Journal of Multivariate Analysis 197 (2023) 105205

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

On singular values of large dimensional lag-τ sample auto-correlation matrices

Zhanting Long [a], Zeng Li [a,*], Ruitao Lin [b], Jiaxin Qiu [c]

[a] Southern University of Science and Technology, China
[b] The University of Texas MD Anderson Cancer Center, United States of America
[c] The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 10 March 2022
Received in revised form 25 May 2023
Accepted 26 May 2023
Available online 1 June 2023

AMS 2020 subject classifications:
primary 60B20
secondary 62H25

Keywords:
Auto-correlation matrix
Auto-covariance matrix
Largest eigenvalue
Limiting spectral distribution
Random matrix theory

ABSTRACT

We study the limiting behavior of singular values of a lag-τ sample auto-correlation matrix $\mathbf{R}_\tau^\epsilon$ of large dimensional vector white noise process, the error term $\boldsymbol{\epsilon}$ in the high-dimensional factor model. We establish the limiting spectral distribution (LSD) that characterizes the global spectrum of $\mathbf{R}_\tau^\epsilon$, and derive the limit of its largest singular value. All the asymptotic results are derived under the high-dimensional asymptotic regime where the data dimension and sample size go to infinity proportionally. Under mild assumptions, we show that the LSD of $\mathbf{R}_\tau^\epsilon$ is the same as that of the lag-τ sample auto-covariance matrix. Based on this asymptotic equivalence, we additionally show that the largest singular value of $\mathbf{R}_\tau^\epsilon$ converges almost surely to the right end point of the support of its LSD. Based on these results, we further propose two estimators of total number of factors with lag-τ sample auto-correlation matrices in a factor model. Our theoretical results are fully supported by numerical experiments as well.

1. Introduction

Consider a sequence of p-dimensional stationary random vectors $\{\mathbf{y}_i\}$ that has a factor structure and can be represented as

$$\mathbf{y}_i = \boldsymbol{\mu} + \mathbf{B}\mathbf{f}_i + \boldsymbol{\epsilon}_i, \quad i \in \{1, \ldots, n\}, \tag{1}$$

where $\{\mathbf{f}_i\}$ is a sequence of k-dimensional latent factor vectors, and $\{\boldsymbol{\epsilon}_i\}$ is a sequence of unobservable stochastic error vectors of independent and identically distributed (i.i.d.) components with zero mean and unit variance, independent with $\{\mathbf{f}_i\}$. Determining the number of factors k is a core problem for the factor model, and it possesses many challenges in the high-dimensional setting. Bai and Ng [2] first proposed a consistent estimator for static factor models. Hallin and Liška [14] developed an information criterion for dynamic factor models. Lam and Yao [18] studied the factor model for high-dimensional time series based on lagged auto-covariance matrices. Fan et al. [12] proposed an estimator based on sample correlation matrices to overcome the issue of the heterogeneous scales of the observed variables. In this paper, we study the lagged sample auto-correlation matrix for two reasons. On one hand, we believe that compared with the sample covariance matrix alone, the auto-correlation matrices of different lags may contain more information on k. Our ultimate goal is to investigate whether or not borrowing information from the auto-correlation matrices of different lags would make the final inference on the unknown number of factors more accurate or efficient. On the other hand, as

* Corresponding author.
E-mail address: liz9@sustech.edu.cn (Z. Li).

https://doi.org/10.1016/j.jmva.2023.105205

with Fan et al. [12], the lag-τ auto-correlation matrix overcomes the disadvantage of the heterogeneity among different components by self-normalization.

Mathematically, given the sequence of random vectors $\{\mathbf{y}_i\}$, we denote the population covariance matrix, the lag-τ (with τ being a fixed positive integer) auto-covariance, and auto-correlation matrices of $\{\mathbf{y}_i\}$ as $\boldsymbol{\Sigma}_0^{\mathbf{y}} = \mathrm{cov}(\mathbf{y}_i)$, $\boldsymbol{\Sigma}_\tau^{\mathbf{y}} = \mathrm{cov}(\mathbf{y}_i, \mathbf{y}_{i+\tau})$ and $\boldsymbol{\Omega}_\tau^{\mathbf{y}} = \mathrm{corr}(\mathbf{y}_i, \mathbf{y}_{i+\tau})$, respectively. Similarly, the population auto-covariance or auto-correlation matrices can be defined for sequences $\{\boldsymbol{\epsilon}_i\}$ and $\{\mathbf{f}_i\}$ by way of analogy. For example, $\boldsymbol{\Sigma}_\tau^{\mathbf{f}} = \mathrm{cov}(\mathbf{f}_i, \mathbf{f}_{i+\tau})$ is the lag-τ auto-covariance of $\{\mathbf{f}_i\}$. Let the superscript "$\top$" denote the transpose of a vector or matrix. It is known that the lag-τ auto-correlation matrix

$$\boldsymbol{\Omega}_\tau^{\mathbf{y}} = [\mathrm{diag}(\boldsymbol{\Sigma}_0^{\mathbf{y}})]^{-1/2}\boldsymbol{\Sigma}_\tau^{\mathbf{y}}[\mathrm{diag}(\boldsymbol{\Sigma}_0^{\mathbf{y}})]^{-1/2} = [\mathrm{diag}(\boldsymbol{\Sigma}_0^{\mathbf{y}})]^{-1/2}(\mathbf{B}\boldsymbol{\Sigma}_\tau^{\mathbf{f}}\mathbf{B}^\top)[\mathrm{diag}(\boldsymbol{\Sigma}_0^{\mathbf{y}})]^{-1/2},$$

exactly has k non-null singular values. As a result, based on the i.i.d observed data sample $\mathbf{y}_1, \ldots, \mathbf{y}_n$, the number of factors k can be naturally estimated via the singular values of sample version of the lag-τ auto-correlation matrix

$$\mathbf{R}_\tau^{\mathbf{y}} = [\mathrm{diag}(\mathbf{S}_0^{\mathbf{y}})]^{-1/2}\mathbf{S}_\tau^{\mathbf{y}}[\mathrm{diag}(\mathbf{S}_0^{\mathbf{y}})]^{-1/2}.$$

Note that, the lag-τ sample auto-covariance matrix is given by

$$\begin{aligned}\mathbf{S}_\tau^{\mathbf{y}} &= \frac{1}{n-1}\sum_{i=1}^n(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_{i+\tau} - \bar{\mathbf{y}})^\top = \mathbf{B}\left(\frac{1}{n-1}\sum_{i=1}^n(\mathbf{f}_i - \bar{\mathbf{f}})(\mathbf{f}_{i+\tau} - \bar{\mathbf{f}})^\top\right)\mathbf{B}^\top \\ &+ \mathbf{B}\left(\frac{1}{n-1}\sum_{i=1}^n(\mathbf{f}_i - \bar{\mathbf{f}})(\boldsymbol{\epsilon}_{i+\tau} - \bar{\boldsymbol{\epsilon}})^\top\right) + \left(\frac{1}{n-1}\sum_{i=1}^n(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})(\mathbf{f}_{i+\tau} - \bar{\mathbf{f}})^\top\right)\mathbf{B}^\top \\ &+ \frac{1}{n-1}\sum_{i=1}^n(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})(\boldsymbol{\epsilon}_{i+\tau} - \bar{\boldsymbol{\epsilon}})^\top \triangleq \mathbf{P}_\tau^{\mathbf{B}} + \mathbf{S}_\tau^{\boldsymbol{\epsilon}},\end{aligned} \tag{2}$$

where for a sequence $\{\mathbf{a}_i\} = \{\mathbf{y}_i\}$, $\{\boldsymbol{\epsilon}_i\}$, or $\{\mathbf{f}_i\}$, $\bar{\mathbf{a}} = \sum_{i=1}^n \mathbf{a}_i/n$ and by convention $\mathbf{a}_i = \mathbf{a}_{n+i}$ for $i \in \{1, \ldots, \tau\}$. Since $\mathbf{P}_\tau^{\mathbf{B}}$ is of rank k, the lag-τ sample auto-covariance matrix of $\{\mathbf{y}_i\}$, $\mathbf{S}_\tau^{\mathbf{y}}$, can be treated as a finite rank perturbation of the lag-τ sample auto-covariance matrix of $\{\boldsymbol{\epsilon}_i\}$, $\mathbf{S}_\tau^{\boldsymbol{\epsilon}}$, which is of rank $p \gg k$. Consequently, under certain circumstances, the lag-τ sample auto-correlation matrix of $\{\mathbf{y}_i\}$, $\mathbf{R}_\tau^{\mathbf{y}}$, is also a finite rank perturbation of the lag-τ sample auto-correlation matrix of $\{\boldsymbol{\epsilon}_i\}$, $\mathbf{R}_\tau^{\boldsymbol{\epsilon}}$, where

$$\mathbf{R}_\tau^{\boldsymbol{\epsilon}} = [\mathrm{diag}(\mathbf{S}_0^{\boldsymbol{\epsilon}})]^{-1/2}\mathbf{S}_\tau^{\boldsymbol{\epsilon}}[\mathrm{diag}(\mathbf{S}_0^{\boldsymbol{\epsilon}})]^{-1/2}.$$

Hence $\mathbf{R}_\tau^{\mathbf{y}}$ follows the spike model pattern which is well studied in the random matrix theory (RMT), see, Johnstone [17], Baik and Silverstein [9], Bai and Yao [4] and Benaych-Georges and Nadakuditi [10]. In fact, based on these observations, we proposed two estimators of total number of factors using sample auto-correlation matrices in the application section. Simulation experiments show that both estimators have satisfactory numerical performances.

In order to estimate total number of factors k, a clear picture is needed for the asymptotic behavior of the singular values of $\mathbf{R}_\tau^{\mathbf{y}}$, which are effected by the finite rank matrix and $\mathbf{R}_\tau^{\boldsymbol{\epsilon}}$. As a result, studying the sample auto-correlation matrix of $\{\boldsymbol{\epsilon}_i\}$, $\mathbf{R}_\tau^{\boldsymbol{\epsilon}}$, takes the first step to identify the number of factors in factor analysis. In this paper, we study the limiting singular value distribution and the limit of the largest singular value of $\mathbf{R}_\tau^{\boldsymbol{\epsilon}}$ under the high-dimensional setting where the dimension p and sample size n are assumed to be of the same order.

Because the eigenvalues of certain large random matrices play a critical role in many multivariate statistical analyses, limiting spectral properties of various matrix models has been widely studied using the RMT. In this paper, we use the tools of RMT to study the limiting spectral properties of the lag-τ sample auto-correlation matrix. There is rich literature on LSD and extreme eigenvalues of large-dimensional matrices. As a pioneering work, Wigner [27,28] discovered LSD for a large dimensional Wigner matrix and the limiting distribution is known as the semicircle law. Marčenko and Pastur [21] found that the empirical spectral distribution of sample covariance matrix converges to the Marčenko–Pastur law under mild conditions. Considering the product of random matrices, Yin and Krishnaiah [32], and Yin [29] investigated the LSD of $\mathbf{S}_n\mathbf{A}$, where $\mathbf{S}_n$ is sample covariance matrix and $\mathbf{A}$ is a positive definite matrix. Bai et al. [1] exhibited the existence of LSD of $\mathbf{S}_n\mathbf{H}$ where $\mathbf{H}$ is an arbitrary Hermitian matrix, and also investigated the LSD of $\mathbf{S}_n\mathbf{W}$ where $\mathbf{W}$ is a Wigner matrix. Yin et al. [30] and Bai et al. [7] showed the existence of the LSD of multivariate F-matrix. Bai et al. [8], Wachter [25] and Silverstein [22] derived the explicit form of the LSD of multivariate F-matrix. The form of $\mathbf{H}+\mathbf{X}\mathbf{D}\mathbf{X}^\top$, where $\mathbf{H}$ is a Hermitian matrix, $\mathbf{D}$ is diagonal, and $\mathbf{X}$ contains independent columns, has been studied by Silverstein and Bai [23]. Bose and Mitra [11] derived the LSD of a circulant matrix. The limiting distributions of eigenvalues of sample correlation matrices were discovered by Jiang [16]. For a high-dimensional time series structure, Li et al. [19] investigated the limiting singular value distribution of sample auto-covariance matrices. Most results are derived via the tools of the Stieltjes transform and moment method.

As for the limiting behavior of extreme eigenvalues, the first known result was established by Geman [13], who showed that the largest eigenvalue of a sample covariance matrix convergences to a limit almost surely under a growth condition on all the moments. Yin et al. [31] improved this result under the existence of the fourth moment. For the Wigner matrix, Bai and Yin [5] found the sufficient and necessary conditions for the almost sure convergence of the largest eigenvalue. Jiang [16] showed the largest eigenvalue of a sample correlation almost surely convergences to the right edge of its LSD support. Vu [24] derived the upper bound for the spectral norm of symmetric random matrices with independent entries. Wang and Yao [26] established the convergence of the largest singular value of a sample auto-covariance matrix based on graph theory.

The results derived in this paper heavily rely on the pioneer work of Jiang [16] and Li et al. [19]. In particular, Jiang [16] showed that LSD for the sample correlation matrix $\mathbf{R}_0^\epsilon$ is the same as that for the sample covariance matrix $\mathbf{S}_0^\epsilon$ and also established the convergence of the largest eigenvalue of $\mathbf{R}_0^\epsilon$. Indeed, inspired by Jiang [16], we try to relate the asymptotic results of singular values of $\mathbf{R}_\tau^\epsilon$ to $\mathbf{S}_\tau^\epsilon$ for fixed $\tau \geq 1$. Since $\mathbf{R}_\tau^\epsilon$ is not symmetric, we equivalently investigate the limiting behavior of eigenvalues of $\mathbf{R}_\tau^* = \mathbf{R}_\tau^\epsilon(\mathbf{R}_\tau^\epsilon)^\top$. We show that LSD for $\mathbf{R}_\tau^*$ is the same as LSD for $\mathbf{S}_\tau^* = \mathbf{S}_\tau^\epsilon(\mathbf{S}_\tau^\epsilon)^\top$ in Li et al. [19], mimicking the case of $\mathbf{R}_0^\epsilon$ and $\mathbf{S}_0^\epsilon$ as shown in Jiang [16]. Additionally, we also prove that the largest eigenvalue of $\mathbf{R}_\tau^*$ converges almost surely to the right edge of its LSD support.

