Revista Colombiana de Estadística - Applied Statistics
July 2022, volume 45, issue 2, pp. 373–399
http://doi.org/10.15446/rce.v45n2.98550

Behavior of Some Hypothesis Tests for the Covariance Matrix of High Dimensional Data

Comportamiento de algunas pruebas de hipótesis para la matriz de covarianza de datos de dimensión alta

Didier Cortez-Elizalde[a], Addy Bolivar-Cime[b]

División Académica de Ciencias Básicas, Universidad Juárez Autónoma de Tabasco, Cunduacán, México

Abstract

The study of the structure of the covariance matrix when the dimension of the data is much greater than the sample size (high dimensional data) is a complicated problem, since we have many unknown parameters and few data. Several hypothesis tests for the covariance matrix, in the high dimensional context and in the classical case (where the dimension of the data is less than the sample size), can be found in the literature. It has been of interest to test the null hypothesis that either the covariance matrix of Gaussian data is equal to the identity matrix or proportional to it, considering the classical case as well as the high dimensional context. Since it is important to have a wide comparison between these tests found in the literature, and for some of them it is difficult to have theoretical results about their powers, in this work we compare several tests by simulations, in terms of the size and power of the test. We also present some examples of application with real high dimensional data found in the literature.

***Key words*:** Covariance matrix; High dimensional data; Hypothesis test; Multivariate Gaussian data; Tracy-Widom law.

Resumen

El estudio de la matriz de covarianza cuando la dimensión de los datos es mucho más grande que el tamaño de la muestra (datos de dimensión alta) es un problema complicado, ya que se tiene una gran cantidad de parámetros desconocidos y pocos datos. Se pueden encontrar en la literatura varias pruebas de hipótesis para la matriz de covarianza, en el contexto de datos de dimensión alta y en el caso clásico (donde la dimensión de los datos es

[a]M.Sc. E-mail: dcortezelizalde@gmail.com
[b]Ph.D. E-mail: addy.bolivar@ujat.mx

menor que el tamaño de la muestra). Ha sido de interés probar la hipótesis nula de que la matriz de covarianza de datos Gaussianos es igual a la matriz identidad o proporcional a ella, considerando el contexto clásico así como el de dimensión alta. Ya que es importante tener una amplia comparación entre estas pruebas encontradas en la literatura, y para algunas de ellas es difícil tener resultados teóricos acerca de sus potencias, en este trabajo comparamos varias pruebas mediante simulaciones, en términos del tamaño y la potencia de la prueba. También presentamos algunos ejemplos de aplicación con datos de dimensión alta reales encontrados en la literatura.

***Palabras clave*:** Datos de dimensión alta; Datos Gaussianos multivariados; Ley Tracy-Widom; Matriz de covarianza; Prueba de hipótesis.

1. Introduction

Data with dimension much greater than the sample size (high dimensional data) arise in many fields, such as genomics, document classification, climatology, finance, functional data analysis, among others (see, Hastie et al., 2009; Johnstone, 2001). In the context of high dimensional data, the estimation of the covariance matrix is a difficult problem because we need to estimate many parameters with few data, for that reason the estimation of the covariance matrix and hypothesis tests about it sometimes require statistical techniques different from those of the classical case, where the sample size is greater than the dimension of the data.

It is worth mentioning that hypothesis tests for the covariance matrix are of interest because several multivariate statistical methodologies strongly depend on the structure of the covariance matrix of the data, for example, principal component analysis, classification, comparison of means, etc. Therefore, it is important to check the structure of the covariance matrix by appropriate statistical tests.

Let $X_1, X_2, \ldots, X_N$ be independent random vectors of the multivariate normal distribution $N_p(\mu, \Sigma)$, where the mean μ and the covariance matrix Σ are unknown, and suppose that we are interested in testing

$$H_0 : \Sigma = I_p \quad \text{vs} \quad H_1 : \Sigma \neq I_p, \tag{1}$$

or

$$H_0 : \Sigma = \lambda I_p \quad \text{vs} \quad H_1 : \Sigma \neq \lambda I_p, \tag{2}$$

where λ is unknown and I_p denotes the identity matrix of size $p \times p$. The null hypothesis H_0 in (2) is called *hypothesis of sphericity*. Let $S_n = \sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})'/n$, with $n = N - 1$, be the *sample covariance matrix* of the data, where $\overline{X} = \sum_{i=1}^{N} X_i/N$ is the *sample mean*.

In Anderson (1984) and Muirhead (2005) it is showed that the likelihood ratio test for (1) is based on the statistic

$$\Lambda = \left(\frac{e}{N}\right)^{pN/2} \operatorname{etr}(-A/2)(\det A)^{N/2},$$

with $A = nS_n$, where $\operatorname{etr}(A) := \exp(\operatorname{tr}(A))$ (the exponential of the trace of the matrix A); and the likelihood ratio test for (2) is based on the *ellipticity statistic* given by

$$V = \frac{\det(S_n)}{[\operatorname{tr}(S_n)/p]^p}.$$

For the case $p \geq N$, with probability one, S_n is not of full rank, therefore $\det(S_n) = 0$. Thus, the likelihood ratio tests for (1) and (2) only exist for the case $p < N$. Hence, it has been of interest to propose and analyze tests for (1) and (2) in the context of high dimensional Gaussian data.

Some tests for (1) were proposed by Johnstone (2001), based on the Tracy-Widom distribution; Ledoit & Wolf (2002); Srivastava (2005); Bai et al. (2009); Cai & Ma (2013) and Srivastava et al. (2014). These tests can be applied in the high dimensional and classical contexts, except the test of Bai et al. (2009) which is only for the classical case. Whereas, for (2) some tests were proposed by John (1971); Srivastava (2005); Zou et al. (2014); Srivastava et al. (2014) and Li & Yao (2016). The last tests can be applied in the high dimensional and classical contexts, except the tests of Li & Yao (2016) which is only for the high dimensional case. In the references provided above there are some comparisons by simulations of some tests, however a broader comparison considering those tests is needed. Furthermore, there is very little about the comparison of the power of the tests.

This work is the result of the master's thesis Cortez-Elizalde (2020). In the present manuscript, we describe briefly several tests found in the literature for (1) and (2) considering Gaussian data, in the high dimensional and classical contexts, being the first one our greatest interest. The tests are compared by simulations in terms of the size and power of the test. The purpose of this analysis is to provide a wide comparison between several tests found in the literature for the covariance matrix of Gaussian data. For many tests it is difficult to provide theoretical results about the power of the test, therefore it is important to carry out simulation studies to compare simultaneously the power of several tests, at least in some cases.

This manuscript is divided as follows, in sections 2 and 3 we present some tests found in the literature for (1) and (2), respectively; in section 4 we present a simulation study for the comparison of the tests; in section 5 we show examples of applications with real high dimensional data found in the literature; in section 6 we provide some conclusions. We also include an appendix with some technical results and details of the simulations.

It is well known, that if we have a random sample $X_1, X_2, \ldots, X_N$ from the $N_p(\mu, \Sigma)$, we can obtain from it a random sample $Z_1, Z_2, \ldots, Z_n$, with $n = N - 1$, from the $N_p(0, \Sigma)$. Furthermore, the sample covariance matrix of the X_i's, S_n, satisfies, $nS_n = \sum_{i=1}^n Z_i Z_i'$; see Appendix A for details. For this reason, some authors of the tests for (1) and (2) give their results considering a random sample $Z_1, Z_2, \ldots, Z_n$ from the $N_p(0, \Sigma)$, and taking $\widetilde{S}_n = \sum_{i=1}^n Z_i Z_i'/n$ as the sample covariance matrix. In the tests presented in this work we consider a random sample $X_1, X_2, \ldots, X_N$ from the $N_p(\mu, \Sigma)$, unless otherwise specified, and $\overline{X}$ and S_n, with $n = N - 1$, are its sample mean and sample covariance matrix, respectively. We try to respect, as much as possible, the notation of the original sources to avoid confusion in the description of their results.

2. Tests for $H_0 : \Sigma = I_p$

Suppose that we are interested in testing (1). Note that if we want to test $H_0 : \Sigma = \Sigma_0$ vs $H_1 : \Sigma \neq \Sigma_0$, where Σ_0 is a specific known positive definite covariance matrix, this is equivalent to testing (1), since we can transform X_i to $Y_i = \Sigma_0^{-1/2} X_i$, $i = 1, 2, \ldots, N$, which are independent random vectors with distribution $N_p(\Sigma_0^{-1/2}\mu, \Sigma_0^{-1/2}\Sigma\Sigma_0^{-1/2})$, and we observe that under the null hypothesis the Y_i's are independent random vectors with distribution $N_p(\Sigma_0^{-1/2}\mu, I_p)$. Thus, we can test (1) based on the transformed data.

