#12-98

Hypothesis Testing When the Sample Covariance Matrix Is Singular

May 1998

Olivier Ledoit
Anderson Graduate School of Management
University of California, Los Angeles

Finance Working Paper
sponsored by

CMS CAPITAL MANAGEMENT SCIENCES

Hypothesis Testing When the Sample Covariance Matrix Is Singular

Olivier Ledoit*

May 14, 1998

Abstract

This paper analyzes whether standard covariance matrix tests work when dimensionality is large, and in particular larger than sample size. In the latter case, the singularity of the sample covariance matrix makes likelihood ratio tests degenerate, but other tests based on quadratic forms of sample covariance matrix eigenvalues remain well-defined. We study the consistency property and limiting distribution of these tests as dimensionality and sample size go to infinity together, with their ratio converging to a finite nonzero limit. We find that the existing test for sphericity is robust against high dimensionality, but not the test for equality of the covariance matrix to a given matrix. For the latter test, we develop a new correction to the existing criterion that makes it robust against high dimensionality.

*Anderson Graduate School of Management at UCLA, Box 90095-1481, 110 Westwood Plaza, Los Angeles CA 90095-1481; phone: (310) 206 6562; fax: (310) 206 5455; e-mail: ledoit@ucla.edu; Web site: http://www.anderson.ucla.edu/faculty/olivier.ledoit/. I thank Michael Wolf for very helpful comments and T. W. Anderson for encouragement. Errors are my own.

1 Introduction

Many empirical problems involve large-dimensional covariance matrices. Sometimes the dimensionality p is even larger than the sample size n, which makes the sample covariance matrix S singular (cf. Uhlig (1994)). How to conduct statistical inference in this case? For concreteness we focus on two common tests in this paper: 1) the covariance matrix Σ is proportional to the identity I (sphericity); 2) $\Sigma = I$. The identity can be replaced with any other matrix Σ_0 by multiplying the data by $\Sigma_0^{-1/2}$. Following much of the literature, we assume normality. For both hypotheses the likelihood ratio test criterion is degenerate when p exceeds n (see, e.g., Muirhead (1982) or Anderson (1984)). This steers us towards other test criteria that do not degenerate, such as

$$U = \frac{1}{p}\mathrm{tr}\left[\left(\frac{S}{\frac{1}{p}\mathrm{tr}(S)} - I\right)^2\right] \quad \text{and} \quad V = \frac{1}{p}\mathrm{tr}\left[(S-I)^2\right] \tag{1}$$

where tr denotes the trace. John (1971) proves that U is the locally most powerful invariant test for sphericity, and Nagao (1973) derives V as the equivalent of U for the test of $\Sigma = I$. The asymptotic framework where U and V have been studied assumes that n goes to infinity while p remains fixed. It treats terms of order p/n like terms of order $1/n$, which is inappropriate if p is of the same order of magnitude as n. The robustness of tests based on U and V against high dimensionality is heretofore unknown.

We study the asymptotic behavior of U and V as p and n go to infinity together with the ratio p/n converging to a limit $c \in (0, +\infty)$ called the concentration.[1] The singular case corresponds to a concentration above one. The robustness issue boils down to power and size: Is the test still consistent? Is the n-limiting distribution under the null still a good approximation? Surprisingly, we find opposite answers for U and V. The power and the size of the sphericity test based on U turn out to be robust against p large, and even larger than n. But the test of $\Sigma = I$ based on V is not consistent against every alternative when p goes to infinity with n, and its n-limiting distribution differs from its (n, p)-limiting distribution under the null. This prompts

[1]This framework is rooted in the spectral theory of large-dimensional random matrices, which studies the distribution of the eigenvalues of random matrices as dimensionality goes to infinity. See Silverstein (1986) for an excellent survey. The connection with standard statistical problems of estimation and testing has not been exploited before, except for some preliminary attempts by Wachter (1976) and Silverstein and Combettes (1992) that are limited to graphical aspects.

us to introduce the modified criterion

$$W = \frac{1}{p}\mathrm{tr}\left[(S-I)^2\right] - \frac{p}{n}\left[\frac{1}{p}\mathrm{tr}(S)\right]^2 + \frac{p}{n}. \tag{2}$$

W has the same n-asymptotic properties as V: it is n-consistent and has the same n-limiting distribution as V under the null. We show that, contrary to V, the power and the size of the test based on W are robust against p large, and even larger than n.

In summary, the contributions of this paper are: (i) developing a method to check the robustness of covariance matrix tests against high dimensionality; and (ii) finding two criteria (one old and one new) for commonly used covariance matrix tests that can be used when the sample covariance matrix is singular. Directions for future research include: applying the method to other test criteria; finding limiting distributions under the alternative to compute power; searching for most powerful tests; relaxing the normality assumption.

Section 2 compiles preliminary results. Section 3 shows that the test statistic U for sphericity is robust against large dimensionality. Section 4 shows that the test of $\Sigma = I$ based on V is not. Section 5 introduces a new criterion W that can be used when p is large. Section 6 reports evidence from Monte-Carlo simulations. Proofs are in the Appendix.

2 Preliminaries

The exact sense in which sample size and dimensionality go to infinity together is defined by the following assumptions.

