**Statistica Sinica Preprint No: SS-2025-0292**

| | |
|---|---|
| **Title** | A Two-Way Factor Model for High-dimensional Matrices with Cross-section and Time-Series Effects |
| **Manuscript ID** | SS-2025-0292 |
| **URL** | http://www.stat.sinica.edu.tw/statistica/ |
| **DOI** | 10.5705/ss.202025.0292 |
| **Complete List of Authors** | Qiang Xia,<br>Zhigen Gao,<br>Gaorong Li and<br>Rubing Liang |
| **Corresponding Authors** | Qiang Xia |
| **E-mails** | xiaqiang@scau.edu.cn |
| Notice: Accepted author version. | |

Statistica Sinica

Statistica Sinica: Newly accepted Paper

Statistica Sinica

# A Two-Way Factor Model Framework for High-Dimensional Panel Time Series

Qiang Xia[a] Zhigen Gao[b] Gaorong Li[c] Rubing Liang[a]

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

[b] *Academy for Advanced Interdisciplinary Studies and KLAS, Northeast Normal University*

[c] *School of Statistics, Beijing Normal University, Beijing, China*

*Abstract:* This paper introduces a novel two-way factor model designed to analyze high-dimensional panel time series. The proposed model assumes that the low-dimensional hidden factors are separably influenced by rows and columns. We conduct likelihood inference via a matrix decomposition technique. To obtain maximum likelihood estimates (MLE) of factor loadings and other parameters, we apply the traditional delta method, which relies on the score function and the Hessian matrix. Additionally, we develop fast computational algorithms based on diagonal block matrices to estimate the model's parameters. Under regularity conditions, we establish the theoretical properties of the proposed estimators, including consistency and asymptotic normality. Notably, the proposed approach achieves $\sqrt{T}$-consistency, representing a substantial improvement over the convergence rates established in prior literature. The effectiveness of the methodology is further validated through simulation experiments and real data analysis.

*Keywords:* Two-way factor model, panel data, high dimensionality, cross-section, time series, likelihood inference

Corresponding author (Rubing Liang): `liang ru bing@scau.edu.cn`, South China Agricultural University.

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

Statistica Sinica: Newly accepted Paper

## 1. Introduction

High-dimensional data analysis finds widespread application in various fields. For example, in asset pricing, portfolio allocation, and risk management, understanding the dynamics of a large number of asset returns is crucial. Large panels of data are often analyzed to study economic phenomena, census prediction, and other time-based measurements. To reduce dimensionality, a factor model is commonly used to provide a low-dimensional, parsimonious representation of high-dimensional dynamics. Various studies, such as Stock and Watson (2002), Bai (2003), Forni et al. (2005), Pan and Yao (2008), Lam et al. (2011), Doz et al. (2012), Ahn and Horenstein (2013), Xia et al. (2015), Chang et al. (2015), Ng et al. (2015), and Fan et al. (2022), have highlighted the effectiveness of factor models.

Recently, matrix-variate time series have gained increasing prominence across diverse fields, including economics, meteorology, and ecology. Leveraging the intrinsic matrix structure of panel data, researchers can effectively reduce dimensionality along both row and column dimensions, prompting the emergence of matrix factor models. Recent studies, including Wang et al. (2019), Chen et al. (2020), Yu et al. (2022), Chen and Fan (2023), He et al. (2023), Yuan et al. (2023), He et al. (2024), and Zhang et al. (2025), have extended conventional factor models to accommodate matrix-variate data structures. This methodological advancement not only characterizes temporal dynamics, but also uncovers cross-sectional dependencies among grouped time series. However, when matrix-valued data lack replication, the structure and certain conditions of the aforementioned model undergo changes, leading to corresponding alterations in their theoretical properties. For high-dimensional matrix data, researchers including Hornstein et al. (2019), and Gao et al. (2021) have explored the distributional properties and correlation structures of such data from multiple perspectives. In contrast to the analysis of matrix-variate time series, which focuses primarily on modeling temporal evolution and cross-sectional associations, high-dimensional panel time series still merit more in-depth investigation from both row and column dimensions.

High-dimensional panel time series featuring both cross-sectional and time-series effects are commonly encountered in various applications. Such data encompass not only coincident and leading economic indicators, but also a broad range of variables with time-series properties McCracken and Ng (2015). For example, FRED-MD, a macroeconomic database comprising 134 monthly US indicators in 658 periods from 1960.3 to 2014.12, was developed by McCracken and Ng (2015) presented in Table 1. If one aims to analyze the column vector, denoted as $Y_{\cdot,t} := (Y_{1,t}, Y_{2,t}, \ldots, Y_{123,t})^{\tau}$, which includes all indicators for $t = 1, 2, \ldots, 360$, it can be studied as a vector time series. Similarly, focusing on the row vector $Y_{i,\cdot} := (Y_{i,1}, Y_{i,2}, \ldots, Y_{i,360})$ for $i = 1, 2, \ldots, 123$, enables the analysis of cross-sectional data. Evidently, there exist inherent relationships between all variables in both columns and rows in Table 1, which motivates the exploration of patterns exhibited by indicators and time to provide a systematic explanation of macroeconomic status.

As the dimensions of the rows and columns associated with $Y$ increase, the curse of di-

Statistica Sinica: Newly accepted Paper

Table 1: Presentation of A High-Dimensional Panel Time Series

| Indicators (i) \ Time (t) | $Y_{(i,t)}$ | | | |
|---|---|---|---|---|
| Trades | $Y_{1,1}$ | $Y_{1,2}$ | $\cdots$ | $Y_{1,360}$ |
| Retails | $Y_{2,1}$ | $Y_{2,2}$ | $\cdots$ | $Y_{2,360}$ |
| Initial claims | $Y_{3,1}$ | $Y_{3,2}$ | $\cdots$ | $Y_{3,360}$ |
| New orders | $Y_{4,1}$ | $Y_{4,2}$ | $\cdots$ | $Y_{4,360}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| Consumer sentiment | $Y_{123,1}$ | $Y_{123,2}$ | $\cdots$ | $Y_{123,360}$ |

mensionality becomes non-negligible. Drawing on Gao et al. (2021), a natural approach is to extract independent latent factor processes from the row and column vectors. Specifically, we decompose $Y$ as:

$$Y = X + Z,$$

where the unobservable components $X$ and $Z$ capture distinct structural patterns. Specifically, $i$-th row of $X$ (denoted $X_{i.}$) and $t$-th column of $Z$ (denoted $Z_{.t}$) are modeled using row and column latent factors, respectively:

$$X_{i.} = LF_i + \varsigma_i, \quad i = 1, \dots, N,$$

$$Z_{.t} = \Lambda E_t + \upsilon_t, \quad t = 1, \dots, T.$$

Here, $F_i$ and $E_t$ represent latent row and column factors, respectively; $L$ and $\Lambda$ are factor

Statistica Sinica: Newly accepted Paper

loading matrices; $\varsigma_i$ and $\upsilon_t$ denote the idiosyncratic noise. To account for factor correlations, the time series vector $E_t$ is assumed to follow a VAR(1) model:

$$E_t = \Phi E_{t-1} + \epsilon_t, \quad t = 1, \ldots, T,$$

where $\epsilon_t$ is a white noise. The proposed model incorporates cross-section and time-series effects and yields a two-way factor model (2wFM), that is similar to, but not identical to, the model proposed by Gao et al. (2021). When $L$ equals zero, our model reduces to the factor model presented by Ng et al. (2015), which is a real vector-valued time series factor model with latent factors following a VAR process. Therefore, this paper focuses on high-dimensional panel time series and introduces a two-way factor analysis method to fully exploit its structure.

In the literature, principal component analysis (PCA) or quasi-maximum likelihood (QM-L) methods are widely used as tools for estimating factor loadings and variance parameters in factor models. However, since the PCA method is not applicable in the context of 2wFM settings, Gao et al. (2021) proposed the QML method for estimating all parameters. To simplify the computation of second-order derivatives, they put forward an EM algorithm, which was previously developed by Bai and Li (2012) for high-dimensional factor analysis. Nevertheless, due to the more complex structure of our proposed model, the QML method based on the EM algorithm is no longer suitable. Therefore, the aim of this paper is to resort to the Newton-Raphson algorithm, which incorporates the Fisher information matrix, to estimate the parameters of the proposed model.

Statistica Sinica: Newly accepted Paper

Although directly inverting the Fisher information matrix requires $O(N^3)$ steps [see Lee (1978) for details], rendering the Newton-Raphson algorithm impractical for large $N$, we are fortunate to present a matrix decomposition strategy that enables the use of the Newton-Raphson algorithm without needing to compute all elements of the Fisher information matrix. Under regularity conditions, the Fisher information matrix can be expressed as a quasi-diagonal matrix, and a block alternating maximization strategy is implemented to obtain the MLE for factor loadings and variance parameters.

Furthermore, we investigate the theoretical properties of the estimators under general conditions and establish their consistency and asymptotic normality. Compared with the results of Ng et al. (2015), we demonstrate that our approach achieves $\sqrt{T}$-consistency for the new model, which is superior to the result reported by Ng et al. (2015).

The remainder of this article is organized as follows. Section 2 provides a detailed introduction to model settings, examines the theoretical properties of estimators, and proposes the estimation procedure. In Section 3, we present the simulation results and the analysis of real data. Finally, in Section 4, we provide a brief summary. The Supplementary Material contains all lemmas and the proofs of theorems.

We now introduce some notation that is essential for understanding the rest of this paper.

(a) $\tau$ denotes the transpose of a matrix. $\mathrm{tr}(\cdot)$ and $|\cdot|$ represent the trace and the determinant of a matrix, respectively.

