < r_0)$ and $I_{t,2} = I(z_t \geq r_0)$. Model (3) is a one-regime factor model in [Wang et al. \(2019\)](#) with constraints in the factor process. If we ignore the changes of loadings, the number of factors may be enlarged as $(k_{1,1} + k_{1,2})(k_{2,1} + k_{2,2})$ in (3), the threshold factor model (2) uses fewer factors and achieves greater dimension reduction by introducing regimes.

Remark 2. The approximate factor model in econometrics was built based on the assumption that the factors have an impact on most of the series, and thus the idiosyncratic noise allows serial dependence but only has weak cross-sectional dependence, i.e., $\frac{1}{p} \sum_{i=1}^p \sum_{j=1}^p |\sigma_{ij,t}| < C$ for $t = 1, \dots, T$, where p is the dimension of the vector time series and $\sigma_{ij,t}$ is the (ij) -th entry in the covariance matrix of the noise process at time t ; See Assumption C3 in [Bai \(2003\)](#), Assumption M1.b in [Stock and Watson \(2002\)](#), and more examples in [Bai and Ng \(2002\)](#); [Fan et al. \(2017\)](#). However, under this condition, the common component and noise process are not identifiable when the dimension is finite. [Lam et al. \(2011\)](#) took another setting that the factors capture all serial dependence of the observed process. As a result, the noise is time-independent and can accommodate strong cross-sectional dependence, i.e., $|\sigma_{ij,t}| < C$ for $i, j = 1, \dots, p$ and $t = 1, \dots, T$; See more examples in [Chang et al. \(2015\)](#); [Liu and Chen \(2016\)](#); [Wang et al. \(2019\)](#); [Chen et al. \(2020a\)](#). To make the common component and noise process separable, in this paper, we follow their settings, and further relax their assumption by allowing heteroscedasticity for the noise process.

Since $\mathbf{R}_i$, $\mathbf{C}_i$ and $\mathbf{F}_{t,i}$ are all unobserved, the factor models have ambiguity issues and $\mathbf{R}_i$ and

$\mathbf{C}_i$ are not uniquely defined. Specifically, the model (2) can be re-written as,

$$\mathbf{X}_t = \sum_{i=1}^2 [\mathbf{R}_i \mathbf{U}_i (\mathbf{U}_i^{-1} \mathbf{F}_{t,i} \mathbf{V}_i^{-1}) \mathbf{V}_i \mathbf{C}_i'] I_{t,i} + \mathbf{E}_t, \quad t = 1, \dots, T.$$

The row loading matrix, column loading matrix and factor process can be replaced by $\mathbf{R}_i \mathbf{U}_i$, $\mathbf{C}_i \mathbf{V}_i'$, and $\mathbf{U}_i^{-1} \mathbf{F}_{t,i} \mathbf{V}_i^{-1}$, for any non-singular matrices $\mathbf{U}_i$ and $\mathbf{V}_i$, $i = 1, 2$. However, the column spaces spanned by $\mathbf{R}_i$ and $\mathbf{C}_i$, $\mathcal{M}(\mathbf{R}_i)$ and $\mathcal{M}(\mathbf{C}_i)$, called row loading space and column loading space for regime i respectively, are identifiable. Our aim is to estimate the row and column loading spaces, instead of loading matrices. We can further decompose $\mathbf{R}_i$ and $\mathbf{C}_i$ as follows

$$\mathbf{R}_i = \mathbf{Q}_{1,i} \mathbf{W}_{1,i}, \text{ and } \mathbf{C}_i = \mathbf{Q}_{2,i} \mathbf{W}_{2,i}, \quad (4)$$

where $\mathbf{Q}_{s,i}$ is a $p_s \times k_{s,i}$ orthogonal matrix, and $\mathbf{W}_{s,i}$ is a $k_{s,i} \times k_{s,i}$ non-singular matrix, for $s, i = 1, 2$. s is the index for dimension reduction directions (row/column), and i is the index for regimes. By the definition, we have $\mathcal{M}(\mathbf{R}_i) = \mathcal{M}(\mathbf{Q}_{1,i})$ and $\mathcal{M}(\mathbf{C}_i) = \mathcal{M}(\mathbf{Q}_{2,i})$. In the following, we will estimate the orthonormal representatives of $\mathcal{M}(\mathbf{R}_i)$ and $\mathcal{M}(\mathbf{C}_i)$, $\mathbf{Q}_{1,i}$ and $\mathbf{Q}_{2,i}$, for $i = 1, 2$.

2.2 Estimation of loading spaces when the threshold value is known

If only considering the v -th column of $\mathbf{X}_t$ in model (2), we have

$$\mathbf{x}_{t,v} = \begin{cases} \mathbf{R}_1 \mathbf{F}_{t,1} \mathbf{c}_{1,v} + \mathbf{e}_{t,v} = \mathbf{Q}_{1,1} \mathbf{Z}_{t,1} \mathbf{q}_{2,1,v} + \mathbf{e}_{t,v} & z_t < r_0, \\ \mathbf{R}_2 \mathbf{F}_{t,2} \mathbf{c}_{2,v} + \mathbf{e}_{t,v} = \mathbf{Q}_{1,2} \mathbf{Z}_{t,2} \mathbf{q}_{2,2,v} + \mathbf{e}_{t,v} & z_t \geq r_0, \end{cases} \quad (5)$$

where $\mathbf{Z}_{t,i} = \mathbf{W}_{1,i} \mathbf{F}_{t,i} \mathbf{W}_{2,i}'$ for $i = 1, 2$.

If we use r as the tentative threshold value to split data, let h be a positive integer, and define the second cross moment matrices of the factor process and observed process in different partitions,

$$\begin{aligned} \Omega_{zq,ij,uv}(h, r) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E}[\mathbf{Z}_{t,i} \mathbf{q}_{2,i,u} \mathbf{q}_{2,j,v}' \mathbf{Z}_{t,j}' I_{t,i}(r) I_{t+h,j}(r_0)], \\ \Omega_{x,ij,uv}(h, r) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E}[\mathbf{x}_{t,u} \mathbf{x}_{t+h,v}' I_{t,i}(r)], \end{aligned}$$

for $i, j = 1, 2$, and $u, v = 1, \dots, p_2$, where $I_{t,1}(r) = I(z < r)$ and $I_{t,2}(r) = I(z \geq r)$.

Since the noise process is independent over time, if $r = r_0$, when $h > 0$, we have

$$\Omega_{x,ij,uv}(h, r_0) = \mathbf{Q}_{1,i} [\Omega_{zq,i1,uv}(h, r_0) + \Omega_{zq,i2,uv}(h, r_0)] \mathbf{Q}_{1,i}'. \quad (6)$$

For a pre-determined positive integer h_0 , define

$$\mathbf{M}_{1,i}(r) = \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \boldsymbol{\Omega}_{x,i,uv}(h, r) \boldsymbol{\Omega}_{x,i,uv}(h, r)', \text{ for } i = 1, 2. \quad (7)$$

By equation (6), if $r = r_0$, it follows that

$$\mathbf{M}_{1,i}(r_0) = \mathbf{Q}_{1,i} \left[\sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \left(\sum_{j=1}^2 \boldsymbol{\Omega}_{zq,ij,uv}(h, r_0) \right) \left(\sum_{j=1}^2 \boldsymbol{\Omega}_{zq,ij,uv}(h, r_0) \right)' \right] \mathbf{Q}_{1,i}'. \quad (8)$$

$\mathbf{M}_{1,i}(r_0)$ is a symmetric non-negative definite matrix sandwiched by $\mathbf{Q}_{1,i}$ and $\mathbf{Q}_{1,i}'$. If the matrices in the parentheses of (8) are full rank, the eigenvectors of $\mathbf{M}_{1,i}(r_0)$ corresponding to the nonzero eigenvalues span the row loading space in regime i . Hence, $\mathcal{M}(\mathbf{Q}_{1,i})$ can be estimated by the eigen-decomposition of sample version of $\mathbf{M}_{1,i}(r_0)$. Let $\mathbf{q}_{1,i,k}$ be the unit eigenvector of $\mathbf{M}_{1,i}(r_0)$ corresponding to the k -th largest eigenvalue, and we can now uniquely define $\mathbf{Q}_{1,i}$ by

$$\mathbf{Q}_{1,i} = (\mathbf{q}_{1,i,1}, \dots, \mathbf{q}_{1,i,k_{1,i}}).$$

Now we define the sample version of the above statistics.

$$\hat{\boldsymbol{\Omega}}_{x,i,uv}(h, r) = \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{x}_{t,u} \mathbf{x}_{t+h,v}' I_{t,i}(r), \quad (9)$$

$$\widehat{\mathbf{M}}_{1,i}(r) = \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \hat{\boldsymbol{\Omega}}_{x,i,uv}(h, r) \hat{\boldsymbol{\Omega}}_{x,i,uv}'(h, r), \quad (10)$$

for $i = 1, 2$. Let $\hat{\mathbf{q}}_{1,i,k}$ be the unit eigenvector of $\widehat{\mathbf{M}}_{1,i}(r_0)$ corresponding to the k -th largest eigenvalue. Then if the threshold value r_0 is known, the row loading space in regime i can be estimated by

$$\widehat{\mathcal{M}}(\mathbf{R}_i) = \mathcal{M}(\hat{\mathbf{Q}}_{1,i}),$$

where $\hat{\mathbf{Q}}_{1,i} = (\hat{\mathbf{q}}_{1,i,1}, \dots, \hat{\mathbf{q}}_{1,i,k_1})$. For the column loading spaces, they can be estimated by performing the same procedure on the transposes of $\mathbf{X}_t'$ s to construct $\mathbf{M}_{2,i}(r)$, for $i = 1, 2$.

Remark 3. The estimation procedure here is a generalized version of that in [Liu and Chen \(2020\)](#). However, [Liu and Chen \(2020\)](#) divides the observed process with two tentative threshold values and defines the second cross moment matrices of the observed process in four partitions. Although their model allows overlap between loading spaces in two regimes in Condition 9, the effective sample size is much smaller; See Remark 4. Another advantage of our method is that we can deal with the case when the numbers of factors vary across regimes, while [Liu and Chen \(2020\)](#) only discusses the scenario when the number of factors remains the same.

2.3 Threshold Estimation

In this section, we construct a project-based objective function following the method in [Liu and Chen \(2020\)](#) for threshold value estimation, and provide the estimators for loading spaces when the threshold value is unknown. The theoretical properties of the estimators will also be studied.

Let $\mathbf{B}_{s,i} = (\mathbf{q}_{s,i,k_i+1}, \dots, \mathbf{q}_{s,i,p_i})$ be a $p_s \times (p_s - k_{s,i})$ matrix, whose columns consist of the unit eigenvectors of $\mathbf{M}_{s,i}(r_0)$ corresponding to zero eigenvalues for $s, i = 1, 2$. $\mathcal{M}(\mathbf{B}_{s,i})$ is the complement of loading space $\mathcal{M}(\mathbf{Q}_{s,i})$ and $\mathbf{Q}'_{s,i}\mathbf{B}_{s,i} = \mathbf{0}$, for $s, i = 1, 2$. Define the objective function

$$G(r) = \sum_{s=1}^2 \sum_{i=1}^2 \|\mathbf{B}'_{s,i} \mathbf{M}_{s,i}(r) \mathbf{B}_{s,i}\|_2. \quad (11)$$

By the definition of $\mathbf{M}_{s,i}(r)$, we can tell that $G(r)$ measures the sum of the squared norm of the projections of $\boldsymbol{\Omega}_{x,i,uv}(h, r)$ onto the complement of loading spaces, $\mathcal{M}(\mathbf{B}_{s,i})$, for $h = 1, \dots, h_0$, $u, v = 1, \dots, p_i$, and $s, i = 1, 2$.

If $r = r_0$, the observations in two regimes are correctly classified into different subsets. Then by (8), $\mathbf{M}_{s,i}(r)$ is sandwiched by $\mathbf{Q}_{s,i}$ and $\mathbf{Q}'_{s,i}$. Hence, $G(r) = 0$. However, if $r \neq r_0$, the observations from one regime are misclassified into two subsets, and one of the two subsets is mixed. $\mathbf{M}_{s,i}(r)$ is not sandwiched by $\mathbf{Q}_{s,i}$ and $\mathbf{Q}'_{s,i}$, and the projection is nonzero.

A standard assumption for threshold variable estimation is imposed which is that r_0 is in a known region of the support of z_t , $r_0 \in (\eta_1, \eta_2)$, where η_1 and η_2 are called trimming parameters. We use data corresponding to $z_t \leq \eta_1$ and $z_t \geq \eta_2$ to estimate $\mathcal{M}(\mathbf{B}_{s,1})$ and $\mathcal{M}(\mathbf{B}_{s,2})$, respectively, for $s = 1, 2$. The sample version of $G(r)$ is defined as

$$\hat{G}(r) = \sum_{s=1}^2 \sum_{i=1}^2 \|\hat{\mathbf{B}}'_{s,i}(\eta_i) \hat{\mathbf{M}}_{s,i}(r) \hat{\mathbf{B}}_{s,i}(\eta_i)\|_2,$$

where

$$\hat{\boldsymbol{\Omega}}_{x,i,uv}(h, \eta_i) = \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{x}_{t,u} \mathbf{x}'_{t+h,v} I_{t,i}(\eta_i), \quad \hat{\mathbf{M}}_{1,i}(\eta_i) = \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \hat{\boldsymbol{\Omega}}_{x,i,uv}(h, \eta_i) \hat{\boldsymbol{\Omega}}_{x,i,uv}(h, \eta_i)',$$

$\hat{\mathbf{M}}_{2,i}(\eta_i)$ is defined in a similar way with transposes of $\mathbf{X}_t$'s, $\hat{\mathbf{B}}_{s,i}(\eta_i) = (\hat{\mathbf{q}}_{s,i,k_i+1}(\eta_i), \dots, \hat{\mathbf{q}}_{s,i,p_i}(\eta_i))$, and $\hat{\mathbf{q}}_{s,i,k}(\eta_i)$ is the k -th largest eigenvector of $\hat{\mathbf{M}}_{s,i}(\eta_i)$.

We estimate r_0 by

$$\hat{r} = \arg \min_{r \in \{z_1, \dots, z_T\} \cap (\eta_1, \eta_2)} \hat{G}(r).$$

Remark 4. Compared with the method by [Liu and Chen \(2020\)](#) using both trimming parameters to estimate the complement loading space in each regime, here we estimate $\mathbf{B}_{s,i}$ with only one trimming parameter utilizing more information from data. Let us say η_1 and η_2 are the 10-th and 90-th quantiles of the threshold variable, and then we use 10% of observations to estimate $\mathbf{B}_{s,i}$ in (11), while [Liu and Chen \(2020\)](#) only uses 1%. The effective sample size for the estimation of $\mathbf{B}_{s,i}$ is much smaller for [Liu and Chen \(2020\)](#). The simulation results in Section 3.2 confirm that our methods outperform the methods by [Liu and Chen \(2020\)](#).

The following regularity conditions are needed to obtain the theoretical properties.

Condition A1. The process $(\mathbf{F}_{t,1}, \mathbf{F}_{t,2}, z_t)$ 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 $\{(\mathbf{F}_{t,1}, \mathbf{F}_{t,2}, z_t) : i \leq t \leq j\}$.

