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\textsuperscript{a}, Addy Bolívar-Cime\textsuperscript{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

\textsuperscript{a}M.Sc. E-mail: dcortezelizalde@gmail.com
\textsuperscript{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 \( \mu \) and the covariance matrix \( \Sigma \) are unknown, and suppose that we are interested in testing

\[
H_0 : \Sigma = I_p \quad \text{vs} \quad H_1 : \Sigma \neq I_p,
\]

or

\[
H_0 : \Sigma = \lambda I_p \quad \text{vs} \quad H_1 : \Sigma \neq \lambda I_p,
\]

where \( \lambda \) 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

\[
\Lambda = \left( \frac{e}{N} \right)^{pN/2} \text{etr}(-A/2)(\det A)^{N/2},
\]
with $A = nS_n$, where $\text{etr}(A) := \exp(\text{tr}(A))$ (the exponential of the trace of the matrix $A$); and the likelihood ratio test for (2) is based on the *ellipticity statistic* given by

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

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, $nS_n = \sum_{i=1}^n Z_iZ_i'$; 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 $\bar{S}_n = \sum_{i=1}^n Z_iZ_i'/n$ 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 $\bar{X}$ 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 $\Sigma_0$ 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 $N_p(\Sigma_0^{-1/2}\mu, \Sigma_0^{-1/2}\Sigma_0^{-1/2})$, 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 $\alpha$ likelihood ratio test for (1) rejects $H_0$ if $\Lambda \leq c_\alpha$, where

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

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

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

Observe that this statistic is obtained from $\Lambda$ 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 $\Lambda^*$, or equivalently, of

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

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,

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

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

\[ L^* = \text{tr}S_n - \log(\det S_n) - p. \]

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 \( n \to \infty \)

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

where

\[ m(g) = -\frac{\log(1-y)}{2}, \]
\[ v(g) = -2\log(1-y) - 2y, \]
\[ F^{y_n}(g) = 1 - \frac{y_n - 1}{y_n} \log(1 - y_n), \]

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 \( \tilde{T}_n \) will be called the corrected likelihood ratio statistic.

A level \( \alpha \) test for (1) based on the statistic (5) rejects \( H_0 \) if \( \tilde{T}_n > z_\alpha \), where \( z_\alpha \) is the upper \( \alpha 	imes 100\% \) point of the standard normal distribution \( N(0,1) \). This test will be called the level \( \alpha \) 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 \( V = \text{tr}[(S_n - I_p)^2]/p \), 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)

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

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,

\[ nW - p \overset{d}{\to} \frac{2}{p} \chi^2_{p(p+1)/2} - p. \]

They also proved the following result.

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

\[ nW - p \overset{d}{\to} N(1, 4). \]

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

\[ \frac{l_1 - \mu_{np}}{\sigma_{np}} \overset{d}{\to} F_1, \]

where the center and scaling constants are

\[ \mu_{np} = \left( \sqrt{n - 1 + \sqrt{p}} \right)^2, \quad (6) \]
\[ \sigma_{np} = \left( \sqrt{n - 1 + \sqrt{p}} \right) \left( \frac{1}{\sqrt{n - 1}} + \frac{1}{\sqrt{p}} \right)^{1/3}, \quad (7) \]

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 \( \alpha \) 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 \( \mu_{np} \) and \( \sigma_{np} \) are the center and scaling constants of Theorem 3. We called this test the level \( \alpha \) 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) propose 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 CLR given in section 2.2, when \( p < n \) and \( p/n \to y \in (0, 1) \).

Let \( X_1, \ldots, X_n \) 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

\[
H_1 : \Sigma \in \Theta, \quad \Theta = \{ \Sigma : \| \Sigma - I_p \|_F \geq \epsilon_n \},
\]

where \( \epsilon_n > 0 \) and

\[
\| A \|_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\text{tr}(AA^H)},
\]

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 \( \epsilon_n \); the smaller \( \epsilon_n \) is, the harder it is to distinguish between the two hypotheses.

