The Annals of Statistics
2019, Vol. 47, No. 1, 527–555
https://doi.org/10.1214/18-AOS1699

HIGH-DIMENSIONAL COVARIANCE MATRICES IN ELLIPTICAL DISTRIBUTIONS WITH APPLICATION TO SPHERICAL TEST

BY JIANG HU*,†,1, WEIMING LI‡,2, ZHI LIU§,3 AND WANG ZHOU¶,4

Northeast Normal University, Key Laboratory for Applied Statistics of MOE†, Shanghai University of Finance and Economics‡, University of Macau§ and National University of Singapore¶*

This paper discusses fluctuations of linear spectral statistics of high-dimensional sample covariance matrices when the underlying population follows an elliptical distribution. Such population often possesses high order correlations among their coordinates, which have great impact on the asymptotic behaviors of linear spectral statistics. Taking such kind of dependency into consideration, we establish a new central limit theorem for the linear spectral statistics in this paper for a class of elliptical populations. This general theoretical result has wide applications and, as an example, it is then applied to test the sphericity of elliptical populations.

1. Introduction. Large-scale statistical inference develops rapidly in the last two decades. This type of inference often relies on spectral statistics of certain random matrices in high-dimensional frameworks, where both the dimension p of the observations and the sample size n tend to infinity. Recall that a *linear spectral statistic* (LSS) [Bai and Silverstein (2010)] of a $p \times p$ Hermitian random matrix R_n is of the form

$$\frac{1}{p}\sum_{i=1}^{p} f(\lambda_i) = \int f(x)\, dF^{R_n}(x), \tag{1.1}$$

where $\lambda_1, \dots, \lambda_p$ are the p eigenvalues of R_n, f is a function defined on $\mathbb{R}$ and $F^{R_n} = (1/p)\sum_{i=1}^{p} \delta_{\lambda_i}$ is called the *spectral distribution* (SD) of R_n. Here, δ_a denotes the Dirac measure at the point a. In Ledoit and Wolf (2002) and Schott (2007), most test statistics are actually LSSs of sample covariance matrices. Bai

Received January 2017; revised October 2017.

[1]Supported in part by NSFC (No. 11771073), Science and Technology Development foundation of Jilin (Grant No. 20160520174JH) and the Fundamental Research Funds for the Central Universities.

[2]Supported in part by NSFC (No. 11401037), MOE Project of Humanities and Social Sciences, No. 17YJC790057, and Program of IRTSHUFE.

[3]Supported by FDCT of Macau (No. 078/2016/A3 and No. 202/2017/A3).

[4]Supported in part by the MOE Tier 2 Grant MOE2015-T2-2-039 (R-155-000-171-112) at the National University of Singapore.

MSC2010 subject classifications. Primary 62H15; secondary 62H10.

Key words and phrases. Covariance matrix, high-dimensional data, elliptical distribution, sphericity test.

et al. (2009) made systematic corrections to several classical likelihood ratio tests to overcome the effect of high dimension using LSSs of sample covariance matrices and F-matrices. Later, Bai et al. (2015) derived the CLT for LSSs of a high-dimensional Beta matrix, which can be broadly used in multivariate statistical analysis, such as testing the equality of several covariance matrices, multivariate analysis of variance and canonical correlation analysis; see Anderson (2003). Most recently, based on an LSS of regularized canonical correlation matrices, Yang and Pan (2015) proposed a test for the independence between two large random vectors. Gao et al. (2017) applied LSSs of sample correlation matrices to the complete independence test for p random variables and the equivalence test for p factor loadings in a factor model. Clearly, it is of great interests to investigate the behaviors of LSSs under various circumstances.

Specifically, let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be n observations of $\mathbf{x} \in \mathbb{R}^p$, whose mean is zero and covariance matrix is Σ. The sample covariance matrix is

$$B_n = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j'.$$

Our attention in this paper is focused on the asymptotic properties of LSSs of B_n. The earliest study on this problem dates back to Jonsson (1982), who obtained the *central limit theorem* (CLT) for LSSs of B_n by assuming the population to be standard multivariate normal. A remarkable breakthrough was done in Bai and Silverstein (2004), where the population is allowed to be a linear transform of a vector of independent and identically distributed (i.i.d.) random variables, that is,

$$\mathbf{x} = A\mathbf{z}. \tag{1.2}$$

Here, $A \in \mathbb{R}^{p \times p}$ is a non-random transformation matrix with $\operatorname{rank}(A) = p$, and $\mathbf{z} = (z_1, \dots, z_p)'$ with i.i.d. z_i's satisfying

$$E(z_1) = 0, \qquad E(z_1^2) = 1 \quad \text{and} \quad E(z_1^4) = 3. \tag{1.3}$$

The fourth moment condition was later extended by Pan and Zhou (2008) to $E(z_1^4) < \infty$. Though these assumptions are fairly weak, their requirement of linearly dependent structure in (1.2) still excludes a lot of important distributions. In particular, it excludes almost all distributions from the elliptical family [Fang and Zhang (1990)].

Elliptical distributions were originally introduced by Kelker (1970) to generalize the multivariate normal distributions. A random vector $\mathbf{x}$ with zero mean follows an elliptical distribution if and only if it has a stochastic representation [Fang and Zhang (1990)]:

$$\mathbf{x} = \xi A\mathbf{u}, \tag{1.4}$$

where the matrix $A \in \mathbb{R}^{p \times p}$ is nonrandom with $\operatorname{rank}(A) = p$, $\xi \geq 0$ is a scalar variable representing the radius of $\mathbf{x}$, and $\mathbf{u} \in \mathbb{R}^p$ is the random direction, which is independent of ξ and uniformly distributed on the unit sphere S^{p-1} in $\mathbb{R}^p$, denoted by $\mathbf{u} \sim U(S^{p-1})$ in the sequel. This family of distributions has been widely applied in many areas, such as statistics, economics and finance, which can describe fat (or light) tails and tail dependence among components of a population; see Fang and Zhang (1990) and Gupta, Varga and Bodnar (2013). Evidently such distributions with high order correlations cannot be modeled by the linear transform model in (1.2).

A question raised immediately is that how the nonlinear dependency affects the asymptotic behaviors of LSSs in high-dimensional frameworks? Indeed, Bai and Zhou (2008) proved that the SD F^{B_n} of B_n converges to a common generalized Marčenko–Pastur law almost surely if, for any sequence of symmetric matrices $\{C_p\}$ with bounded spectral norm,

$$\operatorname{Var}(\mathbf{x}' C_p \mathbf{x}) = o(p^2) \tag{1.5}$$

as $p, n \to \infty$. This condition is also sharp for the convergence; see Li and Yao (2018) for an example. What is more, this condition holds for a list of well-known elliptical distributions, such as multivariate normal distributions, multivariate Pearson type II distributions, power exponential distributions and a more general family of multivariate Kotz-type distributions [Kotz (1975)]; see Section 2 for more details. However, the limit of SD is not enough for many procedures of statistical inference, such as confidence interval and hypothesis testing. Therefore, in this paper, we will explore the fluctuations of LSSs of B_n, when the population belongs to elliptical distributions that satisfy the condition (1.5). Compared with the pioneer work of Bai and Silverstein (2004), the main difficulty of the current study lies in the fact that both the radius ξ and direction $\mathbf{u}$ introduce nonlinear dependence to the coordinates of the population $\mathbf{x}$, which cannot be handled through the same way as they did for the linearly dependent structure. Technically, we are facing the following three challenges. First, for paying the cost of dropping linearly dependent structure, we have to add more moment conditions on ξ, because the finite fourth moment of $\xi/\sqrt{p}$ is no longer sufficient for the nonlinear dependence case [see (2.5)]. This is totally different from the linearly dependent structure case. Second, we need to figure out how the dependence of the entries of $\xi\mathbf{u}$ influences the fluctuations of LSSs of B_n (see Remark 2.3). Third, we have to extend many fundamental conclusions in the independent case [Bai and Silverstein (2004)] to accommodate our nonlinearly dependent structure; see Lemmas A.1–A.4 for example.

The structure of this paper is as follows. First, in Section 2, we set up a new CLT for LSSs of B_n under elliptical distributions satisfying (1.5). Then in Section 3, based on the derived results, we theoretically investigate the problem of sphericity test for covariance matrices. This is done by discussing a John's-type test from Tian, Lu and Li (2015) for general alternative models and a likelihood ratio test from Onatski, Moreira and Hallin (2013) for *spiked covariances* under arbitrary elliptical distributions. For illustration, the John-type test is applied to analyze a dataset of weekly stock returns in Section 4. Technical proofs of the main results are gathered in Section 5. Some supporting lemmas are postponed to the Appendix. The paper has also an on-line supplementary file which includes the following materials: (i) CLT for general moment LSSs; (ii) simulations regarding the John-type test; (iii) proofs of some lemmas.

2. High-dimensional theory for eigenvalues of B_n.

This section investigates asymptotic behaviors of the eigenvalues of B_n, referred as *sample eigenvalues*. We begin with proposing an equivalent condition of (1.5) under the settings of the elliptical model in (1.4).

LEMMA 2.1. *Suppose that a p-dimensional random vector $\mathbf{x}$ has a stochastic form $\mathbf{x} = \xi A\mathbf{u}$ as defined in (1.4) with the radius ξ normalized as $E(\xi^2) = p$. If the spectral norm of $\Sigma = AA'$ is uniformly bounded in p, then the following two conditions are equivalent*:

$$\text{(a)} \quad \operatorname{Var}(\mathbf{x}' C_p \mathbf{x}) = o(p^2), \qquad \text{(b)} \quad E(\xi^4) = p^2 + o(p^2),$$

as $p \to \infty$, where $\{C_p\}$ is any sequence of symmetric matrices with bounded spectral norm.

REMARK 2.1. The fourth moment condition (b) together with the normalization $E(\xi^2) = p$ characterize the class of elliptical distributions discussed in this paper. For the normal case, the squared radius $\xi^2 \sim \chi_p^2$, and thus $E(\xi^2) = p$ and $E(\xi^4) = p^2 + 2p$. In general, the typical order of $E(\xi^4)$ is $p^2 + \tau p + o(p)$ with $\tau \geq 0$ a constant. Hence a specific elliptical distribution can be recognized by evaluating the ratio

$$E(\xi^4)/E^2(\xi^2) = 1 + \tau/p + o(p^{-1}), \tag{2.1}$$

as $p \to \infty$. We note that the parameter τ has a nonnegligible effect on the limiting distributions of LSSs of B_n; see Theorem 2.2. The proof of Lemma 2.1 is given in the Supplementary Material [Hu et al. (2018)].

In the following, we provide three examples of elliptical family satisfying condition (2.1). Some commonly seen elliptical distributions are also checked and the results are summarized in Table 1.