Statistica Sinica: Newly accepted Paper

(b) For any matrix $n \times n$ $A = (a_{i,j})$, we define $\text{diag}(A)$ as the vector $(a_{11}, \ldots, a_{nn})^\tau$ and $\text{tr}(A)$ as the sum of its diagonal entries, that is, $\sum_{i=1}^n a_{ii}$. For any $n \times n$ block matrix $M$ where $M_{i,j}$ refers to the $(i, j)$-th block and $i, j = 1, 2, \ldots, n$, we define $\text{trs}(M) = \sum_{i=1}^n M_{ii}$.

(c) The symbol $\otimes$ represents the Kronecker product of matrices, and vec denotes the vectorization operation.

(d) Let $K_{mn}$ denote the $mn \times mn$ commutation matrix satisfying $K_{mn}\text{vec}(B) = \text{vec}(B^\tau)$ for an $m \times n$ matrix $B$. Its key properties are: $K_{mn}K_{nm} = I_{mn}$, and $K_{mn}^\tau = K_{nm}$, where $I_{mn}$ is the identity matrix of order $mn$.

(e) $\mathbf{1}$. is a vector or matrix with all elements equal to 1.

(f) For a vector $z = (z_1, \ldots, z_N)^\tau$, $\|z\| = (\sum_{i=1}^N |z_i|^2)^{1/2}$ is the Euclidean norm of $z$.

## 2. The Two-Way Factor Model and Its MLE

### 2.1 The Two-Way Factor Model for High-Dimensional Panel Time Series

Let us consider a high-dimensional panel time series, i.e.,

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{bmatrix}.$$

Statistica Sinica: Newly accepted Paper



Assuming that each entry of $Y$ comprises three components: a row-specific common factor $F_i$, a column-specific common factor $E_t$ driven by a low-dimensional autoregressive process, and a static idiosyncratic noise component $\varepsilon_{it}$ following a white noise process. This extension builds upon the 2wFM proposed by Gao et al. (2021), incorporating both cross-sectional and time-series effects as follows:

$$y_{it} = F_i^\tau L_t + \Lambda_i^\tau E_t + \varepsilon_{it}, \quad E_t = \Phi E_{t-1} + \epsilon_t, \tag{1.1}$$

where $F_i = (f_{i1}, f_{i2}, \ldots, f_{ir})^\tau$ with $\mathbb{E}(F_i) = \mathbf{0}$ and $\text{Cov}(F_i) = \Sigma_F$, $i = 1, 2, \ldots, N$; $E_t = (e_{1t}, e_{2t}, \ldots, e_{ct})^\tau$ with $\mathbb{E}(E_t) = \mathbf{0}$ and $\text{Cov}(E_t) = \Sigma_E$, $t = 1, 2, \ldots, T$; $L_t = (l_{1t}, l_{2t}, \ldots, l_{rt})^\tau$, $t = 1, 2, \ldots, T$; and $\Lambda_i = (\Lambda_{i1}, \Lambda_{i2}, \ldots, \Lambda_{ic})^\tau$, $i = 1, 2, \ldots, N$. Furthermore, $\varepsilon_{it} \sim \mathbb{WN}(0, \sigma^2)$ and $\epsilon_t \sim \mathbb{WN}_c(\mathbf{0}, \Sigma_\epsilon)$, where $\mathbb{WN}$ denotes white noise. Moreover, $F_i$, $E_t$, $\varepsilon_{it}$, and $\epsilon_t$ are not correlated with each other.

Denote $F = \big(F_1, F_2, \ldots, F_N\big)^\tau \in \mathbb{R}^{N\times r}$, $L = (L_1, L_2, \ldots, L_T)^\tau \in \mathbb{R}^{T\times r}$, $\Lambda = \big(\Lambda_1, \Lambda_2, \ldots, \Lambda_N\big)^\tau \in \mathbb{R}^{N\times c}$, and $E = \big(E_1, E_2, \ldots, E_T\big)^\tau \in \mathbb{R}^{T\times c}$. Then,

$$Y = FL^\tau + \Lambda E^\tau + \varepsilon.$$

Thus,

$$\Omega_Y := \text{cov}\big(\text{vec}(Y)\big) = (L\Sigma_F L^\tau) \otimes I_N + (I_T \otimes \Lambda)\Omega_E(I_T \otimes \Lambda^\tau) + \sigma_\varepsilon^2 I_{TN},$$

Statistica Sinica: Newly accepted Paper



where $\Sigma_F = \mathrm{cov}(F_i)$, $\Sigma_E = \mathrm{cov}(E_t)$, and

$$\Omega_E = \mathrm{cov}\big(\mathrm{vec}(E^\tau)\big) = \begin{bmatrix} \Sigma_E & \Sigma_E \Phi^\tau & \cdots & \Sigma_E \Phi^{\tau T-1} \\ \Phi \Sigma_E & \Sigma_E & \cdots & \Sigma_E \Phi^{\tau T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi^{T-1}\Sigma_E & \Phi^{T-2}\Sigma_E & \cdots & \Sigma_E \end{bmatrix}.$$

**Remark 1.** Note that the time-series vectors $\{E_t, t = 1, \ldots, T\}$ follow a VAR(1) model. Consequently, our model differs from that of Gao et al. (2021). Gao et al. (2021) and Yuan et al. (2023) derived the explicit expression of the determinant of $\Omega_Y$ and subsequently implemented the EM algorithm based on this result. Unfortunately, we cannot derive the explicit expression of the determinant of $\Omega_Y$. Thus, their MLE methodology is inapplicable to our model. Instead, in this paper, we propose using the traditional delta method, which relies heavily on the score function and Hessian matrix, to compute the MLEs of factor loadings and other model parameters. It should be noted that the model of Ng et al. (2015) is a special case of our framework, obtained when either $F \equiv 0$ or $L \equiv 0$.

It is easy to observe that model (1.1) is not identifiable, as it remains unchanged when multiplied by any non-singular matrices $H_1$ and $H_2$:

$$Y = FH_1^{-1}H_1L^\tau + \Lambda H_2 H_2^{-1}E^\tau + \varepsilon.$$

Thus, the two sets of parameters $(L, \Lambda)$ and $(LH_1^\tau, \Lambda H_2)$ are indistinguishable. This issue is similar to that found in classical factor models. To obtain the MLE of factor loadings and

Statistica Sinica: Newly accepted Paper



other parameters of the model, the identification conditions (IC) must be carefully examined, as stated in works such asBai and Li (2012), and Gao et al. (2021). The IC of our model is presented below.

**Assumption 1.** (IC1) $L^\tau L = \sigma_\varepsilon^2 T I_r$, $\Lambda^\tau \Lambda = \sigma_\varepsilon^2 N I_c$;

(IC2) Both $\Sigma_F$ and $\Sigma_E$ are diagonal matrices with non-repeated elements.

**Remark 2.** Under Assumptions (IC1) and (IC2), the factor loadings $L$ and $\Lambda$ are identifiable only up to column sign indeterminacy. The theoretical foundation for this identifiability in two-way factor models is provided by Gao et al. (2021). Although condition (IC2) is restrictive – requiring $\Sigma_E$ to be diagonal with distinct elements – it is imposed for critical reasons. It ensures the identifiability of $E_t$, which in turn guarantees that the VAR(1) coefficient matrix $\Phi$ is uniquely determined. Furthermore, (IC2) underpins the convergence of the iterative maximum likelihood estimation (MLE) procedure. Without this structure, the likelihood function may fail to converge to a local optimum. In practice, the necessary identification is therefore achieved by applying rotational transformations.

## 2.2 Score Function and Fisher Information Matrix

Let $\Theta \triangleq (\sigma_\varepsilon^2, L, \Lambda, \Sigma_F, \Sigma_E, \Phi)$ denote the parameter vector and $S = \mathrm{vec}(Y)\mathrm{vec}(Y)^\tau$. The negative logarithm of the quasi-likelihood function for model (1.1), ignoring the constant term,

Statistica Sinica: Newly accepted Paper



can be written as follows:

$$\mathbb{L}(\Theta) = \frac{1}{2}\left[tr(\Omega_Y^{-1}S) + \log|\Omega_Y|\right],$$

where log denotes the natural logarithm; $\Omega_Y$ is the covariance matrix for the vectorized data $Y$, and $\Omega_Y^{-1}$ is its inverse.

**Remark 3.** When the idiosyncratic noise $\varepsilon$ exhibits weak cross-sectional correlations or heterogeneity, consists of free parameters  a quantity that vastly exceeds the number of parameters associated with the factors and loadings. Due to this high dimensionality, Yuan et al. (2023) developed a working likelihood by assuming homogeneous variances within the likelihood function, motivated by both computational and theoretical considerations. For a detailed discussion regarding the interpretation of the objective function as a misspecified likelihood, see Doz et al. (2012). Subsequently, when the idiosyncratic noise displays heterogeneous variances, establishing rigorous theoretical proofs becomes challenging.

Before presenting the theoretical results, the underlying assumptions are introduced as follows.

**Assumption 2.** (A1) $\frac{T}{N} = y \in (0, \infty)$, where $y$ is a fixed constant independent of $N$;

(A2) $\|L_t\| = O(1)$, $t = 1, \ldots, T$ and $\|\Lambda_i\| = O(1)$, $i = 1, \ldots, N$;

(A3) Each element in $\Sigma_E$ and $\Sigma_F$ is bounded, and there exists a sufficiently large positive constant $C$, such that $C^{-1} < \sigma_\varepsilon^2 < C$;

Statistica Sinica: Newly accepted Paper



(A4) A complex number $z$ satisfies $\det(I_c - \Phi z) = 0$ for $\|z\| > 1$.