Define the statistic

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

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 \( \| \Sigma - I_p \|_F^2 = \text{tr}[(\Sigma - I_p)^2] \). 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
\[ \text{tr}(\Sigma^2) \to \infty \quad \text{and} \quad \text{tr}(\Sigma^4)/\text{tr}^2(\Sigma^2) \to 0 \]
as $n \to \infty$, then
\[ \frac{T_n - \mu_n(\Sigma)}{\sigma_n(\Sigma)} \xrightarrow{d} N(0, 1), \]
where
\[ \mu_n(\Sigma) = \mathbb{E}_\Sigma(T_n) = \text{tr}(\Sigma - I_p)^2, \]
\[ \sigma_n^2(\Sigma) = \text{var}_\Sigma(T_n) = \frac{4}{n(n - 1)}(\text{tr}^2(\Sigma^2) + \text{tr}(\Sigma^4)) + \frac{8}{n} \text{tr}(\Sigma^2(\Sigma - I_p)^2). \]

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 $\alpha$ test for (1) based on the statistic \( T_n \) rejects \( H_0 \) if \( T_n > z_\alpha\sqrt{\frac{p(p + 1)}{n(n - 1)}} \), where \( z_\alpha \) is the upper $\alpha \times 100\%$ point of the standard normal distribution. We called this test the level $\alpha$ 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 $\Sigma$ for testing (1). Specifically, they considered estimators of the squared Frobenius norm (divided by $p$)
\[ \frac{1}{p} \text{tr}[(\Sigma - I_p)^2] = \frac{1}{p}[\text{tr}(\Sigma^2) - 2\text{tr}(\Sigma) + p] = a_2 - 2a_1 + 1, \quad (9) \]
where
\[ a_i = \frac{\text{tr}\Sigma^i}{p}, \quad i = 1, 2, \ldots, \quad (10) \]

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
\[ a) \text{If } p \to \infty, \text{then } a_i \to a_i^0, \quad 0 < a_i^0 < \infty, \quad i = 1, \ldots, 8, \]
\[ b) n = O(p^\delta), \quad 0 < \delta \leq 1, \quad (11) \]
where \( n = O(p^\delta) \) denotes that \( n/p^\delta \) 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

\[
\tilde{a}_1 = \frac{\text{tr}(S_n)}{p}, \quad \tilde{a}_2 = \frac{n^2}{(n-1)(n+2)p} \left[ \text{tr}(S_n^2) - \frac{1}{n} \left( \text{tr}(S_n) \right)^2 \right].
\]

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

\[
T_{2s} = \frac{n}{2} \left( \tilde{a}_{2s} - 2\tilde{a}_1 + 1 \right) \overset{d}{\to} N(0,1).
\]

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

\[
\tilde{a}_2 = \frac{1}{f} \left[ (N-2)\text{tr}(M^2) - N\text{tr}(D^2) + (\text{tr}D)^2 \right],
\]

where \( f = pN(N-1)(N-2)(N-3) \), \( M = Y'Y \), \( Y = (Y_1, \ldots, Y_N) \), \( D = \text{diag}(Y'_1Y_1, \ldots, Y'_NY_N) \), with \( Y_i = X_i - \bar{X} \). 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

\[
T_2 = \frac{n}{2} \left( \tilde{a}_2 - 2\tilde{a}_1 + 1 \right) \overset{d}{\to} N(0,1).
\]

Therefore, a level \( \alpha \) test for \( (1) \) based on \( T_2 \) rejects \( H_0 \) if \( T_2 > z_\alpha \), where \( z_\alpha \) is the upper \( \alpha \times 100\% \) point of the standard normal distribution. We called this test the level \( \alpha \) 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 \( \Sigma_0 \) is a specific known positive definite covariance matrix and \( \lambda \) 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 $\alpha$ likelihood ratio test for (2) rejects $H_0$ if

$$ V \equiv \frac{\det A}{(\text{tr}(A)/p)^p} = \frac{\det S_n}{(\text{tr}(S_n)/p)^p} \leq k_\alpha, \quad (13) $$

where $A = nS_n$ and $k_\alpha$ 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,

$$ \Pr(-n\rho \log V \leq x) \approx \Pr(\chi^2_{f} \leq x), \quad \forall x \in \mathbb{R}. \quad (14) $$

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

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

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 $\alpha$ test for (2) based on the John’s statistic rejects $H_0$ if $(nU - p - 1)/2 > z_\alpha$, where $z_\alpha$ is the upper $\alpha \times 100\%$ point of the standard normal distribution. This test will be called the level $\alpha$ 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 $X = (X_1, \ldots, X_n)$ 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 $p \leq n$ 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 $XX'/p$ which has exactly the same $n$ non-zero eigenvalues with the matrix $S_n = XX'/n$ (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 $\ell_i$, $i = 1, 2, \ldots, p$, are the eigenvalues of the matrix $S_n$. Based on the last expression, Li & Yao (2016) proposed the quasi-likelihood ratio statistic given by

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

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

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

$$L_n - \frac{n}{2} - \frac{n^2}{6p} - \frac{v_4 - 2}{2} \frac{2}{d} \to N(0,1),$$