EXAMPLE 2.1. A p-dimensional centered multivariate Pearson type II distribution has a density function

$$f(\mathbf{x}) = c_p |\Sigma_p|^{-\frac{1}{2}} [1 - \mathbf{x}'\Sigma_p^{-1}\mathbf{x}]^{\frac{\beta}{2}-1}, \tag{2.2}$$

TABLE 1
Some elliptical distributions and their verification of t condition (2.1). The notation "$\perp\!\!\!\perp$" in the last row denotes independence

$\mathbf{x} = \xi A\mathbf{u} \in \mathbb{R}^p$	Distribution of ξ	$E(\xi^4)/E^2(\xi^2)$	Condition (2.1)
Normal	$\xi^2 \sim \chi_p^2$	$1 + \frac{2}{p}$	Holds ($\tau = 2$).
Double exponential	$\xi \sim \text{Gamma}(p, 1)$	$1 + \frac{4p+6}{p(p+1)}$	Holds ($\tau = 4$).
Exponential power	$\xi^{2s} \sim \text{Gamma}(\frac{p}{2s}, \frac{1}{2})$	$1 + \frac{2}{sp} + o(p^{-1})$	Holds ($\tau = \frac{2}{s}$).
Student-t	$\xi^2/p \sim F(p, v), v > 4$	$1 + \frac{2}{v-4} + \frac{2(v-2)}{p(v-4)}$	Not hold.
Normal scale mixture	$\xi^2 = w \cdot v, w \perp\!\!\!\perp v, v \sim \chi_p^2$	$1 + \frac{\text{Var}(w)}{E^2(w)} + \frac{2}{p}\frac{E(w^2)}{E^2(w)}$	Not hold.

where $c_p = \pi^{-p/2}\Gamma[(\beta + p)/2]/\Gamma(\beta/2)$ and $\beta > 0$. The stochastic representation of such a distribution is $\mathbf{x} = \xi\Sigma_p^{1/2}\mathbf{u}$, where ξ^2 follows the distribution $\text{Beta}(p/2, \beta/2)$; see Fang and Zhang (1990). Therefore, we have

$$E(\xi^4)/E^2(\xi^2) = 1 + 2\beta/(p^2 + \beta p + 2p),$$

which verifies the condition in (2.1) with $\tau = 0$.

EXAMPLE 2.2. The family of Kotz-type distributions introduced by Kotz (1975) is an important class of elliptical distributions, which includes normal distributions, exponential power distributions and double exponential distribution as special cases. The density function of a centered Kotz-type random variable $\mathbf{x}$ is

$$f(\mathbf{x}) = c_p|\Sigma_p|^{-\frac{1}{2}}[\mathbf{x}'\Sigma_p^{-1}\mathbf{x}]^{k-1}\exp\{-\beta[\mathbf{x}'\Sigma_p^{-1}\mathbf{x}]^s\}, \tag{2.3}$$

where $c_p = s\beta^\alpha\pi^{-p/2}\Gamma(p/2)\Gamma(\alpha)$ with $\alpha = (k - 1 + p/2)/s > 0$ and $(\beta, s) > 0$. Write $\mathbf{x} = \xi\Sigma_p^{1/2}\mathbf{u}$. The $2s$ power of the radius is $\xi^{2s} = [\mathbf{x}'\Sigma_p^{-1}\mathbf{x}]^s$ which has the characteristic function

$$E(e^{it\xi^{2s}}) = c_p\int e^{it[\mathbf{x}'\Sigma_p^{-1}\mathbf{x}]^s} f(\mathbf{x})\,d\mathbf{x} \propto \int e^{itx}x^{\alpha-1}e^{-\beta x}\,dx, \tag{2.4}$$

where the seconded integral is derived by polar coordinates transformation. This characteristic function implies that ξ^{2s} follows the Gamma distribution $\text{Gamma}(\alpha, \beta)$. Simple calculations reveal that

$$\frac{E(\xi^4)}{E^2(\xi^2)} = \frac{\Gamma(\alpha + 2/s)\Gamma(\alpha)}{\Gamma^2(\alpha + 1/s)} = 1 + \frac{1}{s^2\alpha} + o(\alpha^{-1}),$$

which verifies the condition in (2.1) with $\tau = 2/s$. For the mentioned three special cases, their details are presented in the 2nd–4th rows of Table 1.

EXAMPLE 2.3. Let $\mathbf{x} = \xi A\mathbf{u}$ with $\xi^2 = \sum_{j=1}^{p} y_j^2$ independent of $\mathbf{u}$, where (y_j) is a sequence of i.i.d. random variables with

$$E(y_1^2) = 1 \quad \text{and} \quad E(y_1^4) = \mu_4 < \infty.$$

Then it is simple to check that $E(\xi^2) = p$ and $E(\xi^4) = p^2 + (\mu_4 - 1)p$ which verifies the condition in (2.1) with $\tau = \mu_4 - 1$.

We should note that condition (2.1) excludes some elliptical distributions, such as multivariate Student-t distributions and normal scale mixtures, as shown in the 5th–6th rows of Table 1. Indeed, sample eigenvalues from these distributions do not obey the generalized Marčenko–Pastur law [El Karoui (2009), Li and Yao (2018)], which are then out of the scope of this paper.

Now we are ready to investigate the asymptotic properties of sample eigenvalues in high-dimensional frameworks, under the following assumptions.

ASSUMPTION (A). Both the sample size n and dimension p tend to infinity in such a way that $c_n := p/n \to c \in (0, \infty)$.

ASSUMPTION (B). There are two independent arrays of i.i.d. random variables $(\mathbf{u}_j)_{j\geq 1}$, $\mathbf{u}_1 \sim U(S^{p-1})$, and $(\xi_j)_{j\geq 1}$ satisfying for some $\tau \geq 0$ and $\varepsilon > 0$,

$$E(\xi_1^2) = p, \qquad E(\xi_1^4) = p^2 + \tau p + o(p) \quad \text{and} \quad E\left|\frac{\xi_1^2 - p}{\sqrt{p}}\right|^{2+\varepsilon} < \infty, \tag{2.5}$$

such that for each p and n the observation vectors can be represented as $\mathbf{x}_j = \xi_j A\mathbf{u}_j$, where A is a $p \times p$ matrix.

ASSUMPTION (C). The spectral distribution H_p of the matrix $\Sigma := AA'$ weakly converges to a probability distribution H, as $p \to \infty$, referred to as Population Spectral Distribution (PSD). Moreover, the spectral norm of the sequence (Σ) is uniformly bounded in p.

In the sequel, for any function G of bounded variation on the real line, its Stieltjes transform is defined by

$$m(z) = \int \frac{1}{\lambda - z}\, dG(\lambda), \qquad z \in \mathbb{C} \setminus S_G, \tag{2.6}$$

where S_G stands for the support of G. Then we have the following theorems.

THEOREM 2.1. *Suppose that Assumptions* (a)–(c) *hold. Then, almost surely, the empirical spectral distribution F^{B_n} converges weakly to a probability distribution $F^{c,H}$, whose Stieltjes transform $m = m(z)$ is the only solution to the equation*

$$m = \int \frac{1}{t(1 - c - czm) - z}\, dH(t), \qquad z \in \mathbb{C}^+, \tag{2.7}$$

in the set $\{m \in \mathbb{C} : -(1-c)/z + cm \in \mathbb{C}^+\}$ where $\mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im(z) > 0\}$.

REMARK 2.2. Theorem 2.1 follows from Lemma 2.1 and Theorem 1.1 in Bai and Zhou (2008), and thus we omit its proof here. Let $\underline{m} = \underline{m}(z)$ be the Stieltjes transform of $\underline{F}^{c,H} = cF^{c,H} + (1-c)\delta_0$. Then equation (2.7) can be reexpressed as

$$z = -\frac{1}{\underline{m}} + c\int \frac{t}{1+t\underline{m}}\,dH(t), \qquad z \in \mathbb{C}^+, \tag{2.8}$$

which is the so-called Silverstein equation [Silverstein (1995)].

Let F^{c_n,H_p} be the distribution defined by (2.7) with the parameters (c, H) replaced by (c_n, H_p) and denote $G_n = F^{B_n} - F^{c_n,H_p}$. We next study the fluctuation of centralized LSSs with form

$$\int f(x)\,dG_n(x) = \int f(x)\,d\big[F^{B_n}(x) - F^{c_n,H_p}(x)\big],$$

where f is a function on the real line.

THEOREM 2.2. *Suppose that Assumptions* (a)–(c) *hold. Let* $f_1, \ldots, f_k$ *be functions analytic on an open interval containing*

$$\Big[\liminf_{p\to\infty} \lambda_{\min}^{\Sigma}\delta_{(0,1)}(c)(1-\sqrt{c})^2, \limsup_{p\to\infty} \lambda_{\max}^{\Sigma}(1+\sqrt{c})^2\Big]. \tag{2.9}$$

Then the random vector

$$p\Big(\int f_1(x)\,dG_n(x), \ldots, \int f_k(x)\,dG_n(x)\Big)$$

converges weakly to a Gaussian vector $(X_{f_1}, \ldots, X_{f_k})$, *with mean function*

$$\begin{aligned} EX_f = &-\frac{1}{2\pi \mathrm{i}}\oint_{\mathcal{C}_1} f(z)\int \frac{c(\underline{m}'(z)t)^2}{\underline{m}(z)(1+t\underline{m}(z))^3}\,dH(t)\,dz \\ &- \frac{\tau-2}{2\pi \mathrm{i}}\oint_{\mathcal{C}_1} f(z)\int \frac{(z\underline{m}(z)+1)\underline{m}'(z)t}{(1+t\underline{m}(z))^2}\,dH(t)\,dz \end{aligned}$$

and covariance function

$$\begin{aligned} \operatorname{Cov}(X_f, X_g) = &-\frac{1}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} \frac{f(z_1)g(z_2)\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\,dz_1\,dz_2 \\ &+ c(\tau-2)\int xf'(x)\,dF(x)\int xg'(x)\,dF(x), \end{aligned}$$

$(f, g \in \{f_1, \ldots, f_k\})$, *where the contours* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *are nonoverlapping, closed, counterclockwise orientated in the complex plane and each enclosing the support of the limiting spectral distribution* $F^{c,H}$.

REMARK 2.3. When the population is normal, or rather $\tau = 2$, this theorem coincides with the main result in Bai and Silverstein (2004). It implies that the high order correlation among the components of the population affects both the limiting mean vectors and the covariance matrices of LSSs by additive quantities proportional to $\tau - 2$. This factor can be further decomposed into two parts, τ and -2, which correspond respectively to the effect from the radius ξ and that from the direction $\mathbf{u}$ (considering the case $\xi^2 \equiv p$). It is interesting to see that these two kinds of dependency have opposite effects and they may cancel each other for normal population.

As an application of Theorem 2.2, we consider $\hat{\beta}_{nj} = \int x^j \, dF^{B_n}(x)$, $j = 1, 2$, the first two moments of sample eigenvalues. Theorem 2.2 implies

$$p \begin{pmatrix} \hat{\beta}_{n1} - \beta_{n1} \\ \hat{\beta}_{n2} - \beta_{n2} \end{pmatrix} \xrightarrow{D} N \left(\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{pmatrix} \right), \tag{2.10}$$

where the parameters possess explicit expressions as

$$\beta_{n1} = \gamma_{n1}, \qquad \beta_{n2} = \gamma_{n2} + c_n \gamma_{n1}^2, \qquad v_1 = 0, \qquad v_2 = c\gamma_2 + c(\tau - 2)\gamma_1,$$

$$\psi_{11} = 2c\gamma_2 + c(\tau - 2)\gamma_1^2, \qquad \psi_{12} = 4c\gamma_3 + 4c^2\gamma_1\gamma_2 + 2c(\tau - 2)\gamma_1(c\gamma_1^2 + \gamma_2),$$

$$\psi_{22} = 8c\gamma_4 + 4c^2\gamma_2^2 + 16c^2\gamma_1\gamma_3 + 8c^3\gamma_1^2\gamma_2 + 4c(\tau - 2)(c\gamma_1^2 + \gamma_2)^2,$$

where $\gamma_{nj} = \int t^j \, dH_p(t)$ and $\gamma_j = \int t^j \, dH(t)$ for $j \geq 1$. For LSSs of higher order moments, explicit formulas of their limiting means and covariances are discussed in the Supplementary Material [Hu et al. (2018)].

