# Robust estimation for number of factors in high dimensional factor modeling via Spearman correlation matrix

Jiaxin Qiu
Department of Statistics and Actuarial Science
The University of Hong Kong, Hong Kong, China
qiujx@connect.hku.hk
and
Zeng Li
Department of Statistics and Data Science
Southern University of Science and Technology, Shenzhen, China
liz9@sustech.edu.cn
and
Jianfeng Yao
School of Data Science
The Chinese University of Hong Kong (Shenzhen), Shenzhen, China
jeffyao@cuhk.edu.cn

**Abstract**

Determining the number of factors in high-dimensional factor modeling is essential but challenging, especially when the data are heavy-tailed. In this paper, we introduce a new estimator based on the spectral properties of Spearman sample correlation matrix under the high-dimensional setting, where both dimension and sample size tend to infinity proportionally. Our estimator is robust against heavy tails in either the common factors or idiosyncratic errors. The consistency of our estimator is established under mild conditions. Numerical experiments demonstrate the superiority of our estimator compared to existing methods.

*Keywords:* Factor model; Heavy tails; High dimensionality; Spearman correlation matrix; Spiked eigenvalue.

# 1 Introduction

Factor models are helpful tools for understanding the common dependence among high-dimensional outputs. They are widely used in data analysis in various areas like finance, genomics, and economics. Estimating the total number of factors is one of the most fundamental challenges when applying factor models in practice. This paper focuses on the following factor model:

$$\boldsymbol{y}_i = \boldsymbol{B}\boldsymbol{f}_i + \boldsymbol{\Psi}\boldsymbol{e}_i, \qquad i \in [n] := \{1, 2, \ldots, n\}, \tag{1}$$

where $\{\boldsymbol{y}_i\}$ are the $p$-dimensional observation vectors, $\{\boldsymbol{f}_i\}$ the $K$-dimensional latent common factor vectors, $\{\boldsymbol{e}_i\}$ the $p$-dimensional idiosyncratic error vectors, $\boldsymbol{B}$ the $p \times K$ factor loading matrix, and $\boldsymbol{\Psi}$ a $p \times p$ diagonal matrix. The objective of this paper is to estimate the number of common factors when the observed data is heavy-tailed.

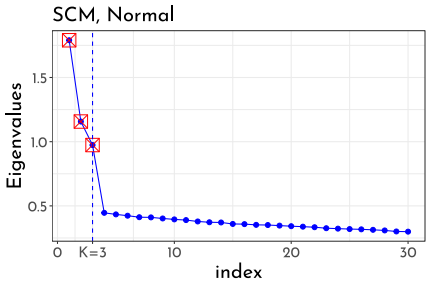
There is a large literature on this estimation problem which can generally be categorized into two types of approaches. The first type is based on information criterion. The seminal work Bai and Ng (2002) proposed several information criteria, which were formulated in many different forms, through modifications of the *Akaike information criterion* (AIC) and the *Bayesian information criterion* (BIC). Hallin and Liška (2007) proposed an information criterion that utilized spectral density matrix estimation. Alessi et al. (2010) modified Bai and Ng (2002)'s criteria by tuning the penalty function to enhance their performance. Kong (2017) employed similar ideas and put forth a local *principal component analysis* (PCA) approach to study a continuous-time factor model with time-varying factor loadings using high-frequency data. Li et al. (2017a) used information criteria akin to those proposed by Bai and Ng (2002) to factor models when the number of factors in-

creases with the cross-section size and time period. The first type of approaches usually require strong signals. The second type of approaches is based on eigenvalue behavior of various types of covariance/correlation matrices. As for sample covariance matrices, Ahn and Horenstein (2013) proposed two estimators by utilizing the ratios of adjacent eigenvalues, namely the *Eigenvalue Ratio* (ER) estimator and the *Growth Ratio* (GR) estimator. Onatski (2010) proposed an alternative *Edge Distribution* (ED) estimator based on the maximum differences between consecutive eigenvalues instead of their ratios. For lagged sample autocovariance matrices, Lam and Yao (2012) developed a ratio-based estimator for factor modeling of multivariate time series. This estimator was further extended by Li et al. (2017b) to accommodate weak factors. As for correlation matrices, Fan et al. (2020) proposed a tuning-free and scale-invariant adjusted correlation thresholding method. This approach has been further extended to time series tensor factor model in Lam (2021) and Chen and Lam (2022).

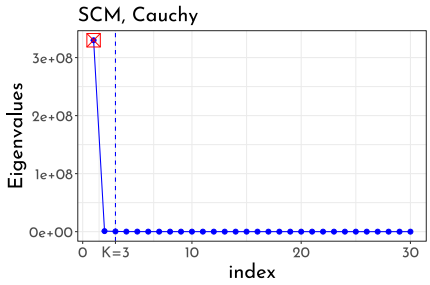
The aforementioned methods have been proved to be inadequate when dealing with heavy-tailed data, and mostly would result in biased or inconsistent estimators. Heavy-tailed data are common in various real-world applications. For instance, prices of stock returns often exhibit heavy tails due to the occurrence of extreme events in the market. However, little literature has focused on estimating number of factors in the context of heavy-tailed data. Assuming a jointly elliptical distribution for both common factors and idiosyncratic errors (as discussed in Fan et al. (2018)), Yu et al. (2019) proposed two estimators utilizing the sample multivariate Kendall's tau matrix. He et al. (2022b) further extended it to the matrix factor model. He et al. (2022a) recovered factor loadings and scores by performing PCA to the multivariate Kendall's tau matrix. It is worth mentioning that Yu et al. (2019)'s method requires that the loading matrix $\boldsymbol{B}$ has a spectral norm of

order $O(n)$, and hence can only detect strong factors. While, in contrast, we consider the more challenging case where $\boldsymbol{B}$ has bounded spectral norm and factors are too weak to be detected by existing methods. We propose an estimator based on *Spearman correlation matrix* (Spearman (1961)) which shows significant improvements over existing methods. Here, we use a toy example to demonstrate the robustness of Spearman correlation matrix. Data are generated following factor model (1) with $K = 3$. The factors and idiosyncratic errors follow either standard normal distribution or standard Cauchy distribution. As shown in Figure 1, when the common factors and the idiosyncratic noise are light-tailed, all four sample covariance/correlation matrices have three spiked eigenvalues, and all factors can be detected. When the data distribution is heavy-tailed, only our method can clearly identify all three factors.

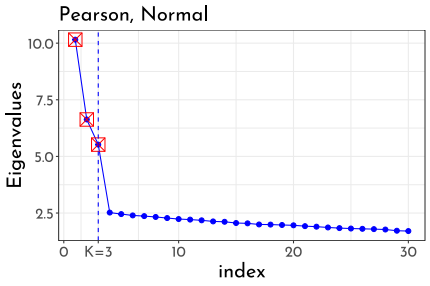
Spearman correlation matrix is defined as the Pearson correlation matrix of the ranks of the data. It is a valuable tool when dealing with heavy-tailed data. However, the nonlinear structure of rank-based correlation brings significant difficulties when analyzing its eigenvalue behavior. We need to resort to tools in *random matrix theory* (RMT). Unfortunately, most existing work in RMT focus on very restrictive settings when data has independent components. Bai and Zhou (2008) showed that its *limiting spectral distribution* (LSD) is the well-known Marčenko-Pastur law. Bao et al. (2015) established the *central limiting theorem* (CLT) for its *linear spectral statistics* (LSS). Bao (2019) showed that the Tracy-Widom law holds for its largest eigenvalues. To the best of our knowledge, the first investigation of Spearman sample correlation matrix for general dependent data was conducted very recently by Wu and Wang (2022), which derived its LSD under the non-paranormal distribution proposed by Liu et al. (2009). Many other spectral properties, including the extreme eigenvalues, CLT for LSS, and spiked eigenvalues for dependent data,



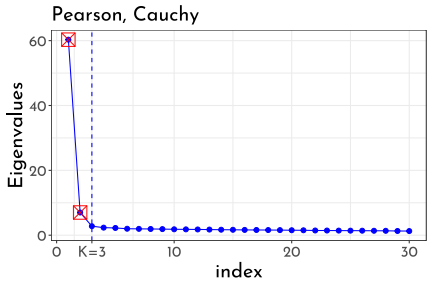

(a) Sample covariance matrix



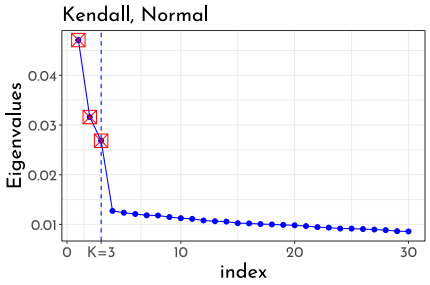

(b) Sample covariance matrix



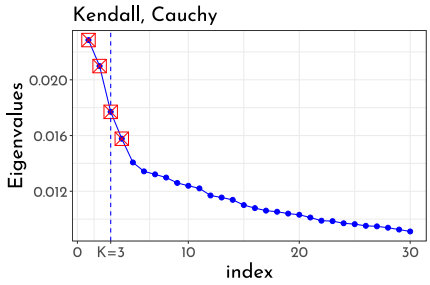

(c) Pearson correlation matrix



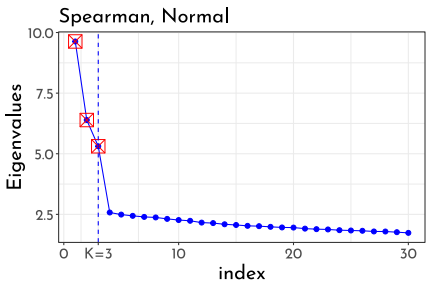

(d) Pearson correlation matrix



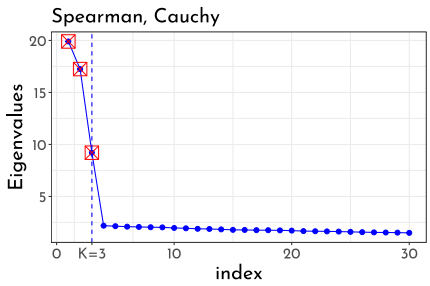

(e) Multivariate Kendall's tau matrix




(f) Multivariate Kendall's tau matrix




(g) Spearman correlation matrix




(h) Spearman correlation matrix

Figure 1: Scatter plots of the first 30 eigenvalues of the sample covariance matrix (SCM), Pearson correlation matrix, Spearman correlation matrix, and multivariate Kendall's tau matrix. Data are generated following the factor model (1) with $K = 3$. The factors and idiosyncratic errors are drawn independently from standard **Normal** distribution (**Left panel**: Figures (a), (c), (e), (g)) or standard **Cauchy** distribution (**Right panel**: Figures (b), (d), (f), (h)). The symbol "⊠" represents the spiked eigenvalues. Further details regarding the matrices $\boldsymbol{B}$ and $\boldsymbol{\Psi}$ can be found in the case (C1) in Section 3.

still remain open. Our work is the first to investigate the eigenvalue behavior of Spearman sample correlation matrix under spike models, and successfully applies the theories to identify weak factors in high-dimensional factor modeling for heavy-tailed data.

To summarize, the main contributions of this paper are two-fold. First, we propose a new estimator based on Spearman sample correlation matrix for the number of common factors in the high-dimensional factor model (1). This estimator is distribution-free and capable of detecting weak factors even when the data is heavy-tailed. Second, we provide a theoretical explanation of phase-transition phenomenon of top eigenvalues of Spearman sample correlation matrix under spike model. From a technical point of view, we investigate this phase-transition theory by establishing the universality of the asymptotic law of a low-dimensional random matrix (see Lemma 6.2 and Remark 2.6 for more details), and our method does not require the commonly used independent component structure.