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

Thus, a level $\alpha$ test for (2) based on the statistic $L_n$ rejects $H_0$ if $L_n - \frac{n}{2} - \frac{n^2}{6p} - \frac{v_4 - 2}{2} > z_\alpha$, where $z_\alpha$ is the upper $\alpha \times 100\%$ point of the standard normal distribution. We called this test the level $\alpha$ 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 \( \lambda \); where the \( \lambda_i \)'s are the eigenvalues of \( \Sigma \). Thus, a measure of sphericity is given by

\[
\frac{a_2}{a_1^2} - 1,
\]

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

\[
T_{1s} = \frac{n}{2} \left( \frac{\hat{a}_2}{\hat{a}_1^2} - 1 \right).
\]

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

\[
T_{1s} \overset{d}{\to} N(0, 1).
\]

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

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

\[
T_1 = \frac{n}{2} \left( \frac{\hat{a}_2}{\hat{a}_1^2} - 1 \right).
\]

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

\[
T_1 \overset{d}{\to} N(0, 1).
\]

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

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

where $\|X\| = (X'X)^{1/2}$ is the Euclidean norm of the vector $X$, $\theta_p$ is the symmetry centre, $\Sigma_p$ is a positive-definite symmetric $p \times p$ scatter matrix, and $g_p$ is a non-negative function of a real variable. The matrix $\Sigma_p$ 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

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

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 $\hat{\theta}_{n,p} = \bar{X}$. Consider the statistic

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

where $\hat{U}_i = U(X_i - \bar{X})$. 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 $\hat{Q}$.

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

$$\frac{\hat{Q} - p\delta_{n,p}}{\sigma_0} \xrightarrow{d} N(0,1)$$

Revista Colombiana de Estadística - Applied Statistics 45 (2022) 373–399
as \( n, p \to \infty \), where \( \tilde{\sigma}_0^2 = 4(p - 1)/[n(n - 1)(p + 2)] \) and

\[
\delta_{n,p} = \frac{1}{n^2} \left( 2 - \frac{2E(R_i^{-2})}{E(R_i^{-1})^2} + \left[ \frac{E(R_i^{-2})}{E(R_i^{-1})^2} \right]^2 \right) + \frac{1}{n^3} \left[ 8E(R_i^{-2})/E(R_i^{-1})^2 - 6 \left( \frac{E(R_i^{-2})}{E(R_i^{-1})^2} \right)^2 + \frac{2E(R_i^{-2})E(R_i^{-3})}{E(R_i^{-1})^3} - \frac{E(2R_i^{-3})}{E(R_i^{-1})^3} \right].
\]

(17)

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

\[
\hat{R}_i = \|X_i - \hat{\theta}_{n,p}\|, \quad \bar{R}_{ix} = \hat{R}_i + \hat{\theta}_{n,p}^* \hat{U}_i - 2^{-1} \hat{R}_i^{-1} \|\hat{\theta}_{n,p}\|^2.
\]

Thus, substituting

\[
E(R_i^{-k})/E(R_i^{-1})^k
\]

by

\[
\frac{n^{k-1} \sum_{i=1}^{n} \bar{R}_{ix}^{-k}}{(\sum_{i=1}^{n} \bar{R}_{ix}^{-1})^k}
\]

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

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

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

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

<table>
<thead>
<tr>
<th>$n, p$</th>
<th>TW</th>
<th>LW</th>
<th>CM</th>
<th>$T_{2s}$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25,500</td>
<td>0.0503</td>
<td>0.0555</td>
<td>0.0537</td>
<td>0.0541</td>
<td>0.0591</td>
</tr>
<tr>
<td>50,500</td>
<td>0.0494</td>
<td>0.0504</td>
<td>0.0509</td>
<td>0.0491</td>
<td>0.0514</td>
</tr>
<tr>
<td>100,500</td>
<td>0.0526</td>
<td>0.0527</td>
<td>0.0483</td>
<td>0.0485</td>
<td>0.0491</td>
</tr>
<tr>
<td>200,500</td>
<td>0.051</td>
<td>0.0539</td>
<td>0.0547</td>
<td>0.0546</td>
<td>0.0546</td>
</tr>
<tr>
<td>400,500</td>
<td>0.0528</td>
<td>0.0501</td>
<td>0.0546</td>
<td>0.0498</td>
<td>0.0508</td>
</tr>
</tbody>
</table>

The empirical powers of the tests for the classical case ($p < n$) and the high dimensional case ($p \geq n$), 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 $\| \Sigma - I_p \|^2_F = \text{tr}(\Sigma - I_p)^2$.

