Journal of Econometrics 249 (2025) 105974

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Tensor time series imputation through tensor factor modelling

Zetai Cen[1], Clifford Lam[*,2]

Department of Statistics, London School of Economics and Political Science, United Kingdom

ARTICLE INFO

Keywords:
Generalized cross-covariance matrix
Tensor unfolding
Core tensor
α-mixing time series variables
Missingness tensor

ABSTRACT

We propose tensor time series imputation when the missing pattern in the tensor data can be general, as long as any two data positions along a tensor fibre are both observed for enough time points. The method is based on a tensor time series factor model with Tucker decomposition of the common component. One distinguished feature of the tensor time series factor model used is that there can be weak factors in the factor loading matrix for each mode. This reflects reality better when real data can have weak factors which drive only groups of observed variables, for instance, a sector factor in a financial market driving only stocks in a particular sector. Using the data with missing entries, asymptotic normality is derived for rows of estimated factor loadings, while consistent covariance matrix estimation enables us to carry out inferences. As a first in the literature, we also propose a ratio-based estimator for the rank of the core tensor under general missing patterns. Rates of convergence are spelt out for the imputations from the estimated tensor factor models. Simulation results show that our imputation procedure works well, with asymptotic normality and corresponding inferences also demonstrated. Re-imputation performances are also gauged when we demonstrate that using slightly larger rank then estimated gives superior re-imputation performances. A Fama–French portfolio example with matrix returns and an OECD data example with matrix of economic indicators are presented and analysed, showing the efficacy of our imputation approach compared to direct vector imputation.

1. Introduction

Large dimensional panel data is easier to obtain than ever thanks to a quickly evolving internet speed and more diverse download platforms. Together with the advancement of statistical analyses for these data over the past decade, researchers also open up more to time series data with higher order, namely, tensor time series data. A prominent example would be order 2 tensor time series, i.e., matrix-valued time series. Wang et al. (2019) proposes a factor model using a Tucker decomposition of the common component in the modelling. An example on monthly import–export volume of products among different countries is given in Chen et al. (2022a), where factor modelling using Tucker decomposition is explored, and generalized to higher order tensors. Focusing on matrix-valued time series, Chang et al. (2023) proposes a tensor-CP decomposition for modelling the data. Zhang (2024) and Chen et al. (2021) propose autoregressive models for matrix-valued time series. For a more comprehensive review on matrix-valued time series analysis, please refer to Tsay (2023).

A less addressed topic in large time series analysis is the treatment of missing data, in particular, imputation of missing data and the corresponding inferences. While there are numerous data-centric methods in various scientific fields for imputing multivariate

* Corresponding author.
E-mail addresses: Z.Cen@lse.ac.uk (Z. Cen), C.Lam2@lse.ac.uk (C. Lam).
[1] Zetai Cen is Ph.D. student, Department of Statistics, London School of Economics.
[2] Clifford Lam is Professor, Department of Statistics, London School of Economics.

https://doi.org/10.1016/j.jeconom.2025.105974
Received 22 March 2024; Received in revised form 15 January 2025; Accepted 8 February 2025
Available online 21 February 2025

time series data (see Chapon et al. (2023) for environmental time series, Kazijevs and Samad (2023) for health time series, Zhao et al. (2023) and Zhang et al. (2021) for using deep-learning related architectures for imputations, to name but a few), almost none of them address statistically how accurate their methods are, and all of them are not for higher order tensor time series. We certainly can line up the variables in a tensor time series to make it a longitudinal panel, but in doing so we lose special structures and insights that can be utilized for forecasting and interpretation of the data. More importantly, transforming a moderate sized tensor to a vector means the length of the vector can be much larger than the sample size, creating curse of dimensionality.

For imputing large panel of time series with statistical analyses, Bai and Ng (2021) defines the concept of TALL and WIDE blocks of data and proposes an iterative TW algorithm in imputing missing values in a large panel, while Cahan et al. (2023) improves the TW algorithm to a Tall-Project (TP) algorithm so that there is no iterations needed. Both papers use factor modelling for the imputations, and derive rates of convergence when all factors are pervasive and the number of factors known. Asymptotic normality for rows of estimated factor loadings and the corresponding practical inferences are also developed. Xiong and Pelger (2023) also bases their imputations on a factor model for a large panel of time series with pervasive factors and number of factors known, and build a method for imputing missing values under very general missing patterns, with asymptotic normality and inferences also developed.

To the best of our knowledge, for tensor time series with order larger than 1 (i.e., at least matrix-valued), there are no theoretical analyses on imputation performances. Imputation methodologies developed on tensor time series are also scattered around very different applications. See Chen et al. (2022b) on traffic tensor data and Pan et al. (2021) for RNA-sequence tensor data for instance.

In view of all the above, as a first in the literature, we aim to develop a tensor imputation method accompanied by theoretical analyses in this paper. Like Cahan et al. (2023), we use factor modelling for tensor time series as a basis for our imputation method. Unlike Cahan et al. (2023), Bai and Ng (2021) or Xiong and Pelger (2023) though, we develop a method that can consistently estimate the number of factors, or the core tensor rank, in a Tucker decomposition-based factor model for the tensor time series with missing values. Our method can be considered a combination of He et al. (2022a) for the tensor factor model, and Xiong and Pelger (2023) for the imputation methodology with general missingness. In Section 3, we introduce two motivating examples and our methodology at the same time. One is the Fama–French portfolio return data with missing entries, to be analysed in Section 5.2. The other is a set of monthly and quarterly OECD Economic indicators, with missingness naturally occurring for the quarterly recorded indicators relative to the monthly ones. We analyse this set of OECD data in Section 5.3.

As a further contribution, we also allow factors to be weak. A weak factor corresponds to a column in a factor loading matrix being sparse, or approximately sparse. This implies that not all units in a tensor have dynamics contributed by all the factors inside the core tensor. Chen and Lam (2024b) allows for weak factors in its analyses, and discovers that there are potentially weak factors in the NYC taxi traffic data, which are going to be analysed in the supplementary materials of this paper. We prove consistency of our imputations under general missingness, and develop asymptotic normality and practical inferences for rows of factor loading matrix estimators, with rates of convergence in all consistency results spelt out. Our method is available in the R package `tensorMiss`, which has used the `Rcpp` package to greatly boost computational speed.

The rest of the paper is organized as follows. Section 2 introduces the notations used in this paper. Section 3 presents the Fama–French portfolio returns data and the OECD data as two motivating examples, before describing the tensor factor model and the imputation methodology we employ. Section 4 lays down the main assumptions for the paper, with consistent estimation and rates of convergence of all factor loading matrix estimators and imputed values presented. Asymptotic normality and the estimators of the corresponding asymptotic covariance matrices for practical inferences are also introduced in Section 4.3, before our proposed ratio-based estimators for the number of factors in Section 4.5. Section 5 presents extensive simulation results for our paper, together with an analysis for the Fama–French portfolio return data in Section 5.2 and an analysis for the OECD economic data in Section 5.3. The NYC taxi traffic data, together with extra simulations are presented in the supplementary materials for the paper. All proofs are in the supplementary materials associated with this paper.

2. Notations

Throughout this paper, we use the lower-case letter, bold lower-case letter, bold capital letter, and calligraphic letter, i.e., $x, \mathbf{x}, \mathbf{X}, \mathcal{X}$, to denote a scalar, a vector, a matrix, and a tensor respectively. We also use $x_i, X_{ij}, \mathbf{X}_{i\cdot}, \mathbf{X}_{\cdot i}$ to denote, respectively, the ith element of $\mathbf{x}$, the (i,j)-th element of $\mathbf{X}$, the ith row (as a column vector) of $\mathbf{X}$, and the ith column of $\mathbf{X}$. We use $\otimes$ to represent the Kronecker product, and $\circ$ the Hadamard product. We use $a \asymp b$ to denote $a = O(b)$ and $b = O(a)$. Hereafter, given a positive integer m, define $[m] := \{1, 2, \dots, m\}$. The ith largest eigenvalue of a matrix $\mathbf{X}$ is denoted by $\lambda_i(\mathbf{X})$. The notation $\mathbf{X} \succcurlyeq 0$ (resp. $\mathbf{X} \succ 0$) means that $\mathbf{X}$ is positive semi-definite (resp. positive definite). We use $\mathbf{X}'$ to denote the transpose of $\mathbf{X}$, and $\mathrm{diag}(\mathbf{X})$ to denote a diagonal matrix with the diagonal elements of $\mathbf{X}$, while $\mathrm{diag}(\{x_1, \dots, x_n\})$ represents the diagonal matrix with $\{x_1, \dots, x_n\}$ on the diagonal.

Norm notations: For a given set, we denote by $|\cdot|$ its cardinality. We use $\|\cdot\|$ to denote the spectral norm of a matrix or the L_2 norm of a vector, and $\|\cdot\|_F$ to denote the Frobenius norm of a matrix. We use $\|\cdot\|_{\max}$ to denote the maximum absolute value of the elements in a vector, a matrix or a tensor. The notations $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the L_1 and L_∞-norm of a matrix respectively, defined by $\|\mathbf{X}\|_1 := \max_j \sum_i |X_{ij}|$ and $\|\mathbf{X}\|_\infty := \max_i \sum_j |X_{ij}|$. Without loss of generality, we always assume the eigenvalues of a matrix are arranged by descending orders, and so are their corresponding eigenvectors.

Tensor-related notations: For the rest of this section, we briefly introduce the notations and operations for tensor data, which will be sufficient for this paper. For more details on tensor manipulations, readers are referred to Kolda and Bader (2009). A multidimensional array with K dimensions is an *order-K* tensor, with its kth dimension termed as *mode-k*. For an order-K tensor $\mathcal{X} = (X_{i_1,\dots,i_K}) \in \mathbb{R}^{I_1 \times \cdots \times I_K}$, a column vector $(X_{i_1,\dots,i_{k-1},i,i_{k+1},\dots,i_K})_{i \in [I_k]}$ represents a *mode-k fibre* for the tensor $\mathcal{X}$. We denote by

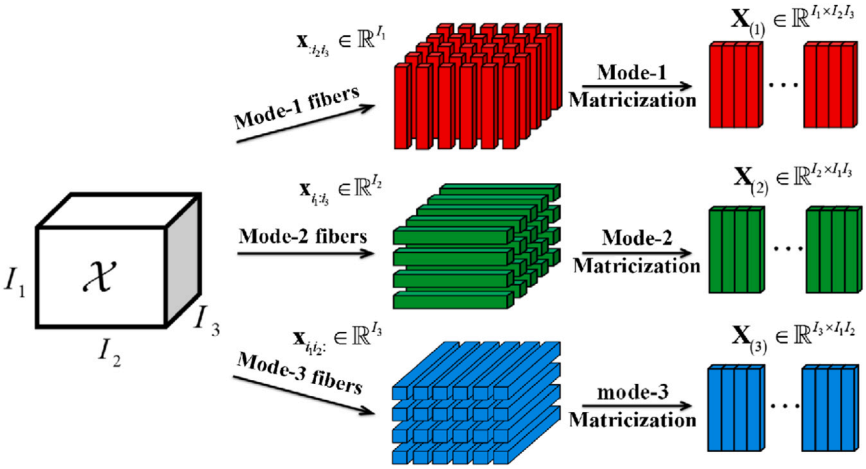

Fig. 1. Illustration of the mode-k fibres and its corresponding unfolding matrix.

$\mathrm{mat}_k(\mathcal{X}) \in \mathbb{R}^{I_k \times I_{\text{-}k}}$ (or sometimes $\mathbf{X}_{(k)}$, with $I_{\text{-}k} := (\prod_{j=1}^{K} I_j)/I_k$) the *mode-k unfolding/matricization* of a tensor, defined by placing all mode-k fibres into a matrix. See Fig. 1 for an illustration (figure from Tao et al. (2019)).

We denote by $\mathcal{X} \times_k \mathbf{A}$ the *mode-k product* of a tensor $\mathcal{X}$ with a matrix $\mathbf{A}$, defined by

$$\mathrm{mat}_k(\mathcal{X} \times_k \mathbf{A}) := \mathbf{A}\,\mathrm{mat}_k(\mathcal{X}).$$

Finally, we use the notation $\mathbf{vec}(\cdot)$ to denote the vectorization of a matrix or the vectorization of the mode-1 unfolding of a tensor.

3. Two motivating examples and the imputation procedure

We first describe two motivating data examples in Sections 3.1 and 3.2, before presenting our imputation procedure for a general order-K mean zero tensor $\mathcal{Y}_t = (\mathcal{Y}_{t,i_1,\ldots,i_K}) \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ for each $t \in [T]$. The two data examples will be analysed in details in Sections 5.2 and 5.3 respectively.

3.1. Example: The Fama–French portfolio returns

This is a set of Fama–French portfolio returns data with missing entries. Stocks are categorized into ten levels of market equity (ME) and ten levels of book-to-equity ratio (BE) which is the book equity for the last fiscal year divided by the end-of-year ME. At the end of June each year, both ME and BE use NYSE deciles as breakpoints, with stocks of NYSE, AMEX and NASDAQ firms allocated accordingly. Moreover, the stocks in each of the 10×10 categories form exactly two portfolios, one being value weighted, and the other of equal weight. Hence, there are two sets of 10 by 10 portfolios with their time series to be studied. We use monthly data from January 1974 to June 2021, so that $T = 570$, and for both value weighted and equal weighted portfolios we have each of our data set as an order-2 tensor $\mathcal{X}_t \in \mathbb{R}^{10\times 10}$ for $t \in [570]$. For more details, please visit

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_100_port_sz.html.

If no stocks are allocated to a category (i.e., intersection of ME and BE categorization) at a timestamp, the corresponding return data is unavailable and hence missing. It is reasonable to argue the missingness might depend on the rows of the loading matrix, i.e., extreme categories tend to contain fewer stocks, but independent of latent factors and noise. The total number of missing entries is 161 and hence the percentage of missing is $161/(10 \times 10 \times 570) = 0.28\%$ for both the value weighted and equal weighted series. However, the irregular missing pattern here can be harmful if we are after a complete case analysis. For full observation after a timestamp, we may only start from July 2009 and hence 74.7% of the data would be ditched. On the other hand, we might ditch four categories to obtain a complete data set but lose the potential insights on the return series of the four categories.

3.2. Example: OECD economic indicators

In this example, we study a group of economic indicators for a selection of countries obtained from the Organization for Economic Co-operation and Development (OECD). The data consists of monthly/quarterly observations of 11 economic indicators: current account balance as percentage of GDP (CA-GDP), consumer price index (CP), merchandise exports (EX), merchandise imports (IM), short-term interest rates (IR3TIB), long-term interest rates (IRLT), interbank rates (IRSTCI), producer price index (PP), production volume (PRVM), retail trade volume (TOVM) and unit labour cost (ULC). They are observed for 17 countries: Belgium (BEL), Canada (CAN), Denmark (DNK), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Italy (ITA), Luxembourg (LUX), Netherlands (NLD), Norway (NOR), Portugal (PRT), Spain (ESP), Sweden (SWE), Switzerland (CHE), United Kingdom (GBR) and United States

(USA), with data spanning from January 1971 to December 2023. We correspond respectively rows and columns to countries and indicators, so that we have our data as an order-2 tensor $\mathcal{Y}_t \in \mathbb{R}^{17\times 11}$ for $t \in [636]$. For more details, see key short-term economic indicators available at https://data.oecd.org/.