Define

$$\begin{aligned} \Sigma_{f,ij}(h) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E}[\text{vec}(\mathbf{F}_{t,i}) \text{vec}(\mathbf{F}_{t+h,j})'], \\ \Sigma_{f,ij}(h, r) &= \frac{\frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E}[\text{vec}(\mathbf{F}_{t,i}) \text{vec}(\mathbf{F}_{t+h,j})' I_{t,i}(r) I_{t+h,j}(r_0)]}{\mathbb{E}[I_{t,i}(r) I_{t+h,j}(r_0)]}. \end{aligned}$$

Condition A2. For any $i = 1, 2$, $u = 1, \dots, k_{1,i}$, $v = 1, \dots, k_{2,i}$, and $t = 1, \dots, T$, $\mathbb{E}(|f_{t,i,uv}|^{4\gamma}) \leq \sigma_f^{4\gamma}$, where σ_f is a positive constant and γ is given in Condition A1. There exists an $h \in [1, h_0]$ such that $\text{rank}(\Sigma_{f,ij}(h)) \geq k_{\max}$, and $\sigma_{k_{\max}}(\Sigma_{f,ij}(h))$ is uniformly bounded, as p_1 and p_2 go to infinity, where $k_{\max} = \max\{k_{1,1}, k_{1,2}, k_{2,1}, k_{2,2}\}$. For $i = 1, 2$, $u = 1, \dots, k_{i,1}$ and $v = 1, \dots, k_{i,2}$, $\frac{1}{T-h} \sum_{t=1}^{T-h} \text{Cov}(\mathbf{f}_{t,i,u}, \mathbf{f}_{t+h,i,u}) \neq \mathbf{0}$, $\frac{1}{T-h} \sum_{t=1}^{T-h} \text{Cov}(\mathbf{f}_{t,i,v}, \mathbf{f}_{t+h,i,v}) \neq \mathbf{0}$.

Condition A3. $\text{Var}(\mathbf{E}_t) = \Sigma_{t,e}$ for $t = 1, \dots, T$. The absolute value of each element in $\Sigma_{t,e}$ remains bounded by σ_e^2 as p_1 and p_2 increase to infinity for $t = 1, \dots, T$. $\text{Cov}(\text{vec}(\mathbf{E}_{t_1}), \text{vec}(\mathbf{F}_{t_2,i})) = \mathbf{0}$ and $\text{Cov}(\text{vec}(\mathbf{E}_{t_1}), \text{vec}(\mathbf{E}_{t_2})) = \mathbf{0}$ for $t_1, t_2 = 1, \dots, T$ and $i = 1, 2$.

Condition A4. For $i = 1, 2$, there exist constants δ_1 and $\delta_2 \in [0, 1]$ such that $\|\mathbf{R}_i\|_2^2 \asymp p_1^{1-\delta_1} \asymp \|\mathbf{R}_i\|_{\min}^2$ and $\|\mathbf{C}_i\|_2^2 \asymp p_2^{1-\delta_2} \asymp \|\mathbf{C}_i\|_{\min}^2$, as p_1 and p_2 go to infinity.

When studying high-dimensional time series, it is common to assume the norm of the loading matrices grows as the dimension increases to infinity. We use the strength of factors δ_i defined in [Lam et al. \(2011\)](#) and [Lam and Yao \(2012\)](#) to measure the growth rate for $s, i = 1, 2$. It reflects the relative growth rate of information about $\mathbf{F}_{t,i}$ carried by $\mathbf{X}_t$ as the dimensions increase, comparing

to the growth rate of the noise process. For example, if $\delta_1 = 0$, the row factors are strong, and observations are fully loaded with signal as p_1 increases. If $\delta_1 = 1$, the row factors are extremely weak, and only noise is added to the observed data as p_1 increases.

Condition A5. $\mathbf{M}_{s,i}$ has $k_{s,i}$ distinct positive eigenvalues for $s, i = 1, 2$.

Conditions A1-A5 are similar to the standard assumptions for factor models (Lam et al., 2011; Lam and Yao, 2012; Wang et al., 2019). Like in Liu and Chen (2020) and Liu and Zhang (2022), here a more general setting is considered, and we only require the factor process to satisfy the mixing condition and the moments of the factor and noise processes to be bounded. Heteroskedasticity in the $\mathbf{F}_{t,1}$, $\mathbf{F}_{t,2}$, or $\mathbf{E}_t$ is allowed not only through their cross-sectional dimension but also the time dimension; See simulation results in Section 3.3.

Condition B1. Assume $r_0 \in (\eta_1, \eta_2)$. z_t is a continuous random variable, and the process $\{z_t\}$ is stationary. The marginal probability of z_t satisfies that $P(z_t \leq \eta_1) > 0$ and $P(z_t \geq \eta_2) > 0$. For the density function of z_t , $f(z_t)$, there exist two positive constants τ_1 and τ_2 such that $\tau_2 \leq f(z_t) \leq \tau_1$ uniformly in $[\eta_1, \eta_2]$. The conditional probability of z_{t+h} given z_t satisfies that $P(z_{t+h} < \eta_1 \mid z_t) > \pi_1 > 0$ and $P(z_{t+h} > \eta_2 \mid z_t) > \pi_2 > 0$ for any $z_t \in (-\infty, \eta_1)$ or $(\eta_2, +\infty)$ and $h = 1, \dots, h_0$.

Condition B2. There exists a positive integer $\tilde{h}_i \in [1, h_0]$ such that $\text{rank}(\Sigma_{f,ij}(\tilde{h}_i, \eta_i)) \geq k_{\max}$ and $\|\Sigma_{f,ij}(\tilde{h}_i, \eta_i)\|_{\min}$ is uniformly bounded above 0, for $i, j = 1, 2$.

Condition B3. $\mathbf{M}_{s,i}(\eta_i)$ has $k_{s,i}$ distinct positive eigenvalues for $s, i = 1, 2$.

We define an indicator function for regime 1, $I^R(\cdot)$. For a constant c , $I^R(r) = 1$ if $r \leq r_0$, and $I^R(c) = 2$ if $r > r_0$. For c_1 and c_2 satisfying $I^R(c_1) = I^R(c_2)$, let

$$\Sigma_{f,ij}(h, c_1, c_2) = \frac{\frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E}[\text{vec}(\mathbf{F}_{t,i}) \text{vec}(\mathbf{F}_{t+h,j})' I(c_1 < z_t < c_2)]}{\mathbb{E}[I(c_1 < z_t < c_2)]}.$$

Condition B4. For any $r \in (\eta_1, r_0)$, there exists an integer $h_1^* \in [1, h_0]$ such that $\text{rank}(\Sigma_{f,1j}(h_1^*, r, r_0)) \geq k_{\max}$, for $j = 1, 2$. For any $r \in (r_0, \eta_2)$, there exists an integer $h_2^* \in [1, h_0]$ such that $\text{rank}(\Sigma_{f,2j}(h_2^*, r_0, r)) \geq k_{\max}$, for $j = 1, 2$. The minimum nonzero singular values of these aforementioned matrices are all uniformly bounded above γ_0 , where $\gamma_0 > 0$.

Condition B5. There exists a constant $v < 1$ such that $\|\mathbf{Q}'_{s,1} \mathbf{Q}_{s,2}\|_2 < v$ for $s = 1, 2$ as p_1 and p_2 grow to infinity.

Condition B1-3 indicate that the estimators for loading spaces are consistent when only data with $\{z_t \in (-\infty, \eta_1)\}$ or $\{z_t \in (\eta_2, +\infty)\}$ are used. Condition B4 guarantees that the cross moment

matrices of $\text{vec}(\mathbf{F}_{t,i})$ with mixed data from two regimes carry the information on the latent factor process in both regimes. Condition B5 is stronger than the Condition 9 in [Liu and Chen \(2020\)](#) which requires the distance of loading space to be greater than 0. This is because we estimate loading spaces with only one trimming parameter. $\|\mathbf{Q}'_{s,1}\mathbf{Q}_{s,2}\|_2 < 1$ ensures that $\dim(\mathcal{M}(\mathbf{Q}_{2,1}) \cap \mathcal{M}(\mathbf{Q}_{s,2})) = 0$ and signals are not cancelled out when we combine data transitioning from regime i to both regimes, $i = 1, 2$.

Before presenting the asymptotic properties of the proposed estimators, we introduce a measure to quantify the distance of two linear spaces, which is first proposed by [Liu and Chen \(2020\)](#). Let $\mathbf{S}_1$ be a $p \times q_1$ full-rank matrix, and $\mathbf{S}_2$ be a $p \times q_2$ full-rank matrix, where $p \geq q_1, q_2$. Let $\mathbf{O}_i$ be an orthogonal representative of $\mathcal{M}(\mathbf{S}_i)$, i.e., $\mathcal{M}(\mathbf{O}_i) = \mathcal{M}(\mathbf{S}_i)$ and $\mathbf{O}_i'\mathbf{O}_i = \mathbf{I}_{q_i}$, for $i = 1, 2$. The distance of $\mathcal{M}(\mathbf{S}_1)$ and $\mathcal{M}(\mathbf{S}_2)$ is defined as

$$\mathcal{D}(\mathcal{M}(\mathbf{S}_1), \mathcal{M}(\mathbf{S}_2)) = \sqrt{1 - \frac{\text{tr}(\mathbf{O}_1\mathbf{O}_1'\mathbf{O}_2\mathbf{O}_2')}{\min\{q_1, q_2\}}}.$$

It is a quantity between 0 and 1. It is 1 if and only if $\mathcal{M}(\mathbf{S}_1) \perp \mathcal{M}(\mathbf{S}_2)$, and is 0 if and only if $\mathcal{M}(\mathbf{S}_1) \subseteq \mathcal{M}(\mathbf{S}_2)$ or $\mathcal{M}(\mathbf{S}_2) \subseteq \mathcal{M}(\mathbf{S}_1)$.

Theorem 1. *Under Conditions A1-A5 and B1-B5, when the numbers of factors are known, if $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2} = o(1)$, as $p_1, p_2, T \rightarrow \infty$, it holds that*

$$P(\hat{r} < r_0 - \epsilon) \leq \frac{Cp_1^{\delta_1}p_2^{\delta_2}}{\epsilon T^{1/2}}, \quad P(\hat{r} > r_0 + \epsilon) \leq \frac{Cp_1^{\delta_1}p_2^{\delta_2}}{\epsilon T^{1/2}},$$

for $\epsilon > 0$.

Theorem 1 shows that the estimator $\hat{r}$ is consistent under some mild conditions. $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2} = o(1)$ is a standard assumption used to prove the consistency for loading space estimation in matrix factor models ([Wang et al., 2019](#)). Hence, it is also needed in model (1) to distinguish two regimes. If all factors are strong with $\delta_1 = \delta_2 = 0$, $P(\hat{r} < r_0 - \epsilon)$ and $P(\hat{r} > r_0 + \epsilon)$ reduce to $O(T^{-1/2})$. The estimation error for r_0 is asymptotically immune to the increase of p_1 and p_2 . When p_1 and p_2 grow, the curse of dimensionality is offset by the information brought by new incoming series. If row and column factors are weak, the convergence rate in Theorem 1 depends on the cross-section dimension. The estimator gets less efficient when p_1 and p_2 increase.

Theorem 1 shows that as long as $T^{1/2}$ increases faster than $p_1^{\delta_1}p_2^{\delta_2}$ asymptotically, $\hat{r}$ converges to r_0 . It is worth noting that the approximate factor model by [Bai \(2003\)](#) requires δ_1 and δ_2 to be less than 1/2 to obtain consistent estimators ([Bates et al., 2013](#); [Bailey et al., 2016](#); [Massacci,](#)

2017). However, [Lam et al. \(2011\)](#) does not put any constraints on the strength of factors. [Lam and Yao \(2012\)](#) and [Liu and Chen \(2020\)](#) did numerical experiments with factor strength equal to 0.5 and 1 respectively, and the results confirm that the estimators are consistent when factor strength is equal to or greater than 0.5. The simulation results in Section 3.2 also proves that our method performs well for finite samples when $\delta_1 = \delta_2 = 0.5$.

The final estimation of loading spaces is obtained using $\hat{r}$ as the threshold value and following the procedure in Section 3.3.

Theorem 2. *Under Conditions A1-A5 and B1-B5, when the numbers of factors are known if $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$, as $p_1, p_2, T \rightarrow \infty$, it holds that*

$$\mathcal{D}(\mathcal{M}(\hat{\mathbf{Q}}_{s,i}(\hat{r})), \mathcal{M}(\mathbf{Q}_{s,i})) = O_p(p_1^{\delta_1} p_2^{\delta_2} T^{-1/2}), \text{ for } s, i = 1, 2.$$

From Theorem 2 we can see that the asymptotics of the loading space estimators are the same with the case when there is only one regime discussed in [Wang et al. \(2019\)](#).

2.4 When the numbers of factors are unknown

Since both the factors and loadings are unobserved, the numbers of factors in each regime need to be estimated. [Lam and Yao \(2012\)](#) proposed a ratio-based estimator, and [Wang et al. \(2019\)](#) extended it to analyze matrix-variate time series. Here we apply it for threshold matrix-variate factor models. Assume that r_0 is in a known interval (η_1, η_2) , and let

$$\hat{k}_{s,i} = \arg \min_{1 \leq k \leq R} \frac{\hat{\lambda}_{s,i,k+1}(\eta_i)}{\hat{\lambda}_{s,i,k}(\eta_i)}, \text{ for } s, i = 1, 2, \quad (12)$$

where $\hat{\lambda}_{s,i,k}(\eta_i)$ is the k -th largest eigenvalue of $\widehat{\mathbf{M}}_{s,i}(\eta_i)$. Note that the eigenvalues practically will go to zero and here we cannot search up to p_1 or p_2 . We use $R = \lfloor p_s/2 \rfloor$ following [Lam and Yao \(2012\)](#) and $R = \lfloor T/2 \rfloor$ when $T < p_s$, for $s = 1, 2$.

Corollary 1. *Under Conditions A1-A5 and B1-B5, if $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$, as $p_1, p_2, T \rightarrow \infty$, it holds that*

$$\begin{aligned} \hat{\lambda}_{s,i,k+1}(\eta_i) / \hat{\lambda}_{s,i,k}(\eta_i) &\asymp 1, \text{ for } k = 1, \dots, k_{s,i} - 1, \\ \hat{\lambda}_{s,i,k_{s,i}+1}(\eta_i) / \hat{\lambda}_{s,i,k_{s,i}}(\eta_i) &= O_p(p_1^{\delta_1} p_2^{\delta_2} T^{-1}) \xrightarrow{p} 0, \text{ for } s, i = 1, 2. \end{aligned}$$

Corollary 1 presents the convergence rates of the ratios of eigenvalues of $\widehat{\mathbf{M}}_{s,i}(\eta_i)$, and indicates that the estimated eigenvalue ratio will drop sharply at $k = k_{s,i}$, which provides a theoretical underpinning for the estimator of $k_{s,i}$. When $k > k_{s,i}$, the eigenvalue $\lambda_{s,i,k}$ is theoretically zero and

thus the property of the ratio $\hat{\lambda}_{s,i,k_{s,i}+1}(\eta_i)/\hat{\lambda}_{s,i,k_{s,i}}(\eta_i)$ is difficult to obtain. Lam and Yao (2012) only proves that the probability to underestimate the number of factors goes to zero asymptotically. Although the consistency of the ratio-based estimator cannot be obtained, the method performs well in numerical experiments; See examples in Chang et al. (2015); Liu and Chen (2016); Wang et al. (2019); Liu and Chen (2020); Liu and Zhang (2022).

Since the probability of overestimating the numbers of factors cannot be proved to converge to 0, we shall discuss the theoretical properties of proposed estimators when the numbers of factors are overestimated in the following. The overestimation issue for the factor number in the approximate factor model is not new and has been considered by Barigozzi et al. (2018) and Barigozzi and Cho (2020). Barigozzi et al. (2018) uses wavelet to detect the structural changes in the observed process, which circumvents the issues of accurate estimation of the number of factors. Barigozzi and Cho (2020) shows that the errors in the principal component estimators for the common component resulting from the overestimated factor number can be non-negligible. Under the settings by Lam et al. (2011), the estimator for the common component may be inconsistent as well when the loading spaces are spuriously enlarged, which needs significant investigation that is beyond the scope of this paper. Even so the proposed estimator for r is still consistent with some regularity conditions which will be shown in the following. The reason is that the classifier used to identify two regimes in (11) is built on $\mathbf{B}_{s,i}$. If factor numbers are overestimated, even though the column space spanned by estimated $\mathbf{B}_{s,i}$ is a proper subspace of $\mathcal{M}(\mathbf{B})_{s,i}$, it still can differentiate data from two regimes under mild conditions.

