DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 01/13

A Jarque-Bera test for sphericity of a large-dimensional covariance matrix

by

Konstantin Glombek

DISKUSSIONSBEITRÄGE ZUR STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

Albertus-Magnus-Platz, D-50923 Köln, Deutschland

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 01/13

A Jarque-Bera test for sphericity of a large-dimensional covariance matrix

by

Konstantin Glombek[1]

Abstract

This article provides a new test for sphericity of the covariance matrix of a d-dimensional multinormal population $X \sim N_d(\mu, \Sigma)$. This test is applicable if the sample size, $n+1$, and d both go to infinity while $d/n \to y \in (0, \infty)$, provided that the limits of $\mathrm{tr}(\Sigma^k)/d, k = 1, \dots, 8$, are finite. The main idea of this test is to check whether the empirical eigenvalue distribution of a suitably standardized sample covariance matrix obeys the semicircle law. Due to similarities of the semicircle law to the normal distribution, the proposed test statistic is of the type of the Jarque-Bera test statistic. Simulation results show that the new sphericity test outperforms the tests from the current literature for certain local alternatives if y is small.

Keywords: Test for covariance matrix, High-dimensional data, Spectral distribution, Semicircle law, Free cumulant, Jarque-Bera test.

AMS Subject Classification: Primary 62H15, Secondary 62H10.

[1]Seminar für Wirtschafts– und Sozialstatistik, Universität zu Köln, D-50923 Köln

1. Introduction

Since modern systems of data processing allow the storage of huge amounts of data, applications of multivariate data analysis do not only face a large sample size, $N = n+1$, but also a large dimension d of the sample. Unfortunately, this setting often leads to wrong results of the asymptotic $n \to \infty$ with d being fixed. Therefore, the asymptotic $n, d \to \infty$ with $d/n \to y \in (0, \infty)$, known as (n, d)-asymptotics, is of special interest for hypothesis tests dealing with high dimensional data.

In this paper, a new test for sphericity of the covariance matrix of a normal population under these (n, d)-asymptotics is proposed. The main idea of this test is to check whether the empirical eigenvalue distribution of a suitably standardized sample covariance matrix obeys the *semicircle law*. Due to similarities of the semicircle law to the normal distribution, the proposed test will be an omnibus test such as the well-known Jarque-Bera test (see Jarque and Bera [1987]).

This article considers the classical situation of having a sample from a d-dimensional normal population $X \sim N_d(\mu, \Sigma)$ with unknown expectation $\mu \in \mathbb{R}^d$ and unknown positive definite covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$. We want to test for sphericity of the covariance matrix, i.e., we consider the hypothesis $H_0 : \Sigma = \sigma^2 I$ against $H_1 : \Sigma \neq \sigma^2 I$ for some unspecified $\sigma^2 > 0$ (I denotes the identity matrix). It is well known that the corresponding likelihood ratio test degenerates if $d > n$, i.e., $y > 1$ (see, e.g., Muirhead [1982], Section 8.3). The recent literature proposes and analyzes tests for this hypothesis that can cope with large-dimensional samples, such as Chen et al. [2010], Fisher et al. [2010], Ledoit and Wolf [2002], Srivastava [2005, 2007], Srivastava et al. [2011]. All these tests utilize ratios of eigenvalue moments of the sample covariance matrix as test statistic.

The squared sum of the skewness and kurtosis of the empirical eigenvalue distribution of a suitably normalized sample covariance matrix is here used as test statistic. A statistic of this kind has been proposed by Jarque and Bera [1987] to test for normality. They derive it using the Lagrange multiplier technique, which means that their test is asymptotically equivalent to the likelihood ratio test that implies maximum local power for large samples. We will see that the new sphericity test has several similarities to the Jarque-Bera test, including large local power.

One of the first applications of the skewness and kurtosis of an empirical eigenvalue distribution in statistics is Ünsalan [2007]. In this article, the ratio of the skewness and kurtosis of the empirical eigenvalue distribution of a certain random matrix is used to measure deviations from the *quarter circle law* in the context of geostatistics. Statistical properties of this method are not given. The main contribution of the present article is to provide a deeper analysis of these statistical properties and to show how these measures can be used for covariance matrix testing.

The next sections are organized as follows: Section 2 provides all necessary preliminaries followed by a description of the statistical setting in Section 3. Based on the skewness and the kurtosis of the *semicircle law*, the null distribution of the proposed test statistic is derived in Section 4. Consistency and power of the new test are investigated in Section 5, followed by a conclusion in Section 6. Selected proofs are provided in Section A.

2. Preliminaries

This section provides the framework for the *semicircle law test*. We begin with basic facts, assumptions and definitions.

2.1. Basic facts and definitions

The sample shall be denoted by $X_1, \ldots, X_N$, $N = n+1$, and is drawn from a multinormal population $X \sim N_d(\mu, \Sigma)$ with unknown expectation $\mu \in \mathbb{R}^d$ and unknown covariance matrix $\Sigma > 0$. We will assume the following:

Assumptions.

1. The (n, d)-asymptotics are given by $n, d \to \infty$ and $d/n \to y \in (0, \infty)$.
2. The limits
$$\frac{1}{d}\mathrm{tr}(\Sigma^k) \to B_k \in (0, \infty), k = 1, \ldots, 8,$$
exist under the above (n, d)-asymptotics.

The unbiased sample covariance matrix is given by

$$S = \frac{1}{n}\sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})^t,$$

where $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$. The eigenvalues of this matrix can be investigated by means of the *empirical spectral distribution function (ESD)* of S which is defined by

$$F^S(x) = \frac{1}{d}\sum_{i=1}^{d} \mathbb{1}_{(-\infty,x]}(\lambda_i),$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of S and

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$$

for a set A. The limiting behavior of F^S is one of the main concerns of the *spectral analysis of large-dimensional random matrices* or *random matrix theory* for short. The first success in finding a limiting distribution of F^S is due to Marčenko and Pastur [1967]. A consequence of their results is the following: If $X \sim N_d(\mu, \sigma^2 I)$, then F^S converges under the (n, d)-asymptotics almost surely to a non-random distribution function $F_y(x)$ consisting of a point mass at the origin in the case of $y > 1$ and a continuous part, namely

$$\mathrm{d}F_y(x) = \left(1 - \frac{1}{y}\right)^+ \mathrm{d}\delta_0(x) + f_y(x)\mathrm{d}x, \tag{1}$$

where

$$f_y(x) = \frac{1}{2\pi x y \sigma^2}\sqrt{(a_+ - x)(x - a_-)}\mathbb{1}_{[a_-,a_+]}(x),$$

$a_\pm = \sigma^2(1 \pm \sqrt{y})^2$ and δ_s is the Dirac delta function in $s \in \mathbb{R}$. This distribution is known as the *Marčenko-Pastur law (MP law).* Here, we can see the so-called *curse of dimensionality* in more detail: Although the expectation of the MP law (1) equals σ^2, the eigenvalues of S vary in the interval $[a_-, a_+]$. Since the length of this interval increases as y becomes larger, the variance of the eigenvalues of S increases, too.

If $y \searrow 0$, the MP law (1) converges to a point mass in σ^2 due to continuity reasons (see also the introduction in Bai and Silverstein [1998]). Since we are interested in a non-degenerated limiting distribution of the ESD of S as $d/n \to 0$, we standardize S appropriately and obtain the *semicircle law.*

Proposition 1. *Let $X \sim N_d(\mu, \sigma^2 I)$. Then, as $n, d \to \infty$ and $d/n \to y = 0$, the ESD of*

$$S^* := \sqrt{\frac{n}{d}}(S - \sigma^2 I)$$

converges almost surely to a non-random continuous distribution function with density

$$w(x) = \frac{1}{2\pi\sigma^4}\sqrt{4\sigma^4 - x^2}\mathbb{1}_{[-2\sigma^2,2\sigma^2]}(x)\ . \tag{2}$$

This distribution is known as the semicircle law.

Proof. The assertion is a special case of Bai and Yin [1988]. □

The semicircle law can also be deduced from the MP law (1) by a linear transform. Let $\nu_1, \ldots, \nu_d$ be the eigenvalues of S^*. Taking limit under the (n, d)-asymptotics yields

$$\nu_i = \frac{1}{\sqrt{y}}(\lambda_i - \sigma^2) \tag{3}$$

from which we calculate the density of the continuous part of the limiting ESD of S^*:

$$\begin{aligned} w_{y,\sigma}(x) :&= \sqrt{y} f_y\left(\sqrt{y}x + \sigma^2\right) \\ &= \frac{1}{2\pi(\sqrt{y}x + \sigma^2)\sigma^2}\sqrt{4\sigma^4 - (\sigma^2\sqrt{y} - x)^2}\mathbb{1}_{[(-2+\sqrt{y})\sigma^2,(2+\sqrt{y})\sigma^2]}(x) \end{aligned}$$

We see that the limit $y \searrow 0$ of $w_{y,\sigma}(x)$ exists for all $x \in \mathbb{R}$ and equals the semicircle law (2). If we include the point mass in (1), then we obtain a limiting distribution function $W_{y,\sigma}$ satisfying

$$\mathrm{d}W_{y,\sigma}(x) = \begin{cases} \dfrac{(y-1)^+}{y^{3/2}}\mathrm{d}\delta_{-\sigma^2/\sqrt{y}}(x) + w_{y,\sigma}(x)\mathrm{d}x, & y \in (0, \infty), \\ w_{0,\sigma}(x)\mathrm{d}x, & y = 0\ . \end{cases}$$

The next proposition helps us to calculate the first four moments of the distribution $W_{y,1}$.