We conduct a small simulation experiment to examine the fluctuations of $\hat{\beta}_{n1}$ and $\hat{\beta}_{n2}$. In the experiment, the PSD H_p is fixed at $H_p = 0.5\delta_1 + 0.5\delta_2$. The distribution of ξ is selected as (1) $\xi \sim k_1 \operatorname{Gamma}(p, 1)$ with $k_1 = 1/\sqrt{p+1}$ and (2) $\xi^2 \sim k_2 \operatorname{Beta}(p/2, 2)$ with $k_2 = p + 4$, which correspond the CLT with $\tau = 4$ and $\tau = 0$, respectively. The factors k_1 and k_2 are selected to satisfy $E(\xi^2) = p$. The dimensional settings are $(p, n, c) = (200, 400, 0.5), (400, 200, 2)$ and the number of independent replications is 10,000. Normal QQ-plots for normalized statistics, that is, $p(\hat{\beta}_{n1} - \beta_{n1})/\sqrt{\psi_{11}}$ and $[p(\hat{\beta}_{n1} - \beta_{n1}) - v_2]/\sqrt{\psi_{22}}$, are displayed in Figure 1. Their asymptotic standard normality is well confirmed in all studied cases.

3. Testing for high-dimensional spherical distributions.

3.1. *John's test and its extension.* In this section, we revisit the sphericity test for covariance matrices in high-dimensional frameworks. For this particular test problem, the underlying population can follow arbitrary elliptical distribution, which may violate the condition in (1.5).

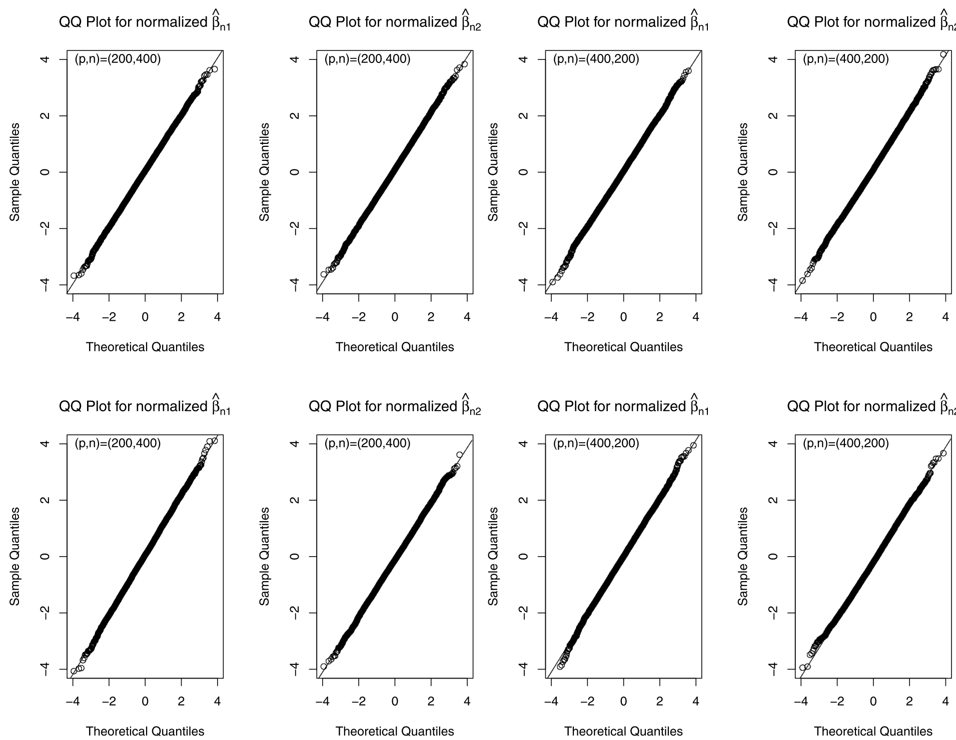

FIG. 1. *Normal QQ-plots for normalized* $\hat{\beta}_{n1}$ *and* $\hat{\beta}_{n2}$ *from* 10,000 *independent replications. Upper panels*: $\xi \sim k_1\,\mathrm{Gamma}(p,1)$ *with* $k_1 = 1/\sqrt{p+1}$. *Lower panels*: $\xi^2 \sim k_2\,\mathrm{Beta}(p/2,2)$ *with* $k_2 = p+4$. *The dimensional settings are* $(p,n,c) = (200, 400, 0.5), (400, 200, 2)$.

The sphericity test on the covariance matrix Σ is

$$H_0 : \Sigma = \sigma^2 I_p \quad \text{vs.} \quad H_1 : \Sigma \neq \sigma^2 I_p, \tag{3.1}$$

where σ^2 is an unknown scalar parameter. When the dimension p is fixed, for normal populations, John (1972) proposed a locally most powerful invariant test statistic to deal with the sphericity hypothesis based on the spectrum of sample covariance matrices. Due to its concise form and broad applicability, this kind of test is quite favorable for high-dimensional situations and has been extensively studied in recent years; see, for example, Ledoit and Wolf (2002), Wang and Yao (2013), Tian, Lu and Li (2015) for the linear transform model in (1.2), while Zou et al. (2014) and Paindaveine and Verdebout (2016) for the elliptical model in (1.4). In particular, the test statistic in Tian, Lu and Li (2015) synthesizes the first four moments of sample eigenvalues, by which it gains extra powers for spike-like alternative covariance matrices. However, this statistic is not valid for general elliptical populations [Li and Yao (2018)]. Hence, we next develop an analogue test procedure with the help of the theoretical results in Section 2, and then compare it numerically with that from Paindaveine and Verdebout (2016).

Since the hypotheses in (3.1) are only concerned with the shape component of Σ, by convention, we transform the original samples into the so-called spatial-sign samples, that is,

$$\check{\mathbf{x}}_j := \sqrt{p}\mathbf{x}_j/\|\mathbf{x}_j\| = \sqrt{p}A\mathbf{u}_j/\|A\mathbf{u}_j\|, \qquad j = 1, \dots, n.$$

Therefore, testing the sphericity of Σ can now be converted to testing the identity of $\check{\Sigma} = E(\check{\mathbf{x}}_1\check{\mathbf{x}}_1')$. This inference can be realized by constructing spectral statistics of $\check{B}_n = \sum_{j=1}^n \check{\mathbf{x}}_j\check{\mathbf{x}}_j'/n$. Specifically, let

$$\alpha_{nj} = p^{-1}\operatorname{tr}(\check{\Sigma}^j) \quad \text{and} \quad \check{\beta}_{nj} = p^{-1}\operatorname{tr}(\check{B}_n^j),$$

$j = 0, 1, 2, \dots$. By verifying the condition in (1.5) for $\check{\mathbf{x}}_1$, one may conclude that Theorem 2.1 also holds for $(\check{\Sigma}, \check{B}_n)$ with all conditions on ξ removed. Then, similar to Tian, Lu and Li (2015), from the fact that $\check{\beta}_{n1} \equiv 1$, one may obtain estimators of α_{n2} and α_{n4} as

$$\check{\alpha}_{n2} = \check{\beta}_{n2} - c_n, \qquad \check{\alpha}_{n4} = \check{\beta}_{n4} - 4c_n\check{\beta}_{n3} - 2c_n(\check{\beta}_{n2})^2 + 10c_n^2\check{\beta}_{n2} - 5c_n^3,$$

respectively, and two simple statistics for the sphericity test as

$$T_1 = \check{\alpha}_{n2} - 1 \quad \text{and} \quad T_2 = \check{\alpha}_{n4} - 1.$$

Moreover, their joint null distribution is directly from (2.10) with $\tau = 0$.

THEOREM 3.1. *Suppose that Assumptions* (a)–(c) [*removing the moment conditions in* (2.5)] *hold. Under the null hypothesis*,

$$n(T_1, T_2) \xrightarrow{D} N_2(\mu, \Omega),$$

where $\mu = (-1, -6 + c)$ *and the covariance matrix* $\Omega = (\omega_{ij})$ *with* $\omega_{11} = 4$, $\omega_{12} = 24$, *and* $\omega_{22} = 8(18 + 12c + c^2)$.

The two statistics T_1 and T_2, together with their null distributions, provide two test procedures for the identity of $\check{\Sigma}$ (thus the sphericity of Σ). The test statistic T_1 agrees with that in Paindaveine and Verdebout (2016), where its null asymptotic distribution is proved to be universal whenever $\min\{n, p\} \to \infty$. For the case where the population mean is unknown, see Zou et al. (2014). The test statistic T_2 is new. Compared with T_1, it is more sensitive to extreme eigenvalues of $\check{\Sigma}$, and thus can serve as a complement of T_1. Parallel to Tian, Lu and Li (2015), a joint statistic of T_1 and T_2 can be constructed as

$$T_m = \max\left\{\frac{nT_1 + 1}{2}, \frac{nT_2 + 6 - c_n}{\sqrt{8(18 + 12c_n + c_n^2)}}\right\},$$

where the two original statistics are both standardized according to their asymptotic null distributions.

THEOREM 3.2. *Suppose that Assumptions* (a)–(c) [*removing the moment conditions in* (2.5)] *hold and let* $\delta_p = p\,\mathrm{tr}(\Sigma^2)/\mathrm{tr}^2(\Sigma) - 1$.

(i) *Under the null hypothesis, for any* $x \in \mathbb{R}$,

$$P(T_m \le x) \to \int_{-\infty}^{x}\int_{-\infty}^{x} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right\} du\,dv,$$

where $\rho = 6/\sqrt{2(18 + 12c + c^2)}$.

(ii) *Under the alternative hypothesis, if* $n\delta_p \to \infty$ *then the power of the test* T_m *goes to* 1 *as* $(n, p) \to \infty$.

The asymptotic null distribution of T_m is an immediate consequence of Theorem 3.1. The consistency of T_m can be proved by showing either the consistency of T_1 or T_2. As the consistency of T_1 has been given in Zou et al. (2014), we omit its proof.

We have run a simulation experiment for the tests T_1, T_2 and T_m to check their finite-sample properties under similar model settings as in Tian, Lu and Li (2015). The results show that all the three tests have satisfactory empirical sizes and powers. In addition, compared with T_1 and T_2, the test T_m exhibits its robustness against different types of alternative models; see the Supplementary Material [Hu et al. (2018)].

3.2. *Sphericity test under spiked alternative model.* The sphericity test T_m applies to general alternative models. However, its consistency requires $n\delta_p \to \infty$ which excludes the well-known *spiked covariance* model [Johnstone (2001)]. For the simplest spiked model, the covariance matrix can be expressed as $\Sigma = \sigma^2(I_p + hvv')$ where σ^2 and I_p are as before, h is a constant, and v is a unit vector in $\mathbb{R}^p$. Both h and v are unknown parameters. Thus the sphericity hypotheses in (3.1) reduce to

$$H_0 : h = 0 \quad \text{vs.} \quad H_1 : h > 0. \tag{3.2}$$

It is obvious that T_m will asymptotically fail to reject such alternatives since $n\delta_p \to 0$. What's more, this testing problem will become more difficult but attractive when the signal h falls below the threshold $\sqrt{c}$; see Berthet and Rigollet (2013), Onatski, Moreira and Hallin (2013, 2014), Donoho and Jin (2015), and references therein. Hence, applying the CLT for LSSs under elliptical distributions, we discuss a test procedure for (3.2) proposed by Onatski, Moreira and Hallin (2013), which was built under normal populations.

