Time-Varying Matrix Factor Models*

Bin Chen^{†1}, Elynn Y. Chen ^{‡2}, Stevenson Bolivar^{§3}, and Rong Chen^{¶3}

¹Department of Economics, University of Rochester ²Stern School of Business, New York University ³Department of Statistics, Rutgers University

April 3, 2024

Abstract

Matrix-variate data of high dimensions are frequently observed in finance and economics, spanning extended time periods, such as the long-term data on international trade flows among numerous countries. To address potential structural shifts and explore the matrix structure's informational context, we propose a time-varying matrix factor model. This model accommodates changing factor loadings over time, revealing the underlying dynamic structure through nonparametric principal component analysis and facilitating dimension reduction. We establish the consistency and asymptotic normality of our estimators under general conditions that allow for weak correlations across time, rows, or columns of the noise. A novel approach is introduced to overcome rotational ambiguity in the estimators, enhancing the clarity and interpretability of the estimated loading matrices. Our simulation study highlights the merits of the proposed estimators and the effective of the smoothing operation. In an application to international trade flow, we investigate the trading hubs, centrality, patterns, and trends in the trading network.

JEL Classifications: C13, C14, C32, C55

Keywords: High-dimensionality, Kernel estimation, Matrix factor models, Matrix time series, Time-varying loadings.

^{*}We thank Zeyu Wang and Yucheng Wong for their excellent research assistance. We also thank John Chao, Giovanni Motta, Mahrad Sharifvaghefi, Mark Watson, and seminar participants at Boston College, University of Maryland, Purdue University, Vanderbilt University, University of California, Riverside, the 2023 NBER-NSF Time Series Conference and the 15th Greater New York Metropolitan Area Econometrics Colloquium for their useful comments and discussions. Any remaining errors are solely ours. R. Chen's research is supported in part by National Science Foundation grants DMS-2027855, DMS-2052949 amd DMS-2319260.

[†]binchen@rochester.edu

[‡]elynn.chen@stern.nyu.edu

[§]sb2230@stat.rutgers.edu

[¶]rongchen@stat.rutgers.edu

1 Introduction

Matrix-variate data are increasingly prevalent across various fields, such as economics, finance, and social networks. Two significant types of such data are particularly noteworthy: Dynamic transportation networks are represented by matrices of monthly import-export volumes between countries, highlighting the complex, weighted, and directional relationships. Dynamic panel data, vital for investor portfolio decisions, are matrices capturing various firm attributes (e.g., stock prices, book-to-market ratio, dividend-to-price ratio) at different times, capturing firm performance and market position.

The field of statistical analysis for matrix-variate data is still emerging, with matrix factor models being pivotal for dimension reduction and uncovering inherent structures. Wang et al. (2019) introduced a matrix factor model for time series, featuring constant matrix factor loadings and time-varying latent factors, estimated through eigenanalysis of auto-cross-covariance matrices. This model has been expanded into a constrained version by Chen et al. (2020) and a threshold variant by Liu & Chen (2019). Contrasting these approaches, Chen & Fan (2023) developed a novel estimation method by analyzing an aggregate of the sample mean matrix and contemporary covariance matrices, proving to be advantageous in several contexts.

While matrix factor models have been proven effective, a common assumption in the literature is that factor loadings remain constant over time – an assumption not always realistic in practice. Changes in policies, preferences, technology, and environmental factors can influence economic variables' relationships, challenging the stability of factor loadings. Empirical evidence, such as that presented by Chen et al. (2022), indicates that ignoring these potential shifts in factor loadings can result in inconsistent estimates and unreliable conclusions.

Recently, the concept of smooth structural changes in economic variables has gar-

nered attention, leading to the development of time-varying time series and panel data models. These models, particularly nonparametric time-varying parameter models introduced by Robinson (1989) and further explored by Cai (2007), Chen & Hong (2012), Zhang & Wu (2012), Inoue et al. (2017), and Chen & Maung (2023), offer a flexible approach to modeling economic variables. These models stand out for their minimal assumptions on the time-varying parameters' functional forms, requiring only that they change smoothly over time. Inspired by the versatility and effectiveness of this approach, we employ this framework to propose time-varying matrix factor models for analyzing matrix time series.

In this paper, we investigate a matrix factor model with loading matrices defined as unknown, smooth, time-varying functions. We aim to estimate these functions and the latent factors using local Principal Component Analysis (PCA), developing a theory of inference that includes consistency, convergence rates, and limiting distributions. This approach accommodates time-varying loadings and weak correlations across time, rows, or columns of the noise matrices. We generalize the eigenvalue ratio-based estimator for latent dimensions (Ahn & Horenstein 2013) to our time-varying context and establish the consistency of the new method.

Traditional methods reduce matrix observations into vectors and apply local PCA to high-dimensional factor models (Su & Wang 2017), losing essential matrix structure and leading to less effective analyses. Our matrix-based estimator is compared against vectorized approaches through simulations to demonstrate its superior ability to preserve and leverage the matrix structure. Additionally, we tackle the rotation ambiguity inherent in matrix factor models by introducing a smoothing process for the time-varying loading spaces, facilitating clearer interpretation.

This paper is organized as follows: Section 2 introduces matrix factor models with time-varying loadings. Section 3 describes our local PCA estimation approach and a

generalized eigenvalue ratio-based estimator for latent dimensions. Section 4 discusses the theoretical foundations, including consistency and asymptotic normality. Section 5 details the smoothing process for time-varying functions. Section 6 assesses our methods through simulation, while Section 7 applies them to international trade data. All mathematical proofs and additional results are provided in the Appendix.

2 Matrix Factor Models with Time-Varying Loadings

Let Y_t , $t \in [T]$, be a matrix-valued time series, where each $Y_t = [Y_{t,ij}]$ is a matrix of size $p \times q$. We consider the following matrix factor models with time-varying factor loadings,

$$\boldsymbol{Y}_t = \boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{C}_t^\top + \boldsymbol{E}_t, \quad t \in [T], \tag{2.1}$$

where F_t is the $k \times r$ common factor matrix, R_t (C_t) is a $p \times k$ ($q \times r$) time-varying row (column) loading matrix and the sequence of matrices E_t is the noise matrix. The noise term E_t is assumed to be uncorrelated with F_t , yet is allowed to be weakly correlated across rows, columns, and times. It generalizes the matrix factor model:

$$\mathbf{Y}_t = \mathbf{R} \mathbf{F}_t \mathbf{C}^\top + \mathbf{E}_t, \tag{2.2}$$

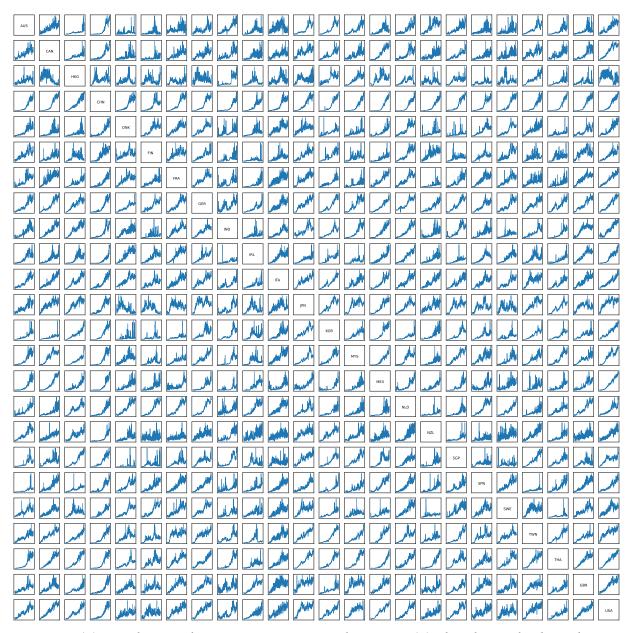
by allowing for structural changes in factor loadings R and C. To cover a wide range of potential smooth temporal variation, we follow the literature on smooth time-varying parameter models (Robinson 1989, Cai 2007, Su & Wang 2017)) and model $R_{t,i}$. ($C_{t,j}$.) as a non-stochastic function of t/T, that is,

$$R_{t,i.} = R_{i.}(t/T)$$
 and $C_{t,j.} = C_{j.}(t/T)$,

where $R_{t,i.}(\cdot)$ ($C_{t,j.}(\cdot)$) is an unknown smooth function of t/T on [0,1] for each i (j). The specification that loading matrices $R_{t,i.}(\cdot)$ and $C_{t,j.}(\cdot)$ are some functions of ratio t/T rather than time t only is a common scaling scheme in the literature for the ease of theoretical analysis (Phillips & Hansen 1990, Robinson 1991, Cai 2007). We note that

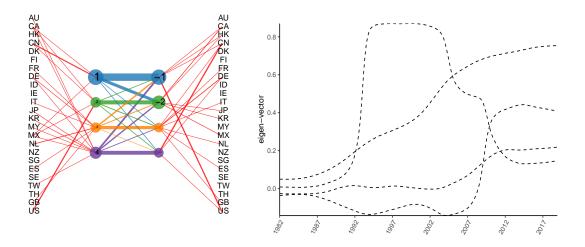
model (2.1) is subject to scale and rotational ambiguity in its individual components R_t , F_t and C_t . Our objective is to estimate sensible representations of R_t and C_t among their equivalent classes and the corresponding F_t , as the signal $S_t = R_t F_t C_t^{\top}$ is uniquely defined.

Figure 1: Time series plots of the value of goods traded among 24 countries.



Notes: (1) sample period: January 1982- December 2018. (2) The plots only show the patterns of the time series while the magnitudes are not comparable between plots because the ranges of the y-axis are different.

Figure 2: (left) Estimated transport network, June, 2016, and (right) the evolution of one row vector of R_t with a 5-year rolling window.



As a motivating example, we introduce the international trade flow time series, with a detailed analysis in Section 7. The data, \mathbf{Y}_t , is represented as a square matrix for each time t, where each element, $y_{t,ij}$, indicates the trade volume from country i (exporter) to country *j* (importer). The time series, spanning from January 1982 to December 2018, depicted in Figure 1, showcases the non-stationary nature of the data, highlighting a globalization trend. Chen & Chen (2023) interpreted the matrix factor model for transport data, viewing the factor process F_t as representing trade volumes among (latent) trading hubs, analogous to trade volumes among groups of countries. This conceptualization suggests that a country's exports pass through these trading hubs before reaching the importing country, with F_t detailing hub-to-hub trade activities. The loading matrices R_t and C_t indicate each country's involvement with the hubs, with elements like $R_{t,11}$ showing country 1's interaction with export hub 1. The cumulative export volume of a country, such as country 1, is calculated as $\sum_{k=1}^{r} \mathbf{R}_{t,1k} \mathbf{F}_{t,k1}$. A visual representation for June 2016 (left figure in Figure 2) demonstrates this model, where hub sizes reflect trade volume F_t and line widths show participation levels R_t and C_t , with minor values (less than 0.15) excluded for clarity. This visualization reveals, for example, significant trade between China's primary export hub and the main import hubs (Hub 1 and 2) used by the US, Canada, Mexico, Japan, Korea, and Taiwan in that month. The right figure in Figure 2 illustrates changes over time in China's export loading vectors, highlighting structural shifts in trading patterns based on a five-year moving analysis.