Assumption 1 (Asymptotics) *Dimensionality and sample size are two increasing integer functions $p = p_k$ and $n = n_k$ of an index $k = 1, 2, \ldots$ such that $\lim_{k\to\infty} p_k = +\infty$, $\lim_{k\to\infty} n_k = +\infty$ and there exists $c \in (0, +\infty)$ such that $\lim_{k\to\infty} p_k/n_k = c$.*

The case where the sample covariance matrix is singular corresponds to a concentration c higher than one.

Assumption 2 (Data-Generating Process) *For each positive integer k, X_k is an $(n_k + 1) \times p_k$ matrix of $n_k + 1$ i.i.d. observations on a system of p_k random variables that are jointly*

normally distributed with mean vector μ_k and covariance matrix Σ_k. Let $\lambda_{1,k}, \ldots, \lambda_{p_k,k}$ denote the eigenvalues of the covariance matrix Σ_k. We suppose that their cross-sectional average $\alpha = \sum_{i=1}^{p_k} \lambda_{i,k}/p_k$ and their cross-sectional dispersion $\delta^2 = \sum_{i=1}^{p_k}(\lambda_{i,k} - \alpha)^2/p_k$ are independent of the index k. Furthermore we require $\alpha > 0$.

S_k is the unbiased sample covariance matrix with entries $s_{ij,k} = \frac{1}{n}\sum_{l=1}^{n+1}(x_{il,k} - m_{i,k})(x_{jl,k} - m_{j,k})$ where $m_{i,k} = \frac{1}{n+1}\sum_{l=1}^{n+1} x_{il,k}$. The assumption that α and δ^2 do not vary with k is not restrictive in practice since the tests would only be applied to one particular covariance matrix with one particular value of α and of δ^2. The null hypothesis of sphericity can be stated as $\delta^2 = 0$, and the null $\Sigma = I$ can be stated as $\delta^2 = 0$ and $\alpha = 1$. We need one more assumption to obtain convergence results under the alternative.

Assumption 3 (Higher Moments) *The cross-sectional third and fourth moment of the eigenvalues of the population covariance matrix $\sum_{i=1}^{p_k}(\lambda_{i,k})^j/p_k$ $(j = 3, 4)$ converge to finite limits, respectively.*

Dependence on k will be omitted when no ambiguity is possible. Much of the mathematical groundwork has already been laid out by research in the spectral theory of large-dimensional random matrices. The fundamental results of interest to us are as follows.

Proposition 1 (Law of Large Numbers) *Under Assumptions 1-3,*

$$\frac{1}{p}\mathrm{tr}(S) \xrightarrow{\mathrm{P}} \alpha \tag{3}$$

$$\frac{1}{p}\mathrm{tr}(S^2) \xrightarrow{\mathrm{P}} (1+c)\alpha^2 + \delta^2 \tag{4}$$

where $\xrightarrow{\mathrm{P}}$ denotes convergence in probability.

All proofs are in the Appendix. This Law of Large Numbers will help us establish whether or not a given test is consistent against every alternative as n and p go to infinity together. The distribution of the test statistic under the null will be found by using this Central Limit Theorem.

Proposition 2 (Central Limit Theorem) *Under Assumptions 1-2, if $\delta^2 = 0$ then*

$$n \times \begin{bmatrix} \frac{1}{p}\mathrm{tr}(S) - \alpha \\ \frac{1}{p}\mathrm{tr}(S^2) - \frac{n+p+1}{n}\alpha^2 \end{bmatrix} \xrightarrow{\mathrm{D}} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2\alpha^2/c & 4\left(1+\frac{1}{c}\right)\alpha^3 \\ 4\left(1+\frac{1}{c}\right)\alpha^3 & 4\left(\frac{2}{c}+5+2c\right)\alpha^4 \end{bmatrix} \right) \tag{5}$$

where $\xrightarrow{\mathrm{D}}$ *denotes convergence in distribution and* $\mathcal{N}$ *the normal distribution.*

3 Sphericity Test

It is well-known that the sphericity test based on U is n-consistent. As for (n,p)-consistency, Proposition 1 implies that, under Assumptions 1-3,

$$U = \frac{\frac{1}{p}\mathrm{tr}(S^2)}{\left[\frac{1}{p}\mathrm{tr}(S)\right]^2} - 1 \xrightarrow{\mathrm{P}} \frac{(1+c)\alpha^2 + \delta^2}{\alpha^2} - 1 = c + \frac{\delta^2}{\alpha^2}. \tag{6}$$

Since c can be approximated by the known quantity p/n, the power of this test to separate the null hypothesis of sphericity $\delta^2/\alpha^2 = 0$ from the alternative $\delta^2/\alpha^2 > 0$ converges to one as n and p go to infinity together: this constitutes an (n,p)-consistent test.

John (1972) shows that, as n goes to infinity while p remains fixed, the limiting distribution of U under the null is given by

$$\frac{np}{2}U \xrightarrow{D} Y_{\frac{1}{2}p(p+1)-1} \qquad \text{or, equivalently,} \tag{7}$$

$$nU - p \xrightarrow{\mathrm{D}} \frac{2}{p}Y_{\frac{1}{2}p(p+1)-1} - p \tag{8}$$

where Y_d denotes a random variable distributed as a χ^2 with d degrees of freedom. It will become apparent after Proposition 4 why we choose to rewrite Equation (7) as Equation (8). This approximation may or may not remain accurate under (n,p)-asymptotics, depending on whether it omits terms of order p/n. To find out, let us start by deriving the (n,p)-limiting distribution of U under the null hypothesis $\delta^2/\alpha^2 = 0$.