The data is naturally missing for three reasons. Firstly, indicator records for some countries at early time periods are unavailable. Secondly, quarterly indicators are only available at the end of each quarter. Finally, these quarterly indicators are sometimes unrecorded. Similar to the Fama–French data, we suppose the missing pattern is dependent on the loading matrices by arguing that relatively less important indicators are only available quarterly. The percentage of missing data is 26.2%, which leads to significantly inefficient use of data if we hope to analyse a set of complete data. The fact that the data is observed at least quarterly in the long run ensures the existence of a lower bound on the proportion of available data, which in turn satisfies Assumption (O1) in Section 4.

3.3. The model and the imputation procedure

The Model: Suppose the order-K mean zero tensor $\mathcal{Y}_t$ is modelled by

$$\mathcal{Y}_t = \mathcal{C}_t + \mathcal{E}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \dots \times_K \mathbf{A}_K + \mathcal{E}_t, \quad t \in [T], \tag{3.1}$$

where $\mathcal{C}_t$ is the common component and $\mathcal{E}_t$ the error tensor. The core tensor is $\mathcal{F}_t \in \mathbb{R}^{r_1\times r_2\times\cdots\times r_K}$, and each mode-$k$ factor loading matrix $\mathbf{A}_k$ has dimension $d_k \times r_k$. See He et al. (2022a) amongst others using the same tensor factor model. Using the QR decomposition, if we can decompose $\mathbf{A}_k = \mathbf{Q}_k\mathbf{Z}_k^{1/2}$ (see Assumption (L1) in Section 4.1 for details), then (3.1) can be written as

$$\begin{aligned}&\mathcal{Y}_t = \mathcal{F}_{Z,t} \times_1 \mathbf{Q}_1 \times_2 \cdots \times_K \mathbf{Q}_K + \mathcal{E}_t, \quad t \in [T], \quad \text{where}\\ &\mathcal{F}_{Z,t} := \mathcal{F}_t \times_1 \mathbf{Z}_1^{1/2} \times_2 \dots \times_K \mathbf{Z}_K^{1/2}.\end{aligned} \tag{3.2}$$

Model (3.1) is an extension to the usual time series factor model ($K = 1$):

$$\mathcal{Y}_t = \mathrm{mat}_1(\mathcal{Y}_t) = \mathrm{mat}_1(\mathcal{F}_t \times_1 \mathbf{A}_1) + \mathrm{mat}_1(\mathcal{E}_t) = \mathbf{A}_1\mathrm{mat}_1(\mathcal{F}_t) + \mathrm{mat}_1(\mathcal{E}_t) = \mathbf{A}_1\mathcal{F}_t + \mathcal{E}_t,$$

and also for a matrix-valued time series factor model ($K = 2$):

$$\mathcal{Y}_t = \mathrm{mat}_1(\mathcal{Y}_t) = \mathbf{A}_1\mathrm{mat}_1(\mathcal{F}_t)\mathbf{A}_2' + \mathrm{mat}_1(\mathcal{E}_t) = \mathbf{A}_1\mathcal{F}_t\mathbf{A}_2' + \mathcal{E}_t.$$

The Imputation Procedure: We only observe partial data. Define the missingness tensor $\mathcal{M}_t = (\mathcal{M}_{t,i_1,\dots,i_K}) \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$ with

$$\mathcal{M}_{t,i_1,\dots,i_K} = \begin{cases} 1, & \text{if } \mathcal{Y}_{t,i_1,\dots,i_K} \text{ is observed;}\\ 0, & \text{otherwise.}\end{cases}$$

Our aim is to recover the value for the common component $\mathcal{C}_{t,i_1,\dots,i_K}$ if $\mathcal{M}_{t,i_1,\dots,i_K} = 0$. Assuming first the number of factors r_k is known for all modes, we want to obtain the estimators of the factor loading matrices, $\widehat{\mathbf{Q}}_k$ for $k \in [K]$, and then the estimated core tensor series $\widehat{\mathcal{F}}_{Z,t}$ for $t \in [T]$. See (3.2) for the definition of $\mathbf{Q}_k$ and $\mathcal{F}_{Z,t}$. We can then estimate the common components at time t by

$$\widehat{\mathcal{C}}_t = \widehat{\mathcal{F}}_{Z,t} \times_1 \widehat{\mathbf{Q}}_1 \times_2 \cdots \times_K \widehat{\mathbf{Q}}_K. \tag{3.3}$$

With (3.3), we can impute $\mathcal{Y}_t$ using

$$\widetilde{\mathcal{Y}}_{t,i_1,\dots,i_K} = \begin{cases} \mathcal{Y}_{t,i_1,\dots,i_K}, & \text{if } \mathcal{M}_{t,i_1,\dots,i_K} = 1;\\ \widehat{\mathcal{C}}_{t,i_1,\dots,i_K}, & \text{if } \mathcal{M}_{t,i_1,\dots,i_K} = 0.\end{cases}$$

We leave the discussion of estimating r_k to Section 4.5. See Section 3.4 in how to obtain $\widehat{\mathbf{Q}}_k$ and Section 3.5 in how to obtain $\widehat{\mathcal{F}}_{Z,t}$.

3.4. Estimation of factor loading matrices

In this paper, we use the following notation:

$$\psi_{k,ij,h} := \left\{ t \in [T] \mid \mathrm{mat}_k(\mathcal{M}_t)_{ih}\mathrm{mat}_k(\mathcal{M}_t)_{jh} = 1 \right\}. \tag{3.4}$$

Hence $\psi_{k,ij,h}$ is the set of time periods where both the ith and jth entries of the hth mode-k fibre are observed, $i, j \in [d_k]$, $h \in [d_{\text{-}k}]$ with $d_{\text{-}k} := d_1 \cdots d_K / d_k$.

Inspired by Xiong and Pelger (2023) for a vector time series panel, our method relies on the reconstruction of the mode-k sample covariance matrix $\mathbf{S}_k$, defined for $i, j \in [d_k]$,

$$(S_k)_{ij} := \frac{1}{T}\sum_{t=1}^{T} \mathrm{mat}_k(\mathcal{Y}_t)_{i\cdot}'\mathrm{mat}_k(\mathcal{Y}_t)_{j\cdot} = \sum_{h=1}^{d_{\text{-}k}} \frac{1}{T}\sum_{t=1}^{T} \mathrm{mat}_k(\mathcal{Y}_t)_{ih}\mathrm{mat}_k(\mathcal{Y}_t)_{jh}. \tag{3.5}$$

With missing entries characterized by $\mathcal{M}_t$ and $\psi_{k,ij,h}$ in (3.4), we can generalize the above to

$$(\widehat{S}_k)_{ij} = \sum_{h=1}^{d_{\text{-}k}} \left\{ \frac{1}{|\psi_{k,ij,h}|} \sum_{t\in\psi_{k,ij,h}} \mathrm{mat}_k(\mathcal{Y}_t)_{ih}\mathrm{mat}_k(\mathcal{Y}_t)_{jh} \right\}. \tag{3.6}$$

Intuitively, the cross-covariance between unit i and j at the hth mode-k fibre is estimated inside the curly bracket in (3.6) using only the corresponding available data. PCA can now be performed on $\widehat{\mathbf{S}}_k$, and $\widehat{\mathbf{Q}}_k$ is obtained as the first r_k eigenvectors of $\widehat{\mathbf{S}}_k$.

3.5. Estimation of the core tensor series

With $\hat{\mathbf{Q}}_k$ available (which is estimating the factor loading space of $\mathbf{Q}_k$, with $\hat{\mathbf{Q}}_k$ having orthonormal columns), we can estimate $\mathcal{F}_{Z,t}$ (equivalently $\mathbf{vec}(\mathcal{F}_{Z,t})$) by observing from model (3.2) that

$$\mathbf{vec}(\mathcal{Y}_t) = \mathbf{Q}_\otimes \mathbf{vec}(\mathcal{F}_{Z,t}) + \mathbf{vec}(\mathcal{E}_t), \quad \text{where } \mathbf{Q}_\otimes := \mathbf{Q}_K \otimes \cdots \otimes \mathbf{Q}_1.$$

If $\mathbf{Q}_\otimes$ is known, then the least squares estimator of $\mathbf{vec}(\mathcal{F}_{Z,t})$ is given by

$$\mathbf{vec}(\mathcal{F}_{Z,t}) = (\mathbf{Q}_\otimes'\mathbf{Q}_\otimes)^{-1}\mathbf{Q}_\otimes'\mathbf{vec}(\mathcal{Y}_t) = \left(\sum_{j=1}^d \mathbf{Q}_{\otimes,j\cdot}\mathbf{Q}_{\otimes,j\cdot}'\right)^{-1}\left(\sum_{j=1}^d \mathbf{Q}_{\otimes,j\cdot}[\mathbf{vec}(\mathcal{Y}_t)]_j\right).$$

With missing data, using the missingness tensor $\mathcal{M}_t$, the above can be generalized to

$$\mathbf{vec}(\hat{\mathcal{F}}_{Z,t}) = \left(\sum_{j=1}^d [\mathbf{vec}(\mathcal{M}_t)]_j \hat{\mathbf{Q}}_{\otimes,j\cdot}\hat{\mathbf{Q}}_{\otimes,j\cdot}'\right)^{-1}\left(\sum_{j=1}^d [\mathbf{vec}(\mathcal{M}_t)]_j \hat{\mathbf{Q}}_{\otimes,j\cdot}[\mathbf{vec}(\mathcal{Y}_t)]_j\right). \tag{3.7}$$

4. Assumptions and theoretical results

We present our assumptions for consistent imputation and estimation of factor loading matrices, with the corresponding theoretical results presented afterwards.

4.1. Assumptions

(O1) (Observation patterns)
1. $\mathcal{M}_t$ is independent of $\mathcal{F}_s$ and $\mathcal{E}_s$ for any $t, s \in [T]$.
2. Given $\mathcal{M}_t$ with $t \in [T]$, for any $k \in [K], i, j \in [d_k], h \in [d_{\text{-}k}]$, there exists a constant ψ_0 such that with probability going to 1, we have

$$\frac{|\psi_{k,ij,h}|}{T} \geq \psi_0 > 0.$$

(M1) (Alpha mixing) *The elements in $\mathcal{F}_t$ and $\mathcal{E}_t$ are α-mixing. A vector process $\{\mathbf{x}_t : t = 0, \pm 1, \pm 2, \ldots\}$ is α-mixing if, for some $\gamma > 2$, the mixing coefficients satisfy the condition that*

$$\sum_{h=1}^\infty \alpha(h)^{1-2/\gamma} < \infty,$$

where $\alpha(h) = \sup_\tau \sup_{A\in\mathcal{H}_{-\infty}^\tau, B\in\mathcal{H}_{\tau+h}^\infty} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ and $\mathcal{H}_\tau^s$ is the σ-field generated by $\{\mathbf{x}_t : \tau \leq t \leq s\}$.

(F1) (Time series in $\mathcal{F}_t$) *There is $\mathcal{X}_{f,t}$ the same dimension as $\mathcal{F}_t$, such that $\mathcal{F}_t = \sum_{q\geq 0} a_{f,q}\mathcal{X}_{f,t-q}$. The time series $\{\mathcal{X}_{f,t}\}$ has i.i.d. elements with mean 0 and variance 1, with uniformly bounded fourth order moments. The coefficients $a_{f,q}$ are such that $\sum_{q\geq 0} a_{f,q}^2 = 1$ and $\sum_{q\geq 0} |a_{f,q}| \leq c$ for some constant c.*

(L1) (Factor strength) *We assume for $k \in [K]$, $\mathbf{A}_k$ is of full column rank and independent of factors and errors series. Furthermore, as $d_k \to \infty$,*

$$\mathbf{Z}_k^{-1/2}\mathbf{A}_k'\mathbf{A}_k\mathbf{Z}_k^{-1/2} \to \boldsymbol{\Sigma}_{A,k}, \tag{4.1}$$

where $\mathbf{Z}_k = \text{diag}(\mathbf{A}_k'\mathbf{A}_k)$ and $\boldsymbol{\Sigma}_{A,k}$ is positive definite with all eigenvalues bounded away from 0 and infinity. We assume $(\mathbf{Z}_k)_{jj} \asymp d_k^{\alpha_{k,j}}$ for $j \in [r_k]$, and $1/2 < \alpha_{k,r_k} \leq \cdots \leq \alpha_{k,2} \leq \alpha_{k,1} \leq 1$.

With Assumption (L1), we can denote $\mathbf{Q}_k := \mathbf{A}_k\mathbf{Z}_k^{-1/2}$ and hence $\mathbf{Q}_k'\mathbf{Q}_k \to \boldsymbol{\Sigma}_{A,k}$. We need $\alpha_{k,j} > 1/2$ in order that the ratio-based estimator of the number of factors in Section 4.5 works.

(E1) (Decomposition of $\mathcal{E}_t$) *We assume K is constant, and*

$$\mathcal{E}_t = \mathcal{F}_{e,t} \times_1 \mathbf{A}_{e,1} \times_2 \cdots \times_K \mathbf{A}_{e,K} + \boldsymbol{\Sigma}_\epsilon \circ \boldsymbol{\epsilon}_t, \tag{4.2}$$

where $\mathcal{F}_{e,t}$ is an order-K tensor with dimension $r_{e,1} \times \cdots \times r_{e,K}$, containing independent elements with mean 0 and variance 1. The order-K tensor $\boldsymbol{\epsilon}_t \in \mathbb{R}^{d_1\times\cdots\times d_K}$ contains independent mean zero elements with unit variance, with the two time series $\{\boldsymbol{\epsilon}_t\}$ and $\{\mathcal{F}_{e,t}\}$ being independent. The order-K tensor $\boldsymbol{\Sigma}_\epsilon$ contains the standard deviations of the corresponding elements in $\boldsymbol{\epsilon}_t$, and has elements uniformly bounded.
Moreover, for each $k \in [K]$, $\mathbf{A}_{e,k} \in \mathbb{R}^{d_k\times r_{e,k}}$ is such that $\left\|\mathbf{A}_{e,k}\right\|_1 = O(1)$. That is, $\mathbf{A}_{e,k}$ is (approximately) sparse.

(E2) (Time series in $\mathcal{E}_t$) *There is* $\mathcal{X}_{e,t}$ *the same dimension as* $\mathcal{F}_{e,t}$, *and* $\mathcal{X}_{\epsilon,t}$ *the same dimension as* ϵ_t, *such that* $\mathcal{F}_{e,t} = \sum_{q\geq 0} a_{e,q}\mathcal{X}_{e,t-q}$ *and* $\epsilon_t = \sum_{q\geq 0} a_{\epsilon,q}\mathcal{X}_{\epsilon,t-q}$, *with* $\{\mathcal{X}_{e,t}\}$ *and* $\{\mathcal{X}_{\epsilon,t}\}$ *independent of each other, and each time series has independent elements with mean* 0 *and variance* 1 *with uniformly bounded fourth order moments. Both* $\{\mathcal{X}_{e,t}\}$ *and* $\{\mathcal{X}_{\epsilon,t}\}$ *are independent of* $\{\mathcal{X}_{f,t}\}$ *from (F1).*
The coefficients $a_{e,q}$ *and* $a_{\epsilon,t}$ *are such that* $\sum_{q\geq 0} a_{e,q}^2 = \sum_{q\geq 0} a_{\epsilon,q}^2 = 1$ *and* $\sum_{q\geq 0} |a_{e,q}|, \sum_{q\geq 0} |a_{\epsilon,q}| \leq c$ *for some constant* c.