The rest of the paper is organized as follows. Section 2 introduces the main theoretical results in this paper, including LSD and limit of the largest singular value of $\mathbf{R}_\tau^\epsilon$. The detailed proofs of the theorems and lemmas are given in Section 3. Section 4 describes the application of estimating total number of factors based on our theoretical results. Simulation experiments are carried out to check the performance of the proposed estimators.

2. Main results

2.1. Preliminary

Let μ be a finite measure on the real line, the Stieltjes transform of μ is defined by

$$m_\mu(z) = \int \frac{1}{x-z}\mu(dx), z \in \mathbb{C} \setminus \Gamma_\mu,$$

where Γ_μ is the support of the finite measure μ on the real line $\mathbb{R}$.

Let $\mathbf{A}_n$ be a $p \times p$ Hermitian matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$, the empirical spectral distribution (ESD) of $\mathbf{A}_n$ is

$$F^{\mathbf{A}_n}(x) = \frac{1}{p}\sum_{j=1}^{p} I\left\{\lambda_j \leq x\right\}, x \in \mathbb{R}.$$

LSD is the limiting distribution of $\left\{F^{\mathbf{A}_n}\right\}_{n\geq 1}$ for a sequence of random matrices $\{\mathbf{A}_n\}_{n\geq 1}$. By the definition of $F^{\mathbf{A}_n}$, the Stieltjes transform of ESD $F^{\mathbf{A}_n}$ is

$$m_{\mathbf{A}_n}(z) = \int \frac{1}{x-z}F^{\mathbf{A}_n}(dx) = \frac{1}{p}\mathrm{tr}\left(\mathbf{A}_n - z\mathbf{I}_p\right)^{-1},$$

where $\mathrm{tr}(\cdot)$ denotes the trace function and $\mathbf{I}_p$ is the p-dimensional identity matrix. With $m_{\mathbf{A}_n}(z)$, the density function of the LSD of $\mathbf{A}_n$ can be obtained by inversion formula,

$$f(u) = \lim_{\epsilon \to 0_+} \Im\mathfrak{m}\, m(u + i\epsilon),$$

where z is substituted by $u + i\epsilon$ and $m(u+i\epsilon)$ is the limit of $m_{\mathbf{A}_n}(u+i\epsilon)$ as $n \to \infty$.

2.2. Limiting spectral distribution

Recall that $\mathbf{y}_i = \boldsymbol{\mu} + \mathbf{B}\mathbf{f}_i + \boldsymbol{\epsilon}_i$, $i \in \{1, \ldots, n\}$, we first focus on the limiting singular value distribution of the lag-τ auto-correlation matrix $\mathbf{R}_\tau^\epsilon$. Equivalently, we consider LSD of the symmetric matrix $\mathbf{R}_\tau^* = \mathbf{R}_\tau^\epsilon(\mathbf{R}_\tau^\epsilon)^\top$.

- Assumption A. $\boldsymbol{\epsilon}_i = \left(\epsilon_{i,1}, \ldots, \epsilon_{i,p}\right)^\top$, $i \in \{1, 2, \ldots, n\}$ are independent p-dimensional random vectors with independent entries satisfying

$$E\left(\epsilon_{i,j}\right) = 0, \quad E\left(\epsilon_{i,j}^2\right) = 1, \quad \sup_{1\leq i\leq n, 1\leq j\leq p} E\left(\left|\epsilon_{i,j}\right|^{4+\delta}\right) < M,$$

  for constant M and positive δ.
- Assumption B. As $p \to \infty$, $n \to \infty$ and $p/n \to y \in (0, \infty)$.

Theorem 1. *Under Assumptions A and B, as $p, n \to \infty$, for fixed $\tau \geq 1$, almost surely the empirical distribution of $F^{\mathbf{R}_\tau^*}$ converges to a deterministic probability function F whose Stieltjes transform $m = m(z)$, $z \in \mathbb{C} \setminus \mathbb{R}$, and satisfies the following equation*

$$z^2y^2m^3 + zy(y-1)m^2 - zm - 1 = 0.$$

The density function of F, $f(u)$, is given by

$$f(u) = \frac{1}{y\pi u}\left\{-u - \frac{5(y-1)^2}{3} + \frac{2^{4/3}\left(3u + (y-1)^2\right)(y-1)}{3d(u)^{1/3}} + \frac{2^{2/3}(y-1)d(u)^{1/3}}{3}\right.$$

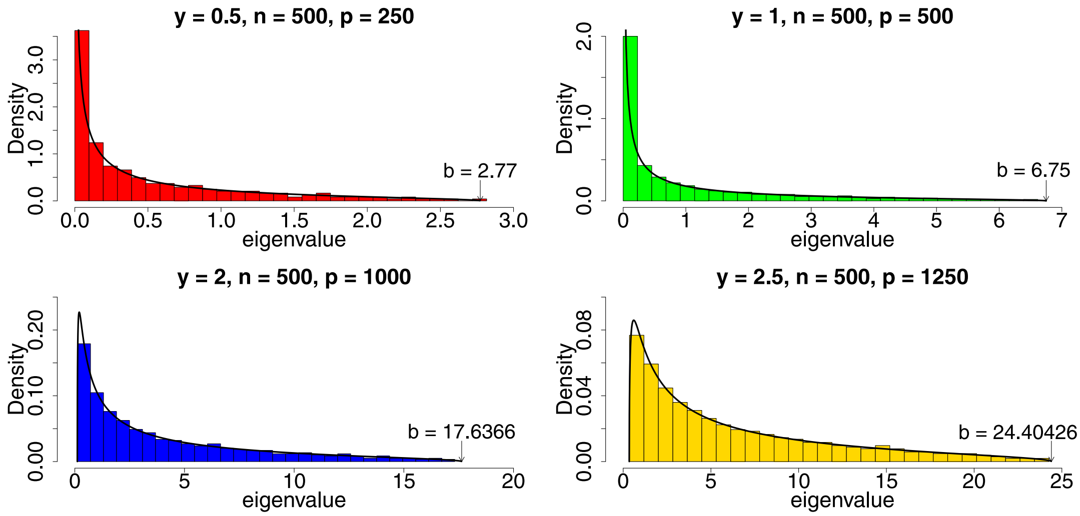

Fig. 1. The histogram of the sample eigenvalues of $\mathbf{R}_\tau^*$ with $\tau = 1$ and the theoretical limiting spectral density function $f(u)$. In all panels, the sample size n is fixed at $n = 500$, and the ratio of the dimensionality to the sample size is set as $y \in \{0.5, 1, 1.5, 2\}$ from top to bottom and left to right, respectively.

$$+\frac{1}{48}\left[-8(y-1)+\frac{2^{4/3}\left(3u+(y-1)^2\right)}{d(u)^{1/3}}+2^{2/3}d(u)^{1/3}\right]^2\Bigg\}^{1/2}, \tag{3}$$

where

$$d(u) = -2(y-1)^3 + 9(1+2y)u + 3\sqrt{3}\sqrt{u\left(-4u^2+(-1+4y(5+2y))u-4y(y-1)^3\right)}.$$

Here, the support of $f(u)$ is $(0, b]$ for $0 < y < 1$, and $[a, b]$ for $y \geq 1$, where

$$a = \frac{1}{8}\left(-1+20y+8y^2-(1+8y)^{3/2}\right), \quad b = \frac{1}{8}\left(-1+20y+8y^2+(1+8y)^{3/2}\right). \tag{4}$$

For the latter case with $y \geq 1$, the density function $f(u)$ has an additional point mass $\left(1-\frac{1}{y}\right)$ at the origin.

Fig. 1 contrasts the ESD of $\mathbf{R}_\tau^*$ (histogram) with $\tau = 1$ and the theoretical limiting density function $f(u)$ (solid line) based on i.i.d. samples from the standard normal distribution with $y \in \{0.5, 1, 2, 2.5\}$ and $n = 500$. It can be seen that the empirical histogram of eigenvalues of $\mathbf{R}_\tau^*$ is consistent with the limiting density function (3) for all (p, n) combinations.

Remark 1. By comparing Theorem 1 with Theorem 2.1 in Li et al. [19], we can see that $\mathbf{R}_1^*$ and $\mathbf{S}_1^*$ have the same LSDs, which is consistent with the results on sample correlation and covariance matrices Jiang [16]. In addition, as shown by Li et al. [19] the singular value distribution of $\mathbf{S}_\tau^\epsilon$ is the same as that of $\mathbf{S}_1^\epsilon$ for any fixed $\tau > 1$. Such results also hold for the singular value distribution of $\mathbf{R}_\tau^\epsilon$.

2.3. Limiting behavior of the largest eigenvalue

Next, we study the limiting behavior of the largest eigenvalue of $\mathbf{R}_\tau^*$. The following theorem shows that the largest eigenvalue converges to the right edge of the support of LSD of $\mathbf{R}_\tau^*$, mimicking the limiting behavior of the largest eigenvalue of $\mathbf{S}_\tau^*$.

Theorem 2. *Suppose that Assumptions A and B hold. Let $\lambda_{\max}(\mathbf{R}_\tau^*)$ be the largest eigenvalue of $\mathbf{R}_\tau^*$, then for fixed $\tau \geq 1$ almost surely,*

$$\lambda_{\max}(\mathbf{R}_\tau^*) \to b, \quad \text{as } p, n \to \infty,$$

where $b = \frac{1}{8}\left(-1+20y+8y^2+(1+8y)^{3/2}\right)$ is the right edge of the support of the LSD of $\mathbf{R}_\tau^*$.

Remark 2. The limit of the largest eigenvalue of $\mathbf{R}_\tau^*$ is equal to that of $\mathbf{S}_\tau^*$.

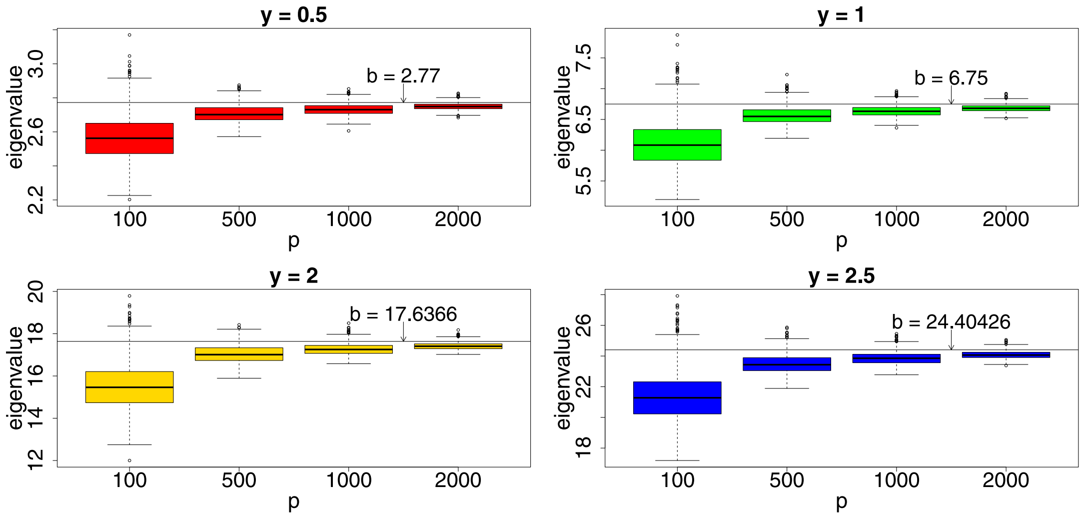

Fig. 2. Boxplot of the largest eigenvalue of lag-1 sample auto-correlation matrices $\mathbf{R}_1^* = \mathbf{R}_1^{\epsilon}(\mathbf{R}_1^{\epsilon})^{\top}$ based on 1000 standard normal samples. In all panels, the horizontal line indicates the right end point of its LSD, and the ratio of the dimensionality to the sample size is set as $y \in \{0.5, 1, 2, 2.5\}$ from top to bottom and left to right, respectively.

Fig. 2 displays the boxplot of the largest eigenvalues of $\mathbf{R}_\tau^*$ with $\tau = 1$ based on 1000 replications of independent and identically distributed samples from the standard normal distribution. We consider four values for the dimension, i.e., $p \in \{100, 500, 1000, 2000\}$, and vary the value of y, i.e., the ratio of the dimensionality to the sample size, from 0.5 to 2.5 in the four panels. In each panel, the horizontal line corresponds to the theoretical right end point b of LSD. From Fig. 2, we can see that the largest eigenvalue of $\mathbf{R}_\tau^*$ converges to the right end point b as both the dimension p and the sample size n increase proportionally.

2.4. Comparison with sample correlation matrix

In the previous sections, we study the lag-τ sample auto-correlation matrix for fixed $\tau \geq 1$. These asymptotic results cannot be directly extended to the case of $\mathbf{R}_0^*$. Because LSD for $\mathbf{R}_0^*$ is no longer the same as in Theorem 1. Unlike $\mathbf{R}_\tau^{\epsilon}$ for fixed $\tau \geq 1$, $\mathbf{R}_0^{\epsilon}$ is a symmetric matrix. The limiting behavior can be directly derived based on the sample correlation matrix $\mathbf{R}_0^{\epsilon}$, and there is no need to consider the eigenvalues of the transformation $\mathbf{R}_0^* = \mathbf{R}_0^{\epsilon}(\mathbf{R}_0^{\epsilon})^{\top}$. Although Jiang [16] already has shown that ESD for $\mathbf{R}_0^{\epsilon}$ converges to the well-known Marčenko–Pastur law, for completeness, we copy the results of $\mathbf{R}_0^{\epsilon}$ below.