2.1. Likelihood ratio test (LRT_1)

Here we suppose $N > p$. As we can see in Muirhead (2005), the level α *likelihood ratio test* for (1) rejects H_0 if $\Lambda \leqslant c_\alpha$, where

$$\Lambda = \left(\frac{e}{N}\right)^{pN/2} \mathrm{etr}(-A/2)(\det A)^{N/2},$$

$A = nS_n$, $n = N - 1$ and c_α is the lower $\alpha \times 100\%$ point of the distribution of Λ. This test is biased, however, doing the following slight modification to the likelihood ratio statistic we obtain an unbiased test

$$\Lambda^* = \left(\frac{e}{n}\right)^{pn/2} \mathrm{etr}(-A/2)(\det A)^{n/2}.$$

Observe that this statistic is obtained from Λ by replacing the sample size N by the degrees of freedom n. Therefore, the likelihood ratio test rejects $H_0 : \Sigma = I_p$ for small enough values of Λ^*, or equivalently, of

$$V^* = \mathrm{etr}(-A/2)(\det A)^{n/2}.$$

When the hypothesis $H_0 : \Sigma = I_p$ is true and n is large, the distribution of $-2\rho\log\Lambda^*$, where $\rho = 1 - (2p^2 + 3p - 1)/(6n(p + 1))$, follows approximately a chi-square distribution with $f = p(p + 1)/2$ degrees of freedom, that is,

$$\mathbb{P}(-2\rho\log\Lambda^* \leq x) \approx \mathbb{P}(\chi_f^2 \leqslant x), \quad \forall x \in \mathbb{R}. \tag{3}$$

Using this approximation, a level α test for (1) rejects H_0 if $-2\rho \log \Lambda^* > \chi_f^2(\alpha)$, where $\chi_f^2(\alpha)$ is the upper $\alpha \times 100\%$ point of the chi-square distribution with f degrees of freedom. This test will be called the level α *likelihood ratio test (LRT_1)* for (1).

2.2. Corrected Likelihood Ratio Test (CLRT)

As it is observed in Bai et al. (2009), the likelihood ratio test has a size much higher than the nominal level in the case when the dimension of the data is very large. For that reason Bai et al. (2009) proposed a correction to the likelihood ratio test statistic using some results of Random Matrix Theory.

Suppose $n > p$. Let

$$L^* = \mathrm{tr}S_n - \log(\det S_n) - p. \tag{4}$$

From (3) we have that

$$T_n = nL^* = -2\log\Lambda^*,$$

converges to the chi-square distribution with $p(p+1)/2$ degrees of freedom under $H_0 : \Sigma = I_p$, when p is fixed and $n \to \infty$. In Bai et al. (2009) it is proved the following result.

Theorem 1. *Suppose $y_n := p/n \to y \in (0,1)$ when $n, p \to \infty$. Let L^* be as in equation (4) and $g(x) = x - \log x - 1$. Then, under H_0 and when $n \to \infty$*

$$\widetilde{T}_n = v(g)^{-1/2}[L^* - p \cdot F^{y_n}(g) - m(g)] \xrightarrow{d} N(0,1), \tag{5}$$

where

$$\begin{aligned} m(g) &= -\frac{\log(1-y)}{2}, \\ v(g) &= -2\log(1-y) - 2y, \\ F^{y_n}(g) &= 1 - \frac{y_n - 1}{y_n}\log(1-y_n), \end{aligned}$$

and "$\xrightarrow{d}$" denotes convergence in distribution. The expressions $m(g)$ and $v(g)$ are the asymptotic mean and variance, respectively, of $G_n(g) = L^* - p \cdot F^{y_n}(g)$ when $n \to \infty$. The statistic $\widetilde{T}_n$ will be called the corrected likelihood ratio statistic.

A level α test for (1) based on the statistic (5) rejects H_0 if $\widetilde{T}_n > z_\alpha$, where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution $N(0,1)$. This test will be called the level α *corrected likelihood ratio test (CLRT)* for (1).

2.3. Ledoit and Wolf (LW) Test

For the testing problem (1) the likelihood ratio test statistic is degenerate when p is greater than n. Nagao (1973) proposed the test statistic $V = \mathrm{tr}[(S_n - I_p)^2]/p$, which does not degenerate in that case. This test statistic for the testing problem (1) is the equivalent of the test statistic U proposed by John (1971) for the testing problem (2) (see section 3.2). In Ledoit & Wolf (2002) it is shown that the power and size of the sphericity test based on U is robust against p large, and even large than n. However, the test of (1) based on V is not consistent against every alternative when p goes to infinity with n. For that reason in Ledoit & Wolf (2002) it is proposed the following test statistic for (1)

$$W = \frac{1}{p}\mathrm{tr}[(S_n - I_p)^2] - \frac{p}{n}\left[\frac{1}{p}\mathrm{tr}S_n\right]^2 + \frac{p}{n},$$

which will be called the *Ledoit & Wolf statistic*. 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 Ledoit & Wolf (2002) is proved that under $H_0 : \Sigma = I_p$, when $n \to \infty$ and p is fixed,

$$nW - p \xrightarrow{d} \frac{2}{p}\chi^2_{p(p+1)/2} - p.$$

They also proved the following result.

Theorem 2. *Suppose that $p/n \to y \in (0, \infty)$ as $n, p \to \infty$, then under H_0*

$$nW - p \xrightarrow{d} N(1, 4).$$

By the last theorem, a level α test for (1) based on the Ledoit & Wolf statistic, rejects H_0 if $(nW - p - 1)/2 > z_\alpha$, where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution. We called this test the level α *Ledoit and Wolf (LW) test* for (1).

2.4. Tracy-Widom (TW) Test

In Johnstone (2001) the case when n and p are large, with $n = n(p)$ and $n/p \to \gamma > 0$ as $p \to \infty$, is considered. Based on Random Matrix Theory, in that work it is obtained the asymptotic distribution of the largest eigenvalue of the sample covariance matrix of Gaussian data, which can be used to give a test for (1).

We denote by $W_p(n, \Sigma)$ the Wishart distribution with n degrees of freedom and covariance matrix Σ of size $p \times p$. The next theorem proposed by Johnstone (2001), gives the asymptotic distribution of the largest eigenvalue of a random matrix with Wishart distribution.

Theorem 3. *Let A with distribution $W_p(n, I_p)$ and let l_1 be the largest eigenvalue of A. If $n/p \to \gamma \geq 1$ as $p \to \infty$, then*

$$\frac{l_1 - \mu_{np}}{\sigma_{np}} \xrightarrow{d} F_1,$$

where the center and scaling constants are

$$\mu_{np} = \left(\sqrt{n-1} + \sqrt{p}\right)^2, \tag{6}$$

$$\sigma_{np} = \left(\sqrt{n-1} + \sqrt{p}\right)\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3}, \tag{7}$$

and F_1 is the distribution function of the Tracy-Widom law of order 1.

The last theorem is stated for the case when $n \geq p$. However, as mentioned in Johnstone (2001), it applies equally well if $n < p$ are both large, simply by reversing the roles of n and p in (6) and (7). Therefore, if we have a random sample of size N from the multivariate normal distribution $N_p(\mu, \Sigma)$ and l_1 is the largest eigenvalue of the sample covariance matrix S_n, with distribution $W_p(n, \Sigma)$, a level α test for (1) rejects H_0 if

$$\frac{nl_1 - \mu_{np}}{\sigma_{np}}$$

is greater than the upper $\alpha \times 100\%$ point of the Tracy-Widom distribution F_1, denoted by $F_1(\alpha)$, where μ_{np} and σ_{np} are the center and scaling constants of Theorem 3. We called this test the level α *Tracy-Widom (TW) test* for (1).

2.5. Cai and Ma (CM) Test

Motivated by a test of Chen et al. (2010), Cai & Ma (2013) proposed a test for (1). The original proposal in Chen et al. (2010) involves higher order symmetric functions of the X_i's than the proposal of Cai & Ma (2013). The test is established in the setting where the dimension $p = p_n \to \infty$ as the sample size $n \to \infty$, and there is no restriction on the limit of p/n. Cai & Ma (2013) proved that the asymptotic power of their proposed test, in a subset of covariance matrices, uniformly dominates that of the CLRT given in section 2.2, when $p < n$ and $p/n \to y \in (0, 1)$.

Let $X_1, \dots, X_n$ be independent random vectors from the normal distribution $N_p(0, \Sigma)$. Cai & Ma (2013) consider to test the hypothesis $H_0 : \Sigma = I_p$ versus the alternative hypothesis

$$H_1 : \Sigma \in \Theta, \quad \Theta = \{\Sigma : \| \Sigma - I_p \|_F \geqslant \epsilon_n\}, \tag{8}$$

where $\epsilon_n > 0$ and

$$\| A \|_F = \sqrt{\left(\sum_{i,j} a_{ij}^2\right)} = \sqrt{\operatorname{tr}(AA^H)},$$

is the Frobenius norm of the matrix $A = (a_{ij})$, where A^H is the conjugate transpose of A. The difficulty of testing between H_0 and H_1 depends on the value of ϵ_n; the smaller ϵ_n is, the harder it is to distinguish between the two hypotheses.

