Statistica Sinica **32** (2022), 2007-2022
doi:https://doi.org/10.5705/ss.202020.0262

# SPECTRAL PROPERTIES OF RESCALED SAMPLE CORRELATION MATRIX

Yanqing Yin, Changcheng Li, Guo-Liang Tian and Shurong Zheng

*Chongqing University, Dalian University of Technology, Southern University of Science and Technology and Northeast Normal University*

*Abstract:* Under the high-dimensional setting that the data dimension and sample size tend to infinity proportionally, we derive the limiting spectral distribution and establish the central limit theorem of the eigenvalue statistics of rescaled sample correlation matrices. In contrast to the existing literature, our proposed spectral properties do not require the Gaussian distribution assumption or the assumption that the population correlation matrix is equal to an identity matrix. The asymptotic mean and variance-covariance in our proposed central limit theorem can be expressed as one-dimensional or two-dimensional contour integrals on a unit circle centered at the origin. Not only is the established central limit theorem of the eigenvalue statistics of the rescaled sample correlation matrices very different to that of sample covariance matrices, it also differs from that of sample correlation matrices with a population correlation matrix equal to an identity matrix. Moreover, to illustrate the spectral properties, we propose three test statistics for the hypothesis testing problem of whether the population correlation matrix is equal to a given matrix. Furthermore, we conduct extensive simulation studies to investigate the performance of our proposed testing procedures.

*Key words and phrases:* Central limit theorem, limiting spectral distribution, random matrix theory, rescaled sample correlation matrix.

## 1. Introduction

With the rapid development of computer science, it is possible to collect, store, and analyze high-dimensional data sets. However, the classical statistical tools often are invalid when presented with such data. The high-dimensional sample correlation matrix is an important random matrix for principal component analysis, factor analysis, and human brain image analysis, among others. Let the sample $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of size $n$ be from a $p$-dimensional population $\mathbf{y}$ with unknown mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and correlation matrix

$$\mathbf{R} = [\mathrm{diag}(\boldsymbol{\Sigma})]^{-1/2}\boldsymbol{\Sigma}[\mathrm{diag}(\boldsymbol{\Sigma})]^{-1/2},$$

Corresponding author: Shurong Zheng, School of Mathematics and Statistics and KLAS, Northeast Normal University, Changchun, Jilin, China. E-mail: zhengsr@nenu.edu.cn.

where $\operatorname{diag}(\mathbf{\Sigma})$ is a diagonal matrix formed by the diagonal elements of $\mathbf{\Sigma}$. The sample covariance matrix and sample correlation matrix are defined respectively as

$$\mathbf{S}_n = (n-1)^{-1}\sum_{i=1}^{n}(\mathbf{y}_i-\bar{\mathbf{y}})(\mathbf{y}_i-\bar{\mathbf{y}})^T,\ \hat{\mathbf{R}}_n = [\operatorname{diag}(\mathbf{S}_n)]^{-1/2}\mathbf{S}_n[\operatorname{diag}(\mathbf{S}_n)]^{-1/2},\ (1.1)$$

with the sample mean $\bar{\mathbf{y}} = n^{-1}\sum_{i=1}^{n}\mathbf{y}_i$.

### 1.1. Existing literature

Many works have studied the spectral properties of the high-dimensional sample covariance matrix $\mathbf{S}_n$; see, for example, Marcenko and Pastur (1967), Silverstein and Choi (1995), Bai and Silverstein (2004), and Zheng, Bai and Yao (2015). However, Gao et al. (2017) showed that the central limit theorem (CLT) of the eigenvalue statistics of the sample correlation matrix $\widehat{\mathbf{R}}_n$ differ from those of the sample covariance matrix $\mathbf{S}_n$, although for $\mathbf{R} = \mathbf{I}_p$. Kullback (1967) found that the CLTs of the eigenvalue statistics of rescaled sample correlation matrices $\widehat{\mathbf{R}}_n\mathbf{R}^{-1}$ are also very different for $\mathbf{R} = \mathbf{I}_p$ and $\mathbf{R} \neq \mathbf{I}_p$. Moreover, Fan, Guo and Zheng (2020) showed that using a sample correlation matrix leads to significant advantages over using a sample covariance matrix in some statistical cases. These facts show that studying high-dimensional sample correlation matrices is important and necessary for random matrix theory.

Under the high-dimensional setting that $p/n \to \rho \in (0,\infty)$, for sample correlation matrices $\widehat{\mathbf{R}}_n$, researchers are often interested in studying the *limiting spectral distribution* (LSD) of $\hat{F}_n(x)$ and the CLT of the eigenvalue statistics $\hat{L}_g$, as follows:

$$\hat{F}_n(x) = p^{-1}\sum_{j=1}^{p}\delta(\lambda_j^{\widehat{\mathbf{R}}_n} \leq x), \quad \hat{L}_g = \sum_{j=1}^{p} g(\lambda_j^{\widehat{\mathbf{R}}_n}),$$

where $\{\lambda_j^{\widehat{\mathbf{R}}_n}, j = 1,\ldots,p\}$ are the eigenvalues of $\widehat{\mathbf{R}}_n$, $\delta(\cdot)$ is an indicator function, and $g(\cdot)$ is an analytic function. For the LSD of $\widehat{\mathbf{R}}_n$, Jiang (2004a) obtained the M–P law of $\widehat{\mathbf{R}}_n$ under the assumption that $\mathbf{R} = \mathbf{I}_p$. Karoui (2009) derived the LSD of $\widehat{\mathbf{R}}_n$ under the elliptical population assumption. For the CLT of $\widehat{\mathbf{R}}_n$, Gao et al. (2017) derived the CLT of the linear spectral statistics for $\mathbf{R} = \mathbf{I}_p$. Mestre and Vallet (2017) derived the CLT of the linear spectral statistics under a Gaussian population with a zero mean vector. Jiang (2019) derived the CLT of $\sum_{j=1}^{p}\log(\lambda_j^{\widehat{\mathbf{R}}_n})$ under Gaussian populations.

Recently, Morales-Jimenez et al. (2019) investigated the asymptotics of the eigenstructure of sample correlation matrices under high-dimensional spiked models; see Aitkin (1969), Jiang (2004b), Li, Liu and Rosalsky (2010), Xiao and Zhou (2010), Cai and Jiang (2011), Cai and Jiang (2012), Shao and Zhou (2014), and the references therein. Research under $\mathbf{R} \neq \mathbf{I}_p$ and non-Gaussian distributions is more challenging. In fact, Jiang (2019) showed that research under $\mathbf{R} \neq \mathbf{I}_p$ is more difficult than that under $\mathbf{R} = \mathbf{I}_p$.

