Biometrika (2026), **113**, 2, asaf093
https://doi.org/10.1093/biomet/asaf093
Advance Access publication 26 December 2025

Tail-robust factor modelling of vector and tensor time series in high dimensions

BY MATTEO BARIGOZZI

Department of Economics, Università di Bologna, Via Zamboni 33, 40126 Bologna, Italy
matteo.barigozzi@unibo.it

HAERAN CHO

School of Mathematics, University of Bristol, Woodland Road, Bristol BS8 1UG, U.K.
haeran.cho@bristol.ac.uk

AND HYEYOUNG MAENG

Department of Mathematical Sciences, Durham University, Stockton Road, Durham DH1 3LE, U.K.
hyeyoung.maeng@durham.ac.uk

SUMMARY

We study the problem of factor modelling vector- and tensor-valued time series in the presence of heavy tails in the data, which produce extreme observations with nonnegligible probability. We propose combining a two-step procedure for tensor decomposition with data truncation, which is easy to implement and does not require an iterative search for a numerical solution. Departing from the light-tail assumptions often adopted in the time series factor-modelling literature, we derive the consistency and asymptotic normality of the proposed estimators while assuming the existence of the $(2 + 2\epsilon)$th moment only, for some $\epsilon \in (0, 1)$. Our rates explicitly depend on ϵ, characterizing the effect of heavy tails and the chosen level of truncation. We also propose a consistent criterion for determining the number of factors. Simulation studies and applications to two macroeconomic datasets demonstrate the strong performance of the proposed estimators.

Some key words: Heavy tail; High dimensionality; Robustness; Tensor factor model.

1. INTRODUCTION

Factor modelling is a popular approach to dimension reduction in high-dimensional time series analysis. It has been successfully applied to various tasks involving large panels of time series, from forecasting macroeconomic variables (Stock & Watson, 2002a) to building low-dimensional indicators of the whole economic activity (Stock & Watson, 2002b). With rapid technological developments, tensor (multi-dimensional array) time series datasets are routinely collected, which are large in volume as well as high in dimensions. Consequently, the demand for learning a parsimonious yet flexible representation that preserves the array structure of the original data is greater than ever, driving renewed interest in factor modelling of tensor-valued time series.

Characterization of a low-rank object for tensor data is not trivial, as extensions of the matrix singular value decomposition are not unique for tensors of higher order. Based on the two popular tensor decompositions (Kolda & Bader, 2009), two approaches to tensor factor analysis exist: Tucker factor models and its special case, CanDecomp/PARAFAC (CP; Han & Zhang, 2022; Chang et al., 2023). We focus on the former, which, having gained vast popularity in the literature on high-dimensional statistics, has been adopted for tensor factor modelling under the assumptions that factors account for pervasive serial (Wang et al., 2019; Chen et al., 2022; Han et al., 2024) or cross-sectional (Yu et al., 2022; Chen & Fan, 2023; Chen & Lam, 2024; He et al., 2024; Zhang et al., 2024; Barigozzi et al., 2025b) dependence. Various estimation methods of Tucker tensor factor models exist, which typically exploit a large gap in the eigenvalues of second-moment matrices, present between the eigenvalues attributed to the factors and the remainder. As such, they naturally involve the principal component (PC) analysis, and in turn, their theoretical properties have been studied assuming the existence of the fourth moment in the data.

Datasets exhibiting tail behaviour that does not warrant such fourth-moment conditions are frequently observed in economics and finance (Cont, 2001; Ibragimov et al., 2015), neuroscience (Eklund et al., 2016) and genomics (Purdom & Holmes, 2005), to name a few. In fact, heavy tailedness is one of the stylised features of high-dimensional data that may arise from the increased chance of extreme events, or the complexity of the data-generating mechanism (Fan et al., 2021). It is well documented that sample estimators for the second moments are highly sensitive to anomalous observations and, consequently, the PC-based methods suffer from heavy tailedness (Kristensen, 2014).

Despite growing interest in factor modelling of tensor time series, few papers address the problem of tail-robust estimation, i.e., estimation of factor structures in the absence of a fourth moment, using the Huber loss (Barigozzi et al., 2023; He et al., 2023a, 2025a; Wang et al., 2023). In this paper, we study the problem of tail-robust estimation of a Tucker-type factor model for tensor-valued time series, where the factor tensor accounts for pervasive cross-sectional dependence. A key ingredient of the proposed method is a data truncation step combined with a projection-based iterative procedure for Tucker decomposition of the tensor data that, while computationally and conceptually simple, leads to estimators that are robust to the presence of heavy tails. Throughout, we address the more challenging case of tensor-valued time series, but the proposed methodology is directly applicable to vector-valued time series corresponding to a trivial, single-mode tensor.

In Table 1, we provide a summary of the theoretical and computational aspects of our proposal, in comparison with existing ones. Theoretically, our contributions are two-fold and, to the best of our knowledge, are new to the literature. First, we characterize the tail behaviour of the data under a weak moment assumption on the existence of the $(2 + 2\epsilon)$th moment for some $\epsilon \in (0, 1)$, through which the effect of heavy tails on the choice of the truncation parameter and the rates of estimation is made explicit; in particular, as $\epsilon \to 1$, our rates nearly match the best rate known in light-tailed situations (namely, $\min\{(np_{-k})^{1/2}, p\}$; see the caption of Table 1). This is further complemented by the asymptotic normality of the proposed estimator, which is also the first of its kind in the tail-robust tensor factor modelling literature. Second, provided that the cross sections are suitably ordered, our

Table 1. *Comparison of Tucker factor modelling methods for matrix- and tensor-valued time series, where factors drive pervasive cross-sectional dependence. For each method, we state the rate of estimation attainable for the mode-k loading space and the conditions on the permitted order of the tensor-valued time series (K), whether temporal and/or spatial dependence is allowed on the idiosyncratic component (see Assumptions* 4 *and* 5*) and the assumption on the tail behaviour. Specifically, the 'tail' column reports the considered range of the largest ν such that the data have a finite νth (centred) moment. Additionally, we give the number of required iterations, where '∞' means that the specific number of iterations has not been given. Here, n denotes the sample size, p_k the dimension of the kth mode, $p = \prod_{k=1}^K p_k$ and $p_{-k} = p/p_k$, and M_n denotes an upper bound on the factor elements; see Assumption* 3(iii)

Method	K	Tail	Temporal	Spatial	Rate of estimation	Iteration #
This paper	$\geqslant 1$	$2+2\epsilon$, $\epsilon \in (0,1)$	✔	✘	$\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}} \vee \frac{1}{p}$	2
			✔	✔	$\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}} \vee \left(\frac{\log^K(np)}{np_{-k}}\right)^{\epsilon/(1+\epsilon)} \vee \frac{1}{p}$	1
Wang et al. (2023)	2	$(2, \infty)$	✘	✘	$\frac{1}{\sqrt{np_{-k}}} \vee \frac{1}{\sqrt{np_k}} \vee \frac{1}{\sqrt{p}}$	∞
He et al. (2023a)	2	$(2, \infty)$	✘	✘	$\frac{1}{\sqrt{np_{-k}}} \vee \frac{1}{\sqrt{np_k}} \vee \frac{1}{\sqrt{p}}$	∞
Barigozzi et al. (2023)	$\geqslant 2$	$(2, 4)$	✘	✘	$\min_{k' \in [K]} \frac{1}{\sqrt{p_{-k'}}}$	∞
Chen & Fan (2023)	2	$[8, \infty)$	✔	✔	$\frac{1}{(np_{-k})^{1/2}} \vee \frac{1}{\sqrt{p_k}}$	0
Yu et al. (2022)	2	$[8, \infty)$	✔	✔	$\frac{1}{\sqrt{np_{-k}}} \vee \frac{1}{np_k} \vee \frac{1}{p}$	1
Barigozzi et al. (2025b)	$\geqslant 1$	$[4, \infty)$	✔	✔	$\frac{1}{\sqrt{np_{-k}}} \vee \frac{1}{np_k} \vee \frac{1}{p}$	1
Zhang et al. (2024)	$\geqslant 1$	$[8, \infty)$	✔	✔	$\frac{1}{\sqrt{np_{-k}}} \vee \frac{1}{p}$	∞
Chen & Lam (2024)	$\geqslant 2$	$[4, \infty)$	✔	✔	$\frac{1}{(np_{-k})^{1/2}} \vee \frac{1}{p} \vee \frac{1}{n}$	∞

theoretical analysis permits both serial and spatial dependencies in the idiosyncratic component within a framework of strongly mixing random fields, moving away from the prevalent approach in robust factor modelling that assumes serial and spatial independence. Computationally, our estimators require at most two iterations without any numerical optimization, and are thus straightforward to compute.

We briefly mention alternative robust approaches to time series factor modelling, where the aims are related yet distinct from ours. There are procedures designed to attain a high breakdown point under Huber's contamination model (Huber, 1964), such as those of Baragona & Battaglia (2007), Alonso et al. (2020) and Trucíos et al. (2021) in the context of time series factor modelling, and Maronna & Yohai (2008), Peña & Yohai (2016) and She et al. (2016) on robust principal component analysis. Methods for quantile factor modelling (Chen et al., 2021; He et al., 2023b) enable the estimation of quantile-dependent factors; in particular, at the quantile level of 0.5, these methods can be considered a form of robust factor analysis. There are also papers on elliptical factor models (Fan et al., 2018; Han & Liu, 2018; He et al., 2022; Qiu et al., 2025), where scatter matrices such as Kendall's tau or Spearman correlation matrices are employed for the estimation of factor structures. All the aforementioned papers consider vector time series factor modelling, with the exception of He et al. (2025b).

We use the following notation throughout. We write $[n] = \{1, \ldots, n\}$ for any positive integer n. For a random variable X and $\nu \geqslant 1$, we write $\|X\|_\nu = \{\mathsf{E}(|X|^\nu)\}^{1/\nu}$. For a matrix $\mathbf{A} = [a_{ii'}, i \in [m], i' \in [n]] \in \mathbb{R}^{m \times n}$, we denote by $\mathbf{A}^{\mathrm{T}}$ its transpose, and by $\mathbf{A}_{i\cdot}$ and $\mathbf{A}_{\cdot i}$ the ith row and column vectors. We write $|\mathbf{A}|_2 = (\sum_{i \in [m]} \sum_{i' \in [n]} |a_{ii'}|^2)^{1/2}$ and denote its spectral norm by $\|\mathbf{A}\|$. By $\mathbf{I}$, we denote an identity matrix. For $\mathcal{X} = [X_{i_1 \ldots i_K}, i_k \in$

$[p_k],\ k \in [K]] \in \mathbb{R}^{p_1 \times \cdots \times p_K}$, its mode-$k$ unfolding matrix, denoted $\mathsf{mat}_k(\mathcal{X})$, is the $p_k \times p_{-k}$ matrix that arranges all the p_{-k} mode-k fibers of $\mathcal{X}$ in its columns. We write the mode-k product of $\mathcal{X}$ with an $m \times p_k$ matrix $\mathbf{A} = [a_{ij}]$ as $\mathcal{X} \times_k \mathbf{A} \in \mathbb{R}^{p_1 \times \cdots \times p_{k-1} \times m \times p_{k+1} \times \cdots \times p_K}$, whose $(i_1, \ldots, i_{k-1}, j, i_{k+1}, \ldots, i_K)$th element is given by $\sum_{i_k=1}^{p_k} X_{i_1 \ldots i_k \ldots i_K} a_{j i_k}$. We denote the vectorization of $\mathcal{X}$ as $\mathsf{vec}(\mathcal{X}) \in \mathbb{R}^p$, which stacks the columns of $\mathsf{mat}_1(\mathcal{X})$, and write $|\mathcal{X}|_2 = |\mathsf{vec}(\mathcal{X})|_2$. We use $\otimes$ to denote the Kronecker product. For two real numbers, we define $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. Given two sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if, for some finite positive constant C, there exists $N \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ such that $|a_n||b_n|^{-1} \leqslant C$ for all $n \geqslant N$, and $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. We write $a_n = o(b_n)$ if, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}_0$ such that $|a_n||b_n|^{-1} \leqslant \varepsilon$ for all $n \geqslant N$. By O_P and o_P, we denote the probabilistic extensions of O and o, respectively.