Define the statistic

$$T_n = \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} h(X_i, X_j),$$

where

$$h(X_i, X_j) = (X_i' X_j)^2 - (X_i' X_i + X_j' X_j) + p.$$

It is important to mention that T_n is an estimator of $\| \Sigma - I_p \|_F^2 = \operatorname{tr}[(\Sigma - I_p)^2]$. The next result given in Cai & Ma (2013) provides the asymptotic distribution of T_n.

Theorem 4. *Suppose that $p \to \infty$ as $n \to \infty$. If the succession of covariance matrices satisfies*

$$\mathrm{tr}(\Sigma^2) \longrightarrow \infty \quad \textit{and} \quad \mathrm{tr}(\Sigma^4)/\mathrm{tr}^2(\Sigma^2) \longrightarrow 0$$

as $n \to \infty$, then

$$\frac{T_n - \mu_n(\Sigma)}{\sigma_n(\Sigma)} \xrightarrow{d} N(0,1),$$

where

$$\mu_n(\Sigma) = \mathbb{E}_\Sigma(T_n) = \mathrm{tr}(\Sigma - I_p)^2,$$
$$\sigma_n^2(\Sigma) = \mathrm{var}_\Sigma(T_n) = \frac{4}{n(n-1)}(\mathrm{tr}^2(\Sigma^2) + \mathrm{tr}(\Sigma^4)) + \frac{8}{n}\mathrm{tr}(\Sigma^2(\Sigma - I_p)^2).$$

Note that the succession of identity matrices $\{I_p\}_{p=1}^{\infty}$ satisfies the assumptions of the last theorem, furthermore $\mu_n(I_p) = 0$ and $\sigma_n^2(I_p) = 4p(p+1)/n(n-1)$. Theorem 4 provides the asymptotic behavior of T_n under H_0. Thus, a level α test for (1) based on the statistic T_n rejects H_0 if $T_n > z_\alpha 2\sqrt{\frac{p(p+1)}{n(n-1)}}$, where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution. We called this test the level α *Cai and Ma (CM) test* for (1).

2.6. Srivastava's Tests (T_{2s}, T_2)

The tests for (1) presented in this section were proposed by Srivastava (2005) and Srivastava et al. (2014). They considered a distance function between the null hypothesis and the alternative hypothesis, and proposed tests based on consistent estimators of this parametric function of the covariance matrix Σ for testing (1). Specifically, they considered estimators of the squared Frobenius norm (divided by p)

$$\frac{1}{p}\mathrm{tr}[(\Sigma - I_p)^2] = \frac{1}{p}[\mathrm{tr}(\Sigma^2) - 2\mathrm{tr}(\Sigma) + p] = a_2 - 2a_1 + 1, \tag{9}$$

where

$$a_i = \frac{\mathrm{tr}\Sigma^i}{p}, \quad i = 1, 2, \ldots. \tag{10}$$

Observe that under the null hypothesis $H_0 : \Sigma = I_p$ we have that (9) is equal to zero. Therefore, a test for (1) can be based on an estimator of (9), where the null hypothesis should be rejected if the observed value of the estimator is greater than some specific amount.

Consider the following assumptions

$$\begin{aligned} &a)\ \textit{If } p \to \infty, \text{then } a_i \to a_i^0,\ 0 < a_i^0 < \infty, \quad i = 1, \ldots, 8. \\ &b)\ n = O(p^\delta), \quad 0 < \delta \leqslant 1, \end{aligned} \tag{11}$$

where $n = O(p^\delta)$ denotes that n/p^δ remains bounded as n and p go to infinity, this includes the case when $(n/p) \to 0$. The following results can be found in Srivastava (2005).

Lemma 1. *Under assumption a), and when $n \to \infty$, unbiased and consistent estimators of a_1 and a_2 are given, respectively, by*

$$\widehat{a}_1 = \frac{\mathrm{tr}S_n}{p}, \quad \widehat{a}_{2s} = \frac{n^2}{(n-1)(n+2)p}\left[\mathrm{tr}S_n^2 - \frac{1}{n}(\mathrm{tr}S_n)^2\right].$$

Theorem 5. *Consider the assumptions (11). Under $H_0 : \Sigma = I_p$, when $n, p \to \infty$, we have*

$$T_{2s} = \frac{n}{2}\left(\widehat{a}_{2s} - 2\widehat{a}_1 + 1\right) \xrightarrow{d} N(0,1).$$

Thus, a level α test for (1) based on T_{2s} rejects H_0 if $T_{2s} > z_\alpha$, where z_α is the upper $\alpha \times 100\%$ point of the normal standard distribution. We called this test the level α *Srivastava's test* T_{2s} for (1).

Srivastava et al. (2014) proposed a different test, but now based on a new unbiased estimator of a_2, given by

$$\widehat{a}_2 = \frac{1}{f}\left[(N-2)n\mathrm{tr}(M^2) - Nn\mathrm{tr}(D^2) + (\mathrm{tr}D)^2\right], \qquad (12)$$

where $f = pN(N-1)(N-2)(N-3)$, $M = \mathbf{Y}'\mathbf{Y}$, $\mathbf{Y} = (Y_1, \ldots, Y_N)$, $D = \mathrm{diag}(Y_1'Y_1, \ldots, Y_N'Y_N)$, with $Y_i = X_i - \overline{X}$. In Srivastava et al. (2014) it is proved the following result.

Theorem 6. *Consider the assumption $N = O(p^\delta)$, $1/2 < \delta < 1$. Under $H_0 : \Sigma = I_p$, when $n, p \to \infty$, we have*

$$T_2 = \frac{n}{2}\left(\widehat{a}_2 - 2\widehat{a}_1 + 1\right) \xrightarrow{d} N(0,1).$$

Therefore, a level α test for (1) based on T_2 rejects H_0 if $T_2 > z_\alpha$, where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution. We called this test the level α *Srivastava's test* T_2 for (1).

3. Tests for $H_0 : \Sigma = \lambda I_p$

In this section we describe briefly some tests for sphericity, that is, tests for (2), where the null hypothesis affirms that the covariance matrix is proportional to the identity matrix.

Note that if we want to test $H_0 : \Sigma = \lambda\Sigma_0$ vs $H_1 : \Sigma \neq \lambda\Sigma_0$, where Σ_0 is a specific known positive definite covariance matrix and λ is unknown, this is equivalent to test (2), by transforming the data to $Y_i = \Sigma_0^{-1/2}X_i$, $i = 1, 2, \ldots, N$, and testing (2) based on the transformed data.

3.1. Likelihood Ratio Test (LRT_2)

In Muirhead (2005) it is shown that the level α *likelihood ratio test* for (2) rejects H_0 if

$$V \equiv \frac{\det A}{(\mathrm{tr}(A)/p)^p} = \frac{\det S_n}{(\mathrm{tr}(S_n)/p)^p} \leqslant k_\alpha, \tag{13}$$

where $A = nS_n$ and k_α is the lower $\alpha \times 100\%$ point of the distribution of V. The statistic V is called *ellipticity statistic*.

When the hypothesis $H_0 : \Sigma = \lambda I_p$ is true, the distribution of $-n\rho \log V$, where $\rho = 1 - (2p^2 + p + 2)/6np$, has approximately a chi-square distribution with $f = (p+2)(p-1)/2$ degrees of freedom when n is large, that is,

$$\mathbb{P}(-n\rho \log V \leqslant x) \approx \mathbb{P}(\chi_f^2 \leqslant x), \quad \forall x \in \mathbb{R}. \tag{14}$$

By this approximation, a level α test for (2) rejects H_0 if $-n\rho \log V > \chi_f^2(\alpha)$, where $\chi_f^2(\alpha)$ is the upper $\alpha \times 100\%$ point of the chi-square distribution with f degrees of freedom. We called this test the level α *likelihood ratio test (LRT_2)* for (2).

3.2. John's (J) Test

For the testing problem (2) the likelihood ratio test is degenerate when p is greater than n. John (1971) proposed to test (2) using the following test statistic, which does not degenerate,

$$U = \frac{1}{p}\mathrm{tr}\left[\left(\frac{S_n}{(1/p)\mathrm{tr}(S_n)} - I_p\right)^2\right] = \frac{(1/p)\mathrm{tr}(S_n^2)}{[(1/p)\mathrm{tr}(S_n)]^2} - 1. \tag{15}$$

We called this statistic the *John's statistic*. John (1972) proved that, when $n \to \infty$ and p is fixed, under H_0

$$nU - p \xrightarrow{d} \frac{2}{p}\chi^2_{p(p+1)/2-1} - p.$$

In Ledoit & Wolf (2002) it is shown that the power and size of the sphericity test based on U is robust against p large, and even larger than n. They proved the following result.

Theorem 7. *Suppose that $p/n \to y \in (0,\infty)$ as $n, p \to \infty$, then under $H_0 : \Sigma = \lambda I_p$*

$$nU - p \xrightarrow{d} N(1,4).$$

Then, a level α test for (2) based on the John's statistic rejects H_0 if $(nU - p - 1)/2 > z_\alpha$, where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution. This test will be called the level α *John's (J) test* for (2).