### 1.2. Our contributions

The existing literature for high-dimensional correlation matrices imposes more restrictive conditions, such as the Gaussian assumption or $\mathbf{R} = \mathbf{I}_p$. As a result, these existing results cannot solve the problem with a non-Gaussian population and $\mathbf{R} \neq \mathbf{I}_p$. Thus, the aim of this study is to examine the spectral properties of high-dimensional **rescaled sample correlation matrices $\widehat{\mathbf{R}}_n\mathbf{R}^{-1}$**. Our contributions to the existing literature are as follows:

(I) This study derives the Stieltjes equation of the LSD of high-dimensional rescaled sample correlation matrices $\widehat{\mathbf{R}}_n\mathbf{R}^{-1}$ under the convergence regime $p/n \to \rho \in (0, \infty)$. The assumption $\mathbf{R} = \mathbf{I}_p$ and spherical distributions are not required. It is interesting that the LSD of $\widehat{\mathbf{R}}_n\mathbf{R}^{-1}$ is just the standard Marčenko–Pastur law with the index $\rho$.

(II) This study establishes the CLT of the eigenvalue statistics of high-dimensional rescaled sample correlation matrices $\widehat{\mathbf{R}}_n\mathbf{R}^{-1}$ under the convergence regime $p/n \to \rho \in (0, \infty)$. Although our proposed CLT closely depends on the population correlation matrix $\mathbf{R}$, the asymptotic mean and asymptotic variance of our proposed CLT have explicit forms that can be expressed as one-dimensional or two-dimensional contour integrals on a unit circle $\{\xi = x + \mathbf{i}\nu : x^2 + \nu^2 = 1\}$. In particular, under the assumption $\mathbf{R} = \mathbf{I}_p$, the asymptotic mean and asymptotic variance of our proposed CLT have much simpler explicit forms, which are independent of the population kurtosis.

The reminder of the paper proceeds as follows. Section 2 presents the limiting spectral distribution of high-dimensional rescaled sample correlation matrices. Section 3 establishes the CLT of the linear spectral statistics of high-dimensional rescaled sample correlation matrices. To illustrate our proposed limiting theorems, Section 4 provides an application to test whether the population correlation matrix is equal to a given matrix. Extensive simulation studies are also presented in Section 4. Then, Section 5 concludes the paper. Proofs are included in the online Supplementary Material.

## 2. Limiting Spectral Distribution

Before studying high-dimensional rescaled sample correlation matrices, we first provide three assumptions.

**Assumption 1.** *Samples satisfy the following independent component structure:*

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{x}_i, \quad i = 1, \ldots, n,$$

*where* $\mathrm{E}\mathbf{y}_i = \boldsymbol{\mu}$, $\boldsymbol{\Gamma} = [\mathrm{diag}(\boldsymbol{\Sigma})]^{1/2}\mathbf{R}^{1/2}$ *and* $\mathbf{x}_i = (x_{1i}, \ldots, x_{pi})^T$.

**Assumption 2.** *Assume that* $\{x_{ji}, j = 1, \ldots, p, i = 1, \ldots, n\}$ *are independent and identically distributed (i.i.d.), with*

$$\mathrm{E}x_{ji} = 0, \mathrm{E}x_{ji}^2 = 1, \quad E(|x_{ji}|^4(\log(|x_{ji}|))^{2+2\epsilon}) < \infty,$$

*for a small positive number* $\epsilon > 0$.

**Assumption 3.** *The convergence regime is* $\rho_n = p/n \to \rho \in (0, +\infty)$.

In fact, if there is an $0 < \varepsilon < 1$ such that the $(4+\varepsilon)$ th moment of $x_{1,1}$ exists, then Assumption B is satisfied. Assumption C requires that the dimension $p$ and the sample size $n$ diverge proportionally.

For simplicity, let $\{\lambda_j, j = 1, \ldots, p\}$ be the eigenvalues of $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$. The *empirical spectral distribution* (ESD) of $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$ is defined as

$$F_n(x) = p^{-1}\sum_{j=1}^{p}\delta(\lambda_j \le x).$$

The following theorem provides the form of the LSD of $F_n(x)$.

**Theorem 1.** *Under Assumptions* 1–3*, the empirical spectral distribution* $F_n(x)$ *of* $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$ *converges almost surely to the Marčenko–Pastur law with the index* $\rho$*, as follows:*

$$f_\rho(x) = \begin{cases} \dfrac{1}{2\pi\rho x}\sqrt{(b_\rho - x)(x - a_\rho)}, & \text{if } a_\rho \le x \le b_\rho, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1}$$

*where* $a_\rho = (1 - \sqrt{\rho})^2$, *and* $b_\rho = (1 + \sqrt{\rho})^2$.

## 3. CLT

Let $\{\lambda_j, j = 1, \ldots, p\}$ be the eigenvalues of $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$. Define the *linear spectral statistic* (LSS) of $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$ as

$$L_g = \sum_{j=1}^{p} g(\lambda_j), \tag{3.1}$$

where $g(\cdot)$ is a given analytic function. We are now interested in the fluctuation of $L_g$. We study the LSS because many commonly used statistics can be expressed as the LSS. For illustration, we provide three examples:

**Example 1.** When $g(x) = x^k$, we have $L_g = \sum_{j=1}^{p} \lambda_j^k = \mathrm{tr}[(\mathbf{R}^{-1}\widehat{\mathbf{R}}_n)^k]$;

**Example 2.** When $g(x) = (x-1)^2$, we have $L_g = \sum_{j=1}^{p}(\lambda_j - 1)^2 = \mathrm{tr}[(\mathbf{R}^{-1}\widehat{\mathbf{R}}_n - \mathbf{I}_p)^2]$;

**Example 3.** When $g(x) = \log x$, we have $L_g = \sum_{j=1}^{p} \log \lambda_j = \log|\mathbf{R}^{-1}\widehat{\mathbf{R}}_n|$.

To establish the CLT of the LSS of $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$, for fixed $K$ and known functions $g_1, \ldots, g_K$, we consider the $K$-dimensional random vector $(W(g_1), \ldots, W(g_K))$, where

$$W(g_\ell) = \sum_{j=1}^{p} g_\ell(\lambda_j) - p \int g_\ell(x) f_{\rho_{n-1}}(x)dx, \ \ell = 1, \ldots, K,$$

and $f_{\rho_{n-1}}(x)$ is defined in (2.1) with $\rho_{n-1} = p/(n-1)$. We also impose the following assumptions for our results concerning the CLT of the LSS of $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$:

**Assumption 4.** *The functions $g_1, \ldots, g_K$ are analytic functions in a domain containing the support set $[a_\rho, b_\rho]$ of the Marčenko–Pastur law in* (2.1)*.*

**Assumption 5.** *Assume that $\{x_{ji}, j = 1, \ldots, p, i = 1, \ldots, n\}$ are i.i.d. with*

$$\mathrm{E}x_{ji} = 0, \ \mathrm{E}x_{ji}^2 = 1, \ \mathrm{E}x_{ji}^4 = \beta_x + 3 + o(1), \ \mathrm{E}(|x_{ji}|^4(\log(|x_{ji}|)^{2+2\epsilon})) < \infty.$$

