

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*}

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

Abstract

In this paper, we study a new two-way factor model for high-dimensional matrix-variate time series. To estimate the number of factors in this two-way factor model, we decompose the series into two parts: one being a non-weakly correlated series and the other being a weakly correlated noise. By comparing the difference between two series, we can construct white-noise based signal statistics to determine the number of factors in row loading matrix (column loading matrix). Furthermore, to mitigate the negative impact on the accuracy of the estimation, which is caused by the interaction between the row loading matrix and the column loading matrix, we propose a transformation so that the transformed model only contains the row loading matrix (column loading matrix). We define sequences of ratios of two test statistics as signal statistics to determine the number of factors and derive the consistence of the estimation. We implement the numerical studies to examine the performance of the new methods.

Key words: Matrix-variate time series, Max-type test, Sum-type test, Two-way factor models

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 author is grateful to Prof. Lixing Zhu and Prof. Xianyang Zhang for many helpful comments and suggestions on an earlier version. First version:2024-1-18; This version: 2024-12-1.

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, Peña and Box (1987), Forni et al. (2000), Bai and Ng (2002), Stock and Watson (2002), Bai (2003), Forni et al. (2005), Hallin and Liska (2007), Pan and Yao (2008), Tsai and Tsay (2010), Lam et al. (2011), Lam and Yao (2012), Ahn and Horenstein (2013), Fan et al. (2013), Chang et al. (2015), and Fan et al. (2022).

In some scientific research fields such as economics, meteorology, and ecology, high-dimensional matrix-variate time series have gained attention. Despite their growing prevalence, the development of statistical methods for analyzing large-scale matrix-form data remains less explored. To achieve the model parsimony, Wang et al. (2019) introduced matrix-variate factor models that preserve the matrix structure of the data, enabling dimension reduction in both row and column directions. Chen et al. (2020) further generalized these models by incorporating domain or prior knowledge through linear constraints. In contrast to the white noise idiosyncratic error matrices studied in these two papers, Yu et al. (2022), Chen and Fan (2023), and He et al. (2023) studied matrix-valued factor models with correlated idiosyncratic errors, where the correlations may be weak between rows, columns, or 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. Otherwise, autocovariance is only a selection to infer the two-way factor model for the non-weakly correlated scenario between the elements of the noise matrix. Meanwhile, continuing the line of Wang et al. (2019), the idiosyncratic errors can be weakly correlated.

Wang et al. (2019) suggested the eigenvalue-ratio (ER) technique to determine the number of row and column factors. However, ER-based estimation is not consistent in theory, and our numerical studies also showed that the estimation is often less accurate, particularly when the matrix-variate time series exhibits weak serial correlations or contains weak factors. As we know, estimation inaccuracy can significantly affect downstream analysis.

Therefore, we propose a novel method to estimate the number of factors in row and column with more accuracy. Taking the factor model of Wang et al. (2019) as an example, we do not directly work on the original time series, while relying on the decomposition of matrix-variate time series: a dynamic component driven by a lower-dimensional matrix-variate time series (the factor process) and a static component consisting of matrix-variate 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. That is, we can transfer the problem for models with serial correlation to the problem for models without serial correlation and the estimation could be easier. This decomposition facilitates both model identification and statistical inference. 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. Let \mathbf{R} be the semi-orthogonal row-loading matrix, and \mathbf{R}^\dagger its orthogonal complement. The matrix-variate time series projected along the columns of \mathbf{R} forms a non-white noise sequence, while its projection along the columns of \mathbf{R}^\dagger results in a white noise process. See (A.1) and (A.2) in the Appendix for details. These observations are of importance because we can determine the number of factors through efficiently examining whether 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 identify 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 Appendix.

Throughout the paper, let \mathbf{I}_p denote the $p \times p$ identity matrix, and A^\top denote the transpose of a matrix $A = (a_{ij})$. Define $|A|_\infty = \max_{i,j} |a_{ij}|$. Let $\|A\|_F$, $\|A\|_2$, 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 $1 \leq i \leq m$ 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 $\{\mathbf{Y}_t : t = 1, \dots, n\}$ represent n observations from a matrix-variate time series with $\mathbb{E}\mathbf{Y}_t = \mathbf{0}$, where \mathbf{Y}_t is a $p \times q$ matrix defined as

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

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

$$\mathbf{Y}_t = \mathbf{R}\mathbf{F}_t\mathbf{C}^\top + \mathbf{E}_t, \quad t = 1, \dots, n, \quad (2.1)$$

where $\mathbf{F}_t = (F_{1,t}, \dots, F_{c,t}) = (F_{1,t}, \dots, F_{r,t})^\top$ is an unobserved $r \times c$ matrix of common fundamental factors; \mathbf{R} is a $p \times r$ row loading matrix; \mathbf{C} is a $q \times c$ column loading matrix; and $\mathbf{E}_t = (E_{1,t}, \dots, E_{q,t}) = (E_{1,t}, \dots, E_{p,t})^\top$ is a $p \times q$ matrix of random errors. In Model (2.1), the common fundamental factors \mathbf{F}_t capture the dynamics and co-movements of \mathbf{Y}_t , while \mathbf{R} and \mathbf{C} reflect the significance of the factors and their interactions. Model (2.1) can also be rewritten as a classical factor model:

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

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

For the vector-valued factor model (3.1), the factors $\text{vec}(\mathbf{F}_t)$ and the loading matrix $\mathbf{C} \otimes \mathbf{R}$ 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 $(\mathbf{R}, \mathbf{F}_t, \mathbf{C})$ and $(\mathbf{R}H_1, H_1^{-1}\mathbf{F}_tH_2^{-1}, \mathbf{C}H_2^\top)$ are equivalent, implying that the factors and loadings are identifiable up to invertible transformations H_1 and H_2 . However, once \mathbf{R} and \mathbf{C} are specified, the factor process \mathbf{F}_t is uniquely determined. This non-uniqueness in \mathbf{R} and \mathbf{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 $\boldsymbol{\Sigma}_f(h) = \text{Cov}(\text{vec}(\mathbf{F}_t), \text{vec}(\mathbf{F}_{t+h}))$, $\boldsymbol{\Sigma}_{ef}(h) = \text{Cov}(\text{vec}(\mathbf{E}_t), \text{vec}(\mathbf{F}_{t+h}))$, $\boldsymbol{\Sigma}_{fe}(h) = \text{Cov}(\text{vec}(\mathbf{F}_t), \text{vec}(\mathbf{E}_{t+h}))$, $\boldsymbol{\Sigma}_e(h) = \text{Cov}(\text{vec}(\mathbf{E}_t), \text{vec}(\mathbf{E}_{t+h}))$, $\boldsymbol{\Sigma}_e = \text{Cov}(\text{vec}(\mathbf{E}_t))$.

Assumption 2.1. (C1) There exist constants $\delta, \omega \in [0, 1]$ such that $\|\mathbf{R}\|_2^2 \asymp p^{1-\delta} \asymp \|\mathbf{R}\|_{\min}^2$

and $\|\mathbf{C}\|_2^2 \asymp q^{1-\omega} \asymp \|\mathbf{C}\|_{\min}^2$.