2. TENSOR TIME SERIES FACTOR MODEL

Under a Tucker decomposition-based factor model, a time series $\{\mathcal{X}_t\}_{t \in [n]}$ of the K-dimensional arrays (tensors) with $\mathcal{X}_t = [X_{i_1 \ldots i_K, t},\ i_k \in [p_k],\ k \in [K]] \in \mathbb{R}^{p_1 \times \cdots \times p_K}$ satisfies

$$\mathcal{X}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t, \quad \text{where} \quad \boldsymbol{\chi}_t = \mathcal{F}_t \times_1 \boldsymbol{\Lambda}_1 \times_2 \cdots \times_K \boldsymbol{\Lambda}_K, \tag{1}$$

with $\mathcal{F}_t = [f_{j_1 \ldots j_K, t},\ j_k \in [r_k],\ k \in [r_k]] \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ and $\boldsymbol{\Lambda}_k = [\lambda_{k, i_k j_k},\ i_k \in [p_k],\ j_k \in [r_k]] \in \mathbb{R}^{p_k \times r_k}$ for some integers $r_k \ll p_k$. Then, the $(i_1, \ldots, i_K)$th element of $\boldsymbol{\chi}_t$ is

$$\chi_{i_1 \ldots i_K, t} = \sum_{j_1 \in [r_1]} \cdots \sum_{j_K \in [r_K]} \lambda_{1, i_1 j_1} \cdots \lambda_{K, i_K j_K} f_{j_1 \ldots j_K, t}.$$

The core tensor $\mathcal{F}_t$ serves as the latent common factor, with its dimensions r_k, $k \in [K]$, fixed and independent of $p_1, \ldots, p_K$ and n. It is loaded by mode-wise loading matrices $\boldsymbol{\Lambda}_k$ to drive pervasive cross-sectional dependence in the tensor data, as specified later. The idiosyncratic component $\boldsymbol{\xi}_t = [\xi_{i_1 \cdots i_K, t}]$ is a tensor-valued time series with $\mathsf{E}(\xi_{i_1 \cdots i_K, t}) = 0$, and is allowed to be serially and cross-sectionally dependent, as specified later. Under (1), the mode-k unfolding of $\boldsymbol{\chi}_t$ admits the following decomposition with $\boldsymbol{\Delta}_k := \bigotimes_{l=K, l \neq k}^{1} \boldsymbol{\Lambda}_l$:

$$\begin{aligned} \mathsf{mat}_k(\boldsymbol{\chi}_t) &= \boldsymbol{\Lambda}_k \mathsf{mat}_k(\mathcal{F}_t) \left(\boldsymbol{\Lambda}_K \otimes \cdots \otimes \boldsymbol{\Lambda}_{k+1} \otimes \Lambda_{k-1} \otimes \cdots \otimes \boldsymbol{\Lambda}_1 \right)^{\mathrm{T}} \\ &= \boldsymbol{\Lambda}_k \mathsf{mat}_k(\mathcal{F}_t) \Delta_k^{\mathrm{T}}. \end{aligned} \tag{2}$$

Example 1. When $K = 1$, model (1) reduces to a vector time series factor model such that $\boldsymbol{\chi}_t = \boldsymbol{\Lambda}_1 \mathcal{F}_t$ with $\mathcal{F}_t \in \mathbb{R}^{r_1}$, and its ith element admits the decomposition $\chi_{it} = \sum_{j \in [r_1]} \lambda_{1, ij} f_{j, t}$ for any $i \in [p_1]$. Typical examples of its application include the modelling of macroeconomic indicators of a given country (McCracken & Ng, 2016; see § D.2 in the Supplementary Material). When $K = 2$, model (1) is a matrix time series factor model, where $\boldsymbol{\chi}_t = \boldsymbol{\Lambda}_1 \mathcal{F}_t \boldsymbol{\Lambda}_2^{\mathrm{T}}$ with $\mathcal{F}_t \in \mathbb{R}^{r_1 \times r_2}$, so that $\mathsf{mat}_1(\boldsymbol{\chi}_t) = \boldsymbol{\chi}_t = \boldsymbol{\Lambda}_1 \mathcal{F}_t \boldsymbol{\Lambda}_2^{\mathrm{T}}$ and $\mathsf{mat}_2(\boldsymbol{\chi}_t) = \boldsymbol{\chi}_t^{\mathrm{T}} = \boldsymbol{\Lambda}_2 \mathcal{F}_t^{\mathrm{T}} \boldsymbol{\Lambda}_1^{\mathrm{T}}$; its (i_1, i_2)th element is written as $\chi_{i_1 i_2, t} = \sum_{j_1 \in [r_1]} \sum_{j_2 \in [r_2]} \lambda_{1, i_1 j_1} \lambda_{2, i_2 j_2} f_{j_1 j_2, t}$ for any $i_k \in [p_k]$, $k \in [2]$. In § 6 below, we analyse a matrix time series consisting of macroeconomic indicators (collected in mode 1, i.e., rows) observed in euro-area countries (collected in mode 2, i.e., columns; Barigozzi et al., 2025a).

3. Tail-robust estimation

3.1. *Overview*

Various methods exist for the estimation of the low-rank approximation of tensor data under a Tucker decomposition; we refer the reader to Kolda & Bader (2009) and Luo & Zhang (2023) for an overview. In the time series factor modelling literature, a popular approach for the estimation under (1) is to obtain a pre-estimator via *higher-order singular value decomposition* (SVD, De Lathauwer et al., 2000a), followed by one or more iterations of mode-wise SVD of the data projected onto the pre-estimated loading space (also termed *higher-order orthogonal iteration*; De Lathauwer et al., 2000b); see, e.g., Yu et al. (2022) and He et al. (2024) for the case of matrix factor modelling, which has been extended to higher-order tensors by Zhang et al. (2024) and Barigozzi et al. (2025b).

To address the heavy-tail behaviour of the data, we propose combining the projection-based iterative procedure with a data truncation step, which, despite its simplicity, has not been explored in the context of tail-robust estimation of factor models of any order $K \geqslant 1$. This allows us to accurately estimate the model, even in the presence of extreme events, such as deep economic recessions, the recent pandemic, wars and the subsequent rise in inflation. We first present the proposed methods for estimating the loadings (§ 3.2) and the core tensor factor (§ 3.3), supposing that the factor numbers are known, with the special case of matrix factor modelling discussed in Example 2. Section 3.4 presents our approach to estimating r_k, $k \in [K]$. The selection of tuning parameters, including the level of truncation, is discussed in § 3.5.

3.2. *Factor loading estimation*

Let us write $X_{\mathbf{i},t} = X_{i_1\ldots i_K,t}$ with $\mathbf{i} = (i_1, \ldots, i_K)^{\mathrm{T}}$. Given a truncation parameter $\tau > 0$, we denote the elementwise truncated data as

$$X^{\mathrm{t}}_{\mathbf{i},t}(\tau) := \mathsf{sign}(X_{\mathbf{i},t}) \cdot (|X_{\mathbf{i},t}| \wedge \tau), \quad \text{i.e.,} \quad X^{\mathrm{t}}_{\mathbf{i},t}(\tau) = \begin{cases} X_{\mathbf{i},t} & \text{if } |X_{\mathbf{i},t}| \leqslant \tau, \\ \mathsf{sign}(X_{\mathbf{i},t}) \cdot \tau & \text{if } |X_{\mathbf{i},t}| > \tau, \end{cases} \tag{3}$$

and $\mathcal{X}^{\mathrm{t}}_t(\tau) = [X^{\mathrm{t}}_{\mathbf{i},t}(\tau), \mathbf{i} \in \prod_{k=1}^K [p_k]]$. Then, the sample counterpart of the mode-k second-moment matrix of $\mathcal{X}_t$, defined as $\boldsymbol{\Gamma}^{(k)} := (np_{-k})^{-1} \sum_{t\in[n]} \mathsf{E}[\mathsf{mat}_k(\mathcal{X}_t)\mathsf{mat}_k(\mathcal{X}_t)^{\mathrm{T}}]$, is given by

$$\hat{\boldsymbol{\Gamma}}^{(k)}(\tau) := \frac{1}{np_{-k}} \sum_{t=1}^n \mathsf{mat}_k\{\mathcal{X}^{\mathrm{t}}_t(\tau)\}\mathsf{mat}_k\{\mathcal{X}^{\mathrm{t}}_t(\tau)\}^{\mathrm{T}}. \tag{4}$$

We denote by $\{\hat{\mu}^{(k)}_j(\tau), \hat{\mathbf{e}}^{(k)}_j(\tau)\}$, $j \in [\min(p_k, n)]$, the pairs of eigenvalues and (normalized) eigenvectors of $\hat{\boldsymbol{\Gamma}}^{(k)}(\tau)$, where $\hat{\mu}^{(k)}_j(\tau)$ are ordered in decreasing order. Under (1), for the identifiability between the latent $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ (asymptotically), and that between the $\boldsymbol{\Lambda}_k$ and $\mathcal{F}_t$, commonly made assumptions are that $p_k^{-1}\boldsymbol{\Lambda}_k^{\mathrm{T}}\boldsymbol{\Lambda}_k = \mathbf{I}_{r_k}$ (see Assumption 1 below) and that the entries of $\boldsymbol{\xi}_t$ are weakly correlated within and across the mode (see Assumptions 4–5). They lead to a gap between the r_k largest eigenvalues of $\boldsymbol{\Gamma}^{(k)}$ that diverge linearly in p_k and the remaining ones. Such observations motivate the choice $\hat{\boldsymbol{\Lambda}}_k(\tau) := \sqrt{p_k}\, \hat{\mathbf{E}}_k(\tau) = \sqrt{p_k}[\hat{\mathbf{e}}^{(k)}_j(\tau), 1 \leqslant j \leqslant r_k]$, as an (initial) estimator of $\boldsymbol{\Lambda}_k$, which amounts to the higher-order SVD performed on the truncated data matrix. For vector time series ($K = 1$), we regard $\hat{\mathbf{\Lambda}}_1(\tau)$ as the final estimator.

For $K \geqslant 2$, the representation in (2) for the mode-k unfolding of $\boldsymbol{\chi}_t$ suggests that further refinement can be achieved by projecting $\mathsf{mat}_k\{\mathcal{X}_t^{\mathsf{t}}(\tau)\}$ onto the column space of $\mathbf{\Delta}_k$ followed by SVD. This leads to the following iterative estimator for some $\imath \geqslant 1$:

$$\check{\mathbf{\Lambda}}_k^{[\imath]}(\tau) := \sqrt{p_k}\, \check{\mathbf{E}}_k^{[\imath]}(\tau) \quad \text{with} \quad \check{\mathbf{E}}_k^{[\imath]}(\tau) = [\check{\mathbf{e}}_j^{(k),[\imath]}(\tau),\, 1 \leqslant j \leqslant r_k]. \tag{5}$$

Here, $\{\check{\mu}_j^{(k),[\imath]}(\tau), \check{\mathbf{e}}_j^{(k),[\imath]}(\tau)\}$, $j \in [\min(p_k, n)]$, denotes a pair of eigenvalues and eigenvectors of

$$\check{\mathbf{\Gamma}}^{(k),[\imath]}(\tau) := \frac{1}{np_{-k}} \sum_{t \in [n]} \mathsf{mat}_k\{\mathcal{X}_t^{\mathsf{t}}(\tau)\}\check{\mathbf{D}}_k^{[\imath-1]}(\tau)\{\check{\mathbf{D}}_k^{[\imath-1]}(\tau)\}^{\mathrm{T}}\mathsf{mat}_k\{\mathcal{X}_t^{\mathsf{t}}(\tau)\}^{\mathrm{T}}, \tag{6}$$

with $\check{\mathbf{D}}_k^{[\imath]}(\tau) := \check{\mathbf{E}}_K^{[\imath]}(\tau) \otimes \cdots \otimes \check{\mathbf{E}}_{k+1}^{[\imath]}(\tau) \otimes \check{\mathbf{E}}_{k-1}^{[\imath]}(\tau) \otimes \cdots \otimes \check{\mathbf{E}}_1^{[\imath]}(\tau)$; for convenience, we sometimes write the initial estimator as $\hat{\mathbf{E}}_k(\tau) = \check{\mathbf{E}}_k^{[0]}(\tau)$.

Algorithm 1 below summarizes the steps for the estimation of factor loadings. Our theoretical investigation shows that at most $\imath = 2$ iterations are sufficient for the resultant estimator to achieve asymptotic normality with a rate comparable to those attainable under stronger moment conditions. If $p_1 = \cdots = p_K = p_0 \asymp n$ and $r_1 = \cdots = r_K = r$, the overall cost to obtain the twice-iterated estimator for given τ is $O(np_0^{K+1} + 2Knrp_0^K)$ (Luo et al., 2021). Section 3.5 below describes a cross-validation (CV) procedure for the selection of τ.

Algorithm 1. Projection-based iterative estimation of $\mathbf{\Lambda}_k$, $k \in [K]$.