Proposition 2 (Bai and Silverstein [2010], Lemma 3.1)**.** *The k-th moment of the MP law (1) for $\sigma^2 = 1$ is given by*

$$MP_k := \sum_{m=0}^{k-1} \frac{1}{m+1} \binom{k}{m} \binom{k-1}{m} y^m \ .$$

As a corollary, we obtain:

Corollary 3. *The first four moments $M_k, k = 1, \ldots, 4$, of the distribution $W_{y,1}$ are given by*

$$M_1 = 0, \qquad M_2 = 1, \qquad M_3 = \sqrt{y}, \qquad M_4 = 2 + y \ .$$

Proof. Due to Equation (3), we have for $y > 0$:

$$\begin{aligned}
M_1 &= \frac{1}{\sqrt{y}}(MP_1 - 1) = 0, \\
M_2 &= \frac{1}{y}(MP_2 - 2MP_1 + 1) = \frac{1}{y}(1 + y - 2 \cdot 1 + 1) = 1, \\
M_3 &= \frac{1}{y^{3/2}}(MP_3 - 3MP_2 + 3MP_1 - 1) \\
&= \frac{1}{y^{3/2}}(1 + 3y + y^2 - 3(1 + y) + 3 \cdot 1 - 1) = \sqrt{y}, \\
M_4 &= \frac{1}{y^2}(MP_4 - 4MP_3 + 6MP_2 - 4MP_1 + 1) \\
&= \frac{1}{y^2}(1 + 6y + 6y^2 + y^3 - 4(1 + 3y + y^2) + 6(1 + y) - 4 \cdot 1 + 1)) = 2 + y \ .
\end{aligned}$$

If $y = 0$, we clearly have $M_1 = M_3 = 0$ due to symmetry. From Lemma 2.1 in Bai and Silverstein [2010], we obtain that $M_2 = 1, M_4 = 2$ for $y = 0$. □

We also obtain from the corollary that the skewness of the distribution $W_{y,\sigma}$ equals $\sqrt{y}$ and the kurtosis of it is given by $2 + y$. In the following, these values will be the null values of the skewness and kurtosis of the limiting ESD of S^*.

2.2. Free cumulants

In this section, all necessary information about *free cumulants* is briefly provided so that the connection between the semicircle law and the normal distribution becomes clearer and the use of the proposed test statistic is justified.

Let $f(t)$ be the characteristic function of some random variable X and $\phi(t) := \ln f(t)$. Assume that X has moments up to order $k \in \mathbb{N}$. Then, the k-th classical cumulant of X, denoted by C_k, can be calculated by

$$C_k = \frac{\phi^{(k)}(0)}{i^k},$$

where $i^2 = -1$ and $\phi^{(k)}$ is the k-th derivative of ϕ. Let us define $A_k := \mathbb{E}(X^k)$. Then, we obtain for $k = 1, \ldots, 4$:

$$\begin{aligned}
C_1 &= A_1, \\
C_2 &= A_2 - A_1^2, \\
C_3 &= A_3 - 3A_1A_2 + 2A_1^3, \\
C_4 &= A_4 - 4A_1A_3 - 3A_2^2 + 12A_1^2A_2 - 6A_1^4 \ .
\end{aligned}$$

This way, the skewness and kurtosis of a non-degenerated X, denoted as $\gamma_1(X)$ and $\gamma_2(X)$, can be expressed in terms of these cumulants as:

$$\begin{aligned}
\gamma_1(X) &= \frac{\mathbb{E}((X - A_1)^3)}{(\mathbb{E}((X - A_1)^2))^{3/2}} = \frac{C_3}{C_2^{3/2}}, \\
\gamma_2(X) &= \frac{\mathbb{E}((X - A_1)^4)}{(\mathbb{E}((X - A_1)^2))^2} = \frac{C_4}{C_2^2} + 3 \ .
\end{aligned}$$

If $X \sim N(\mu, \sigma^2)$, then $\gamma_2(X) = 3$, which is the reason to call $\gamma_1(X) - 3$ *excess kurtosis*. It is well known that the normal distribution is the only distribution having vanishing cumulants for $k > 2$ and that it is the limiting distribution in many central limit theorems.

The notion of *free probability* is introduced in Voiculescu [1985] in order to deal with non-commutative probability spaces. Such spaces are defined as a pair $(\mathcal{A}, \varphi)$, where $\mathcal{A}$ is some unital algebra and $\varphi : \mathcal{A} \to \mathbb{C}$ a linear functional with $\varphi(1) = 1$. One can think of $\mathcal{A}$ as the set of the considered "random variables" and φ as an analogon to the expectation. Indeed, if we choose $(\mathcal{A}, \varphi) = (L^{\infty-}(\Omega, P), \mathbb{E}(\cdot))$, where $(\Omega, \mathcal{F}, P)$ is a classical probability space and $L^{\infty-}(\Omega, P)$ the algebra of real-valued random variables on $(\Omega, \mathcal{F}, P)$ having finite moments of all orders, usual random variables can be embedded in that context. In random matrix theory, $(\mathcal{A}, \varphi) = (\mathcal{M}_d(L^{\infty-}(\Omega, P)), \mathbb{E}(\mathrm{tr}(\cdot)/d))$ is of special interest, where $\mathcal{M}_d(L^{\infty-}(\Omega, P))$ is the algebra of symmetric $d \times d$ matrices with real-valued random entries having finite moments of all order.

A combinatorial setting of how to define cumulants for non-commutative random variables, so-called *free cumulants*, is developed in Nica and Speicher [2006], Part 2. Define $\tilde{A}_k := \varphi(X^k)$ for some $X \in \mathcal{A}$. Then, the first four free cumulants $\tilde{C}_k$ of X are given by:

$$\begin{aligned}
\tilde{C}_1 &= \tilde{A}_1, \\
\tilde{C}_2 &= \tilde{A}_2 - \tilde{A}_1^2, \\
\tilde{C}_3 &= \tilde{A}_3 - 3\tilde{A}_1\tilde{A}_2 + 2\tilde{A}_1^3, \\
\tilde{C}_4 &= \tilde{A}_4 - 4\tilde{A}_1\tilde{A}_3 - 2\tilde{A}_2^2 + 10\tilde{A}_1^2\tilde{A}_2 - 5\tilde{A}_1^4 \ .
\end{aligned}$$

The derivation of $\tilde{C}_k, k = 1, \ldots, 3$ can be found in Nica and Speicher [2006], Lecture 11, and the one of $\tilde{C}_4$ is given in the appendix. While classical and free cumulants agree up to order $k = 3$, they become different for $k > 3$. Since the semicircle law $W_{0,\sigma}$

arises as the limiting distribution in the *free central limit theorem* and, furthermore, is the only distribution whose free cumulants of an order larger than two vanish (see also Nica and Speicher [2006], Part 2), the semicircle law may be regarded as the free analogue to the normal distribution.

Now, we assume that φ is real-valued and positive and that the elements of $\mathcal{A}$ are self-adjoint. These properties are always fulfilled if $(\mathcal{A}, \varphi) = (L^{\infty-}(\Omega, P), \mathbb{E}(\cdot))$ or $(\mathcal{A}, \varphi) = (\mathcal{M}_d(L^{\infty-}(\Omega, P)), \mathbb{E}(\mathrm{tr}(\cdot)/d))$ (see also Nica and Speicher [2006], Part 1, for definitions of these notions). Then, one can define the skewness of $X \in \mathcal{A}$ in terms of free cumulants as

$$\tilde{\gamma}_1(X) := \frac{\tilde{C}_3}{\tilde{C}_2^{\,3/2}},$$

provided that $\tilde{C}_2 > 0$. If $(\mathcal{A}, \varphi) = (L^{\infty-}(\Omega, P), \mathbb{E}(\cdot))$, then the classical notion of skewness coincides with this new one.

Since the kurtosis of the semicircle law $W_{0,\sigma}$ equals 2, it is natural to define the kurtosis of a non-commutative random variable via free cumulants as

$$\tilde{\gamma}_2(X) := \frac{\tilde{C}_4}{\tilde{C}_2^2} + 2,$$

again provided that $\tilde{C}_2 > 0$. This way, excess kurtosis is given by $\tilde{\gamma}_2(X) - 2$. If we deal with classical random variables, i.e., $(\mathcal{A}, \varphi) = (L^{\infty-}(\Omega, P), \mathbb{E}(\cdot))$, then we have $\tilde{\gamma}_2(X) = \gamma_2(X)$ for $X \in L^{\infty-}(\Omega, P)$. Thus, the kurtosis of classical and non-commutative random variables agree in that case. If $(\mathcal{A}, \varphi) = (\mathcal{M}_d(L^{\infty-}(\Omega, P)), \mathbb{E}(\mathrm{tr}(\cdot)/d))$ is considered instead, we first note that we have for $X \in \mathcal{M}_d(L^{\infty-}(\Omega, P))$ that

$$\mathbb{E}\left(\frac{1}{d}\mathrm{tr}(X)\right) = \frac{1}{d}\sum_{i=1}^{d} \mathbb{E}(\lambda_i) = \mathbb{E}(\Lambda),$$

where we consider the eigenvalues $\lambda_1, \ldots, \lambda_d$ of X (which are classical random variables) as a "sample" which is drawn from a population $\Lambda \in L^{\infty-}(\Omega, P)$. The definition of the ESD already suggests considering eigenvalues as a sample. Note that, in contrast to usual samples, this "sample" does not consist of independent random variables (see, e.g., Tulino and Verdú [2004], p. 16, for a literature overview of the Wishart matrix case). Nevertheless, all the unordered eigenvalues of a random matrix have the same distribution (see again Tulino and Verdú [2004] and the references therein for the Wishart matrix case). Therefore, if we assign such an "eigenvalue population" Λ to each $X \in \mathcal{M}_d(L^{\infty-}(\Omega, P))$, we are again in the situation of $(\mathcal{A}, \varphi) = (L^{\infty-}(\Omega, P), \mathbb{E}(\cdot))$.

3. Statistical setting

We will now consider the hypothesis

$$H_0 : \Sigma = \sigma^2 I \quad \text{vs.} \quad H_1 : \Sigma \neq \sigma^2 I \tag{4}$$

for some unspecified $\sigma^2 > 0$. This test problem is commonly known as a test for sphericity. The more general test problem, which considers some positive definite matrix Σ_0 instead of the identity matrix, can always be obtained by using the transformed sample $Y_i = \Sigma_0^{-1/2} X_i, 1 \leq i \leq N$.