(C2) The vector-valued process $(\text{vec}^\top(\mathbf{F}_t), \text{vec}^\top(\mathbf{E}_t))^\top$, 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 $\{(\text{vec}^\top(\mathbf{F}_t), \text{vec}^\top(\mathbf{E}_t))^\top : i \leq t \leq j\}$.

(C3) Let $F_{t,ij}$ be the (i, j) -th entry of \mathbf{F}_t . For all $i = 1, \dots, r$, $j = 1, \dots, c$, and $t = 1, \dots, n$, assume $\mathbb{E}(|F_{t,ij}|^{2\gamma}) \leq \mathcal{C}$ 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 $\boldsymbol{\Sigma}_f(h) \asymp \mathbf{O}(1) \asymp \sigma_l(\boldsymbol{\Sigma}_f(h))$ as $p, q \rightarrow \infty$ with r, c fixed.

(C4) $|\boldsymbol{\Sigma}_e(h)|_\infty = \mathbf{O}(n^{-1})$ for $h \geq 1$.

(C5) The elements of $\boldsymbol{\Sigma}_e$ are bounded as $n, p, q \rightarrow \infty$. Furthermore, for $h \geq 0$, $\|\boldsymbol{\Sigma}_{ef}(h)\|_2 = \mathbf{o}(p^{\frac{1-\delta}{2}} q^{\frac{1-\omega}{2}})$ 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 be weak correlation, which is similar to Chen and Fan (2023). 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 \mathbf{C} is known, the matrix-valued time series \mathbf{Y}_t can be projected to lower-dimensional spaces by setting

$$\mathbf{Y}_t \mathbf{C} / q^{1-\omega} = \mathbf{R} \mathbf{F}_t + \mathbf{E}_t \mathbf{C} / q^{1-\omega} := \mathbf{R} \mathbf{F}_t + \tilde{\boldsymbol{\xi}}_t. \quad (3.1)$$

It is easy to see that Model (3.1) reduces to a new one only including a row loading matrix. When the number of factors is unknown so that \mathbf{C} is unknown, we choose an appropriate matrix $\tilde{\mathbf{C}} = q^{\frac{1-\omega}{2}}(\tilde{Q}_1, \tilde{Q}_2, \dots, \tilde{Q}_m) := (C_{\cdot 1}, \dots, C_{\cdot m})$ with a large positive integer m such that

$$\mathbf{C}^\top \tilde{\mathbf{C}} = q^{\frac{1-\omega}{2}}(\mathbf{I}_c, \mathbf{0}),$$

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

$$\mathbf{X}_t = \mathbf{Y}_t \tilde{\mathbf{C}} / q^{1-\omega} = \mathbf{R} \mathbf{F}_t \mathbf{C}^\top \tilde{\mathbf{C}} / q^{1-\omega} + \mathbf{E}_t \tilde{\mathbf{C}} / q^{1-\omega} := \mathbf{R} \tilde{\mathbf{F}}_t + \tilde{\boldsymbol{\xi}}_t. \quad (3.2)$$

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

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

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

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

$$\mathbf{\Omega}_{\xi f}(h) = \frac{1}{q^{1-\omega}} \sum_{i=1}^c \mathbf{\Sigma}_{\xi f, ii}(h) = \frac{1}{(n-h)q^{1-\omega}} \sum_{t=1}^{n-h} \mathbb{E}(\mathbf{E}_t \mathbf{C} \mathbf{F}_{t+h}^\top).$$

By Assumption 2.1-(C5),

$$\mathbb{E}(\mathbf{R} \mathbf{F}_t \mathbf{C}^\top \tilde{\mathbf{C}} \tilde{\mathbf{C}}^\top \mathbf{E}_{t+h}^\top) = \mathbb{E}(\mathbf{R} \mathbf{F}_t \mathbf{C}^\top \mathbf{E}_{t+h}^\top) = \text{tr}_p \left(\mathbb{E} \left(\text{vec}(\mathbf{R} \mathbf{F}_t) \text{vec}^\top(\mathbf{E}_{t+h} \mathbf{C}) \right) \right) = 0,$$

and

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

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

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

By Assumption 3.1, the matrix $\widetilde{\mathbf{M}}_1$ has rank r . From Equation (3.3), we see that each column of $\widetilde{\mathbf{M}}_1$ is a linear combination of the columns of \mathbf{R} , and thus the eigenspace of $\widetilde{\mathbf{M}}_1$ is the same as the eigenspace of \mathbf{R} , i.e., $\mathcal{M}(\widetilde{\mathbf{M}}_1) = \mathcal{M}(\mathbf{R})$. Therefore, $\mathcal{M}(\mathbf{R})$ can be estimated by the space spanned by the eigenvectors of the sample version of $\widetilde{\mathbf{M}}_1$. Let \mathbf{O}_j be the unit eigenvector corresponding to the j -th largest eigenvalue of $\widetilde{\mathbf{M}}_1$. Define $\mathbf{O} = (\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_r)$, and then obtain $\mathbf{R} = p^{\frac{1-\delta}{2}} \mathbf{O}$.

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

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

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

Remark 3.1. First, $\widehat{\mathbf{C}}$ can be estimated by (S.2) in Appendix. In fact, a small positive value m can also be chosen. Then $\mathbf{C}^\top \tilde{\mathbf{C}} = q^{\frac{1-\omega}{2}} (\mathbf{I}_m, \mathbf{0})^\top$. The result and the relative proof are similar. Thus, we do not discuss them further. Second, omitting to $q^{2-2\omega}$ in practice, we can consider that $\mathbf{\Omega}_x(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} \mathbb{E}(\mathbf{Y}_t \tilde{\mathbf{C}} \tilde{\mathbf{C}}^\top \mathbf{Y}_{t+h}^\top)$, and $\widetilde{\mathbf{M}}_1 = \sum_{h=1}^{h_0} \mathbf{\Omega}_x^\top(h) \mathbf{\Omega}_x(h)$ accordingly. Third,

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. Fourth, 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 $\|\Sigma_e\|_2 = o(p^{1-\delta}q^{1-\omega})$, 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 $\mathbf{O} = (O_1, \dots, O_r) = \mathbf{R}/p^{\frac{1-\delta}{2}}$ is semi-orthogonal, i.e., $\mathbf{O}^\top \mathbf{O} = \mathbf{I}_r$, where $O_i \in \mathbb{R}^{p \times 1}$ for $1 \leq i \leq r$. Let $\mathbf{O}^\dagger = (O_{r+1}, \dots, O_p)$ be the orthogonal complement matrix of \mathbf{O} , so that $(\mathbf{O}, \mathbf{O}^\dagger)^\top (\mathbf{O}, \mathbf{O}^\dagger) = \mathbf{I}_p$. For a better understanding of our approach, we assume that \mathbf{E}_t is white noise firstly. Then we have:

$$O_i^\top \mathbf{X}_t = O_i^\top \mathbf{R} \mathbf{F}_t \tilde{\mathbf{C}} / q^{\frac{1-\omega}{2}} + O_i^\top \mathbf{E}_t \tilde{\mathbf{C}} / q^{\frac{1-\omega}{2}}, \quad i = 1, \dots, r; \quad (3.5)$$

$$O_i^\top \mathbf{X}_t = O_i^\top \mathbf{E}_t \tilde{\mathbf{C}} / q^{\frac{1-\omega}{2}}, \quad i = r+1, \dots, p. \quad (3.6)$$

