LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

M. Rauf Ahmad

Tests for Covariance Matrices, particularly for High-dimensional Data

Technical Report Number 091, 2010
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Tests for Covariance Matrices, Particularly for High Dimensional Data

M. Rauf Ahmad

Department of Medical Informatics, Biometry and Epidemiology

University of Munich, Munich, Germany

Abstract

Test statistics for sphericity and identity of the covariance matrix are presented, when the data are multivariate normal and the dimension, p, can exceed the sample size, n. The test statistics are shown to follow an approximate normal distribution for large p, also when $p >> n$. The statistics are derived under very general conditions, particularly avoiding any strict assumptions on the traces of the unknown covariance matrix. Neither any relationship between n and p is assumed. The accuracy of the statistics is shown through simulation results, particularly emphasizing the case when p can be much larger than n.

Keywords: covariance testing; high dimensional data; sphericity

1 Introduction

The need for estimation and testing of high dimensional covariance matrix in multivariate set up has recently been galvanized by frequently encountered large data sets, particularly, but not necessarily limited to, genetics, microarray, and astronomy. The present manuscript focuses on two tests of covariance matrix when the data are high dimensional. Precisely, suppose

$$\mathbf{X}_k = (X_{k1}, \ldots, X_{kp})' \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

$k = 1, \ldots, n$, are n independent and identically distributed random vectors, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the mean vector and the covariance matrix, respectively. The objective is to test the hypotheses

$$H_{01} : \boldsymbol{\Sigma} = \sigma^2\mathbf{I} \qquad \text{vs} \qquad H_{11} : \boldsymbol{\Sigma} \neq \sigma^2\mathbf{I}$$

and

$$H_{02} : \boldsymbol{\Sigma} = \mathbf{I} \qquad \text{vs} \qquad H_{12} : \boldsymbol{\Sigma} \neq \mathbf{I},$$

particularly when $p > n$, where $\mathbf{I}$ is the identity matrix, and $\sigma^2 > 0$ is some constant. The first of these hypotheses, H_{01}, refers to the well-known sphericity hypothesis, whereas H_{02} is a convenient representation of a more general hypothesis $\mathbf{\Sigma} = \mathbf{\Sigma}_0$, where $\mathbf{\Sigma}_0$ is any known positive definite covariance matrix.

There has been some interesting work on the tests of H_{01} and H_{02}, under high-dimensional set up, over the last few years. Ledoit and Wolf (2002) discuss the validity of the test statistics

$$U = \frac{1}{p}\text{tr}\left(\frac{\mathbf{S}}{\frac{1}{p}\text{tr}(\mathbf{S})} - \mathbf{I}\right)^2, \qquad \text{and} \qquad V = \frac{1}{p}\text{tr}(\mathbf{S} - \mathbf{I})^2, \tag{1}$$

for H_{01} and H_{02}, respectively, under high-dimensional settings, where $\mathbf{S}$ is the sample estimator of $\mathbf{\Sigma}$, and tr denotes the trace. For $n \to \infty$, and p fixed, i.e., under n-asymptotics, the statistics U (John, 1971) and V (Nagao, 1973) provide asymptotically locally most powerful invariant tests, under normality.

Ledoit and Wolf examine the behavior of these statistics when $p > n$, under some additional assumptions, particularly assuming that $\frac{p}{n} \to c \in (0, \infty)$. They show that, under normality, a test based on U is still consistent, even if $p > n$, but the same is not true for V. They then propose a modified version of V which is valid to test H_{02} under new assumptions. Birke and Dette (2005) extend the work of Ledoit and Wolf by considering the same test statistics for the extreme boundaries of concentration, i.e., when $\frac{p}{n} \to c \in [0, \infty]$. Birke and Dette show that the statistics are also valid for extreme cases, although their approximating normal distributions needed to be derived through an approach different from the usual delta method employed by Ledoit and Wolf.

Under similar assumptions on the traces of the covariance matrix, Srivastava (2005) proposed test statistics for spherical, identity, and diagonal covariance matrix. Assuming normality, the test statistics are shown to asymptotically follow a normal distribution when $p > n$. The robustness of the same test statistics to normality is then evaluated in Srivastava et al. (2011) under more strict assumptions on the traces, and putting restrictions on the moments of the underlying distribution. Chen et al (2010) have recently proposed tests for H_{01} and H_{02}, assuming vanishing trace ratios of the unknown covariance matrix. They do not assume normality, but place several restrictions, close to normality, on the moments of the underlying multivariate model.

In the sections to follow, new statistics for the hypotheses H_{01} and H_{02} are proposed under very general conditions, specifically trying to avoid the strict trace assumptions on the covariance matrix, and putting any relationship between n and p. The accuracy of the statistics is shown to remain intact under the most natural settings of the traces of the covariance matrices. The rest of the article is organized as follows.

The test statistics are proposed in the next section. The performance of the statistics is shown through simulation studies in Section 3. Section 4 summarizes the results, and main theoretical derivations are collected in Appendix.