**Remark 4.** For effective analysis of high-dimensional panel time series, selecting equal divergence rates for $N$ and $T$ (A1) ensures the validity of our theory. This approach facilitates differentiation and measurement of the row and column factors, and draws inspiration from the vector factor model described in Bai and Li (2012) and Fan et al. (2008). Typically, $N$ is assumed to be in order no higher than $T$. In this context, conditions (A2) and (A3) closely resemble (C.1) and (C.2) detailed in Bai and Li (2012), respectively. Specifically, (A2) constitutes a modification of (C.1) in Bai and Li (2012), which stipulates that every element in the factor loadings ($L$ and $\Lambda$) row should be of order $O(1)$. Meanwhile, (A4) serves as a stability criterion for the VAR model.

**Theorem 1.** *Assuming that both Assumption 1 and 2 hold, the score function can be expressed as follows:*

$$\begin{aligned}
\frac{\partial \mathbb{L}}{\partial \sigma_\varepsilon^2} &= \frac{1}{2} tr\big(\Omega_Y^{-1} - \Omega_Y^{-1} S \Omega_Y^{-1}\big), \\
\frac{\partial \mathbb{L}}{\partial vec(L)} &= vec\Big\{NL/(T\sigma_\varepsilon^2) - trs\big(K_{TN}\Omega_Y^{-1} S K_{TN}\big)L/(T\sigma_\varepsilon^2)\Big\}, \\
\frac{\partial \mathbb{L}}{\partial vec(\Lambda)} &= vec\Big\{T\Lambda/(N\sigma_\varepsilon^2) - trs\big(\Omega_Y^{-1} S\big)\Lambda/(N\sigma_\varepsilon^2)\Big\},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathbb{L}}{\partial diag(\Sigma_F)} &= \frac{1}{2} diag\Big\{N\Sigma_F^{-1} - \Sigma_F^{-1} L^\tau trs\big(K_{TN} S K_{NT}\big) L \Sigma_F^{-1}/(T^2\sigma_\varepsilon^4)\Big\}, \\
\frac{\partial \mathbb{L}}{\partial vec(\Sigma_E)} &= -\frac{1}{2}\Upsilon_1^\tau vec\Big\{\Omega_E - \Lambda^\tau trs\big(S\big)\Lambda/(N^2\sigma_\varepsilon^4)\Big\}, \\
\frac{\partial \mathbb{L}}{\partial vec(\Phi)} &= -\frac{1}{2}\Upsilon_2^\tau vec\Big\{\Omega_E - \Lambda^\tau trs\big(S\big)\Lambda/(N^2\sigma_\varepsilon^4)\Big\},
\end{aligned}$$

Statistica Sinica: Newly accepted Paper



*where* $\Upsilon = \{\xi_1 \otimes \xi_1, \sum_{t=2}^{T-1} \xi_t \otimes \xi_t, \xi_T \otimes \xi_T, \sum_{t=1}^{T-1}[(\xi_{t+1} \otimes \xi_t) + (\xi_t \otimes \xi_{t+1})K_{cc}\}$, $\Upsilon_1 = \Upsilon \frac{\partial \Gamma^\tau}{\partial vec(\Sigma_E)}$ *and* $\Upsilon_2 = \Upsilon \frac{\partial \Gamma^\tau}{\partial vec(\Phi)}$ *with* $\Gamma = (vec^\tau \Gamma_1, vec^\tau \Gamma_2, vec^\tau \Gamma_3, vec^\tau \Gamma_4)^\tau$ *and* $\xi_t^\tau = (\mathbf{0}_c, \ldots, \mathbf{0}_c, I_c, \mathbf{0}_c, \ldots, \mathbf{0}_c)$ *for* $t = 1, 2, \ldots, T$.

Theorem 1 demonstrates that, given the knowledge of the other parameters, only $L$, $\Lambda$, and $\Sigma_F$ have a standard solution. The row vectors of $NI_T - \operatorname{trs}\big(K_{TN}\Omega_Y^{-1}SK_{TN}\big)$ and $TI_N - \operatorname{trs}\big(\Omega_Y^{-1}S\big)$ are orthogonal to the column vectors of $L$ and $\Lambda$, respectively. However, selecting the appropriate $L$ and $\Lambda$ can be challenging, which is why we employ the Newton-Raphson algorithm to estimate them.

**Theorem 2.** *Under Assumptions 1 and 2, the Fisher information matrix can be written as:*

$$\mathbb{I}(\Theta) = \begin{bmatrix} \mathcal{I}_{\sigma_\varepsilon^2,\sigma_\varepsilon^2} & \mathcal{I}_{\sigma_\varepsilon^2,L} & \mathcal{I}_{\sigma_\varepsilon^2,\Lambda} & \mathcal{I}_{\sigma_\varepsilon^2,\Sigma_F} & \mathcal{I}_{\sigma_\varepsilon^2,\Sigma_E} & \mathcal{I}_{\sigma_\varepsilon^2,\Phi} \\ \mathcal{I}_{L,\sigma_\varepsilon^2} & \mathcal{I}_{L,L} & \mathcal{I}_{L,\Lambda} & \mathcal{I}_{L,\Sigma_F} & \mathcal{I}_{L,\Sigma_E} & \mathcal{I}_{L,\Phi} \\ \mathcal{I}_{\Lambda,\sigma_\varepsilon^2} & \mathcal{I}_{\Lambda,L} & \mathcal{I}_{\Lambda,\Lambda} & \mathcal{I}_{\Lambda,\Sigma_F} & \mathcal{I}_{\Lambda,\Sigma_E} & \mathcal{I}_{\Lambda,\Phi} \\ \mathcal{I}_{\Sigma_F,\sigma_\varepsilon^2} & \mathcal{I}_{\Sigma_F,L} & \mathcal{I}_{\Sigma_F,\Lambda} & \mathcal{I}_{\Sigma_F,\Sigma_F} & \mathcal{I}_{\Sigma_F,\Sigma_E} & \mathcal{I}_{\Sigma_F,\Phi} \\ \mathcal{I}_{\Sigma_E,\sigma_\varepsilon^2} & \mathcal{I}_{\Sigma_E,L} & \mathcal{I}_{\Sigma_E,\Lambda} & \mathcal{I}_{\Sigma_E,\Sigma_F} & \mathcal{I}_{\Sigma_E,\Sigma_E} & \mathcal{I}_{\Sigma_E,\Phi} \\ \mathcal{I}_{\Phi,\sigma_\varepsilon^2} & \mathcal{I}_{\Phi,L} & \mathcal{I}_{\Phi,\Lambda} & \mathcal{I}_{\Phi,\Sigma_F} & \mathcal{I}_{\Phi,\Sigma_E} & \mathcal{I}_{\Phi,\Phi} \end{bmatrix},$$

Statistica Sinica: Newly accepted Paper



*where*

$$\mathcal{I}_{\sigma_\varepsilon^2,\sigma_\varepsilon^2} = \frac{1}{2T} tr\left(\Omega_Y^{-2}\right); \mathcal{I}_{\sigma_\varepsilon^2,L} = \boldsymbol{0}; \mathcal{I}_{\sigma_\varepsilon^2,\Lambda} = \boldsymbol{0}; \mathcal{I}_{\sigma_\varepsilon^2,\Sigma_F} = \boldsymbol{0}; \mathcal{I}_{\sigma_\varepsilon^2,\Sigma_E} = \boldsymbol{0}; \mathcal{I}_{\sigma_\varepsilon^2,\Phi} = \boldsymbol{0};$$

$$\mathcal{I}_{L,\sigma_\varepsilon^2} = \boldsymbol{0}; \mathcal{I}_{L,L} = \frac{1}{T}\Sigma_F \otimes trs\Big(K_{NT}\Omega_Y^{-1}K_{TN}\Big); \mathcal{I}_{L,\Lambda} = \boldsymbol{0}; \mathcal{I}_{L,\Sigma_F} = \boldsymbol{0}; \mathcal{I}_{L,\Sigma_E} = \boldsymbol{0}; \mathcal{I}_{L,\Phi} = \boldsymbol{0};$$

$$\mathcal{I}_{\Lambda,\sigma_\varepsilon^2} = \boldsymbol{0}; \mathcal{I}_{\Lambda,L} = \boldsymbol{0}; \mathcal{I}_{\Lambda,\Lambda} = \Sigma_E \otimes I_N; \mathcal{I}_{\Lambda,\Sigma_F} = \boldsymbol{0}; \mathcal{I}_{\Lambda,\Sigma_E} = \boldsymbol{0}; \mathcal{I}_{\Lambda,\Phi} = \boldsymbol{0};$$

$$\mathcal{I}_{\Sigma_F,\sigma_\varepsilon^2} = \boldsymbol{0}; \mathcal{I}_{\Sigma_F,L} = \boldsymbol{0}; \mathcal{I}_{\Sigma_F,\Lambda} = \boldsymbol{0}; \mathcal{I}_{\Sigma_F,\Sigma_F} = \frac{N}{2T}\Big(\Sigma_F^{-1} \otimes \Sigma_F^{-1}\Big); \mathcal{I}_{\Sigma_F,\Sigma_E} = \boldsymbol{0}; \mathcal{I}_{\Sigma_F,\Phi} = \boldsymbol{0};$$

$$\mathcal{I}_{\Sigma_E,\sigma_\varepsilon^2} = \boldsymbol{0}; \mathcal{I}_{\Sigma_E,L} = \boldsymbol{0}; \mathcal{I}_{\Sigma_E,\Lambda} = \boldsymbol{0}; \mathcal{I}_{\Sigma_E,\Sigma_F} = \boldsymbol{0}; \mathcal{I}_{\Phi,\sigma_\varepsilon^2} = \boldsymbol{0}; \mathcal{I}_{\Phi,L} = \boldsymbol{0}; \mathcal{I}_{\Phi,\Lambda} = \boldsymbol{0}; \mathcal{I}_{\Phi,\Sigma_F} = \boldsymbol{0};$$

$$\mathcal{I}_{\Sigma_E,\Sigma_E} = \frac{1}{2T}\Upsilon_1^\tau\Big(\Omega_E \otimes \Omega_E\Big)\Upsilon_1; \mathcal{I}_{\Phi,\Phi} = \frac{1}{2T}\Upsilon_2^\tau\Big(\Omega_E \otimes \Omega_E\Big)\Upsilon_2; \mathcal{I}_{\Sigma_E,\Phi} = \frac{1}{2T}\Upsilon_1^\tau\Big(\Omega_E \otimes \Omega_E\Big)\Upsilon_2;$$
$$\mathcal{I}_{\Phi,\Sigma_E} = \frac{1}{2T}\Upsilon_2^\tau\Big(\Omega_E \otimes \Omega_E\Big)\Upsilon_1.$$

It can be observed from Theorem 2 that the Fisher information matrix assumes the form of a quasi-block-diagonal matrix, subject to the conditions (IC1) and (IC2). This property facilitates the estimation of MLE using the Newton-Raphson algorithm as it streamlines the computation process.