3 Estimation

The time-varying nature of (2.1) presents a significant challenge in estimating R_t , C_t , and F_t , as there is effectively only one sample point available at each time t. To see this challenge, we first analyze the problem on the population level. Under the time-varying matrix factor model (2.1), the second *uncentered* moment writes

$$\mathbb{E}\left[\boldsymbol{Y}_{t}\boldsymbol{Y}_{t}^{\top}\right] = \boldsymbol{R}_{t}\mathbb{E}\left[\boldsymbol{F}_{t}\boldsymbol{C}_{t}^{\top}\boldsymbol{C}_{t}\boldsymbol{F}_{t}^{\top}\right]\boldsymbol{R}_{t}^{\top} + \mathbb{E}\left[\boldsymbol{E}_{t}\boldsymbol{E}_{t}^{\top}\right],\tag{3.1}$$

$$\mathbb{E}\left[\boldsymbol{Y}_{t}^{\top}\boldsymbol{Y}_{t}\right] = \boldsymbol{C}_{t}\mathbb{E}\left[\boldsymbol{F}_{t}^{\top}\boldsymbol{R}_{t}^{\top}\boldsymbol{R}_{t}\boldsymbol{F}_{t}\right]\boldsymbol{C}_{t}^{\top} + \mathbb{E}\left[\boldsymbol{E}_{t}^{\top}\boldsymbol{E}_{t}\right],\tag{3.2}$$

which aggregate the information of unknown parameters R_t and C_t from both the first moment $\mathbb{E}[Y_t] = R_t \mathbb{E}[F_t] C_t^{\top}$ and the second *centered* moment $\mathbb{E}[(Y_t - \mathbb{E}[Y_t]) (Y_t - \mathbb{E}[Y_t])^{\top}]$.

Noticing that the first term on the right-hand-side of (3.1) is of rank k under some mild conditions and ignoring the second term (as justified by the pervasive assumption below), it is easy to see R_t is the same as the top k eigenvectors of the second moment, up to an affine transformation. Similar observation holds for C_t with (3.2).

This motivates the employment of the spectral method based on the sample estimators of $\mathbb{E}\left[\boldsymbol{Y}_{t}\boldsymbol{Y}_{t}^{\top}\right]$ and $\mathbb{E}\left[\boldsymbol{Y}_{t}^{\top}\boldsymbol{Y}_{t}\right]$ to estimate \boldsymbol{R}_{t} and \boldsymbol{C}_{t} . As model (2.1) is time varying, we use nonparametric kernel smoothing to estimate $\mathbb{E}\left[\boldsymbol{Y}_{t}\boldsymbol{Y}_{t}^{\top}\right]$ and $\mathbb{E}\left[\boldsymbol{Y}_{t}^{\top}\boldsymbol{Y}_{t}\right]$, based on the smoothness assumption of the loading functions $\boldsymbol{R}_{i}.(t/T)$ and $\boldsymbol{C}_{i}.(t/T)$ and the stationary assumptions of the factor process \boldsymbol{F}_{t} , which leads to smoothness of $\mathbb{E}\left[\boldsymbol{Y}_{t}\boldsymbol{Y}_{t}^{\top}\right]$ and $\mathbb{E}\left[\boldsymbol{Y}_{t}^{\top}\boldsymbol{Y}_{t}\right]$ in t/T. Specifically, we define sample scatter matrix $\widehat{\boldsymbol{M}}_{R,t}$ and $\widehat{\boldsymbol{M}}_{C,t}$ as

$$\widehat{\boldsymbol{M}}_{R,t} = \frac{1}{pqT} \sum_{s=1}^{T} K_{h,ts} \boldsymbol{Y}_{s} \boldsymbol{Y}_{s}^{\top} \quad \text{and} \quad \widehat{\boldsymbol{M}}_{C,t} = \frac{1}{pqT} \sum_{s=1}^{T} K_{h,ts} \boldsymbol{Y}_{s}^{\top} \boldsymbol{Y}_{s}$$
(3.3)

where $K_{h,st}$ is a boundary-modified kernel function:

$$K_{h,st} = \begin{cases} h^{-1}k(\frac{s-t}{Th})/\int_{-t/(Th)}^{1}k(u)\,du & if\,t \in [1,\lfloor Th\rfloor], \\ h^{-1}k(\frac{s-t}{Th}) & if\,t \in [\lfloor Th\rfloor, T-\lfloor Th\rfloor], \\ h^{-1}k(\frac{s-t}{Th})/\int_{-1}^{(1-t/T)/h}k(u)\,du & otherwise. \end{cases}$$
(3.4)

The kernel $k(\cdot): [-1,1] \to \mathbb{R}^+$ is a predetermined symmetric probability density function, with h = h(p,q,T) as the bandwidth parameter for the row (column) loading matrix $R_t(C_t)$, and $\lfloor Th \rfloor$ representing the integer part of Th. Common examples of $k(\cdot)$ are the uniform, Epanechnikov, and quartic kernels (Hastie et al. 2009).

To estimate R_t , and C_t , we use \sqrt{p} (\sqrt{q}) multiplied by the eigenvectors corresponding to the k (r) largest eigenvalues of $\widehat{M}_{R,t}$ ($\widehat{M}_{C,t}$). Due to rotational ambiguity, the solution \widehat{R}_t (\widehat{C}_t) is one of the representatives of the column space of R_t (C_t). They do not necessarily constitute a smooth function $R_t = R(t/T)$ as t changes. In Section 5 we present a smoothing procedure to obtain a smooth representation. Furthermore, the realized latent factor F_t and the signal part S_t can be obtained by

$$\widehat{F}_t = \frac{1}{pq} \widehat{R}_t^{\top} Y_t \widehat{C}_t$$
 and $\widehat{S}_t = \frac{1}{pq} \widehat{R}_t \widehat{R}_t^{\top} Y_t \widehat{C}_t \widehat{C}_t^{\top}$.

The above local-PCA-based method assumes that the latent dimensions $k \times r$ are known. However, in practice, we need to estimate k and r as well. We extend the eigenvalue ratio-based estimator, proposed by Ahn & Horenstein (2013). Let $\widehat{\lambda}_{1,t} \geq \widehat{\lambda}_{2,t} \geq \cdots \geq \widehat{\lambda}_{k,t} \geq 0$ be the ordered eigenvalues of $\widehat{M}_{R,t}$. The generalized eigenvalue ratio-based estimator for k can be defined as

$$\widehat{k} = \underset{1 \le j \le k_{max}}{\operatorname{argmax}} \frac{1}{T} \sum_{t=1}^{T} \frac{\widehat{\lambda}_{j,t}}{\widehat{\lambda}_{j+1,t}},$$
(3.5)

and k_{max} is a selected upper bound. We could choose $k_{max} = \lfloor p/2 \rfloor$ or $k_{max} = \lfloor p/3 \rfloor$. The estimator \widehat{r} can be defined analogously with respect to $\widehat{M}_{C,t}$.

4 Theory

Let $R_t^{(1)}\left(C_t^{(1)}\right)$ and $R_t^{(2)}\left(C_t^{(2)}\right)$ denote the first and second order derivative of $R_t\left(C_t\right)$ respectively, and $R_t^{(0)}=R_t$ and $C_t^{(0)}=C_t$. To derive the asymptotic properties of our estimators, we impose the following regularity conditions.

Assumption 1 (\$\alpha\$-mixing). The vector-variate process $\{ \text{VEC}(\pmb{F}_t)^\top, \text{VEC}(\pmb{E}_t)^\top \}^\top$ is a \$\alpha\$-mixing process with mixing coefficients \$\alpha(h) = \sup_{\tau} \sup_{A \in \mathcal{F}_{-\infty}^{\tau}, B \in \mathcal{F}_{\tau+h}^{\infty}} |P(A \cap B) - P(A)P(B)|, where \$\mathcal{F}_{\tau}^{s}\$ is the \$\alpha\$-field generated by $\{ \text{VEC}(\pmb{F}_t)^\top, \text{VEC}(\pmb{E}_t)^\top \}^\top : \tau \leq t \leq s \}$, and the mixing coefficients satisfy the condition that \$\sum_{h=1}^{\infty} \alpha(h)^{1-2/\gamma} < \infty\$, for some \$\gamma > 2\$.

Assumption 2 (Factor and noise matrices). There is a positive constant $C < \infty$ such that

- (a) Factor matrix \mathbf{F}_t is of fixed dimension $k \times r$, and $\max_t \mathbb{E} ||\mathbf{F}_t||^4 \le C$.
- (b) For all $i \in [p]$, $j \in [q]$ and $t \in [T]$, $\mathbb{E}\left[e_{t,ij}\right] = 0$ and $\mathbb{E}|e_{t,ij}|^8 \le C$.
- (c) Factor and noise are uncorrelated, that is, $\mathbb{E}\left[e_{t,ij}f_{s,lh}\right] = 0$ for any $t,s \in [T]$, $i \in [p]$, $j \in [q]$, $l \in [k]$, and $h \in [r]$.

Assumption 3 (Loading matrix). There exists a positive constant $C < \infty$ such that

- (a) For each row of \mathbf{R}_t , $\max_t \|\mathbf{R}_{t,i}\| \le C$, and, as $p, q \to \infty$, we have $\max_t \|p^{-1}\mathbf{R}_t^{\top}\mathbf{R}_t \mathbf{\Omega}_{R,t}\| \to 0$ for some $k \times k$ positive definite matrix $\mathbf{\Omega}_{R,t}$.
- (b) For each row of C_t , $\max_t \|C_{t,i}\| \le C$, and, as $p, q \to \infty$, $\max_t \|q^{-1}C_t^{\top}C_t \Omega_{C,t}\| \to 0$ for some $r \times r$ positive definite matrix $\Omega_{C,t}$.

Assumption 4 (Cross row (column) correlation of noise E_t). There exists some positive constant $C < \infty$ such that,

- (a) Let $\mathbf{U}_E = \mathbb{E}\left[\mathbf{E}_t \mathbf{E}_t^T / q\right]$ and $\mathbf{V}_E = \mathbb{E}\left[\mathbf{E}_t^T \mathbf{E}_t / p\right]$, we assume $\|\mathbf{U}_E\|_1 \le C$ and $\|\mathbf{V}_E\|_1 \le C$.
- (b) For all $i \in [p]$ and $j \in [q]$ and $t \in [T]$, we assume $\sum_{\substack{l \in [p] \\ l \neq i}} \sum_{\substack{h \in [q] \\ h \neq j}} |\mathbb{E}\left[e_{t,ij}e_{t,lh}\right]| \leq C$.

(c) For any row $i, l \in [p]$, any time $t, s \in [T]$, and any column $j \in [q]$,

$$\sum_{\substack{m \in [p]}} \sum_{\substack{h \in [q] \\ h \neq j}} |\text{Cov}\left[\left(e_{t,ij}e_{t,lj}, e_{s,ih}e_{s,mh}\right)\right]| \leq C$$

Similar, for any column $j,h \in [q]$, any time $t \in [T]$, and any row $i \in [p]$,

$$\sum_{m \in [q]} \sum_{\substack{s \in [T]}} \sum_{\substack{l \in [p] \\ l \neq i}} |\text{Cov}\left[\left(e_{t,ij}e_{t,ih}, \ e_{s,lj}e_{s,lm}\right)\right]| \leq C.$$

Assumption 5. $R_{i\cdot}(t/T)$ and $C_{j\cdot}(t/T)$ have continuous derivatives up to the second order. Moreover, there exists m > 2, $1 < a,b < \infty$, 1/a + 1/b = 1, c = 0,1,2 such that, for some positive $C < \infty$, and any $t \in [T]$,

(a) For any
$$l \in [k]$$
, and $i \in [p]$, $\mathbb{E}\left[\left|\frac{1}{\sqrt{q}}\sum_{j=1}^{q}e_{t,ij}\right|^{mb}\right] = \mathcal{O}(1)$, $\mathbb{E}\left[\left\|\frac{1}{\sqrt{q}}\sum_{j=1}^{q}C_{t,j}^{(c)}e_{t,ij}\right\|^{mb}\right] = \mathcal{O}(1)$, and $\max_{t} \mathbb{E}\left[\left\|f_{t,l}\right\|^{ma}\right] \leq C$.