Input: Data $\{\mathcal{X}_t\}_{t \in [n]}$, factor numbers r_k, $k \in [K]$, truncation parameter τ
Perform data truncation and obtain $\mathcal{X}_t^{\mathsf{t}}(\tau) = [X_{\mathbf{i},t}^{\mathsf{t}}(\tau)]$
for $k \in [K]$ **do**
  Compute $\hat{\mathbf{\Gamma}}^{(k)}(\tau)$ using $\mathsf{mat}_k\{\mathcal{X}_t^{\mathsf{t}}(\tau)\}$ as in (4)
  Obtain $\hat{\mathbf{E}}_k(\tau) = \check{\mathbf{E}}_k^{[0]}(\tau) = [\hat{\mathbf{e}}_j^{(k)}(\tau), j \in [r_k]]$ from the SVD of $\hat{\mathbf{\Gamma}}^{(k)}(\tau)$
Set $\imath \leftarrow 1$
while $\imath \leqslant 2$ **do**
  for $k \in [K]$ **do**
    Compute $\check{\mathbf{\Gamma}}^{(k),[\imath]}(\tau)$ using $\mathsf{mat}_k\{\mathcal{X}_t^{\mathsf{t}}(\tau)\}\check{\mathbf{D}}_k^{[\imath-1]}(\tau)$, as in (6)
    Obtain $\check{\mathbf{\Lambda}}_k^{[\imath]}(\tau) = \sqrt{p_k}[\check{\mathbf{e}}_j^{(k),[\imath]}(\tau), j \in [r_k]]$ from the SVD of $\check{\mathbf{\Gamma}}^{(k),[\imath]}(\tau)$
  Set $\imath \leftarrow \imath + 1$
Output: $\check{\mathbf{\Lambda}}_k^{[\imath]}(\tau)$, $k \in [K]$, $\imath \in \{0, 1, 2\}$.

Remark 1. We propose truncating the elements of $\mathcal{X}_t$ in order to lessen the influence of extreme observations on the estimators of the second-moment matrices. In the literature on high-dimensional vector time series analysis, Wang & Tsay (2023) studied the estimation of (auto)covariance matrices and showed the near-minimax optimality for the truncation-based estimator (see also Remark 3 below). Ke et al. (2019) and Fan et al. (2021) considered spectrum-wise truncation estimators that truncate ℓ_2 or ℓ_4 norms of the random vectors; their theoretical consistency is established under a fourth-moment condition. Alternatively to the proposed truncation, which is applied symmetrically around zero to each $X_{\mathbf{i},t}$, we may adopt winsorization or elementwise Huber regression; we refer the reader to Zhang (2021), where the consistency of the latter estimator is established under a condition comparable to Assumption 3 below. Unlike in the M-estimation framework, the truncation step in (3) allows for explicitly decomposing the estimation errors, which facilitates the theoretical analysis of our proposed estimators.

3.3. *Core tensor factor estimation*

For the estimation of the core tensor $\mathcal{F}_t$, we adopt an analogue of the PC estimator in vector time series factor modelling that takes the form of a cross-sectional weighted average of $X_{\mathbf{i},t}$, $\mathbf{i} \in \prod_{k=1}^{K}[p_k]$, with the weights determined by the estimators of $\mathbf{\Lambda}_k$. Empirically, such an estimator may be sensitive to some $X_{\mathbf{i},t}$ taking extremely large values. Therefore, we consider an estimator of $\mathcal{F}_t$ by taking the weighted average of the truncated observations with a truncation parameter $\kappa > 0$, i.e.,

$$\hat{\mathcal{F}}_t(\tau,\kappa) := \frac{1}{p}\mathcal{X}_t^{\mathrm{t}}(\kappa) \times_1 \check{\mathbf{\Lambda}}_1(\tau)^{\mathcal{T}} \times_2 \cdots \times_K \check{\mathbf{\Lambda}}_K(\tau)^{\mathrm{T}} = \frac{1}{p}\mathcal{X}_t^{\mathrm{t}}(\kappa) \mathop{\times}_{k=1}^{K} \check{\mathbf{\Lambda}}_k(\tau)^{\mathrm{T}}, \tag{7}$$

where $\check{\mathbf{\Lambda}}_k(\tau) = \check{\mathbf{\Lambda}}_k^{[\iota]}(\tau)$ for some $\iota \in \{1,2\}$. Theoretically, the estimator without any additional truncation, namely $\hat{\mathcal{F}}_t(\tau) \equiv \hat{\mathcal{F}}_t(\tau,\infty)$, achieves consistency both in terms of a pointwise error or its ℓ_2 aggregation over time; see Theorem 4 below. However, numerically, there is strong evidence supporting the truncation of observations prior to cross-sectional aggregation, which we explore on simulated datasets in § 5 below.

Example 2. In the special case of a matrix factor model ($K = 2$), the estimation steps given in § 3.2–§ 3.3 can be described without using tensor-related operators. For the loading matrix estimation, we first perform the elementwise truncation with $\tau > 0$ as the truncation parameter, and obtain $\mathcal{X}_t^{\mathrm{t}}(\tau) = [X_{ii',t}^{\mathrm{t}}, i \in [p_1], i' \in [p_2]]$ with $X_{ii',t}^{\mathrm{t}} = \mathsf{sign}(X_{ii',t}) \cdot (|X_{ii',t}| \wedge \tau)$. Then, we estimate the two second-moment matrices of $\mathcal{X}_t$, defined as $\mathbf{\Gamma}^{(1)} = (np_2)^{-1}\sum_{t\in[n]} \mathsf{E}(\mathcal{X}_t\mathcal{X}_t^{\mathrm{T}})$ and $\mathbf{\Gamma}^{(2)} = (np_1)^{-1}\sum_{t\in[n]} \mathsf{E}(\mathcal{X}_t^{\mathcal{T}}\mathcal{X}_t)$, by $\hat{\mathbf{\Gamma}}^{(1)}(\tau) = (np_2)^{-1}\sum_{t\in[n]} \mathcal{X}_t^{\mathrm{t}}(\tau)\mathcal{X}_t^{\mathrm{t}}(\tau)^{\mathrm{T}}$ and $\hat{\mathbf{\Gamma}}^{(2)}(\tau) = (np_1)^{-1}\sum_{t\in[n]} \mathcal{X}_t^{\mathrm{t}}(\tau)^{\mathrm{T}}\mathcal{X}_t^{\mathrm{t}}(\tau)$, respectively. From this, the initial estimator $\hat{\mathbf{\Lambda}}_1(\tau)$ of the row loadings $\mathbf{\Lambda}_1$ is obtained from the r_1 leading eigenvectors of $\hat{\mathbf{\Gamma}}^{(1)}(\tau)$, multiplied by $\sqrt{p_1}$, and the column loadings are estimated by $\hat{\mathbf{\Lambda}}_2(\tau)$ in the same way. Then, at the ιth iteration for $\iota \in \{1,2\}$, we project $\mathcal{X}_t^{\mathrm{t}}(\tau)$ onto the column space of $\check{\mathbf{\Lambda}}_2^{[\iota-1]}(\tau)$ (with $\check{\mathbf{\Lambda}}_2^{[0]}(\tau) = \hat{\mathbf{\Lambda}}_2(\tau)$), compute

$$\check{\mathbf{\Gamma}}^{(1),[\iota]}(\tau) = \frac{1}{np_2}\sum_{t\in[n]} \mathcal{X}_t^{\mathrm{t}}(\tau)\left(\frac{\check{\mathbf{\Lambda}}_2^{[\iota-1]}(\tau)}{\sqrt{p_2}}\right)\left(\frac{\check{\mathbf{\Lambda}}_2^{[\iota-1]}(\tau)}{\sqrt{p_2}}\right)^{\mathrm{T}} \mathcal{X}_t^{\mathrm{t}}(\tau)^{\mathrm{T}}$$

and derive $\check{\mathbf{\Lambda}}_1^{[\iota]}(\tau)$ as its r_1 leading eigenvectors multiplied by $\sqrt{p_1}$. We obtain $\check{\mathbf{\Lambda}}_2^{[\iota]}(\tau)$ analogously. For estimating $\mathcal{F}_t \in \mathbb{R}^{r_1\times r_2}$, we project $\mathcal{X}_t$ after truncation onto the estimated row and column loadings as $\hat{\mathcal{F}}_t(\tau,\kappa) = (p_1p_2)^{-1}\check{\Lambda}_1^{[\iota]}(\tau)^{\mathrm{T}}\mathcal{X}_t^{\mathrm{t}}(\kappa)\check{\mathbf{\Lambda}}_2^{[\iota]}(\tau)$ for some $\kappa > 0$.

3.4. *Factor-number estimation*

Commonly adopted estimators of the factor number exploit the presence of a large gap in the eigenvalues of $\boldsymbol{\Gamma}^{(k)}$ attributed to the presence of factors; see, among others, Bai & Ng (2002), Alessi et al. (2010), Onatski (2010), Ahn & Horenstein (2013) and Fan et al. (2022) in the vector setting, and Han et al. (2022) and Chen & Lam (2024) in the tensor setting. Here, we propose to infer r_k, $k \in [K]$, by screening the ratio of eigenvalues as

$$\hat{r}_k(\tau) := \arg\max_{1 \leqslant j \leqslant \bar{r}_k} \{\check{\mu}_{j+1}^{(k)}(\tau) + \rho\}^{-1} \check{\mu}_j^{(k)}(\tau)$$

with some fixed $1 \leqslant \bar{r}_k \leqslant p_k - 1$, where, for simplicity, we write $\check{\mu}_j^{(k)}(\tau)$ to denote the jth largest eigenvalue of the first-iteration estimator $\check{\boldsymbol{\Gamma}}^{(k),[1]}(\tau)$. We add a small constant ρ in the denominator to ensure that the ratio is well defined. The calculation of $\check{\mu}_j^{(k)}(\tau)$ itself requires projecting the truncated data onto the low-dimensional space spanned by the pre-estimators of $\boldsymbol{\Lambda}_k$. Therefore, we adopt the approach of Barigozzi et al. (2025b) and iteratively update the factor-number estimators and the second-moment matrices of the projected data; see Algorithm 2 below for a full description. Proposition 1 below shows that Algorithm 2 converges after one iteration, provided that $\bar{r}_k$ is chosen sufficiently large.

3.5. *Tuning parameter selection*

We propose selecting the truncation parameters τ and κ via CV. First, a sequence of possible truncation-parameter values are generated as $\mathbf{t} = \{\mathrm{t}_m, m \in [M]\}$, where $\mathrm{t}_1 = \max_{\mathbf{i},t} |X_{\mathbf{i},t}|$, $\mathrm{t}_M = \mathrm{median}_{\mathbf{i},t}|X_{\mathbf{i},t}|$ and t_m, $2 \leqslant m \leqslant M-1$, are chosen such that $\{\log(\mathrm{t}_m), m \in [M]\}$ is a sequence of equi-distanced elements. We then partition the data into L parts with the index sets $\mathcal{I}_\ell = \{\lceil n/L \rceil (\ell - 1) + 1, \ldots, \min(\lceil n/L \rceil \ell, n)\}$, $\ell \in [L]$, and compute the CV measure as

$$\mathrm{CV}(\mathrm{t}_l) := \sum_{k=1}^{K} \sum_{\ell=1}^{L} \left(1 - \frac{1}{r_k} \mathrm{tr}[\check{\mathbf{E}}_{-\ell,k}^{[\iota]}(\mathrm{t}_l)\{\check{\mathbf{E}}_{-\ell,k}^{[\iota]}(\mathrm{t}_l)\}^{\mathrm{T}} \check{\mathbf{E}}_{\ell,k}^{[\iota]}(\mathrm{t}_l)\{\check{\mathbf{E}}_{\ell,k}^{[\iota]}(\mathrm{t}_l)\}^{\mathrm{T}}]\right),$$

where $\check{\mathbf{E}}_{-\ell,k}^{[\iota]}(\mathrm{t}_l)$ (respectively $\check{\mathbf{E}}_{\ell,k}^{[\iota]}(\mathrm{t}_l)$) denotes the estimator of $\boldsymbol{\Lambda}_k/\sqrt{p_k}$ in (5) obtained with $\mathcal{X}_t$, $t \in [n] \setminus \mathcal{I}_\ell$ (respectively $t \in \mathcal{I}_\ell$). Then, we find that $\tau_{\mathrm{CV}} = \arg\min_{\mathrm{t}_l \in \mathbf{t}} \mathrm{CV}(\mathrm{t}_l)$ for the subsequent estimation steps. The factor estimator in (7) involves the sum of p cross sections of $\mathcal{X}_t$. Therefore, the construction of a CV procedure for the selection of κ requires partitioning the cross sections while preserving the tensor structure, which may result in substantial data loss. Empirically, we find that setting $\kappa = \tau_{\mathrm{CV}}$ leads to good numerical performance, and we adopt this choice in all our numerical experiments, which are performed with $\iota = 2$, $L = 3$ and $M = 50$.

For factor-number estimation, we set $\bar{r}_k = \min(\lfloor p_k/2 \rfloor, 20)$ and $\rho = 1/\check{\mu}_1^{(k),[1]}(\tau)$. The estimation of the factor numbers faces the additional difficulty that there is an interplay between the choice of truncation parameter τ and the choice of r_k, $k \in [K]$. In practice, we integrate the CV step into the iterative procedure outlined in Algorithm 2 below; specifically, we set $\tau = \mathrm{t}_1$, the largest value from the grid, and run Algorithm 2 to estimate r_k. These estimators are fed into the CV selection and vice versa, until the estimators of r_k stabilize. In the Supplementary Material, when analysing the real datasets, we explore an approach that identifies a 'region of stability' for the factor-number estimators over varying truncation parameters.