The hypothesis is tested by considering the eigenvalues $\nu_1, \dots, \nu_d$ of S^* as a secondary "sample" derived from $X_1, \dots, X_N$. Under the null, this secondary sample stems from a population with distribution $W_{y,\sigma}$ which is what we are going to test for. Since σ^2 is unspecified, the eigenvalues $\nu_i = \sqrt{n/d}(\lambda_i - \sigma^2)$ cannot be observed because their centralization is unknown. Furthermore, the scaling of the eigenvalues λ_i is also unknown. Therefore, we use location and scale invariant test statistics, which give the same values when inserting either the λ_i or the ν_i. Two possible test statistics are the skewness and kurtosis of the ESD of S^*. We have seen in Section 2.2 that the skewness and kurtosis of the limiting ESD of S^* can be defined as usual and well interpreted in terms of free cumulants if $y = 0$. In this case, the skewness and kurtosis of the limiting ESD of S^* under the null are given by 0 and 2. Regarding Corollary 3, these null values become $\sqrt{y}$ and $2 + y$ respectively for $y > 0$. So, under the null, the skewness and kurtosis of the ESD of S^* must converge to these limits. Note that we have $y \approx d/n > 0$ in any practical situation. Hence, we define the moments, variance, skewness and kurtosis of the ESD of S^* by

$$\overline{\nu}_k := \frac{1}{d}\sum_{i=1}^{d} \nu_i^k, \qquad s_\nu^2 := \overline{\nu}_2 - (\overline{\nu}_1)^2,$$

$$\gamma_1(\nu) := \frac{1}{d}\sum_{i=1}^{d} \left(\frac{\nu_i - \overline{\nu}_1}{s_\mu}\right)^3, \qquad \gamma_2(\nu) := \frac{1}{d}\sum_{i=1}^{d} \left(\frac{\nu_i - \overline{\nu}_1}{s_\mu}\right)^4 .$$

Similarly, we define the same statistics with respect to the $\lambda_i, 1 \leq i \leq d$:

$$\overline{\lambda}_k := \frac{1}{d}\sum_{i=1}^{d} \lambda_i^k, \qquad s_\lambda^2 := \overline{\lambda}_2 - (\overline{\lambda}_1)^2,$$

$$\gamma_1(\lambda) := \frac{1}{d}\sum_{i=1}^{d} \left(\frac{\lambda_i - \overline{\lambda}_1}{s_\lambda}\right)^3, \qquad \gamma_2(\lambda) := \frac{1}{d}\sum_{i=1}^{d} \left(\frac{\lambda_i - \overline{\lambda}_1}{s_\lambda}\right)^4 .$$

Note that $\gamma_i(\nu) = \gamma_i(\lambda), i = 1, 2$. Therefore, the statistics $\gamma_i(\nu), i = 1, 2$, can be calculated without knowing σ^2.

As will be seen in the following, the estimators $\overline{\lambda}_k, k = 1, \dots, 4$, are only n-consistent for the true value B_k but not (n, d)-consistent if $y > 0$. This inconsistency leads to the MP law under the null and to the limiting ESD $W_{y,\sigma}$ of S^*, which is exactly what we want to exploit in the following. It has to be pointed out that the principle of the following test is not to compare some (n, d)-consistent estimator for a parameter of the true limiting eigenvalue distribution of Σ (i.e., the limit of F^Σ under the (n, d)-asymptotics) with a theoretic null value. Such a comparison would not be reasonable in our situation as the limiting eigenvalue distribution of Σ under the null is just a point mass. Due to

Proposition 1, the skewness and kurtosis of this limiting eigenvalue distribution equal 0 and 2, respectively. But (n, d)-consistent estimators for the skewness and kurtosis of the true limiting eigenvalue distribution of Σ must have a degenerated distribution under the null (shown in the appendix). Instead, using the statistics above leads to testing for the *empirical phenomenon* of the distribution $W_{y,\sigma}$ arising under the null. This distribution will be rejected if either the skewness or the kurtosis of the ESD of S^* deviates significantly from $\sqrt{y} \approx \sqrt{d/n}$ or $2 + y \approx 2 + d/n$, respectively. Thus, the estimators $\gamma_1(\nu)$ and $\gamma_2(\nu)$ should be understood as estimators for the skewness and kurtosis of the limiting ESD and not of the true limiting eigenvalue distribution.

It has been pointed out in Section 2.2 that the semicircle law is quite similar to the normal distribution which is why we are going to test for the semicircle law in the same manner as for the normal distribution. It will be shown that a test statistic of the form

$$SL := b\left(\left(\frac{\gamma_1(\nu) - z_1}{\sqrt{\mathbb{V}ar(\gamma_1(\nu))}}\right)^2 + \left(\frac{\gamma_2(\nu) - z_2}{\sqrt{\mathbb{V}ar(\gamma_2(\nu))}}\right)^2\right),$$

where $z_1 \approx \sqrt{d/n}, z_2 \approx 2 + d/n$ and b is a suitable blow-up factor, converges weakly to a χ^2 distribution with two degrees of freedom under the null as $n, d \to \infty$ and $d/n \to y = 0$. Jarque and Bera [1987] propose a statistic of this kind to test for normality. In the following, we show several similarities of the new test to theirs including large local power.

In the next section, we derive the null distribution of SL and see how to extend this distribution to the case of $y > 0$.

4. Distribution of the test statistic

First, the following law of large numbers is provided.

Lemma 4. *Let $X \sim N_d(\mu, \Sigma)$. Then, we have the following probability (n, d)-limits:*

- $\overline{\lambda}_1 \to B_1$
- $\overline{\lambda}_2 \to yB_1^2 + B_2$
- $\overline{\lambda}_3 \to y^2B_1^3 + 3yB_1B_2 + B_3$
- $\overline{\lambda}_4 \to y^3B_1^4 + 6y^2B_1^2B_2 + 4yB_1B_3 + 2yB_2^2 + B_4$

Proof. Yin [1986] shows in Formula 4.14 that these limits hold even almost surely. $\square$

This law of large numbers shows that the estimators $\overline{\lambda}_k, k \geq 2$, are not (n, d)-consistent for B_k unless $y = 0$. As already mentioned, we exploit this fact when using the statistics from Section 3 to test for the distribution $W_{y,\sigma}$. Hence, we consider the $\overline{\lambda}_k$ as what they are: (n, d)-consistent estimators for the limits given in Lemma 4. Note that the limits of Lemma 4 equal the first four moments of the MP law in Proposition 2 if $\Sigma = I$, i.e., $B_i = 1, i = 1, \dots, 4$. A direct calculation from Lemma 4 gives the following corollary.

Corollary 5. *Let $X \sim N_d(\mu, \Sigma)$. Then, we have the following probability limits under the (n,d)-asymptotics:*

- $\gamma_1(\nu) \to \dfrac{y^2B_1^3 + 3yB_1B_2 + B_3 - 3B_1(yB_1^2 + B_2) + 2B_1^3}{((y-1)B_1^2 + B_2)^{3/2}} =: l_{\gamma_1}$
- $\gamma_2(\nu) \to \dfrac{1}{((y-1)B_1^2 + B_2)^2} (y^3B_1^4 + 6y^2B_1^2B_2 + 4yB_1B_3 + 2yB_2^2 + B_4$ $-4B_1(y^2B_1^3 + 3yB_1B_2 + B_3) + 6B_1^2(yB_1^2 + B_2) - 3B_1^4) =: l_{\gamma_2}$

We see that the denominators of the limits l_{γ_1} and l_{γ_2} equal zero if and only if $y = 0$ and $B_2 - B_1^2 = 0$. Due to the Cauchy-Schwarz inequality, the latter condition is equivalent to the null hypothesis. Under the null, the numerator and denominator of l_{γ_1} become $\sigma^6 y^2$ and $(\sigma^2 y)^{3/2}$ so that the denominator cancels down. Similarly, the denominator of l_{γ_2} also cancels down under the null hypothesis. Thus, the probability limits of $\gamma_i(\nu), i = 1, 2$, are finite under the null and the alternative for $y \geq 0$.

If $\Sigma = \sigma^2 I$, then l_{γ_1} and l_{γ_2} equal $\sqrt{y}$ and $2 + y$, respectively. But the null hypothesis $\Sigma = \sigma^2 I$ may not be necessary for obtaining these limits, which can imply certain inconsistencies of the tests as will be seen in Section 5.1. However, it will be shown that the sphericity tests based on SL and $\gamma_2(\nu)$ are consistent on the whole alternative if $y \geq 1$. Further, consistency of the SL test for $0 < y < 1$ can also be conjectured.

Now, we compute the asymptotic null distribution of $(\gamma_1(\nu), \gamma_2(\nu))$.

Theorem 6. *Let $X \sim N_d(\mu, \sigma^2 I)$. Then, as $n, d \to \infty$ and $d/n \to y \in (0, \infty)$, the random vector*

$$d \begin{pmatrix} \gamma_1(\nu) - z_1 \\ \gamma_2(\nu) - z_2 \end{pmatrix},$$

where

$$z_1 = \left(\frac{d}{d+1}\right)^{3/2} \left(\sqrt{\frac{d}{n}} + \frac{3}{\sqrt{nd}}\right),$$

$$z_2 = \left(\frac{d}{d+1}\right)^2 \left(2 + \frac{d}{n} + \frac{5}{d} + \frac{6}{n}\right),$$

converges weakly to a normally distributed random vector with mean zero and covariance matrix

$$\begin{pmatrix} 6 + 9y & 24\sqrt{y}(1+y) \\ 24\sqrt{y}(1+y) & 8 + 96y + 64y^2 \end{pmatrix}.$$

Proof. See the appendix. □

We see that the centralizations of $\gamma_1(\nu)$ and $\gamma_2(\nu)$, z_1 and z_2, both converge to the null values $\sqrt{y}$ and $2 + y$ under the (n,d)-asymptotics. The reason for using these centralizations, and not $\sqrt{d/n}$ and $2 + d/n$ as one might expect, is that the empirical moments $\overline{\lambda}_k, k \geq 2$, are biased from their probability limits given by Lemma 4 (with $y = d/n$). Since this bias is of the order $\mathcal{O}(n^{-1})$, it is not negligible when blown up by a factor of d (see the appendix for more details). As mentioned in Section 3, we are only interested in the null hypothesis (4). Thus, a bias correction, which would be helpful if null matrices other than a multiple of the identity were of interest, is not necessary.

Let us define

$$\begin{aligned} \gamma_1^*(\nu) &:= \frac{d\,(\gamma_1(\nu) - z_1)}{\sqrt{6 + 9d/n}}, \\ \gamma_2^*(\nu) &:= \frac{d\,(\gamma_2(\nu) - z_2)}{\sqrt{8 + 96d/n + 64(d/n)^2}}, \\ SL &:= \left(\gamma_1^*(\nu)\right)^2 + \left(\gamma_2^*(\nu)\right)^2 \ . \end{aligned}$$

Under the null, $\gamma_i^*(\nu), i = 1, 2$, are approximately standard normal for large n, d. We reject the null if either $|\gamma_1^*(\nu)| > \Phi^{-1}(1-\alpha/2)$ or $|\gamma_2^*(\nu)| > \Phi^{-1}(1-\alpha/2)$, where Φ is the standard normal cdf and $\alpha \in (0, 1)$ the level of the test. From Theorem 6, we see that $\gamma_1^*(\nu)$ and $\gamma_2^*(\nu)$ become independent as $y \approx d/n$ approaches zero. Thus, SL converges under the null weakly to a χ^2 distribution with two degrees of freedom as $n, d \to \infty$ and $d/n \to 0$. Since $W_{0,\sigma}$ equals the semicircle law, we obtain another similarity of SL to the Jarque-Bera test statistic. However, it is not clear whether the asymptotic distributions of $\gamma_i^*(\nu), i = 1, 2$, and SL exist under the alternative for $y = 0$. For this reason, we will not allow for $y = 0$. Nevertheless, the null distribution of SL will be close to a χ^2 distribution with two degrees of freedom and the ESD of S^* under the null close to the semicircle law as $y \approx d/n > 0$ becomes small.

