The Annals of Statistics
2013, Vol. 41, No. 3, 1204–1231
DOI: 10.1214/13-AOS1100

ASYMPTOTIC POWER OF SPHERICITY TESTS FOR HIGH-DIMENSIONAL DATA

BY ALEXEI ONATSKI, MARCELO J. MOREIRA[1] AND MARC HALLIN[2]

University of Cambridge, FGV/EPGE, and Université libre de Bruxelles and Princeton University

This paper studies the asymptotic power of tests of sphericity against perturbations in a single unknown direction as both the dimensionality of the data and the number of observations go to infinity. We establish the convergence, under the null hypothesis and contiguous alternatives, of the log ratio of the joint densities of the sample covariance eigenvalues to a Gaussian process indexed by the norm of the perturbation. When the perturbation norm is larger than the *phase transition threshold* studied in Baik, Ben Arous and Péché [*Ann. Probab.* **33** (2005) 1643–1697] the limiting process is degenerate, and discrimination between the null and the alternative is asymptotically certain. When the norm is below the threshold, the limiting process is nondegenerate, and the joint eigenvalue densities under the null and alternative hypotheses are mutually contiguous. Using the asymptotic theory of statistical experiments, we obtain asymptotic power envelopes and derive the asymptotic power for various sphericity tests in the contiguity region. In particular, we show that the asymptotic power of the Tracy–Widom-type tests is trivial (i.e., equals the asymptotic size), whereas that of the eigenvalue-based likelihood ratio test is strictly larger than the size, and close to the power envelope.

1. Introduction. Recently, there has been much interest in testing sphericity in a high-dimensional setting. Various tests have been proposed and analyzed in Ledoit and Wolf (2002), Srivastava (2005), Birke and Dette (2005), Schott (2006), Bai et al. (2009), Fisher, Sun and Gallagher (2010), Chen, Zhang and Zhong (2010) and Berthet and Rigollet (2012). In many studies, a distinct interesting alternative to the null of sphericity is the existence of a low-dimensional structure or signal in the data. Detecting such a structure has been the focus of recent studies in various applied fields including population and medical genetics [Patterson, Price and Reich (2006)], econometrics [Onatski (2009, 2010)], wireless communication

Received August 2011; revised June 2012.

[1]Supported by CNPq and NSF Grant SES-0819761.

[2]Supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft, the IAP research network grant P7/06 of the Belgian Federal Government (Belgian Science Policy), and a Discovery grant of the Australian Research Council.

MSC2010 subject classifications. Primary 62H15, 62B15; secondary 41A60.

Key words and phrases. Sphericity tests, large dimensionality, asymptotic power, spiked covariance, contiguity, power envelope, steepest descent, contour integral representation.

[Bianchi et al. (2011)], chemometrics [Kritchman and Nadler (2008)] and signal processing [Perry and Wolfe (2010)].

Most of the existing sphericity tests are based on the eigenvalues of the sample covariance matrix, which constitute the maximal invariant statistic with respect to orthogonal transformations of the data. The asymptotic power of such tests depends on the asymptotic behavior of the sample covariance eigenvalues under the alternative hypothesis. When the alternative is a rank-k perturbation of the null, the corresponding population covariance matrix is proportional to a sum of the identity matrix and a matrix of rank k. Johnstone (2001) calls such a situation "spiked covariance."

The asymptotic behavior of the sample covariance eigenvalues in "spiked covariance" models of increasing dimension is well studied. Consider the simplest case, when $k = 1$. If the largest population covariance eigenvalue is above the "phase transition" threshold studied in Baik, Ben Arous and Péché (2005), then the largest sample covariance eigenvalue remains separated from the rest of the eigenvalues, which are asymptotically "packed together as in the support of the Marchenko–Pastur density" [Baik and Silverstein (2006)]. Since the largest eigenvalue separates from the "bulk," it is easy to detect a signal.

If the largest population covariance eigenvalue is at or below the threshold, the empirical distribution of the sample covariance eigenvalues still converges to the Marchenko–Pastur distribution, but the largest sample covariance eigenvalue now converges to the upper boundary of its support, both under the null of sphericity and the "spiked" alternative [Silverstein and Bai (1995) and Baik and Silverstein (2006)]. Hence, the signal detection becomes problematic. At the threshold, the null and the alternative hypotheses lead to different asymptotic distributions for the centered and normalized largest sample covariance eigenvalue [Bloemendal and Virág (2012) and Mo (2012)], which implies some asymptotic detection power. However, below the threshold, the difference disappears with the joint distribution of any finite number of the centered and normalized largest sample covariance eigenvalues converging to the multivariate Tracy–Widom law under both the null and the alternative [Johnstone (2001), Baik, Ben Arous and Péché (2005), El Karoui (2007) and Féral and Péché (2009)].

This similarity in the asymptotic behavior of covariance eigenvalues under the null and the alternative prompts Nadakuditi and Edelman (2008) and Nadakuditi and Silverstein (2010) to call the transition threshold "the fundamental asymptotic limit of sample-eigenvalue-based detection." They claim that no reliable signal detection is possible below that limit in the asymptotic sense. This asymptotic impossibility is also pointed out and discussed in several other recent studies, including Patterson, Price and Reich (2006), Hoyle (2008), Nadler (2008), Kritchman and Nadler (2009) and Perry and Wolfe (2010).

In this paper, we analyze the capacity of statistical tests to detect a one-dimensional signal with the corresponding population covariance eigenvalue below the "impossibility threshold," showing that the terminology "impossibility threshold" is overly pessimistic. We establish that the eigenvalue region below the threshold actually is the region of mutual contiguity [in the sense of Le Cam (1960)] of the joint distributions of the sample covariance eigenvalues under the null and under the alternative. We obtain the limit in distribution of the log likelihood ratio process inside this contiguity region and derive the asymptotic power envelope for sample-eigenvalue-based detection tests.

The power envelope is larger than size for local alternatives and monotonically tends to one as the signal's population eigenvalue approaches the threshold from below. Hence, the detection of a signal with high asymptotic probability is quite possible even in cases where the largest population covariance eigenvalue is smaller than the threshold, especially when the distance from the threshold remains small.

In the contiguity region, the log likelihood ratio is asymptotically equivalent to a simple statistic related to the Stieltjes transform of the empirical distribution of the sample covariance eigenvalues. The reason the asymptotic behavior of this statistic differs under the null and under the alternative despite the apparent similarity of eigenvalue behaviors just mentioned is that it is not based merely on a contrast between the largest and the rest of the eigenvalues. The information about the presence of the signal exploited by this statistic is hidden in the small deviations of the empirical distribution of the eigenvalues from its Marchenko–Pastur limit.

Let us examine our setting and our results in more detail. Suppose that data consist of n independent observations of p-dimensional real-valued vectors X_t distributed according to the Gaussian law with mean zero and covariance matrix $\sigma^2(I_p + hvv')$, where I_p is the p-dimensional identity matrix, σ and h are scalars and v is a p-dimensional vector with Euclidean norm one. We are interested in the asymptotic power of the tests of the null hypothesis $H_0 : h = 0$ against the alternative $H_1 : h > 0$ based on the eigenvalues of the sample covariance matrix of the data when both n and p go to infinity. The vector v is an unspecified nuisance parameter indicating the direction of the perturbation of sphericity. In contrast to Berthet and Rigollet (2012), who study signal detection in a similar setting where the vector v is sparse, we do not constrain v in any way except normalizing its Euclidean norm to one.

We consider the cases of known and unknown σ^2. For the sake of brevity, in the rest of this Introduction, we discuss only the case of unknown σ^2, which, in practice, is also more relevant. Let λ_j be the jth largest sample covariance eigenvalue, let $\mu_j = \lambda_j/(\lambda_1 + \cdots + \lambda_p)$ be its normalized version and let $\mu = (\mu_1, \ldots, \mu_{m-1})$, where $m = \min(n, p)$. We begin our analysis with a study of the asymptotic properties of the likelihood ratio process $L(h; \mu)$ defined as the ratio of the density of μ when $h \neq 0$ to that when $h = 0$. We represent $L(h; \mu)$ in the form of an integral over a contour in the complex plane and use the Laplace approximation method and recent results from the large random matrix theory to derive an asymptotic expansion of $L(h; \mu)$ as $p, n \to \infty$ so that $p/n \to c \in (0, \infty)$, which we throughout abbreviate into $p, n \to_c \infty$.

We show that, for any $\bar{h}$ such that $0 < \bar{h} < \sqrt{c}$, $\ln L(h;\mu)$ converges in distribution under the null to a Gaussian process $\mathcal{L}(h;\mu)$ on $h \in [0, \bar{h}]$ with

$$\mathrm{E}[\mathcal{L}(h;\mu)] = \tfrac{1}{4}[\ln(1 - c^{-1}h^2) + c^{-1}h^2]$$

and

$$\mathrm{Cov}(\mathcal{L}(h_1;\mu), \mathcal{L}(h_2;\mu)) = -\tfrac{1}{2}[\ln(1 - c^{-1}h_1h_2) + c^{-1}h_1h_2].$$

By Le Cam's first lemma [see van der Vaart (1998), page 88], this implies that the joint distributions of the normalized sample covariance eigenvalues under the null and under the alternative are mutually contiguous for any $h \in [0, \bar{h}]$. We also show that these joint distributions are not mutually contiguous for any $h > \sqrt{c}$.

Since $\mathcal{L}(h;\mu)$, as a likelihood ratio process, is not of the LAN Gaussian shift type, local asymptotic normality does not hold, and the asymptotic optimality analysis of tests of $H_0 : h = 0$ against $H_1 : h > 0$ is difficult. However, an asymptotic power envelope is easy to construct using the Neyman–Pearson lemma along with Le Cam's third lemma. We show that, for tests of asymptotic size α, the maximum achievable power against a specific alternative $h = h_1$ is $1 - \Phi[\Phi^{-1}(1-\alpha) - \sqrt{-\frac{1}{2}(\ln(1 - c^{-1}h_1^2) + c^{-1}h_1^2)}]$, where Φ, as usual, denotes the standard normal distribution function.

Using our result on the limiting distribution of $\ln L(h;\mu)$ and Le Cam's third lemma, we compute the asymptotic powers of several previously proposed tests of sphericity and of the likelihood ratio (LR) test based on μ. We find that the power of the LR test comes close to the asymptotic power envelope. The LR test outperforms the test proposed by John (1971) and studied in Ledoit and Wolf (2002), as well as Srivastava (2005) and the test proposed by Bai et al. (2009). The asymptotic powers of the tests based on the largest sample covariance eigenvalue, such as the tests proposed by Bejan (2005), Patterson, Price and Reich (2006), Kritchman and Nadler (2009), Onatski (2009), Bianchi et al. (2011) and Nadakuditi and Silverstein (2010), equals the tests' asymptotic size for alternatives in the contiguity region.

The rest of the paper is organized as follows. Section 2 provides a representation of the likelihood ratio in terms of a contour integral. Section 3 applies Laplace's method to obtain an asymptotic approximation to the contour integral. Section 4 uses that approximation to establish the convergence of the log likelihood ratio process to a Gaussian process. Section 5 provides an analysis of the asymptotic power of various sphericity tests and derives the asymptotic power envelope. Section 6 concludes. Proofs are given in the Appendix; the more technical ones are relegated to the Supplementary Appendix [Onatski, Moreira and Hallin (2013)].

2. Likelihood ratios as contour integrals. Let X be a $p \times n$ matrix with i.i.d. real Gaussian $N(0, \sigma^2(I_p + hvv'))$ columns. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ be the ordered eigenvalues of $\frac{1}{n}XX'$ and let $\lambda = (\lambda_1, \ldots, \lambda_m)$, where $m = \min\{n, p\}$. Finally, let $\mu = (\mu_1, \ldots, \mu_{m-1})$, where $\mu_j = \lambda_j/(\lambda_1 + \cdots + \lambda_p)$.