(R1) (Further rate assumptions) *We assume that, with* $d := d_1 \cdots d_K$ *and* $g_s := \prod_{k=1}^K d_k^{\alpha_{k,1}}$,

$$d g_s^{-2} T^{-1} d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})+1} = o(1), \quad d g_s^{-1} T^{-1} d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})} = o(1), \quad d g_s^{-1} d_k^{\alpha_{k,1}-\alpha_{k,r_k}-1/2} = o(1).$$

Assumption (O1) means that the missing mechanism is independent of the factors and the noise series, which is also assumed in Xiong and Pelger (2023) for the purpose of identification. It also means that the missing pattern can depend on the K factor loading matrices, allowing for a wide variety of missing patterns that can vary over time and units in different dimensions. Condition 2 of (O1) implies that the number of time periods that any two individual units are both observed are at least proportional to T, which simplifies proofs and presentations, and is also used in Xiong and Pelger (2023). Assumption (M1) is a standard assumption in vector time series factor models, which facilitates proofs using central limit theorem for time series without losing too much generality. Assumption (F1), (E1) and (E2) are exactly the corresponding assumptions in Chen and Lam (2024b), allowing for serial correlations in the factor series, and serial and cross-sectional dependence within and among the error tensor fibres. These three assumptions facilitate the proof of asymptotic normality in Section 4.3, and boil down to similar assumptions in Chen and Fan (2023) for matrix time series and in Barigozzi et al. (2023) for general tensor time series (see Proposition 1 in the supplementary materials for the technical details). Together with Assumption (M1), we implicitly restrict the general linear processes in (F1) and (E2) to be, for instance, of short rather than long dependence.

Assumption (L1) is quite different from assumptions in other papers on factor models, in the sense that we allow for the existence of weak factors alongside the pervasive ones. Chen and Lam (2024b) adapted the same assumption, which allows each column of $\mathbf{A}_k$ to be completely dense (i.e., a pervasive factor) or sparse to a certain extent. A diagonal entry in $\mathbf{Z}_k$ then records how dense a column really is, and the corresponding strength of factors defined. Assumption (L1) is similar to, yet technically more general than, Assumption 1(iii) in Onatski (2012) which requires $\boldsymbol{\Sigma}_{A,k}$ to be diagonal while the normalization on the factor series is essentially the same as ours. If all factors are pervasive, (4.1) can be read as $d_k^{-1}\mathbf{A}_k'\mathbf{A}_k \to \boldsymbol{\Sigma}_{A,k}$ which is akin to Assumption 3 of Chen and Fan (2023) for $K = 2$. Modelling with weak factors is closer to reality, and empirical evidence can be found in economics and finance, etc. For instance, apart from a pervasive market factor, there can be weaker sector factors in a large selection of stock returns (Trzcinka, 1986). More recent work on factor models specifically focuses on weak factors with real data examples confirming the existence of weak factors, such as Freyaldenhoven (2022) and Chen and Lam (2024a).

Finally, Assumption (R1) gives the technical rates needed for the proof of various theorems in the paper because of the existence of weak factors. If all factors are pervasive (i.e., $\alpha_{k,j} = 1$), then the conditions are automatically satisfied. Suppose $K = 2$, $T \asymp d_1 \asymp d_2$ and the strongest factors are all pervasive (i.e., $\alpha_{k,1} = 1$), then we need $\alpha_{k,r_k} > 1/2$ for (R1) to be satisfied. This condition is the same as the one remarked right after we stated Assumption (L1). A factor with $\alpha_{k,j}$ close to 0.5 presents a significantly weak factor with only more than $d_k^{1/2}$ of elements are non-zero in the corresponding column of $\mathbf{A}_k$.

Remark 1. With the missing entries imputed using the estimated common components $\widehat{C}_{t,i_1,\ldots,i_K}$, we have a completed data set which could be used for re-estimation and hence re-imputation. The convergence could be shown empirically to be accelerated by such a procedure. The rate improvement would be from the difference between T and $\psi_0 T$, where ψ_0 is the lowest proportion of observation among all entries from Assumption (O1). We omit the lengthy proofs as eventually the rates only differ by a constant, but we note here that re-imputation can indeed improve our imputation, which is essentially credited to the more observations used when we have an initially good imputation.

4.2. Consistency: factor loadings and imputed values

We present consistency results in this section. For $k \in [K], j \in [d_k]$, define

$$\mathbf{H}_{k,j} := \widehat{\mathbf{D}}_k^{-1} \sum_{i=1}^{d_k} \widehat{\mathbf{Q}}_{k,i\cdot} \sum_{h=1}^{d_{\text{-}k}} \frac{1}{|\psi_{k,ij,h}|} \sum_{t\in\psi_{k,ij,h}} \Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{k,hm} \mathrm{mat}_k(\mathcal{F}_{Z,t})_{\cdot m}\Big)' \mathbf{Q}_{k,i\cdot} \Big(\sum_{m=1}^{r_{\text{-}k}} \Lambda_{k,hm} \mathrm{mat}_k(\mathcal{F}_{Z,t})_{\cdot m}\Big)', \tag{4.3}$$

$$\mathbf{H}_k^a := \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{D}}_k^{-1} \widehat{\mathbf{Q}}_k' \mathbf{Q}_k \mathrm{mat}_k(\mathcal{F}_{Z,t}) \boldsymbol{\Lambda}_k' \boldsymbol{\Lambda}_k \mathrm{mat}_k(\mathcal{F}_{Z,t})', \tag{4.4}$$

where $\widehat{\mathbf{D}}_k := \widehat{\mathbf{Q}}_k' \widehat{\mathbf{S}}_k \widehat{\mathbf{Q}}_k$ is a diagonal matrix of eigenvalues of $\widehat{\mathbf{S}}_k$ defined in (3.6). Hence $\mathbf{H}_{k,j} = \mathbf{H}_k^a$ if there are no missing entries, i.e., $|\psi_{k,ij,h}| = T$ for each $k \in [K], i, j \in [d_k]$ and $h \in [d_{\text{-}k}]$. Furthermore, each $\mathbf{H}_{k,j}$ and $\mathbf{H}_k^a$ can be shown asymptotically invertible (see Lemma 3 and 4 in the supplementary materials).

We first present a consistency result for the factor loading matrix estimator $\widehat{\mathbf{Q}}_k$ of $\mathbf{Q}_k$. In particular, our theoretical rates are shown in the presence of potential weak factors. To compare with results in similar literature, we will end this section with a simplified result. Readers interested in the rates under only pervasive factors can go straight to Corollary 1.

Theorem 1. *Under Assumptions (O1), (M1), (F1), (L1), (E1), (E2) and (R1), for any $k \in [K]$, we have*

$$\frac{1}{d_k}\sum_{j=1}^{d_k}\left\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_{k,j}\mathbf{Q}_{k,j\cdot}\right\|^2 = O_P\left(d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})-1}\left(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\right)\frac{d^2}{g_s^2}\right) = o_P(1),$$

where g_s is defined in Assumption (R1). Furthermore, define $\eta := 1-\psi_0$ with ψ_0 from Assumption (O1), then

$$\frac{1}{d_k}\sum_{j=1}^{d_k}\left\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{Q}_{k,j\cdot}\right\|^2 = \frac{1}{d_k}\left\|\widehat{\mathbf{Q}}_k - \mathbf{Q}_k\mathbf{H}_k^{a\,\prime}\right\|_F^2$$
$$= O_P\left(d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})-1}\left\{\left(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\right)\frac{d^2}{g_s^2}+\min\left(\frac{1}{T},\frac{\eta^2}{(1-\eta)^2}\right)\right\}\right) = o_P(1).$$

The proof of the theorem can be found in the supplementary materials of this paper. The two results in Theorem 1 coincide with each other if $\eta = 0$, i.e., there are no missing values.

We present the two results in the theorem to highlight the difficulty of obtaining consistency when there are missing values. Since a factor loading matrix is not uniquely defined, in the second result in Theorem 1 we are estimating how close $\widehat{\mathbf{Q}}_k$ is to a version of $\mathbf{Q}_k$ in Frobenius norm, namely $\mathbf{Q}_k\mathbf{H}_k^a$, which is still defining the same factor loading space as $\mathbf{Q}_k$ does. With missing data, such a feat is complicated, in the sense that for the jth row of $\mathbf{Q}_k$, $\widehat{\mathbf{Q}}_{k,j\cdot}$, there corresponds an $\mathbf{H}_{k,j}$ different from $\mathbf{H}_k^a$ in general, so that $\widehat{\mathbf{Q}}_{k,j\cdot}$ is close to $\mathbf{H}_{k,j}\mathbf{Q}_{k,j\cdot}$. The extra rate $\min(1/T,\eta^2/(1-\eta)^2)$ in the second result is essentially measuring how close each $\mathbf{H}_{k,j}$ is to $\mathbf{H}_k^a$. See Lemma 3 in the supplementary materials as well.

Theorem 2. *Under the Assumptions in Theorem 1, suppose we further have $d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}} = o(d_{\text{-}k})$. Define*

$$g_\eta := \min\left(\frac{1}{T},\frac{\eta^2}{(1-\eta)^2}\right),\quad g_w := \prod_{k=1}^K d_k^{\alpha_{k,r_k}}.$$

Then we have the following.

1. *The error of the estimated factor series has rate*

$$\left\|\mathbf{vec}(\widehat{\mathcal{F}}_{Z,t}) - \left(\mathbf{H}_K^{a\prime}\otimes\cdots\otimes\mathbf{H}_1^{a\prime}\right)^{-1}\mathbf{vec}(\mathcal{F}_{Z,t})\right\|^2$$
$$= O_P\left(\max_{k\in[K]}\left\{T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1}+d^2g_s^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1}+g_\eta g_s d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\right\}+\frac{d}{g_w}\right).$$

2. *For any $k\in[K], i_k\in[d_k], t\in[T]$, the squared individual imputation error is*

$$(\widehat{C}_{t,i_1,\ldots,i_K} - C_{t,i_1,\ldots,i_K})^2$$
$$= O_P\left(\max_{k\in[K]}\left\{T^{-1}dd_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1}g_w^{-1}+d^2g_s^{-1}g_w^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1}+g_\eta g_s g_w^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\right\}+\frac{d}{g_w^2}\right).$$

3. *The average imputation error is given by*

$$\frac{1}{Td}\sum_{t=1}^T\sum_{i_1,\ldots,i_K=1}^{d_1,\ldots,d_K}(\widehat{C}_{t,i_1,\ldots,i_K} - C_{t,i_1,\ldots,i_K})^2$$
$$= O_P\left(\max_{k\in[K]}\left\{T^{-1}d_k^{3\alpha_{k,1}-2\alpha_{k,r_k}}g_s^{-1}+dg_s^{-1}d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}-1}+d^{-1}g_\eta g_s d_k^{2\alpha_{k,1}-3\alpha_{k,r_k}+1}\right\}+\frac{1}{g_w}\right).$$

The proof can be found in the supplementary materials, which utilizes some rates from the proof of Theorem 3 in the supplementary materials (without the need for extra rate restrictions like Theorem 3 though). The complication of missing data comes explicitly from the rate g_η. The average squared imputation error in result 3 improves upon individual squared error in result 2 when weak factors exist, with degree of improvements larger when the difference in strength of factors is larger.

Our rate can be considered a generalization to a general order tensor, with general factor strengths and missing data, see the comparison of our results with others' below Corollary 1. Such generalizations have certainly revealed that when there are weak factors, especially when the strongest and weakest factor strengths are quite different, those rates of convergence greatly suffer.

Corollary 1 (*Simplified Theorems 1 and 2 Under Pervasive Factors*)**.** *Let Assumption (O1), (M1), (F1), (L1), (E1) and (E2) hold. If all factors are pervasive such that $\alpha_{k,j} = 1$ for all $k\in[K], j\in[r_k]$, then with the renormalized loading and core factor estimators defined as $\widehat{\mathbf{A}}_k = \sqrt{d_k}\,\widehat{\mathbf{Q}}_k$ and $\widehat{\mathcal{F}}_t = \widehat{\mathcal{F}}_{Z,t}/\sqrt{d}$, we have the following:*

1. *The (renormalized) loading estimator is consistent such that for any $k\in[K]$,*

$$\frac{1}{d_k}\sum_{j=1}^{d_k}\left\|\widehat{\mathbf{A}}_{k,j\cdot} - \mathbf{H}_{k,j}\mathbf{A}_{k,j\cdot}\right\|^2 = O_P\left(\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}\right) = o_P(1),$$

$$\frac{1}{d_k}\sum_{j=1}^{d_k}\left\|\widehat{\mathbf{A}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{A}_{k,j\cdot}\right\|^2 = O_P\left\{\frac{1}{Td_{\text{-}k}}+\frac{1}{d_k}+\min\left(\frac{1}{T},\frac{\eta^2}{(1-\eta)^2}\right)\right\} = o_P(1).$$

2. *The (renormalized) core factor estimator is consistent such that for any $t \in [T]$,*

$$\Big\|\mathbf{vec}(\widehat{\mathcal{F}}_t) - \big(\mathbf{H}_K^{a'} \otimes \cdots \otimes \mathbf{H}_1^{a'}\big)^{-1}\mathbf{vec}(\mathcal{F}_t)\Big\|^2 = O_P\Big\{\max_{k\in[K]}\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k^2}\Big) + \min\Big(\frac{1}{T}, \frac{\eta^2}{(1-\eta)^2}\Big)\Big\}.$$

3. *The imputation is consistent both for each entry and on average (with the same rate), such that for any $k \in [K], i_k \in [d_k], t \in [T]$,*

$$(\widehat{C}_{t,i_1,\ldots,i_K} - C_{t,i_1,\ldots,i_K})^2 = O_P\Big\{\max_{k\in[K]}\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k^2}\Big) + \min\Big(\frac{1}{T}, \frac{\eta^2}{(1-\eta)^2}\Big) + \frac{1}{d}\Big\}.$$

When $K = 1$ with missing data, result 1 has rate $1/\min(d_1, T)$, which is the same as the rate in Theorem 1 of Xiong and Pelger (2023). If $K = 2$ and $\eta = 0$ (i.e, no missing values), result 1 has rate $1/\min(d_k, Td_{\text{-}k})$, which is consistent with Theorem 1 of Chen and Fan (2023), for example. For a general order-K tensor without missing data (i.e., $\eta = 0$), our Lemma 5 in the supplementary materials states that

$$\|\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{Q}_{k,j\cdot}\|^2 = O_P\Big(\frac{1}{Td} + \frac{1}{d_k^3}\Big), \ \text{ implying } \ \frac{1}{d_k}\|\widehat{\mathbf{A}}_k - \mathbf{A}_k\mathbf{H}_k^a\|_F^2 = O_P\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k^2}\Big),$$

which aligns with Theorem 3.1 of He et al. (2022a) or Barigozzi et al. (2023).

If $K \geq 2$ and $\eta = 0$, result 3 has rate

$$\max_{k\in[K]}\Big(\frac{1}{Td_{\text{-}k}} + \frac{1}{d_k^2}\Big) + \frac{1}{d} \asymp \frac{1}{\min(Td_{\text{-}1}, \ldots, Td_{\text{-}K}, d_1^2, \ldots, d_K^2)}.$$

This rate is the same as the result in Theorem 4 of Chen and Fan (2023) for $K = 2$, which is a rate for estimating the common component. On the other hand, if η is a constant and $K = 1$, then result 3 becomes $d_1^{-1} + T^{-1} \asymp 1/\min(d_1, T)$, which is the same rate as result 3 of Theorem 2 in Xiong and Pelger (2023).

