Statistica Sinica Preprint No: SS-2025-0212

Title	Determining the Number of Factors in Two-way Factor Model of High-dimensional Matrix-variate Time Series: a White-noise Based Method for Serial Correlation Models
Manuscript ID	SS-2025-0212
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0212
Complete List of Authors	Qiang Xia, W.K. Li and Rubing Liang
Corresponding Authors	Qiang Xia
E-mails	xiaqiang@scau.edu.cn
Notice: Accepted author version.	

Statistica Sinica

Determining The Number of Factors in Two-Way Factor Model of High-Dimensional Matrix-Variate Time Series: A White-Noise based Method for Serial Correlation Models

Qiang Xia[a] Wai Keung Li[b] Rubing Liang [a]

[a] *College of Mathematics and Informatics, South China Agricultural University*

[b] *Department of Mathematics and Information Technology, Education University of Hong Kong*

Abstract: This study proposes a novel two-way factor modeling framework for high-dimensional matrix-variate time series. Motivated by the objective of identifying white noise components, we develop two ratio-based estimators leveraging the element-wise maximum norm and Frobenius norm of sample auto-covariance matrices to determine the dimensions of row and column factor spaces. To reduce the impact of cross-row and cross-column factor strength heterogeneity, the original matrix factor model is reparameterized as a reduced-form model containing only a row loading matrix or a column loading matrix. We then investigate the refined ratio-based methods developed under this reparameterized framework. Under regularity conditions, we establish the theoretical properties of the proposed methods, demonstrating their consistency in estimating the number of factors. Through Monte Carlo simulations and a real data application, we validate the finite-sample performance of the proposed methods and compare them with existing alternatives.

Keywords: Matrix-variate time series, Max-type test, Sum-type test, Two-way factor models.

Corresponding author: liang_ru_bing@scau.edu.cn, South China Agricultural University.

*Author's ORCID ID: Qiang Xia, https://orcid.org/0000-0003-2477-1742

1. Introduction

The analysis of high-dimensional time series data is becoming increasingly critical in the era of big data. For example, time-based data from census prediction and economic studies involve the analysis of high-dimensional time series using large-panel data. Monitoring environmental indices across various locations has further contributed to the growing high-dimensionality of environmental time series data. To address the complex structure of high-dimensional time series, factor models are commonly employed to provide a low-dimensional, parsimonious representation. For further reference, see, for vector-valued data, Bai (2003), Forni et al. (2005), Pan and Yao (2008), Lam et al. (2011), Lam and Yao (2012), and Fan et al. (2022).

In fields such as economics, meteorology, and ecology, high-dimensional matrix-variate time series have gained increasing attention in recent years. Despite their growing prevalence, the development of statistical methodologies for analyzing large-scale matrix-form data remains relatively underexplored. To enhance model parsimony, Wang et al. (2019) proposed matrix-variate factor models that retain the intrinsic matrix structure of the data, facilitating dimensionality reduction along both row and column dimensions. Building on this framework, Chen et al. (2020) extended these models by integrating domain-specific or prior knowledge via linear constraints. In contrast to the white noise idiosyncratic error structures examined in the aforementioned studies, Yu et al. (2022), Chen and Fan (2023), and He et al. (2023) investigated matrix-valued factor models featuring correlated idiosyncratic errors, with weak correlations observed across rows, columns, or individual entries.

Under these two different model structures, the researchers investigated the estimation of factor loading matrices and used the ratios of their eigenvalues to identify the unknown number of factors. Specifically, Wang et al. (2019) and Chen et al. (2020) assumed that the noise matrix follows white noise, using its autocovariance structure that disregards contemporaneous correlations for estimation procedures, which allows strong contemporaneous correlations among the elements of white noise. In contrast, Yu et al. (2022) and Chen and Fan (2023) employed concurrent covariance for their estimation procedures, allowing weak temporal correlations while assuming weak correlations among the elements of the noise matrix. However, when concurrent covariance is used to estimate the loading matrix as well as the number of factors in the factor model, weak correlations between the elements of the noise matrix need to be allowed. We will rely solely on the autocovariance to infer the two-way factor model when the elements of the noise matrix are not weakly correlated. Meanwhile, continuing the line of Wang et al. (2019), the idiosyncratic errors can be weakly correlated and asymptotically negligible under certain conditions. Wang et al. (2019) suggested the eigenvalue-ratio (ER) technique to determine the number of row and column factors. However, the eigenvalue-ratio (ER) estimator lacks theoretical consistency. Furthermore, our simulation evidence indicates that it suffers from substantial finite-sample bias, particularly under weak serial correlation or weak factor strength. As we know, estimation inaccuracy can significantly affect downstream analysis.

Therefore, we propose a novel method to estimate the number of factors in the row and column with more accuracy. Taking the factor model proposed by Wang et al. (2019) as an example, the matrix-variate time series is decomposed into two components: a dynamic component driven by a lower-dimensional factor process and a static component composed of matrix white noise. Since the white noise series exhibits no serial correlations, the decomposition is unique, meaning that for a given finite sample size, the dimension of the factor process and the factor loading spaces can be identified. Rather than working directly with the original time series, we instead rely on projecting the matrix-variate time series onto these components – specifically, either the dynamic component (primarily driven by the lower-dimensional factor process) or the static component (consisting of matrix white noise). Let $\boldsymbol{R}$ be the semi-orthogonal row-loading matrix, and $\boldsymbol{R}^{\dagger}$ its orthogonal complement. The matrix-variate time series projected along the columns of $\boldsymbol{R}$ forms a non-white noise sequence, while its projection along the columns of $\boldsymbol{R}^{\dagger}$ results in a white noise process. See (A.1) and (A.2) in the Supplementary Material for details. These observations are of importance because we can determine the number of factors through efficiently examining whether or not the projected series is white noise series. When idiosyncratic errors are weakly correlated, this novel approach is still effective between non-weakly correlated series and weakly correlated noise.

In this paper, for the new factor model with weakly correlated idiosyncratic errors, we use the maximum absolute autocovariances and the sum of squared autocovariance matrices of the component series (after projecting onto the estimated factor loading spaces) to construct respectively two sequences of ratios as signal functions. Their empirical versions are defined as the signal statistics. The estimators are the maximizers that are greater than thresholds for the signal statistics. We also note that, based on empirical findings, the strengths of the row and column factors interact with each other through the structure of the matrix factor model. This deteriorates the performance of the respective estimators of the number of row and column factors. To improve accuracy of the estimation, we make, when we focus on the row-loading matrix, a transformation for the model under study so that the transformed model only has a row loading matrix. Afterward, we determine the number of factors using the ratio estimation, which is proved to be consistent. The same principle applies to the transformed model only having column load matrix. Our numerical studies indicate the advantages of the transformation method when the matrix-variate time series exhibits weak serial correlations or contains weak factors.

The remainder of the paper is organized as follows. Section 2 introduces matrix-variate factor models and the new ratio-based estimators for the number of factors in rows and columns. Section 3 examines the theoretical properties of the proposed estimators. Section 4 presents simulation results and a real-data application. Section 5 concludes the research in this paper. All technical details are provided in the Supplementary Material.

Throughout the paper, we use $\boldsymbol{I}_p$ for the $p \times p$ identity matrix, and $A^\top$ for the transpose of a matrix $A = (a_{ij})$. Define $|A|_\infty = \max_{i,j} |a_{ij}|$ and $\sigma_i(A)$ as the i-th largest singular value of A. The symbols $\|A\|_F$, $\|A\|_2$, and $\|A\|_{\min}$ denote the Frobenius norm, the maximum eigenvalue, and the minimum eigenvalue of $AA^\top$ or $A^\top A$, respectively. Let a_n and b_n be two sequences of numbers, and write $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. Suppose A is a $km \times km$ square matrix with m row and column partitions. Let A_{ii} for $1 \le i \le m$ denote the $k \times k$ block matrices along the diagonal of A. We introduce the operation $\text{trs}_k(A) = \sum_{i=1}^m A_{ii}$. Finally, let $\text{vec}(\cdot)$ denote the vectorization operator.

2. Methodology

2.1 Two-way Factor Models

Let $\{\boldsymbol{Y}_t : t = 1, \ldots, n\}$ represent n observations from a matrix-variate time series with $\mathbb{E}\boldsymbol{Y}_t = \mathbf{0}$, where $\boldsymbol{Y}_t$ is a $p \times q$ matrix defined as

$$\boldsymbol{Y}_t = \left(Y_{\cdot 1,t}, \ldots, Y_{\cdot q,t}\right) = \begin{pmatrix} Y_{1\cdot,t} \\ \vdots \\ Y_{p\cdot,t} \end{pmatrix} = \begin{pmatrix} y_{11,t} & \cdots & y_{1q,t} \\ \vdots & \ddots & \vdots \\ y_{p1,t} & \cdots & y_{pq,t} \end{pmatrix}.$$

Wang et al. (2019) proposed a factor model for matrix-valued time series $\boldsymbol{Y}_t$, given by

$$\boldsymbol{Y}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}^\top + \boldsymbol{E}_t, \quad t = 1, \ldots, n, \tag{2.1}$$

where $\boldsymbol{F}_t = (F_{\cdot 1,t}, \ldots, F_{\cdot c,t}) = (F_{1\cdot,t}, \ldots, F_{r\cdot,t})^\top$ is an unobserved $r \times c$ matrix of common fundamental factors; $\boldsymbol{R}$ is a $p \times r$ row loading matrix; $\boldsymbol{C}$ is a $q \times c$ column loading matrix; and $\boldsymbol{E}_t = (E_{\cdot 1,t}, \ldots, E_{\cdot q,t}) = (E_{1\cdot,t}, \ldots, E_{p\cdot,t})^\top$ is a $p \times q$ matrix of random errors. In Model (2.1), the common fundamental factors $\boldsymbol{F}_t$ capture the dynamics and co-movements of $\boldsymbol{Y}_t$, while $\boldsymbol{R}$

and $\boldsymbol{C}$ reflect the significance of the factors and their interactions. Model (2.1) can also be rewritten as a classical factor model:

$$\text{vec}(\boldsymbol{Y}_t) = (\boldsymbol{C} \otimes \boldsymbol{R})\text{vec}(\boldsymbol{F}_t) + \text{vec}(\boldsymbol{E}_t), \quad t = 1, \ldots, n, \tag{2.2}$$

where $\boldsymbol{C} \otimes \boldsymbol{R}$ is the $pq \times rc$ loading matrix, with rc being the total number of factors; $\text{vec}(\boldsymbol{F}_t)$ represents the vectorized latent factors; and $\text{vec}(\boldsymbol{E}_t)$ denotes the vectorized error term.