As explained in the Introduction, our goal is to study the asymptotic power of the eigenvalue-based tests of $H_0 : h = 0$ against $H_1 : h > 0$. If σ^2 is known, the model is invariant with respect to orthogonal transformations, and the maximal invariant statistic is λ. Therefore, we consider tests based on λ. If σ^2 is unknown (which, strictly speaking, is what is meant by "sphericity"), the model is invariant with respect to orthogonal transformations and multiplications by nonzero scalars, and the maximal invariant is μ. Hence, we consider tests based on μ. Note that the distribution of μ does not depend on σ^2, whereas if σ^2 is known, we can always normalize λ dividing it by σ^2. Therefore, in what follows, we will assume that $\sigma^2 = 1$ without loss of generality.

Let us denote the joint density of $\lambda_1, \dots, \lambda_m$ as $p(\lambda; h)$ and that of $\mu_1, \dots, \mu_{m-1}$ as $p(\mu; h)$. The following proposition gives explicit formulas for $p(\lambda; h)$ and $p(\mu; h)$.

PROPOSITION 1. *Let $\mathcal{S}(r)$ be the $(r-1)$-dimensional unit sphere, and let $(\mathrm{d}x_r)$ be the invariant measure on $\mathcal{S}(r)$ normalized so that the total measure is one. Further, let $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$ and $M = \mathrm{diag}(\mu_1, \dots, \mu_p)$. Then*

$$p(\lambda; h) = \frac{\gamma(n, p, \lambda)}{(1+h)^{n/2}} \int_{\mathcal{S}(p)} e^{(n/2)(h/(1+h))x_p' \Lambda x_p} (\mathrm{d}x_p) \tag{2.1}$$

and

$$\begin{aligned} p(\mu; h) = {} & \frac{\delta(n, p, \mu)}{(1+h)^{n/2}} \\ & \times \int_0^\infty y^{(np-2)/2} e^{-ny/2} \int_{\mathcal{S}(p)} e^{(n/2)(yh/(1+h))x_p' M x_p} (\mathrm{d}x_p)\, \mathrm{d}y, \end{aligned} \tag{2.2}$$

where $\gamma(n, p, \lambda)$ and $\delta(n, p, \mu)$ depend only on n and p, and on λ and μ, respectively.

The spherical integrals in (2.1) and (2.2) can be represented in the form of a confluent hypergeometric function ${}_1F_1$ of matrix argument [Hillier (2001), page 4]. For example, for the integral in (2.1),

$$\int_{\mathcal{S}(p)} e^{(n/2)(h/(1+h))x_p' \Lambda x_p} (\mathrm{d}x_p) = {}_1F_1\left(\frac{1}{2}, \frac{p}{2}; \frac{n}{2}\frac{h}{1+h}\Lambda\right).$$

Butler and Wood (2002) develop Laplace approximations to functions ${}_1F_1$ but do not analyze the asymptotic behavior of the approximation errors. The next lemma derives an alternative representation of the spherical integrals in Proposition 1. This representation has the form of a contour integral of a single complex variable, and our asymptotic analysis will be based on the Laplace approximation to such an integral.

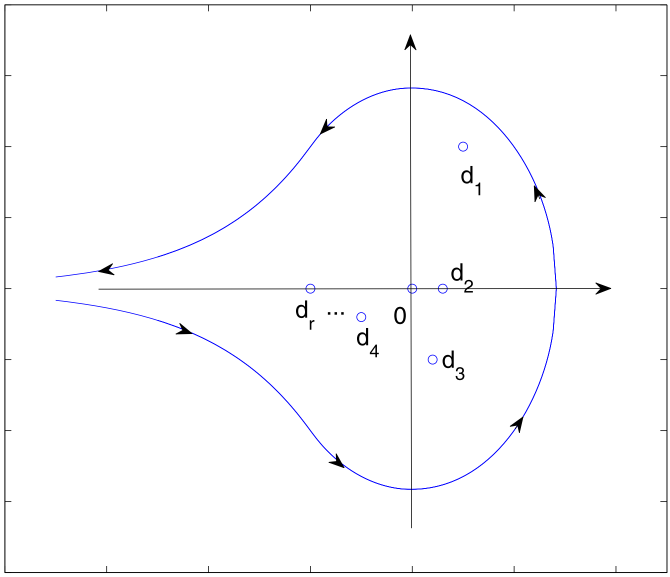

FIG. 1. *Contour of integration $\mathcal{K}$ in* (2.3).

LEMMA 2. *Let* $D = \operatorname{diag}(d_1, \dots, d_r)$, *where* d_j *are arbitrary complex numbers. Further, let* $\mathcal{K}$ *be a contour in the complex plane starting at* $-\infty$, *encircling counter-clockwise the points* $0, d_1, \dots, d_r$, *and going back to* $-\infty$. *Such a contour is shown in Figure* 1. *We have*

$$\int_{\mathcal{S}(r)} e^{x_r' D x_r}(\mathrm{d}x_r) = \frac{\Gamma(r/2)}{2\pi i} \oint_{\mathcal{K}} e^s \prod_{j=1}^{r} (s - d_j)^{-1/2}\, \mathrm{d}s. \tag{2.3}$$

PROOF. The integral on the left-hand side of (2.3) is the expected value of $\exp(\frac{y_1^2 d_1 + \cdots + y_r^2 d_r}{y_1^2 + \cdots + y_r^2})$, where $y_1, \dots, y_r$ are independent standard normal random variables. The variables $u_j = \frac{y_j^2}{y_1^2 + \cdots + y_r^2}$, $j = 1, \dots, r$, have Dirichlet distribution $\mathcal{D}(k_1, \dots, k_r)$ with parameters $k_1 = \cdots = k_r = \frac{1}{2}$. Denoting the expectation operator with respect to such a distribution as $\mathrm{E}_{\mathcal{D}}$, we have

$$\int_{\mathcal{S}(r)} e^{x_r' D x_r}(\mathrm{d}x_r) = \mathrm{E}_{\mathcal{D}} \exp(u_1 d_1 + \cdots + u_r d_r). \tag{2.4}$$

Now, expanding the exponent in the latter expression into power series and taking expectations term by term yields

$$\mathrm{E}_{\mathcal{D}} \exp(u_1 d_1 + \cdots + u_r d_r) = \sum_{k=0}^{\infty} \frac{\mathrm{E}_{\mathcal{D}}(u_1 d_1 + \cdots + u_r d_r)^k}{k!}. \tag{2.5}$$

The Dirichlet average of $(u_1 d_1 + \cdots + u_r d_r)^k$ is well studied. By Theorem 3.1 of Dickey (1983),

$$
\begin{aligned}
&\mathrm{E}_{\mathcal{D}}\big[(u_1 d_1 + \cdots + u_r d_r)^k\big] \\
&\quad = \sum_{\substack{m_1,\dots,m_r \geq 0 \\ m_1+\cdots+m_r=k}} \frac{k!}{m_1! \cdots m_r!} \frac{(1/2)_{m_1} \cdots (1/2)_{m_r}}{(r/2)_k} d_1^{m_1} \cdots d_r^{m_r},
\end{aligned}
\tag{2.6}
$$

where $(k)_s = k(k+1)\cdots(k+s-1)$ is Pochhammer's notation for the shifted factorial.

Combining (2.6) with (2.5) and (2.4), we get

$$
\begin{aligned}
\int_{\mathcal{S}(r)} e^{x_r' D x_r} (\mathrm{d}x_r) &= \sum_{m_1,\dots,m_r \geq 0} \frac{(1/2)_{m_1} \cdots (1/2)_{m_r}}{(r/2)_{m_1+\cdots+m_r}} \frac{d_1^{m_1} \cdots d_r^{m_r}}{m_1! \cdots m_r!} \\
&= {}_r\Phi(1/2, \dots, 1/2; r/2; d_1, \dots, d_r),
\end{aligned}
\tag{2.7}
$$

where the last equality is the definition of the confluent form of the Lauricella F_D function, denoted as ${}_r\Phi(\cdot)$. The functions ${}_r\Phi(\cdot)$ were introduced by Erdelyi (1937) and are discussed by Srivastava and Karlsson (1985). In probability and statistics, they were recently used to study the mean of a Dirichlet process [see Lijoi and Regazzini (2004) and references therein].

Erdelyi (1937), formula (8,6), establishes the following contour integral representation of ${}_r\Phi(\cdot)$:

$$
\begin{aligned}
&{}_r\Phi(k_1, \dots, k_r; t; d_1, \dots, d_r) \\
&\quad = \frac{\Gamma(t)}{2\pi i} \oint_{\mathcal{K}} e^s s^{-t+k_1+\cdots+k_r} \prod_{j=1}^{r} (s - d_j)^{-k_j} \, \mathrm{d}s.
\end{aligned}
\tag{2.8}
$$

Lemma 2 follows from equalities (2.7) and (2.8). $\square$

The contour integral representation given in Lemma 2 has been derived independently by Mo (2012) and Wang (2012), who use it to study the largest sample covariance eigenvalue when the corresponding population eigenvalue equals the critical threshold or lies above it. Our proof effectively takes advantage of old results of Dickey (1983) and Erdelyi (1937), and thus is different from the proofs in the above mentioned papers.

Using Lemma 2 and Proposition 1, we derive contour integral representations for the likelihood ratios $L(h; \lambda) = p(\lambda; h)/p(\lambda; 0)$ and $L(h; \mu) = p(\mu; h)/p(\mu; 0)$. The quantity $L(h; \lambda)$ is the likelihood ratio based on λ as opposed to the entire data X. Similarly, $L(h; \mu)$ is the likelihood ratio based on μ.

LEMMA 3. *Let $\mathcal{K}$ be a contour in the complex plane that starts at $-\infty$, then encircles counter-clockwise the sample covariance eigenvalues $\lambda_1, \dots, \lambda_p$, and* *goes back to* $-\infty$. *In addition, we require that for any* $z \in \mathcal{K}$, $\operatorname{Re} z < \frac{1+h}{h} S$, *where* $\operatorname{Re} z$ *denotes the real part of* $z \in \mathbb{C}$ *and* $S = \lambda_1 + \cdots + \lambda_p$. *Then*,

$$L(h; \lambda) = k_1 \left(\frac{2}{n}\right)^{(p-2)/2} \frac{1}{2\pi i} \oint_{\mathcal{K}} e^{(n/2)(h/(1+h))z} \prod_{j=1}^{p} (z - \lambda_j)^{-1/2} \, \mathrm{d}z \tag{2.9}$$

and

$$\begin{aligned} L(h; \mu) = k_2 & \frac{S^{(p-2)/2}}{2\pi i} \\ & \times \oint_{\mathcal{K}} e^{-((np-p+2)/2) \ln(1-(h/(1+h))(z/S))} \prod_{j=1}^{p} (z - \lambda_j)^{-1/2} \, \mathrm{d}z, \end{aligned} \tag{2.10}$$

where $k_1 = h^{-(p-2)/2}(1+h)^{(p-n-2)/2}\Gamma(p/2)$ *and* $k_2 = k_1 \frac{\Gamma((np-p+2)/2)}{\Gamma(np/2)}$.

Close inspection of the proof of Lemma 3 reveals that the right-hand side of (2.10) depends on λ only through μ. Although it is possible to express $L(h; \mu)$ as an explicit function of μ, the implicit form given in (2.10) is convenient because it allows us to use similar methods for the asymptotic analysis of the two likelihood ratios.

In the next two sections, we perform an asymptotic analysis of $L(h; \lambda)$ and $L(h; \mu)$ that relies on the Laplace approximation of the contour integrals in Lemma 3 after those contours have been suitably deformed without changing the value of the integrals.

