

A note on mean testing for high dimensional multivariate data under non-normality

M. Rauf Ahmad, Dietrich von Rosen and Martin Singull

Linköping University Post Print



N.B.: When citing this work, cite the original article.

This is the pre-reviewed version of the following article:

M. Rauf Ahmad, Dietrich von Rosen and Martin Singull, A note on mean testing for high dimensional multivariate data under non-normality, 2013, Statistica neerlandica (Print), (67), 1, 81-99.

It has been published in final form at:

<http://dx.doi.org/10.1111/j.1467-9574.2012.00533.x>

Copyright: Wiley-Blackwell

<http://eu.wiley.com/WileyCDA/Brand/id-35.html>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-87379>

A note on mean testing for high dimensional multivariate data under non-normality

M. Rauf Ahmad^{1,2*}, Dietrich von Rosen^{1,3}, Martin Singull³

¹ *Department of Energy and Technology, Swedish University of Agricultural Sciences, Uppsala, Sweden*

² *Department of Statistics, Uppsala University, Uppsala, Sweden*

³ *Department of Mathematics, Linköping University, Sweden*

Abstract

A test statistic is considered for testing a hypothesis for the mean vector for multivariate data, when the dimension of the vector, p , may exceed the number of vectors, n , and the underlying distribution need not necessarily be normal. With $n, p \rightarrow \infty$, and under mild assumptions, but without assuming any relationship between n and p , the statistic is shown to asymptotically follow a chi-square distribution. A by product of the paper is the approximate distribution of a quadratic form, based on the reformulation of the well-known Box's approximation, under high-dimensional set up. Using a classical limit theorem, the approximation is further extended to an asymptotic normal limit under the same high dimensional set up. The simulation results, generated under different parameter settings, are used to show the accuracy of the approximation for moderate n and large p .

Keyword: Non-normality; High dimensionality; Box's approximation

1 Introduction

Suppose $\mathbf{X}_k = (X_{k1}, \dots, X_{kp})'$, $k = 1, \dots, n$, are independent, identically distributed random vectors. Assume

$$\mathbf{X}_k \sim \mathcal{F} \tag{1}$$

where \mathcal{F} denotes a p -variate distribution with $E(\mathbf{X}_k) = \boldsymbol{\mu}$ and $\text{Cov}(\mathbf{X}_k) = \boldsymbol{\Sigma}$, $\boldsymbol{\Sigma} > 0$. We are interested to test the hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ when $p > n$ and \mathcal{F} is not necessarily normal.

Recently, there have been some attempts to address the issue of high dimensionality, particularly for mean vector testing, with or without normality assumption. LÄUTER (2004), for example, develops one sample tests based on

* *Corresponding author.* Email: rauf.ahmad@slu.se. Tel.: 0046-18-673807

spherical distributions using principal components of the design matrix; see also LÄUTER *et al.* (1998). The test is constructed without subjecting the dimension or sample size to any serious restrictions; see also SRIVASTAVA and DU (2008). Under normality, a modified version of Hotelling's T^2 , using the Moore-Penrose inverse, is considered by SRIVASTAVA (2007), whereas a test for non-normal random vectors is given in SRIVASTAVA (2009).

Some attempts to extend the one sample case to two or more independent groups have also been made. The earliest developments for two independent samples can be traced back to DEMPSTER (1958), which was then more elaborated and discussed in detail in BAI and SARANADASA (1996). BATHKE (2002) deals with the classical ANOVA set up for large number of independent groups, whereas a completely nonparametric approach is given in BATHKE and HARRAR (2008). A comprehensive review of different multivariate approaches is given in BATHKE, HARRAR and AHMAD (2009). For a comparison of high dimensional mean testing statistics, under normality assumption, see KROPF *et al.* (2009).

Again under normality, a statistic is considered in AHMAD (2008, Ch. 2) to test a hypothesis of the form $H'_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, where \mathbf{T} is any matrix such that H'_0 is a general linear hypothesis, see also AHMAD *et al.* (2008). For $\mathbf{T} = \mathbf{I}$, where \mathbf{I} is the identity matrix, H'_0 reduces to H_0 , the hypothesis of our interest (see also Section 5). The test statistic is based on quadratic and bilinear forms composed of the random vectors \mathbf{X}_k , and is shown to follow a scaled chi-square distribution when $n \rightarrow \infty$ and p is fixed where p may exceed n . The present manuscript evaluates the same test statistic to test H_0 under the standard high dimensional, (n, p) -asymptotics, i.e., assuming both n and p to be large inclusive of the case when $p > n$, or even $p \gg n$. We, however, let n and p grow arbitrarily, and do not assume any relationship between them like, for example, $\frac{p}{n} \rightarrow c \in (0, \infty)$. Under very mild and practical assumptions, we use the asymptotic theory of U -statistics to show that the statistic still follows the same chi-square distribution asymptotically. Indirectly, it helps us establish the distribution of a quadratic form under high-dimensional set up, and hence re-state the well-known Box's approximation (BOX, 1954), based on the representation of a quadratic form as a weighted sum of single degree-of-freedom independent chi-square random variables.

The test statistic and its asymptotic distribution under the new set up are given in the next section. In Section 3, the performance of the modified approximation is evaluated through simulation studies for both test size and power. Section 4 is dedicated to a few important remarks regarding the proposed statistic. The results are summarized in Section 5.

2 The test statistic

For the model given in (1), $\bar{\mathbf{X}} = \frac{1}{n} \sum_{k=1}^n \mathbf{X}_k$ is an unbiased estimator of $\boldsymbol{\mu}$ and, under $H_0 : \boldsymbol{\mu} = \mathbf{0}$,

$$\mathbf{S} = \frac{1}{n} \sum_{k=1}^n \mathbf{X}_k \mathbf{X}_k' \quad (2)$$

is an unbiased estimators of $\boldsymbol{\Sigma}$. To test H_0 , consider the statistic (AHMAD *et al.*, 2008)

$$T = \frac{Q}{E_1}, \quad (3)$$

where

$$Q = n\bar{\mathbf{X}}'\bar{\mathbf{X}} \quad \text{and} \quad E_1 = \frac{1}{n} \sum_{k=1}^n A_k = \text{tr}(\mathbf{S}) \quad (4)$$

with $A_k = \mathbf{X}_k' \mathbf{X}_k$. Obviously, Q is the quadratic form of means, and E_1 is the mean of n independent quadratic forms. Since, $E(A_k) = \text{tr}(\boldsymbol{\Sigma})$, without normality assumption (SEARLE, 1971, p 55), where tr denotes the trace, therefore, E_1 is an unbiased estimator of $\text{tr}(\boldsymbol{\Sigma})$, and so is Q , by the same argument. Further moments of Q and E_1 , without normality assumption, are summarized in the following proposition for further reference.

Proposition 2.1. *For Q and E_1 , as defined in (4), we have*

$$E(Q) = \text{tr}(\boldsymbol{\Sigma}) = E(E_1) \quad (5)$$

$$\text{Var}(Q) = \text{Var}(E_1) + \frac{2(n-1)}{n} \text{tr}(\boldsymbol{\Sigma}^2) \quad (6)$$

$$\text{Cov}(Q, E_1) = \text{Var}(E_1). \quad (7)$$

The proof is quite trivial and is therefore omitted. In fact, except computing $\text{Var}(E_1)$ explicitly, the proof of Proposition 5 in AHMAD *et al.* (2008) holds without normality assumption (see also Section 4.2). Two main results determine the proof: $E(A_{kl}^2) = \text{tr}(\boldsymbol{\Sigma}^2)$ and $\text{Cov}(A_{kl}, A_m) = 0$, $k \neq l$, which can be trivially proved to hold without normality assumption (see also RAO and KLEFFE, 1988, p 38).