**Assumption 6.** *Assume*

$$a_g = \lim_{p\to\infty} p^{-1} \sum_{k=1}^{p}\sum_{h=1}^{p} g_{kh}^3 \mathbf{e}_h^T \mathbf{R}^{-1/2}\mathbf{e}_k, \qquad a_{\mathbf{R}} = \lim_{p\to\infty} p^{-1} \sum_{k,\ell=1}^{p} \mathbf{e}_k^T \mathbf{R}^{-1}\mathbf{e}_\ell r_{k\ell}^3,$$

$$c_g = \lim_{p\to\infty} p^{-1} \sum_{k=1}^{p}\sum_{h=1}^{p} g_{kh}^4, \qquad d_{\mathbf{R}} = \lim_{p\to\infty} p^{-1}\mathrm{tr}(\mathbf{R}^2),$$

$$h_{\mathbf{R}} = \lim_{p\to\infty} p^{-1} \sum_{k,\ell=1}^{p} \mathbf{e}_k^T \mathbf{R}^{-1}\mathbf{e}_\ell r_{k\ell} \sum_{h=1}^{p} g_{\ell h}^2 g_{kh}^2,$$

*where* $\mathbf{R}^{1/2} = (g_{kh})$, $\mathbf{R} = (r_{kh})$ *and* $\mathbf{e}_j$ *is the* $j$*th column of the* $p \times p$ *identity matrix, for* $j = 1, \dots, p$.

**Remark 1.** Here, two examples are given to show that Assumption F can be satisfied. (i). When $\mathbf{R} = \mathbf{I}_p$, it is easy to see that $\mathbf{R}^{1/2} = \mathbf{R}^{-1/2} = \mathbf{R}^{-1} = \mathbf{I}_p$. Thus, we have $a_g = a_{\mathbf{R}} = c_g = d_{\mathbf{R}} = h_{\mathbf{R}} = 1$. (ii). When $\mathbf{R} = (1-\tau)\mathbf{I}_p + \tau\mathbf{1}\mathbf{1}^T$, with $\tau \in (-1, 1)$, we have

$$\begin{aligned}
\mathbf{R}^{-1} &= \frac{1}{1-\tau}\mathbf{I}_p - \frac{\tau}{(1-\tau)(1+(p-1)\tau)}\mathbf{1}\mathbf{1}^T, \\
\mathbf{R}^{1/2} &= \sqrt{1-\tau}\mathbf{I}_p + \frac{\sqrt{1+(p-1)\tau} - \sqrt{(1-\tau)}}{p}\mathbf{1}\mathbf{1}^T, \\
\mathbf{R}^{-1/2} &= \frac{1}{\sqrt{1-\tau}}\mathbf{I}_p - \frac{\sqrt{1+(p-1)\tau} - \sqrt{(1-\tau)}}{p\sqrt{1+(p-1)\tau}\sqrt{(1-\tau)}}\mathbf{1}\mathbf{1}^T.
\end{aligned}$$

If $\tau = C/\sqrt{p}$, then $a_g = a_{\mathbf{R}} = c_g = h_{\mathbf{R}} = 1$ and $d_{\mathbf{R}} = 1 + C^2$.

The following theorem states our main results concerning the CLT of the LSS of $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$.

**Theorem 2.** *Under Assumptions* 1 *and* 3–6*, the random vector* $(W(g_1), \dots, W(g_K))$ *converges weakly to a multivariate Gaussian random vector* $(X_{g_1}, \dots, X_{g_K})$ *with*

$$\mathrm{E}X_{g_\ell} = -\frac{1}{2\pi\mathbf{i}}\oint_{\mathcal{C}} g_\ell(z)\mathrm{E}M(z)\,dz$$

*and*

$$\mathrm{Cov}(X_{g_{\ell_1}}, X_{g_{\ell_2}}) = -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} g_{\ell_1}(z_1)g_{\ell_2}(z_2)\mathrm{Cov}(M(z_1), M(z_2))\,dz_2\,dz_1,$$

*for* $\ell, \ell_1, \ell_2 \in \{1, \dots, K\}$*, where* $\mathcal{C}, \mathcal{C}_1$*, and* $\mathcal{C}_2$ *are three non-overlapping contours including* $[a_\rho, b_\rho]$*,* $\mathcal{C}_1$*, and* $\mathcal{C}_2$*, the contour integral* $\oint$ *is anticlockwise, and* $\mathrm{E}M(z)$ *and* $\mathrm{Cov}(M(z_1), M(z_2))$ *are calculated as follows:*

$$\begin{aligned}
\mathrm{E}M(z) =& \frac{\rho\underline{s}^3(z)[1+\underline{s}(z)]^{-3}}{[1-\rho\underline{s}^2(z)(1+\underline{s}(z))^{-2}]^2} + \frac{\beta_x\rho\underline{s}(z)\underline{s}'(z)}{[1+\underline{s}(z)]^3} \\
&- \frac{\underline{s}'(z)}{[1+\underline{s}(z)]^2}\frac{[10 - 2a_{\mathbf{R}} + \beta_x(4a_g + c_g - h_{\mathbf{R}})]\rho}{4} \\
&+ \frac{\underline{s}'(z)}{[1+\underline{s}(z)]^3}\frac{[6 - 2a_{\mathbf{R}} + \beta_x(4a_g - c_g - h_{\mathbf{R}})]\rho}{2},
\end{aligned} \tag{3.2}$$

*and*

$$\begin{aligned}\mathrm{Cov}(M(z_1),M(z_2)) = 2\left[\frac{\underline{s}'(z_1)\underline{s}'(z_2)}{(\underline{s}(z_2)-\underline{s}(z_1))^2}-\frac{1}{(z_2-z_1)^2}\right] \\ +2(d_{\mathbf{R}}-2)\rho\frac{\underline{s}'(z_1)\underline{s}'(z_2)}{(1+\underline{s}(z_1))^2(1+\underline{s}(z_2))^2},\end{aligned} \tag{3.3}$$

*where ′ is the derivative notation, and* $\underline{s}(z)$ *is the unique solution to* $z=-\underline{s}^{-1}(z)+\rho(1+\underline{s}(z))^{-1}$*, which leads to* $\underline{s}'(z)=\underline{s}^2(z)/\{1-\rho\underline{s}^2(z)[1+\underline{s}(z)]^{-2}\}$*.*

To simplify the mean and the covariance, we derive the following corollary, where the contour becomes a unit circle centered at the origin.