Proposition 1 (*[16]*)**.** *Suppose $\boldsymbol{\epsilon}_i = \left(\epsilon_{1i}, \ldots, \epsilon_{pi}\right)^{\top}$, $i \in \{1, 2, \ldots, n\}$ are independent p-dimensional random vectors with entries satisfying $E(\epsilon_{ij}) = 0$, $E\left(|\epsilon_{ji}|^2\right) < \infty$. Let $p/n \to y \in (0, \infty)$, then, almost surely, $F^{\mathbf{R}_0^{\epsilon}}$ converges to a deterministic probability distribution with density function*

$$f_y(u) = \begin{cases} \frac{1}{2\pi u y}\sqrt{(b-u)(u-a)}, & \text{if } a \leq u \leq b, \\ 0, & \text{otherwise,} \end{cases}$$

and a point mass with value $1 - 1/y$ at $x = 0$ if $y > 1$,where $a = \left(1 - \sqrt{y}\right)^2$ and $b = \left(1 + \sqrt{y}\right)^2$.

Fig. 3 contrasts LSD for $\mathbf{R}_\tau^*$ (solid red curve) versus LSD for the sample correlation matrix $\mathbf{R}_0^{\epsilon}$ (solid blue curve), and ESD for $\mathbf{R}_\tau^*$ (light red histogram) with $\tau = 1$ versus ESD for $\mathbf{R}_0^{\epsilon}$ (light blue histogram) based on i.i.d. samples from the standard normal distribution with $y \in \{0.5, 1, 2, 2.5\}$ and $n = 500$. Clearly, the figure shows that LSD (or ESD) of $\mathbf{R}_\tau^*$ has different shapes with that of $\mathbf{R}_0^{\epsilon}$ for all (p, n) combinations.

3. Proofs

In this section, we provide the proofs of Theorems 1 and 2. Actually, our results rely on the results of the lag-τ sample auto-covariance matrix, which has been derived by Li et al. [19]. The strategy of our LSD proof is to show that LSD for $\mathbf{R}_\tau^*$ is the same as LSD for $\mathbf{S}_\tau^*$. Meanwhile, since the largest eigenvalue of $\mathbf{S}_\tau^*$ has been studied by Wang and Yao [26], we show that the largest eigenvalues of $\mathbf{R}_\tau^*$ and $\mathbf{S}_\tau^*$ converge to the same limit.

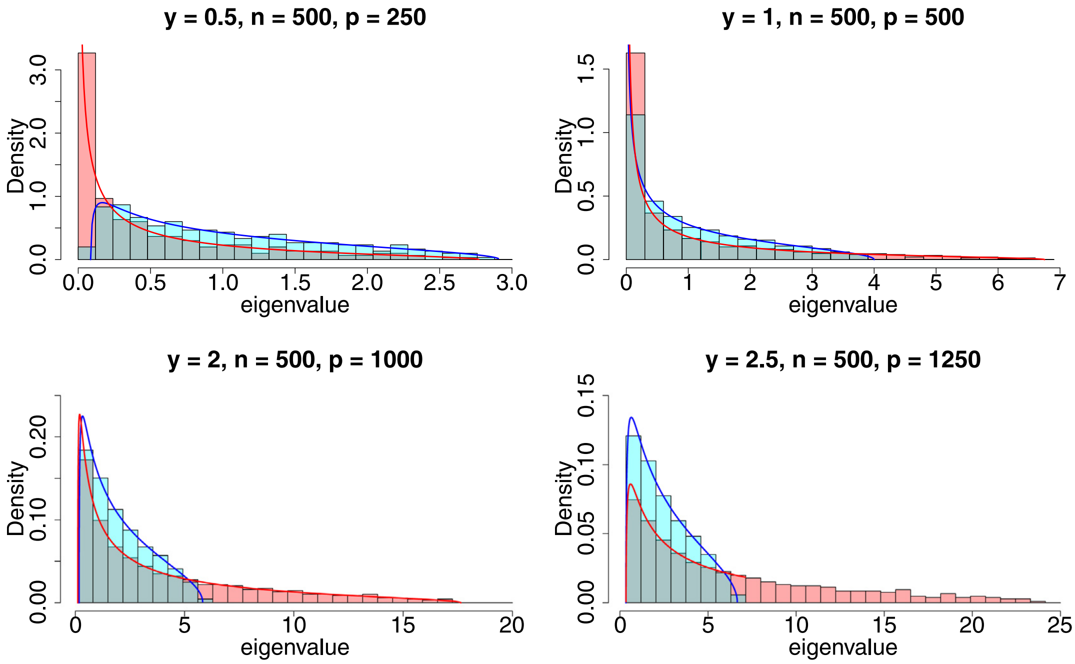

Fig. 3. Histograms of the sample eigenvalues of the lag-1 sample auto-correlation matrix $\mathbf{R}_\tau^*$ with $\tau = 1$ (light red) and the sample correlation matrix $\mathbf{R}_0^\epsilon$ (light blue). Theoretical density functions of the LSDs of $\mathbf{R}_\tau^*$ (red) and $\mathbf{R}_0^\epsilon$ (blue) are exhibited in lines. In all panels, the sample size n is fixed at $n = 500$, and the ratio of the dimensionality to the sample size is set as $y \in \{0.5, 1, 1.5, 2\}$ from top to bottom and left to right, respectively.

3.1. Standardization

We first introduce a standardization procedure. It is known that, for the sample covariance matrix

$$\mathbf{S}_0^\epsilon = \frac{1}{N}\sum_{i=1}^{n}(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})^\top$$

where $N = n - 1$ is the adjusted sample size, if we consider the non-centered sample covariance matrix

$$\tilde{\mathbf{S}}_0^\epsilon = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^\top,$$

with $\mathrm{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}$, the asymptotic results for eigenvalues of $\mathbf{S}_0^\epsilon$ partially hold for matrix $\tilde{\mathbf{S}}_0^\epsilon$. Specifically, as for the first order result, $\mathbf{S}_0^\epsilon$ and $\tilde{\mathbf{S}}_0^\epsilon$ share the same LSD, i.e., the Marčenko–Pastur distribution F_y with index $y = \lim p/n$. We found that similar results apply for sample auto-covariance and auto-correlation matrices. Specifically, denote the non-centered sample auto-correlation and auto-covariance matrices as

$$\tilde{\mathbf{R}}_\tau^\epsilon = [\mathrm{diag}(\tilde{\mathbf{S}}_0^\epsilon)]^{-1/2}\tilde{\mathbf{S}}_\tau^\epsilon[\mathrm{diag}(\tilde{\mathbf{S}}_0^\epsilon)]^{-1/2}, \qquad \tilde{\mathbf{S}}_\tau^\epsilon = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_{i+\tau}^\top$$

and

$$\tilde{\mathbf{R}}_\tau^* = \tilde{\mathbf{R}}_\tau^\epsilon(\tilde{\mathbf{R}}_\tau^\epsilon)^\top, \qquad \tilde{\mathbf{S}}_\tau^* = \tilde{\mathbf{S}}_\tau^\epsilon(\tilde{\mathbf{S}}_\tau^\epsilon)^\top.$$

We first show in the following lemmas that the centered sample auto-covariance and auto-correlation matrices, $\mathbf{S}_\tau^*$, $\mathbf{R}_\tau^*$ and their corresponding non-centered versions $\tilde{\mathbf{S}}_\tau^*$, $\tilde{\mathbf{R}}_\tau^*$, share same first-order results, including LSD and limit of the largest eigenvalue.

Lemma 1. *Under the assumptions in Theorem 1, for fixed $\tau \geq 1$, as $p, n \to \infty$, the empirical spectral distribution $F^{\tilde{\mathbf{S}}_\tau^*}$ almost surely converges to the same LSD as $F^{\mathbf{S}_\tau^*}$.*

Lemma 2. *Under the assumptions in Theorem 1, for fixed $\tau \geq 1$, as $p, n \to \infty$, the largest eigenvalue of $\tilde{\mathbf{S}}_\tau^*$, $\lambda_{\max}(\tilde{\mathbf{S}}_\tau^*)$ almost surely converges to the same limit as that of $\lambda_{\max}(\mathbf{S}_\tau^*)$.*

Lemma 3. *Under the assumptions in Theorem 1, for fixed $\tau \geq 1$, as $p, n \to \infty$, the empirical spectral distribution $F^{\tilde{\mathbf{R}}_\tau^*}$ almost surely converges to the same LSD as $F^{\mathbf{R}_\tau^*}$, the distribution with a density function given by (3).*

Lemma 4. *Under the assumptions in Theorem 1, for fixed $\tau \geq 1$, as $p, n \to \infty$, the largest eigenvalue of $\tilde{\mathbf{R}}_\tau^*$, $\lambda_{\max}(\tilde{\mathbf{R}}_\tau^*)$ almost surely converges to the same limit as that of $\lambda_{\max}(\mathbf{R}_\tau^*)$.*

The proof of these four lemmas are provided in Section 3.2. Based on these asymptotic equivalence results, we only need to study the non-centered lag-τ sample auto-covariance matrix $\tilde{\mathbf{S}}_\tau^\epsilon$ and the non-centered lag-τ sample auto-correlation matrix $\tilde{\mathbf{R}}_\tau^\epsilon$ to complete the proofs of Theorems 1 and 2.

3.2. Proofs of theorems and lemmas

Proof of Theorem 1. Theorem 1 follows from Lemmas 1 and 3 and the following Lemma 5.

Lemma 5. *Under the assumptions in Theorem 1, let $L(\cdot,\cdot)$ be the Levy distance, for fixed $\tau \geq 1$, as $p, n \to \infty$, we have*

$$L^4\left(F^{\tilde{\mathbf{R}}_\tau^*}, F^{\tilde{\mathbf{S}}_\tau^*}\right) \to 0, \quad a.s.$$

Proof. First we consider the case $\tau = 1$. Suppose $\boldsymbol{\epsilon}_j^0 = \left(\epsilon_{1,j}, \ldots, \epsilon_{n,j}\right)^\top$, $\boldsymbol{\epsilon}_j^1 = \left(\epsilon_{2,j}, \ldots, \epsilon_{n+1,j}\right)^\top$, then we can define the non-centered sample auto-correlation matrix $\tilde{\mathbf{R}}_1^\epsilon$ and the non-centered sample auto-covariance matrix $\tilde{\mathbf{S}}_1^\epsilon$ as follows:

$$\tilde{\mathbf{R}}_1^\epsilon = \mathbf{X}_0^\top \mathbf{X}_1, \quad \tilde{\mathbf{S}}_1^\epsilon = \frac{1}{n}\mathbf{E}_0^\top \mathbf{E}_1,$$

where $\mathbf{X}_0 = \left(\frac{\boldsymbol{\epsilon}_1^0}{\|\boldsymbol{\epsilon}_1^0\|}, \ldots, \frac{\boldsymbol{\epsilon}_p^0}{\|\boldsymbol{\epsilon}_p^0\|}\right)$, $\mathbf{X}_1 = \left(\frac{\boldsymbol{\epsilon}_1^1}{\|\boldsymbol{\epsilon}_1^0\|}, \ldots, \frac{\boldsymbol{\epsilon}_p^1}{\|\boldsymbol{\epsilon}_p^0\|}\right)$, $\mathbf{E}_0 = \left(\boldsymbol{\epsilon}_1^0, \ldots, \boldsymbol{\epsilon}_p^0\right)$ and $\mathbf{E}_1 = \left(\boldsymbol{\epsilon}_1^1, \ldots, \boldsymbol{\epsilon}_p^1\right)$.

By the difference inequality, we have

$$L^4\left(F^{\tilde{\mathbf{R}}_1^*}, F^{\tilde{\mathbf{S}}_1^*}\right) \leq \frac{2}{p}\mathrm{tr}\left(\left(\tilde{\mathbf{R}}_1^\epsilon - \tilde{\mathbf{S}}_1^\epsilon\right)\left(\tilde{\mathbf{R}}_1^\epsilon - \tilde{\mathbf{S}}_1^\epsilon\right)^\top\right) \cdot \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{R}}_1^* + \tilde{\mathbf{S}}_1^*\right) := 2 \cdot W_1 \cdot W_2.$$

For $W_2 = \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{R}}_1^* + \tilde{\mathbf{S}}_1^*\right)$, we need to prove $W_2 \to C_1$ *a.s.*, where C_1 is a positive constant. Note that

$$\begin{aligned}
\frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right) &= \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^\top \boldsymbol{\epsilon}_k^1 {\boldsymbol{\epsilon}_k^1}^\top \boldsymbol{\epsilon}_j^0}{n^2} = \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\left(\sum_{i=1}^{n}\epsilon_{i,j}\epsilon_{i+1,k}\right)^2}{n^2} \\
&= \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\sum_{i=1}^{n}\epsilon_{i,j}^2\epsilon_{i+1,k}^2}{n^2} + \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\sum_{i_1\neq i_2}^{n}\epsilon_{i_1,j}\epsilon_{i_1+1,k}\epsilon_{i_2,j}\epsilon_{i_2+1,k}}{n^2} := Q_1 + Q_2.
\end{aligned}$$

For the term Q_1, if $p/n \to y > 0$ we have

$$Q_1 = \frac{p}{n}\cdot\frac{1}{np^2}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{i=1}^{n}\epsilon_{i,j}^2\epsilon_{i+1,k}^2 \to y, \quad a.s.$$

based on the law of large numbers.