4.3. Inference on the factor loadings

We establish asymptotic normality of the factor loadings for inference purposes. In Section 4.4 we present the covariance matrix estimator for practical use of our asymptotic normality result. First, we define

$$\mathbf{H}_k^{a,*} := \mathrm{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k})^{1/2} \cdot \mathbf{D}_k^{-1/2}\Upsilon_k'\mathbf{Z}_k^{1/2}, \tag{4.5}$$

where $\mathbf{D}_k := \mathrm{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k})\,\mathrm{diag}\{\lambda_1(\mathbf{A}_k'\mathbf{A}_k), \ldots, \lambda_{r_k}(\mathbf{A}_k'\mathbf{A}_k)\}$, and Υ_k is the eigenvector matrix of $\mathrm{tr}(\mathbf{A}_{\text{-}k}'\mathbf{A}_{\text{-}k}) \cdot g_s^{-1}d_k^{\alpha_{k,1}-\alpha_{k,r_k}}\mathbf{Z}_k^{1/2}\boldsymbol{\Sigma}_{A,k}\mathbf{Z}_k^{1/2}$. It turns out $\mathbf{H}_k^{a,*}$ is the probability limit of $\mathbf{H}_k^a$ defined in (4.4). Before presenting our results, we need three additional assumptions.

(L2) (Eigenvalues) *For any $k \in [K]$, the eigenvalues of the $r_k \times r_k$ matrix $\boldsymbol{\Sigma}_{A,k}\mathbf{Z}_k$ from Assumption (L1) are distinct.*

(AD1) *Define $\omega_B := d_{\text{-}k}^{-1}d_k^{2\alpha_{k,r_k}-3\alpha_{k,1}}g_s^2$ and the following,*

$$\Xi_{k,j} := \operatorname*{plim}_{T,d_1,\ldots,d_K\to\infty} \mathrm{Var}\Big(\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{k,ij,h}|}\sum_{t\in\psi_{k,ij,h}}\mathrm{mat}_k(\mathcal{E}_t)_{jh}(\mathbf{A}_{\text{-}k})_{h\cdot}'\mathrm{mat}_k(\mathcal{F}_t)'\mathbf{A}_{k,i\cdot}\Big),$$

then we assume $T\omega_B \cdot \Big\|\mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\Xi_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}\Big\|_F$ is of constant order.

(AD2) *Define the filtration $\mathcal{G}^T := \sigma(\cup_{s=1}^T\mathcal{G}_s)$ with $\mathcal{G}_s := \sigma(\{\mathcal{M}_{t,i_1,\ldots,i_K} \mid t \leq s\}, \mathbf{A}_1, \ldots, \mathbf{A}_K)$, and*

$$\Delta_{F,k,ij,h} := \frac{1}{|\psi_{k,ij,h}|}\sum_{t\in\psi_{k,ij,h}}\mathrm{mat}_k(\mathcal{F}_t)\mathbf{v}_{k,h}\mathbf{v}_{k,h}'\mathrm{mat}_k(\mathcal{F}_t)' - \frac{1}{T}\sum_{t=1}^T\mathrm{mat}_k(\mathcal{F}_t)\mathbf{v}_{k,h}\mathbf{v}_{k,h}'\mathrm{mat}_k(\mathcal{F}_t)',$$

where $\mathbf{v}_{k,h} := [\otimes_{l\in[K]\setminus\{k\}}\mathbf{A}_l]_{h\cdot}$. With $\mathbf{Q}_k$ being the normalized mode-k factor loading defined below Assumption (L1), we have for every $k \in [K], j \in [d_k]$, for a function $h_{k,j} : \mathbb{R}^{r_k} \to \mathbb{R}^{r_k\times r_k}$,

$$\sqrt{Td_k^{\alpha_{k,r_k}}} \cdot \mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\sum_{i=1}^{d_k}\mathbf{Q}_{k,i\cdot}\mathbf{A}_{k,i\cdot}'\sum_{h=1}^{d_{\text{-}k}}\Delta_{F,k,ij,h}\mathbf{A}_{k,j\cdot} \to \mathcal{N}(\mathbf{0}, \mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}h_{k,j}(\mathbf{A}_{k,j\cdot})(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}) \quad \mathcal{G}^T\text{-stably}.$$

Assumption (AD1) guarantees a part of the covariance matrix of the asymptotic normality in Theorem 3 is of constant order. It can be regarded as a lower bound condition which is necessary for the dominance of a certain term involved in the asymptotic normality. Since we show the upper bound of $T\omega_B \cdot \Big\|\mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\Xi_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}\Big\|_F$ is of constant order in the proof of Theorem 3 in the supplementary materials, this assumption is not particularly strong.

Assumption (AD2) is required since the missing data creates a discrepancy term $\Delta_{F,k,ij,h}$ as defined in the assumption. This assumption is also parallel to Assumption G3.5 in Xiong and Pelger (2023). We demonstrate how this assumption is satisfied with Assumption (O1), (F1), (L1) and two additional but simpler assumptions in Proposition 1 in Section 4.6.

Theorem 3. *Let all the assumptions under Theorem 2 hold, in addition to (L2), (AD1) and (AD2) above. With r_k fixed and $d_k, T \to \infty$ for $k \in [K]$, suppose also $Td_{\text{-}k} = o(d_k^{\alpha_{k,1}+\alpha_{k,r_k}})$. We have*

$$\sqrt{Td_k^{\alpha_{k,r_k}}} \cdot (\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{k,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}(Td_k^{\alpha_{k,r_k}} \cdot \Xi_{k,j} + h_{k,j}(\mathbf{A}_{j\cdot}))(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}).$$

Furthermore, if $Td^{-1}g_s^2 g_\eta d_k^{1+\alpha_{k,1}-3\alpha_{k,r_k}} = o(1)$ is also satisfied, then

$$\sqrt{T\omega_B} \cdot (\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a \mathbf{Q}_{k,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, T\omega_B \cdot \mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\Xi_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}).$$

If all factors are pervasive, the rate condition $Td_{\text{-}k} = o(d_k^{\alpha_{k,1}+\alpha_{k,r_k}})$ reduces to $Td_{\text{-}k} = o(d_k^2)$, which is equivalent to the condition needed for asymptotic normality in Bai (2003) for $K = 1$ and Chen and Fan (2023) for $K = 2$. The first asymptotic normality result is compatible to Theorem 2.1 of Xiong and Pelger (2023) when all factors are pervasive. In their Theorem 2.1, the $\Gamma_{\Lambda,j}^{obs}$ is in fact of rate N^{-1}, so that the normalizing rate is $\sqrt{TN}$, which is exactly $\sqrt{Td_1}$ in our first result when $K = 1$.

Suppose all factors are pervasive. The rate condition $Td^{-1}g_s^2 g_\eta d_k^{1+\alpha_{k,1}-3\alpha_{k,r_k}} = o(1)$ is automatically satisfied when there is no missing data, i.e., $\eta = 0$ so that $g_\eta = 0$. If so, the rate of convergence is $\sqrt{T\omega_B} = \sqrt{Td}$, which is compatible to Theorem 2.1, Theorem 2.2 of Chen and Fan (2023) and Theorem 3.2 of Barigozzi et al. (2023) (after our normalization to their factor loading matrices). The condition is also satisfied when there is only finite number of missing data points, so that $\eta \asymp T^{-1}$ and $g_\eta \asymp T^{-2}$, and $d_1, d_2 = o(T)$ for $K = 2$.

Remark 2. We do not establish asymptotic normality for the estimated factor series and common components. The reason is that for a tensor with $K > 1$, the decomposition in the estimated factor series and the common components cannot be dominated by terms that are asymptotically normal. This is also the reason why Chen and Fan (2023) does not include asymptotic normality for the estimated factor series and common components. Barigozzi et al. (2023) constructs asymptotic normality for the core factor built upon their projection estimator $\widetilde{\mathcal{F}}_t$, which is sensible as the projecting loading estimator already has an improved rate. In comparison, the rate of any PCA-type estimators, such as the one in Chen and Fan (2023) for matrix data and the one in our case for general tensors, is insufficient for a potentially asymptotically Gaussian term to be dominating. The main goal of this work is to impute missing entries, and existing methods on tensor factor models using Tucker decomposition should be applicable with all missing entries replaced by the consistent imputations.

4.4. Estimation of the asymptotic covariance matrix

In order to carry out inferences for the factor loadings using Theorem 3, we need to estimate the asymptotic covariance matrix for $\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{Q}_{k,j\cdot}$. To this end, we use the heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West, 1987) based on $\{\widehat{\mathbf{Q}}_k, \text{mat}_k(\widehat{\mathcal{C}}_t), \text{mat}_k(\widehat{\mathcal{E}}_t)\}_{t\in[T]}$, where

$$\text{mat}_k(\widehat{\mathcal{C}}_t) = (\widehat{\mathbf{Q}}_k)\text{mat}_k(\widehat{\mathcal{F}}_{Z,t})(\widehat{\mathbf{Q}}_K \otimes \cdots \otimes \widehat{\mathbf{Q}}_{k+1} \otimes \widehat{\mathbf{Q}}_{k-1} \otimes \cdots \otimes \widehat{\mathbf{Q}}_1)', \quad \text{mat}_k(\widehat{\mathcal{E}}_t) := \text{mat}_k(\mathcal{Y}_t) - \text{mat}_k(\widehat{\mathcal{C}}_t).$$

With a tuning parameter β such that $\beta \to \infty$ and $\beta/(Td_k^{\alpha_{k,r_k}})^{1/4} \to 0$, we define two HAC estimators

$$\widehat{\Sigma}_{HAC} := \mathbf{D}_{k,0,j} + \sum_{\nu=1}^{\beta}\left(1 - \frac{\nu}{1+\beta}\right)\left(\mathbf{D}_{k,\nu,j} + \mathbf{D}_{k,\nu,j}'\right),$$

$$\widehat{\Sigma}_{HAC}^{\Delta} := \mathbf{D}_{k,0,j}^{\Delta} + \sum_{\nu=1}^{\beta}\left(1 - \frac{\nu}{1+\beta}\right)\left(\mathbf{D}_{k,\nu,j}^{\Delta} + (\mathbf{D}_{k,\nu,j}^{\Delta})'\right), \quad \text{where}$$

$$\mathbf{D}_{k,\nu,j} := \sum_{t=1+\nu}^{T}\left(\sum_{i=1}^{d_k}\Big(\frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{D}}_k^{-1}\widehat{\mathbf{Q}}_k'\widehat{\mathbf{C}}_{(k),s}\widehat{\mathbf{C}}_{(k),s,i\cdot}\Big)\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{k,ij,h}|}\widehat{\mathbf{E}}_{(k),t,jh}\widehat{\mathbf{C}}_{(k),t,ih}\cdot\mathbb{1}\{t\in\psi_{k,ij,h}\}\right)$$
$$\cdot\left(\sum_{i=1}^{d_k}\Big(\frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{D}}_k^{-1}\widehat{\mathbf{Q}}_k'\widehat{\mathbf{C}}_{(k),s}\widehat{\mathbf{C}}_{(k),s,i\cdot}\Big)\sum_{h=1}^{d_{\text{-}k}}\frac{1}{|\psi_{k,ij,h}|}\widehat{\mathbf{E}}_{(k),t-\nu,jh}\widehat{\mathbf{C}}_{(k),t-\nu,ih}\cdot\mathbb{1}\{t-\nu\in\psi_{k,ij,h}\}\right)',$$

$$\mathbf{D}_{k,\nu,j}^{\Delta} := \sum_{t=1+\nu}^{T}\left[\sum_{i=1}^{d_k}\Big(\frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{D}}_k^{-1}\widehat{\mathbf{Q}}_k'\widehat{\mathbf{C}}_{(k),s}\widehat{\mathbf{C}}_{(k),s,i\cdot}\Big)\sum_{h=1}^{d_{\text{-}k}}\Big(\frac{1}{|\psi_{k,ij,h}|}\widehat{\mathbf{C}}_{(k),t,ih}\widehat{\mathbf{C}}_{(k),t,jh}\cdot\mathbb{1}\{t\in\psi_{k,ij,h}\}\right.$$
$$\left.-\frac{1}{T}\widehat{\mathbf{C}}_{(k),t,ih}\widehat{\mathbf{C}}_{(k),t,jh}\Big)\right]\cdot\left[\sum_{i=1}^{d_k}\Big(\frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{D}}_k^{-1}\widehat{\mathbf{Q}}_k'\widehat{\mathbf{C}}_{(k),s}\widehat{\mathbf{C}}_{(k),s,i\cdot}\Big)\right.$$
$$\left.\cdot\sum_{h=1}^{d_{\text{-}k}}\Big(\frac{1}{|\psi_{k,ij,h}|}\widehat{\mathbf{C}}_{(k),t-\nu,ih}\widehat{\mathbf{C}}_{(k),t-\nu,jh}\cdot\mathbb{1}\{t-\nu\in\psi_{k,ij,h}\}-\frac{1}{T}\widehat{\mathbf{C}}_{(k),t-\nu,ih}\widehat{\mathbf{C}}_{(k),t-\nu,jh}\Big)\right]',$$

where $\widehat{\mathbf{C}}_{(k),s} := \text{mat}_k(\widehat{\mathcal{C}}_s)$ and $\widehat{\mathbf{E}}_{(k),s} := \text{mat}_k(\widehat{\mathcal{E}}_s)$.

Theorem 4. *Let all the assumptions under Theorem 2 hold, in addition to (L2), (AD1) and (AD2) above. With r_k fixed and $d_k, T \to \infty$ for $k \in [K]$, suppose also the rate for the individual common component imputation error in result 2 of Theorem 2 is $o(1)$, together with*

$Td_{\text{-}k} = o(d_k^{\alpha_{k,1}+\alpha_{k,r_k}})$ *and* $d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}[(Td_{\text{-}k})^{-1} + d_k^{-1}]d^2 g_s^{-2} = o(1)$. *Then*

1. $\widehat{\mathbf{D}}_k^{-1}\widehat{\boldsymbol{\Sigma}}_{HAC}\widehat{\mathbf{D}}_k^{-1}$ *is consistent for* $\mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}\Xi_{k,j}(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}$;
2. $\widehat{\mathbf{D}}_k^{-1}\widehat{\boldsymbol{\Sigma}}_{HAC}^{\Delta}\widehat{\mathbf{D}}_k^{-1}$ *is consistent for* $(Td_k^{\alpha_{k,r_k}})^{-1}\mathbf{D}_k^{-1}\mathbf{H}_k^{a,*}h_{k,j}(\mathbf{A}_{k,j\cdot})(\mathbf{H}_k^{a,*})'\mathbf{D}_k^{-1}$;
3. $(\widehat{\boldsymbol{\Sigma}}_{HAC} + \widehat{\boldsymbol{\Sigma}}_{HAC}^{\Delta})^{-1/2}\widehat{\mathbf{D}}_k(\widehat{\mathbf{Q}}_{k,j\cdot} - \mathbf{H}_k^a\mathbf{Q}_{k,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_{r_k})$.