**Corollary 1.** *Under Assumptions* 1 *and* 3–6*, the random vector* $(W(g_1),\ldots,W(g_K))$ *converges weakly to a multivariate Gaussian random vector* $(X_{g_1},\ldots,X_{g_K})$*, with*

$$\begin{aligned}\mathrm{E}X_{g_\ell} &= \lim_{r\to1^+}\frac{1}{2\pi\mathbf{i}}\oint\limits_{\|\xi\|=1} g_\ell(1+\sqrt{\rho}\xi+\sqrt{\rho}\xi^{-1}+\rho)\left(\frac{\xi}{\xi^2-r^{-2}}-\frac{1}{\xi}\right)d\xi \\ &\quad+\frac{4a_{\mathbf{R}}-12+\beta_x(4-8a_g+2c_g+2h_{\mathbf{R}})}{8\pi\mathbf{i}}\oint\limits_{\|\xi\|=1}\frac{g_\ell(1+\sqrt{\rho}\xi+\sqrt{\rho}\xi^{-1}+\rho)}{\xi^3}d\xi \\ &\quad+\frac{[2a_{\mathbf{R}}-2+\beta_x(-4a_g+3c_g+h_{\mathbf{R}})]\sqrt{\rho}}{8\pi\mathbf{i}}\oint\limits_{\|\xi\|=1}\frac{g_\ell(1+\sqrt{\rho}\xi+\sqrt{\rho}\xi^{-1}+\rho)}{\xi^2}d\xi,\end{aligned}$$

*and*

$$\begin{aligned}&\mathrm{Cov}(X_{g_{\ell_1}},X_{g_{\ell_2}}) \\ &= \lim_{r\to1^+}\frac{-1}{2\pi^2}\oint\limits_{\|\xi_1\|=1}\oint\limits_{\|\xi_2\|=1}\frac{g_{\ell_1}(\|1+\sqrt{\rho}\xi\|^2)g_{\ell_2}(\|1+\sqrt{\rho}\xi\|^2)}{(\xi_1-r\xi_2)^2}d\xi_2d\xi_1 \\ &\quad-\frac{(d_{\mathbf{R}}-2)}{2\pi^2}\oint\limits_{\|\xi_1\|=1}\frac{g_{\ell_1}(\|1+\sqrt{\rho}\xi_1\|^2)}{\xi_1^2}d\xi_1\oint\limits_{\|\xi_2\|=1}\frac{g_{\ell_2}(\|1+\sqrt{\rho}\xi_2\|^2)}{\xi_2^2}d\xi_2,\end{aligned}$$

*for* $\ell,\ell_1,\ell_2\in\{1,\ldots,K\}$*, where the integral* $\oint$ *is anticlockwise, and* $\|1+\sqrt{\rho}\xi\|^2=1+\sqrt{\rho}\xi+\sqrt{\rho}\xi^{-1}+\rho$*, for* $\xi$ *satisfying* $\|\xi\|=1$*.*

We provide some examples to illustrate the application of Theorem 2.

**Example 4.** When $\mathbf{R}=\mathbf{I}_p$, it is easy to see that $a_g=a_{\mathbf{R}}=c_g=d_{\mathbf{R}}=h_{\mathbf{R}}=1$. Then, $\mathrm{E}X_{g_\ell}$ and $\mathrm{Cov}(X_{g_{\ell_1}},X_{g_{\ell_2}})$ can be further simplified as

$$\mathrm{E}X_{g_\ell} = \lim_{r\to 1^+} \frac{1}{2\pi\mathbf{i}} \oint\limits_{\|\xi\|=1} g_\ell(1+\sqrt{\rho}\xi+\sqrt{\rho}\xi^{-1}+\rho)\left(\frac{\xi}{\xi^2-r^{-2}}-\frac{1}{\xi}-\frac{2}{\xi^3}\right)d\xi,$$

$$\begin{aligned}\mathrm{Cov}(X_{g_{\ell_1}},X_{g_{\ell_2}}) &= \lim_{r\to 1^+}\frac{-1}{2\pi^2}\oint\limits_{\|\xi_1\|=1}\oint\limits_{\|\xi_2\|=1}\frac{g_{\ell_1}(\|1+\sqrt{\rho}\xi\|^2)g_{\ell_2}(\|1+\sqrt{\rho}\xi\|^2)}{(\xi_1-r\xi_2)^2}d\xi_2 d\xi_1\\ &\quad+\frac{1}{2\pi^2}\oint\limits_{\|\xi_1\|=1}\frac{g_{\ell_1}(\|1+\sqrt{\rho}\xi_1\|^2)}{\xi_1^2}d\xi_1\oint\limits_{\|\xi_2\|=1}\frac{g_{\ell_2}(\|1+\sqrt{\rho}\xi_2\|^2)}{\xi_2^2}d\xi_2,\end{aligned}$$

where $\ell,\ell_1,\ell_2\in\{1,\ldots,K\}$.

**Example 5.** Letting $g_\ell(x)=x^\ell$ for $\ell=1,2,3,4$ and $g_5(x)=\log x$, we have the centering terms:

$$\begin{aligned}&\int g_1(x)f_{\rho_{n-1}}(x)dx=1,\quad \int g_2(x)f_{\rho_{n-1}}(x)dx=1+\rho_{n-1},\\ &\int g_3(x)f_{\rho_{n-1}}(x)dx=1+3\rho_{n-1}+\rho_{n-1}^2,\\ &\int g_4(x)f_{\rho_{n-1}}(x)dx=1+6\rho_{n-1}+6\rho_{n-1}^2+\rho_{n-1}^3,\\ &\int g_5(x)f_{\rho_{n-1}}(x)dx=\frac{\rho_{n-1}-1}{\rho_{n-1}}\log(1-y_{n-1})-1,\rho_{n-1}<1,\end{aligned}$$

the mean

$$\begin{aligned}\mathrm{E}X_{g_1} &= (-0.5+0.5a_{\mathbf{R}})\rho+\beta_x(-a_g+0.75c_g+0.25h_{\mathbf{R}})\rho,\\ \mathrm{E}X_{g_2} &= (-3+2a_{\mathbf{R}})\rho+(-1+a_{\mathbf{R}})\rho^2+\beta_x(1-4a_g+2c_g+h_{\mathbf{R}})\rho\\ &\quad+\beta_x(-2a_g+0.5h_{\mathbf{R}}+1.5c_g)\rho^2,\\ \mathrm{E}X_{g_3} &= 1.5(-1+a_{\mathbf{R}})\rho^3+1.5(-7+5a_{\mathbf{R}})\rho^2+1.5(-5+3a_{\mathbf{R}})\rho\\ &\quad+\beta_x[0.75(-4a_g+3c_g+h_{\mathbf{R}})\rho^3+(3-15a_g+8.25c_g+3.75h_{\mathbf{R}})\rho^2]\\ &\quad+\beta_x(3-9a_g+3.75c_g+2.25h_{\mathbf{R}})\rho,\\ \mathrm{E}X_{g_4} &= (-2+2a_{\mathbf{R}})\rho(1+\rho)^3+6(-3+2a_{\mathbf{R}})\rho(1+\rho)^2-2a_{\mathbf{R}}\rho^2+6(1-a_{\mathbf{R}})\rho\\ &\quad+\beta_x[(-4a_g+3c_g+h_{\mathbf{R}})\rho(1+\rho)^3+3(2-4a_g+c_g+h_{\mathbf{R}})\rho(1+\rho)^2]\\ &\quad+\beta_x[3(-4a_g+3c_g+h_{\mathbf{R}})\rho^2(1+\rho)+2(2-4a_g+c_g+h_{\mathbf{R}})\rho^2],\\ \mathrm{E}X_{g_5} &= 0.5\log(1-\rho)+\rho+0.5\beta_x(c_g-1)\rho,\end{aligned}$$

