Journal of Multivariate Analysis 102 (2011) 1090–1103

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Some tests for the covariance matrix with fewer observations than the dimension under non-normality

Muni S. Srivastava [a], Tõnu Kollo [b], Dietrich von Rosen [c,*]

[a] Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario, M5S 3G3, Canada

[b] Mathematical Statistics, University of Tartu, J. Liivi 2-513 50409, Tartu, Estonia

[c] Department of Energy and Technology, Swedish University of Agricultural Sciences, Box 7032, SE-750 07 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 2 October 2009
Available online 21 March 2011

AMS subject classification:
62H15
62H10

ABSTRACT

This article analyzes whether some existing tests for the $p \times p$ covariance matrix Σ of the N independent identically distributed observation vectors work under non-normality. We focus on three hypotheses testing problems: (1) testing for sphericity, that is, the covariance matrix Σ is proportional to an identity matrix I_p; (2) the covariance matrix Σ is an identity matrix I_p; and (3) the covariance matrix is a diagonal matrix. It is shown that the tests proposed by Srivastava (2005) for the above three problems are robust under the non-normality assumption made in this article irrespective of whether $N \leq p$ or $N \geq p$, but $(N, p) \to \infty$, and N/p may go to zero or infinity. Results are asymptotic and it may be noted that they may not hold for finite (N, p).

1. Introduction

Quantitative measurements of thousands of genes' expressions are obtained through DNA microarrays. Since these observations on the genes are on the same subject, they are not independently distributed. Thus, if there are measurements on p genes, it has a $p \times p$ covariance matrix Σ. The number of subjects on which these measurements are obtained, say N, are often very few, that is $N \ll p$. The analysis of such data sets requires new developments of multivariate theory, many of which have recently been obtained in the literature. The analysis is, however, simplified considerably if the $p \times p$ covariance matrix Σ satisfies either of the following three hypotheses:

$$H_1 : \Sigma = \lambda I_p, \quad \lambda > 0,$$
$$H_2 : \Sigma = I_p,$$
$$H_3 : \Sigma = D = \operatorname{diag}(d_1, \dots, d_p),$$

where D is a $p \times p$ diagonal matrix with diagonal elements $d_1, \dots, d_p$. For example, if either the hypothesis H_1 or H_2 holds, then most of the univariate results can be used to analyze the data. If the hypothesis H_3 holds, then a standardized version of the univariate test statistics can be used. In microarray data analysis of genes, it is invariably assumed, implicitly or explicitly, that the genes are independently distributed to carry out the analysis; that is, the analysis is carried out under the hypothesis H_3. The frequently applied false discovery rate (FDR) of the Benjamini and Hochberg [1] procedure can be controlled at the specified level only if the hypothesis H_3 is true, or if the covariance matrix Σ is of the intraclass correlation structure with positive correlation provided that the data are normally distributed; but so far no satisfactory test is available for testing

* Corresponding author.
E-mail addresses: srivasta@utstat.toronto.edu (M.S. Srivastava), Tonu.Kollo@ut.ee (T. Kollo), Dietrich.von.Rosen@bt.slu.se, dietrich.von.rosen@slu.se (D. von Rosen).

0047-259X/$ – see front matter © 2011 Published by Elsevier Inc.
doi:10.1016/j.jmva.2011.03.003

the intraclass correlation structure when $N \leq p$. Since $N \ll p$, it is not known how to ascertain the multivariate normality of the data. Thus, it would be desirable to have tests for which the significance levels can be controlled with or without the assumption of normality of the data; that is, to have robust tests.

When p is finite and N is large, it may not be important or necessary to obtain robust tests as the level of significance can be maintained at the specified level by using the bootstrap methods of Beran and Srivastava [2], Nagao and Srivastava [12] for the covariance matrix. For this reason, most studies considered selecting a test that has better power among the available tests. For example, Chan and Srivastava [4], and Nagao and Srivastava [12] compared the power of the LRT with that of LBIT defined in Section 4 for testing sphericity. Further details and references concerning the tests H_1–H_3 are given in Sections 4–6. It may be noted that when $N/p \to$ constant, the testing problems H_1 and H_2 have been considered by Ledoit and Wolf [8] and the problem H_3 by Schott [14]. Robustness of these tests has yet to be considered.

For $N \leq p$ and both N and p going to infinity, bootstrap theory is not yet available. Thus, it is desirable to obtain robust tests for this situation. Our objective in this paper is to show that the tests proposed by Srivastava [16] are robust for the model described below.

It is assumed that the p-dimensional observation vectors $\mathbf{x}_1, \ldots, \mathbf{x}_N$ on N subjects are independently identically distributed (*i.i.d.*) with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma = CC'$, where C is a $p \times p$ non-singular matrix, that is Σ is a positive definite (p.d.) matrix. Moreover, we shall assume that the N *i.i.d.* observation vectors $\mathbf{x}_i$ of dimension p can be written as

$$\mathbf{x}_i = \boldsymbol{\mu} + C\mathbf{z}_i, \tag{1.1}$$
$$E(\mathbf{z}_i) = \mathbf{0}, \qquad \mathrm{Cov}(\mathbf{z}_i) = I_p, \quad i = 1, \ldots, N.$$

For testing the hypothesis H_3 of diagonality of the covariance matrix Σ, we shall, however, assume that under H_3, $C = \mathrm{diag}(d_1^{1/2}, \ldots, d_p^{1/2}) = D^{1/2}$.

Instead of normality of $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})$, $i = 1, \ldots, N$, we shall assume that not only that $\mathbf{z}_i$ are *i.i.d.*, but that z_{ij} are *i.i.d.* for all i and j with

$$E(z_{ij}^r) = \gamma_r, \quad r = 3, \ldots, 8, \text{ with } \gamma_4 = \gamma. \tag{1.2}$$

Under normality, $\gamma_3 = \gamma_5 = \gamma_7 = 0$, $\gamma = 3$, $\gamma_6 = 15$, and $\gamma_8 = 105$. Unbiased estimators of $\boldsymbol{\mu}$ and Σ are respectively given by

$$\bar{\mathbf{x}} = N^{-1}\sum_{i=1}^{N}\mathbf{x}_i \quad \text{and} \quad S = \frac{1}{n}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})', \quad n = N - 1. \tag{1.3}$$

When $N \leq p$, the sample covariance matrix S is singular and no likelihood ratio test (LRT) is available for any of the three hypotheses. Thus, we consider the following tests proposed by Srivastava [16] for the hypotheses H_1, H_2, H_3. Let

$$\hat{\delta}_1 = \mathrm{tr}S/p, \qquad \hat{\delta}_2 = c_n[\mathrm{tr}S^2 - n^{-1}(\mathrm{tr}S)^2]/p, \tag{1.4}$$

$$\hat{\delta}_{20} = c_n\sum_{i=1}^{p} s_{ii}^2/p, \quad \text{and} \quad \hat{\delta}_{40} = \sum_{i=1}^{p} s_{ii}^4/p, \quad S = (s_{ij}), \tag{1.5}$$

where

$$c_n = \frac{n^2}{(n-1)(n+2)}. \tag{1.6}$$

Then for testing the hypothesis H_1, known in the literature as the 'Sphericity' hypothesis, we consider the test statistic given by

$$T_1 = \frac{\hat{\delta}_2}{\hat{\delta}_1^2} - 1; \tag{1.7}$$

for the hypothesis H_2, the test statistic is given by

$$T_2 = \hat{\delta}_2 - 2\hat{\delta}_1 + 1; \tag{1.8}$$

and for the hypothesis H_3, the test statistic is given by

$$T_3 = \frac{\left(\hat{\delta}_2/\hat{\delta}_{20}\right) - 1}{\left(1 - \frac{1}{p}\left(\hat{\delta}_{40}/\hat{\delta}_{20}^2\right)\right)^{1/2}}. \tag{1.9}$$

Let

$$\delta_i = \frac{1}{p}\mathrm{tr}\Sigma^i, \quad i = 1, \ldots, 4, \qquad \delta_{20} = \frac{1}{p}\sum_{i=1}^{p}\sigma_{ii}^2, \qquad \delta_{40} = \frac{1}{p}\sum_{i=1}^{p}\sigma_{ii}^4. \tag{1.10}$$

We make the following assumption for the consistency of the statistics T_1, T_2, and T_3.

Assumption A. As $p \to \infty$, $\delta_2 \to \delta_2^o$, $p^{-1}\delta_4 \to 0$, and $\gamma = 3 + O(p^{-\epsilon})$, $\epsilon > 0$.