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

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

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

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

$$\check{\Gamma}_i(h) := \left(\check{\gamma}_{i,kl}(h) \right) = \mathfrak{R}_i^\top \Sigma_X(h) \mathfrak{R}_i. \quad (3.7)$$

Referring to Chang et al. (2017) and Li et al. (2019), the max-type and sum-type tests can

be proposed below to identify whether the sequence $\{\mathfrak{R}_i^\top \mathbf{X}_t : t = 1, 2, \dots\}$ is white noise or not. The max-type test is constructed based on the maximum norm of the auto-covariance matrix:

$$\check{T}_{i,n} := \max_{1 \leq h \leq K} n^{1/2} \left| \check{\Gamma}_i(h) \right|_\infty = \max_{1 \leq h \leq K} n^{1/2} \left| \mathfrak{R}_i^\top \Sigma_X(h) \mathfrak{R}_i \right|_\infty, \quad (3.8)$$

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

$$\check{G}_{i,n} := \sum_{1 \leq h \leq K} \text{Tr} \left(\check{\Gamma}_i(h)^\top \check{\Gamma}_i(h) \right) = \sum_{1 \leq h \leq K} \left\| \check{\Gamma}_i(h) \right\|_F^2 = \sum_{1 \leq h \leq K} \sum_j \check{\sigma}_j^2 \left(\check{\Gamma}_i(h) \right), \quad (3.9)$$

where Tr denotes the trace operation for a square matrix and $\check{\sigma}_j^2$ denotes the j -th largest singular value of $\check{\Gamma}_i(h)$. Large values of $\check{T}_{i,n}$ and $\check{G}_{i,n}$ indicate a potential departure from the assumption of white noise in the sequence $\{\mathfrak{R}_i^\top \mathbf{X}_t\}$. 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 $\check{T}_{i,n}$ and $\check{T}_{i+1,n}$ is likely to reach its maximum value, as $\{\mathfrak{R}_{r+1}^\top \mathbf{X}_t\}$ is a white noise sequence, while $\{\mathfrak{R}_r^\top \mathbf{X}_t\}$ is a non-white noise sequence. When \mathbf{E}_t is weakly correlated, the same phenomenon can be observed. It is merely that $\check{T}_{i,n} = \check{G}_{i,n} = 0$ under the assumption of white noise for \mathbf{E}_t , while $\check{T}_{i,n} > 0$ and $\check{G}_{i,n} > 0$ under the assumption of weak correlation for \mathbf{E}_t .

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

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

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

$$\hat{r}_{\text{MR}} = \arg \max_{1 \leq i \leq p/2} \frac{\hat{T}_{i,n}}{\hat{T}_{i+1,n}} =: \arg \max_{1 \leq i \leq p/2} \text{MR}(i), \quad (3.10)$$

$$\hat{r}_{\text{SR}} = \arg \max_{1 \leq i \leq p/2} \frac{\hat{G}_{i,n} - \hat{G}_{i+1,n}}{\hat{G}_{i+1,n} - \hat{G}_{i+2,n}} =: \arg \max_{1 \leq i \leq p/2} \text{SR}(i), \quad (3.11)$$

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

Remark 3.2. Similarly to R_{MR} and R_{SR} defined in (3.10) and (3.11), two statistics using autocorrelations and cross-correlations were respectively employed by Chang et al. (2018) for time series principal component analysis and by Han et al. (2024) for decorrelation bilinear transformation. For \hat{r}_{MR} and \hat{r}_{SR} , if the random errors are weak correlation, both $\hat{T}_{i,n}$ and $\hat{G}_{i,n}$ are greater than zero. Then their consistency can be ensured. However, because $\widetilde{\mathbf{M}}_1$ is not full rank, the ER method based on $\widetilde{\mathbf{M}}_1$ 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 the suitable $\widetilde{\mathbf{R}} = p^{\frac{1-\delta}{2}}(Q_1, Q_2, \dots, Q_m) = (R_1, \dots, R_m)$ with a large positive integer m , and define:

$$\begin{aligned} \Omega_{\tilde{x}}(h) &:= \frac{1}{(n-h)p^{2-2\delta}} \sum_{t=1}^{n-h} \mathbb{E}(\mathbf{Y}_t^\top \widetilde{\mathbf{R}} \widetilde{\mathbf{R}}^\top \mathbf{Y}_{t+h}), \\ \Omega_{\tilde{f}}(h) &:= \frac{1}{n-h} \sum_{t=1}^{n-h} \mathbb{E}(\mathbf{F}_t^\top \mathbf{F}_{t+h}), \\ \Omega_{\tilde{\xi}\tilde{f}}(h) &:= \frac{1}{(n-h)p^{1-\delta}} \sum_{t=1}^{n-h} \mathbb{E}(\mathbf{E}_t^\top \mathbf{R} \mathbf{F}_{t+h}). \end{aligned}$$

Further, define the matrix:

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

and its sample version:

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

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

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

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

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

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

The max-type test statistic is given by:

$$\check{T}_{i,n} := \max_{1 \leq h \leq K} n^{1/2} |\check{\Gamma}_i(h)|_\infty = \max_{1 \leq h \leq K} n^{1/2} \left| \mathfrak{C}_i^\top \Sigma_{\widetilde{\mathbf{X}}}(h) \mathfrak{C}_i \right|_\infty, \quad (3.15)$$

and the sum-type test statistic is:

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

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

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

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

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

$$\hat{c}_{\text{MR}} = \arg \max_{1 \leq i \leq q/2} \frac{\hat{T}_{i,n}}{\hat{T}_{i+1,n}} =: \arg \max_{1 \leq i \leq q/2} \text{MR}(i), \quad (3.17)$$

$$\hat{c}_{\text{SR}} = \arg \max_{1 \leq i \leq q/2} \frac{\hat{G}_{i,n} - \hat{G}_{i+1,n}}{\hat{G}_{i+1,n} - \hat{G}_{i+2,n}} =: \arg \max_{1 \leq i \leq q/2} \text{SR}(i), \quad (3.18)$$

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

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 $\{\mathbf{Y}_t\}_{t \leq T}$, maximum number m

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

- 1: Initial estimators $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$ from $\hat{\mathbf{M}}_1$ and $\hat{\mathbf{M}}_2$;
 - 2:
 - Calculate $\hat{\mathbf{M}}_1$ using $\hat{\mathbf{C}}$, then obtain \hat{r}_{MR} and \hat{r}_{SR} from equations (3.10) and (3.11);
 - Calculate $\hat{\mathbf{M}}_2$ using $\hat{\mathbf{R}}$, then obtain \hat{c}_{MR} and \hat{c}_{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.