Before moving forward, let us introduce some notations that will be used throughout this paper. We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. We adopt the convention of using regular letters for scalars, and bold-face letters for vectors or matrices. For any matrix $\boldsymbol{A}$, we denote its $(i, j)$-th entry by $A_{ij}$, its transpose by $\boldsymbol{A}^\intercal$, its trace by $\mathrm{tr}(\boldsymbol{A})$, its $j$-th largest eigenvalue by $\lambda_j(\boldsymbol{A})$ (when the eigenvalues of $\boldsymbol{A}$ are real), its spectral norm by $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_1(\boldsymbol{A}\boldsymbol{A}^\intercal)}$, and its element-wise maximum norm by $\|\boldsymbol{A}\|_{\max} = \max_{i,j} |A_{ij}|$. We use $\mathrm{diag}(\boldsymbol{A})$ to denote the diagonal matrix of $\boldsymbol{A}$ (replacing all off-diagonal entries with zero). For a sequence of random variables $\{X_n\}_{n=1}^\infty$ and a corresponding set of nonnegative real numbers $\{a_n\}_{n=1}^\infty$, we write $X_n = O_P(a_n)$ if $X_n/a_n = O_P(1)$ (bounded in probability), and we write $X_n = o_P(a_n)$ if $X_n/a_n \to 0$ in probability. For any univariate function $f$, we denote $f(\boldsymbol{A}) = [f(A_{ij})]$ as a matrix with $f$ applied on each entry of $\boldsymbol{A}$. Throughout this paper, $C$ stands for some positive constant whose value is not important and may change

from line to line. The notation "$i_1 \neq i_2 \neq \cdots \neq i_m$" indicates that the $m$ indices $\{i_\ell\}_{\ell=1}^m$ are pairwise different. All limits are for $n \to \infty$, unless explicitly stated otherwise.

The rest of this article is organized as follows. Section 2 proposes a new estimator of the number of common factors in the factor model (1). The consistency of our estimator is proved based on spectral properties of Spearman sample correlation matrix. Section 3 offers comprehensive simulation experiments, comparing our estimator with others. In Section 4, we evaluate the performance of the proposed estimator on a real dataset. A brief discussion is given in Section 5. Section 6 presents some important technical lemmas and proofs. Auxiliary lemmas and technical proofs are relegated to the supplementary material.

# 2 Main results

## 2.1 Spearman correlation matrix

For $p$-dimensional i.i.d. data sample $\{\boldsymbol{y}_i\}_{i=1}^n$, we denote the ranks of the data as follows:

$$\boldsymbol{Y}_n = \begin{pmatrix} \boldsymbol{y}_1^\intercal \\ \vdots \\ \boldsymbol{y}_n^\intercal \end{pmatrix} = \underbrace{\begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix}}_{\text{raw data matrix}} \Longrightarrow \underbrace{\begin{pmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{np} \end{pmatrix}}_{\text{ranks matrix}},$$

where $r_{ij} = \sum_{\ell=1}^n \mathbb{1}\{y_{\ell j} \leqslant y_{ij}\}$ is the rank of $y_{ij}$ among $\{y_{\ell j}\}_{\ell=1}^n$, and $\mathbb{1}\{\cdot\}$ denotes the indicator function. Spearman correlation matrix of the raw data matrix $\boldsymbol{Y}_n$ is the Pearson correlation matrix of the ranks matrix. Define the normalized ranks matrix

$$\boldsymbol{R} = \sqrt{\frac{12}{n^2-1}}\Big(r_{ij} - \frac{n+1}{2}\Big)_{n\times p}, \tag{2}$$

and let $\boldsymbol{R}_i^\intercal$ be the $i$-th row of the matrix $\boldsymbol{R}$. The Spearman sample correlation matrix of $\boldsymbol{Y}_n$ is

$$\boldsymbol{\rho}_n = \frac{1}{n}\boldsymbol{R}^\intercal\boldsymbol{R} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{R}_i\boldsymbol{R}_i^\intercal. \tag{3}$$

The *empirical spectral distribution* (ESD) of $\boldsymbol{\rho}_n$ is referred to as a random measure $F^{\boldsymbol{\rho}_n} = p^{-1}\sum_{j=1}^{p}\delta_{\lambda_j(\boldsymbol{\rho}_n)}$, where $\delta_{\lambda_j(\boldsymbol{\rho}_n)}$ is the Dirac mass at the point $\lambda_j(\boldsymbol{\rho}_n)$. The limit of $F^{\boldsymbol{\rho}_n}$ is called *limiting spectral distribution* (LSD). Under the assumption that the components of $\boldsymbol{y}_i$ are i.i.d., Bai and Zhou (2008) proved that the LSD of $\boldsymbol{\rho}_n$ is the well-known Marčenko-Pastur distribution. Recently, Wu and Wang (2022) extended this result to the non-paranormal distribution. In this study, we further extend their findings to encompass the scale mixture of normal distributions (see Definition 2.1), as stated in Lemma 6.1.

Throughout this paper, we assume that both common factors and idiosyncratic errors follow continuous distributions. Therefore, with probability one, there are no ties among $\{y_{ij}, i \in [n]\}$ for each $j$. For any $j \in [p]$ and $i, \ell \in [n]$ with $i \neq \ell$, we have $\mathbb{1}\{y_{\ell j} \leqslant y_{ij}\} = \frac{1}{2} + \frac{1}{2}\mathrm{sign}(y_{ij} - y_{\ell j})$, where $\mathrm{sign}(\cdot)$ denotes the sign function. Hence, we have

$$r_{ij} - \frac{n+1}{2} = 1 + \frac{1}{2}\sum_{i\neq\ell}\big\{1 + \mathrm{sign}(y_{ij} - y_{\ell j})\big\} - \frac{n+1}{2} = \frac{1}{2}\sum_{i\neq\ell}\mathrm{sign}(y_{ij} - y_{\ell j}). \tag{4}$$

For two sample vectors $\boldsymbol{y}_i$ and $\boldsymbol{y}_\ell$, we define the sign vector

$$\boldsymbol{A}_{i\ell} = \mathrm{sign}(\boldsymbol{y}_i - \boldsymbol{y}_\ell) = \big(\mathrm{sign}(y_{i1} - y_{\ell 1}), \ldots, \mathrm{sign}(y_{ip} - y_{\ell p})\big)^\intercal.$$

Then, from (2) and (4), we can rewrite the Spearman sample correlation matrix (3) as

$$\boldsymbol{\rho}_n = \frac{3}{n(n^2-1)}\sum_{i=1}^{n}\sum_{\ell_1,\ell_2\neq i}\boldsymbol{A}_{i\ell_1}\boldsymbol{A}_{i\ell_2}^\intercal.$$

The application of sign transformations to the data introduces an intractable nonlinear

correlation structure. To address this challenge, we utilize *Hoeffding's decomposition* (Hoeffding (1948)) to handle the nonlinear correlation within $\boldsymbol{A}_{i\ell}$. By employing this decomposition, we can identify the dominant term of $\boldsymbol{\rho}_n$. Let $\boldsymbol{A}_i := \mathbb{E}(\boldsymbol{A}_{i\ell} \mid \boldsymbol{y}_i)$, the Hoeffding's decomposition of $\boldsymbol{A}_{i\ell}$ can be expressed as follows:

$$\boldsymbol{A}_{i\ell} = \boldsymbol{A}_i - \boldsymbol{A}_\ell + \boldsymbol{\varepsilon}_{i\ell}, \tag{5}$$

where $\boldsymbol{\varepsilon}_{i\ell} := \boldsymbol{A}_{i\ell} - \boldsymbol{A}_i + \boldsymbol{A}_\ell$. Note that $\mathbb{E}\boldsymbol{A}_{i\ell} = \mathbb{E}\boldsymbol{A}_i = \boldsymbol{0}$, and the covariance matrix of $\boldsymbol{A}_i$ is $\mathbb{E}(\boldsymbol{A}_i\boldsymbol{A}_i^\intercal)$. With Hoeffding's decomposition defined in (5), we have

$$\begin{aligned}\boldsymbol{\rho}_n = \frac{3}{n(n^2-1)}\Bigg\{&\sum_{i=1}^n \sum_{\ell_1,\ell_2\neq i} (\boldsymbol{A}_i - \boldsymbol{A}_{\ell_1})(\boldsymbol{A}_i - \boldsymbol{A}_{\ell_2})^\intercal + \sum_{i=1}^n \sum_{\ell_1,\ell_2\neq i} (\boldsymbol{A}_i - \boldsymbol{A}_{\ell_1})\boldsymbol{\varepsilon}_{i\ell_2}^\intercal \\ &+ \sum_{i=1}^n \sum_{\ell_1,\ell_2\neq i} \boldsymbol{\varepsilon}_{i\ell_1}(\boldsymbol{A}_i - \boldsymbol{A}_{\ell_2})^\intercal + \sum_{i=1}^n \sum_{\ell_1,\ell_2\neq i} \boldsymbol{\varepsilon}_{i\ell_1}\boldsymbol{\varepsilon}_{i\ell_2}^\intercal \Bigg\}.\end{aligned}$$

It will be shown that the cross-terms in the above identity are negligible. We can then focus on the first term,

$$\begin{aligned}&\frac{3}{n(n^2-1)} \sum_{i=1}^n \sum_{\ell_1,\ell_2\neq i} (\boldsymbol{A}_i - \boldsymbol{A}_{\ell_1})(\boldsymbol{A}_i - \boldsymbol{A}_{\ell_2})^\intercal \\ &= \frac{n-2}{n+1} \cdot \frac{3}{n} \sum_{i=1}^n \Bigg\{ \frac{1}{(n-1)(n-2)} \sum_{\ell_1\neq\ell_2\neq i} (\boldsymbol{A}_i - \boldsymbol{A}_{\ell_1})(\boldsymbol{A}_i - \boldsymbol{A}_{\ell_2})^\intercal \Bigg\} + \frac{3}{n+1}\boldsymbol{\tau}_n,\end{aligned} \tag{6}$$

where $\boldsymbol{\tau}_n := \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\ell\neq i} (\boldsymbol{A}_i - \boldsymbol{A}_\ell)(\boldsymbol{A}_i - \boldsymbol{A}_\ell)^\intercal$ is the sample marginal Kendall's tau correlation matrix (Bandeira et al. (2017), Li et al. (2022)). The second term, $3\boldsymbol{\tau}_n/(n+1)$, and the cross-terms in (6) are negligible. Hence, the leading order term of $\boldsymbol{\rho}_n$ is

$$\boldsymbol{W}_n = \frac{3}{n} \sum_{i=1}^n \boldsymbol{A}_i\boldsymbol{A}_i^\intercal. \tag{7}$$

Under certain assumptions, we can show that the spectrum of $\boldsymbol{\rho}_n$ can be approximated

by that of $\boldsymbol{W}_n$ (see Lemma 2.2). This allows us to study the spectral properties of $\boldsymbol{\rho}_n$ via those of $\boldsymbol{W}_n$. Under the scale mixture of normals framework defined in Section 2.2, we found that the population covariance matrix $\boldsymbol{\Sigma}_\rho := \mathbb{E}\boldsymbol{W}_n$ has a finite-rank perturbation structure and it has $K$ spiked eigenvalues (see Lemma 2.3). Thus naturally $\boldsymbol{W}_n$ has $K$ relatively large eigenvalues too. Subsequently, we can estimate the number of factors based on the top eigenvalues of $\boldsymbol{W}_n$ or $\boldsymbol{\rho}_n$. By establishing the phase-transition theory of spiked eigenvalues of $\boldsymbol{W}_n$, the consistency of the new estimator follows.

## 2.2 Phase transition theory

From the perspective of RMT, the spectrum of $\boldsymbol{W}_n$ relies on the structure of $\boldsymbol{\Sigma}_\rho = \mathbb{E}\boldsymbol{W}_n = \mathbb{E}\boldsymbol{\rho}_n$. Although from factor model (1), it is clear that $\boldsymbol{\Sigma}_y := \mathrm{Cov}(\boldsymbol{y}_i) = \boldsymbol{B}\boldsymbol{B}^\intercal + \boldsymbol{\Psi}$ has finite-rank-$K$ perturbation structure, the relationship between $\boldsymbol{\Sigma}_\rho$ and $\boldsymbol{\Sigma}_y$ is unclear. The structure of $\boldsymbol{\Sigma}_\rho$ changes for different distributions of $\boldsymbol{y}_i$. Therefore, extra distribution assumption of $\boldsymbol{y}_i$ is needed to maintain the finite-rank perturbation structure of $\boldsymbol{\Sigma}_\rho$. Specifically, we assume that both the common factors and the idiosyncratic errors follow a *scale mixture of normal distributions*, defined as follows:

**Definition 2.1** (Scale mixture of normals, Andrews and Mallows (1974))**.** A $p$-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_p)^\intercal$ follows a *scale mixture of normal distributions* (SMN) if $\boldsymbol{X}$ has the stochastic representation:

$$\boldsymbol{X} \overset{\mathsf{d}}{=} \sqrt{W}\boldsymbol{Z}, \tag{8}$$

where $W$ is a scalar-valued random variable with positive support, and $\boldsymbol{Z}$ follows $p$-dimensional normal distribution $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ independent of $W$, where $\boldsymbol{\Sigma}$ is a positive semi-definite matrix. The notation "$X \overset{\mathsf{d}}{=} Y$" in (8) means $X$ and $Y$ have the same distribution.