Under Assumption A, it is shown that $\hat{\delta}_1$ and $\hat{\delta}_2$ are consistent estimators of δ_1 and δ_2 as $(N, p) \to \infty$.

Next, we state the approximative distributions of the test statistics T_1, T_2, and T_3 under the null hypotheses when (N, p) is large. Moreover, we suppose that all presented matrix manipulations are valid when p (n) goes to ∞. The theorems will be proved in the subsequent sections. Let $\Phi(\cdot)$ denote the cdf of a standard normal random variable, $N(0, 1)$, and P_0 denote the distribution under the null hypotheses H_1, H_2, H_3, respectively, for the three test statistics.

Theorem 1.1. *Under the model (1.1)–(1.2) and Assumption A, for large (N, p),*

$$P_0((n/2)T_1 \le t_1) \approx \Phi(t_1),$$

where $\Phi(\cdot)$ denotes the cdf of a standard normal random variable, $N(0, 1)$, and P_0 denotes the distribution under the hypothesis H_1.

Theorem 1.2. *Under the model (1.1)–(1.2) and Assumption A, for large (N, p),*

$$P_0((n/2)T_2 \le t_2) \approx \Phi(t_2),$$

where P_0 denotes the distribution under the hypothesis H_2.

Theorem 1.3. *Under the model (1.1)–(1.2) and Assumption A, for large (N, p),*

$$P_0((n/2)T_3 \le t_3) \approx \Phi(t_3),$$

where P_0 denotes the distribution under the hypothesis H_3.

The approximative distributions for $T_1 - T_3$ which are presented in Theorems 1.1–1.3 are the same as those obtained under normality assumption in [16]. Thus, the tests based on T_1, T_2 or T_3 are robust tests.

To obtain the distribution of the test statistic T_1 and T_2, we need to obtain the joint distribution of $\hat{\delta}_1$ and $\hat{\delta}_2$ under the model (1.1)–(1.2). To prove robustness, we need only obtain the joint distribution of $\hat{\delta}_1$ and $\hat{\delta}_2$ under the null hypotheses H_1 and H_2. Since the statistic T_1 is invariant under the scalar transformation $\mathbf{x}_i \to c\mathbf{x}_i$, $c \ne 0$, we shall assume without loss of generality that $\lambda = 1$. Thus, the results of the following theorem are applicable to both the statistics T_1 and T_2.

Theorem 1.4. *Let (1.1), (1.2), and $\Sigma = I_p$ hold. Then, the joint distribution of $\hat{\delta}_1$ and $\hat{\delta}_2$ displayed in (1.4), for (N, p) large, is approximately given by*

$$(np)^{1/2}\Omega^{-1/2}\begin{pmatrix}\hat{\delta}_1 - 1\\ \hat{\delta}_2 - 1\end{pmatrix} \approx N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, I_2\right),$$

where

$$\Omega = \begin{pmatrix}\gamma - 1 & 2(\gamma - 1)\\ 2(\gamma - 1) & 4(\gamma - 1) + 4\dfrac{p}{n}\end{pmatrix} = \Omega^{1/2}\Omega^{1/2}, \tag{1.11}$$

and I_2 is the identity matrix of size 2×2.

Note that

$$(np)^{-1}\Omega = \begin{pmatrix}\dfrac{\gamma - 1}{np} & \dfrac{2(\gamma - 1)}{np}\\ \dfrac{2(\gamma - 1)}{np} & \dfrac{4(\gamma - 1)}{np} + \dfrac{4}{n^2}\end{pmatrix},$$

is the asymptotic covariance matrix of $(\widehat{\delta}_1, \widehat{\delta}_2)$ which exists for all values of n and p without any condition on n and p, i.e. it goes to zero as $(n, p) \to \infty$, proving the consistency of $\widehat{\delta}_1$ and $\widehat{\delta}_2$. It may also be noted that all the above three tests are robust when $\gamma = 3 + O(p^{-\epsilon})$, $\epsilon > 0$.

The organization of the paper is as follows. In Section 2, we give some preliminary results needed to prove Theorem 1.4, which is proven in Section 3. The proofs of Theorems 1.1–1.3 are given in Sections 4–6, respectively. In particular, in Section 6 some of the notion and ideas of Section 2 will be repeated but now it is focused on T_3 instead of T_1 and T_2.

2. Some preliminary results

In this section, we present some preliminary results. We first comment on the constant $c_n = n^2/(n-1)(n+2)$ multiplied to the random variable $1/p(\mathrm{tr}S^2 - n^{-1}(\mathrm{tr}S)^2)$ in (1.4) and to $1/p\sum_i s_{ii}^2$ in (1.5). Under normality assumption, we get from [16, p. 261]

$$E\left[\frac{1}{p}(\mathrm{tr}S^2 - n^{-1}(\mathrm{tr}S)^2) - \delta_2\right] = \frac{(n-1)(n+2)}{n^2}\delta_2 - \delta_2 = (n^{-1} - 2n^{-2})\delta_2.$$

Thus,

$$nE\left[\frac{1}{p}(\mathrm{tr}S^2 - n^{-1}(\mathrm{tr}S)^2) - \delta_2\right] = \delta_2 + O(n^{-1})$$

which goes to δ_2, a constant, as (N, p) becomes large. That is, the bias does not go to zero and asymptotic normality cannot hold. On the other hand,

$$nE\left[\frac{c_n}{p}(\mathrm{tr}S^2 - n^{-1}(\mathrm{tr}S)^2) - \delta_2\right] = 0,$$

i.e., the bias is zero, and asymptotic normality has been shown in [16].

Now, we consider the model given in (1.1) and (1.2) under $\Sigma = I$. Let

$$G = I_N - \frac{1}{N}\mathbf{1}\mathbf{1}',$$

where $\mathbf{1} = (1, \dots, 1)'$ is an N-vector of ones. Then, since $\Sigma = I_p$ under H_1, we may write S as

$$S = \frac{1}{n}ZGZ', \quad Z' = (\mathbf{w}_1, \dots, \mathbf{w}_p) : N \times p,$$

where $\mathbf{w}_i$ are i.i.d. $N_N(\mathbf{0}, I_N)$, $i = 1, \dots, p$. Note that for $G = (g_{ij}), g_{ii} = n/N, G = G^2, \mathrm{tr}G = N - 1 = n$. Thus, using Theorem 2.1(a) given in the next section, we get

$$\begin{aligned}
E\left[\frac{1}{p}\mathrm{tr}S^2 - \frac{1}{np}(\mathrm{tr}S)^2\right] &= \frac{1}{pn^2}E\left[\mathrm{tr}(ZGZ'ZGZ') - \frac{1}{n}(\mathrm{tr}GZ'Z)^2\right] \\
&= \frac{1}{pn^2}E\left[(1 - 1/n)\sum_{i=1}^{p}(\mathbf{w}_i'G\mathbf{w}_i)^2 + \sum_{i\neq j}^{p}(\mathbf{w}_i'G\mathbf{w}_j\mathbf{w}_j'G\mathbf{w}_i - n^{-1}\mathbf{w}_i'G\mathbf{w}_i\mathbf{w}_j'G\mathbf{w}_j)\right] \\
&= \frac{1}{pn^2}\left(\frac{(n-1)p}{n}\left((\gamma - 3)\sum_{i=1}^{N}g_{ii}^2 + 2\mathrm{tr}G^2 + (\mathrm{tr}G)^2\right) + \sum_{i\neq j}^{p}(\mathrm{tr}G - n^{-1}(\mathrm{tr}G)^2)\right) \\
&= \frac{n-1}{n^3}((\gamma - 3)Nn^2/N^2 + 2n + n^2) \\
&= \frac{n-1}{n^3}\left(\frac{(\gamma - 3)n^2}{N} + n(n+2)\right) = \frac{n-1}{n^2}\left(\frac{(\gamma - 3)n}{N} + (n+2)\right).
\end{aligned}$$

Hence, under Assumption A,

$$E[\widehat{\delta}_2] = \frac{c_n}{p}E[(\mathrm{tr}S^2 - n^{-1}(\mathrm{tr}S)^2)] = 1 + \frac{(\gamma - 3)n}{N(n+2)} = 1 + O(N^{-1}p^{-\epsilon}), \quad \epsilon > 0.$$

Thus, the bias goes to zero at the rate of $O(N^{-1}p^{-\epsilon})$. We may note that for showing its consistency, the factor c_n, or whether we use S with divisor n or N do not make any difference. Similarly, for obtaining the variances of $\widehat{\delta}_1$ and $\widehat{\delta}_2$. It is, however, notationally more convenient to consider

$$S^* = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})' \tag{2.1}$$

as an estimator of Σ and work with the estimators

$$\widehat{\delta}_1^\star = \frac{1}{p}\mathrm{tr}S^\star, \tag{2.2}$$

and

$$\widehat{\delta}_2^\star = \frac{c_N}{p}(\mathrm{tr}{S^\star}^2 - N^{-1}(\mathrm{tr}S^\star)^2), \tag{2.3}$$

where

$$c_N = \frac{N^2}{n(N+2)},$$

in place of $\widehat{\delta}_1$ and $\widehat{\delta}_2$. It may be noted that

$$c_n - c_N = O(n^{-2}),$$

and hence, we may use c_n in place of c_N.