Proposition 3 *Under the assumptions of Proposition 2,*

$$nU - p \xrightarrow{\mathrm{D}} \mathcal{N}(1,4). \tag{9}$$

Now we can compare Equations (8) and (9).

Proposition 4 *Suppose that, for every k, the random variable $Y_{\frac{1}{2}p_k(p_k+1)+a}$ is distributed as a χ^2 with $\frac{1}{2}p_k(p_k+1)+a$ degrees of freedom, where a is a constant integer. Then its limiting distribution under Assumption 1 verifies*

$$\frac{2}{p_k}Y_{\frac{1}{2}p_k(p_k+1)+a} - p_k \xrightarrow{\mathrm{D}} \mathcal{N}(1,4). \tag{10}$$

Using Proposition 4 with $a=-1$ shows that the n-limiting distribution given by Equation (8) is still correct under (n,p)-asymptotics.

The conclusion of our analysis of the sphericity test based on U is the following: the existing n-asymptotic theory (where p is fixed) remains valid if p goes to infinity with n, even for the case $p>n$.

4 Test that a Covariance Matrix Is the Identity

As n goes to infinity with p fixed, $S \xrightarrow{P} \Sigma$, therefore $V \xrightarrow{P} \frac{1}{p}\mathrm{tr}\left[(\Sigma-I)^2\right]$. This shows that the test of $\Sigma = I$ based on V is n-consistent. As for (n,p)-consistency, Proposition 1 implies that, under Assumptions 1-3,

$$V = \frac{1}{p}\mathrm{tr}(S^2) - \frac{2}{p}\mathrm{tr}(S) + 1 \xrightarrow{\mathrm{P}} (1+c)\alpha^2 + \delta^2 - 2\alpha + 1 = c\alpha^2 + (\alpha-1)^2 + \delta^2. \tag{11}$$

Since $\frac{1}{p}\mathrm{tr}[(\Sigma-I)^2] = (\alpha-1)^2+\delta^2$ is a squared measure of distance between the population covariance matrix and the identity, the null hypothesis can be rewritten as $(\alpha-1)^2+\delta^2=0$, and the alternative as $(\alpha-1)^2+\delta^2>0$. The problem is that the probability limit of the test criterion V is not directly a function of $(\alpha-1)^2+\delta^2$: it involves another term, $c\alpha^2$, which contains the nuisance parameter α^2. Therefore the test based on V may sometimes be powerless to separate the null from the alternative. More specifically, when the triplet (c,α,δ) verifies

$$c\alpha^2 + (\alpha-1)^2 + \delta^2 = c, \tag{12}$$

the test criterion V has the same probability limit under the null as under the alternative. The clearest counter-examples are those where $\delta^2 = 0$, because Proposition 2 allows us to compute the limit of the power of the test against such alternatives. When $\delta^2 = 0$ the solution to Equation (12) is $\alpha = \frac{1-c}{1+c}$.

Proposition 5 *Under Assumptions 1-2, if $c \in (0,1)$ and there exists a finite d such that $\frac{p}{n} = c + \frac{d}{n} + o\left(\frac{1}{n}\right)$ then the power of the one-sided test of any positive significance level based on V to reject the null $\Sigma = I$ when the alternative $\Sigma = \frac{1-c}{1+c} I$ is true converges to a limit strictly below one.*

We see that the n-consistency of the test based on V does not extend to (n,p)-asymptotics.

Nagao (1973) shows that, as n goes to infinity while p remains fixed, the limiting distribution of V under the null is given by

$$\frac{np}{2} V \xrightarrow{D} Y_{\frac{1}{2}p(p+1)} \qquad \text{or, equivalently,} \tag{13}$$

$$nV - p \xrightarrow{\mathrm{D}} \frac{2}{p} Y_{\frac{1}{2}p(p+1)} - p \tag{14}$$

where, as before, Y_d denotes a random variable distributed as a χ^2 with d degrees of freedom. It is not immediately apparent whether this approximation remains accurate under (n,p)-asymptotics. The (n,p)-limiting distribution of V under the null hypothesis $(\alpha - 1)^2 + \delta^2 = 0$ is derived in Equation (41) in the Appendix as part of the proof of Proposition 5:

$$nV - p \xrightarrow{\mathrm{D}} \mathcal{N}(1, 4 + 8c). \tag{15}$$

Using Proposition 4 with $a = 0$ shows that the n-limiting distribution given by Equation (14) is incorrect under (n,p)-asymptotics.

The conclusion of our analysis of the test of $\Sigma = I$ based on V is the following: the existing n-asymptotic theory (where p is fixed) breaks down when p goes to infinity with n, including the case $p > n$.

5 Test that a Covariance Matrix Is the Identity: New Criterion

The ideal would be to find a simple modification of V that would have the same n-asymptotic properties and better (n,p)-asymptotic properties (in the spirit of U). This is why we introduce the new criterion

$$W = \frac{1}{p}\mathrm{tr}\left[(S-I)^2\right] - \frac{p}{n}\left[\frac{1}{p}\mathrm{tr}(S)\right]^2 + \frac{p}{n}. \tag{16}$$

As n goes to infinity with p fixed, $W \xrightarrow{P} \frac{1}{p}\mathrm{tr}\left[(\Sigma - I)^2\right]$, therefore the test of $\Sigma = I$ based on W is n-consistent. As for (n,p)-consistency, Proposition 1 implies that, under Assumptions 1-3,

$$W \xrightarrow{\mathrm{P}} c\alpha^2 + (\alpha-1)^2 + \delta^2 - c\alpha^2 + c = c + (\alpha-1)^2 + \delta^2. \tag{17}$$

Since c can be approximated by the known quantity p/n, the power of the test based on W to separate the null hypothesis $(\alpha-1)^2 + \delta^2 = 0$ from the alternative $(\alpha-1)^2 + \delta^2 > 0$ converges to one as n and p go to infinity together: the test based on W is (n,p)-consistent.