Let

$$\hat{G}(r, k_{1,1}, k_{1,2}, k_{2,1}, k_{2,2}) = \sum_{s=1}^2 \sum_{i=1}^2 \|\hat{\mathbf{B}}'_{s,i,k_{s,i}}(\eta_i) \hat{\mathbf{M}}_{s,i}(r) \hat{\mathbf{B}}_{s,i,k_{s,i}}(\eta_i)\|_2, \quad (13)$$

where $\hat{\mathbf{B}}_{s,i,k_{s,i}}(\eta_i) = (\hat{\mathbf{q}}_{s,i,k_{s,i}+1}(\eta_i), \dots, \hat{\mathbf{q}}_{s,i,p_s}(\eta_i))$, for $s, i = 1, 2$. When the numbers of factors are unknown, we estimate r_0 by

$$\tilde{r} = \arg \min_{r \in \{z_1, \dots, z_T\} \cap (\eta_1, \eta_2)} \hat{G}(r, \hat{k}_{1,1}, \hat{k}_{1,2}, \hat{k}_{2,1}, \hat{k}_{2,2}). \quad (14)$$

If the numbers of factors are overestimated, Condition B6 ensures that the two spuriously enlarged loading spaces $\mathcal{M}(\mathbf{Q}_{s,1}^*)$ and $\mathcal{M}(\mathbf{Q}_{s,2}^*)$ are still differentiable for $s = 1, 2$.

Condition B6. When $\hat{k}_{s,i} > k_{s,i}$ there exists a positive constant $\tilde{d}$ such that $\mathcal{D}(\mathcal{M}(\mathbf{Q}_{s,1}^*), \mathcal{M}(\mathbf{Q}_{s,2})) > \tilde{d}$ and $\mathcal{D}(\mathcal{M}(\mathbf{Q}_{s,1}), \mathcal{M}(\mathbf{Q}_{s,2}^*)) > \tilde{d}$, as p_s goes to infinity, where $\mathbf{Q}_{s,i}^* = (\mathbf{Q}_{s,i}, \mathbf{U}_{s,i})$ is a $p_s \times \hat{k}_{s,i}$ matrix, for $s, i = 1, 2$ and any $p_s \times (\hat{k}_{s,i} - k_{s,i})$ matrix $\mathbf{U}_{s,i}$ such that $\dim(\mathcal{M}(\mathbf{U}_{s,i}) \cap \mathcal{M}(\mathbf{Q}_{s,i})) = 0$,

Theorem 3. *Under Conditions A1-A5 and B1-B6, if $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$, $\hat{k}_{s,i} \geq k_{s,i}$ for $s, i = 1, 2$, as $p_1, p_2, T \rightarrow \infty$, it holds that*

$$P(\tilde{r} < r_0 - \epsilon) \leq \frac{C p_1^{\delta_1} p_2^{\delta_2}}{\epsilon T^{1/2}}, \quad P(\tilde{r} > r_0 + \epsilon) \leq \frac{C p_1^{\delta_1} p_2^{\delta_2}}{\epsilon T^{1/2}},$$

for $\epsilon > 0$.

The loading spaces are estimated using $\hat{k}_{s,i}$ as the number of factors and $\tilde{r}$ as the threshold value,

$$\tilde{\mathbf{Q}}_{s,i}(\tilde{r}, \hat{k}_{s,i}) = (\hat{\mathbf{q}}_{s,i,1}(\tilde{r}), \dots, \hat{\mathbf{q}}_{s,i,\hat{k}_{s,i}}(\tilde{r})), \text{ for } s, i = 1, 2. \quad (15)$$

Define $\tilde{\mathbf{Q}}_{s,i}(\tilde{r})$ as the matrix which consists of the first $k_{s,i}$ columns of $\tilde{\mathbf{Q}}_{s,i}(\tilde{r}, \hat{k}_{s,i})$. The following theorem indicates that its spanned space converges to the true loading space as fast as $\mathcal{M}(\hat{\mathbf{Q}}_{s,i}(\hat{r}))$ shown in Theorem 2 asymptotically,

$$\tilde{\mathbf{Q}}_{s,i}(\tilde{r}) = (\hat{\mathbf{q}}_{s,i,1}(\tilde{r}), \dots, \hat{\mathbf{q}}_{s,i,k_{s,i}}(\tilde{r})), \text{ for } s, i = 1, 2.$$

Theorem 4. *Under Conditions A1-A5 and B1-B6, if $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$, $\hat{k}_{s,i} \geq k_{s,i}$ for $s, i = 1, 2$, as $p_1, p_2, T \rightarrow \infty$, it holds that*

$$\mathcal{D}(\mathcal{M}(\tilde{\mathbf{Q}}_{s,i}(\tilde{r})), \mathcal{M}(\mathbf{Q}_{s,i})) = O_p(p_1^{\delta_1} p_2^{\delta_2} T^{-1/2}), \text{ for } s, i = 1, 2.$$

Theorems 3 and 4 tell that when the numbers of factors are overestimated, our estimators perform asymptotically as good as those when the number of factors are correctly estimated under mild conditions. The consistency of the estimators shown in Theorems 3 and 4 helps us build a complete theoretical framework for the estimation of threshold value and loading spaces since the numbers of factors could be overestimated theoretically (Lam and Yao, 2012). One may argue that in practice we do not know whether the numbers of factors are correctly estimated or overestimated. However, compared with the estimators proposed by Massacci (2017) in which the theoretical properties were investigated based on the assumption that the number of factor is known or correctly estimated, here the proposed estimators are more robust.

2.5 Model identification for models with multiple regimes

In this section, we will consider the threshold matrix-variate factor models with multiple regimes, and develop an algorithm to identify the regimes and estimate multiple threshold values. The

following m -regime threshold matrix-variate factor model is considered:

$$\mathbf{X}_t = \begin{cases} \mathbf{R}_1 \mathbf{F}_{t,1} \mathbf{C}'_1 + \mathbf{E}_t & z_t < r_0, \\ \mathbf{R}_2 \mathbf{F}_{t,2} \mathbf{C}'_2 + \mathbf{E}_t & r_0 \leq z_t < r_1, \\ \dots & \\ \mathbf{R}_m \mathbf{F}_{t,m} \mathbf{C}'_m + \mathbf{E}_t & r_{m-2} \leq z_t < r_{m-1} = +\infty, \end{cases} \quad (16)$$

for $t = 1, \dots, T$, where $0 < m \ll T$. Let $\mathbf{Q}_{1,i}$ and $\mathbf{Q}_{2,i}$ be the orthonormal representatives of $\mathcal{M}(\mathbf{R}_i)$ and $\mathcal{M}(\mathbf{C}_i)$ for $i = 1, \dots, m$. For any $i \neq j$, $\|\mathbf{Q}'_{1,i} \mathbf{Q}_{1,j}\|_2 < 1$ or $\|\mathbf{Q}'_{2,i} \mathbf{Q}_{2,j}\|_2 < 1$. When $m = 1$, model (16) is a one-regime matrix-variate factor model introduced by Wang et al. (2019); When $m = 2$, model (16) is the one we discuss in Section 2.1. Our aim is to estimate m , threshold values $\{r_i \mid i = 0, \dots, m-2\}$, and loading spaces in different regimes. Motivated by the idea that the numbers of factors will be overestimated when fitting data from multiple regimes with one-regime factor models (Chen et al., 2014; Wu, 2021; Liu and Zhang, 2022), we consider fitting data from multiple regimes with one-regime models first. We shall use the simple example with $m = 2$ to illustrate the idea. Without loss of generality, we assume row loading spaces are different with $\|\mathbf{Q}'_{1,1} \mathbf{Q}_{1,2}\|_2 < 1$. Define

$$\mathbf{\Omega}_{x,uv}(h) = \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{E}(\mathbf{x}_{t,u} \mathbf{x}'_{t+h,v}), \quad \mathbf{M} = \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbf{\Omega}_{x,uv}(h) \mathbf{\Omega}_{x,uv}(h)', \quad (17)$$

$$\widehat{\mathbf{\Omega}}_{x,uv}(h) = \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{x}_{t,u} \mathbf{x}'_{t+h,v}, \quad \widehat{\mathbf{M}} = \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \widehat{\mathbf{\Omega}}_{x,uv}(h) \widehat{\mathbf{\Omega}}_{x,uv}(h)'.$$

With the following conditions, we shall show that $\widehat{\mathbf{M}}$ may lead to an overestimated number of row factors.

Condition B7. The nonzero eigenvalues of $\mathbf{M}$ are distinct.

Corollary 2. Under Conditions A1-A5 and B1-B7, if $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} = o(1)$, as $p_1, p_2, T \rightarrow \infty$, it holds that

$$\begin{aligned} \widehat{\lambda}_{k+1}/\widehat{\lambda}_k &\asymp 1, \quad \text{for } k = 1, \dots, k_{1,1} + k_{1,2} - 1, \\ \widehat{\lambda}_{k_1+k_2+1}/\widehat{\lambda}_{k_1+k_2} &= O_p(p_1^{2\delta_1} p_2^{2\delta_2} T^{-1}) \xrightarrow{p} 0, \end{aligned}$$

where $\widehat{\lambda}_k$ is the eigenvector of $\widehat{\mathbf{M}}$ corresponding to the k -th largest eigenvalue.

Corollary 2 indicates that the ratio of the estimated eigenvalues will drop sharply at $k = k_1 + k_2$ if we combine data from two regimes, which has the same pattern with one-regime models shown

in Corollary 1 in [Lam and Yao \(2012\)](#) for vector time series and Theorem 2 in [Wang et al. \(2019\)](#) for matrix time series. Motivated by this observation, we extend the algorithm developed by [Wu \(2021\)](#) and propose a regime-switching identification procedure which applies to matrix time series and can deal with the case that the numbers of factors are different across regimes.

First we divide the range of the threshold variable $\{z_t\}$ into J subintervals, $S_1, \dots, S_J$, where J is a pre-specified integer satisfying $m \ll J \ll T$ and $S_j = (s_{j-1}, s_j]$. Specifically, let $s_j = z_{(t_j)}$ for $j = 0, \dots, J$, where $t_j = \lfloor \frac{jT}{J} \rfloor$ and $z_{(t_j)}$ is the (t_j) -th order statistic of z_t . For each subinterval, we define

$$\hat{\Omega}_{x,j,uv}^J(h) = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{t,u} \mathbf{x}_{t+h,v}' I(s_{j-1} < z_t \leq s_j), \quad \widehat{\mathbf{M}}_{1,j}^J = \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \hat{\Omega}_{x,j,uv}^J(h) \hat{\Omega}_{x,j,uv}^J(h)',$$

for $j = 1, \dots, J$. $\widehat{\mathbf{M}}_{2,j}^J$ can be defined in a similar way with transposes of $\mathbf{X}_t$'s. The numbers of row and column factors in the j -th subinterval can be estimated by

$$\hat{k}_{1,j}^J = \arg \min_{1 \leq k \leq R} \frac{\hat{\lambda}_{1,j,k+1}^J}{\hat{\lambda}_{1,j,k}^J}, \quad \hat{k}_{2,j}^J = \arg \min_{1 \leq k \leq R} \frac{\hat{\lambda}_{2,j,k+1}^J}{\hat{\lambda}_{2,j,k}^J},$$

where $\hat{\lambda}_{s,j,k}^J$ is the k -th largest eigenvalue of $\widehat{\mathbf{M}}_{s,j}^J$, for $s = 1, 2$ and $j = 1, \dots, J$. We define an indicator function I_j^J to track the possible changes of the numbers of factors among these subintervals. Let $I_j^J = 1$ if $\hat{k}_{1,j}^J \neq \hat{k}_{1,j-1}^J$ or $\hat{k}_{2,j}^J \neq \hat{k}_{2,j-1}^J$; Otherwise $I_j^J = 0$. There are three situations that can happen for I_j^J :

- (i) When $I_{j-1}^J = 0$, $I_j^J = 1$, $I_{j+1}^J = 1$ and $I_{j+2}^J = 0$, the regime switching happens in the interior of S_j .
- (ii) When $I_{j-1}^J = 0$ and $I_j^J = 0$, no regime switching happens in the interior of S_j , or the regime switching may happen in the left end of S_j or the right end of S_{j-1} and the numbers of factors are the same in two regimes.
- (iii) When $I_{j-1}^J = 0$, $I_j^J = 1$, and $I_{j+1}^J = 0$, the regime switching happens in the left end of S_j or the right end of S_{j-1} , and the numbers of factors are different in two regimes.

The case when $I_{j-1}^J = 1$ is not included here because it shall be considered for I_{j-1}^J ; the case when $I_{j-1}^J = 0$, $I_j^J = 1$, $I_{j+1}^J = 1$, and $I_{j+2}^J = 1$ is not discussed since we assume $m \gg J$ implying that there is at most one regime switching in two consecutive subintervals.

For case (i), the complement loading spaces in (11) can be estimated using eigenvectors of $\widehat{\mathbf{M}}_{s,j-1}^J$ corresponding to the smallest $(p_s - \hat{k}_{s,j-1}^J)$ eigenvalues and eigenvectors of $\widehat{\mathbf{M}}_{s,j+1}^J$ corresponding to the smallest $(p_s - \hat{k}_{s,j+1}^J)$ eigenvalues for $s = 1, 2$, and then the estimators for the

threshold value and loading spaces can be obtained by the method in Section 2.3. For case (ii), to detect if the regime switching happens in the left end of S_j or the right end of S_{j-1} , we use the midpoints of the J subintervals to re-divide the range of z_t into $J+1$ subintervals, $S_j^* = (s_{j-1}^*, s_j^*]$ for $j = 1, \dots, J+1$, where $s_j^* = z_{(t_j^*)}$, $t_0^* = 0$, $t_j^* = \lfloor \frac{(2j-1)T}{2J} \rfloor$ for $j = 1, \dots, J$, and $t_{J+1}^* = T$. Then we estimate the numbers of row and column factors in the j -th subinterval, denoted by $\hat{k}_{1,j}^{J*}$ and $\hat{k}_{2,j}^{J*}$. We define an indicator function I_j^{J*} to identify if the numbers of factors in S_j^{J*} and S_j^J are different. $I_j^{J*} = 1$ if $\hat{k}_{1,j}^{J*} \neq \hat{k}_{1,j}^J$ or $\hat{k}_{2,j}^{J*} \neq \hat{k}_{2,j}^J$; $I_j^{J*} = 0$ if $\hat{k}_{1,j}^{J*} = \hat{k}_{1,j}^J$ and $\hat{k}_{2,j}^{J*} = \hat{k}_{2,j}^J$. There are two situations that can happen for I_j^{J*} in case (ii):

- (a) when $I_j^{J*} = 0$, there is no regime switching in the left end of S_j or the right end of S_{j-1} .
- (b) when $I_j^{J*} = 1$, regime switching happens in the left end of S_j or the right end of S_{j-1} .

For case (b) and case (iii) discussed above, we can estimate the complement loading space with eigenvectors of $\widehat{\mathbf{M}}_{s,j-1}^{J*}$ corresponding to the smallest $(p_s - \hat{k}_{s,j-1}^{J*})$ eigenvalues and eigenvectors of $\widehat{\mathbf{M}}_{s,j+1}^{J*}$ corresponding to the smallest $(p_s - \hat{k}_{s,j+1}^{J*})$ eigenvalues for $s = 1, 2$, then we follow the procedure described in Section 2.3 to estimate the threshold value and loading spaces in these regimes.

The proposed identification procedure is build on the ratio estimator by Lam and Yao (2012). Since the eigenvalues go to zero theoretically, Lam and Yao (2012) was not able to prove the consistency of the estimators. Therefore, we cannot derive an explicit asymptotic expression for the identification of different regimes. However, the simulation results in Section 3.3 show that the proposed procedure performs very well.