(b) For any
$$h \in [r]$$
, and $j \in [q]$, $\mathbb{E}\left[\left|\frac{1}{\sqrt{p}}\sum_{i=1}^{p}e_{t,ij}\right|^{mb}\right] = \mathcal{O}(1)$, $\mathbb{E}\left[\left\|\frac{1}{\sqrt{p}}\sum_{i=1}^{p}\mathbf{R}_{t,i}^{(c)}e_{t,ij}\right\|^{mb}\right] = \mathcal{O}(1)$, and $\max_{t} \mathbb{E}\left[\left\|\mathbf{f}_{t,t}\right\|^{ma}\right] \leq C$.

$$(c) \ \mathbb{E}\left[|\tfrac{1}{\sqrt{pq}}\sum_{i=1}^p\sum_{j=1}^q e_{t,ij}|^{mb}\right] = \mathcal{O}(1) \ and \ \mathbb{E}\left[\left\|\tfrac{1}{\sqrt{pq}}\sum_{i=1}^p\sum_{j=1}^q \boldsymbol{R}_{t,i}^{(c)}\boldsymbol{C}_{t,j}^{(c)\top}e_{t,ij}\right\|^{mb}\right] = \mathcal{O}(1).$$

Assumption 6. $k: [-1,1] \to \mathbb{R}^+$ is a symmetric and Lipschitz continuous probability density function such that $\int_{-1}^1 k(u) du = 1$, $\int_{-1}^1 uk(u) du = 0$, and $\int_{-1}^1 u^2 k(u) du < \infty$.

The α -mixing condition in Assumption 1 allows weak temporal correlations for both the factors and noises. Assumption 2 mainly imposes moment conditions on the errors, factors, and their interactions. Assumption 3 is an extension of the pervasive assumption (Stock & Watson 2002) to the matrix variate data. It ensures that each row and column of the factor matrix F_t has a nontrivial contribution to the variance of rows and columns of Y_t . Assumptions 1-3 are standard in the literature on factor models and similar assumptions have been imposed in Chen & Fan (2023) and Yu et al. (2022).

Assumption 1 addresses temporal dependence, while the dimensions p and q influence our estimator's convergence rates through cross-row and cross-column dependencies. To leverage information across rows and columns, we introduce Assumptions 4 and 5, which are met if errors E_t are i.i.d. across rows and columns at any time t, independent from factor F_t , and meet specific moment conditions. These assumptions accommodate mild cross-row, cross-column, and temporal dependencies. Assumption 6 is standard for kernel regression and is fulfilled by typical second-order kernels like Epanechnikov, uniform, and quartic kernels.

Let $\widehat{V}_{R,t} \in \mathbb{R}^{k \times k}$ and $\widehat{V}_{C,t} \in \mathbb{R}^{r \times r}$ be the time-varying diagonal matrices consisting of the first k and r largest eigenvalues of $\widehat{M}_{R,t}$ ($\widehat{M}_{C,t}$) in a descending order respectively. By the definition of our estimators \widehat{R}_t and \widehat{C}_t , we have

$$\widehat{R}_t = \widehat{M}_{R,t} \widehat{R}_t \widehat{V}_{R,t}^{-1}$$
 and $\widehat{C}_t = \widehat{M}_{C,t} \widehat{C}_t \widehat{V}_{C,t}^{-1}$.

Similar to traditional factor models, the latent factor matrix F_t and loading matrices R_t and C_t cannot be identified separately but can be estimated through an invertible matrix transformation. There are invertible matrices $H_{R,t}$ and $H_{C,t}$ making \widehat{R}_t and \widehat{C}_t consistent estimators of $R_tH_{R,t}$ and $C_tH_{C,t}$, respectively.

$$\boldsymbol{H}_{R,t} = \frac{1}{Tpq} \sum_{s=1}^{T} K_{h,ts} \boldsymbol{F}_{s} \boldsymbol{C}_{t}^{\top} \boldsymbol{C}_{t} \boldsymbol{F}_{s}^{\top} \boldsymbol{R}_{t}^{\top} \widehat{\boldsymbol{R}}_{t} \widehat{\boldsymbol{V}}_{R,t}^{-1}, \quad \text{and} \quad \boldsymbol{H}_{C,t} = \frac{1}{Tpq} \sum_{s=1}^{T} K_{h,ts} \boldsymbol{F}_{s} \boldsymbol{R}_{t}^{\top} \boldsymbol{R}_{t} \boldsymbol{F}_{s}^{\top} \boldsymbol{C}_{t}^{\top} \widehat{\boldsymbol{C}}_{t} \widehat{\boldsymbol{V}}_{C,t}^{-1}.$$

$$(4.1)$$

Theorem 1. Suppose Assumptions 1-6 are satisfied. As k, r fixed, $h \to 0$, $Th^3 \to \infty$, and $T, p, q \to \infty$, we have

$$\frac{1}{p}\left\|\widehat{\boldsymbol{R}}_{t}-\boldsymbol{R}_{t}\boldsymbol{H}_{R,t}\right\|_{F}^{2}=\mathcal{O}_{p}\left(\frac{1}{p^{2}}+\frac{1}{Tqh}+h^{4}\right),\quad and \quad \frac{1}{q}\left\|\widehat{\boldsymbol{C}}_{t}-\boldsymbol{C}_{t}\boldsymbol{H}_{C,t}\right\|_{F}^{2}=\mathcal{O}_{p}\left(\frac{1}{q^{2}}+\frac{1}{Tph}+h^{4}\right).$$

Consequently,

$$\frac{1}{p}\left\|\widehat{\boldsymbol{R}}_{t}-\boldsymbol{R}\boldsymbol{H}_{R,t}\right\|^{2}=\mathcal{O}_{p}\left(\frac{1}{p^{2}}+\frac{1}{Tqh}+h^{4}\right),\quad and\quad \frac{1}{q}\left\|\widehat{\boldsymbol{C}}_{t}-\boldsymbol{C}\boldsymbol{H}_{C,t}\right\|^{2}=\mathcal{O}_{p}\left(\frac{1}{q^{2}}+\frac{1}{Tph}+h^{4}\right).$$

Theorem 1 details the convergence rates of our nonparametric estimators, \widehat{R}_t and \widehat{C}_t . By leveraging the local vector factor model approach to vectorized data (Su & Wang (2017)), we explore the dynamics of matrix-valued data with:

$$\operatorname{VEC}(\boldsymbol{Y}_t) = \boldsymbol{\Xi}_t \operatorname{VEC}(\boldsymbol{F}_t) + \operatorname{VEC}(\boldsymbol{E}_t), \tag{4.2}$$

where $\operatorname{vec}(\boldsymbol{Y}_t) \in \mathbb{R}^{pq}$ and $\operatorname{vec}(\boldsymbol{F}_t) \in \mathbb{R}^{kr}$, respectively. This method, while bypassing the tensor structure in the loading matrix, leads to convergence rates of $\frac{1}{pq} \left\| \widehat{\boldsymbol{\Xi}}_t - \boldsymbol{\Xi}_t \boldsymbol{H}_t \right\|^2 = \mathcal{O}_p \left(\frac{1}{p^2 q^2} + \frac{1}{Th} + h^4 \right)$, with \boldsymbol{H}_t being an orthonormal matrix¹. However, modeling $\operatorname{vec}(\boldsymbol{Y}_t)$ as $(\boldsymbol{C}_t \otimes \boldsymbol{R}_t) \cdot \operatorname{vec}(\boldsymbol{F}_t) + \operatorname{vec}(\boldsymbol{E}_t)$ within our specific framework offers improved convergence rates due to its unique structure. Estimation of \boldsymbol{R}_t and \boldsymbol{C}_t from $\boldsymbol{\Xi}_t$ using Kronecker product decomposition introduces potential inaccuracies (Cai et al. 2022).

To analyze the limiting distribution of the nonparametric estimators \widehat{R}_t and \widehat{C}_t , we introduce

$$\Sigma_{FR,t}^{(c_1,c_2)} = E\left(\boldsymbol{F}_t^{\top} \frac{\boldsymbol{R}_t^{(c_1)\top} \boldsymbol{R}_t^{(c_2)}}{p} \boldsymbol{F}_t\right), \text{ and } \Sigma_{FC,t}^{(c_1,c_2)} = E\left(\boldsymbol{F}_t \frac{\boldsymbol{C}_t^{(c_1)\top} \boldsymbol{C}_t^{(c_2)}}{q} \boldsymbol{F}_t^{\top}\right), \tag{4.3}$$

and

$$\Sigma_{R,t}^{(c)} = \frac{\boldsymbol{R}_t^{\top} \boldsymbol{R}_t^{(c)}}{p}$$
, and $\Sigma_{C,t}^{(c)} = \frac{\boldsymbol{C}_t^{\top} \boldsymbol{C}_t^{(c)}}{q}$,

where c_1 , c_2 and c=0,1,2, $\Sigma_{FR,t}^{(0,0)}=\Sigma_{FR,t}$ and $\Sigma_{FC,t}^{(0,0)}=\Sigma_{FC,t}$

Assumption 7. The matrices $\Omega_{R,t}\Sigma_{FC,t}$ and $\Omega_{C,t}\Sigma_{FR,t}$ are positive definite with distinct eigenvalues for each t.

Theorem 2. Under Assumptions 1-7, as k, r fixed, $h \to 0$, $Th^3 \to \infty$, and $T, p, q \to \infty$,

(i) For row loading \mathbf{R}_t , if $\sqrt{qTh}/p \to 0$, for each $i \in [p]$ and $t \in [T]$, we have

$$\sqrt{qTh}\left(\widehat{\boldsymbol{R}}_{t,i.} - \boldsymbol{H}_{R,t}^{\top} \boldsymbol{R}_{t,i.} - \boldsymbol{B}_{R_{t,i}}\right) \rightarrow^{d} N(0, \nu_{0} \Sigma_{R_{t,i}}),$$

where

¹The bias terms are missing in Su & Wang (2017) but have been corrected in Su & Wang (2020).

$$\boldsymbol{B}_{R_{t,i}} = \frac{1}{2} \boldsymbol{V}_{R,t}^{-1} h^2 \mu_2 \sum_{c_1 + c_2 + c_3 + c_4 = 2} \boldsymbol{H}_{R,t}^{\top} \boldsymbol{\Sigma}_{R,t}^{(c_1)} \boldsymbol{\Sigma}_{FC,t}^{(c_2,c_3)} \boldsymbol{R}_{t,i.}^{(c_4)} + o_P(h^2),$$

$$\Sigma_{R_{t,i}} = \boldsymbol{V}_{R,t}^{-1} \boldsymbol{Q}_{R,t} \boldsymbol{\Phi}_{R,t,i} \boldsymbol{Q}_{R,t}^{\top} \boldsymbol{V}_{R,t}^{-1},$$

$$\mathbf{\Phi}_{R,t,i} = \underset{q,T \to \infty}{\text{plim}} \frac{1}{qT} \sum_{r=1}^{T} \sum_{s=1}^{T} \mathbb{E} \left[\mathbf{F}_{r} \mathbf{C}_{t}^{\top} \mathbf{e}_{r,i} \cdot \mathbf{e}_{s,i}^{\top} \cdot \mathbf{C}_{t} \mathbf{F}_{s}^{\top} \right], \tag{4.4}$$

$$\begin{split} \nu_0 &= \int_{-1}^1 k^2(u) du, \ \mu_2 = \int_{-1}^1 u^2 k(u) du, \ \mathbf{Q}_{R,t} = \mathbf{V}_{R,t}^{1/2} \mathbf{\Psi}_{R,t}^\top \boldsymbol{\Sigma}_{FC,t}^{-1/2}, \ \mathbf{V}_{R,t} \ \text{is a diagonal matrix} \\ \text{whose entries are the eigenvalues of } \boldsymbol{\Sigma}_{FC,t} \ \text{in decreasing order, } \boldsymbol{\Psi}_{R,t} \ \text{is the corresponding} \\ \text{eigenvector matrix such that } \boldsymbol{\Psi}_{R,t}^\top \boldsymbol{\Psi}_{R,t} = \mathbf{I}, \ \text{and } \boldsymbol{\Sigma}_{FC,t} \ \text{is defined in (4.3)}. \end{split}$$