The following proposition shows that W has the same n-limiting distribution as V under the null.

Proposition 6 *As n goes to infinity with p fixed, the limiting distribution of W under the null hypothesis $(\alpha-1)^2 + \delta^2 = 0$ is the same as for V:*

$$\frac{np}{2}W \xrightarrow{D} Y_{\frac{1}{2}p(p+1)} \qquad \textit{or, equivalently,} \tag{18}$$

$$nW - p \xrightarrow{\mathrm{D}} \frac{2}{p}Y_{\frac{1}{2}p(p+1)} - p \tag{19}$$

where Y_d denotes a random variable distributed as a χ^2 with d degrees of freedom.

To find out whether this approximation remains accurate under (n,p)-asymptotics, we derive the (n,p)-limiting distribution of W under the null.

Proposition 7 *Under Assumptions 1-2, if $(\alpha-1)^2 + \delta^2 = 0$ then*

$$nW - p \xrightarrow{\mathrm{D}} \mathcal{N}(1,4). \tag{20}$$

Using Proposition 4 with $a = 0$ shows that the n-limiting distribution given by Equation (19) is still correct under (n, p)-asymptotics.

The conclusion of our analysis of the test of $\Sigma = I$ based on W is the following: the n-asymptotic theory developed for V is directly applicable to W, and it remains valid (for W but not V) if p goes to infinity with n, even in the case $p > n$.

6 Monte-Carlo Simulations

Monte-Carlo simulations are used to find the size and power of the test statistics U, V and W for $p, n = 4, 8, \ldots, 256$. In each case we run $10,000$ simulations. The alternative against which power is computed has to be "scalable" in the sense that it can be represented by population covariance matrices of any dimension $p = 4, 8, \ldots, 256$. The simplest alternative we can think of is to set half of the population eigenvalues equal to one, and the other ones equal to a half.

Table 1 reports the size of the sphericity test based on U. One-sided 95% confidence interval are constructed from the χ^2 n-limiting distribution in Equation (8). We see that the quality of this approximation does not get worse when p gets large: it can be relied upon even when $p > n$. This is what we expected given Proposition 4.

Table 2 shows the power of the sphericity test based on U against the alternative described above. We see that the power does not become lower when p gets large: power stays high even when $p > n$. This confirms the (n, p)-consistency result derived from Equation (6). The table indicates that the power seems to depend predominantly on n. For fixed sample size, the power of the test is often increasing in p, which is somewhat unexpected.

Using the same methodology as in Table 1, we report in Table 3 the size of the test for $\Sigma = I$ based on V. We see that the χ^2 n-limiting distribution under the null in Equation (14) is a poor approximation for large p. This is what we expected given the discussion surrounding Equation (15).

Using the same methodology as in Table 2, we report in Table 4 the power of the test based on V against the alternative described above. Given the discussion surrounding Equation (12), we anticipate that this test will not be powerful when $c = [(\alpha - 1)^2 + \delta^2]/(1 - \alpha^2) = 2/7$. Indeed

we observe that, in the cells where p/n exceeds the critical value $2/7$, this test does not have much power to reject the alternative.

Using the same methodology as in Table 1, we report in Table 5 the size of the test for $\Sigma = I$ based on W. We see that the χ^2 approximation in Equation (19) for the null distribution does not get worse when p gets large: it can be relied upon even when $p > n$. This is what we expected given the discussion surrounding Equation (15).

Using the same methodology as in Table 2, we report in Table 6 the power of the test based on W against the alternative described above. We see that the power does not become lower when p gets large: power stays high even when $p > n$. This confirms the (n, p)-consistency result derived from Equation (17). As with U, the table indicates that the power seems to depend predominantly on n, and to be increasing in p for fixed n.

Overall, these Monte-Carlo simulations confirm the finite-sample relevance of the asymptotic results obtained in Sections 3, 4 and 5.

p n	4	8	16	32	64	128	256
4	0.01	0.03	0.04	0.05	0.05	0.05	0.05
8	0.03	0.04	0.04	0.05	0.05	0.05	0.05
16	0.04	0.05	0.05	0.05	0.05	0.05	0.05
32	0.05	0.05	0.05	0.05	0.05	0.05	0.05
64	0.05	0.05	0.05	0.05	0.05	0.05	0.05
128	0.05	0.05	0.05	0.05	0.05	0.05	0.05
256	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Table 1: Size of Sphericity Test Based on U. One-sided 95% confidence intervals are constructed using the χ^2 approximation. Actual size converges to nominal size as dimensionality p goes to infinity with sample size n. Results come from $10,000$ Monte-Carlo Simulations.

p n	4	8	16	32	64	128	256
4	0.02	0.06	0.15	0.37	0.76	0.98	1.00
8	0.05	0.09	0.18	0.42	0.85	1.00	1.00
16	0.06	0.11	0.20	0.48	0.90	1.00	1.00
32	0.08	0.13	0.22	0.50	0.93	1.00	1.00
64	0.09	0.13	0.24	0.52	0.95	1.00	1.00
128	0.09	0.14	0.23	0.53	0.95	1.00	1.00
256	0.09	0.14	0.24	0.54	0.96	1.00	1.00