3 Simulation

In this section we present the performance of the proposed methods for synthetic data sets. For all settings, the reported results are based on 200 replications, $h_0 = 1$, and the threshold process z_t follows an independent Gaussian process $\mathcal{N}(0, 1)$.

3.1 Estimation of numbers of factors

We study the performance of the proposed estimators for the factor numbers in Section 2.4. There are two regimes. The dimension of common factors in both regimes is 2×2 with $k_{1,1} = k_{1,2} = k_{2,1} = k_{2,2} = 2$. The factor at time t is generated by 4 independent autoregressive (AR) process of

order 1 with AR coefficient 0.9 and innovation variance 4. The threshold value is 0. η_1 and η_2 are the 10-th and 90-th percentiles of $\{z_t\}$. The error process $\mathbf{E}_t$ is a white noise process with mean $\mathbf{0}$ and a Kronecker product covariance structure, that is, $\text{Cov}(\text{vec}(\mathbf{E}_t)) = \mathbf{\Gamma}_2 \otimes \mathbf{\Gamma}_1$, where $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are of sizes $p_1 \times p_1$ and $p_2 \times p_2$ respectively, and both have diagonal entries of 1 and off-diagonal entries equal to σ_γ . Here, $\sigma_\gamma = 0.1$. Table 1 reports the relative frequency that the numbers of factors are correctly estimated. We can tell that as the sample size increases, the frequency that $\hat{k}_{1,1} = \hat{k}_{1,2} = \hat{k}_{2,1} = \hat{k}_{2,2} = 2$ increases as well. When the factors are strong, the estimation is at its best. When there are weak factors, the estimation is worse but still accurate.

Table 1: The relative frequency that $\hat{k}_{1,1} = \hat{k}_{1,2} = \hat{k}_{2,1} = \hat{k}_{2,2} = 2$ with different combinations of (T, p_1, p_2) and factor strength in Section 3.1

T	p_1	p_2	$\delta_1 = 0, \delta_2 = 0$	$\delta_1 = 0.2, \delta_2 = 0$	$\delta_1 = 0.2, \delta_2 = 0.2$
200	20	20	0.910	0.835	0.710
	40	40	0.985	0.950	0.760
400	20	20	0.995	0.985	0.920
	40	40	1.000	0.995	0.980
800	20	20	1.000	1.000	1.000
	40	40	1.000	1.000	1.000

3.2 Estimation of threshold value and loading spaces

In this section, we examine the performance of the propose estimators for threshold value and loading spaces in Section 2.4 when there are two regimes. The dimension of the latent factors in both regime is fixed at 3×3 with $k_{1,1} = k_{1,2} = k_{2,1} = k_{2,2} = 3$. We simulate the latent factor process from 9 independent AR models of order one. The AR coefficients are -0.8, 0.8, 0.9, -0.7, -0.9, 0.8, 0.7, 0.8, 0.7, and the innovation variance is 4. The threshold value, η_1 and η_2 are the same with these in Section 3.2. The distribution of error $\mathbf{E}_t$ has the same structure with that in Section 3.2 and $\sigma_\gamma = 0.2$.

Since Section 3.1 shows that the numbers of factors can be estimated precisely, in the following we set different values for $\hat{k}$'s to better demonstrate the performance of the proposed method when the numbers of factors are correctly estimated or overestimated. We run simulations with different combinations of (T, p_1, p_2) and factor strength. When there are weak factors, the strength of weak factors are 0.3. Table 2 shows the mean and standard deviation (in the parentheses) of

the absolute error $|\hat{r} - r_0|$. It can be seen that the threshold value r_0 is estimated with higher accuracy when the latent dimensions are correctly estimated shown in the first four rows of Table 2. We can also tell that the performance of our method improves as the sample size increases from 200 to 400 and the factors strength increases in all settings. However, the increase of p_1 and p_2 has different impact on three settings. When row and column factors are strong with $\delta_1 = \delta_2 = 0$, the estimation does not change much as p_1 and p_2 grow and is immune to the curse of dimensionality. When row or/and column factors are weak, the estimation performance gets worse as p_1 and p_2 increases. Table 2 also reports the estimation results for the threshold value when $\hat{k}_{1,1} = \hat{k}_{1,2} = \hat{k}_{2,1} = \hat{k}_{2,2} = 4$ are overestimated, which confirms the theoretical results in Section 2.4 that the estimator for r_0 is consistent when the numbers of factors are overestimated. The last four rows in Table 2 show the estimation errors when the method in Liu and Chen (2020) is extended for matrix time series analysis. Specifically, we define

$$\begin{aligned}\hat{\Omega}_{x,ij,uv}(h, r_1, r_2) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{x}_{t,u} \mathbf{x}_{t+h,v}' I_{t,i}(r_i) I_{t+h,j}(r_j), \\ \hat{\mathbf{M}}_{1,i}(r_1, r_2) &= \sum_{h=1}^{h_0} \sum_{j=1}^2 \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \hat{\Omega}_{x,ij,uv}(h, r_1, r_2) \hat{\Omega}_{x,ij,uv}(h, r_1, r_2)',\end{aligned}$$

for $i, j = 1, 2$. $\hat{\mathbf{M}}_{2,i}$ can be constructed similarly with transposes of $\mathbf{X}_t$'s. The threshold value is estimated by

$$\hat{r}_{LCR} = \arg \min \|\hat{\mathbf{B}}_{s,i}(\eta_1, \eta_2)' \hat{\mathbf{M}}_{s,i}(r, r) \hat{\mathbf{B}}_{s,i}(\eta_1, \eta_2)\|_2,$$

where the columns of $\hat{\mathbf{B}}_{s,i}(\eta_1, \eta_2)$ are eigenvectors of $\hat{\mathbf{M}}_{s,i}(\eta_1, \eta_2)$ corresponding to the $(p_s - k_{s,i})$ smallest eigenvalues. $\hat{\mathbf{Q}}_{s,i}$ is estimated by the eigenvectors of $\hat{\mathbf{M}}_{s,i}(\hat{r}_{LCR}, \hat{r}_{LCR})$ corresponding to the $k_{s,i}$ largest eigenvalues. From Table 2, it can be seen that our method outperforms the one by Liu and Chen (2020).

Table 3 shows the mean and standard deviation of estimation errors for loading spaces under different combinations of (T, p_1, p_2) and factor strength. When the numbers of factors are correctly specified or overestimated, our method can estimate loading spaces precisely. Compared the proposed estimators with the ones in Liu and Chen (2020), ours performs much better when sample size is relatively small or there are weak factors. When sample size is large ($T = 400$) and factors are strong, two methods can both estimate $\mathbf{B}_{s,i}$ in (11) very well, so the estimation results are very close. More comparison results with different values of η_1 and η_2 are presented in Appendix 2.

Table 2: Mean and standard deviation (in the parentheses) of the absolute error $|\hat{r} - r_0|$ with different combinations of (T, p_1, p_2) and factor strength in Section 3.2

Method	$(\hat{k}_{1,1}, \hat{k}_{1,2}, \hat{k}_{2,1}, \hat{k}_{2,2})$	T	p_1	p_2	$\delta_1 = 0, \delta_2 = 0$	$\delta_1 = 0.3, \delta_2 = 0$	$\delta_1 = 0.3, \delta_2 = 0.3$
Our method	(3,3,3,3) correctly specified	200	20	20	0.017 (0.018)	0.027 (0.030)	0.085 (0.076)
			40	40	0.018 (0.016)	0.037 (0.034)	0.180 (0.109)
		400	20	20	0.009 (0.010)	0.016 (0.017)	0.048 (0.041)
			40	40	0.011 (0.012)	0.023 (0.022)	0.093 (0.065)
	(4,4,4,4) overestimated	200	20	20	0.013 (0.013)	0.013 (0.013)	0.017(0.016)
			40	40	0.013 (0.014)	0.013 (0.014)	0.016 (0.015)
		400	20	20	0.006 (0.006)	0.007 (0.007)	0.009 (0.009)
			40	40	0.006 (0.006)	0.007 (0.006)	0.009 (0.012)
LCR	(3,3,3,3) correctly specified	200	20	20	0.213 (0.294)	0.284 (0.286)	0.431 (0.243)
			40	40	0.222 (0.306)	0.299 (0.283)	0.552 (0.246)
		400	20	20	0.025 (0.058)	0.054 (0.071)	0.205 (0.119)
			40	40	0.021 (0.040)	0.084 (0.089)	0.313 (0.150)

Note: LCR denotes the method in [Liu and Chen \(2020\)](#) extended to analyze matrix time series data

Then we will present the performance of our method when the strength level of weak factors is 0.5. T is set as $p_1 p_2$, $2p_1 p_2$, $4p_1 p_2$. Table 4 shows the mean and standard deviation (in the parentheses) of the absolute error $|\hat{r} - r_0|$. When there are strong factors in at least one regime ($\delta_1 = \delta_2 = 0$ or $\delta_1 = 0.5$ and $\delta_2 = 0$), as T grows with p_1 and p_2 , the estimate error decreases. When $\delta_1 = \delta_2 = 0.5$, it is not surprising that the estimation gets worse as T grows with p_1 and p_2 . However, if T grows faster than $(p_1 p_2)$, our estimation improves as T increases which is aligned with results shown in Theorem 1. Table 5 shows the mean and standard deviation (in the parentheses) of the estimation errors for loading spaces. It can be seen that if the strength of weak factors is 0.5, our method can estimate loading spaces precisely when sample size is large.

3.3 Regime switching identification

In this example, we investigate the performance of multiple threshold estimation discussed in Section 2.5. We will show that the proposed procedure performs well even when the factor and noise processes are non-stationary. $T = 500, 1000$, p_1 and p_2 are set to be 20 and 40, and δ_1 and δ_2 are set to be 0 and 0.2. When T is 500, $J = 10$; when $T = 1000$, $J = 15$. If the factor process is stationary, each time series in the factor process follows an independent AR(1) model with AR coefficient 0.9. If the factor process is non-stationary, $\mathbf{F}_{t,i}$ is a 2×2 matrix for $t = 1, \dots, T$ and

Table 3: Mean and standard deviation (in the parentheses) of $\mathcal{D}(\mathcal{M}(\hat{\mathbf{Q}}_{s,i}), \mathcal{M}(\mathbf{Q}_{s,i}))$ with different combinations of (T, p_1, p_2) and factor strength in Section 3.2

	$(\hat{k}_{1,1}, \hat{k}_{1,2}, \hat{k}_{2,1}, \hat{k}_{2,2})$	T	p_1	p_2		$\delta_1 = 0, \delta_2 = 0$		$\delta_1 = 0.3, \delta_2 = 0$		$\delta_1 = 0.3, \delta_2 = 0.3$	
						regime 1	regime 2	regime 1	regime 2	regime 1	regime 2
Our method	(3,3,3,3)	200	20	20	row	0.017 (0.005)	0.017 (0.004)	0.029 (0.008)	0.029 (0.009)	0.056 (0.022)	0.057 (0.032)
					column	0.018 (0.004)	0.017 (0.004)	0.029 (0.008)	0.028 (0.008)	0.058 (0.023)	0.056 (0.024)
		40	40	40	row	0.012 (0.002)	0.011 (0.002)	0.022 (0.005)	0.022 (0.005)	0.066 (0.031)	0.063 (0.031)
					column	0.012 (0.002)	0.012 (0.002)	0.022 (0.005)	0.022 (0.004)	0.061 (0.029)	0.064 (0.034)
		400	20	20	row	0.012 (0.003)	0.011 (0.002)	0.019 (0.005)	0.019 (0.004)	0.035 (0.011)	0.034 (0.011)
					column	0.013 (0.003)	0.012 (0.003)	0.021 (0.005)	0.019 (0.004)	0.038 (0.013)	0.035 (0.011)
	(4,4,4,4)	200	20	20	row	0.015 (0.003)	0.014 (0.003)	0.023 (0.005)	0.023 (0.004)	0.035 (0.008)	0.035 (0.007)
					column	0.015 (0.004)	0.015 (0.004)	0.023 (0.005)	0.023 (0.005)	0.036 (0.008)	0.035 (0.008)
		40	40	40	row	0.010 (0.001)	0.010 (0.001)	0.017 (0.003)	0.017 (0.002)	0.029 (0.004)	0.028 (0.004)
					column	0.010 (0.002)	0.010 (0.002)	0.017 (0.003)	0.017 (0.003)	0.029 (0.004)	0.029 (0.004)
		400	20	20	row	0.010 (0.002)	0.010 (0.002)	0.016 (0.003)	0.015 (0.003)	0.025 (0.005)	0.024 (0.005)
					column	0.010 (0.002)	0.010 (0.002)	0.016 (0.003)	0.016 (0.003)	0.026 (0.005)	0.025 (0.005)
LCR	(3,3,3,3)	200	20	20	row	0.066 (0.144)	0.069 (0.145)	0.066 (0.147)	0.069 (0.147)	0.151 (0.168)	0.153 (0.171)
					column	0.073 (0.152)	0.069 (0.141)	0.101 (0.164)	0.095 (0.148)	0.167 (0.179)	0.157 (0.162)
		40	40	40	row	0.065 (0.148)	0.064 (0.151)	0.079 (0.147)	0.082 (0.152)	0.163 (0.169)	0.175 (0.179)
					column	0.075 (0.166)	0.072 (0.160)	0.088 (0.163)	0.098 (0.169)	0.184 (0.192)	0.210 (0.199)
		400	20	20	row	0.011 (0.004)	0.011 (0.013)	0.018 (0.024)	0.019 (0.024)	0.051 (0.060)	0.045 (0.040)
					column	0.013 (0.005)	0.013 (0.025)	0.019 (0.009)	0.021 (0.031)	0.056 (0.061)	0.057 (0.068)
	(4,4,4,4)	200	20	20	row	0.008 (0.005)	0.008 (0.002)	0.018 (0.015)	0.020 (0.016)	0.054 (0.064)	0.057 (0.065)
					column	0.010 (0.026)	0.008 (0.002)	0.020 (0.029)	0.025 (0.054)	0.071 (0.094)	0.088 (0.114)
		40	40	40	row	0.007 (0.001)	0.007 (0.001)	0.012 (0.002)	0.012 (0.002)	0.020 (0.003)	0.021 (0.003)
					column	0.007 (0.001)	0.007 (0.001)	0.012 (0.002)	0.012 (0.002)	0.021 (0.003)	0.021 (0.003)
		400	20	20	row	0.007 (0.001)	0.007 (0.001)	0.012 (0.002)	0.012 (0.002)	0.020 (0.003)	0.021 (0.003)
					column	0.007 (0.001)	0.007 (0.001)	0.012 (0.002)	0.012 (0.002)	0.021 (0.003)	0.021 (0.003)

Note: LCR denotes the method in [Liu and Chen \(2020\)](#) extended to analyze matrix time series data

Table 4: Mean and standard deviation (in the parentheses) of the absolute error $|\hat{r} - r_0|$ with different combinations of (T, p_1, p_2) and factor strength in Section 3.2

δ_1, δ_2	p_1	p_2	$T = p_1 p_2$	$T = 2p_1 p_2$	$T = 4p_1 p_2$
$\delta_1 = 0, \delta_2 = 0$	10	10	0.029 (0.030)	0.018 (0.019)	0.009 (0.010)
	10	20	0.018 (0.019)	0.011 (0.012)	0.008 (0.010)
	20	20	0.009 (0.010)	0.007 (0.007)	0.005 (0.005)
$\delta_1 = 0.5, \delta_2 = 0$	10	10	0.063 (0.063)	0.044 (0.047)	0.024 (0.030)
	10	20	0.038 (0.040)	0.024 (0.025)	0.017 (0.017)
	20	20	0.028 (0.029)	0.022 (0.021)	0.015 (0.014)
$\delta_1 = 0.5, \delta_2 = 0.5$	10	10	0.214 (0.185)	0.166 (0.137)	0.120 (0.100)
	10	20	0.275 (0.172)	0.203 (0.118)	0.143 (0.086)
	20	20	0.378 (0.144)	0.281 (0.114)	0.159 (0.084)