Then, from Theorem 2.9 given later at the end of this section, $\hat{\delta}_1$ and $\hat{\delta}_2$, given in (1.4), can be approximated in probability by (2.2) and (2.3), respectively. Moreover, in order to prove the consistency of $\hat{\delta}_1^*$ and $\hat{\delta}_2^*$, we need some results on quadratic forms, stated in the following subsection.

2.1. Moments of quadratic forms

Theorem 2.1. *Let $\mathbf{u} = (u_1, \ldots, u_p)'$ where u_i are i.i.d. with mean 0, variance 1, fourth moment γ, sixth moment γ_6 and eighth moment γ_8. Then for any $A = (a_{ij})$ and $B = (b_{ij})$ symmetric matrices of size $p \times p$,*

(a)

$$E[\mathbf{u}'A\mathbf{u}]^2 = (\gamma - 3)\sum_{i=1}^{p} a_{ii}^2 + 2\operatorname{tr} A^2 + (\operatorname{tr} A)^2,$$

(b)

$$\operatorname{Var}[\mathbf{u}'A\mathbf{u}] = (\gamma - 3)\sum_{i=1}^{p} a_{ii}^2 + 2\operatorname{tr} A^2,$$

(c)

$$E[(\mathbf{u}'A\mathbf{u})(\mathbf{u}'B\mathbf{u})] = (\gamma - 3)\sum_{i=1}^{p} a_{ii}b_{ii} + 2\operatorname{tr}(AB) + (\operatorname{tr} A)(\operatorname{tr} B),$$

(d)

$$\operatorname{Cov}[(\mathbf{u}'A\mathbf{u}), (\mathbf{u}'B\mathbf{u})] = (\gamma - 3)\sum_{i=1}^{p} a_{ii}b_{ii} + 2\operatorname{tr}(AB),$$

(e)

$$\operatorname{Var}[(\mathbf{u}'\mathbf{u})^2] = p(\gamma_8 - \gamma^2) + 4p(p-1)(\gamma_6 - \gamma) + 4(p-1)(p-2)(p-3)(\gamma - 1),$$

(f)

$$E[\mathbf{u}'\mathbf{u}]^3 = p\gamma_6 + 3p(p-1)\gamma + p(p-1)(p-2).$$

Theorem 2.2. *Let u_i and v_j be independently and identically distributed with mean 0, variance 1 and fourth moment γ, $i, j = 1, \ldots, p$. Then for $\mathbf{u} = (u_1, \ldots, u_p)'$, and $\mathbf{v} = (v_1, \ldots, v_p)'$, and any $p \times p$ symmetric matrix $B = (b_{ij})$,*

$$\operatorname{Var}[\mathbf{u}'\mathbf{B}\mathbf{v}]^2 = (\gamma - 3)^2 \sum_{i=1}^{p}\sum_{j=1}^{p} b_{ij}^4 + 6(\gamma - 3)\sum_{i=1}^{p}\left(\sum_{j=1}^{p} b_{ij}^2\right)^2 + 6\operatorname{tr} B^4 + 2(\operatorname{tr} B^2)^2.$$

2.2. Consistency of $\hat{\delta}_1^*$

For the sake of convenience of presentation, we shall not distinguish between δ_i and $\delta_i^o = \lim_{p\to\infty} \delta_i$, $i = 1, \ldots, 4$. From (1.1), $S^* = N^{-1}\sum_{i=1}^{N} C\mathbf{z}_i\mathbf{z}_i'C'$. Let $B = C'C = (b_{ij})$. Then

$$E[\hat{\delta}_1^*] = \frac{N}{Np}E[\mathbf{z}_i'B\mathbf{z}_i] = \frac{\operatorname{tr} B}{p} = \delta_1$$

$$\operatorname{Var}[\hat{\delta}_1^*] = \frac{N}{N^2p^2}\operatorname{Var}[\mathbf{z}_i'B\mathbf{z}_i] = \frac{1}{Np}\left((\gamma - 3)\sum_{i=1}^{p}\frac{b_{ii}^2}{p} + 2\frac{\operatorname{tr} B^2}{p}\right).$$

Thus, under Assumption A, $\operatorname{Var}[\hat{\delta}_1^*] = O((Np)^{-1})$, and $\hat{\delta}_1^*$ is a consistent estimator of δ_1. Furthermore,

$$E[(\hat{\delta}_1^*)^2] = \delta_1^2 + \operatorname{Var}[\hat{\delta}_1^*] = \delta_1^2 + O((Np)^{-1}). \tag{2.4}$$

Now

$$S = n^{-1}CZGZ'C' = n^{-1}(1 - 1/N)CZZ'C' - \frac{1}{nN}\sum_{i\neq j}^{N} C\mathbf{z}_i\mathbf{z}_j'C'.$$

Hence,

$$\widehat{\delta}_1 = \widehat{\delta}_1^{\star} - \frac{1}{npN}\sum_{i\neq j}^{N} \mathbf{z}_i' B \mathbf{z}_j,$$

$$E[\widehat{\delta}_1] = E[\widehat{\delta}_1^{\star}], \quad \mathrm{Var}(\widehat{\delta}_1) = O((Np)^{-1}),$$

and

$$E[\hat{\delta}_1^2] = \delta_1^2 + O((Np)^{-1}), \tag{2.5}$$

and the next theorem has been established.

Theorem 2.3.

$$E[(\hat{\delta}_1^*)^2] = \delta_1^2 + O((Np)^{-1}) = E[\hat{\delta}_1^2].$$

2.3. *Variance of $\hat{\delta}_2^*$ under the hypotheses H_1 and H_2*

The proposed statistic T_1 is invariant under the scalar transformations $\mathbf{x}_i \to c\mathbf{x}_i,\ c \neq 0$. Thus, we may assume without any loss of generality that $\Sigma = I$ under the hypothesis H_1, the same as for the hypothesis H_2. Hence, all the results in this subsection are obtained under the assumption that $\Sigma = I_p$. When $\Sigma = I_p$, the observation matrix can be expressed in two ways:

$$Z = (z_{ij}) = (\mathbf{z}_1, \ldots, \mathbf{z}_N) = (\mathbf{w}_1, \ldots, \mathbf{w}_p)' = (w_{ij}). \tag{2.6}$$

Under H_1 and H_2 all the elements z_{ij} or w_{ij} are *i.i.d.* with mean 0 and variance 1. Thus,

$$E[\mathbf{w}_i] = \mathbf{0}, \qquad \mathrm{Cov}[\mathbf{w}_i] = I_N,$$

since $\mathbf{w}_i$ is an N-dimensional random vector. We shall now express $\hat{\delta}_2^*$ in terms of $\mathbf{w}_i$ as $B = I$ under H_1 and H_2. Thus under H_1 or H_2,

$$S^* = \frac{1}{N}ZZ' = \frac{1}{N}(\mathbf{w}_1, \ldots, \mathbf{w}_p)'(\mathbf{w}_1, \ldots, \mathbf{w}_p). \tag{2.7}$$

To evaluate the variance of $\hat{\delta}_2^*$, we rewrite $\hat{\delta}_2^*$ in terms of the random vectors $\mathbf{w}_i,\ i = 1, \ldots, p$. That is,

$$\hat{\delta}_2^* = q_1 + q_2, \tag{2.8}$$

where

$$q_1 = \frac{nc_N}{N^3 p}\sum_{i=1}^{p} v_{ii}^2, \quad v_{ii} = \mathbf{w}_i'\mathbf{w}_i, \tag{2.9}$$

$$q_2 = \frac{2c_N}{N^2 p}\sum_{i<j}^{p}\left(v_{ij}^2 - \frac{1}{N}v_{ii}v_{jj}\right), \quad v_{ij} = \mathbf{w}_i'\mathbf{w}_j. \tag{2.10}$$