(ii) For column row loading C_t , if $\sqrt{pTh}/q \to 0$, for each $j \in [q]$ and $t \in [T]$, we have

$$\sqrt{pTh}\left(\widehat{\boldsymbol{C}}_{t,j.} - \boldsymbol{H}_{C,t}^{\top} \boldsymbol{C}_{t,j.} - \boldsymbol{B}_{C_{t,j}}\right) \rightarrow^{d} N(0, \nu_{0} \Sigma_{C_{t,j}}),$$

where

$$\boldsymbol{B}_{C_{t,j}} = \frac{1}{2} \boldsymbol{V}_{C,t}^{-1} h^2 \mu_2 \sum_{c_1 + c_2 + c_3 + c_4 = 2} \boldsymbol{H}_{C,t}^{\top} \boldsymbol{\Sigma}_{C,t}^{(c_1)} \boldsymbol{\Sigma}_{FR,t}^{(c_2,c_3)} \boldsymbol{C}_{t,j.}^{(c_4)} + o_P(h^2),$$

and

$$\Sigma_{C_{t,i}} = V_{C,t}^{-1} Q_{C,t} \Phi_{C,t,i} Q_{C,t}^{\top} V_{C,t}^{-1}, \tag{4.5}$$

 $\Phi_{C,t,j}$ is defined similarly to $\Phi_{R,t,i}$, $Q_{C,t} = V_{C,t}^{1/2} \Psi_{C,t}^{\top} \Sigma_{FR,t}^{-1/2}$, $V_{C,t}$ is a diagonal matrix whose entries are the eigenvalues of $\Sigma_{FR,t}$ in decreasing order, $\Psi_{C,t}$ is the corresponding eigenvector matrix such that $\Psi_{C,t}^{\top} \Psi_{C,t} = \mathbf{I}$, and $\Sigma_{FR,t}$ is defined in (4.3).

Theorem 2 establishes the asymptotic normality of \widehat{R}_t and \widehat{C}_t estimators. Both in the interior and boundary regions, the asymptotic bias $B_{R_{t,i}}$ and $B_{C_{t,j}}$ is linked to h^2 , attributed to employing a boundary-adjusted kernel function as specified in (3.4). The pointwise asymptotic mean squared error (AMSE) of $\widehat{R}_{t,i}$ can be easily determined by

$$AMSE_{R} = \frac{h^{4}\mu_{2}^{2}}{4} \left\| \sum_{c_{1}+c_{2}+c_{3}+c_{4}=2} \mathbf{V}_{R,t}^{-1} \mathbf{H}_{R,t}^{\top} \mathbf{\Sigma}_{R,t}^{(c_{1})} \mathbf{\Sigma}_{FC,t}^{(c_{2},c_{3})} \mathbf{R}_{t,i.}^{(c_{4})} \right\|^{2} + \frac{\text{Tr}\left(\nu_{0} \mathbf{\Sigma}_{R_{t,i}}\right)}{qTh}, \tag{4.6}$$

and hence the optimal bandwidth for $\widehat{R}_{t,i}$ is

$$h_{R}^{opt} = (qT)^{-\frac{1}{5}} \left(\frac{\text{Tr}\left(\nu_{0} \Sigma_{R_{t,i}}\right)}{\mu_{2}^{2} \left\| \sum_{c_{1}+c_{2}+c_{3}+c_{4}=2} V_{R,t}^{-1} H_{R,t}^{\top} \Sigma_{R,t}^{(c_{1})} \Sigma_{FC,t}^{(c_{2},c_{3})} R_{t,i.}^{(c_{4})} \right\|^{2}} \right)^{\frac{1}{5}}.$$
 (4.7)

The optimal bandwidth for $\widehat{C}_{t,i}$ can be derived similarly.

Theorem 3. Suppose Assumptions 1-7 are satisfied. As k, r fixed, $h \to 0$, $Th^3 \to \infty$, and $T, p, q \to \infty$, we have

$$\widehat{\boldsymbol{F}}_{t} - \boldsymbol{H}_{R,t}^{-1} \boldsymbol{F}_{t} \boldsymbol{H}_{C,t}^{-1} = \mathcal{O}_{p} \left(\frac{1}{\min(p,q)} + h^{2} \right),$$

$$\widehat{\boldsymbol{S}}_{t,ij} - \boldsymbol{S}_{t,ij} = \mathcal{O}_{p} \left(\frac{1}{\min(p,q,\sqrt{pTh},\sqrt{qTh})} + h^{2} \right),$$

for any $1 \le i \le p$ and $1 \le j \le q$.

Theorem 3 derives the convergence rate of the estimated latent factor \widehat{F}_t and signal \widehat{S}_t . In order to achieve consistency, we require dimensions p and q approach infinity. That is because we need sufficient information accumulation to distinguish the signal S_t from the noise E_t at each time point t. Theorems 2 and 3 present the asymptotic properties when the dimension of the latent matrix factor $k \times r$ is assumed to be known. In practice we can estimate k and r via the generalized eigenvalue ratio-based estimators defined in (3.5). The asymptotic validity of k and k is justified in Theorem 4 below.

Theorem 4. Suppose Assumptions 1-7 are satisfied and k_{max} is a predetermined constant no smaller than k or r. Then

$$P(\widehat{k}=k) \rightarrow 1$$
 and $P(\widehat{r}=r) \rightarrow 1$

as $h \to 0$, $Th^3 \to \infty$, $T, p, q \to \infty$, where \hat{k} and \hat{r} are defined in (3.5).

Theorem 4 derives the consistency of rank estimators \hat{k} and \hat{r} based on the eigenvalue ratios. This can be viewed as a generalization of Theorem 1 of Ahn & Horenstein (2013) from constant parameter factor models to time-varying matrix factor models.

5 Smooth Loading Functions

In this section, we present an approach to obtaining a 'smooth' estimation of the loading functions $\mathbf{R}_t = \mathbf{R}(t/T)$ ($\mathbf{C}_t = \mathbf{C}(t/T)$), which is important for interpretation and visualization. We present the method for \mathbf{R}_t , and that for \mathbf{C}_t is similar.

First we note that the estimator \widehat{R}_t obtained in Section 3 are not necessarily smooth over t/T, which is incurred by the fact that \widehat{R}_t are estimated separately for each time t, the inherit rotational ambiguity associated with PCA, and the possible eigenvalue coalescing at time t-1 and t. Specifically, the columns of \widehat{R}_t are \sqrt{p} multiplied by the top k eigenvectors of $\widehat{M}_{R,t}$ corresponding to the $k \times 1$ eigenvalue vector $\widehat{\alpha}_t = [\widehat{\alpha}_{t,1}, \cdots, \widehat{\alpha}_{t,r}]^{\top}$ in descending order. The matrix $M_{R,t}$ and $\widehat{M}_{R,t}$ are both smooth functions of t, under the assumption that R_t and C_t are smooth. When $\widehat{M}_{R,t}$ and $\widehat{M}_{R,t-1}$ are close, the eigenvectors \widehat{R}_t and \widehat{R}_{t-1} and the eigenvalues $\widehat{\alpha}_t$ and $\widehat{\alpha}_{t-1}$ would be close most of the time, except in two situations: (1) a **sign change** in the eigenvalue/eigenvector when $\widehat{\alpha}_{t-1,i}$ and $\widehat{\alpha}_{t,i}$ changes signs; and (2) an **order switching** in the eigenvalues when the eigenvector $\widehat{R}_{\cdot j}((t-1)/T)$ is closer to $\widehat{R}_{\cdot j+1}(t/T)$ than to $\widehat{R}_{\cdot j}(t/T)$. This happens when the eigenvalues at t-1 and t corresponding to the smooth eigenvectors switch order from t to t and vice versa. We assume the switching only involves two adjacent eigenvalues.

Motta et al. (2023) introduce the $\sqrt{2}$ -signed and $\sqrt{2}$ -bridged methods for tracking smoothly evolving covariance matrices in independent normal observations. The $\sqrt{2}$ -signed technique assigns eigenvector signs at time t based on closeness to those at t-1, starting with an order determined by singular values at t=1. The $\sqrt{2}$ -bridged approach

identifies coalescing areas where two eigenvalues nearly interchange in magnitude.

Our setting is different from that considered in Motta et al. (2023) as we consider cross-moment $M_{R,t}$ with a potentially underlying time varying mean, and there is a strong serial dependence induced by the smoothing operation in the construction of $\widehat{M}_{R,t}$. Empirical study shows that the approach by Motta et al. (2023) indeed encounters difficulty in our situation. We propose an alternative method for identifying coalescing regions, a crucial input for the second stage of the procedure introduced by Motta et al. (2023). We treat the problem as a **jump point detection** problem under nonparametric smoothing framework. Specifically we use left and right kernels to estimate the underlying smooth functions on the left and right of the potential jump points and use them to construct a test statistics that is sensitive to jumps.

Define the one sided kernel $K_{h^*,st}^* = 2h^{*-1}k(\frac{s-t}{Th^*})$, and correspondingly

$$\widehat{\boldsymbol{M}}_{R,t}^{+} = \sum_{s=t}^{t+\lfloor h^*T \rfloor} K_{h^*,st}^* \mathbf{Y}_s \mathbf{Y}_s^{\top}, \text{ and } \widehat{\boldsymbol{M}}_{R,t}^{-} = \sum_{s=t-\lfloor h^*T \rfloor}^{s=t-1} K_{h^*,st}^* \mathbf{Y}_s \mathbf{Y}_s^{\top}.$$