3. Laplace approximation. In this section, we derive the Laplace approximations to the contour integrals in Lemma 3. Laplace's method for contour integrals is discussed, for example, in Chapter 4 of Olver (1997). The method describes an asymptotic approximation to a contour integral $\oint_{\mathcal{K}} e^{-nf(z)} g(z) \, \mathrm{d}z$ as $n \to \infty$, where $f(z)$ and $g(z)$ are analytic functions of z. The approximation is usually based on the part of the contour integral coming from a neighborhood of some point $z_0 \in \mathcal{K}$, where z_0 is such that $\frac{\mathrm{d}}{\mathrm{d}z} f(z_0) = 0$ and $\operatorname{Re} f(z_0) = \min_{z \in \mathcal{K}} \operatorname{Re} f(z)$. For such a point to exist, one might need to deform the contour so that, by Cauchy's theorem, the value of the integral does not change. Typically, the deformation is chosen so that $\operatorname{Re}(-f(z))$ declines in the fastest way possible as z goes away from z_0 along the contour. For this reason, the method is called the *method of steepest descent*.

The contour integrals in (2.9) and (2.10) can be represented in the Laplace form with a deterministic function $f(z)$ and a random function $g(z)$ that converges to a log-normal random process on the contour as $p, n \to_c \infty$. To see this, note that the logarithm of the multiple product in (2.9) and (2.10) equals $-\frac{1}{2} \sum_{j=1}^{p} \ln(z - \lambda_j)$. For each z, this expression is a special form of the linear spectral statistic

$\sum_{j=1}^{p} \varphi(\lambda_j)$ studied by Bai and Silverstein (2004). According to the central limit theorem (Theorem 1.1) established in that paper, the random variable

$$\Delta_p(z) = \sum_{j=1}^{p} \ln(z - \lambda_j) - p \int \ln(z - \lambda)\, \mathrm{d}\mathcal{F}_p(\lambda) \tag{3.1}$$

converges in distribution to a normal random variable when $p, n \to_c \infty$. Here $\mathcal{F}_p(\lambda)$ is the cumulative distribution function of the Marchenko–Pastur distribution with a mass of $\max(0, 1 - c_p^{-1})$ at zero and density

$$\psi_p(x) = \frac{1}{2\pi c_p x} \sqrt{(b_p - x)(x - a_p)}, \tag{3.2}$$

where $c_p = p/n$, $a_p = (1 - \sqrt{c_p})^2$ and $b_p = (1 + \sqrt{c_p})^2$.

Such a convergence suggests the following choices of $f(z)$ and $g(z)$ in the Laplace forms of the integrals in (2.9) and (2.10):

$$f(z) = -\frac{1}{2}\left(\frac{h}{1+h} z - c_p \int \ln(z - \lambda)\, \mathrm{d}\mathcal{F}_p(\lambda)\right) \tag{3.3}$$

and

$$g(z) = \begin{cases} \exp\left\{-\frac{1}{2}\Delta_p(z)\right\}, & \text{for (2.9)}, \\ \exp\left\{-\frac{np - p + 2}{2} \ln\left(1 - \frac{h}{1+h}\frac{z}{S}\right) - \frac{n}{2}\frac{h}{1+h} z - \frac{1}{2}\Delta_p(z)\right\}, & \\ \qquad \text{for (2.10)}. & \end{cases} \tag{3.4}$$

As mentioned above, a particularly useful deformation of $\mathcal{K}$ passes through the point $z = z_0(h)$ where $\frac{\mathrm{d}}{\mathrm{d}z} f(z) = 0$. Taking the derivative of the right-hand side of (3.3), we see that $z_0(h)$ must satisfy

$$\frac{h}{1+h} + c_p m_p\big(z_0(h)\big) = 0, \tag{3.5}$$

where $m_p(z) = \int \frac{1}{\lambda - z}\, \mathrm{d}\mathcal{F}_p(\lambda)$ is the Stieltjes transform of the Marchenko–Pastur distribution with parameter c_p. The properties of $m_p(z)$ are well studied. In particular, the analytic expression for $m_p(z)$ is known; see, for example, equation (2.3) in Bai (1993). For $z \neq 0$, which lies outside the support of $\mathcal{F}_p(\lambda)$, we have

$$m_p(z) = \frac{-z - c_p + 1 + \sqrt{(z - c_p - 1)^2 - 4c_p}}{2c_p z}, \tag{3.6}$$

where the branch of the square root is chosen so that the real and the imaginary parts of $\sqrt{(z - c_p - 1)^2 - 4c_p}$ have the same signs as the real and the imaginary parts of $z - c_p - 1$, respectively.

Substituting (3.6) into (3.5) and solving for $z_0(h)$ when $h \in (0, \sqrt{c_p})$, we get

$$z_0(h) = \frac{(1+h)(c_p+h)}{h}. \tag{3.7}$$

When $h \geq \sqrt{c_p}$, there are no solutions to (3.5) that lie outside the support of $\mathcal{F}_p(\lambda)$. When $h = \sqrt{c_p}$, the right-hand side of (3.7) equals $(1+\sqrt{c_p})^2$, which lies exactly on the boundary of the support of $\mathcal{F}_p(\lambda)$. When $h > \sqrt{c_p}$, (3.7) provides a solution to (3.5) only when the branch of the square root in (3.6) is chosen differently. As can be verified using (3.3) and (3.6), in such a case, $\frac{\mathrm{d}}{\mathrm{d}z} f(z)$ is strictly negative at $z = z_0(h)$ given by (3.7).

As $c_p \to c$, any fixed h that is smaller than $\sqrt{c}$ eventually satisfies the inequality $h < \sqrt{c_p}$, so that $\frac{\mathrm{d}}{\mathrm{d}z} f(z) = 0$ at $z = z_0(h)$. Therefore, for $h < \sqrt{c}$, we will deform the contour $\mathcal{K}$ into a contour K that passes through $z_0(h)$. We define K as $K = K_+ \cup K_-$, where K_- is the complex conjugate of K_+ and $K_+ = K_1 \cup K_2$ with

$$K_1 = \{z_0(h) + it : 0 \leq t \leq 3z_0(h)\} \tag{3.8}$$

and

$$K_2 = \{x + 3iz_0(h) : -\infty < x \leq z_0(h)\}. \tag{3.9}$$

Figure 2 illustrates the choice of K.

A proof of the following technical lemma is relegated to the Supplementary Appendix [Onatski, Moreira and Hallin (2013)].

LEMMA 4. *Suppose that our null hypothesis is true, and let $\bar{h}$ be any fixed number such that $0 < \bar{h} < \sqrt{c}$. Deforming contour $\mathcal{K}$ into K leaves the value of the integrals (2.9) and (2.10) in Lemma 3 unchanged for all $h \in (0, \bar{h}]$ with probability approaching one as $p, n \to_c \infty$.*

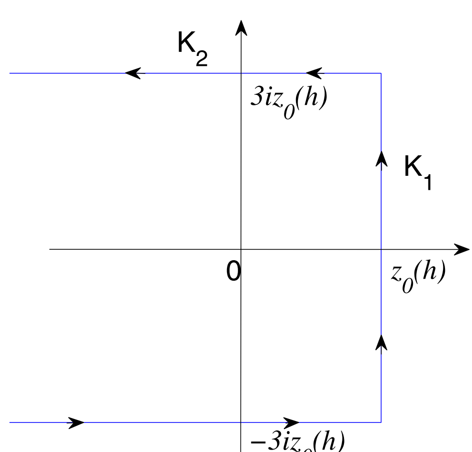

FIG. 2. *Deformation K of contour $\mathcal{K}$.*

We now derive, uniform (over $h \in (0, \bar{h}]$), Laplace approximations to the integrals (2.9) and (2.10) in Lemma 3. First, we introduce additional notation. When $f(z)$ and $g(z)$ are analytic at $z_0 = z_0(h)$, let f_s and g_s with $s = 0, 1, \dots$ be the coefficients in the power series representations

$$f(z) = \sum_{s=0}^{\infty} f_s (z - z_0)^s, \qquad g(z) = \sum_{s=0}^{\infty} g_s (z - z_0)^s. \tag{3.10}$$

When $f(z)$ and $g(z)$ are not analytic at z_0, let the coefficients f_s and g_s be arbitrary numbers for all s.

The following lemma is a generalization of the well-known Watson lemma for contour integrals; see Olver (1997), page 118. Theorem 7.1 in Olver (1997), page 127, derives a similar generalization for the case when $f(z)$ and $g(z)$ are fixed deterministic analytic functions. In contrast to Olver's theorem, our lemma allows $g(z)$ to be a random function, and $f(z)$ to depend on parameter h, and obtains a uniform approximation over $h \in (0, \bar{h}]$. The proof is relegated to the Supplementary Appendix [Onatski, Moreira and Hallin (2013)].

LEMMA 5. *Under the conditions of Lemma 4, for any $h \in (0, \bar{h}]$ and any positive integer m, as $p, n \to_c \infty$, we have*

$$\oint_K e^{-nf(z)} g(z)\, \mathrm{d}z = 2e^{-nf_0} \left[\sum_{s=0}^{m-1} \Gamma\left(s + \frac{1}{2}\right) \frac{a_{2s}}{n^{s+1/2}} + \frac{O_p(1)}{hn^{m+1/2}} \right], \tag{3.11}$$

where $O_p(1)$ is uniform in $h \in (0, \bar{h}]$. The coefficients a_s in (3.11) can be expressed through f_s and g_s defined above. In particular, we have

$$a_0 = \frac{g_0}{2f_2^{1/2}} \quad \text{and} \quad a_2 = \left\{ 4g_2 - \frac{6f_3 g_1}{f_2} + \left(\frac{15f_3^2}{2f_2^2} - \frac{6f_4}{f_2} \right) g_0 \right\} \frac{1}{8f_2^{3/2}}. \tag{3.12}$$

As we explained above, $z_0(h)$ is not a critical point of $f(z)$ when $h > \sqrt{c_p}$. This leads to a situation where the Laplace method for the integral $\oint_K e^{-nf(z)} g(z)\, \mathrm{d}z$ delivers a rather crude approximation. Fortunately, our asymptotic analysis tolerates crude approximations when $h > \sqrt{c_p}$. The following lemma, which is proven in the Supplementary Appendix [Onatski, Moreira and Hallin (2013)], is sufficient for our purposes.

LEMMA 6. *Let $\tilde{h} > \sqrt{c}$, and denote by $K(\tilde{h})$ the corresponding contour, as defined in (3.8) and (3.9). Under the null hypothesis, deforming the contour $\mathcal{K}$ into $K(\tilde{h})$ leaves the value of the integrals in Lemma 3 unchanged for all $h \in [\tilde{h}, \infty)$ with probability approaching one as $p, n \to_c \infty$. Further, for any $h \in [\tilde{h}, \infty)$,*

$$\oint_{K(\tilde{h})} e^{-nf(z)} g(z)\, \mathrm{d}z = e^{-nf(z_0(\tilde{h}))} O_p(1), \tag{3.13}$$

where $O_p(1)$ is uniform over $h \in [\tilde{h}, \infty)$.

Neither Lemma 5 nor Lemma 6 addresses interesting cases with h in a neighborhood of $\sqrt{c}$. In such cases, $z_0(h)$ would be close to the upper boundary of the support of the Marchenko–Pastur distribution. This may lead to the nonanalyticity of $f(z)$ and $g(z)$ on K and a more complicated asymptotic behavior of $g(z)$. We leave the analysis of cases where h may approach $\sqrt{c}$ for future research.