Let $\mathbf{w}$ be a random vector having the same distribution as $\mathbf{w}_i$, and $v = \mathbf{w}'\mathbf{w}$. Then, from Theorem 2.1(a), (b)

$$E[q_1] = \frac{nc_N}{N^3 p}E[v^2] = \frac{nc_N}{N^2}(N + 2 + \gamma - 3) = 1 + \frac{\gamma - 3}{N+2} = 1 + O(N^{-1}p^{-\epsilon}), \tag{2.11}$$

$$\begin{aligned} \mathrm{Var}[q_1] &= \frac{n^2 c_N^2}{p^2 N^6}\sum_{i=1}^{p}\mathrm{Var}[v_{ii}^2] = \frac{1}{N^2(N+2)^2 p}\mathrm{Var}[v_{ii}] \\ &= N^{-2}(N+2)^{-2}p^{-1}(N(\gamma_8 - \gamma^2) + 4Nn(\gamma_6 - \gamma) + 4n(N-2)(N-3)(\gamma - 1)). \end{aligned} \tag{2.12}$$

Theorem 2.4. *Let q_1 be given in (2.9). Then, under the hypothesis $H_1(H_2)$, and Assumption A*

$$E[q_1] = 1 + O(N^{-1}p^{-\epsilon}), \quad \epsilon > 0,$$

and

$$\mathrm{Var}[q_1] = 4(\gamma - 1)(Np)^{-1}(1 + O(N^{-1}p^{-1})).$$

Let

$$u_{ij} = v_{ij}^2 - \frac{1}{N} v_{ii} v_{jj} = (\mathbf{w}_i' \mathbf{w}_j \mathbf{w}_j' \mathbf{w}_i) - \frac{1}{N} (\mathbf{w}_i' \mathbf{w}_i)(\mathbf{w}_j' \mathbf{w}_j). \tag{2.13}$$

Then (2.10) equals

$$q_2 = \frac{2c_N}{N^2 p} \sum_{i<j}^{p} u_{ij}, \tag{2.14}$$

and $E[q_2] = 0$. Hence, under the Assumption A,

$$E[\widehat{\delta}_2^\star] = E[q_1] + E[q_2] = 1 + \frac{\gamma - 3}{N+2} = 1 + O(N^{-1} p^{-\epsilon}), \quad \epsilon > 0.$$

To calculate the variance of q_2, we first evaluate

$$\mathrm{Cov}[u_{ij}, u_{ik}] = E\left[((\mathbf{w}_j' \mathbf{w}_i)^2 - N^{-1}(\mathbf{w}_i' \mathbf{w}_i)(\mathbf{w}_j' \mathbf{w}_j))((\mathbf{w}_k' \mathbf{w}_i)^2 - N^{-1}(\mathbf{w}_i' \mathbf{w}_i)(\mathbf{w}_k' \mathbf{w}_k))\right], \quad i \neq j \neq k.$$

Since,

$$E[(\mathbf{w}_i' \mathbf{w}_j \mathbf{w}_j' \mathbf{w}_i)(\mathbf{w}_i' \mathbf{w}_k \mathbf{w}_k' \mathbf{w}_i)] = E[(\mathbf{w}_i' \mathbf{w}_i)^2],$$
$$-\frac{1}{N} E[(\mathbf{w}_i' \mathbf{w}_j \mathbf{w}_j' \mathbf{w}_i)(\mathbf{w}_i' \mathbf{w}_i)(\mathbf{w}_k' \mathbf{w}_k)] = -E[(\mathbf{w}_i' \mathbf{w}_i)^2],$$
$$-\frac{1}{N} E[(\mathbf{w}_i' \mathbf{w}_i)(\mathbf{w}_j' \mathbf{w}_j)(\mathbf{w}_i' \mathbf{w}_k \mathbf{w}_k' \mathbf{w}_i)] = -E[(\mathbf{w}_i' \mathbf{w}_i)^2],$$
$$\frac{1}{N^2} E[(\mathbf{w}_i' \mathbf{w}_i)(\mathbf{w}_j' \mathbf{w}_j)(\mathbf{w}_i' \mathbf{w}_i)(\mathbf{w}_k' \mathbf{w}_k)] = E[(\mathbf{w}_i' \mathbf{w}_i)^2],$$

it follows that

$$\mathrm{Cov}[u_{ij}, u_{ik}] = 0, \quad i \neq j \neq k. \tag{2.15}$$

Hence,

$$\mathrm{Var}[q_2] = \frac{4c_N^2}{N^4 p^2} \sum_{i<j}^{p} \mathrm{Var}[u_{ij}] = \frac{2p(p-1)c_N^2}{N^4 p^2} \mathrm{Var}[u_{ij}].$$

Thus, we need to evaluate $\mathrm{Var}[u_{ij}] = E[u_{ij}^2]$, since $E[u_{ij}] = 0$. Let $A_j = (a_{ik}(j)) = \mathbf{w}_j \mathbf{w}_j'$, $\mathbf{w}_j = (w_{j1}, \ldots, w_{jN})'$. Then, for $i \neq j$,

$$u_{ij}^2 = v_{ij}^4 - \frac{2}{N} v_{ij}^2 v_{ii} v_{jj} + \frac{1}{N^2} v_{ii}^2 v_{jj}^2, \quad \text{and} \quad v_{ij}^4 = (\mathbf{w}_i' \mathbf{w}_j \mathbf{w}_j' \mathbf{w}_i)^2 = (\mathbf{w}_i' A_j \mathbf{w}_i)^2.$$

Hence, for $i \neq j$

$$E[v_{ij}^4] = E[E(\mathbf{w}_i' A_j \mathbf{w}_i)|A_j]^2 = 3N^2 + N(\gamma^2 - 3).$$

Next, we evaluate

$$E[v_{ij}^2 v_{ii} v_{jj}] = E[\mathbf{w}_i' A_j \mathbf{w}_i \mathbf{w}_i' \mathbf{w}_i \mathrm{tr} A_j] = N(N + \gamma - 1)^2.$$

Finally,

$$E[v_{ii}^2 v_{jj}^2] = E[\mathbf{w}_i' \mathbf{w}_i]^2 E[\mathbf{w}_j' \mathbf{w}_j]^2 = N^2(N + \gamma - 1)^2.$$

Hence,

$$\mathrm{Var}[u_{ij}] = (N-1)((\gamma - 1)^2 + 2N),$$

and we get the following theorem.

Theorem 2.5. *Let $\mathbf{w}_1, \ldots, \mathbf{w}_p$ be i.i.d. with mean $\mathbf{0}$ and covariance I_N, and fourth moment γ. Then, the mean and variance of q_2 in (2.14) is given by*

$$E[q_2] = 0,$$
$$\mathrm{Var}[q_2] = \frac{4}{N^4 p^2} \frac{p(p-1)c_N^2}{2} (N-1)((\gamma - 1)^2 + 2N) \approx \frac{4}{N^2} \left(1 + \frac{(\gamma - 1)^2}{2N}\right).$$

We may also prove

Theorem 2.6. *Let q_1 and q_2 be given by (2.9) and (2.14), respectively. Then,* $\mathrm{Cov}[q_1, q_2] = 0$.

Theorem 2.7. *Let $\hat{\delta}_1^*$ and q_2 be given by (2.2) and (2.14), respectively. Then,* $\mathrm{Cov}[\hat{\delta}_1^*, q_2] = 0$.

Using the results obtained above, we get the following theorem.

Theorem 2.8. *The variance of $\delta_2^\star$ is given by*

$$\mathrm{Var}[\delta_2^\star] = \mathrm{Var}[q_1] + \mathrm{Var}[q_2].$$

Theorem 2.9. *As $(N, p) \to \infty$;*

$$\text{(a) } \frac{1}{p}(\mathrm{tr}(S^\star)^2 - \mathrm{tr}S^2) = o_p(N^{-2}),$$

$$\text{(b) } \widehat{\delta_2} - \widehat{\delta_2^\star} = O_p(N^{-2}).$$

Since $S = S^\star - \frac{1}{nN}\sum_{j\neq k}^N \mathbf{x}_j\mathbf{x}_k'$,

$$\frac{1}{p}(\mathrm{tr}(S^\star)^2 - \mathrm{tr}S^2) = \frac{2}{nNp}\sum_{j\neq k}^N \mathbf{x}_k' S^\star \mathbf{x}_j - \frac{1}{n^2N^2p}\mathrm{tr}\left(\sum_{j\neq k}^N \mathbf{x}_j\mathbf{x}_k'\right)^2.$$

Note that $\mathbf{x}_j' S^\star \mathbf{x}_k$ is a linear combination of the terms $S_{lm}^\star$ where as $(N, p) \to \infty$, $S_{lm}^\star \to \sigma_{lm}$ in probability. Thus, the first term on the right-hand side is equal to $\frac{2}{nNp}\sum_{j\neq k}^N \mathbf{x}_j' \Sigma \mathbf{x}_k$ in probability, with variance given by $\frac{8}{n^2N^2p^2}\mathrm{tr}\Sigma^4$. Since $\frac{1}{p^2}\mathrm{tr}\Sigma^4 \to 0$, it is of order $o_p(N^{-2})$. The second term is of even lower order. Thus (a) is proved. The proof of (b) follows from Theorem 2.3.