The extra rate assumption $d_k^{2(\alpha_{k,1}-\alpha_{k,r_k})}[(Td_{\text{-}k})^{-1} + d_k^{-1}]d^2 g_s^{-2} = o(1)$ ensures that we have Frobenius norm consistency for $\widehat{\mathbf{Q}}_k$ from Theorem 1. The imputation error from result 2 of Theorem 2 also has rate going to 0 when all factors are pervasive, for instance. With result 3 in particular, we can perform inferences on any rows of $\widehat{\mathbf{Q}}_k$. Practical performances of result 3 is demonstrated in Section 5.1.3. The reason that we need two HAC estimators is that similar to Theorem 1, there is a component for missing data, arising from the fact that $\mathbf{H}_{k,j}$ is different from $\mathbf{H}_k^a$ for each $j \in [d_k]$ in general.

4.5. Estimation of number of factors

The reconstructed mode-k sample covariance matrix $\widehat{\mathbf{S}}_k$ is in fact estimating a complete-sample version of a matrix $\mathbf{R}_k^*$, where

$$\mathbf{R}_k^* := \frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}_k \text{mat}_k(\mathcal{F}_{Z,t})\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k \text{mat}_k(\mathcal{F}_{Z,t})'\mathbf{Q}_k', \tag{4.6}$$

and $\mathcal{F}_{Z,t}$ and $\boldsymbol{\Lambda}_k$ are defined in (3.2). It turns out that we have $\lambda_j(\widehat{\mathbf{S}}_k) \asymp_P \lambda_j(\mathbf{R}_k^*)$ for $j \in [r_k]$, and

$$\lambda_j(\mathbf{R}_k^*) \asymp_P d_k^{\alpha_{k,j}-\alpha_{k,1}} g_s, \quad g_s := \prod_{k=1}^{K} d_k^{\alpha_{k,1}} \text{ as defined in (R1).}$$

We have the following theorem.

Theorem 5. *Let Assumption (O1), (M1), (F1), (L1), (E1), (E2) and (R1) hold. Moreover, assume*

$$\begin{cases} dg_s^{-1}d_k^{\alpha_{k,1}-\alpha_{k,r_k}}[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}] = o(d_k^{\alpha_{k,j+1}-\alpha_{k,j}}), & j \in [r_k - 1] \text{ with } r_k \geq 2; \\ dg_s^{-1}[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}] = o(1), & r_k = 1. \end{cases}$$

Then $\widehat{r}_k$ *is a consistent estimator of* r_k, *where*

$$\widehat{r}_k := \arg\min_{\ell}\left\{\frac{\lambda_{\ell+1}(\widehat{\mathbf{S}}_k) + \xi}{\lambda_{\ell}(\widehat{\mathbf{S}}_k) + \xi}, \ \ell \in [\lfloor d_k/2 \rfloor]\right\}, \quad \xi \asymp d[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}]. \tag{4.7}$$

The extra rate assumption is satisfied, for instance, when all factors corresponding to $\mathbf{A}_k$ are pervasive. An eigenvalue-ratio estimator is considered in Lam and Yao (2012) and Ahn and Horenstein (2013), while a perturbed eigenvalue ratio estimator is considered in Pelger (2019). However, all of these estimators are for a vector time series factor model. Our estimator $\widehat{r}_k$ in (4.7) extracts eigenvalues from $\widehat{\mathbf{S}}_k$, which is not necessarily positive semi-definite. The addition of ξ can make $\widehat{\mathbf{S}}_k + \xi\mathbf{I}_{d_k}$ positive semi-definite, while stabilizing the estimator. We naturally assume that $r_k < d_k/2$, which is a very reasonable assumption for all applications of factor models. In fact, our recommended choice of ξ is

$$\xi = \frac{1}{5}d[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}].$$

The requirement that $\xi \asymp d[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}]$ ensures that $\xi = o_P(\lambda_{r_k}(\widehat{\mathbf{S}}_k))$ from our rate assumption in the theorem. Our simulations in Section 5.1.2 suggest that this proposal works very well.

4.6. *How Assumption (AD2) can be implied

This section details how Assumption (AD2) can be implied from simpler assumptions. Readers can skip this part and go straight to the next section for a more integral reading experience. We begin by presenting a proposition.

Proposition 1. *Let Assumption (O1), (F1), (L1) hold. For a given* $k \in [K], j \in [d_k]$, *assume also the following:*

1. *The mode-*k *factor is strong enough such that* $\alpha_{k,r_k} > 4/5$, *and* $d_k^{\alpha_{k,1}-\alpha_{k,r_k}}T^{-\epsilon/2} = o(1)$ *with some* $\epsilon \in (0, 1)$.
2. *There exists some* $\psi_{k,ij}$ *such that* $\psi_{k,ij} = \psi_{k,ij,h}$ *for any* $i \in [d_k], h \in [d_{\text{-}k}]$. *Furthermore, there exists* $\omega_{\psi,k,j}$ *such that for any* $t \in [T]$, *as* $d_k, T \to \infty$,
$$d_k^{-2}\sum_{i=1}^{d_k}\sum_{l=1}^{d_k}\Big(\frac{T \cdot \mathbb{1}\{t \in \psi_{k,ij}\}}{|\psi_{k,ij}|} - 1\Big)\Big(\frac{T \cdot \mathbb{1}\{t \in \psi_{k,lj}\}}{|\psi_{k,lj}|} - 1\Big) \xrightarrow{p} \omega_{\psi,k,j}.$$

With the above, Assumption (AD2) is satisfied.

Condition 1 and 2 in Proposition 1 are on the factor strength and missingness pattern, respectively. Condition 1 is trivially satisfied if all factors are pervasive. Condition 2 can be easily satisfied by assuming that in $\mathrm{mat}_k(\mathcal{Y}_t)$, all the elements in each row are missing at random with probability $1 - p_0$. We then have for each $t \in [T]$, as $d_k, T \to \infty$,

$$
\begin{aligned}
& d_k^{-2} \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} \Big(\frac{T \cdot \mathbb{1}\{t \in \psi_{k,ij}\}}{|\psi_{k,ij}|} - 1 \Big) \Big(\frac{T \cdot \mathbb{1}\{t \in \psi_{k,lj}\}}{|\psi_{k,lj}|} - 1 \Big) \\
&= d_k^{-2} \sum_{i=1}^{d_k} \sum_{l=1}^{d_k} \Big(\frac{T^2 \cdot \mathbb{1}\{t \in \psi_{k,ij}\} \cdot \mathbb{1}\{t \in \psi_{k,lj}\}}{|\psi_{k,ij}| \cdot |\psi_{k,lj}|} - \frac{T \cdot \mathbb{1}\{t \in \psi_{k,ij}\}}{|\psi_{k,ij}|} - \frac{T \cdot \mathbb{1}\{t \in \psi_{k,lj}\}}{|\psi_{k,lj}|} + 1 \Big) \\
& \xrightarrow{p} p_0^{-1} - 1,
\end{aligned}
$$

which is $\omega_{\psi,k,j}$. Similar to Assumption S3.2 in Xiong and Pelger (2023), the value of $\omega_{\psi,k,j}$ can be regarded as a measure of missingness complexity. It is a parameter related to the variance of the stable convergence, and tends to increase when there is a larger portion of data missing.

5. Numerical results

5.1. Simulation

We demonstrate the empirical performance of our estimators in this section. Note that we do not have comparisons to other imputation methods since to the best of our knowledge, there are no other general imputation methods available for $K > 1$ apart from tensor completion methods for very specific applications as mentioned in the introduction. However, we will make comparisons with an alternative approach to impute tensor time series combining Xiong and Pelger (2023) and Chen and Lam (2024b), as demonstrated in Section 5.1.4. Under different missing patterns which will be described later, we investigate the performance of the factor loading matrix estimators, the imputation, and the estimator of the number of factors. We also demonstrate asymptotic normality as described in Theorem 3, followed by an example plot of a statistical power function using result 3 of Theorem 4. Throughout this section, each simulation experiment of a particular setting is repeated 1000 times.

For the data generating process, we use model (3.1) together with Assumption (E1), (E2) and (F1). More precisely, the elements in $\mathcal{F}_t$ are independent standardized AR(5) with AR coefficients 0.7, 0.3, −0.4, 0.2, and −0.1. The elements in $\mathcal{F}_{e,t}$ and ϵ_t are generated similarly, but their AR coefficients are (−0.7, −0.3, −0.4, 0.2, 0.1) and (0.8, 0.4, −0.4, 0.2, −0.1) respectively. The standard deviation of each element in ϵ_t is generated by i.i.d. $|\mathcal{N}(0,1)|$.

For each $k \in [K]$, each factor loading matrix $\mathbf{A}_k$ is generated independently with $\mathbf{A}_k = \mathbf{U}_k \mathbf{B}_k$, where each entry of $\mathbf{U}_k \in \mathbb{R}^{d_k \times r_k}$ is i.i.d. $\mathcal{N}(0,1)$, and $\mathbf{B}_k \in \mathbb{R}^{r_k \times r_k}$ is diagonal with the jth diagonal entry being $d_k^{-\zeta_{k,j}}$, $0 \le \zeta_{k,j} \le 0.5$. Pervasive (strong) factors have $\zeta_{k,j} = 0$, while weak factors have $0 < \zeta_{k,j} \le 0.5$. Each entry of $\mathbf{A}_{e,k} \in \mathbb{R}^{d_k \times r_{e,k}}$ is i.i.d. $\mathcal{N}(0,1)$, but has independent probability of 0.95 being set exactly to 0. We set $r_{e,k} = 2$ for all $k \in [K]$ throughout the section.

To investigate the performance with missing data, we consider four missing patterns:

- (M-i) Random missing with probability 0.05.
- (M-ii) Random missing with probability 0.3.
- (M-iii) The missing entries have index $(t, i_1, \dots, i_K)$, where

$$0.5T \le t \le T, \quad 1 \le i_k \le 0.5 d_k \text{ for all } k \in [K].$$

- (M-iv) Conditional random missingness such that the unit with index j along mode-1 is missing with probability 0.2 if $(\mathbf{A}_1)_{j,1} \ge 0$, and with probability 0.5 if $(\mathbf{A}_1)_{j,1} < 0$.

To test how robust our imputation is under heavy-tailed distribution, we consider two distributions for the innovation process in generating $\mathcal{F}_t$, $\mathcal{F}_{e,t}$ and ϵ_t: (1) i.i.d. $\mathcal{N}(0,1)$; (2) i.i.d. t_3.

5.1.1. Accuracy in the factor loading matrix estimators and imputations

For both the factor loading matrix estimators and the imputations, since our procedure for vector time series ($K = 1$) is essentially the same as that in Xiong and Pelger (2023), we show here only the performance for $K = 2, 3$. We use the column space distance

$$\mathcal{D}(\mathbf{Q}, \widehat{\mathbf{Q}}) = \left\| \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}' - \widehat{\mathbf{Q}}(\widehat{\mathbf{Q}}'\widehat{\mathbf{Q}})^{-1}\widehat{\mathbf{Q}}' \right\|$$

for any given $\mathbf{Q}, \widehat{\mathbf{Q}}$, which is a commonly used measure in the literature. For measuring the imputation accuracy, we report the relative mean squared errors (MSE) defined by

$$\text{relative MSE}_{\mathcal{S}} = \frac{\sum_{j \in \mathcal{S}} (\widehat{C}_j - C_j)^2}{\sum_{j \in \mathcal{S}} C_j^2}, \tag{5.1}$$

where $\mathcal{S}$ either denotes the set of all missing, all observed, or all available units.

We consider the following simulation settings:

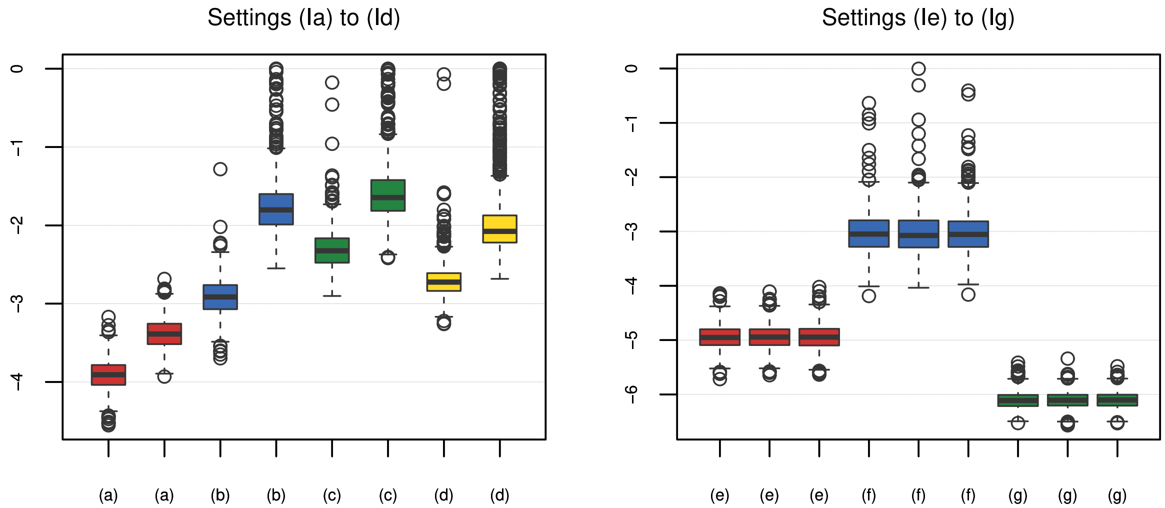

Fig. 2. Plot of the column space distance $\mathcal{D}(\mathbf{Q}_k, \hat{\mathbf{Q}}_k)$ (in log-scale) for $k \in [K]$ for missing pattern (M-i), with $K = 2$ on the left panel and $K = 3$ on the right. The horizontal axis is indexed from (a) to (g) to represent Settings (Ia) to (Ig), with the kth boxplot of each setting corresponding to the kth factor loading matrix $\mathbf{Q}_k$ therein. Performances on other missing patterns are very similar, and are omitted.

(Ia) $K = 2, T = 100, d_1 = d_2 = 40, r_1 = 1, r_2 = 2$. All factors are pervasive with $\zeta_{k,j} = 0$ for all k, j. All innovation processes in constructing $\mathcal{F}_t, \mathcal{F}_{e,t}$ and ϵ_t are i.i.d. standard normal, and missing pattern is (M-i).
(Ib) Same as (Ia), but one factor is weak with $\zeta_{k,1} = 0.2$ for all $k \in [K]$.
(Ic) Same as (Ia), but all innovation processes are i.i.d. t_3, and all factors are weak with $\zeta_{k,j} = 0.2$ for all k, j.
(Id) Same as (Ic), but $T = 200, d_1 = d_2 = 80$.
(Ie) $K = 3, T = 80, d_1 = d_2 = d_3 = 20, r_1 = r_2 = r_3 = 2$. All factors are pervasive with $\zeta_{k,j} = 0$ for all k, j. All innovation processes in constructing $\mathcal{F}_t, \mathcal{F}_{e,t}$ and ϵ_t are i.i.d. standard normal, and missing pattern is (M-i).
(If) Same as (Ie), but all factors are weak with $\zeta_{k,j} = 0.2$ for all k, j.
(Ig) Same as (If), but $T = 200, d_1 = d_2 = d_3 = 40$.

Settings (Ia) to (Id) have $K = 2$, and settings (Ie) to (Ig) have $K = 3$. They all have missing pattern (M-i), but we have considered all settings with missing patterns (M-ii) to (M-iv), with performance of the factor loading matrix estimators very similar to those with missing pattern (M-i). Hence we are only presenting the results for settings (Ia) to (Ig) in Fig. 2 for the missing pattern (M-i). The imputation results for the above settings are collected in Table 1, together with those under different missing patterns.

