# New Invariant and Consistent Chi-squared Type Goodness-of-fit Tests for Multivariate Normality and a Related Comparative Simulation Study

Voinov, V.[1], Pya, N.[1], Makarov, R.[1], Voinov, Y. [2]

[1] KIMEP University, 2 Abay Ave., 050010, Almaty, Kazakhstan

[2] Institute for Mathematics and Mathematical Modeling, Ministry of Education and Science of the Republic of Kazakhstan, 125 Pushkin Str., 050010 Almaty, Kazakhstan

**Abstract**

New chi-squared type invariant and consistent goodness-of-fit tests for multivariate normality are introduced. The tests are based on the Karhunen-Loève transformation of a multi-dimensional sample from a population. This transformation diagonalizes the sample covariance matrix. Then a modification of Moore and Stubblebine technique for construction Wald's type chi-squared tests was used. A comparison of simulated powers of these tests and some other well known tests with respect to seven different symmetrical multivariate alternatives is given. The simulation study demonstrates that the power of the proposed McCull test almost does not depend on the number of grouping cells. The test shows an advantage over other chi-squared type tests and is more powerful than several known tests against some alternatives.

**Key Words:** Chi-squared goodness-of-fit tests, invariant and consistent tests, multivariate normality, symmetric alternatives, power of tests

## Introduction

The assumption of multivariate normality (MVN) is of great importance for applied multivariate statistics, e.g., for multivariate analysis of variance, discriminant analysis, canonical correlation and factor analysis, analysis of regression residuals, residuals in time-series models, etc. To check that assumption only most powerful tests should be used, because tests with low power cannot discriminate for sure between null hypothesis of MVN and supposed alternatives. Several examples of such a situation in the univariate case were, e.g., given in Voinov et al (2013), Sec. 3.10.1, and Sec. 3.9. Henze (2002), p.468 stated that those tests of MVN must be invariant (with respect to affine transformations of sample data) and consistent (probability to fall into rejection region under alternative must tend to 1 if sample size $n \to \infty$).

Two recent surveys of the state of art are known: Henze (2002), and Mecklin and Mundfrom (2004). Those surveys list tenth of publications devoted to testing of MVN. It is almost impossible to find out and site all existing papers on tests of MVN. We would like to mention here papers that have not been listed in those surveys (Small (1978), Follmann (1996), Bartozynski et al (1997), Royston (1992), Hwu et al (2002)), and several recent publications: Keselman (2005), Mecklin and Mundfrom (2005), Székely and Rizzo (2005), von Eye (2005), Farrell et al (2007), Doornik and Hansen (2008)), Sürücü (2006), Villasenor Alva and Estrada (2009), Cardoso De Oliveira and Ferreira

[1] Email: voinovv@mail.ru. This research has been supported by the grant 0760/GF of the Ministry of Education and Science of the Republic of Kazakhstan.

(2010), Hanusz and Tarasińska (2012), Joenssen and Vogel (2012), Batsidis and Zografos (2013), and Batsidis et al (2013).

In this article we introduce new invariant and consistent chi-squared goodness-of-fit tests for MVN. Section 1 is devoted to theoretical background of our proposal. In Section 2 we derive closed form expressions for our tests. In Section 3 a Monte-Carlo power comparison study of different MVN tests is described. A conclusion and recommendations are given in Section 4.

## 1. Preliminaries

The main goal of this research is a construction of a Wald's type chi-squared goodness-of-fit tests for the composite hypothesis that a set $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ of $n$ independent identically distributed (iid) $p$ - dimensional random vectors does not contradict the following joint probability density function

$$f(\boldsymbol{x}\,|\,\boldsymbol{\theta}) = (2\pi)^{-p/2}\,|\,\boldsymbol{\Sigma}\,|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{\mu}$ is the $p$ - vector of means and $\boldsymbol{\Sigma}$ is a nonsingular $p\times p$ covariance matrix. Let a given vector of unknown parameters be

$$\boldsymbol{\theta} = (\mu_1,\ldots,\mu_p,\sigma_{11},\sigma_{12},\sigma_{22},\ldots,\sigma_{1j},\sigma_{2j},\ldots,\sigma_{jj},\ldots,\sigma_{pp})^T,$$

where the elements of the matrix $\boldsymbol{\Sigma}$ are arranged column-wise by taking the elements of the upper-triangular submatrix of $\boldsymbol{\Sigma}$. Since parameter $\boldsymbol{\theta}$ is considered to be unknown, we shall estimate it by the maximum likelihood method. The maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ is the vector $(\bar{\boldsymbol{X}},\boldsymbol{S})^T$, where $\bar{\boldsymbol{X}}$ is the vector of sample means, and $\boldsymbol{S}$ is the sample covariance matrix. Given $\hat{\theta}_n$, and constants $0\le c_0<\cdots<c_r=\infty$, the $r$ grouping cells can be defined as (Moore and Stubblebine (1981))

$$E_{in}(\hat{\boldsymbol{\theta}}_n) = \left\{\mathbf{X}\in\mathbf{R}^p : c_{i-1}\le(\mathbf{X}-\bar{\mathbf{X}})^T\mathbf{S}^{-1}(\mathbf{X}-\bar{\mathbf{X}})<c_i\right\},\ i=1,\ldots,r.$$