The formulation of T is a modification of the Wald statistic, $n\bar{\mathbf{X}}'\hat{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{X}}$, for testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ when p can exceed n , where $\hat{\boldsymbol{\Sigma}}$ is the maximum likelihood estimator of $\boldsymbol{\Sigma}$, under normality (see, for example, SERFLING, 1980, p 155). Note that, the Wald statistic (or Hotelling's T^2 statistic) is constructed by standardizing the norm of the mean, $\|\bar{\mathbf{X}}\|^2$, with an inverse of covariance matrix as a scaling factor. Since, for $p > n$, the problem stems from singularity of $\hat{\boldsymbol{\Sigma}}$, any statistic aimed at testing a hypothesis on the location parameter can not be constructed with such a standardization since the estimated covariance matrix is not invertible. The modification in (3) is, therefore, based on taking $\hat{\boldsymbol{\Sigma}}$ out of the Wald statistic and reformulating the statistic by norming it with the trace, instead of inverse, of the covariance matrix.

How much is the off-set we have to offer for the modification? For large n , and large but fixed p such that p may exceed n , a detailed study of T for a one-sample high-dimensional test for the mean vector in a repeated measures set up is given in AHMAD *et al.* (2008), where \mathcal{F} is assumed multivariate normal. It is shown, theoretically and also supplemented with simulation results, that

$$T \xrightarrow{\mathcal{D}} \frac{\chi_f^2}{f} \quad (8)$$

as $n \rightarrow \infty$, with $f = E_2/E_3$ as an estimator of $[\text{tr}(\mathbf{\Sigma})]^2/\text{tr}(\mathbf{\Sigma}^2)$, where

$$E_2 = \frac{1}{n(n-1)} \sum_{\substack{k=1 \\ k \neq l}}^n \sum_{l=1}^n A_k A_l, \quad E_3 = \frac{1}{n(n-1)} \sum_{\substack{k=1 \\ k \neq l}}^n \sum_{l=1}^n A_{kl}^2, \quad (9)$$

with A_k and A_l being quadratic forms, as defined above, and $A_{kl} = \mathbf{X}'_k \mathbf{X}_l$ is a bilinear form. The estimators, E_1 , E_2 and E_3 , are shown to be unbiased and consistent for the respective traces. Further, the chi-square approximation of T is based on the following representation theorem of a quadratic form, combined with the Box's approximation (Box, 1954).

Theorem 2.2. *Let $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ and let \mathbf{M} be any symmetric, positive semi-definite matrix with r non-zero eigenvalues, $r \leq p$. Then*

$$A = \mathbf{X}' \mathbf{M} \mathbf{X} \sim \sum_{i=1}^r \lambda_i C_i$$

where λ_i are the eigenvalues of $\mathbf{M}\mathbf{\Sigma}$ and the $C_i \sim \chi_1^2$ are independent.

As the Box's approximation calls for (Box, 1954, Theorem 3.1), we equate first two moments of A with those of a scaled, $g\chi_f^2$ distribution, i.e.

$$\begin{aligned} gf &= \sum_{i=1}^r \lambda_i \\ 2g^2f &= 2 \sum_{i=1}^r \lambda_i^2 + \left(\sum_{i=1}^r \lambda_i \right)^2. \end{aligned}$$

This gives

$$f = \frac{[\text{tr}(\mathbf{M}\mathbf{\Sigma})]^2}{\text{tr}(\mathbf{M}\mathbf{\Sigma})^2} \quad \text{and} \quad g = \frac{\text{tr}(\mathbf{M}\mathbf{\Sigma})^2}{\text{tr}(\mathbf{M}\mathbf{\Sigma})}, \quad (10)$$

where $\sum_{i=1}^r \lambda_i = \text{tr}(\mathbf{M}\mathbf{\Sigma})$ and $\sum_{i=1}^r \lambda_i^2 = \text{tr}(\mathbf{M}\mathbf{\Sigma})^2$. For more details, and a two-sample extension of T , see AHMAD (2008).

As the main objective of the present work, we evaluate T by relaxing the normality assumption, and show that its convergence to chi-square distribution still holds, and this approximation is valid even when $p \rightarrow \infty$, not just for fixed

p . This approximation of T , under high-dimensional, or (n, p) -asymptotics (see FUJIKOSHI *et al.*, 2010, Chs. 5 & 8), is established using the theory of U -statistics. Thereby, as a side objective, we show that Theorem 2.2 also holds under high-dimensional set up. Theorem 2.6 gives the asymptotic distribution of T . For the proof of this theorem, we need to set some assumptions.

Assumption 2.3. $E(X_{ks}^4) \leq \gamma < \infty, \forall s = 1, \dots, p$, for some finite constant γ independent of p .

Assumption 2.4. For $p \rightarrow \infty$, let $\frac{\text{tr}(\mathbf{\Sigma})}{p} = O(1)$.

Assumption 2.5. For $p \rightarrow \infty$, let $\frac{\text{tr}(\mathbf{\Sigma}^2)}{p^2} = O(\delta)$, where $0 < \delta \leq 1$.

Assumption 2.3 states that the moments of the elements of random vector \mathbf{X}_k , up to order 4, exist and are finite. The finite fourth moment assumption is a somewhat natural replacement of normality, particularly when dealing with the computations which involve second moment of quadratic forms; see for example, ATIQUILLAH (1962), RAO and KLEFFE (1988, Ch. 2), WIENS (1992), KNIGHT (1985), and SRIVASTAVA (2009), to mention a few. Moreover, as will be shown in the proof of Theorem 2.6, Assumptions 2.3 and 2.4 help us establish the convergence of $\frac{1}{p}E_1$, whereas Assumptions 2.4 and 2.5 are needed to ensure that the asymptotic distribution of the U -statistic, used to prove Theorem 2.6, remains non-degenerate, even when p grows large.

Assumption 2.5 deviates from its usually adopted form in the literature for high dimensional set up where it is assumed that $\text{tr}(\mathbf{\Sigma}^2)/p = O(1)$. In most cases it is assumed with even higher powers of $\mathbf{\Sigma}$, at least up to 4; see for example LEDOIT and WOLF (2002), SRIVASTAVA (2007), FISHER *et al.* (2010), and CHEN *et al.* (2010). But it can be easily verified that $\text{tr}(\mathbf{\Sigma}^2)/p$ shows a very bad behavior for many practical covariance structures when the dimension grows large. The ratio $\text{tr}(\mathbf{\Sigma}^2)/p$ actually diverges to infinity very quickly in certain most important cases. For our purposes, i.e., to prove Theorem 2.6, it suffices to assume $\text{tr}(\mathbf{\Sigma}^2)/p^2 > 0$, but for most of practical cases, it can be shown that $\text{tr}(\mathbf{\Sigma}^2)/p^2 = O(1)$, as p gets moderately large.

As a typical example, let $\mathbf{\Sigma} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$ be the compound symmetric matrix, where \mathbf{I} is the identity matrix, \mathbf{J} is the matrix of 1s, and ρ is a constant. Assume $\rho = 0.5$, so that $\mathbf{\Sigma} > 0$. Then, $\text{tr}(\mathbf{\Sigma}^i) = O(p^i)$, $i = 1, 2$, so that $\text{tr}(\mathbf{\Sigma}^2)/p^2 = O(1)$, for $p \rightarrow \infty$. Clearly, for $\rho = 0$, so that $\mathbf{\Sigma} = \mathbf{I}$, the ratio $\text{tr}(\mathbf{\Sigma}^2)/p^2$ will eventually vanish as $p \rightarrow \infty$. This is an extreme case which does not allow any correlational structure of the variables. In fact, it can be shown that Assumption 2.5 is conveniently satisfied for a variety of practically useful covariance matrices, for example, compound symmetric, autoregressive of order 1, or unstructured matrix (see also WOLFINGER, 1996), except for the cases when the covariance matrix is singular or near singular.