We can see from Fig. 2 that the factor loading matrix estimators perform worse when there are weak factors or when the distribution of the innovation processes is fat-tailed. However, larger dimensions ameliorate the worsen performance. The increase in the loading space distance from $k = 1$ to $k = 2$ in settings (Ia) to (Id) is due to more factors along mode-2, which naturally incurs more errors compared to smaller r_k. In comparison, the loading space errors shown in the right panel of Fig. 2 are in line for all modes due to the same number of factors along each mode.

From Table 1, we can see that missing pattern (M-iii) is uniformly more difficult in all settings for imputation. This is understandable as there is a large block of data missing in setting (M-iii), so that we obtain less information towards the "centre" of the missing block. This is also the reason why under (M-iii), the imputation performance for the missing set is worse than the observed set, unlike for other missing patterns where all imputation performances are close.

Random missing in (M-i) and (M-ii) are relatively easier for our imputation procedure to handle. Note that if the TALL-WIDE algorithm in Bai and Ng (2021) were to be extended to the case for $K > 1$, it can handle missing pattern (M-iii), but not (M-i) and (M-ii). The design of our method allows us to handle a wider variety of missing patterns, including random missingness. We want to stress that we have made attempts to generalize the TALL-WIDE algorithm to impute high-order time series data for comparisons, yet the method is almost impossible to use in tensor data. The generalization is also too complicated, and hence is not shown here.

5.1.2. *Performance for the estimation of the number of factors*

In this section, we demonstrate the performance of our ratio estimator $\hat{r}_k$ in (4.7) for estimating r_k for $K = 1, 2, 3$. For each $k \in [K]$, we set the value of ξ in Theorem 5 as $\xi = d[(Td_{\text{-}k})^{-1/2} + d_k^{-1/2}]/5$. We have tried a wide range of values other than $1/5$ for ξ in all settings, but $1/5$ is working the best in the vast majority of settings, see simulation results on the sensitivity of different ξ in the supplementary materials. Hence we do not recommend treating it as a tuning parameter in this section for saving computational time.

We present the results under a fully observed scenario and a missing data scenario for each of the following setting:

Table 1
Relative MSE for settings (Ia) to (Ig), reported for $\mathcal{S}$ as the set containing respectively observed (obs), missing (miss), and all (all) units. For $K = 3$, all results presented are multiplied by 10^4.

Setting		K=2			K=3			
Missing Pattern	$\mathcal{S}$	(Ia)	(Ib)	(Ic)	(Id)	(Ie)	(If)	(Ig)
(M-i)	obs	.002	.020	.066	.039	2.61	120	.293
	miss	.002	.020	.066	.039	2.63	121	.294
	all	.002	.020	.066	.039	2.61	120	.293
(M-ii)	obs	.003	.025	.079	.045	5.97	154	.702
	miss	.003	.025	.079	.045	6.06	155	.703
	all	.003	.025	.079	.045	6.00	154	.702
(M-iii)	obs	.004	.025	.079	.048	6.64	136	1.75
	miss	.009	.036	.107	.061	14.7	164	4.02
	all	.005	.026	.083	.050	7.19	138	1.89
(M-iv)	obs	.004	.027	.086	.047	7.75	173	.888
	miss	.004	.028	.088	.047	8.49	179	.964
	all	.004	.027	.086	.047	8.00	175	.914

Table 2
Results for setting (IIa). Each column reports the mean and SD (subscripted, in bracket) of the estimated number of factors over 1000 replications, followed by the correct proportion of the estimates. The estimator $\hat{r}$ is our proposed estimator; $\hat{r}_{\mathrm{re},0}$ and $\hat{r}_{\mathrm{re},1}$ are similar but used imputed data where the imputation is done using the number of factors as $\hat{r}$ and $\hat{r}+1$, respectively; $\hat{r}_{\mathrm{iTIP,re},0}$ and $\hat{r}_{\mathrm{iTIP,re},1}$ are iTIP-ER on imputed data (using $\hat{r}$ and $\hat{r}+1$ respectively); $\hat{r}_{\mathrm{full}}$ and $\hat{r}_{\mathrm{iTTP,full}}$ are our estimator and iTIP-ER on fully observed data (in green), respectively.

Setting (IIa) (True $r_1 = 2$)							
Missing Pattern	$\hat{r}$	$\hat{r}_{\mathrm{re},0}$	$\hat{r}_{\mathrm{re},1}$	$\hat{r}_{\mathrm{iTIP,re},0}$	$\hat{r}_{\mathrm{iTIP,re},1}$	$\hat{r}_{\mathrm{full}}$	$\hat{r}_{\mathrm{iTIP,full}}$
	Mean$_{\mathrm{(SD)}}$						
(M-ii)	$1.98_{(.13)}$	$1.98_{(.13)}$	$2.00_{(.06)}$	$1.97_{(.18)}$	$1.97_{(.22)}$	$1.99_{(.10)}$	$1.92_{(.28)}$
(M-iii)	$1.92_{(.27)}$	$1.93_{(.26)}$	$1.97_{(.20)}$	$1.90_{(.30)}$	$1.92_{(.31)}$		
(M-iv)	$1.98_{(.14)}$	$1.98_{(.14)}$	$2.01_{(.08)}$	$1.97_{(.17)}$	$1.98_{(.24)}$		
	Correct Proportion						
(M-ii)	.982	.982	.996	.967	.949	.99	.917
(M-iii)	.921	.93	.96	.901	.898		
(M-iv)	.979	.979	.993	.97	.943		

(IIa) $K = 1, T = d_1 = 80, r_1 = 2$. All factors are pervasive with $\zeta_{1,j} = 0$ for all j. All innovation processes involved are i.i.d. standard normal. We try missing patterns (M-ii), (M-iii), and (M-iv).
(IIb) Same as (IIa), but one factor is weak with $\zeta_{1,1} = 0.1$ and the missing pattern is only (M-ii).
(IIc) Same as (IIb), but factors are weak with $\zeta_{1,1} = 0.1$ and $\zeta_{1,2} = 0.15$.
(IId) Same as (IIc), but $T = 160$.
(IIIa) $K = 2, T = d_1 = d_2 = 40, r_1 = 2, r_2 = 3$. All factors are pervasive with $\zeta_{k,j} = 0$ for all k, j. All innovation processes involved are i.i.d. standard normal, and the missing pattern is (M-ii).
(IIIb) Same as (IIIa), but all factors are weak with $\zeta_{k,j} = 0.1$ for all k, j.
(IIIc) Same as (IIIb), but $T = d_1 = d_2 = 80$.
(IVa) $K = 3, T = d_1 = d_2 = d_3 = 20, r_1 = 2, r_2 = 3, r_3 = 4$. All factors are pervasive with $\zeta_{k,j} = 0$ for all k, j. All innovation processes involved are i.i.d. standard normal, and the missing pattern is (M-ii).
(IVb) Same as (IVa), but all innovation processes are i.i.d. t_3.
(IVc) Same as (IVa), but $T = 40$.

Since estimating the number of factors with missing data is new to the literature, it is of interest to explore the accuracy of the estimator under different missing patterns. Hence we explore different missing patterns in setting (IIa). Extensive experiments (not shown here) on the imputation accuracy using misspecified number of factors show that underestimation is harmful, while slight overestimation hardly worsen the performance of the imputations. Thus, for each of the above settings, we also compare the performance using re-imputation and iTIP-ER by Han et al. (2022), where the re-imputation is done by using both $\hat{r}_k$ and $\hat{r}_k + 1$ to avoid information loss due to underestimating the number of factors, see Tables 2 and 3.

From both Tables 2 and 3, it is easy to see that our proposed method generally gives more accurate estimates than iTIP-ER, and it is clear that the re-imputation estimate is at least as good as the initial estimate. In fact, $\hat{r}_{\mathrm{re},1}$ outperforms $\hat{r}_{\mathrm{full}}$ which is based on full observation.

Table 3
Results for settings (II), (III), and (IV), excluding (IIa). Refer to Table 2 for the definitions of different estimators. The missing pattern concerned in all settings is (M-ii).

	Correct Proportion						
Setting	$\hat{r}$	$\hat{r}_{\text{re},0}$	$\hat{r}_{\text{re},1}$	$\hat{r}_{\text{iTIP,re},0}$	$\hat{r}_{\text{iTIP,re},1}$	$\hat{r}_{\text{full}}$	$\hat{r}_{\text{iITP,full}}$
$K = 1$ (True $r_1 = 2$)							
(IIb)	.556	.556	.886	.526	.765	.633	.53
(IIc)	.626	.626	.762	.594	.668	.67	.539
(IId)	.791	.791	.817	.794	.837	.812	.767
$K = 2$ (True $(r_1, r_2) = (2, 3)$)							
(IIIa)	1	1	1	.995	.995	1	.994
(IIIb)	.978	.978	.987	.985	.989	.981	.986
(IIIc)	.999	.999	1	1	.996	.999	1
$K = 3$ (True $(r_1, r_2, r_3) = (2, 3, 4)$)							
(IVa)	1	1	1	.987	.987	1	.988
(IVb)	.996	.996	.999	.991	.991	1	.991
(IVc)	1	1	1	.999	1	1	1

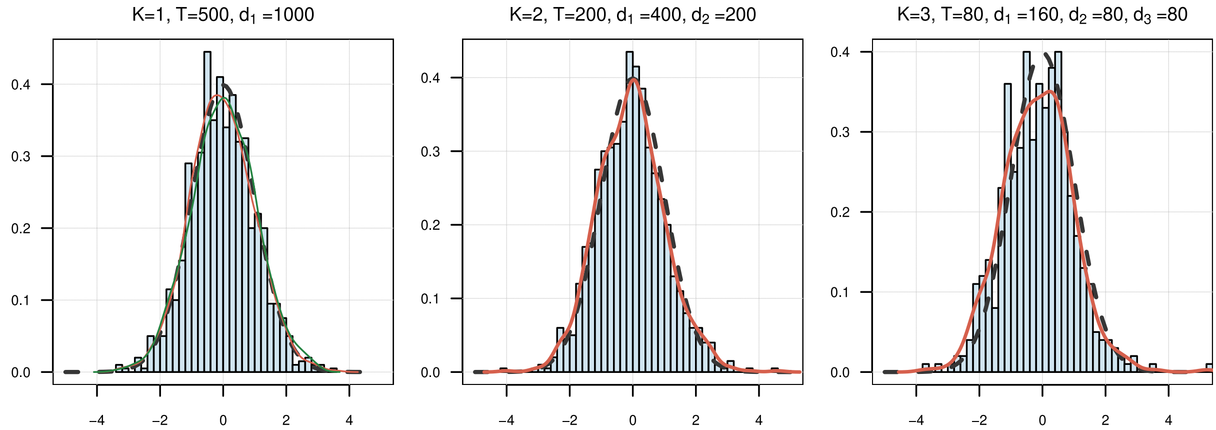

Fig. 3. Histograms of the first entry of $(\hat{\boldsymbol{\Sigma}}_{HAC} + \hat{\boldsymbol{\Sigma}}^{\Delta}_{HAC})^{-1/2}\hat{\mathbf{D}}_1(\hat{\mathbf{Q}}_{1,1\cdot} - \mathbf{H}_1^a \mathbf{Q}_{1,1\cdot})$. In each panel, the curve (in red) is the empirical density, and the other curve (in green) in the left panel depicts the empirical density of the second entry of $(\hat{\boldsymbol{\Sigma}}_{HAC} + \hat{\boldsymbol{\Sigma}}^{\Delta}_{HAC})^{-1/2}\hat{\mathbf{D}}_1(\hat{\mathbf{Q}}_{1,1\cdot} - \mathbf{H}_1^a \mathbf{Q}_{1,1\cdot})$. The density curve for $\mathcal{N}(0,1)$ (in black, dotted) is also superimposed on each histogram.

5.1.3. Asymptotic normality

We present the asymptotic normality results for $K = 1, 2, 3$ respectively. When the data is a vector time series ($K = 1$), our approach is similar to Xiong and Pelger (2023), but their proposed covariance estimator for the asymptotic normality includes information at lag 0 only (i.e., the estimator of the asymptotic variance of the loading estimator), while we use the HAC-type estimator facilitating more serial information. For all K considered, we present the result on $(\hat{\mathbf{Q}})_{11}$, with the parameter β of our HAC-type estimator set as $\lfloor \frac{1}{5}(Td_1)^{1/4} \rfloor$. We use (M-i) as the missing pattern for all settings.

The data generating process is similar to the ones for assessing the factor loading matrix estimators and imputations, but the parameters are slightly adjusted. All elements in $\mathcal{F}_t$, $\mathcal{F}_{e,t}$, and ϵ_t are now independent standardized AR(1) with AR coefficients 0.05, and we use i.i.d. $\mathcal{N}(0,1)$ as the innovation process. We stress that we include contemporary and serial dependence among the noise variables through our construction following Assumption (E1) and (E2), while most existing literatures demonstrating asymptotic normality display results only for i.i.d. Gaussian noise.

We assume all factors are pervasive in this section. For all $K = 1, 2, 3$, given d_1, we set $T, d_i = d_1/2$, $i \neq 1$. We generate a two-factor model for $K = 1$, and a one-factor model for $K = 2, 3$. For the settings $(K, d_1) = (1, 1000), (2, 400), (3, 160)$, we consider $(\hat{\boldsymbol{\Sigma}}_{HAC} + \hat{\boldsymbol{\Sigma}}^{\Delta}_{HAC})^{-1/2}\hat{\mathbf{D}}_1(\hat{\mathbf{Q}}_{1,1\cdot} - \mathbf{H}_1^a \mathbf{Q}_{1,1\cdot})$. In particular, we plot the histograms of the first and second entry in Fig. 3, whereas the corresponding QQ plots are presented in Fig. 4.

The plots in Fig. 3 provide empirical support to Theorem 3 and result 3 of Theorem 4. For $K = 3$, there are some heavy-tail issues, as seen in the bump at the right tail in the histogram (confirmed by its corresponding QQ plot). The QQ plot for $K = 2$ also hints on this, but the tail is thinned as the dimension increases. Our simulation is similar to that in Chen and Fan (2023) for $K = 2$, but we allow partial data unobserved and we generalize to any tensor order K. We remark that the convergence rate of the HAC-type estimator is not completely satisfactory, such that relatively large dimension is needed, and it becomes less feasible for some applications. We leave the improvements of the HAC-type estimator to future work.