For the vector-valued factor model (2.2), the factors $\text{vec}(\boldsymbol{F}_t)$ and the loading matrix $\boldsymbol{C} \otimes \boldsymbol{R}$ are identifiable only up to an invertible linear transformation. A similar identification issue arises in the matrix-valued factor model (2.1). Specifically, suppose H_1 and H_2 are two invertible matrices of sizes $r \times r$ and $c \times c$, respectively. Then the triplets $(\boldsymbol{R}, \boldsymbol{F}_t, \boldsymbol{C})$ and $(\boldsymbol{R}H_1, H_1^{-1}\boldsymbol{F}_t H_2^{-1}, \boldsymbol{C}H_2^{\top})$ are equivalent, implying that the factors and loadings are identifiable up to invertible transformations H_1 and H_2. However, once $\boldsymbol{R}$ and $\boldsymbol{C}$ are specified, the factor process $\boldsymbol{F}_t$ is uniquely determined. This non-uniqueness in $\boldsymbol{R}$ and $\boldsymbol{C}$ can be advantageous, as it allows for the selection of H_1 and H_2 to simplify estimation. For further details on identification issues in factor models, see Bai and Li (2012).

2.2 Regularity Conditions

To facilitate inference for the proposed model, we adopt the following regularity conditions, referring to Wang et al. (2019). For $h \geq 0$, define $\boldsymbol{\Sigma}_f(h) = \text{Cov}\big(\text{vec}(\boldsymbol{F}_t), \text{vec}(\boldsymbol{F}_{t+h})\big)$, $\boldsymbol{\Sigma}_{ef}(h) = \text{Cov}\big(\text{vec}(\boldsymbol{E}_t), \text{vec}(\boldsymbol{F}_{t+h})\big)$, $\boldsymbol{\Sigma}_{fe}(h) = \text{Cov}\big(\text{vec}(\boldsymbol{F}_t), \text{vec}(\boldsymbol{E}_{t+h})\big)$, $\boldsymbol{\Sigma}_e(h) = \text{Cov}\big(\text{vec}(\boldsymbol{E}_t), \text{vec}(\boldsymbol{E}_{t+h})\big)$, $\boldsymbol{\Sigma}_e = \text{Cov}\big(\text{vec}(\boldsymbol{E}_t)\big)$.

Assumption 1. (C1) There exist constants $\delta, \omega \in [0,1]$ such that $\|\boldsymbol{R}\|_2^2 \asymp p^{1-\delta} \asymp \|\boldsymbol{R}\|_{\min}^2$ and $\|\boldsymbol{C}\|_2^2 \asymp q^{1-\omega} \asymp \|\boldsymbol{C}\|_{\min}^2$.

(C2) The vector-valued process $\left(\text{vec}^\top(\boldsymbol{F}_t), \text{vec}^\top(\boldsymbol{E}_t)\right)^\top$, is α-mixing, with mixing coefficients $\alpha(h)$ satisfying $\sum_{h=1}^\infty \alpha(h)^{1-2/\gamma} < \infty$ for some $\gamma > 2$, 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 $\mathcal{F}_i^j$ denotes the σ-field generated by $\{\left(\text{vec}^\top(\boldsymbol{F}_t), \text{vec}^\top(\boldsymbol{E}_t)\right)^\top : i \le t \le j\}$.

(C3) Let $F_{t,ij}$ be the (i,j)-th entry of $\boldsymbol{F}_t$. For any $i = 1, \ldots, r$, $j = 1, \ldots, c$, and $t = 1, \ldots, n$, assume $\mathbb{E}(|F_{t,ij}|^{2\gamma}) \le \mathcal{C}$ for some constant $\mathcal{C} > 0$. Moreover, there exists $h \in [0, h_0]$ such that $\text{rank}(\boldsymbol{\Sigma}_f(h)) = l = \max(r,c)$, and $\Sigma_f(h) \asymp \mathbf{O}(1) \asymp \sigma_l(\Sigma_f(h))$ as $p, q \to \infty$ with r, c fixed.

(C4) For $h \ge 1$, $|\boldsymbol{\Sigma}_e(h)|_\infty = \mathbf{O}(n^{-1})$. Let $E_{t,kl}$ denote the (k,l)-th entry of $\boldsymbol{E}_t$ for any $k = 1, \ldots, p$, $l = 1, \ldots, q$, where $\mathbb{E}(|E_{t,kl}|^{2\gamma}) \le \mathcal{C}$.

(C5) The elements of $\boldsymbol{\Sigma}_e$ are bounded as $n, p, q \to \infty$. Furthermore, for $h \ge 0$, $\|\boldsymbol{\Sigma}_{ef}(h)\|_2 = \mathbf{o}(p^{\frac{1-\delta}{2}} q^{\frac{1-\omega}{2}})$ and $\boldsymbol{\Sigma}_{fe}(h) = 0$.

Conditions (C1)-(C3) are similar to those in Wang et al. (2019). When $\delta = \omega = 0$, the row and column factors are referred to as strong factors. Conversely, when $\delta, \omega \neq 0$, the factors are classified as weak factors. Wang et al. (2019) assumed that the random errors are white noise, which is relaxed by condition (C4). Condition (C4) implies that the random errors can have weak correlation. Wang et al. (2019) assumed that the random errors are not correlated with the factors. Condition (C5) is less restrictive, requiring only that the future white noise components are uncorrelated with the factors up to the present.

3. The Two-Step Procedure for Identifying The Numbers of Factors

Based on the structure of Model (2.1), the strength of the factors in the rows and columns interact with each other, we consider a two-step procedure as follows.

3.1 Transformed Factor Models

Following the approach in Yu et al. (2022), if $\boldsymbol{C}$ is known and $\boldsymbol{C}^\top \boldsymbol{C}/q^{1-\omega} = \boldsymbol{I}_c$, the matrix-valued time series $\boldsymbol{Y}_t$ can be projected to lower-dimensional spaces by setting

$$\boldsymbol{Y}_t \boldsymbol{C}/q^{1-\omega} = \boldsymbol{R}\boldsymbol{F}_t + \boldsymbol{E}_t \boldsymbol{C}/q^{1-\omega} := \boldsymbol{R}\boldsymbol{F}_t + \boldsymbol{\xi}_t. \tag{3.1}$$

It is easy to see that Model (2.1) reduces to a new one with only a row loading matrix. When the number of factors is unknown so that $\boldsymbol{C}$ is unknown, we choose an appropriate matrix $\widetilde{\boldsymbol{C}} = q^{\frac{1-\omega}{2}}(\tilde{Q}_1, \tilde{Q}_2, \ldots, \tilde{Q}_m) := (C_{\cdot 1}, \ldots, C_{\cdot m})$ with a large positive integer m and $\tilde{Q}_i$ defined after (A.8) in the Supplementary Material such that

$$\boldsymbol{C}^\top \widetilde{\boldsymbol{C}} = q^{1-\omega}(\boldsymbol{I}_c, \boldsymbol{0}),$$

where $\boldsymbol{I}_\cdot$ denotes an identity matrix. The matrix-valued time series $\boldsymbol{Y}_t$ is then projected to a lower-dimensional space by setting

$$\boldsymbol{X}_t = \boldsymbol{Y}_t\widetilde{\boldsymbol{C}}/q^{1-\omega} = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}^\top\widetilde{\boldsymbol{C}}/q^{1-\omega} + \boldsymbol{E}_t\widetilde{\boldsymbol{C}}/q^{1-\omega} := \boldsymbol{R}\widetilde{\boldsymbol{F}}_t + \widetilde{\boldsymbol{\xi}}_t. \tag{3.2}$$

For $i, j = 1, 2, \ldots, c$, define the following covariance matrices:

$$\begin{aligned}
\boldsymbol{\Sigma}_{x,ij}(h) &= \frac{1}{n-h}\sum_{t=1}^{n-h}\mathrm{Cov}(\boldsymbol{Y}_t C_{\cdot i}, \boldsymbol{Y}_{t+h}C_{\cdot j}),\\
\boldsymbol{\Sigma}_{f,ij}(h) &= \frac{1}{n-h}\sum_{t=1}^{n-h}\mathrm{Cov}(f_{t,\cdot i}, f_{t+h,\cdot j}),\\
\boldsymbol{\Sigma}_{\xi f,ij}(h) &= \frac{1}{n-h}\sum_{t=1}^{n-h}\mathrm{Cov}(\boldsymbol{E}_t C_{\cdot i}, F_{t+h,\cdot j}),\\
\boldsymbol{\Sigma}_{\xi,ij}(h) &= \frac{1}{n-h}\sum_{t=1}^{n-h}\mathrm{Cov}(\boldsymbol{E}_t C_{\cdot i}, \boldsymbol{E}_t C_{\cdot j}).
\end{aligned}$$

Since $C_{\cdot i}^\top C_{\cdot j} = 0$ for $i \neq j$, and slightly differing from Wang et al. (2019), we further define:

$$\begin{aligned}
\boldsymbol{\Omega}_x(h) &= \frac{1}{q^{2-2\omega}}\sum_{i=1}^{c}\boldsymbol{\Sigma}_{x,ii}(h) = \frac{1}{(n-h)q^{2-2\omega}}\sum_{t=1}^{n-h}\mathbb{E}(\boldsymbol{Y}_t\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{C}}^\top\boldsymbol{Y}_{t+h}^\top),\\
\boldsymbol{\Omega}_f(h) &= \sum_{i=1}^{c}\boldsymbol{\Sigma}_{f,ii}(h) = \frac{1}{n-h}\sum_{t=1}^{n-h}\mathbb{E}(\boldsymbol{F}_t\boldsymbol{F}_{t+h}^\top),\\
\boldsymbol{\Omega}_{\xi f}(h) &= \frac{1}{q^{1-\omega}}\sum_{i=1}^{c}\boldsymbol{\Sigma}_{\xi f,ii}(h) = \frac{1}{(n-h)q^{1-\omega}}\sum_{t=1}^{n-h}\mathbb{E}(\boldsymbol{E}_t\boldsymbol{C}\boldsymbol{F}_{t+h}^\top),\\
\boldsymbol{\Omega}_\xi(h) &= \sum_{i=1}^{c}\boldsymbol{\Sigma}_{\xi,ii}(h) = \frac{1}{n-h}\sum_{t=1}^{n-h}\mathbb{E}(\boldsymbol{E}_t\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{C}}^\top\boldsymbol{E}_{t+h}^\top).
\end{aligned}$$