We do admit here that there are practical cases where this ratio may converge to 0 for very large p , but this is mostly the case when $\mathbf{\Sigma}$ is singular, and the more intense the singularity of $\mathbf{\Sigma}$ is, the more rapid is its convergence to zero; see also AHN *et al.* (2007). Since we want to avoid zero convergence case for

this ratio, we feel much safer to have Assumption 2.5 with divisor p^2 , instead of p . For a detailed study on the rationale of the assumptions frequently adopted in the current high-dimensional literature, see AHMAD *et al.* (2011).

Now, we prove the main theorem on the asymptotic distribution of T .

Theorem 2.6. *Assume $n, p \rightarrow \infty$. Then, under Assumptions 2.3 - 2.5, the test statistic, T , as given in Equation (3), is distributed as χ_f^2/f with f estimated as E_2/E_3 , where E_2 and E_3 are given in (9). Further, under the same set up, and using Hájek-Šidák Lemma (see Lemma 2.7 below), it can be shown that,*

$$\frac{T - E(T)}{\sqrt{\widehat{\text{Var}}(T)}} \xrightarrow{\mathcal{D}} N(0, 1),$$

where $E(T) = 1$, and $\widehat{\text{Var}}(T) = 2E_3/E_2$ is the sample estimator of $\text{Var}(T) = 2\text{tr}(\Sigma^2)/[\text{tr}(\Sigma)]^2$.

Proof Following the discussion around Theorem 2.2, we essentially need to prove that T can still be represented as a weighted sum of independent chi-square random variables, even when $p \rightarrow \infty$ and the distribution is not normal. For this, we first note that

$$\begin{aligned} n\overline{\mathbf{X}}'\overline{\mathbf{X}} &= \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n A_{kl} = \frac{1}{n} \sum_{k=1}^n A_k + \frac{1}{n} \sum_{k=1}^n \sum_{\substack{l=1 \\ k \neq l}}^n A_{kl} \\ &= E_1 + \frac{1}{n} \sum_{k=1}^n \sum_{\substack{l=1 \\ k \neq l}}^n A_{kl}, \end{aligned}$$

where $A_k = \mathbf{X}'_k \mathbf{X}_k$ and $A_{kl} = \mathbf{X}'_k \mathbf{X}_l$. Then, we can write T in (3) as

$$T = 1 + \frac{E_0}{\frac{1}{p}E_1}, \quad (11)$$

with

$$\begin{aligned} E_0 &= \frac{1}{n} \sum_{k=1}^n \sum_{\substack{l=1 \\ k \neq l}}^n \frac{1}{p} A_{kl} \\ &= (n-1) \left(\frac{1}{n(n-1)} \sum_{k=1}^n \sum_{\substack{l=1 \\ k \neq l}}^n \frac{1}{p} A_{kl} \right) = (n-1)U_n, \end{aligned} \quad (12)$$

where U_n is a U -statistic with the symmetric kernel $\frac{1}{p}A_{kl}$, $k \neq l$; see for example SHAO (2003, Ch. 3), and LEHMANN (1999, Ch. 6). The asymptotic theory of U -statistics, introduced by Hoeffding (1948), is quite rich and provides some

nice results, also for random processes with dependent random variables; see for example DENKER and KELLER (1983). A comprehensive review of the asymptotic results of U -statistics is given in DEHLING (2006); see also KOROLJUK and BOROVSKICH (1994). A general form of multivariate U -statistics is dealt with in SUGIURA (1965). For recent examples on the use of U -statistics theory in high-dimensional inference, see AHMAD *et al.* (2011), and ZHANG and CHEN (2011), whereas a brief motivation for the use of (degenerate) U -statistics theory to deal with high-dimensional asymptotics is given in Section 4 (Remark 1).

We shall exploit this theory and show that U_n , being an average of bilinear forms composed of independent random vectors, is approximately chi-square distributed even if $p \rightarrow \infty$. This result, apart from being helpful in proving Theorem 2.6, is of great interest on its own. The steps of our proof closely follow the assertions of SHAO (2003, p 174ff.) and SERFLING (1980, Ch. 5).

First, note the normalization of the kernel of U_n by p which is essential to fix the approximating distribution based on Assumptions 2.4 and 2.5. For convenience, we can consider $B_{kl} = \mathbf{Y}_k' \mathbf{Y}_l$, where $\mathbf{Y}_j = \mathbf{X}_j / \sqrt{p}$, $j = k, l$, so that $E(\mathbf{Y}_j) = \mathbf{0}$, under H_0 , and $\text{Var}(\mathbf{Y}_j) = \frac{1}{p} \Sigma$. Now, if we let $\lambda_1, \dots, \lambda_p$ to be the eigenvalues of Σ , then Assumptions 2.4 and 2.5 clearly refer to the eigenvalues of $\frac{1}{p} \Sigma$, since $E(\mathbf{Y}_j' \mathbf{Y}_j) = \frac{1}{p} \text{tr}(\Sigma)$ and $E(\mathbf{Y}_k' \mathbf{Y}_l)^2 = \frac{1}{p^2} \text{tr}(\Sigma^2)$. For further reference, let us denote the p -scaled eigenvalues of $\frac{1}{p} \Sigma$ as $\nu_j = \lambda_j / p$, $j = 1, \dots, p$.

We begin by showing that the denominator of T , $\frac{1}{p} E_1$, converges, in probability, to $\sum_j \nu_j$ which, by Assumption 2.4, is uniformly bounded away from 0 and ∞ . To see this, note that $E(E_1) = \text{tr}(\Sigma)$, also valid without normality assumption. Further, by Assumption 2.3, $\text{Var}(E_1/p) \leq \gamma/n \rightarrow 0$, as $n \rightarrow \infty$ since, by Cauchy-Schwarz inequality

$$E(A_k^2) = \sum_{s=1}^p E(X_{ks}^4) + \sum_{s \neq t} \sum E(X_{ks}^2 X_{kt}^2) \leq \gamma p^2. \quad (13)$$

The required convergence then follows immediately by Chebychev inequality. Now, we deal with E_0 in Equation (12).

Following SERFLING (1980, Ch. 5), we define $h(\mathbf{Y}_1, \dots, \mathbf{Y}_m) = \mathbf{Y}_k' \mathbf{Y}_l$, so that $h_c(\mathbf{Y}_1, \dots, \mathbf{Y}_c) = E(h(\mathbf{Y}_1, \dots, \mathbf{Y}_m) | \mathbf{Y}_1 = \mathbf{y}_1, \dots, \mathbf{Y}_c = \mathbf{y}_c)$, with $1 \leq c \leq m$. Further, $\xi_c = \text{Var}[h_c(\mathbf{Y}_1, \dots, \mathbf{Y}_c)]$, with $\xi_0 = 0$. For the present case, $m = 2$ and $c = 1, 2$, so that $h_2(\cdot) = h_m(\cdot) = h(\cdot)$. Now, since, under H_0 , $E(\mathbf{Y}_k) = \mathbf{0}$, it implies that $E(B_{kl}) = 0$, and $E(B_{kl}^2) = \frac{1}{p^2} \text{Var}(B_{kl}) = \text{tr}(\Sigma^2)$ (see also Proposition 2.1). Then, it is just trivial to verify that

$$E[h_1(\mathbf{Y}_1)] = 0 = E[h_2(\mathbf{Y}_1, \mathbf{Y}_2)], \quad \xi_1 = 0, \quad \text{and} \quad \xi_2 = \frac{\text{tr}(\Sigma^2)}{p^2}, \quad (14)$$

so that the variance of U_n becomes (LEHMANN, 1999, p 368)

$$\text{Var}(U_n) = \frac{1}{\binom{n}{2}} \sum_{c=1}^2 \binom{2}{c} \binom{n-2}{2-c} \xi_c = \frac{2}{n(n-1)} \xi_2, \quad (15)$$

same as can be directly computed (see Equation (6) of Proposition 2.1). Given Equation (15), the three conditions on $\text{Var}(U_n)$, as given in SERFLING (1980, Lemma A, p 183), can be immediately verified using Assumption 2.5. For the present case with $m = 2$ and $\xi_1 = 0$, Equation (15) gives

- (i) $0 \leq \text{Var}(U_n) \leq \frac{2}{n}\xi_2$
- (ii) $\frac{2}{n}\xi_2 \leq \frac{2}{n-1}\xi_2$
- (iii) $\text{Var}(U_n) = O(n^{-2})$.