There is another nice interpretation of the test based on SL: This test leads us to testing whether the third and fourth free cumulant of the limiting ESD of S^* can be zero in the case of small $y \approx d/n$. This can be seen from

$$l_{\gamma_1} - \sqrt{y} \approx \frac{\tilde{C}_3}{\tilde{C}_2^{3/2}}$$

and

$$l_{\gamma_2} - (2 + y) \approx \frac{\tilde{C}_4}{\tilde{C}_2^2},$$

if y is small, where $\tilde{C}_k$ is the k-th free cumulant of the limiting ESD of S^*. Note that the original Jarque-Bera test can be seen as a test whether the third and fourth classical cumulant of some underlying distribution can be zero.

The asymptotic null distribution of SL for $y > 0$ is given by the next theorem.

Theorem 7. *Let $X \sim N_d(\mu, \sigma^2 I)$. Then, under the (n, d)-asymptotics, SL converges weakly to a random variable with distribution function*

$$F_{SL}(x) = \int_0^x \left(\Phi\left(\frac{\sqrt{x - z} - a\sqrt{z}}{\sqrt{1 - a^2}} \right) + \Phi\left(\frac{\sqrt{x - z} + a\sqrt{z}}{\sqrt{1 - a^2}} \right) - 1 \right) f_1(z)\ \mathrm{d}z,$$

where Φ is the standard normal distribution function,

$$a = \frac{24\sqrt{y}(1 + y)}{\sqrt{6 + 9y}\sqrt{8 + 96y + 64y^2}}$$

and $f_1(z)$ is the density of the χ^2 distribution with one degree of freedom.

Proof. See the appendix. $\square$

So, we reject the null hypothesis for large n, d if $SL > F_{SL}^{-1}(1-\alpha)$, where $y \approx d/n$. Note that $0 < a < 1$ for every $y \in (0, \infty)$ so that there will not be a division by zero. But one may consider even the limit $y \to \infty$ or, equivalently, $a \to 1$. Note that

$$\lim_{a\to 1} \Phi\left(\frac{\sqrt{x-z} - a\sqrt{z}}{\sqrt{1-a^2}}\right) = \begin{cases} 1, & x > 2z \\ 0, & x < 2z \end{cases}$$

and

$$\lim_{a\to 1} \Phi\left(\frac{\sqrt{x-z} + a\sqrt{z}}{\sqrt{1-a^2}}\right) = 1 \ .$$

By applying the dominated convergence theorem, we obtain:

$$\lim_{a\to 1} F_{SL}(x) = \int_0^{x/2} f_1(z)\ \mathrm{d}z$$

Thus, the quantiles of F_{SL} for $y = \infty$ are twice the quantiles of the χ^2 distribution with one degree of freedom. We see that the distribution of SL under the null does not degenerate as $d/n \to \infty$. Considering the statistics $\gamma_1^*(\nu)$ and $\gamma_2^*(\nu)$, that is a surprising result. However, it is not clear whether the distribution of SL exists under the alternative for $y = \infty$. We will therefore not consider this case any further.

Remark. The blow-up factor for the statistics $\gamma_i^*(\nu), i = 1, 2$, is chosen to be d and not n. A blow-up by n is obtained by multiplying these statistics by a factor of $n/d \approx 1/y$. This would lead to a multiplication of all variances by a factor of $(n/d)^2 \approx (1/y)^2$ so that all variances would become arbitrarily large as d/n becomes small. Since we want to consider small values of d/n, the d-blow-up is therefore more suitable. Further, the eigenvalues are considered as some kind of a sample. Therefore, the "sample size", which is usually the blow-up factor, is d rather than n.

5. Properties of the test

In this section, we want to analyze consistency and power of the proposed test. We begin with consistency.

5.1. Conditions for consistency

Consistency is usually proved by investigating the distribution of the test statistic under the alternative in order to compute the power function. However, this is a rather complicated strategy in our situation because the parameter to be tested is described in a non-linear way and, furthermore, has an infinite asymptotic dimension. Instead, we adopt a method introduced by Ledoit and Wolf [2002]. This method is advantageous as it does not require any knowledge about the distribution of the statistics under the alternative but about their probability limits in this case. The idea is to investigate whether the probability limits

$$\gamma_1(\nu) - z_1 \to 0, \qquad\qquad \gamma_2(\nu) - z_2 \to 0,$$

or, equivalently,

$$\gamma_1(\nu) \to \sqrt{y}, \qquad\qquad \gamma_2(\nu) \to 2 + y,$$

only hold under the null. From Corollary 5, we know the probability limits of $\gamma_i(\nu)$, $i = 1, 2$, which coincide with the limits above under the null. But if the probability limits of Corollary 5 become the limits above for some covariance matrix from the alternative, then the corresponding test will have an inconsistency. Note that the test based on SL is (n, d)-consistent if either the test based on $\gamma_1^*(\nu)$ or the one based on $\gamma_2^*(\nu)$ is (n, d)-consistent. Thus, we seek for sufficient restrictions for the alternative so that the equations

$$l_{\gamma_1} = \sqrt{y}, \qquad\qquad l_{\gamma_2} = 2 + y,$$

are only solvable for the null hypothesis. Note that the limits l_{γ_i} consist of the limits B_k which have been defined in Section 2.

In the following, we will make use of the Cauchy-Schwarz inequality which implies that $B_{2k} \geq B_k^2, k > 0$, and $B_{2k} = B_k^2$ if and only if $\Sigma = \sigma^2 I$. This way, the null hypothesis is equivalent to $B_{2k} = B_k^2$ for some $k > 0$.

Now, we look for representations of the equations $l_{\gamma_1} - \sqrt{y} = 0$ and $l_{\gamma_2} - (2 + y) = 0$ which consist of non-negative summands which are zero-valued under the null. Further, at least one summand has to be a multiple of the term $B_{2k} - B_k^2$ for some $k > 0$ which ensures that these equations will then only be solvable for a covariance matrix from the null hypothesis.

According to Corollary 5, the probability limit of $\gamma_1(\nu)$ under the alternative can be expressed as:

$$\begin{aligned}
l_{\gamma_1} =& \sqrt{y} \\
\Rightarrow 0 =& l_{\gamma_1}^2 - y \\
\Leftrightarrow 0 =& (y-1)(3B_1B_2(B_3 - B_1B_2) + 2B_1B_3(B_2 - B_1^2) + B_2(B_1B_3 - B_2^2)) \\
& + (1-y)^2(6B_1^2(B_2 - B_1^2)^2 + 2B_1^3(B_3 - B_1^3)) \\
& + 3y(1-y)^2B_1^4(B_2 - B_1^2) + B_3^2 - B_2^3
\end{aligned}$$

Note that $B_3 - B_1B_2 \geq 0, B_3 - B_1^3 \geq 0, B_3^2 - B_2^3 \geq 0$ because of Jensen's inequality. Thus, we have that the sphericity test based on $\gamma_1^*(\nu)$ is (n, d)-consistent in the case of $y > 1$ if

$$B_1B_3 - B_2^2 \geq 0 \tag{5}$$

holds. This restriction is not as strict as it may seem because it is fulfilled by many well-known distributions on the positive real line, e.g., the χ^2, Pareto, exponential and Poisson distribution as well as the MP law.

Now, we take a look at the probability limit of the kurtosis:

$$\begin{aligned} l_{\gamma_2} =& 2 + y \\ \Leftrightarrow 0 =& 4(y-1)B_1(B_3 - B_1B_2) + 4(y-1)^2B_1^2(B_2 - B_1^2) \\ &+ (y-1)(B_2 - B_1^2)^2 + B_4 - B_2^2 \end{aligned}$$

Thus, the sphericity test based on the kurtosis is (n, d)-consistent on the whole alternative if $y \geq 1$. The SL test is therefore also (n, d)-consistent on the whole alternative if $y \geq 1$.

In contrast, finding restrictions for the alternative in the case of $0 < y < 1$ is very difficult. One can derive various sufficient restrictions so that either $l_{\gamma_1}^2 = y$ or $l_{\gamma_2} = 2+y$ can only be fulfilled under the null. As far as the author knows, all these restrictions are complicated and difficult to interpret. This is why this question will be left open for future research. But it seems that there is some duality between the tests based on $\gamma_1^*(\nu)$ and $\gamma_2^*(\nu)$: If one of them lacks consistency, the other one does not. All in all, it is conceivable that the SL test is also consistent if $0 < y < 1$. We will come back to this point in the next subsection.

There are representations of the equations $l_{\gamma_1}^2 = y$ and $l_{\gamma_2} = 2 + y$ from which we can derive restrictions for the alternative so that consistency is achieved for every $y > 0$. We have:

$$\begin{aligned} l_{\gamma_1}^2 =& y \\ \Leftrightarrow 0 =& (B_3 - 3B_1B_2 + 2B_1^3)^2 + 2y^2B_1^3(B_3 - 3B_1B_2 + 2B_1^3) \\ &+ 6y(B_2 - B_1^2)(B_1B_3 - B_2^2 + B_1^4 - B_1^2B_2) \\ &+ 3y(B_2 - B_1^2)(B_2 + (y-1)B_1^2)^2 + 2y(B_2 - B_1^2)^3, \\ l_{\gamma_2} =& 2 + y \\ \Leftrightarrow 0 =& B_4 - 4B_1B_3 - 2B_2^2 + 10B_1^2B_2 - 5B_1^4 \\ &+ 4yB_1(B_3 - 3B_1B_2 + 2B_1^3) + y(B_2 - B_1^2)^2 + 4y^2B_1^2(B_2 - B_1^2) \end{aligned}$$

We see that the skewness test is (n, d)-consistent for all $y > 0$ if the skewness of the true limiting eigenvalue distribution is non-negative and

$$B_1B_3 - B_2^2 + B_1^4 - B_1^2B_2 \geq 0, \tag{6}$$

which is slightly stricter than Restriction (5). Nevertheless, all distributions mentioned above that fulfill Restriction (5) also fulfill Restriction (6). Thus, a wide range of limiting eigenvalue distributions for the alternative is still allowed.