3.3. Quasi-Likelihood Ratio Test (QLRT)

Let $X_1, X_2, \ldots, X_n$ be independent random vectors of the multivariate normal distribution $N_p(0, \Sigma)$, and let $\mathbf{X} = (X_1, \ldots, X_n)$ be the matrix of size $p \times n$ whose columns are the vectors X_i, $i = 1, 2, \ldots, n$. The likelihood ratio test for (2), denoted by LRT_2 and described in section 3.1, requires $p \leqslant n$ because when $p > n$, $p - n$ eigenvalues of the sample covariance matrix S_n are zero and therefore V in (13) is equal to zero. Li & Yao (2016) proposed an extension of the LRT_2 for the case when $p > n$, by considering the matrix $\mathbf{X}'\mathbf{X}/p$ which has exactly the same n non-zero eigenvalues with the matrix $S_n = \mathbf{X}\mathbf{X}'/n$ (up to some scaling). Their results are in the ultra-dimensional asymptotic setting $p \gg n$, where $p/n \to \infty$ and $n \to \infty$.

The quantity $-n \log V$ that appears in the left side of (14) can be expressed as

$$n \log \left[\frac{\left(\frac{1}{p} \sum_{i=1}^{p} \ell_i \right)^p}{\prod_{i=1}^{p} \ell_i} \right],$$

where ℓ_i, $i = 1, 2, \ldots, p$, are the eigenvalues of the matrix S_n. Based on the last expression, Li & Yao (2016) proposed the *quasi-likelihood ratio statistic* given by

$$\mathcal{L}_n = \frac{p}{n} \log \left[\frac{\left(\frac{1}{n} \sum_{i=1}^{n} \tilde{\lambda}_i \right)^n}{\prod_{i=1}^{n} \tilde{\lambda}_i} \right],$$

where $\tilde{\lambda}_i$, $i = 1, 2, \ldots, n$, are the eigenvalues of the matrix $\mathbf{X}'\mathbf{X}/p$. They presented the next theorem.

Theorem 8. *Suppose that $p/n \to \infty$ and $n \to \infty$, then under H_0*

$$\mathcal{L}_n - \frac{n}{2} - \frac{n^2}{6p} - \frac{v_4 - 2}{2} \xrightarrow{d} N(0, 1),$$

where v_4 is the fourth moment of the standard normal distribution.

Thus, a level α test for (2) based on the statistic $\mathcal{L}_n$ rejects H_0 if $\mathcal{L}_n - \frac{n}{2} - \frac{n^2}{6p} - \frac{v_4 - 2}{2} > z_\alpha$, where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution. We called this test the level α *quasi-likelihood ratio test (QLRT)* for (2).

3.4. Srivastava's Tests (T_{1s}, T_1)

The next tests for (2) were proposed by Srivastava (2005) and Srivastava et al. (2014). Let a_i, for $i = 1, 2, \ldots,$ given by (10). Srivastava (2005) showed, using the

Cauchy-Schwarz inequality, that

$$\frac{a_2}{a_1^2} = \frac{\sum_{i=1}^{p} \lambda_i^2/p}{(\sum_{i=1}^{p} \lambda_i/p)^2} \geq 1,$$

with equality holding if and only if $\lambda_i = \lambda$, for $i = 1, 2, \ldots, p$ and some constant λ; where the λ_i's are the eigenvalues of Σ. Thus, a measure of sphericity is given by

$$\frac{a_2}{a_1^2} - 1, \tag{16}$$

which is equal to zero if and only if $\lambda_i = \lambda$, for $i = 1, 2, \ldots, p$. Therefore, a test for (2) can be based on an estimator of (16), where the null hypothesis should be rejected if the observed value of the estimator is greater than some specific amount.

Consider the unbiased estimators of a_1 and a_2 of lemma 1. Define

$$T_{1s} = \frac{n}{2}\left(\frac{\widehat{a}_{2s}}{\widehat{a}_1^2} - 1\right).$$

We have the following result given by Srivastava (2005).

Theorem 9. *Consider the assumptions (11). Under $H_0 : \Sigma = \lambda I_p$, when $n, p \to \infty$, we have*

$$T_{1s} \xrightarrow{d} N(0, 1).$$

Thus, a level α test for (2) based on T_{1s} rejects H_0 if $T_{1s} > z_\alpha$, where z_α the upper $\alpha \times 100\%$ point of the standard normal distribution. This test will be called the level α *Srivastava's test* T_{1s} for (2).

Consider $\widehat{a}_2$ given by (12). Substituting $\widehat{a}_2$ by $\widehat{a}_{2s}$ in T_{1s} we get

$$T_1 = \frac{n}{2}\left(\frac{\widehat{a}_2}{\widehat{a}_1^2} - 1\right).$$

The next result is provided by Srivastava et al. (2014).

Theorem 10. *Consider $N = O(p^\delta)$, $1/2 < \delta < 1$. Under $H_0 : \Sigma = \lambda I_p$, when $n, p \to \infty$, we have*

$$T_1 \xrightarrow{d} N(0, 1).$$

Therefore, a level α test for (2) based on T_1 rejects H_0 if $T_1 > z_\alpha$, where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution. We called this test the level α *Srivastava's test* T_1 for (2).

3.5. Zou's (Z) Test

The next test was proposed by Zou et al. (2014). They proposed a test for (2) considering a random sample from a p-variate elliptical distribution, in the high dimensional context. Their proposal is a modification of the sign test statistic (Hallin & Paindaveine (2006)), which is defined by mimicking John's test statistic given by (15), considering the multivariate sign function.

Let $X_1, \ldots, X_n$ be random vectors from a p-variate *elliptical distribution* with density function of the form

$$\det(\Sigma_p)^{-1/2} g_p(\|\Sigma_p^{-1/2}(X - \theta_p)\|),$$

where $\|X\| = (X'X)^{1/2}$ is the Euclidean norm of the vector X, θ_p is the symmetry centre, Σ_p is a positive-definite symmetric $p \times p$ scatter matrix, and g_p is a non-negative function of a real variable. The matrix Σ_p that describes the covariances between the p variables can be expressed as $\Sigma_p = \sigma_p \Lambda_p$, where $\sigma_p = \sigma(\Sigma_p)$ is a scale parameter and $\Lambda_p = \sigma_p^{-1}\Sigma_p$ is a shape matrix. The scale parameter is assumed to satisfy $\sigma(I_p) = 1$ and $\sigma(a\Sigma_p) = a\sigma(\Sigma_p)$ for all $a > 0$. We are interested in testing $H_0 : \Sigma_p = \lambda I_p$, which is equivalent to $\Lambda_p = I_p$.

The *multivariate sign function* is defined as

$$U(X) = \|X\|^{-1} X, \quad \text{for all } X \neq \mathbf{0}.$$

The observed signs for X_i, $i = 1, 2, \ldots, n$, are

$$U_i = U(X_i - \theta_p).$$

Let $X_1, \ldots, X_n$ be random vectors from the multivariate normal distribution $N_p(\mu, \Sigma)$. The multivariate normal distribution is an elliptical distribution, for which $\theta_p = \mu$ and its estimator is given by $\widehat{\theta}_{n,p} = \overline{X}$. Consider the statistic

$$\tilde{Q} = \frac{p}{n(n-1)} \sum_{i \neq j} (\widehat{U}_i' \widehat{U}_j')^2 - 1,$$

where $\widehat{U}_i = U(X_i - \overline{X})$. This statistic is a modified version of the sign test statistic given in Hallin & Paindaveine (2006). Let $R_i = \| X_i - \mu \|$ and consider the following assumption.

Assumption 1. The moments $E(R_i^{-k})$ for $k = 1, 2, \ldots, 4$ exist for large enough p; and $E(R_i^{-k})/E(R_i^{-1})^k \to d_k \in [1, \infty)$ as $p \to \infty$, where the d_k are constants, for $k = 2, 3, 4$.

In the supplementary material of Zou et al. (2014) is verified Assumption 1 for the multivariate normal distribution, the multivariate t distribution, and mixtures of multivariate normal distributions. The following result of Zou et al. (2014) gives the asymptotic distribution of $\tilde{Q}$.