In Onatski, Moreira and Hallin (2013), the authors discussed a likelihood ratio test based on the joint distribution of sample eigenvalues from normal populations. This test was especially designed for the local alternative $H_1 : h \in (0, \sqrt{c})$ and the employed statistic was approximated by a special LSS. In our settings, this LSS can be formulated as

$$T_{\mathrm{LR}}(s) = \int \ln\bigl(z(s) - x\bigr)\, dF^{\check{B}_n}(z) - \int \ln\bigl(z(s) - x\bigr)\, dF^{c_n,\delta_1}(x), \tag{3.3}$$

where $s \in (0, \bar{s})$ is a testing parameter and $z(s) = (1 + s)(c_n + s)/s$. The upper bound $\bar{s}$ of s is chosen as $\bar{s} = \sqrt{c_n}$ for $h \in [0, \sqrt{c_n}]$ and $\bar{s} = h^{-1}c_n$ for $h \in (\sqrt{c_n}, \infty)$ such that $z(s)$ is larger than the limit of $\lambda_{\max}(\check{B}_n)$, the largest sample eigenvalues. Applying Theorem 2.2, one may get the asymptotic distribution of $T_{\mathrm{LR}}(s)$ under general elliptical distributions.

THEOREM 3.3. *Suppose that Assumptions* (a)–(c) [*removing the moment conditions in* (2.5)] *hold. Under the null hypothesis, for any fixed* $s \in (0, \bar{s})$,

$$pT_{\mathrm{LR}}(s) \xrightarrow{D} N\bigl(\mu_s, \sigma_s^2\bigr), \tag{3.4}$$

where the respective mean and variance functions are

$$\mu_s = \frac{1}{2}\ln\bigl(1 - c^{-1}s^2\bigr) + c^{-1}s^2 \quad \text{and} \quad \sigma_s^2 = -2\ln\bigl(1 - c^{-1}s^2\bigr) - 2c^{-1}s^2.$$

The proof of Theorem 3.3 is given in the Supplementary Material [Hu et al. (2018)]. Given a value of $s \in (0, \bar{s})$ and a significance level $\alpha \in (0, 1)$, the test $T_{\mathrm{LW}}(s)$ rejects H_0 if $pT_{\mathrm{LW}}(s) < \sigma_s\Phi^{-1}(\alpha) + \mu_s$, where $\Phi(x)$ denotes the standard normal distribution function. Unlike Onatski, Moreira and Hallin (2013), the theoretical power of this test is not available at present since $pT_{\mathrm{LR}}(s)$ is not a likelihood ratio statistic in elliptical distributions. Another reason is that Theorem 2.2 is inapplicable to this situation since the spatial-sign sample is not anymore elliptically distributed under H_1.

Let us take a step back and only consider the testing problem in elliptical distributions satisfying (2.1). For simplicity, we assume σ is known and set $\sigma = 1$, so that the test $T_{\mathrm{LR}}(s)$ is still valid by simply substituting the sample covairance B_n into $\check{B}_n$, that is,

$$\widetilde{T}_{\mathrm{LR}}(s) = \int \ln\bigl(z(s) - x\bigr)\, dF^{B_n}(z) - \int \ln\bigl(z(s) - x\bigr)\, dF^{c_n,\delta_1}(x),$$

whose asymptotic distribution under both the null and alternative hypotheses is described in the following theorem.

THEOREM 3.4. *Suppose that Assumptions* (a)–(c) *hold. Let* h_0 *be the true value of* h *and* $\sigma = 1$, *then for any fixed* $s \in (0, \bar{s})$,

$$p\widetilde{T}_{\mathrm{LR}}(s) \xrightarrow{D} N\bigl(\tilde{\mu}_s(h_0), \tilde{\sigma}_s^2\bigr), \tag{3.5}$$

TABLE 2
Number of stocks in each NAICS sector

Sector	1	2	3	4	5	6	7	8	9	10	11
Number of stocks	30	32	189	17	36	14	37	65	14	17	11

where the respective mean and variance functions are

$$\tilde{\mu}_s(h) = \frac{1}{2}\ln\bigl(1 - c^{-1}s^2\bigr) + (1 - \tau/2)c^{-1}s^2 + \ln\bigl(1 - c^{-1}sh\bigr),$$
$$\tilde{\sigma}_s^2 = -2\ln\bigl(1 - c^{-1}s^2\bigr) - (2 - \tau)c^{-1}s^2.$$

This theorem is a direct conclusion of Theorem 2.2. It proof is similar to that of Theorem 3.3 and we thus omit it here. From Theorem 3.4, the power function of $\widetilde{T}_{\mathrm{LW}}(s)$ is

$$P_{H_1}\bigl(\widetilde{T}_{\mathrm{LW}}(s) \text{ reject } H_0\bigr) = \Phi\biggl[\Phi^{-1}(\alpha) - \frac{\tilde{\mu}_s(h_0) - \tilde{\mu}_s(0)}{\tilde{\sigma}_s}\biggr], \qquad h_0 > 0.$$

For normal populations ($\tau = 2$), this power function reaches its maximum at $s = h_0$, which agrees with (5.1) in Proposition 9 of Onatski, Moreira and Hallin (2013). In general, the maximizer may not locate at h_0. An interesting case is $\tau = 0$, for which the power function tends to 1 as $s \to 0^+$. This is from the fact that

$$\tilde{\mu}_s(h_0) - \tilde{\mu}_s(0) = -c^{-1}sh_0 + o(s) \quad \text{and} \quad \sigma_s^2 = 2c^{-2}s^4 + o\bigl(s^4\bigr).$$

At this time, $\widetilde{T}_{\mathrm{LW}}(s)$ will successfully detect any positive h_0 as long as s is close to zero.

4. An empirical study. For illustration, we apply the test procedure based on T_m to analyze weekly returns of the stocks from S&P 500. The tests T_{LR} and $\widetilde{T}_{\mathrm{LR}}$ are not included in this analysis since there is a lack of evidence to fit the data using the simplest spiked model. According to the North American Industry Classification System (NAICS), which is used by business and government to classify business establishments, the 500 stocks can be divided into 20 sectors. Nine of them are removed from our analysis since their numbers of stocks are all less than 10. The remaining 11 sectors as well as their numbers of stocks are listed in Table 2. Usually the stocks in the same sector are correlated, and the stocks in different sectors are uncorrelated. So it is expected that the weekly returns of stocks in the same sector are not spherically distributed, and it is interesting to see if the weekly returns of stocks in different sectors are spherically distributed. In the following, we apply T_m to stocks in the same sector and stocks in different sectors, respectively.

The original data are the closing prices or the bid/ask average of these stocks for the trading days in the first half of 2013, that is, from 1 January 2013 to 30 June

2013, with total 124 trading days. This dataset is downloadable from the Center for Research in Security Prices Daily Stock in Wharton Research Data Services. The testing model is established as follows. Denote p_l as the number of stocks in the lth sector, $u_{ij}(l)$ as the price of the ith stock in the lth sector on Wednesday of the jth week. The reason that we choose Wednesday's price here is to avoid the weekend effect in stock market. Thus we get 22 returns for each stock. In order to meet the condition of the proposed test, the original data $u_{ij}(l)$ should be transformed by logarithmic difference, which is a very commonly used procedure in finance. There are a number of theoretical and practical advantages of using logarithmic returns. One of them is that the sequence of logarithmic returns are independent of each other for big time scales [e.g., ≥1 day, see Rama (2001)]. Denote $x_{ij}(l) = \ln(u_{i(j+1)}(l)/u_{ij}(l))$, $j = 1, \dots, 21$ and $X(l) = (x_{ij}(l))_{p_l \times n}$, where $n = 21$ is the sample size.

Now applying T_m to the dataset $X(l)$, $l = 1, \dots, 11$, respectively, we obtain 11 p-values, which are all below 10^{-9}. Therefore, we have strong evidence to believe that stocks in the same sector are not spherically distributed. This is consistent with our intuition. Next, we consider stocks in different sectors. Specifically, we choose one stock from each sector to make up a group of 11 cross-sectoral stocks and then test whether these stocks are spherically distributed. Because there are about 9.79×10^{15} different groups, we just randomly draw 1,000,000 groups from them to analyze. It turns out that the largest p-value is 0.3889, 231 p-values are bigger than 0.05, and 69 p-values are bigger than 0.1. These results again demonstrate that, when the number of stocks is not very small, it is hard to say weekly logarithmic returns for the stocks are spherically distributed. It is also very interesting to analyze these spherically distributed stocks in different sectors, which have almost the same variances.

5. Proof of Theorem 2.2.

The proof of Theorem 2.2 relies on analyzing the resolvent of the sample covariance matrix B_n and the general strategy follows the approach in Bai and Silverstein (2004). Also see Bai et al. (2015) and Gao et al. (2017) for recent developments. However, as we are dealing with the new model equipped with nonlinear dependency, all technical steps of implementing this strategy have to be updated, or at least revalidated. They are presented in this section.

5.1. *Sketch of the proof of Theorem* 2.2. Let $v_0 > 0$ be arbitrary, x_r any number greater than the right end point of interval (2.9) and x_l any negative number if the left end point of (2.9) is zero; otherwise, choose $x_l \in (0, \liminf_{p\to\infty} \lambda_{\min}^{\Sigma}(1 - \sqrt{c})^2)$. Let $\mathcal{C}_u = \{x \pm i v_0 : x \in [x_l, x_r]\}$ and define a contour $\mathcal{C}$

$$\mathcal{C} = \{x + iv : x \in \{x_r, x_l\}, v \in [-v_0, v_0]\} \cup \mathcal{C}_u. \tag{5.1}$$

By definition, this contour encloses a rectangular region in the complex plane containing the support of the LSD $F^{c,H}$. Denote by $m_n(z)$ and $m_{F^{c_n,H_p}}(z)$ the Stieltjes transforms of the ESD F^{B_n} and the LSD F^{c_n,H_p}, respectively. Their companion Stieltjes transforms are given by

$$\underline{m}_n(z) = -\frac{1-c_n}{z} + c_n m_n(z) \quad \text{and} \quad \underline{m}_{F^{c_n,H_p}}(z) = -\frac{1-c_n}{z} + c_n m_{F^{c_n,H_p}}(z).$$

With these notation, we define an empirical process on $\mathcal{C}$ as

$$M_n(z) = p\big[m_n(z) - m_{F^{c_n,H_p}}(z)\big] = n\big[\underline{m}_n(z) - \underline{m}_{F^{c_n,H_p}}(z)\big], \quad z \in \mathcal{C}.$$

Since $f_\ell, \ell = 1, \dots, k$, in Theorem 2.2 are analytic on an open region containing the interval (2.9) (thus analytic on the region enclosed by $\mathcal{C}$), by Cauchy's integral formula, we have for any k complex numbers $a_1, \dots, a_k$,

$$\sum_{\ell=1}^{k} p a_\ell \int f_\ell(x)\, dG_n(x) = -\sum_{\ell=1}^{k} \frac{a_\ell}{2\pi i} \oint_{\mathcal{C}} f_\ell(z) M_n(z)\, dz,$$

when all sample eigenvalues fall in the interval (x_l, x_r), which is correct with overwhelming probability. In order to remove the small probability event that some sample eigenvalues fall outside the interval, we need a truncated version of $M_n(z)$, denoted by $\widehat{M}_n(z)$. Specifically, let $\{\varepsilon_n\}$ be a sequence decreasing to zero satisfying $\varepsilon_n > n^{-a}$ for some $a \in (3/4, 1)$. The truncated process $\widehat{M}_n(z)$ for $z = x + iv \in \mathcal{C}$ is given by

$$\widehat{M}_n(z) = \begin{cases} M_n(z), & z \in \mathcal{C}_n, \\ M_n\big(x + in^{-1}\varepsilon_n\big), & x \in \{x_l, x_r\}, v \in \big[0, n^{-1}\varepsilon_n\big], \\ M_n\big(x - in^{-1}\varepsilon_n\big), & x \in \{x_l, x_r\}, v \in \big[-n^{-1}\varepsilon_n, 0\big], \end{cases} \tag{5.2}$$

where

$$\mathcal{C}_n = \mathcal{C}_u \cup \big\{x \pm iv : x \in \{x_l, x_r\}, v \in \big[n^{-1}\varepsilon_n, v_0\big]\big\},$$

on which $\widehat{M}_n(z)$ agrees with $M_n(z)$, is a regularized set of $\mathcal{C}$ excluding a small segment near the real line. Then we have the following.