Note that, since $\xi_1 = 0$, Condition (iii) can also be verified through a corollary as stated on page 185 in Serfling. Finally, using the projection method of U -statistics, Condition (iii) and Equation (15) imply that

$$\text{Var}(nU_n) \rightarrow 2\xi_2$$

as $n \rightarrow \infty$ (SERFLING, 1980, p 189); see also SHAO (2003, Lemma 3.2, page 180). Clearly, under Assumption 2.5, $\text{Var}(nU_n)$ is uniformly bounded, as $p, n \rightarrow \infty$.

Now, the theorem in SERFLING (1980, p 194) gives the approximating distribution of U_n , and hence eventually of the statistic T . We note that, the factor n in nU_n is already available as a multiplier in Equation (12). Hence, with $m = 2$ and $E(U_n) = 0$, we have

$$nU_n \xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} \nu_j C_j - \sum_{j=1}^{\infty} \nu_j, \quad (16)$$

under Assumption 2.5, where C_j 's are independent chi-square random variables, each with single degree of freedom, and they are weighted with the eigenvalues ν_j such that $\sum_j \nu_j^2 = \xi_2$ (see SHAO, 2003, Theorem 3.5(ii), p 180). Using (16), along with the convergence of $\frac{1}{p}E_1$, by Slutsky theorem (VAN DER VAART, 1998, Lemma 2.8, p 11), T in Equation (11) can be written as

$$T - 1 \xrightarrow{\mathcal{D}} \frac{W}{K}, \quad (17)$$

with $W = \sum_{j=1}^{\infty} \nu_j(C_j - 1)$ and $K = \sum_{j=1}^{\infty} \nu_j$, as $n, p \rightarrow \infty$, where K , under Assumption 2.4, is uniformly bounded away from 0 and infinity. For simplicity, let $w_j = \nu_j/K$ such that (17) can be represented as

$$T - 1 \xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} w_j(C_j - 1) = \sum_{j=1}^{\infty} w_j C_j - 1, \quad (18)$$

which represents T similar to the representation of a quadratic form in Theorem 2.2, where $\sum_{j=1}^{\infty} w_j = 1$. This gives an exact analog of Theorem 2.2 for the case when $p \rightarrow \infty$, and hence a similar version of Box's approximation can also be

carried through. Since, $E(W) = 0$ and $\text{Var}(W) = 2 \sum_{j=1}^{\infty} \nu_j^2$, therefore,

$$E(T) = 1 \quad (19)$$

$$\text{Var}(T) = \frac{2 \sum_{j=1}^{\infty} \nu_j^2}{\left(\sum_{j=1}^{\infty} \nu_j \right)^2} = 2 \sum_{j=1}^{\infty} w_j^2, \quad (20)$$

which are exactly the same moments as obtained for T under normality with fixed p , i.e., the moments of χ_f^2/f for the approximation in (8).

Now, it is also worth mentioning that, if we let p fixed but continue to relax normality, we can simply go through the same procedure as explained above such that the validity of the approximation (17) comes immediately. In fact, with p fixed, it is relatively easier to show the approximation since the asymptotic convergence of U_n is directly applicable for $n \rightarrow \infty$ without worrying about bounding the expressions for large p . This shows that the original approximation of T as given in AHMAD *et al.* (2008) for fixed p under normality, i.e., (8), is valid without normality assumption whether p is kept fixed or is allowed to grow with n . These results are summarized in Section 4 (see Remark 2). Clearly, using the properties of chi-square distribution, the approximation for fixed p can also be expressed as

$$\frac{T-1}{\sqrt{2/f}} \xrightarrow{\mathcal{D}} N(0, 1), \quad (21)$$

as $n \rightarrow \infty$. A similar approximation for the present case, with $p \rightarrow \infty$, can be obtained using the following Hájek-Šidák Lemma (see JIANG, 2010, Example 6.6, p 183; HÁJEK *et al.*, 1999, p 184) which eventually helps us prove the second part of Theorem 2.6.

Lemma 2.7. *Let X_1, X_2, \dots be iid random variables with mean 0 and variance 1. Let a_{nj} , $1 \leq j \leq n$, be a sequence of constants such that $\max_j a_{nj}^2 \rightarrow 0$ as $n \rightarrow \infty$. Then*

$$\sum_{j=1}^n a_{nj} X_j \xrightarrow{\mathcal{D}} N(0, 1),$$

as $n \rightarrow \infty$.

Consider Approximation (18). Since, $E(C_j) = 1$, $\text{Var}(C_j) = 2$, therefore, in the notations of Lemma 2.7, $X_j = (C_j - 1)/\sqrt{2}$. Define $a_j = w_j/\sqrt{\sum_j w_j^2}$ such that $\sum_j a_j^2 = 1$ and $\max_j a_j^2 \rightarrow 0$ as $p \rightarrow \infty$. Then, by Lemma 2.7

$$\sum_j a_j X_j = \frac{\sum_j w_j (C_j - 1)}{\sqrt{2 \sum_j w_j^2}} \xrightarrow{\mathcal{D}} N(0, 1),$$

as $p \rightarrow \infty$. Obviously, to make T practically workable, we need to estimate unknown parameters used in the approximation. From $w_j = \nu_j/\sum_j \nu_j$, and from Equation (20), we need to estimate $\sum_j \nu_j$, $\sum_j \nu_j^2$, $(\sum_j \nu_j)^2$, where the estimators must be unbiased, and consistent even when $p \rightarrow \infty$. We know

Table 1: QUANTILES OF T , GIVEN IN EQUATION (3), FOR EXPONENTIAL DISTRIBUTION

CS		p						
n	$1 - \alpha$	10	20	50	100	200	500	1000
10	0.90	0.8810	0.8812	0.8775	0.8783	0.8797	0.8773	0.8795
	0.95	0.9298	0.9246	0.9205	0.9205	0.9198	0.9188	0.9214
	0.99	0.9792	0.9741	0.971	0.9725	0.968	0.9679	0.9683
20	0.90	0.8994	0.8954	0.8952	0.9000	0.8921	0.8891	0.8925
	0.95	0.9385	0.9334	0.9325	0.9341	0.9319	0.9246	0.9268
	0.99	0.9754	0.9741	0.9746	0.9694	0.9702	0.9685	0.9675
50	0.90	0.9077	0.8976	0.9030	0.9003	0.9006	0.9007	0.9030
	0.95	0.9466	0.9385	0.9408	0.9364	0.9376	0.9367	0.9415
	0.99	0.9799	0.976	0.9752	0.9722	0.9736	0.9740	0.9758
UN		p						
n	α	10	20	50	100	200	500	1000
10	0.90	0.9177	0.922	0.9215	0.9169	0.9062	0.9125	0.9164
	0.95	0.9669	0.9678	0.967	0.9635	0.9592	0.9591	0.9617
	0.99	0.9968	0.9981	0.9976	0.9952	0.995	0.9939	0.9957
20	0.90	0.9081	0.9049	0.9085	0.9082	0.9009	0.9065	0.9078
	0.95	0.9593	0.958	0.9577	0.9579	0.9531	0.9569	0.9555
	0.99	0.9937	0.9943	0.994	0.9938	0.9938	0.9920	0.9927
50	0.90	0.9075	0.9037	0.9090	0.9038	0.9038	0.9073	0.9020
	0.95	0.9561	0.9522	0.9529	0.9518	0.9547	0.9529	0.9504
	0.99	0.9932	0.9901	0.9901	0.9906	0.9913	0.9914	0.9920