Guionnet and Maïda (2005) study the asymptotic behavior of spherical integrals using large deviation techniques. Their Theorems 3 and 6 imply Lemma 6 and can be used to obtain the first term in the asymptotic expansion of Lemma 5.

4. Asymptotic behavior of the likelihood ratios. In this section, we discuss the asymptotic behavior of the likelihood ratios $L(h;\lambda)$ and $L(h;\mu)$. First, let us focus on the case where $h \le \bar{h}$. In the Appendix, we use Lemmas 4 and 5 to derive the following theorem.

THEOREM 7. *Suppose that the null hypothesis is true* ($h = 0$). *Let* $\bar{h}$ *be any fixed number such that* $0 < \bar{h} < \sqrt{c}$ *and let* $C[0, \bar{h}]$ *be the space of real-valued continuous functions on* $[0, \bar{h}]$ *equipped with the supremum norm. Then as* p, $n \to_c \infty$, *we have, uniformly in* $h \in (0, \bar{h}]$

$$L(h;\lambda) = e^{-[\Delta_p(z_0(h)) - \ln(1-h^2/c_p)]/2} + O_p\bigl(n^{-1}\bigr) \tag{4.1}$$

and

$$L(h;\mu) = e^{-[\Delta_p(z_0(h)) - \ln(1-h^2/c_p) - h^2/(2c_p) + (h/c_p)(S-p)]/2} + O_p\bigl(n^{-1}\bigr). \tag{4.2}$$

Furthermore, $\ln L(h;\lambda)$ *and* $\ln L(h;\mu)$, *viewed as random elements of* $C[0, \bar{h}]$, *converge weakly to* $\mathcal{L}(h;\lambda)$ *and* $\mathcal{L}(h;\mu)$ *with Gaussian finite-dimensional distributions such that, for any* $h_1, \dots, h_r \in [0, \bar{h}]$,

$$\mathrm{E}\bigl(\mathcal{L}(h_j;\lambda)\bigr) = \tfrac{1}{4}\ln\bigl(1 - c^{-1}h_j^2\bigr), \tag{4.3}$$

$$\mathrm{Cov}\bigl(\mathcal{L}(h_j;\lambda), \mathcal{L}(h_k;\lambda)\bigr) = -\tfrac{1}{2}\ln\bigl(1 - c^{-1}h_j h_k\bigr), \tag{4.4}$$

$$\mathrm{E}\bigl(\mathcal{L}(h_j;\mu)\bigr) = \tfrac{1}{4}\bigl[\ln\bigl(1 - c^{-1}h_j^2\bigr) + c^{-1}h_j^2\bigr] \tag{4.5}$$

and

$$\mathrm{Cov}\bigl(\mathcal{L}(h_j;\mu), \mathcal{L}(h_k;\mu)\bigr) = -\tfrac{1}{2}\bigl[\ln\bigl(1 - c^{-1}h_j h_k\bigr) + c^{-1}h_j h_k\bigr]. \tag{4.6}$$

The log likelihood ratio processes studied in Theorem 7 are not of the standard *locally asymptotically normal* form. This is because they cannot be represented as $\varphi_1(h)W + \varphi_2(h)$, where $\varphi_1(h)$ and $\varphi_2(h)$ are some deterministic functions of h, and W is a standard normal random variable. Indeed, had the representation $\varphi_1(h)W + \varphi_2(h)$ been possible, the covariance of the limiting log likelihood process at h_1 and h_2 would have been $\varphi_1(h_1)\varphi_1(h_2)$. Hence, for $\mathcal{L}(h;\lambda)$, for instance, we would have had $\varphi_1(h) = \sqrt{-\frac{1}{2}\ln(1 - c^{-1}h^2)}$ and $\varphi_1(h_1)\varphi_1(h_2) = -\frac{1}{2}\ln(1 - c^{-1}h_1 h_2)$, which cannot be true for all $0 < h_1 < \sqrt{c}$ and $0 < h_2 < \sqrt{c}$.

The quantity $\Delta_p(z_0(h))$ plays an important role in the limits of experiments. The likelihood ratio processes are well approximated by simple functions of $\Delta_p(z_0(h))$ and S, which are easy to compute from the data and are asymptotically Gaussian by the central limit theorem of Bai and Silverstein (2004). Recalling the definition (3.1) of $\Delta_p(z_0(h))$, we see that asymptotically, all statistical information about parameter h is contained in the deviations of the sample covariance eigenvalues $\lambda_1, \dots, \lambda_p$ from $\lim_{n,p\to\infty} z_0(h) = \frac{(1+h)(h+c)}{h}$. Although the latter limit does not have an obvious interpretation when $h < \sqrt{c}$, it is the probability limit of λ_1 under alternatives with $h > \sqrt{c}$; see, for example, Baik and Silverstein (2006).

Let us now consider cases where $h > \sqrt{c}$. We prove the following theorem in the Appendix.

THEOREM 8. *Suppose that the null hypothesis is true* ($h = 0$), *and let H be any fixed number such that $\sqrt{c} < H < \infty$. Then as $p, n \to_c \infty$, the following holds. For any $h \in [H, \infty)$, the likelihood ratios $L(h; \lambda)$ and $L(h; \mu)$ converge to zero*; *more precisely*, *there exists $\delta > 0$ that depends only on H such that*

$$L(h; \lambda) = O_p(e^{-n\delta}) \quad \textit{and} \quad L(h; \mu) = O_p(e^{-n\delta}). \tag{4.7}$$

Note that Theorem 7 and Le Cam's first lemma [see van der Vaart (1998), page 88] imply that the joint distributions of $\lambda_1, \dots, \lambda_m$ (as well as those of $\mu_1, \dots, \mu_{m-1}$) under the null and under the alternative are mutually contiguous for any $h \in [0, \sqrt{c})$. In contrast, Theorem 8 shows that mutual contiguity is lost for $h > \sqrt{c}$. For such h, consistent tests (as $p, n \to_c \infty$) exist at any probability level $\alpha > 0$.

In a similar setting, Nadakuditi and Edelman (2008) call the number of "signal eigenvalues" of the population covariance matrix that exceed $1 + \sqrt{c}$ the "effective number of identifiable signals" [see also Nadakuditi and Silverstein (2010)]. Theorems 7 and 8 shed light on the formal statistical content of this concept. The "identifiable signals" are detected with probability approaching one in large samples (irrespective of the probability level $\alpha > 0$ at which identification tests are performed). Other signals still can be detected, but the probability of detecting them will never approach one (whatever the probability level $\alpha < 1$).

5. Asymptotic power analysis. Theorem 7 can be used to study "local" powers of the tests for detecting signals in noise. The nonstandard form of the limit of log likelihood ratio processes in our setting makes it hard to develop tests with optimal local power properties. However, using the Neyman–Pearson lemma and Le Cam's third lemma, we can analytically derive the local asymptotic power envelope and compare local asymptotic powers of specific tests to this envelope.

It is convenient to reparametrize our problem to $\theta = \sqrt{-\ln(1 - h^2/c)}$. As h varies in the region of contiguity $[0, \sqrt{c})$, θ spans the entire half-line $[0, \infty)$. Note that the asymptotic mean and autocovariance functions of the log likelihood ratios derived in the previous section depend on h only through $h/\sqrt{c} = \sqrt{1 - e^{-\theta^2}}$. Therefore, under the new parametrization, they depend only on θ. Loosely speaking, θ and $\sqrt{p/n} \sim \sqrt{c}$ play the classical roles of a "local parameter" and a contiguity rate, respectively.

Let $\beta(\theta_1; \lambda)$ and $\beta(\theta_1; \mu)$ be the asymptotic powers of the asymptotically most powerful λ- and μ-based tests of size α of the null $\theta = 0$ against the alternative $\theta = \theta_1$. The following proposition is proven in the Appendix.

PROPOSITION 9. *Let Φ denote the standard normal distribution function. Then*

$$\beta(\theta_1; \lambda) = 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \frac{\theta_1}{\sqrt{2}}\right] \tag{5.1}$$

and

$$\beta(\theta_1; \mu) = 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \sqrt{\tfrac{1}{2}\left(\theta_1^2 - 1 + e^{-\theta_1^2}\right)}\right]. \tag{5.2}$$

Plots of the asymptotic power envelopes $\beta(\theta_1; \lambda)$ and $\beta(\theta_1; \mu)$ against θ_1 for asymptotic size $\alpha = 0.05$ are shown in the left panel of Figure 3. The power loss of the μ-based tests relative to the λ-based tests is due to the nonspecification of σ^2. In contrast to λ-based tests, μ-based tests may achieve the corresponding power envelope even when σ^2 is unknown.

The right panel of Figure 3 shows the envelopes as functions of the original parameter h normalized by $\sqrt{c}$. We see that the alternatives that can theoretically

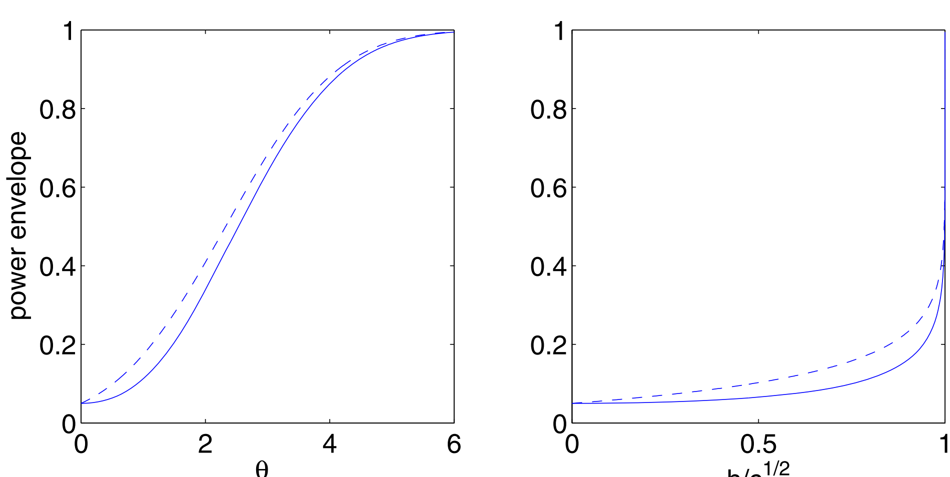

FIG. 3. *The maximal asymptotic power of the λ-based tests (dashed lines) and μ-based tests (solid lines) of $\theta = 0$ against $\theta = \theta_1$. Left panel: θ-parametrization. Right panel: h-parametrization.*

be detected with high probability are concentrated near the threshold $h = \sqrt{c}$. The strong nonlinearity of the θ-parametrization should be kept in mind while interpreting the figures that follow.

It is interesting to compare the power envelopes to the asymptotic powers of the likelihood ratio (LR) and weighted average power (WAP) tests. The λ-based LR and WAP tests of $\theta = 0$ against the alternative $\theta \in (0, M]$, where $M < \infty$, would reject the null if and only if, respectively, $2\sup_{\theta\in(0,M]} \ln L(\theta; \lambda)$ and $\ln \int_0^M L(\theta; \lambda) W(\mathrm{d}\theta)$ are sufficiently large. The power of a WAP test would, of course, depend on the choice of the weighting measure $W(\mathrm{d}\theta)$. The μ-based LR and WAP tests are defined similarly. Theorem 7 and Le Cam's third lemma suggest a straightforward procedure for the numerical evaluation of the corresponding asymptotic power functions.