Table 5: Mean and standard deviation (in the parentheses) of $\mathcal{D}(\mathcal{M}(\hat{\mathbf{Q}}_{s,i}), \mathcal{M}(\mathbf{Q}_{s,i}))$ with different combinations of (T, p_1, p_2) and factor strength in Section 3.2

δ_1, δ_2	p_1	p_2		$T = p_1 p_2$		$T = 2p_1 p_2$		$T = 4p_1 p_2$	
				regime 1	regime 2	regime 1	regime 2	regime 1	regime 2
$\delta_1 = 0, \delta_2 = 0$	10	10	row	0.040 (0.016)	0.038 (0.012)	0.027 (0.008)	0.027 (0.009)	0.019 (0.007)	0.018 (0.006)
			col	0.039 (0.015)	0.037 (0.013)	0.027 (0.012)	0.027 (0.009)	0.019 (0.007)	0.019 (0.007)
	10	20	row	0.019 (0.008)	0.018 (0.006)	0.013 (0.005)	0.013 (0.004)	0.009 (0.003)	0.009 (0.003)
			column	0.025 (0.004)	0.025 (0.004)	0.018 (0.004)	0.018 (0.004)	0.013 (0.003)	0.013 (0.003)
	20	20	row	0.012 (0.002)	0.011 (0.002)	0.008 (0.002)	0.008 (0.002)	0.006 (0.001)	0.006 (0.001)
			column	0.012 (0.002)	0.012 (0.002)	0.008 (0.002)	0.008 (0.002)	0.006 (0.001)	0.006 (0.001)
$\delta_1 = 0.5, \delta_2 = 0$	10	10	row	0.087 (0.057)	0.082 (0.047)	0.055 (0.024)	0.054 (0.025)	0.038 (0.019)	0.034 (0.013)
			column	0.086 (0.059)	0.081 (0.053)	0.056 (0.032)	0.055 (0.022)	0.037 (0.018)	0.037 (0.016)
	10	20	row	0.042 (0.022)	0.041 (0.041)	0.028 (0.014)	0.026 (0.011)	0.018 (0.007)	0.017 (0.005)
			column	0.048 (0.008)	0.049 (0.009)	0.033 (0.008)	0.034 (0.009)	0.023 (0.006)	0.023 (0.006)
	20	20	row	0.028 (0.007)	0.027 (0.008)	0.019 (0.005)	0.019 (0.005)	0.013 (0.003)	0.013 (0.003)
			column	0.030 (0.009)	0.028 (0.007)	0.019 (0.005)	0.019 (0.005)	0.013 (0.003)	0.013 (0.003)
$\delta_1 = 0.5, \delta_2 = 0.5$	10	10	row	0.251 (0.140)	0.245 (0.130)	0.165 (0.101)	0.162 (0.105)	0.106 (0.076)	0.097 (0.078)
			col	0.257 (0.148)	0.261 (0.148)	0.162 (0.114)	0.176 (0.119)	0.101 (0.075)	0.106 (0.078)
	10	20	row	0.242 (0.132)	0.248 (0.142)	0.154 (0.110)	0.144 (0.111)	0.080 (0.057)	0.073 (0.057)
			column	0.242 (0.075)	0.240 (0.089)	0.126 (0.075)	0.137 (0.089)	0.072 (0.036)	0.075 (0.039)
	20	20	row	0.226 (0.138)	0.214 (0.131)	0.107 (0.075)	0.102 (0.068)	0.050 (0.027)	0.048 (0.023)
			column	0.260 (0.136)	0.232 (0.140)	0.119 (0.076)	0.113 (0.087)	0.051 (0.026)	0.050 (0.026)

$i = 1, \dots, m$.

$$f_{t,1} = -\frac{0.1t}{T} + 0.9f_{t-1,1} + \epsilon_{t,1}, \quad f_{t,2} = -1 - \frac{3t}{T}, \quad f_{t,3} = 1 + \frac{3t}{T}, \quad f_{t,4} = \frac{0.1t}{T} - 0.9f_{t-1,4} + \epsilon_{t,2},$$

where $\epsilon_{t,u}$ are independent and $\mathcal{N}(0, 4)$ for $t = 1, \dots, T$ and $u = 1, \dots, 2$.

The noise process is a time-independent Gaussian process with mean $\mathbf{0}$ and a Kronecker product covariance structure, that is, $\text{Cov}(\text{vec}(\mathbf{E}_t)) = \mathbf{\Gamma}_{t,2} \otimes \mathbf{\Gamma}_{t,1}$, where $\mathbf{\Gamma}_{t,1}$ and $\mathbf{\Gamma}_{t,2}$ are of sizes $p_1 \times p_1$ and $p_2 \times p_2$ respectively. If $\{\mathbf{E}_t\}$ is stationary, $\mathbf{\Gamma}_{t,1}$ and $\mathbf{\Gamma}_{t,2}$ both have diagonal entries of 1 and off-diagonal entries of 0.1 for $t = 1, \dots, T$. If the noise process is non-stationary, $\mathbf{\Gamma}_{t,1}$ and $\mathbf{\Gamma}_{t,2}$ both have diagonal entries of $\sqrt{1 + \sin(2\pi t/T)}$ and off-diagonal entries of 0.1 for $t = 1, \dots, T$.

Three settings are considered:

1. $m = 2$ and there is only one threshold with $r_0 = 0$. Both the factor and noise processes are stationary. In regime 1, $k_{1,1} = 1$ and $k_{1,2} = 2$; In regime 2, $k_{2,1} = k_{2,2} = 2$.
2. $m = 3$ and there are two thresholds with $r_0 = -0.5$ and $r_1 = 0.5$. The factor process is stationary and the noise process is non-stationary.

3. $m = 1$. Both the factor and noise processes are non-stationary.

Table 6 shows the relative frequency to identify the regime switching(s) correctly with different combinations of (T, p_1, p_2) and factor strength, and Table 7 reports the absolute error of threshold value estimation conditional on the all regime switchings are correctly detected. When the factors are strong in at least one direction, our method can identify the regime switchings with a very high relative frequency and estimate their locations very well. When both row and column factors are weak with $\delta_1 = \delta_2 = 0.2$, the performance of our methods gets slightly worse, but the proposed algorithm still can identify the regime switchings in most cases.

Table 6: Relative frequency to correctly detect the regime switchings with different combinations of (T, p_1, p_2) and factor strength under three settings in Section 3.3

T	p_1	p_2		$\delta_1 = 0$	$\delta_1 = 0.2$	$\delta_1 = 0.2$
				$\delta_2 = 0$	$\delta_2 = 0$	$\delta_2 = 0.2$
500	20	20	Setting 1	0.935	0.860	0.745
			Setting 2	0.985	0.970	0.900
			Setting 3	0.955	0.915	0.860
	40	40	Setting 1	0.990	0.960	0.810
			Setting 2	0.995	1.000	0.970
			Setting 3	0.965	0.955	0.940
1000	20	20	Setting 1	0.965	0.955	0.875
			Setting 2	1.000	0.995	0.940
			Setting 3	0.985	0.960	0.935
	40	40	Setting 1	0.995	0.990	0.990
			Setting 2	1.000	1.000	0.965
			Setting 3	0.995	0.990	0.975

4 Application to Real Data

We apply the proposed method to the Fama-French 10 by 10 return series. A universe of stocks is grouped into 100 portfolios, according to ten levels of market capital and ten levels of equity ratio. We analyze their monthly returns from November 1980 to October 2020 with $T = 480$

Table 7: Mean and standard deviation (in the parentheses) of absolute error of threshold value estimation conditional on that all regime switching are correctly detected in Section 3.3

T	p_1	p_2			$\delta_1 = 0$	$\delta_1 = 0.2$	$\delta_1 = 0.2$
					$\delta_2 = 0$	$\delta_2 = 0$	$\delta_2 = 0.2$
500	20	20	Setting 1		0.007(0.008)	0.009(0.012)	0.008(0.009)
			Setting 2	Threshold 1	0.006(0.007)	0.007(0.008)	0.007(0.008)
				Threshold 2	0.006(0.006)	0.007(0.007)	0.007(0.008)
	40	40	Setting 1		0.006(0.005)	0.008(0.009)	0.012(0.011)
			Setting 2	Threshold 1	0.006(0.005)	0.008(0.008)	0.009(0.008)
				Threshold 2	0.007(0.006)	0.007(0.007)	0.009(0.009)
1000	20	20	Setting 1		0.004(0.003)	0.004(0.004)	0.005(0.006)
			Setting 2	Threshold 1	0.004(0.004)	0.004(0.004)	0.004(0.004)
				Threshold 2	0.003(0.003)	0.004(0.003)	0.004(0.004)
	40	40	Setting 1		0.004(0.003)	0.005(0.005)	0.006(0.006)
			Setting 2	Threshold 1	0.003(0.003)	0.003(0.004)	0.005(0.005)
				Threshold 2	0.003(0.003)	0.004(0.004)	0.005(0.005)

and $p_1 = p_2 = 10$. More detailed information about this data set is available here, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Here we let $h_0 = 1$ and $J = 12$, and use the monthly log return of S&P 500 index as the threshold variable. Figure 1 plots the estimate numbers of factors in subintervals $\{S_j\}$ and $\{S_j^*\}$. Although the number of row factors remains the same in the support of z_t , the number of column factors changes in S_2 and S_5 . It indicates that there are two regime switchings caused changes in the column loading spaces, happening in the interior of S_2 and S_5 , and the number of column factors does not change across regimes. With method in Section 2.5, we have $\hat{r}_0 = -0.026$ and $\hat{r}_1 = -0.002$. Table 8 reports the distance of estimated loading spaces, and we can tell that the column loading spaces are well apart, and the row loading spaces are closer to each other, which is aligned with the observation that the estimated number of row factors remain the same in all subintervals shown in Figure 1.

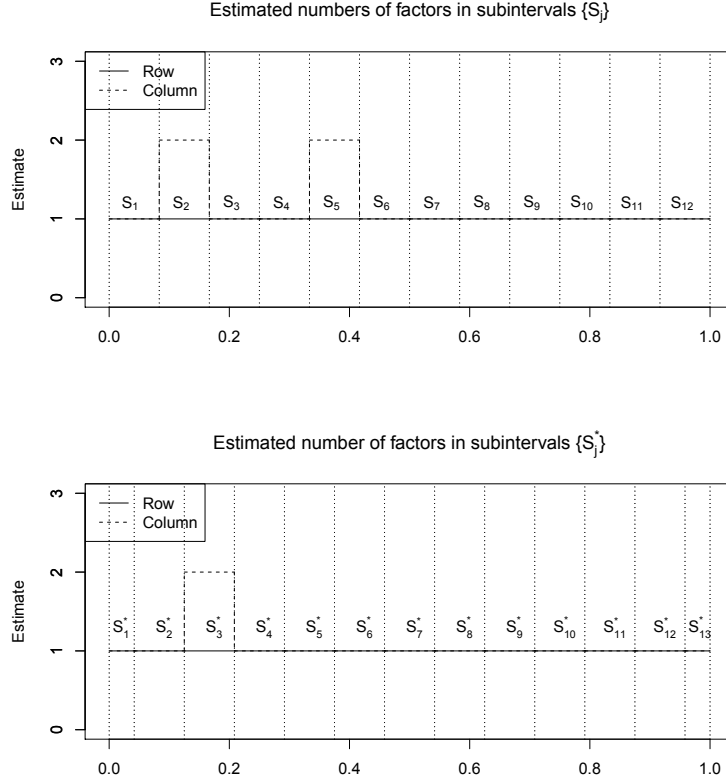


Figure 1: Estimated numbers of factors in subintervals $\{S_j\}$ and $\{S_j^*\}$ for data in Section 4.

Table 8: Distance of estimated loading spaces for Section 4

Row loading space	Regime 1	Regime 2	Regime 3
Regime 1	—	0.170	0.110
Regime 2	0.170	—	0.167
Regime 3	0.110	0.167	—
Column loading space	Regime 1	Regime 2	Regime 3
Regime 1	—	0.519	0.238
Regime 2	0.519	—	0.325
Regime 3	0.238	0.325	—

5 Conclusion

In this article, we extend the threshold factor models for multivariate time series to matrix-variate data and propose a threshold factor model for high-dimensional matrix-variate time series, where loading spaces change across regimes controlled by a threshold variable. The methods to estimate threshold value, loading spaces in different regimes, and the numbers of factors are developed. The procedure to identify regime switching is discussed. Compared with the existing results on threshold factor models which often assume that the number of factors does not change over the sampling period, the proposed method can be applied to the case when the numbers of factors vary across regime. Another distinguished feature of our method is that the factor or noise process is not necessarily to be stationary within regime which greatly enhances the application scope of threshold factor models, while most current research results are obtained based on the assumption that the observed process is stationary within regime. The simulated and real data examples all confirm that the proposed method can detect regime switchings effectively and estimate threshold values and loading spaces well.

Address of corresponding author:

Xialu Liu

Department of Management Information Systems

San Diego State University

5500 Campanile Dr, San Diego, CA 92182, USA

Email: xialu.liu@sdsu.edu

Appendix 1: Lemmas and Proofs

In this section, only the theoretical results for $s = 1$ are demonstrated, since those for $s = 2$ are similar. Moreover, we mainly focus on the proofs when $r > r_0$ and $\epsilon > 0$ because those for $r \leq r_0$ or $\epsilon < 0$ can be obtained in a similar fashion.

In the following we assume c_1 and c_2 are constants satisfying $I^R(c_1) = I^R(c_2)$. Define

$$\begin{aligned}
I_t(c_1, c_2) &= I(c_1 < z_t < c_2), \\
\mathbf{S}_t &= \sum_{i=1}^2 \mathbf{R}_i \mathbf{F}_{t,i} \mathbf{C}'_i I_{t,i}, \\
\boldsymbol{\Omega}_{s,ij,uv}(h, r) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E} [\mathbf{s}'_{t,u} \mathbf{s}_{t+h,v} I_{t,i}(r) I_{t+h,j}], \\
\hat{\boldsymbol{\Omega}}_{s,ij,uv}(h, r) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{s}'_{t,u} \mathbf{s}_{t+h,v} I_{t,i}(r) I_{t+h,j}, \\
\hat{\boldsymbol{\Omega}}_{se,ij,uv}(h, r) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{s}'_{t,u} \mathbf{e}_{t+h,v} I_{t,i}(r) I_{t+h,j}, \\
\hat{\boldsymbol{\Omega}}_{es,ij,uv}(h, r) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{e}'_{t,u} \mathbf{s}_{t+h,v} I_{t,i}(r) I_{t+h,j}, \\
\hat{\boldsymbol{\Omega}}_{e,ij,uv}(h, r) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{e}'_{t,u} \mathbf{e}_{t+h,v} I_{t,i}(r) I_{t+h,j}, \\
\boldsymbol{\Omega}_{fc,ij,uv}(h, c_1, c_2) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E} [\mathbf{F}_{t,i} \mathbf{c}_{i,u} \mathbf{c}'_{j,v} \mathbf{F}'_{t+h,j} I_t(c_1, c_2) I_{t+h,j}], \\
\hat{\boldsymbol{\Omega}}_{fc,ij,uv}(h, c_1, c_2) &= \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{F}_{t,i} \mathbf{c}_{i,u} \mathbf{c}'_{j,v} \mathbf{F}'_{t+h,j} I_t(c_1, c_2) I_{t+h,j}.
\end{aligned}$$

Lemma 1. Let $f_{t,i,qu}$ denote the (qu) -th entry in $\mathbf{F}_{t,i}$. Under Conditions A1-A2 and B1, for any $q, u = 1, 2, \dots, k_1$, and $\ell, v = 1, \dots, k_2$, it holds that

$$\begin{aligned}
\mathbb{E} \left\{ \frac{1}{T} \sum_{t=1}^{T-h} [f_{t,i,q\ell} f_{t+h,j,uv} I_t(c_1, c_2) I_{t+h,j} - \mathbb{E}(f_{t,i,q\ell} f_{t+h,j,uv} I_t(c_1, c_2) I_{t+h,j})] \right\}^2 \\
\leq \frac{(3h + 8\alpha) \rho_{c_1, c_2} \sigma_f^4}{T},
\end{aligned}$$

and

$$\left| \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E}[f_{t,i,q\ell} f_{t+h,j,uv} I_t(c_1, c_2) I_{t+h,j}] \right| = \rho_{c_1, c_2} \sigma_f^2,$$

where $\alpha = \sum_{u=1}^{\infty} \alpha(u)^{1-2/\gamma}$, and $c_1 < c_2$ can be any real numbers in (η_1, η_2) , $-\infty$, or $+\infty$. $\rho_{c_1, c_2} = 1$ if at least one of c_1 and c_2 is $-\infty$ or $+\infty$, and $\rho_{c_1, c_2} = \tau_1(c_2 - c_1)$ if c_1 and c_2 are both real numbers, where τ_1 is given in Condition B1.