For the term Q_2,

$$\begin{aligned}
\mathrm{E}\left(Q_2\right) &= \frac{1}{pn^2}\mathrm{E}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{i_1\neq i_2}^{n}\epsilon_{i_1,j}\epsilon_{i_1+1,k}\epsilon_{i_2,j}\epsilon_{i_2+1,k} \\
&= \frac{1}{pn^2}\sum_{j\neq k}^{p}\sum_{i_1\neq i_2}^{n}\mathrm{E}\left(\epsilon_{i_1,j}\right)\mathrm{E}\left(\epsilon_{i_1+1,k}\right)\mathrm{E}\left(\epsilon_{i_2,j}\right)\mathrm{E}\left(\epsilon_{i_2+1,k}\right) + \frac{1}{pn^2}\mathrm{E}\sum_{j=k}^{p}\sum_{i_1\neq i_2}^{n}\epsilon_{i_1,j}\epsilon_{i_1+1,k}\epsilon_{i_2,j}\epsilon_{i_2+1,k} \\
&= \frac{1}{pn^2}\mathrm{E}\sum_{j=1}^{p}\sum_{\substack{i_1\neq i_2 \\ i_1\neq i_2+1}}^{n}\epsilon_{i_1,j}\epsilon_{i_1+1,j}\epsilon_{i_2,j}\epsilon_{i_2+1,j} + 2\frac{1}{pn^2}\mathrm{E}\sum_{j=1}^{p}\sum_{i_1=i_2+1}^{n}\epsilon_{i_1,j}\epsilon_{i_1+1,j}\epsilon_{i_2,j}\epsilon_{i_2+1,j} \\
&= 2\frac{1}{pn^2}\mathrm{E}\sum_{j=1}^{p}\sum_{i_1=i_2+1}^{n}\epsilon_{i_1,j}^2\epsilon_{i_1+1,j}\epsilon_{i_2,j} = 0.
\end{aligned}$$

$$\begin{aligned}\mathrm{Var}\,(Q_2) &= \frac{1}{p^2n^4}\mathrm{E}\left(\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{i_1\neq i_2}^{n}\epsilon_{i_1,j}\epsilon_{i_1+1,k}\epsilon_{i_2,j}\epsilon_{i_2+1,k}\right)^2 = \frac{1}{p^2n^4}\mathrm{E}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{i_1\neq i_2}^{n}\epsilon_{i_1,j}^2\epsilon_{i_1+1,k}^2\epsilon_{i_2,j}^2\epsilon_{i_2+1,k}^2\\
&= \frac{1}{p^2n^4}\sum_{j\neq k}^{p}\sum_{i_1\neq i_2}^{n}\mathrm{E}\left(\epsilon_{i_1,j}^2\right)\mathrm{E}\left(\epsilon_{i_1+1,k}^2\right)\mathrm{E}\left(\epsilon_{i_2,j}^2\right)\mathrm{E}\left(\epsilon_{i_2+1,k}^2\right) + \frac{1}{p^2n^4}\mathrm{E}\sum_{j=k}^{p}\sum_{i_1\neq i_2}^{n}\epsilon_{i_1,j}^2\epsilon_{i_1+1,k}^2\epsilon_{i_2,j}^2\epsilon_{i_2+1,k}^2\\
&= O\left(\frac{1}{n^2}\right) + \frac{1}{p^2n^4}\sum_{j=1}^{p}\sum_{\substack{i_1\neq i_2\\ i_1\neq i_2+1}}^{n}\mathrm{E}\left(\epsilon_{i_1,j}^2\right)\mathrm{E}\left(\epsilon_{i_1+1,j}^2\right)\mathrm{E}\left(\epsilon_{i_2,j}^2\right)\mathrm{E}\left(\epsilon_{i_2+1,j}^2\right) + 2\frac{1}{p^2n^4}\mathrm{E}\sum_{j=1}^{p}\sum_{i_1=i_2+1}^{n}\epsilon_{i_1,j}^4\epsilon_{i_1+1,j}^2\epsilon_{i_2,j}^2\\
&= O\left(\frac{1}{n^2}\right) + O\left(\frac{1}{pn^2}\right) + O\left(\frac{1}{pn^3}\right) = O\left(\frac{1}{n^2}\right).\end{aligned}$$

According to Chebyshev's inequality, for any $\epsilon > 0$

$$P(|Q_2| > \epsilon) \le \frac{\mathrm{Var}\,(Q_2)}{\epsilon^2} = O\left(\frac{1}{n^2}\right),$$

which is summable. Hence, based on Borel–Cantelli lemma, $Q_2 \to 0$, *a.s.* Thus, we have

$$\frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right) \to y, \quad a.s.$$

For $\frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{R}}_1^*\right)$, we obtain that

$$\begin{aligned}\left|\frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{R}}_1^*\right) - \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right)\right| &= \left|\frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\boldsymbol{\epsilon}_j^{0\top}\boldsymbol{\epsilon}_k^1\boldsymbol{\epsilon}_k^{1\top}\boldsymbol{\epsilon}_j^0}{\|\boldsymbol{\epsilon}_j^0\|^2\|\boldsymbol{\epsilon}_k^0\|^2} - \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\boldsymbol{\epsilon}_j^{0\top}\boldsymbol{\epsilon}_k^1\boldsymbol{\epsilon}_k^{1\top}\boldsymbol{\epsilon}_j^0}{n^2}\right|\\
&\le \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\boldsymbol{\epsilon}_j^{0\top}\boldsymbol{\epsilon}_k^1\boldsymbol{\epsilon}_k^{1\top}\boldsymbol{\epsilon}_j^0}{n^2}\left|\frac{n^2}{\|\boldsymbol{\epsilon}_j^0\|^2\|\boldsymbol{\epsilon}_k^0\|^2} - 1\right|\\
&= \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\boldsymbol{\epsilon}_j^{0\top}\boldsymbol{\epsilon}_k^1\boldsymbol{\epsilon}_k^{1\top}\boldsymbol{\epsilon}_j^0}{n^2}\left|\left(\frac{n}{\|\boldsymbol{\epsilon}_j^0\|^2} - 1\right)\left(\frac{n}{\|\boldsymbol{\epsilon}_k^0\|^2} - 1\right) + \frac{n}{\|\boldsymbol{\epsilon}_j^0\|^2} - 1 + \frac{n}{\|\boldsymbol{\epsilon}_k^0\|^2} - 1\right|\\
&\le \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\boldsymbol{\epsilon}_j^{0\top}\boldsymbol{\epsilon}_k^1\boldsymbol{\epsilon}_k^{1\top}\boldsymbol{\epsilon}_j^0}{n^2}\left[\left|\left(\frac{n}{\|\boldsymbol{\epsilon}_j^0\|^2} - 1\right)\left(\frac{n}{\|\boldsymbol{\epsilon}_k^0\|^2} - 1\right)\right| + \left|\frac{n}{\|\boldsymbol{\epsilon}_j^0\|^2} - 1\right| + \left|\frac{n}{\|\boldsymbol{\epsilon}_k^0\|^2} - 1\right|\right]\\
&\le \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right)\cdot\left[\left(\max_{1\le j\le p}\left|\frac{n}{\|\boldsymbol{\epsilon}_j^0\|^2} - 1\right|\right)^2 + 2\max_{1\le j\le p}\left|\frac{n}{\|\boldsymbol{\epsilon}_j^0\|^2} - 1\right|\right].\end{aligned}$$

Since $\mathrm{E}|\epsilon_{1,1}|^4 < \infty$, by the Lemma 2 from Bai and Yin [6], we know

$$\max_{1\le j\le p}\left|\frac{\sum_{i=1}^n \epsilon_{i,j}^2}{n} - 1\right| \to 0, \quad a.s.,$$

and this implies that

$$\max_{1\le j\le p}\left|\frac{n}{\sum_{i=1}^n \epsilon_{i,j}^2} - 1\right| \to 0, \quad a.s. \tag{5}$$

Since $\frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right)$ converges to a constant y which has been shown above, we have

$$\left|\frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{R}}_1^*\right) - \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right)\right| \to 0, \quad a.s.$$

It follows that $\frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{R}}_1^*\right) \to y$ a.s., and then

$$W_2 = \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{R}}_1^*\right) + \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right) \to 2y, \quad a.s.$$

For the term of W_1, we have

$$W_1 = \frac{1}{p}\mathrm{tr}\left(\left(\tilde{\mathbf{R}}_1^{\epsilon} - \tilde{\mathbf{S}}_1^{\epsilon}\right)\left(\tilde{\mathbf{R}}_1^{\epsilon} - \tilde{\mathbf{S}}_1^{\epsilon}\right)^{\top}\right) = \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n^2}$$
$$- \frac{2}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n\cdot\|\boldsymbol{\epsilon}_j^0\|\|\boldsymbol{\epsilon}_k^0\|} + \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{\|\boldsymbol{\epsilon}_j^0\|^2\|\boldsymbol{\epsilon}_k^0\|^2} := Q_3 - 2Q_4,$$

where

$$Q_3 = \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{\|\boldsymbol{\epsilon}_j^0\|^2\|\boldsymbol{\epsilon}_k^0\|^2} - \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n^2}, \quad Q_4 = \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n\cdot\|\boldsymbol{\epsilon}_j^0\|\|\boldsymbol{\epsilon}_k^0\|} - \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n^2}.$$

For Q_4,

$$\left|\frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n\cdot\|\boldsymbol{\epsilon}_j^0\|\|\boldsymbol{\epsilon}_k^0\|} - \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n^2}\right| \le \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n^2}\left|\frac{n}{\|\boldsymbol{\epsilon}_j^0\|\|\boldsymbol{\epsilon}_k^0\|} - 1\right|$$
$$= \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{{\boldsymbol{\epsilon}_j^0}^{\top}\boldsymbol{\epsilon}_k^1{\boldsymbol{\epsilon}_k^1}^{\top}\boldsymbol{\epsilon}_j^0}{n^2}\left|\left(\frac{\sqrt{n}}{\sqrt{\|\boldsymbol{\epsilon}_j^0\|^2}} - 1\right)\left(\frac{\sqrt{n}}{\sqrt{\|\boldsymbol{\epsilon}_k^0\|^2}} - 1\right) + \frac{\sqrt{n}}{\sqrt{\|\boldsymbol{\epsilon}_j^0\|^2}} - 1 + \frac{\sqrt{n}}{\sqrt{\|\boldsymbol{\epsilon}_k^0\|^2}} - 1\right|$$
$$\le \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right)\cdot\left[\left(\max_{1\le j\le p}\left|\frac{\sqrt{n}}{\sqrt{\|\boldsymbol{\epsilon}_j^0\|^2}} - 1\right|\right)^2 + 2\max_{1\le j\le p}\left|\frac{\sqrt{n}}{\sqrt{\|\boldsymbol{\epsilon}_j^0\|^2}} - 1\right|\right].$$

According to Lemma 2 of Bai and Yin [6], we have

$$\max_{1\le j\le p}\left|\frac{\sqrt{n}}{\sqrt{\sum_{i=1}^{n}\epsilon_{i,j}^2}} - 1\right| \to 0, \quad a.s. \tag{6}$$

Therefore $Q_4 \to 0$ a.s. Given that the following result

$$Q_3 = \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{R}}_1^*\right) - \frac{1}{p}\mathrm{tr}\left(\tilde{\mathbf{S}}_1^*\right) \to 0, \quad a.s.$$

has been proved, we have

$$W_1 = \frac{1}{p}\mathrm{tr}\left(\left(\tilde{\mathbf{R}}_1^{\epsilon} - \tilde{\mathbf{S}}_1^{\epsilon}\right)\left(\tilde{\mathbf{R}}_1^{\epsilon} - \tilde{\mathbf{S}}_1^{\epsilon}\right)^{\top}\right) \to 0, \quad a.s.$$

Together with W_1, W_2, we obtain

$$L^4\left(F^{\tilde{\mathbf{R}}_1^*}, F^{\tilde{\mathbf{S}}_1^*}\right) \to 0, \quad a.s.$$

The procedure of the proof will not change for any given positive integer τ. Therefore, we have

$$L^4\left(F^{\tilde{\mathbf{R}}_\tau^*}, F^{\tilde{\mathbf{S}}_\tau^*}\right) \to 0, \quad a.s. \quad \square$$

Proof of Theorem 2. Theorem 2 follows from Lemmas 2, 4, 6 and Theorem 4.1 from Wang and Yao [26].

Lemma 6. *Under the assumptions in Theorem 1, let $\lambda_{\max}(\tilde{\mathbf{S}}_\tau^*)$ and $\lambda_{\max}(\tilde{\mathbf{R}}_\tau^*)$ be the largest eigenvalues of $\tilde{\mathbf{S}}_\tau^*$ and $\tilde{\mathbf{R}}_\tau^*$, respectively. As $p, n \to \infty$, we have*

$$\left|\sqrt{\lambda_{\max}(\tilde{\mathbf{R}}_\tau^*)} - \sqrt{\lambda_{\max}(\tilde{\mathbf{S}}_\tau^*)}\right| \to 0, \quad a.s.$$

Proof. Denote $\boldsymbol{\epsilon}_j^0 = \left(\epsilon_{1,j}, \ldots, \epsilon_{n,j}\right)^{\top}$, $\boldsymbol{\epsilon}_j^\tau = \left(\epsilon_{\tau,j}, \ldots, \epsilon_{n+\tau,j}\right)^{\top}$. Rewrite

$$\tilde{\mathbf{R}}_\tau^{\epsilon} = \frac{1}{n}\mathbf{D}\mathbf{E}_0^{\top}\mathbf{E}_\tau\mathbf{D}, \quad \tilde{\mathbf{S}}_\tau^{\epsilon} = \frac{1}{n}\mathbf{E}_0^{\top}\mathbf{E}_\tau,$$

where $\mathbf{D} = \mathrm{diag}\left(\frac{\sqrt{n}}{\|\boldsymbol{\epsilon}_1^0\|}, \ldots, \frac{\sqrt{n}}{\|\boldsymbol{\epsilon}_p^0\|}\right)$, $\mathbf{E}_0 = \left(\boldsymbol{\epsilon}_1^0, \ldots, \boldsymbol{\epsilon}_p^0\right)$ and $\mathbf{E}_\tau = \left(\boldsymbol{\epsilon}_1^\tau, \ldots, \boldsymbol{\epsilon}_p^\tau\right)$.