## 2.3 Asymptotic properties of MLE $\widehat{\Theta}$

In order to investigate the asymptotic characteristics of $\widehat{\Theta}$, it is essential to establish the following regularity condition that describes the dependence of time series.

**Assumption 3.** The common factors $\{F_i\}$ or the time series $\{(y_{\cdot t}, E_t\}$ is stationary and $\alpha$-mixing, where the mixing coefficients $\alpha(\cdot)$ satisfy the condition $\sum_{l=1}^{\infty} \alpha(l)^{1-2/\gamma} < \infty$ for some

Statistica Sinica: Newly accepted Paper



$\gamma > 2$, where

$$\alpha(l) = \sup_j \sup_{U \in \mathcal{F}_{-\infty}^j, V \in \mathcal{F}_{j+l}^\infty} |P(U \cap V) - P(U)P(V)| \tag{1.2}$$

and $\mathcal{F}_k^j$ is the $\sigma$-field generated by $\{F_i : j \le i \le k\}$ or $\{(y_{\cdot t}, E_t) : j \le t \le k\}$. Furthermore, $\mathbb{E}(|F_i|^4) < \infty$ and $\mathbb{E}(|E_t|^4) < \infty$ elementwise; $\mathbb{E}(\varepsilon_{i,t}^4) < \infty$.

**Theorem 3.** *Under the assumptions stated in Assumptions 1, 2, and 3, as $T, N \to \infty$, with a probability approaching one, $\widehat{\Theta}$ possesses average $\sqrt{T}$-consistency. Here, average consistency refers to*

$$\begin{aligned}
&|\widehat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| = O_P(T^{-1}), \quad \frac{1}{T} tr\Big[\Big(\widehat{L} - L\Big)\Big(\widehat{L} - L\Big)^\tau\Big] = O_P(T^{-1}), \\
&\frac{1}{N} tr\Big[\Big(\widehat{\Lambda} - \Lambda\Big)\Big(\widehat{\Lambda} - \Lambda\Big)^\tau\Big] = O_P(T^{-1}), \quad tr\Big[\Big(\widehat{\Sigma}_F - \Sigma_F\Big)\Big(\widehat{\Sigma}_F - \Sigma_F\Big)^\tau\Big] = O_P(T^{-1}), \\
&tr\Big[\Big(\widehat{\Sigma}_E - \Sigma_E\Big)\Big(\widehat{\Sigma}_E - \Sigma_E\Big)^\tau\Big] = O_P(T^{-1}), \quad \|vec\big(\widehat{\Phi} - \Phi\big)\| = O_P(T^{-1/2}).
\end{aligned}$$

As $T$ approaches infinity, the model proposed by Ng et al. (2015) demonstrates that the MLE for their parameters is consistent and driven to be $O_P(T^{-\alpha})$ with $0 < \alpha < 1/2$. On the other hand, Theorem 3 establishes the $\sqrt{T}$-consistency of the proposed estimators in an average sense, which surpasses the convergence rate established in Ng et al. (2015).

**Theorem 4.** *Under Assumptions 1, 2 and 3, as $T, N \to \infty$,*

$$\sqrt{NT}\big(\widehat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\big) \to \mathbf{N}\big(0, \mathbb{E}(\varepsilon_{it}^4) - \sigma_\varepsilon^{\dagger 4}\big),$$

$$\sqrt{N}\big(\widehat{L}_t - L_t\big) \to \mathbf{N}_r\Big(\boldsymbol{0}, \sigma_\varepsilon^2 \Sigma_F^{-1}\Big), \quad t = 1, \ldots, T,$$

Statistica Sinica: Newly accepted Paper



$$\sqrt{N}diag(\widehat{\Sigma}_F - \Sigma_F) \to \mathbf{N}_r\Big(\boldsymbol{0}, \mathbb{E}\big(diag(F_iF_i^\tau)^2\big) - \Sigma_F^{\dagger 2}\Big),$$

$$\sqrt{T}(\widehat{\Lambda}_i - \Lambda_i) \to \mathbf{N}_c\Big(\boldsymbol{0}, \sigma_\varepsilon^2\ \Sigma_E^{-1}\Big), \quad i = 1, \ldots, N,$$

$$\sqrt{T}\Big(vec(\widehat{\Sigma}_E - \Sigma_E)^\tau, vec(\widehat{\Phi} - \Phi)^\tau\Big)^\tau \to \mathbf{N}_{2c^2}\Big(\boldsymbol{0}, \mathbf{J}\mathbf{\Pi}\mathbf{J}^\tau\Big),$$

*where* $\mathbf{E} = \Big(vec^\tau\big(E_1E_1^\tau - \Sigma_E\big), \sum_{t=2}^{T-1} vec^\tau\big(E_tE_t^\tau - \Sigma_E\big), vec^\tau\big(E_TE_T^\tau - \Sigma_E\big), 2\sum_{t=1}^{T-1} vec^\tau\big(E_tE_{t+1}^\tau - \Sigma_E\Phi^\tau\big)\Big)^\tau,$

$\mathbf{\Pi} = Cov(\mathbf{E})$, $\mathcal{J} = T\begin{pmatrix} \mathcal{I}_{\Sigma_E,\Sigma_E} & \mathcal{I}_{\Sigma_E,\Phi} \\ \mathcal{I}_{\Phi,\Sigma_E} & \mathcal{I}_{\Phi,\Phi} \end{pmatrix}$, *and* $\mathbf{J} = \frac{1}{2}\mathcal{J}^{-1}\Big(\frac{\partial vec(\Gamma)}{\partial vec^\tau(\Sigma_E)}, \frac{\partial vec(\Gamma)}{\partial vec^\tau(\Phi)}\Big)^\tau\Big|_{\Sigma_E,\Phi}$ *with* $\Sigma_E = \Phi\Sigma_E\Phi^\tau + \Sigma_\epsilon$.

Assuming that Assumptions 1, 2, and 3 are satisfied, we can establish asymptotic normality. Furthermore, the MLEs for $\sigma_\varepsilon^2$, $L$, $\Sigma_F$, and $\Lambda$ are independent of each other, given their mutual independence. However, the MLEs for $\Sigma_E$ and $\Phi$ are not independent.

### 2.4 Block Newton-Raphson algorithm for MLE of Θ

Let $\widehat{\Theta}$ be the MLE of $\Theta$. The score function is denoted by $\boldsymbol{\nabla}\mathbb{Q}(x|\widehat{\Theta}, \Theta) = \boldsymbol{\nabla}\mathbb{L}\Big(\Theta + x\big(\widehat{\Theta} - \Theta\big)\Big)$. By substituting the Hessian matrix, the Newton-Raphson algorithm can be used to obtain the MLE of $\Theta$ by utilizing the Fisher information matrix. Furthermore, due to the Fisher information matrix's beneficial structure, the block Newton-Raphson algorithm for MLE of $\Theta$ can be implemented using Theorem 5.

**Theorem 5.** *Based on the assumptions provided in Assumptions 1, 2, and 3, it can be inferred*

Statistica Sinica: Newly accepted Paper



*that*

$$\widehat{\Theta} = \Theta^{\dagger} - \mathbb{I}^{-1}(\Theta^{\dagger})\boldsymbol{\nabla}\mathbb{Q}(0|\widehat{\Theta}, \Theta^{\dagger}) + \mathbf{O}_P\big(N^{1/2}T^{-3/2} + T^{-1}\big)\mathbf{1}_{1+Tr+Nc+r^2+2c^2},$$

*where* $\Theta^{\dagger}$ *is the truth-value of* $\Theta$.

It should be noted that according to Theorem 5, the Fisher information matrix is a quasi-block-diagonal matrix. The parameter space $\Theta$ is divided into five groups, namely $\sigma^2$, $L$, $\Lambda$, $\Sigma_F$, and $(\Sigma_E, \Phi)$. To compute their MLEs, a block Newton-Raphson algorithm is proposed using equation (A.12) as demonstrated in the proof of Theorem 5. More specifically, updating each parameter group proceeds as follows:

(1) Initialize $\sigma_\varepsilon^{2(0)}$, $L^{(0)}$, $\Lambda^{(0)}$, $\Sigma_F^{(0)}$, and $(\Sigma_E^{(0)}, \Phi^{(0)})$.