Consider, for example, the λ-based LR test statistic. According to Theorem 7, its asymptotic distribution under the null equals the distribution of $2\sup_{\theta\in(0,M]} X_\theta$, where X_θ is a Gaussian process with $\mathrm{E}(X_\theta) = -\theta^2/4$ and $\mathrm{Cov}(X_{\theta_1}, X_{\theta_2}) = -\frac{1}{2}\ln(1 - \sqrt{(1 - e^{-\theta_1^2})(1 - e^{-\theta_2^2})})$. According to Le Cam's third lemma, under a specific alternative $\theta = \theta_1 \le M$, the asymptotic distribution of the LR statistic equals the distribution of $2\sup_{\theta\in(0,M]} \tilde{X}_\theta$, where $\tilde{X}_\theta$ is a Gaussian process with the same covariance function as that of X_θ, but with a different mean: $\mathrm{E}(\tilde{X}_\theta) = \mathrm{E}(X_\theta) + \mathrm{Cov}(X_\theta, X_{\theta_1})$.

Therefore, to numerically evaluate the asymptotic power function of the λ-based LR test, we simulate 500,000 observations of X_θ on a grid of 1000 equally spaced points in $\theta \in [0, M = 6]$, where $M = 6$ is chosen as the upper limit of the grid because it is large enough for the power envelopes to rich the value of 99%. For each observation, we save its supremum on the grid, and use the empirical distribution of two times the suprema as the approximate asymptotic distribution of the likelihood ratio statistic under the null. We denote this distribution as $\hat{F}_0$. Its 95% quantile equals 4.3982.

For each θ_1 on the grid, we repeat the simulation for process $\tilde{X}_\theta$ to obtain the approximate asymptotic distribution of the likelihood ratio statistic under the alternative $\theta = \theta_1$, which we denote as $\hat{F}_1$. We use the value of $\hat{F}_1$ at the 95% quantile of $\hat{F}_0$ as a numerical approximation to the asymptotic power at θ_1 of the λ-based LR test with asymptotic size 0.05.

Figure 4 shows the resulting asymptotic power curve of the LR test (solid line) along with the asymptotic power envelope (dotted line). It also shows the asymptotic power of the WAP test with $W(\mathrm{d}\theta)$ equal to the uniform measure on $[0, 6]$ (dashed line). The left and right panels correspond to λ- and μ-based tests, respectively.

The asymptotic powers of the LR and WAP tests both come close to the power envelope. The LR and WAP power functions are so close that they are difficult to distinguish clearly. The asymptotic power of the WAP test appears to be larger than that of the LR test for all θ_1 in the $[0, 6]$ range, except for relatively large θ_1.

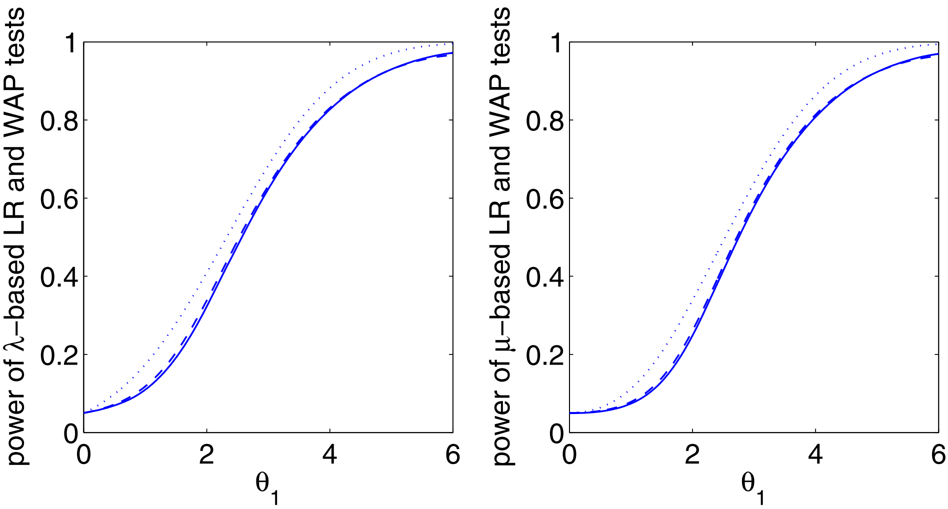

FIG. 4. *The asymptotic power envelope* (*dotted line*), *the asymptotic power of the LR test* (*solid line*), *and the asymptotic power of the WAP test with uniform weighting measure on* $\theta \in [0, 6]$ (*dashed line*). *Left panel*: λ*-based tests and envelope*. *Right panel*: μ*-based tests and envelope*.

Hence, the LR test still may be admissible. More accurate numerical analysis is needed to shed further light on this issue.

In the remaining part of this section, we consider some of the tests that have been proposed previously in the literature, and, in Proposition 10, derive their asymptotic power functions. We focus on four examples. Three of them are inspired by the "classical" fixed-p theory, while the fourth is more directly based on results from the large random matrix theory.

The problem of testing the hypothesis of sphericity has a long history, and has generated a considerable body of literature, which we only very briefly summarize here. The classical fixed-p Gaussian analysis of the various problems considered here goes back to Mauchly (1940), who first derived the Gaussian likelihood ratio test for sphericity. The (Gaussian) locally most powerful invariant (under shift, scale and orthogonal transformations) test was obtained by John (1971, 1972) and by Sugiura (1972), with adjusted versions resisting elliptical violations of the Gaussian assumptions proposed in Hallin and Paindaveine (2006), where a Le Cam approach is adopted under a general elliptical setting. Ledoit and Wolf (2002) propose two extensions (for the unknown and known scale problems, resp.) of John's test, while Bai et al. (2009) adapt Mauchly's (1940) likelihood ratio test.

EXAMPLE 1 [John's (1971) test of sphericity]. John (1971) proposes testing the sphericity hypothesis $\theta = 0$ against general alternatives using the test statistic $U = \frac{1}{p}\operatorname{tr}[(\frac{\hat{\Sigma}}{(1/p)\operatorname{tr}(\hat{\Sigma})} - I_p)^2]$, where $\hat{\Sigma}$ is the sample covariance matrix of the data. He shows that, when $n > p$, such a test is locally most powerful invariant. Studying

John's test when $p/n \to c \in (0, \infty)$, Ledoit and Wolf (2002) prove that, under the null, $nU - p \xrightarrow{d} N(1, 4)$. Hence, the test with asymptotic size α rejects the null hypothesis of sphericity if $\frac{1}{2}(nU - p - 1) > \Phi^{-1}(1 - \alpha)$.

EXAMPLE 2 [The Ledoit and Wolf (2002) test of $\Sigma = I$]. Ledoit and Wolf (2002) propose using $W = \frac{1}{p}\operatorname{tr}[(\hat{\Sigma} - I)^2] - \frac{p}{n}[\frac{1}{p}\operatorname{tr}\hat{\Sigma}]^2 + \frac{p}{n}$ as a test statistic for testing the hypothesis that the population covariance matrix is a unit matrix. Under the null, $nW - p \xrightarrow{d} N(1, 4)$. As in the previous example, the null hypothesis is rejected at asymptotic size α if $\frac{1}{2}(nW - p - 1) > \Phi^{-1}(1 - \alpha)$.

EXAMPLE 3 [The "corrected" LRT of Bai et al. (2009)]. When $n > p$, Bai et al. (2009) propose a corrected version of the likelihood ratio statistic $\mathrm{CLR} = \operatorname{tr}\hat{\Sigma} - \ln\det\hat{\Sigma} - p - p(1 - (1 - \frac{n}{p})\ln(1 - \frac{p}{n}))$ based on the entire data, as opposed to λ or μ only, to test the equality of the population covariance matrix to the identity matrix against general alternatives. Under the null, $\mathrm{CLR} \xrightarrow{d} N(-\frac{1}{2}\ln(1 - c), -2\ln(1 - c) - 2c)$. The null hypothesis is rejected whenever $\mathrm{CLR} + \frac{1}{2}\ln(1 - c) > \sqrt{-2\ln(1 - c) - 2c}\,\Phi^{-1}(1 - \alpha)$.

More directly inspired by the asymptotic theory of random matrices, several authors have recently proposed and studied various tests based on λ_1 or μ_1: see Bejan (2005), Patterson, Price and Reich (2006), Kritchman and Nadler (2009), Onatski (2009), Bianchi et al. (2011) and Nadakuditi and Silverstein (2010). We refer to these tests, which reject H_0 for large values of λ_1 or μ_1, as Tracy–Widom-type tests.

EXAMPLE 4 (Tracy–Widom-type tests). Asymptotic critical values of such tests are obtained using the fact, established by Johnstone (2001), that under the null,

$$n^{2/3}c^{1/6}(1 + \sqrt{c})^{-4/3}\bigl(\lambda_1 - (1 + \sqrt{c})^2\bigr) \xrightarrow{d} \mathrm{TW}, \tag{5.3}$$

where TW denotes the Tracy–Widom law of the first kind. The null hypothesis is rejected when λ_1 or μ_1 exceeds the adequate Tracy–Widom quantile.

Consider the tests described in Examples 1, 2, 3 and 4, and denote by $\beta_{\mathrm{J}}(\theta_1)$, $\beta_{\mathrm{LW}}(\theta_1)$, $\beta_{\mathrm{CLR}}(\theta_1)$, and $\beta_{\mathrm{TW}}(\theta_1)$ their respective asymptotic powers at asymptotic level α. The following proposition is established in the Appendix.

PROPOSITION 10. *Denote* $1 - e^{-\theta_1^2}$ *as* $\psi(\theta_1)$. *The asymptotic power functions of the tests described in Examples 1–4 satisfy, for any* $\theta_1 > 0$,

$$\beta_{\mathrm{TW}}(\theta_1) = \alpha, \tag{5.4}$$

$$\beta_{\mathrm{J}}(\theta_1) = \beta_{\mathrm{LW}}(\theta_1) = 1 - \Phi\bigl(\Phi^{-1}(1 - \alpha) - \tfrac{1}{2}\psi(\theta_1)\bigr) \tag{5.5}$$

and

$$\beta_{\mathrm{CLR}}(\theta_1) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\sqrt{c\psi(\theta_1)} - \ln(1+\sqrt{c\psi(\theta_1)})}{\sqrt{-2\ln(1-c) - 2c}}\right). \tag{5.6}$$

With the important exception of Srivastava (2005), (5.4)–(5.6) are the first results on the asymptotic power of those tests against contiguous alternatives. Srivastava (2005) analyzes the asymptotic power of tests similar to those in Examples 1 and 2. His Theorems 3.1 and 4.1 can be used to establish (5.5).

From Proposition 10, we see that the local asymptotic power of the Tracy–Widom-type tests is trivial. As shown by Baik, Ben Arous and Péché (2005) in the complex data case and by Féral and Péché (2009) in the real data case, the convergence (5.3) holds not only under the null, but also under any alternative of the form $h = h_0 < \sqrt{c}$. Under the "local" parametrization adopted in this section, such alternatives have the form $\theta = \theta_1 > 0$. It can be shown that the Tracy–Widom-type tests are consistent against noncontiguous alternatives $h = h_1 > \sqrt{c}$. However, such a consistency is likely to be also a property of the LR tests based on μ or on λ. If this holds true, the LR tests asymptotically dominate the Tracy–Widom-type tests. A more detailed analysis of the optimality properties of LR tests is the subject of ongoing research.