By Assumption 1-(C5),

$$\mathbb{E}(\boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}^\top\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{C}}^\top\boldsymbol{E}_{t+h}^\top) = q^{1-\omega}\mathbb{E}(\boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}^\top\boldsymbol{E}_{t+h}^\top) = \mathrm{trs}_p\left(\mathbb{E}\left(\mathrm{vec}(\boldsymbol{R}\boldsymbol{F}_t)\mathrm{vec}^\top(\boldsymbol{E}_{t+h}\boldsymbol{C})\right)\right) = 0,$$

where the definition of trs_p is provided in Filipiak et al. (2018). By Assumption 1-(C1), (C3) and (C4), it holds that $\left\|\boldsymbol{R}\mathbb{E}(\boldsymbol{F}_t\boldsymbol{F}_{t+h}^\top)\boldsymbol{R}^\top\right\|_2 \asymp p^{1-\delta}$ and

$$\begin{aligned}\left\|\mathbb{E}(\boldsymbol{E}_t\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{C}}^\top\boldsymbol{E}_{t+h}^\top)\right\|_2 &= \left\|\mathrm{trs}_p\big(\mathbb{E}\mathrm{vec}(\boldsymbol{E}_t\widetilde{\boldsymbol{C}})\mathrm{vec}^\top(\boldsymbol{E}_{t+h}\widetilde{\boldsymbol{C}})\big)\right\|_2 = \left\|\mathrm{trs}_p\big((\widetilde{\boldsymbol{C}}^\top\otimes\boldsymbol{I}_p)\boldsymbol{\Sigma}_e(h)(\widetilde{\boldsymbol{C}}\otimes\boldsymbol{I}_p)\big)\right\|_2\\ &\leq c\left\|\big((\widetilde{\boldsymbol{C}}^\top\otimes\boldsymbol{I}_p)\boldsymbol{\Sigma}_e(h)(\widetilde{\boldsymbol{C}}\otimes\boldsymbol{I}_p)\right\|_2 = \mathbf{O}(pq^{2-\omega}n^{-1}).\end{aligned}$$

Comparing with $q^{2-2\omega}\boldsymbol{R}\mathbb{E}(\boldsymbol{F}_t\boldsymbol{F}_{t+h}^\top)\boldsymbol{R}^\top$, then $\mathbb{E}(\boldsymbol{E}_t\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{C}}^\top\boldsymbol{E}_{t+h}^\top)$ is asymptotically negligible. Thus we have

$$\mathbb{E}(\boldsymbol{Y}_t\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{C}}^\top\boldsymbol{Y}_{t+h}^\top) \approx q^{2-2\omega}\boldsymbol{R}\mathbb{E}(\boldsymbol{F}_t\boldsymbol{F}_{t+h}^\top)\boldsymbol{R}^\top + q^{1-\omega}\mathbb{E}(\boldsymbol{E}_t\boldsymbol{C}\boldsymbol{F}_{t+h}^\top)\boldsymbol{R}^\top.$$

For a pre-determined integer $h_0 \geq 1$, we consider the main non-zero lags and define:

$$\widetilde{\boldsymbol{M}}_1 = \sum_{h=1}^{h_0}\boldsymbol{\Omega}_x^\top(h)\boldsymbol{\Omega}_x(h) \approx \boldsymbol{R}\left(\sum_{h=1}^{h_0}\left(\boldsymbol{R}\boldsymbol{\Omega}_f(h)+\boldsymbol{\Omega}_{\xi f}(h)\right)^\top\left(\boldsymbol{R}\boldsymbol{\Omega}_f(h)+\boldsymbol{\Omega}_{\xi f}(h)\right)\right)\boldsymbol{R}^\top. \tag{3.3}$$

From Equation (3.3), we see that each column of $\widetilde{\boldsymbol{M}}_1$ is a linear combination of the columns of $\boldsymbol{R}$, and thus the eigenspace of $\widetilde{\boldsymbol{M}}_1$ is the same as the eigenspace of $\boldsymbol{R}$, i.e., $\mathcal{M}(\widetilde{\boldsymbol{M}}_1) = \mathcal{M}(\boldsymbol{R})$. Therefore, $\mathcal{M}(\boldsymbol{R})$ can be estimated by the space spanned by the eigenvectors of the sample version of $\widetilde{\boldsymbol{M}}_1$. Let O_j be the unit eigenvector corresponding to the j-th largest eigenvalue of $\widetilde{\boldsymbol{M}}_1$. Define $\boldsymbol{O} = (O_1, O_2, \ldots, O_r)$, and then obtain $\boldsymbol{R} = p^{\frac{1-\delta}{2}}\boldsymbol{O}$.

We construct the sample versions of these quantities to estimate the factor loading matrix as follows. Let

$$\widehat{\widetilde{\boldsymbol{M}}}_1 = \sum_{h=1}^{h_0}\widehat{\boldsymbol{\Omega}}_x^\top(h)\widehat{\boldsymbol{\Omega}}_x(h), \tag{3.4}$$

where $\widehat{\mathbf{\Omega}}_x(h) = \frac{1}{(n-h)q^{2-2\omega}} \sum_{t=1}^{n-h} \mathbf{Y}_t \widehat{\widetilde{\boldsymbol{C}}} \widehat{\widetilde{\boldsymbol{C}}}^\top \mathbf{Y}_{t+h}^\top$. Then, $\mathcal{M}(\boldsymbol{R})$ can be estimated by $\mathcal{M}(\widehat{\boldsymbol{O}})$, where $\widehat{\boldsymbol{O}} = (\hat{O}_1, \ldots, \hat{O}_r)$, and $\hat{O}_1, \ldots, \hat{O}_r$ are the unit eigenvectors of $\widehat{\widetilde{\boldsymbol{M}}}_1$ corresponding to its r largest eigenvalues. As a result, $\widehat{\boldsymbol{R}} = p^{\frac{1-\delta}{2}} \widehat{\boldsymbol{O}}$.

Remark 1. First, $\widehat{\widetilde{\boldsymbol{C}}}$ can be estimated by (A.8) in the Supplementary Material. **Second**, in fact, a small positive value m can also be chosen. Then $\boldsymbol{C}^\top \widetilde{\boldsymbol{C}} = q^{1-\omega}(\boldsymbol{I}_m, \mathbf{0})^\top$. The result and the relative proof are similar. Thus, we do not discuss them further. **Third**, omitting $q^{2-2\omega}$ in practice, we can consider $\mathbf{\Omega}_x(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} \mathbb{E}(\mathbf{Y}_t \widetilde{\boldsymbol{C}} \widetilde{\boldsymbol{C}}^\top \mathbf{Y}_{t+h}^\top)$, and $\widetilde{\boldsymbol{M}}_1 = \sum_{h=1}^{h_0} \mathbf{\Omega}_x^\top(h) \mathbf{\Omega}_x(h)$ accordingly. **Fourth**, according to Wang et al. (2019), the selection of h_0 is not sensitive to the estimations of the number of factors and the loading matrix. In this paper, we set $h_0 = 2$ in all simulation experiments and the application. **Fifth**, the autocovariance is used to construct statistics for estimating the loading matrix and the number of factors in rows and columns. If we employ the concurrent covariance, the additional condition, that is $\|\mathbf{\Sigma}_e\|_2 = \mathbf{o}(p^{1-\delta} q^{1-\omega})$, should be satisfied.

3.2 Max-type and Sum-type Estimators

For Model (3.2), as discussed in Section 2.1, we can assume without loss of generality that $\boldsymbol{O} = (O_1, \ldots, O_r) = \boldsymbol{R}/p^{\frac{1-\delta}{2}}$ is semi-orthogonal, i.e., $\boldsymbol{O}^\top \boldsymbol{O} = \boldsymbol{I}_r$, where $O_i \in \mathbb{R}^{p\times 1}$ for $1 \le i \le r$. Let $\boldsymbol{O}^\dagger = (O_{r+1}, \ldots, O_p)$ be the orthogonal complement matrix of $\boldsymbol{O}$, so that $(\boldsymbol{O}, \boldsymbol{O}^\dagger)^\top (\boldsymbol{O}, \boldsymbol{O}^\dagger) = \boldsymbol{I}_p$. For a better understanding of our approach, we assume that $\boldsymbol{E}_t$ is white noise, then we have:

$$O_i^\top \boldsymbol{X}_t = O_i^\top \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}\widetilde{\boldsymbol{C}}/q^{1-\omega} + O_i^\top \boldsymbol{E}_t\widetilde{\boldsymbol{C}}/q^{1-\omega}, \quad i = 1, \ldots, r; \tag{3.5}$$

$$O_i^\top \boldsymbol{X}_t = O_i^\top \boldsymbol{E}_t\widetilde{\boldsymbol{C}}/q^{1-\omega}, \quad i = r+1, \ldots, p. \tag{3.6}$$

This implies that $O_i^\top \boldsymbol{X}_t$ in (3.5) is not a white noise sequence for $i = 1, \ldots, r$, while $O_i^\top \boldsymbol{X}_t$ in (3.6) is a white noise sequence for $i = r+1, \ldots, p$. This observation offers the motivation to develop a new approach to estimate the number of factors r by examining the uncorrelatedness of the sequences $O_i^\top \boldsymbol{X}_t$ for $i = 1, \ldots, p$.