that, $\frac{1}{p}E_1$, $\frac{1}{p^2}E_2$ and $\frac{1}{p^2}E_3$ are unbiased estimators of these three traces (see (9)). Then, we can replace the sums in the approximation by their estimators if we can prove that they are also consistent with respect to the traces they estimate. This consistency for $\frac{1}{p}E_1$ is already shown, which immediately leads to the consistency of $\frac{1}{p^2}E_2$, under Assumptions 2.3-2.5, by the independence of quadratic forms involved in E_2 . For $\frac{1}{p^2}E_3$, we note that,

$$\text{Cov}(A_{kl}^2, A_{rs}^2) \leq \text{Var}(A_{kl}^2) \leq \text{E}(A_{kl}^4) \leq [\text{E}(A_k^2)]^2,$$

by Cauchy-Schwarz inequality, and then the result follows from (13) and Assumption 2.5. Summing up the results, we have

$$\frac{T - \text{E}(T)}{\sqrt{\widehat{\text{Var}}(T)}} \xrightarrow{\mathcal{D}} N(0, 1),$$

where $\widehat{\text{Var}}(T) = 2E_3/E_2$, when $n, p \rightarrow \infty$, as needed to be proved.

3 Simulations

The test statistic is investigated through simulations under a variety of parameter settings, for both test size and power. Tables 1 and 2 report estimated quantiles of T for standard exponential and uniform distributions. The

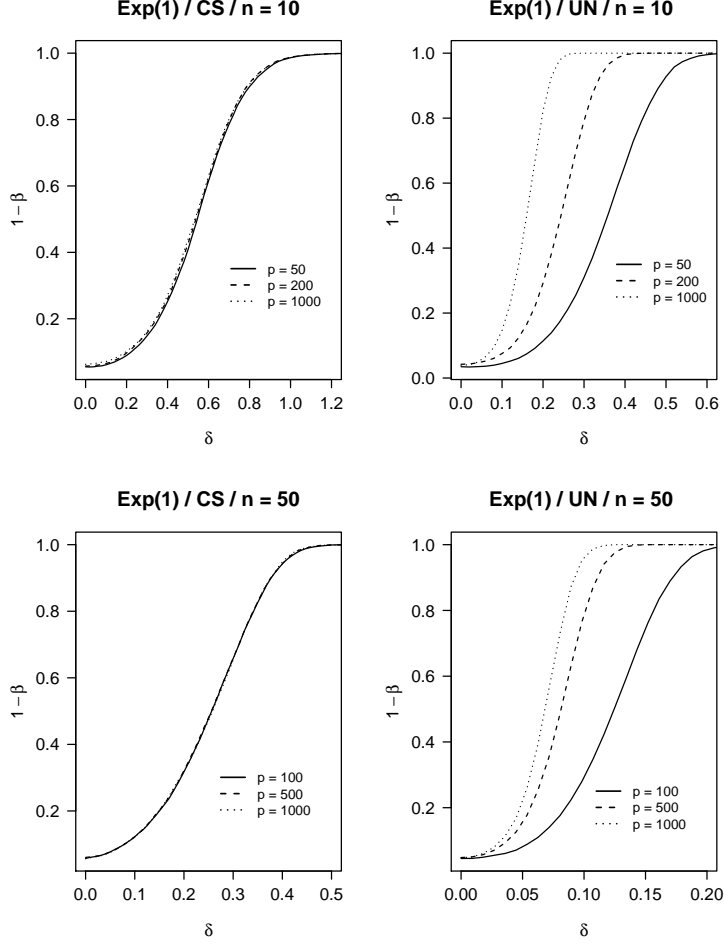


Figure 1: POWER CURVES OF T : EXPONENTIAL DISTRIBUTION

nominal quantiles, $1 - \alpha = 0.90, 0.95, 0.99$, are estimated for three moderate sample sizes, $n = 10, 20, 50$, each n combined with several dimensions, $p = 10, 20, 50, 100, 200, 500, 1000$. Further, two covariance structures are imposed on Σ , viz. compound symmetry (CS) and unstructured (UN). The CS structure is defined as $\Sigma = \sigma^2 [(1 - \rho)\mathbf{I} + \rho\mathbf{J}]$, where \mathbf{I} is the identity matrix, \mathbf{J} is the matrix of ones, and $\sigma^2 > 0$ and ρ are constants, whereas the UN pattern refers to $\Sigma = (\sigma_{ij})_{i,j=1}^p$. For the simulations reported here, it is assumed that $\sigma^2 = 1$, $\rho = 0.5$ for CS, and $\sigma_{ij} = 1(1)p$ ($i = j$), $\rho_{ij} = (i - 1)/p$ ($i > j$), for UN.

As is evident from Table 1, the statistic is slightly liberal under CS structure and slightly conservative under UN structure for exponential distribution for $n = 10$, but the accuracy improves significantly for only a moderate increase

Table 2: QUANTILES OF T , GIVEN IN EQUATION (3), FOR UNIFORM DISTRIBUTION

CS n	$1 - \alpha$	p						
		10	20	50	100	200	500	1000
10	0.90	0.9083	0.9053	0.9082	0.903	0.9073	0.901	0.9036
	0.95	0.9526	0.9467	0.9483	0.9457	0.9456	0.9439	0.9468
	0.99	0.9916	0.9873	0.9862	0.9863	0.9854	0.9871	0.9855
20	0.90	0.9094	0.9094	0.9094	0.9118	0.9131	0.9029	0.9026
	0.95	0.9511	0.9468	0.9498	0.9490	0.9465	0.9448	0.9421
	0.99	0.9874	0.9862	0.9850	0.9856	0.9827	0.9825	0.9815
50	0.90	0.9150	0.9111	0.9119	0.9149	0.9113	0.9106	0.9074
	0.95	0.9540	0.9496	0.9481	0.9510	0.9462	0.9458	0.9444
	0.99	0.9873	0.9857	0.9843	0.9842	0.9828	0.9818	0.9804

UN n	α	p						
		10	20	50	100	200	500	1000
10	0.90	0.9067	0.9079	0.9069	0.9149	0.9097	0.9126	0.9104
	0.95	0.9582	0.9581	0.9590	0.9619	0.9606	0.9598	0.9574
	0.99	0.9938	0.9949	0.9958	0.9964	0.9947	0.9935	0.995
20	0.90	0.9075	0.9090	0.9025	0.9053	0.9065	0.9020	0.9078
	0.95	0.9573	0.9564	0.9546	0.9537	0.955	0.9545	0.9554
	0.99	0.9935	0.9923	0.9944	0.9923	0.9922	0.9919	0.9927
50	0.90	0.9077	0.9039	0.8971	0.8996	0.8974	0.9023	0.903
	0.95	0.9528	0.9532	0.9507	0.9533	0.9483	0.9538	0.9536
	0.99	0.9907	0.9909	0.9906	0.9917	0.9901	0.9916	0.9902

in the sample size, i.e. for $n = 20$ and $n = 50$. For uniform distribution (Table 2), the results are relatively better for n as small as 10, and improves for increasing n . Further, for any fixed n , the accuracy for both distributions remains intact for increasing p which shows the high-dimensional consistency of the test statistic.