Lastly, we demonstrate an example of statistical testing for the above one-factor model for $K = 2$. More precisely, we want to test the null hypothesis $\mathcal{H}_0 : \mathbf{Q}_{1,11} = 0$ with a two-sided test. A 5% significance level is used so that we reject the null if

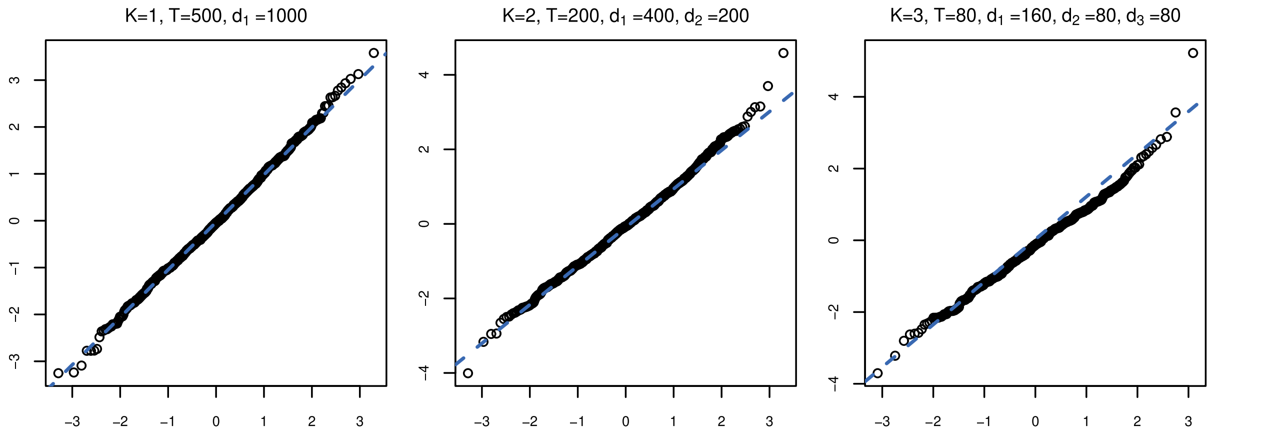

Fig. 4. QQ plots of the first entry of $(\widehat{\boldsymbol{\Sigma}}_{HAC} + \widehat{\boldsymbol{\Sigma}}_{HAC}^{\Delta})^{-1/2}\widehat{\mathbf{D}}_1(\widehat{\mathbf{Q}}_{1,1\cdot} - \mathbf{H}_1^a\mathbf{Q}_{1,1\cdot})$. The horizontal and vertical axes are theoretical and empirical quantiles respectively.

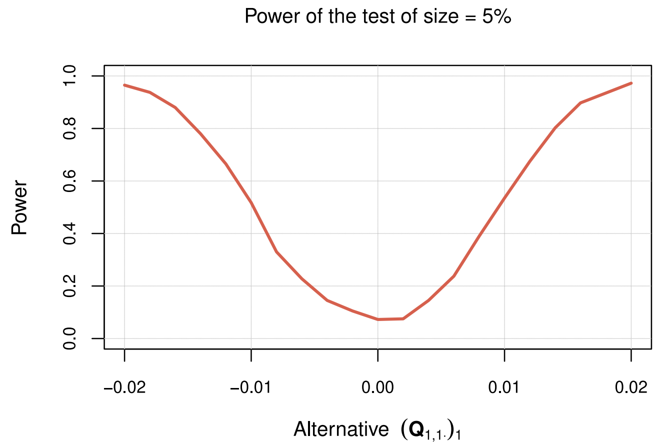

Fig. 5. Statistical power of testing the null hypothesis $\mathcal{H}_0 : (\mathbf{Q}_{1,1\cdot})_1 = \mathbf{Q}_{1,11} = 0$ against the general alternative. The null is rejected when $|(\widehat{\boldsymbol{\Sigma}}_{HAC} + \widehat{\boldsymbol{\Sigma}}_{HAC}^{\Delta})^{-1/2} \widehat{\mathbf{D}}_1\widehat{\mathbf{Q}}_{1,11}| > 1.96$.

$(\widehat{\boldsymbol{\Sigma}}_{HAC} + \widehat{\boldsymbol{\Sigma}}_{HAC}^{\Delta})^{-1/2}\widehat{\mathbf{D}}_1\widehat{\mathbf{Q}}_{1,11}$ is not in $[-1.96, 1.96]$. Each experiment is repeated 400 times and the power function for $\mathbf{Q}_{1,11}$ ranging from -0.02 to 0.02 is presented in Fig. 5. The power function is approximately symmetric, and suggests that our test can successfully reject the null if the true value for $\mathbf{Q}_{1,11}$ is away from 0. When $\mathbf{Q}_{1,11} = 0$, the false positive probability is 7.25% which is slightly higher than the designated size of test. This is due to the slow convergence of the HAC estimators, and an increase in dimensions would improve this.

5.1.4. Comparison with an iterative vectorization-based approach

We compare our proposed tensor factor-based imputation method with the following procedure.

Iterative vectorization-based imputation

1. Given an order-K tensor with missing entries, $\mathcal{Y}_t \in \mathbb{R}^{d_1\times\cdots\times d_K}$ for $t \in [T]$, obtain $\mathbf{y}_t = \textbf{vec}(\mathcal{Y}_t) \in \mathbb{R}^d$ for all time stamps. Impute the vector time series $\{\mathbf{y}_t\}_{t\in[T]}$ by Xiong and Pelger (2023) and denote by the tensorized imputation data $\{\widehat{\mathcal{Y}}_{\text{vec},t}\}_{t\in[T]}$.
2. Replace missing entries in $\mathcal{Y}_t$ by the corresponding entries in $\widehat{\mathcal{Y}}_{\text{vec},t}$. For the resulting time series, estimate the loading matrices, core factors and hence the common components by Chen and Lam (2024b). Denote the series of estimated common components by $\{\widehat{\mathcal{Y}}_{\text{preavg},t}\}_{t\in[T]}$.
3. Iterate from step 2, except that we replace missingness of $\mathcal{Y}_t$ by $\widehat{\mathcal{Y}}_{\text{preavg},t}$ from the previous iteration.

The above algorithm is a natural way of leveraging the vector imputation of Xiong and Pelger (2023) to tensor time series, and the iteration step is akin to Appendix A of Stock and Watson (2002). For demonstration, all innovation processes in constructing $\mathcal{F}_t, \mathcal{F}_{e,t}$ and ϵ_t are i.i.d. standard normal, and all factors are pervasive. In particular, the following settings are considered:

(Va) $K = 2, T = 20, d_1 = d_2 = 40, r_1 = r_2 = 2$, and missing pattern is (M-ii).

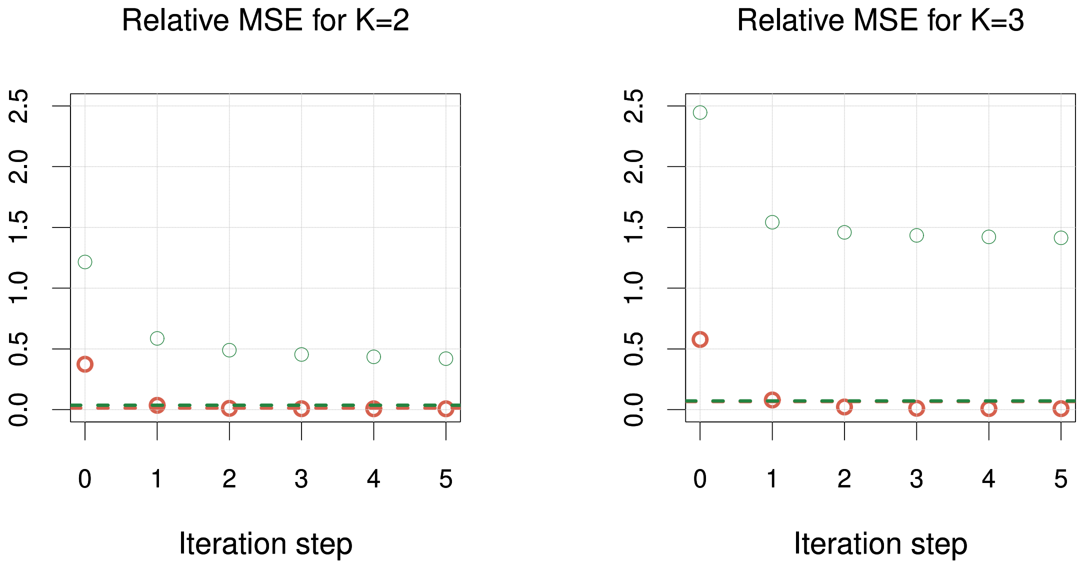

Fig. 6. Plot of the relative MSE for settings (Va) to (Vd), averaged over 1000 replications. Setting (Va), (Vb), (Vc) and (Vd) are represented by the symbols in *red on the left panel*, *green on the left panel*, *red on the right panel* and *green on the right panel*, respectively. Dashed lines denote our tensor-based approach (without iteration), while points denote the iterative vectorization-based method with step 0 corresponding to the initial imputation.

(Vb) Same as (Va), except that the missing pattern is (M-iii).
(Vc) $K = 3, T = 10, d_1 = d_2 = d_3 = 10, r_1 = r_2 = r3 = 2$, and missing pattern is (M-ii).
(Vd) Same as (Vc), except that the missing pattern is (M-iii).

The results for settings (Va) to (Vd) are shown in Fig. 6. From both panels, our proposed method (in dashed lines) performs better than the direct vectorized imputation. One intuition can be the following. Suppose we have a matrix-valued time series $\mathbf{Y}_t \in \mathbb{R}^{d_1 \times d_2}$ for $t \in [T]$, and assume $d_1 \asymp d_2$ and the data is asymptotically observed with the rate $\eta \asymp 1/\sqrt{Td_1}$. According to Corollary 1, the squared imputation error has rate $1/(Td_1) + 1/d_1^2$. In comparison, if we choose to vectorize the data and impute, the squared error rate is $1/T + 1/d_1^2$ which is inflated.

The performance of the vectorization-based imputation can be further improved by iterative imputation in the context of tensor data. However, Fig. 6 demonstrates the low efficiency of such iterative method if the missing pattern is unbalanced to a certain extent. We also point out that the computation time of the initial vectorized imputations can be significantly larger than the our proposed method if the order of the data is large. In fact, the computational complexity (given the number of factors) of direct vectorized imputation is (ignoring the cost of vectorization and unfolding) $O(Td^2 + d^3)$, while our proposed method is $O(K \max_{k\in[K]}\{Tdd_k + d_k^3\})$, which can be of significantly smaller order than d^3.

5.2. Real data analysis: Fama–French portfolio returns

We analyse the set of Fama–French portfolio returns data described in Section 3.1. With sufficient observed samples of each category along its time series, Assumption (O1) in Section 4 can be satisfied and our imputation approach is applicable under such missing pattern. Since the market factor is pervasive in financial returns, we remove the market effect by modelling the data with CAPM as

$$\mathbf{vec}(\mathcal{X}_t) = \mathbf{vec}(\bar{\mathcal{X}}) + \boldsymbol{\beta}(r_t - \bar{r}) + \mathbf{vec}(\mathcal{Y}_t),$$

where $\mathbf{vec}(\mathcal{X}_t) \in \mathbb{R}^{100}$ is the vectorized returns at time t, $\mathbf{vec}(\bar{\mathcal{X}})$ is the sample mean of $\mathbf{vec}(\mathcal{X}_t)$, $\boldsymbol{\beta}$ is the coefficient vector to be estimated, r_t is the return of the NYSE composite index at time t, $\bar{r}$ is the sample mean of r_t, and $\mathbf{vec}(\mathcal{Y}_t)$ is the CAPM residual. We compute the sample mean using only the observed data, and more sophisticated methods could be studied in the future. The least squares solution is

$$\hat{\boldsymbol{\beta}} = \frac{\sum_{t=1}^T (r_t - \bar{r})\{\mathbf{vec}(\mathcal{X}_t) - \mathbf{vec}(\bar{\mathcal{X}})\}}{\sum_{t=1}^T (r_t - \bar{r})^2}.$$

Hence for the rest of this section, we focus on the matrix series $\{\hat{\mathcal{Y}}_t\}_{t\in[570]}$ with $\hat{\mathcal{Y}}_t \in \mathbb{R}^{10\times 10}$, constructed from the estimated CAPM residual $\{\mathbf{vec}(\mathcal{X}_t) - \mathbf{vec}(\bar{\mathcal{X}}) - \hat{\boldsymbol{\beta}}(r_t - \bar{r})\}_{t\in[570]}$.

To estimate the rank of the core factors, we first use our proposed rank estimator to obtain initial estimates $(\hat{r}_1, \hat{r}_2) = (1, 1)$ for both series, followed by re-estimating the rank based on the imputed series using $(\hat{r}_1 + r_*, \hat{r}_2 + r_*)$ with some pre-specified integer r_* to

Table 4
Rank estimators for Fama–French portfolios. Miss-ER represents the rank re-estimated by our proposed eigenvalue-ratio estimator for missing data.

	initial		Miss-ER		BCorTh		iTIP-ER		RTFA-ER	
	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_1$	$\hat{r}_2$
Value Weighted	1	1	1	1	2	1	1	1	1	2
Equal Weighted	1	1	1	1	2	1	1	1	1	2

Table 5
Estimated loading matrices $\hat{\mathbf{Q}}_1$ and $\hat{\mathbf{Q}}_2$ for the value weighted portfolio series, after varimax rotation and scaling (entries rounded to the nearest integer). Magnitudes larger than 9 are in red to highlight units with heavy loadings. All null hypotheses of a row of $\mathbf{Q}_1$ or $\mathbf{Q}_2$ being zero (see (5.2)) are rejected at 5% significance level.

	ME1	ME2	ME3	ME4	ME5	ME6	ME7	ME8	ME9	ME10
Factor 1	−15	−14	−9	−7	−6	−3	−1	0	2	3
Factor 2	5	3	−3	−6	−7	−9	−10	−11	−10	−10
	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8	BE9	BE10
Factor 1	2	−1	−2	−3	−5	−6	−7	−8	−10	−18
Factor 2	16	12	9	7	5	3	3	2	1	−7

capture any omitted weak factors. We have seen in Tables 2 and 3 where such rank re-estimation with $r_* = 1$ is stable and accurate. However, factors can be empirically too weak to detect in the initial estimation under various missing patterns, see the NYC taxi tensor data analysis in the supplementary materials as an extra example. According to previous studies by e.g. Wang et al. (2019), we choose $r_* = 3$ here to ensure sufficient information of factors is carried in the imputation, at the cost of including more noise. For re-estimation, in addition to our eigenvalue-ratio estimator, we also experiment BCorTh by Chen and Lam (2024b), iTIP-ER by Han et al. (2022) and RTFA-ER by He et al. (2022b). The results are presented in Table 4. To ease demonstration, we use $(2,2)$ as the core factor rank for both series hereafter.

With the chosen rank, we perform imputation which is further refined by re-imputation. The results are similar on the two portfolio series, so we only present the one for the value weighted series. The estimated loading matrices are presented in Table 5, after a varimax rotation and scaling. It is clear from the entries in red that on the size factor (i.e., ME loading), ME1 and ME2 form one group ("small size") and ME7 to ME10 form another group ("large size"). On the book-to-equity factor (i.e., BE loading), BE1 and BE2 form a group and BE9 and BE10 form another, which can be interpreted as "undervalued" and "overvalued" respectively. This grouping effect is similarly seen in Table 9 and 10 in Wang et al. (2019).

Moreover, we apply our Theorems 3 and 4 to test if any rows of the loading matrices are zero. For each $k \in [2], i \in [10]$, we test

$$\mathcal{H}_0 : \mathbf{Q}_{k,i\cdot} = \mathbf{0}, \quad \mathcal{H}_1 : \mathbf{Q}_{k,i\cdot} \neq \mathbf{0}. \tag{5.2}$$

The above can be tested since $\mathbf{H}_k^a \mathbf{Q}_{k,i\cdot} = \mathbf{0}$ under the null, and no matter what varimax rotations we use, it retains its meaning. For instance, if $\mathbf{Q}_{1,i\cdot} = \mathbf{0}$, then it means that the ith category of the row factor (here, the ith Market Equity category) is useless in explaining any data variability.