2 The proposed statistics

Let $\mathbf{X}_k = (X_{k1}, \ldots, X_{kp})' \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$, $k = 1, \ldots, n$, be the model, as stated above. Without loss of generality, we assume $\boldsymbol{\mu} = \mathbf{0}$.

Consider the hypotheses H_{01} and H_{02}, and the corresponding statistics U and V in (1). They are computed using the sample covariance matrix $\mathbf{S}$ as a plug-in estimator of $\boldsymbol{\Sigma}$ in

$$\frac{1}{p}\text{tr}\left(\frac{\boldsymbol{\Sigma}}{\frac{1}{p}\text{tr}(\boldsymbol{\Sigma})} - \mathbf{I}\right)^2 = \frac{\frac{1}{p}\text{tr}(\boldsymbol{\Sigma}^2)}{\left[\frac{1}{p}\text{tr}(\boldsymbol{\Sigma})\right]^2} - 1 = \frac{p\text{tr}(\boldsymbol{\Sigma}^2)}{[\text{tr}(\boldsymbol{\Sigma})]^2} - 1 \tag{2}$$

and

$$\frac{1}{p}\text{tr}(\boldsymbol{\Sigma} - \mathbf{I})^2 = \frac{1}{p}\text{tr}(\boldsymbol{\Sigma}^2) - \frac{2}{p}\text{tr}(\boldsymbol{\Sigma}) + 1, \tag{3}$$

respectively. To study the behavior of (2) and (3), we need estimators of $\text{tr}(\boldsymbol{\Sigma})$, $[\text{tr}(\boldsymbol{\Sigma})]^2$, and $\text{tr}(\boldsymbol{\Sigma}^2)$, and for the use of the statistics for high dimensional set up, the estimators must be consistent for large p, even if $p > n$. For $\mathbf{X}_k$, $k = 1, \ldots, n$, define $A_k = \mathbf{X}_k'\mathbf{X}_k$ as a quadratic form, and $A_{kl} = \mathbf{X}_k'\mathbf{X}_l$, $k \neq l$, as a symmetric bilinear form. The estimators of the three traces, $\text{tr}(\boldsymbol{\Sigma})$, $[\text{tr}(\boldsymbol{\Sigma})]^2$, and $\text{tr}(\boldsymbol{\Sigma}^2)$, are given in the following definition.

Definition 2.1. *Let A_k and A_{kl} be as defined above. Then*

$$E_1 = \frac{1}{n}\sum_{k=1}^{n} A_k, \tag{4}$$

$$E_2 = \frac{1}{n(n-1)} \underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}}_{k\neq l} A_k A_l, \tag{5}$$

$$E_3 = \frac{1}{n(n-1)} \underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}}_{k\neq l} A_{kl}^2. \tag{6}$$

are the estimators of $tr(\boldsymbol{\Sigma})$, $[tr(\boldsymbol{\Sigma})]^2$, and $tr(\boldsymbol{\Sigma}^2)$, respectively.

Note that, E_1, E_2 and E_3 are moment estimators of the respective traces, as opposed to the plug-in estimators using sample covariance matrix $\mathbf{S}$, given in (1). The desired properties of the estimators are established in Lemma 2.2 below. The estimators are already discussed in Ahmad et al (2008, Equation 6), in the context of presenting a statistic for mean testing in high dimensional longitudinal data, where it is proved that the estimators are unbiased and consistent, and the consistency remains intact even if the dimension exceeds the sample size. It is also interesting to note that the estimators in Equations (4)-(6) are very closely related to the estimators used in Chen et al (2010) for the construction of test statistics for H_{01} and H_{02}; see also Section 4.

The following lemma, proved in Appendix B, summarizes the basic properties of the estimators given in Definition 2.1.

Lemma 2.2. *Let the estimators E_1, E_2 and E_3, be as given in Definition 2.1. Then*

$$\begin{aligned}
E(E_1) &= tr(\boldsymbol{\Sigma}), \qquad Var(E_1) = \frac{2}{n} tr(\boldsymbol{\Sigma}^2),\\
E(E_2) &= [tr(\boldsymbol{\Sigma})]^2, \quad Var(E_2) = \frac{8}{n(n-1)} \left[(n-1) tr(\boldsymbol{\Sigma}^2)[tr(\boldsymbol{\Sigma})]^2 + \left[tr(\boldsymbol{\Sigma}^2)\right]^2\right]\\
E(E_3) &= tr(\boldsymbol{\Sigma})^2, \qquad Var(E_3) = \frac{4}{n(n-1)} \left[(2n-1) tr(\boldsymbol{\Sigma}^4) + \left[tr(\boldsymbol{\Sigma}^2)\right]^2\right],\\
Cov(E_1, E_2) &= \frac{4}{n} tr(\boldsymbol{\Sigma}) tr(\boldsymbol{\Sigma}^2),\\
Cov(E_1, E_3) &= \frac{4}{n} tr(\boldsymbol{\Sigma}^3),\\
Cov(E_2, E_3) &= \frac{8}{n(n-1)} \left[tr(\boldsymbol{\Sigma}^4) + (n-1) tr(\boldsymbol{\Sigma}) tr(\boldsymbol{\Sigma}^3)\right],
\end{aligned}$$

where $\frac{Cov(E_i,E_j)}{E(E_i)E(E_j)} \leq O\left(\frac{1}{n}\right)$, $\forall\ i, j$, whether $i = j$ or $i \neq j$, i.e., the (co)variance ratios are uniformly bounded in p.