Proof: Similar to the proof of Lemma 1 in [Liu and Chen \(2020\)](#), we can obtain the conclusions. ■

Lemma 2. For $i, j = 1, 2$, $u, v = 1, \dots, p_2$, it holds that

$$\|\mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)\|_2^2 \leq \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{E}[\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j}] \right\|_F^2 \|\mathbf{c}_{i,u}\|_2^2 \cdot \|\mathbf{c}_{j,v}\|_2^2,$$

and

$$\begin{aligned} & \|\hat{\mathbf{\Omega}}_{fc,ij,uv}(h, c_1, c_2) - \mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)\|_2^2 \\ & \leq \left\| \frac{1}{T} \sum_{t=1}^{T-h} [\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j} - \mathbf{E}(\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j})] \right\|_F^2 \|\mathbf{c}_{i,u}\|_2^2 \cdot \|\mathbf{c}_{j,v}\|_2^2, \end{aligned}$$

where $c_1 < c_2$ can be real numbers, $-\infty$, or $+\infty$.

Proof: By the definition and properties of Kronecker product, we have

$$\begin{aligned} & \|\mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)\|_2^2 \\ & \leq \|\mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)\|_F^2 = \|\text{vec}(\mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2))\|_2^2 \\ & = \left\| \frac{1}{T} \sum_{t=1}^{T-h} \text{vec}(\mathbf{E}(\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j})) \right\|_2^2 \\ & = \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{E}[\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j}] \text{vec}(\mathbf{c}_{i,u} \mathbf{c}_{j,v}') \right\|_2^2 \\ & \leq \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{E}[\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j}] \right\|_F^2 \|\text{vec}(\mathbf{c}_{i,u} \mathbf{c}_{j,v}')\|_2^2 \\ & \leq \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{E}[\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j}] \right\|_F^2 \|\mathbf{c}_{i,u}\|_2^2 \cdot \|\mathbf{c}_{j,v}\|_2^2 \\ & \leq \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{E}[\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j}] \right\|_F^2 \|\mathbf{c}_{i,u}\|_2^2 \cdot \|\mathbf{c}_{j,v}\|_2^2. \end{aligned}$$

The other inequality can be proven similarly. ■

Lemma 3. Under Conditions A1-A2, A4 and B1, for $i, j = 1, 2$, it holds that

$$\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbf{E} \|\hat{\mathbf{\Omega}}_{fc,ij,uv}(h, c_1, c_2) - \mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)\|_2^2 \leq (3h + 8\alpha) \rho_{c_1, c_2} k_1 k_2^3 a_0^4 \sigma_f^2 p_2^{2-2\delta_2} T^{-1},$$

and

$$\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)\|_2^2 \leq \rho_{c_1, c_2}^2 k_1 k_2^3 a_0^4 \sigma_f^4 p_2^{2-2\delta_2},$$

where $k_1 = \max\{k_{1,1}, k_{1,2}\}$, $k_2 = \max\{k_{2,1}, k_{2,2}\}$, a_0 satisfies $\|\mathbf{C}_i\|_2 \leq a_0 p_2^{1/2-\delta_2/2}$ for $i = 1, 2$, and $c_1 < c_2$ can be real numbers in (η_1, η_2) , $-\infty$ or $+\infty$.

Proof: Condition A4 implies that there exists a positive constant a_0 such that $\|\mathbf{C}_i\|_2 \leq a_0 p_2^{1/2-\delta_2/2}$ for $i = 1, 2$. By Lemma 1 and Lemma 2, it follows

$$\begin{aligned} & \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbb{E} \|\hat{\mathbf{\Omega}}_{fc,ij,uv}(h, c_1, c_2) - \mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)\|_2^2 \\ &= \left(\sum_{u=1}^{p_2} \|\mathbf{c}_{i,u}\|_2^2 \right) \left(\sum_{v=1}^{p_2} \|\mathbf{c}_{j,v}\|_2^2 \right) \\ & \quad \cdot \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T-h} [\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j} - \mathbb{E}(\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2))] \right\|_F^2 \\ &\leq \|\mathbf{C}_i\|_F^2 \|\mathbf{C}_j\|_F^2 \cdot \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T-h} [\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j} - \mathbb{E}(\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j})] \right\|_F^2 \\ &\leq k_2^2 \|\mathbf{C}_i\|_2^2 \|\mathbf{C}_j\|_2^2 \cdot \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T-h} [\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j} - \mathbb{E}(\mathbf{F}_{t+h,j} \otimes \mathbf{F}_{t,i} I_t(c_1, c_2) I_{t+h,j})] \right\|_F^2 \\ &\leq (3h + 8\alpha) \rho_{c_1, c_2} k_1 k_2^3 a_0^4 \sigma_f^4 p_2^{2-2\delta_2} T^{-1}. \end{aligned}$$

We can also obtain the bound of $\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)\|_2^2$ with Lemma 1 and Lemma 2 in a similar way. ■

Lemma 4. Under Conditions A1-A4 and B1, for $i = 1, 2$ and any $\epsilon \in (\eta_1 - r_0, \eta_2 - r_0)$, it holds that

$$\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbb{E} \|\hat{\mathbf{\Omega}}_{x,i,uv}(h, r_0 + \epsilon) - \mathbf{\Omega}_{x,i,uv}(h, r_0 + \epsilon)\|_2^2 = O_p(p_1^2 p_2^2 T^{-1}).$$

Proof: By Condition A4, Lemmas 1 and 3, when $\epsilon > 0$ and $i = j = 1$, we have

$$\begin{aligned}
& \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbb{E} \|\hat{\mathbf{\Omega}}_{s,11,uv}(h, r_0 + \epsilon) - \mathbf{\Omega}_{s,11,uv}(h, r_0 + \epsilon)\|_2^2 \\
& \leq \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} 4 \|\mathbf{R}_1\|_2^4 \cdot \mathbb{E} \|(\hat{\mathbf{\Omega}}_{fc,11,uv}(h, -\infty, r_0) - \mathbf{\Omega}_{fc,11,uv}(h, -\infty, r_0))\|_2^2 \\
& \quad + \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} 4 \|\mathbf{R}_1\|_2^2 \cdot \mathbb{E} \|\hat{\mathbf{\Omega}}_{fc,12,uv}(h, -\infty, r_0) - \mathbf{\Omega}_{fc,12,uv}(h, -\infty, r_0)\|_2^2 \cdot \|\mathbf{R}_2\|_2^2 \\
& \quad + \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} 4 \|\mathbf{R}_2\|_2^2 \cdot \mathbb{E} \|\hat{\mathbf{\Omega}}_{fc,21,uv}(h, r_0, r_0 + \epsilon) - \mathbf{\Omega}_{fc,21,uv}(h, r_0, r_0 + \epsilon)\|_2^2 \cdot \|\mathbf{R}_1\|_2^2 \\
& \quad + \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} 4 \|\mathbf{R}_2\|_2^4 \cdot \mathbb{E} \|\hat{\mathbf{\Omega}}_{fc,22,uv}(h, r_0, r_0 + \epsilon) - \mathbf{\Omega}_{fc,22,uv}(h, r_0, r_0 + \epsilon)\|_2^2 \\
& \leq C p_1^{2-2\delta_1} p_2^{2-2\delta_2} T^{-1}.
\end{aligned}$$

For the interaction of the common component and noise and the noise term, we use the same trick,

$$\begin{aligned}
& \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbb{E} \|\hat{\mathbf{\Omega}}_{se,11,uv}(h, r_0 + \epsilon)\|_2^2 \\
& \leq 2 \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{R}_1\|_2^2 \cdot \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{F}_{t,1} \mathbf{c}_{1,u} \mathbf{e}'_{t+h,v} I(z_t < r_0) I(z_{t+h} < r_0) \right\|_2^2 \\
& \quad + 2 \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{R}_2\|_2^2 \cdot \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{F}_{t,2} \mathbf{c}_{2,u} \mathbf{e}'_{t+h,v} I(r_0 < z_t < r_0 + \epsilon) I(z_{t+h} < r_0) \right\|_2^2 \\
& \leq 2 \|\mathbf{R}_1\|_2^2 \left(\sum_{v=1}^{p_2} \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{e}_{t+h,v} \otimes \mathbf{F}_{t,1} I(z_t < r_0) \right\|_2^2 \right) \left(\sum_{u=1}^{p_2} \|\mathbf{c}_{1,u}\|_2^2 \right) \\
& \quad + 2 \|\mathbf{R}_2\|_2^2 \left(\sum_{v=1}^{p_2} \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T-h} \mathbf{e}_{t+h,v} \otimes \mathbf{F}_{t,2} I(r_0 < z_t < r_0 + \epsilon) \right\|_2^2 \right) \left(\sum_{u=1}^{p_2} \|\mathbf{c}_{2,u}\|_2^2 \right) \\
& \leq C p_1^{1-\delta_1} p_2^{1-\delta_2} \left[\frac{1}{T^2} \sum_{v=1}^{p_2} \sum_{u=1}^{p_1} \sum_{q=1}^{k_{1,1}} \sum_{v=1}^{k_{2,1}} \sum_{t=1}^{T-h} \mathbb{E} (e_{t+h,uv}^2 f_{t,1,qv}^2 I(z_t < r_0)) \right] \\
& \quad + C p_1^{1-\delta_1} p_2^{1-\delta_2} \left[\frac{1}{T^2} \sum_{v=1}^{p_2} \sum_{u=1}^{p_1} \sum_{q=1}^{k_{1,2}} \sum_{v=1}^{k_{2,2}} \sum_{t=1}^{T-h} \mathbb{E} (e_{t+h,uv}^2 f_{t,2,qv}^2 I(r_0 < z_t < r_0 + \epsilon)) \right] \\
& \leq C p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1}, \tag{18}
\end{aligned}$$

$$\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbb{E} \|\hat{\mathbf{\Omega}}_{ez,11,uv}(h, r_0 + \epsilon)\|_2^2 \leq C p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1},$$

$$\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbb{E} \|\hat{\mathbf{\Omega}}_{e,11,uv}(h, r_0 + \epsilon)\|_2^2 \leq C p_1^2 p_2^2 T^{-1}.$$

It follows

$$\begin{aligned}
& \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbb{E} \|\hat{\boldsymbol{\Omega}}_{x,1,uv}(h, r_0 + \epsilon) - \boldsymbol{\Omega}_{x,1,uv}(h, r_0 + \epsilon)\|_2^2 \\
& \leq \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \left(4\mathbb{E} \|\hat{\boldsymbol{\Omega}}_{s,11,uv}(h, r_0 + \epsilon) - \boldsymbol{\Omega}_{s,11,uv}(h, r_0 + \epsilon)\|_2^2 + 4\mathbb{E} \|\hat{\boldsymbol{\Omega}}_{se,11,uv}(h, r_0 + \epsilon)\|_2^2 \right. \\
& \quad \left. + 4\mathbb{E} \|\hat{\boldsymbol{\Omega}}_{es,11,uv}(h, r_0 + \epsilon)\|_2^2 + 4\mathbb{E} \|\hat{\boldsymbol{\Omega}}_{e,11,uv}(h, r_0 + \epsilon)\|_2^2 \right) \leq Cp_1^2 p_2^2 T^{-1}.
\end{aligned}$$

■

Lemma 5. *Under Conditions A1-A5, and B5 for $\epsilon \in (\eta_1 - r_0, \eta_2 - r_0)$, we have*

$$\begin{aligned}
\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{B}'_{1,1}(\eta_1) \boldsymbol{\Omega}_{x,1,uv}(h, r_0 + \epsilon)\|_2^2 &= \begin{cases} 0 & \epsilon \leq 0, \\ O(\epsilon^2 p_1^{1-2\delta_1} p_2^{2-2\delta_2}) & \epsilon > 0, \end{cases} \\
\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{B}'_{1,2}(\eta_1) \boldsymbol{\Omega}_{x,2,uv}(h, r_0 + \epsilon)\|_2^2 &= \begin{cases} O(\epsilon^2 p_1^{1-2\delta_1} p_2^{2-2\delta_2}) & \epsilon \leq 0, \\ 0 & \epsilon > 0. \end{cases}
\end{aligned}$$

Proof: From the definition we can easily reach the conclusion following the proof of Lemma 4 in [Liu and Chen \(2020\)](#). ■

Lemma 6. *Under Conditions A1-A5 and B4-B5, for any $\epsilon \in (\eta_1 - r_0, 0)$,*

$$\lambda_{k_{1,1}} \left(\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \boldsymbol{\Omega}_{fc,1j,uv}(h_1^*, r_0 + \epsilon, r_0) \boldsymbol{\Omega}_{fc,1j,uv}(h_1^*, r_0 + \epsilon, r_0)' \right) \geq C\epsilon^2 p_2^{2-2\delta_2},$$

and for any $\epsilon \in (0, \eta_2 - r_0)$,

$$\lambda_{k_{1,2}} \left(\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \boldsymbol{\Omega}_{fc,2j,uv}(h_2^*, r_0, r_0 + \epsilon) \boldsymbol{\Omega}_{fc,2j,uv}(h_2^*, r_0, r_0 + \epsilon)' \right) \geq C\epsilon^2 p_2^{2-2\delta_2},$$

for $j = 1, 2$, where $\lambda_k(\mathbf{H})$ is the k -th largest eigenvalue of $\mathbf{H}$.

Proof: By definition and properties of the Kronecker product, we have

$$\begin{aligned}
& \boldsymbol{\Omega}_{fc,ij,uv}(h, c_1, c_2) \\
&= \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E}[(\mathbf{c}'_{i,u} \otimes \mathbf{I}_{k_{1,i}}) \text{vec}(\mathbf{F}_{t,i}) \text{vec}(\mathbf{F}_{t+h,j})' (\mathbf{c}_{j,v} \otimes \mathbf{I}_{k_{1,j}}) I_t(c_1, c_2) I_{t+h,j}] \\
&= P(c_1 < z_t < c_2) (\mathbf{c}'_{i,u} \otimes \mathbf{I}_{k_{1,i}}) \boldsymbol{\Sigma}_{f,ij}(h, c_1, c_2) (\mathbf{c}_{j,v} \otimes \mathbf{I}_{k_{1,j}}).
\end{aligned}$$

Under Conditions A1-A3, following the proof of Lemma 5 in Wang et al. (2019), we can obtain

$$\begin{aligned} & \frac{1}{P(c_1 < z_t < c_2)^2} \cdot \lambda_{k_{1,i}} \left(\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2) \mathbf{\Omega}_{fc,ij,uv}(h, c_1, c_2)' \right) \\ & \geq \lambda_{k_{1,i}} \left((\mathbf{C}_j \otimes \mathbf{I}_{k_{1,i}}) \mathbf{\Sigma}_{f,ij}(h, c_1, c_2)' (\mathbf{C}_i' \mathbf{C}_i \otimes \mathbf{I}_{k_{1,i}}) \mathbf{\Sigma}_{f,ij}(h, c_1, c_2) (\mathbf{C}_j' \otimes \mathbf{I}_{k_{1,j}}) \right). \end{aligned}$$

Since $\mathbf{C}_i' \mathbf{C}_i$ is a $k_{2,i} \times k_{2,i}$ symmetric positive definite matrix, we can find a $k_{2,i} \times k_{2,i}$ positive definite matrix $\mathbf{U}_i$ such that $\mathbf{C}_i' \mathbf{C}_i = \mathbf{U}_i \mathbf{U}_i'$ and $\|\mathbf{U}_i\|_2 \geq \|\mathbf{U}_i\|_{\min} \geq Cp_2^{1/2-\delta_2/2}$, for $i = 1, 2$. With the property of the Kronecker product, it can be seen that $\sigma_1(\mathbf{U}_i \otimes \mathbf{I}_{k_{1,i}}) \geq \sigma_{(k_{1,i}, k_{2,i})}(\mathbf{U}_i \otimes \mathbf{I}_{k_{1,i}}) \geq Cp_2^{1/2-\delta_2/2}$. By Theorem 9 in Merikoski and Kumar (2004), Lemma 3, and Condition B4, we have $\sigma_{k_{1,2}}(\mathbf{\Sigma}_{f,2j}(h_2^*, r_0, r_0 + \epsilon)(\mathbf{U}_2 \otimes \mathbf{I}_{k_1})) \geq Cp_2^{1/2-\delta_2/2}$.