Assumption 3.1. $\hat{\mathbf{M}}_1$ has r distinct positive eigenvalues; $\hat{\mathbf{M}}_2$ has c distinct positive eigenvalues.

Theorem 3.1. Under Assumptions 2.1 and 3.1 and $p^\delta q^\omega n^{-1/2} = o(1)$, we have

$$\frac{\hat{T}_{r+1,n}}{\hat{T}_{r,n}} = O_p(p^{2\delta} q^{2\omega} n^{-1}); \quad \frac{\hat{T}_{c+1,n}}{\hat{T}_{c,n}} = O_p(p^{2\delta} q^{2\omega} n^{-1}).$$

Theorem 3.2. Under Assumptions 2.1 and 3.1 and $p^\delta q^\omega n^{-1/2} = o(1)$, we have

$$\frac{\hat{G}_{r+1,n} - \hat{G}_{r+2,n}}{\hat{G}_{r,n} - \hat{G}_{r+1,n}} = O_p(p^{2\delta} q^{2\omega} n^{-1}); \quad \frac{\hat{G}_{c+1,n} - \hat{G}_{c+2,n}}{\hat{G}_{c,n} - \hat{G}_{c+1,n}} = O_p(p^{2\delta} q^{2\omega} n^{-1}).$$

Theorem 3.3. Suppose Assumptions 2.1 and 3.1 hold and $p^\delta q^\omega n^{-1/2} = o(1)$. Then, the MR, SR, $\widetilde{\text{MR}}$ and $\widetilde{\text{SR}}$ estimators (i.e., \hat{r}_{MR} in (3.10), \hat{r}_{SR} in (3.11), \hat{c}_{MR} in (3.17) and \hat{c}_{SR} in (3.18)) satisfy that

$$\begin{aligned} P(\hat{r}_{\text{SR}} = r) &\rightarrow 1; & P(\hat{c}_{\text{SR}} = c) &\rightarrow 1; \\ P(\hat{r}_{\text{MR}} = r) &\rightarrow 1; & P(\hat{c}_{\text{MR}} = c) &\rightarrow 1. \end{aligned}$$

Remark 3.3. First, when $\delta = \omega < 0.5$ and $n \asymp pq$, the condition for weak factors is

$p^\delta q^\omega n^{-1/2} = o(1)$; 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. From Theorem 3.2 and Theorem 3.1, the two-step estimators can achieve the same convergence rate as the counterparts in Theorem 7.1 and Theorem 7.2 of the Appendix.

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 \mathbf{Y}_t 's from model (2.1), where the dimensions of the latent factor process \mathbf{F}_t are chosen to be $r = c = 3$. The entries of \mathbf{F}_t are independent and follow AR(1) processes, with errors generated from the standard normal distribution. That is:

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

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) Weak correlations: $a = 0.1$;
- (2) Moderate correlations: $a = 0.5$;
- (3) Strong correlations: $a = 0.9$.

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

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

- (1) $\mathbf{\Sigma}_1 = \mathbf{I}_p$ and $\mathbf{\Sigma}_2 = \mathbf{I}_q$;
- (2) $\mathbf{\Sigma}_1 = (\sigma_{1,ij})$, where $\sigma_{1,ij} = 0.1$ for $i \neq j$ and $\sigma_{1,ii} = 1$, and $\mathbf{\Sigma}_2 = (\sigma_{2,ij})$, 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 1-6, where the results are presented in the form $x(y|z)$. Here, the estimators ER_o , SR_o , and MR_o are defined based on model (2.1), as detailed in Appendix.

Tables 1-3 show the case where Σ_1 and Σ_2 are diagonal. Based on model (2.1), the estimated results demonstrate that the proposed approaches perform excellently. 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 correlations. Both SR and MR estimators consistently yield better results than the ER estimator.

When the elements in Σ_1 and Σ_2 are small but equal, this case illustrates that the elements of \mathbf{E}_t are weakly correlated. Based on model (2.1), Tables 4-6 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 the ER estimator. Specifically, the SR estimator consistently outperforms the others, with its improved power being most notable. At the same time, the MR estimator clearly outperforms the ER estimator.

In summary, Tables 1-6 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 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 months. 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, \dots, D_L , and fit a factor model using each of the D_{-l} samples to obtain $\hat{\mathbf{R}}_{-l}$ and $\hat{\mathbf{C}}_{-l}$, 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 $\mathbf{S}_{t,l} = \mathbf{R}\mathbf{F}_t\mathbf{C}^\top$. The estimated loading spaces are then used to obtain $\hat{\mathbf{S}}_{t,l} = \hat{\mathbf{R}}_{-l}\hat{\mathbf{R}}_{-l}^\top\mathbf{Y}_t\hat{\mathbf{C}}_{-l}\hat{\mathbf{C}}_{-l}^\top$ for each $l = 1, \dots, L$.

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

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

Here, we set $L = 5$ and ensure that the time length of each dataset D_l is 175. We compute RSS for the matrix factor model under two different scenarios: when $r = 1$ and $c = 2$, the value RSS is 6.673, while when $r = 1$ and $c = 4$, the value 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.

References

- Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors, *Econometrica* **81**: 1203–1227.
- 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.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**: 191–221.
- 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. Y., Guo, B. and Yao, Q. (2015). High dimensional stochastic regression with latent factors, endogeneity and nonlinearity, *Journal of Econometrics* **189**: 297–312.
- Chang, J., Yao, Q. and Zhou, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations, *Biometrika* **104**(1): 31–41.
- 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.
- Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements, *Journal of the Royal Statistical Society: Series B* **75**: 603–680.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic-factor model: identification and estimation, *The Review of Economics and Statistics* **82**: 540–554.
- 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.
- Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model, *Journal of the American Statistical Association* **102**: 603–617.

- Han, Y., Chen, R., Zhang, C. H. and Yao, Q. (2024). Simultaneous decorrelation of matrix time series, *Journal of the American Statistical Association* **119**: 957–969.
- 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.
- Peña, D. and Box, G. (1987). Identifying a simplifying structure in time series, *Journal of the American statistical Association* **82**: 836–843.
- Stock, J. and Watson, M. (2002). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**: 1167–1179.
- Tsai, H. and Tsay, R. S. (2010). Constrained factor models, *Journal of the American Statistical Association* **105**: 1593–1605.
- 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 matrixfactor models, *Journal of Econometrics* **229**: 201–217.