Table 2: Power of Sphericity Test Based on U. Null hypothesis is rejected when test statistic exceeds the one-sided 95% cutoff point obtained from the χ^2 approximation. Data are generated under the alternative where half of the population eigenvalues are equal to one, and the other ones are equal to a half. Power converges to one as dimensionality p goes to infinity with sample size n. Results come from $10,000$ Monte-Carlo Simulations.

p n	4	8	16	32	64	128	256
4	0.12	0.10	0.08	0.07	0.06	0.05	0.05
8	0.18	0.15	0.11	0.09	0.08	0.07	0.06
16	0.25	0.21	0.15	0.12	0.09	0.07	0.06
32	0.31	0.27	0.21	0.17	0.13	0.09	0.07
64	0.35	0.33	0.29	0.22	0.17	0.13	0.09
128	0.40	0.38	0.34	0.29	0.23	0.17	0.12
256	0.43	0.41	0.38	0.34	0.28	0.22	0.17

Table 3: Size of Equality Test Based on V. One-sided 95% confidence intervals are constructed using the χ^2 approximation. Actual size does not converge to nominal size as dimensionality p goes to infinity with sample size n. Results come from $10,000$ Monte-Carlo Simulations.

p n	4	8	16	32	64	128	256
4	0.04	0.03	0.03	0.11	0.76	1.00	1.00
8	0.03	0.02	0.02	0.03	0.56	1.00	1.00
16	0.02	0.01	0.00	0.00	0.05	1.00	1.00
32	0.00	0.00	0.00	0.00	0.00	0.14	1.00
64	0.00	0.00	0.00	0.00	0.00	0.00	0.56
128	0.00	0.00	0.00	0.00	0.00	0.00	0.00
256	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4: Power of Equality Test Based on U. Null hypothesis is rejected when test statistic exceeds the one-sided 95% cutoff point obtained from the χ^2 approximation. Data are generated under the alternative where half of the population eigenvalues are equal to one, and the other ones are equal to a half. Power does not converge to one as dimensionality p goes to infinity with sample size n. Results come from $10,000$ Monte-Carlo Simulations.

p n	4	8	16	32	64	128	256
4	0.03	0.04	0.05	0.05	0.05	0.05	0.05
8	0.04	0.05	0.05	0.05	0.05	0.05	0.05
16	0.05	0.05	0.05	0.05	0.05	0.05	0.05
32	0.05	0.05	0.05	0.05	0.05	0.05	0.05
64	0.05	0.05	0.06	0.05	0.05	0.05	0.05
128	0.06	0.06	0.06	0.05	0.05	0.05	0.05
256	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Table 5: Size of Equality Test Based on W. One-sided 95% confidence intervals are constructed using the χ^2 approximation. Actual size converges to nominal size as dimensionality p goes to infinity with sample size n. Results come from $10,000$ Monte-Carlo Simulations.

p n	4	8	16	32	64	128	256
4	0.02	0.02	0.06	0.37	0.93	1.00	1.00
8	0.02	0.03	0.07	0.43	0.98	1.00	1.00
16	0.02	0.03	0.08	0.51	1.00	1.00	1.00
32	0.02	0.03	0.09	0.53	1.00	1.00	1.00
64	0.02	0.03	0.09	0.56	1.00	1.00	1.00
128	0.02	0.03	0.08	0.57	1.00	1.00	1.00
256	0.02	0.03	0.09	0.58	1.00	1.00	1.00

Table 6: Power of Equality Test Based on W. Null hypothesis is rejected when test statistic exceeds the one-sided 95% cutoff point obtained from the χ^2 approximation. Data are generated under the alternative where half of the population eigenvalues are equal to one, and the other eigenvalues are equal to a half. Power converges to one as dimensionality p goes to infinity with sample size n. Results come from $10,000$ Monte-Carlo Simulations.

Appendix

Proof of Proposition 1 The proof of this proposition is contained inside the proof of the main theorem of Yin and Krishnaiah (1983). Even though their main theorem is derived under assumptions on *all* the cross-sectional moments of the eigenvalues of the population covariance matrix, careful inspection of their proof reveals that convergence in probability of the first two moments requires only assumptions up to the fourth moment. The formulas for the limits come from Yin and Krishnaiah's second equation on the top of page 504. □

Proof of Proposition 2 Changing α simply amounts to rescaling $\frac{1}{p}\mathrm{tr}(S)$ by α and $\frac{1}{p}\mathrm{tr}(S^2)$ by α^2, therefore we can assume without loss of generality that $\alpha = 1$. Jonsson's (1982) Theorem 4.1 shows that, under the assumptions of Proposition 2,

$$\begin{bmatrix} \dfrac{n}{n+p}\Big\{\mathrm{tr}(S) - \mathrm{E}[\mathrm{tr}(S)]\Big\} \\ \dfrac{n^2}{(n+p)^2}\Big\{\mathrm{tr}(S^2) - \mathrm{E}[\mathrm{tr}(S^2)]\Big\} \end{bmatrix} \tag{21}$$

converges in distribution to a bivariate normal. Since $p/n \to c \in (0, +\infty)$, this implies that