Similar to proof of Lemma 5 in Wang et al. (2019), we have

$$\begin{aligned} & \frac{1}{p_\epsilon^2} \lambda_{k_{1,2}} \left(\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbf{\Omega}_{fc,2j,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{\Omega}_{fc,2j,uv}(h_2^*, r_0, r_0 + \epsilon)' \right) \\ & \geq \left[\sigma_{k_{1,2}} \left((\mathbf{U}_2' \otimes \mathbf{I}_{k_1}) \mathbf{\Sigma}_{f,2j}(h_2^*, r_0, r_0 + \epsilon) (\mathbf{U}_2 \otimes \mathbf{I}_{k_1}) \right) \right]^2 \geq Cp_2^{2-2\delta_2}, \end{aligned}$$

where $p_\epsilon = P(r_0 < z_t < r_0 + \epsilon, I^R(z_{t+h}) = j)$. The conclusion follows. $\blacksquare$

Lemma 7. Under Conditions A1-A5 and B1-B5, for $\epsilon \in (\eta_1 - r_0, \eta_2 - r_0)$, when the numbers of factors are known, we have $G(r_0) = 0$, and

$$G(r_0 + \epsilon) \geq Cp_1^{2-2\delta_1} p_2^{2-2\delta_2}, \text{ if } \epsilon \neq 0.$$

Proof: When $\epsilon > 0$, by Theorem 9 in Merikoski and Kumar (2004) and Lemmas 3 and 6, we have

$$\begin{aligned} & \|\mathbf{B}_{1,1}' \mathbf{M}_{1,1}(r_0 + \epsilon) \mathbf{B}_{1,1}\|_2 \geq \left\| \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbf{B}_{1,1}' \mathbf{\Omega}_{x,1,uv}(h_2^*, r_0 + \epsilon) \mathbf{\Omega}_{x,1,uv}(h_2^*, r_0 + \epsilon)' \mathbf{B}_{1,1} \right\|_2 \\ & = \left\| \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbf{B}_{1,1}' \mathbf{R}_2 \left[\mathbf{\Omega}_{fc,21,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{R}_1' + \mathbf{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{R}_2' \right] \right. \\ & \quad \cdot \left[\mathbf{R}_1 \mathbf{\Omega}_{fc,21,uv}(h_2^*, r_0, r_0 + \epsilon)' + \mathbf{R}_2 \mathbf{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon)' \right] \mathbf{R}_2' \mathbf{B}_{1,1} \left. \right\|_2 \\ & = \|\mathbf{B}_{1,1}' \mathbf{R}_2\|_2^2 \cdot \left\| \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \left[\mathbf{\Omega}_{fc,21,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{R}_1' + \mathbf{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{R}_2' \right] \right. \\ & \quad \cdot \left[\mathbf{R}_1 \mathbf{\Omega}_{fc,21,uv}(h_2^*, r_0, r_0 + \epsilon)' + \mathbf{R}_2 \mathbf{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon)' \right] \left. \right\|_{\min} \\ & = \|\mathbf{B}_{1,1}' \mathbf{R}_2\|_2^2 \cdot \min_{\|\mathbf{u}\| \in \mathbb{R}^{k_{1,2}}} \left\{ \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbf{u}' \left[\mathbf{\Omega}_{fc,21,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{R}_1' + \mathbf{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{R}_2' \right] \right. \\ & \quad \cdot \left[\mathbf{R}_1 \mathbf{\Omega}_{fc,21,uv}(h_2^*, r_0, r_0 + \epsilon)' + \mathbf{R}_2 \mathbf{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon)' \right] \mathbf{u} \left. \right\}. \end{aligned} \tag{19}$$

Let

$$\mathbf{v}_j = \mathbf{W}_{1,j} \boldsymbol{\Omega}_{fc,2j,uv}(h_2^*, r_0, r_0 + \epsilon)' \mathbf{u},$$

where $\mathbf{W}_{1,j}$ is defined in (4), for $j = 1, 2$. Condition B5 implies that

$$\begin{aligned} & \mathbf{u}' [\boldsymbol{\Omega}_{fc,21,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{R}'_1 + \boldsymbol{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon) \mathbf{R}'_2] \\ & \quad \cdot [\mathbf{R}_1 \boldsymbol{\Omega}_{fc,21,uv}(h_2^*, r_0, r_0 + \epsilon)' + \mathbf{R}_2 \boldsymbol{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon)'] \mathbf{u} \\ &= \|\mathbf{Q}_{1,1} \mathbf{v}_1\|_2^2 + 2\mathbf{v}'_1 \mathbf{Q}'_{1,2} \mathbf{Q}_{1,2} \mathbf{v}_2 + \|\mathbf{Q}_{1,2} \mathbf{v}_2\|_2^2 \geq \|\mathbf{Q}_{1,1} \mathbf{v}_1 - v \mathbf{Q}_{1,2} \mathbf{v}_2\|_2^2 + (1 - v^2) \|\mathbf{Q}_{1,2} \mathbf{v}_2\|_2^2 \\ &\geq (1 - v^2) \|\mathbf{Q}_{1,2} \mathbf{v}_2\|_2^2 = (1 - v^2) \|\mathbf{R}_2 \boldsymbol{\Omega}_{fc,2j,uv}(h_2^*, r_0, r_0 + \epsilon)' \mathbf{u}\|_2^2. \end{aligned} \quad (20)$$

Lemma 5 in [Liu and Chen \(2020\)](#) tells us that $\|\mathbf{B}_{1,1} \mathbf{R}_2\|_2^2 \geq Cp_1^{1-\delta_1}$, therefore, with (19) and (20), we have

$$\begin{aligned} & \|\mathbf{B}'_{1,1} \mathbf{M}_{1,1}(r_0 + \epsilon) \mathbf{B}_{1,1}\|_2 \\ & \geq \|\mathbf{B}'_{1,1} \mathbf{R}_2\|_2^2 \cdot \min_{\|\mathbf{u}\| \in \mathbb{R}^{k_{1,2}}} \left(\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{R}_2 \boldsymbol{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon)' \mathbf{u}\|_2^2 \right) \\ & \geq \|\mathbf{B}'_{1,1} \mathbf{R}_2\|_2^2 \cdot \|\mathbf{R}\|_{\min}^2 \lambda_{k_{1,2}} \left(\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \boldsymbol{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon) \boldsymbol{\Omega}_{fc,22,uv}(h_2^*, r_0, r_0 + \epsilon)' \right) \\ & \geq C\epsilon^2 p_1^{2-2\delta_1} p_2^{2-2\delta_2}. \end{aligned}$$

Thus,

$$G(r_0 + \epsilon) \geq C\epsilon^2 p_1^{2-2\delta_1} p_2^{2-2\delta_2}, \text{ if } \epsilon > 0.$$

It can be shown that $G(r_0 + \epsilon) \geq \epsilon^2 p_1^{2-2\delta_1} p_2^{2-2\delta_2}$ when $\epsilon < 0$ and $G(r_0) = 0$ by definition and Lemmas 5 and 6. ■

Lemma 8. *Under Condition A1-A5 and B1-B5, if $r_0 \in (\eta_1 - r_0, \eta_2 - r_0)$ and $r_0 \neq 0$, and the numbers of factors are known, we have*

$$\mathbb{E}|\hat{G}(r_0 + \epsilon) - G(r_0 + \epsilon)| \leq C_1 p_1^2 p_2^2 T^{-1} + C_2 \epsilon p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2},$$

and

$$\mathbb{E}|\hat{G}(r_0) - G(r_0)| \leq Cp_1^2 p_2^2 T^{-1}.$$

Proof: Since $r_0 \in (\eta_1, \eta_2)$, by the definition, $\mathcal{M}(\mathbf{B}_{s,i}) = \mathcal{M}(\mathbf{B}_{s,i}(\eta_i))$ for $s, i = 1, 2$. Hence, there

exists a $(p_s - k_s) \times (p_s - k_s)$ orthogonal matrix $\mathbf{\Gamma}_{s,i}$ such that $\mathbf{B}_{s,i} = \mathbf{B}_{s,i}(\eta_i)\mathbf{\Gamma}_{s,i}$. Then we have

$$\begin{aligned}
& \left| \sum_{i=1}^2 \left(\|\widehat{\mathbf{B}}'_{1,i}(\eta_i)\widehat{\mathbf{M}}_{1,i}(r)\widehat{\mathbf{B}}_{1,i}(\eta_i)\|_2 - \|\mathbf{B}'_{1,i}\mathbf{M}_{1,i}(r)\mathbf{B}_{1,i}\|_2 \right) \right| \\
&= \left| \sum_{i=1}^2 \left(\|\widehat{\mathbf{B}}'_{1,i}(\eta_i)\widehat{\mathbf{M}}_{1,i}(r)\widehat{\mathbf{B}}_{1,i}(\eta_i)\|_2 - \|\mathbf{B}'_{1,i}(\eta_i)\mathbf{M}_{1,i}(r)\mathbf{B}_{1,i}(\eta_i)\|_2 \right) \right| \\
&\leq \sum_{h=1}^{h_0} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \left(\|\widehat{\mathbf{B}}'_{1,i}(\eta_i)\widehat{\mathbf{\Omega}}_{x,ij,uv}(h,r) - \mathbf{B}'_{s,i}(\eta_i)\mathbf{\Omega}_{x,ij,uv}(h,r)\|_2^2 \right. \\
&\quad \left. + 2\|\mathbf{B}'_{1,i}(\eta_i)\mathbf{\Omega}_{x,ij,uv}(h,r)[\widehat{\mathbf{\Omega}}'_{x,ij,uv}(h,r)\widehat{\mathbf{B}}_{1,i}(\eta_i) - \mathbf{\Omega}'_{x,ij,uv}(h,r)\mathbf{B}_{s,i}(\eta_i)]\|_2 \right) \\
&\leq \sum_{h=1}^{h_0} \sum_{i=1}^2 \sum_{j=1}^2 \left[\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \left(\|\widehat{\mathbf{B}}_{1,i}(\eta_i)\|_2 \|\widehat{\mathbf{\Omega}}_{x,ij,uv}(h,r) - \mathbf{\Omega}_{x,ij,uv}(h,r)\|_2 \right. \right. \\
&\quad \left. \left. + \|\widehat{\mathbf{B}}_{1,i}(\eta_i) - \mathbf{B}_{1,i}(\eta_i)\|_2 \|\mathbf{\Omega}_{x,ij,uv}(h,r)\|_2 \right)^2 + 2 \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{B}_{1,i}(\eta_i)\mathbf{\Omega}_{x,ij,uv}(h,r)\|_2 \right. \\
&\quad \left. \cdot \|\widehat{\mathbf{\Omega}}_{x,ij,uv}(h,r) - \mathbf{\Omega}_{x,ij,uv}(h,r)\|_2 \|\widehat{\mathbf{B}}_{1,i}(\eta_i)\|_2 + 2 \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{B}_{1,i}(\eta_i) \cdot \mathbf{\Omega}_{x,ij,uv}(h,r)\mathbf{\Omega}_{x,ij,uv}(h,r)'\|_2 \right. \\
&\quad \left. \|\widehat{\mathbf{B}}_{1,i}(\eta_i) - \mathbf{B}_{1,i}(\eta_i)\|_2 \right] \\
&= \sum_{i=1}^2 \sum_{j=1}^2 [L_{i,j,1}(r) + L_{i,j,2}(r) + L_{i,j,3}(r)]. \tag{21}
\end{aligned}$$

When $\epsilon > 0$, by Cauchy-Schwarz inequality and Lemmas 4-5,

$$\begin{aligned}
& \mathbb{E}(L_{1,1,1}(r_0 + \epsilon)) \\
&\leq 2 \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \mathbb{E} \|\widehat{\mathbf{\Omega}}_{x,1,uv}(h, r_0 + \epsilon) - \mathbf{\Omega}_{x,1,uv}(h, r_0 + \epsilon)\|_2^2 \\
&\quad + 2\mathbb{E} \|\widehat{\mathbf{B}}_{1,1}(\eta_i) - \mathbf{B}_{1,1}(\eta_i)\|_2^2 \left(\sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \|\mathbf{\Omega}_{x,1,uv}(h, r_0 + \epsilon)\|_2^2 \right) \leq O(p_1^2 p_2^2 T^{-1}), \\
&\mathbb{E}(L_{1,1,2}(r_0 + \epsilon)) = O(\epsilon p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}), \\
&\mathbb{E}(L_{1,1,3}(r_0 + \epsilon)) = O(\epsilon p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}), \\
&\mathbb{E}(L_{1,2,1}(r_0 + \epsilon)) = O(p_1^2 p_2^2 T^{-1}), \quad \mathbb{E}(L_{1,2,2}(r_0 + \epsilon)) = O(\epsilon p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}), \\
&\mathbb{E}(L_{1,2,3}(r_0 + \epsilon)) = O(\epsilon p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}), \\
&\mathbb{E}(L_{2,1,1}(r_0 + \epsilon)) \leq O(p_1^2 p_2^2 T^{-1}), \quad L_{2,1,2}(r_0 + \epsilon) = 0, \quad L_{2,1,3}(r_0 + \epsilon) = 0, \\
&\mathbb{E}(L_{2,2,1}(r_0 + \epsilon)) \leq O(p_1^2 p_2^2 T^{-1}), \quad L_{2,2,2}(r_0 + \epsilon) = 0, \quad L_{2,2,3}(r_0 + \epsilon) = 0.
\end{aligned}$$

It follows from (21),

$$\mathbb{E} \left| \sum_{i=1}^2 \left(\|\widehat{\mathbf{B}}'_{1,i}(\eta_i)\widehat{\mathbf{M}}_{1,i}(r)\widehat{\mathbf{B}}_{1,i}(\eta_i)\|_2 - \|\mathbf{B}'_{1,i}\mathbf{M}_{1,i}(r)\mathbf{B}_{1,i}\|_2 \right) \right| \leq O(p_1^2 p_2^2 T^{-1}) + O(\epsilon p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}).$$

Similarly we can establish the rate of convergence for $\sum_{i=1}^2 (\|\widehat{\mathbf{B}}'_{2,i}(\eta_i)\widehat{\mathbf{M}}_{2,i}(r)\widehat{\mathbf{B}}_{2,i}(\eta_i)\|_2 - \|\mathbf{B}'_{2,i}\mathbf{M}_{2,i}(r)\mathbf{B}_{2,i}\|_2)$.