In the next section, the test statistics for H_{01} and H_{02} are constructed, based on Definition 2.1 and the results of Lemma 2.2.

2.1 Test statistics for H_{01} and H_{02}

We use estimators given in Definition 2.1 to define the estimator of Equation (2) as

$$T_1 = \frac{pE_3}{E_2} - 1 \;=\; \widehat{\psi} - 1, \tag{7}$$

where $\widehat{\psi} = \frac{pE_3}{E_2}$ estimates $\psi = \frac{p\mathrm{tr}(\boldsymbol{\Sigma}^2)}{[\mathrm{tr}(\boldsymbol{\Sigma})]^2}$. To compute asymptotic distribution of T_1, we need to compute the moments of T_1. As T_1 involves a ratio of two correlated estimators, exact moments of T_1 can not be computed. However, a reasonable approximation of the first two moments can be obtained using the bivariate delta method of moments (Lehmann, 1999). The moment approximation is based on a bivariate Taylor expansion, and the remainders of the two moments vanish for large n (Casella and Berger, 2002, Ch. 5; Stuart and Ord, 1994, Ch. 10); see also Ahmad et al (2008).

Since, our main focus is on the application of T_1 under high-dimensional set up, we approach the problem of asymptotic normality from a slightly different, and relatively simpler way. For this, we re-write T_1 as

$$\frac{T_1 + 1}{\psi} \;=\; \frac{p\mathrm{E}_3}{\mathrm{E}_2} \cdot \frac{[\mathrm{tr}(\boldsymbol{\Sigma})]^2}{p\mathrm{tr}(\boldsymbol{\Sigma}^2)} \;=\; \frac{\mathrm{E}_3/\mathrm{tr}(\boldsymbol{\Sigma}^2)}{\mathrm{E}_2/[\mathrm{tr}(\boldsymbol{\Sigma})]^2}. \tag{8}$$

Following the same lines, to compute the test statistic for H_{02}, we plug in the relevant estimators from Definition 2.1 into Equation (3), and get

$$T_2 \;=\; \frac{1}{p} E_3 - \frac{2}{p} E_1 + 1. \tag{9}$$

Clearly, T_2 is an unbiased estimator of Equation (3), or pT_2 is an unbiased estimator of $\mathrm{tr}(\mathbf{\Sigma}-\mathbf{I})^2$. As T_2 is simply a linear combination of the estimators, computation of the moments of T_2 is trivial. A closer look at Equation (9) clues to the fact that the asymptotic normality of T_1 and T_2 can be shown simultaneously. We write,

$$\begin{aligned} T_2 - \frac{1}{p}\mathrm{tr}(\mathbf{\Sigma}-\mathbf{I})^2 &= \frac{1}{p}[\mathrm{E}_3 - \mathrm{tr}(\mathbf{\Sigma}^2)] - \frac{2}{p}[\mathrm{E}_1 - \mathrm{tr}(\mathbf{\Sigma})] \\ &= \frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{p}\left(\frac{\mathrm{E}_3}{\mathrm{tr}(\mathbf{\Sigma}^2)} - 1\right) - \frac{2\mathrm{tr}(\mathbf{\Sigma})}{p}\left(\frac{\mathrm{E}_1}{\mathrm{tr}(\mathbf{\Sigma})} - 1\right) \end{aligned} \tag{10}$$

Now, the ratios of estimators to their corresponding traces in Equations (8) and (10) are in the form of U-statistics (Lehmann, 1999, Ch. 6). The asymptotic theory of U-statistics is, therefore, used to establish the approximate normality of the test statistics. The proofs of the following two theorems follow from the asymptotic theory of U-statistics, and are omitted; for details, see Koroljuk and Borovskich (1994), and Lehmann (1999, Ch. 6).

Theorem 2.3. *Let T_1 be as defined in Equation (7). Then,*

$$\sigma^{-1}\left(\frac{T_1+1}{\psi} - 1\right) \xrightarrow{\mathcal{D}} N(0,1), \tag{11}$$

as $p, n \to \infty$, where σ^2 is the variance of the statistic (see Equation 8). In particular, under H_{01},

$$\tfrac{n}{2}T_1 \xrightarrow{\mathcal{D}} N(0,1).$$

Theorem 2.4. *Let T_2 be as defined in Equation (7). Then,*

$$\sigma^{-1}\left[T_2 - \frac{1}{p}tr(\mathbf{\Sigma}-\mathbf{I})^2\right] \xrightarrow{\mathcal{D}} N(0,1), \tag{12}$$

as $p, n \to \infty$, where $\sigma^2_{T_2}$ is the variance of the statistic (see Equation 10). In particular, under H_{02},

$$\tfrac{n}{2}T_2 \xrightarrow{\mathcal{D}} N(0,1).$$

Some remarks regarding the two test statistics are in order. First, the null distributions of both statistics are same. Further, these null distributions are same as established by other researchers, for the same hypotheses; see, for example, Ledoit and Wolf (2002), Srivastava (2005), Chen et al (2010). But, it can be emphasized that the distributions presented in Theorems 2.3 and 2.4 do not depend on any strict assumptions, for example regarding traces of the covariance matrix, or any relationship between n and p. The distributions are derived under general conditions, particularly focusing the case when p can far exceed n, as the practical high-dimensional situations demand. In this context, it must be noted, as is also verified through simulations in Section 3, that the approximating distributions are essentially developed for $p \to \infty$, disregarding how large n is.