The asymptotic power functions of the tests from Examples 1, 2 and 3 are nontrivial. Figure 5 compares these power functions to the corresponding power envelopes. Since John's test is invariant with respect to orthogonal transformations and scalings, $\beta_{\mathrm{J}}(\theta_1)$ is compared to the power envelope $\beta(\theta_1; \mu)$. The asymptotic power functions $\beta_{\mathrm{LW}}(\theta_1)$ and $\beta_{\mathrm{CLR}}(\theta_1)$ are compared to the power envelope

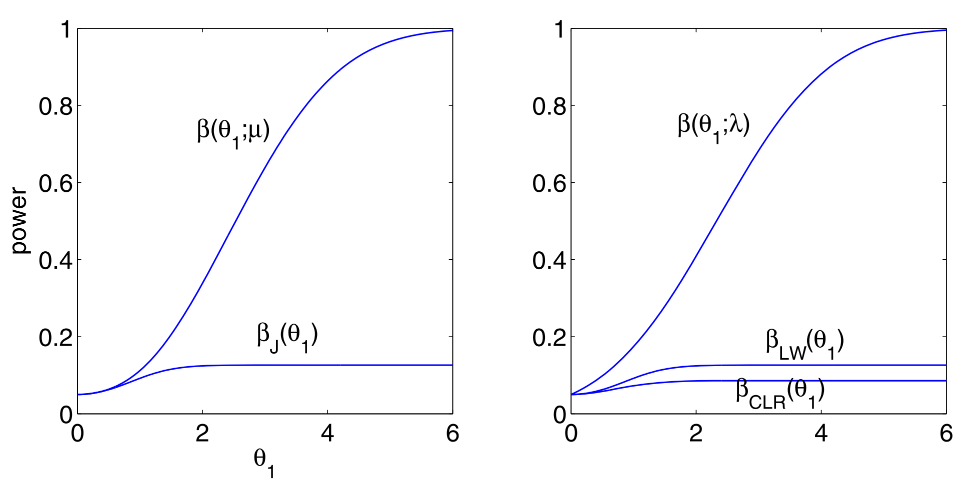

FIG. 5. *Asymptotic powers* ($\beta_{\mathrm{J}}, \beta_{\mathrm{LW}}, \beta_{\mathrm{CLR}}$) *of the tests described in Examples* 1 (*John*), 2 (*Ledoit and Wolf*) *and* 3 (*Bai et al.*).

$\beta(\theta_1; \lambda)$ because the Ledoit–Wolf test of $\Sigma = I$ and the "corrected" likelihood ratio test are invariant only with respect to orthogonal transformations.

Interestingly, whereas $\beta_{\mathrm{J}}(\theta_1)$ and $\beta_{\mathrm{LW}}(\theta_1)$ depend only on α and θ_1, $\beta_{\mathrm{CLR}}(\theta_1)$ depends also on c. As c converges to one, $\beta_{\mathrm{CLR}}(\theta_1)$ converges to α, which corresponds to the case of trivial power. As c converges to zero, $\beta_{\mathrm{CLR}}(\theta_1)$ converges to $\beta_{\mathrm{J}}(\theta_1)$. In Figure 5, we provide the plot of $\beta_{\mathrm{CLR}}(\theta_1)$ that corresponds to $c = 0.5$.

The left panel of Figure 5 shows that the power function of John's test is very close to the power envelope $\beta(\theta_1; \mu)$ in the vicinity of $\theta_1 = 0$. Such behavior is consistent with the fact that John's test is locally most powerful invariant. However, for large θ_1, the asymptotic power functions of all the tests from Examples 1, 2 and 3 are lower than the corresponding asymptotic power envelopes. We should stress here that these tests have power against general alternatives as opposed to the "spiked" alternatives that maintain the assumption that the population covariance matrix of data has the form $\sigma^2(I_p + hvv')$.

For the "spiked" alternatives, the λ- and μ-based LR tests may be more attractive. However, implementing these tests requires some care. A "quick-and-dirty" approach would be to approximate $\ln L(\theta; \lambda)$ and $\ln L(\theta; \mu)$ by the simple but asymptotically equivalent expressions from (4.1) and (4.2), compute two times their maxima on a grid over $\theta \in (0, M]$, and compare them with critical values obtained by simulation as for the construction of Figure 4. Unfortunately, in finite samples, this simple approach will lead to a numerical breakdown whenever $z_0(h(\theta))$ happens to be less than the largest sample covariance eigenvalue for some $\theta \leq M$. In addition, since the asymptotic approximation derived in Theorem 7 is not uniform over entire half-line $\theta \in [0, \infty)$, its quality will depend on the choice of M. For relatively large M, the asymptotic behavior of the LR test implemented as above may poorly match its finite sample behavior.

Instead, we recommend implementing the LR tests without using the asymptotic approximations. The finite sample log likelihood ratios $\ln L(\theta; \lambda)$ and $\ln L(\theta; \mu)$ can be computed using the contour integral representations (2.9) and (2.10). Choosing the contour of integration so that the sample covariance eigenvalues remain to its left will eliminate the numerical breakdown problem associated with the asymptotic tests. Furthermore, under the Gaussianity assumption, the finite sample distributions of the log likelihood ratios are pivotal. Hence, the exact critical values can be computed via Monte Carlo simulations as follows: simulate many replications of data under the null. For each replication, compute the log likelihood ratio and store two times its maximum. Use the 95% quantile of the empirical distribution of the stored values as a numerical approximation for the exact critical value of the test. The finite sample properties of such a test are left as an important topic for future research.

6. Conclusion. In this paper, we study the asymptotic power of tests for the existence of rank-one perturbations of sphericity as both the dimensionality of the data and the number of observations go to infinity. Focusing on tests that are invariant with respect to orthogonal transformations and rescaling, we establish the convergence of the log ratio of the joint densities of the sample covariance eigenvalues under the alternative and null hypotheses to a Gaussian process indexed by the norm of the perturbation.

When the perturbation norm is larger than the phase transition threshold studied in Baik, Ben Arous and Péché (2005), the limiting log-likelihood process is degenerate and the joint eigenvalue distributions under the null and alternative hypotheses are asymptotically mutually singular, so that the discrimination between the null and the alternative is asymptotically certain. When the norm is below the threshold, the limiting log-likelihood process is nondegenerate and the joint eigenvalue distributions under the null and alternative hypotheses are mutually contiguous. Using the asymptotic theory of statistical experiments, we obtain power envelopes and derive the asymptotic size and power for various eigenvalue-based tests in the region of contiguity.

Several questions are left for future research. First, we only considered rank-one perturbations of the spherical covariance matrices. It would be desirable to extend the analysis to finite-rank perturbations. Such an extension will require a more complicated technical analysis. Second, it would be interesting to extend our analysis to the asymptotic regime $p, n \to \infty$ with $p/n \to \infty$ or $p/n \to 0$. In the context of sphericity tests, such asymptotic regimes have been recently studied in Birke and Dette (2005). Third, a thorough analysis of the finite sample properties of the proposed LR tests would clarify the related practical implementation issues. Fourth, our Lemma 5 can be used to derive higher-order asymptotic approximations to the likelihood ratios, which may improve finite-sample performances of asymptotic tests. Finally, it would be of considerable interest to relax the Gaussian assumptions, for example, into elliptical ones, preferably with unspecified radial densities, on the model (in a fixed-p context) of Hallin and Paindaveine (2006).

APPENDIX

A.1. Proof of Proposition 1. For the joint density $p(\lambda; h)$ of $\lambda_1, \dots, \lambda_m$, we have

$$
\begin{aligned}
p(\lambda; h) = \tilde{\gamma} \frac{\prod_{i=1}^{m} \lambda_i^{(|p-n|-1)/2} \prod_{i<j}^{m} (\lambda_i - \lambda_j)}{(1+h)^{n/2}} \\
\times \int_{\mathcal{O}(p)} e^{-(n/2)\operatorname{tr}(\Pi Q' \Lambda Q)} (\mathrm{d}Q),
\end{aligned}
\tag{A.1}
$$

where $\tilde{\gamma}$ depends only on n and p, $\Pi = \operatorname{diag}((1+h)^{-1}, 1, \dots, 1)$, $\mathcal{O}(p)$ is the set of all $p \times p$ orthogonal matrices and $(\mathrm{d}Q)$ is the invariant measure on the orthogonal group $\mathcal{O}(p)$ normalized to make the total measure unity. When $n \geq p$, (A.1) is a special case of the density given in James (1964), page 483. When $n < p$, (A.1) follows from Theorems 2 and 6 in Uhlig (1994).

Let $\Psi = \mathrm{diag}(\frac{h}{1+h}, 0, \dots, 0)$ be a $p \times p$ matrix. Since $\Pi = I_p - \Psi$, we have $\mathrm{tr}(\Pi Q'\Lambda Q) = \mathrm{tr}\,\Lambda - \mathrm{tr}(\Psi Q'\Lambda Q)$, and we can rewrite (A.1) as

$$p(\lambda; h) = \tilde{\gamma} \frac{\prod_{i=1}^m \lambda_i^{(|p-n|-1)/2} \prod_{i<j}^m (\lambda_i - \lambda_j) e^{-(n/2)\,\mathrm{tr}\,\Lambda}}{(1+h)^{n/2}} \times \int_{\mathcal{O}(p)} e^{(n/2)\,\mathrm{tr}(\Psi Q'\Lambda Q)} (\mathrm{d}Q). \tag{A.2}$$

Note that $\mathrm{tr}(\Psi Q'\Lambda Q) = \mathrm{tr}(Q\Psi Q'\Lambda) = \frac{h}{1+h} x_p'\Lambda x_p$, where x_p is the first column of Q. When Q is uniformly distributed over $\mathcal{O}(p)$, its first column x_p is uniformly distributed over $\mathcal{S}(p)$. Therefore, we have

$$p(\lambda; h) = \tilde{\gamma} \frac{\prod_{i=1}^m \lambda_i^{(|p-n|-1)/2} \prod_{i<j}^m (\lambda_i - \lambda_j) e^{-(n/2)\,\mathrm{tr}\,\Lambda}}{(1+h)^{n/2}} \times \int_{\mathcal{S}(p)} e^{(n/2)(h/(1+h))x_p'\Lambda x_p} (\mathrm{d}x_p), \tag{A.3}$$

which establishes (2.1). Now, let $y = \lambda_1 + \cdots + \lambda_p$ so that $\mu_j = \lambda_j / y$. Note that $\mathrm{tr}\,\Lambda = y$, $\mathrm{tr}\,M = \mu_1 + \cdots + \mu_p = 1$, and that the Jacobian of the coordinate change from $\lambda_1, \dots, \lambda_m$ to $\mu_1, \dots, \mu_{m-1}, y$ equals y^{m-1}. Changing variables in (A.3), and integrating y out, we obtain (2.2).

A.2. Proof of Lemma 3. Using (2.3) in the ratio of the right-hand side of (2.1) with $h > 0$ to that with $h = 0$, and changing the variable of integration from s to $z = \frac{1+h}{h}\frac{2}{n}s$, we get (2.9). Further, from (2.2), we have

$$\begin{aligned} p(\mu; 0) &= \delta(n, p, \mu) \int_0^\infty y^{np/2-1} e^{-ny/2}\, \mathrm{d}y \\ &= \delta(n, p, \mu) \left(\frac{2}{n}\right)^{np/2} \Gamma\left(\frac{np}{2}\right). \end{aligned} \tag{A.4}$$

For $h > 0$, using (2.3) in (2.2), we get

$$p(\mu; h) = \frac{\delta(n, p, \mu)}{(1+h)^{n/2}} \frac{\Gamma(p/2)}{2\pi i} \times \int_0^\infty \oint_{\tilde{\mathcal{K}}} y^{(np-2)/2} e^{s-ny/2} \prod_{j=1}^p \left(s - \frac{n}{2}\frac{yh}{1+h}\mu_j\right)^{-1/2} \mathrm{d}s\, \mathrm{d}y,$$

where $\tilde{\mathcal{K}}$ is a contour starting at $-\infty$, encircling counter-clockwise the points 0, $\frac{ny}{2}\frac{h}{1+h}\mu_1, \dots, \frac{ny}{2}\frac{h}{1+h}\mu_m$ and going back to $-\infty$. Since $\frac{h}{1+h}\mu_j < 1$ by construction, we may and will choose $\tilde{\mathcal{K}}$ so that for any $s \in \tilde{\mathcal{K}}$, $\mathrm{Re}\, s < \frac{ny}{2}$. Changing variables of integration from y and s to $w = \frac{ny}{2}$ and $z = s\frac{1+h}{hw}S$, where S is any positive constant, and dividing by the right-hand side of (A.4), we obtain

$$L(h;\mu) = \frac{S^{(p-2)/2}(1+h)^{(p-n-2)/2}\Gamma(p/2)}{h^{(p-2)/2}\Gamma(np/2)2\pi i} \times \int_0^\infty \oint_{\mathcal{K}} w^{np/2-p/2} e^{(wh/(1+h))(z/S)-w} \prod_{j=1}^p (z - S\mu_j)^{-1/2}\,\mathrm{d}z\,\mathrm{d}w,$$

where $\mathcal{K}$ is a contour starting at $-\infty$, encircling counter-clockwise the points 0, $S\mu_1, \ldots, S\mu_m$, and going back to $-\infty$. In addition, for any $z \in \mathcal{K}$, $\operatorname{Re} z < \frac{1+h}{h} S$. Such a choice of $\mathcal{K}$ guarantees that the integrand in the above double integral is absolutely integrable on $[0,\infty) \times \mathcal{K}$, so that Fubini's theorem can be used to justify the interchange of the order of the integrals. Changing the order of the integrals and setting $S = \lambda_1 + \cdots + \lambda_p$, we obtain (2.10).