Then when $\epsilon > 0$, we have

$$\begin{aligned} & |\widehat{G}(r_0 + \epsilon) - G(r_0 + \epsilon)| \\ & \leq \left| \sum_{s=1}^2 \sum_{i=1}^2 \left(\|\widehat{\mathbf{B}}'_{s,i}(\eta_i)\widehat{\mathbf{M}}_{s,i}(r)\widehat{\mathbf{B}}_{s,i}(\eta_i)\|_2 - \|\mathbf{B}'_{s,i}\mathbf{M}_{s,i}(r)\mathbf{B}_{s,i}\|_2 \right) \right| \\ & = O(p_1^2 p_2^2 T^{-1}) + O(\epsilon p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}). \end{aligned}$$

■

Proof of Theorem 1. Following the proof of Theorem 2 in [Liu and Chen \(2020\)](#), we can reach the conclusion. ■

Proof of Theorem 2. Following the proof of Theorem 3 in [\(Liu and Chen, 2020\)](#), the conclusions can be reached. ■

Proof of Corollary 1. Similar to proof of and Corollary 1 in [Liu and Chen \(2016\)](#).

Proof of Theorem 3. Following the proof of Theorem 4 in [Liu and Chen \(2020\)](#), we can reach the conclusion.

Proof of Theorem 4. Similar to proof of Theorem 2.

Proof of Corollary 2: By definition of $\mathbf{M}$ in (17), we have

$$\mathbf{M} = \sum_{h=1}^{h_0} \sum_{u=1}^{p_2} \sum_{v=1}^{p_2} \left\{ \sum_{i=1}^2 \left[\mathbf{Q}_{1,i} \left(\sum_{j=1}^2 \Omega_{zq,ij,uv}(h, r_0) \right) \right] \right\} \left\{ \sum_{i=1}^2 \left[\mathbf{Q}_{1,i} \left(\sum_{j=1}^2 \Omega_{zq,ij,uv}(h, r_0) \right) \right] \right\}'.$$

Similar to the proof of Corollary 2 in [Liu and Zhang \(2022\)](#), we can show that

$$\sigma_{k_{1,1}+k_{1,2}} \left(\sum_{i=1}^2 \left[\mathbf{Q}_{1,i} \left(\sum_{j=1}^2 \Omega_{zq,ij,uv}(h, r_0) \right) \right] \right) \geq C p_1^{1-\delta_1} p_2^{1-\delta_2}.$$

Following the proof of Lemma 6, it can be proved that

$$\lambda_{k_{1,1}+k_{1,2}}(\mathbf{M}) \geq C p_1^{2-2\delta_1} p_2^{2-2\delta_2}.$$

Then we can reach the conclusions following the proof of Corollary 1 in [Lam and Yao \(2012\)](#). ■

Appendix 2

We compare the proposed method with the one by [Liu and Chen \(2020\)](#) using different values of η_1 and η_2 in Table 9 and Table 10. Our method beats the one by [Liu and Chen \(2020\)](#). When $\eta_1 = 0.15$, $\eta_2 = 0.85$, and sample size is large, there are enough observations used to estimate $\mathbf{B}_{s,i}$ in (11) and the difference between estimators of two approaches is much smaller.

Table 9: Mean and standard deviation (in the parentheses) of the absolute error $|\hat{r} - r_0|$ for two methods with different combinations of (T, p_1, p_2) , factor strength, and (η_1, η_2)

(η_1, η_2)	T	p_1	p_2	Method	$\delta_1 = 0, \delta_2 = 0$	$\delta_1 = 0.3, \delta_2 = 0$	$\delta_1 = 0.3, \delta_2 = 0.3$
(0.05, 0.95)	100	10	10	Our method	0.036 (0.038)	0.066 (0.078)	0.138 (0.120)
				LCR	0.839 (0.248)	0.850 (0.249)	0.851 (0.255)
		20	20	Our method	0.028 (0.024)	0.059 (0.063)	0.208 (0.167)
				LCR	0.888 (0.242)	0.904 (0.250)	0.920 (0.252)
		200	10	Our method	0.018 (0.020)	0.032 (0.034)	0.062 (0.062)
				LCR	0.795 (0.216)	0.809 (0.212)	0.821 (0.212)
		20	20	Our method	0.016 (0.016)	0.027 (0.028)	0.107 (0.090)
				LCR	0.872 (0.235)	0.879 (0.225)	0.886 (0.221)
	200	10	10	Our method	0.026 (0.026)	0.032 (0.034)	0.060 (0.065)
				LCR	0.168 (0.251)	0.222 (0.254)	0.295 (0.246)
		20	20	Our method	0.027 (0.023)	0.033 (0.030)	0.074 (0.073)
				LCR	0.168 (0.270)	0.214 (0.266)	0.362 (0.255)
		200	10	Our method	0.016 (0.018)	0.023 (0.025)	0.035 (0.038)
				LCR	0.025 (0.030)	0.048 (0.054)	0.040 (0.043)
(0.15, 0.85)	100	10	10	Our method	0.016 (0.018)	0.023 (0.025)	0.035 (0.038)
				LCR	0.025 (0.030)	0.048 (0.054)	0.040 (0.043)
		20	20	Our method	0.015 (0.014)	0.020 (0.022)	0.043 (0.043)
				LCR	0.019 (0.021)	0.049 (0.061)	0.156 (0.112)

Note: LCR denotes the method in [Liu and Chen \(2020\)](#) extended to analyze matrix time series data

References

- Bai, J. (2003), “Inferential theory for factor models of large dimensions.” *Econometrica*, 71, 135–171.
- Bai, J. and Ng, S. (2002), “Determining the number of factors in approximate factor models.” *Econometrica*, 70, 191–221.
- Bailey, N., Kapetanios, G., and Pesaran, M. (2016), “Exponent of Cross-Sectional Dependence: Estimation and Inference.” *Journal of Applied Econometrics*, 31, 926–960.
- Baltagi, B., Kao, C., and Wang, F. (2017), “Identification and estimation of a large factor model with structural instability.” *Journal of Econometrics*, 197, 87–100.
- Barigozzi, M. and Cho, H. (2020), “Consistent estimation of high-dimensional factor models when the factor number is over-estimated.” *Electronic Journal of Statistics*, 14, 2892–2921.

Table 10: Mean and standard deviation (in the parentheses) of $\mathcal{D}(\mathcal{M}(\hat{\mathbf{Q}}_{s,i}), \mathcal{M}(\mathbf{Q}_{s,i}))$ for two methods with different combinations of (T, p_1, p_2) , factor strength, and (η_1, η_2)

(η_1, η_2)	T	p_1	p_2	Method		$\delta_1 = 0, \delta_2 = 0$		$\delta_1 = 0.3, \delta_2 = 0$		$\delta_1 = 0.3, \delta_2 = 0.3$	
(0.05, 0.95)	100	10	10	Our method	row	0.035 (0.012)	0.035 (0.016)	0.056 (0.025)	0.059 (0.050)	0.107 (0.070)	0.101 (0.072)
				column	0.037 (0.015)	0.036 (0.013)	0.059 (0.031)	0.059 (0.035)	0.117 (0.098)	0.106 (0.070)	
			LCR	row	0.307 (0.248)	0.267 (0.254)	0.321 (0.238)	0.279 (0.244)	0.346 (0.218)	0.305 (0.227)	
				column	0.313 (0.254)	0.268 (0.251)	0.326 (0.242)	0.284 (0.240)	0.352 (0.224)	0.317 (0.223)	
		20	20	Our method	row	0.023 (0.007)	0.022 (0.005)	0.031 (0.019)	0.038 (0.011)	0.104 (0.078)	0.090 (0.051)
				column	0.022 (0.005)	0.022 (0.006)	0.040 (0.014)	0.039 (0.012)	0.099 (0.065)	0.100 (0.070)	
			LCR	row	0.296 (0.275)	0.306 (0.277)	0.309 (0.264)	0.321 (0.266)	0.340 (0.239)	0.265 (0.243)	
				column	0.295 (0.272)	0.303 (0.274)	0.306 (0.261)	0.319 (0.261)	0.343 (0.234)	0.365 (0.235)	
	200	10	10	Our method	row	0.025 (0.009)	0.023 (0.010)	0.037 (0.015)	0.036 (0.017)	0.060 (0.032)	0.059 (0.037)
				column	0.025 (0.015)	0.025 (0.009)	0.038 (0.026)	0.037 (0.015)	0.064 (0.050)	0.062 (0.039)	
			LCR	row	0.290 (0.248)	0.232 (0.244)	0.305 (0.242)	0.237 (0.235)	0.327 (0.227)	0.253 (0.221)	
				column	0.287 (0.244)	0.235 (0.244)	0.303 (0.237)	0.242 (0.237)	0.328 (0.223)	0.261 (0.224)	
		20	20	Our method	row	0.015 (0.004)	0.015 (0.004)	0.024 (0.007)	0.024 (0.006)	0.049 (0.020)	0.048 (0.022)
				column	0.016 (0.003)	0.015 (0.004)	0.026 (0.007)	0.026 (0.007)	0.054(0.024)	0.054 (0.030)	
(0.15, 0.85)	100	10	10	Our method	row	0.035 (0.012)	0.034 (0.015)	0.052 (0.020)	0.051 (0.025)	0.085 (0.043)	0.081 (0.045)
				column	0.036 (0.014)	0.035 (0.012)	0.055 (0.025)	0.053 (0.020)	0.094 (0.069)	0.085 (0.041)	
			LCR	row	0.068 (0.110)	0.079 (0.130)	0.101 (0.124)	0.104 (0.130)	0.162 (0.145)	0.149 (0.135)	
				column	0.066 (0.102)	0.082 (0.135)	0.094 (0.106)	0.113 (0.138)	0.175 (0.155)	0.169 (0.154)	
		20	20	Our method	row	0.023 (0.007)	0.022 (0.005)	0.039 (0.015)	0.036 (0.010)	0.078 (0.045)	0.071 (0.032)
				column	0.022 (0.005)	0.022 (0.006)	0.037 (0.010)	0.037 (0.010)	0.072 (0.026)	0.073 (0.030)	
			LCR	row	0.054 (0.118)	0.076 (0.157)	0.076 (0.117)	0.100 (0.159)	0.151 (0.143)	0.168 (0.163)	
				column	0.052 (0.111)	0.079 (0.155)	0.071 (0.108)	0.103 (0.154)	0.144 (0.128)	0.168 (0.155)	
	200	10	10	Our method	row	0.025 (0.009)	0.023 (0.010)	0.036 (0.014)	0.035 (0.016)	0.056 (0.026)	0.055 (0.032)
				column	0.025 (0.015)	0.025 (0.009)	0.038 (0.026)	0.036 (0.015)	0.061 (0.049)	0.058 (0.030)	
			LCR	row	0.025 (0.009)	0.024 (0.010)	0.039 (0.020)	0.036 (0.017)	0.055 (0.025)	0.054 (0.031)	
				column	0.025 (0.016)	0.025 (0.009)	0.040 (0.032)	0.038 (0.017)	0.060 (0.050)	0.057 (0.029)	
		20	20	Our method	row	0.015 (0.004)	0.015 (0.003)	0.023 (0.007)	0.024 (0.006)	0.044 (0.015)	0.043 (0.017)
				column	0.016 (0.004)	0.015 (0.004)	0.026 (0.006)	0.025 (0.006)	0.048 (0.017)	0.046 (0.017)	
LCR	row	0.015 (0.004)	0.015 (0.003)	0.026 (0.010)	0.026 (0.012)	0.062 (0.051)	0.055 (0.029)				
		column	0.016 (0.005)	0.015 (0.004)	0.028 (0.011)	0.027 (0.020)	0.069 (0.057)	0.064 (0.057)			

Note: LCR denotes the method in [Liu and Chen \(2020\)](#) extended to analyze matrix time series data

Barigozzi, M., Cho, H., and Fryzlewicz, P. (2018), “Simultaneous multiple change-point and factor analysis for high-dimensional time series.” *Journal of Econometrics*, 206, 187–225.

Bates, B. J., Plagborg-Møller, M., Stock, J. H., and Watson, M. W. (2013), “Consistent factor estimation in dynamic factor models with structural instability,” *Journal of Econometrics*, 177, 289–304.

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

Chen, E. Y. and Chen, R. (2019), “Modeling Dynamic Transport Network with Matrix Factor

- Models: with an Application to International Trade Flow,” *arXiv preprint arXiv:1901.00769*.
- Chen, E. Y. and Fan, J. (2021), “Statistical inference for high-dimensional matrix-variate factor models,” *Journal of the American Statistical Association*, 1–18.
- Chen, E. Y., Tsay, R. S., and Chen, R. (2020a), “Constrained Factor Models for High-Dimensional Matrix-Variate Time Series,” *Journal of the American Statistical Association*, 115, 775–793.
- Chen, E. Y., Yun, X., Chen, R., and Yao, Q. (2020b), “Modeling Multivariate Spatial-Temporal Data with Latent Low-Dimensional Dynamics,” *arXiv preprint arXiv:2002.01305*.
- Chen, L., Dolado, J., and Gonzalo, J. (2014), “Detecting big structural breaks in large factor models.” *Journal of Econometrics*, 180, 30–48.
- Chen, R. (1995), “Threshold variable selection in open-loop threshold autoregressive models.” *Journal of Time Series Analysis*, 16, 461–481.
- Fan, J., Xue, L., and Yao, J. (2017), “Sufficient forecasting using factor models.” *Journal of Econometrics*, 201, 292–306.
- Fan, J. and Yao, Q. (2008), *Nonlinear time series: nonparametric and parametric methods*, Springer Science & Business Media.
- Kim, C. J. and Nelson, C. R. (1998), “Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching.” *The Review of Economics and Statistics*, 80, 188–201.
- Lam, C. and Yao, Q. (2012), “Factor modeling for high-dimensional time series: inference for the number of factors,” *Annals of Statistics*, 40, 694–726.
- Lam, C., Yao, Q., and Bathia, N. (2011), “Estimation of latent factors for high-dimensional time series,” *Biometrika*, 98, 901–918.
- Liu, X. and Chen, R. (2016), “Regime-switching factor models for high-dimensional time series.” *Statistica Sinica*, 26, 1427–1451.
- (2020), “Threshold factor models for high-dimensional time series.” *Journal of Econometrics*, 216, 53–70.

- Liu, X. and Zhang, T. (2022), “Estimating Change-Point Latent Factor Models for High-Dimensional Time Series.” *Journal of Statistical Planning and Inference*, 217, 69–91.
- Ma, S. and Su, L. (2018), “Estimation of large dimensional factor models with an unknown number of breaks.” *Journal of Econometrics*, 207, 1–29.
- Massacci, D. (2017), “Least squares estimation of large dimensional threshold factor models.” *Journal of Econometrics*, 197, 101–129.
- Merikoski, J. K. and Kumar, R. (2004), “Inequalities for spreads of matrix sums and products.” *Applied Mathematics E-Notes*, 4, 150–159.
- Stock, J. H. and Watson, M. W. (2002), “Forecasting using principal components from a large number of predictors.” *Journal of the American Statistical Association*, 97, 1167–1179.
- Tong, H. (1990), *Nonlinear time series: a dynamic system approach*, Clarendon Press, Oxford.
- Tong, H. and Lim, K. (1980), “Threshold autoregression, limit cycles and cyclical data,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 42, 245–292.
- Tsay, R. (1989), “Testing and modeling threshold autoregressive process.” *Journal of the American Statistical Association*, 84, 231–240.
- (1998), “Testing and modeling multivariate threshold models.” *Journal of the American Statistical Association*, 93, 1188–1202.
- Wang, D., Liu, X., and Chen, R. (2019), “Factor Models for Matrix-Valued High-Dimensional Time Series,” *Journal of Econometrics*, 208, 231–248.
- Wu, J. (2021), “Estimation of high dimensional factor model with multiple threshold-type regime shifts,” *Computational Statistics and Data Analysis*, 157.
- Wu, S. and Chen, R. (2007), “Threshold variable selection and threshold variable driven switching autoregressive models,” *Statistica Sinica*, 17, 241–264.
- Yu, R., Cheng, D., and Liu, Y. (2015), “Accelerated online low rank tensor learning for multivariate spatiotemporal streams,” in *International Conference on Machine Learning*, PMLR, pp. 238–247.