A similar performance of the statistic is also observed in power computations, as reported in Figures 1 (Exponential distribution) and 2 (Uniform distribution). The upper panel of each figure represents power curves for $n = 10$ with $p = 50, 200, 1000$ and the lower panel of each figure represents power curves for $n = 50$ with $p = 100, 500, 1000$. Further, within each of these panels, the left (right) part depicts power curves for CS (UN) covariance pattern. The nominal quantile for all power computations is fixed at 0.95, and δ represents the alternative hypothesis under which the power is computed. Note that this alternative hypothesis can be represented as $\boldsymbol{\mu} = \delta_i \mathbf{p}^{-1}$, where $\boldsymbol{\mu} = \mathbf{E}(\mathbf{X})$, $\mathbf{p}^{-1} = \left(\frac{1}{p}, \frac{2}{p}, \dots, \frac{p}{p}\right)'$, and δ_i is the i th value of δ .

We observe that the power of the test statistic is not only high, but increases for increasing dimension. We observe a discernably better performance under UN pattern than under CS pattern, particularly regarding increase in power for increasing dimension, which is somewhat clear given the eigenstructure of the two covariance patterns. Another significant difference, also present in the estimated test sizes, is the relatively better performance of the statistic for uniform distribution than for exponential distribution. This is also comprehensible tak-

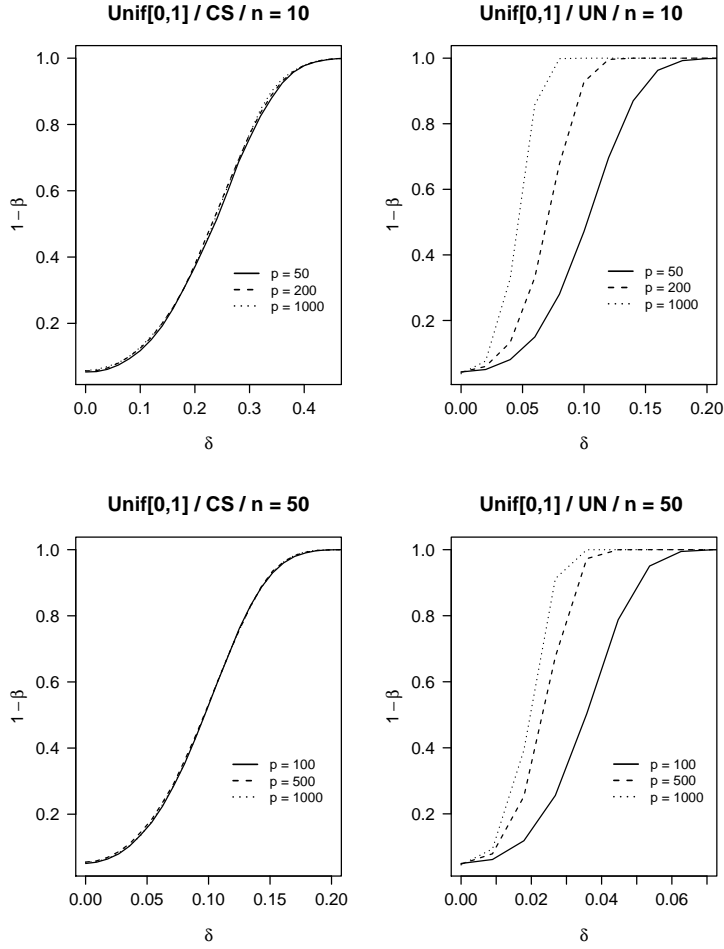


Figure 2: POWER CURVES OF T : UNIFORM DISTRIBUTION

ing into account the fact that exponential distribution represents one of the most serious violations of normality.

4 Some Remarks

4.1 High dimensional asymptotics and U -statistics

As the convergence of the statistic to a weighted sum of independent chi-square random variables, as shown in (16), is the linchpin of the proof, it deserves a closer look. It is an interesting extension of the finite dimensional case of Theorem 2.2 to an infinite dimensional case that is the subject of this paper. This

fundamental result is based on an orthogonal expansion of the kernel $\frac{1}{p}h(\mathbf{X}_k, \mathbf{X}_l)$ of the degenerate U -statistic, U_n , in terms of orthogonal functions, represented as

$$h(\mathbf{X}_k, \mathbf{X}_l) = \sum_{j=1}^{\infty} \nu_j e_j(\mathbf{X}_k) e_j(\mathbf{X}_l) \quad (22)$$

where ν_j 's and e_j 's are, respectively, the eigenvalues and orthonormal eigenfunctions of $h(\mathbf{X}_k, \mathbf{X}_l)$. The validity of the approximation of T in (16) rests on certain conditions imposed on $h(\mathbf{X}_k, \mathbf{X}_l)$ in the expansion (22). First, from (14), we have $\xi_1 = 0$ and $\xi_2 < \infty$, where the former is known as the condition of (first order) degeneracy of the kernel (LEE, 1990, p 78), and the latter ensures, under Assumption 2.5, that the kernel is square-integrable, i.e. that the \mathcal{L}_2 norm of its eigenvalues is finite. A brief but nice exposition of the theory is given in VAN DER VAART (1998, Ch. 12). For more details, see DENKER (1985), SERFLING (1980, Ch. 5), LEE (1990, Ch. 3), and KOROLJUK and BOROVSKICH (1994, Ch. 4).

A particularly interesting feature of the test statistic is that the degenerate kernel it is based upon is of order 2, and as VAN DER VAART (1998, p 168) puts it, the orthogonal expansion, like (22), of such a second order kernel is the most straightforward to comprehend, and which is guaranteed by the Hilbert-Schmidt theory of self-adjoint linear operators (see, for example, REED and SIMON, 1980, Theorem VI.16, p 203). A detailed study of the theory of operators and their properties is given, for example, in DUNFORD and SCHWARTZ (1963), and KREYSZIG (1978).

Finally, the convergence of the expansion in (22) is in mean square, in the sense that

$$\mathbb{E} \left(h(\mathbf{X}_k, \mathbf{X}_l) - \sum_{j=1}^p \nu_j e_j(\mathbf{X}_k) e_j(\mathbf{X}_l) \right)^2 = \sum_{j=p+1}^{\infty} \nu_j^2 \rightarrow 0,$$

as $p \rightarrow \infty$, i.e., the approximation corresponds to at most countably many eigenvalues (VAN DER VAART, 1998, p 168). This L_2 -convergence helps us establish (16) as a modification to the result of Theorem 2.2 for $p \rightarrow \infty$ and under non-normality. A more thorough treatment of this theory for its use in high dimensional inference is the subject of another manuscript.

4.2 Box's approximation

As briefly explained after Equations (19) and (20), the approximate limit distribution of the test statistic T (Theorem 2.6) is also valid if we keep normality assumption relaxed but assume p fixed, where p may also exceed n . This shows that the validity of T , and hence that of Box's approximation (Theorem 2.2) is valid for both n -asymptotics and for (n, p) -asymptotics, and whether \mathcal{F} is assumed multivariate normal (as shown in AHMAD *et al.*, 2008), or \mathcal{F} represents any multivariate distribution with finite fourth moment, as proved here under Assumptions 2.3-2.5.