3. Proof of Theorem 1.4

To establish the joint asymptotic normality of k statistics

$$t_{i,p}^{(n)} = \sum_{j=1}^p x_{ij}^{(n)}, \quad i = 1, \ldots, k$$

we consider an arbitrary linear combination

$$t_p^{(n)} = c_1 t_{1,p}^{(n)} + \cdots + c_k t_{k,p}^{(n)} = \sum_{j=1}^p \sum_{i=1}^k c_i x_{ij}^{(n)} \equiv \sum_{j=1}^p y_j^{(n)}$$

where without any loss of generality $c_1^2 + \cdots + c_k^2 = 1$, and $y_j^{(n)} = \sum_{i=1}^k c_i x_{ij}^{(n)}$. Here, $x_{ij}^{(n)}$ is a sequence of random variables which may depend on n. From the definition of multivariate normality, see [19], the joint normality for all $c_1, \ldots, c_k$ will follow if the normality of $t_p^{(n)}$ is established. Let $F_l^{(n)}$ be the σ-algebra generated by the random variables $(x_{1j}^{(n)}, \ldots, x_{kj}^{(n)})$, $j = 1, \ldots, l$, $l = 1, \ldots, p$. Then $F_0 \subset F_1^{(n)} \subset \cdots \subset F_p^{(n)} \subset F$, where $(\Lambda, F, \mathcal{P})$ is the probability space and $F_0 = \{\emptyset, \Lambda\}$; $\emptyset$ being the null set, and Λ the whole set.

Theorem 3.1. *Let $x_{ij}^{(n)}$ be a sequence of random variables, and $y_j^{(n)} = \sum_{i=1}^k c_i x_{ij}^{(n)}$, $j = 1, \ldots, p$. If*

(i) $E[y_j^{(n)} | F_{j-1}^{(n)}] = 0$,

(ii) $\lim_{(N,p)\to\infty} E[(y_j^{(n)})^2] < \infty$,

(iii) $\sum_{j=0}^p E[(y_j^{(n)})^2 | F_{j-1}^{(n)}] \xrightarrow{p} \sigma_0^2$, *as* $(n, p) \to \infty$,

(iv) $L \equiv \sum_{j=0}^p E[(y_j^{(n)})^2 I(|y_j^{(n)}| > \epsilon) | F_{j-1}^{(n)}] \xrightarrow{p} 0$, *as* $(n, p) \to \infty$,

then $t_p^{(n)} = \sum_{j=1}^p y_j^{(n)} \xrightarrow{d} N(0, \sigma_0^2)$, *as* $(n, p) \to \infty$.

The proof of this theorem follows from Theorem 4 of Shiryayev [15, p. 511], since the first two conditions imply that $\{x_j^{(n)}, F_j^{(n)}\}$ forms a sequence of integrable martingale differences. The condition (iv) is known as Lindeberg's condition. To verify this condition, we note that from Markov's and Cauchy–Schwarz inequalities

$$P[L > \delta] \leq \sum_{j=0}^p E[(y_j^{(n)})^4]/\delta\epsilon^2.$$

Thus,

$$E[(y_j^{(n)})^4] \le k^3 \sum_{i=1}^{k} c_i^4 E[(x_{ij}^{(n)})^4] \le k^3 \sum_{i=1}^{k} E[(x_{ij}^{(n)})^4].$$

Hence, if

$$\sum_{j=1}^{p} E[(x_{ij}^{(n)})^4] \to 0,$$

for all $i = 1, \ldots, k$, the Lindeberg condition is satisfied.

Because of the invariance of the statistic T_1 under a scalar transformation, we shall assume as before that $\Sigma = I$ and thus $B = I$ in both the hypotheses H_1 and H_2. We first consider the joint distribution of $\hat{\delta}_1^*$ and q_1 defined in (2.2) and (2.9) respectively, under $\Sigma = I_p$. Let $\boldsymbol{\xi}_i = (\xi_{1i}, \xi_{2i})'$ where $\xi_{1i} = N^{-\frac{1}{2}}(\mathbf{w}_i'\mathbf{w}_i - N)$, $\xi_{2i} = N^{-\frac{3}{2}}[(\mathbf{w}_i'\mathbf{w}_i)^2 - N^2 - N(\gamma - 1)]$, $i = 1, \ldots, p$ and $\mathbf{w}_i$ is as in Section 2. Then the vectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_p$ are *i.i.d.* with mean $\mathbf{0}$ and covariance matrix Ω_1 given by

$$\Omega_1 = \begin{pmatrix} \gamma - 1 & 2(\gamma - 1) \\ 2(\gamma - 1) & 4(\gamma - 1) \end{pmatrix}.$$

Hence,from the multivariate central limit theorem $(1/\sqrt{p}) \sum_{i=1}^{p} \boldsymbol{\xi}_i \longrightarrow N_2(\mathbf{0}, \Omega_1)$, irrespective of whether N goes to infinity and then p goes to infinity or p goes to infinity and then N goes to infinity. Since

$$\hat{\delta}_1^* = \frac{1}{p\sqrt{N}} \sum_{i=1}^{p} \xi_{1i} + 1, \quad \text{and} \quad q_1 = \frac{1}{p\sqrt{N}} \sum_{i=1}^{p} \xi_{2i} + 1 + \frac{\gamma - 1}{N},$$

we get the following theorem.

Theorem 3.2. *The asymptotic distribution of $\hat{\delta}_1^*$ and q_1 is bivariate normal given by*

$$\sqrt{Np}\,\Omega_1^{-1/2} \begin{pmatrix} \hat{\delta}_1^* - 1 \\ q_1 - 1 \end{pmatrix} \xrightarrow{d} N_2(0, I_2)$$

as $(N, p) \to \infty$ in any manner.

It remains to find the distribution of q_2, to obtain the joint distribution of $\hat{\delta}_1^*$ and $\hat{\delta}_2^*$. Note that from (2.14),

$$Nq_2c_n^2 = \sum_{j=2}^{p} \eta_j \equiv \frac{2}{Np} \sum_{j=2}^{p} \sum_{i=1}^{j-1} u_{ij}.$$

Let F_j be the σ-algebra generated by the random vectors $\mathbf{w}_1, \ldots, \mathbf{w}_j$. Letting $\mathbf{w}_0 = 0$, and $F_0 = (\emptyset, \Lambda) = F_{-1}$, where $\emptyset$ is the empty set and Λ is the whole space, we find that $F_0 \subset F_1 \subset \cdots \subset F_p \subset F$. Also,

$$E[\eta_j | F_{j-1}] = 0.$$

Then

$$\begin{aligned} E[\eta_j^2 | F_{j-1}] &= \frac{4}{N^2p^2} \left(\sum_{i=1}^{j-1} E(u_{ij}^2 | F_{j-1}) + 2 \sum_{k<l}^{j-1} E(u_{kj}u_{lj} | F_{j-1}) \right) \\ &\equiv \frac{4}{N^2p^2} \left(\sum_{i=1}^{j-1} b_{iN} + 2 \sum_{k<l}^{j-1} c_{klN} \right) \end{aligned}$$

and

$$E[\eta_j^2] = \frac{4}{N^2p^2}((j-1)b_N + (j-1)(j-2)h_n), \quad j \le p,$$

where

$$b_N = E[b_{iN}] = E[u_{ij}^2] = N(N-1)\left(2 + \frac{(\gamma - 1)^2}{N}\right),$$

and

$$h_N = E[c_{klN}] = E[u_{kj}u_{lj}] = \text{Cov}[u_{kj}, u_{lj}] = 0, \quad k < l < j,$$

giving

$$E[\eta_j^2] = \frac{4N(N-1)}{N^2p^2}(j-1)\left(2+\frac{(\gamma-1)^2}{N}\right) < \infty, \quad j \le p.$$