Let $\mathfrak{R}_i = (O_i, \ldots, O_p)$, for $i = 1, \ldots, p$. The estimator is based on statistics for checking the white noise assumption on the sequence $\{\mathfrak{R}_i^\top \boldsymbol{X}_t : t = 1, 2, \ldots\}$. Define the sample covariance matrix for $\boldsymbol{X}_t$ as

$$\boldsymbol{\Sigma}_X(h) := \frac{1}{n-h}\sum_{t=1}^{n-h} \mathbb{E}(\boldsymbol{X}_t\boldsymbol{X}_{t+h}^\top) \in \mathbb{R}^{p\times p}.$$

The auto-covariance matrix for $\{\mathfrak{R}_i^\top \boldsymbol{X}_t\}$ at lag h is given by

$$\breve{\Gamma}_i(h) := \left(\breve{\gamma}_{i,kl}(h)\right) = \mathfrak{R}_i^\top \boldsymbol{\Sigma}_X(h)\mathfrak{R}_i. \tag{3.7}$$

Referring to Chang et al. (2017), Chang et al. (2018) and Li et al. (2019), the max-type and sum-type tests can be proposed below to identify whether the sequence $\{\mathfrak{R}_i^\top \boldsymbol{X}_t : t = 1, 2, \ldots\}$ is white noise or not. The max-type test is constructed based on the maximum norm of the auto-covariance matrix:

$$\breve{T}_{i,n} := \max_{1\le h\le K} n^{1/2}\left|\breve{\Gamma}_i(h)\right|_\infty = \max_{1\le h\le K} n^{1/2}\left|\mathfrak{R}_i^\top \boldsymbol{\Sigma}_X(h)\mathfrak{R}_i\right|_\infty, \tag{3.8}$$

where K is a user-specified integer. The sum-type test is constructed based on the Frobenius norm of the auto-covariance matrix:

$$\breve{G}_{i,n} := \sum_{1\le h\le K} \mathrm{Tr}\left(\breve{\Gamma}_i(h)^\top \breve{\Gamma}_i(h)\right) = \sum_{1\le h\le K} \left\|\breve{\Gamma}_i(h)\right\|_F^2 = \sum_{1\le h\le K}\sum_j \breve{\sigma}_j^2\left(\breve{\Gamma}_i(h)\right), \tag{3.9}$$

where Tr denotes the trace operation for a square matrix and $\breve{\sigma}_j^2$ denotes the square of the j-th largest singular value of $\breve{\Gamma}_i(h)$. Significant magnitudes of $\breve{T}_{i,n}$ and $\breve{G}_{i,n}$ provide evidence against the null hypothesis of white noise in the projected series, suggesting the presence of residual dynamic structure. To estimate the number of factors r, we sequentially examine the values of the statistics as i grows. Intuitively, when $i = r$, the gap between $\breve{T}_{i,n}$ and $\breve{T}_{i+1,n}$ is likely to reach its maximum value, as $\{\mathfrak{R}_{r+1}^\top \boldsymbol{X}_t\}$ is a white noise sequence, while $\{\mathfrak{R}_r^\top \boldsymbol{X}_t\}$ is a non-white noise sequence. When $\boldsymbol{E}_t$ is weakly correlated, the same phenomenon can be observed. It is merely that $\breve{T}_{i,n}$=$\breve{G}_{i,n}$=0 under the assumption of white noise for $\boldsymbol{E}_t$, while $\breve{T}_{i,n} > 0$ and $\breve{G}_{i,n} > 0$ under the assumption of weak correlation for $\boldsymbol{E}_t$.

By Lemma 10 in the Supplementary Material, $\breve{T}_{i,n}$ and $\breve{G}_{i,n}$ are monotonically decreasing with respect to i. Based on (3.3), we can construct the feasible max-type and sum-type statistics as:

$$\begin{aligned}\widehat{\breve{T}}_{i,n} &= \max_{1\le h\le K} n^{1/2}\left|\widehat{\breve{\Gamma}}_i(h)\right|_\infty, \\ \widehat{\breve{G}}_{i,n} &= \sum_{1\le h\le K} \mathrm{Tr}\left(\widehat{\breve{\Gamma}}_i(h)^\top \widehat{\breve{\Gamma}}_i(h)\right),\end{aligned}$$

where $\widehat{\breve{\Gamma}}_i(h) = \widehat{\mathfrak{R}}_i^\top \widehat{\boldsymbol{\Sigma}}_X(h)\widehat{\mathfrak{R}}_i$. By Lemma 10, $\widehat{\breve{T}}_{i,n}$ and $\widehat{\breve{G}}_{i,n}$ also decrease monotonically with respect to i. The following ratio estimators for r are proposed based on an enhanced elbow criterion:

$$\widehat{r}_{\text{MR}} = \arg\max_{1\le i\le p/2} \frac{\widehat{\breve{T}}_{i,n}}{\widehat{\breve{T}}_{i+1,n}} =: \arg\max_{1\le i\le p/2} \text{MR}(i), \tag{3.10}$$

$$\widehat{r}_{\text{SR}} = \arg\max_{1\le i\le p/2} \frac{\widehat{\breve{G}}_{i,n} - \widehat{\breve{G}}_{i+1,n}}{\widehat{\breve{G}}_{i+1,n} - \widehat{\breve{G}}_{i+2,n}} =: \arg\max_{1\le i\le p/2} \text{SR}(i), \tag{3.11}$$

where $\text{MR}(i) = \widehat{\breve{T}}_{i,n}/\widehat{\breve{T}}_{i+1,n}$ and $\text{SR}(i) = (\widehat{\breve{G}}_{i,n} - \widehat{\breve{G}}_{i+1,n})/(\widehat{\breve{G}}_{i+1,n} - \widehat{\breve{G}}_{i+2,n})$.

Remark 2. The estimators of r that we propose are defined as the maximizers of $\text{MR}(i)$ and $\text{SR}(i)$, respectively. In contrast, the counterpart estimator $\widehat{r}_{\text{ER}}$ is defined as the maximizer of the eigenvalue ratio, given by

$$\widehat{r}_{\text{ER}} = \arg\max_{1\le i\le p/2} \frac{\widehat{\lambda}_i(\widehat{\widetilde{\boldsymbol{M}}}_1)}{\widehat{\lambda}_{i+1}(\widehat{\widetilde{\boldsymbol{M}}}_1)} =: \arg\max_{1\le i\le p/2} \text{ER}(i),$$

where $\widehat{\lambda}_i(\widehat{\widetilde{\boldsymbol{M}}}_1)$ denotes the i-th largest eigenvalue of the matrix $\widehat{\widetilde{\boldsymbol{M}}}_1$. For $\widehat{r}_{\text{MR}}$ and $\widehat{r}_{\text{SR}}$, if the random errors are of weak correlation, both $\widehat{\breve{T}}_{i,n}$ and $\widehat{\breve{G}}_{i,n}$ are greater than zero. Then their consistency can be ensured. However, because $\widetilde{\boldsymbol{M}}_1$ is not in full rank, the ER method based on $\widetilde{\boldsymbol{M}}_1$ is not consistent.

Similarly, we can estimate the number of column factors c. As the procedure is almost the same, we only introduce the notation and estimators. Select a suitable $\widetilde{\boldsymbol{R}} = p^{\frac{1-\delta}{2}}(Q_1, Q_2, \ldots, Q_m) = (R_{\cdot 1}, \ldots, R_{\cdot m})$ with a large positive integer m, where Q_i is the eigenvector of $\boldsymbol{M}_1$ in (A.5) of the Supplementary Material, and define:

$$\boldsymbol{\Omega}_{\tilde{x}}(h) := \frac{1}{(n-h)p^{2-2\delta}} \sum_{t=1}^{n-h} \mathbb{E}(\boldsymbol{Y}_t^\top \widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{R}}^\top \boldsymbol{Y}_{t+h}),$$

$$\boldsymbol{\Omega}_{\tilde{f}}(h) := \frac{1}{n-h} \sum_{t=1}^{n-h} \mathbb{E}(\boldsymbol{F}_t^\top \boldsymbol{F}_{t+h}),$$

$$\boldsymbol{\Omega}_{\tilde{\xi}\tilde{f}}(h) := \frac{1}{(n-h)p^{1-\delta}} \sum_{t=1}^{n-h} \mathbb{E}(\boldsymbol{E}_t^\top \boldsymbol{R}\boldsymbol{F}_{t+h}).$$

Further, define the matrix:

$$\widetilde{\boldsymbol{M}}_2 = \sum_{h=1}^{h_0} \boldsymbol{\Omega}_{\tilde{x}}^\top(h)\boldsymbol{\Omega}_{\tilde{x}}(h) \approx \boldsymbol{C}\left(\sum_{h=1}^{h_0}\left(\boldsymbol{C}\boldsymbol{\Omega}_{\tilde{f}}(h)+\boldsymbol{\Omega}_{\tilde{\xi}\tilde{f}}(h)\right)^\top\left(\boldsymbol{C}\boldsymbol{\Omega}_{\tilde{f}}(h)+\boldsymbol{\Omega}_{\tilde{\xi}\tilde{f}}(h)\right)\right)\boldsymbol{C}^\top, \quad (3.12)$$

and its sample version:

$$\widehat{\widetilde{\boldsymbol{M}}}_2 = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Omega}}_{\tilde{x}}^\top(h)\widehat{\boldsymbol{\Omega}}_{\tilde{x}}(h), \quad (3.13)$$

where $\widehat{\boldsymbol{\Omega}}_{\tilde{x}}(h) = \frac{1}{(n-h)q^{2-2\omega}} \sum_{t=1}^{n-h} \boldsymbol{Y}_t^\top \widehat{\widetilde{\boldsymbol{R}}}\widehat{\widetilde{\boldsymbol{R}}}^\top \boldsymbol{Y}_{t+h}$. Let $\tilde{O}_j$ and $\hat{\tilde{O}}_j$ be the unit eigenvectors corresponding to the j-th largest eigenvalue of $\widetilde{\boldsymbol{M}}_2$ and $\widehat{\widetilde{\boldsymbol{M}}}_2$, respectively. We define $\widetilde{\boldsymbol{O}} = (\tilde{O}_1, \tilde{O}_2, \ldots, \tilde{O}_c)$ and $\widehat{\widetilde{\boldsymbol{O}}} = (\hat{\tilde{O}}_1, \hat{\tilde{O}}_2, \ldots, \hat{\tilde{O}}_c)$, and then obtain $\boldsymbol{C} = q^{\frac{1-\omega}{2}}\widetilde{\boldsymbol{O}}$ and $\widehat{\boldsymbol{C}} = q^{\frac{1-\omega}{2}}\widehat{\widetilde{\boldsymbol{O}}}$.