Under the conditions of Theorem 1, according to Theorem 4.1 from Wang and Yao [26], we have

$$\lambda_{\max}(\tilde{\mathbf{S}}_\tau^*) \to b, \quad a.s., \tag{7}$$

where $b = \frac{1}{8}\left(-1 + 20y + 8y^2 + (1+8y)^{3/2}\right)$ is the right end point of the support of the LSD of $\mathbf{R}_\tau^*$. Our target is to show that

$$\left|\sqrt{\lambda_{\max}(\tilde{\mathbf{R}}_\tau^*)} - \sqrt{\lambda_{\max}(\tilde{\mathbf{S}}_\tau^*)}\right| \to 0, \quad a.s. \tag{8}$$

For any matrix $\mathbf{A}$, we denote $\|\mathbf{A}\|_2$ as the spectrum norm of $\mathbf{A}$, which is defined as the square root of the largest eigenvalue of $\mathbf{A}\mathbf{A}^\top$. By Corollary 7.3.8 from Horn and Johnson [15], we have

$$\left|\sqrt{\lambda_{\max}(\tilde{\mathbf{R}}_\tau^*)} - \sqrt{\lambda_{\max}(\tilde{\mathbf{S}}_\tau^*)}\right| \le \|\tilde{\mathbf{R}}_\tau^\epsilon - \tilde{\mathbf{S}}_\tau^\epsilon\|_2. \tag{9}$$

Meanwhile the spectrum norm satisfies the triangle inequality and $\|\mathbf{AC}\|_2 \le \|\mathbf{A}\|_2 \cdot \|\mathbf{C}\|_2$ for any $\mathbf{A}$ and $\mathbf{C}$, then we have

$$\begin{aligned}
\|\tilde{\mathbf{R}}_\tau^\epsilon - \tilde{\mathbf{S}}_\tau^\epsilon\|_2 &= \left\|\frac{1}{n}\mathbf{D}\mathbf{E}_0^\top\mathbf{E}_\tau\mathbf{D} - \frac{1}{n}\mathbf{E}_0^\top\mathbf{E}_\tau\right\|_2 = \left\|\frac{1}{n}\mathbf{D}\mathbf{E}_0^\top\mathbf{E}_\tau\mathbf{D} - \frac{1}{n}\mathbf{D}\mathbf{E}_0^\top\mathbf{E}_\tau + \frac{1}{n}\mathbf{D}\mathbf{E}_0^\top\mathbf{E}_\tau - \frac{1}{n}\mathbf{E}_0^\top\mathbf{E}_\tau\right\|_2 \\
&\le \left\|\frac{1}{n}\mathbf{D}\mathbf{E}_0^\top\mathbf{E}_\tau\mathbf{D} - \frac{1}{n}\mathbf{D}\mathbf{E}_0^\top\mathbf{E}_\tau\right\|_2 + \left\|\frac{1}{n}\mathbf{D}\mathbf{E}_0^\top\mathbf{E}_\tau - \frac{1}{n}\mathbf{E}_0^\top\mathbf{E}_\tau\right\|_2 \\
&= \left\|\frac{1}{n}\left(\mathbf{D} - \mathbf{I} + \mathbf{I}\right)\mathbf{E}_0^\top\mathbf{E}_\tau\left(\mathbf{D} - \mathbf{I}\right)\right\|_2 + \left\|\frac{1}{n}\left(\mathbf{D} - \mathbf{I}\right)\mathbf{E}_0^\top\mathbf{E}_\tau\right\|_2 \\
&\le \left\|\frac{1}{n}\left(\mathbf{D} - \mathbf{I}\right)\mathbf{E}_0^\top\mathbf{E}_\tau\left(\mathbf{D} - \mathbf{I}\right)\right\|_2 + 2\left\|\frac{1}{n}\left(\mathbf{D} - \mathbf{I}\right)\mathbf{E}_0^\top\mathbf{E}_\tau\right\|_2 \\
&\le \left\|\frac{1}{n}\mathbf{E}_0^\top\mathbf{E}_\tau\right\|_2 \cdot \|\mathbf{D} - \mathbf{I}\|_2^2 + 2\left\|\frac{1}{n}\mathbf{E}_0^\top\mathbf{E}_\tau\right\|_2 \cdot \|\mathbf{D} - \mathbf{I}\|_2 .
\end{aligned} \tag{10}$$

Since $\mathrm{E}|\epsilon_{1,1}|^4 < \infty$, by Lemma 2 of Bai and Yin [6], we know that

$$\max_{1\le j\le p}\left|\frac{\|\epsilon_j^0\|^2}{n} - 1\right| \to 0, \quad a.s.,$$

which implies

$$\|\mathbf{D} - \mathbf{I}\|_2 = \max_{1\le j\le p}\left|\frac{\sqrt{n}}{\|\epsilon_j^0\|} - 1\right| \to 0, \quad a.s. \tag{11}$$

This together with (7) and (10) proves (8). □

Proof of Lemma 1. By using the fact $\tilde{\mathbf{S}}_\tau^\epsilon = \mathbf{S}_\tau^\epsilon + \bar{\epsilon}\bar{\epsilon}^\top$ and Theorem A.44 in Bai and Silverstein [3], we have

$$\left\|F^{\tilde{\mathbf{S}}_\tau^*} - F^{\mathbf{S}_\tau^*}\right\| = \left\|F^{\tilde{\mathbf{S}}_\tau^\epsilon(\tilde{\mathbf{S}}_\tau^\epsilon)^\top} - F^{\mathbf{S}_\tau^\epsilon(\mathbf{S}_\tau^\epsilon)^\top}\right\| \leqslant \frac{1}{p}\mathrm{rank}\big(\tilde{\mathbf{S}}_\tau^\epsilon - \mathbf{S}_\tau^\epsilon\big) = \frac{1}{p} \to 0,$$

where $\|f\| = \sup_x |f(x)|$. Hence, $F^{\tilde{\mathbf{S}}_\tau^*}$ almost surely converges to $F^{\mathbf{S}_\tau^*}$. □

Proof of Lemma 2. Let $s_1(\mathbf{A})$ denote the largest singular value of matrix $\mathbf{A}$. By Corollary 7.3.8 from Horn and Johnson [15] and the strong law of large number, we have

$$\left|s_1(\tilde{\mathbf{S}}_\tau^\epsilon) - s_1(\mathbf{S}_\tau^\epsilon)\right| \leqslant \left\|\tilde{\mathbf{S}}_\tau^\epsilon - \mathbf{S}_\tau^\epsilon\right\|_2 = \|\bar{\epsilon}\bar{\epsilon}^\top\|_2 = \sqrt{\bar{\epsilon}^\top\bar{\epsilon}} \to 0, \quad a.s. \tag{12}$$

By this equation and (7) imply that

$$\left|s_1(\tilde{\mathbf{S}}_\tau^\epsilon) + s_1(\mathbf{S}_\tau^\epsilon)\right| \to 2\sqrt{b}, \quad a.s.$$

Hence,

$$\left|\lambda_{\max}(\tilde{\mathbf{S}}_\tau^*) - \lambda_{\max}(\mathbf{S}_\tau^*)\right| \leqslant \left|s_1(\tilde{\mathbf{S}}_\tau^\epsilon) - s_1(\mathbf{S}_\tau^\epsilon)\right| \cdot \left|s_1(\tilde{\mathbf{S}}_\tau^\epsilon) + s_1(\mathbf{S}_\tau^\epsilon)\right| \to 0, \quad a.s.$$

This completes the proof of Lemma 2. □

Proof of Lemma 3. Recall that $\mathbf{D} = [\mathrm{diag}(\mathbf{S}_0^\epsilon)]^{-1/2}$. Denote $\tilde{\mathbf{D}} = [\mathrm{diag}(\mathbf{S}_0^\epsilon + \bar{\epsilon}\bar{\epsilon}^\top)]^{-1/2}$. By definition, we can write

$$\tilde{\mathbf{R}}_\tau^\epsilon = \tilde{\mathbf{D}}(\mathbf{S}_\tau^\epsilon + \bar{\epsilon}\bar{\epsilon}^\top)\tilde{\mathbf{D}} = \hat{\mathbf{R}}_\tau^\epsilon + \tilde{\mathbf{D}}(\bar{\epsilon}\bar{\epsilon}^\top)\tilde{\mathbf{D}},$$

where $\hat{\mathbf{R}}_\tau^\epsilon = \tilde{\mathbf{D}}\mathbf{S}_\tau^\epsilon\tilde{\mathbf{D}}$. By Theorem A.44 in Bai and Silverstein [3], we have

$$\left\|F^{\tilde{\mathbf{R}}_\tau^*} - F^{\hat{\mathbf{R}}_\tau^\epsilon(\hat{\mathbf{R}}_\tau^\epsilon)^\top}\right\| \leqslant \frac{1}{p}\mathrm{rank}(\tilde{\mathbf{D}}\bar{\boldsymbol{\epsilon}}\bar{\boldsymbol{\epsilon}}^\top\tilde{\mathbf{D}}) = \frac{1}{p} \to 0.$$

It suffices to prove that

$$L\left(F^{\hat{\mathbf{R}}_\tau^\epsilon(\hat{\mathbf{R}}_\tau^\epsilon)^\top}, F^{\mathbf{R}_\tau^\epsilon(\mathbf{R}_\tau^\epsilon)^\top}\right) \to 0, \qquad a.s. \tag{13}$$

By Theorem A.47 in Bai and Silverstein [3], we have

$$L\left(F^{\hat{\mathbf{R}}_\tau^\epsilon(\hat{\mathbf{R}}_\tau^\epsilon)^\top}, F^{\mathbf{R}_\tau^\epsilon(\mathbf{R}_\tau^\epsilon)^\top}\right) \leqslant 2\|\mathbf{R}_\tau^\epsilon\|_2 \cdot \|\mathbf{R}_\tau^\epsilon - \hat{\mathbf{R}}_\tau^\epsilon\|_2 + \|\mathbf{R}_\tau^\epsilon - \hat{\mathbf{R}}_\tau^\epsilon\|_2^2.$$

In view of this inequality, to prove (13), we need to show that $\|\mathbf{R}_\tau^\epsilon\|_2$ is bounded and $\|\mathbf{R}_\tau^\epsilon - \hat{\mathbf{R}}_\tau^\epsilon\|_2$ converges to zero almost surely.

First, we consider $\|\mathbf{R}_\tau^\epsilon\|_2$. By basic norm inequality, we have $\|\mathbf{R}_\tau^\epsilon\|_2 = \|\mathbf{D}\mathbf{S}_\tau^\epsilon\mathbf{D}\|_2 \leqslant \|\mathbf{S}_\tau^\epsilon\|_2 \cdot \|\mathbf{D}\|_2^2$. It follows from (7) and (12) that

$$\|\mathbf{S}_\tau^\epsilon\|_2 \to \sqrt{b}, \quad a.s. \tag{14}$$

By using (6), we have

$$\|\mathbf{D}\|_2 = \max_{1\leqslant j\leqslant p}\left|\frac{\sqrt{n}}{\sqrt{\sum_{i=1}^n \epsilon_{i,j}^2}}\right| \leqslant \max_{1\leqslant j\leqslant p}\left|\frac{\sqrt{n}}{\sqrt{\sum_{i=1}^n \epsilon_{i,j}^2}} - 1\right| + 1 \to 1, \quad a.s. \tag{15}$$

Hence, we conclude that $\|\mathbf{R}_\tau^\epsilon\|_2$ is bounded almost surely.