From the definition, it follows that (η_k, F_k) is a sequence of integrable martingale differences. To prove the asymptotic normality of Nq_2, we apply Theorem 3.1. We note that

$$E\left[\sum_{j=2}^{p} E[\eta_j^2|F_{j-1}]\right] = \sum_{j=2}^{p} E[\eta_j^2] = \frac{2N(N-1)}{N^2p^2}p(p-1)\left(2+\frac{(\gamma-1)^2}{N}\right).$$

Thus

$$\lim_{(N,p)\to\infty} E\left[\sum_{j=2}^{p} E[\eta_j^2|F_{j-1}]\right] = 4,$$

and in Theorem 3.1(iii) $\sigma_0^2 = 4$. We will show that $\upsilon^2 = \mathrm{Var}\left[\sum_{j=2}^{p} E[\eta_j^2|F_{j-1}]\right] \to 0$, as $(N,p)\to\infty$, and find that

$$\upsilon^2 = \mathrm{Var}\left[\frac{4}{N^2p^2}\sum_{j=2}^{p}\left(\sum_{i=1}^{j-1} b_{iN} + 2\sum_{k<l}^{j-1} c_{klN}\right)\right],$$

where

$$\begin{aligned} b_{iN} &= E[u_{ij}^2|F_{j-1}], \quad i<j \\ &= E\left[(\mathbf{w}_j'A_i\mathbf{w}_j)^2 - \frac{2}{N}(\mathbf{w}_j'A_i\mathbf{w}_j)\upsilon_{jj}\upsilon_{ii} + \frac{1}{N^2}\upsilon_{ii}^2(\mathbf{w}_j'\mathbf{w}_j)^2|F_{j-1}\right], \end{aligned}$$

with $A_i = \mathbf{w}_i\mathbf{w}_i' = (a_{rl}(i)) : N\times N$. Using Theorem 2.1, yields

$$\begin{aligned} b_{iN} &= (\gamma-3)\sum_{r=1}^{N} a_{rr}^2(i) + 3(\mathbf{w}_i'\mathbf{w}_i)^2 - \frac{2}{N}\left((\gamma-3)\sum_{l=1}^{N} a_{ll}(i) + 2\mathbf{w}_i'\mathbf{w}_i + N\mathbf{w}_i'\mathbf{w}_i\right)(\mathbf{w}_i'\mathbf{w}_i) \\ &\quad + \frac{1}{N^2}((\gamma-3)N + 2N + N^2)(\mathbf{w}_i'\mathbf{w}_i)^2 \\ &= d(\mathbf{w}_i'\mathbf{w}_i)^2 + (\gamma-3)\left(\sum_{k=1}^{N} w_{ik}^4\right), \quad d = \left(2 - \frac{\gamma-1}{N}\right). \end{aligned}$$

Thus, to show that the variance of $4(N^2p^2)^{-1}\left(\sum_{j=2}^{p}\sum_{i=1}^{j-1} b_{iN}\right)$ goes to zero, it will be sufficient to show that the variance of $4d(N^2p^2)^{-1}\sum_{j=2}^{p}\sum_{i=1}^{j-1}(\mathbf{w}_i'\mathbf{w}_i)^2$, as well as the variance of $4(\gamma-3)(N^2p^2)^{-1}\sum_{j=2}^{p}\sum_{i=1}^{j-1}\left(\sum_{k=1}^{N} w_{ik}^4\right)$ go to zero. Clearly,

$$\begin{aligned} \mathrm{Var}\left[\frac{4d}{N^2p^2}\sum_{j=2}^{p}\sum_{i=1}^{j-1}\mathbf{w}_i'\mathbf{w}_i\right] &= \frac{16d^2}{N^4p}\mathrm{Var}\left[\sum_{j=1}^{p-1}(p-j)(\mathbf{w}_j'\mathbf{w}_j)\right] \\ &\le \frac{16d^2}{N^4p}((\gamma-3)N + N^2) \to 0 \quad \text{as } (N,p)\to\infty. \end{aligned}$$

Similarly, in order to show that $\upsilon^2 \to 0$, we need to show that

$$\mathrm{Var}\left[\frac{8}{N^2p^2}\sum_{j=2}^{p}\sum_{k<l}^{j-1} c_{klN}\right] = \frac{8^2}{N^4p^4}\mathrm{Var}\left[\sum_{1\le k<l}^{p-1}(p-l-1)c_{klN}\right] \to 0.$$

For this, we calculate c_{klN} which after some manipulations can be shown to equal

$$c_{klN} = E[u_{kj}u_{lj}|F_{j-1}] = (\gamma-3)\sum_{r=1}^{N} w_{rr}^2(k)w_{rr}^2(l) + 2\left(\upsilon_{kl}^2 - \frac{\gamma-1}{N}\upsilon_{kk}\upsilon_{ll}\right), \quad k<l<j.$$

Thus,

$$\begin{aligned} \frac{64}{N^4p^4}\mathrm{Var}\left[\sum_{1\le k<l}^{p-1}(p-l-1)c_{klN}\right] &\le \frac{64}{N^4p^2}\mathrm{Var}\left[\sum_{1\le k<l}^{p-1} c_{klN}\right] \\ &= \frac{64}{N^4p^2}\mathrm{Var}\left[\sum_{1\le k<l}^{p-1}\left\{(\gamma-3)\sum_{r=1}^{N} w_{rr}^2(k)w_{rr}^2(l) + 2\left(\upsilon_{kl}^2 - \frac{\gamma-1}{N}\upsilon_{kk}\upsilon_{ll}\right)\right\}\right]. \end{aligned}$$

We need to show that the variance of each of the terms goes to zero. Clearly, the first term is of the order $O(N^{-3})$. Similarly, from the results of Section 2, the second term is of the order $O(N^{-2})$ and the third term is of the order $O(N^{-3})$. Hence, we have shown that condition (iii) is satisfied.

Next, we show that

$$\sum_{k=0}^{p} E[\eta_k^4] \to 0 \quad \text{as } (N,p) \to \infty.$$

For this, we note that $\eta_j = 2(Np)^{-1} \sum_{i=1}^{j-1} u_{ij}$, and hence,

$$\begin{aligned} c_n^{-4} N^4 p^4 \sum_{j=0}^{p} E[\eta_j^4] &= 16E \sum_{j=2}^{p} \left[\sum_{i=1}^{j-1} u_{ij}^4 + 6 \sum_{k<l}^{j-1} u_{kj}^2 u_{lj}^2 \right] \\ &= 16E \left[\sum_{j=2}^{p} \sum_{i=1}^{j-1} E[u_{ij}^4 | F_{j-1}] + 6 \sum_{k<l}^{j-1} E[u_{kj}^2 u_{lj}^2 | F_{j-1}] \right]. \end{aligned}$$

Now

$$\begin{aligned} u_{ij}^4 &= \left((\mathbf{w}_j' A_i \mathbf{w}_j)^2 - \frac{2}{N} (\mathbf{w}_j' A_i \mathbf{w}_j) v_{jj} v_{ii} + \frac{1}{N^2} v_{ii}^2 (\mathbf{w}_j' \mathbf{w}_j)^2 \right)^2 \\ &= (\mathbf{w}_j' A_i \mathbf{w}_j)^4 + \frac{4}{N^2} (\mathbf{w}_j' A_i \mathbf{w}_j)^2 v_{jj}^2 v_{ii}^2 + \frac{1}{N^4} v_{ii}^4 (\mathbf{w}_j' \mathbf{w}_j)^4 - \frac{4}{N} (\mathbf{w}_j' A_i \mathbf{w}_j)^3 v_{jj} v_{ii} \\ &\quad + \frac{2}{N^2} (\mathbf{w}_j' A_i \mathbf{w}_j)^2 (\mathbf{w}_j' \mathbf{w}_j)^2 v_{ii}^2 - \frac{4}{N^3} (\mathbf{w}_j' A_i \mathbf{w}_j) (\mathbf{w}_j' \mathbf{w}_j)^2 v_{ii}^3 v_{jj}. \end{aligned}$$

It can be shown that the leading term in u_{ij}^4 is $(\mathbf{w}_j' A_i \mathbf{w}_j)^4$, and

$$E[(\mathbf{w}_j' A_i \mathbf{w}_j)^4] \le E[(\mathbf{w}_j' \mathbf{w}_j)^4 (\mathbf{w}_i' \mathbf{w}_i)^4] \le E[v_{ii}^4 v_{jj}^4] = O(N^4).$$

Hence, $E[u_{ij}^4] = O(N^4)$.