Theorem 11. *Under $H_0 : \Sigma = \lambda I_p$ and Assumption 1, if $p = O(n^2)$, then*

$$\frac{\tilde{Q} - p\delta_{n,p}}{\tilde{\sigma}_0} \xrightarrow{d} N(0, 1)$$

as $n, p \to \infty$*, where* $\widetilde{\sigma}_0^2 = 4(p-1)/[n(n-1)(p+2)]$ *and*

$$\begin{aligned}\delta_{n,p} =& \frac{1}{n^2}\left(2 - \frac{2\mathbb{E}(R_i^{-2})}{\mathbb{E}(R_i^{-1})^2} + \left[\frac{\mathbb{E}(R_i^{-2})}{\mathbb{E}(R_i^{-1})^2}\right]^2\right) \\ &+ \frac{1}{n^3}\left[\frac{8\mathbb{E}(R_i^{-2})}{\mathbb{E}(R_i^{-1})^2} - 6\left(\frac{\mathbb{E}(R_i^{-2})}{\mathbb{E}(R_i^{-1})^2}\right)^2 + \frac{2\mathbb{E}(R_i^{-2})\mathbb{E}(R_i^{-3})}{\mathbb{E}(R_i^{-1})^5} - \frac{\mathbb{E}(2R_i^{-3})}{\mathbb{E}(R_i^{-1})^3}\right].\end{aligned} \tag{17}$$

The unknown quantities in $\delta_{n,p}$ are $\mathbb{E}(R_i^{-2})/\mathbb{E}(R_i^{-1})^2$ and $\mathbb{E}(R_i^{-3})/\mathbb{E}(R_i^{-1})^3$, which can be estimated as follows. Let

$$\widehat{R}_i = \|X_i - \widehat{\theta}_{n,p}\|, \quad \widehat{R}_{i*} = \widehat{R}_i + \widehat{\theta}'_{n,p}\widehat{U}_i - 2^{-1}\widehat{R}_i^{-1}\|\widehat{\theta}_{n,p}\|^2.$$

Thus, substituting

$$\mathbb{E}(R_i^{-k})/\mathbb{E}(R_i^{-1})^k$$

by

$$\frac{n^{k-1}\sum_{i=1}^n \widehat{R}_{i*}^{-k}}{(\sum_{i=1}^n \widehat{R}_{i*}^{-1})^k}$$

in (17) we obtain an estimator of $\delta_{n,p}$, denoted by $\hat{\hat{\delta}}_{n,p}$.

Therefore, a level α test for (2) based on the statistic $\tilde{Q}$ rejects H_0 if $(\tilde{Q} - p\hat{\delta}_{n,p})/\tilde{\sigma}_0 > z_\alpha$, where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution. We called this test the level α *Zou's (Z) test* for (2).

4. Simulation Study

In this section we present a simulation study to compare the tests presented before, in terms of the size and power of the test. For both hypothesis testing problems (1) and (2), we considered $M = 10,000$ random samples of size $N = n+1$ from the p-variate standard normal distribution. We considered several values of n and p. For the case $n > p$, we fixed $n = 500$ and took five values of p less than n; and for the case $p \geqslant n$, we fixed $p = 500$ and took five values of n less than p. We considered the significance level $\alpha = 0.05$ for all the tests, and we calculated the empirical size of each test, given by the proportion of rejections of H_0 with the test. If a test is good in terms of the size of the test, its empirical size should be very close to the significance level.

To evaluate the power of the tests, for some values of p and n, we compute the empirical power of each test, given by the proportion of rejections of H_0 under H_1 with the test, considering random samples from the multivariate normal distribution with zero mean and covariance matrix in a subset of matrices satisfying the alternative hypothesis H_1. The covariance matrices of this subset have the form

$$\Sigma = I_p + hvv',$$

where h is a positive scalar and v is a unit vector. Observe that this covariance matrix is a slightly deviation from the identity matrix when h is very small, and it becomes very different from the identity when h increases. The vector v was randomly generated in the following way: we first generated a random vector from the p-variate standard normal distribution, then we divided each entry by the norm of the random vector in order to obtain a unit vector. The scalar h varied in a range of adequate values. This range is chosen in such a way that we can observe how the values of the empirical powers are close to one when h increases. Since the empirical powers may change for different values of p and n, the range of values of h may also vary.

For the tests that consider a random sample from the $N_p(\mu, \Sigma)$, we compute the test statistics with the original data. For the tests that consider a random sample from the $N_p(0, \Sigma)$, which are the CM test and the QLRT, we first transform the data to a random sample of size $n = N - 1$ from the $N_p(0, \Sigma)$, using the transformation of multivariate normal data given in the Appendix A. We also present in Appendix A the general ideas of the algorithms used in the simulations to compute the empirical sizes and empirical powers of the tests. We used the software R (`https://www.r-project.org`) to perform the simulation study.

4.1. Simulations for $H_0 : \Sigma = I_p$

The considered tests for (1) are: LRT_1, TW, LW, CLRT, CM, T_{2s}, T_2. For the classical case ($p < n$) all the tests can be applied, and for the high dimensional case ($p \geq n$) all the tests, except LRT_1 and CLRT, can be applied.

4.1.1. Empirical Sizes of the Tests

The results of the empirical sizes of the tests are shown in tables 1 and 2. In these tables we observe the following:

1. The empirical sizes of TW, LW, CM, T_{2s} and T_2 are very close to the significance level, for the cases $p < n$ and $p \geq n$, this means that these tests are good in terms of the size of the test.

2. On the other hand, LRT_1 is not good, since its empirical sizes are far from the considered significance level. When p approaches to n the empirical size approaches to 1, however this test can be good if n is large enough with respect to p.

3. CLRT corrects the problem of LRT_1, since its empirical sizes are very close to the considered significance level, actually when p approaches to n the behavior of the empirical size of CLRT is better.

4.1.2. Empirical powers of the tests

The results of the empirical powers of the tests by varying h, are presented in the tables 3 and 4. The figures 1 and 2 illustrate the behavior of the empirical

Table 1: Empirical sizes of the tests for $H_0 : \Sigma = I_p$, case $p < n$.

n,p	LRT_1	TW	CLRT	LW	CM	T_{2s}	T_2
500,25	0.0740	0.0498	0.0555	0.0541	0.0526	0.0546	0.0552
500,50	0.2240	0.0480	0.0526	0.0538	0.0519	0.0541	0.0535
500,100	0.9753	0.0520	0.0526	0.0539	0.0527	0.0525	0.0531
500,200	1	0.0517	0.0513	0.0493	0.0500	0.0491	0.0500
500,400	1	0.0536	0.0509	0.0492	0.0487	0.0495	0.0503

Table 2: Empirical sizes of the tests for $H_0 : \Sigma = I_p$, case $p \geqslant n$.

n,p	TW	LW	CM	T_{2s}	T_2
25,500	0.0503	0.055	0.0537	0.0541	0.0591
50,500	0.0494	0.0504	0.0509	0.0491	0.0514
100,500	0.0526	0.0527	0.0483	0.0485	0.0491
200,500	0.051	0.0549	0.0547	0.0546	0.0546
400,500	0.0528	0.0501	0.0516	0.0498	0.0508

powers of the tests for the classical case ($p < n$) and the high dimensional case ($p \geqslant n$), respectively. We observe the following:

1. For $p < n$ and $p \geq n$, the empirical powers of the tests are good when h increases, since they are equal to or approximately one.

2. For $p < n$ the empirical power of the TW test is bigger than the corresponding to the rest of the tests, except LRT_1. For $p \geq n$ the empirical power of the TW test has the best behavior.

3. Despite that the empirical power of LRT_1 is one for all values of h when $n = 500$ and $p = 200$, this test is bad in terms of the size (see Table 1), therefore it is not recommended in this case.

4. Even when the empirical power of CLRT is the smallest, it has an acceptable behavior.

5. The empirical powers of LW, CM, T_{2s} and T_2 are very close between them that their graphs are almost indistinguishable for both cases, $p < n$ and $p \geq n$. This may be due to the fact that the test statistics of these tests are based on estimators of the squared Frobenius norm $\| \Sigma - I_p \|_F^2 = \mathrm{tr}[(\Sigma - I_p)^2]$.

4.2. Simulations for $H_0 : \Sigma = \lambda I_p$

The considered tests for (2) are: LRT_2, J, QLRT, T_{1s}, T_1 and Z. For the case $n > p$ all the tests, except QLRT, can be applied; whereas for the case $p \geq n$ all the test, except LRT_2, can be applied.

TABLE 3: Empirical powers of the tests for $H_0 : \Sigma = I_p$, case $p < n$.

$n = 500$, $p = 50$							
h	LRT_1	TW	CLRT	LW	CM	T_{2s}	T_2
0.3	0.3494	0.1884	0.104	0.1266	0.1215	0.1255	0.1263
0.6	0.6777	0.9047	0.3514	0.5456	0.5373	0.5413	0.5425
0.9	0.9405	0.9997	0.7731	0.9543	0.9514	0.9524	0.9526
1.2	0.9964	1	0.9729	0.9994	0.9992	0.9993	0.9993
1.5	1	1	0.9986	1	1	1	1
$n = 500$, $p = 200$							
h	LRT_1	TW	CLRT	LW	CM	T_{2s}	T_2
0.5	1	0.1172	0.0751	0.095	0.0948	0.0945	0.0955
1	1	0.8746	0.1596	0.3575	0.3535	0.3538	0.3554
1.5	1	0.9998	0.3455	0.8389	0.8363	0.8361	0.8371
2	1	1	0.6014	0.9945	0.9943	0.9944	0.9943
2.5	1	1	0.8295	0.9999	0.9999	0.9999	0.9999

TABLE 4: Empirical powers of the tests for $H_0 : \Sigma = I_p$, case $p \geqslant n$.