Table 1: Relative frequency estimates for $P(\hat{r} = 3)$ and $P(\hat{c} = 3)$ with 200 reduplicate samples for the AR coefficients with case (1) and Σ_1 and Σ_2 with case (1)

(δ, ω)		$(0.5, 0.5)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{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	ER	SR	MR	ER	SR	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	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{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	ER	SR	MR	ER	SR	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	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{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	ER	SR	MR	ER	SR	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 reduplicate samples for the AR coefficients with case (2) and Σ_1 and Σ_2 with case (1)

(δ, ω)		(0.5, 0.5)					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.520(95 1)	0.790(41 1)	0.545(90 1)	0.530(94 0)	0.790(42 0)	0.575(85 0)
	800	0.500(99 1)	0.775(45 0)	0.525(95 0)	0.460(106 2)	0.775(44 1)	0.515(96 1)
(20,40)	200	0.495(100 1)	0.680(64 0)	0.555(89 0)	0.525(95 0)	0.945(11 0)	0.720(56 0)
	800	0.535(92 1)	0.710(58 0)	0.550(89 1)	0.525(90 5)	0.955(9 0)	0.775(44 1)
(40,40)	200	0.585(83 0)	0.845(31 0)	0.665(67 0)	0.555(89 0)	0.830(34 0)	0.640(72 0)
	800	0.530(94 0)	0.905(10 0)	0.650(70 0)	0.545(91 0)	0.885(23 0)	0.620(76 0)
(p, q)	n	ER	SR	MR	ER	SR	MR
(20,20)	200	0.585(79 4)	0.905(17 2)	0.685(63 0)	0.565(83 4)	0.895(20 1)	0.700(60 0)
	800	0.560(82 6)	0.920(13 3)	0.630(72 2)	0.530(92 2)	0.905(16 3)	0.690(59 3)
(20,40)	200	0.600(79 1)	0.850(30 0)	0.690(62 0)	0.565(87 0)	0.975(4 1)	0.815(37 0)
	800	0.625(73 2)	0.900(20 0)	0.710(57 1)	0.675(62 3)	0.970(6 0)	0.850(29 1)
(40,40)	200	0.655(69 0)	0.955(8 1)	0.780(44 0)	0.605(79 0)	0.925(15 0)	0.785(42 1)
	800	0.590(82 0)	0.975(4 1)	0.810(38 0)	0.615(77 0)	0.965(7 0)	0.740(51 1)
(δ, ω)		(0, 0.5)					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.850(30 0)	0.965(7 0)	0.890(22 0)	0.830(33 1)	0.985(3 0)	0.885(22 1)
	800	0.800(40 0)	0.975(5 0)	0.820(26 0)	0.815(37 0)	0.990(2 0)	0.935(13 0)
(20,40)	200	0.855(28 1)	0.960(8 0)	0.900(20 0)	0.815(37 0)	1.000(0 0)	0.920(12 0)
	800	0.830(34 0)	0.920(16 0)	0.925(15 0)	0.815(37 0)	1.000(0 0)	0.920(12 0)
(40,40)	200	0.815(37 0)	0.995(1 0)	0.940(12 0)	0.840(32 0)	1.000(0 0)	0.940(12 0)
	800	0.845(31 0)	0.995(1 0)	0.950(10 0)	0.875(25 0)	0.985(3 0)	0.945(11 0)
(p, q)	n	ER	SR	MR	ER	SR	MR
(20,20)	200	0.850(26 4)	0.995(1 0)	0.925(15 0)	0.835(31 2)	1.000(0 0)	0.900(18 2)
	800	0.810(36 2)	0.985(3 0)	0.895(20 1)	0.825(34 1)	0.995(0 1)	0.940(11 1)
(20,40)	200	0.865(26 1)	0.980(3 1)	0.925(15 0)	0.825(34 2)	1.000(0 0)	0.910(12 2)
	800	0.830(34 0)	0.980(4 0)	0.955(9 0)	0.825(32 3)	1.000(0 0)	0.915(12 1)
(40,40)	200	0.840(32 0)	0.995(1 0)	0.945(10 1)	0.850(30 0)	1.000(0 0)	0.950(9 1)
	800	0.850(30 0)	1.000(0 0)	0.945(11 0)	0.880(24 0)	1.000(0 0)	0.945(11 0)
(δ, ω)		(0, 0)					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.930(12 0)	1.000(0 0)	0.990(2 0)	0.910(16 2)	1.000(0 0)	0.965(5 2)
	800	0.930(12 2)	1.000(0 0)	0.990(2 0)	0.935(13 0)	1.000(0 0)	0.975(5 0)
(20,40)	200	0.975(5 0)	1.000(0 0)	1.000(0 0)	0.945(11 0)	1.000(0 0)	0.975(5 0)
	800	0.970(6 0)	1.000(0 0)	1.000(0 0)	0.940(10 0)	1.000(0 0)	0.805(39 0)
(40,40)	200	0.960(8 0)	1.000(0 0)	1.000(0 0)	0.980(3 1)	1.000(0 0)	0.995(10)
	800	0.970(6 0)	0.995(1 0)	0.770(46 0)	0.965(7 0)	1.000(0 0)	0.980(2 2)
(p, q)	n	ER	SR	MR	ER	SR	MR
(20,20)	200	0.930(10 2)	1.000(0 0)	0.995(0 1)	0.905(12 7)	1.000(0 0)	0.970(4 2)
	800	0.930(10 4)	1.000(0 0)	0.990(2 0)	0.930(12 2)	1.000(0 0)	0.975(5 0)
(20,40)	200	0.975(5 0)	1.000(0 0)	1.000(0 0)	0.940(11 1)	1.000(0 0)	0.975(4 1)
	800	0.970(6 0)	1.000(0 0)	1.000(0 0)	0.945(9 2)	1.000(0 0)	0.980(1 3)
(40,40)	200	0.960(8 0)	1.000(0 0)	1.000(0 0)	0.980(3 1)	1.000(0 0)	0.990(0 2)
	800	0.970(6 0)	1.000(0 0)	1.000(0 0)	0.965(6 1)	1.000(0 0)	0.985(1 2)

Table 3: Relative frequency estimates for $P(\hat{r} = 3)$ and $P(\hat{c} = 3)$ with 200 reduplicate samples for the AR coefficients with case (3) and Σ_1 and Σ_2 with case (1)