Algorithm 2. Iterative estimation of r_k, $k \in [K]$.

Input: Data $\{\mathcal{X}_t\}_{t\in[n]}$, $\rho > 0$, maximum allowed factor numbers $\bar{r}_k$, $k \in [K]$, truncation parameter τ, maximum number of iterations N

for $k \in [K]$ **do**
  Obtain $\{\hat{\mu}_k^{(k)}(\tau), \hat{\mathbf{e}}_j^{(k)}(\tau)\}$, $j \geqslant 1$, from $\hat{\boldsymbol{\Gamma}}^{(k)}(\tau)$ computed as in (4)
  Initialize $\hat{r}_k^{(0)}(\tau) = \bar{r}_k$

Initialize $m = 1$

while $m \leqslant N$ **do**
  for $k \in [K]$ **do**
    Obtain the eigenvalues $\breve{\mu}_j^{(k),(m)}(\tau)$, $j \geqslant 1$, of $\breve{\boldsymbol{\Gamma}}^{(k),(m)}(\tau) \equiv \breve{\boldsymbol{\Gamma}}^{(k),(m),[1]}(\tau)$, computed as in (6), with $\breve{\mathbf{E}}_k^{(m),[0]}(\tau) = [\hat{\mathbf{e}}_j^{(k)}(\tau),\ 1 \leqslant j \leqslant \hat{r}_k^{(m-1)}(\tau)]$
    Find the ratio-based estimator

$$\hat{r}_k^{(m)}(\tau) = \arg\max_{1\leqslant j\leqslant \bar{r}_k} \breve{\mu}_j^{(k),(m)}(\tau)\{\breve{\mu}_{j+1}^{(k),(m)}(\tau) + \rho\}^{-1} \tag{8}$$

  if $\hat{r}_k^{(m-1)}(\tau) = \hat{r}_k^{(m)}(\tau)$ for all $k \in [K]$ **then**
    Set $\hat{r}_k^{(M)}(\tau) = \hat{r}_k^{(m)}(\tau)$, $k \in [K]$, and $m \leftarrow M + 1$
  else Set $m \leftarrow m + 1$

Output: $\hat{r}_k^{(M)}(\tau)$, $k \in [K]$.

4. Theory

4.1. *Assumptions*

We introduce the assumptions that ensure (asymptotic) identifiability of the factor model, as well as characterizing the tail behaviour of $\{\mathcal{X}_t\}_{t\in[n]}$ and the dependence therein.

Assumption 1. For all $k \in [K]$, $\boldsymbol{\Lambda}_k = [\lambda_{k,ij},\ i \in [p_k],\ j \in [r_k]]$ satisfy

(i) $p_k^{-1}\boldsymbol{\Lambda}_k^{\mathrm{T}}\boldsymbol{\Lambda}_k = \mathbf{I}_{r_k}$ for all $p_k \geqslant r_k$, and
(ii) $\max_{i\in[p_k]}\max_{j\in[r_k]}|\lambda_{k,ij}| \leqslant \bar{\lambda} < \infty$.

Assumption 2. Suppose that $\{\mathcal{F}_t\}_{t\in[n]}$ is a sequence of deterministic K-dimensional arrays of dimensions $r_1 \times \cdots \times r_K$ that are independent of $p_1, \ldots, p_K$ and n. For each $k \in [K]$, there exists a positive definite matrix $\boldsymbol{\Gamma}_f^{(k)} \in \mathbb{R}^{r_k\times r_k}$ such that $\|n^{-1}\sum_{t\in[n]}\mathsf{mat}_k(\mathcal{F}_t)\mathsf{mat}_k(\mathcal{F}_t)^{\mathrm{T}} - \boldsymbol{\Gamma}_f^{(k)}\| = o(1)$ as $n \to \infty$, and the eigenvalues of $\boldsymbol{\Gamma}_f^{(k)}$ are distinct from one another.

Under the Tucker factor model in (1), $\boldsymbol{\Lambda}_k$ and $\mathcal{F}_t$ are not identifiable in that, for any invertible $\mathbf{R} \in \mathbb{R}^{r_k\times r_k}$, we have $\mathsf{mat}_k(\boldsymbol{\chi}_t) = \boldsymbol{\Lambda}_k\mathbf{R}\mathbf{R}^{-1}\mathsf{mat}_k(\mathcal{F}_t)\boldsymbol{\Delta}_k$ (see (2)). Assumptions 1–2 are the tensor analogues of the conditions found in Bai & Ng (2013), and similar conditions are frequently found in the literature; see, e.g., Stock & Watson (2002a), Yu et al. (2022) and Barigozzi et al. (2025b) in vector, matrix and tensor settings. These assumptions ensure that our loading estimators approximate $\boldsymbol{\Lambda}_k$ up to an (asymptotically) orthogonal matrix; see

Theorems 1–3 below. Following the classical approach in factor analysis (Amemiya et al., 1987; Anderson, 2003), we treat factors as being deterministic; see also Bai & Li (2012) and Onatski (2012) in the context of vector-valued time series and Barigozzi et al. (2023) and He et al. (2023a) in matrix or tensor settings. This is a technical assumption adopted to handle the dependence in the data after a nonlinear transform is applied to tackle the heavy tailedness; see Remark 2 below for its possible relaxation.

Assumption 3. There exist constants $\epsilon \in (0, 1)$ and $\omega, C > 0$ such that

(i) $\max_{\mathbf{i} \in \prod_{k=1}^K [p_k]} \max_{t \in \mathbb{Z}} \|\xi_{\mathbf{i},t}\|_{2+2\epsilon} \leqslant \omega$, where $\|\xi_{\mathbf{i},t}\|_{2+2\epsilon} = \{\mathsf{E}(|\xi_{\mathbf{i},t}|^{2+2\epsilon})\}^{1/(2+2\epsilon)}$,
(ii) for all $n \geqslant 1$, $n^{-1} \sum_{t=1}^n |\mathcal{F}_t|_2^\nu \leqslant \omega^\nu$ for $\nu \in \{1+\epsilon, 2, 2+2\epsilon\}$,
(iii) for all $n \geqslant 1$, there exists $M_n > 0$ that satisfies $\max(\max_{t \in [n]} |\mathcal{F}_t|_2, \omega) \leqslant M_n$ and may diverge polynomially with $n \to \infty$ (see (11) below).

Assumption 3(i) relaxes the widely found requirements of the existence of the fourth or higher moments of $\xi_{\mathbf{i},t}$ (see the references in Table 1), or even (sub-)Gaussianity (Chen et al., 2022; Han & Zhang, 2022). Assumption 3(ii) places an analogous condition on the factors while accommodating for their deterministic nature. We do not rule out the situation where $|\mathcal{F}_t|_2$ takes an extreme value at some time-point (see Assumption 3(iii)), whereas Huber loss-based estimation requires the much stronger condition of bounded $\max_t |\mathcal{F}_t|_2$ (Barigozzi et al., 2023; Wang et al., 2023; He et al., 2025a). The rates of estimation we derive make explicit their dependence on ϵ, M_n and $|\mathcal{F}_t|_2$; see Theorems 1–4 below.

Before presenting the condition on the cross-sectional and serial dependence in $\{\mathcal{X}_t\}_{t \in [n]}$, we introduce two definitions. First, for two σ-algebras $\mathcal{A}$ and $\mathcal{B}$, we denote by $\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathsf{P}(A \cap B) - \mathsf{P}(A)\mathsf{P}(B)|$ the α-mixing coefficient. Second, we define the semi-distance between two sets $\mathcal{E}_i \in \mathbb{N}^K \times \mathbb{Z}$, $i = 1, 2$, by

$$\rho(\mathcal{E}_1, \mathcal{E}_2) = \sup \left\{ \max \left(\max_{1 \leqslant k \leqslant K} |i_k - j_k|, \, |t - u| \right) : (\mathbf{i}, t) \in \mathcal{E}_1, \, (\mathbf{j}, u) \in \mathcal{E}_2 \right\}.$$

Then, we make either of the following two assumptions on serial and cross-sectional dependence in $\boldsymbol{\xi} = \{\boldsymbol{\xi}_{\mathbf{i},t}, \, (\mathbf{i}, t) \in \mathbb{N}^K \times \mathbb{Z}\}$.

Assumption 4.

(i) For any $\mathbf{i}, \mathbf{i}' \in \prod_{k=1}^K [p_k]$, the two time series $\{\xi_{\mathbf{i},t}\}_{t \in \mathbb{Z}}$ and $\{\xi_{\mathbf{i}',t}\}_{t \in \mathbb{Z}}$ are independently distributed when $\mathbf{i} \neq \mathbf{i}'$. Also, letting $\alpha_{\mathbf{i}}(m) = \sup_{t \in \mathbb{Z}} \alpha(\sigma\{\xi_{\mathbf{i},u}, \, u \leqslant t\}, \sigma\{\xi_{\mathbf{i},v}, \, v \geqslant t + m\})$ for each $\mathbf{i} \in \prod_{k=1}^K [p_k]$, we have $\alpha_{\mathbf{i}}(m) \leqslant \exp(-c_0 m)$ for all $\mathbf{i}$, for some constant $c_0 > 0$.
(ii) For all $n \geqslant 1$ and $k \in [K]$, there exists some constant $c_\epsilon > 0$ such that

$$\frac{1}{n} \sum_{t,u \in [n]} |\mathcal{F}_t|_2^{1+\epsilon} |\mathcal{F}_u|_2^{1+\epsilon} \exp\left(-\frac{c_0 \epsilon |t - u|}{1 + \epsilon} \right) \leqslant \omega^{2+2\epsilon} \, c_\epsilon,$$

$$\frac{1}{n} \sum_{t,u \in [n]} |\mathcal{F}_t|_2^{1+\epsilon} |\mathcal{F}_u|_2^{1+\epsilon} \exp\left(-\frac{c_0 |t - u|}{3 \log(np_{-k})} \right) \leqslant \omega^{2+2\epsilon} \, c_\epsilon \log(np_{-k}).$$

Assumption 5.

(i) Let $\boldsymbol{\xi}$ be a measurable random field with its strong mixing coefficient defined as $\alpha(m, \ell_1, \ell_2) = \sup_{\mathcal{E}_1, \mathcal{E}_2} [\alpha\{\mathcal{S}(\mathcal{E}_1), \mathcal{S}(\mathcal{E}_2)\}; \, |\mathcal{E}_i| \leqslant \ell_i, \, i = 1, 2, \, \rho(\mathcal{E}_1, \mathcal{E}_2) \geqslant m]$, where, for each set $\mathcal{E}_i \subset \mathbb{N}^K \times \mathbb{Z}$, $|\mathcal{E}_i|$ denotes its cardinality and $\mathcal{S}(\mathcal{E}_i)$ is the σ-algebra generated by $\{\boldsymbol{\xi}_{\mathbf{i},t}, (\mathbf{i}, t) \in \mathcal{E}_i\}$. Then, there exists some constant $c_0 > 0$ such that $\alpha(m) := \alpha(m, \infty, \infty) \leqslant \exp(-c_0 m)$ for all $m \geqslant 0$.

(ii) For all $n \geqslant 1$, there exists some constant $c_\epsilon > 0$ such that

$$\frac{1}{n}\sum_{t,u\in[n]} |\mathcal{F}_t|_2^{1+\epsilon}|\mathcal{F}_u|_2^{1+\epsilon} \exp\left(-\frac{c_0\epsilon|t-u|}{K(1+\epsilon)}\right) \leqslant \omega^{2+2\epsilon}\, c_\epsilon K,$$

$$\frac{1}{n}\sum_{t,u\in[n]} |\mathcal{F}_t|_2^{1+\epsilon}|\mathcal{F}_u|_2^{1+\epsilon} \exp\left(-\frac{c_0|t-u|}{3K\log(np)}\right) \leqslant \omega^{2+2\epsilon}\, c_\epsilon K\log(np).$$

As seen in Table 1, in the existing literature, the analysis of tail-robust factor modelling methods is restricted to the case of temporal and spatial independence, and weak cross-sectional assumptions are permitted only under stronger assumptions on the tail behaviour. We make a first attempt at tail-robust factor modelling of serially and cross-sectionally dependent tensor data under Assumption 5(i), which, adopting the notion of a strongly mixing random field (Doukhan, 1994), presupposes that, across the multiple modes, the cross sections of tensors are arranged in a meaningful order. A natural setting for such an assumption is when the data are collected on a spatial grid, e.g., as in neuroimaging applications (Tzourio-Mazoyer et al., 2002). Assumption 4(i), a special instance of Assumption 5(i), permits serial dependence in $\{\boldsymbol{\xi}_{\mathbf{i},t}\}_{t\in\mathbb{Z}}$ while imposing independence across $\mathbf{i}$.