and the variance-covariance

$$\mathrm{Var}(X_{g_1})=2\rho(d_{\mathbf{R}}-1),\quad \mathrm{Var}(X_{g_2})=4\rho^2+8\rho\left(1+\rho\right)^2\left(d_{\mathbf{R}}-1\right),$$

$$\begin{aligned}
\operatorname{Var}(X_{g_3}) &= 6\rho^3 + 36\rho^2(1+\rho)^2 + 9\rho[(1+\rho)^2+\rho]^2[2(d_{\mathbf{R}}-1)],\\
\operatorname{Var}(X_{g_4}) &= 8\rho^4 + 96\rho^3(1+\rho)^2 + 16\rho^2[2\rho+3(1+\rho)^2]^2\\
&\quad +16\rho(1+\rho)^2[3\rho+(1+\rho)^2]^2[2(d_{\mathbf{R}}-1)],\\
\operatorname{Var}(X_{g_5}) &= -2\log(1-\rho) - 2\rho + \rho[2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_1}, X_{g_2}) &= 2\rho(1+\rho)[2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_1}, X_{g_3}) &= 3\rho[(1+\rho)^2+\rho][2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_1}, X_{g_4}) &= 4\rho[3(1+\rho)\rho+(1+\rho)^3][2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_2}, X_{g_3}) &= 12\rho^2(1+\rho) + 6\rho\,(1+\rho)\,[(1+\rho)^2+\rho][2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_2}, X_{g_4}) &= 8\rho^2[2\rho+3(1+\rho)^2]\\
&\quad +8\rho(1+\rho)[3(1+\rho)\rho+(1+\rho)^3][2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_3}, X_{g_4}) &= 24(1+\rho)\rho^3 + 24\rho^2(1+\rho)[2\rho+3(1+\rho)^2]\\
&\quad +12\rho[3(1+\rho)\rho+(1+\rho)^3][(1+\rho)^2+\rho][2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_1}, X_{g_5}) &= \rho[2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_2}, X_{g_5}) &= -2\rho^2 + 2\rho(1+\rho)[2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_3}, X_{g_5}) &= 2\rho^3 - 6\rho^2(1+\rho) + 3\rho\left((1+\rho)^2+\rho\right)[2(d_{\mathbf{R}}-1)],\\
\operatorname{Cov}(X_{g_4}, X_{g_5}) &= -2\rho^4 + 8(1+\rho)^2\rho^3 - 4\rho^2\left(2\rho+3(1+\rho)^2\right)\\
&\quad +4\rho(1+\rho)\left(3\rho+(1+\rho)^2\right)[2(d_{\mathbf{R}}-1)].
\end{aligned}$$

**Example 6.** Let $g_\ell(x) = x^\ell$ for $\ell = 1,2,3,4$, and $g_5(x) = \log x$. When $\mathbf{R} = \mathbf{I}_p$, we have the mean

$$\begin{aligned}
&\mathrm{E}X_{g_1} = 0,\ \ \mathrm{E}X_{g_2} = -\rho,\ \ \mathrm{E}X_{g_3} = -3\rho^2 - 3\rho,\\
&\mathrm{E}X_{g_4} = -6\rho(1+\rho)^2 - 2\rho^2,\ \mathrm{E}X_{g_5} = 0.5\log(1-\rho) + \rho,
\end{aligned}$$

and the covariance

$$\begin{aligned}
\operatorname{Var}(X_{g_1}) &= 0,\quad \operatorname{Var}(X_{g_2}) = 4\rho^2,\quad \operatorname{Var}(X_{g_3}) = 6\rho^3 + 36\rho^2(1+\rho)^2,\\
\operatorname{Var}(X_{g_4}) &= 8\rho^4 + 96\rho^3(1+\rho)^2 + 16\rho^2[2\rho+3(1+\rho)^2]^2,\\
\operatorname{Var}(X_{g_5}) &= -2\log(1-\rho) - 2\rho,\quad \operatorname{Cov}(X_{g_1}, X_{g_j}) = 0,\ j = 2,3,4,5\\
\operatorname{Cov}(X_{g_2}, X_{g_3}) &= 12\rho^2(1+\rho),\quad \operatorname{Cov}(X_{g_2}, X_{g_4}) = 8\rho^2[2\rho+3(1+\rho)^2],\\
\operatorname{Cov}(X_{g_2}, X_{g_5}) &= -2\rho^2,\quad \operatorname{Cov}(X_{g_3}, X_{g_5}) = 2\rho^3 - 6\rho^2(1+\rho),\\
\operatorname{Cov}(X_{g_3}, X_{g_4}) &= 24(1+\rho)\rho^3 + 24\rho^2(1+\rho)[2\rho+3(1+\rho)^2],\\
\operatorname{Cov}(X_{g_4}, X_{g_5}) &= -2\rho^4 + 8(1+\rho)^2\rho^3 - 4\rho^2\left(2\rho+3(1+\rho)^2\right).
\end{aligned}$$

## 4. An Application and Simulation Studies

### 4.1. An application

We apply our proposed CLT of eigenvalue statistics of high-dimensional rescaled sample correlation matrices, to study the hypothesis testing problem that the population correlation matrix is equal to a given matrix:

$$H_0 : \mathbf{R} = \mathbf{R}_0 \ \ v.s. \ \ H_1 : \mathbf{R} \neq \mathbf{R}_0, \tag{4.1}$$

where $\mathbf{R}_0$ is a prespecified matrix. Let $\widehat{\mathbf{R}}_n$ be the sample correlation matrix. Under the null hypothesis $H_0$, Kullback (1967) showed

$$T_K = (n-1)\{\mathrm{tr}(\mathbf{R}_0^{-1}\widehat{\mathbf{R}}_n) - \log(\mathbf{R}_0^{-1}\widehat{\mathbf{R}}_n) - p\} \to \chi^2_{p(p-1)/2}.$$

Based on the different distance between $\widehat{\mathbf{R}}_n$ and $\mathbf{R}_0$, our three proposed test statistics are as follows:

$$T_1 = \mathrm{tr}[(\mathbf{R}_0^{-1}\widehat{\mathbf{R}}_n - \mathbf{I}_p)^4], \ T_2 = \mathrm{tr}[(\mathbf{R}_0^{-1}\widehat{\mathbf{R}}_n - \mathbf{I}_p)^2], \ T_3 = \log|\mathbf{R}_0^{-1}\widehat{\mathbf{R}}_n|.$$

By Example 5, asymptotic distributions of $T_1$, $T_2$, and $T_3$ are derived in the following theorem.