Next, we consider $\|\mathbf{R}_\tau^\epsilon - \hat{\mathbf{R}}_\tau^\epsilon\|_2$. By definition, we write

$$\|\mathbf{R}_\tau^\epsilon - \hat{\mathbf{R}}_\tau^\epsilon\|_2 = \left\|\mathbf{D}\mathbf{S}_\tau^\epsilon\mathbf{D} - \tilde{\mathbf{D}}\mathbf{S}_\tau^\epsilon\tilde{\mathbf{D}}\right\|_2 = \left\|\mathbf{D}\mathbf{S}_\tau^\epsilon\mathbf{D} - \mathbf{D}\mathbf{S}_\tau^\epsilon\tilde{\mathbf{D}} + \mathbf{D}\mathbf{S}_\tau^\epsilon\tilde{\mathbf{D}} - \tilde{\mathbf{D}}\mathbf{S}_\tau^\epsilon\tilde{\mathbf{D}}\right\|_2 \leqslant (\|\mathbf{D}\|_2 + \|\tilde{\mathbf{D}}\|_2) \cdot \|\mathbf{S}_\tau^\epsilon\|_2 \cdot \|\mathbf{D} - \tilde{\mathbf{D}}\|_2.$$

From the estimation (15) and the fact $\|\tilde{\mathbf{D}}\|_2 \leqslant \|\mathbf{D}\|_2$, we obtain that both $\|\mathbf{D}\|_2$ and $\|\tilde{\mathbf{D}}\|_2$ are bounded almost surely. Note that

$$\begin{aligned}\|\mathbf{D} - \tilde{\mathbf{D}}\|_2 &= \max_{1\leqslant j\leqslant p}\left(\frac{1}{\sqrt{(\mathbf{S}_0^\epsilon)_{jj}}} - \frac{1}{\sqrt{(\mathbf{S}_0^\epsilon)_{jj} + (\bar{\boldsymbol{\epsilon}}\bar{\boldsymbol{\epsilon}}^\top)_{jj}}}\right) \leqslant \max_{1\leqslant j\leqslant p}\left(\frac{\sqrt{(\mathbf{S}_0^\epsilon)_{jj} + (\bar{\boldsymbol{\epsilon}}\bar{\boldsymbol{\epsilon}}^\top)_{jj}} - \sqrt{(\mathbf{S}_0^\epsilon)_{jj}}}{(\mathbf{S}_0^\epsilon)_{jj}}\right)\\ &\leqslant \max_{1\leqslant j\leqslant p}\frac{\sqrt{(\bar{\boldsymbol{\epsilon}}\bar{\boldsymbol{\epsilon}}^\top)_{jj}}}{(\mathbf{S}_0^\epsilon)_{jj}} = \max_{1\leqslant j\leqslant p}\frac{\sqrt{(\sum_{i=1}^n \epsilon_{i,j})^2}}{\sum_{i=1}^n \epsilon_{i,j}^2},\end{aligned}$$

where the second inequality follows from $\sqrt{x+y} \leqslant \sqrt{x} + \sqrt{y}$ for any $x, y \geqslant 0$. This shows that $\|\mathbf{D} - \tilde{\mathbf{D}}\|_2 \to 0$ is equivalent to $\max_{1\leqslant j\leqslant p}[(\sum_{i=1}^n \epsilon_{i,j})/(\sum_{i=1}^n \epsilon_{i,j}^2)]^2 \to 0$. By the triangular inequality, we have

$$\max_{1\leqslant j\leqslant p}\left(\frac{\sum_{i=1}^n \epsilon_{i,j}}{\sum_{i=1}^n \epsilon_{i,j}^2}\right)^2 \leqslant \max_{1\leqslant j\leqslant p}\left|\frac{1}{\sum_{i=1}^n \epsilon_{i,j}^2}\right| + \max_{1\leqslant j\leqslant p}\left|\frac{\sum_{i\neq i'} \epsilon_{i,j}\epsilon_{i',j}}{\left(\sum_{i=1}^n \epsilon_{i,j}^2\right)^2}\right|. \tag{16}$$

From (5), we obtain the first term on the RHS of (16) converges to zero almost surely. By Lemma 2 in Bai and Yin [6], we have

$$\max_{1\leqslant j\leqslant p}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_{i,j}\right|^2 \to 0, \quad a.s., \qquad \max_{1\leqslant j\leqslant p}\left|\frac{1}{n^2}\sum_{i=1}^n \epsilon_{i,j}^2\right| \to 0, \quad a.s.,$$

which imply that

$$\max_{1\leqslant j\leqslant p}\left|\frac{1}{n^2}\sum_{i\neq i'} \epsilon_{i,j}\epsilon_{i',j}\right| \to 0, \quad a.s.$$

This estimation, together with

$$\max_{1\leqslant j\leqslant p}\left|\frac{\frac{1}{n^2}\sum_{i\neq i'} \epsilon_{i,j}\epsilon_{i',j}}{\left(\frac{1}{n}\sum_{i=1}^n \epsilon_{i,j}^2\right)^2} - \frac{1}{n^2}\sum_{i\neq i'} \epsilon_{i,j}\epsilon_{i',j}\right| \leqslant \max_{1\leqslant j\leqslant p}\left|\frac{1}{\left(\frac{1}{n}\sum_{i=1}^n \epsilon_{i,j}^2\right)^2} - 1\right| \cdot \max_{1\leqslant j\leqslant p}\left|\frac{1}{n^2}\sum_{i\neq i'} \epsilon_{i,j}\epsilon_{i',j}\right|$$

and (5), implies that the second term on the RHS of (16) also converges to zero almost surely. Hence, we conclude that $\|\mathbf{R}_\tau^\epsilon - \hat{\mathbf{R}}_\tau^\epsilon\|_2 \to 0\, a.s.$ This completes the proof of Lemma 3. □

Proof of Lemma 4. By Weyl's theorem and the norm inequality $\|\mathbf{A}\mathbf{A}^\top - \mathbf{B}\mathbf{B}^\top\|_2 \leqslant 2\|\mathbf{A}\|_2 \cdot \|\mathbf{A}-\mathbf{B}\|_2 + \|\mathbf{A}-\mathbf{B}\|_2^2$, we have

$$\left|\lambda_{\max}(\mathbf{R}_\tau^*) - \lambda_{\max}(\hat{\mathbf{R}}_\tau^*)\right| \leqslant \left\|\mathbf{R}_\tau^* - \hat{\mathbf{R}}_\tau^*\right\|_2 \leqslant 2\|\mathbf{R}_\tau^\epsilon\|_2 \cdot \|\mathbf{R}_\tau^\epsilon - \hat{\mathbf{R}}_\tau^\epsilon\|_2 + \|\mathbf{R}_\tau^\epsilon - \hat{\mathbf{R}}_\tau^\epsilon\|_2^2. \tag{17}$$

In the proof of Lemma 3, we show that the RHS of this inequality converges to zero almost surely. Thus,

$$\left|\lambda_{\max}(\mathbf{R}_\tau^*) - \lambda_{\max}(\hat{\mathbf{R}}_\tau^*)\right| \to 0, \quad a.s. \tag{18}$$

An argument similar to the one used in (17) shows that

$$\left|\lambda_{\max}(\hat{\mathbf{R}}_\tau^*) - \lambda_{\max}(\tilde{\mathbf{R}}_\tau^*)\right| \leqslant \left\|\hat{\mathbf{R}}_\tau^* - \tilde{\mathbf{R}}_\tau^*\right\|_2 \leqslant 2\|\hat{\mathbf{R}}_\tau^\epsilon\|_2 \cdot \|\hat{\mathbf{R}}_\tau^\epsilon - \tilde{\mathbf{R}}_\tau^\epsilon\|_2 + \|\hat{\mathbf{R}}_\tau^\epsilon - \tilde{\mathbf{R}}_\tau^\epsilon\|_2^2.$$

By definition of $\hat{\mathbf{R}}_\tau^\epsilon$, we obtain

$$\|\hat{\mathbf{R}}_\tau^\epsilon\|_2 \leqslant \|\mathbf{S}_\tau^\epsilon\|_2 \cdot \|\tilde{\mathbf{D}}\|_2^2 \leqslant \|\mathbf{S}_\tau^\epsilon\|_2 \cdot \|\mathbf{D}\|_2^2,$$

which, together with (14) and (15), implies that $\|\hat{\mathbf{R}}_\tau^\epsilon\|_2$ is bounded almost surely. Since $\tilde{\mathbf{R}}_\tau^\epsilon = \hat{\mathbf{R}}_\tau^\epsilon + \tilde{\mathbf{D}}(\bar{\boldsymbol{\epsilon}}\bar{\boldsymbol{\epsilon}}^\top)\tilde{\mathbf{D}}$, we have

$$\|\hat{\mathbf{R}}_\tau^\epsilon - \tilde{\mathbf{R}}_\tau^\epsilon\|_2 \leqslant \|\bar{\boldsymbol{\epsilon}}\bar{\boldsymbol{\epsilon}}^\top\|_2 \cdot \|\tilde{\mathbf{D}}\|_2^2 \leqslant \sqrt{\bar{\boldsymbol{\epsilon}}^\top\bar{\boldsymbol{\epsilon}}} \cdot \|\mathbf{D}\|_2^2.$$

By the strong law of large number and (15), we obtain that $\|\hat{\mathbf{R}}_\tau^\epsilon - \tilde{\mathbf{R}}_\tau^\epsilon\|_2$ converges to zero almost surely. Hence,

$$\left|\lambda_{\max}(\hat{\mathbf{R}}_\tau^*) - \lambda_{\max}(\tilde{\mathbf{R}}_\tau^*)\right| \to 0, \quad a.s.,$$

which, together with (18), completes the proof. □

3.3. Conjecture about the smallest eigenvalue of $\mathbf{R}_\tau^*$

We conjecture that the smallest eigenvalue of $\mathbf{R}_\tau^*$ has similar asymptotic behavior to the largest eigenvalue in Theorem 2. Specifically, we conjecture that, under the same assumptions as in Theorem 2, $\lambda_{\min}(\mathbf{R}_\tau^*)$, namely

$$\lambda_{\min}(\mathbf{R}_\tau^*) = \begin{cases} \text{the smallest eigenvalue of } \mathbf{R}_\tau^*, & \text{if } p \leqslant n, \\ \text{the } (p-n+1)\text{th smallest eigenvalue of } \mathbf{R}_\tau^*, & \text{if } p > n, \end{cases}$$

will converge to the left edge of the support of the LSD of $\mathbf{R}_\tau^*$.

First of all, the standardization procedure applies for $\lambda_{\min}(\mathbf{R}_\tau^*)$. Lemmas 2 and 4 can be extended to the smallest eigenvalue case, which means that we can turn our attention to the non-centered sample auto-correlation matrix $\lambda_{\min}(\tilde{\mathbf{R}}_\tau^*)$. Secondly, $\lambda_{\min}(\tilde{\mathbf{R}}_\tau^*)$ shares the same limit with that of the non-centered sample auto-covariance matrix $\lambda_{\min}(\tilde{\mathbf{S}}_\tau^*)$. By Corollary 7.3.8 from Horn and Johnson [15], we have a stronger result than (9):

$$\max_{1\leqslant i\leqslant p}\left|\sqrt{\lambda_i(\tilde{\mathbf{R}}_\tau^*)} - \sqrt{\lambda_i(\tilde{\mathbf{S}}_\tau^*)}\right| \leqslant \|\tilde{\mathbf{R}}_\tau^\epsilon - \tilde{\mathbf{S}}_\tau^\epsilon\|_2,$$

which, together with (7), (10) and (11), implies that

$$\left|\sqrt{\lambda_{\min}(\tilde{\mathbf{R}}_\tau^*)} - \sqrt{\lambda_{\min}(\tilde{\mathbf{S}}_\tau^*)}\right| \to 0, \quad a.s.,$$

where and $\lambda_{\min}(\tilde{\mathbf{S}}_\tau^*)$ is defined in the same way. This together with (7) shows that $\left|\lambda_{\min}^{1/2}(\tilde{\mathbf{R}}_\tau^*) + \lambda_{\min}^{1/2}(\tilde{\mathbf{S}}_\tau^*)\right|$ is bounded almost surely. Hence,

$$\left|\lambda_{\min}(\tilde{\mathbf{R}}_\tau^*) - \lambda_{\min}(\tilde{\mathbf{S}}_\tau^*)\right| = \left|\sqrt{\lambda_{\min}(\tilde{\mathbf{R}}_\tau^*)} - \sqrt{\lambda_{\min}(\tilde{\mathbf{S}}_\tau^*)}\right| \cdot \left|\sqrt{\lambda_{\min}(\tilde{\mathbf{R}}_\tau^*)} + \sqrt{\lambda_{\min}(\tilde{\mathbf{S}}_\tau^*)}\right| \to 0, \quad a.s.$$

By estimations above, we conclude that

$$\left|\lambda_{\min}(\mathbf{R}_\tau^*) - \lambda_{\min}(\mathbf{S}_\tau^*)\right| \to 0, \quad a.s.,$$

that is, $\lambda_{\min}(\mathbf{R}_\tau^*)$ and $\lambda_{\min}(\mathbf{S}_\tau^*)$ share the same limit.

To the best of our knowledge, the asymptotic limit of $\lambda_{\min}(\mathbf{S}_\tau^*)$ has not been rigorously derived in the current literature. In the spirit of the well-known Bai–Yin Law (Bai and Yin [6], the smallest and largest eigenvalues of large-dimensional sample covariance matrix converge to the left and right endpoint of the Marcĕnko–Pastur law, respectively), we conjecture that $\lambda_{\min}(\mathbf{S}_\tau^*)$ (also $\lambda_{\min}(\mathbf{R}_\tau^*)$) will converge to the left endpoint of the support of LSD of $\mathbf{S}_\tau^*$, which is 0 for $0 < y < 1$ and $(-1+20y+8y^2-(1+8y)^{3/2})/8$ for $y \geqslant 1$. Fig. 4 compares the boxplot of smallest singular value with its corresponding limit, i.e. the left endpoint of the LSD. Empirical evidence in Fig. 4 convinces us of the correctness of this conjecture. The potential route to prove this result is through the moment method used in Bai and Yin [6] and Wang and Yao [26]. We will leave it for future study.