3 Simulation results

Table 1: Estimated Quantiles for T_1 and T_2

(T_1)		p							
n	$1-\alpha$	5	10	20	50	100	200	500	1000
10	0.90	0.913	0.900	0.900	0.906	0.908	0.900	0.904	0.906
	0.95	0.949	0.959	0.956	0.951	0.956	0.950	0.951	0.953
	0.99	0.986	0.987	0.985	0.987	0.989	0.985	0.989	0.989
50	0.90	0.916	0.913	0.908	0.900	0.903	0.900	0.900	0.900
	0.95	0.954	0.955	0.954	0.952	0.95	0.950	0.952	0.949
	0.99	0.983	0.989	0.990	0.990	0.989	0.990	0.989	0.989
(T_2)		p							
n	$1-\alpha$	5	10	20	50	100	200	500	1000
10	0.90	0.916	0.909	0.900	0.905	0.908	0.904	0.904	0.905
	0.95	0.952	0.948	0.949	0.951	0.951	0.950	0.951	0.952
	0.99	0.980	0.978	0.983	0.986	0.986	0.988	0.987	0.989
50	0.90	0.919	0.902	0.902	0.900	0.900	0.898	0.899	0.900
	0.95	0.956	0.953	0.950	0.947	0.950	0.948	0.949	0.948
	0.99	0.984	0.985	0.989	0.988	0.990	0.988	0.992	0.989

Table 2: Power of Test for T_1 and T_2

(T_1)		p							
$\boldsymbol{\Sigma}$	n	5	10	20	50	100	200	500	1000
CS	10	0.759	0.937	0.989	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AR(1)	10	0.707	0.844	0.907	0.955	0.967	0.970	0.970	0.972
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(T_2)		p							
$\boldsymbol{\Sigma}$	n	5	10	20	50	100	200	500	1000
CS	10	0.650	0.894	0.980	0.999	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AR(1)	10	0.578	0.762	0.867	0.940	0.956	0.967	0.971	0.973
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 1 reports estimated quantiles for T_1 (upper panel) and T_2 (lower panel) for different pairs of n and p, and for three nominal quantiles 0.9, 0.95 and 0.99. The small-n, large-p accuracy of both test statistics can be clearly evidenced from the estimated values for $n = 10$ where $p = 1000$. As expected, the accuracy of the statistics increases for increasing n, against any p, but this accuracy is not damaged for any fixed n when p is allowed to increase. This validates the high-dimensional consistency of the estimators used to construct the test statistics. We observe an increasing stability of the estimated quantiles for increasing n and p, where T_1 slightly outperforms T_2.

Table 2 reports power values for the two tests for the same pairs of n and p as used for quantile estimation, and under compound symmetry (CS) and first-order autoregressive, AR(1), covariance structures, where the nominal quantile is fixed at 0.95. A CS covariance structure is defined as $\mathbf{\Sigma} = \sigma^2 [(1-\rho)\mathbf{I} + \rho\mathbf{J}]$, where $\mathbf{I}$ is identity matrix, $\mathbf{J}$ is a matrix of 1s, and σ^2 and ρ are appropriate constants. A covariance structure is AR(1) if $\text{Cov}(X_k, X_l) = \sigma^2\rho^{|k-l|}$, $\forall\ k, l$. For the computations reported in Table 2, it is assumed that $\sigma^2 = 1$, $\rho = 0.5$ for CS, and $\sigma^2 = 1$, $\rho = 0.6$ for AR(1).

It is observed that the power of both test statistics increases for increasing p for any n, and also for increasing n against any p. The highest power is generated by the compound symmetric covariance pattern, followed by autoregressive pattern. In general, it can be concluded that the test statistics have high power for moderate n, say 10 or more, for any $p \lesseqgtr n$. It must be noted that the power pattern of the statistics for autoregressive covariance structure does not vary significantly by changing ρ. The assumed $\rho = 0.6$ was taken only as a moderate example value, but it was verified that similar results are produced by smaller or larger values of the correlation coefficient.

The Autoregressive structure gives relatively low power, particularly for small sample size, and power increases slowly for increasing n and p, but still for n as moderate as 10 the power is high and increases for increasing dimension. The compound symmetric structure is the closest violation of the null hypothesis since any compound symmetric matrix can be orthonormally transformed to a spherical matrix.