$$n \times \begin{bmatrix} \frac{1}{p}\mathrm{tr}(S) - \mathrm{E}\left[\frac{1}{p}\mathrm{tr}(S)\right] \\ \frac{1}{p}\mathrm{tr}(S^2) - \mathrm{E}\left[\frac{1}{p}\mathrm{tr}(S^2)\right] \end{bmatrix} \tag{22}$$

also converges in distribution to a bivariate normal. $\frac{1}{p}\mathrm{tr}(S)$ is the cross-sectional average of the diagonal elements of the unbiased sample covariance matrix, therefore its expectation is equal to one. John (1972, Lemma 2) shows that the expectation of $\frac{1}{p}\mathrm{tr}(S)$ is equal to $\frac{n+p+1}{n}$. So far we have established that

$$n \times \begin{bmatrix} \frac{1}{p}\mathrm{tr}(S) - 1 \\ \frac{1}{p}\mathrm{tr}(S^2) - \frac{n+p+1}{n} \end{bmatrix} \tag{23}$$

converges in distribution to a bivariate normal. Since this limiting bivariate normal has mean zero, the only task left is to compute its covariance matrix. This can be done by taking the limit of the covariance matrix of the expression in Equation (23). Using once again the moments

computed by John (1972, Lemma 2), we find that

$$\begin{aligned}
\mathrm{Var}\left[\frac{n}{p}\mathrm{tr}(S)\right] &= \mathrm{E}\left[\left(\frac{n}{p}\mathrm{tr}(S)\right)^2\right] - \left(\mathrm{E}\left[\frac{n}{p}\mathrm{tr}(S)\right]\right)^2 = \frac{n(np+2)}{p} - n^2 = 2\frac{n}{p} \longrightarrow \frac{2}{c} \\
\mathrm{Var}\left[\frac{n}{p}\mathrm{tr}(S^2)\right] &= \mathrm{E}\left[\left(\frac{n}{p}\mathrm{tr}(S^2)\right)^2\right] - \left(\mathrm{E}\left[\frac{n}{p}\mathrm{tr}(S^2)\right]\right)^2 \\
&= \frac{pn^3 + (2p^2+2p+8)n^2 + (p^3+2p^2+21p+20)n + 8p^2 + 20p + 20}{pn} \\
&\quad - (n+p+1)^2 \\
&= \frac{8n}{p} + \frac{20p^2+20p}{p^2} + \frac{8p^3+20p^2+20p}{p^2 n} \\
&\longrightarrow \frac{8}{c} + 20 + 8c.
\end{aligned}$$

Finally we have to find the covariance term. Let s_{ij} denote the entry (i,j) of the unbiased sample covariance matrix S. We have:

$$\begin{aligned}
\mathrm{E}[\mathrm{tr}(S)\mathrm{tr}(S^2)] &= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} \mathrm{E}[s_{ii}s_{jl}^2] \\
&= p(p-1)(p-2)\mathrm{E}[s_{11}s_{23}^2] + p(p-1)\mathrm{E}[s_{11}s_{22}^2] + 2p(p-1)\mathrm{E}[s_{11}s_{12}^2] + p\mathrm{E}[s_{11}^3] \\
&= \frac{p(p-1)(p-2)}{n} + p(p-1)\frac{n+2}{n} + 2p(p-1)\frac{n+2}{n^2} + p\frac{(n+2)(n+4)}{n^2} \qquad (24) \\
&= p^2 + \frac{p^3+p^2+4p}{n} + \frac{4p^2+4p}{n^2}.
\end{aligned}$$

The moment formulas that appear in Equation (24) are computed in the same fashion as in the proof of Lemma 2 by John (1972). This enables us to compute the limiting covariance term as

$$\begin{aligned}
\mathrm{Cov}\left[\frac{n}{p}\mathrm{tr}(S), \frac{n}{p}\mathrm{tr}(S^2)\right] &= \frac{n^2}{p^2}\mathrm{E}[\mathrm{tr}(S)\mathrm{tr}(S^2)] - \mathrm{E}\left[\frac{n}{p}\mathrm{tr}(S)\right] \times \mathrm{E}\left[\frac{n}{p}\mathrm{tr}(S^2)\right] \qquad (25) \\
&= n^2 + n\frac{p^2+p+4}{p} + \frac{4p+4}{p} - n(n+p+1) \qquad (26) \\
&= 4\frac{n}{p} + 4 + \frac{4}{p} \qquad (27) \\
&\longrightarrow 4\left(1+\frac{1}{c}\right). \qquad (28)
\end{aligned}$$

This completes the proof of Proposition 2. □

Proof of Proposition 3 Define the function $f(x,y) = \frac{y}{x^2} - 1$. Then $U = f\left(\frac{1}{p}\mathrm{tr}(S), \frac{1}{p}\mathrm{tr}(S^2)\right)$. Proposition 2 implies that, by the delta method,

$$n\left[U - f\left(\alpha, \frac{n+p+1}{n}\alpha^2\right)\right] \xrightarrow{\mathrm{D}} \mathcal{N}(0, \lim A), \text{ where}$$

$$A = \begin{bmatrix} \frac{\partial f}{\partial x}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) \\ \frac{\partial f}{\partial y}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) \end{bmatrix}^{\top} \begin{bmatrix} 2\alpha^2/c & 4\left(1+\frac{1}{c}\right)\alpha^3 \\ 4\left(1+\frac{1}{c}\right)\alpha^3 & 4\left(\frac{2}{c}+5+2c\right)\alpha^4 \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) \\ \frac{\partial f}{\partial y}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) \end{bmatrix}$$

and $\top$ denotes the transpose. Notice that

$$f\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) = \frac{p+1}{n} \tag{29}$$

$$\frac{\partial f}{\partial x}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) = -2\frac{n+p+1}{n\alpha} \tag{30}$$

$$\frac{\partial f}{\partial y}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) = \frac{1}{\alpha^2}. \tag{31}$$