Let

$$g_i = E[u_{ij}^4 | F_{j-1}], \quad i < j,$$

and

$$h_{kl} = E[u_{kj}^2 u_{lj}^2 | F_{j-1}], \quad k < l.$$

Then,

$$\begin{aligned} \sum_{j=2}^{p} E[\eta_j^4] &= \frac{c_n^4 16}{N^4 p^4} \left(\sum_{j=1}^{p-1} (p-j) E[g_j] + 6 \sum_{1 \le k < l}^{p-1} (p-l-1) h_{kl} \right) \\ &\le \frac{c_n^4 16}{N^4 p^3} \left(\sum_{j=1}^{p-1} E[g_j] + 6 \sum_{1 \le k < l}^{p-1} E[h_{kl}] \right) = O(p^{-2}) + O(p^{-1}), \end{aligned}$$

from Theorem 2.1. Thus, the Lindeberg condition is also satisfied. Hence, as $(N,p) \to \infty$,

$$Nq_2 \to N(0,4),$$

or equivalently, q_2 is approximately normally distributed as normal with mean 0 and variance $4/N^2$.

We shall now apply Theorem 3.1 again to obtain the joint normality of $\hat{\delta}_1^*$, q_1, and q_2. In the notation of Theorem 3.1, let

$$t_{1,p}^{(n)} = \sum_{i=1}^{p} \left(\frac{\xi_{1i}}{\sqrt{p}} \right), \qquad t_{2,p}^{(n)} = \sum_{i=1}^{p} \left(\frac{\xi_{2i}}{\sqrt{p}} \right), \qquad t_{3,p}^{(n)} = \sum_{i=1}^{p} \eta_i.$$

It is easy to check that

$$\sum_{i=1}^{p} E\left[\left(\frac{\xi_{1i}}{\sqrt{p}} \right)^4 \right] \quad \text{and} \quad \sum_{i=1}^{p} E\left[\left(\frac{\xi_{2i}}{\sqrt{p}} \right)^4 \right]$$

go to zero as $(N,p) \to \infty$ while we have already shown that $\sum_{i=1}^{p} E[\eta_i^4] \to 0$ as $(N,p) \to \infty$. Similarly, the convergence can be satisfied. Hence, we have

$$\begin{pmatrix} (Np)^{-1}\Omega_1 & 0 \\ 0 & 4/N^2 \end{pmatrix}^{-1/2} \begin{pmatrix} \hat{\delta}_1^* - 1 \\ q_1 - 1 \\ q_2 \end{pmatrix} \sim N_3(0, I_3).$$

Hence

$$\sqrt{Np}\Omega^{-1/2}\begin{pmatrix}\hat{\delta}_1^* - 1\\ \hat{\delta}_2^* - 1\end{pmatrix} \sim N_2(0, I_2),$$

where Ω is defined in (1.11). This proves Theorem 1.4. The corresponding results for $\widehat{\delta}_1$ and $\widehat{\delta}_2$ are obtained by replacing N by n.

4. Robustness of the sphericity test: proof of Theorem 1.1

In this section, we first discuss various tests available for testing the hypothesis of 'sphericity' H_1. When $N > p$, the likelihood ratio test (LRT) is based on the ratio of the arithmetic mean to the geometric mean of the eigenvalues of the sample covariance matrix S. The power of the LRT is a monotonically increasing function of the ratio of the eigenvalues of Σ, see [3]. Another test, known in the literature as the locally best invariant test (LBIT) was originally proposed by Nagao [10] but it was John [6] and Sugiura [20] who showed that it is the LBIT. It is based on the statistic

$$U = \left[\frac{\mathrm{tr}S^2}{p\hat{\delta}_1^2}\right] - 1.$$

It may be noted that $\left(\frac{\mathrm{tr}S^2}{p}\right)$ is a consistent estimator of $\left(\frac{\mathrm{tr}\Sigma^2}{p}\right)$, if $\left(\frac{p}{N}\right) \to 0$. Thus, when $\frac{p}{N} \to c \neq 0$, Ledoit and Wolf [8] considered the statistic $U - \frac{p}{N}$ and using the asymptotic result of Jonsson [7] gave its (N, p) asymptotic null-distribution under the Assumption A and the assumption that $\frac{p}{N} \to c$ as $(N, p) \to \infty$. The (N, p) asymptotic non-null distribution of $U - p/n$ can be obtained from Corollary 2.1 of Srivastava [16].

It may be noted that the statistic U exists irrespective of whether $N \leq p$ or $N > p$. Next, we define a measure of sphericity. From Cauchy–Schwarz inequality, we have for a $p \times p$ positive definite matrix Σ,

$$\frac{\delta_2}{\delta_1^2} = \frac{(\mathrm{tr}(\Sigma^2)/p)}{(\mathrm{tr}\,\Sigma/p)^2} \geq 1. \tag{4.16}$$

The equality holds if and only if (iff) all the eigenvalues of Σ are equal to some unknown constant, say λ. That is, iff $\Sigma = \lambda I_p$. Thus, as in [16], a measure of sphericity may be defined by

$$m_s = \left[\frac{(\mathrm{tr}(\Sigma^2)/p)}{(\mathrm{tr}\,\Sigma/p)^2} - 1\right]; \tag{4.17}$$

the larger the value of m_s, the larger the deviation from the sphericity as $m_s = 0$ under sphericity. The statistic T_1 defined in Section 1 is a consistent estimator of m_s. It may be noted that the statistic T_1 is invariant under the scalar transformation $\mathbf{x}_i \to a\mathbf{x}_i,\ a \neq 0$. Thus, without any loss of generality, we may assume that $\lambda = 1$ in obtaining the distribution of T_1.

We use Theorem 1.4 to obtain the distribution of T_1 under the hypothesis H_1 as $(N, p) \to \infty$. Under H_1, $\hat{\delta}_1$ and $\hat{\delta}_2$ are consistent estimators of δ_1 and δ_2, respectively. Now

$$\frac{\partial T_1}{\partial \hat{\delta}_1} = -2\frac{\hat{\delta}_2}{\hat{\delta}_1^3}, \qquad \frac{\partial T_1}{\partial \hat{\delta}_2} = \frac{1}{\hat{\delta}_1^2}.$$

Thus $(np)^{-1}(-2, 1)\Omega(-2, 1)' = 4n^{-2}$.

Hence, under H_1, $n^{-1}T_1 \xrightarrow{d} N(0, 4)$ as $(N, p) \to \infty$, proving Theorem 1.1, as well as showing that the test statistic T_1 for testing sphericity is robust.

5. A robust test for testing that Σ is an identity matrix: Proof of Theorem 1.2

Despite the monotonicity property of the power function of the LRT for this problem established by Nagao [9] and Das and Gupta [5], it cannot be considered since $N \leq p$. Thus, we consider a test based on a consistent estimator of the distance function that measures the departure of the hypothesis from the alternative, namely,

$$m_I = \frac{1}{p}\mathrm{tr}(\Sigma - I)^2 = \delta_2 - 2\delta_1 + 1.$$

Thus, Rao [13], and independently Nagao [11] proposed a test statistic

$$RN = \frac{1}{p}\mathrm{tr}S^2 - 2\hat{\delta}_1 + 1,$$

for testing the hypothesis that $\Sigma = I_p$. Ledoit and Wolf [8] modified it to

$$LW = RN - \frac{p}{n}\hat{\delta}_1^2,$$

and obtained its null distribution as normal, under the condition that

$$\lim_{(N,p)\to\infty} \frac{p}{N} = c > 0.$$

Using consistent estimators of δ_1 and δ_2, Srivastava [16] proposed a test based on the statistic

$$T_2 = \hat{\delta}_2 - 2\hat{\delta}_1 + 1,$$

and obtained its null as well as non-null distribution as $(N, p) \to \infty$. In this article, we show that T_2 is a robust test under the non-normality model (1.1)–(1.2). To obtain the distribution T_2, we use Theorem 1.4. Since

$$\frac{\partial T_2}{\partial \hat{\delta}_1} = -2, \qquad \frac{\partial T_2}{\partial \hat{\delta}_2} = 1,$$

we have

$$(np)^{-1}(-2, 1)'\Omega(-2, 1)' = 4n^{-2}.$$

Thus as $(N, p) \to \infty$, $N^{-1}T_2 \xrightarrow{d} N(0, 4)$, and hence proving Theorem 1.2 and the robustness of the test statistic T_2 as it does not depend on γ, γ_3, $\gamma_5 - \gamma_8$, it is the same distribution as given by Srivastava [16] under the assumption of normality.