4 Summary and conclusions

Test statistics for sphericity and identity of high dimensional covariance matrix are presented, under normality. The statistics, based on unbiased and consistent estimators, follow approximate normal distribution, and are also valid when the data are not high dimensional. The statistics are computed under very general conditions, and do not require any specific assumptions regarding the underlying covariance matrix. Further, no relationship between n and p is assumed. Simulation results show that the statistics accurately control test size and have high power even when the dimension is much larger than the sample size. The power properties of the statistics are demonstrated to remain intact under a variety of alternative hypothesis. The general behavior of the statistics is that they are accurate, both for size control and power, for a moderate sample size and any dimension.

Acknowledgement

The work on this manuscript was done during the author's short research stay at IBE, LMU, Munich. For this, the author is indebted to Prof. Anne-Laure Boulesteix for providing an opportunity, and a friendly, working atmosphere at the institute. The author is also grateful to DAAD (German Academic Exchange Service) for financial support of this research work.

A Some basic results

For the proof of Lemma 2.2, we need the following basic results.

Theorem A.1. *(Mathai and Provost (1992, Ch. 3); Mathai et al (1995, Ch. 2)) Let* $\mathbf{u}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$*, and* $\mathbf{v}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ $\mathbf{\Sigma} > 0$*, be random vectors, and* $\mathbf{A}$ *be any symmetric matrix. Define* $Q = \mathbf{u}'\mathbf{A}\mathbf{u}$ *as a quadratic form, and* $B = \mathbf{u}'\mathbf{A}\mathbf{v}$ *as a symmetric bilinear form. Then, the rth cumulants of Q and B are given as*

$$\begin{aligned}\kappa_r(Q) &= 2^{r-1}(r-1)!tr(\mathbf{A\Sigma})^r, \quad r = 1, 2, \ldots, \\ \kappa_r(B) &= \frac{1}{2}(r-1)!tr(\mathbf{A\Sigma})^r, \quad r = 1, 2, \ldots,\end{aligned}$$

respectively, where tr denotes the trace. Particularly, for r = 1 and 2, we have

$$\begin{aligned}E(Q) &= tr(\mathbf{A\Sigma}), \quad Var(Q) = 2tr(\mathbf{A\Sigma})^2 \\ E(B) &= 0, \qquad\quad Var(B) = tr(\mathbf{A\Sigma})^2,\end{aligned}$$

assuming $\mathbf{u}$ *and* $\mathbf{v}$ *are independent (see also Ahmad et al, 2008).*

Lemma A.2. *(Magnus, 1978, Lemma 6.2, page 209) Let* $\mathbf{u}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$*,* $\mathbf{\Sigma} > 0$ *be a random vector, and* $\mathbf{A}$ *and* $\mathbf{B}$ *be two symmetric matrices. Define* $\mathbf{u}'\mathbf{A}\mathbf{u}$ *and* $\mathbf{u}'\mathbf{B}\mathbf{u}$ *be the quadratic forms. Then, the mean and variance of the product* $z = \mathbf{u}'\mathbf{A}\mathbf{u} \cdot \mathbf{u}'\mathbf{B}\mathbf{u}$ *are given as*

$$\begin{aligned}E(z) &= tr(\mathbf{A\Sigma})tr(\mathbf{B\Sigma}) + 2tr(\mathbf{A\Sigma B\Sigma}) \\ Var(z) &= 32tr\left[(\mathbf{A\Sigma})^2(\mathbf{B\Sigma})^2\right] \\ &+ 16\left[tr(\mathbf{A\Sigma B\Sigma})^2 + tr(\mathbf{A\Sigma})tr\{(\mathbf{A\Sigma})(\mathbf{B\Sigma})^2\} + tr(\mathbf{B\Sigma})tr\{(\mathbf{A\Sigma})^2(\mathbf{B\Sigma})\}\right] \\ &+ 4\left[tr(\mathbf{A\Sigma})^2 tr(\mathbf{B\Sigma})^2 + [tr(\mathbf{A\Sigma B\Sigma})]^2 + tr(\mathbf{A\Sigma})tr(\mathbf{B\Sigma})tr(\mathbf{A\Sigma B\Sigma})\right] \\ &+ 2\left[[tr(\mathbf{A\Sigma})]^2 tr(\mathbf{B\Sigma})^2 + [tr(\mathbf{B\Sigma})]^2 tr(\mathbf{A\Sigma})^2\right].\end{aligned}$$

B Proof of Lemma 2.2

For unbiasedness and consistency of the estimators, see Ahmad et al (2008, Appendix B). Then, we work on the covariances. We have

$$\mathrm{Cov}(E_2, E_3) = \frac{1}{n^2(n-1)^2} \underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}}_{k \neq l} \underbrace{\sum_{r=1}^{n}\sum_{s=1}^{n}}_{r \neq s} \mathrm{Cov}(A_k A_l, A_{rs}^2), \qquad (13)$$

where, from Theorem A.1, $\mathrm{E}(A_k A_l) = [\mathrm{tr}(\mathbf{\Sigma})]^2$, $\mathrm{E}(A_{rs}^2) = \mathrm{tr}(\mathbf{\Sigma}^2)$, for $k \neq l$ and $r \neq s$. The covariance vanishes when $k \neq r, l \neq s$. The remaining cases are: $k = r, l = s; k = l, r \neq s; k \neq r, l = s; k \neq r, l \neq s$. For $k = l, r \neq s$