The approximation for fixed p , of course, does not heavily depend on Assumptions 2.3-2.5. Under normality, none of these assumptions is needed for either Box's approximation or the asymptotic distribution of T , as shown in AHMAD *et al.* (2008), whereas Assumptions 2.3 and 2.4 are obviously needed when normality is relaxed. In the later case, under H_0 , the multivariate central limit theorem (LEHMANN, 1999, Theorem 5.4.4, p 313) gives asymptotic normality of $\sqrt{n}\bar{\mathbf{X}}$, and the continuous mapping theorem (JIANG, 2010, Theorem 2.12, p 30) yields the asymptotic chi-square distribution $n\bar{\mathbf{X}}'\bar{\mathbf{X}} \approx \chi_p^2$, as $n \rightarrow \infty$. Alternatively, the asymptotic theory of U -statistics can be used in which case the expressions in Section 2 are clearly much easier to deal with for fixed p case than for arbitrarily growing p .

For example, the statistic T in Equation (11) reduces to $T = 1 + E_0/E_1$, without the divisor p which does not influence the asymptotic limits, and all the computations go through. In this sense, we consider the approximation of T for fixed p as a special case of the one presented in Section 2. For reference, we summarize the results in the following theorems, including the result based on normality assumption (AHMAD, 2008, AHMAD *et al.*, 2008), leaving the detailed derivations aside. The first theorem is for Box's approximation, and the second for the test statistic T . Without loss of generality, we continue to assume $\boldsymbol{\mu} = \mathbf{0}$.

Theorem 4.1. (Theorem 2.2, revisited) *Consider model (1), and let $Q = n\bar{\mathbf{X}}'\bar{\mathbf{X}}$ be as defined in Section 2. We have the following.*

1. *Assume p fixed and $n \rightarrow \infty$. If (a) \mathcal{F} is multivariate normal, or (b) \mathcal{F} is as in (1) for which Assumptions 2.3-2.4 are satisfied, then,*

$$n\bar{\mathbf{X}}'\bar{\mathbf{X}} \approx \sum_{j=1}^p \lambda_j C_j,$$

where λ_j are the eigenvalues of $\boldsymbol{\Sigma}$ and C_j are iid χ_1^2 .

2. *Assume $n \rightarrow \infty$ and $p \rightarrow \infty$, and \mathcal{F} is as in (1) for which Assumptions 2.3-2.5 are satisfied. Then (see Equation (16))*

$$n\bar{\mathbf{X}}'\bar{\mathbf{X}} \approx \sum_{j=1}^{\infty} \nu_j C_j,$$

where C_j are iid χ_1^2 and ν_j are the eigenvalues of $\frac{1}{p}\boldsymbol{\Sigma}$.

Theorem 4.2. (The test statistic) *Consider model (1). Let $T = Q/E_1$ be the test statistic as defined in Equation (3), and E_2, E_3 be as defined in (9), with $\hat{f} = E_2/E_3$. We have the following.*

1. *Assume p fixed and $n \rightarrow \infty$. If (a) \mathcal{F} is multivariate normal, or (b) \mathcal{F} is as in (1) for which Assumptions 2.3-2.4 are satisfied, then,*

$$fT \xrightarrow{\mathcal{D}} \chi_f^2,$$

where $f = [\sum_{j=1}^p \lambda_j]^2 / \sum_{j=1}^p \lambda_j^2$ is estimated as \hat{f} .

2. Assume $n \rightarrow \infty$ and $p \rightarrow \infty$, and \mathcal{F} is as in (1) for which Assumptions 2.3-2.5 are satisfied. Then

$$fT \xrightarrow{\mathcal{D}} \chi_f^2,$$

where $f = [\sum_{j=1}^{\infty} \nu_j]^2 / \sum_{j=1}^{\infty} \nu_j^2$ is estimated as \hat{f} .

We notice that, for fixed p case, an explicit expression for $\text{Var}(E_1)$ is not furnished in Proposition 2.1. In fact, a closer look at the proof of Proposition 5 in AHMAD *et al.* (2008) reveals that this variance is not needed to compute the moments of the test statistic using the delta method, as they originally used. However, in order to study the properties of estimators, one may need certain moments, and there is a plethora of literature on the moments of quadratic forms without normality assumption. Under the assumption of independence of X_{ks} , $s = 1, \dots, p$, the variance of a quadratic form is given in ATIQULLAH (1962) without proof, a general proof of which is given in SEBER and LEE (2003, p 10) and a proof for the central case is given in PLACKET (1960, p 16); see also OHTAKI (1990, Corollary A.2). In fact, Theorem A.1 in OHTAKI (1990) is a very general result for the moments of quadratic forms replacing normality assumption with the assumption of finite moments of order up to order 4. For example, a substitution of $n = 1$ in this theorem trivially yields the multivariate version of Atiquallah's result which is more fitting for the computation of $\text{Var}(E_1)$ directly. A brief sketch of the proof of the general theorem is given by Ohtaki (see page 89 of the manuscript) which shows that the proof is extremely straightforward. For more related work, see the references given in the context of Assumption 2.3 and the references cited therein.

4.3 Testing a general linear hypothesis

Although, for simplicity, the test statistic is discussed in the context of testing the usual multivariate hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$, one can, however, also use it to test a general linear hypothesis with certain hypothesis matrix defining the appropriate contrasts on the elements of the vector \mathbf{X}_k . For example, we may test $H_0 : \mathbf{H}\boldsymbol{\mu} = \mathbf{0}$, where \mathbf{H} is a hypothesis matrix. For a unique representation, H_0 can be replaced with an equivalent hypothesis $H_0 : \mathbf{P}\boldsymbol{\mu} = \mathbf{0}$, where $\mathbf{P} = \mathbf{H}(\mathbf{H}'\mathbf{H})^- \mathbf{H}'$, and $(\cdot)^-$ denotes a g -inverse. The two-fold advantages of using \mathbf{P} are that it significantly reduces the computational burden due to it being a projection matrix (symmetric and idempotent), and that it can help us impose a factorial structure on the random vector \mathbf{X}_k , so that the use of the statistic \mathbf{T} can be extended to other hypotheses of interest, for example repeated measures profile analysis, or factorial structures for between effects, etc. Then, for this one-sample case, the hypothesis matrix \mathbf{P} is either of the form $\mathbf{P} = \mathbf{I} - \frac{1}{p}\mathbf{J}$ (when there is no structure), where \mathbf{I} is the identity matrix, and \mathbf{J} is the matrix of 1s, or it is of the form $\mathbf{P} \otimes \mathbf{C}$, where \mathbf{C} is a matrix of contrasts regarding the imposed structure, and \otimes denotes the Kronecker product. Hence, \mathbf{C} may be a single matrix, or itself a Kronecker product of several other similar matrices, depending upon the number of factors involved and the type of between-factor

hypotheses of interest. But in either case, \mathbf{C} does not involve the dimension p . This immediately leads us to see that, for $p \rightarrow \infty$, the statistic can be used to test either of these hypotheses. For some examples of the hypothesis matrix \mathbf{P} used to test multi-factorial structure in a design, see AHMAD (2008).

5 Summary and conclusions

A statistic for testing the mean vector is presented when the dimension of the multivariate vector, p , may exceed the number of such vectors, n , and the underlying distribution need not necessarily be multivariate normal. Using the asymptotic theory of one-sample first-order degenerate U -statistics, the test statistic is shown to follow an approximate chi-square distribution, and based on a classical asymptotic result, eventually to an approximate normal distribution. The main result of the paper, Theorem 2.6, leads to a modified version of the Box's approximation for non-normal and high-dimensional multivariate data.

It is further illustrated, through simulation studies, that the statistic performs accurately, both in size estimation and power, for a moderate n and a very large p , and that the statistic maintains these properties for the most commonly used covariance structures. It is shown in simulation results, that the largeness of n for the validity of the approximation of the test statistic, for practical purposes, refers only to moderate sample sizes and the accuracy of the test statistic remains intact even when p grows much larger than n .