(2) Given $\sigma_\varepsilon^{2(m)}$, $L^{(m)}$, $\Lambda^{(m)}$, $\Sigma_F^{(m)}$, and $(\Sigma_E^{(m)}, \Phi^{(m)})$, update $\sigma_\varepsilon^2$ to $\sigma_\varepsilon^{2(m+1)}$ by

$$\begin{aligned}\sigma_\varepsilon^{2(m+1)} &= \sigma_\varepsilon^{2(m)} - \mathcal{I}^{-1}_{\sigma_\varepsilon^2,\sigma_\varepsilon^2}\frac{\boldsymbol{\partial}\mathbb{L}}{\boldsymbol{\partial}\sigma_\varepsilon^2}/T\big|_{\sigma_\varepsilon^{2(m)},L^{(m)},\Lambda^{(m)},\Sigma_F^{(m)},\Sigma_E^{(m)},\Phi^{(m)}} \\ &= \sigma_\varepsilon^{2(m)} - tr\big(\Omega_Y^{-1} - \Omega_Y^{-1}S\Omega_Y^{-1}\big)/tr\big(\Omega_Y^{-2}\big)\big|_{\sigma_\varepsilon^{2(m)},L^{(m)},\Lambda^{(m)},\Sigma_F^{(m)}}.\end{aligned}$$

(3) Given $\sigma_\varepsilon^{2(m+1)}$, $L^{(m)}$, $\Lambda^{(m)}$, $\Sigma_F^{(m)}$, and $(\Sigma_E^{(m)}, \Phi^{(m)})$, update $L$ to $L^{(m+1)}$ by

$$\mathrm{vec}(L^{(m+1)}) = \mathrm{vec}(L^{(m)}) - \mathcal{I}^{-1}_{L,L}\frac{\boldsymbol{\partial}\mathbb{L}}{\boldsymbol{\partial}\mathrm{vec}(L)}/T\big|_{\sigma_\varepsilon^{2(m+1)},L^{(m)},\Lambda^{(m)},\Sigma_F^{(m)},\Sigma_E^{(m)},\Phi^{(m)}},$$

$$L^{(m+1)} = L^{(m)} - \big\{L - \mathrm{trs}\big[K_{NT}\big(\Omega_Y^{-1}S\big)K_{TN}\big]L/N\big\}\Sigma_F^{-1}/T\big|_{\sigma_\varepsilon^{2(m+1)},L^{(m)},\Lambda^{(m)},\Sigma_F^{(m)}}.$$

Statistica Sinica: Newly accepted Paper



(4) Given $\sigma_\varepsilon^{2(m+1)}$, $L^{(m+1)}$, $\Lambda^{(m)}$, $\Sigma_F^{(m)}$, and $(\Sigma_E^{(m)}, \Phi^{(m)})$, update $\Lambda$ to $\Lambda^{(m+1)}$ by

$$\mathrm{vec}(\Lambda^{(m+1)}) = \mathrm{vec}(\Lambda^{(m)}) - \mathcal{I}_{\Lambda,\Lambda}^{-1}\frac{\partial \mathbb{L}}{\partial \mathrm{vec}(\Lambda)}/T\Big|_{\sigma_\varepsilon^{2(m+1)},L^{(m+1)},\Lambda^{(m)},\Sigma_F^{(m)},\Sigma_E^{(m)},\Phi^{(m)}},$$

$$\Lambda^{(m+1)} = \Lambda^{(m)} - \big\{\Lambda - \mathrm{trs}\big[\big(\Omega_Y^{-1}S\big)\big]\Lambda/T\big\}\Sigma_E^{-1}/N\Big|_{\sigma_\varepsilon^{2(m+1)},L^{(m+1)},\Lambda^{(m)},\Sigma_F^{(m)},\Sigma_E^{(m)}}.$$

(5) Given $\sigma_\varepsilon^{2(m+1)}$, $L^{(m+1)}$, $\Lambda^{(m+1)}$, $\Sigma_F^{(m)}$, and $(\Sigma_E^{(m)}, \Phi^{(m)})$, update $\Sigma_F$ to $\Sigma_F^{(m+1)}$ by

$$\mathrm{diag}(\Sigma_F^{(m+1)}) = \mathrm{diag}(\Sigma_F^{(m)}) - \mathcal{I}_{\Sigma_F,\Sigma_F}^{-1}\frac{\partial \mathbb{L}}{\partial \mathrm{diag}(\Sigma_F)}/T\Big|_{\sigma_\varepsilon^{2(m+1)},L^{(m+1)},\Lambda^{(m+1)},\Sigma_F^{(m)},\Sigma_E^{(m)},\Phi^{(m)}},$$

$$\Sigma_F^{(m+1)} = \mathrm{diag}\big[L^{(m+1)\tau}\mathrm{trs}\big(K_{Nr}SK_{rN}\big)L^{(m+1)}\big]/N.$$

(6) Denote $VC = (\Sigma_E, \Phi)$, $VC^{(m)} = (\Sigma_E^{(m)}, \Phi^{(m)})$, and $\mathcal{S}_{VC} = \mathrm{vec}(\frac{\partial \mathbb{L}}{\partial \mathrm{vec}(\Sigma_E)}^\tau, \frac{\partial \mathbb{L}}{\partial \mathrm{vec}(\Phi)}^\tau)^\tau$. Given $\sigma_\varepsilon^{2(m+1)}$, $L^{(m+1)}$, $\Lambda^{(m+1)}$, $\Sigma_F^{(m+1)}$, and $VC^{(m)}$, update $VC$ to $VC^{(m+1)}$ by

$$\mathrm{vec}(VC^{(m+1)}) = \mathrm{vec}(VC^{(m)}) - \mathcal{J}^{-1}\mathcal{S}_{VC}\Big|_{\sigma_\varepsilon^{2(m+1)},\Lambda^{(m+1)},\Sigma_E^{(m)},\Phi^{(m)}}.$$

**Remark 5.** Since

$$\mathcal{J}^{-1} = \frac{1}{T}\begin{pmatrix} \big(\mathcal{I}_{\Sigma_E,\Sigma_E} - \mathcal{I}_{\Sigma_E,\Phi}\mathcal{I}_{\Phi,\Phi}^{-1}\mathcal{I}_{\Phi,\Sigma_E}\big)^{-1} & -\mathcal{I}_{\Sigma_E,\Sigma_E}^{-1}\mathcal{I}_{\Sigma_E,\Phi}\big(\mathcal{I}_{\Phi,\Phi} - \mathcal{I}_{\Phi,\Sigma_E}\mathcal{I}_{\Sigma_E,\Sigma_E}^{-1}\mathcal{I}_{\Sigma_E,\Phi}\big)^{-1} \\ -\mathcal{I}_{\Phi,\Phi}^{-1}\mathcal{I}_{\Phi,\Sigma_E}\big(\mathcal{I}_{\Sigma_E,\Sigma_E} - \mathcal{I}_{\Sigma_E,\Phi}\mathcal{I}_{\Phi,\Phi}^{-1}\mathcal{I}_{\Phi,\Sigma_E}\big)^{-1} & \big(\mathcal{I}_{\Phi,\Phi} - \mathcal{I}_{\Phi,\Sigma_E}\mathcal{I}_{\Sigma_E,\Sigma_E}^{-1}\mathcal{I}_{\Sigma_E,\Phi}\big)^{-1} \end{pmatrix},$$

it can be directly utilized in step (6) of the block Newton-Raphson algorithm.

**Remark 6.** Since $\mathbb{E}(Y^\tau Y) = L\mathbb{E}(F^\tau F)L^\tau + \underline{N\sigma_\varepsilon^2\mathbb{E}(EE^\tau)} + \mathbb{E}(\varepsilon\varepsilon^\tau)$, the underlined term constitutes a main component of $\mathbb{E}(Y^\tau Y)$ and is non-negligible. This indicates that principal component analysis (PCA) cannot be directly employed to estimate $L$. A similar result holds

Statistica Sinica: Newly accepted Paper

for the estimation of $\Lambda$. However, to obtain initial values, it can be rough to use PCA of the sample covariance matrix to obtain an initial estimate of the factor matrices and loadings. Under the identification condition, two steps can be taken to compute the initial values of all parameters. In the first step, equation $Y = \Lambda E^\tau + \xi$ is considered, where $\xi = FL^\tau + \varepsilon$. The eigenvectors corresponding to the $c$ largest eigenvalues of the matrix $YY^\tau$ are then used to estimate the factor matrix $\widehat{E}_0$ and the corresponding factor loadings $\widehat{\Lambda}_0$ [See Bai (2003) for details]. In the second step, we define $Y^*$ as $Y - \widehat{\Lambda}_0 \widehat{E}_0^\tau = FL^\tau + \varepsilon$. Subsequently, $\widehat{L}_0$ and $\widehat{F}_0$ can be estimated using $Y^{*\tau} Y^*$. Next, $\widehat{\Phi}_0$ can be computed by obtaining the least squares estimate of $\widehat{E}_0$. Finally, we can choose an initial estimate of $\sigma^2$ using the formula $\widehat{\sigma}_0^2 = \mathrm{tr}\Big((Y - \widehat{F}_0 \widehat{L}_0^\tau - \widehat{\Lambda}_0 \widehat{E}_0^\tau)(Y - \widehat{F}_0 \widehat{L}_0^\tau - \widehat{\Lambda}_0 \widehat{E}_0^\tau)^\tau\Big)/(NT)$.

## 3. Simulation and Application

In this section, we will begin by conducting simulation experiments to demonstrate the efficacy of estimating $\Theta$. Furthermore, the novel model will be employed to analyze a real-world application.

### 3.1 Simulation Experiments for Estimation Accuracy

To illustrate the asymptotic properties discussed in the previous section, we perform several simulation experiments. We generate data from model (1.1) with $r = c = 2$, as outlined in the main text. The sample size $T$ varies between $50, 100, 200, 500$, and $1000$, while the

Statistica Sinica: Newly accepted Paper



dimensions $N$ are set at $0.5T, T$, and $1.5T$. All elements $T \times r$ of $L$ and elements $N \times c$ of $\Lambda$ are independently drawn from the uniform distribution on the interval $[-1, 1]$. The columns of $L$ and $\Lambda$ are then rescaled by $T^{1/2}$ and $N^{1/2}$, respectively, to satisfy the identification condition (IC) of model (1.1). To replicate the simulation for various $T$ and $N$, we repeat the experiment 200 times for two cases: one under the Normal distribution and the other under the Student-$t$ distribution.