Our motivation for using this scale mixture of normals is two-fold. First, the scale mix-

ture of normals contains heavy-tailed distributions, such as Student's $t$ distribution. If $W$ follows the inverse Gamma distribution $\mathsf{invGamma}(\nu/2, \nu/2)$ with probability density function $g_W(w) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} w^{-(\nu/2+1)} \exp\{-\nu/(2w)\}$, then $\boldsymbol{X}$ follows the $p$-dimentional Student's $t$ distribution $t_\nu(\mathbf{0}, \boldsymbol{\Sigma})$ with location parameter $\mathbf{0}$, scale matrix $\boldsymbol{\Sigma}$, and degrees of freedom $\nu$, the probability density function of which is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\Gamma\big((\nu+p)/2\big)}{\Gamma(\nu/2)\nu^{p/2}\pi^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \left(1 + \frac{1}{\nu}\boldsymbol{x}^\intercal \boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right)^{-(\nu+p)/2}.$$

A number of well-known distributions can be written as scale mixtures of normals. We refer the readers to Section 2 of Heinen and Valdesogo (2020) for more examples. The second motivation is for technical advantage. From the fact that $\boldsymbol{X} \mid (W = w) \sim \mathcal{N}_p(\mathbf{0}, w\boldsymbol{\Sigma})$, we can relate the Spearman sample correlation matrix of $\boldsymbol{X}$ to the scale matrix $\boldsymbol{\Sigma}$ using Grothendieck's identity (see Lemma S1.5). Heinen and Valdesogo (2020) derived an explicit expression for Spearman correlation of bivariate scale mixture of normals. We extend this result to a more complicated bivariate population (see Lemma S1.6) and utilize it to examine the structure of $\boldsymbol{\Sigma}_\rho$ (see Lemma 2.3). This direct connection to the scale matrix $\boldsymbol{\Sigma}_\rho$ is a fundamental step in the analysis of our proposed estimator.

Furthermore, we need the following assumptions:

**Assumption (A1).** *As $n \to \infty$, $p = p(n) \to \infty$ and $p/n = c_n \to c \in (0, \infty)$.*

**Assumption (A2).** *All pairs $\{(\boldsymbol{f}_i^\intercal, \boldsymbol{e}_i^\intercal)^\intercal\}_{i=1}^n$ are i.i.d., and $\boldsymbol{f}_i$ is independent of $\boldsymbol{e}_i$, and both of them follow the scale mixture of normal distributions. Suppose that $w_f$ and $w_e$ are two independent random variables with positive support. The common factor $\boldsymbol{f}_i$ has a stochastic representation $\boldsymbol{f}_i \overset{\mathsf{d}}{=} \sqrt{w_i^f}\boldsymbol{x}_i$, where $w_i^f$ is an independent copy of $w_f$, and $\boldsymbol{x}_i$ follows $K$-dimensional standard normal distribution. The idiosyncratic error $\boldsymbol{e}_i$ has a stochastic representation $\boldsymbol{e}_i \overset{\mathsf{d}}{=} (\sqrt{w_{i1}^e}z_{i1}, \ldots, \sqrt{w_{ip}^e}z_{ip})^\intercal$, where $\{w_{ij}^e\}_{j=1}^p$ are i.i.d. copies of $w_e$, and $(z_{i1}, \ldots, z_{ip})^\intercal$ follows $p$-dimensional standard normal distribution.*

**Assumption (A3).** *The loading matrix* $\boldsymbol{B}$ *is normalized by the constraint* $\boldsymbol{B}^\intercal\boldsymbol{B} = \boldsymbol{I}_K$. *All the entries of* $\boldsymbol{B}$ *are of order* $O(1/\sqrt{p})$.

**Assumption (A4).** *The matrix* $\boldsymbol{\Psi}$ *is diagonal with entries of order* $O(1)$.

Assumption (A1) is common in the RMT literature. Assumption (A2) pertains to the distribution of the common factors and the idiosyncratic errors, and allows them to be heavy-tailed. This assumption is crucial for examining the structure of the population covariance matrix $\boldsymbol{\Sigma}_\rho$. The constraint $\boldsymbol{B}^\intercal\boldsymbol{B} = \boldsymbol{I}_K$ in (A3) is a commonly used identifiability condition, see Bai and Li (2012). It follows that by singular value decomposition, we can represent $\boldsymbol{B}$ as $\boldsymbol{U}\boldsymbol{V}^\intercal$, where $\boldsymbol{U} \in \mathbb{R}^{p\times K}$ and $\boldsymbol{V} \in \mathbb{R}^{K\times K}$, and both have orthonormal columns. The column vectors of $\boldsymbol{U}$ and $\boldsymbol{V}$ are unit vectors in $\mathbb{R}^p$ and $\mathbb{R}^K$, respectively. Thus, the condition $B_{ij} = O(1/\sqrt{p})$ in (A3), for any $i \in [p]$ and $j \in [K]$, is not overly restrictive. This condition facilitates our technical proofs. Assumption (A4) is standard in the factor models literature.

In what follows, we develop some important spectral properties of Spearman sample correlation matrix. First, we show that the spectrum of $\boldsymbol{\rho}_n$ can be approximated by that of $\boldsymbol{W}_n$, as stated in the following lemma.

**Lemma 2.2.** *Under Assumptions (A1) – (A4), for any* $j \in [p]$, $|\lambda_j(\boldsymbol{\rho}_n) - \lambda_j(\boldsymbol{W}_n)| = o_P(1)$ *as* $n \to \infty$.

The proof of this lemma is provided in the supplementary material. From this Lemma, we can investigate the properties of $\boldsymbol{\rho}_n$ through its surrogate $\boldsymbol{W}_n$. As $\boldsymbol{\Sigma}_\rho$ represents the expectation of $\boldsymbol{W}_n$, examining the structure of $\boldsymbol{\Sigma}_\rho$ provides us with valuable insights of $\boldsymbol{W}_n$. The spike structure of $\boldsymbol{\Sigma}_\rho$ is illustrated in the following lemma.

**Lemma 2.3** (Finite-rank perturbation)**.** *Under Assumptions (A1) – (A4), we have*

$$\left\|\boldsymbol{\Sigma}_\rho - \left\{\operatorname{diag}\left(\boldsymbol{I}_p - \gamma\boldsymbol{\Psi}^{-1}\boldsymbol{B}\boldsymbol{B}^\intercal\boldsymbol{\Psi}^{-1}\right) + \gamma\boldsymbol{\Psi}^{-1}\boldsymbol{B}\boldsymbol{B}^\intercal\boldsymbol{\Psi}^{-1}\right\}\right\|_2 = o(1) \tag{9}$$

*as* $n \to \infty$, *where* $\gamma = (6/\pi)\mathbb{E}\big[w^f/\{(w_1^e + w_2^e)^{1/2}(w_3^e + w_4^e)^{1/2}\}\big]$, *and* $w_j^e$, $j = 1, 2, 3, 4$, *are independent copies of* $w_e$.

**Remark 2.4.** *Note that both* $\boldsymbol{\Sigma}_\rho$ *and its approximation in* (9) *are correlation-type matrices, and all the diagonal entries equal to one. Consequently, the average of their eigenvalues are both one, and their bulk eigenvalues are clustered around one.*

The proof of this lemma is provided in the supplementary material. In this lemma, we derive a "consistent" approximation of the population covariance matrix $\boldsymbol{\Sigma}_\rho$. From Weyl's lemma (Lemma S1.1), the spectrum of the matrix $\boldsymbol{\Sigma}_\rho$ can be approximated by that of a rank-$K$ perturbation of a diagonal matrix. Intuitively, $\boldsymbol{\Sigma}_\rho$ would have at most $K$ relatively larger eigenvalues. As for the sample counterpart, at most $K$ spiked sample eigenvalues of $\boldsymbol{\rho}_n$ would lay outside the support of its LSD. Naturally, by counting the number of spiked eigenvalues of $\boldsymbol{\rho}_n$, we can obtain a promising estimator of total number of factors. However, a very important yet intuitive observation here is that, for $1 \leqslant j \leqslant K$, $\lambda_j(\boldsymbol{\rho}_n)$ is not always far away from the bulk eigenvalues $\{\lambda_j(\boldsymbol{\rho}_n), K+1 \leqslant j \leqslant p\}$. It depends on whether the signal $\lambda_j(\boldsymbol{\Sigma}_\rho)$ is strong enough. If $\lambda_j(\boldsymbol{\Sigma}_\rho)$ is too weak, $\lambda_j(\boldsymbol{\rho}_n)$ would lie on the boundary of the support of bulk eigenvalues. This phenomenon is commonly referred to as the *phase-transition phenomenon*, which is described in the following theorem.

**Theorem 2.5** (Phase transition)**.** *For the high-dimensional factor model* (1)*, assume that Assumptions (A1) – (A4) hold, and the ESD of* $\boldsymbol{\Sigma}_\rho$ *tends to a proper probability measure* $H$ *as* $n \to \infty$. *Denote* $\psi(\alpha) = \alpha + c\int \frac{t\alpha}{\alpha - t}\,\mathrm{d}H(t)$, *we have*

**(a)** *For* $1 \leqslant j \leqslant K$ *satisfying* $\psi'\big(\lambda_j(\boldsymbol{\Sigma}_\rho)\big) > 0$, *the* $j$*-th sample eigenvalue of* $\boldsymbol{\rho}_n$ *converges almost surely to* $\psi\big(\lambda_j(\boldsymbol{\Sigma}_\rho)\big)$, *which is outside the support of the LSD of* $\boldsymbol{\rho}_n$.

**(b)** *For* $1 \leqslant j \leqslant K$ *satisfying* $\psi'\big(\lambda_j(\boldsymbol{\Sigma}_\rho)\big) \leqslant 0$, *the* $j$*-th sample eigenvalue of* $\boldsymbol{\rho}_n$ *converges almost surely to the right endpoint of the support of the LSD of* $\boldsymbol{\rho}_n$.

The proof of this theorem is provided in Section 6.