Let $\mathfrak{C}_i = (\tilde{O}_i, \ldots, \tilde{O}_q)$ for $i = 1, \ldots, q$, and define the covariance matrix for $\widetilde{\boldsymbol{X}}_t$ as:

$$\boldsymbol{\Sigma}_{\widetilde{X}}(h) := \frac{1}{n-h} \sum_{t=1}^{n-h} \mathbb{E}(\widetilde{\boldsymbol{X}}_t \widetilde{\boldsymbol{X}}_{t+h}^\top) \in \mathbb{R}^{q\times q}.$$

The auto-covariance matrix for $\{\mathfrak{C}_i^\top \widetilde{\boldsymbol{X}}_t\}$ at lag h is:

$$\check{\Gamma}_i(h) := \mathfrak{C}_i^\top \boldsymbol{\Sigma}_{\widetilde{X}}(h)\mathfrak{C}_i. \quad (3.14)$$

The max-type test statistic is given by:

$$\check{T}_{i,n} := \max_{1\le h\le K} n^{1/2} \left|\check{\Gamma}_i(h)\right|_\infty = \max_{1\le h\le K} n^{1/2} \left|\mathfrak{C}_i^\top \boldsymbol{\Sigma}_{\tilde{X}}(h)\mathfrak{C}_i\right|_\infty , \tag{3.15}$$

and the sum-type test statistic is:

$$\check{G}_{i,n} := \sum_{1\le h\le K} \mathrm{Tr}\left(\check{\Gamma}_i^\top(h)\check{\Gamma}_i(h)\right) = \sum_{1\le h\le K} \left\|\check{\Gamma}_i(h)\right\|_F^2 = \sum_{1\le h\le K}\sum_j \check{\sigma}_j^2\left(\check{\Gamma}_i(h)\right), \tag{3.16}$$

where $\check{\sigma}_j^2$ denotes the square of the j-th largest singular value of $\check{\Gamma}_i(h)$.

The sample versions of the max-type and sum-type statistics are:

$$\begin{aligned}\widehat{\check{T}}_{i,n} &= \max_{1\le h\le K} n^{1/2} \left|\widehat{\check{\Gamma}}_i(h)\right|_\infty ,\\ \widehat{\check{G}}_{i,n} &= \sum_{1\le h\le K} \mathrm{Tr}\left(\widehat{\check{\Gamma}}_i^\top(h)\widehat{\check{\Gamma}}_i(h)\right),\end{aligned}$$

where $\widehat{\check{\Gamma}}_i(h) = \widehat{\mathfrak{C}}_i^\top \widehat{\boldsymbol{\Sigma}}_{\tilde{X}}(h)\widehat{\mathfrak{C}}_i$. The ratio-based estimators for the number of column factors c are:

$$\widehat{c}_{\mathrm{MR}} = \arg\max_{1\le i\le q/2} \frac{\widehat{\check{T}}_{i,n}}{\widehat{\check{T}}_{i+1,n}} =: \arg\max_{1\le i\le q/2} \widetilde{\mathrm{MR}}(i), \tag{3.17}$$

$$\widehat{c}_{\mathrm{SR}} = \arg\max_{1\le i\le q/2} \frac{\widehat{\check{G}}_{i,n} - \widehat{\check{G}}_{i+1,n}}{\widehat{\check{G}}_{i+1,n} - \widehat{\check{G}}_{i+2,n}} =: \arg\max_{1\le i\le q/2} \widetilde{\mathrm{SR}}(i), \tag{3.18}$$

where $\widetilde{\mathrm{MR}}(i) = \frac{\widehat{\check{T}}_{i,n}}{\widehat{\check{T}}_{i+1,n}}$, and $\widetilde{\mathrm{SR}}(i) = \frac{\widehat{\check{G}}_{i,n}-\widehat{\check{G}}_{i+1,n}}{\widehat{\check{G}}_{i+1,n}-\widehat{\check{G}}_{i+2,n}}$.

Remark 3. Similarly to the maximizers of $\widetilde{\mathrm{MR}}(j)$ and $\widetilde{\mathrm{SR}}(j)$, the estimators of c as maximizing $\widetilde{\mathrm{ER}}(j)$ is also defined as follows:

$$\widehat{c}_{\mathrm{ER}} = \arg\max_{1\le i\le q/2} \frac{\widehat{\lambda}_i(\widehat{\widetilde{\boldsymbol{M}}}_2)}{\widehat{\lambda}_{i+1}(\widehat{\widetilde{\boldsymbol{M}}}_2)} =: \arg\max_{1\le i\le q/2} \widetilde{\mathrm{ER}}(i),$$

where $\widehat{\lambda}_i(\widehat{\widetilde{\boldsymbol{M}}}_2)$ denotes the i-th largest eigenvalue of $\widehat{\widetilde{\boldsymbol{M}}}_2$. Additionally, based on model (2.1), we also propose estimators for r and c by taking the maximizers of $\mathrm{ER}_o(i)$, $\mathrm{MR}_o(i)$, $\mathrm{SR}_o(i)$, $\widetilde{\mathrm{ER}}_o(i)$, $\widetilde{\mathrm{MR}}_o(i)$ and $\widetilde{\mathrm{SR}}_o(i)$, respectively. These estimators, denoted respectively as $\widehat{r}_{\mathrm{ER}_o}$, $\widehat{r}_{\mathrm{MR}_o}$, $\widehat{r}_{\mathrm{SR}_o}$, $\widehat{c}_{\mathrm{ER}_o}$, $\widehat{c}_{\mathrm{MR}_o}$ and $\widehat{c}_{\mathrm{SR}_o}$, follow a similar naming convention as in the Supplementary Material of this paper.

Finally, we suggest the following Algorithm 1 to specify the numbers of row and column factors.

Algorithm 1 Two-step procedure for specifying the numbers of factors

Input: Data matrices $\{\boldsymbol{Y}_t\}_{t\le T}$, maximum number m

Output: Numbers of row and column factors $\widehat{r}_{\mathrm{MR}}$, $\widehat{r}_{\mathrm{SR}}$, $\widehat{c}_{\mathrm{MR}}$, $\widehat{c}_{\mathrm{SR}}$

1: Initial estimators $\widehat{\widehat{\boldsymbol{R}}}$ and $\widehat{\widehat{\boldsymbol{C}}}$ from $\widehat{\boldsymbol{M}}_1$ in (A.5) and $\widehat{\boldsymbol{M}}_2$ in (A.8), respectively;

2:
- Calculate $\widehat{\widetilde{\boldsymbol{M}}}_1$ using $\widehat{\widehat{\boldsymbol{C}}}$, then obtain $\widehat{r}_{\mathrm{MR}}$ and $\widehat{r}_{\mathrm{SR}}$ from equations (3.10) and (3.11);
- Calculate $\widehat{\widetilde{\boldsymbol{M}}}_2$ using $\widehat{\widehat{\boldsymbol{R}}}$, then obtain $\widehat{c}_{\mathrm{MR}}$ and $\widehat{c}_{\mathrm{SR}}$ from equations (3.17) and (3.18).

3.3 Asymptotic Properties of The Estimators

The properties of the proposed estimators for the two-step method are investigated under the asymptotic region where p, q and n all tend to infinity while r and c are fixed. The rates of convergence for the ratios of the max-type and sum-type statistics as well as the consistency of the proposed estimators are established in the following theories.

Theorem 1. *Under Assumptions 1 and* $p^{\delta}q^{\omega}n^{-1/2} = \boldsymbol{o}(1)$*, we have*

$$\frac{\widehat{\breve{T}}_{r+1,n}}{\widehat{\breve{T}}_{r,n}} = \boldsymbol{O}_p\big(p^{2\delta}q^{2\omega}n^{-1}\big); \quad \frac{\widehat{\breve{T}}_{c+1,n}}{\widehat{\breve{T}}_{c,n}} = \boldsymbol{O}_p\big(p^{2\delta}q^{2\omega}n^{-1}\big).$$

Theorem 2. *Under Assumptions 1 and* $p^{\delta}q^{\omega}n^{-1/2} = \boldsymbol{o}(1)$*, we have*

$$\frac{\widehat{\breve{G}}_{r+1,n} - \widehat{\breve{G}}_{r+2,n}}{\widehat{\breve{G}}_{r,n} - \widehat{\breve{G}}_{r+1,n}} = \boldsymbol{O}_p\big(p^{2\delta}q^{2\omega}n^{-1}\big); \quad \frac{\widehat{\breve{G}}_{c+1,n} - \widehat{\breve{G}}_{c+2,n}}{\widehat{\breve{G}}_{c,n} - \widehat{\breve{G}}_{c+1,n}} = \boldsymbol{O}_p\big(p^{2\delta}q^{2\omega}n^{-1}\big).$$

Theorem 3. *Suppose Assumptions 1 and* $p^{\delta}q^{\omega}n^{-1/2} = \boldsymbol{o}(1)$ *hold. Then, as* $|\boldsymbol{\Sigma}_e(h)|_{\infty} > 0$*, the* MR, SR, $\widetilde{\mathrm{MR}}$ *and* $\widetilde{\mathrm{SR}}$ *estimators (i.e.,* $\widehat{r}_{\mathrm{MR}}$ *in (3.10),* $\widehat{r}_{\mathrm{SR}}$ *in (3.11),* $\widehat{c}_{\mathrm{MR}}$ *in (3.17) and* $\widehat{c}_{\mathrm{SR}}$ *in (3.18)) satisfy that*

$$P(\widehat{r}_{\mathrm{SR}} = r) \to 1; \quad P(\widehat{c}_{\mathrm{SR}} = c) \to 1;$$
$$P(\widehat{r}_{\mathrm{MR}} = r) \to 1; \quad P(\widehat{c}_{\mathrm{MR}} = c) \to 1.$$

Remark 4. First, when $\delta = \omega < 0.5$ and $n \asymp pq$, the condition for weak factors is $p^{\delta}q^{\omega}n^{-1/2} = \mathbf{o}(1)$; when $\delta = \omega = 0$, the convergence rates for strong factors are all n^{-1}. Second, all proposed ratios are the smallest in position r or c, which is just the number of factors in the row or column. As the estimation procedures can be the same, we consider the row loading matrix in the following. The detailed estimation for the column loading matrix is omitted in the theoretical investigation. From Theorem 1 and Theorem 2, the two-step estimators can achieve the same convergence rate as the counterparts in Theorem 1 and Theorem 2 of the Supplementary Material.