**Theorem 3.** *Under Assumptions* 1 *and* 3–6 *and under* $H_0$*, we have*

$$\begin{aligned}
&\sigma_1^{-1}(T_1 - \mu_1) \to N(0,1), \ 0 < \rho_{n-1},\\
&\sigma_2^{-1}(T_2 - \mu_2) \to N(0,1), \ 0 < \rho_{n-1},\\
&\sigma_3^{-1}(T_3 - \mu_3) \to N(0,1), \ 0 < \rho_{n-1} < 1,
\end{aligned}$$

*where*

$$\begin{aligned}
\mu_1 &= p\rho_{n-1}^3 + 2p\rho_{n-1}^2 + \mathrm{E}X_{g_4} - 4\mathrm{E}X_{g_3} + 6\mathrm{E}X_{g_2} - 4\mathrm{E}X_{g_1},\\
\mu_2 &= p\rho_{n-1} + \mathrm{E}X_{g_2} - 2\mathrm{E}X_{g_1},\\
\mu_3 &= p(\rho_{n-1} - 1)(\rho_{n-1})^{-1}\log(1-\rho_{n-1}) - p + \mathrm{E}X_{g_5},\\
\sigma_1^2 &= \mathrm{Var}(X_{g_4}) + 16\,\mathrm{Var}(X_{g_3}) + 36\,\mathrm{Var}(X_{g_2}) + 16\,\mathrm{Var}(X_{g_1}) - 8\mathrm{Cov}(X_{g_3}, X_{g_4})\\
&\qquad +12\mathrm{Cov}(X_{g_2}, X_{g_4}) - 8\mathrm{Cov}(X_{g_1}, X_{g_4}) - 48\mathrm{Cov}(X_{g_2}, X_{g_3})\\
&\qquad +32\mathrm{Cov}(X_{g_1}, X_{g_3}) - 48\mathrm{Cov}(X_{g_1}, X_{g_2}),\\
\sigma_2^2 &= \mathrm{Var}(X_{g_2}) - 4\mathrm{Cov}(X_{g_1}, X_{g_2}) + 4\,\mathrm{Var}(X_{g_1}), \quad \sigma_3^2 = \mathrm{Var}(X_{g_5}),
\end{aligned}$$

*where* $\rho_{n-1} = p/(n-1)$, $\mathrm{E}X_{g_\ell}$, $\mathrm{Var}(X_{g_\ell})$, $\mathrm{Cov}(X_{g_{\ell_1}}, X_{g_{\ell_2}})$ *with* $\ell, \ell_1, \ell_2 \in \{1,2,3,4,5\}$ *are as in Example* 5*, where* $\rho$ *can be replaced by* $\rho_{n-1}$.

Then, the rejection region of the test statistic $T_i$ at the test level 5% is

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_n : \sigma_i^{-1}|T_i - \mu_i| > q_{0.975}\},$$

for $i = 1, 2, 3$, where $q_{0.975}$ is the 97.5% quantile of $N(0, 1)$.

### 4.2. Simulation studies

In this subsection, the results from extensive simulation studies are presented to evaluate the performance of our three proposed test statistics. Without loss of generality, assume that the sample $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is drawn from a population with mean zero and covariance matrix $\mathbf{\Sigma}$. The dimension is taken as $p = 100, 200, 400$, and $\rho = p/n$ is taken as $0.1, 0.5, 0.8$. We consider two models:

**Model 1:** $\mathbf{\Sigma} = U(\mathbf{I}_p + D)U^T + \theta \mathbf{1}_p^T \mathbf{1}_p$, where $\mathbf{1}_p$ is the $p$-dimensional vector with all elements being one, $U$ is the eigenvector matrix of $\mathbf{Z}^T\mathbf{Z}$ with all the elements of $\mathbf{Z} = (z_{ij})_{i,j=1,\ldots,p}$ being i.i.d. from $N(0, 1)$, and $D = \operatorname{diag}(d_{11}, \ldots, d_{pp})$ is a diagonal matrix with $d_{11}, \ldots, d_{pp}$ being i.i.d. from the uniform distribution $U(0, 1)$;

**Model 2:** $\mathbf{\Sigma} = (s_{i,j,\theta})_{p\times p}$, where

$$s_{i,j,\theta} = 2(1 - p^{-1/2})^{|i-j|} + \theta p^{-1/2}\delta_{\{i=j\}},$$

with $\delta_{\{\cdot\}}$ being an indicator function, for $i, j = 1, \ldots, p$.

In both models, we set $\theta = 0$ to evaluate the empirical size, set $\theta = 0.01, 0.02$ to examine the empirical power. The nominal test size is $\alpha = 5\%$. Tables 1 and 4 present the empirical sizes for Gaussian and non-Gaussian distributions under Model 1 and Model 2. Tables 2–3 and Tables 5–6 present the empirical power of the Gaussian and non-Gaussian distributions when $\theta = 0.01$ and $\theta = 0.02$ under Model 1 and Model 2, respectively. Our proposed tests $T_1, T_2, T_3$ are compared with $T_K$ proposed in Kullback (1967) and $T_G$ ($T_G$ is a special case of $T_2$ with $\mathbf{R} = \mathbf{I}$) proposed in Gao et al. (2017). For each setting, we run the simulation 10,000 times.

The simulation results show that for Models 1–2, $T_K$ becomes invalid for large $p$ because $T_K$ is proposed for fixed $p$. $T_G$ is invalid for $\mathbf{R} \neq \mathbf{I}$ because $T_G$ is proposed for $\mathbf{R} = \mathbf{I}$. $T_1$, $T_2$, and $T_3$ have empirical sizes close to the nominal test size of 5% for small or large dimension $p$. In particular, $T_1$ may have slightly higher empirical test sizes, because $T_1$ contains the sums of the fourth powers of the sample eigenvalues of $\hat{\mathbf{R}}_n\mathbf{R}_0^{-1}$. Then, if there are some large top eigenvalues, the variance of $T_1$ may be large, which will influence the finite-sample performance

Table 1. Percentages for the empirical size for Gaussian and Non-Gaussian populations under Model 1.

| Empirical sizes for $\theta = 0$ under Model 1 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | | Normal | | | | | Gamma | | | | |
| $\rho$ | $p$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ |
| 0.1 | 100 | 4.58 | 4.81 | 5.01 | 56.0 | 0.00 | 4.90 | 5.04 | 5.45 | 56.1 | 0.00 |
| | 200 | 4.91 | 5.01 | 5.06 | 96.7 | 0.00 | 5.24 | 4.98 | 4.78 | 96.8 | 0.00 |
| | 400 | 5.14 | 5.00 | 5.14 | 100 | 0.00 | 4.80 | 5.03 | 4.92 | 100 | 0.00 |
| 0.5 | 100 | 4.43 | 4.67 | 5.02 | 100 | 0.00 | 4.72 | 5.51 | 5.45 | 100 | 0.00 |
| | 200 | 4.98 | 4.78 | 4.76 | 100 | 0.00 | 4.93 | 5.18 | 5.24 | 100 | 0.00 |
| | 400 | 4.74 | 4.62 | 4.98 | 100 | 0.00 | 4.92 | 5.05 | 4.93 | 100 | 0.00 |
| 0.8 | 100 | 4.44 | 4.77 | 5.16 | 100 | 0.00 | 4.43 | 4.88 | 5.43 | 100 | 0.00 |
| | 200 | 4.77 | 4.84 | 4.98 | 100 | 0.00 | 4.79 | 5.25 | 5.44 | 100 | 0.00 |
| | 400 | 4.91 | 4.88 | 5.10 | 100 | 0.00 | 4.96 | 5.03 | 5.56 | 100 | 0.00 |