**Remark 2.6.** *From Lemma 2.2, we can investigate the asymptotic behavior of spiked eigenvalues of* $\boldsymbol{\rho}_n$ *via those of* $\boldsymbol{W}_n$. *Although* $\boldsymbol{W}_n = (3/n)\sum_{i=1}^n \boldsymbol{A}_i\boldsymbol{A}_i^\intercal$ *is a Wishart-type random matrix, the nonlinear correlation structure of* $\boldsymbol{A}_i$ *makes it difficult to directly apply the current phase-transition analysis techniques. The reason is that the vectors* $\{\boldsymbol{A}_i\}_{i=1}^n$ *do not follow the commonly used* independent component structure *as in Bai and Yao (2008, 2012) and Jiang and Bai (2021). Specifically, the vector* $\boldsymbol{A}_i$ *cannot be written as* $\boldsymbol{A}_i = \boldsymbol{\Sigma}^{1/2}\boldsymbol{x}_i$, *where* $\boldsymbol{\Sigma}$ *is non-negative definite and all elements of* $\boldsymbol{x}_i \in \mathbb{R}^p$ *are i.i.d. with zero mean and unit variance. To remove the independent component structure assumption, we first show that replacing* $\{\sqrt{3}\boldsymbol{A}_i\}_{i=1}^n$ *in* $\boldsymbol{W}_n$ *with i.i.d.* $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_\rho)$ *random vectors does not change the asymptotic behavior of spiked eigenvalues of* $\boldsymbol{W}_n$ *(see Lemma 6.2 and Section 6.2 for more details). To guarantee the feasibility of this replacement, we establish concentration properties related to certain quadratic forms and their higher-order moments under the nonlinear correlation structure (see Lemma S1.7). Then since Gaussian random vectors naturally follows the independent component structure, we can directly apply the phase transition theory in Bai and Yao (2008, 2012) and Jiang and Bai (2021).*

## 2.3 Estimation of the number of factors

With the phase-transition theory in Theorem 2.5, we now propose our new estimator of number of factors. As stated in Theorem 2.5, if $\psi'\big(\lambda_j(\boldsymbol{\Sigma}_\rho)\big) \leqslant 0$ for some $j \in [K]$, the corresponding sample eigenvalue $\lambda_j(\boldsymbol{\rho}_n)$ will converge to the right endpoint of the support of the LSD of $\boldsymbol{\rho}_n$, which is also the limit of the largest noise eigenvalue $\lambda_{K+1}(\boldsymbol{\rho}_n)$. Hence, such weak factors will be merged into the noise component, making their signal undetectable. By taking this into account, we define the *number of significant factors* as

$$K_0 = \#\big\{j \in [K] : \psi'\big(\lambda_j(\boldsymbol{\Sigma}_\rho)\big) > 0\big\}, \tag{10}$$

where the notation $\#\mathcal{S}$ denotes the cardinality number of the set $\mathcal{S}$. By Theorem 2.5, the leading $K_0$ eigenvalues of $\boldsymbol{\rho}_n$ will lay outside the support of its LSD .

The LSD of $\boldsymbol{\rho}_n$, denoted by $F_{c,H}$, is the generalized Marčenko-Pastur law as stated in Lemma 6.1. Let $\mathsf{supp}(F_{c,H})$ denote the support of $F_{c,H}$. The Stieltjes transform of $F_{c,H}$ is defined as $m(x) = \int \frac{1}{t-x}\,\mathrm{d}F_{c,H}(t)$ for $x \in \mathbb{R} \setminus \mathsf{supp}(F_{c,H})$. Its first-order derivative $m'(x)$ is also only defined outside $\mathsf{supp}(F_{c,H})$, and can be extended as a function mapping the entire real line $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$ as follows:

$$m'(x) = \begin{cases} \int \frac{1}{(t-x)^2}\,\mathrm{d}F_{c,H}(t), & \text{if } x \in \mathbb{R} \setminus \mathsf{supp}(F_{c,H}), \\ +\infty, & \text{if } x \in \mathsf{supp}(F_{c,H}). \end{cases} \tag{11}$$

This implies that $m'(\lambda_j(\boldsymbol{\rho}_n))$ takes either finite or infinite values, depending on whether $\lambda_j(\boldsymbol{\rho}_n)$ is a spiked eigenvalue or a bulk eigenvalue. Based on this observation, we utilize the derivative of Stieltjes transform defined in (11) to identify all the spiked eigenvalues and estimate total number of significant factors. Let $K_{\max}$ be a predetermined upper bound on the true number of significant factors, $K_0$. As the LSD $H$ of $\boldsymbol{\Sigma}_\rho$ is unknown, we cannot obtain the explicit expression of $m'(x)$. Therefore, we utilize

$$\widehat{m}'_{n,j}(x) = \frac{1}{p-j} \sum_{\ell=j+1}^{p} \frac{1}{\{x - \lambda_\ell(\boldsymbol{\rho}_n)\}^2} \tag{12}$$

to estimate $m'(x)$ for $1 \leqslant j \leqslant K_{\max}$. Intuitively, if $\lambda_j(\boldsymbol{\rho}_n)$ lies within the bulk spectrum of $\boldsymbol{\rho}_n$, we would expect $\widehat{m}'_{n,j}\big(\lambda_j(\boldsymbol{\rho}_n)\big)$ to be very large. On the contrary, if $\lambda_j(\boldsymbol{\rho}_n)$ is a spiked eigenvalue, $\widehat{m}'_{n,j}\big(\lambda_j(\boldsymbol{\rho}_n)\big)$ should be relatively small. This phenomenon bears similarities to

the behavior of $m'(x)$ described in the equation (11). Actually, it will be shown that

$$\widehat{m}'_{n,j}\big(\lambda_j(\boldsymbol{\rho}_n)\big) = \begin{cases} O(1), & \text{for } 1 \leqslant j \leqslant K_0, \\ O(p), & \text{for } K_0+1 \leqslant j \leqslant K_{\max}, \end{cases}$$

as $n \to \infty$. Hence a natural estimator of the number of significant factors is

$$\widehat{K}_{\mathrm{SR}} = \underset{1 \leqslant j \leqslant K_{\max}}{\arg\max} \frac{\widehat{m}'_{n,j+1}\big(\lambda_{j+1}(\boldsymbol{\rho}_n)\big)}{\widehat{m}'_{n,j}\big(\lambda_j(\boldsymbol{\rho}_n)\big)}, \tag{13}$$

where the "S" in subscript stands for Stieltjes transform, and the "R" stands for Ratio. The consistency of this estimator is established in the following theorem, and its proof is postponed to Section 6.

**Theorem 2.7** (Consistency of $\widehat{K}_{\mathrm{SR}}$)**.** *For the high-dimensional factor model* (1)*, assume that Assumptions (A1) – (A4) hold. Let* $K_0$ *be the number of significant factors defined in* (10) *and* $\widehat{K}_{\mathrm{SR}}$ *be the proposed estimator defined in* (12) – (13)*. Then, we have*

$$\lim_{n\to\infty} \mathbb{P}(\widehat{K}_{\mathrm{SR}} = K_0) = 1.$$

# 3 Simulation studies

In this section, we conduct some simulations to examine the finite sample performance of the proposed estimator. We compare with several outstanding estimators in the current literature, including the ACT estimator (Fan et al. (2020)), and the MKER and MKTCR estimators (Yu et al. (2019)). Specifically, the competing estimators are defined as follows:

- ACT estimator (Fan et al. (2020)):

  Let $\{\boldsymbol{y}_i\}_{i=1}^n$ be an i.i.d. sample from the factor model (1). The sample covariance

matrix $\boldsymbol{S}_n$ and sample correlation matrix $\boldsymbol{P}_n$ from $\{\boldsymbol{y}_i\}_{i=1}^n$ are, respectively,

$$\boldsymbol{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^\intercal, \qquad \boldsymbol{P}_n = \big[\mathrm{diag}(\boldsymbol{S}_n)\big]^{-1/2}\boldsymbol{S}_n\big[\mathrm{diag}(\boldsymbol{S}_n)\big]^{-1/2},$$

where $\overline{\boldsymbol{y}} = n^{-1}\sum_{i=1}^n \boldsymbol{y}_i$ is the sample mean. Based on the spectral properties of $\boldsymbol{P}_n$, Fan et al. (2020) proposed an estimator to estimate the factor number as follows:

$$\widehat{K}_{\mathrm{ACT}} = \max\Big\{j : \widehat{\alpha}_j(\boldsymbol{P}_n) > 1 + \sqrt{p/(n-1)}\Big\}, \tag{14}$$

where $\{\widehat{\alpha}_j(\boldsymbol{P}_n)\}_{j=1}^p$ are bias correction of sample eigenvalues of $\boldsymbol{P}_n$, defined by $\widehat{\alpha}_j(\boldsymbol{P}_n) = -1/\underline{m}_{n,j}\big(\lambda_j(\boldsymbol{P}_n)\big)$ with $\underline{m}_{n,j}(x) = -(1-c_j)/x + c_j m_{n,j}(x)$, $c_j = (p-j)/(n-1)$, and

$$m_{n,j}(x) = \frac{1}{p-j}\left[\sum_{\ell=j+1}^{p}\frac{1}{\lambda_\ell(\boldsymbol{P}_n) - x} + \frac{1}{\big\{3\lambda_j(\boldsymbol{P}_n) + \lambda_{j+1}(\boldsymbol{P}_n)\big\}/4 - x}\right].$$

- MKER and MKTCR estimators (Yu et al. (2019)):

  The sample multivariate Kendall's tau matrix is defined by

$$\boldsymbol{K}_n = \frac{2}{n(n-1)}\sum_{1\leqslant i<\ell\leqslant n}\frac{(\boldsymbol{y}_i - \boldsymbol{y}_\ell)(\boldsymbol{y}_i - \boldsymbol{y}_\ell)^\intercal}{\|\boldsymbol{y}_i - \boldsymbol{y}_\ell\|^2}.$$

  Based on the eigenvalues $\{\lambda_j(\boldsymbol{K}_n)\}_{j=1}^p$, Yu et al. (2019) constructed the MKER estimator,

$$\widehat{K}_{\mathrm{MKER}} = \underset{1\leqslant j\leqslant K_{\max}}{\arg\max}\frac{\lambda_j(\boldsymbol{K}_n)}{\lambda_{j+1}(\boldsymbol{K}_n)},$$

  and the MKTCR estimator,

$$\widehat{K}_{\mathrm{MKTCR}} = \underset{1\leqslant j\leqslant K_{\max}}{\arg\max}\frac{\ln\{1+\lambda_j(\boldsymbol{K}_n)/V_{j-1}\}}{\ln\{1+\lambda_{j+1}(\boldsymbol{K}_n)/V_j\}},$$

  where $V_j = \sum_{i=j+1}^{\min(p,n)}\lambda_i(\boldsymbol{K}_n)$, $j \in \{0,1,\ldots,\min(p,n)-1\}$.

Our simulation studies consider various combinations of dimension and sample size, namely $(p, n) = (50, 100)$, $(100, 200)$, $(150, 300)$, and $(200, 400)$, which all have the ratio $p/n = 1/2$. We take the true number of common factors $K = 3$ and set the possible maximum value of the number of common factors $K_{\max} = 10$. Results are based on 1000 replicates and reported in the form $a(b \,|\, c)$, in which $a, b, c$ are the percentages (%) of true estimates, overestimates, and underestimates, respectively. The notation "ave($\widehat{K}$)" denotes mean estimators for the case of the largest dimensions $(p, n) = (200, 400)$.

Recalling our distribution assumption (A2) for both common factors $\boldsymbol{f}_i$ and idiosyncratic errors $\boldsymbol{e}_i$, we generate $\boldsymbol{f}_i$ and $\boldsymbol{e}_i$ by

$$\boldsymbol{f}_i = \sqrt{w_i^f}\boldsymbol{x}_i, \qquad e_{ij} = \sqrt{w_{ij}^e} z_{ij},$$

where $e_{ij}$ denotes the $j$-th component of the noise vector $\boldsymbol{e}_i$, $\{\boldsymbol{x}_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_K(\boldsymbol{0}, \boldsymbol{I}_K)$, and $\{z_{ij}, i \in [n], j \in [p]\} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. We employ four different scenarios to generate sample data for $w_i^f$ and $w_{ij}^e$:

1. (Normal population, see Table 1) Let $w_i^f = w_{ij}^e = 1$ for all $i \in [n]$ and $j \in [p]$;

2. (Uniform and Chi-squared population, see Table 1) Let $\{w_i^f\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathsf{Uniform}(0, 1)$ and $\{w_{ij}^e, i \in [n], j \in [p]\} \overset{\text{i.i.d.}}{\sim} \chi^2(1)$;

3. (Student's $t(2)$ population, see Table 2) Let $\{w_i^f\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathsf{invGamma}(1, 1)$ and $\{w_{ij}^e, i \in [n], j \in [p]\} \overset{\text{i.i.d.}}{\sim} \mathsf{invGamma}(1, 1)$, where $\mathsf{invGamma}(\alpha, \beta)$ denotes the inverse Gamma distribution with shape parameter $\alpha$ and scale parameter $\beta$. In this scenario, both $\boldsymbol{f}_i$ and $e_{ij}$ follow (multivariate) Student's $t(2)$ distributions;

4. (Cauchy population, see Table 2) Let $\{w_i^f\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathsf{invGamma}(1/2, 1/2)$ and $\{w_{ij}^e, i \in [n], j \in [p]\} \overset{\text{i.i.d.}}{\sim} \mathsf{invGamma}(1/2, 1/2)$. In this scenario, both $\boldsymbol{f}_i$ and $e_{ij}$ follow (multivariate) Cauchy distributions.