The kurtosis test is (n, d)-consistent for all $y > 0$ if the true limiting eigenvalue distribution exhibits a non-negative third and fourth free cumulant or, equivalently, a non-negative skewness and a kurtosis which is greater than or equal to 2. Thus, if one restricts the alternative according to these restrictions, then these “one-sided tests” based on $\gamma_1^*(\nu)$ (where Restriction (6) is additionally fulfilled) and $\gamma_2^*(\nu)$ are always (n, d)-consistent regardless of y.

5.2. Simulation results

In this subsection, the finite sample properties of the proposed tests are investigated by simulation.

5.2.1. Size and Power

The QQ-plots in Fig. 1 illustrate that the normal approximation of the distribution of the statistics $\gamma_i^*(\nu), i = 1, 2$, under the null fits quite well for small n, d. These plots are made by choosing $n = 100, d = 50, y = d/n$ and simulating $10,000$ realizations of $\gamma_1^*(\nu)$ and $\gamma_2^*(\nu)$ under $X \sim N_d(0, I)$.

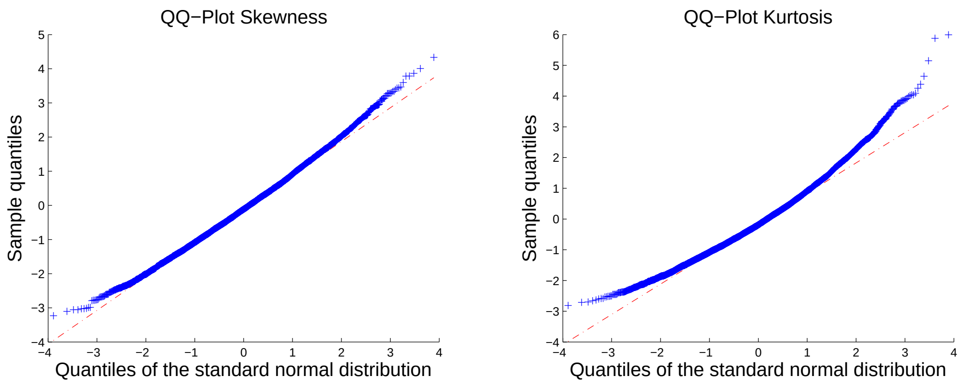

Figure 1: Normal QQ-Plots of $\gamma_1^*(\nu)$ (left) and $\gamma_2^*(\nu)$ (right) for $y = 0.5$ under H_0

It can be seen that $\gamma_1^*(\nu)$ converges faster to a standard normal than $\gamma_2^*(\nu)$. However, the normal approximation for $\gamma_2^*(\nu)$ holds if we do not go too far into the tail of the distribution of $\gamma_2^*(\nu)$.

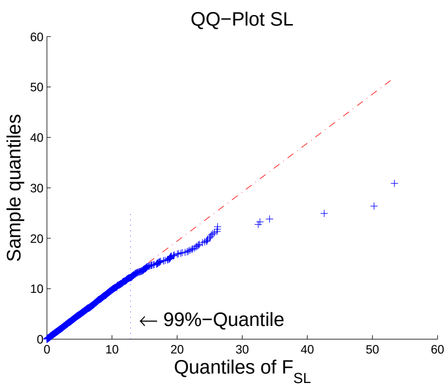

Figure 2: QQ-Plot of SL for $y = 0.5$ under H_0

Next, we have a look at the distribution of the statistic SL which shall be approximated by the distribution given in Theorem 7. Again, we choose $n = 100, d = 50$, $y = d/n$ and simulate $10,000$ realizations of SL under H_0. The QQ-plot in Fig. 2 shows the result of this simulation. We see that the kurtosis affects the weak convergence of SL. But the approximation of the distribution of SL by F_{SL} still remains valid if we do not go too far into the tail again. Note that $F_{SL}^{-1}(0.99) = 12.8514$ for $y = 0.5$ so that all relevant quantiles are well approximated by the asymptotic quantile function F_{SL}^{-1}. Thus, $\gamma_1^*(\nu)$ assures that the weak convergence of SL under the null is of a certain speed while $\gamma_2^*(\nu)$ produces power, as we will see in the following.

	n				
d	25	50	100	200	500
25	0.0464	0.0464	0.0489	0.0474	0.0439
50	0.0541	0.0481	0.0527	0.0515	0.0475
100	0.0523	0.0520	0.0531	0.0523	0.0516
200	0.0537	0.0533	0.0529	0.0534	0.0528
500	0.0557	0.0531	0.0532	0.0506	0.0543

Table 1: Actual sizes of the sphericity test based on $\gamma_1^*(\nu)$

	n				
d	25	50	100	200	500
25	0.0369	0.0403	0.0413	0.0419	0.0454
50	0.0525	0.0498	0.0498	0.0562	0.0498
100	0.0555	0.0560	0.0544	0.0549	0.0536
200	0.0552	0.0535	0.0568	0.0556	0.0565
500	0.0565	0.0555	0.0537	0.0511	0.0537

Table 2: Actual sizes of the sphericity test based on $\gamma_2^*(\nu)$

	n				
d	25	50	100	200	500
25	0.0432	0.0448	0.0461	0.0436	0.0425
50	0.0541	0.0494	0.0519	0.0525	0.0505
100	0.0545	0.0544	0.0532	0.0538	0.0529
200	0.0545	0.0535	0.0555	0.0538	0.0555
500	0.0566	0.0542	0.0542	0.0516	0.0532

Table 3: Actual sizes of the sphericity test based on SL

Next, the actual sizes of the sphericity tests for finite n, d are obtained by simulation. These simulations work as follows: Choose some n, d, a theoretical test level $\alpha \in (0, 1)$ and a large number $m \in \mathbb{N}$. Then, draw m samples of size n from $N_d(0, \sigma^2 I)$ (w.l.o.g. choose $\sigma^2 = 1$) and obtain m realizations of one of the test statistics under the null. Count the number of rejections (see Section 4 for when to reject) and divide this number by m. The result should be near to the theoretical value of α. Tables 1-3 report the results of these simulations for different n, d after $m = 10{,}000$ repetitions and setting $\alpha = 0.05$. We see that the approximations of almost all null distributions are near to the true ones, leading to actual test sizes which are close to the theoretical value of $\alpha = 0.05$.

Thus, in contrast to the slow weak convergence of the Jarque-Bera test statistic to the χ_2^2 distribution (see, e.g., Bowman and Shenton [1975]), the weak convergence of the statistic SL appears to be quite fast.

A comparison of Tables 1 and 2 shows again that $\gamma_1^*(\nu)$ converges faster in distribution to a standard normal than $\gamma_2^*(\nu)$. The case of $n = d = 25$ for $\gamma_2^*(\nu)$, which only exhibits a rejection probability of 0.0369, illustrates this point, which is in line with the results of Fisher et al. [2010] who also use a test statistic based on the fourth empirical eigenvalue moment. The weak convergence of SL is affected by that, but still remains fast enough to obtain reasonable results in Table 3.

Now, we investigate the power of the new sphericity test by Monte Carlo simulation and compare the new test with the ones of Fisher et al. [2010] and John [1971]. It is shown in John [1971] that the sphericity test based on his statistic U is the locally most powerful invariant test. Ledoit and Wolf [2002] further show that this test is applicable under (n, d)-asymptotics. Note that this test agrees with the sphericity test by Srivastava [2005] up to some bias correction. Another (n, d)-consistent sphericity test is introduced by Fisher et al. [2010]. They demonstrate that the test based on their statistic T performs well if $y \geq 1$ and if the alternative is chosen to be near sphericity. They define a near sphericity matrix as

$$\Sigma = \begin{pmatrix} \Theta & \mathbf{0}^t \\ \mathbf{0} & I \end{pmatrix}, \tag{7}$$

where $\Theta \in \mathbb{R}^{k \times k}, k << d$, is a diagonal matrix with diagonal entries unequal to 1, $\mathbf{0} \in \mathbb{R}^{(d-k)\times k}$ is a matrix of zeros and I is the $(d-k) \times (d-k)$ identity matrix. The following tables show the results of Monte Carlo simulations which have been made according to the same principle as explained above, except that Σ is chosen as a near sphericity matrix. We will see that the new test performs very well in this near sphericity case for small values of $y = d/n$. The test level for all following tests is chosen as $\alpha = 0.05$.

	statistic				
d/n=0.01	$\gamma_1^*(\nu)$	$\gamma_2^*(\nu)$	SL	T	U
25/2,500	0.5630	0.6734	0.7094	0.6383	0.5381
50/5,000	0.7287	0.8724	0.8918	0.7025	0.5894
100/10,000	0.8233	0.9619	0.9671	0.7368	0.6140
200/20,000	0.8793	0.9936	0.9933	0.7606	0.6259
500/50,000	0.9110	0.9992	0.9991	0.7695	0.6381

Table 4: Power of the sphericity tests under near sphericity with $k = 1, \Theta = 1.2$

Table 4 reports the results of a Monte Carlo simulation with $k = 1, \Theta = 1.2$, $y = d/n = 0.01$ after $m = 10,000$ repetitions. We observe (n, d)-consistency of all new tests and see that all of them (except the skewness test for $d/n = 25/2,500$) outperform the tests based on T and U. As has already been mentioned, the sphericity test based on the kurtosis leads to more power than the one based on the skewness. Moreover, the test based on SL is even more powerful than the one using $\gamma_2^*(\nu)$. Thus, the squared sum of $\gamma_1^*(\nu)$ and $\gamma_2^*(\nu)$ seems to be overadditive with respect to power. Note that the test problem still can be viewed as a large-dimensional one if $y > 0$ is small (see Bai and Zheng [2007]).

Table 5 reports the results of a further simulation with $k = 2, \Theta = \text{diag}(1.2, 1.3)$, $y = d/n = 0.025$ after $m = 10,000$ repetitions. Except in the case of $d/n = 25/1,000$, we observe once again that the SL test outperforms the other tests. The overadditivity property of the SL test can also be seen.

	statistic				
d/n=0.025	$\gamma_1^*(\nu)$	$\gamma_2^*(\nu)$	SL	T	U
25/1,000	0.5560	0.5803	0.6272	0.7679	0.6695
50/2,000	0.7701	0.8496	0.8714	0.8635	0.7480
100/4,000	0.8835	0.9636	0.9684	0.9026	0.7882
200/8,000	0.9346	0.9930	0.9938	0.9249	0.8109
500/20,000	0.9645	0.9997	0.9995	0.9379	0.8237

Table 5: Power of the sphericity tests under near sphericity with $k = 2, \Theta = \text{diag}(1.2, 1.3)$

The results of Tables 4 and 5 indicate that the SL test is asymptotically locally very powerful for small y. Furthermore, this test seems to become even more superior to the tests from the literature the smaller y becomes.