Acknowledgment

The authors are thankful to the editor, the associate editor, and to two anonymous referees for their helpful suggestions and comments which led to a substantial improvement of the original draft of this manuscript.

References

- AHMAD, M. RAUF (2008), *Analysis of high-dimensional repeated measures designs: The one- and two-sample test statistics* (Ph.D. Thesis), Cuvillier Verlag, Göttingen, Germany.
(Online: <http://webdoc.sub.gwdg.de/diss/2008/ahmad/ahmad.pdf>)
- AHMAD, M. RAUF, C. WERNER and E. BRUNNER (2008), Analysis of high dimensional repeated measures designs: The one sample case, *Computational Statistics and Data Analysis*, **53**, 416-427.
- AHMAD, M. RAUF, M. OHLSON and D. VON ROSEN (2011), Tests for covariance matrices, particularly for high-dimensional data, *Scandinavian Journal of Statistics*. Submitted.

- AHN, J., J. S. MARRON, K. M. MULLER and A. Y. CHI (2007), The high-dimension, low-sample-size geometric representation holds under mild conditions, *Biometrika*, **94**, 760-766.
- ATIQULLAH, M. (1962), The estimation of residual variance in quadratically balanced least-squares problems and the robustness of the \mathcal{F} -test, *Biometrika*, **49**, 83-91.
- BAI, Z. and H. SARANADASA (1996), Effect of high dimension: By an example of a two sample problem, *Statistica Sinica*, **6**, 311-329.
- BATHKE, A. (2002), ANOVA for a large number of treatments, *Mathematical Methods of Statistics*, **11**, 118-132.
- BATHKE, A. and S. W. HARRAR (2008), Nonparametric methods in multivariate factorial designs for large number of factor levels, *Journal of Statistical Planning and Inference*. **138**, 588-610.
- BATHKE, A., S. W. HARRAR and M. RAUF AHMAD (2009), Some contributions to the analysis of multivariate data, *Biometrical Journal*, **51**, 285-303.
- BOX, G. E. P. (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**, 290-302.
- CHEN, S. X., L-X. ZHANG and P-S. ZHONG (2010), Tests for high dimensional covariance matrices, *Journal of the American Statistical Association*, **105**, 810-819.
- DEHLING, H. (2006), *Limit theorems for dependent U-statistics*. In: Bertail, P, Doukhan, P, and Soulier, P (Eds): *Dependence in probability and statistics*, Springer, New York.
- DENKER, M. (1985), *Asymptotic distribution theory in nonparametric statistics*, Vieweg Advanced Lectures, Vieweg, Braunschweig.
- DENKER, M, and G. KELLER (1983), On U -statistics and v. Mises' statistics for weakly dependent processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **64**, 505-522.
- DEPMSTER, A. P. (1958). A high dimensional two sample significance test, *Annals of Mathematical Statistics*, **29**, 995-1010.
- DUNFORD, N. and J. T. SCHWARTZ (1967), *Linear operators*. Part II: *Spectral theory - Self-adjoint operators in Hilbert space*, Wiley, New York.
- FISHER, T. J., X. SUN and C. M. GALLAGHER (2010), A new test for sphericity of the covariance matrix for high dimensional data. *Journal of Multivariate Analysis*, **101**, 2554-2570.

- FUJIKOSHI, Y., V. V. ULYANOV and R. SHIMIZU (2010), *Multivariate statistics: High-dimensional and large-sample approximations*, Wiley, New York.
- HÁJEK, J., Z. ŠIDÁK and P. K. SEN (1999), *Theory of rank tests*, Academic Press, San Diego.
- HOEFFDING, W. (1948), A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics*, **19**, 293-325.
- JIANG, J. (2010), *Large sample techniques for statistics*, Springer, New York.
- KNIGHT, J. L. (1995), The joint characteristic function of linear and quadratic forms of non-normal variables, *Sankhyā: The Indian Journal of Statistics, Series A*, **41**, 231-238.
- KOROLJUK, V. S. and Y. V. BOROVSKICH (1994), *Theory of U-statistics*, Kluwer Academic Press, Dordrecht.
- KREYSZIG, E. (1978), *Introductory functional analysis with applications*, Wiley, New York.
- KROPF, S., D. KOSE, J. LÄUTER and D. VON ROSEN (2009), Comparison of exact parametric tests for high-dimensional data, *Computational Statistics & Data Analysis*, **53**, 776-787.
- LÄUTER, J. (2004), Two new multivariate tests, in particular for a high dimension, *Acta et Commentationes Universitatis Tartuensis de Mathematica*, **8**, 179-186.
- LÄUTER, J., E. GLIMM and S. KROPF (1998), Multivariate tests based on left-spherically distributed linear scores, *The Annals of Statistics*, **26**, 1972-1988. Corrections: **27**, 1441.
- 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**, 1081-1102.
- LEE, A. J. (1990), *U-statistics: Theory and Practice*, CRC Press, Boca Raton.
- LEHMANN, E. L. (1999), *Elements of large-sample theory*, Springer, New York.
- OHTAKI, M. (1990), Some estimators of covariance matrix in multivariate non-parametric regression and their applications, *Hiroshima Math J*, **20**, 63-91.
- PINHEIRO, A., P. K. SEN and H. P. PINHEIRO (2009), Decomposability of high-dimensional diversity measures: Quasi- U -statistics, martingales, and nonstandard asymptotics, *Journal of Multivariate Analysis*, **100**, 1645-1656.
- PLACKETT, R. L. (1960), *Principles of regression analysis*, Oxford University Press, London.

- RAO, C. R. and J. KLEFFE (1988), *Estimation of variance components and applications*, Elsevier science publishers, Amsterdam.
- REED, M. and B. SIMON (1980), *Methods of modern mathematical physics*, Vol. I: Functional analysis, Academic Press, San Diego, CA.
- SEARLE, S. R. (1971), *Linear models*, Wiley, New York.
- SEBER, G. A. F. and A. J. LEE (2003), *Linear regression analysis*, 2nd ed., Wiley, New York.
- SERFLING, R. J. (1980), *Approximation theorems of mathematical statistics*, Wiley, New York.
- SHAO, J. (2003), *Mathematics statistics*, 2nd. ed., Springer, New York.
- SRIVASTAVA, M. S. (2007), Multivariate theory for analyzing high dimensional data, *Journal of Japan Statistical Society*, **37**, 53-86.
- SRIVASTAVA, M. S. (2009), A test for the mean vector with fewer observations than the dimension under non-normality, *Journal of Multivariate Analysis*, **100**, 518-532.
- SRIVASTAVA, M. S. and M. DU (2008), A test for the mean vector with fewer observations than the dimension, *Journal of Multivariate Analysis*, **99**, 386-402.
- SUGIURA, N. (1965), Multisample and multivariate nonparametric tests based on U -statistics and their asymptotic efficiencies, *Osaka J Math*, **2**, 385-426.
- VAN DER VAART, A. W. (1998), *Asymptotic statistics*, Cambridge University Press, Cambridge.
- WIENS, D. P. (1992), On moments of quadratic forms in non-spherically distributed random variables, *Statistics*, **23**, 265-270.
- WOLFINGER, R. D. (1996), Heterogenous variance: Covariance structures for repeated measures, *Journal of Agricultural, Biological, and Environmental Statistics*, **1**, 205-230.
- ZHANG, P-S. and S. X. CHEN (2011), Tests for high-dimensional regression coefficients with factorial designs, *Journal of the American Statistical Association*, **106**(493), 260-274.