Furthermore, we consider three different cases for the loading matrix $\boldsymbol{B} = (B_{ij})_{p\times K}$ and the matrix $\boldsymbol{\Psi}$ as follows:

(C1) For any $j \in [K]$, let $B_{ij} = 1/\sqrt{p}$ for $i \in [K]$, and let $B_{ij} = a_{ij}/\sqrt{p-j}$ for $i \in \{K+1, \ldots, p\}$, where $a_{ij} = -1$ if $i = rj$ or $a_{ij} = 1$ if $i \neq rj$, $r \in \mathbb{N}^+$. Let $\Psi_{ii} = 0.4$ for $i \in [p]$.

(C2) For any $j \in [K]$, let $\sqrt{p}B_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $i \in [0.5p]$, and let $\sqrt{p}B_{ij} \overset{\text{i.i.d.}}{\sim} \mathsf{Uniform}(-1,1)$ for $i \in \{0.5p+1, \ldots, p\}$. Let $\{\Psi_{ii}\}_{i=1}^{p} \overset{\text{i.i.d.}}{\sim} \mathsf{Uniform}(0,1)$.

(C3) For any $1 \leqslant j \leqslant 2$, let $\sqrt{p}B_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,2)$ for $i \in [0.5p]$, and let $\sqrt{p}B_{ij} \overset{\text{i.i.d.}}{\sim} \mathsf{Uniform}(-1,1)$ for $i \in \{0.5p+1, \ldots, p\}$. For any $2 < j \leqslant K$, let $\sqrt{p}B_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $i \in [p]$. Let $\{\Psi_{ii}\}_{i=1}^{p} \overset{\text{i.i.d.}}{\sim} \mathsf{Uniform}(0,1)$.

The simulation results are reported in Tables 1 - 2 and Figures 2 - 5. In the case of light-tailed data (see Tables 1 and Figures 2 - 3), both the ACT and SR estimators demonstrate satisfactory performance. However, the MKER and MKTCR estimators do not perform well in cases (C2) and (C3). One possible reason is that they were originally developed under the assumptions that the loading matrix $\boldsymbol{B}$ has diverging eigenvalues and data are generated from an elliptical distribution (see Assumptions 2.1 and 2.3 in Yu et al. (2019)), which are not fulfilled in these scenarios. In the case of heavy-tailed data (see Tables 2 and Figures 4 - 5), our SR estimator performs significantly better than the other estimators. The three competing estimators, ACT, MKER, and MKTCR, struggle to accurately determine the number of common factors in most cases. Specifically, the ACT estimator has a tendency for overestimation, while the MKER and the MKTCR estimators tend to underestimate. On the contrary, the correct identification rate of our method SR approaches 100% as both $p$ and $n$ increase, as depicted in Figures 2 - 5.

Table 1: Percentages (%) of estimated number of common factors in 1000 simulations. Entries of common factors and idiosyncratic errors are generated from **light-tailed** distributions. The results are reported in the form $a(b|c)$, in which $a, b, c$ are the percentages of true estimates, overestimates, and underestimates, respectively. The notation "ave($\widehat{K}$)" denotes mean estimators for the case of the largest dimensions $(p, n) = (200, 400)$.

| **Case** | $p$ | $n$ | **ACT** | **MKER** | **MKTCR** | **SR** |
|---|---|---|---|---|---|---|
| | | | | **Normal population** | | |
| C1 | 50 | 100 | 100(0\|0) | 93.7(0\|6.3) | 96.9(0\|3.1) | 78.5(21\|0.5) |
| | 100 | 200 | 99.9(0.1\|0) | 99.1(0\|0.9) | 99.5(0\|0.5) | 85.8(14\|0.2) |
| | 150 | 300 | 100(0\|0) | 99.8(0\|0.2) | 99.9(0\|0.1) | 88.3(11.5\|0.2) |
| | 200 | 400 | 99.9(0.1\|0) | 99.7(0\|0.3) | 99.7(0\|0.3) | 89.2(10.8\|0) |
| | avg($\widehat{K}$) | | 3.001 | 2.997 | 2.997 | 3.397 |
| C2 | 50 | 100 | 99.7(0.3\|0) | 10.9(44.6\|44.5) | 11.2(43.7\|45.1) | 62.5(31\|6.5) |
| | 100 | 200 | 99.8(0.2\|0) | 11.8(28.3\|59.9) | 12(29.8\|58.2) | 81.2(18.8\|0) |
| | 150 | 300 | 99.4(0.6\|0) | 7.4(15.3\|77.3) | 7.6(16.3\|76.1) | 93.3(6.7\|0) |
| | 200 | 400 | 99(1\|0) | 7.4(11.9\|80.7) | 7.4(12.6\|80) | 96.8(3.2\|0) |
| | avg($\widehat{K}$) | | 3.01 | 1.879 | 1.919 | 3.118 |
| C3 | 50 | 100 | 99.9(0.1\|0) | 21(5.9\|73.1) | 23.3(8.2\|68.5) | 64.7(33.8\|1.5) |
| | 100 | 200 | 100(0\|0) | 32.7(7.6\|59.7) | 34.8(9.5\|55.7) | 94.5(5.4\|0.1) |
| | 150 | 300 | 99.6(0.4\|0) | 39.6(3.8\|56.6) | 41.6(4.4\|54) | 96.3(3.6\|0.1) |
| | 200 | 400 | 99.7(0.3\|0) | 35.3(2.9\|61.8) | 36.5(3.2\|60.3) | 98(2\|0) |
| | avg($\widehat{K}$) | | 3.003 | 2.047 | 2.079 | 3.083 |
| | | | | **Uniform and Chi-squared population** | | |
| C1 | 50 | 100 | 92(0\|8) | 46.3(0.2\|53.5) | 50.8(0.5\|48.7) | 72.3(26.9\|0.8) |
| | 100 | 200 | 99.4(0.3\|0.3) | 55.1(0\|44.9) | 59(0\|41) | 85(14.9\|0.1) |
| | 150 | 300 | 99.1(0.9\|0) | 58.9(0\|41.1) | 62.4(0\|37.6) | 88.4(11.5\|0.1) |
| | 200 | 400 | 98.4(1.6\|0) | 59.7(0\|40.3) | 63.5(0\|36.5) | 89.5(10.5\|0) |
| | avg($\widehat{K}$) | | 3.016 | 2.466 | 2.523 | 3.428 |
| C2 | 50 | 100 | 83.6(0.5\|15.9) | 10.5(45.4\|44.1) | 11.7(42.7\|45.6) | 53.1(39.7\|7.2) |
| | 100 | 200 | 99(1\|0) | 11(28.8\|60.2) | 11(30.4\|58.6) | 88.7(10.4\|0.9) |
| | 150 | 300 | 98.4(1.6\|0) | 11.6(29.6\|58.8) | 11.7(30.3\|58) | 91.9(8.1\|0) |
| | 200 | 400 | 98.7(1.3\|0) | 12.4(24.5\|63.1) | 12.1(26.2\|61.7) | 94.6(5.3\|0.1) |
| | avg($\widehat{K}$) | | 3.013 | 2.676 | 2.766 | 3.203 |
| C3 | 50 | 100 | 89(0.6\|10.4) | 15.6(29.2\|55.2) | 16.5(29.4\|54.1) | 59.5(37.7\|2.8) |
| | 100 | 200 | 99.6(0.4\|0) | 12(27.8\|60.2) | 11.7(29.1\|59.2) | 93.1(6.9\|0) |
| | 150 | 300 | 98.8(1.2\|0) | 12.4(22\|65.6) | 12.4(23.2\|64.4) | 93.2(6.7\|0.1) |
| | 200 | 400 | 99.1(0.9\|0) | 12.7(27.7\|59.6) | 12.9(29.3\|57.8) | 97.8(2.2\|0) |
| | avg($\widehat{K}$) | | 3.009 | 2.8 | 2.891 | 3.079 |

Table 2: Percentages (%) of estimated number of common factors in 1000 simulations. Entries of common factors and idiosyncratic errors are generated from **heavy-tailed** distributions. The results are reported in the form $a(b|c)$, in which $a, b, c$ are the percentages of true estimates, overestimates, and underestimates, respectively. The notation "ave($\widehat{K}$)" denotes mean estimators for the case of the largest dimensions $(p, n) = (200, 400)$.

| **Case** | $p$ | $n$ | **ACT** | **MKER** | **MKTCR** | **SR** |
|---|---|---|---|---|---|---|
| | | | | $t(2)$ **population** | | |
| C1 | 50 | 100 | 89.3(1.2\|9.5) | 36.4(1.7\|61.9) | 40(2.6\|57.4) | 84.9(14.4\|0.7) |
| | 100 | 200 | 93.8(4.5\|1.7) | 39.6(0.5\|59.9) | 44.2(0.7\|55.1) | 93.3(6.7\|0) |
| | 150 | 300 | 87.6(11\|1.4) | 45.8(0.5\|53.7) | 48(0.5\|51.5) | 94.1(5.8\|0.1) |
| | 200 | 400 | 82.8(16.5\|0.7) | 42.9(0.4\|56.7) | 44.8(0.6\|54.6) | 95.3(4.6\|0.1) |
| | avg($\widehat{K}$) | | 3.174 | 2.184 | 2.223 | 3.157 |
| C2 | 50 | 100 | 75.6(1.7\|22.7) | 10.3(34.8\|54.9) | 11.3(35.2\|53.5) | 67.8(28.2\|4) |
| | 100 | 200 | 91.8(7.2\|1) | 11(19.9\|69.1) | 11.7(20.6\|67.7) | 92(7.7\|0.3) |
| | 150 | 300 | 88.3(11.3\|0.4) | 10.2(16.3\|73.5) | 10.2(17.6\|72.2) | 94.5(5.5\|0) |
| | 200 | 400 | 81(18.7\|0.3) | 9.6(15\|75.4) | 10.3(16.4\|73.3) | 96.8(3.2\|0) |
| | avg($\widehat{K}$) | | 3.198 | 2.103 | 2.194 | 3.121 |
| C3 | 50 | 100 | 80.6(1.3\|18.1) | 9.8(15\|75.2) | 10.5(16.6\|72.9) | 73.3(24.9\|1.8) |
| | 100 | 200 | 93.6(5.7\|0.7) | 10.7(11.7\|77.6) | 11.4(12.6\|76) | 94.3(5.4\|0.3) |
| | 150 | 300 | 89.5(10.2\|0.3) | 13(9.6\|77.4) | 13.1(11.1\|75.8) | 95(5\|0) |
| | 200 | 400 | 85.9(14.1\|0) | 10(8.7\|81.3) | 10.5(9.4\|80.1) | 98.6(1.3\|0.1) |
| | avg($\widehat{K}$) | | 3.157 | 1.808 | 1.865 | 3.043 |
| | | | | **Cauchy population** | | |
| C1 | 50 | 100 | 24.9(2\|73.1) | 11.8(8.9\|79.3) | 12.6(10.1\|77.3) | 84.7(13.9\|1.4) |
| | 100 | 200 | 39.5(16\|44.5) | 8(7.2\|84.8) | 8.2(8\|83.8) | 93.7(5.9\|0.4) |
| | 150 | 300 | 34.5(31.5\|34) | 8.3(7.2\|84.5) | 8.7(7.7\|83.6) | 94.8(5.2\|0) |
| | 200 | 400 | 29.3(45.2\|25.5) | 9.6(7.7\|82.7) | 9.9(8.2\|81.9) | 95.4(4.6\|0) |
| | avg($\widehat{K}$) | | 3.433 | 1.727 | 1.764 | 3.174 |
| C2 | 50 | 100 | 19.2(2.4\|78.4) | 11(21.9\|67.1) | 11.6(22.3\|66.1) | 71.7(25.7\|2.6) |
| | 100 | 200 | 34.7(17.8\|47.5) | 12.3(16.1\|71.6) | 12.7(17.6\|69.7) | 93.7(6.2\|0.1) |
| | 150 | 300 | 37.8(37.4\|24.8) | 13.2(13.1\|73.7) | 13.3(14\|72.7) | 95.4(4.6\|0) |
| | 200 | 400 | 30.6(52\|17.4) | 10.7(16.9\|72.4) | 10.7(17.8\|71.5) | 95.7(4.3\|0) |
| | avg($\widehat{K}$) | | 3.675 | 2.196 | 2.234 | 3.171 |
| C3 | 50 | 100 | 24.4(3.4\|72.2) | 12.3(19\|68.7) | 12.6(20.1\|67.3) | 75.9(22.3\|1.8) |
| | 100 | 200 | 46.6(22.1\|31.3) | 12.6(18.8\|68.6) | 12.7(20\|67.3) | 95.6(4.4\|0) |
| | 150 | 300 | 39.9(35.1\|25) | 10.4(14.4\|75.2) | 10.5(15.7\|73.8) | 95.6(4.3\|0.1) |
| | 200 | 400 | 30.7(54.5\|14.8) | 8.4(12.4\|79.2) | 8.4(13.1\|78.5) | 98.6(1.3\|0.1) |
| | avg($\widehat{K}$) | | 3.745 | 1.929 | 1.979 | 3.056 |