LEMMA 5.1. *Under the same assumptions in Theorem 2.2, we have for any* $\ell > 0$,

$$\oint_{\mathcal{C}} f_\ell M_n(z)\, dz = \int_{\mathcal{C}_n} f_\ell(z) \widehat{M}_n(z)\, dz + o_p(1).$$

The proof of this lemma will be put in the Supplementary Material Hu et al. (2018). Hence, Theorem 2.2 follows by similar arguments on pages 562–563 in Bai and Silverstein (2004) and the following lemma.

LEMMA 5.2. *Under Assumptions* (a)–(c), *the random process* $\widehat{M}_n(\cdot)$ *converges weakly to a two-dimensional Gaussian process* $M(\cdot)$ *with the mean function*

$$EM(z) = \int \frac{c(\underline{m}'(z)t)^2\,dH(t)}{\underline{m}(z)(1+t\underline{m}(z))^3} + (\tau-2)\int \frac{(z\underline{m}(z)+1)\underline{m}'(z)t\,dH(t)}{(1+t\underline{m}(z))^2} \tag{5.3}$$

and covariance function

$$\operatorname{Cov}\bigl(M(z_1), M(z_2)\bigr) = \frac{2\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2} - \frac{2}{(z_1-z_2)^2} + \frac{\tau-2}{c}\bigl(\underline{m}(z_1)+z_1\underline{m}'(z_1)\bigr)\bigl(\underline{m}(z_2)+z_2\underline{m}'(z_2)\bigr), \tag{5.4}$$

where $z, z_1, z_2 \in \mathcal{C}$.

The proof of this lemma is the main task of this section and can be achieved by four steps as described below. Notice that in the proof we will use several inequalities frequently, which are presented as lemmas in the Appendix. We will show how and where to use these lemmas in the following. Write for $z \in \mathcal{C}_n$,

$$\begin{aligned}\widehat{M}_n(z) &= p\bigl[m_n(z) - Em_n(z)\bigr] + p\bigl[Em_n(z) - m_{F^{c_n,H_p}}(z)\bigr] \\ &:= M_n^{(1)}(z) + M_n^{(2)}(z).\end{aligned}$$

Step 1: Truncation and rescaling of ξ. This step regularizes the variables $\{\xi_j\}$ in $B_n = \sum_{j=1}^n \xi_j^2 A\mathbf{u}_j\mathbf{u}_j' A'/n$ such that $\{\xi_j\}$ have proper bound for finite (n, p) while maintaining the limiting distribution of the LSSs. Compared with the proof in Bai and Silverstein (2004), this result is entirely new since their model does not include a radius variable at all. Moreover, our key inequalities (Lemmas A.2–A.4) are all built on this regularization, and thus their proofs have to be updated to accommodate the elliptical model.

Step 2: Finite dimensional convergence of $M_n^{(1)}(z)$ in distribution. This step finds the joint limiting distribution of an r-dimensional vector $(M_n^{(1)}(z_\ell))_{1\le\ell\le r}$ by the martingale CLT. Lemmas A.2 and A.3 are used to simplify the expression of the martingale difference and verify Lindeberg's condition, respectively. The (limiting) covariance function is calculated based on Lemma A.1 with the help of Lemma A.3. A new finding here is that the nonlinear dependency comes up with an extra term in the covariance function (Lemma A.1), which results in a novel procedure of proving the convergence of this term.

Step 3: Tightness of $M_n^{(1)}(z)$ on $\mathcal{C}_n$. This step presents the basic idea for establishing the tightness. A key element is the uniform boundedness of $E\|(B_n - zI)^{-q}\|$ for $q > 0$ which is derived by Lemma A.4. By virtue of this and Lemmas A.2–A.4, the tightness can be verified following similar arguments in Bai and Silverstein (2004).

Step 4: Convergence of $M_n^{(2)}(z)$. This final step mainly calculates the limit of $M_n^{(2)}(z)$, the limiting mean function of the LSSs. Akin to deriving the covariance function in Step 2, the nonlinear effect again brings an additional quantity to the mean function. Hence, most work in this part is to handle this new quantity. Note that Lemma A.4 is useful in this step to obtain several convergence results and uniform boundedness on $\mathcal{C}_n$.

These detailed four steps are presented in the next subsection. We note that when simplifying $M_n^{(1)}(z)$ and $M_n^{(2)}(z)$, one more straightforward method is used [see (5.11) and (5.32), resp.], which are different from Bai and Silverstein (2004).

5.2. *Truncation and rescaling of the* ξ*-variable.* From the moment condition $E|(\xi_1^2 - p)/\sqrt{p}|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$ in Assumption (b), we can choose a sequence of $\delta_n > 0$ such that

$$\delta_n \to 0, \qquad \delta_n n^{1/2} \to \infty, \qquad \delta_n^{-2} p^{-1} E\big[(\xi_1^2 - p)^2 I_{\{|\xi_1^2 - p| \geq \delta_n p\}}\big] \to 0. \tag{5.5}$$

Let $\hat{B}_n = \sum_{j=1}^n \hat{\mathbf{x}}_j \hat{\mathbf{x}}_j'/n$ where $\hat{\mathbf{x}}_j = \hat{\xi}_j A \mathbf{u}_j$ with $\hat{\xi}_j = \xi_j I_{\{|\xi_j^2 - p| < \delta_n p\}}$. We then have

$$\begin{aligned} P(\hat{B}_n \neq B_n) &\leq n P\big(|\xi_1^2 - p| \geq \delta_n p\big) \\ &\leq \delta_n^{-2} n p^{-2} E\big[(\xi_1^2 - p)^2 I_{\{|\xi_1^2 - p| \geq \delta_n p\}}\big] \to 0. \end{aligned} \tag{5.6}$$

Define $\tilde{B}_n = \sum_{j=1}^n \tilde{\mathbf{x}}_j \tilde{\mathbf{x}}_j'/n$ where $\tilde{\mathbf{x}}_j = \tilde{\xi}_j A \mathbf{u}_j$ with $\tilde{\xi}_j = \hat{\xi}_j/\sigma_n$ and $\sigma_n^2 = E(\hat{\xi}_1^2)/p$. By the assumptions in (5.5),

$$\begin{aligned} |1 - \sigma_n^2| &= E\big(\xi_1^2/p - 1\big) I_{\{|\xi_1^2 - p| \geq \delta_n p\}} + E I_{\{|\xi_1^2 - p| \geq \delta_n p\}} \\ &\leq \delta_n^{-1}\big(1 + \delta_n^{-1}\big) p^{-2} E\big(\xi_1^2 - p\big)^2 I_{\{|\xi_1^2 - p| \geq \delta_n p\}} = o\big(p^{-1}\big). \end{aligned} \tag{5.7}$$

Therefore, we have

$$E\big(\tilde{\xi}_1^2\big) = p \quad \text{and} \quad E\big(\tilde{\xi}_1^4\big) = \frac{1}{\sigma_n^4}\big(E\big(\xi_1^4\big) - E\xi_1^4 I_{\{|\xi_1^2 - p| \geq \delta_n p\}}\big) = p^2 + \tau p + o(p).$$

On the other hand, write $\mathbf{u}_j = \mathbf{z}_j/\|\mathbf{z}_j\|$ where, and in the following $\mathbf{z}_j \sim N(0, I_p)$ and $\|\cdot\|$ denotes the spectral norm for a matrix, or L_2 norm for a vector. By the strong law of large numbers, for any fixed $0 < \eta < 1$, we have $\max\{\|\mathbf{z}_j\|^2/p : j = 1, \ldots, n\} \geq 1 - \eta$ holds almost surely for large p. Hence we have for large p,

$$\|\tilde{B}_n\| = \left\| \frac{1}{n} \sum_{j=1}^n \frac{\tilde{\xi}_j^2/p}{\|\mathbf{z}_j\|^2/p} A \mathbf{z}_j \mathbf{z}_j' A' \right\| \leq \frac{(1 + \delta_n)}{(1 - \eta)\sigma_n^2} \|\Sigma\| \left\| \frac{1}{n} \sum_{j=1}^n \mathbf{z}_j \mathbf{z}_j' \right\|$$

almost surely. Thus, from Yin, Bai and Krishnaiah (1988), we know that $\limsup_n \lambda_{\max}^{\tilde{B}_n}$ (and similarly $\limsup_n \lambda_{\max}^{\hat{B}_n}$) are almost surely bounded by $\limsup_p \|\Sigma\|(1 + \sqrt{c})^2$.

Let $\tilde{G}_n(x)$ and $\hat{G}_n(x)$ be the analogues of $G_n(x)$ with the matrix B_n replaced by $\tilde{B}_n$ and $\hat{B}_n$, respectively. From the arguments in Bai and Silverstein (2004) and (5.7), we can get for $f(x,z) = 1/(x-z)$ $(z \in \mathcal{C})$, almost surely,

$$\begin{aligned} p^2 \left| \int f(x)\, d\hat{G}_n(x) - \int f(x)\, d\tilde{G}_n(x) \right|^2 & \\ \leq \left(\sum_{j=1}^p K \left|\lambda_j^{\hat{B}_n} - \lambda_j^{\tilde{B}_n}\right| \right)^2 & \\ \leq 4K^2 \sum_{j=1}^p \left(\sqrt{\lambda_j^{\hat{B}_n}} - \sqrt{\lambda_j^{\tilde{B}_n}}\right)^2 \sum_{j=1}^p \left(\lambda_j^{\hat{B}_n} + \lambda_j^{\tilde{B}_n}\right) & \\ \leq 4K^2 p \left(\lambda_{\max}^{\hat{B}_n} + \lambda_{\max}^{\tilde{B}_k}\right) \frac{1}{n} \sum_{j=1}^n (\hat{\mathbf{x}}_j - \tilde{\mathbf{x}}_j)'(\hat{\mathbf{x}}_j - \tilde{\mathbf{x}}_j) & \\ = 4K^2 p \left(\lambda_{\max}^{\hat{B}_n} + \lambda_{\max}^{\tilde{B}_k}\right) \frac{(1-\sigma_n^2)^2}{\sigma_n^2(1+\sigma_n^2)} \operatorname{tr}(\hat{B}_n) \to 0, & \end{aligned}$$

where K is an upper bound of $|f_x'(x,z)|$. As a consequence of this and (5.6),

$$M_n(z) = p \int f(x)\, dG_n(x) = p \int f(x)\, d\tilde{G}_n(x) + o_p(1).$$

Therefore, we only need to find the limiting distribution of $\int f(x)\, d\tilde{G}_n(x)$. For simplicity, we still use B_n, $\mathbf{x}_j$, ξ_j instead of $\tilde{B}_n$, $\tilde{\mathbf{x}}_j$, $\tilde{\xi}_j$, respectively, and assume that

$$(5.8) \quad \forall j, \qquad |\xi_j^2 - p| < \delta_n p, \qquad E(\xi_1^2) = p, \qquad E(\xi_1^4) = p^2 + \tau p + o(p),$$

in the sequel.