Let $\widehat{R}_{i,t}^+/\sqrt{p}$ and $\widehat{R}_{i,t}^-/\sqrt{p}$ be the *i*-th eigenvectors of $\widehat{M}_{R,t}^+$ and $\widehat{M}_{R,t}^-$, respectively. Our objective is to detect if there is a switching in the eigenvalues/eigenvectors between t-1 and t. Assuming switching only occurs between two adjacent eigenvalues, we take a sequential approach and define the test statistics, for $i=1,\ldots r-1$:

$$T_{i,t} = \left\| \widehat{\mathbf{R}}_{i,t}^{-} \widehat{\mathbf{R}}_{i,t}^{-\top} - \widehat{\mathbf{R}}_{i,t}^{+} \widehat{\mathbf{R}}_{i,t}^{+\top} \right\|^{2} + \left\| \widehat{\mathbf{R}}_{i+1,t}^{-} \widehat{\mathbf{R}}_{i+1,t}^{-\top} - \widehat{\mathbf{R}}_{i+1,t}^{+} \widehat{\mathbf{R}}_{i+1,t}^{+\top} \right\|^{2} - \left\| \widehat{\mathbf{R}}_{i:i+1,t}^{-} \widehat{\mathbf{R}}_{i:i+1,t}^{-\top} - \widehat{\mathbf{R}}_{i:i+1,t}^{+} \widehat{\mathbf{R}}_{i:i+1,t}^{+\top} \right\|^{2},$$
(5.1)

where $\widehat{\mathbf{R}}_{j:j+1,t} \in \mathbb{R}^{p \times 2}$ is a submatrix of $\widehat{\mathbf{R}}_t$ composed of its i-th and i + 1-th columns only. The use of $\mathbf{R}_{j,t}^{\pm}\mathbf{R}_{j,t}^{\pm}^{\top}$ instead of $\mathbf{R}_{j,t}^{\pm}$ serves the purpose of circumventing the sign switching and also enhancing the differences. The intuition of the last term in $T_{i,t}$ is

$$\left\|\mathbf{R}_{t}^{-}\mathbf{R}_{t}^{-\top}-\mathbf{R}_{t}^{+}\mathbf{R}_{t}^{+\top}\right\|^{2}=\left\|\sum_{j=1}^{r}\left(\mathbf{R}_{j,t}^{-}\mathbf{R}_{j,t}^{-\top}-\mathbf{R}_{j,t}^{+}\mathbf{R}_{j,t}^{+\top}\right)\right\|^{2}\leq\sum_{j=1}^{r}\left\|\mathbf{R}_{j,t}^{-}\mathbf{R}_{j,t}^{-\top}-\mathbf{R}_{j,t}^{+}\mathbf{R}_{j,t}^{+\top}\right\|^{2},$$

with the equality reached when there is no switching. The last term in (5.1) actually

serves two purposes. First, the last term is not sensitive on whether there is indeed a switch between i-th and i+1-th columns, but it tends to capture the time varying differences of the linear space spanned the two columns between the left and right side of t, hence partially compensating the difference in the first two terms due to the same effect. It makes (5.1) more sensitive to switching but less sensitive to the time varying nature of the functions around t. Second, it makes the test statistic less sensitive to the situation when the switching is actually between the i-1-th and i-th eigenvalues/eigenvectors. In this case, $\|\widehat{\mathbf{R}}_{i,t}^{-}\widehat{\mathbf{R}}_{i,t}^{-}\widehat{\mathbf{R}}_{i,t}^{+}\|^2$ will be large, but it may be largely cancelled by the last term in (5.1). It is interesting to note that, using the orthonormality property of the matrices, $T_{j,t}$ can be rewritten as $T_{i,t} = 2\left[\left(\widehat{\mathbf{R}}_{i,t}^{-}\widehat{\mathbf{R}}_{i+1,t}^{+}\right)^2 + \left(\widehat{\mathbf{R}}_{i+1,t}^{-}\widehat{\mathbf{R}}_{i,t}^{+}\right)^2\right]$, showing that $T_{i,t}$ actually measures the distance (angle) after swapping the i-th and (i+1)-th columns.

When there is a switching between $R_{i,t}$ and $R_{i+1,t}$, it has impact on their sample versions in a much wider time range, due to the smoothing operation. In addition, when two eigenvalues are the same, their corresponding eigenvectors are not uniquely identified. Hence the estimators $\widehat{R}_{i,t}$ would be very inaccurate when t is close to the actual switching point. Hence corrections and re-estimation are needed in a region around the switching point. We determine the significant regions (coalescing regions) that are impacted by the switching using the continuous subsets of $T_{i,t}$ that exceeds a certain thresholds thr_i for $i \in [r-1]$, similar to an outlier detection procedure. We define these coalescing regions as $P_m = (a_m, b_m] \subset [T]$ and its associated switching column κ_m such that for all $t \in P_m$, $T_{\kappa_m,t} > thr_{k_m}$ and $T_{\kappa_m,a_{m-1}} \leq thr_{\kappa_m}$ and $T_{\kappa_m,b_{m+1}} \leq thr_{\kappa_m}$. Here, $m=1,2,\cdots,M$ where M is the total number of regions. We use the thresholds $thr_i = \max\{1.5IQR_i,1.5IQR\}$, where IQR_j is the interquartile range of $T_{j,t}$ and IQR is the global interquartile range employing the data set $\{T_{j,t}: j \in [r-1]; t = [T]\}$. It is possible to develop a bootstrap type of procedures similar to that of Motta et al. (2023) though it is much more involved due to our complex setting. Empirically we found the ad hoc

procedure using IQR is quite reasonable.

Next, using the coalescing regions P_m and its associated switching order κ_m , we follow the approach of Motta et al. (2023) to perform a switching and smoothing operation. Specifically, starting from i=1, we identify all regions P_m 's such that $\kappa_m=i$, perform the switching of $\widehat{R}_{i,t}$ outside these intervals, and apply smoothing splines to interpolate the functions inside the P_m regions. The interpolated function are further adjusted using Gram-Schmidt orthogonalization to obtain the final estimates (Motta et al. 2023).

Note that the smoothed loading functions are still subject to **global** rotation ambiguity. To obtained more interpretable representation, we perform a global varimax (Kaiser 1958) rotation on the $(dT) \times r$ matrix by stacking the smoothed \widetilde{R}_t , $t \in [T]$ matrices.

6 Simulation

In this section, we study the numerical performance of the proposed time-varying matrix-valued approach. Throughout, the matrix observations \mathbf{Y}_t 's are generated according to model (2.1). The dimension of the latent factor matrix \mathbf{F}_t is fixed at $k \times r = 2 \times 2$. The values of p, q, and T vary in different settings. We simulate $\text{vec}(\mathbf{F}_t)$ from the following Vector Auto-Regressive model of order one (VAR(1) model):

$$\operatorname{vec}(\boldsymbol{F}_t) = \boldsymbol{\Phi} \cdot \operatorname{vec}(\boldsymbol{F}_{t-1}) + \boldsymbol{\varepsilon}_t, \tag{6.1}$$

where the AR coefficient matrix $\mathbf{\Phi} = 0.1 \cdot \mathbf{I}_4$ and $\text{Var}[(\varepsilon_t)] = 0.99 \cdot \mathbf{I}_4$. Thus, $\text{Var}[\text{VEC}(\mathbf{F}_t)] = \mathbf{I}_4$. We simulate noise \mathbf{E}_t also from VAR(1),

$$VEC(\boldsymbol{E}_t) = \boldsymbol{\Gamma} \cdot VEC(\boldsymbol{E}_{t-1}) + \boldsymbol{u}_t, \tag{6.2}$$

where $\Gamma = \psi \cdot I_{pq}$ and $\text{Var}[u_t] = (1 - \psi^2)I_{pq}$. Thus, $\text{Var}[\text{Vec}(E_t)] = I_{pq}$. We choose $\psi = 0.1$ and then increase to $\psi = 0.5$ to examine how temporal dependence may affect our results. The true loading matrices R_t and C_t are generated as:

- DGP1: The first column of \mathbf{R}_t : $\mathbf{R}_{t,.1} = \mathbf{R}_{.1}$, the second column of \mathbf{R}_t : $\mathbf{R}_{t,.2} = \mathbf{R}_{.2} + \mathbf{G}_t$; the first column of \mathbf{C}_t : $\mathbf{C}_{t,.1} = \mathbf{C}_{.1}$, the second column of \mathbf{C}_t : $\mathbf{C}_{t,.2} = \mathbf{C}_{.2} + \mathbf{H}_t$. The time-invariant loading matrices $\mathbf{R}_{.j}$ and $\mathbf{C}_{.j}$ are independently sampled from $\mathcal{U}(-1,1)$ for j=1,2, $\mathbf{G}_t = \mathbf{G}\left(\frac{t}{T}\right) = 2\left(\frac{t}{T}\right) + \exp\left[-16\left(\frac{t}{T} 0.5\right)^2\right] 1$ and $\mathbf{H}_t = 0.2 \exp\left[-0.7 + 3.5\left(\frac{t}{T}\right)\right]$.
- DGP2: The first column of \mathbf{R}_t : $\mathbf{R}_{t,.1} = \mathbf{R}_{.1} + \mathbf{G}_t$, the second column of \mathbf{R}_t : $R_{t,i2} = \mathbf{F} (10t/T; 2; 5i/p + 2)$, i = 1,...,p; the first column of \mathbf{C}_t : $\mathbf{C}_{t,.1} = \mathbf{C}_{.1} + \mathbf{H}_t$, the second column of \mathbf{C}_t : $C_{t,j2} = \mathbf{F} (10t/T; 2; 5j/q + 2)$, j = 1,...,q. The time-invariant loading matrices $\mathbf{R}_{.j}$ and $\mathbf{C}_{.j}$ are independently sampled from $\mathcal{N} (0,1)$ for j = 1,2, b = 2, and $\mathbf{F} (\tau; \kappa; \gamma) = \{1 + \exp[-\kappa(\tau \gamma)]\}^{-1}$.

We first study the performance of our proposed approach on estimating the time-varying loading matrices R_t and C_t . We consider three pairs of (p,q) combinations: (20,20), (100,20), or (100,100). The sample size T is selected as 100, 200, and 400, and simulation results are based on 100 repetitions. For nonparametric estimation, we use the Epanechnikov kernel and Silverman's rule-of-thumb (RoT) bandwidth $h_R = c(qT)^{-\frac{1}{5}}$ and $h_C = c(pT)^{-\frac{1}{5}}$, where $c = \frac{2.345}{\sqrt{12}}$, for estimating R_t and C_t respectively. This simple RoT bandwidth attains the optimal rate for local smoothing as shown in Theorem 2. We also try the RoT bandwidth with different scaling parameters, and the simulation shows that our estimation results are not very sensitive to the bandwidth selection. To evaluate the accuracy of our estimator, we use the average column space distance

$$\overline{\mathcal{D}}_{\left(\widehat{A},A\right)} = \frac{1}{T} \sum_{t=1}^{T} \left\| \widehat{A}_{t} \left(\widehat{A}_{t}^{\top} \widehat{A}_{t} \right)^{-1} \widehat{A}_{t}^{\top} - A_{t} \left(A_{t}^{\top} A_{t} \right)^{-1} A_{t}^{\top} \right\|. \tag{6.3}$$

Table 1 reports the mean and standard deviation of the average column space distance. For both DGPs considered, $\overline{\mathcal{D}}_{(\widehat{R},R)}$ and $\overline{\mathcal{D}}_{(\widehat{C},C)}$ decrease with the increase of p,q and T. The estimation results are robust to the the temporal dependence of noise E_t .