The factor $E_t$ is generated as a 2-dimensional VAR(1). In the first case, the two components of $\epsilon_t$ follow independent normal distributions $\mathbb{N}(0, 5)$ and $\mathbb{N}(0, 4)$, respectively. In the second case, they follow independent Student $t$ distributions $\mathbb{T}(3)$. The factor $F_i$ is generated as a 2-dimensional vector, with its components independently following $\mathbb{N}(0, 3)$ and $\mathbb{N}(0, 1)$ in the first case, and $\mathbb{T}(4)$ and $\mathbb{T}(10)$ in the second case, respectively. In the proposed model (1.1), the values $\{\varepsilon_{it}\}$ are independent across $i$ and $t$. In the first case, they follow normal distributions $\mathbb{N}(0, 1)$, while in the second case, they follow standardized Student-$t$ distributions $\mathbb{T}(5)$. We set the autoregressive coefficient matrix as $\Phi = (\phi_1, \phi_2)^\tau \in \mathbb{R}^{2\times 2}$, where $\phi_1 = (-0.6, 0)^\tau$ and $\phi_2 = (0, 0.4)^\tau$ in the first case and $\phi_1 = (0.7, 0)^\tau$ and $\phi_2 = (0, -0.4)^\tau$ in the second case.

With a sample size of 200, Tables 2-5 demonstrate reliable inference under the Normal and Students $t$ distributions. As $N$ and $T$ increase, Tables 2 and 4 present the estimated means and standard deviations of all parameters, showing that the means converge to their true values while the standard deviations decrease progressively.

To evaluate the accuracy of the parameter estimates, we compute the mean squared error

Statistica Sinica: Newly accepted Paper



(MSE) and mean absolute error (MAE) between the estimated parameters and their true counterparts, as reported in Tables 3 and 5. These tables further reveal that both MSE and MAE remain remarkably small and diminish gradually with increasing row and column dimensions.

Taken together, the findings summarized in simulation results (Tables 2-5) confirm that the block NewtonCRaphson in estimating $\Theta$ algorithm delivers stable and accurate estimates across varying sample sizes and distributions.

### 3.2 Simulation Experiments for initial estimates by a two-step PCA procedure

Although PCA is unsuitable as a primary method, we employ a two-step PCA procedure only to generate initial estimates for the MLE algorithm. This is motivated by computational stability and benchmarking utility. MLE for high-dimensional models often suffers from local optima; PCA provides a quick, rough estimate of $L$ and $\Lambda$ to start the iterative optimization. As shown in Tables 6 and 7, the two-step PCA estimates (with factor numbers predetermined) serve as adequate initial values – their proximity to the true parameters ensures the MLE converges efficiently.

To highlight the practical advantages of MLE, we compare it with the two-step PCA procedure (used here as a standalone estimator, not just for initialization) in Tables 6 and 7. The findings indicate that MLE exhibits lower estimation error and robustness to model complexity. The MLE method attains significantly lower MSE and MAE than the two-step PCA

Statistica Sinica: Newly accepted Paper



Table 2: Performance Evaluation of Simulation Study Results Based on 200 Replicates under Normal distribution

| | | $\sigma_\varepsilon^2$ | $\Sigma_{F,11}$ | $\Sigma_{F,22}$ | $\Sigma_{E,11}$ | $\Sigma_{E,22}$ | $\Phi_{11}$ | $\Phi_{22}$ |
|---|---|---|---|---|---|---|---|---|
| $T$ | $N$ | 1 | 3 | 1 | 7.8125 | 4.7619 | -0.6 | 0.4 |
| | $0.5T$ | 0.9446(0.2461) | 3.2781(1.0461) | 1.4204(1.0842) | 8.0557(1.4494) | 5.1314(1.7032) | -0.5443(0.2690) | 0.4103(0.3166) |
| 50 | $T$ | 0.9464(0.1172) | 3.1359(0.8208) | 1.3044(0.8820) | 8.0991(1.5081) | 5.2258(1.7854) | -0.5359(0.2749) | 0.4391(0.2711) |
| | $1.5T$ | 0.9669(0.1291) | 3.1289(0.5245) | 1.2328(0.6345) | 8.1163(1.4018) | 5.3203(1.8491) | -0.5369(0.2848) | 0.4646(0.2756) |
| | $0.5T$ | 0.9472(0.0312) | 3.1189(0.6112) | 1.1258(0.2682) | 8.2173(1.5269) | 5.5657(1.8035) | -0.5578(0.2556) | 0.4983(0.2514) |
| 100 | $T$ | 0.9657(0.0224) | 3.1786(0.4399) | 1.1044(0.2025) | 8.3272(1.4952) | 5.6816(1.9640) | -0.5998(0.2366) | 0.4861(0.2391) |
| | $1.5T$ | 0.9714(0.0254) | 3.1890(0.3498) | 1.1229(0.1902) | 8.2503(1.5531) | 5.4351(1.9541) | -0.5607(0.2531) | 0.4839(0.2370) |
| | $0.5T$ | 0.9698(0.0098) | 3.1500(0.4596) | 1.0840(0.1492) | 7.8146(1.5829) | 4.9963(1.7831) | -0.5337(0.1982)) | 0.4207(0.0790) |
| 200 | $T$ | 0.9811(0.0072) | 3.1290(0.3224) | 1.0787(0.1255) | 7.7287(1.4338) | 5.1683(1.6001) | -0.5260(0.1933) | 0.4406(0.2093) |
| | $1.5T$ | 0.9843(0.0058) | 3.0757(0.2524) | 1.0793(0.1047) | 7.7345(1.1625) | 5.2624(1.2125) | -0.5824(0.0617) | 0.4683(0.1533) |
| | $0.5T$ | 0.9885(0.0040) | 3.0883(0.2753) | 1.0543(0.1039) | 7.9439(0.8857) | 5.3416(1.2273) | -0.5951(0.0332) | 0.4562(0.1451) |
| 500 | $T$ | 0.9925(0.0027) | 3.0558(0.1797) | 1.0318(0.0673) | 7.8959(0.7988) | 5.3478(1.1780) | -0.5980(0.2292) | 0.4747(0.1341) |
| | $1.5T$ | 0.9995(0.0025) | 3.0271(0.1598) | 1.0199(0.0560) | 7.8708(0.7063) | 4.7550(1.0410) | -0.6009(0.0151) | 0.4019(0.1262) |
| | $0.5T$ | 0.9937(0.0019) | 3.0454(0.2082) | 1.0186(0.0684) | 7.8780(0.6691) | 5.1902(1.2178) | -0.5983(0.0229) | 0.4447(0.1711) |
| 1000 | $T$ | 1.0001(0.0014) | 3.0197(0.1296) | 1.0121(0.0501) | 7.8657(0.5089) | 4.7466(0.6745) | -0.5995(0.0091) | 0.3943(0.0764) |
| | $1.5T$ | 1.0002(0.0012) | 3.0125(0.1160) | 1.0127(0.0372) | 7.8382(0.4687) | 4.8332(0.7318) | -0.5991(0.0098) | 0.4066(0.0851) |

Statistica Sinica: Newly accepted Paper



Table 3: MSE and MAE for Estimating Results of The Simulation Study Based on 200 Replicates under Normal distribution

| | | $\sigma_\varepsilon^2$ | $L$ | $\Sigma_F$ | $\Lambda$ | $\Sigma_E$ | $\Phi$ |
|---|---|---|---|---|---|---|---|
| $T$ | $N$ | | | MSE | (MAE) | | |
| 50 | $0.5T$ | 0.0633(0.1367) | 0.4534(0.3704) | 1.2563(0.6564) | 0.2191(0.2062) | 2.5863(1.1056) | 0.0874(0.1978) |
| | $T$ | 0.0165(0.0752) | 0.2394(0.2362) | 0.7778(0.4756) | 0.1393(0.1467) | 2.8663(1.2353) | 0.0770(0.1911) |
| | $1.5T$ | 0.0177(0.0716) | 0.2060(0.2004) | 0.3726(0.3526) | 0.1331(0.1363) | 2.8809(1.2602) | 0.0822(0.1971) |
| 100 | $0.5T$ | 0.0039(0.0572) | 0.0876(0.1430) | 0.2366(0.3536) | 0.0655(0.0827) | 3.1833(1.3681) | 0.0696(0.1876) |
| | $T$ | 0.0016(0.0376) | 0.1105(0.1229) | 0.1381(0.2615) | 0.0849(0.0809) | 3.5866(1.4183) | 0.0600(0.1752) |
| | $1.5T$ | 0.0014(0.0332) | 0.0866(0.1008) | 0.1043(0.2331) | 0.0532(0.0654) | 3.4243(1.4506) | 0.0641(0.1783) |
| 200 | $0.5T$ | 0.0010(0.0302) | 0.1057(0.1081) | 0.1309(0.2537) | 0.0110(0.0319) | 2.8559(1.2615) | 0.0454(0.1437) |
| | $T$ | 0.0004(0.0189) | 0.0331(0.0586) | 0.0709(0.1904) | 0.0505(0.0461) | 2.3826(1.1528) | 0.0439(0.1391) |
| | $1.5T$ | 0.0003(0.0157) | 0.0123(0.0418) | 0.0432(0.1595) | 0.0307(0.0370) | 1.5320(0.8765) | 0.0160(0.0790) |
| 500 | $0.5T$ | 0.0001(0.0115) | 0.0423(0.0567) | 0.0485(0.1643) | 0.0302(0.0283) | 1.3164(0.7988) | 0.0126(0.0623) |
| | $T$ | 6e-05(0.0075) | 0.0611(0.0538) | 0.0203(0.1052) | 0.0302(0.0279) | 1.1829(0.7499) | 0.0121(0.0624) |
| | $1.5T$ | 7e-06(0.0020) | 0.0409(0.0398) | 0.0148(0.0875) | 0.0104(0.0189) | 0.7890(0.7066) | 0.0080(0.0549) |
| 1000 | $0.5T$ | 4e-05(0.0063) | 0.0611(0.0537) | 0.0251(0.1136) | 0.0201(0.0193) | 1.0543(0.6696) | 0.0158(0.0619) |
| | $T$ | 2e-06(0.0012) | 0.0206(0.0285) | 0.0098(0.0696) | 0.0002(0.0104) | 0.3567(0.4471) | 0.0029(0.0346) |
| | $1.5T$ | 1e-06(0.0009) | 0.0290(0.0441) | 0.0075(0.0617) | 0.0002(0.0104) | 0.3786(0.4790) | 0.0037(0.0381) |