5.3. *Finite dimensional convergence of $M_n^{(1)}(z)$ in distribution.* We will show in this part that for any positive integer r and any complex numbers $z_1, \ldots, z_r \in \mathcal{C}_n$, the random vector

$$[M_n^{(1)}(z_1), \ldots, M_n^{(1)}(z_r)]$$

converges to a $2r$-dimensional Gaussian vector. Because of Assumption (c), without loss of generality, we may assume $\|\Sigma\| \leq 1$ for all p. We will denote by K any constant appearing in inequalities and it may take different values at different places.

We first define some quantities which are frequently used in the sequel:

$$r_j = (1/\sqrt{n})\mathbf{x}_j, \qquad D(z) = B_n - zI, \qquad D_j(z) = D(z) - r_j r_j',$$

$$D_{ij}(z) = D(z) - r_i r_i' - r_j r_j', \qquad \varepsilon_j(z) = r_j' D_j^{-1}(z) r_j - \frac{1}{n} \operatorname{tr} \Sigma D_j^{-1}(z),$$

$$\zeta_j(z) = r_j' D_j^{-2}(z) r_j - \frac{1}{n} \operatorname{tr} \Sigma D_j^{-2}(z), \qquad \beta_j(z) = \frac{1}{1 + r_j' D_j^{-1}(z) r_j},$$

$$\bar{\beta}_j(z) = \frac{1}{1 + n^{-1} \operatorname{tr} \Sigma D_j^{-1}(z)}, \qquad b_n(z) = \frac{1}{1 + n^{-1} E \operatorname{tr} \Sigma D_j^{-1}(z)}.$$

Note that, for any $z = u + iv \in \mathbb{C}^+$, the last three quantities are bounded in absolute value by $|z|/v$. Moreover, $D^{-1}(z)$ and $D_j^{-1}(z)$ satisfy

$$D^{-1}(z) - D_j^{-1}(z) = -D_j^{-1}(z) r_j r_j' D_j^{-1}(z) \beta_j(z). \tag{5.9}$$

From Lemma 2.6 in Silverstein and Bai (1995), for any $p \times p$ matrix B,

$$\left|\operatorname{tr}\left(D^{-1}(z) - D_j^{-1}(z)\right) B\right| \le \frac{\|B\|}{v}. \tag{5.10}$$

Let $E_0(\cdot)$ denote expectation, and $E_j(\cdot)$ the conditional given the σ-field generated by $r_1, \ldots, r_j$. Using the martingale decomposition, we can express $M_n^{(1)}(z)$ as

$$\begin{aligned}
&\sum_{j=1}^{n} (E_j - E_{j-1}) \operatorname{tr} D^{-1}(z) \\
&\quad = \sum_{j=1}^{n} (E_j - E_{j-1}) \operatorname{tr}\left[D^{-1}(z) - D_j^{-1}(z)\right] \\
&\quad = -\sum_{j=1}^{n} (E_j - E_{j-1}) \beta_j(z) r_j' D_j^{-2} r_j = \sum_{j=1}^{n} (E_j - E_{j-1}) \frac{d \log(\beta_j(z)/\bar{\beta}_j(z))}{dz},
\end{aligned}$$

where the second equality uses the identity (5.9). By the fact that

$$\beta_j(z) = \bar{\beta}_j(z) - \bar{\beta}_j(z)\beta_j(z)\varepsilon_j(z) = \bar{\beta}_j(z) - \bar{\beta}_j^2 \varepsilon_j(z) + \bar{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z),$$

we have

$$M_n^{(1)}(z) = \frac{d}{dz} \sum_{j=1}^{n} (E_j - E_{j-1}) \log\left[1 - \bar{\beta}_j(z)\varepsilon_j(z) + \bar{\beta}_j(z)\beta_j(z)\varepsilon_j^2(z)\right]. \tag{5.11}$$

Notice that for all $j > 0$ and any $z \in \mathcal{C}_n$, $\bar{\beta}_j(z)\varepsilon_j(z)$ and $\bar{\beta}_j(z)\beta_j(z)\varepsilon_j^2(z)$ are almost surely away from 1 when n is large enough. In addition, by Lemma A.2 and Burkholder's inequality [Lemma 2.1 in Bai and Silverstein (2004)], we have

$$E\left|\sum_{j=1}^{n} (E_j - E_{j-1}) \bar{\beta}_j(z)\beta_j(z)\varepsilon_j^2(z)\right|^2 = O\left(\delta_n^2\right) \to 0.$$

Therefore, applying Taylor expansion,

$$M_n^{(1)}(z) = -\frac{d}{dz}\sum_{j=1}^{n}(E_j - E_{j-1})\bar{\beta}_j(z)\varepsilon_j(z) + o_p(1)$$
$$= -\frac{d}{dz}\sum_{j=1}^{n} E_j\big(\bar{\beta}_j(z)\varepsilon_j(z)\big) + o_p(1).$$

For any $\epsilon > 0$,

$$\sum_{j=1}^{n} E\left|E_j\frac{d}{dz}\varepsilon_j(z)\bar{\beta}_j(z)\right|^2 I_{(|E_j\frac{d}{dz}\varepsilon_j(z)\bar{\beta}_j(z)|\geq\epsilon)}$$
$$\leq \frac{1}{\epsilon^2}\sum_{j=1}^{n} E\left|E_j\frac{d}{dz}\varepsilon_j(z)\bar{\beta}_j(z)\right|^4$$
$$\leq \frac{K}{\epsilon^2}\sum_{j=1}^{n}\left(\frac{|z|^4 E|\zeta_j(z)|^4}{v^4} + \frac{|z|^8 p^4 E|\varepsilon_j(z)|^4}{v^{16}n^4}\right),$$

which tends to zero according to Lemma A.3, and thus Lindeberg's condition is verified. Therefore, from the martingale CLT [Theorem 35.12 Billingsley (1995)], the random vector $(M_n^{(1)}(z_j))$ tends to a $2r$-dimensional zero-mean Gaussian vector $(M(z_j))$ with covariance function $\mathrm{Cov}(M(z_1), M(z_2))$ being

$$\lim_{n\to\infty}\sum_{j=1}^{n}\frac{\partial^2}{\partial z_1\,\partial z_2}E_{j-1}\big(E_j\varepsilon_j(z_1)\bar{\beta}_j(z_1)\cdot E_j\varepsilon_j(z_2)\bar{\beta}_j(z_2)\big), \tag{5.12}$$

provided that this limit exits in probability. The same argument on page 571 of Bai and Silverstein (2004) implies that it suffices to show

$$\sum_{j=1}^{n} E_{j-1}\prod_{k=1}^{2} E_j\bar{\beta}_j(z_k)\varepsilon_j(z_k) \tag{5.13}$$

converges in probability. In addition, by the martingale decomposition,

$$\begin{aligned} &E\big|\bar{\beta}_j(z) - b_n(z)\big|^2 \\ &\quad = \big|b_n(z)\big|^2 n^{-2} E\left|\bar{\beta}_1(z)\sum_{k=2}^{n}(E_k - E_{k-1})\,\mathrm{tr}\big(D_1^{-1}(z) - D_{1k}^{-1}(z)\big)\right|^2 \qquad (5.14)\\ &\quad \leq K|z|^4 v^{-6} n^{-1}, \end{aligned}$$

where the inequality is from (5.10).Thus it is sufficient to study the convergence of

$$b_n(z_1)b_n(z_2)\sum_{j=1}^{n} E_{j-1}\big(E_j\varepsilon_j(z_1)E_j\varepsilon_j(z_2)\big), \tag{5.15}$$

whose second mixed partial derivative yields the limit of (5.12). Applying Lemma A.1, we know that

$$(5.15) = n\left(\frac{E\xi^4}{p(p+2)} - 1\right)T_1 + \frac{2E\xi^4}{p(p+2)}T_2, \tag{5.16}$$

where

$$T_1 = b_n(z_1)b_n(z_2)\frac{1}{n^3}\sum_{j=1}^{n} \operatorname{tr}[\Sigma E_j D_j^{-1}(z_1)] \operatorname{tr}[\Sigma E_j D_j^{-1}(z_2)],$$

$$T_2 = b_n(z_1)b_n(z_2)\frac{1}{n^2}\sum_{j=1}^{n} \operatorname{tr}[\Sigma E_j D_j^{-1}(z_1)\Sigma E_j D_j^{-1}(z_2)].$$

We note that the statistic T_2 has the same form as equation (2.8) in Bai and Silverstein (2004), which has been handled under their model. Following their calculations and using Lemmas A.2–A.3 instead, one may get

$$T_2 \xrightarrow{\text{i.p.}} \int_0^{a(z_1,z_2)} \frac{1}{1-z}\,dz, \tag{5.17}$$

where

$$a(z_1,z_2) = \int \frac{c\underline{m}(z_1)\underline{m}(z_2)t^2\,dH(t)}{(1+t\underline{m}(z_1))(1+t\underline{m}(z_2))} = 1 + \frac{\underline{m}(z_1)\underline{m}(z_2)(z_1-z_2)}{\underline{m}(z_2)-\underline{m}(z_1)}$$

and

$$\frac{\partial^2 T_2}{\partial z_1\,\partial z_2} \xrightarrow{\text{i.p.}} \frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2} - \frac{1}{(z_1-z_2)^2}. \tag{5.18}$$

Now we derive the limit of T_1 and its second mixed partial derivative, which is new compared with the linear transform model. Denote

$$\beta_{ij}(z) = \left(1 + r_i' D_{ij}^{-1}(z) r_i\right)^{-1}, \qquad b_1(z) = \left(1 + n^{-1}E\operatorname{tr}\Sigma D_{12}^{-1}(z)\right)^{-1}.$$

By similar proofs of (5.14) and equation (4.3) of Bai and Silverstein (1998), one may get $|b_1(z) - b_n(z)| \le Kn^{-1}$ and $|b_n(z) - E\beta_1(z)| \le Kn^{-1/2}$, respectively. Also, by equation (2.2) of Silverstein (1995) and discussions in Section 5 of Bai and Silverstein (1998), we obtain

$$E\beta_1(z) = -zE\underline{m}_n(z) \quad \text{and} \quad \left|E\underline{m}_n(z) - \underline{m}_{F^{c_n,H_p}}(z)\right| \le Kn^{-1},$$

respectively. Therefore, we get

$$\left|b_1(z) + z\underline{m}_{F^{c_n,H_p}}(z)\right| \le Kn^{-1/2}. \tag{5.19}$$

With the quantity $b_1(z)$, we define a nonrandom matrix $L(z)$ for the purpose of replacing $D_j(z)$ in T_1,

$$L(z) = -zI + \frac{n-1}{n}b_1(z)\Sigma,$$

which satisfies

$$\|L(z)\|^{-1} \le \frac{|b_1^{-1}(z)|}{\Im(zb_1^{-1}(z))} \le \frac{|b_1^{-1}(z)|}{\Im(z)} \le \frac{1+p/(nv)}{v}. \tag{5.20}$$