Remark 2. We may allow for stochastic $\mathcal{F}_t$ and relax Assumption 2 by replacing the above assumptions with those imposed on the moments of $\mathcal{F}_t$ and the conditional distribution of $\boldsymbol{\xi}_t$ given $\mathcal{F} := \{\mathcal{F}_t\}_{t\in[n]}$; e.g., Assumption 3 becomes

(i) $\max_{\mathbf{i}\in\prod_{k=1}^K[p_k]}\max_{t\in\mathbb{Z}}[\mathsf{E}(|\xi_{\mathbf{i},t}|^{2+2\epsilon} \mid \mathcal{F})]^{1/(2+2\epsilon)} \leqslant \omega$,

(ii) $\max_{\mathbf{j}\in\prod_{k=1}^K[r_k]}\max_{t\in[n]}[\mathsf{E}(|\mathcal{F}_{\mathbf{j},t}|_2^{2+2\epsilon})]^{1/(2+2\epsilon)} \leqslant \omega$.

The second conditions in Assumptions 4–5 are akin to the condition bounding $n^{-1}\mathsf{E}(|\sum_{t\in[n]}|\mathcal{F}_t|_2^{1+\epsilon}|^2)$ for stochastic $\mathcal{F}_t$, and can thus be replaced by a strong mixing condition placed on $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$.

Under the above assumptions, the latent common and idiosyncratic components of the Tucker factor model are asymptotically identifiable as $\min(p_1, \ldots, p_K) \to \infty$, since the leading r_k eigenvalues of $\boldsymbol{\Gamma}^{(k)}$ (mode-k second-moment matrix of $\mathcal{X}_t$) are distinct and diverge linearly in p_k, while the remaining ones are bounded; we refer the reader to the Supplementary Material for the precise statement of asymptotic identifiability of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$.

4.2. *Asymptotic properties*

To investigate the theoretical properties of the proposed estimators, we define

$$\tau_{n,p}^{(k)} := \begin{cases} \omega\left(\dfrac{np_{-k}}{\log(np_{-k})}\right)^{1/(2+2\epsilon)} & \text{under Assumption 4,} \\ \omega\left(\dfrac{np_{-k}}{\log^K(np)}\right)^{1/(2+2\epsilon)} & \text{under Assumption 5,} \end{cases} \tag{9}$$

for each $k \in [K]$, where $\epsilon \in (0,1)$ and $\omega > 0$ are defined in Assumption 3, and

$$\psi_{n,p}^{(k)} = \begin{cases} \omega^2 \left(\frac{\log(np_{-k})}{np_{-k}} \right)^{\epsilon/(1+\epsilon)} = (\tau_{n,p}^{(k)})^2 \cdot \frac{\log(np_{-k})}{np_{-k}} & \text{under Assumption 4,} \\ \omega^2 \left(\frac{\log^K(np)}{np_{-k}} \right)^{\epsilon/(1+\epsilon)} = (\tau_{n,p}^{(k)})^2 \cdot \frac{\log^K(np)}{np_{-k}} & \text{under Assumption 5.} \end{cases} \tag{10}$$

Also, let us write $\bar{\psi}_{n,p}^{(k)} = \sum_{k' \in [K] \setminus \{k\}} \psi_{n,p}^{(k')}$ and $\bar{\psi}_{n,p} = \sum_{k \in [K]} \psi_{n,p}^{(k)}$. The following three theorems study the initial and the iteratively projected estimators of $\mathbf{\Lambda}_k$. For vector time series, Theorem 1 gives the final estimation rate, while Theorems 2–3 investigate the consistency and asymptotic distribution of the iterative estimators for tensor time series with $K \geqslant 2$.

Theorem 1 (Initial loading estimator). *Suppose that Assumptions 1, 2 and 3 hold. We set $\tau \asymp \tau_{n,p}^{(k)}$ for each $k \in [K]$ as in (9), and assume that M_n in Assumption 3(iii) satisfies*

$$M_n \max_{k \in [K]} (\tau_{n,p}^{(k)})^{-1} \log^{1/2\epsilon}(np_{-k}) = o(1). \tag{11}$$

Then, there exists a matrix $\hat{\mathbf{H}}_k \in \mathbb{R}^{r_k \times r_k}$ satisfying $\hat{\mathbf{H}}_k^T \hat{\mathbf{H}}_k = \mathbf{I}_{r_k} + o_P(1)$ (in the elementwise sense) such that, as $\min(n, p_1, \ldots, p_K) \to \infty$,

$$\frac{1}{\sqrt{p_k}} \|\hat{\mathbf{\Lambda}}_k(\tau) - \mathbf{\Lambda}_k \hat{\mathbf{H}}_k\| = \begin{cases} O_P\left(\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}} \vee \frac{1}{p_k} \vee \frac{\psi_{n,p}^{(k)}}{\sqrt{p_k}} \right) & \text{under Assumption 4,} \\ O_P\left(\psi_{n,p}^{(k)} \vee \frac{1}{p_k} \right) & \text{under Assumption 5.} \end{cases}$$

Theorem 2 (First iteration loading estimator). *Suppose that Assumptions 1, 2 and 3 hold, that $\tau \asymp \tau_{n,p}^{(k)}$ as in (9) for each $k \in [K]$ and that (11) is met. Then, there exists some $\check{\mathbf{H}}_k^{[1]} \in \mathbb{R}^{r_k \times r_k}$ satisfying $(\check{\mathbf{H}}_k^{[1]})^T \check{\mathbf{H}}_k^{[1]} = \mathbf{I}_{r_k} + o_P(1)$ such that, as $\min(n, p_1, \ldots, p_K) \to \infty$,*

$$\frac{1}{\sqrt{p_k}} \|\check{\mathbf{\Lambda}}_k^{[1]}(\tau) - \mathbf{\Lambda}_k \check{\mathbf{H}}_k^{[1]}\|$$
$$= \begin{cases} O_P\left[\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}} \vee \frac{1}{p} \vee \bar{\psi}_{n,p}^{(k)} \left(\frac{\psi_{n,p}^{(k)}}{\sqrt{p_k}} + \frac{M_n^{1-\epsilon}}{\sqrt{n}} \right) \vee \frac{\bar{\psi}_{n,p}}{\sqrt{p}} \right] & \text{under Assumption 4,} \\ O_P\left(\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}} \vee \frac{1}{p} \vee \psi_{n,p}^{(k)} \vee \frac{M_n^{1-\epsilon} \bar{\psi}_{n,p}^{(k)}}{\sqrt{n}} \right) & \text{under Assumption 5.} \end{cases}$$

Theorem 3 (Second iteration loading estimator). *Suppose that Assumptions 1, 2, 3 and 4 hold, that $\tau \asymp \tau_{n,p}^{(k)}$ as in (9) for each $k \in [K]$ and that (11) is met. Additionally, assume that*

$$\frac{\psi_{n,p}^{(k)}}{\sqrt{p}} \vee \bar{\psi}_{n,p}^{(k)} \left(\frac{\psi_{n,p}^{(k)}}{\sqrt{p_k}} + \frac{M_n^{1-\epsilon}}{\sqrt{n}} + \frac{1}{\sqrt{p}} \right) = O\left(\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}} \vee \frac{1}{p} \right) \tag{12}$$

for all $k \in [K]$ as $\min(n, p_1, \ldots, p_K) \to \infty$.

(i) *There exists some* $\check{\mathbf{H}}_k^{[2]} \in \mathbb{R}^{r_k \times r_k}$ *satisfying* $(\check{\mathbf{H}}_k^{[2]})^{\mathrm{T}} \check{\mathbf{H}}_k^{[2]} = \mathbf{I}_{r_k} + o_P(1)$ *such that*

$$\frac{1}{\sqrt{p_k}} \| \check{\mathbf{\Lambda}}_k^{[2]}(\tau) - \mathbf{\Lambda}_k \check{\mathbf{H}}_k^{[2]} \| = O_P\left(\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}} \vee \frac{1}{p} \right).$$

(ii) *Furthermore, assume that* $M_n = M \in (0, \infty)$ *for all* $n \geqslant 1$*, and that* $(np_{-k})^{1/2} = o(p)$ *for given* $k \in [K]$*. Then, for any* $i \in [p_k]$*, we have, as* $\min(n, p_1, \ldots, p_K) \to \infty$,

$$(np_{-k})^{1/2} \{ \check{\mathbf{\Lambda}}_{k,i\cdot}^{[2]}(\tau) - \mathbf{\Lambda}_{k,i\cdot} \check{\mathbf{H}}_k^{[2]} \}^{\mathrm{T}} \to \mathcal{N}_{r_k}\{\mathbf{0}, \mathbf{\Phi}_i^{(k)}(\tau)\},$$

where

$$\mathbf{\Phi}_i^{(k)}(\tau) := (\mathbf{\Gamma}_f^{(k)})^{-1} \left(\frac{1}{np_{-k}} \sum_{t,u \in [n]} \mathsf{mat}_k(\mathcal{F}_t) \mathbf{\Delta}_k^{\mathrm{T}} \mathbf{\Psi}_{i,tu}^{(k)}(\tau) \mathbf{\Delta}_k \mathsf{mat}_k(\mathcal{F}_u)^{\mathrm{T}} \right) (\mathbf{\Gamma}_f^{(k)})^{-1}$$

and $\mathbf{\Psi}_{i,tu}^{(k)}(\tau) := \mathrm{diag}[\mathsf{cov}\{X_{k,i\ell,t}^{\mathfrak{t}}(\tau), X_{k,i\ell,u}^{\mathfrak{t}}(\tau)\}, \ell \in [p_{-k}]]$ *with* $X_{k,i\ell,t}^{\mathfrak{t}}(\tau)$ *denoting the element of* $\mathsf{mat}_k\{\mathcal{X}_t^{\mathfrak{t}}(\tau)\}$*, such that* $\|\mathbf{\Phi}_i^{(k)}(\tau)\| = O(M^{2-2\epsilon}\omega^{2+2\epsilon})$.

Most notably, the above results make explicit the effect of heavy tails on the rates of estimation through ϵ (see the definition of $\psi_{n,p}^{(k)}$ in (10)) and M_n (defined in Assumption 3(iii)), which is distinguished from the existing work on tail-robust factor modelling. Under the mild condition in (11), which permits M_n to grow polynomially in n, we have $M_n^{1-\epsilon}/(np_{-k})^{1/2} = o(1)$ as $\min(n, p_1, \ldots, p_K) \to \infty$. As an illustration, let $r_k = 1$ for all $k \in [K]$, and suppose that $\{\mathcal{F}_t\}_{t \in [n]}$ is a sequence of independent and identically distributed regularly varying random variables with index $\alpha > 0$. Then we can relate M_n to the data-generating process as $M_n \asymp n^{1/\alpha}$ (Mikosch & Račkauskas, 2010). With $\alpha = 2 + 2\epsilon$, the condition in (11) is readily met, e.g., if $K \geqslant 2$ and $p_k \asymp n^{\gamma}$ for all k with any $\gamma > 0$. Comparing Theorems 1 and 2 shows that the first-iteration estimator reduces the rate of the estimation error attributed to the latency of the factor-driven common component thanks to the projection step, from p_k^{-1} to p^{-1}. Theorem 2 reveals that, as $\epsilon \to 1$, the rates attained by $\check{\mathbf{\Lambda}}_k^{[1]}(\tau)$ are comparable to those available in light-tailed settings. Also, their dependence on the cross-sectional dimensions is sharper (p^{-1}) than those attainable by Huber loss-based methods ($p^{-1/2}$ or $p_k^{-1/2}$), even under the more general Assumption 5 permitting temporal and spatial dependence; see Table 1.

In Theorem 3, analysing $\check{\mathbf{\Lambda}}_k^{[2]}(\tau)$, we impose (12), a mild condition that is readily met if, e.g., $\log(np) = o\{\min(n, p_k)\}$ and $\max_{k' \in [K] \setminus \{k\}} p_{k'} = o(np_k)$ when $\epsilon = 1$. This is for ease of presentation, since any gain in the rate of estimation from the extra iterations is limited to the reduction of nonleading terms. In fact, under Assumption 4 and (12), the rates reported in Theorems 2 and 3 for $\check{\mathbf{\Lambda}}_k^{[\iota]}(\tau)$, $[\iota] \in \{1, 2\}$, are comparable. This is in line with Luo et al. (2021) who, investigating the problem of Tucker decomposition reconstruction, remarked that 'tensor reconstruction error rate of the higher-order orthogonal iteration with only one iteration is optimal' (see their Remark 6). At the same time, the second iteration allows us to establish the asymptotic normality of $\check{\mathbf{\Lambda}}_{k,i\cdot}^{[2]}(\tau)$, which is a first result of its kind in the context of tail-robust tensor factor modelling. The necessity for an extra iteration stems from the fact that, for $\check{\mathbf{\Lambda}}_{k,i\cdot}^{[1]}(\tau)$, the projection involved in (6) is from the initial estimator that attains a suboptimal rate of estimation compared to that involved in $\check{\mathbf{\Lambda}}^{[2]}_{k,i\cdot}(\tau)$. We refer the reader to the Supplementary Material for a numerical investigation into the convergence behaviour of $\check{\mathbf{\Lambda}}^{[\iota]}_{k}(\tau)$.