Replacing the last two expressions into the formula for A yields

$$A = 8\frac{(n+p+1)^2}{cn^2} - 16\left(1+\frac{1}{c}\right)\frac{n+p+1}{n} + 4\left(\frac{2}{c}+5+2c\right) \tag{32}$$

$$\longrightarrow 8\frac{(1+c)^2}{c} - 16\left(1+\frac{1}{c}\right)(1+c) + 4\left(\frac{2}{c}+5+2c\right) = 4 \tag{33}$$

This completes the proof of Proposition 3. □

Proof of Proposition 4 Let $z_1, z_2, \ldots$ denote a sequence of i.i.d. standard normal random variables. Then $Y_{\frac{1}{2}p_k(p_k+1)+a}$ has the same distribution as $z_1^2 + \ldots + z^2_{p_k(p_k+1)/2+a}$. Since $\mathrm{E}[z_1^2] = 1$ and $\mathrm{Var}[z_1^2] = 2$, the Lindeberg-Lévy central limit theorem implies that

$$\sqrt{p_k(p_k+1)/2+a}\left[\frac{Y_{\frac{1}{2}p_k(p_k+1)+a}}{p_k(p_k+1)/2+a} - 1\right] \xrightarrow{\mathrm{D}} \mathcal{N}(0,2). \tag{34}$$

Multiplying the left-hand side by $\sqrt{p_k(p_k+1)+2a}/p_k$, which converges to one, does not affect the limit, therefore

$$\frac{\sqrt{2}}{p_k}Y_{\frac{1}{2}p_k(p_k+1)+a} - \frac{p_k+1}{\sqrt{2}} + \frac{a\sqrt{2}}{p_k} \xrightarrow{\mathrm{D}} \mathcal{N}(0,2). \tag{35}$$

Subtracting from the left-hand side $a\sqrt{2}/p_k$, which converges to zero, does not affect the limit,

therefore

$$\frac{\sqrt{2}}{p_k} Y_{\frac{1}{2}p_k(p_k+1)+a} - \frac{p_k+1}{\sqrt{2}} \xrightarrow{\text{D}} \mathcal{N}(0,2). \tag{36}$$

Rescaling Equation (36) yields Equation (10). □

Proof of Proposition 5 Define the function $g(x,y) = y - 2x + 1$. Then $V = g\left(\frac{1}{p}\text{tr}(S), \frac{1}{p}\text{tr}(S^2)\right)$. Proposition 2 implies that, by the delta method,

$$n\left[V - g\left(\alpha, \frac{n+p+1}{n}\alpha^2\right)\right] \xrightarrow{\text{D}} \mathcal{N}(0, \lim B), \text{ where}$$

$$B = \begin{bmatrix} \frac{\partial g}{\partial x}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) \\ \frac{\partial g}{\partial y}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) \end{bmatrix}^\top \begin{bmatrix} 2\alpha^2/c & 4\left(1+\frac{1}{c}\right)\alpha^3 \\ 4\left(1+\frac{1}{c}\right)\alpha^3 & 4\left(\frac{2}{c}+5+2c\right)\alpha^4 \end{bmatrix} \begin{bmatrix} \frac{\partial g}{\partial x}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) \\ \frac{\partial g}{\partial y}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) \end{bmatrix}.$$

Notice that

$$g\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) = (\alpha-1)^2 + \frac{p+1}{n}\alpha^2 \tag{37}$$

$$\frac{\partial g}{\partial x}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) = -2 \tag{38}$$

$$\frac{\partial g}{\partial y}\left(\alpha, \frac{n+p+1}{n}\alpha^2\right) = 1. \tag{39}$$

Replacing the last two expressions into the formula for B yields

$$B = 8\frac{\alpha^2}{c} - 16\left(1+\frac{1}{c}\right)\alpha^3 + 4\left(\frac{2}{c}+5+2c\right)\alpha^4. \tag{40}$$

First let us find the (n,p)-limiting distribution of V under the null. Setting α equal to one yields $g\left(1, \frac{n+p+1}{n}\right) = \frac{p+1}{n}$ and $B = 4 + 8c$ hence, under the null,

$$n\left(V - \frac{p+1}{n}\right) \xrightarrow{\text{D}} \mathcal{N}(0, 4+8c). \tag{41}$$