With the same simulated data, we compare the proposed time-varying matrix-valued approach and the time-varying vector-valued approach in Su & Wang (2017) through the estimation accuracy of the total loading matrix $\Xi_t = C_t \otimes R_t$. In what follows, the sub-

Table 1: Estimation of loading matrices

		T = 100		T = 200		T = 400		
DGP	(p,q)	$\overline{\mathcal{D}}_{\left(\widehat{R},R ight)}$	$\overline{\mathcal{D}}_{\left(\widehat{C},C ight)}$	$\overline{\mathcal{D}}_{\left(\widehat{R},R ight)}$	$\overline{\mathcal{D}}_{\left(\widehat{C},C ight)}$	$\overline{\mathcal{D}}_{\left(\widehat{R},R ight)}$	$\overline{\mathcal{D}}_{\left(\widehat{C},C ight)}$	
		$\psi = 0.1$						
	(20, 20)	1.09(0.21)	0.98(0.15)	0.98(0.12)	0.55(0.07)	0.58(0.08)	0.50(0.04)	
	(100, 20)	0.98(0.12)	0.55(0.07)	0.73(0.08)	0.40(0.05)	0.41(0.04)	0.35(0.02)	
1	(100, 100)	0.58(0.08)	0.50(0.04)	0.52(0.04)	0.29(0.04)	0.29(0.02)	0.26(0.02)	
		$\psi=0.5$						
	(20, 20)	1.14(0.22)	1.04(0.17)	1.01(0.12)	0.58(0.08)	0.59(0.08)	0.52(0.04)	
	(100, 20)	1.01(0.12)	0.58(0.08)	0.75(0.08)	0.42(0.06)	0.42(0.04)	0.37(0.02)	
	(100, 100)	0.59(0.08)	0.52(0.04)	0.54(0.05)	0.30(0.04)	0.30(0.03)	0.27(0.02)	
	$\psi = 0.1$							
2	(20, 20)	2.46(0.20)	2.15(0.20)	2.27(0.18)	1.94(0.16)	2.11(0.14)	1.79(0.16)	
	(100, 20)	2.35(0.20)	1.72(0.17)	2.14(0.14)	1.57(0.13)	1.95(0.11)	1.42(0.10)	
	(100, 100)	1.91(0.20)	1.47(0.13)	1.73(0.13)	1.29(0.08)	1.61(0.10)	1.15(0.09)	
	$\psi = 0.5$							
	(20, 20)	2.48(0.20)	2.24(0.20)	2.30(0.18)	2.05(0.15)	2.13(0.14)	1.88(0.16)	
	(100, 20)	2.38(0.19)	1.80(0.17)	2.17(0.14)	1.64(0.13)	1.97(0.11)	1.50(0.11)	
	(100,100)	1.93(0.19)	1.55(0.13)	1.75(0.13)	1.37(0.08)	1.62(0.10)	1.23(0.09)	

Note: Mean and standard deviation in parentheses of $\overline{\mathcal{D}}_{(\widehat{R},R)}$ and $\overline{\mathcal{D}}_{(\widehat{C},C)}$ from 100 iterations.

scripts mat and vec denote our approach and Su & Wang (2017)'s method, respectively. The loading space $\widehat{\Xi}_{mat,t}$ is computed as $\widehat{\Xi}_{mat,t} = \widehat{C}_t \otimes \widehat{R}_t$, where \widehat{C}_t and \widehat{R}_t are our non-parametric estimators. For the vector-valued approach, we apply Su & Wang (2017)'s method to the vectorized observations $\text{VEC}(Y_t)$, $t=1,2,\cdots,T$ as in (4.2) to obtain $\widehat{\Xi}_{vec,t}$. The results for the estimation accuracy of Ξ_t measured by $\overline{\mathcal{D}}_{mat}$ and $\overline{\mathcal{D}}_{vec}$ are reported in Table 2, which shows that the matrix approach efficiently improves the estimation accuracy over the vector-valued approach. We only consider three pairs of small (p,q) combinations: (10,10), (20,10), or (20,20) because the estimation of time-varying vectorized factor models is extremely time-consuming. Nevertheless we expect the same pattern of estimation comparison would continue to hold for large p and q.

We next evaluate the performance of the matrix-valued approach on estimating the number of factors k and r. Table 3 shows the relative frequencies of estimated rank pairs over 100 iterations. The three pairs (2,1), (1,2) and (2,2) have high appearances in all of the combinations of (p,q,T). The row for the true rank pair (2,2) is highlighted. It shows that the relative frequency of correctly estimating the true rank pair increases with the increase of (p,q,T) and approaches unity.

Table 2: Comparison of estimation accuracy

		T =	T = 200			
DGP	(p,q)	$\overline{\mathcal{D}}_{max}$	$\overline{\mathcal{D}}_{vec}$	$\overline{\mathcal{D}}_{max}$	$\overline{\mathcal{D}}_{vec}$	
			$\psi = 0.1$			
	(10, 10)	0.23(0.06)	0.78(0.10)	0.13(0.02)	0.62(0.10)	
	(20, 10)	0.17(0.04)	0.65(0.11)	0.15(0.02)	0.75(0.07)	
1	(20, 20)	0.19(0.03)	0.76(0.09)	0.11(0.02)	0.60(0.08)	
1			$\psi = 0.5$			
	(10, 10)	0.25(0.07)	0.90(0.09)	0.14(0.03)	0.79(0.12)	
	(20, 10)	0.18(0.05)	0.79(0.13)	0.16(0.03)	0.91(0.06)	
	(20, 20)	0.20(0.04)	0.91(0.06)	0.12(0.02)	0.78(0.12)	
			$\psi = 0.1$			
	(10, 10)	0.34(0.03)	0.66(0.11)	0.31(0.02)	0.58(0.12)	
	(20, 10)	0.32(0.02)	0.64(0.10)	0.29(0.01)	0.54(0.08)	
2	(20, 20)	0.29(0.02)	0.61(0.09)	0.27(0.02)	0.51(0.07)	
2			$\psi = 0.5$			
	(10, 10)	0.35(0.03)	0.70(0.12)	0.32(0.03)	0.61(0.13)	
	(20, 10)	0.33(0.02)	0.69(0.11)	0.29(0.02)	0.58(0.09)	
	(20, 20)	0.30(0.02)	0.65(0.10)	0.27(0.02)	0.55(0.09)	

Note: Mean and standard deviation in parentheses of $\overline{\mathcal{D}}_{max}$ and $\overline{\mathcal{D}}_{vec}$ from 100 iterations.

Table 3: Estimation of latent dimensions

		p, q = 20, 20			p, q = 100, 20			p, q = 100, 100		
DGP	$(\widehat{k},\widehat{r})$	T = 100	T = 200	T = 400	T = 100	T = 200	T = 400	T = 100	T = 200	T = 400
						$\psi = 0.1$				
	(2,1)	0.13	0.06	0.04	0.02	0.03	0.01	0.00	0.00	0.00
	(1, 2)	0.05	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00
	(2,2)	0.80	0.90	0.93	0.98	0.97	0.99	1.00	1.00	1.00
1	other	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1		$\psi = 0.5$								
	(2,1)	0.12	0.08	0.03	0.02	0.02	0.00	0.00	0.00	0.00
	(1,2)	0.09	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.00
	(2,2)	0.77	0.87	0.94	0.98	0.98	1.00	1.00	1.00	1.00
	other	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		$\psi = 0.1$								
	(2,1)	0.21	0.16	0.14	0.21	0.11	0.06	0.00	0.00	0.00
	(1, 2)	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(2,2)	0.74	0.81	0.86	0.79	0.89	0.94	1.00	1.00	1.00
2	other	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2		$\psi = 0.5$								
	(2,1)	0.22	0.16	0.15	0.20	0.12	0.07	0.00	0.00	0.00
	(1,2)	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(2,2)	0.73	0.80	0.85	0.80	0.88	0.93	1.00	1.00	1.00
	other	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: Relative frequency of estimated rank pair $(\widehat{k},\widehat{r})$ from 100 iterations. The row with the true rank pair (2,2) is highlighted.

We evaluated the smoothing operation's empirical performance on the load matrices as described in Section 5. To closely mirror the switching behavior observed in the real data example in Section 7, we generated data for controlled experiments:

$$\mathbf{X}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{C}_t \text{ and } \mathbf{Y}_t = \alpha \mathbf{X}_t + \mathbf{E}_t, \tag{6.4}$$

where F_t and E_t follows the simulation setting above in (6.1) and (6.2), with a 4×4 F_t factor process. The parameter α^2 denotes the signal-to-noise ratio (SNR) (i.e. $(1 + \alpha^2)^{-1}$ is the relative size of the noise process with respect to \mathbf{Y}_t), after \mathbf{X}_t is normalized so $\mathrm{var}(\mathrm{vec}(\mathbf{X}_t)) = \mathrm{var}(\mathrm{vec}(\mathbf{X}_t)) = \mathbf{I}_{pq}$.

The time-varying matrices \mathbf{R}_t and \mathbf{C}_t are constructed similarly to those in Motta et al. (2023). Specifically, we first generate the time-varying matrix \mathbf{Q}_t whose j-th column is

$$\mathbf{Q}_{\cdot j,t} = \begin{bmatrix} \sin\left((j-1)\pi w \frac{t}{T} - \phi\right)\sqrt{2/3} \\ \cos\left((j-1)\pi w \frac{t}{T} - \phi\right)\sqrt{2/3} \\ \exp\left((j-1)\pi w \frac{t}{T} - \phi\right)\sqrt{2/3} \\ \exp\left(-(j-1)\pi w \frac{t}{T} - \phi\right)\sqrt{2/3}. \end{bmatrix}$$

$$(6.5)$$

where w = 0.75 and $\phi = 0.2$. Let $\overline{\mathbf{R}}_t = \mathbf{U}_R \mathbf{Q}_{R,t} \left(\mathbf{Q}_{R,t}^{\top} \mathbf{Q}_{R,t} \right)^{-1/2}$ and $\overline{\mathbf{C}}_t = \mathbf{U}_C \mathbf{Q}_{C,t} \left(\mathbf{Q}_{C,t}^{\top} \mathbf{Q}_{R,t} \right)^{-1/2}$, where $\mathbf{U}_R \in \mathbb{R}^{p \times k}$ and $\mathbf{U}_C \in \mathbb{R}^{q \times r}$ are orthonormal matrices randomly generated. Then the matrices \mathbf{R}_t and \mathbf{C}_t are generated as $\mathbf{R}_t = \overline{\mathbf{R}}_t \boldsymbol{\Lambda}_{R,t}^{1/2}$ and $\mathbf{C}_t = \overline{\mathbf{C}}_t \boldsymbol{\Lambda}_{C,t}^{1/2}$, ensuring that $\overline{\mathbf{R}}_t^{\top} \overline{\mathbf{R}}_t = \mathbf{I}_k$, $\overline{\mathbf{C}}_t^{\top} \overline{\mathbf{C}}_t = \mathbf{I}_r$. This normalization is to ensure that α^2 in (6.4) is the true signal-to-noise ratio.

We set the eigenvalues function matrices to explore one and two coalescing points in separate analyzes. Specifically, let $\Lambda_{R,t}$ be a 4×4 diagonal matrix, with diagonal elements being $3.5-1.5cos(\pi t/T)$, $3-1.5cos(\pi t/T)$, $0.5(1+cos(\pi t/T))$, and $0.5(1-cos(\pi t/T))$. Similarly, let $\Lambda_{C,t}$ be a 4×4 diagonal matrix, with diagonal elements being $4-cos(\pi t/T)$, $3-1.5cos(\pi t/T)$, $1.25(1+cos(\pi t/T))$, and $0.5+0.5cos(1.5\pi t/T)$. The functions are shown in Figure 3. We show the results of estimating R_t and C_t , using $\mathbf{X}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{C}_t^{\top}$ and $\mathbf{X}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{C}_t^{\top}$, to have $E[\mathbf{X}_t^{\top} \mathbf{X}_t] = k\gamma(0)\mathbf{R}_t \Lambda_{R,t} \mathbf{R}_t^{\top}$. and $E[\mathbf{X}_t \mathbf{X}_t^{\top}] = r\gamma(0)\mathbf{C}_t \Lambda_{C,t} \mathbf{C}_t^{\top}$, respectively. Here, $\gamma(0)$ is the variance of an AR process with autoregressive parameter 0.1 and noise variance 1. In this context, we designate these conditions as Scenario 0. In Scenario 1 we add a deterministic trend to F_t , using $\mathbf{F}_t^{(1)} = \mathbf{F}_t^{(0)} + \frac{\sqrt{k}t}{T}\mathbf{M}$ where \mathbf{M} is a randomly generated orthonormal matrix. In Scenario 2 we simply use $\mathbf{F}_t^{(2)} = \frac{t}{T}\mathbf{M}$ as the factor process. We are interested in these two scenarios because the factor process in

our trade volume application is clearly nonstationary, while the method in Motta et al. (2023) critically rely on the stationary processes. In both scenarios the factor processes are scaled to achieve the desired Signal-to-Noise Ratio.