Statistica Sinica: Newly accepted Paper



Table 4: Performance Evaluation of Simulation Study Results Based on 200 Replicates under Student-$t$ distribution

| | | $\sigma_\varepsilon^2$ | $\Sigma_{F,11}$ | $\Sigma_{F,22}$ | $\Sigma_{E,11}$ | $\Sigma_{E,22}$ | $\Phi_{11}$ | $\Phi_{22}$ |
|---|---|---|---|---|---|---|---|---|
| $T$ | $N$ | 1 | 2 | 1.25 | 5.8824 | 3.5714 | 0.7 | -0.4 |
| 50 | $0.5T$ | 0.9798(0.3332) | 2.3683(2.4146) | 1.6367(1.1496) | 6.6592(1.8693) | 4.2476(1.8506) | 0.6775(0.2343) | -0.4414(0.3256) |
| | $T$ | 0.9968(0.2537) | 2.2215(1.2039) | 1.7057(1.1257) | 6.2376(1.7601) | 4.1540(1.7267) | 0.6179(0.3226) | -0.5011(0.2914) |
| | $1.5T$ | 0.9912(0.2277) | 2.4325(1.5417) | 1.6160(0.9920) | 5.8755(1.8605) | 3.8434(1.7554) | 0.7032(0.2009) | -0.4487(0.3124) |
| 100 | $0.5T$ | 0.9912(0.2448) | 2.2257(1.0211) | 1.4857(0.6644) | 6.2125(1.5389) | 4.2218(1.4871) | 0.6878(0.1804) | -0.4826(0.3076) |
| | $T$ | 0.9764(0.1342) | 2.2122(0.8628) | 1.4937(0.9936) | 5.7009(1.3624) | 3.8712(1.3647) | 0.6814(0.1841) | -0.4866(0.2711) |
| | $1.5T$ | 1.0099(0.2099) | 2.3667(1.2184) | 1.4571(0.7130) | 5.6356(1.2783) | 3.6486(1.3035) | 0.6460(0.2663) | -0.4429(0.2920) |
| 200 | $0.5T$ | 0.9844(0.1143) | 2.0621(0.7496) | 1.3940(0.5485) | 5.7713(0.9878) | 3.7165(1.1944) | -0.5337(0.1982)) | -0.4334(0.2635) |
| | $T$ | 0.9908(0.0966) | 2.1357(0.7688) | 1.3149(0.1935) | 5.5577(0.8449) | 3.5735(1.0769) | 0.6824(0.1531) | -0.4943(0.2234) |
| | $1.5T$ | 0.9876(0.0322) | 2.1001(0.738) | 1.4222(0.8398) | 5.6769(0.9315) | 3.4118(0.9795) | 0.6831(0.1237) | -0.4836(0.1594) |
| 500 | $0.5T$ | 0.9889(0.0098) | 2.0906(0.5210) | 1.2993(0.1658) | 5.6829(0.7011) | 3.4376(0.8287) | 0.6927(0.1139) | -0.4603(0.1362) |
| | $T$ | 0.9927(0.0059) | 2.0594(0.5155) | 1.2845(0.0992) | 5.6726(0.6946) | 3.4849(0.8610) | 0.6991(0.0567) | -0.4632(0.1140) |
| | $1.5T$ | 0.9986(0.0051) | 2.0270(0.2422) | 1.2712(0.0838) | 5.7443(0.951) | 3.0649(0.9316) | 0.7020(0.0373) | -0.3988(0.1391) |
| 1000 | $0.5T$ | 0.9941(0.0043) | 2.0326(0.3276) | 1.2701(0.0949) | 5.7714(0.5482) | 3.6262(0.7239) | 0.6938(0.0797) | -0.4664(0.0933) |
| | $T$ | 1.0001(0.0024) | 2.0124(0.2337) | 1.2553(0.0662) | 5.6986(0.7773) | 3.1728(0.8483) | 0.7030(0.0299) | -0.3959(0.0987) |
| | $1.5T$ | 0.9999(0.0022) | 2.0157(0.2003) | 1.2596(0.0507) | 5.6970(0.0.9356) | 3.0913(0.8318) | 0.6980(0.0788) | -0.3825(0.1028) |

Statistica Sinica: Newly accepted Paper



Table 5: MSE and MAE for Estimating Results of The Simulation Study Based on 200 Replicates under Student-$t$ distribution

| | | $\sigma_\varepsilon^2$ | $L$ | $\Sigma_F$ | $\Lambda$ | $\Sigma_E$ | $\Phi$ |
|---|---|---|---|---|---|---|---|
| $T$ | $N$ | | | MSE | (MAE) | | |
| 50 | $0.5T$ | 0.1108(0.1893) | 0.6016(0.4390) | 1.7956(0.7961) | 0.3676(0.2932) | 3.9726(1.5217) | 0.0812(0.1935) |
| | $T$ | 0.0640(0.1249) | 0.4638(0.3245) | 1.4799(0.6503) | 0.2494(0.2161) | 3.2574(1.3762) | 0.1025(0.2127) |
| | $1.5T$ | 0.0516(0.1030) | 0.3292(0.2537) | 1.8326(0.6270) | 0.2682(0.2144) | 3.2923(1.3824) | 0.0698(0.1862) |
| 100 | $0.5T$ | 0.0597(0.1054) | 0.3658(0.2650) | 0.7916(0.4877) | 0.2064(0.1576) | 2.5444(1.1816) | 0.0667(0.1718) |
| | $T$ | 0.0184(0.0518) | 0.1556(0.1450) | 0.9138(0.4409) | 0.1026(0.1002) | 1.9115(1.0391) | 0.0573(0.1610) |
| | $1.5T$ | 0.0439(0.0695) | 0.1884(0.1438) | 1.0803(0.4393) | 0.1250(0.1127) | 1.6918(0.9659) | 0.0801(0.1703) |
| 200 | $0.5T$ | 0.0134(0.0429) | 0.1412(0.1271) | 0.4415(0.3259) | 0.0731(0.0704) | 1.2120(0.8361) | 0.0433(0.1341) |
| | $T$ | 0.0093(0.0290) | 0.0567(0.0719) | 0.3240(0.2759) | 0.0388(0.0518) | 0.9849(0.7371) | 0.0411(0.1275) |
| | $1.5T$ | 0.0025(0.0190) | 0.0741(0.0727) | 0.3023(0.2856) | 0.0865(0.0726) | 0.9428(0.7370) | 0.0239(0.0923) |
| 500 | $0.5T$ | 0.0002(0.0123) | 0.0721(0.0696) | 0.0457(0.1404) | 0.0302(0.0284) | 1.1517(0.7541) | 0.0107(0.0571) |
| | $T$ | 9e-05(0.0079) | 0.0411(0.0458) | 0.1394(0.1683) | 0.0404(0.0358) | 0.6346(0.5845) | 0.0101(0.0607) |
| | $1.5T$ | 8e-06(0.0021) | 0.0308(0.0358) | 0.0134(0.0825) | 0.0303(0.0274) | 0.8115(0.7042) | 0.0074(0.0533) |
| 1000 | $0.5T$ | 5e-05(0.0062) | 0.0212(0.0372) | 0.0586(0.1539) | 0.0202(0.0212) | 0.4179(0.4561) | 0.0097(0.0528) |
| | $T$ | 6e-06(0.0019) | 0.0106(0.0243) | 0.0294(0.1128) | 0.0302(0.0244) | 0.7550(0.6656) | 0.0053(0.0444) |
| | $1.5T$ | 5e-06(0.0017) | 0.0304(0.0293) | 0.0242(0.0941) | 0.0204(0.0225) | 0.8766(0.6943) | 0.0059(0.0492) |

Statistica Sinica: Newly accepted Paper



Table 6: Performance Evaluation of Simulation Study Results Based on 200 Replicates under Normal distribution Using a Two-Step PCA Procedure

| | | $\sigma_\varepsilon^2$ | $\Sigma_{F,11}$ | $\Sigma_{F,22}$ | $\Sigma_{E,11}$ | $\Sigma_{E,22}$ | $\Phi_{11}$ | $\Phi_{22}$ |
|---|---|---|---|---|---|---|---|---|
| $T$ | $N$ | 1 | 3 | 1 | 7.8125 | 4.7619 | -0.6 | 0.4 |
| 50 | $T$ | 0.8431(0.0269) | 2.6165(0.5504) | 0.9256(0.1881) | 8.1874(2.0932) | 4.8511(0.9082) | -0.3935(0.3307) | 0.1882(0.2487) |
| 100 | $T$ | 0.9210(0.0135) | 2.7900(0.3869) | 0.9563(0.1323) | 8.0293(1.4916) | 4.8479(0.7052) | -0.5215(0.1809) | 0.3216(0.1487) |
| 200 | $T$ | 0.9600(0.0072) | 2.8756(0.3065) | 0.9845(0.1058) | 7.8649(1.1392) | 4.8604(0.5093) | -0.5657(0.1052) | 0.3584(0.0897) |
| 500 | $T$ | 0.9842(0.0028) | 2.9647(0.1866) | 0.9873(0.0625) | 7.8814(0.7143) | 4.8027(0.3546) | -0.5896(0.0418) | 0.3915(0.0423) |
| 1000 | $T$ | 0.9841(0.0027) | 2.9481(0.1921) | 0.9948(0.0591) | 7.8809(0.6272) | 4.7766(0.3253) | -0.5916(0.0365) | 0.3860(0.0406) |

procedure (which excludes $\Sigma_E$ and $\Phi$). For example, when estimating the loading matrix $L$ or $\Lambda$, MLE reduces MSE or MAE obviously. The advantage of MLE is most pronounced, where PCA's confounding of temporal/cross-sectional factors leads to severe bias.