Remark 3.

(i) In the case of vector time series ($K = 1$), we have $p_{-1} = 1$ and the first-stage (and final) estimator $\hat{\mathbf{\Lambda}}_1(\tau)$ amounts to the popularly adopted PC-based estimator combined with data truncation. Consequently, as $\epsilon \to 1$, the rates in Theorem 1 match (up to a logarithmic factor in the case of spatial dependence) the rates derived in light-tailed settings; see Bai (2003, Theorem 2). Also, an intermediate result (Proposition B.1 in the Supplementary Material) shows that $\hat{\mathbf{\Gamma}}^{(k)}(\tau)$ is near-minimax optimal for vector time series; see Wang & Tsay (2023, Proposition 7).

(ii) In Theorem 3, we make the more stringent condition in (ii) that requires a fixed upper bound on $|\mathcal{F}_t|_2$. If we suppose that $\mathrm{vec}(\mathcal{F}_t)$ is on an ℓ_2 ball of radius M_n, we can relax the condition and still establish the asymptotic normality of $\check{\mathbf{\Lambda}}^{[2]}_{k,i\cdot}(\tau)$ with the rate of convergence suitably adjusted, namely to $(np_{-k})^{1/2}M_n^{-1+\epsilon}$.

THEOREM 4 (TENSOR FACTOR ESTIMATOR). *Suppose that Assumptions 1, 2 and 3 hold, that $\tau \asymp \tau^{(k)}_{n,p}$ is set as in (9) in producing $\check{\mathbf{\Lambda}}^{[\iota]}_{k}(\tau)$ for each $k \in [K]$ and that (11) is met. With $\iota = 2$ under Assumption 4 and $\iota = 1$ under Assumption 5, then, as $\min(n, p_1, \dots, p_K) \to \infty$, the following statements hold.*

(i) *For each* $t \in [n]$,

$$\left|\hat{\mathcal{F}}_t(\tau) - \mathcal{F}_t \mathop{\times}_{k=1}^{K} (\check{\mathbf{H}}^{[\iota]}_k)^{-1}\right|_2 = \begin{cases} O_P\left[(|\mathcal{F}_t|_2 + \omega)\left(\sum_{k\in[K]} \frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}} \vee \frac{1}{\sqrt{p}}\right)\right] & \text{under Assumption 4 and (12)}, \\ O_P\left[(|\mathcal{F}_t|_2 + \omega)\left(\sum_{k\in[K]} \frac{M_n^{1-\epsilon}}{\sqrt{np_{-k'}}} \vee \bar{\psi}_{n,p} \vee \frac{1}{\sqrt{p}}\right)\right] & \text{under Assumption 5}, \end{cases} \tag{13}$$

$$\frac{1}{n}\sum_{t\in[n]}\left|\hat{\mathcal{F}}_t(\tau) - \mathcal{F}_t \mathop{\times}_{k=1}^{K} (\check{\mathbf{H}}^{[\iota]}_k)^{-1}\right|_2^2 = \begin{cases} O_P\left[\omega^2\left\{\sum_{k\in[K]} \left(\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}}\right)^2 \vee \frac{1}{p}\right\}\right] & \text{under Assumption 4 and (12)}, \\ O_P\left[\omega^2\left\{\sum_{k\in[K]} \left(\frac{M_n^{1-\epsilon}}{(np_{-k})^{1/2}}\right)^2 \vee \bar{\psi}^2_{n,p} \vee \frac{1}{p}\right\}\right] & \text{under Assumption 5}. \end{cases} \tag{14}$$

(ii) *Let Assumption 4 hold, and assume that* $\max_{k\in[K]} p_k = o(n)$ *and* $M_n = M \in (0,\infty)$ *for all* $n \geqslant 1$. *Then, for given* $t \in [n]$, *as* $\min(n, p_1, \ldots, p_K) \to \infty$,

$$\sqrt{p}[\mathsf{vec}\{\hat{\mathcal{F}}_t(\tau)\} - \{\check{\mathbf{H}}^{[2]}\}^{-1}\mathsf{vec}(\mathcal{F}_t)] \to \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Upsilon}_t)$$

with $\check{\mathbf{H}}^{[2]} = \check{\mathbf{H}}_K^{[2]} \otimes \cdots \otimes \check{\mathbf{H}}_1^{[2]}$. *Here,* $\boldsymbol{\Upsilon}_t = \mathbf{E}_\chi^{\mathrm{T}}\mathsf{cov}\{\mathsf{vec}(\boldsymbol{\xi}_t)\}\mathbf{E}_\chi$, *where* $\mathbf{E}_\chi := \bigotimes_{k=K}^{1} \mathbf{E}_{\chi,k}$ *with* $\mathbf{E}_{\chi,k}$ *containing the* r_k *largest leading eigenvectors of* $\boldsymbol{\Gamma}_\chi^{(k)} := \boldsymbol{\Lambda}_k\{n^{-1}\sum_{t\in[n]} \mathsf{mat}_k(\mathcal{F}_t)\mathsf{mat}_k(\mathcal{F}_t)^\top\}\boldsymbol{\Lambda}_k^{\mathrm{T}}$, *and satisfies* $\|\boldsymbol{\Upsilon}_t\| \leqslant \omega^2$.

Theorem 4 shows that, when the interest lies in estimating the factor tensor at a given t, the rate of estimation scales with $|\mathcal{F}_t|_2$, which may grow with $\min(n, p_1, \ldots, p_K)$, unlike when investigating the averaged ℓ_2 error over all $t \in [n]$. When $|\mathcal{F}_t|_2 = O(1)$, the results in (13) are comparable to those found in light-tail settings; see Barigozzi et al. (2025b, Theorem 3.6), who reported the rate $O_P(p^{-1/2})$ when n is sufficiently large. In heavy-tailed situations, Barigozzi et al. (2023, Theorem 3.3) derived a rate of $O_P(\max_{k\in[K]} p_{-k}^{-1})$, compared to which (14) reports a far more competitive rate. A careful inspection of the proof of Theorem 4 reveals that the rate does not change when we additionally truncate $\mathcal{X}_t$ with a truncation parameter $\kappa = \kappa_p \to \infty$ as $\min(p_1, \ldots, p_K) \to \infty$. In § 5 below, we show that, numerically, additional truncation contributes to reducing the error in estimating the common component, and thus we advocate such an approach in practice.

Finally, the following proposition shows that, provided that the initial estimator $\hat{r}_k^{(0)} = \bar{r}_k$ is chosen appropriately large, Algorithm 2 iteratively gives a consistent estimator of r_k.

Proposition 1 (Factor-number estimator). *Suppose that either Assumptions 1, 2, 3 and 4 hold or Assumption 5 holds, that* $\tau \asymp \tau_{n,p}^{(k)}$ *is set as in* (9) *and that* (11) *is met. Also, let* $\rho = \rho_{n,p} \to 0$ *satisfy, as* $\min(n, p_1, \ldots, p_K) \to \infty$,

$$\sum_{k'\in[K]} \left(\psi_{n,p}^{(k')} \vee \frac{1}{p_{k'}}\right) = O(\rho_{n,p}).$$

Then, provided that $r_k \leqslant r_k^{(m-1)} \leqslant \bar{r}_k$, *the estimator in* (8) *satisfies* $\mathsf{P}\{\hat{r}_k^{(m)}(\tau) = r_k\} \to 1$.

5. Simulation studies

In this section, we focus on the tensor-valued time series scenarios and defer the vector case to the Supplementary Material. Following Barigozzi et al. (2023), we generate tensor time series with $K = 3$ and $(r_1, r_2, r_3) = (3, 3, 3)$, according to the following three scenarios while varying $n \in \{100, 200, 500\}$: (T1) $(p_1, p_2, p_3) = (10, 10, 10)$, (T2) $(p_1, p_2, p_3) = (100, 10, 10)$ and (T3) $(p_1, p_2, p_3) = (20, 30, 40)$. For the generation of $\mathcal{F}_t$ and $\boldsymbol{\xi}_t$, we consider Gaussian and t_3 distributions, and introduce both temporal and spatial dependence. We additionally consider the situations where either the factors or the idiosyncratic components are contaminated by outliers at random, setting $\varrho \in \{0, 0.1, 0.5, 1\} \times 10^{-2}$ as the proportion of outliers out of the nr entries of $\{\mathcal{F}_t\}_{t\in[n]}$ or the np entries of $\{\boldsymbol{\xi}_t\}_{t\in[n]}$. This exercise is motivated by Raymaekers & Rousseeuw (2024), who, while noting the differences between the models for (cellwise) outliers and heavy tails and the objectives thereof, also remarked that 'estimators for heavy-tailed data can still perform reasonably well under cellwise contamination'.

In comparison with our proposed estimator ($\check{\mathbf{\Lambda}}_k^{[2]}(\tau)$, referred to as 'Trunc'), we include the iterative projection procedure of Barigozzi et al. (2025b) (iPE) and the pre-averaging-based estimator of Chen & Lam (2024) (PreAve), both of which are designed for light-tailed situations and included as a benchmark only, to gauge the advantage of adopting Trunc in the presence of heavy tails and outliers. We also consider the Huber loss-based estimator of Barigozzi et al. (2023) (RTFA), which has been shown to perform competitively against a variety of existing methods for tensor factor analysis (Chen et al., 2022; Zhang et al., 2024), both under light- and heavy-tailed settings. When investigating the performance in common-component estimation, we examine the role of additional truncation in tensor factor estimation by comparing Trunc (truncation applied to factor estimation with $\kappa = \tau$) and 'noTrunc' ($\kappa = \infty$); see the discussion below Theorem 4. When investigating the loading and common-component estimation performance, we treat the factor numbers as known, and the performance of the factor-number estimator is examined separately in the Supplementary Material, along with complete descriptions and results.

As a representative example, Fig. 1 reports some estimation results obtained under (T3). In loading estimation, all methods show marginally improved performance with growing n, and the modewise error is larger for the mode k corresponding to smaller p_k (hence larger np_{-k}), confirming the theoretical findings. Comparing the performance under (T1) ($p = 10^3$), (T2) ($p = 10^4$) and (T3) ($p = 2.4 \times 10^4$), estimation errors decrease considerably as p increases. Under Gaussianity with no outliers, iPE and RTFA perform the best, but with the introduction of heavy tails and idiosyncratic outliers, Trunc performs as well as, or better than, iPE, RTFA and PreAve according to all metrics.

Noteworthy differences are observed when the outliers are present in $\boldsymbol{\xi}_t$: Trunc is not affected by the growing proportion of outliers, thanks to the data-driven truncation-parameter selection via CV, which is effective in loading-space estimation as well as factor estimation (and, together, the estimation of the common component).

RTFA, while attaining tail robustness in loading-space estimation via Huber loss minimization, is sensitive to the presence of anomalous observations when estimating the common component. This is attributed to the fact that RTFA (as well as iPE and PreAve) simply takes a weighted average of the raw data for the estimation of $\mathcal{F}_t$, unlike Trunc, and the benefit of data truncation in factor estimation is also apparent when comparing Trunc and noTrunc, see the bottom panel of Fig. 1. The local estimation error is greater than the global one, which conforms to Theorem 4(i), showing that the estimation error at a given t scales with $|\mathcal{F}_t|_2$. Overall, the performance of Trunc in most settings does not deviate far from its performance in the Gaussian setting without any outlier. We defer the case in which outliers are present in $\mathcal{F}_t$ to the Supplementary Material; while it is not possible to recover the uncontaminated factors, as all cross sections of $\mathcal{X}_t$ are contaminated by the outliers, reasonable recovery of loadings and $\boldsymbol{\chi}_t$ (post-contamination) is achieved by most methods.

Finally, Fig. 2 verifies the asymptotic normality of $\check{\mathbf{\Lambda}}_k^{[2]}(\tau)$ derived in Theorem 3; see the Supplementary Material for the complete results.