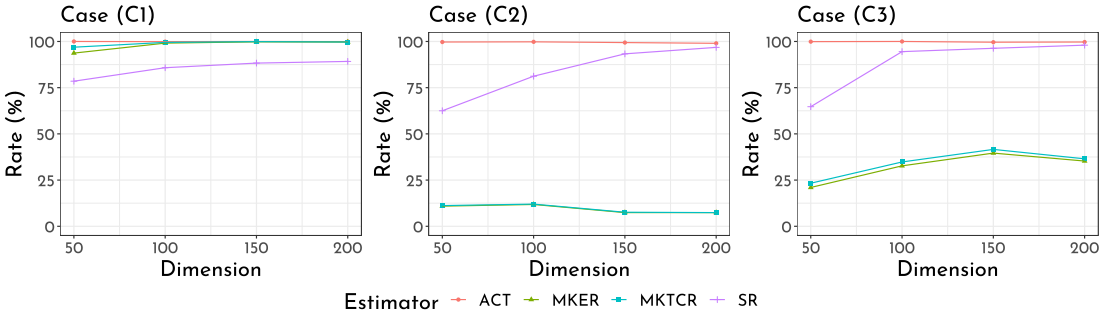



Figure 2: Correct identification rate of four estimators. Entries of common factors and idiosyncratic errors are generated from the standard normal distribution.

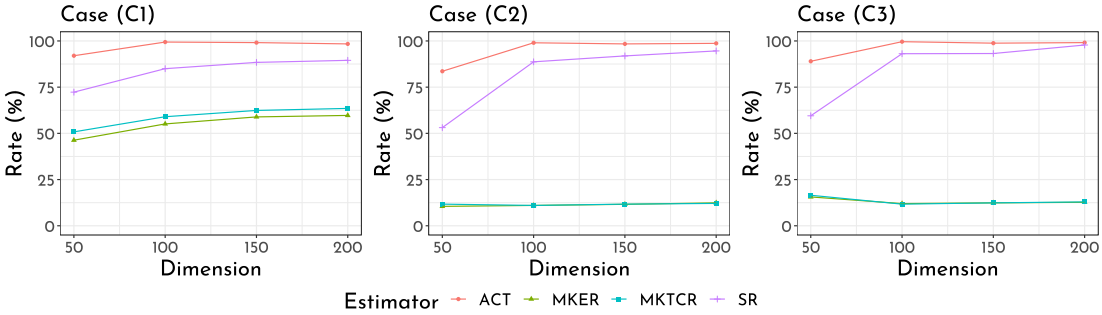



Figure 3: Correct identification rate of four estimators. Entries of common factors are generated from a scale mixture of normals with $\{w_i^f\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathsf{Uniform}(0,1)$, and entries of idiosyncratic errors are generated from a scale mixture of normals with $\{w_{ij}^e, i \in [n], j \in [p]\} \overset{\text{i.i.d.}}{\sim} \chi^2(1)$.

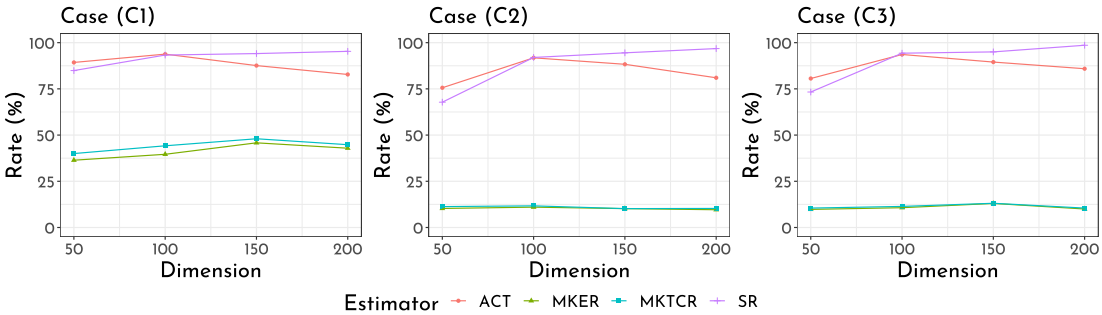



Figure 4: Correct identification rate of four estimators. Entries of common factors and idiosyncratic errors are generated from Student's $t(2)$ distribution.

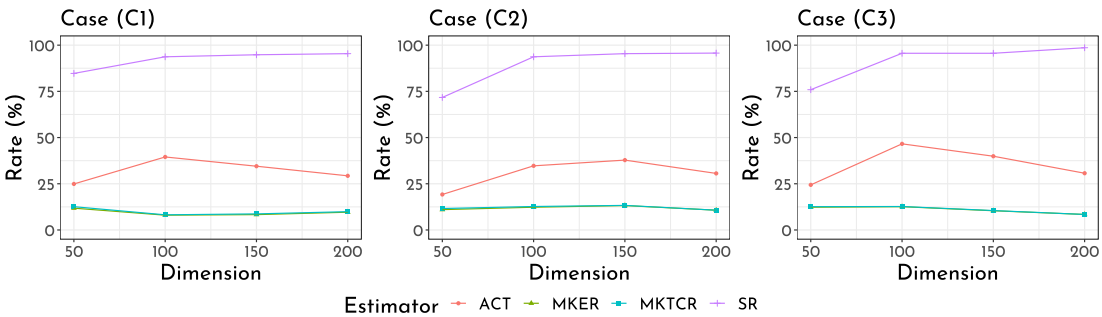


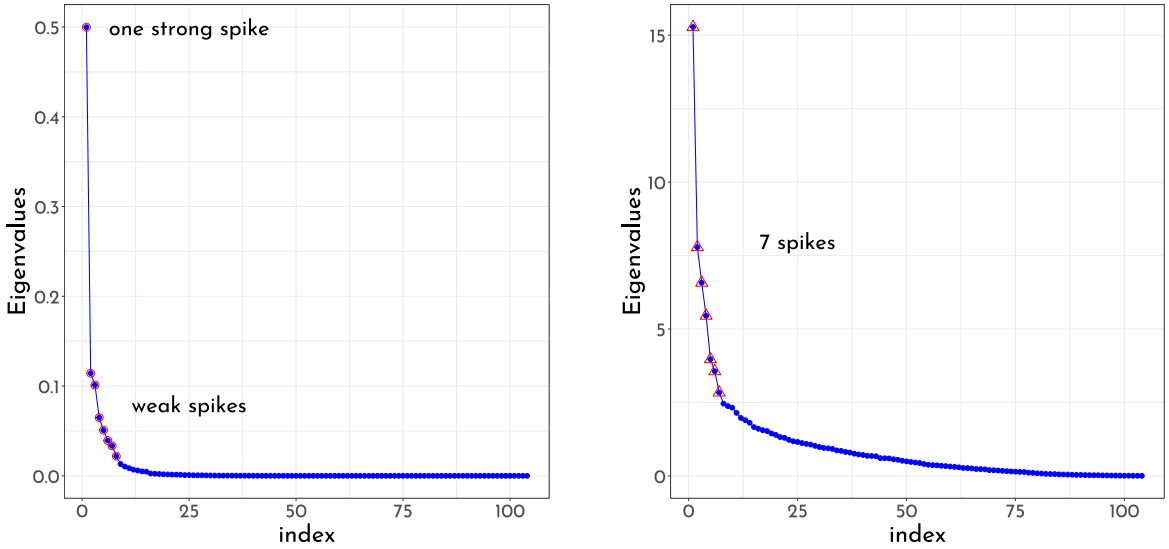

Figure 5: Correct identification rate of four estimators. Entries of common factors and idiosyncratic errors are generated from the standard Cauchy distribution.

# 4 Real data analysis

In this section, we analyze the monthly macroeconomic dataset (FRED-MD, McCracken and Ng (2016)) from March 1959 to January 2023. The data can be downloaded from the website http://research.stlouisfed.org/econ/mccracken/fred-md/, and includes the monthly series of 128 macroeconomic variables. Following McCracken and Ng (2016), the series with missing values are removed and the remaining dataset is transformed to a stationary form. After this preprocessing procedure, the data dimension is $p = 105$ and the sample size is $n = 767$. McCracken and Ng (2016)'s recommendation to remove outliers has not been implemented in our data analysis, as we believe that data with heavy-tailed distributions will inevitably contain extreme observations that cannot be circumvented. Since our estimator is tailored to heavy-tailed observations, we directly use it to identify number of factors.

We set $K_{\max} = 20$ and employ four methods for estimation. The results are as follows: $\widehat{K}_{\mathrm{ACT}} = 13$, $\widehat{K}_{\mathrm{MKER}} = \widehat{K}_{\mathrm{MKTCR}} = 1$, and $\widehat{K}_{\mathrm{SR}} = 7$. The dataset reveals that more than 67% of the macroeconomic variables exhibit a sample kurtosis that exceeds 9, which is the theoretical kurtosis of the Student's $t(5)$ distribution. This indicates that the dataset is probably heavy-tailed. The moment conditions which guarantee the consistency of ACT estimator are not satisfied in this dataset. Nonetheless, as shown in the simulation studies in Section 3, ACT has similar performance with our estimator when data is light-tailed, while it tends to overestimate when data is heavy-tailed. Same story happens for this real dataset. Both our SR estimator and Yu et al. (2019)'s MKER and MKTCR estimators are based on eigenvalues of certain type of sample correlation matrices as plotted in Figure 6. From Figure 6(a), it is evident that the multivariate Kendall's tau matrix exhibits one "strong" spike and several "weak" spikes. However, both the MKER and the MKTCR estimators only detect the strong spike while ignore the weaker spikes. They potentially

underestimate the total number of factors, similarly as shown in the simulation studies in Section 3. On the other hand, Figure 6(b) illustrates that our SR estimator has successfully detected all seven spikes of Spearman sample correlation matrix. Therefore, $\widehat{K}_{\mathrm{SR}} = 7$ is a more persuasive estimation for this dataset.




(a) Multivariate Kendall's tau matrix

(b) Spearman correlation matrix

Figure 6: Scatter plots of all the eigenvalues of multivariate Kendall's tau matrix and Spearman correlation matrix generated from the real dataset. Both the MKER and the MKTCR estimators only detect one "strong" spiked eigenvalue of the multivariate Kendall's tau matrix, and neglect several "weak" spikes. Our SR estimator detects all seven spikes of Spearman correlation matrix.

# 5 Discussions

In summary, we propose a novel estimator to identify weak factors in high-dimensional factor models when data is heavy-tailed. We demonstrate that, under certain assumptions, the number of spiked eigenvalues of Spearman sample correlation matrix is consistent with total number of significant factors. Our estimator is constructed based on this observation, and its consistency is proved under mild assumptions. From the perspective of RMT, we investigate the eigenstructure of Spearman sample correlation matrix under spike models and establish the phase-transition theory of its spiked eigenvalues. Simulation results demonstrate that our proposed estimator outperforms competing methods in various scenarios,

especially with heavy-tailed observations. However, our SR estimator does not perform well when the sample size is not large enough, such as $(p, n) = (50, 100)$. The possible reason is that the estimation of $m'(x)$ is inaccurate when the sample size is small. A more accurate estimator for $m'(x)$ would improve the accuracy of our SR estimator. Furthermore, it is worth extending our method for factor modeling in high-dimensional time series (Lam and Yao (2012), Li et al. (2017b)) and tensor data (Lam (2021), Chen and Lam (2022)). These extensions are beyond the scope of the current paper, and we leave them to future work.