4. Numerical Studies

4.1 Simulation Experiments

We conduct simulation experiments to compare the proposed method with the eigenvalue ratio (ER) method suggested by Wang et al. (2019). Specifically, we simulate $\boldsymbol{Y}_t$'s from model (2.1), where the dimensions of the latent factor process $\boldsymbol{F}_t$ are chosen to be $r = c = 3$. The entries of $\boldsymbol{F}_t$ are independent and follow AR(1) processes, with errors generated from the standard normal distribution. That is:

$$F_{ij,t} = a_{ij} F_{ij,t-1} + \varepsilon_{ij,t}, \quad i = 1, \ldots, r; \quad j = 1, \ldots, c; \quad t = 1, \ldots, n,$$

where $a_{ij} = \omega_{ij} a$, with ω_{ij} being i.i.d. random variables such that $P(\omega_{ij} = \pm 1) = 0.5$. We consider three scenarios for a:

(1) Non-correlations: $a = 0$;

(2) Weak correlations: $a = 0.1$;

(3) Moderate correlations: $a = 0.5$;

(4) Strong correlations: $a = 0.9$.

The entries of $\boldsymbol{R}$ and $\boldsymbol{C}$ are independently sampled from the uniform distribution on $(-1, 1)$, and scaled by $p^{\delta/2}$ and $q^{\omega/2}$, respectively. Following Wang et al. (2019), we consider the idiosyncratic errors $\boldsymbol{E}_t = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\varepsilon}_t\boldsymbol{\Sigma}_2^{1/2}$, where $\boldsymbol{\varepsilon}_t$ follows a normal distribution with $\mathbb{E}(\boldsymbol{\varepsilon}_t) = 0$ and $\mathrm{Cov}(\mathrm{vec}(\boldsymbol{\varepsilon}_t)) = \boldsymbol{I}_{pq}$.

We explore the following two cases for the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$:

(1) $\boldsymbol{\Sigma}_1 = \boldsymbol{I}_p$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{I}_q$;

(2) $\boldsymbol{\Sigma}_1 = (\boldsymbol{\sigma}_{1,ij})$, where $\sigma_{1,ij} = 0.1$ for $i \neq j$ and $\sigma_{1,ii} = 1$, and $\boldsymbol{\Sigma}_2 = (\boldsymbol{\sigma}_{2,ij})$, where $\sigma_{2,ij} = 0.1$ for $i \neq j$ and $\sigma_{2,ii} = 1$.

We consider three combinations of (δ, ω), namely $(0.5, 0.5)$, $(0.5, 0)$, and $(0, 0)$. For each combination of δ and ω, the dimensions (p, q) are chosen as $(20, 20)$, $(20, 40)$, and $(40, 40)$. The sample size n is set to either 200 or 800. Since $K \geq 2$, the proposed estimators perform well, and we choose $K = 3$ for the simulations and one application in this paper. The simulation results are based on 200 Monte Carlo replications.

We report the relative frequencies of occurrences where $\hat{r} = r$ or $\hat{c} = c$ (denoted by x), the frequency of $\hat{r} < r$ or $\hat{c} < c$ (denoted by y), and the frequency of $\hat{r} > r$ or $\hat{c} > c$ (denoted by z) among the 200 replications in Tables 1 and 2, and Tables 1-6 in the Supplementary Material, where the results are presented in the form $x(y|z)$. Here, the estimators $\widehat{r}_{ER_o}$, $\widehat{r}_{SR_o}$, $\widehat{r}_{MR_o}$, $\widehat{c}_{ER_o}$, $\widehat{c}_{SR_o}$, $\widehat{c}_{MR_o}$ are based on model (2.1) directly, as detailed in Subsections 1.1 and 1.2 of the Supplementary Material, respectively.

Table 1, and Tables 1-3 in the Supplementary Material show the case where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are diagonal. For the estimators based on model (2.1), the estimated results demonstrate that the performance of the proposed approaches is excellent. The SR_o estimator is the most powerful method for determining the number of factors in both rows and columns among the three estimators, while the ER_o estimator is less effective in this scenario. Moreover, all estimators based on the transformed model (3.2) perform better than their counterparts based on model (2.1), especially when the latent factors exhibit weak to moderate auto-correlations. They are also better for weak factors but comparable with each other under strong factors, i.e., $(\delta, \omega) = (0, 0)$. Both SR and MR estimators consistently yield better results than the ER estimator.

When the off diagonal elements in $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are small but equal, this case illustrates that the elements of $\boldsymbol{E}_t$ are weakly correlated. Based on model (2.1), Table 2, and Tables 4-6 in the Supplementary Material show that the ER_o estimator performs poorly, and the estimators SR_o and MR_o do not perform satisfactorily under weak factors and small correlations. However, based on the transformed model (3.2), the proposed estimators achieve significantly better results than those based on (2.1) and the ER estimator. Specifically, the SR estimator consistently outperforms the others, with its improved power being most notable. The estimators based on (3.1) also perform better their counterparts based on (2.1) under the same autocorrelation and strength pattern of the latent factors as those in Tables **??**-**??**. At the same time, the MR estimator clearly outperforms the ER estimator.

In summary, Tables 1-2, and Tables 1-6 in the Supplementary Material indicate that the performance of all estimators improves significantly as the strength or serial correlations of the latent factors increase, and none of the estimators is sensitive to the sample size. Furthermore, the proposed methods are more powerful than the ER method in all cases based on the transformed models. Comparing with the MR estimator, the SR estimator is more robust.

4.2 Real Data Analysis

We employ the matrix factor model to analyze a ten-by-ten return series comprising market equity (ME) and the Book-to-Market ratio (BE/ME). The dataset consists of 100 portfolios, constructed by intersecting ten portfolios based on ME and ten portfolios based on BE/ME. The analysis is performed on the daily returns of these portfolios, collected from January 6, 2021, to June 28, 2024, covering a total of 875 days. For further details, please refer to *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*.

Firstly, we utilize the ER_o, SR_o, and MR_o methods to estimate the numbers of row and column factors in the matrix-variate factor model. As illustrated in Figure 1, the ER_o estimator suggests one row factor and two column factors (i.e., $r = 1$, $c = 2$); the SR_o method indicates one row factor and one column factor (i.e., $r = 1$, $c = 1$); and the MR_o approach implies one row factor and four column factors (i.e., $r = 1$, $c = 4$). Next, based on the transformed matrix factor model, we apply the ER, SR, and MR methods to determine the numbers of row and column factors. As shown in Figure 2, the ER estimator yields the same results as before, i.e., $r = 1$, $c = 2$, while both the MR and SR methods indicate one row factor and four column factors (i.e., $r = 1$, $c = 4$).

To compare models with different numbers of factors, we use L-fold cross-validation. Specifically, we divide the entire dataset D into L subsamples, $D_1, \ldots, D_L$, and fit a factor model using each of the D_{-l} samples to obtain $\widehat{\boldsymbol{R}}_{-l}$ and $\widehat{\boldsymbol{C}}_{-l}$, where D_{-l} denotes the dataset with the l-th subsample removed. For each data subset D_l, we define the dynamic signal part of the factor model as $\boldsymbol{S}_{t,l} = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}^\top$. The estimated loading spaces are then used to obtain $\widehat{\boldsymbol{S}}_{t,l} = \widehat{\boldsymbol{R}}_{-l}\widehat{\boldsymbol{R}}_{-l}^\top\boldsymbol{Y}_t\widehat{\boldsymbol{C}}_{-l}\widehat{\boldsymbol{C}}_{-l}^\top$ for each $l = 1, \ldots, L$.

Next, we compute the out-of-sample residuals $\boldsymbol{Y}_t - \widehat{\boldsymbol{S}}_{t,l}$ and calculate the adjusted residual sum of squares (RSS) over the L folds for comparison:

$$RSS = \frac{1}{pq}\sum_{l=1}^{L}\sum_{t\in D_l}\|\boldsymbol{Y}_t - \widehat{\boldsymbol{S}}_{t,l}\|_F.$$

Here, we set $L = 5$ and ensure that the time length of each dataset D_l is 175. We compute value of RSS for the matrix factor model under two different scenarios: when $r = 1$ and $c = 2$, the value of RSS is 6.673, while when $r = 1$ and $c = 4$, the value of RSS is 6.455. These results indicate that both SR and MR estimators perform better than ER estimator, suggesting that incorporating the appropriate number of factors in the matrix factor model leads to more accurate out-of-sample predictions.

5. Conclusion

In this paper, we proposed two new ratio-based estimators for determining the number of row and column factors in two-way high-dimensional matrix-variate time series factor models.

These estimators were developed based on the element-wise maximum norm and the Frobenius norm of the sample auto-covariance matrices of the time series. To mitigate the influence of the strength of the factors between rows and columns, we transformed the matrix factor model into a new one only including a row or column loading matrix. The improved ratio-based methods, based on this transformed model, were then investigated.

We analyzed the theoretical properties of the proposed approaches under regularity conditions and demonstrated that they provide consistent estimators for the number of factors. Through Monte Carlo simulations and a real data analysis, we evaluated the finite-sample performance of the proposed methods and compared them with the traditional eigenvalue-based ratio approach. Our results show that the proposed estimators outperform the classical method, offering significant improvements in estimation accuracy.

In summary, we believe that the methods introduced in this paper represent a valuable contribution to the toolbox for high-dimensional matrix-variate time series factor models, providing more reliable and accurate estimation techniques in practice.

Acknowledgments

The authors are grateful to the Co-Editor, the Associate Editor, and two anonymous referees for their insightful comments and constructive suggestions, which significantly improved the quality of this work. We also thank Prof. Lixing Zhu, Prof. Xianyang Zhang, and Dr. Zhaoxing Gao for their valuable feedback on an earlier version of the manuscript. This research was supported in part by the National Natural Science Foundation of China (No. 12171161) and the National Statistical Scientific Research Projects (No. 2025LZ029).

References

Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica* **71**: 135–171.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension, *The Annals of Statistics* **40**: 436–465.

Chang, J., Guo, B. and Yao, Q. (2018). Principal component analysis for second-order stationary vector time series, *The Annals of Statistics* **46**: 2094–2124.

Chang, J., Yao, Q. and Zhou, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations, *Biometrika* **104**(1): 111–127.

Chen, E. Y. and Fan, J. (2023). Statistical inference for high-dimensional matrix-variate factor model, *Journal of the American Statistical Association* **118**: 1038–1055.

Chen, E. Y., Tsay, R. S. and Chen, R. (2020). Constrained factor models for high-dimensional matrix-variate time series, *Journal of the American Statistical Association* **115**: 775–793.