(δ, ω)		$(0.5, 0.5)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.825(35 0)	0.960(8 0)	0.915(17 0)	0.795(41 0)	0.960(8 0)	0.925(15 0)
	800	0.820(36 0)	0.985(3 0)	0.960(8 0)	0.850(30 0)	0.995(1 0)	0.935(13 0)
(20,40)	200	0.840(32 0)	0.965(7 0)	0.950(10 0)	0.800(40 0)	0.990(2 0)	0.935(13 0)
	800	0.840(32 0)	0.975(5 0)	0.955(9 0)	0.805(39 0)	0.990(2 0)	0.945(11 0)
(40,40)	200	0.810(38 0)	0.980(4 0)	0.970(6 0)	0.855(29 0)	0.975(5 0)	0.940(12 05)
	800	0.855(29 0)	0.990(2 0)	0.965(7 0)	0.805(39 0)	0.990(2 0)	0.935(13 0)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.825(32 3)	1.000(0 0)	0.940(11 1)	0.810(37 1)	0.980(4 0)	0.955(9 0)
	800	0.835(33 0)	1.000(3 0)	0.975(5 0)	0.860(26 2)	0.995(1 0)	0.960(8 0)
(20,40)	200	0.840(32 0)	0.985(3 0)	0.970(6 0)	0.800(39 1)	0.990(2 0)	0.930(13 1)
	800	0.845(31 0)	0.995(1 0)	0.975(5 0)	0.810(38 0)	0.995(0 1)	0.940(12 0)
(40,40)	200	0.825(35 0)	0.985(3 0)	0.975(5 0)	0.865(27 0)	0.990(2 0)	0.940(12 05)
	800	0.870(26 0)	1.000(0 0)	0.965(6 1)	0.825(34 1)	0.990(2 0)	0.935(12 1)
(δ, ω)		$(0, 0.5)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.940(12 0)	1.000(0 0)	0.985(3 0)	0.915(17 0)	1.000(0 0)	0.995(1 0)
	800	0.960(8 0)	0.990(2 0)	0.980(3 1)	0.940(12 0)	1.000(0 0)	0.990(2 0)
(20,40)	200	0.955(9 0)	0.985(3 0)	0.970(6 0)	0.915(17 0)	1.000(0 0)	0.970(4 2)
	800	0.955(9 0)	0.985(3 0)	0.990(2 0)	0.950(10 0)	1.000(0 0)	1.000(0 0)
(40,40)	200	0.985(3 0)	1.000(0 0)	0.995(1 0)	0.970(6 0)	1.000(0 0)	0.985(1 2)
	800	0.970(6 0)	1.000(0 0)	0.990(1 1)	0.955(9 0)	1.000(0 0)	0.990(2 0)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.935(11 2)	1.000(0 0)	0.985(2 1)	0.915(16 1)	1.000(0 0)	0.990(1 1)
	800	0.950(7 3)	0.995(1 0)	0.980(2 2)	0.930(11 3)	1.000(0 0)	0.990(2 0)
(20,40)	200	0.960(7 1)	0.990(3 0)	0.975(5 0)	0.910(16 2)	1.000(0 0)	0.965(5 2)
	800	0.965(7 0)	0.990(2 0)	0.990(1 1)	0.950(10 0)	1.000(0 0)	1.000(0 0)
(40,40)	200	0.985(3 0)	1.000(0 0)	0.995(1 0)	0.970(6 0)	1.000(0 0)	0.980(3 15)
	800	0.975(5 0)	1.000(0 0)	0.995(0 0)	0.965(7 0)	1.000(0 0)	0.985(1 2)
(δ, ω)		$(0, 0)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.985(2 1)	1.000(0 0)	1.000(1 0)	0.980(4 0)	0.995(1 0)	0.985(2 1)
	800	0.990(2 0)	0.995(1 0)	0.995(1 0)	0.990(1 3)	1.000(0 0)	1.000(0 0)
(20,40)	200	0.985(2 1)	1.000(0 0)	0.995(1 0)	0.980(4 0)	1.000(0 0)	1.000(0 0)
	800	0.990(2 0)	1.000(1 0)	0.995(1 0)	0.990(2 0)	1.000(0 0)	1.000(0 0)
(40,40)	200	0.980(3 1)	1.000(0 0)	1.000(0 0)	0.990(2 0)	1.000(0 0)	1.000(0 0)
	800	1.000(0 0)	1.000(0 0)	1.000(0 0)	1.000(0 0)	1.000(0 0)	1.000(0 0)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.980(2 2)	1.000(0 0)	1.000(0 0)	0.980(4 0)	0.995(1 0)	0.985(2 1)
	800	0.985(2 3)	0.995(1 0)	0.995(1 0)	0.980(1 3)	1.000(0 0)	1.000(0 0)
(20,40)	200	0.985(2 1)	1.000(0 0)	1.000(0 0)	0.970(4 2)	1.000(0 0)	1.000(0 0)
	800	0.985(1 2)	1.000(0 0)	0.995(1 0)	0.985(2 1)	1.000(0 0)	1.000(0 0)
(40,40)	200	0.975(3 2)	0.975(5 0)	0.790(42 0)	0.990(1 1)	1.000(0 0)	1.000(0 0)
	800	1.000(0 0)	1.000(0 0)	1.000(0 0)	1.000(0 0)	1.000(0 0)	1.000(0 0)

Table 4: Relative frequency estimates for $P(\hat{r} = 3)$ and $P(\hat{c} = 3)$ with 200 reduplicate samples for the AR coefficients with case (1) and Σ_1 and Σ_2 with case (2)

(δ, ω)		$(0.5, 0.5)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{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	ER	SR	MR	ER	SR	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	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{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	ER	SR	MR	ER	SR	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	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{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	ER	SR	MR	ER	SR	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)

Table 5: Relative frequency estimates for $P(\hat{r} = 3)$ and $P(\hat{c} = 3)$ with 200 reduplicate samples for the AR coefficients with case (2) and Σ_1 and Σ_2 with case (2)