# 6 Proof of Theorems 2.5 and 2.7

## 6.1 Some technical lemmas

In this section, we propose two technical lemmas in preparation for proving Theorems 2.5 and 2.7. The proofs of these lemmas are relegated to supplementary material.

Lemma 6.1 provides the LSD of $\boldsymbol{\rho}_n$ and $\boldsymbol{W}_n$, extending the result of Wu and Wang (2022). Their result is restricted to the non-paranormal distribution, and our study considers the case where the data follows a scale mixture of normal distributions, as indicated in Assumption (A2).

**Lemma 6.1** (Limiting spectral distribution)**.** *For the high-dimensional factor model* (1)*, assume that Assumptions (A1) – (A4) hold, and the ESD of* $\boldsymbol{\Sigma}_\rho = \mathbb{E}\boldsymbol{W}_n$ *tends to a proper probability measure* $H$ *as* $n \to \infty$*. Then, with probability one, both* $F^{\boldsymbol{\rho}_n}$ *and* $F^{\boldsymbol{W}_n}$ *tend to a non-random probability distribution* $F_{c,H}$*, the Stieltjes transform* $m = m(z)$ $(z \in \mathbb{C}^+)$ *of which is the unique solution to the equation*

$$m = \int \frac{1}{t(1 - c - czm) - z} \, \mathrm{d}H(t). \tag{15}$$

The following Lemma 6.2 concerns the limiting behavior of $\boldsymbol{\Omega}_K(\cdot, \cdot)$ defined in (18),

which plays a crucial role in the proof of Theorem 2.5.

**Lemma 6.2.** *Let* $\boldsymbol{X} = (X_{ij})_{n\times p} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\intercal$ *be an* $n \times p$ *random matrix. Assume that* $\boldsymbol{X}$ *satisfies Assumption (A1) and the following assumptions:*

*(B1) The vectors* $\{\boldsymbol{x}_\ell\}_{\ell=1}^n$ *are i.i.d., but the entries of each* $\boldsymbol{x}_\ell$ *are not necessarily i.i.d.*

*(B2) (Moment condition) For any* $i, j, s, t \in [p]$ *with* $i \neq j \neq s \neq t$*, we have*

$$\mathbb{E}X_{1i} = 0, \quad \mathbb{E}X_{1i}^2 = 1, \quad \mathbb{E}X_{1i}X_{1j} = 0, \quad \mathbb{E}X_{1i}^4 = O(1),$$

$$\mathbb{E}X_{1i}^2X_{1j}X_{1s} = O(p^{-1}), \quad \mathbb{E}X_{1i}X_{1j}X_{1s}X_{1t} = O(p^{-2}).$$

*(B3) (Weak dependency) For any* $p \times p$ *symmetric matrix* $\boldsymbol{T}$ *with bounded spectral norm, we have* $\mathrm{Var}(\boldsymbol{x}_1^\intercal \boldsymbol{T}\boldsymbol{x}_1) = o(p^2)$ *as* $p \to \infty$.

*(B4) (Concentration) For any convex 1-Lipschitz (with respect to the Euclidean norm) function* $F$ *from* $\mathbb{R}^p$ *to* $\mathbb{R}$*, let* $m_F$ *denote a median of* $F$*,*

$$\mathbb{P}\big(|F(\boldsymbol{x}_1) - m_F| > t\big) \leqslant C\exp\big\{-c(p)t^2\big\},$$

*where* $C$ *and* $c(p)$ *are independent of* $F$*, and* $C$ *is independent of* $p$*. We allow* $c(p)$ *to be a constant or to go to zero with* $p$ *like* $p^{-\alpha}$*,* $0 \leqslant \alpha < 1$.

*Moreover, let* $\boldsymbol{Y} = (Y_{ij})_{n\times p} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^\intercal$ *be a random matrix independent of* $\boldsymbol{X}$*, satisfying Assumptions (A1) and (B1) – (B4) with* $X_{ij}$ *and* $\boldsymbol{x}_\ell$ *replaced by* $Y_{ij}$ *and* $\boldsymbol{y}_\ell$*, respectively. Then,* $\boldsymbol{\Omega}_K(\lambda, \boldsymbol{X})$ *and* $\boldsymbol{\Omega}_K(\lambda, \boldsymbol{Y})$ *have the same limiting distribution, where* $\boldsymbol{\Omega}_K(\cdot,\cdot)$ *is defined in* (18).

**Remark 6.3.** *Assumption (B4) is from El Karoui (2009). In (El Karoui, 2009, p. 2386), the author gave some examples of distributions satisfying Assumption (B4), such as:*

- *Gaussian random vectors with covariance matrix* $\mathbf{\Sigma}_p$ *and* $c(p) = 1/\|\mathbf{\Sigma}_p\|_2$ *(according to Theorem 2.7 in Ledoux (2001)).*

- *Random vectors with entries bounded by* $1/\sqrt{c(p)}$ *(according to Lemma S1.3).*

## 6.2 Proof of Theorem 2.5

From Lemma 2.2, we investigate the phase-transition theory of spiked eigenvalues of $\boldsymbol{\rho}_n$ by those of $\boldsymbol{W}_n = (3/n)\sum_{i=1}^n \boldsymbol{A}_i\boldsymbol{A}_i^\intercal$. Recall that $\mathbf{\Sigma}_\rho = \mathbb{E}\boldsymbol{W}_n$. Define the spectral decomposition of $\mathbf{\Sigma}_\rho^{1/2}$ as

$$\mathbf{\Sigma}_\rho^{1/2} = \boldsymbol{U}\begin{pmatrix} \boldsymbol{D}_1^{1/2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_2^{1/2} \end{pmatrix}\boldsymbol{U}^\intercal,$$

where $\boldsymbol{U}$ is a $p \times p$ orthogonal matrix, $\boldsymbol{D}_1$ is the diagonal matrix consisting of the $K$ spiked population eigenvalues, and $\boldsymbol{D}_2$ is the diagonal matrix consisting of the remaining $p - K$ non-spiked eigenvalues. Let $\widetilde{\boldsymbol{A}}_i := \sqrt{3}\mathbf{\Sigma}_\rho^{-1/2}\boldsymbol{A}_i$ denote a transformed version of $\sqrt{3}\boldsymbol{A}_i$. It is obvious that $\widetilde{\boldsymbol{A}}_i$ is isotropic, that is, $\mathrm{Cov}(\widetilde{\boldsymbol{A}}_i) = \boldsymbol{I}_p$. By using these notations and the spectral decomposition of $\mathbf{\Sigma}_\rho^{1/2}$, we can write the characteristic equation as

$$0 = |\lambda\boldsymbol{I}_p - \boldsymbol{W}_n| = \left|\lambda\boldsymbol{I}_p - \boldsymbol{U}\begin{pmatrix} \boldsymbol{D}_1^{1/2} & \\ & \boldsymbol{D}_2^{1/2} \end{pmatrix}\boldsymbol{U}^\intercal\widetilde{\boldsymbol{W}}_n\boldsymbol{U}\begin{pmatrix} \boldsymbol{D}_1^{1/2} & \\ & \boldsymbol{D}_2^{1/2} \end{pmatrix}\boldsymbol{U}^\intercal\right|, \qquad (16)$$

where $\widetilde{\boldsymbol{W}}_n := n^{-1}\boldsymbol{\mathcal{A}}^\intercal\boldsymbol{\mathcal{A}}$ with $\boldsymbol{\mathcal{A}} := (\widetilde{\boldsymbol{A}}_1, \ldots, \widetilde{\boldsymbol{A}}_n)^\intercal$. Let $\boldsymbol{Q} = \boldsymbol{U}^\intercal\widetilde{\boldsymbol{W}}_n\boldsymbol{U}$ and partition it as

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{U}_1^\intercal\widetilde{\boldsymbol{W}}_n\boldsymbol{U}_1 & \boldsymbol{U}_1^\intercal\widetilde{\boldsymbol{W}}_n\boldsymbol{U}_2 \\ \boldsymbol{U}_2^\intercal\widetilde{\boldsymbol{W}}_n\boldsymbol{U}_1 & \boldsymbol{U}_2^\intercal\widetilde{\boldsymbol{W}}_n\boldsymbol{U}_2 \end{pmatrix},$$

where $\boldsymbol{U}_1$ is the submatrix formed by the first $K$ columns of $\boldsymbol{U}$, and $\boldsymbol{U}_2$ is the remaining submatrix. Plugging this identity into (16) yields that

$$
\begin{aligned}
0 &= \left| \lambda \boldsymbol{I}_p - \begin{pmatrix} \boldsymbol{D}_1^{1/2}\boldsymbol{Q}_{11}\boldsymbol{D}_1^{1/2} & \boldsymbol{D}_1^{1/2}\boldsymbol{Q}_{12}\boldsymbol{D}_2^{1/2} \\ \boldsymbol{D}_2^{1/2}\boldsymbol{Q}_{21}\boldsymbol{D}_1^{1/2} & \boldsymbol{D}_2^{1/2}\boldsymbol{Q}_{22}\boldsymbol{D}_2^{1/2} \end{pmatrix} \right| \\
&= \left| \lambda \boldsymbol{I}_{p-K} - \boldsymbol{D}_2^{1/2}\boldsymbol{Q}_{22}\boldsymbol{D}_2^{1/2} \right| \\
&\quad \times \left| \lambda \boldsymbol{I}_K - \boldsymbol{D}_1^{1/2}\boldsymbol{Q}_{11}\boldsymbol{D}_1^{1/2} - \boldsymbol{D}_1^{1/2}\boldsymbol{Q}_{12}\boldsymbol{D}_2^{1/2}(\lambda \boldsymbol{I}_{p-K} - \boldsymbol{D}_2^{1/2}\boldsymbol{Q}_{22}\boldsymbol{D}_2^{1/2})^{-1}\boldsymbol{D}_2^{1/2}\boldsymbol{Q}_{21}\boldsymbol{D}_1^{1/2} \right|,
\end{aligned}
$$

where the last equality follows from the formula $\det\left(\begin{smallmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{smallmatrix}\right) = \det(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}) \cdot \det(\boldsymbol{D})$. Since we only consider the spiked eigenvalues, we have $|\lambda \boldsymbol{I}_{p-K} - \boldsymbol{D}_2^{1/2}\boldsymbol{Q}_{22}\boldsymbol{D}_2^{1/2}| \neq 0$, and

$$
\begin{aligned}
0 &= \left| \lambda \boldsymbol{D}_1^{-1} - \boldsymbol{Q}_{11} - \boldsymbol{Q}_{12}\boldsymbol{D}_2^{1/2}(\lambda \boldsymbol{I}_{p-K} - \boldsymbol{D}_2^{1/2}\boldsymbol{Q}_{22}\boldsymbol{D}_2^{1/2})^{-1}\boldsymbol{D}_2^{1/2}\boldsymbol{Q}_{21} \right| \\
&= \left| \lambda \boldsymbol{D}_1^{-1} - \frac{1}{n}\boldsymbol{U}_1^\intercal \boldsymbol{\mathcal{A}}^\intercal \left[ \boldsymbol{I}_n + \frac{1}{n}\boldsymbol{\mathcal{A}}\boldsymbol{U}_2\boldsymbol{D}_2^{1/2}\left( \lambda \boldsymbol{I}_{p-K} - \frac{1}{n}\boldsymbol{D}_2^{1/2}\boldsymbol{U}_2^\intercal \boldsymbol{\mathcal{A}}^\intercal \boldsymbol{\mathcal{A}}\boldsymbol{U}_2\boldsymbol{D}_2^{1/2} \right)^{-1} \boldsymbol{D}_2^{1/2}\boldsymbol{U}_2^\intercal \boldsymbol{\mathcal{A}}^\intercal \right] \boldsymbol{\mathcal{A}}\boldsymbol{U}_1 \right| \\
&= \left| \lambda \boldsymbol{D}_1^{-1} - \frac{\lambda}{n}\boldsymbol{U}_1^\intercal \boldsymbol{\mathcal{A}}^\intercal \left( \lambda \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{\mathcal{A}}\boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{U}_2^\intercal \boldsymbol{\mathcal{A}}^\intercal \right)^{-1} \boldsymbol{\mathcal{A}}\boldsymbol{U}_1 \right| \\
&= \left| \lambda \boldsymbol{D}_1^{-1} - \frac{\lambda}{n}\mathrm{tr}\left\{ \left( \lambda \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{\mathcal{A}}\boldsymbol{\Gamma}\boldsymbol{\mathcal{A}}^\intercal \right)^{-1} \right\} \boldsymbol{I}_K + n^{-1/2}\boldsymbol{\Omega}_K(\lambda, \boldsymbol{\mathcal{A}}) \right|, \qquad (17)
\end{aligned}
$$