6. Robustness of the diagonality test T_3: proof of Theorem 1.3

When the observations are normally distributed, the LRT is based on the determinant of the sample correlation matrix:

$$R = (r_{ij}),\ r_{ii} = 1,\ r_{ij} = \frac{s_{ij}}{(s_{ii}s_{jj})^{1/2}},$$

provided that $N > p$. When $N \le p$, the determinant of R does not exist. By defining the distance function as the sum of squared correlations $\rho_{ij}^2 = \frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}}$, $\sum_{i<j}\rho_{ij}^2$, which is zero iff $\rho_{ij} = 0$, Srivastava [16,17] proposed a test based on the normalized version of its consistent estimator. Schott [14] also gave its distribution under the condition that $\frac{p}{N} \to c$. However, since the convergence to normality is slow, Srivastava [16,17] proposed a test based on Fisher's transformation, and obtained its (N, p) asymptotic distribution. Srivastava [16] defined another distance function to measure the departure from the hypothesis H_3. It is given by

$$m_d = \frac{\operatorname{tr}\Sigma^2}{\sum_{i=1}^p \sigma_{ii}^2} - 1, \quad \Sigma = (\sigma_{ij}),$$

which is zero if and only if $\rho_{ij} = 0$. Under normality, a test based on its consistent estimator is given by the test statistic T_3 defined in Section 1. The (N, p) asymptotic distribution is given in [16] and its power compared in [17] with the test based on Fisher's transformation and shown to be at least as good as based on the Fisher's transformation. In this section, we show that this test T_3 defined in Section 1 is robust under the model (1.1)–(1.2). As in Section 2, we can for the asymptotic distribution purposes, consider $\hat{\delta}_2^*$ based on S^* instead of S, and N in place of $N-1$ and may show that

$$\hat{\delta}_2^* \approx \hat{\delta}_{20}^* + 2c_N \sum_{i<j}^p \left(s_{ij}^{*2} - \frac{1}{N}s_{ii}^*s_{jj}^* \right),$$

where $\hat{\delta}_{20}^* = p^{-1}c_N \sum_{i=1}^p s_{ii}^{*2}$.

Under the hypothesis H_3, $\Sigma = D$ with $C = D^{1/2}$. Hence, if $\mathbf{w}_i$ are *i.i.d.* with mean $\mathbf{0}$, covariance I_n, with fourth moment γ and the existence of eight moments, we can write

$$s_{ij}^* = d_i d_j \mathbf{w}_i'\mathbf{w}_j \quad \text{for all } i, j = 1, \ldots, p.$$

Let

$$q_3^* = \frac{2}{p}\sum_{i<j}^p \left(s_{ij}^{*2} - \frac{1}{N}s_{ii}^*s_{jj}^* \right) \equiv \frac{2}{N^2 p}\sum_{i<j}^p d_i d_j u_{ij},$$

with $E[u_{ij}] = 0$, and $\operatorname{Cov}[u_{ij}, u_{ik}] = 0,\ i \ne j \ne k$. Hence, following as in Theorem 2.4,

$$\operatorname{Var}(q_3^*) = \frac{4}{N^4p^2}\sum_{i<j}^p d_i^2 d_j^2 \operatorname{Var}[u_{ij}] = \frac{4}{N^2}(\delta_{20}^2 - p^{-1}\delta_{40}) + O(N^{-3}).$$

We now show that $\hat{\delta}_{20}^*$ and $\hat{\delta}_{40}^*$ are consistent estimators of $\delta_{20} = p^{-1}\sum_{i=1}^p \sigma_{ii}^2$ and $\delta_{40} = p^{-1}\sum_{i=1}^p \sigma_{ii}^4$, respectively under the hypothesis H_3 when $C = D^{1/2} = \text{diag}\left(d_1^{\frac{1}{2}}, \ldots, d_p^{\frac{1}{2}}\right)$; see Eq. (1.4) for the definition of their estimators. In terms of the *i.i.d.* random vector $\mathbf{w}_i$,

$$\hat{\delta}_{20} = \frac{c_N}{pN^2}\sum_{i=1}^p d_i^2(\mathbf{w}_i'\mathbf{w}_i)^2,$$

and its variance is given by

$$\text{Var}(\hat{\delta}_{20}) = \frac{1}{pN^4}\text{Var}(\mathbf{w}_i'\mathbf{w}_i)^2\left(\sum_{i=1}^p \frac{d_i^4}{p}\right) = O(N^{-1}p^{-1})$$

from Assumption A and Theorem 2.1(e). Since $E(\hat{\delta}_{20}) = \delta_{20}[1 + O(N^{-1})]$, $\hat{\delta}_{20}$ is a consistent estimator of δ_{20}. Similarly, it can be shown that $\hat{\delta}_{40}$ is a consistent estimator of δ_{40}. Let

$$\eta_k^* = \frac{2}{Np}d_k\sum_{i=1}^{k-1} d_i u_{ik}.$$

Then following the steps of Section 3, it can be shown that $\{\eta_k^*, F_k\}$ is a sequence of integrable martingale difference satisfying the convergence condition and Lindeberg's condition, i.e. Theorem 3.1(iii), (iv). Thus, Theorem 1.3 follows and thus the test statistic T_3 is shown to be robust.

Tests based on correlations and Fisher's z-transformation are given by Schott [14] and Srivastava [16,17]. While their robustness may be studied later, numerical comparison given in [17,18] show that the test T_3 performs better than the test based on correlation, specially for small n.

Acknowledgments

We wish to thank the two referees for their kind suggestions which led to the improvement in presentation. The research by M.S. Srivastava was supported by NSERC of Canada.

References

[1] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Statist. Soc. Ser. B 57 (1995) 289–300.
[2] R. Beran, M.S. Srivastava, Bootstrap tests and confidence regions for functions of a covariance matrix, Ann. Statist. 13 (1985) 95–115.
[3] E.M. Carter, M.S. Srivastava, Monotonicity of the power functions of the modified likelihood ratio criteria for the homogeneity of variances and of the sphericity test, J. Multivariate Anal. 7 (1977) 229–233.
[4] Y.M. Chan, M.S. Srivastava, Comparison of powers for the sphericity test using both the asymptotic distribution and the bootstrap, Commun. Stat. 17 (1988) 671–690.
[5] S. DasGupta, Properties of power functions of some tests concerning dispersion matrices of multivariate normal distributions, Ann. Math. Statist. 40 (1969) 697–701.
[6] S. John, Some optimal multivariate tests, Biometrika 58 (1971) 123–127.
[7] D. Jonsson, Some limit theorems for the eigenvalues of a sample covariance matrix, J. Multivariate Anal. 12 (1982) 1–38.
[8] O. Ledoit, M. Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, Ann. Statist. 30 (2002) 1081–1102.
[9] H. Nagao, Monotonicity of the modified likelihood ratio test for a covariance matrix, J. Sci. Hiroshima Univ. Ser. A-I 31 (1967) 147–150.
[10] H. Nagao, Asymptotic expansions of some test criteria for homogeneity of variances and covariance matrices from normal populations, J. Sci. Hiroshima Univ. Ser. A-I 34 (1970) 153–247.
[11] H. Nagao, On some test criteria for covariance matrix, Ann. Math. Statist. 1 (1973) 700–709.
[12] H. Nagao, M.S. Srivastava, On the distribution of some test criteria for a covariance matrix under local alternatives and bootstrap approximations, J. Multivariate Anal. 43 (1992) 331–350.
[13] C.R. Rao, Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, Math. Proc. Cambridge Philos. Soc. 44 (1948) 50–57.
[14] J.R. Schott, Testing for complete independence in high dimensions, Biometrika 92 (2005) 951–956.
[15] A.N. Shiryayev, Probability, Springer-Verlag, New York, 1984.
[16] M.S. Srivastava, Some tests concerning the covariance matrix in high-dimensional data, J. Japan Statist. Soc. 35 (2005) 251–272.
[17] M.S. Srivastava, Some tests criteria for the covariance matrix with fewer observations than the dimension, Acta Comment. Univ. Tartu. Math. 10 (2006) 77–93.
[18] M.S. Srivastava, A review of multivariate theory for high dimensional data with fewer observations, in: A. Sengupta (Ed.), Advances in Multivariate Statistical Methods, World Scientific Publishing Co. Pte. Ltd., Singapore, 2009, pp. 25–51.
[19] M.S. Srivastava, C.G. Khatri, An Introduction to Multivariate Statistics, North Holland, New York, 1979.
[20] N. Sugiura, Locally best invariant test for sphericity and the limiting distributions, Ann. Math. Statist. 43 (1972) 1312–1316.