$$\begin{aligned}\mathrm{E}(A_k A_l A_{ks}^2) &= \mathrm{E}(A_l)\mathrm{E}(A_k A_{ks}^2) = \mathrm{tr}(\mathbf{\Sigma})\mathrm{E}[\mathbf{X}_k'\mathbf{X}_k\mathbf{X}_k'\mathbf{X}_s\mathbf{X}_k'\mathbf{X}_s] \\ &= \mathrm{tr}(\mathbf{\Sigma})\mathrm{E}[\mathrm{tr}\{\mathbf{X}_k\mathbf{X}_k'\mathbf{X}_k\mathbf{X}_k'\mathbf{X}_s\mathbf{X}_s'\}] = \mathrm{tr}(\mathbf{\Sigma})\mathrm{E}[\mathbf{X}_k'\mathbf{X}_k\mathbf{X}_k'\mathbf{\Sigma}\mathbf{X}_k] \\ &= 2\mathrm{tr}(\mathbf{\Sigma})\mathrm{tr}(\mathbf{\Sigma}^3) + [\mathrm{tr}(\mathbf{\Sigma})]^2\mathrm{tr}(\mathbf{\Sigma}^2),\end{aligned}$$

so that $\mathrm{Cov}(A_k A_l, A_{ks}^2) = 2\mathrm{tr}(\mathbf{\Sigma})\mathrm{tr}(\mathbf{\Sigma}^3)$. For $k = r, l = s$, write

$$A_{kl}^2(A_k + A_l)^2 \;=\; A_{kl}^2 A_k^2 + A_{kl}^2 A_l^2 + 2A_{kl}^2 A_k A_l,$$

so that

$$\mathrm{E}(A_k A_l A_{kl}^2) \;=\; \frac{1}{2}\Big[\mathrm{E}[A_{kl}^2(A_k + A_l)^2] - \mathrm{E}(A_{kl}^2 A_k^2) - \mathrm{E}(A_{kl}^2 A_l^2)\Big]. \tag{14}$$

Let $\mathbf{Z} = (\mathbf{X}_k \mathbf{X}_l)'$, with $\mathrm{Cov}(\mathbf{Z}) = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \end{pmatrix} = \mathbf{I}_2 \otimes \mathbf{\Sigma} = \mathbf{V}$(say). Then $A_{kl} = \frac{1}{2}\mathbf{Z}'\mathbf{A}\mathbf{Z}$, $A_k + A_l = \mathbf{Z}'\mathbf{B}\mathbf{Z}$, $A_k = \mathbf{Z}'\mathbf{C}\mathbf{Z}$, and $A_l = \mathbf{Z}'\mathbf{D}\mathbf{Z}$, where

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix},\ \mathbf{B} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},\ \mathbf{C} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},\ \text{and } \mathbf{D} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$

see Mathai (1992), and Mathai et al (1995, p 19). We need the following traces.

$\mathrm{tr}(\mathbf{AV}) = 0$; $\mathrm{tr}(\mathbf{BV}) = \mathrm{tr}(\mathbf{V}) = 2\mathrm{tr}(\mathbf{\Sigma})$; $\mathrm{tr}(\mathbf{CV}) = \mathrm{tr}(\mathbf{\Sigma}) = \mathrm{tr}(\mathbf{DV})$;

$\mathrm{tr}(\mathbf{AV})^2 = 2\mathrm{tr}(\mathbf{\Sigma}^2) = \mathrm{tr}(\mathbf{BV})^2$; $\mathrm{tr}(\mathbf{CV})^2 = \mathrm{tr}(\mathbf{\Sigma}^2) = \mathrm{tr}(\mathbf{DV})^2$;

$\mathrm{tr}(\mathbf{AVBV}) = 0 = \mathrm{tr}(\mathbf{AVCV}) = \mathrm{tr}(\mathbf{AVDV})$;

$\mathrm{tr}(\mathbf{AVBV})^2 = 2\mathrm{tr}(\mathbf{\Sigma}^4)$; $\mathrm{tr}(\mathbf{AVCV})^2 = 0 = \mathrm{tr}(\mathbf{AVDV})^2$;

$\mathrm{tr}[(\mathbf{AV})^2(\mathbf{BV})] = 2\mathrm{tr}(\mathbf{\Sigma}^3)$; $\mathrm{tr}[(\mathbf{AV})(\mathbf{BV})^2] = 0$; $\mathrm{tr}[(\mathbf{AV})^2(\mathbf{BV})^2] = 2\mathrm{tr}(\mathbf{\Sigma}^4)$;

$\mathrm{tr}[(\mathbf{AV})^2(\mathbf{CV})] = \mathrm{tr}(\mathbf{\Sigma}^3)$; $\mathrm{tr}[(\mathbf{AV})(\mathbf{CV})^2] = 0$; $\mathrm{tr}[(\mathbf{AV})^2(\mathbf{CV})^2] = \mathrm{tr}(\mathbf{\Sigma}^4)$;