Next, we compare the tests for larger values of $y = d/n$. Table 6 provides the results of a power simulation after $10,000$ repetitions. The parameters are chosen as $k = 3$, $\Theta = \text{diag}(2, 2, 2), y = d/n = 0.5$. The overadditivity property of the SL test can only be observed for $d = 50, 100$ and the kurtosis test is now outperformed by the skewness test for $d < 100$. While the test based on T outperforms the new ones in smaller dimensions, the kurtosis and SL tests become comparable to this test in larger dimensions. Further, we see that the test based on U is asymptotically less powerful than each of the new ones.

	statistic				
d/n=0.5	$\gamma_1^*(\nu)$	$\gamma_2^*(\nu)$	SL	T	U
25/50	0.4118	0.3843	0.4059	0.5471	0.5671
50/100	0.7048	0.7006	0.7128	0.8097	0.7264
100/200	0.8636	0.8810	0.8826	0.9309	0.8158
200/400	0.9431	0.9638	0.9591	0.9775	0.8631
500/1,000	0.9814	0.9951	0.9933	0.9960	0.8927

Table 6: Power of the sphericity tests under near sphericity with $k = 3, \Theta = \text{diag}(2, 2, 2)$

Now, we look at the case where the sample covariance matrix becomes singular, i.e., $y = d/n > 1$, and set $k = 1, \Theta = 4, y = d/n = 2$. This case has also been considered in Fisher et al. [2010]. From Table 7, we observe (n, d)-consistency of the new tests again and see that the overadditivity property of the SL test is no longer given. Further, the new tests outperform the test based on U, but are less powerful than the test based on T.

	statistic				
d/n=2	$\gamma_1^*(\nu)$	$\gamma_2^*(\nu)$	SL	T	U
50/25	0.5222	0.5566	0.5450	0.5593	0.5033
100/50	0.6738	0.7247	0.7079	0.7646	0.5845
200/100	0.7908	0.8525	0.8321	0.8999	0.6489
500/250	0.8830	0.9445	0.9250	0.9786	0.6854

Table 7: Power of the sphericity tests under near sphericity with $k = 1, \Theta = 4$

The results of this section indicate that the new tests locally dominate the tests from the current literature for small $y = d/n$, i.e., when the ESD of S^* under the null is close to the semicircle law. In this case, the test based on SL becomes overadditive with respect to power compared to the tests based on $\gamma_1^*(\nu)$ and $\gamma_2^*(\nu)$. Further, the statistic SL has an interesting structure concerning the role of its summands: While $(\gamma_1^*(\nu))^2$ ensures that the weak convergence of SL under the null is of a certain speed, $(\gamma_2^*(\nu))^2$ produces power. If $0 << y < 1$, the tests based on $\gamma_2^*(\nu)$ and SL are still comparable to the existing tests, but the overadditivity property of SL cannot be assured any more. If $y \geq 1$, the new tests underperform the tests from the literature.

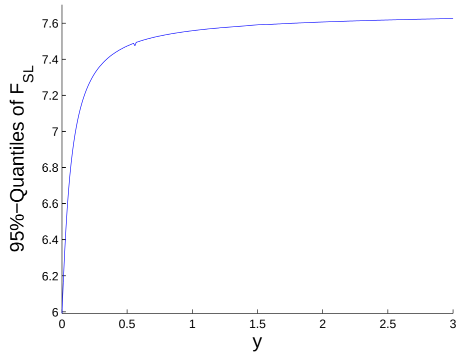

Figure 3: 95% Quantiles of SL as a function of y

These observations are in line with Fig. 3 which shows a plot of the 95% quantiles of the distribution of SL as a function of y. As long as the quantiles are about 6, we have the overadditivity property of SL. The steep slope of the curve in a neighborhood of $y = 0$ leads to rapidly increasing quantiles, which goes hand in hand with the observed loss of the overadditivity property. As y increases, the curve becomes flat and converges to twice the 95% quantile of the χ^2 distribution with one degree of freedom (see Section 4). This convergence comes with the loss of the power superiority of the SL test.

5.2.2. Limitations and further consistency properties of the sphericity tests

In Section 5.1, (n, d)-consistency of the proposed sphericity tests was investigated. While the tests based on $\gamma_2^*(\nu)$ and SL were shown to be (n, d)-consistent for $y \geq 1$, the case of $0 < y < 1$ was left open. Now, this case shall be further investigated by simulation results. Table 8 shows the simulated power of the new tests compared to the tests of Fisher et al. [2010] and John [1971]. Again, the test level is chosen as $\alpha = 0.05$ and the alternative is of the form (7), but with $k = d/2, \Theta = 1.1I, y = d/n = 0.025$ (if $d = 25$, the dimension of Θ is chosen as 12). Such an alternative could be regarded as being "further away" from sphericity.

	statistic				
d/n=0.025	$\gamma_1^*(\nu)$	$\gamma_2^*(\nu)$	SL	T	U
25/1,000	0.0490	0.0503	0.0529	0.2917	0.2946
50/2,000	0.0623	0.0577	0.0646	0.6928	0.7072
100/4,000	0.1112	0.0563	0.0892	0.9959	0.9965
200/8,000	0.3122	0.0544	0.2049	1.0000	1.0000
500/20,000	0.9630	0.0572	0.9187	1.0000	1.0000

Table 8: Power of the sphericity tests with $k = d/2, \Theta = 1.1I$

We see that the kurtosis test lacks consistency, while the skewness test is consistent. Consequently, the SL test is also consistent. The power of the new tests appear to be poor compared to the tests based on T and U.

Table 9 reports the simulated power of the new tests under the alternative (7) with $k = d/2, \Theta = 0.5I, y = d/n = 0.05$. Whereas the skewness test is biased for $d/n = 25/500$, the kurtosis test leads to power of more than 90% so that the SL test gains power of almost 80%. We even see that the kurtosis test and thus the SL test are comparable to the tests based on T and U if $d > 25$.

	statistic				
d/n=0.05	$\gamma_1^*(\nu)$	$\gamma_2^*(\nu)$	SL	T	U
25/500	0.0054	0.9217	0.7875	1.0000	1.0000
50/1,000	0.2026	1.0000	1.0000	1.0000	1.0000
100/2,000	0.9525	1.0000	1.0000	1.0000	1.0000
200/4,000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 9: Power of the sphericity tests with $k = d/2, \Theta = 0.5I$

These examples indicate that the skewness and kurtosis tests complement each other concerning (n, d)-consistency. Hence, one may expect that the SL test is also (n, d)-consistent on the whole alternative for $0 < y < 1$. Further, we have seen that the tests based on T and U outperform the new ones if the alternative is further away from sphericity. In contrast, the performance of the new SL test is superior to the other ones if y is small and the alternative is of a local kind such as the previously discussed near sphericity case. We therefore observe another similarity of the SL test to the Jarque-Bera test: It seems that the SL test is asymptotically locally optimal for small y.

6. Conclusion

This article provides a new approach to test for sphericity of the covariance matrix of a multinormal population under (n, d)-asymptotics. The main idea of this approach is to consider the empirical eigenvalue distribution of a suitably standardized sample covariance matrix which tends to a kind of semicircle law under the null. Since the semicircle law is very similar to the normal distribution, this paper proposes to use a test statistic which is of the type of the well known Jarque-Bera test statistic. This new test can be nicely interpreted in terms of free cumulants and its test statistic exhibits an asymptotic distribution which is comparable to that of the original Jarque-Bera test. It is shown that the new sphericity test is (n, d)-consistent if the limiting ratio of dimension to sample size, y, is greater than or equal to one. Further simulations indicate that (n, d)-consistency for $0 < y < 1$ can also be conjectured. Future research will be dedicated to the proof of this conjecture and the derivation of the distribution of the test statistic under the alternative. Lastly, a comprehensive simulation study shows that the new sphericity test seems to be asymptotically locally optimal if y becomes small.

Acknowledgments

The author is very grateful to Karl Mosler and Dietrich von Rosen for their advice and valuable suggestions. Many thanks also go to Harriet Bergmann, Florian Faulenbach, Philipp Immenkötter and Christof Wiechers for their comments on this manuscript.

A. Proofs

In this appendix, all skipped proofs are provided. We begin with the derivation of the fourth free cumulant $\tilde{C}_4$ from Section 2.2 which will be done by using the moment-cumulant formula (11.7) from p. 176 in Nica and Speicher [2006]. It is given by

$$\tilde{C}_4 = \sum_{\sigma \in NC(4)} \varphi_\sigma(X)\mu(\sigma, 1_4),$$

where $(\mathcal{A}, \varphi)$ is a non-commutative probability space, $X \in \mathcal{A}$, $NC(4)$ the set of all non-crossing partitions of $\{1, 2, 3, 4\}$, $1_4 = \{(1, 2, 3, 4)\} \in NC(4)$, $\varphi_\sigma(X) = \prod_{V \in \sigma} \varphi(X^{|V|})$. Here, $V \in \sigma$ denotes a block of the partition σ, $|V|$ its size and $\mu(\sigma, 1_4)$ the *Möbius function* on $\{1, 2, 3, 4\}$ (see also Nica and Speicher [2006], Lecture 10). Note that the only crossing partition of $\{1, 2, 3, 4\}$ is $\{(1, 3), (2, 4)\}$ so that $NC(4) = \{0_4, \tau_1, \ldots, \tau_{12}, 1_4\}$, where

$$\begin{aligned}
0_4 &= \{(1),(2),(3),(4)\}, & \tau_1 &= \{(1,2),(3),(4)\}, & \tau_2 &= \{(1),(2),(3,4)\},\\
\tau_3 &= \{(1),(2,3),(4)\}, & \tau_4 &= \{(1,4),(2),(3)\}, & \tau_5 &= \{(1,3),(2),(4)\},\\
\tau_6 &= \{(1),(2,4),(3)\}, & \tau_7 &= \{(1,2),(3,4)\}, & \tau_8 &= \{(1,4),(2,3)\},\\
\tau_9 &= \{(1,2,3),(4)\}, & \tau_{10} &= \{(1),(2,3,4)\}, & \tau_{11} &= \{(1,3,4),(2)\},\\
\tau_{12} &= \{(1,2,4),(3)\}\ .
\end{aligned}$$