6. EURO-AREA MACROECONOMIC DATA

We analyse the EA-MD (Barigozzi et al., 2025a), a collection of 37 macroeconomic indicators that have been collected at a monthly frequency for eight euro-area (EA) countries (Austria, Belgium, Germany, Greece, Spain, France, Italy and the Netherlands). The data

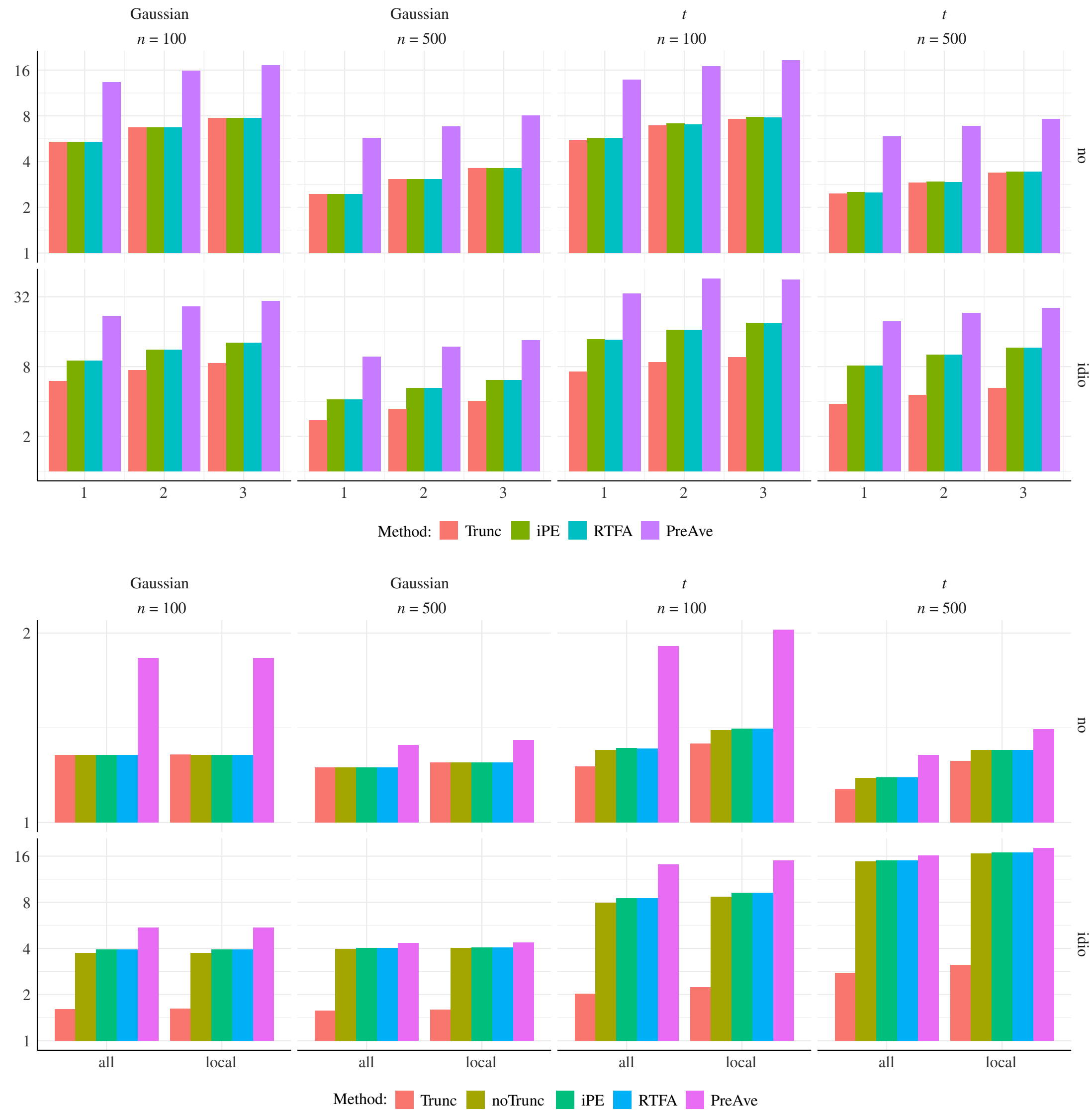

Fig. 1. Scenario (T3) with $(p_1, p_2, p_3) = (20, 30, 40)$. Top: loading estimation errors, measured as in (C.2) in the Supplementary Material, for each mode (x axis) for Trunc, iPE, RTFA and PreAve averaged over 100 realizations per setting, over varying $n \in \{100, 500\}$ and distributions for $\mathcal{F}_t$ and $\boldsymbol{\xi}_t$ (Gaussian and t_3). Bottom: common-component estimation errors measured as in (C.3) in the Supplementary Material with $\mathcal{T} = [n]$ ('all') and $\mathcal{T} = \{n - 10 + 1, \ldots, n\}$ ('local') (x axis), where we additionally include 'noTrunc' (see the text). Within each plot, we consider the cases of no outlier ('no'), and when the outliers are in 0.5% of the entries of $\boldsymbol{\xi}_t$ ('idio'). The y axis is on a log scale, and all errors have been scaled for better presentation.

form a matrix-valued ($K = 2$) time series of dimensions $(p_1, p_2) = (8, 37)$ and span the period from 2002-02 to 2023-09 ($n = 257$) without any missing observations. All time series are individually transformed to stationarity based on standard unit-root tests, as suggested by Barigozzi et al. (2025a), and are further centred and standardized using the median and the mean absolute deviation, respectively. In this dataset, the COVID-19 pandemic is likely to play a major role since such a large outlier is likely to drive most of the co-movement due to the short sample size n, and thus possibly bias the factor analysis, which motivates the use of our robust approach.

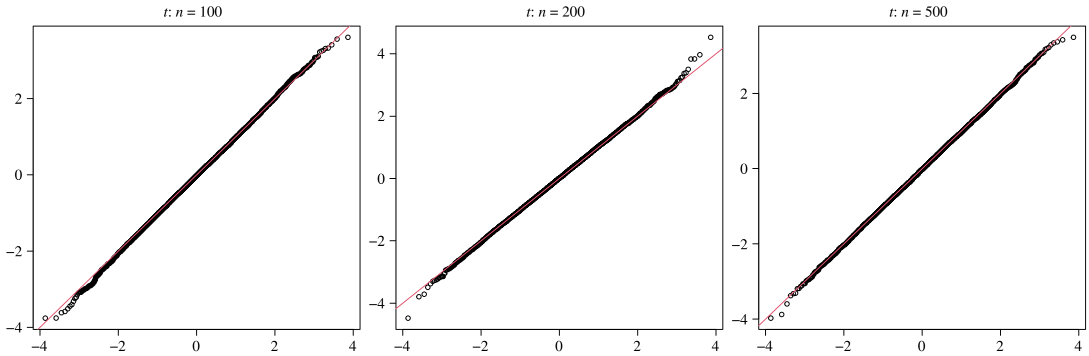

Fig. 2. Scenario (T3): plots of the sample quantiles of the scaled and centred entries of $\check{\mathbf{\Lambda}}_k^{[2]}(\tau)$ (y axis) against the quantiles from the standard normal distribution (x axis) over varying $n \in \{100, 200, 500\}$ (left to right) when the data are generated from the t_3 distribution. In each plot, the $y = x$ line is superimposed. See the Supplementary Material for full details.

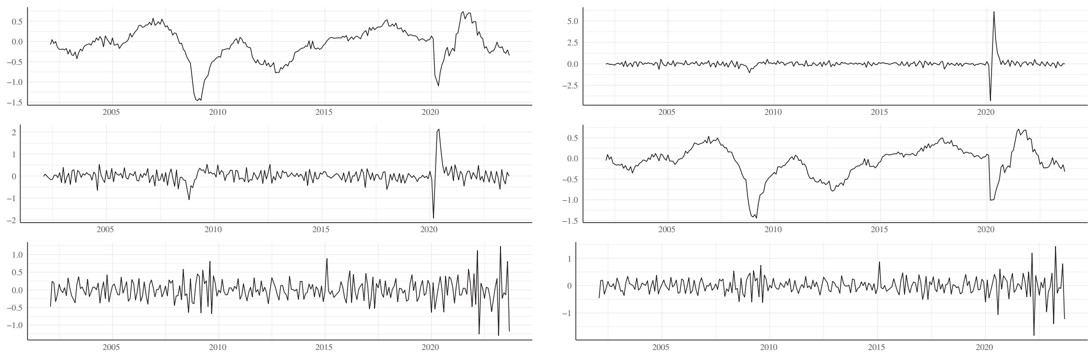

Fig. 3. EA-MD: factor time series $\hat{f}_{1j,t}(\tau, \kappa)$ for $j = 1, 2, 3$ (top to bottom) with (left) and without (right) truncation.

Table 2. *EA-MD: we report the mean, median and standard deviation of the one-step-ahead forecasting errors across the eight countries over time for IPMN and HICPNEF, with and without truncation. Additionally, we report the percentage of instances in which the forecasting error of Trunc is smaller than that of noTrunc for each indicator ('Perc'). See (D.1) in the Supplementary Material for the definition of the forecasting error*

	IPMN (in 10^{-2})				HICPNEF (in 10^{-3})			
	Mean	Median	SE	Perc	Mean	Median	SE	Perc
Trunc	1.524	1.242	1.222	56.548	4.247	2.706	4.248	54.762
noTrunc	1.632	1.323	1.422	–	4.273	2.765	4.184	–

We adopt the ratio-based estimator discussed in § 3.4 with $\bar{r}_k = \min(\lfloor p_k/2 \rfloor, 20)$, which, for most choices of the truncation parameter, returns $(\hat{r}_1, \hat{r}_2) = (1, 3)$, where the 'regions of stability' are achieved over varying τ (see Fig. D.1 in the Supplementary Material). This agrees with the output from the method of Barigozzi et al. (2023), while that of Chen & Lam (2024) returns $\hat{r}_1 = 1$ and $\hat{r}_2 \in \{2, 3, 4\}$ due to the random projections adopted therein.

Setting $(\hat{r}_1, \hat{r}_2) = (1, 3)$, we investigate the efficacy of truncation by comparing the proposed truncation-based estimator ('Trunc') against the non-tail-robust counterpart

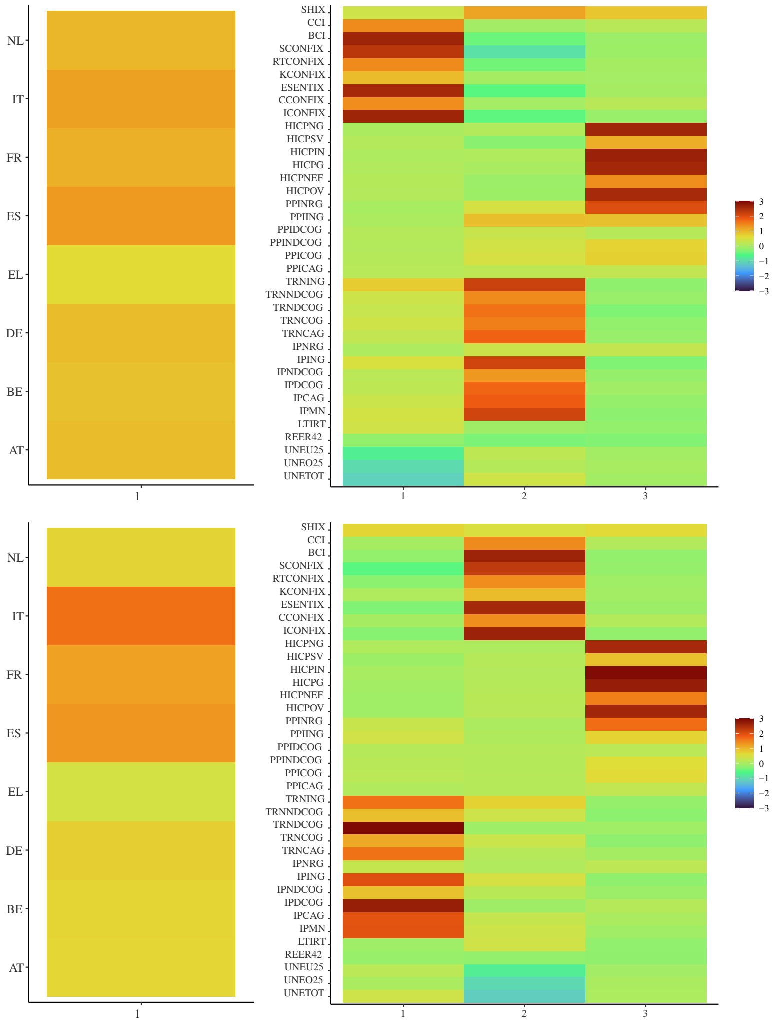

Fig. 4. EA-MD: estimated loading matrices $\check{\mathbf{\Lambda}}_k^{[2]}(\tau)$ for $k = 1$ (left, $(p_1, \hat{r}_1) = (8, 1)$) and $k = 2$ (right, $(p_2, \hat{r}_2) = (37, 3)$) with and without truncation (top to bottom).