$\mathrm{tr}[(\mathbf{AV})^2(\mathbf{DV})] = \mathrm{tr}(\mathbf{\Sigma}^3)$; $\mathrm{tr}[(\mathbf{AV})(\mathbf{DV})^2] = 0$; $\mathrm{tr}[(\mathbf{AV})^2(\mathbf{DV})^2] = \mathrm{tr}(\mathbf{\Sigma}^4)$

Since many of the traces vanish, the moments of Lemma A.2 reduce to the following simple forms.

$$\begin{aligned} \mathrm{E}(z) &= 0 \\ \mathrm{E}(z^2) &= 32\mathrm{tr}\left[(\mathbf{A\Sigma})^2(\mathbf{B\Sigma})^2\right] + 16\left[\mathrm{tr}(\mathbf{A\Sigma B\Sigma})^2 + \mathrm{tr}(\mathbf{B\Sigma})\mathrm{tr}\{(\mathbf{A\Sigma})^2(\mathbf{B\Sigma})\}\right] \\ &\quad + 4\mathrm{tr}(\mathbf{A\Sigma})^2\mathrm{tr}(\mathbf{B\Sigma})^2 + 2[\mathrm{tr}(\mathbf{B\Sigma})]^2\mathrm{tr}(\mathbf{A\Sigma})^2. \end{aligned} \tag{15}$$

Using these results, we obtain the following moments.

$$\begin{aligned} \mathrm{E}(\mathbf{Z}'\mathbf{AZ}\cdot\mathbf{Z}'\mathbf{BZ})^2 &= 96\mathrm{tr}(\mathbf{\Sigma}^4) + 64\mathrm{tr}(\mathbf{\Sigma}^3)\mathrm{tr}(\mathbf{\Sigma}) + 16\left[\mathrm{tr}(\mathbf{\Sigma}^2)\right]^2 + 16[\mathrm{tr}(\mathbf{\Sigma})]^2\mathrm{tr}(\mathbf{\Sigma}^2) \\ \mathrm{E}(\mathbf{Z}'\mathbf{AZ}\cdot\mathbf{Z}'\mathbf{CZ})^2 &= 32\mathrm{tr}(\mathbf{\Sigma}^4) + 16\mathrm{tr}(\mathbf{\Sigma}^3)\mathrm{tr}(\mathbf{\Sigma}) + 8\left[\mathrm{tr}(\mathbf{\Sigma}^2)\right]^2 + 4[\mathrm{tr}(\mathbf{\Sigma})]^2\mathrm{tr}(\mathbf{\Sigma}^2) \\ \mathrm{E}(\mathbf{Z}'\mathbf{AZ}\cdot\mathbf{Z}'\mathbf{DZ})^2 &= \mathrm{E}(\mathbf{Z}'\mathbf{AZ}\cdot\mathbf{Z}'\mathbf{CZ})^2 \end{aligned}$$

which gives, from Equation (14),

$$\mathrm{E}(A_k A_l A_{kl}^2) \;=\; 4\mathrm{tr}(\mathbf{\Sigma}^4) + 4\mathrm{tr}(\mathbf{\Sigma}^3)\mathrm{tr}(\mathbf{\Sigma}) + [\mathrm{tr}(\mathbf{\Sigma})]^2\mathrm{tr}(\mathbf{\Sigma}^2), \tag{16}$$

so that $\mathrm{Cov}(A_k A_l, A_{kl}^2) = 4\mathrm{tr}(\mathbf{\Sigma}^4) + 4\mathrm{tr}(\mathbf{\Sigma}^3)\mathrm{tr}(\mathbf{\Sigma})$. Finally,

$$\mathrm{Cov}(E_2, E_3) \;=\; \frac{8}{n(n-1)}\Big[\mathrm{tr}(\mathbf{\Sigma}^4) + (n-1)\mathrm{tr}(\mathbf{\Sigma}^3)\mathrm{tr}(\mathbf{\Sigma})\Big]. \tag{17}$$

Further,

$$\frac{\mathrm{Cov}(E_2, E_3)}{[\mathrm{tr}(\mathbf{\Sigma})]^2\mathrm{tr}(\mathbf{\Sigma}^2)} \quad \le \quad \frac{8}{n-1} \quad = \; O\left(\frac{1}{n}\right).$$