Now, the values of the Möbius function are computed. Due to Proposition 10.15 in Nica and Speicher [2006], we have $\mu(0_4, 1_4) = -5$. From Remark 10.9 in this book, we obtain

$$\begin{aligned}
&\mu(\tau_1,1_4) = \mu(\tau_2,1_4) = \mu(\tau_3,1_4) = \mu(\tau_4,1_4) = 2,\\
&\mu(\tau_5,1_4) = \mu(\tau_6,1_4) = 1,\\
&\mu(\tau_7,1_4) = \mu(\tau_8,1_4) = \mu(\tau_9,1_4) = \mu(\tau_{10},1_4) = \mu(\tau_{11},1_4) = \mu(\tau_{12},1_4) = -1,\\
&\mu(1_4,1_4) = 1\ .
\end{aligned}$$

All in all, we have:

$$\begin{aligned}
\tilde{C}_4 = &\sum_{\sigma\in NC(4)} \varphi_\sigma(X)\mu(\sigma,1_4)\\
=&\varphi_{0_4}(X)\mu(0_4,1_4) + \varphi_{\tau_1}(X)\mu(\tau_1,1_4) + \varphi_{\tau_2}(X)\mu(\tau_2,1_4)\\
&+ \varphi_{\tau_3}(X)\mu(\tau_3,1_4) + \varphi_{\tau_4}(X)\mu(\tau_4,1_4) + \varphi_{\tau_5}(X)\mu(\tau_5,1_4)\\
&+ \varphi_{\tau_6}(X)\mu(\tau_6,1_4) + \varphi_{\tau_7}(X)\mu(\tau_7,1_4) + \varphi_{\tau_8}(X)\mu(\tau_8,1_4)\\
&+ \varphi_{\tau_9}(X)\mu(\tau_9,1_4) + \varphi_{\tau_{10}}(X)\mu(\tau_{10},1_4) + \varphi_{\tau_{11}}(X)\mu(\tau_{11},1_4)\\
&+ \varphi_{\tau_{12}}(X)\mu(\tau_{12},1_4) + \varphi_{1_4}(X)\mu(1_4,1_4)\\
=&\varphi^4(X)(-5) + \varphi(X^2)\varphi^2(X)2 + \varphi(X^2)\varphi^2(X)2 + \varphi(X^2)\varphi^2(X)2\\
&+ \varphi(X^2)\varphi^2(X)2 + \varphi(X^2)\varphi^2(X) + \varphi(X^2)\varphi^2(X) + \varphi^2(X^2)\\
&+ \varphi^2(X^2) + \varphi(X^3)\varphi(X)(-1) + \varphi(X^3)\varphi(X)(-1)\\
&+ \varphi(X^3)\varphi(X)(-1) + \varphi(X^3)\varphi(X)(-1) + \varphi(X^4)\\
=&\varphi(X^4) - 4\varphi(X)\varphi(X^3) - 2\varphi^2(X^2) + 10\varphi^2(X)\varphi(X^2) - 5\varphi^4(X),
\end{aligned}$$

which is the result of Section 2.2.

Proof of Theorem 6

Now, we come to the skipped proofs of Section 4 and begin with the one from Theorem 6. We need the following theorem.

Theorem 8. *Let $X \sim N_d(\mu, I)$. Then, as $n, d \to \infty$ and $d/n \to y \in (0,\infty)$, the vector*

$$d\begin{pmatrix}\overline{\lambda}_1 - c_1\\ \overline{\lambda}_2 - c_2\\ \overline{\lambda}_3 - c_3\\ \overline{\lambda}_4 - c_4\end{pmatrix},$$

where

$$
\begin{aligned}
c_1 &= 1, \\
c_2 &= 1 + \frac{d}{n} + \frac{1}{n}, \\
c_3 &= 1 + 3\frac{d}{n} + \left(\frac{d}{n}\right)^2 + 3\frac{1}{n} + 3\frac{d}{n^2}, \\
c_4 &= 1 + 6\frac{d}{n} + 6\left(\frac{d}{n}\right)^2 + \left(\frac{d}{n}\right)^3 + 6\frac{1}{n} + 17\frac{d}{n^2} + 6\frac{d^2}{n^3},
\end{aligned}
$$

converges in distribution to a normally distributed vector $Z = (Z_1, Z_2, Z_3, Z_4)^t$ with mean zero, variances

$$
\begin{aligned}
&\mathbb{V}ar(Z_1) = 2y, \qquad \mathbb{V}ar(Z_2) = 4y(2y^2 + 5y + 2), \\
&\mathbb{V}ar(Z_3) = 6y(3y^4 + 24y^3 + 46y^2 + 24y + 3), \\
&\mathbb{V}ar(Z_4) = 8y(4y^6 + 66y^5 + 300y^4 + 485y^3 + 300y^2 + 66y + 4)
\end{aligned}
$$

and covariances

$$
\begin{aligned}
&\mathbb{C}ov(Z_1, Z_2) = 4y(y + 1), \qquad \mathbb{C}ov(Z_1, Z_3) = 6y(y^2 + 3y + 1), \\
&\mathbb{C}ov(Z_1, Z_4) = 8y(y^3 + 6y^2 + 6y + 1), \mathbb{C}ov(Z_2, Z_3) = 12y(y^3 + 5y^2 + 5y + 1), \\
&\mathbb{C}ov(Z_2, Z_4) = 8y(2y^4 + 17y^3 + 32y^2 + 17y + 2), \\
&\mathbb{C}ov(Z_3, Z_4) = 24y(y^5 + 12y^4 + 37y^3 + 37y^2 + 12y + 1) \ .
\end{aligned}
$$

Proof. This result is firstly stated for a standard normal population in Arharov [1971] and corrected by Jonsson [1982] without specifying the mean and the covariances in general. Bai and Silverstein [2004] and Lytova and Pastur [2009] generalize this result and give explicit formulas for the mean and covariances. The calculation of the centralization and the covariances can be found in Bai and Silverstein [2004], Section 5. Since the determinant of the covariance matrix equals $384y^{10}$, this matrix is non-singular unless $y = 0$. □

If we compare $\mathbb{E}(\overline{\lambda}_k)$ with the moments of the MP law MP_k (with $y = d/n$), we see that the estimators $\overline{\lambda}_k, k \geq 2$, are biased from these (null) moments and that the bias is of the order $\mathcal{O}(n^{-1})$.

If $X \sim N_d(\mu, \Sigma)$, then the asymptotic covariances in Theorem 8 will contain the limits $B_i, i = 1, \ldots, 8$ (see also Fisher et al. [2010] and Srivastava [2005] for a derivation method). Thus, the existence of these limits ensures that the distribution of the test statistics under the alternative exists as well.

Now, we obtain Theorem 6 by an application of the delta method.

Proof. Set

$$
f(x_1, x_2, x_3, x_4) = \\
\left(\frac{1}{(x_2 - x_1^2)^{3/2}} \left(x_3 - 3x_1x_2 + 2x_1^3\right), \frac{1}{(x_2 - x_1^2)^2} \left(x_4 - 4x_1x_3 + 6x_1^2x_2 - 3x_1^4\right) \right) \ .
$$

Note that $f(\overline{\lambda}_1, \overline{\lambda}_2, \overline{\lambda}_3, \overline{\lambda}_4) = (\gamma_1(\lambda), \gamma_2(\lambda)) = (\gamma_1(\nu), \gamma_2(\nu))$ and

$$\begin{aligned} f\left(c_1, c_2, c_3, c_4\right) &= \left(\left(\frac{d}{d+1}\right)^{3/2}\left(\sqrt{\frac{d}{n}} + \frac{3}{\sqrt{nd}}\right), \left(\frac{d}{d+1}\right)^{2}\left(2 + \frac{d}{n} + \frac{5}{d} + \frac{6}{n}\right)\right) \\ &= \left(z_1, z_2\right) . \end{aligned}$$

The derivative of f in (c_1, c_2, c_3, c_4) is given by

$$\begin{aligned} & f'(c_1, c_2, c_3, c_4) \\ &= \begin{pmatrix} \frac{3\sqrt{n}(dn+n+d-1)}{(d+1)^{5/2}} & -\frac{3\sqrt{n}(2dn+d^2+2n+3d)}{2(d+1)^{5/2}} & \left(\frac{n}{d+1}\right)^{3/2} & 0 \\ \frac{4(2d^2n-dn^2-n^2+2d^2+5dn-3d)}{(d+1)^3} & \frac{2(3dn^2-2d^2n-d^3+3n^2-5dn-6d^2)}{(d+1)^3} & -\frac{4n^2}{(d+1)^2} & \frac{n^2}{(d+1)^2} \end{pmatrix} \\ &\to \begin{pmatrix} \frac{3}{y^{3/2}} & -\frac{6+3y}{2y^{3/2}} & \frac{1}{y^{3/2}} & 0 \\ \frac{8y-4}{y^2} & \frac{6-4y-2y^2}{y^2} & -\frac{4}{y^2} & \frac{1}{y^2} \end{pmatrix} \end{aligned}$$

as $d/n \to y$. The given covariance matrix is obtained from the calculation of the limit of $[f'(c_1, c_2, c_3, c_4)]\mathbb{C}ov(Z)[f'(c_1, c_2, c_3, c_4)]^t$. □

Remark. We immediately see from Corollary 4 that an (n, d)-consistent estimator for B_2 is given by $\overline{\lambda}_2 - y\overline{\lambda}_1^2$, where $y = d/n$, leading to

$$\tilde{s}_\lambda^2 := \overline{\lambda}_2 - (y+1)\overline{\lambda}_1^2$$

as an (n, d)-consistent estimator for the variance $B_2 - B_1^2$. Unfortunately, $\tilde{s}_\lambda^2$ can become negative so that a root of it is possibly complex. So, an (n, d)-consistent estimator for the skewness of the true limiting eigenvalue distribution would have a complex distribution which is not what we are aiming for.

Next, an (n, d)-consistent estimator for the third centralized eigenvalue moment $B_3 - 3B_1B_2 + 2B_1^3$ is given by

$$\hat{B}_3^c := \overline{\lambda}_3 - 3(y+1)\overline{\lambda}_1\overline{\lambda}_2 + (2y^2 + 3y + 2)\overline{\lambda}_1^3,$$

which leads to

$$\frac{\hat{B}_3^c}{\tilde{s}_\lambda^{3/2}}$$

as an (n, d)-consistent estimator for the skewness of the true limiting eigenvalue distribution. In order to obtain its null distribution applying the delta method, one has to consider the function

$$\tilde{f}(x_1, x_2, x_3) := \frac{1}{(x_2 - (y+1)x_1^2)^{3/2}}(x_3 - 3(y+1)x_1x_2 + (2y^2 + 3y + 2)x_1^3) .$$

But the derivative $\tilde{f}'(c_1, c_2, c_3)$ does not exist under the (n, d)-asymptotics. Similar considerations show that the asymptotic null distribution of an (n, d)-consistent estimator for the kurtosis of the true limiting eigenvalue distribution also does not exist.