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 $p < n$.

\begin{tabular}{cccccccc}
\hline
 & $n=500$, $p=50$ & & & & & & \\
\hline
$h$ & $LRT_1$ & TW & CLRT & LW & CM & $T_{2s}$ & $T_2$ \\
\hline
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 & 0.7331 & 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 & \\
\hline
\end{tabular}

\begin{tabular}{cccccccc}
\hline
 & $n=500$, $p=200$ & & & & & & \\
\hline
$h$ & $LRT_1$ & TW & CLRT & LW & CM & $T_{2s}$ & $T_2$ \\
\hline
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 \\
\hline
\end{tabular}

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

\begin{tabular}{cccccc}
\hline
 & $p=500$, $n=50$ & & & & & \\
\hline
$h$ & TW & LW & CM & $T_{2s}$ & $T_2$ \\
\hline
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 & 0.4658 & 0.4686 & 0.4716 \\
7.6 & 0.9577 & 0.7576 & 0.745 & 0.7551 & 0.7514 \\
9.5 & 0.9932 & 0.9339 & 0.9096 & 0.916 & 0.9148 \\
\hline
\end{tabular}

\begin{tabular}{cccccc}
\hline
 & $p=500$, $n=200$ & & & & & \\
\hline
$h$ & TW & LW & CM & $T_{2s}$ & $T_2$ \\
\hline
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 \\
\hline
\end{tabular}

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 $p \geq n$, 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 $n$. 

Revista Colombiana de Estadística - Applied Statistics 45 (2022) 373–399
Figure 1: Empirical powers of the tests for $H_0 : \Sigma = I_p$, case $p < n$.

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

3. For $p \geq n$, 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 $p < n$.

<table>
<thead>
<tr>
<th>$n, p$</th>
<th>$LRT_2$</th>
<th>$J$</th>
<th>$T_{1a}$</th>
<th>$T_1$</th>
<th>$Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500, 25</td>
<td>0.0475</td>
<td>0.0493</td>
<td>0.0499</td>
<td>0.0497</td>
<td>0.052</td>
</tr>
<tr>
<td>500, 50</td>
<td>0.0505</td>
<td>0.0515</td>
<td>0.052</td>
<td>0.0519</td>
<td>0.0534</td>
</tr>
<tr>
<td>500, 100</td>
<td>0.0576</td>
<td>0.0526</td>
<td>0.0515</td>
<td>0.0522</td>
<td>0.0501</td>
</tr>
<tr>
<td>500, 200</td>
<td>0.3673</td>
<td>0.0488</td>
<td>0.0493</td>
<td>0.049</td>
<td>0.0491</td>
</tr>
<tr>
<td>500, 400</td>
<td>1</td>
<td>0.0497</td>
<td>0.0491</td>
<td>0.0496</td>
<td>0.0502</td>
</tr>
</tbody>
</table>
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 $p \geq n$. 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 ($p \geq n$) 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 $p < n$.

<table>
<thead>
<tr>
<th>$n = 500$, $p = 50$</th>
<th>$h$</th>
<th>$LRT_2$</th>
<th>$J$</th>
<th>$T_{1s}$</th>
<th>$T_1$</th>
<th>$Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.0987</td>
<td>0.1161</td>
<td>0.1159</td>
<td>0.1158</td>
<td>0.1108</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.3323</td>
<td>0.5133</td>
<td>0.5133</td>
<td>0.5115</td>
<td>0.4691</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.7453</td>
<td>0.9423</td>
<td>0.9427</td>
<td>0.9423</td>
<td>0.9139</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>0.9649</td>
<td>0.9989</td>
<td>0.999</td>
<td>0.9989</td>
<td>0.9978</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>0.998</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 500$, $p = 200$</th>
<th>$h$</th>
<th>$LRT_2$</th>
<th>$J$</th>
<th>$T_{1s}$</th>
<th>$T_1$</th>
<th>$Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.4466</td>
<td>0.0921</td>
<td>0.0927</td>
<td>0.0931</td>
<td>0.0924</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.6174</td>
<td>0.3445</td>
<td>0.3441</td>
<td>0.3443</td>
<td>0.3309</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>0.8065</td>
<td>0.8287</td>
<td>0.8286</td>
<td>0.8287</td>
<td>0.8052</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.9349</td>
<td>0.9935</td>
<td>0.9934</td>
<td>0.9936</td>
<td>0.9905</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>0.9849</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$p = 500$, $n = 50$</th>
<th>$h$</th>
<th>$J$</th>
<th>QLR $T_1$</th>
<th>$T_{1s}$</th>
<th>$T_1$</th>
<th>$Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9</td>
<td>0.0745</td>
<td>0.0841</td>
<td>0.0699</td>
<td>0.0744</td>
<td>0.0722</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>0.1865</td>
<td>0.1672</td>
<td>0.1813</td>
<td>0.1863</td>
<td>0.1716</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>0.4546</td>
<td>0.3466</td>
<td>0.4573</td>
<td>0.4582</td>
<td>0.4249</td>
<td></td>
</tr>
<tr>
<td>7.6</td>
<td>0.7424</td>
<td>0.5834</td>
<td>0.7436</td>
<td>0.7401</td>
<td>0.7031</td>
<td></td>
</tr>
<tr>
<td>9.5</td>
<td>0.908</td>
<td>0.782</td>
<td>0.9106</td>
<td>0.91</td>
<td>0.8834</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p = 500$, $n = 200$</th>
<th>$h$</th>
<th>$J$</th>
<th>QLR $T_1$</th>
<th>$T_{1s}$</th>
<th>$T_1$</th>
<th>$Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>0.0929</td>
<td>0.9714</td>
<td>0.092</td>
<td>0.0925</td>
<td>0.0923</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>0.319</td>
<td>0.9902</td>
<td>0.3151</td>
<td>0.3156</td>
<td>0.3083</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>0.7653</td>
<td>0.9979</td>
<td>0.7656</td>
<td>0.7653</td>
<td>0.7434</td>
<td></td>
</tr>
<tr>
<td>4.8</td>
<td>0.9768</td>
<td>0.9995</td>
<td>0.9775</td>
<td>0.9767</td>
<td>0.9695</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.9992</td>
<td>1</td>
<td>0.9993</td>
<td>0.9994</td>
<td>0.999</td>
<td></td>
</tr>
</tbody>
</table>