Table 2. Percentages for the empirical power for Gaussian and Non-Gaussian populations with $\theta = 0.01$ under Model 1.

| Empirical powers for $\theta = 0.01$ under Model 1 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | | Normal | | | | | Gamma | | | | |
| $\rho$ | $p$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ |
| 0.1 | 100 | 93.6 | 58.5 | 15.3 | 95.0 | 0 | 93.6 | 58.3 | 14.6 | 94.8 | 0 |
| | 200 | 100 | 100 | 92.3 | 100 | 98.6 | 100 | 100 | 92.2 | 100 | 98.4 |
| | 400 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 11.24 | 8.52 | 5.71 | 100 | 0 | 11.42 | 8.40 | 5.52 | 100 | 0 |
| | 200 | 84.2 | 45.7 | 11.6 | 100 | 0 | 83.8 | 45.3 | 11.8 | 100 | 0 |
| | 400 | 100 | 99.9 | 53.9 | 100 | 12.4 | 100 | 100 | 53.6 | 100 | 13.5 |
| 0.8 | 100 | 8.45 | 6.86 | 5.99 | 100 | 0 | 8.41 | 6.80 | 5.73 | 100 | 0 |
| | 200 | 40.0 | 20.4 | 6.88 | 100 | 0 | 42.1 | 21.4 | 6.86 | 100 | 0 |
| | 400 | 100 | 97.9 | 18.1 | 100 | 0 | 100 | 97.8 | 18.8 | 100 | 0 |

Table 3. Percentages for the empirical power for Gaussian and Non-Gaussian populations $\theta = 0.02$ under Model 1.

| Empirical powers for $\theta = 0.02$ under Model 1 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | | Normal | | | | | Gamma | | | | |
| $\rho$ | $p$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ |
| 0.1 | 100 | 100 | 100 | 91.4 | 100 | 94.6 | 100 | 100 | 91.5 | 100 | 94.5 |
| | 200 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 400 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 78.1 | 47.7 | 11.0 | 100 | 0 | 77.5 | 47.0 | 10.8 | 100 | 0 |
| | 200 | 100 | 99.9 | 51.2 | 100 | 14.9 | 100 | 99.9 | 51.5 | 100 | 15 |
| | 400 | 100 | 100 | 99 | 100 | 100 | 100 | 100 | 99.9 | 100 | 100 |
| 0.8 | 100 | 47.9 | 27.8 | 7.09 | 100 | 0 | 48.3 | 27.6 | 6.91 | 100 | 0 |
| | 200 | 99.9 | 96.3 | 17.3 | 100 | 0.05 | 99.9 | 96.2 | 17.2 | 100 | 0.01 |
| | 400 | 100 | 100 | 73.2 | 100 | 99.9 | 100 | 100 | 73.4 | 100 | 99.9 |

Table 4. Percentages for the empirical size for Gaussian and Non-Gaussian populations under Model 2.

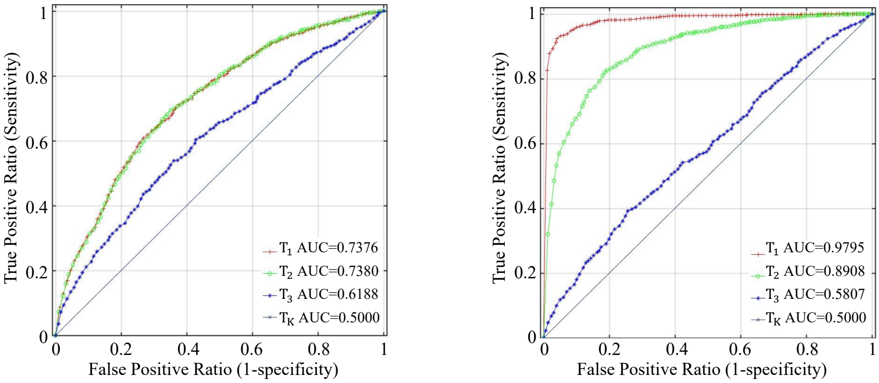
| Empirical sizes for $\theta = 0$ under Model 2 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | | Normal | | | | | Gamma | | | | |
| $\rho$ | $p$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ |
| 0.1 | 100 | 6.10 | 5.72 | 4.89 | 80.1 | 0.09 | 6.21 | 5.88 | 4.91 | 85.9 | 0.10 |
| | 200 | 5.58 | 5.61 | 4.99 | 99.6 | 0.19 | 5.78 | 5.62 | 5.18 | 99.9 | 0.31 |
| | 400 | 5.57 | 5.23 | 5.01 | 100 | 0.39 | 5.24 | 5.14 | 4.72 | 100 | 0.64 |
| 0.5 | 100 | 7.12 | 6.11 | 5.16 | 100 | 1.34 | 7.33 | 6.14 | 5.72 | 100 | 1.34 |
| | 200 | 5.97 | 5.42 | 5.04 | 100 | 2.30 | 6.23 | 5.68 | 5.27 | 100 | 2.69 |
| | 400 | 5.71 | 5.40 | 5.13 | 100 | 3.13 | 5.47 | 5.19 | 5.06 | 100 | 3.66 |
| 0.8 | 100 | 8.60 | 6.79 | 5.18 | 100 | 1.94 | 7.62 | 6.31 | 5.46 | 100 | 1.68 |
| | 200 | 6.63 | 5.76 | 5.27 | 100 | 3.16 | 6.22 | 5.72 | 5.24 | 100 | 3.10 |
| | 400 | 6.14 | 5.53 | 5.00 | 100 | 6.43 | 5.87 | 5.29 | 4.92 | 100 | 6.41 |

Table 5. Percentages for the empirical power for Gaussian and Non-Gaussian populations with $\theta = 0.01$ under Model 2.

| Empirical powers for $\theta = 0.01$ under Model 2 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | | Normal | | | | | Gamma | | | | |
| $\rho$ | $p$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ |
| 0.1 | 100 | 31.0 | 27.0 | 25.9 | 85.7 | 2.67 | 30.3 | 25.3 | 25.9 | 89.4 | 3.07 |
| | 200 | 62.2 | 58.4 | 61.4 | 99.8 | 21.3 | 61.1 | 57.4 | 59.9 | 99.9 | 23.4 |
| | 400 | 96.3 | 95.4 | 96.89 | 100 | 84.4 | 96.7 | 96.0 | 97.0 | 100 | 87.9 |
| 0.5 | 100 | 15.8 | 12.7 | 8.90 | 100 | 4.86 | 15.3 | 12.3 | 9.07 | 100 | 4.36 |
| | 200 | 22.5 | 19.7 | 15.9 | 100 | 15.4 | 22.3 | 19.8 | 16.9 | 100 | 16.1 |
| | 400 | 45.5 | 44.1 | 40.7 | 100 | 50.2 | 44.9 | 42.5 | 39.8 | 100 | 50.6 |
| 0.8 | 100 | 14.9 | 11.2 | 7.13 | 100 | 4.66 | 14.7 | 11.2 | 7.86 | 100 | 4.54 |
| | 200 | 18.3 | 15.6 | 11.9 | 100 | 13.4 | 17.6 | 15.0 | 11.7 | 100 | 13.1 |
| | 400 | 32.7 | 30.7 | 26.6 | 100 | 39.2 | 32.3 | 30.3 | 26.4 | 100 | 39.9 |