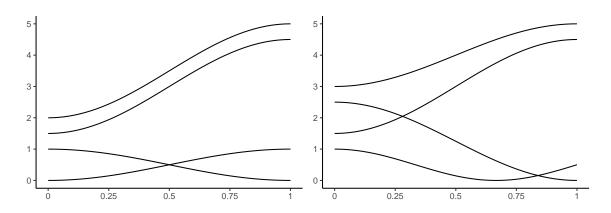


Figure 3: Diagonal entries of the matrices $\Lambda_{R,\frac{t}{T}}$ (Left panel) and $\Lambda_{C,\frac{t}{T}}$ (Right panel)

We illustrate the procedure by varying the dimensions $(p,q) \in \{(10,10),(10,20),(20,20)\}$, the sample size $T \in \{200,400,1000\}$, the error size $(1+\alpha^2)^{-1} \in \{5\%,10\%\}$, and the AR parameter $\psi \in \{0.1,0.5\}$. For each scenario, we obtain 50 replicates and classify the identified regions concerning coalescing points, namely 0.5 for the eigenvalues of $M_{R,t}$ and $\{0.28,0.83\}$ for that of $M_{C,t}$. Figure 4 shows the number of times the true coalescing points are in the identified coalescing regions. If an identified coalescing region does not contain any of the true coalescing points, it is classified as "False Positive". Within each box, the first two bars are for the 'true positive' and 'false positive' identifications of the single coalescing point in $M_{R,t}$. The third and fourth bars are for the 'true positive' identifications of the two coalescing point in $M_{C,t}$ and the last bar is for the 'false positive' of $M_{C,t}$.

In addition, we also obtained the results using that in Motta et al. (2023), where local stationarity is assumed. Here we use the cross-moment $M_{R,t}$ and $M_{C,t}$, instead of covariance matrix. These matrices are then utilized to generate bootstrap samples through the expression $\widehat{\mathbf{R}}t\widehat{\boldsymbol{\Lambda}}^{1/2}R$, $t\mathbf{Z}_t$, where \mathbf{Z}_t denotes a Gaussian white noise process

with zero mean and variance **I**. The purpose of developing these bootstrap samples is to delineate the region where eigenvalues coalesce, with the 5th percentile as the bounds to identify the coalescing regions.

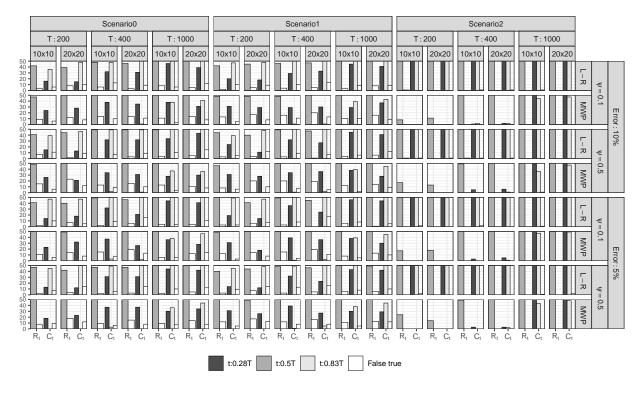


Figure 4: Number of times identified coalescing points on each scenario. Varying the dimensions $(p,q) \in (10,10), (10,20), (20,20)$, the sample size $T \in 200,400,1000$, the size of the noise $(1+\alpha^2)^{-1} \in \{5\%,10\%\}$, and the AR parameter $\psi \in \{0.1,0.5\}$.

Figure 4 shows that both approaches yield an increase in correct identifications and a decrease in false positives as the sample size increases. Comparing the Left-Right approach (labelled as L-R) and that of Motta et al. (2023) (labelled as MVP) (e.g. firts row and second row), it is seen that in relative small sample sizes the left-right detection procedure in Section 5 performs better than that of Motta et al. (2023) for C_t where there are two switchings. Both procedures are similar when there is only only crossing for R_t . The procedure of Motta et al. (2023) has difficulty in Scenario 2 where a deterministic trend is present while the left-right detection procedure seems to benefit from the presence of the trend.

7 Empirical Application

Understanding international trade flow patterns and their evolution is crucial for policy-making, economic forecasting, and firm optimization. International trade data, forming a dynamic sequence of matrix variates, closely mirrors network structures, offering insights into interaction patterns, evolutionary significance, and natural actor groupings within the network. Theoretical and empirical research on international trade networks has advanced significantly (Chaney 2016), with the Global Vector Autoregressive (GVAR) approach being particularly notable for analyzing trade network interactions (Chudik & Pesaran 2016). For instance, Bussiére et al. (2012) and Greenwood-Nimmo et al. (2012) utilized GVAR models to study demand shocks' effects on global imbalances and forecast trade imbalances across 33 countries, respectively. However, GVAR's reliance on a pre-determined weighting matrix to address dimensionality issues is a limitation. This study introduces a time-varying matrix factor model for international trade data, estimating time-varying loading matrices and latent factors concurrently.

7.1 Data and sample

We use monthly multilateral import and export volumes of commodity goods among 24 countries and regions from January 1982 to December 2018. The data come from the International Monetary Fund Direction of Trade Statistics (IMF-DOTS). The countries and regions included in the study are Australia(AU), Canada(CA), China Mainland(CN), Denmark(DK), Finland(FI), France(FR), Germany(DE), Hong Kong(HK), Indonesia(ID), Ireland(IE), Italy(IT), Japan(JP), Korea(KR), Malaysia(MY), Mexico(MX), Netherlands(NL), New Zealand(NZ), Singapore(SG), Spain(ES), Sweden(SE), Taiwan(TW), Thailand(TH), United Kingdom(GB), and the United States(US).

In our study, we utilize the time-varying matrix factor model (2.1) to examine in-

ternational trade flow data, where Y_t is depicted as a 24×24 matrix, with rows and columns representing exporting and importing countries, respectively. We rely on import data valued in U.S. dollars for their presumed accuracy over export figures (Linnemann 1967). It's noteworthy that Taiwan's trade data, absent in the IMF-DOTS, is inferred from its trading partners' export information. To mitigate the effects of significant one-off transactions and trade anomalies, we calculate trade flows using three-month averages instead of monthly snapshots.

7.2 Estimation Results

As described in Section 2, the $k \times r$ latent factors in model (2.1) can be interpreted as the transport volumes of a condensed trading network among k export hubs and r import hubs in our application. We use the generalized ratio-based method in (3.5) as well as the screen plot to select (4,4) as the number of dimensions of the latent hubs. For better visualization and interpretation, all estimated R_t and C_t are normalized so the sum of absolute values of each column is 1.

Figure 5 shows the heatmap of the initial estimates of R_t , corresponding to the descending order of the eigenvalues. The first two eigenvectors (loading vectors) show quite smooth changes. However, the third and fourth eigenvectors show abrupt changes. Figure 6 shows the test statistics $T_{i,t}$ in (5.1), where the large values indicate possible switching between i-th and (i + 1)-th eigenvalues. It confirms no switching between the first two columns in R_t , but reveals switching between columns 3 and 4 around Year 2010, and two switches between columns 4 and 5 around Years 1993 and 2004. Note that column 5 (corresponding to the fifth largest eigenvalue) was deemed insignificant by our model selection procedure. It appears that the significance of the 5th eigenvalue of $M_{R,t}$ gradually increases, while the significance of the 4-th eigenvalue of $M_{R,t}$ gradually decreases to insignificance, and they switch during the period.

Due to the existence of switching, caution is required in labeling. Given the significant increase in import-export volume due to globalization, and thus increasing signal strengths, we adopt labels based on the later periods. That is, we name factor k as the one associated with the largest k-th eigenvalue of $M_{R,T}$ where T is the end of sample period. Figure 7 shows the estimated six largest eigenvalues over time, suggesting potential switching around 2010 when the third and fourth eigenvalues are quite close. In the beginning when the signal is relatively weak, there are a number of potential switchings.

Using this labeling system, Figure 6 shows a switch between Factors 3 and 4 around Year 2010 (the third subfigure), implying that in the year **after** 2010, Factor 3 corresponds to the 3rd largest eigenvalue of $M_{R,t}$ and its corresponding eigenvectors (as factor loading). Just before 2010, Factor 3 corresponds to the 4th largest eigenvalue, and it switches again around 2004 (the fourth subfigure), meaning that just before 2004, Factor 3 corresponds to the 5th largest eigenvalue. The top figure in Figure 8 shows the identified coalescing region for Factors 3 in these two time periods, along with the initial estimates of the corresponding loading vectors **after** the two switches outside the coalescing regions. The bottom figure shows the results after interpolating the loadings within the coalescing regions and smoothing outside the regions, following the procedure used in Motta et al. (2023). Figure 9 shows the same for Factors 4, indicating a switch with Factor 3 around 2010. Both figures show that the smoothing operation in Section 5 works well.

Since the signal is stronger in the later periods, we apply a global varimax rotation only using the smoothed eigenfunctions between 2003 to 2018. The rotation is then applied to the entire sample period. The final estimates of the four loading vector functions \mathbf{R}_t are shown in Figure 10. Similar estimation approaches result in the final estimate of the loading vector functions of \mathbf{C}_t , shown in Figure 11. Note that after varimax

rotation, the strength of the factor may not be in descending order anymore.

Figure 5: Trading volume example: Initial estimate of R_t .

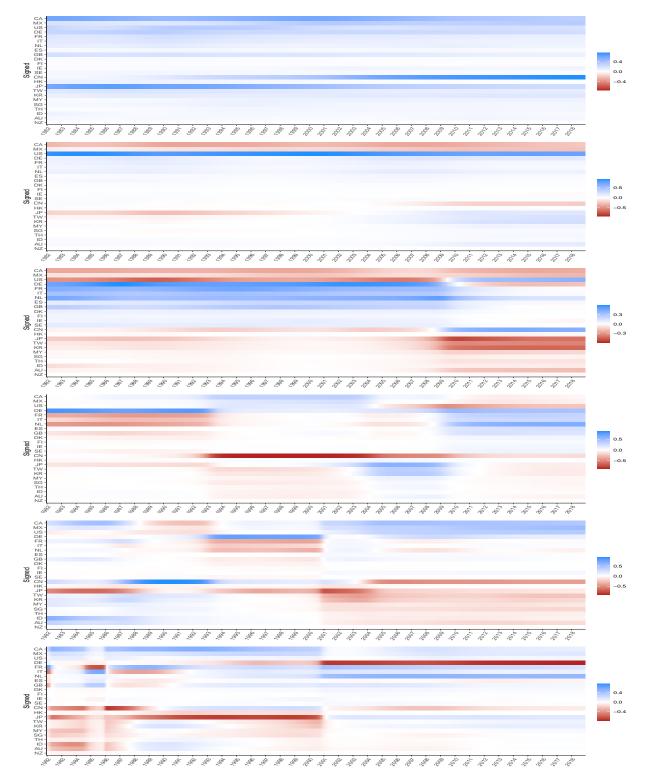


Figure 6: Trading Volume Example: Test statistics $T_{i,t}$ of R_t . Horizontal line is the 0.95 upper IQR line.