A.3. Proof of Theorem 7. First, let us formulate the following technical lemma. Its proof is in the Supplementary Appendix [Onatski, Moreira and Hallin (2013)].

LEMMA 11. (i) *If* $h < \sqrt{c_p}$, $f_0 = -\frac{1}{2}(c_p + (1-c_p)\ln(1+h) - c_p \ln \frac{c_p}{h})$.
(ii) *If* $h > \sqrt{c_p}$, $f_0 = -\frac{1}{2}(h + c_p + (1-c_p)\ln(c_p+h) - \frac{c_p}{h} - \ln h)$.

Below, we prove Theorem 7 for $L(h;\mu)$. The proof for $L(h;\lambda)$ is similar but simpler, and we omit it to save space. As follows from Lemmas 4 and 5, the integral in (2.10) can be represented as $2e^{-nf_0}[\Gamma(\frac{1}{2})\frac{a_0}{n^{1/2}} + \frac{O_p(1)}{hn^{3/2}}]$ uniformly in $h \in (0,\bar{h}]$. Therefore, and since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, we can write

$$L(h;\mu) = \frac{k_2 S^{(p-2)/2}}{\sqrt{n\pi}\,i} e^{-nf_0}\left[a_0 + h^{-1} O_p\left(\frac{1}{n}\right)\right], \tag{A.5}$$

where $k_2 = h^{-(p-2)/2}(1+h)^{(p-n-2)/2}\frac{(n-1)p}{2}\Gamma(\frac{(n-1)p}{2})\Gamma(\frac{p}{2})\Gamma^{-1}(\frac{np}{2})$. Using Stirling's approximation $\Gamma(r) = e^{-r}r^r(\frac{2\pi}{r})^{1/2}(1 + O(r^{-1}))$ with $r = \frac{p}{2}$, $\frac{np}{2}$ and $\frac{(n-1)p}{2}$, and the fact that $\ln(n-1) = \ln n - n^{-1} - \frac{1}{2}n^{-2} + O(n^{-3})$, we find, after algebraic simplifications, that

$$\frac{k_2}{\sqrt{n\pi}} = h^{-(p-2)/2}(1+h)^{(p-n-2)/2} \times e^{-((p-2)/2)\ln n - p/2 + c_p/4 + \ln c_p/2}\bigl(1 + O(n^{-1})\bigr). \tag{A.6}$$

Using (A.6) and Lemma 11(i), we obtain

$$\frac{k_2 S^{(p-2)/2}}{\sqrt{n\pi}\,i} e^{-nf_0} h^{-1} O_p\left(\frac{1}{n}\right) = \frac{1}{1+h}\left(\frac{S}{p}\right)^{(p-2)/2} e^{c_p/4 - \ln c_p/2} O_p\left(\frac{1}{n}\right),$$

which, together with the fact that $S - p = O_p(1)$, implies that

$$\frac{k_2 S^{(p-2)/2}}{\sqrt{n\pi} i} e^{-nf_0} h^{-1} O_p\left(\frac{1}{n}\right) = O_p\left(\frac{1}{n}\right) \tag{A.7}$$

uniformly over $h \in (0, \bar{h}]$.

Now, as can be verified using (3.3) and (3.6), if $h < \sqrt{c_p}$, then

$$f_2 = -\frac{h^2}{4(1+h)^2(c_p - h^2)}. \tag{A.8}$$

Therefore, using (3.12), we obtain

$$a_0 = i\frac{(1+h)(c_p - h^2)^{1/2}}{h} g_0. \tag{A.9}$$

Using (3.4), (A.6), (A.9) and Lemma 11(i) in (A.5), after algebraic simplifications and rearrangements of terms, we get

$$\begin{aligned} &\ln\left[\frac{k_2 S^{(p-2)/2} e^{-nf_0} a_0}{\sqrt{n\pi} i}\right] \\ &\quad = \frac{1}{2}\ln\left(1 - \frac{h^2}{c_p}\right) + \frac{c_p}{4} + \frac{p-2}{2}\ln\left(\frac{S}{p}\right) \\ &\qquad - \frac{n}{2}\frac{hz_0(h)}{1+h} - \frac{np - p + 2}{2}\ln\left(1 - \frac{h}{1+h}\frac{z_0(h)}{S}\right) - \frac{1}{2}\Delta_p(z_0(h)). \end{aligned} \tag{A.10}$$

Finally, using the fact that $S - p = O_p(1)$, we obtain $\ln(S/p) = (S-p)/p + O_p(p^{-2})$ and

$$\begin{aligned} \ln\left(1 - \frac{h}{1+h}\frac{z_0(h)}{S}\right) &= -\frac{h}{1+h}\frac{z_0(h)}{p} - \frac{1}{2}\left(\frac{hz_0(h)}{(1+h)p}\right)^2 \\ &\quad + \frac{h}{1+h}\frac{z_0(h)}{p^2}(S - p) + O_p(p^{-3}). \end{aligned}$$

The latter two equalities, (A.10) and the fact that $\frac{h}{1+h} z_0(h) = h + c_p$ entail

$$\begin{aligned} &\frac{k_2 S^{(p-2)/2} e^{-nf_0} a_0}{\sqrt{n\pi} i} \\ &\quad = e^{-\{\Delta_p(z_0(h)) - \ln(1 - h^2/c_p) + (h/c_p)(S-p) - h^2/(2c_p) + O_p(p^{-1})\}/2}, \end{aligned} \tag{A.11}$$

which, together with (A.7), imply formula (4.2).

Now, let us prove the convergence of $\ln L(h;\mu)$ to $\mathcal{L}(h;\mu)$. By (4.2), the joint convergence of $\ln L(h_j;\mu)$ with $j = 1, \dots, r$ to a Gaussian vector is equivalent to the convergence of $(S - p, \Delta_p(z_0(h_1)), \dots, \Delta_p(z_0(h_r)))$ to a Gaussian vector. A proof of the following technical lemma, based on Theorem 1.1 of Bai and Silverstein (2004), is given in the Supplementary Appendix [Onatski, Moreira and Hallin (2013)].

LEMMA 12. *Suppose that the null hypothesis holds. Then, as $p, n \to_c \infty$, the vector $(S - p, \Delta_p(z_0(h_1)), \dots, \Delta_p(z_0(h_r)))$ converges in distribution to a Gaussian vector $(\eta, \xi_1, \dots, \xi_r)$ with*

$$\mathrm{E}\eta = 0, \qquad \mathrm{Var}(\eta) = 2c, \qquad \mathrm{Cov}(\eta, \xi_j) = -2h_j,$$
$$\mathrm{Cov}(\xi_j, \xi_k) = -2\ln\bigl(1 - c^{-1}h_j h_k\bigr) \quad \text{and} \quad \mathrm{E}\xi_j = \tfrac{1}{2}\ln\bigl(1 - c^{-1}h_j^2\bigr).$$

Lemma 12 and (4.2) imply that $\mathrm{E}[\mathcal{L}(h_j; \mu)] = -\frac{1}{2}\mathrm{E}\xi_j + \frac{1}{2}\ln(1 - c^{-1}h_j^2) + \frac{1}{4}c^{-1}h_j^2 = \frac{1}{4}[\ln(1 - c^{-1}h_j^2) + c^{-1}h_j^2]$ and

$$\begin{aligned}\mathrm{Cov}\bigl[\mathcal{L}(h_j; \mu), \mathcal{L}(h_k; \mu)\bigr] &= \frac{1}{4}\mathrm{Cov}(\xi_j, \xi_k) + \frac{h_k}{4c}\mathrm{Cov}(\xi_j, \eta) \\ &\quad + \frac{h_j}{4c}\mathrm{Cov}(\xi_k, \eta) + \frac{h_j h_k}{4c^2}\mathrm{Var}(\eta) \\ &= -\frac{1}{2}\ln\bigl(1 - c^{-1}h_j h_k\bigr) - \frac{h_j h_k}{2c},\end{aligned}$$

which establishes (4.5) and (4.6).

To complete the proof of Theorem 7, we need to note that the tightness of $L(h; \mu)$, viewed as a random element of the space $C([0, \bar{h}])$, as $p, n \to_c \infty$, follows from formula (4.2) and the fact that $S - p$ and $\Delta_p(z_0(h))$, are $O_p(1)$, uniformly in $h \in (0, \bar{h}]$. This uniformity is a consequence of Lemma A2 proven in the Supplementary Appendix [Onatski, Moreira and Hallin (2013)].

A.4. Proof of Theorem 8. As in the proof of Theorem 7, we will focus on the case of the likelihood ratio based on μ. The proof for $L(h; \lambda)$ is similar. According to Lemma 6 and formula (2.10), for any $\tilde{h} > \sqrt{c}$, we have $L(h; \mu) = k_2 S^{(p-2)/2} e^{-nf(z_0(\tilde{h}))} O_p(1)$. Using (A.6) and the fact that $(\frac{S}{p})^p = (1 + \frac{S-p}{p})^p = (1 + \frac{O_p(1)}{p})^p = O_p(1)$, we can write

$$L(h; \mu) = e^{(n/2)(c_p \ln(c_p(1+h)/h) - \ln(1+h) - c_p - 2f(z_0(\tilde{h})))} O_p\bigl(n^{1/2}\bigr). \tag{A.12}$$

Noting that $\tilde{h} > \sqrt{c_p}$ for sufficiently large n and p, and using Lemma 11(ii) and the fact that $\frac{\tilde{h}}{1+\tilde{h}} z_0(\tilde{h}) = \tilde{h} + c_p$, we get $-2f(z_0(\tilde{h})) = (1 - c_p)\ln(c_p + \tilde{h}) - \frac{c_p}{\tilde{h}} - \ln\tilde{h} + \frac{h}{1+h} z_0(\tilde{h})$. Substituting the latter expression in (A.12) and simplifying, we obtain

$$L(h; \mu) = e^{(n/2)R(h, \tilde{h}, c_p)} O_p\bigl(n^{1/2}\bigr), \tag{A.13}$$

where $O_p(\cdot)$ is uniform in $h \in [\tilde{h}, \infty)$ and $R(h, \tilde{h}, c_p) = (1 - c_p)\ln(c_p + \tilde{h}) - \frac{c_p}{\tilde{h}} - \ln\tilde{h} + \frac{h}{1+h} z_0(\tilde{h}) - (1 - c_p)\ln(1 + h) - c_p \ln h + c_p \ln c_p - c_p$.