(δ, ω)		$(0.5, 0.5)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.500(97 3)	0.670(66 0)	0.535(91 2)	0.400(118 2)	0.680(63 1)	0.540(91 1)
	800	0.525(95 0)	0.690(55 7)	0.500(97 3)	0.515(96 1)	0.725(54 1)	0.510(97 1)
(20,40)	200	0.415(114 3)	0.540(90 2)	0.510(98 0)	0.510(98 0)	0.830(31 3)	0.630(73 1)
	800	0.470(106 0)	0.535(92 1)	0.485(103 0)	0.550(89 1)	0.770(39 7)	0.570(84 2)
(40,40)	200	0.490(102 0)	0.490(96 6)	0.460(105 3)	0.480(104 0)	0.545(88 3)	0.555(87 2)
	800	0.500(100 0)	0.565(78 9)	0.585(81 2)	0.450(110 0)	0.485(91 12)	0.500(98 2)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.555(85 4)	0.870(26 0)	0.620(76 0)	0.535(91 2)	0.905(18 1)	0.635(69 4)
	800	0.530(91 3)	0.840(29 3)	0.675(63 2)	0.585(81 2)	0.880(22 2)	0.650(67 3)
(20,40)	200	0.500(99 0)	0.750(49 1)	0.615(76 1)	0.620(74 2)	0.985(1 2)	0.790(41 1)
	800	0.610(78 0)	0.795(41 0)	0.610(78 0)	0.675(65 0)	0.960(0 1)	0.755(47 2)
(40,40)	200	0.550(90 0)	0.830(31 3)	0.660(68 0)	0.575(85 0)	0.840(30 2)	0.705(58 1)
	800	0.655(69 0)	0.905(19 0)	0.735(51 2)	0.565(86 1)	0.875(23 2)	0.670(65 1)
(δ, ω)		$(0, 0.5)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.790(41 1)	0.975(5 0)	0.895(21 0)	0.815(37 0)	0.985(3 0)	0.915(17 0)
	800	0.795(41 0)	0.970(5 1)	0.890(21 1)	0.840(30 0)	0.950(10 0)	0.910(18 0)
(20,40)	200	0.880(24 0)	0.940(12 0)	0.895(21 0)	0.785(42 1)	0.995(0 1)	0.915(15 2)
	800	0.830(34 0)	0.955(9 0)	0.915(17 0)	0.790(42 0)	0.990(2 0)	0.895(21 0)
(40,40)	200	0.855(29 0)	1.000(0 0)	0.995(1 0)	0.850(30 0)	0.990(2 0)	0.910(18 0)
	800	0.890(22 0)	0.980(4 0)	0.945(11 0)	0.855(29 0)	1.000(0 0)	0.925(15 0)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.785(42 1)	0.995(1 0)	0.900(18 2)	0.830(34 0)	0.995(1 0)	0.930(13 1)
	800	0.795(41 0)	0.985(3 0)	0.910(17 1)	0.845(28 3)	0.975(5 0)	0.925(15 0)
(20,40)	200	0.885(23 0)	0.970(6 0)	0.910(18 0)	0.780(43 1)	0.995(0 1)	0.925(13 2)
	800	0.850(30 0)	0.975(5 0)	0.935(13 0)	0.810(37 1)	0.995(1 0)	0.910(18 0)
(40,40)	200	0.855(28 1)	1.000(0 0)	0.995(1 0)	0.850(30 0)	0.995(1 0)	0.925(15 0)
	800	0.900(20 0)	1.000(0 0)	0.950(10 0)	0.870(26 0)	1.000(0 0)	0.920(14 2)
(δ, ω)		$(0, 0)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.935(13 0)	1.000(0 0)	0.970(5 1)	0.920(16 0)	0.990(2 0)	0.970(5 1)
	800	0.905(18 1)	1.000(0 0)	0.985(2 1)	0.950(8 2)	0.990(2 0)	0.980(4 0)
(20,40)	200	0.970(6 0)	1.000(0 0)	0.995(1 0)	0.950(8 2)	1.000(0 0)	0.990(1 1)
	800	0.975(5 0)	0.995(1 0)	0.995(1 0)	0.950(9 1)	1.000(0 0)	0.990(2 0)
(40,40)	200	0.955(9 0)	1.000(0 0)	0.980(2 0)	0.970(6 0)	1.000(0 0)	0.995(1 0)
	800	0.965(6 0)	1.000(0 0)	1.000(0 0)	0.960(8 0)	1.000(0 0)	0.990(1 1)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.930(11 3)	1.000(0 0)	0.970(5 1)	0.920(14 2)	1.000(0 0)	0.970(5 1)
	800	0.905(17 2)	1.000(0 0)	0.985(2 1)	0.950(7 3)	1.000(0 0)	0.980(4 0)
(20,40)	200	0.970(5 1)	1.000(0 0)	0.995(1 0)	0.950(8 2)	1.000(0 0)	0.995(0 1)
	800	0.975(5 0)	0.995(1 0)	0.995(1 0)	0.950(8 2)	1.000(0 0)	0.990(2 0)
(40,40)	200	0.950(9 0)	1.000(0 0)	0.985(1 2)	0.970(6 0)	1.000(0 0)	0.995(1 0)
	800	0.965(6 0)	1.000(0 0)	1.000(0 0)	0.960(8 0)	1.000(0 0)	0.990(1 1)

Table 6: Relative frequency estimates for $P(\hat{r} = 3)$ and $P(\hat{c} = 3)$ with 200 reduplicate samples for the AR coefficients with case (3) and Σ_1 and Σ_2 with case (2)

(δ, ω)		$(0.5, 0.5)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.820(35 1)	0.975(5 0)	0.940(12 0)	0.785(42 1)	0.965(7 0)	0.930(14 0)
	800	0.845(31 0)	0.990(12 0)	0.935(13 0)	0.770(46 0)	0.950(10 0)	0.935(13 0)
(20,40)	200	0.810(38 0)	0.925(15 0)	0.925(15 0)	0.825(35 0)	0.990(2 0)	0.905(19 0)
	800	0.850(30 0)	0.950(10 0)	0.925(15 0)	0.825(35 0)	0.975(5 0)	0.930(14 0)
(40,40)	200	0.845(31 0)	0.970(6 0)	0.895(21 0)	0.825(35 0)	0.975(5 0)	0.915(16 1)
	800	0.825(35 0)	0.985(3 0)	0.905(19 0)	0.825(35 0)	0.995(1 0)	0.935(13 0)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.820(33 3)	0.990(2 0)	0.950(8 2)	0.810(37 1)	0.985(2 1)	0.930(14 0)
	800	0.850(30 0)	0.995(1 0)	0.950(10 0)	0.785(43 0)	0.975(5 0)	0.940(12 0)
(20,40)	200	0.820(36 0)	0.965(7 0)	0.950(10 0)	0.835(32 1)	0.995(1 0)	0.910(16 2)
	800	0.855(29 0)	0.975(5 0)	0.955(9 0)	0.830(32 2)	1.000(0 0)	0.930(14 0)
(40,40)	200	0.845(31 0)	0.985(3 0)	0.910(18 0)	0.840(32 0)	0.995(1 0)	0.945(10 1)
	800	0.835(33 0)	1.000(3 0)	0.920(16 0)	0.860(26 2)	0.995(1 0)	0.940(11 1)
(δ, ω)		$(0, 0.5)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.935(12 1)	0.995(1 0)	0.990(2 0)	0.940(12 0)	1.000(0 0)	0.990(2 0)
	800	0.970(5 1)	1.000(0 0)	0.985(3 0)	0.945(11 0)	0.990(1 1)	0.985(2 1)
(20,40)	200	0.940(12 0)	0.990(2 0)	0.985(3 0)	0.900(19 1)	1.000(0 0)	0.975(4 1)
	800	0.960(8 0)	0.995(1 0)	0.990(2 0)	0.950(10 0)	1.000(0 0)	0.980(3 1)
(40,40)	200	0.940(11 1)	0.995(1 0)	0.985(2 1)	0.950(10 0)	0.990(2 0)	0.985(3 0)
	800	0.955(9 0)	0.995(0 0)	0.990(2 0)	0.955(9 0)	0.995(1 0)	0.990(2 0)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.930(12 2)	0.995(1 0)	0.990(2 0)	0.940(9 3)	1.000(0 0)	0.990(2 0)
	800	0.975(4 1)	1.000(0 0)	0.985(3 0)	0.940(10 2)	0.995(0 1)	0.985(2 1)
(20,40)	200	0.945(11 0)	0.995(1 0)	0.990(1 0)	0.910(16 2)	1.000(0 0)	0.970(3 3)
	800	0.965(7 0)	1.000(0 0)	0.990(1 1)	0.950(10 0)	1.000(0 0)	0.980(2 2)
(40,40)	200	0.940(11 1)	1.000(0 0)	0.985(2 1)	0.950(10 0)	1.000(0 0)	0.980(2 2)
	800	0.975(5 0)	1.000(0 0)	0.990(2 0)	0.955(9 0)	0.995(1 0)	0.990(2 0)
(δ, ω)		$(0, 0)$					
(p, q)	n	ER_o	SR_o	MR_o	\widetilde{ER}_o	\widetilde{SR}_o	\widetilde{MR}_o
(20,20)	200	0.985(2 1)	0.995(1 0)	0.990(2 0)	0.980(4 0)	1.000(0 0)	1.000(0 0)
	800	0.970(5 1)	0.995(1 0)	0.995(1 0)	0.970(5 1)	0.995(1 0)	0.995(1 0)
(20,40)	200	0.985(2 1)	1.000(0 0)	1.000(0 0)	0.985(2 1)	1.000(0 0)	0.990(1 1)
	800	0.990(2 0)	1.000(1 0)	0.995(1 0)	0.975(1 2)	1.000(0 0)	1.000(0 0)
(40,40)	200	0.995(1 0)	1.000(0 0)	1.000(0 0)	0.990(2 0)	1.000(0 0)	0.995(0 1)
	800	0.990(2 0)	1.000(0 0)	1.000(0 0)	0.995(0 0)	1.000(0 0)	1.000(0 0)
(p, q)	n	ER	SR	MR	\widetilde{ER}	\widetilde{SR}	\widetilde{MR}
(20,20)	200	0.985(1 2)	0.995(1 0)	0.990(2 0)	0.980(4 0)	1.000(0 0)	1.000(0 0)
	800	0.975(1 2)	1.000(0 0)	1.000(0 0)	0.970(4 2)	0.995(1 0)	0.995(1 0)
(20,40)	200	0.985(2 1)	1.000(0 0)	1.000(0 0)	0.985(2 3)	1.000(0 0)	0.995(0 1)
	800	0.990(1 1)	1.000(0 0)	0.995(1 0)	0.970(1 5)	1.000(0 0)	1.000(0 0)
(40,40)	200	0.995(1 0)	1.000(0 0)	1.000(0 0)	0.990(2 0)	1.000(0 0)	1.000(0 0)
	800	0.995(1 0)	1.000(0 0)	1.000(0 0)	0.995(1 0)	1.000(0 0)	1.000(0 0)