It turns out that at 5% significance level, we cannot reject any null hypotheses for $\mathbf{Q}_{1,i\cdot} = \mathbf{0}$ or $\mathbf{Q}_{2,i\cdot} = \mathbf{0}$, meaning that individual market equity and book-to-equity ratio categories are tested to be meaningful in explaining some variations of the data. See the NYC Taxi traffic data analysis in the supplementary materials for some similar null hypotheses not rejected. We remark that, since the dimensions of our data are not very large, the accuracy of the asymptotic normality and the HAC estimators are weakened, and there can be false positives as a result.

Lastly, two imputation examples for the category (ME10, BE10) are displayed in Fig. 7. From the timestamps on which the portfolio series is observed, we see that the estimated series (in green) does capture some patterns of fluctuations on the true CAPM residual series (in red) and hence can be a good reference for the CAPM residual of portfolios consisted of large size, overvalued stocks. This is certainly more revealing than a naive imputation using zeros or local means. From the above discussions, the estimated factors can be potentially used to replace the Fama–French size factor (SMB) and book-to-equity factor (HML) in a Fama–French factor model for asset pricing, factor trading etc., with a more sophisticated further analysis of the data.

5.3. Real data analysis: OECD economic indicators for countries

We analyse the OECD economic data described in Section 3.2. After investigating the estimated number of factors (Table 6) in a similar re-imputation approach as in Section 5.2, we decide to use $(\hat{r}_1, \hat{r}_2) = (3,3)$ for the rest of this section due to the potentially weak factors suggested by iTIP-ER and RTFA-ER. The estimated loading matrices for countries are presented in Table 7 after a varimax rotation and scaling, with entries highlighted in red to facilitate interpretation. The first factor is mainly formed by European countries except the Northern European ones which, together with Canada, form the third factor. Such regional grouping effects are also confirmed in the second factor which mainly consists of the United States, and the fact that Germany loads also heavily on this factor suggests their similar economic patterns as large economic entities. For the estimated loading for indicators reported in Table 8, CP, PRVM and TOVM form the first factor ("consumption factor"), PP and ULC form the second ("production factor"),

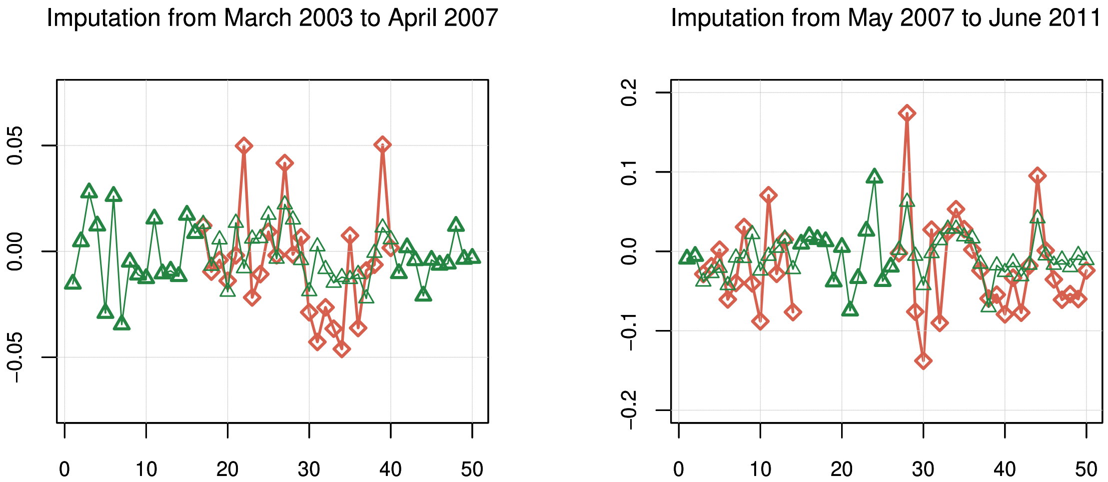

Fig. 7. Two 50-day examples for the value weighted series in the category (ME10, BE10), with horizontal axis of both panels indexed by each day of the selected period. Green triangles denote the estimated series and red squares denote the observed true series. Bold symbols represent the imputed series which consists of the observed series whenever available and the estimated series otherwise.

Table 6
Rank estimators for economic indicators. Refer to Table 4 for the definitions of different estimators.

	initial		Miss-ER		BCorTh		iTIP-ER		RTFA-ER	
	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_1$	$\hat{r}_2$
OECD	1	1	1	1	1	2	4	5	3	3

Table 7
Estimated loading matrix $\hat{\mathbf{Q}}_1$ on three country factors for the OECD data, after varimax rotation and scaling (entries rounded to the nearest integer). Magnitudes larger than 9 are in red to highlight units with heavy loadings. All null hypotheses of a row of $\mathbf{Q}_1$ being zero (see (5.3)) are rejected at 5% significance level.

	BEL	CAN	DNK	FIN	FRA	DEU	GRC	ITA	LUX	NLD	NOR	PRT	ESP	SWE	CHE	GBR	USA
1	−10	4	−3	−7	−8	−8	−9	−7	−1	−2	1	−1	−10	0	−15	−13	1
2	−1	−6	2	5	−2	−12	7	−1	2	−7	0	2	1	−2	0	−1	−24
3	1	−12	−8	−5	-2	3	−6	−4	−11	-6	−12	−11	−2	−10	5	4	−1

Table 8
Estimated loading matrix $\hat{\mathbf{Q}}_2$ on three indicator factors for OECD data, after varimax rotation and scaling (entries rounded to the nearest integer). Magnitudes larger than 9 are in red to highlight units with heavy loadings. All null hypotheses of a row of $\mathbf{Q}_2$ being zero (see (5.3)) are rejected at 5% significance level.

	CA-GDP	CP	EX	IM	IR3TIB	IRLT	IRSTCI	PP	PRVM	TOVM	ULC
1	0	−20	1	3	0	0	0	0	−20	−11	−1
2	0	−6	2	3	1	1	1	20	1	9	18
3	0	9	18	22	−2	−2	−2	1	−4	−2	-1

and EX and IM form the third ("international trade factor").

Moreover, we apply Theorems 3 and 4 to test if a particular row in the two factor loading matrices is zero, meaning that if a country (if a row in $\mathbf{Q}_1$ is $\mathbf{0}$) or an economic indicator (if a row in $\mathbf{Q}_2$ is $\mathbf{0}$) cannot explain any variations in the data. The meaning here is independent of the varimax rotation performed. For each $k \in [2], i \in [d_k], j \in [3]$ with $(d_1, d_2) = (17, 11)$, we form the hypothesis

$$\mathcal{H}_0 : \mathbf{Q}_{k,i\cdot} = \mathbf{0}, \quad \mathcal{H}_1 : \mathbf{Q}_{k,i\cdot} \neq \mathbf{0}. \tag{5.3}$$

Similar to the Fama–French data analysis, all null hypotheses of a row of $\mathbf{Q}_1$ or $\mathbf{Q}_2$ being zero are rejected at 5% significance level. It means that all individual country and economic indicator are tested to be meaningful categories in explaining some variations of the data. Similar to a reminder in Section 5.2, there could be false positives due to the fact that the dimension of the data is not very large.

In Fig. 8, we present two examples of the imputed series overlaid on the observed series. One panel plots ULC of the United States and the other plots PP of the United Kingdom. ULC is a quarterly observed index and the peak pattern in-between each reported timestamp suggests potentially high labour cost in the United States from 1971 to 1975. The PP data in our OECD data is unavailable for the United Kingdom until December 2008. Our imputation implies a gradual increase of the PP before the data is reported, which is reasonable by the impact of the financial crisis. Lastly, we compare between our tensor imputation (matrix imputation for this

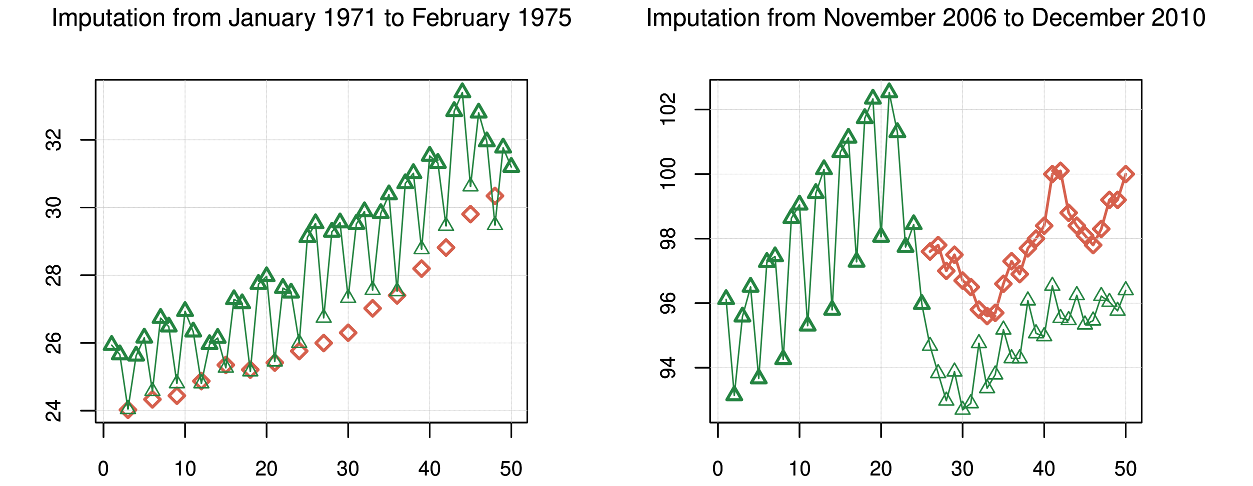

Fig. 8. 50-day examples for unit labour cost of the United States (left panel) and production price index of the United Kingdom (right panel), with horizontal axis of both panels indexed by each day of the selected period. Refer to Fig. 7 for the explanations of different symbols.

Table 9
Comparison of different models for the OECD data. The total sum of squares of the observation is 324,402,709.

	Factor	RSS	# factors	# parameters
Matrix model	(3,3)	7,087,373	9	84
Matrix model	(4,4)	4,542,956	16	112
Matrix model	(5,5)	3,066,851	25	140
Matrix model	(6,6)	1,973,321	36	168
Vector model	2	8,240,976	2	374
Vector model	3	3,954,554	3	561
Vector model	4	2,093,001	4	748

example) and the vectorized imputation using Xiong and Pelger (2023). We use different models to perform imputations whose results are summarized in Table 9 similar to Wang et al. (2019), except that the reported residual sum of squares are computed on the observed entries. Although we require a larger number of factors in general for matrix models, the imputation by matrix models with less parameters can perform better than those by vector models with a much larger number of parameters. This is consistent with the conclusion of Table 11 in Wang et al. (2019).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Proofs of all the theorems in this paper can be found in the supplement of this paper. Instruction in using our R package `tensorMiss` can be found here.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2025.105974.

References

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81 (3), 1203–1227.
Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71 (1), 135–171.
Bai, J., Ng, S., 2021. Matrix completion, counterfactuals, and factor analysis of missing data. J. Am. Stat. Assoc. 116 (536), 1746–1763.
Barigozzi, M., He, Y., Li, L., Trapani, L., 2023. Statistical inference for large-dimensional tensor factor model by iterative projections. arXiv:2206.09800v2.
Cahan, E., Bai, J., Ng, S., 2023. Factor-based imputation of missing values and covariances in panel data of large dimensions. J. Econometrics 233 (1), 113–131.
Chang, J., He, J., Yang, L., Yao, Q., 2023. Modelling matrix time series via a tensor CP-decomposition. J. R. Stat. Soc. Ser. B Stat. Methodol. 85 (1), 127–148.
Chapon, A., Ouarda, T.B., Hamdi, Y., 2023. Imputation of missing values in environmental time series by d-vine copulas. Weather. Clim. Extrem. 41, 100591.
Chen, E.Y., Fan, J., 2023. Statistical inference for high-dimensional matrix-variate factor models. J. Amer. Statist. Assoc. 118 (542), 1038–1055.
Chen, W., Lam, C., 2024a. Factor strength estimation in vector and matrix time series factor models. arXiv:2405.07294v1.
Chen, W., Lam, C., 2024b. Rank and factor loadings estimation in time series tensor factor model by pre-averaging. Ann. Stat. 52 (1), 364–391.

Chen, X., Lei, M., Saunier, N., Sun, L., 2022b. Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. IEEE Trans. Intell. Transp. Syst. 23 (8), 12301–12310.
Chen, R., Xiao, H., Yang, D., 2021. Autoregressive models for matrix-valued time series. J. Econometrics 222 (1, Part B), 539–560, Annals Issue: Financial Econometrics in the Age of the Digital Economy.
Chen, R., Yang, D., Zhang, C.-H., 2022a. Factor models for high-dimensional tensor time series. J. Amer. Statist. Assoc. 117 (537), 94–116.
Freyaldenhoven, S., 2022. Factor models with local factors — determining the number of relevant factors. J. Econometrics 229 (1), 80–102.
Han, Y., Zhang, C.H., Chen, R., 2022. Rank determination in tensor factor model. Electron. J. Stat. 16, 1726–1803.
He, Y., Li, L., Trapani, L., 2022a. Statistical inference for large-dimensional tensor factor model by weighted/unweighted projection. arXiv:2206.09800.
He, Y., Wang, Y., Yu, L., Zhou, W., Zhou, W.-X., 2022b. Matrix kendall's tau in high-dimensions: A robust statistic for matrix factor model. arXiv:2207.09633.
Kazijevs, M., Samad, M.D., 2023. Deep imputation of missing values in time series health data: A review with benchmarking. J. Biomed. Informatics 144, 104440.
Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Rev. 51 (3), 455–500.
Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: Inference for the number of factors. Ann. Statist. 40 (2), 694–726.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorre- lation consistent covariance matrix. Econometrica 55 (3), 703–708.
Onatski, A., 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. J. Econometrics 168 (2), 244–258.
Pan, X., Li, Z., Qin, S., Yu, M., Hu, H., 2021. Sclrtc: imputation for single-cell rna-seq data via low-rank tensor completion. BMC Genomics 22 (860).
Pelger, M., 2019. Large-dimensional factor modeling based on high-frequency observations. J. Econometrics 208 (1), 23–42, Special Issue on Financial Engineering and Risk Management.
Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. J. Bus. Econom. Statist. 20 (2), 147–162.
Tao, M., Su, J., Wang, L., 2019. Land cover classification of polsar image using tensor representation and learning. J. Appl. Remote. Sens. 13 (1), 016516.
Trzcinka, C., 1986. On the number of factors in the arbitrage pricing model. J. Financ. 41, 347–368.
Tsay, R.S., 2023. Matrix-variate time series analysis: A brief review and some new developments. Int. Stat. Rev. http://dx.doi.org/10.1111/insr.12558.
Wang, D., Liu, X., Chen, R., 2019. Factor models for matrix-valued high-dimensional time series. J. Econometrics 208 (1), 231–248, Special Issue on Financial Engineering and Risk Management.
Xiong, R., Pelger, M., 2023. Large dimensional latent factor modeling with missing observations and applications to causal inference. J. Econometrics 233 (1), 271–301.
Zhang, H.-F., 2024. Additive autoregressive models for matrix valued time series. J. Time Series Anal. 45 (3), 398–420.
Zhang, Y., Zhou, B., Cai, X., Guo, W., Ding, X., Yuan, X., 2021. Missing value imputation in multivariate time series with end-to-end generative adversarial networks. Inform. Sci. 551, 67–82.
Zhao, J., Rong, C., Lin, C., Dang, X., 2023. Multivariate time series data imputation using attention-based mechanism. Neurocomputing 542, 126238.