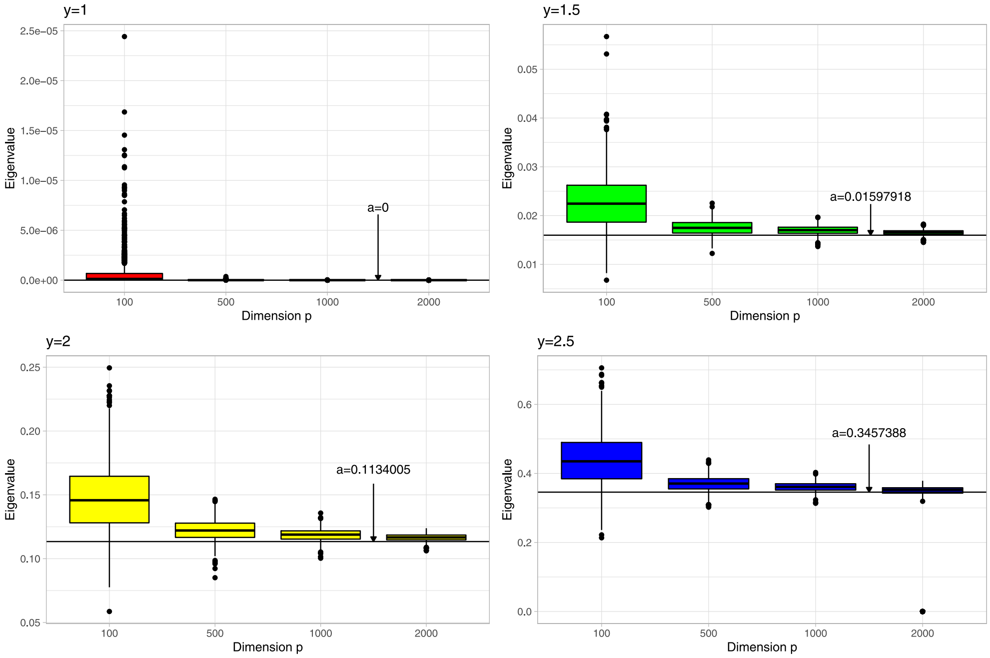

Fig. 4. Boxplot of the smallest eigenvalue of $\mathbf{R}_1^* = \mathbf{R}_1^\epsilon(\mathbf{R}_1^\epsilon)^\top$ based on 1000 Gaussian samples. In all panels, the horizontal line indicates the left endpoint of its LSD, and the ratio of the dimension to the sample size is set as $y \in \{1, 1.5, 2, 2.5\}$ from top to bottom and left to right, respectively.

4. Application

4.1. Estimation of total number of factors

High-dimensional factor models have met a large success in data analysis across many scientific fields such as psychology, economics and signal processing, to name a few. Their appeal mainly relies on their capability to reduce the generally high dimensionality of data to much lower-dimensional common factor components. Determining total number of factors is a central problem in high dimensional factor modeling.

Consider a factor model for high-dimensional time series proposed by Lam and Yao [18]: for $1 \leqslant i \leqslant n$, let $\mathbf{y}_i$ denote the p-dimensional vector observed at time i. It consists of two parts, a low-dimensional common latent factor sequence $\mathbf{f}_i$ and an idiosyncratic component $\boldsymbol{\epsilon}_i$:

$$\mathbf{y}_i = \mathbf{B}\mathbf{f}_i + \boldsymbol{\epsilon}_i, \quad i \in \{1, 2, \dots, n\}. \tag{19}$$

Here $\mathbf{B}$ is a $p \times k$ factor loading matrix satisfying $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_k$, $\{\boldsymbol{\epsilon}_i\}$ is a p-dimensional white noise with $\mathbb{E}(\boldsymbol{\epsilon}_i) = 0$, $\mathrm{Var}(\boldsymbol{\epsilon}_i) = \sigma^2\mathbf{I}_p$. The k-dimensional factor sequence $\{\mathbf{f}_i\}$ is a time sequence which is temporally correlated.

Our goal is to develop an estimator of the number of factors k in the model (19) using sample auto-correlation matrix of the observed sequence $\{\mathbf{y}_i,\ 1 \le i \le n\}$. In the current literature, there are some other estimators of k using different sample covariance/correlation matrices . To name a few, Li et al. [20] proposed an estimator based on eigenvalues of the lag-1 sample auto-covariance matrix $\mathbf{S}_1^{\mathbf{y}}$. Denote the eigenvalue ratios (ER) between consecutive eigenvalues of $\mathbf{M} = \mathbf{S}_1^{\mathbf{y}}(\mathbf{S}_1^{\mathbf{y}})^\top$ as

$$\theta_j = \lambda_{j+1}(\mathbf{M})/\lambda_j(\mathbf{M}), \qquad j \in \{1, 2, \dots, p-1\},$$

Li et al. [20] proposed the following estimator:

$$\hat{k}_{\mathrm{ER_{acov}}} = \left\{\text{first } j \geqslant 1 \text{ such that } \theta_j > 1 - d_n\right\} - 1,$$

where $0 < d_n < 1$ is a positive constant. They proved that $\hat{k}_{\mathrm{ER_{acov}}}$ is a consistent estimator for k under certain mild conditions. It is possible that two consecutive spike eigenvalues are very close to each other and their ratio exceeds the threshold $1 - d_n$, which may lead to underestimation. For the sake of robustness, Li et al. [20] also proposed an equivalent reinforced estimator

$$\hat{k}^*_{\mathrm{ER_{acov}}} = \left\{\text{first } j \geqslant 1 \text{ such that } \theta_j > 1 - d_n \text{ and } \theta_{j+1} > 1 - d_n\right\} - 1. \tag{20}$$

More recently, Fan et al. [12] developed a tuning-free scale-invariant adjusted correlation thresholding (ACT) method based on the sample correlation matrix $\mathbf{R}_0^{\mathbf{y}}$. For any given $1 \leqslant j \leqslant p$, define

$$m_{n,j}(z) = \frac{1}{p-j}\left[\sum_{\ell=j+1}^{p} \frac{1}{\lambda_\ell(\mathbf{R}_0^{\mathbf{y}}) - z} + \frac{1}{(3\lambda_j(\mathbf{R}_0^{\mathbf{y}}) + \lambda_{j+1}(\mathbf{R}_0^{\mathbf{y}}))/4 - z}\right],$$

$$\underline{m}_{n,j}(z) = -(1 - c_{j,n-1})z^{-1} + c_{j,n-1} m_{n,j}(z), \qquad \lambda_j^C(\mathbf{R}_0^{\mathbf{y}}) = -1/\underline{m}_{n,j}\big(\lambda_j(\mathbf{R}_0^{\mathbf{y}})\big),$$

where $c_{j,n-1} = (p-j)/(n-1)$. Fan et al. [12] developed the following estimator of total number of factors:

$$\hat{k}_{\mathrm{ACT}} = \max\Big\{j : \lambda_j^C(\mathbf{R}_0^{\mathbf{y}}) > 1 + \sqrt{p/(n-1)}\Big\}. \tag{21}$$

Now we aim to develop an estimator of k using sample auto-correlation matrix. As mentioned in the introduction, the sample auto-correlation matrix $\mathbf{R}_\tau^{\mathbf{y}}$ of $\mathbf{y}_i$ is a finite rank perturbation of $\mathbf{R}_\tau^{\epsilon}$ of $\boldsymbol{\epsilon}_i$. Take $\tau = 1$, then we can use the eigenvalues of $\tilde{\mathbf{M}} = \mathbf{R}_1^{\mathbf{y}}(\mathbf{R}_1^{\mathbf{y}})^\top$ to estimate the number of factors k. The basic idea is to count the number of spike eigenvalues of $\tilde{\mathbf{M}}$ which are larger than the bulk eigenvalues of $\mathbf{R}_1^{\epsilon}(\mathbf{R}_1^{\epsilon})^\top$. Specifically, take the canonical form where $\mathbf{B} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{0}_{p-k} \end{pmatrix}$ as an example. Since

$$\begin{aligned}
\mathbf{S}_0^{\mathbf{y}} &= \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top = \mathbf{B}\left(\frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{f}_i - \bar{\mathbf{f}})(\mathbf{f}_i - \bar{\mathbf{f}})^\top\right)\mathbf{B}^\top + \mathbf{B}\left(\frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{f}_i - \bar{\mathbf{f}})(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})^\top\right) \\
&\quad + \left(\frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})(\mathbf{f}_i - \bar{\mathbf{f}})^\top\right)\mathbf{B}^\top + \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})^\top \triangleq \mathbf{P}_0^{\mathbf{B}} + \mathbf{S}_0^{\epsilon},
\end{aligned}$$

only the first k diagonal elements of $\mathrm{diag}(\mathbf{P}_0^{\mathbf{B}})$ are non-zero. Then $\mathrm{diag}(\mathbf{P}_0^{\mathbf{B}}) = \mathrm{diag}(\mathbf{S}_0^{\mathbf{y}}) - \mathrm{diag}(\mathbf{S}_0^{\epsilon})$ is a rank-k diagonal matrix. This implies that the difference $\boldsymbol{\Delta} := \mathbf{D}_{\mathbf{y}} - \mathbf{D}_{\epsilon}$ has finite rank k, where $\mathbf{D}_{\epsilon} = [\mathrm{diag}(\mathbf{S}_0^{\epsilon})]^{-1/2}$ and $\mathbf{D}_{\mathbf{y}} = [\mathrm{diag}(\mathbf{S}_0^{\mathbf{y}})]^{-1/2}$. Similarly, $\mathbf{S}_1^{\mathbf{y}} = \mathbf{P}_1^{\mathbf{B}} + \mathbf{S}_1^{\epsilon}$ and only the first k diagonal elements of $\mathrm{diag}(\mathbf{P}_1^{\mathbf{B}})$ are non-zero. Therefore, the lag-1 sample auto-correlation matrix

$$\mathbf{R}_1^{\mathbf{y}} = \mathbf{D}_{\mathbf{y}}\mathbf{P}_1^{\mathbf{B}}\mathbf{D}_{\mathbf{y}} + (\mathbf{D}_{\epsilon} + \boldsymbol{\Delta})\mathbf{S}_1^{\epsilon}(\mathbf{D}_{\epsilon} + \boldsymbol{\Delta}) = \underbrace{\mathbf{D}_{\mathbf{y}}\mathbf{P}_1^{\mathbf{B}}\mathbf{D}_{\mathbf{y}} + \mathbf{D}_{\epsilon}\mathbf{S}_1^{\epsilon}\boldsymbol{\Delta} + \boldsymbol{\Delta}\mathbf{S}_1^{\epsilon}\mathbf{D}_{\epsilon} + \boldsymbol{\Delta}\mathbf{S}_1^{\epsilon}\boldsymbol{\Delta}}_{\text{finite rank}} + \mathbf{D}_{\epsilon}\mathbf{S}_1^{\epsilon}\mathbf{D}_{\epsilon},$$

is a finite rank perturbation of $\mathbf{R}_1^{\epsilon} = \mathbf{D}_{\epsilon}\mathbf{S}_1^{\epsilon}\mathbf{D}_{\epsilon}$. Thus $\tilde{\mathbf{M}} = \mathbf{R}_1^{\mathbf{y}}(\mathbf{R}_1^{\mathbf{y}})^\top$ and $\mathbf{R}_1^{*} = \mathbf{R}_1^{\epsilon}(\mathbf{R}_1^{\epsilon})^\top$ share the same LSD while asymptotically $\tilde{\mathbf{M}}$ has k extra spike eigenvalues. We can use the right endpoint of this LSD as a threshold for filtering these spike eigenvalues. The right endpoint is determined by our Theorem 2.

In this way, we introduce our estimator of total number of factors:

$$\hat{k}_{\mathrm{RE}} = \max\Big\{j : \lambda_j(\mathbf{R}_1^{*}) > b_{c_n} + h_n\Big\}, \tag{22}$$

where $b_{c_n} = \big(-1 + 20c_n + 8c_n^2 + (1 + 8c_n)^{3/2}\big)/8$ (see Theorems 1 and 2) with $c_n = p/n$ and $h_n > 0$ is a tuning parameter dealing with the fluctuation of the largest bulk (non-spike) eigenvalue. The subscript of $\hat{k}_{\mathrm{RE}}$ represents the abbreviation of Right Endpoint.

Remark 3 (*Calibration of Tuning Parameter h_n*). Practically, we need to choose an appropriate tuning parameter h_n. It is very likely that the asymptotic distribution of $n^{2/3}(\lambda_{k+1}(\tilde{\mathbf{M}}) - b_{c_n})$ is the same as that of $n^{2/3}(\lambda_1(\mathbf{R}_1^{*}) - b_{c_n})$. Using the similarity, we propose an *a priori* calibration of h_n:

(i) For any given pair (p, n), the empirical distribution of $n^{2/3}(\lambda_1(\mathbf{R}_1^{*}) - b_{c_n})$ is obtained by sampling a large number (say 2000) of independent replications of standard Gaussian vectors $\boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \mathbf{I}_p)$. Its lower 99.5% quantile $q_{p,n,99.5\%}$ is obtained from this empirical distribution.

(ii) Using the approximation

$$\Pr\Big\{n^{2/3}(\lambda_{k+1}(\tilde{\mathbf{M}}) - b_{c_n}) \leqslant q_{p,n,99.5\%}\Big\} \approx \Pr\Big\{n^{2/3}(\lambda_1(\mathbf{R}_1^{*}) - b_{c_n}) \leqslant q_{p,n,99.5\%}\Big\} = 99.5\%,$$

we calibrate h_n at the value $h_n = n^{-2/3} \cdot q_{p,n,99.5\%}$.

This tuned value of h_n is used for all given pairs of (p, n) in the simulation studies in Section 4.2.