Now,

$$\mathrm{Cov}(E_3, E_1) \;\; = \;\; \frac{1}{n^2(n-1)} \underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}}_{k \,\neq\, (l,m)}{}_{l\neq m} \mathrm{Cov}(A_k, A_{lm}^2), \tag{18}$$

where $\mathrm{Cov}(A_k, A_{lm}^2) = \mathrm{E}(A_k A_{lm}^2) - \mathrm{tr}(\mathbf{\Sigma})\mathrm{tr}(\mathbf{\Sigma}^2)$, from Theorem A.1. The covariance is zero for $k \neq l \neq m$. The other two cases, $k = l \neq m$ and $k = m \neq l$ yield the same result. Then, for $k = l \neq m$, we have

$$\begin{aligned} \mathrm{E}(A_k A_{lm}^2) \;\; &= \;\; \mathrm{E}(\mathbf{X}_k'\mathbf{X}_k\mathbf{X}_k'\mathbf{X}_m\mathbf{X}_k'\mathbf{X}_m) \;\; = \;\; \mathrm{E}[\mathrm{tr}(\mathbf{X}_k\mathbf{X}_k'\mathbf{X}_k\mathbf{X}_k'\mathbf{X}_m\mathbf{X}_m')] \\ &= \;\; \mathrm{E}(\mathbf{X}_k'\mathbf{X}_k\mathbf{X}_k'\mathbf{\Sigma}\mathbf{X}_k) \;\; = \;\; 2\mathrm{tr}(\mathbf{\Sigma}^3) + \mathrm{tr}(\mathbf{\Sigma})\mathrm{tr}(\mathbf{\Sigma}^2), \end{aligned}$$

from Lemma A.2, so that $\mathrm{Cov}(A_k, A_{lm}^2) = 2\mathrm{tr}(\mathbf{\Sigma}^3)$, and from Equation (20)

$$\mathrm{Cov}(E_3, E_1) \;\; = \;\; \frac{1}{n^2(n-1)}\left[2n(n-1)\{2\mathrm{tr}(\mathbf{\Sigma}^3)\}\right] \;\; = \;\; \frac{4}{n}\mathrm{tr}(\mathbf{\Sigma}^3), \tag{19}$$

where,

$$\frac{\mathrm{Cov}(E_3, E_1)}{\mathrm{tr}(\mathbf{\Sigma}^2)\mathrm{tr}(\mathbf{\Sigma})} \quad = \quad \frac{4}{n}\left(\frac{\mathrm{tr}(\mathbf{\Sigma}^3)}{\mathrm{tr}(\mathbf{\Sigma}^2)\mathrm{tr}(\mathbf{\Sigma})}\right) \;\; \le \;\; \frac{4}{n} \;\; = \;\; O\left(\frac{1}{n}\right).$$

Finally,

$$\mathrm{Cov}(E_1, E_2) \;\; = \;\; \frac{1}{n^2(n-1)} \underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}}_{k \,\neq\, (l,m)}{}_{l\neq m} \mathrm{Cov}(A_k, A_l A_m), \tag{20}$$

where $\mathrm{Cov}(A_k, A_l A_m) = \mathrm{E}(A_k A_l A_m) - [\mathrm{tr}(\mathbf{\Sigma})]^3$, from Theorem A.1. Then, working on the same lines as for $\mathrm{Cov}(E_3, E_1)$, it can be shown that

$$\mathrm{Cov}(E_1, E_2) \;\; = \;\; \frac{4}{n}\mathrm{tr}(\mathbf{\Sigma}^2)\mathrm{tr}(\mathbf{\Sigma}), \tag{21}$$

so that,

$$\frac{\mathrm{Cov}(E_1, E_2)}{\mathrm{tr}(\mathbf{\Sigma})[\mathrm{tr}(\mathbf{\Sigma})]^2} \;\; \le \;\; \frac{4}{n} \;\; = \;\; O\left(\frac{1}{n}\right).$$

References

Ahmad, MR, C Werner, and E Brunner (2008). Analysis of high dimensional repeated measures designs: The one sample case. *Computational Statistics & Data Analysis*, **53**, 416-427.

Birke, M, and H Dette (2005). A note on testing the covariance matrix for large dimension. *Statistics and Probability Letters*, **74**, 281-289.

Box, GEP (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems I: Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, **25**(2), 290-302.

Casella, G, and RL Berger (2002). *Statistical inference.* 2nd Ed., Duxbury Press, Pacific Grove, CA.

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

Fujikoshi, Y, VV Ulyanov, and R Shimizu (2010). *Multivariate statistics: High-dimensional and large-sample approximations.* Wiley, New York.

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

Koroljuk, VS, and YV Borovskich (1994). *Theory of U-statistics.* Kluwer Academic Press, Dordrecht.

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

Lehmann, EL (1999). *Elements of large-sample theory.* Springe, New York.

Magnus, JR (1978). The moments of products of quadratic forms in normal variables. *Statistical Neerlandica*, **32**, 201-210.

Mathai, AM (1992). On bilinear forms in normal variables. *Ann. Inst. Statist. Math.*, **44**(4), 769-779.

Mathai, AM, and SB Provost (1992). *Quadratic forms in random variables.* Marcel Dekker Inc., New York.

Mathai, AM, SB Provost. and T Hayakawa (1995). *Bilinear forms and zonal polynomials.* Lecture Notes in Statistics, No. 102. Springer, New York.

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

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

Srivastava, MS, T Kollo, and D von Rosen (2011). Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, **102**, 1090-1103.

Stuart, A. and J. K. Ord (1994). *Kendall's advanced theory of statistics*, Vol. 1: Distribution theory. 6th Edition. Arnold pubs., London.

Winer, BJ, DR Brown, and KM Michels (1991). *Statistical principles in experimental design.* Third edition. McGraw-Hill, New York.