Fan, J., Guo, J. and Zheng, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding, *Journal of the American Statistical Association* **117**: 852–61.

Filipiak, K., Klein, D. and Vojtkova, E. (2018). The properties of partial trace and block trace operators of partitioned matrices, *Electronic Journal of Linear Algebra* **33**: 3–15.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). The generalized dynamic-factor model: one-sided estimation and forecasting, *Journal of the American Statistical Association* **100**: 830–840.

He, Y., Trapani, L., Kong, X. and Yu, L. (2023). One-way or two-way factor model for matrix sequences?, *Journal of Econometrics* **235**: 1981–2004.

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

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

Li, Z., Lam, C., Yao, J. and Yao, Q. (2019). On testing for high-dimensional white noise, *The Annals of Statistics* **47**(6): 3382–3412.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors, *Biometrika* **95**: 356–379.

Wang, D., Liu, X. and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series, *Journal of Econometrics* **208**: 231–248.

Yu, L., He, Y., Kong, X. and Zhang, X. (2022). Projected estimation for large-dimensional matrix factor models, *Journal of Econometrics* **229**: 201–217.

Table 1: Relative frequency estimates for $P(\hat{r}=3)$ and $P(\hat{c}=3)$ with 200 replicate samples for the AR coefficients with case (2) and $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ with case (1)

(δ,ω)	$(0.5,0.5)$						
(p,q)	n	$\widehat{r}_{ER_o}$	$\widehat{r}_{SR_o}$	$\widehat{r}_{MR_o}$	$\widehat{c}_{ER_o}$	$\widehat{c}_{SR_o}$	$\widehat{c}_{MR_o}$
(20,20)	200	0.360(128|0)	0.640(70|2)	0.415(114|3)	0.340(130|2)	0.640(70|2)	0.370(124|2)
	800	0.445(110|1)	0.665(88|3)	0.475(103|2)	0.465(106|1)	0.685(62|1)	0.445(111|0)
(20,40)	200	0.355(170|1)	0.480(103|1)	0.345(131|0)	0.470(106|0)	0.910(18|0)	0.580(83|1)
	800	0.420(114|2)	0.515(97|0)	0.380(123|1)	0.465(107|0)	0.860(26|2)	0.570(86|0)
(40,40)	200	0.430(113|1)	0.770(46|0)	0.535(92|1)	0.390(122|0)	0.790(42|0)	0.565(87|0)
	800	0.485(103|0)	0.760(48|0)	0.520(96|0)	0.415(117|0)	0.755(49|0)	0.500(100|0)
(p,q)	n	$\widehat{r}_{ER}$	$\widehat{r}_{SR}$	$\widehat{r}_{MR}$	$\widehat{c}_{ER}$	$\widehat{c}_{SR}$	$\widehat{c}_{MR}$
(20,20)	200	0.445(109|2)	0.885(23|0)	0.570(85|1)	0.410(112|6)	0.800(35|5)	0.520(94|2)
	800	0.545(88|3)	0.875(21|4)	0.560(84|4)	0.495(96|5)	0.895(19|2)	0.630(71|2)
(20,40)	200	0.555(87|2)	0.825(34|1)	0.570(86|0)	0.570(86|0)	0.970(6|0)	0.745(50|1)
	800	0.545(91|0)	0.855(28|1)	0.595(81|0)	0.545(88|3)	0.955(5|4)	0.715(56|1)
(40,40)	200	0.520(95|1)	0.935(13|0)	0.720(56|0)	0.510(98|0)	0.940(11|1)	0.690(61|15)
	800	0.575(85|0)	0.970(5|1)	0.705(59|0)	0.485(102|1)	0.965(7|0)	0.690(62|0)
(δ,ω)	$(0,0.5)$						
(p,q)	n	$\widehat{r}_{ER_o}$	$\widehat{r}_{SR_o}$	$\widehat{r}_{MR_o}$	$\widehat{c}_{ER_o}$	$\widehat{c}_{SR_o}$	$\widehat{c}_{MR_o}$
(20,20)	200	0.750(48|2)	0.950(9|0)	0.845(29|2)	0.795(40|1)	0.980(4|0)	0.890(22|0)
	800	0.770(44|2)	0.965(7|0)	0.875(25|0)	0.805(38|1)	0.985(3|0)	0.875(24|1)
(20,40)	200	0.770(46|0)	0.925(15|0)	0.895(21|0)	0.720(56|0)	1.000(0|0)	0.885(23|0)
	800	0.750(50|0)	0.920(16|0)	0.870(26|0)	0.735(53|0)	1.000(0|0)	0.915(17|0)
(40,40)	200	0.830(34|0)	0.990(2|0)	0.920(16|0)	0.820(36|0)	0.995(1|0)	0.930(14|05)
	800	0.840(32|0)	0.990(2|0)	0.935(13|0)	0.820(36|0)	0.990(2|0)	0.920(16|0)
(p,q)	n	$\widehat{r}_{ER}$	$\widehat{r}_{SR}$	$\widehat{r}_{MR}$	$\widehat{c}_{ER}$	$\widehat{c}_{SR}$	$\widehat{c}_{MR}$
(20,20)	200	0.755(45|4)	0.980(2|2)	0.895(17|4)	0.800(39|1)	1.000(0|0)	0.910(16|2)
	800	0.775(43|2)	0.990(2|0)	0.900(19|1)	0.815(35|2)	0.990(2|0)	0.900(19|1)
(20,40)	200	0.790(42|0)	0.980(4|0)	0.890(22|0)	0.715(57|0)	1.000(0|0)	0.905(17|2)
	800	0.755(49|0)	0.970(6|0)	0.895(21|0)	0.755(46|3)	1.000(0|0)	0.910(18|0)
(40,40)	200	0.830(34|0)	1.000(0|0)	0.960(16|0)	0.825(35|0)	1.000(0|0)	0.930(13|1)
	800	0.855(29|0)	1.000(0|0)	0.940(12|0)	0.840(32|0)	0.995(1|0)	0.925(15|0)
(δ,ω)	$(0,0)$						
(p,q)	n	$\widehat{r}_{ER_o}$	$\widehat{r}_{SR_o}$	$\widehat{r}_{MR_o}$	$\widehat{c}_{ER_o}$	$\widehat{c}_{SR_o}$	$\widehat{c}_{MR_o}$
(20,20)	200	0.910(18|0)	1.000(0|0)	0.985(3|0)	0.910(18|0)	1.000(0|0)	0.980(2|0)
	800	0.900(19|1)	0.990(2|0)	0.970(6|0)	0.905(19|0)	0.985(3|0)	0.975(5|0)
(20,40)	200	0.965(7|0)	1.000(0|0)	0.985(3|0)	0.910(18|0)	1.000(0|0)	0.970(6|0)
	800	0.940(12|0)	1.000(0|0)	0.985(3|0)	0.915(16|1)	1.000(0|0)	0.965(5|2)
(40,40)	200	0.925(15|0)	1.000(0|0)	0.965(6|1)	0.955(9|0)	1.000(0|0)	0.985(3|0)
	800	0.960(8|0)	1.000(0|0)	0.985(3|0)	0.935(13|0)	1.000(0|0)	0.970(4|2)
(p,q)	n	$\widehat{r}_{ER}$	$\widehat{r}_{SR}$	$\widehat{r}_{MR}$	$\widehat{c}_{ER}$	$\widehat{c}_{SR}$	$\widehat{c}_{MR}$
(20,20)	200	0.910(13|5)	1.000(0|0)	0.990(2|0)	0.910(17|1)	1.000(0|0)	0.980(2|0)
	800	0.900(18|1)	0.990(2|0)	0.970(5|1)	0.900(18|2)	0.995(1|0)	0.975(5|0)
(20,40)	200	0.965(6|1)	1.000(0|0)	0.990(2|0)	0.910(17|1)	1.000(0|0)	0.965(5|2)
	800	0.945(10|1)	1.000(0|0)	0.995(1|0)	0.915(15|2)	1.000(0|0)	0.955(5|4)
(40,40)	200	0.925(15|0)	1.000(0|0)	0.965(6|1)	0.955(9|0)	1.000(0|0)	0.990(2|0)
	800	0.960(7|1)	1.000(0|0)	0.995(1|0)	0.935(13|0)	1.000(0|0)	0.965(3|4)

Table 2: Relative frequency estimates for $P(\hat{r}=3)$ and $P(\hat{c}=3)$ with 200 replicate samples for the AR coefficients with case (2) and $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ with case (2)