All in all, the method of constructing a test based on a test statistic which is (n, d)-consistent for the skewness and kurtosis of the true limiting eigenvalue distribution fails. This is why this article proposes to use estimators for the skewness and kurtosis of the limiting ESD.

Proof of Theorem 7

The derivation of the asymptotic distribution of SL requires two small lemmas.

Lemma 9. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $A, B, C \in \mathcal{A}$ with $B \cap C = \emptyset$ and $P(B) > 0$ or $P(C) > 0$. Then:*

$$P(A|B \cup C) = \frac{P(B)}{P(B) + P(C)} P(A|B) + \frac{P(C)}{P(B) + P(C)} P(A|C)$$

Proof. This proof is elementary and therefore omitted. □

Lemma 10. *Let (X, Y) be a two dimensional continuously distributed random vector. Further, let the marginal distribution of Y be symmetric around 0 and X be $\mathcal{A} - \mathcal{B}$ measurable. Then:*

$$P(X \in B|\{Y = s\} \cup \{Y = -s\}) = \frac{1}{2} P(X \in B|Y = s) + \frac{1}{2} P(X \in B|Y = -s),$$

where $B \in \mathcal{B}, s \in \mathbb{R}$.

Proof. The assertion is obviously true for $s = 0$. Therefore, let $s \neq 0$. Set $B_\varepsilon := \{s - \varepsilon < Y \leq s + \varepsilon\}, C_\varepsilon := \{-s - \varepsilon < Y \leq -s + \varepsilon\}$ for $\varepsilon > 0$. Note that $P(B_\varepsilon) = P(C_\varepsilon)$ because of the symmetry of Y. Further, if ε is sufficiently small, then $B_\varepsilon \cap C_\varepsilon = \emptyset$. Thus, we have:

$$\begin{aligned}
&P(X \in B|\{Y = s\} \cup \{Y = -s\}) = \lim_{\varepsilon \searrow 0} P(X \in B|B_\varepsilon \cup C_\varepsilon) \\
&\overset{\text{Lem. 9}}{=} \lim_{\varepsilon \searrow 0} \left[\frac{P(B_\varepsilon)}{P(B_\varepsilon) + P(C_\varepsilon)} P(X \in B|B_\varepsilon) + \frac{P(C_\varepsilon)}{P(B_\varepsilon) + P(C_\varepsilon)} P(X \in B|C_\varepsilon) \right] \\
&= \frac{1}{2} P(X \in B|Y = s) + \frac{1}{2} P(X \in B|Y = -s)
\end{aligned}$$

□

Now, we can prove Theorem 7.

Proof. We obtain from the law of total probability:

$$\begin{aligned}
F_{SL}(x) = P(SL \leq x) &= P\left((\gamma_1^*(\nu))^2 + (\gamma_2^*(\nu))^2 \leq x\right) \\
&= \int_0^\infty \underbrace{P\left((\gamma_1^*(\nu))^2 + (\gamma_2^*(\nu))^2 \leq x \mid (\gamma_2^*(\nu))^2 = z\right)}_{=0 \text{ if } z > x} f_1(z) \, \mathrm{d}z \\
&= \int_0^x P\left((\gamma_1^*(\nu))^2 + (\gamma_2^*(\nu))^2 \leq x \mid (\gamma_2^*(\nu))^2 = z\right) f_1(z) \, \mathrm{d}z
\end{aligned}$$

Now, we notice that $(\gamma_1^*(\nu), \gamma_2^*(\nu))$ has an asymptotic joint normal distribution because of Theorem 6. The parameters of this normal distribution are asymptotically given by

$$\mathbb{E}\left(\gamma_1^*(\nu)\right) \to 0, \mathbb{E}\left(\gamma_2^*(\nu)\right) \to 0, \mathbb{V}ar\left(\gamma_1^*(\nu)\right) \to 1, \mathbb{V}ar\left(\gamma_2^*(\nu)\right) \to 1,$$
$$\mathbb{C}ov\left(\gamma_1^*(\nu), \gamma_2^*(\nu)\right) \to a = \frac{24\sqrt{y}(1+y)}{\sqrt{6+9y}\sqrt{8+96y+64y^2}} .$$

Thus, $\gamma_1^*(\nu)$ given $\gamma_2^*(\nu) = \xi$ is also asymptotically normally distributed with mean

$$\mathbb{E}\left(\gamma_1^*(\nu)\right) + \frac{\mathbb{C}ov\left(\gamma_1^*(\nu), \gamma_2^*(\nu)\right)}{\mathbb{V}ar\left(\gamma_2^*(\nu)\right)}\left(\xi - \mathbb{E}\left(\gamma_2^*(\nu)\right)\right) \to a\xi$$

and variance

$$\mathbb{V}ar\left(\gamma_1^*(\nu)\right)\left(1 - \frac{\mathbb{C}ov^2\left(\gamma_1^*(\nu), \gamma_2^*(\nu)\right)}{\mathbb{V}ar\left(\gamma_1^*(\nu)\right)\mathbb{V}ar\left(\gamma_2^*(\nu)\right)}\right) \to 1 - a^2$$

From these considerations and Lemma 10, we have for $x \leq z$:

$$\begin{aligned}
&P\left((\gamma_1^*(\nu))^2 + (\gamma_2^*(\nu))^2 \leq x \mid (\gamma_2^*(\nu))^2 = z\right) \\
&\quad = P\left((\gamma_1^*(\nu))^2 \leq x - z \mid (\gamma_2^*(\nu))^2 = z\right) \\
&\quad = P\left(-\sqrt{x-z} \leq \gamma_1^*(\nu) \leq \sqrt{x-z} \mid \{\gamma_2^*(\nu) = \sqrt{z}\} \cup \{\gamma_2^*(\nu) = -\sqrt{z}\}\right) \\
&\quad = \frac{1}{2}P\left(-\sqrt{x-z} \leq \gamma_1^*(\nu) \leq \sqrt{x-z} \mid \gamma_2^*(\nu) = \sqrt{z}\right) \\
&\qquad\qquad + \frac{1}{2}P\left(-\sqrt{x-z} \leq \gamma_1^*(\nu) \leq \sqrt{x-z} \mid \gamma_2^*(\nu) = -\sqrt{z}\right) \\
&\quad \to \frac{1}{2}\left[\Phi\left(\frac{\sqrt{x-z} - a\sqrt{z}}{\sqrt{1-a^2}}\right) - \Phi\left(\frac{-\sqrt{x-z} - a\sqrt{z}}{\sqrt{1-a^2}}\right)\right] \\
&\qquad\qquad + \frac{1}{2}\left[\Phi\left(\frac{\sqrt{x-z} + a\sqrt{z}}{\sqrt{1-a^2}}\right) - \Phi\left(\frac{-\sqrt{x-z} + a\sqrt{z}}{\sqrt{1-a^2}}\right)\right] \\
&\quad = \frac{1}{2}\left[\Phi\left(\frac{\sqrt{x-z} - a\sqrt{z}}{\sqrt{1-a^2}}\right) - \left(1 - \Phi\left(\frac{\sqrt{x-z} + a\sqrt{z}}{\sqrt{1-a^2}}\right)\right)\right] \\
&\qquad\qquad + \frac{1}{2}\left[\Phi\left(\frac{\sqrt{x-z} + a\sqrt{z}}{\sqrt{1-a^2}}\right) - \left(1 - \Phi\left(\frac{\sqrt{x-z} - a\sqrt{z}}{\sqrt{1-a^2}}\right)\right)\right] \\
&\quad = \Phi\left(\frac{\sqrt{x-z} - a\sqrt{z}}{\sqrt{1-a^2}}\right) + \Phi\left(\frac{\sqrt{x-z} + a\sqrt{z}}{\sqrt{1-a^2}}\right) - 1
\end{aligned}$$

So, the proof of this theorem is completed. □

References

L. V. Arharov. Limit theorems for the characteristic roots of a sample covariance matrix. *Sov. Math. Dokl.*, 12:1206–1209, 1971.

Z. D. Bai and Y. Q. Yin. Convergence to the semicircle law. *Ann. Probab.*, 16(2): 863–875, 1988.

Zhidong Bai and Jack W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

Zhidong Bai and Jack W. Silverstein. CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann. Probab.*, 32(1):553–605, 2004.

Zhidong Bai and Jack W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer, New York, 2nd edition, 2010.

Zhidong Bai and Shurong Zheng. Large dimensional data analysis with applications of random matrix theory. In *4th International Congress of Chinese Mathematicians*, 2007.

K. O. Bowman and L. R. Shenton. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika*, 62(2):243–250, 1975.

Song Xi Chen, Li-Xin Zhang, and Ping-Shou Zhong. Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.*, 105(490):810–819, 2010.

Thomas J. Fisher, Xiaoqian Sun, and Colin M. Gallagher. A new test for sphericity of the covariance matrix for high dimensional data. *J. Multivariate Anal.*, 101(10): 2554–2570, 2010.

Carlos M. Jarque and Anil K. Bera. A test for normality of observations and regression residuals. *Internat. Statist. Rev.*, 55(2):163–172, 1987.

S. John. Some optimal multivariate tests. *Biometrika*, 58(1):123–127, 1971.

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

Olivier Ledoit and Michael Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, 30(4):1081–1102, 2002.

A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 2009.

V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.*, 72(4):457–483, 1967.

Robb John Muirhead. *Aspects of Multivariate Statistical Theory.* Wiley & Sons, New York, 1st edition, 1982.

Alexandru Nica and Roland Speicher. *Lectures on the Combinatorics of Free Probability.* Cambridge University Press, Cambridge, 1st edition, 2006.

Muni S. Srivastava. Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, 35(2):251–272, 2005.

Muni S. Srivastava. Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.*, 37(1):53–86, 2007.

Muni S. Srivastava, Tonu Kollo, and Dietrich von Rosen. Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *J. Multivariate Anal.*, 102(6):1090–1103, 2011.

Antonia M. Tulino and Sergio Verdú. *Random Matrices and Wireless Communications.* Foundations and Trends in Communications and Information Theory 1, 2004. URL `http://dx.doi.org/10.1561/0100000001`.

Cem Ünsalan. Measuring land development in urban regions using graph theoretical and conditional statistical features. *IEEE Trans. Geoscience and Remote Sensing*, 45 (12):3989–3999, 2007.

Dan Voiculescu. Symmetries of some reduced free product C*-algebras. In *Operator algebras and their connections with topology and ergodic theory*, Lecture Notes in Mathematics, vol. 1132, pages 556–588, Berlin, 1985. Springer.

Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20(1):50–68, 1986.