By the identity $r_i'D_j^{-1}(z) = \beta_{ij}(z)r_i'D_{ij}^{-1}(z)$, we get their difference

$$D_j^{-1}(z) - L^{-1}(z) = b_1(z)R_1(z) + R_2(z) + R_3(z), \tag{5.21}$$

where

$$R_1(z) = -\sum_{i\neq j} L^{-1}(z)(r_ir_i' - n^{-1}\Sigma)D_{ij}^{-1}(z),$$
$$R_2(z) = -\sum_{i\neq j}(\beta_{ij}(z) - b_1(z))L^{-1}(z)r_ir_i'D_{ij}^{-1}(z),$$
$$R_3(z) = -n^{-1}b_1(z)L^{-1}(z)\Sigma\sum_{i\neq j}(D_{ij}^{-1}(z) - D_j^{-1}(z)).$$

For any $p \times p$ (nonrandom) matrix M, from (5.10), (5.20) and Lemma A.3, we get

$$\begin{aligned} &E|\operatorname{tr} R_1(z)M| \\ &\le nE^{1/2}|r_1'D_{12}^{-1}(z)ML^{-1}(z)r_1 - n^{-1}\operatorname{tr}\Sigma D_{12}^{-1}(z)ML^{-1}(z)|^2 \\ &\le n^{1/2}K\|M\|\frac{(1+p/(nv))}{v^2}, \end{aligned} \tag{5.22}$$

$$\begin{aligned} &E|\operatorname{tr} R_2(z)M| \\ &\le nE^{1/2}(|\beta_{12}(z) - b_1(z)|^2)E^{1/2}|r_1'D_{12}^{-1}ML^{-1}(z)r_1|^2 \\ &\le n^{1/2}K\|M\|\frac{|z|^2(1+p/(nv))}{v^5}, \end{aligned} \tag{5.23}$$

$$|\operatorname{tr} R_3(z)M| \le \|M\|\frac{|z|(1+p/(nv))}{v^3}. \tag{5.24}$$

Hence, plugging (5.21) into T_1 and applying the inequalities (5.19), (5.22)–(5.24), we have

$$\begin{aligned} \prod_{k=1}^2 \operatorname{tr} E_jD_j^{-1}(z_k)\Sigma &= \prod_{k=1}^2 \operatorname{tr} L^{-1}(z_k)\Sigma + Q_1(z_1,z_2) \\ &= p^2\prod_{k=1}^2 \frac{1}{z_k}\int\frac{t\,dH_p(t)}{1+t\underline{m}_{F^{c_n,H_p}}(z_k)} + Q_2(z_1,z_2), \end{aligned}$$

where $E|Q_k(z_1, z_2)| \le Kn^{3/2}$, $k = 1, 2$. We thus get

$$T_1 = \prod_{k=1}^{2} \underline{m}_{F^{c_n,H_p}}(z_k) \int \frac{c_n t\, dH_p(t)}{1 + t\underline{m}_{F^{c_n,H_p}}(z_k)} + o_p(1) \xrightarrow{\text{i.p.}} \prod_{k=1}^{2} \bigl(1 + z_k \underline{m}(z_k)\bigr)$$

whose second mixed partial derivative is

$$\partial^2 T_1/(\partial z_1\, \partial z_2) \xrightarrow{\text{i.p.}} \bigl(\underline{m}(z_1) + z_1\underline{m}'(z_1)\bigr)\bigl(\underline{m}(z_2) + z_2\underline{m}'(z_2)\bigr). \tag{5.25}$$

The result in (5.25) can be obtained by Vitali's convergence theorem [Lemma 2.3 in Bai and Silverstein (2004)].

Collecting results in (5.16), (5.18) and (5.25), we finally get

$$\begin{aligned}\operatorname{Cov}\bigl(M(z_1), M(z_2)\bigr) &= \bigl(\underline{m}(z_1) + z_1\underline{m}'(z_1)\bigr)\bigl(\underline{m}(z_2) + z_2\underline{m}'(z_2)\bigr) \\ &\quad + 2\underline{m}'(z_1)\underline{m}'(z_2)/\bigl(\underline{m}(z_1) - \underline{m}(z_2)\bigr)^2 - 2/(z_1 - z_2)^2,\end{aligned}$$

which completes the proof of Step 1.

5.4. *Tightness of* $M_n^{(1)}(z)$. The tightness of $M_n^{(1)}(z)$ can be established by verifying the moment condition (12.51) of Billingsley (1968), that is,

$$\sup_{n, z_1, z_2 \in \mathcal{C}_n} E\bigl|M_n^{(1)}(z_1) - M_n^{(1)}(z_2)\bigr|^2/|z_1 - z_2|^2 < \infty. \tag{5.26}$$

By the martingale decomposition and the equality

$$m_n(z_1) - m_n(z_2) = (z_1 - z_2)p^{-1} \operatorname{tr}\bigl(D^{-1}(z_1)D^{-1}(z_2)\bigr),$$

to show (5.26), it is sufficient to prove the absolute second moment of

$$\sum_{j=1}^{n} (E_j - E_{j-1}) \operatorname{tr}\bigl[D^{-1}(z_1)D^{-1}(z_2)\bigr]$$

is bounded uniformly. We first show the uniformly boundedness of $E\|D^{-q}(z)\|$ on $\mathcal{C}$ for any fixed $q > 0$. Note that $D^{-1}(z)$ is bounded on $z \in \mathcal{C}_u$. While for $z \in \mathcal{C}_l \cup C_r$, applying Lemma A.4 with suitable large s, we have uniformly

$$E\|D^{-1}(z)\|^q \le K + \frac{1}{v^q} P\bigl(\|B_n\| > \eta_r \text{ or } \lambda_{\min}^{B_n} < \eta_l\bigr) \le K + o(1),$$

where $\limsup_p \|\Sigma\|(1 + \sqrt{c})^2 < \eta_r < x_r$ and $x_l < \eta_l < \liminf_p \lambda_{\min}^{\Sigma}(1 - \sqrt{c})^2$. Analogously, $E\|D_j^{-1}(z)\|^q$ has the same order, and we thus get

$$\max\bigl\{E\|D^{-1}(z)\|^q, E\|D_j^{-1}(z)\|^q, E\|D_{ij}^{-1}(z)\|^q\bigr\} \le K_q. \tag{5.27}$$

Then (5.26) can be obtained by the same procedure in Section 3 of Bai and Silverstein (2004), applying Lemmas A.2–A.4 together with (5.27). We omit the details.

5.5. *Convergence of* $M_n^{(2)}(z)$. Next, we will show that for $z \in \mathcal{C}_n$, $\{M_n^{(2)}(z)\}$ converges to (5.3), is bounded and forms a uniformly equicontinuous family.

We first introduce some auxiliary results, which can be verified by applying Lemma A.4 in our theoretical framework through a similar proof of the same statements in Bai and Silverstein (2004). First of all, we note that

$$\sup_{z\in\mathcal{C}_n}\big|E\underline{m}_n(z)-\underline{m}(z)\big|\to 0 \quad\text{and}\quad \sup_{n,z\in\mathcal{C}_n}\big\|V^{-1}(z)\big\|<\infty, \tag{5.28}$$

where $V(z)=E\underline{m}_n(z)\Sigma+I$. Then, for any nonrandom $p\times p$ matrix M,

$$E\big|\operatorname{tr} D^{-1}(z)M-E\operatorname{tr} D^{-1}(z)M\big|^2\le K\|M\|^2. \tag{5.29}$$

Next, there exists a number $\theta\in(0,1)$ such that for all n large enough

$$\sup_{z\in\mathcal{C}_n}\left|c_n\int\frac{(tE\underline{m}_n(z))^2}{(1+tE\underline{m}_n(z))^2}\,dH_p(t)\right|<\theta. \tag{5.30}$$

Lastly, from (4.12) of Bai and Silverstein (2004) and (5.2) in Bai and Silverstein (1998), we have that

$$M_n^{(2)}(z)=-\frac{\underline{m}_{F^{c_n,H_p}}(z)Q_n(z)}{(1-\int\frac{c_nE\underline{m}_n(z)\underline{m}_{F^{c_n,H_p}}(z)t^2\,dH_p(t)}{(1+tE\underline{m}_n(z))(1+t\underline{m}_{F^{c_n,H_p}}(z))})}, \tag{5.31}$$

where

$$\begin{aligned}Q_n(z)&=n\biggl(c_n\int\frac{dH_p(t)}{1+tE\underline{m}_n(z)}+zc_nEm_n(z)\biggr)\\&=nE\beta_1(z)\bigl(r_1'D_1^{-1}(z)V^{-1}(z)r_1-n^{-1}E\operatorname{tr}V^{-1}(z)\Sigma D^{-1}(z)\bigr).\end{aligned} \tag{5.32}$$

From (5.30) and an analog inequality involving $\underline{m}_{F^{c_n,H_p}}(z)$, the denominator of (5.31) is bounded away from zero. Therefore, we need only to study the limit of $Q_n(z)$ for $z\in\mathcal{C}_n$.

For simplicity, we suppress the variable z from expressions in the sequel when there is no confusion. Let $\varrho_1:=\varrho_1(z)=r_1'D_1^{-1}r_1-(1/n)E\operatorname{tr}\Sigma D_1^{-1}$. By the equality,

$$\beta_1=b_n-b_n\beta_1\varrho_1=b_n-b_n^2\varrho_1+b_n^2\beta_1\varrho_1^2, \tag{5.33}$$

we have $Q_n=Q_n^{(1)}+Q_n^{(2)}+Q_n^{(3)}$, where

$$\begin{aligned}Q_n^{(1)}&=b_nE\bigl(\operatorname{tr}D_1^{-1}V^{-1}\Sigma-\operatorname{tr}V^{-1}\Sigma D^{-1}\bigr),\\Q_n^{(2)}&=-nb_n^2E\varrho_1\bigl(r_1'D_1^{-1}V^{-1}r_1-n^{-1}\operatorname{tr}D_1^{-1}V^{-1}\Sigma\bigr),\\Q_n^{(3)}&=nb_n^2E\beta_1\varrho_1^2\bigl(r_1'D_1^{-1}V^{-1}r_1-n^{-1}E\operatorname{tr}V^{-1}\Sigma D^{-1}\bigr).\end{aligned}$$

For $Q_n^{(1)}$, apply (5.9) and (5.33) again,

$$\begin{aligned}
&E \operatorname{tr} V^{-1}\Sigma\big(D_1^{-1} - D^{-1}\big) \\
&\quad = E\beta_1 r_1' D_1^{-1} V^{-1}\Sigma D_1^{-1} r_1 \\
&\quad = b_n n^{-1} E \operatorname{tr} D_1^{-1} V^{-1}\Sigma D_1^{-1} - b_n E\beta_1 \varrho_1 r_1' D_1^{-1} V^{-1}\Sigma D_1^{-1} r_1.
\end{aligned} \tag{5.34}$$