matrix $\Sigma_0$ 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
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 \( \hat{\lambda}^{-1/2} \), where \( \hat{\lambda} = \text{tr}(S_n)/p \), 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.

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

<table>
<thead>
<tr>
<th>Test</th>
<th>Value of the test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>LW</td>
<td>80.13608</td>
</tr>
<tr>
<td>TW</td>
<td>2005.514</td>
</tr>
<tr>
<td>$T_2$s</td>
<td>5859.268</td>
</tr>
<tr>
<td>$T_2$</td>
<td>5826.955</td>
</tr>
<tr>
<td>CM</td>
<td>361737.7</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Test</th>
<th>Value of the test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>231.2596</td>
</tr>
<tr>
<td>QLRT</td>
<td>5263.1241</td>
</tr>
<tr>
<td>$T_1$s</td>
<td>23836.63</td>
</tr>
<tr>
<td>$T_1$</td>
<td>23679.95</td>
</tr>
<tr>
<td>Z</td>
<td>181.0721</td>
</tr>
</tbody>
</table>

\textbf{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.

<table>
<thead>
<tr>
<th>Test</th>
<th>Value of the test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>LW</td>
<td>188.7283</td>
</tr>
<tr>
<td>TW</td>
<td>676.3819</td>
</tr>
<tr>
<td>$T_{2s}$</td>
<td>2491.531</td>
</tr>
<tr>
<td>$T_2$</td>
<td>2459.416</td>
</tr>
<tr>
<td>CM</td>
<td>514277.4</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Test</th>
<th>Value of the test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>315.2971</td>
</tr>
<tr>
<td>QLR T</td>
<td>2988.593</td>
</tr>
<tr>
<td>$T_{1s}$</td>
<td>6417.741</td>
</tr>
<tr>
<td>$T_1$</td>
<td>6334.869</td>
</tr>
<tr>
<td>Z</td>
<td>176.9723</td>
</tr>
</tbody>
</table>

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 QLR T 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


**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 \( nS_n = \sum_{i=1}^n Z_iZ'_i \). 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 \( \Lambda \) is a \((N-1) \times (N-1)\) diagonal matrix such that \( D'D = O\Lambda O' \). Define \( \tilde{B} = DE \). Then the matrix

\[
B = \begin{bmatrix}
\tilde{B}' \\
A'
\end{bmatrix}
\]

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, \( Z_1, Z_2, \ldots, Z_{N-1} \) have distribution \( N_p(0, \Sigma) \) and \( Z_N \) has distribution \( N_p(\sqrt{N}\mu, \Sigma) \). Furthermore, \( nS_n = \sum_{i=1}^n Z_iZ'_i \).

**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

\[
\Sigma = I_p + hvv',
\]

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 \( \bar{v} \).

2. Let \( h = (h_1, h_2, \ldots, h_5)' \), where \( h_i = i \cdot r/5 \), 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 \( v = \bar{v} \) 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 \( v = \bar{v} \) 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 \( v = \bar{v} \) and \( h = h_i \), for \( i = 2, 3, 4, 5 \).