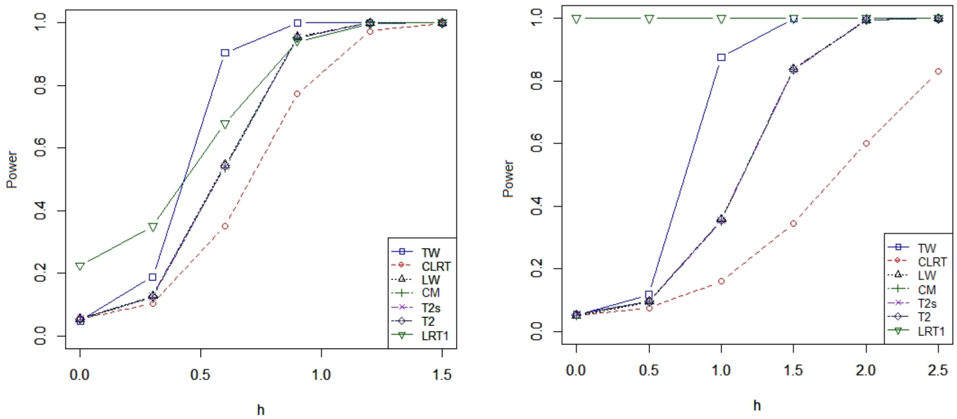
$p = 500$, $n = 50$					
h	TW	LW	CM	T_{2s}	T_2
1.9	0.0911	0.077	0.0739	0.0732	0.0763
3.8	0.3753	0.196	0.1881	0.1877	0.1935
5.7	0.7866	0.4707	0.4638	0.4686	0.4716
7.6	0.9577	0.7576	0.745	0.7551	.7514
9.5	0.9932	0.9139	0.9096	0.916	0.9148
$p = 500$, $n = 200$					
h	TW	LW	CM	T_{2s}	T_2
1.2	0.1086	0.0943	0.0935	0.0945	0.0945
2.4	0.796	0.3261	0.3241	0.322	0.3211
3.6	0.9985	0.7751	0.7703	0.7735	0.7726
4.8	1	0.9788	0.9771	0.979	0.9784
6	1	0.9994	0.9993	0.9994	0.9994

4.2.1. Empirical Sizes of the Tests

The results of the empirical sizes of the tests are in the tables 5 and 6. We observe the following:

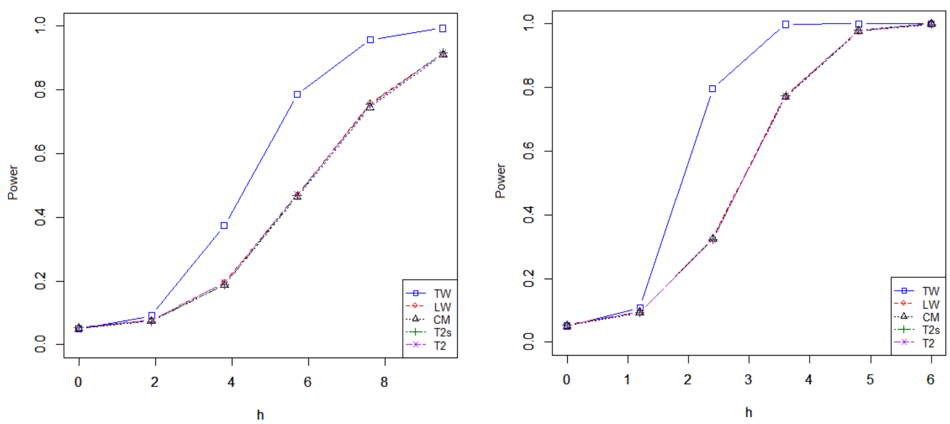
1. The empirical sizes of J, T_{1s}, T_1 y Z are very close to the significance level, for $p < n$ and $p \geqslant n$, therefore they have a good behavior in terms of the size of the test.

2. On the other hand, for $p < n$, LRT_2 has empirical size close to the significance level when n is large enough with respect to p, but it has a bad behavior when p is close to n.



(a) $p = 50, n = 500$ (b) $p = 200, n = 500$

Figure 1: Empirical powers of the tests for $H_0 : \Sigma = I_p$, case $p < n$.



(a) $p = 500, n = 50$ (b) $p = 500, n = 200$

Figure 2: Empirical powers of the tests for $H_0 : \Sigma = I_p$, case $p \geqslant n$.

3. For $p \geqslant n$, QLRT has good behavior of the empirical size only when p is large enough with respect to n. When n approaches to p the empirical size approaches to one.

Table 5: Empirical sizes of the tests for $H_0 : \Sigma = \lambda I_p$, case $p < n$.

n, p	LRT_2	J	T_{1s}	T_1	Z
500,25	0.0475	0.0493	0.0492	0.0497	0.052
500,50	0.0505	0.0515	0.052	0.0519	0.0534
500,100	0.0676	0.0526	0.0515	0.0522	0.0501
500,200	0.3673	0.0488	0.0493	0.049	0.0491
500,400	1	0.0497	0.0491	0.0496	0.0502

TABLE 6: Empirical sizes of the tests for $H_0 : \Sigma = \lambda I_p$, case $p \geqslant n$.

n, p	J	QLRT	T_{1s}	T_1	Z
25,500	0.054	0.0615	0.0537	0.0586	0.0604
50,500	0.0498	0.0629	0.049	0.051	0.0519
100,500	0.0519	0.1313	0.0481	0.0489	0.0511
200,500	0.0545	0.9593	0.0546	0.0543	0.0536
400,500	0.0499	1	0.0497	0.0504	0.052

4.2.2. Empirical Powers of the Tests

The results of the empirical powers of the tests by varying h, are presented in the tables 7 and 8. The figures 3 and 4 illustrate the behavior of the empirical powers of the tests for the cases $p < n$ and $p \geqslant n$. We observe the following:

1. For $p < n$ and $p \geq n$, the empirical powers of the tests are good when h increases, since they are equal or approximately one.
2. For the cases $p < n$ and $p \geq n$, the empirical powers of J, T_{1s}, T_1 and Z are very close between them that their graphs are almost indistinguishable. This may be due to the fact that the test statistics of J, T_{1s} and T_1 are based on estimators of the measure of sphericity $a_2/a_1^2 - 1$, with a_i given by expression (10), and the test statistic of the Z test is similar to the statistic of the J test.
3. For the case $n = 500$ and $p = 50$, the empirical powers of J, T_{1s}, T_1 and Z are bigger than the corresponding to LRT_2, however the empirical power of the last one is good when h increases.
4. When $n = 500$, $p = 200$ and $h = 0.5, 1$, LRT_2 has bigger empirical power than J, T_{1s}, T_2 and Z, however LRT_2 has a bad behavior in terms of the size of the test (see Table 5), hence it is not recommendable in this case.
5. When $p = 500$, $n = 50$ and $h \geq 3.8$, the empirical powers of J, T_{1s}, T_1 and Z are bigger than the corresponding to QLRT.
6. When $p = 500$ and $n = 200$, for all the considered values of h, QLRT has the biggest empirical power, however the size of this test is bad (see Table 6), therefore this test is not recommendable in this case.

5. Examples of Application

In this section we apply the tests for (1) and (2) in the high dimensional context ($p \geqslant n$) to two sets of DNA microarray data found in the literature. It is worth to mention that, in real applications with DNA microarray data, where there is an expert of the dataset that can suggest a form for the covariance matrix, frequently it is of interest to test $H_0 : \Sigma = \Sigma_0$ vs $H_1 : \Sigma \neq \Sigma_0$, for some specific covariance

Table 7: Empirical powers of the tests for $H_0 : \Sigma = \lambda I_d$, case $p < n$.

$n = 500$, $p = 50$					
h	LRT_2	J	T_{1s}	T_1	Z
0.3	0.0987	0.1161	0.1159	0.1158	0.1108
0.6	0.3323	0.5133	0.5133	0.5115	0.4691
0.9	0.7453	0.9423	0.9427	0.9423	0.9139
1.2	0.9649	0.9989	0.999	0.9989	0.9978
1.5	0.998	1	1	1	1
$n = 500$, $p = 200$					
h	LRT_2	J	T_{1s}	T_1	Z
0.5	0.4466	0.0921	0.0927	0.0931	0.0924
1	0.6174	0.3445	0.3441	0.3443	0.3309
1.5	0.8065	0.8287	0.8286	0.8287	0.8052
2	0.9349	0.9935	0.9934	0.9936	0.9905
2.5	0.9849	0.9999	0.9999	0.9999	0.9999

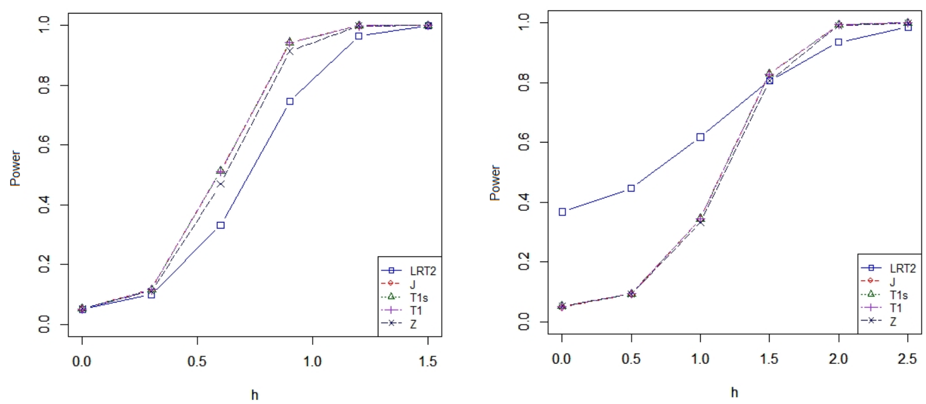
Table 8: Empirical powers of the tests for $H_0 : \Sigma = \lambda I_p$, case $p \geqslant n$.