(δ,ω)	$(0.5, 0.5)$						
(p,q)	n	$\widehat{r}_{ER_o}$	$\widehat{r}_{SR_o}$	$\widehat{r}_{MR_o}$	$\widehat{c}_{ER_o}$	$\widehat{c}_{SR_o}$	$\widehat{c}_{MR_o}$
(20,20)	200	0.335(132|1)	0.540(85|7)	0.330(130|4)	0.360(125|3)	0.540(85|7)	0.360(127|1)
	800	0.330(132|2)	0.660(81|7)	0.365(127|0)	0.395(120|1)	0.635(69|4)	0.465(106|1)
(20,40)	200	0.335(131|2)	0.395(118|3)	0.375(124|1)	0.375(125|0)	0.600(68|12)	0.455(106|3)
	800	0.365(126|1)	0.370(122|4)	0.365(124|3)	0.435(111|2)	0.660(51|17)	0.420(111|5)
(40,40)	200	0.360(125|3)	0.310(118|20)	0.315(119|20)	0.345(129|2)	0.325(103|32)	0.325(132|3)
	800	0.335(129|4)	0.370(107|19)	0.410(111|7)	0.345(129|2)	0.330(107|27)	0.330(123|11)
(p,q)	n	$\widehat{r}_{ER}$	$\widehat{r}_{SR}$	$\widehat{r}_{MR}$	$\widehat{c}_{ER}$	$\widehat{c}_{SR}$	$\widehat{c}_{MR}$
(20,20)	200	0.425(111|4)	0.820(32|4)	0.535(89|4)	0.480(101|3)	0.790(36|6)	0.570(83|3)
	800	0.440(108|4)	0.825(31|4)	0.550(88|2)	0.510(97|1)	0.855(28|1)	0.550(82|8)
(20,40)	200	0.445(109|2)	0.690(60|2)	0.510(96|2)	0.520(95|1)	0.890(18|4)	0.670(65|1)
	800	0.460(108|0)	0.655(69|0)	0.490(101|0)	0.565(86|1)	0.930(9|5)	0.675(65|0)
(40,40)	200	0.470(102|4)	0.635(59|14)	0.480(102|2)	0.510(98|0)	0.680(51|13)	0.510(95|3)
	800	0.510(96|2)	0.740(44|8)	0.620(71|5)	0.455(126|3)	0.760(41|7)	0.500(95|5)
(δ,ω)	$(0, 0.5)$						
(p,q)	n	$\widehat{r}_{ER_o}$	$\widehat{r}_{SR_o}$	$\widehat{r}_{MR_o}$	$\widehat{c}_{ER_o}$	$\widehat{c}_{SR_o}$	$\widehat{c}_{MR_o}$
(20,20)	200	0.725(55|0)	0.925(14|1)	0.835(32|1)	0.735(53|0)	0.955(9|0)	0.855(29|0)
	800	0.785(43|0)	0.965(6|1)	0.885(23|0)	0.765(47|0)	0.945(11|0)	0.835(33|0)
(20,40)	200	0.820(36|0)	0.930(14|0)	0.840(32|0)	0.725(55|0)	0.995(0|1)	0.845(31|0)
	800	0.835(33|0)	0.960(8|0)	0.885(23|0)	0.710(58|0)	0.970(6|0)	0.885(23|0)
(40,40)	200	0.815(37|0)	0.975(5|0)	0.915(16|1)	0.825(35|0)	0.985(3|0)	0.895(21|0)
	800	0.885(23|0)	0.990(0|0)	0.890(22|0)	0.820(35|0)	0.985(3|0)	0.925(15|0)
(p,q)	n	$\widehat{r}_{ER}$	$\widehat{r}_{SR}$	$\widehat{r}_{MR}$	$\widehat{c}_{ER}$	$\widehat{c}_{SR}$	$\widehat{c}_{MR}$
(20,20)	200	0.725(53|2)	0.960(7|1)	0.840(28|4)	0.740(52|0)	0.990(2|0)	0.910(18|0)
	800	0.785(41|2)	0.990(1|1)	0.915(15|2)	0.785(41|2)	0.955(9|0)	0.880(24|0)
(20,40)	200	0.830(34|0)	0.960(8|0)	0.900(20|0)	0.730(53|1)	0.995(1|0)	0.885(22|1)
	800	0.835(33|0)	0.980(4|0)	0.910(17|1)	0.725(53|2)	0.985(2|1)	0.890(21|1)
(40,40)	200	0.835(33|0)	1.000(0|0)	0.915(15|2)	0.830(33|1)	0.995(1|0)	0.905(19|0)
	800	0.880(21|3)	1.000(0|0)	0.905(19|0)	0.820(35|0)	1.000(0|0)	0.940(12|0)
(δ,ω)	$(0, 0)$						
(p,q)	n	$\widehat{r}_{ER_o}$	$\widehat{r}_{SR_o}$	$\widehat{r}_{MR_o}$	$\widehat{c}_{ER_o}$	$\widehat{c}_{SR_o}$	$\widehat{c}_{MR_o}$
(20,20)	200	0.890(21|1)	0.995(1|0)	0.945(11|0)	0.910(16|2)	0.980(4|0)	0.930(14|0)
	800	0.925(14|1)	0.990(2|0)	0.960(6|2)	0.920(15|1)	0.995(1|0)	0.980(4|0)
(20,40)	200	0.970(6|0)	0.990(2|0)	0.990(2|0)	0.950(10|0)	1.000(0|0)	0.965(4|3)
	800	0.935(13|0)	0.985(3|0)	0.980(4|0)	0.935(12|1)	1.000(0|0)	0.955(7|2)
(40,40)	200	0.980(4|0)	0.995(1|0)	0.990(2|0)	0.945(11|0)	1.000(0|0)	0.990(0|2)
	800	0.960(8|0)	0.995(1|0)	0.995(1|0)	0.940(12|0)	1.000(0|0)	0.995(1|0)
(p,q)	n	$\widehat{r}_{ER}$	$\widehat{r}_{SR}$	$\widehat{r}_{MR}$	$\widehat{c}_{ER}$	$\widehat{c}_{SR}$	$\widehat{c}_{MR}$
(20,20)	200	0.890(19|3)	1.000(0|0)	0.970(6|0)	0.910(15|3)	0.990(2|0)	0.950(10|0)
	800	0.925(13|2)	0.995(1|0)	0.995(1|0)	0.920(14|2)	0.995(1|0)	0.985(3|0)
(20,40)	200	0.970(6|0)	0.995(1|0)	0.995(1|0)	0.950(9|1)	1.000(0|0)	0.960(5|3)
	800	0.940(11|1)	0.995(1|0)	0.980(4|0)	0.930(11|3)	1.000(0|0)	0.960(6|2)
(40,40)	200	0.975(3|2)	0.995(1|0)	0.990(2|0)	0.945(11|0)	1.000(0|0)	0.990(0|2)
	800	0.960(8|0)	0.995(1|0)	0.995(1|0)	0.940(12|0)	1.000(0|0)	0.995(1|0)

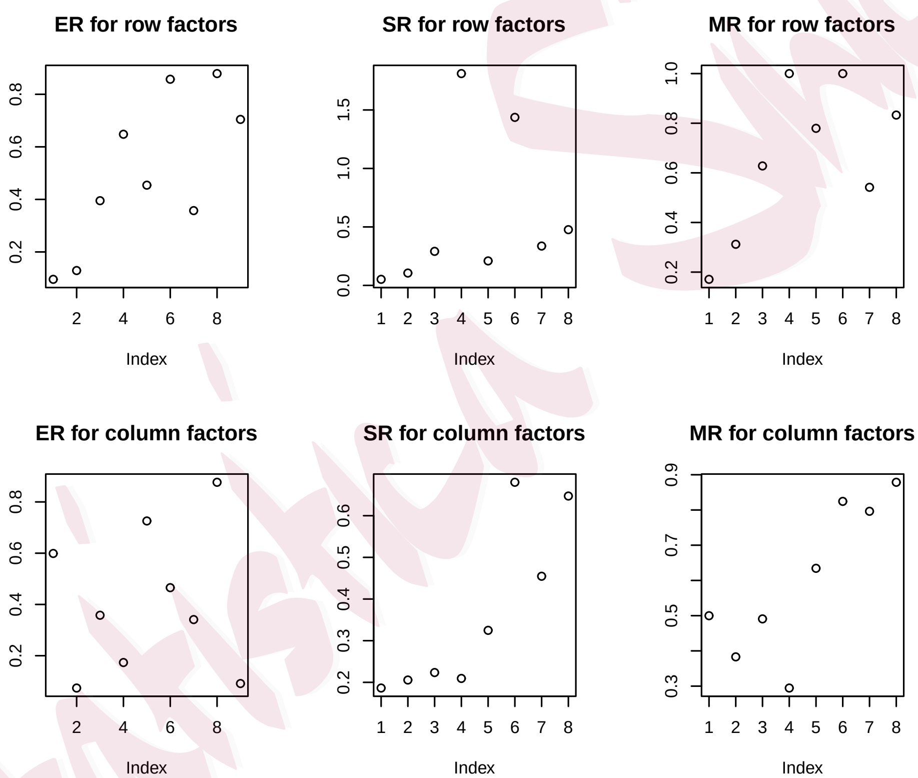

Figure 1: The number of row and column factors is determined by $\widehat{r}_{ER_o}$ and $\widehat{c}_{ER_o}$, $\widehat{r}_{SR_o}$ and $\widehat{c}_{SR_o}$, and $\widehat{r}_{MR_o}$ and $\widehat{c}_{MR_o}$, estimators for the real data set under the model (2.1)

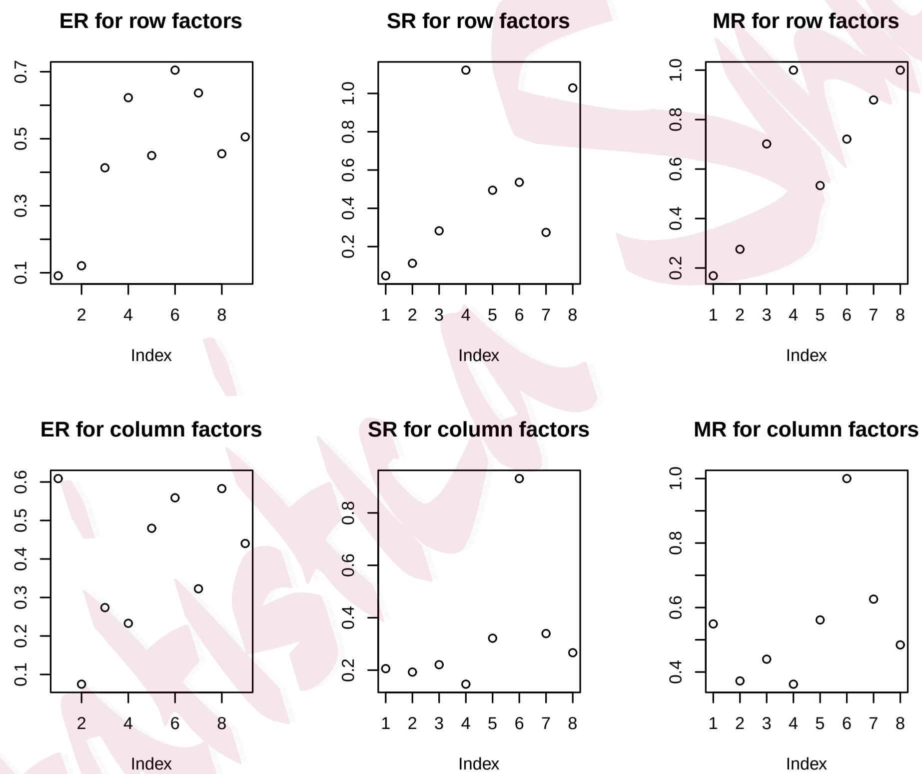

Figure 2: The number of row and column factors is determined by $\widehat{r}_{ER}$ and $\widehat{c}_{ER}$, $\widehat{r}_{SR}$ and $\widehat{c}_{SR}$, and $\widehat{r}_{MR}$ and $\widehat{c}_{MR}$ estimators for the real data set under the transformed model