As $n, p \to \infty$, $R(h, \tilde{h}, c_p) \to R(h, \tilde{h}, c)$ uniformly over $(h, \tilde{h}) \in [\sqrt{c}, H]^2$. On the other hand, $R(h, \tilde{h}, c)$ is continuous on $(h, \tilde{h}) \in [\sqrt{c}, H]^2$, $R(\sqrt{c}, \sqrt{c}, c) = 0$, and $\frac{\mathrm{d}}{\mathrm{d}h} R(h, \tilde{h}, c) = (1+h)^{-2}(\frac{(1+\tilde{h})(c+\tilde{h})}{\tilde{h}} - \frac{(1+h)(c+h)}{h}) < 0$ for all h and $\tilde{h}$ such that $\sqrt{c} \le \tilde{h} < h \le H$. Therefore, for any $H > \sqrt{c}$, there exist $\tilde{h}$ and δ such that $\sqrt{c} < \tilde{h} \le H$, $\delta > 0$ and $R(H, \tilde{h}, c) < -3\delta$; and thus, for sufficiently large n and p, $R(H, \tilde{h}, c_p) < -3\delta$. Now, $\frac{\mathrm{d}}{\mathrm{d}h} R(h, \tilde{h}, c_p) = (1+h)^{-2}(z_0(\tilde{h}) - z_0(h)) < 0$ for all $h > \tilde{h}$, as long as $\tilde{h} \ge \sqrt{c_p}$. Hence, for sufficiently large n and p, $R(h, \tilde{h}, c_p) < -3\delta$ for all $h > \tilde{h}$. Using (A.13), we get $|L(h; \mu)| \le e^{-3n\delta/2} O_p(n^{1/2}) = O_p(e^{-n\delta})$ uniformly over $h \in [H, \infty)$.

A.5. Proof of Proposition 9. For brevity, we derive only the asymptotic power envelope for the case of μ-based tests. According to the Neyman–Pearson lemma, the most powerful test of the null $\theta = 0$ against a particular alternative $\theta = \theta_1$ is the test which rejects the null when $\ln L(\theta_1; \mu)$ is larger than some critical value C. It follows from Theorem 7 that, for such a test to have asymptotic size α, C must be $C = \sqrt{V(\theta_1)}\Phi^{-1}(1-\alpha) + m(\theta_1)$, where $m(\theta_1) = (-\theta_1^2 + 1 - e^{-\theta_1^2})/4$ and $V(\theta_1) = (\theta_1^2 - 1 + e^{-\theta_1^2})/2$ are obtained from (4.5) and (4.6) by the reparametrization $\theta = \sqrt{-\ln(1 - h^2/c)}$. Now, according to Le Cam's third lemma and Theorem 7, under $\theta = \theta_1$, $\ln L(\theta_1; \mu) \xrightarrow{d} N(m(\theta_1) + V(\theta_1), V(\theta_1))$. Therefore, the asymptotic power $\beta(\theta_1; \mu)$ of the asymptotically most powerful test of $\theta = 0$ against $\theta = \theta_1$ is (5.2).

A.6. Proof of Proposition 10. As shown by Baik, Ben Arous and Péché (2005) in the complex case and by Féral and Péché (2009) in the real case, the convergence (5.3) takes place not only under the null, but also under alternatives $h = h_1$ with $h_1 < \sqrt{c}$, yielding $\theta = \theta_1 < \infty$ under the parametrization $\theta = \sqrt{-\ln(1 - h^2/c)}$. Hence, (5.4) follows.

Formulas (5.5) and (5.6) can be established using conceptually similar steps. To save space, below we only establish formula (5.6). The following technical lemma is proven in the Supplementary Appendix [Onatski, Moreira and Hallin (2013)].

LEMMA 13. *Let* CLR *be the "corrected" likelihood ratio statistic as defined in Example 3. Then, under the null, as $p, n \to_c \infty$, the vector* $(\mathrm{CLR}, \Delta_p(z_0(h)))$ *converges in distribution to a Gaussian vector* (ζ_1, ζ_2) *with* $\mathrm{Cov}(\zeta_1, \zeta_2) = -2h + 2\ln(1+h)$.

Lemma 13 and (4.2) imply the convergence in distribution of the vector $(\mathrm{CLR}, \ln L(h; \lambda))$ to a Gaussian vector $(\zeta_1, -\frac{1}{2}\zeta_2)$. From Bai et al. (2009), we know that, under the null, $\mathrm{CLR} \xrightarrow{d} N(-\frac{1}{2}\ln(1-c), -2\ln(1-c) - 2c)$. By Le

Cam's third lemma, under the alternative $h = h_1$, CLR converges to a Gaussian random variable with the same variance but with mean equal to $-\frac{1}{2}\ln(1 - c) + \mathrm{Cov}(\zeta_1, -\frac{1}{2}\zeta_2) = -\frac{1}{2}\ln(1-c) + h - \ln(1+h)$ evaluated at $h = h_1$. Therefore, the power of the "corrected" likelihood ratio test of asymptotic size α equals $1 - \Phi(\Phi^{-1}(1-\alpha) - \frac{h_1 - \ln(1+h_1)}{\sqrt{-2\ln(1-c)-2c}})$. Using the reparametrization $\theta_1 = \sqrt{-\ln(1 - h_1^2/c)}$, we get (5.6).

Acknowledgments. This work started when the first two authors worked at and the third author visited Columbia University. We would like to thank Tony Cai, the Associate Editor, Nick Patterson and an anonymous referee for helpful and encouraging comments.

SUPPLEMENTARY MATERIAL

Supplementary Appendix (DOI: 10.1214/13-AOS1100SUPP; .pdf). The Supplementary Appendix contains proofs of Lemmas 4, 5, 6, 11, 12 and 13.

REFERENCES

BAI, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.* **21** 649–672. MR1217560

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

BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575

BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

BEJAN, A. I. (2005). Largest eigenvalues and sample covariance matrices. Tracy–Widom and Painleve II; computational aspects and realization in S-plus with applications. Unpublished manuscript, Univ. Warwick.

BERTHET, Q. and RIGOLLET, P. (2012). Optimal detection of sparse principal components in high dimension. Available at arXiv:1202.5070.

BIANCHI, P., DEBBAH, M., MAIDA, M. and NAJIM, J. (2011). Performance of statistical tests for single source detection using random matrix theory. *IEEE Trans. Inform. Theory* **57** 2400–2419.

BIRKE, M. and DETTE, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. MR2189467

BLOEMENDAL, A. and VIRÁG, B. (2012). Limits of spiked random matrices I. *Probab. Theory Related Fields*. To appear. DOI:10.1007/s00440-012-0443-2. Available at arXiv:1011.1877v2.

BUTLER, R. W. and WOOD, A. T. A. (2002). Laplace approximations for hypergeometric functions with matrix argument. *Ann. Statist.* **30** 1155–1177. MR1926172

CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

DICKEY, J. M. (1983). Multiple hypergeometric functions: Probabilistic interpretations and statistical uses. *J. Amer. Statist. Assoc.* **78** 628–637. MR0721212

EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592

ERDELYI, A. (1937). Beitrag zur theorie der konfluenten hypergeometrischen funktionen von mehreren veranderlichen. *Sitzungsberichte*, *Akademie der Wissenschaften in Wien*, *Abteilung IIa*, *Mathematisch-Naturwissenschaftliche Klasse* **146** 431–467.

FÉRAL, D. and PÉCHÉ, S. (2009). The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case. *J. Math. Phys.* **50** 073302, 33. MR2548630

FISHER, T. J., SUN, X. and GALLAGHER, C. M. (2010). A new test for sphericity of the covariance matrix for high dimensional data. *J. Multivariate Anal.* **101** 2554–2570. MR2719881

GUIONNET, A. and MAÏDA, M. (2005). A Fourier view on the R-transform and related asymptotics of spherical integrals. *J. Funct. Anal.* **222** 435–490. MR2132396

HALLIN, M. and PAINDAVEINE, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. MR2329465

HILLIER, G. (2001). The density of a quadratic form in a vector uniformly distributed on the n-sphere. *Econometric Theory* **17** 1–28. MR1863565

HOYLE, D. C. (2008). Automatic PCA dimension selection for high dimensional data and small sample sizes. *J. Mach. Learn. Res.* **9** 2733–2759.

JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475–501. MR0181057

JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568

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

KRITCHMAN, S. and NADLER, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems* **94** 19–32.

KRITCHMAN, S. and NADLER, B. (2009). Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Trans. Signal Process.* **57** 3930–3941. MR2683143

LE CAM, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.* **3** 37–98. MR0126903

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

LIJOI, A. and REGAZZINI, E. (2004). Means of a Dirichlet process and multiple hypergeometric functions. *Ann. Probab.* **32** 1469–1495. MR2060305

MAUCHLY, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Statist.* **11** 204–209. MR0002084

MO, M. Y. (2012). Rank 1 real Wishart spiked model. *Comm. Pure Appl. Math.* **65** 1528–1638. MR2969495

NADAKUDITI, R. R. and EDELMAN, A. (2008). Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Trans. Signal Process.* **56** 2625–2638. MR1500236

NADAKUDITI, R. R. and SILVERSTEIN, J. W. (2010). Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE Journal of Selected Topics in Signal Processing* **4** 468–480.

NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36** 2791–2817. MR2485013

OLVER, F. W. J. (1997). *Asymptotics and Special Functions*. AK Peters, Wellesley, MA. MR1429619

ONATSKI, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070

ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econom. Statist.* **92** 1004–1016.

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Supplement to "Asymptotic power of sphericity tests for high-dimensional data." DOI:10.1214/13-AOS1100SUPP.

PATTERSON, N., PRICE, A. L. and REICH, D. (2006). Population structure and eigenanalysis. *PLoS Genetics* **2** 2074–2093.

PERRY, P. O. and WOLFE, P. J. (2010). Minimax rank estimation for subspace tracking. *IEEE Journal of Selected Topics in Signal Processing* **4** 504–513.

SCHOTT, J. R. (2006). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *J. Multivariate Anal.* **97** 827–843. MR2256563

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

SRIVASTAVA, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.* **35** 251–272. MR2328427

SRIVASTAVA, H. M. and KARLSSON, P. W. (1985). *Multiple Gaussian Hypergeometric Series*. Ellis Horwood, Chichester. MR0834385

SUGIURA, N. (1972). Locally best invariant test for sphericity and the limiting distributions. *Ann. Math. Statist.* **43** 1312–1316. MR0311032

UHLIG, H. (1994). On singular Wishart and singular multivariate beta distributions. *Ann. Statist.* **22** 395–405. MR1272090

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

WANG, D. (2012). The largest eigenvalue of real symmetric, Hermitian and Hermitian self-dual random matrix models with rank one external source, Part I. *J. Stat. Phys.* **146** 719–761. MR2916094

A. ONATSKI
FACULTY OF ECONOMICS
UNIVERSITY OF CAMBRIDGE
SIDGWICK AVENUE
CAMBRIDGE, CB3 9DD
UNITED KINGDOM
E-MAIL: ao319@cam.ac.uk

M. J. MOREIRA
ESCOLA DE PÓS-GRADUAÇÃO EM ECONOMIA
FUNAÇÃO GETULIO VARGAS (FGV/EPGE)
PRAIA DE BOTAFOGO, 190-SALA 1100
RIO DE JANEIRO-RJ 22250-900
BRAZIL
E-MAIL: mjmoreira@fgv.br

M. HALLIN
ECARES
UNIVERSITÉ LIBRE DE BRUXELLES CP 114/04
50, AVENUE F.D. ROOSEVELT
B-1050 BRUXELLES
BELGIUM
E-MAIL: mhallin@ulb.ac.be