$p = 500$, $n = 50$					
h	J	QLRT	T_{1s}	T_1	Z
1.9	0.0745	0.0841	0.0699	0.0744	0.0722
3.8	0.1865	0.1672	0.1813	0.1863	0.1716
5.7	0.4546	0.3466	0.4573	0.4582	0.4249
7.6	0.7424	0.5834	0.7436	0.7401	0.7031
9.5	0.908	0.782	0.9106	0.91	0.8834
$p = 500$, $n = 200$					
h	J	QLRT	T_{1s}	T_1	Z
1.2	0.0929	0.9714	0.092	0.0925	0.0923
2.4	0.319	0.9902	0.3151	0.3156	0.3083
3.6	0.7653	0.9979	0.7656	0.7653	0.7434
4.8	0.9768	0.9995	.9775	0.9767	0.9695
6	0.9992	1	0.9993	0.9994	0.999

matrix Σ_0 fixed by the expert. As mentioned in previous sections, to test the last null hypothesis we can transform the data and test (1) based on the transformed dataset. Since the objective of this work is not the study of some specific datasets, and we do not have an expert of the considered data found in the literature, we only consider the null hypotheses in (1) and (2) to illustrate the implementation of the tests and to observe their behavior.

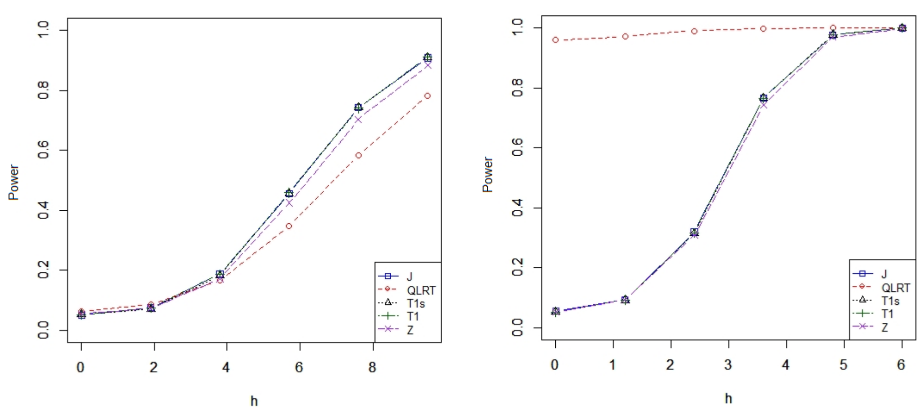
For the high dimensional context we considered the following tests: TW, J, LW, QLRT, CM, T_{1s}, T_{2s}, T_1, T_2 and Z. The p-value of each test was calculated. We took the significance level $\alpha = 0.05$.

Since all of the considered tests for the high dimensional context take a random sample from the $N_p(\mu, \Sigma)$, except QLRT that consider a random sample from the $N_p(0, \Sigma)$, for the last test we applied the transformation of multivariate normal



(a) $p = 50, n = 500$ (b) $p = 200, n = 500$

FIGURE 3: Empirical powers of the tests for $H_0 : \Sigma = \lambda I_p$, case $p < n$.



(c) $p = 500, n = 50$ (d) $p = 500, n = 200$

FIGURE 4: Empirical powers of the tests for $H_0 : \Sigma = \lambda I_p$, case $p \geqslant n$.

data given in Appendix A before the application of this test. Because the real datasets might not be scaled to have covariance matrix equal to the identity matrix, the datasets are scaled by $\widehat{\lambda}^{-1/2}$, where $\widehat{\lambda} = \mathrm{tr}(S_n)/p$, before the application of the tests for (1), as suggested in Ma (2012).

5.1. DLBCL Data

We consider the DNA microarray data of Rosenwald et al. (2002), which correspond to patients with diffuse large B-cell lymphoma (DLBCL). This dataset considers 7399 genes and 240 patients. The values of the test statistics for the hypothesis testing problems (1) and (2) are shown in the tables 9 and 10, respectively.

For each test in the tables 9 and 10 the value of the statistic was very large, and then the p-value was approximately zero. Due to the p-values of all the tests are approximately zero, we are strongly rejecting both null hypotheses. Therefore, considering, say, the significance level $\alpha = 0.05$, we have statistical evidence to reject $H_0 : \Sigma = I_p$ and $H_0 : \Sigma = \lambda I_p$ with all the considered tests.

Since the hypotheses $H_0 : \Sigma = I_p$ and $H_0 : \Sigma = \lambda I_p$ were rejected, we conclude that the population covariance matrix of the data has not these structures. This was expected, because it is known that there exists correlation between the genes of the same individual.

Table 9: Values of the test statistics for $H_0 : \Sigma = I_p$ considering the DLBCL data.

Test	Value of the test statistic
LW	80.19308
TW	2035.514
T_{2s}	5859.268
T_2	5820.955
CM	361737.7

Table 10: Values of the test statistics for $H_0 : \Sigma = \lambda I_p$ considering the DLBCL data.

Test	Value of the test statistic
J	231.2596
QLRT	5263.1241
T_{1s}	23836.63
T_1	23679.95
Z	181.0721

5.2. NCI60 Data

We now consider the NCI microarray data of Ross et al. (2000). The data contains expression levels on 6830 genes from 64 cancer cell lines. The values of the tests statistics for testing (1) and (2) are shown in the tables 11 and 12, respectively.

For each test in the tables 11 and 12 the value of the statistic was very large, and then the p-value was approximately zero. Due to the p-values of all the tests are approximately zero, we are strongly rejecting both null hypotheses. Therefore, considering, say, the significance level $\alpha = 0.05$, we have statistical evidence to reject $H_0 : \Sigma = I_p$ and $H_0 : \Sigma = \lambda I_p$ with all the considered tests.

Therefore, we conclude that the population covariance matrix of the data has not these structures. As in the last example, this conclusion was expected, because it is known that there exists correlation between the genes of the same individual.

Table 11: Values of the test statistics for $H_0 : \Sigma = I_p$ considering the NCI60 data.

Test	Value of the test statistic
LW	188.7283
TW	676.3819
T_{2s}	2491.531
T_2	2459.416
CM	514277.4

Table 12: Values of the test statistics for $H_0 : \Sigma = \lambda I_p$ considering the NCI60 data.

Test	Value of the test statistic
J	315.2971
QLRT	2988.593
T_{1s}	6417.741
T_1	6334.869
Z	176.9723

6. Conclusions

In this work we studied tests for the covariance matrix of multivariate Gaussian data. Our main interest was the case when the dimension of the data is greater than or equal to the sample size (high dimensional case), however we also studied tests for the case when the sample size is greater than the dimension of the data (classical case). We considered the null hypotheses $H_0 : \Sigma = I_p$ and $H_0 : \Sigma = \lambda I_p$.

The simulation study of this work to analyze the behavior of the tests for $H_0 : \Sigma = I_p$, indicates that in the considered settings, for the classical case ($p < n$) and the high dimensional case ($p \geq n$), the tests TW, CLRT, LW, CM, T_{2s} and T_2 have a good behavior in terms of the size of the test, since the empirical sizes were very close to the considered significance level. For the case $p < n$, LRT_1 had empirical sizes far away from the considered significance level when p was close to n, and the empirical sizes of CLRT have a better behavior than those of LRT_1. In terms of the power of the test, the TW test was superior to the other considered tests. Despite the empirical powers of the tests CLRT, LW, CM, T_{2s} and T_2 are smaller than that of TW, these tests are good alternatives for the considered settings. It is important to mention that although in some cases LRT_1 had empirical power bigger than the corresponding to the TW test, this test is not recommendable in those cases, since it has a bad behavior in terms of the size of the test. Therefore, for the considered settings, the TW test is a very good alternative for testing $H_0 : \Sigma = I_p$, in both cases, $p < n$ and $p \geq n$, since it had good results in terms of size and power of the test.