By Lemma (A.2), Hölder's inequality and the fact that $r_1' D_1^{-1} V^{-1}\Sigma D_1^{-1} r_1$, b_n and β_1 are all bounded for $z \in \mathcal{C}_n$, the second term in equation (5.34) is $o(1)$. Analogously, we can get that $Q_n^{(3)} = o(1)$. Together with applying Lemma A.1 to $Q_n^{(1)}$, we finally obtain that

$$\begin{aligned}
Q_n &= -b_n^2 n^{-1}\left(\frac{E\xi^4}{p(p+2)} - 1\right) E \operatorname{tr} D_1^{-1}\Sigma E \operatorname{tr} D_1^{-1} V^{-1}\Sigma \\
&\quad - b_n^2 n^{-1}\left(\frac{2E\xi^4}{p(p+2)} - 1\right) E \operatorname{tr} D_1^{-1} V^{-1}\Sigma D_1^{-1}\Sigma + o(1) \\
&:= -(\tau - 2) c_n^{-1} b_n^2 Q_n^{(4)} - b_n^2 Q_n^{(5)} + o(1),
\end{aligned} \tag{5.35}$$

where $Q_n^{(4)} = n^{-2} E \operatorname{tr} D^{-1}\Sigma E \operatorname{tr} D^{-1} V^{-1}\Sigma$ and $Q_n^{(5)} = n^{-1} E \operatorname{tr} D^{-1} V^{-1}\Sigma \times D^{-1}\Sigma$. The limit of $Q_n^{(5)}$ can be obtained by a similar approach to deriving (4.13)–(4.22) in Bai and Silverstein (2004). It turns out that

$$Q_n^{(5)} = \frac{c_n}{z^2}\int \frac{t^2\, dH_p(t)}{(1 + tE\underline{m}_n)^3}\left(1 - c_n \int \frac{(tE\underline{m}_n)^2\, dH_p(t)}{(1 + tE\underline{m}_n)^2}\right)^{-1} + o(1). \tag{5.36}$$

The quantity $Q_n^{(4)}$ is new under the elliptical model. To study its limit, similar to (5.21), we approximate the matrix $D^{-1}(z)$ by

$$\tilde{L} = -zI + b_n\Sigma.$$

Notice that

$$b_n = E\beta_1 + O\big(n^{-1/2}\big) = -zE\underline{m}_n + O\big(n^{-1/2}\big) \to -z\underline{m},$$

as $n \to \infty$. By (5.28), it follows that $\tilde{L}$ is nonsingular and $\|\tilde{L}^{-1}\|$ is bounded. Then, analogous to (5.21)–(5.24), we have

$$D^{-1} - \tilde{L}^{-1} = b_n \tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3, \tag{5.37}$$

where

$$|E \operatorname{tr} \tilde{R}_1 M| \le n^{1/2} K, \qquad |E \operatorname{tr} \tilde{R}_2 M| \le n^{1/2} K\big(E\|M\|^4\big)^{1/4}, \tag{5.38}$$

$$|\operatorname{tr} \tilde{R}_3 M| \le K\big(E\|M\|^2\big)^{1/2} \tag{5.39}$$

for any $p \times p$ nonrandom matrix M with bounded norm. From (5.37)–(5.38), we have that

$$n^{-1} E \operatorname{tr} D^{-1} \Sigma = -\frac{c_n}{z} \int \frac{t\, dH_p(t)}{1 + tE\underline{m}_n} + o(1), \tag{5.40}$$

$$n^{-1} E \operatorname{tr} D^{-1} V^{-1} \Sigma = -\frac{c_n}{z} \int \frac{t\, dH_p(t)}{(1 + tE\underline{m}_n)^2} + o(1). \tag{5.41}$$

Equations (5.40) and (5.41) imply that

$$Q_n^{(4)} = \frac{c_n^2}{z^2} \int \frac{t\, dH_p(t)}{1 + tE\underline{m}_n} \int \frac{t\, dH_p(t)}{(1 + tE\underline{m}_n)^2} + o(1). \tag{5.42}$$

Combining (5.31), (5.35), (5.36) and (5.42), we finally get

$$\begin{aligned} M_n^{(2)}(z) &= \biggl[(\tau - 2) c_n \int \frac{t \underline{m}_{F^{c_n, H_p}}\, dH_p(t)}{1 + tE\underline{m}_n} \int \frac{t \underline{m}_n^2\, dH_p(t)}{(1 + tE\underline{m}_n)^2} \\ &\qquad + c_n \int \frac{t^2 \underline{m}_{F^{c_n, H_p}} \underline{m}_n^2\, dH_p(t)}{(1 + tE\underline{m}_n)^3} \biggl(1 - c_n \int \frac{(tE\underline{m}_n)^2\, dH_p(t)}{(1 + tE\underline{m}_n)^2} \biggr)^{-1} \biggr] \\ &\qquad \times \biggl(1 - c_n \int \frac{E\underline{m}_n \underline{m}_{F^{c_n, H_p}} t^2\, dH_p(t)}{(1 + tE\underline{m}_n)(1 + t\underline{m}_{F^{c_n, H_p}})} \biggr)^{-1} + o(1) \\ &\to (\tau - 2) \int \frac{(z\underline{m} + 1) \underline{m}' t\, dH(t)}{(1 + t\underline{m})^2} + c \int \frac{(\underline{m}' t)^2\, dH(t)}{\underline{m}(1 + t\underline{m})^3}, \end{aligned}$$

where $\underline{m}' = \underline{m}'(z)$ denotes the derivative of $\underline{m}(z)$ with respective to z.

The boundedness and uniform equicontinuity for $z \in \mathcal{C}_n$ can be verified directly following the arguments on pages 592–593 of Bai and Silverstein (2004). So we omit them here. Then the proof of Theorem 2.2 is complete.

APPENDIX

These lemmas can be viewed as extensions of independent cases. Their proofs are postponed to the Supplementary Material [Hu et al. (2018)].

LEMMA A.1. *Let* $\mathbf{x} = \xi \mathbf{u}$ *where* ξ *and* $\mathbf{u}$ *are defined in Assumption* (b). *Then, for any* $p \times p$ *complex matrices* C *and* $\tilde{C}$,

$$\begin{aligned} & E(\mathbf{x}' C \mathbf{x} - \operatorname{tr} C)(\mathbf{x}' \tilde{C} \mathbf{x} - \operatorname{tr} \tilde{C}) \\ & \quad = \frac{E\xi^4}{p(p+2)} (\operatorname{tr} C \operatorname{tr} \tilde{C} + \operatorname{tr} C\tilde{C}' + \operatorname{tr} C\tilde{C}) - \operatorname{tr} C \operatorname{tr} \tilde{C}. \end{aligned} \tag{A.1}$$

LEMMA A.2. *Let* $\mathbf{x} = \xi \mathbf{u}$ *where* ξ *satisfies* (5.8), *independent of* $\mathbf{u} \sim U(S^{p-1})$, *then for any* $p \times p$ *complex matrix* C *and* $q \ge 2$,

$$E|\mathbf{x}' C \mathbf{x} - \operatorname{tr} C|^q \le K \|C\|^q \delta_n^{q-2} p^{q-1}, \tag{A.2}$$

where K *is a positive constant depending only on* q.

LEMMA A.3. *Let* $r = \xi \mathbf{u}/\sqrt{n}$ *where* ξ *satisfies* (5.8), *independent of* $\mathbf{u} \sim U(S^{p-1})$. *Then, for any nonrandom* $p \times p$ *matrix* C_k, $k = 1, \dots, q_1$ *and* $\tilde{C}_l$, $l = 1, \dots, q_2$, $q_1, q_2 \geq 0$,

$$\left| E\left(\prod_{k=1}^{q_1} r' C_k r \prod_{l=1}^{q_2} (r' \tilde{C}_l r - n^{-1} \operatorname{tr} \tilde{C}_l) \right) \right| \leq K n^{-(1 \wedge q_2)} \delta_n^{(q_2 - 2) \vee 0} \prod_{k=1}^{q_1} \|C_k\| \prod_{l=1}^{q_2} \|\tilde{C}_l\|,$$

where K *is a positive constant depending on* q_1 *and* q_2.

LEMMA A.4. *Suppose* (5.8) *holds. Then, for any positive* s,

$$P(\|B_n\| > \eta_r) = o(n^{-s}),$$

whenever $\eta_r > \limsup_{p \to \infty} \|\Sigma\|(1 + \sqrt{c})^2$. *If* $0 < \liminf_{p \to \infty} \lambda_{\min}^{\Sigma} I_{(0,1]}(c)$ *then*

$$P(\lambda_{\min}^{B_n} < \eta_l) = o(n^{-s}),$$

whenever $0 < \eta_l < \liminf_{p \to \infty} \lambda_{\min}^{\Sigma} I_{(0,1)}(c)(1 - \sqrt{c})^2$.

Acknowledgment. The authors are grateful to Professor Zhidong Bai for discussions on the innovation of high-dimensional elliptical models, and to Professor Jianfeng Yao for discussions on spiked covariance models. They also would like to thank the editors and referees whose careful and detailed comments have led to great improvements of the manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "High-dimensional covariance matrices in elliptical distributions with application to spherical test" (DOI: 10.1214/18-AOS1699SUPP; .pdf). This supplementary material gives a general result for the CLT of the moments of sample eigenvalues, proofs of Theorem 3.3 and Lemmas 2.1, 5.1, A.1–A.4, and additional simulations for assessing the tests T_1, T_2 and T_m.

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. MR1990662

BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051

BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792

BAI, Z. D. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

BAI, Z. D. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

BAI, Z. D., JIANG, D. D., YAO, J.-F. and ZHENG, S. R. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

BAI, Z. D., HU, J., PAN, G. M. and ZHOU, W. (2015). Convergence of the empirical spectral distribution function of Beta matrices. *Bernoulli* **21** 1538–1574. MR3352053

BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. MR1324786

DONOHO, D. and JIN, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* **30** 1–25. MR3317751

EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248

FANG, K. T. and ZHANG, Y. T. (1990). *Generalized Multivariate Analysis*. Springer, Berlin. MR1079542

GAO, J. T., HAN, X., PAN, G. M. and YANG, Y. R. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402

GUPTA, A. K., VARGA, T. and BODNAR, T. (2013). *Elliptically Contoured Models in Statistics and Portfolio Theory*, 2nd ed. Springer, New York. MR3112145

HU, J., LI, W. M., LIU, Z. and ZHOU, W. (2019). Supplement to "High-dimensional covariance matrices in elliptical distributions with application to spherical test." DOI:10.1214/18-AOS1699SUPP.

JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. MR0312619

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

JONSSON, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** 1–38. MR0650926

KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–438. MR0287628

KOTZ, S. (1975). Multivariate distributions at a cross-road. In *Statistical Distributions in Scientific Work* **1** 247–270.

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

LI, W. M. and YAO, J. F. (2018). On structure testing for component covariance matrices of a high-dimensional mixture. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 293–318.

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485

PAINDAVEINE, D. and VERDEBOUT, T. (2016). On high-dimensional sign tests. *Bernoulli* **22** 1745–1769. MR3474832

PAN, G. M. and ZHOU, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244

RAMA, C. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quant. Finance* **1** 223–236.

SCHOTT, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput. Statist. Data Anal.* **51** 6535–6542. MR2408613

SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408

SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534

TIAN, X. T., LU, Y. T. and LI, W. M. (2015). A robust test for sphericity of high-dimensional covariance matrices. *J. Multivariate Anal.* **141** 217–227. MR3390068

WANG, Q. W. and YAO, J. F. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.* **7** 2164–2192. MR3104916

YANG, Y. R. and PAN, G. M. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.* **43** 467–500. MR3316187

YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344

ZOU, C. L., PENG, L. H., FENG, L. and WANG, Z. J. (2014). Multivariate sign-based high-dimensional tests for sphericity. *Biometrika* **101** 229–236. MR3180668

J. HU
KLASMOE AND SCHOOL OF MATHEMATICS & STATISTICS
NORTHEAST NORMAL UNIVERSITY
CHANGCHUN 130024
CHINA
E-MAIL: huj156@nenu.edu.cn

W. LI
SCHOOL OF STATISTICS AND MANAGEMENT
SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS
SHANGHAI 200433
CHINA
E-MAIL: li.weiming@shufe.edu.cn

Z. LIU
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MACAU
MACAU
CHINA
E-MAIL: liuzhi@umac.mo

W. ZHOU
DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117546
SINGAPORE
E-MAIL: stazw@nus.edu.sg