('noTrunc' with $\tau = \kappa = \infty$). The CV procedure in § 3.5 returns $\tau_{\mathcal{CV}} \approx 5.306$. Assumptions 1 and 2 allow for the identification of factors up to an orthogonal transformation, but they lack economic meaning; that is, our estimation method can be used for exploratory factor analysis, but not for confirmatory factor analysis in general. Therefore, we apply the Varimax rotation (Mardia et al., 1979) to $\check{\mathbf{\Lambda}}_2^{[2]}(\tau) \in \mathbb{R}^{37\times 3}$, which results in a rotation matrix close to the identity matrix. With $\hat{r}_1 = 1$, the column vector $\check{\mathbf{\Lambda}}_1^{[2]}(\tau)$ does not suffer from the identification issue.

The noTrunc approach over-represents the potential outliers around COVID-19, which results in the leading factor estimates (top right of Fig. 3) with sharp peaks around early 2020. While the effect of these outliers is also observable from the second factor returned by Trunc, its extent is limited, and the leading factor returned by Trunc clearly represents the economic cycle characterised by lower-frequency oscillations typical of the business cycle, with downturns during recessions. We may regard the results from Trunc as more realistic;

indeed, the main drivers of macroeconomic datasets are often found to be those factors associated with the real economic activity, such as industrial production, which also drive the business cycle (Barigozzi et al., 2025a).

Overall, the loading matrices estimated with or without truncation exhibit similar patterns (see Fig. 4); namely, the elements of $\check{\mathbf{\Lambda}}_1^{[2]}(\tau)$ are of the same sign across the eight countries, and the columns of $\check{\mathbf{\Lambda}}_2^{[2]}(\tau)$ exhibit clusterings based on the grouping of the macroeconomic indicators. At the same time, recalling that each column of $p_2^{-1/2}\check{\mathbf{\Lambda}}_2^{[2]}(\tau)$ has the unit ℓ_2-norm, truncation of extreme observations enables us to better recover the loadings of smaller magnitude, particularly among the indicators associated with industrial production (IDs starting with IP); see also Table D.1 in the Supplementary Material, which shows that the loading estimates from noTrunc are more extreme.

These results have implications for forecasting performance, as confirmed in the following forecasting exercise, which shows that the forecasts from Trunc, being driven by the business cycle factor, are marginally superior to those from noTrunc. Denoting by $N = 236$ the number of observations prior to 2022-01, we sequentially produce one-step-ahead forecasts for the two indicators, the growth rate of the industrial production manufacturing index (IPMN) and the difference in core consumer price inflation (HICPNEF), at $t \in \{N+1, \ldots, n\}$ (recall that $n = 257$), each time using all preceding observations as the training data, see the Supplementary Material for complete details. Table 2 reports a summary of the forecasting errors for each indicator, where Trunc marginally outperforms noTrunc for both indicators across all metrics. As expected from Fig. 4, where the loading estimates differ the most for France and Italy, these countries contribute most to the difference between the two estimators' forecasting performance.

ACKNOWLEDGEMENT

Barigozzi's research was funded by the Italian Ministry of Education and Research (PRIN2022, 2022H2STF2 001) and Cho gratefully acknowledges the support from the Engineering and Physical Sciences Research Council (EP/Z531327/1).

SUPPLEMENTARY MATERIAL

The Supplementary Material includes additional information on asymptotic identifiability, complete simulation results and proofs. An implementation of the proposed methods is available at `https://github.com/haeran-cho/robustTFM`.

REFERENCES

AHN, S. C. & HORENSTEIN, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–27.

ALESSI, L., BARIGOZZI, M. & CAPASSO, M. (2010). Improved penalization for determining the number of factors in approximate static factor models. *Statist. Prob. Lett.* **80**, 1806–13.

ALONSO, A. M., GALEANO, P. & PEÑA, D. (2020). A robust procedure to build dynamic factor models with cluster structure. *J. Economet.* **216**, 35–52.

AMEMIYA, Y., FULLER, W. A. & PANTULA, S. G. (1987). The asymptotic distriautions of some estimators for a factor analysis model. *J. Mult. Anal.* **22**, 51–64.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Hoboken, NJ: John Wiley and Sons.

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

Bai, J. & Li, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40**, 436–65.

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

Bai, J. & Ng, S. (2013). Principal components estimation and identification of static factors. *J. Economet*. **176**, 18–29.

Baragona, R. & Battaglia, F. (2007). Outliers in dynamic factor models. *Electron. J. Statist.* **1**, 392–432.

Barigozzi, M., Lissona, C. & Tonni, L. (2025a). Large datasets for the Euro Area and its member countries and the dynamic effects of the common monetary policy. *arXiv:* 2410.05082v2.

Barigozzi, M., He, Y., Li, L. & Trapani, L. (2025b). Statistical inference for large-dimensional tensor factor models by iterative projections. *arXiv:* 2206.09800v3.

Barigozzi, M., He, Y., Li, L. & Trapani, L. (2023). Robust tensor factor analysis. *arXiv:* 2303.18163v2.

Chang, J., He, J., Yang, L. & Yao, Q. (2023). Modelling matrix time series via a tensor CP-decomposition. *J. R. Statist. Soc. B* **85**, 127–48.

Chen, E. Y. & Fan, J. (2023). Statistical inference for high-dimensional matrix-variate factor models. *J. Am. Statist. Assoc*. **118**, 1038–55.

Chen, L., Dolado, J. J. & Gonzalo, J. (2021). Quantile factor models. *Econometrica* **89**, 875–910.

Chen, R., Yang, D. & Zhang, C.-H. (2022). Factor models for high-dimensional tensor time series. *J. Am. Statist. Assoc*. **117**, 94–116.

Chen, W. & Lam, C. (2024). Rank and factor loadings estimation in time series tensor factor model by pre-averaging. *Ann. Statist.* **52**, 364–91.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quant. Finan*. **1**, 223–36.

De Lathauwer, L., De Moor, B. & Vandewalle, J. (2000a). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl*. **21**, 1253–78.

De Lathauwer, L., De Moor, B. & Vandewalle, J. (2000b). On the best rank-1 and rank-(r_1, r_2, …, r_n) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl*. **21**, 1324–42.

Doukhan, P. (1994). *Mixing: Properties and Examples*. New York: Springer.

Eklund, A., Nichols, T. E. & Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Nat. Acad. Sci*. **113**, 7900–5.

Fan, J., Guo, J. & Zheng, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *J. Am. Statist. Assoc*. **117**, 852–61.

Fan, J., Liu, H. & Wang, W. (2018). Large covariance estimation through elliptical factor models. *Ann. Statist*. **46**, 1383.

Fan, J., Wang, W. & Zhu, Z. (2021). A shrinkage principle for heavy-tailed data: high-dimensional robust low-rank matrix recovery. *Ann. Statist*. **49**, 1239.

Han, F. & Liu, H. (2018). ECA: high-dimensional elliptical component analysis in non-Gaussian distributions. *J. Am. Statist. Assoc*. **113**, 252–68.

Han, Y., Chen, R., Yang, D. & Zhang, C.-H. (2024). Tensor factor model estimation by iterative projection. *arXiv:* 2006.02611v3.

Han, Y., Chen, R. & Zhang, C.-H. (2022). Rank determination in tensor factor model. *Electron. J. Statist.* **16**, 1726–803.

Han, Y. & Zhang, C.-H. (2022). Tensor principal component analysis in high dimensional CP models. *IEEE Trans. Info. Theory* **69**, 1147–67.

He, Y., Kong, X., Yu, L. & Zhang, X. (2022). Large-dimensional factor analysis without moment constraints. *J. Bus. Econ. Statist*. **40**, 302–12.

He, Y., Kong, X., Yu, L., Zhang, X. & Zhao, C. (2024). Matrix factor analysis: From least squares to iterative projection. *J. Bus. Econ. Statist*. **42**, 322–34.

He, Y., Kong, X.-B., Liu, D. & Zhao, R. (2023a). Robust statistical inference for large-dimensional matrix-valued time series via iterative Huber regression. *arXiv:* 2306.03317.

He, Y., Li, L., Liu, D. & Zhou, W.-X. (2025a). Huber principal component analysis for large-dimensional factor models. *J. Economet*. **249**, 105993.

He, Y., Wang, Y., Yu, L., Zhou, W. & Zhou, W.-X. (2025b). A new non-parametric kendall's tau for matrix-valued elliptical observations. *Bernoulli* **31**, 3331–55.

He, Y., Xin-Bing, K., Long, Y. & Peng, Z. (2023b). Quantile factor analysis for large-dimensional time series with statistical guarantee. *arXiv:* 2006.08214v4.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73–101.

Ibragimov, M., Ibragimov, R. & Walden, J. (2015). *Heavy-tailed Distributions and Robustness in Economics and Finance*. Cham: Springer.

Ke, Y., Minsker, S., Ren, Z., Sun, Q. & Zhou, W.-X. (2019). User-friendly covariance estimation for heavy-tailed distributions. *Statist. Sci.* **34**, 454–71.

Kolda, T. G. & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51**, 455–500.

KRISTENSEN, J. T. (2014). Factor-based forecasting in the presence of outliers: are factors better selected and estimated by the median than by the mean? *Studies Nonlin. Dyn. Econom*. **18**, 309–38.
LUO, Y., RASKUTTI, G., YUAN, M. & ZHANG, A. R. (2021). A sharp blockwise tensor perturbation bound for orthogonal iteration. *J. Mach. Learn. Res*. **22**, 1–48.
LUO, Y. & ZHANG, A. R. (2023). Low-rank tensor estimation via Riemannian Gauss-Newton: statistical optimality and second-order convergence. *J. Mach. Learn. Res*. **24**, 1–48.
MARDIA, K. V., KENT, J. T. & TAYLOR, C. C. (1979). *Multivariate Analysis*. New York: John Wiley and Sons.
MARONNA, R. A. & YOHAI, V. J. (2008). Robust low-rank approximation of data matrices with elementwise contamination. *Technometrics* **50**, 295–304.
MCCRACKEN, M. W. & NG, S. (2016). FRED-MD: a monthly database for macroeconomic research. *J. Bus. Econ. Statist*. **34**, 574–89.
MIKOSCH, T. & RAČKAUSKAS, A. (2010). The limit distribution of the maximum increment of a random walk with regularly varying jump size distribution. *Bernoulli* **16**, 1016–38.
ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Statist*. **92**, 1004–16.
ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Economet*. **168**, 244–58.
PEÑA, D. & YOHAI, V. J. (2016). Generalized dynamic principal components. *J. Am. Statist. Assoc*. **111**, 1121–31.
PURDOM, E. & HOLMES, S. P. (2005). Error distribution for gene expression data. *Statist. Appl. Genet. Mol*. **4**, 16.
QIU, J., LI, Z. & YAO, J. (2025). Robust estimation for number of factors in high dimensional factor modeling via spearman correlation matrix. *J. Am. Statist. Assoc*. **120**, 1139–51.
RAYMAEKERS, J. & ROUSSEEUW, P. J. (2024). Challenges of cellwise outliers. *Economet. Statist*., doi: 10.1016/j.ecosta.2024.02.002
SHE, Y., LI, S. & WU, D. (2016). Robust orthogonal complement principal component analysis. *J. Am. Statist. Assoc*. **111**, 763–71.
STOCK, J. H. & WATSON, M. W. (2002a). Forecasting using principal components from a large number of predictors. *J. Am. Statist. Assoc*. **97**, 1167–79.
STOCK, J. H. & WATSON, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *J. Bus. Econ. Statist*. **20**, 147–62.
TRUCÍOS, C., MAZZEU, J. H., HOTTA, L. K., PEREIRA, P. L. V. & HALLIN, M. (2021). Robustness and the general dynamic factor model with infinite-dimensional space: identification, estimation, and forecasting. *Int. J. Forecast*. **37**, 1520–34.
TZOURIO-MAZOYER, N., LANDEAU, B., PAPATHANASSIOU, D., CRIVELLO, F., ETARD, O., DELCROIX, N., MAZOYER, B. & JOLIOT, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–89.
WANG, D., LIU, X. & CHEN, R. (2019). Factor models for matrix-valued high-dimensional time series. *J. Economet*. **208**, 231–48.
WANG, D. & TSAY, R. S. (2023). Rate-optimal robust estimation of high-dimensional vector autoregressive models. *Ann. Statist.* **51**, 846–77.
WANG, Y., ZHU, Y., SUN, Q. & QIN, L. (2023). Adaptively robust high-dimensional matrix factor analysis under Huber loss function. *J. Statist. Plan. Infer*. **231**, 106137.
YU, L., HE, Y., KONG, X. & ZHANG, X. (2022). Projected estimation for large-dimensional matrix factor models. *J. Economet*. **229**, 201–17.
ZHANG, D. (2021). Robust estimation of the mean and covariance matrix for high dimensional time series. *Statist. Sinica* **31**, 797–820.
ZHANG, X., LI, G. & LIU, C. C. (2024). Tucker tensor factor models for high-dimensional higher-order tensor observations. *arXiv:* 2206.02508v4.

[*Received on* 25 *February* 2025. *Accepted on* 4 *December* 2025]