The simulations to evaluate the tests for $H_0 : \Sigma = \lambda I_p$, indicated that in the considered settings, for the classical and the high dimensional cases, the tests J, T_{1s}, T_1 and Z have a good behavior in terms of the size of the test, since the empirical sizes were very close to the considered significance level. Whereas, LRT_2 and QLRT in some cases had empirical sizes far away from the considered significance level. LRT_2 had a good behavior only when n is large enough with respect to p, and QLRT had a good behavior only when p is large enough with respect to n. The empirical powers of the tests J, T_{1s}, T_1 and Z had a good behavior for both cases, $p < n$ and $p \geq n$. The empirical powers of LRT_1 and QLRT in some cases were bigger than the corresponding to the other considered tests, however they had a bad behavior in terms of the size of the test, thus they are not recommendable in those cases. Therefore, for the considered settings, the tests J, T_{1s}, T_1 and Z are very good alternatives for testing $H_0 : \Sigma = \lambda I_p$ in both cases, $p < n$ and $p \geq n$, since they had good results in terms of size and power of the test.

The results obtained in this work are useful to have a better knowledge of the behavior of several tests of the literature for the covariance matrix of multivariate Gaussian data, by comparing the tests simultaneously in terms of size and power of the test.

On the other hand, we applied the tests to real data found in the literature. For the examples of application we considered two sets of DNA microarray data, and with all the tests we rejected the two null hypotheses $H_0 : \Sigma = I_p$ and $H_0 : \Sigma = \lambda I_p$, considering the significance level $\alpha = 0.05$. Therefore, we concluded that the covariance matrices of these datasets are not equal nor proportional to the identity matrix. This conclusion was expected, since it is well known that there exist high correlations between the genes of the same individual. These examples show the usefulness of the considered tests to analyze the structure of the covariance matrix of DNA microarray data.

Acknowledgments

The authors thank the Editor and the anonymous reviewers for their comments and valuable suggestions, which helped greatly to improve this manuscript.

[Received: September 2021 — Accepted: May 2022]

References

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc.

Bai, Z., Jiang, D., Yao, J. F. & Zheng, S. (2009), 'Corrections to LRT on large-dimensional covariance matrix by RMT', *The Annals of Statistics* **37**(6B), 3822–3840.

Cai, T. T. & Ma, Z. (2013), 'Optimal hypothesis testing for high dimensional covariance matrices', *Bernoulli* **19**(5B), 2359–2388.

Chen, S. X., Zhang, L.-X. & Zhong, P.-S. (2010), 'Tests for high-dimensional covariance matrices', *Journal of the American Statistical Association* **105**(490), 810–819.

Cortez-Elizalde, D. (2020), Pruebas de hipótesis para la matriz de covarianza poblacional de datos de dimensión alta, Tesis de Maestría, Universidad Juárez Autónoma de Tabasco, División Académica de Ciencias Básicas, Cunduacán, México.

Hallin, M. & Paindaveine, D. (2006), 'Semiparametrically efficient rank-based inference for shape. i. optimal rank-based tests for sphericity', *The annals of statistics* **34**(6), 2707–2756.

Hastie, T., Tibshirani, R. & Friedman, J. H. (2009), *Elements of statistical learning: data mining, inference, and prediction*, Springer.

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.

Johnstone, I. M. (2001), 'On the Distribution of the Largest Eigenvalue in Principal Components Analysis', *The Annals of Statistics* **29**(2), 295–327.

Ledoit, O. & Wolf, M. (2002), 'Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size', *The Annals of Statistics* **30**(4), 1081–1102.

Li, Z. & Yao, J. (2016), 'Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size', *Electronic Journal of Statistics* **10**(2), 2973–3010.

Ma, Z. (2012), 'Accuracy of the tracy–widom limits for the extreme eigenvalues in white wishart matrices', *Bernoulli* **18**(1), 322–359.

Muirhead, R. J. (2005), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc.

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

Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., Fisher, R. I. & et al. (2002), 'The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma', *New England Journal of Medicine* **346**(25), 1937–1947.

Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P. & et al. (2000), 'Systematic variation in gene expression patterns in human cancer cell lines', *Nature genetics* **24**(3), 227–235.

Srivastava, M. S. (2005), 'Some tests concerning the covariance matrix in high dimensional data', *Journal of the Japan Statistical Society* **35**(2), 251–272.

Srivastava, M. S., Yanagihara, H. & Kubokawa, T. (2014), 'Tests for covariance matrices in high dimension with less sample size', *Journal of Multivariate Analysis* **130**, 289–309.

Zou, C., Peng, L., Feng, L. & Wang, Z. (2014), 'Multivariate sign-based high-dimensional tests for sphericity', *Biometrika* **101**(1), 229–236.

Appendix A. Details of the Simulations

Transformation of Multivariate Normal Data

By Anderson (1984, pp. 76), if $X_1, X_2, \ldots, X_N$ is a random sample from the $N_p(\mu, \Sigma)$, we can obtain from it a random sample $Z_1, Z_2, \ldots, Z_n$, with $n = N - 1$, from the $N_p(0, \Sigma)$, such that the sample covariance matrix of the X_i's, S_n, satisfies $nS_n = \sum_{i=1}^{n} Z_i Z_i'$. The procedure to obtain the new random sample is the following:

1. Take an $N \times N$ orthogonal matrix $B = (b_{i,j})$ with the last row equal to $A' = (1/\sqrt{N}, 1/\sqrt{N}, \ldots, 1/\sqrt{N})'$. To do this, let $C = [e_1 \ e_2 \ \cdots \ e_{N-1}]$ be the matrix whose j-th column is the p-dimensional unit vector e_j, with one in the j-th entry and zeros in the rest, for $j = 1, 2, \ldots, N-1$. Define $D = C - AA'C$ and $E = O\Lambda^{-1/2}$, where O is an $(N-1) \times (N-1)$ orthogonal matrix and Λ is a $(N-1) \times (N-1)$ diagonal matrix such that $D'D = O\Lambda O'$. Define $\widetilde{B} = DE$. Then the matrix

$$B = \begin{bmatrix} \widetilde{B}' \\ A' \end{bmatrix}$$

is an $N \times N$ orthogonal matrix with the last row equal to A'.

2. Define $Z_i = \sum_{j=1}^{N} b_{i,j} X_j$, for $j = 1, 2, \ldots, N$. Then the Z_i's are independent, $Z_1, Z_1, \ldots, Z_{N-1}$ have distribution $N_p(0, \Sigma)$ and Z_N has distribution $N_p(\sqrt{N}\mu, \Sigma)$. Furthermore, $nS_n = \sum_{i=1}^{n} Z_i Z_i'$.

Algorithms for the Simulations

Now we present the general ideas of the algorithms used on the simulations to compute the empirical sizes and empirical powers of the tests.

Consider the tests for the hypothesis testing problem (1) (or (2)). For specific values of n and p, the following algorithm was used to compute the empirical sizes of the tests.

Algorithm 1. (Empirical sizes of the tests)

1. Generate a random sample of size $N = n + 1$ from the p-variate standard normal distribution.

2. For the tests that consider random samples from the $N_p(\mu, \Sigma)$, compute the test statistics with the original data. For the tests that consider random

samples from the $N_p(0, \Sigma)$, transform the original data to a random sample of size $n = N - 1$ from the $N_p(0, \Sigma)$, as described in the previous section, then compute the test statistics with the new random sample.

3. Taking the significance level $\alpha = 0.05$, record for each test whether the null hypothesis was rejected.
4. Repeat the last steps $M = 10,000$ times, and take the proportion of times that each test was rejected. These proportions are the empirical sizes of the tests.

As mentioned before, to evaluate the power of the tests we considered covariance matrices of the form

$$\Sigma = I_p + hvv', \tag{18}$$

where h is a positive scalar and v is a unit vector. The vector v is randomly generated and the scalar h varies in a range of values, say, in the interval $[0, r]$, for some $r > 0$. For specific values of n, p and r, the following algorithm was used to compute the empirical powers of the tests.

Algorithm 2. (Empirical powers of the tests)

1. Generate a random vector from the p-variate standard normal distribution, then compute its norm and divide by it each entry of the vector. Call the resulting unit vector $\widetilde{v}$.
2. Let $\mathbf{h} = (h_1, h_2, \ldots, h_5)'$, where $h_i = i * r/5$, for $i = 1, 2, \ldots, 5$.
3. For $i = 1$, generate a random sample of size $N = n + 1$ from the p-variate normal distribution with zero mean and covariance matrix (18), with $v = \widetilde{v}$ and $h = h_i$.
4. For the tests that consider random samples from the $N_p(\mu, \Sigma)$, compute the test statistics with the original data. For the tests that consider random samples from the $N_p(0, \Sigma)$, transform the original data to a random sample of size $n = N - 1$ from the $N_p(0, \Sigma)$, as described in the previous section, then compute the test statistics with the new random sample.
5. Taking the significance level $\alpha = 0.05$, record for each test whether the null hypothesis was rejected.
6. Repeat $M = 10,000$ times the steps 3–5, and take the proportion of times that each test was rejected. These proportions are the empirical powers of the tests for the covariance matrix (18), with $v = \widetilde{v}$ and $h = h_i$.
7. For $i = 2, 3, 4, 5$, repeat the steps 3–6 to obtain the empirical powers of the tests for the covariance matrix (18), with $v = \widetilde{v}$ and $h = h_i$, for $i = 2, 3, 4, 5$.