In short, while PCA can initialize MLE, it fails as a primary method due to structural bias. The simulation results confirm that MLE's likelihood-based approach better captures the two-way factor structure, yielding more accurate and efficient estimates.

### 3.3 One Real Data Example

The use of factors derived from a large panel data set as explanatory variables has been shown in several recent works to be highly valuable in VAR modeling, as discussed in a comprehensive analysis and review of the literature by Bernanke et al. (2005). In this section, we estimate the factor matrices and factor loadings for our model using the FRED-MD dataset, which is available at $http://research.stlouisfed.org/econ/mccracken/fred-md/$ and includes 123

Statistica Sinica: Newly accepted Paper



Table 7: MSE and MAE of Estimation Results from a Simulation Study with 200 Replicates Under a Normal Distribution Using a Two-Step PCA Procedure

| | | $\sigma_\varepsilon^2$ | $L$ | $\Sigma_F$ | $\Lambda$ | $\Sigma_E$ | $\Phi$ |
|---|---|---|---|---|---|---|---|
| $T$ | $N$ | | | MSE | (MAE) | | |
| 50 | $T$ | 0.0253(0.1569) | 1.9271(0.9676) | 0.2485(0.3670) | 2.0607(1.0280) | 2.7646(1.1925) | 0.1289(0.2470) |
| 100 | $T$ | 0.0065(0.0796) | 1.9456(0.9425) | 0.1073(0.2377) | 1.7807(0.8878) | 1.4884(0.9192) | 0.0491(0.1404) |
| 200 | $T$ | 0.0016(0.0400) | 1.9066(0.9030) | 0.0602(0.1790) | 2.1234(0.9886) | 0.7810(0.6623) | 0.0109(0.0646) |
| 500 | $T$ | 0.0003(0.0158) | 1.8794(0.8638) | 0.0200(0.1013) | 1.8615(0.8520) | 0.3196(0.4305) | 0.0019(0.0337) |
| 1000 | $T$ | 0.0003(0.0159) | 1.9503(0.8934) | 0.0215(0.1003) | 1.7631(0.8110) | 0.2508(0.3834) | 0.0016(0.0320) |

monthly time series spanning the period from January 1991 to December 2020. To ensure the reliability of our results, the variables in the data set are first transformed into a stationary form and outliers are removed according to the approach described in McCracken and Ng (2015). Additionally, we utilize double-demeaned variables, as recommended by Ahn and Horenstein (2013), to eliminate variable-specific and time effects. Subsequently, we conduct likelihood inference to model this subset of the transformed data.

To obtain the initial values of $\Theta$, a two-step PCA of the sample covariance matrix is utilized. Following the approach suggested by Ahn and Horenstein (2013), the eigenvalue ratio method indicates that $r = c = 2$. Subsequently, the Newton-Raphson block algorithm is used to calculate $\widehat{\Theta}$, and the final estimates are reported in Table 8. The results reveal a significant reduction in dimensionality for both cross-sectional and time-series effects.

The variances of $F_i$ differ noticeably from those of $E_t$. To showcase the utility and benefits

Statistica Sinica: Newly accepted Paper

Table 8: The Estimating Results of The Real Data Analysis

| $\sigma_\varepsilon^2$ | $\Sigma_{F,11}$ | $\Sigma_{F,22}$ | $\Sigma_{E,11}$ | $\Sigma_{E,22}$ | $\Phi_{11}$ | $\Phi_{12}$ | $\Phi_{21}$ | $\Phi_{22}$ |
|---|---|---|---|---|---|---|---|---|
| 0.0130 | 10.2967 | 4.2579 | 0.7012 | 0.6632 | 0.4452 | 0.0239 | 0.2802 | 0.8766 |

of our approach, we perform an out-of-sample forecast and compare it with the 2wFM of Gao et al. (2021). We utilize a $T \times N$ matrix, represented by $\left(y_{i,t}\right)_{1:T,1:N}$, which contains the re-centered data. We divide the data set into two segments: the first part, $\left(y_{i,t}\right)_{1:T-3,1:N-3}$, serves as the training set, while the remaining portion is used as the test set.

Let $\left(\widehat{y}_{i,t}\right)_{T-2:T,N-2:N}$ represent the predicted values, where $i = 1,2,3$ and $t = 1,2,3$. The squared prediction error is then calculated as $\mathcal{E}_{i,t} \triangleq \|\widehat{y}_{N-3+i,T-3+t} - y_{N-3+i,T-3+t}\|^2$. The average mean squared prediction error (MSPE) is obtained by summing all $\mathcal{E}_{i,t}$ values and dividing by $3(N+T-3)$, resulting in MSPE. It is notable that the predictions of new observations in the 2wFM framework rely solely on the mean level. The proposed model yields an MSPE of 0.0475, representing a 8.8% reduction relative to the benchmark 2wFM (MSPE = 0.0521), indicating improved forecasting performance. These results indicate that explicit AR modeling of the factor series facilitates the prediction and results in satisfactory forecasting. Collectively, our approach delivers robust empirical validation through real data analysis.

## 4. Conclusion

In this paper, we present a novel model-based method for analyzing high-dimensional panel time series. Unlike the two-way factor model (2wFM) approach proposed by Gao et al. (2021),

Statistica Sinica: Newly accepted Paper

our method incorporates a two-way factor structure that accounts for both cross-sectional and time-series effects. We investigate likelihood inference to derive asymptotic theories for the model. Our approach assumes that the low-dimensional hidden factors of both rows and columns are influenced by distinct factors. This enables us to incorporate a dimensionality-reduction structure in a vector-valued factor model and establish a new framework for separating row and column factors.

A method known as the traditional delta method, which is heavily based on the score function and the Hessian matrix obtained through a matrix decomposition technique, is recommended for computing the MLE of factor loadings and other model parameters. Under standard identifiability conditions, the consistency and asymptotic normality of the proposed estimators are established. The proposed fast computational algorithms for estimation are well demonstrated through simulation experiments and real data analysis using a diagonal block matrix.

Consistently determining the number of factors in a 2wFM is of utmost importance. Bai and Ng (2002) developed an information criterion for determining the number of factors in a vector factor model. However, for 2wFM, further investigation is required to develop an information criterion approach based on residuals and the number of factors $(r, c)$.

Statistica Sinica: Newly accepted Paper



**Acknowledgments**

We sincerely thank Co-Editor Prof. John Stufken, the Associate Editor, and two referees for their thorough review and perceptive comments that greatly strengthened this manuscript. We are also grateful to Prof. Jianhua Guo for his helpful suggestions. This work was supported in part by the National Statistical Scientific Research Projects (No. 2025LZ029), the National Natural Science Foundation of China (Nos. 12171161 and No. 12271046), and the Technology Development Plan Project of Jilin Province, China (No. 20240101022JJ).

## 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.

Bernanke, B., Bovian, J. and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach, *Quarterly Journal of Economics* **120**: 387–422.

Statistica Sinica: Newly accepted Paper



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.

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.

Doz, C., Giannone, D. and Reichlin, L. (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models, *The Review of Economics and Statistics* **94**: 1014–1024.

Fan, J., Fan, Y. and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model, *Journal of Econometrics* **147**: 186–197.

Fan, J., Guo, J. and Zheng, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding, *Journal of the American Statistical Association* **117**(538): 852–861.

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.

Gao, Z., Yuan, C., Jing, B., Wei, H. and Guo, J. (2021). A two-way factor model for high-dimensional matrix data, *arXiv.org* .

Statistica Sinica: Newly accepted Paper



He, Y., Kong, X., Yu, L., Zhang, X. and Zhao, C. (2024). Matrix factor analysis: From least squares to iterative projection, *Journal of Business and Economic Statistics* **42**(1): 322–334.

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.

Hornstein, M., Fan, R., Shedden, K. and Zhou, S. (2019). Joint mean and covariance estimation with unreplicated matrix-variate data, *Journal of the American Statistical Association* **114**(526): 682–696.

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

Lee, S. (1978). The gauss-newton algorithm for the weighted least squares factor analysis, *The Statistician* **27**: 103114.

McCracken, M. and Ng, S. (2015). Fred-md: a monthly database for macroeconomic research, *Journal of Business and Economic Statistics* **34**: 574–589.

Ng, C. T., Yau, C. Y. and Chan, N. H. (2015). Likelihood inferences for high-dimensional factor analysis of time series with applications in finance, *Journal of Computational and Graphical Statistics* **24**: 866–884.

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

Statistica Sinica: Newly accepted Paper



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.

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

Xia, Q., Xu, W. and Zhu, L. X. (2015). Consistently determining the number of factors in multivariate volatility modelling, *Statistica Sinica* **25**: 1025–1044.

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

Yuan, C., Gao, Z., He, X., Huang, W. and Guo, J. (2023). Two-way dynamic factor models for high-dimensional matrix-valued time series, *Journal of the Royal Statistical Society: Series B* **85**: 1517–1537.

Zhang, X., Liu, C. C., Guo, J., Yuen, K. C. and Welsh, A. H. (2025). Modeling and learning on high-dimensional matrix-variate sequences, *Journal of the American Statistical Association* **120**: 419–434.