Now let us find the (n,p)-limiting distribution of V under the alternative. Setting α equal to $\frac{1-c}{1+c}$ yields

$$g\left(\frac{1-c}{1+c}, \frac{n+p+1}{n} \times \frac{(1-c)^2}{(1+c)^2}\right) = \left(\frac{1-c}{1+c} - 1\right)^2 + \frac{p+1}{n}\left(\frac{1-c}{1+c}\right)^2$$

$$= \frac{p+1}{n} - \frac{c(2-c)}{(1+c)^2} \times \frac{d+1}{n} + o\left(\frac{1}{n}\right)$$

and

$$\begin{aligned} B &= \frac{8}{c}\left(\frac{1-c}{1+c}\right)^2 - 16\left(1+\frac{1}{c}\right)\left(\frac{1-c}{1+c}\right)^3 + 4\left(\frac{2}{c}+5+2c\right)\left(\frac{1-c}{1+c}\right)^4 \\ &= 4(1-c)^2\frac{1+5c^2+2c^3}{(1+c)^4} \end{aligned}$$

hence, under the alternative,

$$n\left(V - \frac{p+1}{n}\right) \xrightarrow{\mathrm{D}} \mathcal{N}\left(\frac{c(2-c)(d+1)}{(1+c)^2}, 4(1-c)^2\frac{1+5c^2+2c^3}{(1+c)^4}\right). \tag{42}$$

Therefore the power of a one-sided test of significance level $\theta > 0$ to reject the null $\Sigma = I$ when the alternative $\Sigma = \frac{1-c}{1+c}I$ is true converges to:

$$1 - \Phi\left(\frac{\Phi^{-1}(\theta)\sqrt{4+8c} - \frac{c(2-c)(d+1)}{(1+c)^2}}{\sqrt{4(1-c)^2\frac{1+5c^2+2c^3}{(1+c)^4}}}\right) < 1 \tag{43}$$

where Φ denotes the standard normal c.d.f. □

Proof of Proposition 6 The moments given in John's (1972) Lemma 2 yield

$$\begin{aligned} \mathrm{E}[n^2(V-W)^2] &= \frac{1}{p^2}\mathrm{E}[\{\mathrm{tr}(S)\}^4] - 2\mathrm{E}[\{\mathrm{tr}(S)\}^2] + p^2 && (44) \\ &= \frac{(np+2)(np+4)(np+6)}{pn^3} - 2\frac{p(np+2)}{n} + p^2 && (45) \\ &= \frac{8p}{n} + \frac{44}{n^2} + \frac{48}{pn^3} && (46) \end{aligned}$$

Therefore $n(V-W)$ converges to zero in quadratic mean, hence in probability, as n goes to infinity for p fixed. This implies that $\frac{1}{2}np \times W$ has the same n-limiting distribution under the null as V. □

Proof of Proposition 7 Define $h(x,y) = y - 2x + 1 - \frac{p}{n}x^2 + \frac{p}{n}$. Then $W = h\left(\frac{1}{p}\mathrm{tr}(S), \frac{1}{p}\mathrm{tr}(S^2)\right)$.

Proposition 2 implies that, by the delta method,

$$n\left[W - h\left(1, \frac{n+p+1}{n}\right)\right] \xrightarrow{\mathrm{D}} \mathcal{N}(0, \lim C), \text{ where}$$

$$C = \begin{bmatrix} \frac{\partial h}{\partial x}\left(1, \frac{n+p+1}{n}\right) \\ \frac{\partial h}{\partial y}\left(1, \frac{n+p+1}{n}\right) \end{bmatrix}^{\top} \begin{bmatrix} \frac{2}{c} & 4\left(1+\frac{1}{c}\right) \\ 4\left(1+\frac{1}{c}\right) & 4\left(\frac{2}{c}+5+2c\right) \end{bmatrix} \begin{bmatrix} \frac{\partial h}{\partial x}\left(1, \frac{n+p+1}{n}\right) \\ \frac{\partial h}{\partial y}\left(1, \frac{n+p+1}{n}\right) \end{bmatrix}.$$

Notice that

$$h\left(1, \frac{n+p+1}{n}\right) = \frac{p+1}{n} \tag{47}$$

$$\frac{\partial h}{\partial x}\left(1, \frac{n+p+1}{n}\right) = -2\frac{n+p}{n} \tag{48}$$

$$\frac{\partial h}{\partial y}\left(1, \frac{n+p+1}{n}\right) = 1. \tag{49}$$

Replacing the last two expressions into the formula for C yields

$$C = 8\frac{(n+p)^2}{cn^2} - 16\left(1+\frac{1}{c}\right)\frac{n+p}{n} + 4\left(\frac{2}{c}+5+2c\right) \tag{50}$$

$$\longrightarrow 8\frac{(1+c)^2}{c} - 16\left(1+\frac{1}{c}\right)(1+c) + 4\left(\frac{2}{c}+5+2c\right) = 4 \tag{51}$$

This completes the proof of Proposition 7. □

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York, 2nd edition.

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

John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, 59(1):169–173.

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

Muirhead, R. J. (1984). *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, New York.

Nagao, H. (1973). On some test criteria for covariance matrix. *Annals of Statistics*, 1(4):700–709.

Silverstein, J. (1986). Eigenvalues and eigenvectors of large dimensional sample covariance matrices. *Contemporary Mathematics*, 50:153–159.

Silverstein, J. W. and Combettes, P. L. (1992). Signal detection via spectral theory of large dimensional random matrices. *IEEE Transactions on Signal Processing*, 40:2100–2105.

Uhlig, H. (1994). On singular Wishart and singular multivariate Beta distributions. *Annals of Statistics*, 22(1):395–405.

Wachter, K. W. (1976). Probability plotting points for principal components. In *Proceedings of the Ninth Interface Symposium on Computer Science and Statistics*, pages 299–308. Hoaglin and Welsch, eds.

Yin, Y. Q. and Krishnaiah, P. R. (1983). A limit theorem for the eigenvalues of product of two random matrices. *Journal of Multivariate Analysis*, 13:489–507.