Table 6. Percentages for the empirical power for Gaussian and Non-Gaussian populations $\theta = 0.02$ under Model 2.

| Empirical powers for $\theta = 0.02$ under Model 2 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | | Normal | | | | | Gamma | | | | |
| $\rho$ | $p$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ | $T_1$ | $T_2$ | $T_3$ | $T_K$ | $T_G$ |
| 0.1 | 100 | 75.9 | 71.0 | 73.2 | 94.4 | 23.1 | 76.0 | 70.4 | 72.9 | 95.6 | 25.6 |
| | 200 | 99.4 | 99.0 | 99.3 | 99.9 | 90.9 | 99.5 | 99.2 | 99.4 | 100 | 92.9 |
| | 400 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | 100 | 31.8 | 27.4 | 20.1 | 100 | 13.0 | 31.4 | 26.5 | 20.6 | 100 | 12.7 |
| | 200 | 57.7 | 54.6 | 49.1 | 100 | 47.9 | 58.7 | 55.1 | 49.5 | 100 | 49.4 |
| | 400 | 93.7 | 93.3 | 91.6 | 100 | 95.1 | 93.2 | 92.9 | 91.2 | 100 | 95.1 |
| 0.8 | 100 | 26.2 | 20.9 | 14.0 | 100 | 11.0 | 25.2 | 20.2 | 13.7 | 100 | 10.5 |
| | 200 | 44.7 | 40.4 | 32.0 | 100 | 37.6 | 43.3 | 39.5 | 32.2 | 100 | 36.6 |
| | 400 | 81.2 | 79.7 | 74.5 | 100 | 86.1 | 80.1 | 78.9 | 73.8 | 100 | 85.7 |



(a) Model 1

(b) Model 2

Figure 1. ROC cures for Models 1–2 with $p = 400, \rho = 0.5, \theta = 0.005$.

of $T_1$ when $n$ and $p$ are not large enough. In fact, the empirical sizes of $T_1$ are close to 5% as $n$ and $p$ become large.

For detailed comparisons, we plot the receiver operating characteristic (ROC) curve for four tests $T_1, T_2, T_3, T_K$ when $p = 400$, $\rho = 0.5$, and $\theta = 0.02$ in Figure 1 for Model 1 and Model 2. From the ROC curve, $T_1$ and $T_2$ perform better than other tests for Model 1, and $T_1$ performs better than other tests for Model 2.

## 5. Conclusion and Discussion

We have investigated the spectral properties of high-dimensional rescaled sample correlation matrices. Under the framework that the dimension and the sample size tend to infinity proportionally, we proved that the LSS of $\mathbf{R}^{-1}\widehat{\mathbf{R}}_n$ have Gaussian fluctuations under some mild assumptions. By multiplying $\mathbf{R}^{-1}$, we relaxed the commonly used assumption in random matrix theory that the spectral norm of $\mathbf{R}$ is bounded in $p$. Furthermore, in contrast to the existing literature, we do not need to assume that $\mathbf{R} = \mathbf{I}_p$ or Gaussian populations. We also provided some useful examples of the LSS for rescaled sample correlation matrices. An application was proposed for hypothesis testing on population correlation matrices, and simulations were conducted to investigate the performance of the proposed test statistics. Future work will focus on the spectral properties of $\widehat{\mathbf{R}}_n$.

## Supplementary Material

The online Supplementary Material contains detailed proofs of Theorem 1, Theorem 2, Corollary 1, and Example 5.

## Acknowledgments

We thank the associate editor, referees, and editor for their constructive and insightful comments. Yanqing Yin is supported by NSFC grant 11801234 and the Natural Science Foundation of Jiangsu Province (BK20181000). Shurong Zheng (the corresponding author) was supported by NSFC grant 12071066 and 11690012. Guoliang Tian is supported by NSFC 12171225.

## References

Aitkin, M. A. (1969). Some tests for correlation matrices. *Biometrika* **56**, 443–446.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann. Probab.* **32**, 553–605.

Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39**, 1496–1525.

Cai, T. T. and Jiang, T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *J. Multivariate Anal.* **107**, 24–39.

Fan, J. Q., Guo, J. H. and Zheng, S. R. (2020). Estimating number of factors by adjusted eigenvalues thresholding. *J. Amer. Statist. Assoc.* DOI: 10.1080/01621459.2020.1825448.

Gao, J. T., Han, X., Pan, G. M. and Yang, Y. R. (2017). High-dimensional correlation matrices: The central limit theorem and its application. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **79**, 677–693.

Jiang, T. F. (2004a). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhya* **66**, 35–48.

Jiang, T. F. (2004b). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14**, 865–880.

Jiang, T. F. (2019). Determinant of sample correlation matrix with application. *Ann. Appl. Probab.* **29**, 1356–1397.

Karoui, N. E. (2009). Concentration of measure and spectra of random matrices: With applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19**, 2362–2405.

kullback, S (1967). On testing correlation matrices. *J. Roy. Statist. Soc. Ser. C* **16**, 80–85.

Li, D., Liu, W. D. and Rosalsky, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields* **148**, 5–35.

Marcenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb* **1**, 457–483.

Mestre, X. and Vallet, P. (2017). Correlation tests and linear spectral statistics of the sample correlation matrix. *IEEE Transactions on Information Theory* **63**, 4585–4618.

Morales-Jimenez, D., Johnstone, I. M., Mckay, M. R. and Yang, J. (2019). Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models. *arXiv:1810.10214v3*

Shao, Q. M. and Zhou, W. X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab.* **42**, 623–648.

Silverstein, J. and Choi, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* **54**, 295–309.

Xiao, H. and Zhou, W. (2010). Almost sure limit of the smallest eigenvalue of some sample correlation matrices. *J. Theoret. Probab.* **23**, 1–20.

Zheng, S. R, Bai, Z. D. and Yao, J. F. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43**, 546–591.

Yanqing Yin

School of Mathematics and Statistics, Chongqing University, Chongqing, China.

E-mail: yinyq799@nenu.edu.cn

Changcheng Li

School of Mathematical Sciences, Dalian University of Technology, Dalian, Liaoning, China.

E-mail: lichangcheng@dlut.edu.cn

Guo-Liang Tian

Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, China.

E-mail: tiangl@sustech.edu.cn

Shurong Zheng

School of Mathematics & Statistics and KLAS, Northeast Normal University, Changchun, Jilin, China.

E-mail: zhengsr@nenu.edu.cn

(Received July 2020; accepted March 2021)