where $\boldsymbol{\Gamma} = \boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{U}_2^\intercal$ and

$$
\boldsymbol{\Omega}_K(\lambda, \boldsymbol{\mathcal{A}}) = \frac{\lambda}{\sqrt{n}} \left[ \mathrm{tr}\left\{ \left( \lambda \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{\mathcal{A}}\boldsymbol{\Gamma}\boldsymbol{\mathcal{A}}^\intercal \right)^{-1} \right\} \boldsymbol{I}_K - \boldsymbol{U}_1^\intercal \boldsymbol{\mathcal{A}}^\intercal \left( \lambda \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{\mathcal{A}}\boldsymbol{\Gamma}\boldsymbol{\mathcal{A}}^\intercal \right)^{-1} \boldsymbol{\mathcal{A}}\boldsymbol{U}_1 \right]. \qquad (18)
$$

From Lemma 6.2 and Remark 6.3, if $\widetilde{\boldsymbol{A}}_i$ satisfies Assumptions (B1) – (B4), we can replace entries of $\boldsymbol{\mathcal{A}}$ in (17) by the standard Gaussian entries without changing the phase-transition theory of the spiked eigenvalues. Then, our Theorem 2.5 follows from Lemma 3.1 and Theorems 4.1 – 4.2 in Bai and Yao (2012).

It remains to prove that random vector $\widetilde{\boldsymbol{A}}_i$ satisfies Assumptions (B1) – (B4), which shows that our Lemma 6.2 applies. It is obvious that $\widetilde{\boldsymbol{A}}_i$ satisfies Assumption (B1). By

using Lemma 2.3, we conclude that $\liminf_{p\to\infty} \lambda_p(\boldsymbol{\Sigma}_\rho^{1/2}) > 0$, and thus $\boldsymbol{\Sigma}_\rho^{-1/2}$ is bounded in spectral norm. Combining this information and the fact that each component of the random vector $\boldsymbol{A}_i$ follows $\mathsf{Uniform}(-1,1)$ distribution, we conclude that each component of the random vector $\widetilde{\boldsymbol{A}}_i$ has bounded fourth moment. This, together with Lemma 2.3 and similar calculations in Section S2.3.1, implies that $\widetilde{\boldsymbol{A}}_i$ satisfies the moment condition (B2). From (S2.16) and the fact that $\|\boldsymbol{\Sigma}_\rho^{-1/2}\|_2 = O(1)$, we have, for any $p\times p$ symmetry matrix $\boldsymbol{T}$ with bounded spectral norm, $\mathrm{Var}(\widetilde{\boldsymbol{A}}_i^\intercal \boldsymbol{T}\widetilde{\boldsymbol{A}}_i) = \mathrm{Var}(3\boldsymbol{A}_i^\intercal \boldsymbol{\Sigma}_\rho^{-1/2}\boldsymbol{T}\boldsymbol{\Sigma}_\rho^{-1/2}\boldsymbol{A}_i) = o(p^2)$. Hence, $\widetilde{\boldsymbol{A}}_i$ satisfies Assumption (B3). By Assumptions (A3) and (A4), Lemma 2.3, and the fact that each component of $\boldsymbol{A}_i$ follows $\mathsf{Uniform}(-1,1)$, we conclude that each component of $\widetilde{\boldsymbol{A}}_i$ is bounded, and thus satisfies the concentration assumption (B4), according to Lemma S1.3. Therefore, the random vector $\widetilde{\boldsymbol{A}}_i$ satisfies Assumptions (B1) – (B4), which shows that our Lemma 6.2 applies. This completes the proof of Theorem 2.5.

## 6.3 Proof of Theorem 2.7

From Theorem 2.5, we have, with probability one,

$$\begin{aligned}\widehat{m}'_{n,j}\big(\lambda_j(\boldsymbol{\rho}_n)\big) &= \frac{1}{p-j}\sum_{\ell=j+1}^{p}\frac{1}{\big\{\lambda_j(\boldsymbol{\rho}_n)-\lambda_\ell(\boldsymbol{\rho}_n)\big\}^2}\\ &\leqslant \frac{1}{\big\{\lambda_j(\boldsymbol{\rho}_n)-\lambda_{j+1}(\boldsymbol{\rho}_n)\big\}^2} \to \big\{\psi\big(\lambda_j(\boldsymbol{\Sigma}_\rho)\big)-\psi\big(\lambda_{j+1}(\boldsymbol{\Sigma}_\rho)\big)\big\}^{-2} \asymp 1,\end{aligned}$$

for any $j\in[K_0]$, where the function $\psi(\cdot)$ is defined in Theorem 2.5. Thus,

$$\frac{\widehat{m}'_{n,j+1}\big(\lambda_{j+1}(\boldsymbol{\rho}_n)\big)}{\widehat{m}'_{n,j}\big(\lambda_j(\boldsymbol{\rho}_n)\big)} \asymp 1, \qquad \text{for } j\in[K_0-1]. \tag{19}$$

The eigenvalues $\{\lambda_j(\boldsymbol{\rho}_n), K_0+1 \leqslant j \leqslant K_{\max}\}$ are bulk eigenvalues, and then we have $\lambda_j(\boldsymbol{\rho}_n) - \lambda_{j+1}(\boldsymbol{\rho}_n) = O(p^{-1})$. Thus, for $K_0+1 \leqslant j \leqslant K_{\max}$, we obtain

$$\widehat{m}'_{n,j}\big(\lambda_j(\boldsymbol{\rho}_n)\big) = \frac{1}{p-j}\sum_{\ell=j+1}^{p}\frac{1}{\big\{\lambda_j(\boldsymbol{\rho}_n)-\lambda_\ell(\boldsymbol{\rho}_n)\big\}^2} \geqslant \frac{1}{p-j}\frac{1}{\big\{\lambda_j(\boldsymbol{\rho}_n)-\lambda_{j+1}(\boldsymbol{\rho}_n)\big\}^2} = O(p).$$

Therefore,

$$\frac{\widehat{m}'_{n,K_0+1}\big(\lambda_{K_0+1}(\boldsymbol{\rho}_n)\big)}{\widehat{m}'_{n,K_0}\big(\lambda_{K_0}(\boldsymbol{\rho}_n)\big)} \to \infty. \tag{20}$$

Moreover, for $K_0+1 \leqslant j \leqslant K_{\max}$, we have

$$\begin{aligned}\frac{\widehat{m}'_{n,j+1}\big(\lambda_{j+1}(\boldsymbol{\rho}_n)\big)}{\widehat{m}'_{n,j}\big(\lambda_j(\boldsymbol{\rho}_n)\big)} &= \frac{p-j}{p-j-1}\frac{\sum_{\ell=j+2}^{p}\big\{\lambda_{j+1}(\boldsymbol{\rho}_n)-\lambda_\ell(\boldsymbol{\rho}_n)\big\}^{-2}}{\sum_{\ell=j+1}^{p}\big\{\lambda_j(\boldsymbol{\rho}_n)-\lambda_\ell(\boldsymbol{\rho}_n)\big\}^{-2}} \\ &\asymp \frac{\sum_{\ell=j+2}^{p}\big\{\lambda_{j+1}(\boldsymbol{\rho}_n)-\lambda_\ell(\boldsymbol{\rho}_n)\big\}^{-2}}{\sum_{\ell=j+1}^{p}\big\{\lambda_j(\boldsymbol{\rho}_n)-\lambda_\ell(\boldsymbol{\rho}_n)\big\}^{-2}} \asymp 1.\end{aligned} \tag{21}$$

Combining (19) – (21), we complete the proof of Theorem 2.7.

## SUPPLEMENTARY MATERIAL

This supplementary material contains some auxiliary lemmas and the technical proofs of Lemmas 2.2, 2.3, 6.1, 6.2, S1.6, and S1.7.

# References


Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters*, 80(23-24):1806–1813.

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1):99–102.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1):436–465.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, Z. and Yao, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'IHP Probabilités et statistiques*, 44(3):447–474.

Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177.

Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statistica Sinica*, 18(2):425–442.

Bandeira, A. S., Lodhia, A., and Rigollet, P. (2017). Marčenko-Pastur law for Kendall's tau. *Electronic Communications in Probability*, 22.

Bao, Z. (2019). Tracy–widom limit for spearman's rho. Technical report.

Bao, Z., Lin, L.-C., Pan, G., and Zhou, W. (2015). Spectral statistics of large dimensional spearman's rank correlation matrix and its application. *The Annals of Statistics*, 43(6):2588–2623.

Chen, W. and Lam, C. (2022). Rank and factor loadings estimation in time series tensor factor model by pre-averaging. *arXiv:2208.04012*.

El Karoui, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *The Annals of Applied Probability*, 19(6):2362 – 2405.

Fan, J., Guo, J., and Zheng, S. (2020). Estimating number of factors by adjusted eigenvalues thresholding. *Journal of the American Statistical Association*, 117(538):852–861.

Fan, J., Liu, H., and Wang, W. (2018). Large covariance estimation through elliptical factor models. *The Annals of statistics*, 46(4):1383–1414.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

He, Y., Kong, X., Yu, L., and Zhang, X. (2022a). Large-dimensional factor analysis without moment constraints. *Journal of Business & Economic Statistics*, 40(1):302–312.

He, Y., Wang, Y., Yu, L., Zhou, W., and Zhou, W.-X. (2022b). Matrix Kendall's tau in high-dimensions: A robust statistic for matrix factor model. *arXiv preprint arXiv:2207.09633*.

Heinen, A. and Valdesogo, A. (2020). Spearman rank correlation of the bivariate Student t and scale mixtures of normal distributions. *Journal of Multivariate Analysis*, 179.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325.

Jiang, D. and Bai, Z. (2021). Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices. *Bernoulli*, 27(1):274–294.

Kong, X.-B. (2017). On the number of common factors with high-frequency data. *Biometrika*, 104(2):397–410.

Lam, C. (2021). Rank determination for time series tensor factor model using correlation thresholding. Technical report, Working paper LSE.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, 40(2).

Ledoux, M. (2001). *The concentration of measure phenomenon*, volume 89. American Mathematical Society.

Li, H., Li, Q., and Shi, Y. (2017a). Determining the number of factors when the number of factors can increase with sample size. *Journal of Econometrics*, 197(1):76–86.

Li, Z., Wang, C., and Wang, Q. (2022). On eigenvalues of a high dimensional Kendall's rank correlation matrix with dependence. *Science China Mathematics*, to appear.

Li, Z., Wang, Q., and Yao, J. (2017b). Identifying the number of factors from singular values of a large sample auto-covariance matrix. *The Annals of Statistics*, 45(1):257 – 288.

Liu, H., Lafferty, J., and Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10(80):2295–2328.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92(4):1004–1016.

Spearman, C. (1961). The proof and measurement of association between two things. *The American Journal of Psychology*, 15:72–101.

Wu, Z. and Wang, C. (2022). Limiting spectral distribution of large dimensional Spearman's rank correlation matrices. *Journal of Multivariate Analysis*, 191.

Yu, L., He, Y., and Zhang, X. (2019). Robust factor number specification for large-dimensional elliptical factor model. *Journal of Multivariate analysis*, 174.