The estimated probability to fall into each cell is

$$p_{in}(\hat{\boldsymbol{\theta}}_n) = \int_{E_{in}(\hat{\boldsymbol{\theta}}_n)} f(\mathbf{x}\,|\,\hat{\boldsymbol{\theta}}_n)d\mathbf{x},\ i=1,\ldots,r.$$

If $c_i$ are the $i/r$ points, $i=1,\ldots,r,$ of $\chi^2(p)$ distribution that is the limit distribution of $(\mathbf{X}-\bar{\mathbf{X}})^T\mathbf{S}^{-1}(\mathbf{X}-\bar{\mathbf{X}})$, then the cells are equiprobable and $p_{in}(\hat{\boldsymbol{\theta}}_n)=1/r,\ i=1,\ldots,r.$ Denote $\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n)$ a vector of standardized cell frequencies with components

$$V_i = V_{in}(\hat{\boldsymbol{\theta}}_n) = \frac{(N_{in}-n/r)}{\sqrt{n/r}},\ i=1,\ldots,r,$$

where observed frequency $N_{in}$ is the number of random vectors $\mathbf{X}_1,\ldots,\mathbf{X}_n$ falling into $E_{in}(\hat{\boldsymbol{\theta}}_n),\ i=1,\ldots,r.$ Moore and Stubblebine (1981) showed that in this case, the limiting covariance matrix of standardized frequencies $\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n)$ is $\boldsymbol{\Sigma}_l = \mathbf{I}-\mathbf{q}\mathbf{q}^T-\mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$, where $\mathbf{B}$ is the $r\times m$ matrix with elements

$$B_{ij} = \frac{1}{\sqrt{p_i(\boldsymbol{\theta})}}\frac{\partial p_i(\boldsymbol{\theta})}{\partial\theta_j},\ i=1,\ldots,r,\ j=1,\ldots,m,$$

$\mathbf{q}$ is a $r$-vector with its entries as $1/\sqrt{r}$ , $\mathbf{J}=\mathbf{J}(\boldsymbol{\theta})$ is the Fisher information matrix for one observation, and $m = p + p(p+1)/2$ is the number of unknown parameters. The $m \times m$ Fisher information matrix can be evaluated as (Moore and Stubblebine (1981))

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1} \end{bmatrix},$$

where $\mathbf{Q}$ is the $p(p+1)/2 \times p(p+1)/2$ covariance matrix of $\mathbf{w}$ (a vector of the entries of $\sqrt{n}\mathbf{S}$ arranged column-wise by taking the upper triangular elements)

$$\mathbf{w} = (s_{11}, s_{12}, s_{22},\ s_{13},\ s_{23}, s_{33}, \ldots, s_{pp})^T .$$

The elements of $\mathbf{Q}$ can be written precisely as (Press (1972))

$$\text{Var}(s_{ij}) = \sigma_{ij}^2 + \sigma_{ii}\sigma_{jj},\ i, j = 1, \ldots, p,\ i \le j,$$

$$\text{Cov}(s_{ij}, s_{kl}) = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk},\ i, j, k, l = 1, \ldots, p,\ i \le j,\ k \le l,$$

where $\sigma_{ij},\ i, j = 1, \ldots, p,$ are elements of $\boldsymbol{\Sigma}$. Based on the above Moore and Stubblebine (1981) suggested to use the Nikulin-Rao-Robson (NRR) statistic

$$Y_n^2 = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n) + \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)\mathbf{B}_n(\mathbf{J}_n - \mathbf{B}_n^T\mathbf{B}_n)^{-1}\mathbf{B}_n^T\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n), \tag{1}$$

where $\mathbf{B}_n$ and $\mathbf{J}_n$ are MLEs of $\mathbf{B}$ and $\mathbf{J}$.

The two-dimensional case of (1) has been thoroughly considered in Voinov (2013a), and Voinov (2013b). It was shown that the limiting covariance matrix $\boldsymbol{\Sigma}_l = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$ of standardized frequencies $\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n)$ in this case depends on unknown parameters (on elements of $\boldsymbol{\Sigma}$ in particular) and, hence, in accordance with Lemma 9 of Khatri (1968), $Y_n^2$ in (1) cannot be distributed as chi-squared. Moreover, the statistic $Y_n^2$ is not distribution-free and cannot be used for testing bivariate normality in principle.