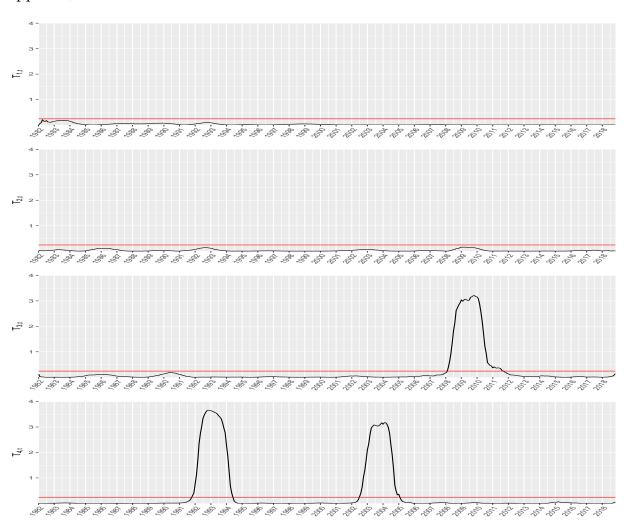


Figure 7: Estimated eigenvalue functions

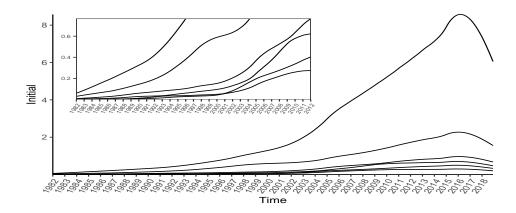


Figure 8: Trading Volume Example: (top) Identified coalesing regions R_t related to Factors with the original estimated after switching outside the coalescing region. (bottom) Result with interpolated loadings of Factor 3 within the coalescing region.

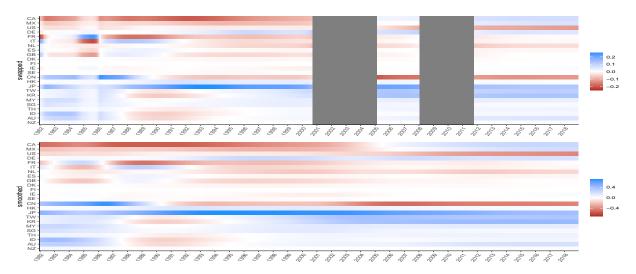
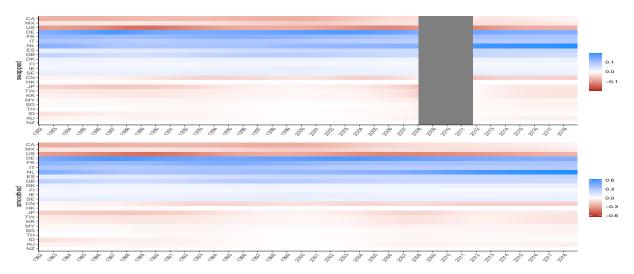


Figure 9: Trading Volume Example: (top) Identified coalesing region R_t between Factors 4 and 5, and switching outside the coalescing region. (bottom) Result with interpolated loadings within the coalescing region.



7.3 Interpretation of results

Here we adopt the interpretation of the model described in Section 2. Figures 10 and 11 provide insights into the dynamics of the globalization. The latent Export Hub 1 (top figure in Figure 10) can be viewed as a China-dominated export hub, as China has the maximum loading overall. Before 2001, Japan had a higher loading on this hub

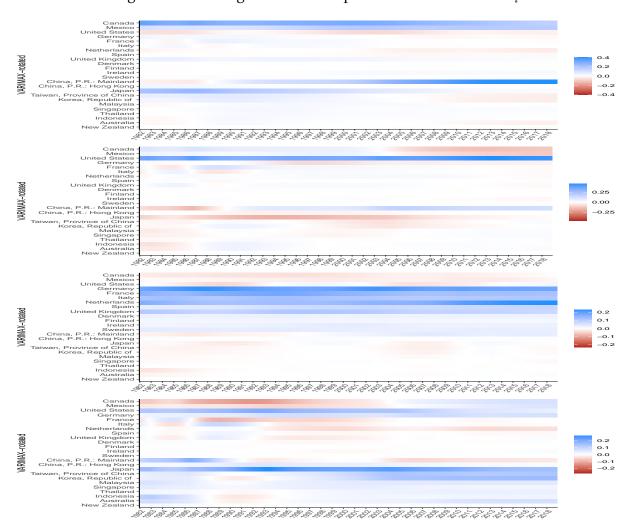


Figure 10: Trading volume example: Final estimates of R_t .

than China, indicating a shift in exporting centrality within the Asian economy. China's loading on this hub experiences significant growth, particularly following its accession to the WTO in 2001. Despite this, Japan, Canada, Hong Kong and Taiwan also remain significant contributors to this export hub.

The latent Export Hub 2 can be viewed as a USA-dominated export hub as evidenced by the substantially larger loading of the USA compared to other countries. While China also participates in this hub, the United States maintains dominance throughout the sample period. The latent Export Hub 3 can be interpreted as a hub dominated by European countries such as Germany, France, Italy, Netherlands, and Spain. Last, the

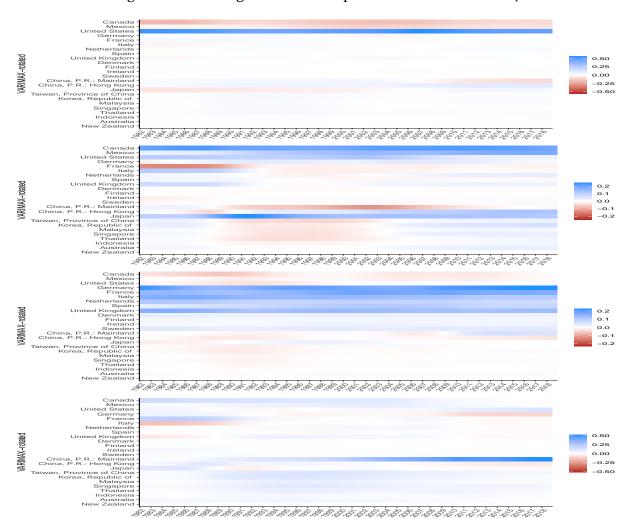


Figure 11: Trading volume example: Final estimates of C_t .

latent Export Hub 4 was mainly used by APEC countries, in which Japan dominated throughout the period, while Korea, Taiwan and Hong Kong emerge as active participants post-2000. United States also participated in this hub, albeit to a lesser extent.

On the import side, United States dominated Import Hub 1 (top figure in Figure 11). Canada, Japan and Mexico used Import Hub 2. The European countries used Import Hub 3, and China dominated in Import Hub 4.

8 Conclusion

Modeling high-dimensional matrix-valued time series has attracted increasing attention recently. We contribute to this literature by proposing a new time-varying matrix factor model that may be more applicable in real applications, establishing the asymptotic properties of a local PCA estimator for the model, and proposing a generalized eigenvalue-ratio estimator for latent dimensions. Our proposed estimators are intuitively appealing and straightforward to compute. They fully preserve the matrix structure and explore the local variation of the loading matrices. Moreover, our results are obtained under very general conditions that allow for weak correlations across time, rows, and columns. A smoothing operation is used to obtain a smooth representation of the time-varying loading matrices in order to facilitate easy interpretation of the model. Simulations have shown that the nonparametric estimator for factor loadings and the generalized eigenvalue ratio-based estimator perform well in finite samples. The empirical application concerning the international trade flow reveals some interesting time-varying patterns.

References

- Ahn, S. & Horenstein, A. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica* **81**(5), 1203–1227.
- Bussiére, M., Chudik, A. & Sestieri, G. (2012), 'Modelling global trade flows: results from a gvar model', *Globalization and Monetary Policy Institute Working Paper 119*.
- Cai, C., Chen, R. & Xiao, H. (2022), 'KoPA: Automated kronecker product approximation', *Journal of Machine Learning Research* **23**, 1–44.
- Cai, Z. (2007), 'Trending time-varying coefficient time series models with serially correlated errors', *Journal of Econometrics* **136**(1), 163–188.
- Chaney, T. (2016), 'Networks in international trade', *The Oxford Handbook of the Economics of Networks* pp. 754–775.
- Chen, B. & Hong, Y. (2012), 'Testing for smooth structural changes in time series models via nonparametric regression', *Econometrica* **80**(3), 1157–1183.

- Chen, B. & Maung, K. (2023), 'Time-varying forecast combination for high-dimensional data', *Journal of Econometrics* **237**(2), 1–21.
- Chen, E. Y. & Chen, R. (2023), 'Modeling dynamic transport network with matrix factor models: an application to international trade flow', *Journal of Data Science* 21(3), 490–507.
- Chen, E. Y. & Fan, J. (2023), 'Statistical inference for high-dimensional matrix-variate factor models', *Journal of the American Statistical Association* **118**(542), 1038–1055.
- Chen, E. Y., Tsay, R. S. & Chen, R. (2020), 'Constrained factor models for high-dimensional matrix-variate time series', *Journal of the American Statistical Association* **115**(530), 775–793.
- Chen, R., Yang, D. & Zhang, C.-H. (2022), 'Factor Models for High-Dimensional Tensor Time Series', *Journal of the American Statistical Association* **117**, 94–132.
- Chudik, A. & Pesaran, M. H. (2016), 'Theory and practice of gvar modelling', *Journal of Economic Surveys* pp. 165–197.
- Greenwood-Nimmo, M., Nguyen, V. & Shin, Y. (2012), 'Probabilistic forecasting of output, growth, inflation and the balance of trade in a gvar framework', *Journal of Applied Econometrics* pp. 554–573.
- Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. (2009), *The elements of statistical learning: data mining, inference, and prediction*, Vol. 2, Springer.
- Inoue, A., Lu, J. & Rossi, B. (2017), 'Window selection for out-of-sample forecasting with time-varying parameters', *Journal of econometrics* **196**(1), 55–67.
- Kaiser, H. F. (1958), 'The varimax criterion for analytic rotation in factor analysis', *Psychometrika* **23**(3), 187–200.
- Linnemann, H. (1967), 'An econometric study of international trade flows', *The Economic Journal* 77, 366–368.
- Liu, X. & Chen, E. (2019), 'Helping effects against curse of dimensionality in threshold factor models for matrix time series', *arXiv e-prints* p. arXiv:1904.07383.
- Motta, G., Wu, W. B. & Pourahmadi, M. (2023), ' $\sqrt{2}$ -estimation for smooth eigenvectors of matrix-valued functions', *Biometrika* **110**, 1077–1098.
- Phillips, P. & Hansen, B. (1990), 'Statistical inference in instrumental variables regression with i(1) processes', *Rev. Econ. Stud* **57**, 99–125.
- Robinson, P. M. (1989), Nonparametric estimation of time-varying parameters, *in* 'Statistical Analysis and Forecasting of Economic Structural Change', Springer, pp. 253–264.
- Robinson, P. M. (1991), Time-varying nonlinear regression, *in* 'Economic Structural Change Analysis and Forecasting', Springer, pp. 179–190.

- Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**(460), 1167–1179.
- Su, L. & Wang, X. (2017), 'On time-varying factor models: Estimation and testing', *Journal of econometrics* **198**(1), 84–101.
- Su, L. & Wang, X. (2020), 'Corrigendum to "on time-varying factor models: Estimation and testing", working paper.
- Wang, D., Liu, X. & Chen, R. (2019), 'Factor models for matrix-valued high-dimensional time series', *Journal of Econometrics* **208**(1), 231–248.
- Yu, L., He, Y., Kong, X. & Zhang, X. (2022), 'Projected estimation for large-dimensional matrix factor models', *Journal of Econometrics* **229**, 201–217.
- Zhang, T. & Wu, W. B. (2012), 'Inference of time-varying regression models', *The Annals of Statistics* **40**(3), 1376–1402.