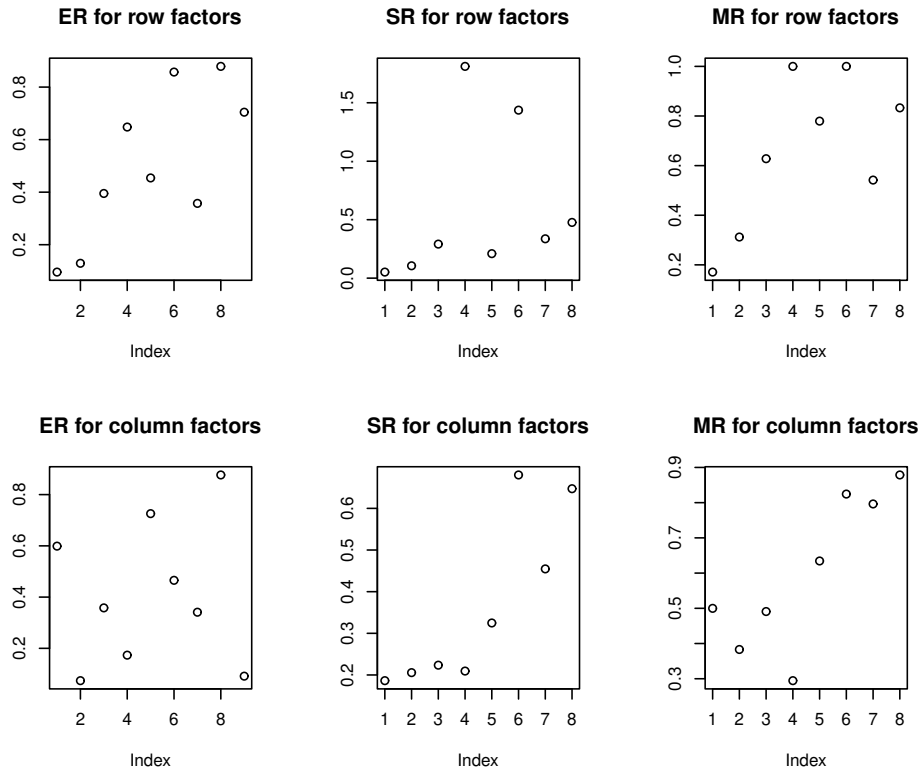


Figure 1: The number of row and column factors by ER_o , SR_o and MR_o estimators for the real data set

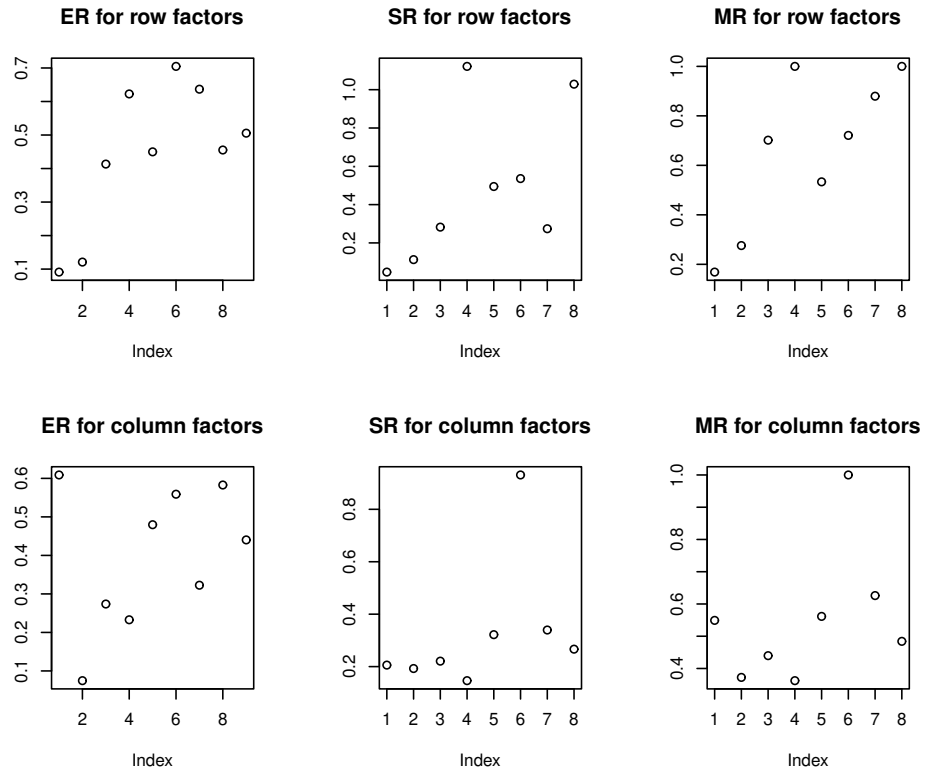


Figure 2: The number of row and column factors by ER, SR and MR estimators for the real data set