Moreover, we propose another ratio-based estimator of total number of factors k. Denote

$$\tilde{\theta}_j = \lambda_{j+1}(\tilde{\mathbf{M}})/\lambda_j(\tilde{\mathbf{M}}), \qquad j \in \{1, 2, \ldots, p-1\}.$$

as the sequence of ratios between consecutive eigenvalues of $\tilde{\mathbf{M}}$, we propose

$$\hat{k}_{\mathrm{ER_{acor}}} = \Big\{\text{first } j \geqslant 1 \text{ such that } \tilde{\theta}_j > 1 - \tilde{d}_n\Big\} - 1,$$

where $0 < \tilde{d}_n < 1$ is a positive constant. The subscript of $\hat{k}_{\mathrm{ER_{acor}}}$ represents the abbreviation of Eigenvalue Ratio of sample auto-correlation matrix. A robust version of this estimator is given by

$$\hat{k}^*_{\mathrm{ER_{acor}}} = \left\{\text{first } j \geqslant 1 \text{ such that } \tilde{\theta}_j > 1 - \tilde{d}_n \text{ and } \tilde{\theta}_{j+1} > 1 - \tilde{d}_n\right\} - 1. \tag{23}$$

Remark 4 (*Calibration of the Tuning Parameter d_n and $\tilde{d}_n$*). Similarly with $\hat{k}_{\mathrm{RE}}$, we need to choose tuning parameters d_n and $\tilde{d}_n$. $\tilde{d}_n$ is the same with d_n and we take the calibration of d_n as an example. We use the asymptotic distribution of $n^{2/3}\left(\frac{\lambda_2(\mathbf{M}_\epsilon)}{\lambda_1(\mathbf{M}_\epsilon)} - 1\right)$ to approximate that of $n^{2/3}\left(\frac{\lambda_{k+2}(\mathbf{M})}{\lambda_{k+1}(\mathbf{M})} - 1\right)$, where $\mathbf{M}_\epsilon = \mathbf{S}_1^\epsilon(\mathbf{S}_1^\epsilon)^\top$. Specifically,

1. For any given pair (p, n), the empirical distribution of $n^{2/3}\left(\frac{\lambda_2(\mathbf{M}_\epsilon)}{\lambda_1(\mathbf{M}_\epsilon)} - 1\right)$ is obtained by sampling a large number (say 2000) of independent replications of standard Gaussian vectors $\boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \mathbf{I}_p)$. Its lower 0.5% quantile $\tilde{q}_{p,n,0.5\%}$ is obtained from this empirical distribution.
2. Using the approximation
$$\Pr\left\{n^{2/3}\left(\frac{\lambda_{k+2}(\mathbf{M})}{\lambda_{k+1}(\mathbf{M})} - 1\right) \leqslant \tilde{q}_{p,n,0.5\%}\right\} \approx \Pr\left\{n^{2/3}\left(\frac{\lambda_{k+2}(\mathbf{M}_\epsilon)}{\lambda_{k+1}(\mathbf{M}_\epsilon)} - 1\right) \leqslant \tilde{q}_{p,n,0.5\%}\right\} = 0.5\%,$$
we calibrate d_n at the value $d_n = n^{-2/3} \cdot |\tilde{q}_{p,n,0.5\%}|$.

This tuned value of d_n is used for all given pairs of (p, n) in the simulation studies in Section 4.2.

4.2. Numerical performance

In this section, we conduct some simulation experiments to examine the finite-sample performance of our estimators $\hat{k}_{\mathrm{RE}}$ and $\hat{k}^*_{\mathrm{ER_{acor}}}$ defined in (22) and (23). We compare these two estimators with two other estimators in the current literature, $\hat{k}^*_{\mathrm{ER_{acov}}}$ from Li et al. [20] and $\hat{k}_{\mathrm{ACT}}$ from Fan et al. [12], which are introduced in (20) and (21) respectively.

We adopt the same simulation settings as in Lam and Yao [18] and Li et al. [20] where the p-dimensional random vectors $\{\mathbf{y}_i, 1 \le i \le n\}$ are generated by

$$\begin{aligned} \mathbf{y}_i &= \mathbf{B}\mathbf{f}_i + \boldsymbol{\epsilon}_i, \qquad \boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \mathbf{I}_p), \\ \mathbf{f}_i &= \boldsymbol{\Theta}\mathbf{f}_{i-1} + \mathbf{e}_t, \qquad \mathbf{e}_i \sim N_k(\mathbf{0}, \boldsymbol{\Gamma}). \end{aligned} \tag{24}$$

Here $\mathbf{B}$ is a $p \times k$ factor loading matrix which can be decomposed as $\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are matrices of sizes $p \times k$ and $k \times k$ with orthogonal unit columns and $\mathbf{D}$ is a $k \times k$ diagonal matrix. We consider the following two scenarios:

(I) $k = 2$:

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \boldsymbol{\Theta} = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.8 \end{pmatrix}, \qquad \boldsymbol{\Gamma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(II) $k = 3$:

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\Theta} = \begin{pmatrix} 0.7 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.9 \end{pmatrix}, \qquad \boldsymbol{\Gamma} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The dimensions and sample sizes are taken to be $p \in \{100, 200, 400, 600\}$ and $n = \{0.5p, 2p\}$. For each scenario, we report the empirical percentages of true estimation ($\hat{k} = k$), underestimation ($\hat{k} < k$) and overestimation ($\hat{k} > k$) of the number of factors k based on 1000 replications in Tables 1 and 2. As shown in Tables 1 and 2, our newly proposed estimators $\hat{k}^*_{\mathrm{ER_{acor}}}$ and $\hat{k}_{\mathrm{RE}}$ have comparable performance with $\hat{k}^*_{\mathrm{ER_{acov}}}$ and $\hat{k}_{\mathrm{ACT}}$. All estimators are consistent and converge quickly when $n = 2p$ and slower when $n = p/2$.

CRediT authorship contribution statement

Zhanting Long: Software, Validation, Writing – original draft, Data curation, Formal analysis, Investigation. **Zeng Li:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Ruitao Lin:** Conceptualization, Project administration. **Jiaxin Qiu:** Validation, Writing – original draft, Formal analysis, Investigation.

Acknowledgments

We thank the Editor, Associate Editor and referees for careful reading and insightful comments. Zeng Li's research is supported by NSFC (National Natural Science Foundation of China) Grant No. 12101292, NSFC, China Grant No. 12031005 and Shenzhen Fundamental Research Program JCYJ20220818100602005. Ruitao Lin's research is partially supported by the NIH, United States grant 1R01CA261978.

Table 1
Comparison of four estimators $\hat{k}_{ACT}$, $\hat{k}^*_{ER_{acov}}$, $\hat{k}^*_{ER_{acor}}$ and $\hat{k}_{RE}$ in the factor model (24). Data is generated following scenarios (I) with true number of factors $k = 2$. This table reports the empirical percentages (%) of accurate estimation ($\hat{k} = k$), underestimation ($\hat{k} < k$) and overestimation ($\hat{k} > k$) of four estimators based on 1000 replications. Accuracy rates are highlighted in bold letters.

p	Percentages	$n = 2p$				$n = p/2$			
		ACT	ER^*_{acov}	ER^*_{acor}	RE	ACT	ER^*_{acov}	ER^*_{acor}	RE
100	$\hat{k} = k$	**98.8**	**98.4**	**98**	**97.3**	**27.4**	**22.3**	**20.2**	**11.1**
	$\hat{k} < k$	0.6	0.9	0.9	2.7	72.5	77.6	79.8	88.9
	$\hat{k} > k$	0.6	0.7	1.1	0	0.1	0.1	0	0
200	$\hat{k} = k$	**97.9**	**99.5**	**99.3**	**100**	**50.3**	**52.8**	**50.3**	**36.8**
	$\hat{k} < k$	0	0	0	0	48.4	47	49.4	63.2
	$\hat{k} > k$	2.1	0.5	0.7	0	1.3	0.2	0.3	0
400	$\hat{k} = k$	**96.5**	**100**	**99.7**	**100**	**75.1**	**82.4**	**81.6**	**75.3**
	$\hat{k} < k$	0	0	0	0	20.9	16.9	17.9	24.7
	$\hat{k} > k$	3.5	0	0.3	0	4	0.7	0.5	0
600	$\hat{k} = k$	**96.6**	**98.9**	**99.3**	**100**	**87.1**	**93.2**	**92.9**	**92.2**
	$\hat{k} < k$	0	0	0	0	9.4	6.5	6.8	7.8
	$\hat{k} > k$	3.4	1.1	0.7	0	3.5	0.3	0.3	0

Table 2
Comparison of four estimators $\hat{k}_{ACT}$, $\hat{k}^*_{ER_{acov}}$, $\hat{k}^*_{ER_{acor}}$ and $\hat{k}_{RE}$ in the factor model (24). Data is generated following scenarios (II) with true number of factors $k = 3$. This table reports the empirical percentages (%) of true estimation ($\hat{k} = k$), underestimation ($\hat{k} < k$) and overestimation ($\hat{k} > k$) of four estimators based on 1000 replications. Accuracy rates are highlighted in bold letters.

p	Percentages	$n = 2p$				$n = p/2$			
		ACT	ER^*_{acov}	ER^*_{acor}	RE	ACT	ER^*_{acov}	ER^*_{acor}	RE
100	$\hat{k} = k$	**99.9**	**99.2**	**98.5**	**99.7**	**32.5**	**32**	**30.8**	**8**
	$\hat{k} < k$	0.1	0	0	0.3	67.5	66.7	68.1	92
	$\hat{k} > k$	0	0.8	1.5	0	0	1.3	1.1	0
200	$\hat{k} = k$	**100**	**99.4**	**99.4**	**100**	**77.3**	**80.2**	**79**	**56.8**
	$\hat{k} < k$	0	0	0	0	22	18.9	19.7	43.2
	$\hat{k} > k$	0	0.6	0.6	0	0.7	0.9	1.3	0
400	$\hat{k} = k$	**99.8**	**99.3**	**98.8**	**100**	**97**	**98.9**	**98.1**	**96.8**
	$\hat{k} < k$	0	0	0	0	1.2	0.8	1	3.2
	$\hat{k} > k$	0.2	0.7	1.2	0	1.8	0.3	0.9	0
600	$\hat{k} = k$	**98.7**	**99.5**	**99.3**	**100**	**98.2**	**98.5**	**99**	**99.6**
	$\hat{k} < k$	0	0	0	0	0.3	0.1	0.1	0.4
	$\hat{k} > k$	1.3	0.5	0.7	0	1.5	1.4	0.9	0

References

[1] Z. Bai, B. Miao, B. Jin, On limit theorem for the eigenvalues of product of two random matrices, J. Multivariate Anal. 98 (1) (2007) 76–101.
[2] J. Bai, S. Ng, Determining the number of factors in approximate factor models, Econometrica 70 (1) (2002) 191–221.
[3] Z. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, second ed., Springer, 2010.
[4] Z. Bai, J. Yao, Central limit theorems for eigenvalues in a spiked population model, Ann. L'IHP Probab. Stat. 44 (2008) 447–474.
[5] Z. Bai, Y. Yin, Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix, Ann. Probab. (1988) 1729–1741.
[6] Z. Bai, Y. Yin, Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, Ann. Probab. (1993) 1275–1294.
[7] Z. Bai, Y. Yin, P. Krishnaiah, On limiting spectral distribution of product of two random matrices when the underlying distribution is isotropic, J. Multivariate Anal. 19 (1) (1986) 189–200.
[8] Z. Bai, Y. Yin, P. Krishnaiah, On the limiting empirical distribution function of the eigenvalues of a multivariate f matrix, Theory Probab. Appl. 32 (3) (1988) 490–500.
[9] J. Baik, J.W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, J. Multivariate Anal. 97 (6) (2006) 1382–1408.
[10] F. Benaych-Georges, R.R. Nadakuditi, The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices, Adv. Math. 227 (1) (2011) 494–521.
[11] A. Bose, J. Mitra, Limiting spectral distribution of a special circulant, Statist. Probab. Lett. 60 (1) (2002) 111–120.
[12] J. Fan, J. Guo, S. Zheng, Estimating number of factors by adjusted eigenvalues thresholding, J. Amer. Statist. Assoc. (2020) 1–10.
[13] S. Geman, A limit theorem for the norm of random matrices, Ann. Probab. (1980) 252–261.
[14] M. Hallin, R. Liška, Determining the number of factors in the general dynamic factor model, J. Amer. Statist. Assoc. 102 (478) (2007) 603–617.

[15] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 2012.
[16] T. Jiang, The limiting distributions of eigenvalues of sample correlation matrices, Sankhyā: Indian J. Stat. (2004) 35–48.
[17] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. (2001) 295–327.
[18] C. Lam, Q. Yao, Factor modeling for high-dimensional time series: inference for the number of factors, Ann. Statist. 40 (2) (2012) 694–726.
[19] Z. Li, G. Pan, J. Yao, On singular value distribution of large-dimensional autocovariance matrices, J. Multivariate Anal. 137 (2015) 119–140.
[20] Z. Li, Q. Wang, J. Yao, Identifying the number of factors from singular values of a large sample auto-covariance matrix, Ann. Statist. 45 (1) (2017) 257–288.
[21] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues for some sets of random matrices, Math. USSR-Sb. 1 (4) (1967) 457.
[22] J.W. Silverstein, The limiting eigenvalue distribution of a multivariate f matrix, SIAM J. Math. Anal. 16 (3) (1985) 641–646.
[23] J.W. Silverstein, Z. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal. 54 (2) (1995) 175–192.
[24] V.H. Vu, Spectral norm of random matrices, Combinatorica 27 (6) (2007) 721–736.
[25] K.W. Wachter, The limiting empirical measure of multiple discriminant ratios, Ann. Statist. 8 (5) (1980) 937–957.
[26] Q. Wang, J. Yao, Moment approach for singular values distribution of a large auto-covariance matrix, Ann. L'Inst. Henri Poincaré, Probab. Stat. 52 (4) (2016) 1641–1666.
[27] E.P. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, Ann. of Math. 62 (3) (1955) 548–564.
[28] E.P. Wigner, On the distribution of the roots of certain symmetric matrices, Ann. of Math. 67 (2) (1958) 325–327.
[29] Y. Yin, Limiting spectral distribution for a class of random matrices, J. Multivariate Anal. 20 (1) (1986) 50–68.
[30] Y. Yin, Z. Bai, P. Krishnaiah, Limiting behavior of the eigenvalues of a multivariate f matrix, J. Multivariate Anal. 13 (4) (1983) 508–516.
[31] Y. Yin, Z. Bai, P. Krishnaiah, On the limit of the largest eigenvalue of the large dimensional sample covariance matrix, Probab. Theory Related Fields 78 (4) (1988) 509–521.
[32] Y. Yin, P. Krishnaiah, A limit theorem for the eigenvalues of product of two random matrices, J. Multivariate Anal. 13 (4) (1983) 489–507.