Producing a simulation study of the statistic (1) we observed that the limit distribution of (1) will follow the chi-squared distribution with $r-1$ degrees of freedom if and only if $\boldsymbol{\Sigma}$ is a diagonal matrix.

Consider briefly a two-dimensional case when $p = 2$ for a diagonal $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$.

The Fisher's information matrix $\mathbf{J}(\boldsymbol{\theta})$ for one observation, and $\mathbf{Q}$ are

$$\mathbf{J} = \begin{pmatrix} 1/\sigma_{11} & 0 & 0 & 0 \\ 0 & 1/\sigma_{22} & 0 & 0 \\ 0 & 0 & 1/(2\sigma_{11}^2) & 0 \\ 0 & 0 & 0 & 1/(2\sigma_{22}^2) \end{pmatrix},\ \mathbf{Q} = \begin{pmatrix} 2\sigma_{11}^2 & 0 \\ 0 & 2\sigma_{22}^2 \end{pmatrix}.$$

In the blocked form the matrix $\mathbf{B} = (\mathbf{0} : \tilde{\mathbf{B}})$, where $\tilde{\mathbf{B}} = \begin{pmatrix} d_1/\sigma_{11} & d_1/\sigma_{22} \\ d_2/\sigma_{11} & d_2/\sigma_{22} \\ \cdots & \cdots \\ d_r/\sigma_{11} & d_r/\sigma_{22} \end{pmatrix}$, then

$\tilde{\mathbf{B}}^T\tilde{\mathbf{B}} = r\sum d_i^2 \begin{pmatrix} 1/\sigma_{11}^2 & 1/(\sigma_{11}\sigma_{22}) \\ 1/(\sigma_{11}\sigma_{22}) & 1/\sigma_{22}^2 \end{pmatrix}$, where $d_i$, $i$=1,...,$r$, are defined by Moore and Stubblebine (1981), p.720. In the blocked form the matrix $\mathbf{J} - \mathbf{B}^T\mathbf{B} = \begin{pmatrix} \mathbf{\Sigma}^{-1} & 0 \\ 0 & \mathbf{D}_2 \end{pmatrix}$, where

$$\mathbf{D}_2 = \begin{pmatrix} \frac{1-2r\sum d_i^2}{2\sigma_{11}^2} & -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{22}} \\ -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{22}} & \frac{1-2r\sum d_i^2}{2\sigma_{22}^2} \end{pmatrix}.$$

Evidently that $(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1} = \begin{pmatrix} \mathbf{\Sigma} & 0 \\ 0 & \mathbf{D}_2^{-1} \end{pmatrix}$ and $\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T = \tilde{\mathbf{B}}\mathbf{D}_2^{-1}\tilde{\mathbf{B}}^T$.

$$\mathbf{D}_2^{-1} = \frac{4}{1-4r\sum d_i^2} \begin{pmatrix} \frac{(1-2r\sum d_i^2)\sigma_{11}^2}{2} & r\sum d_i^2\sigma_{11}\sigma_{22} \\ r\sum d_i^2\sigma_{11}\sigma_{22} & \frac{(1-2r\sum d_i^2)\sigma_{22}^2}{2} \end{pmatrix}, \qquad (2)$$

and the NRR statistic in (1) becomes

$$Y_n^2 = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n) + \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)\tilde{\mathbf{B}}_n\mathbf{D}_{2n}^{-1}\tilde{\mathbf{B}}_n^T\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n), \qquad (3)$$

where $\tilde{\mathbf{B}}_n$ and $\mathbf{D}_{2n}^{-1}$ are MLEs of $\tilde{\mathbf{B}}$ and $\mathbf{D}_2^{-1}$ respectively. Since

$\tilde{\mathbf{B}}\mathbf{D}_2^{-1} = \frac{2\sqrt{r}}{1-4r\sum d_i^2} \begin{pmatrix} d_1\sigma_{11} & d_1\sigma_{22} \\ d_2\sigma_{11} & d_2\sigma_{22} \\ \cdots & \cdots \\ d_r\sigma_{11} & d_r\sigma_{22} \end{pmatrix}$, then $\tilde{\mathbf{B}}\mathbf{D}_2^{-1}\tilde{\mathbf{B}}^T = \frac{4r}{1-4r\sum d_i^2} \begin{pmatrix} d_1^2 & d_1d_2 & \cdots & d_1d_r \\ \cdots & \cdots & \cdots & \cdots \\ d_rd_1 & d_rd_2 & \cdots & d_r^2 \end{pmatrix}$.

From all these it follows that

$\mathbf{V}^T\tilde{\mathbf{B}}\mathbf{D}_2^{-1}\tilde{\mathbf{B}}^T\mathbf{V} = \frac{4r\left(\sum V_i d_i\right)^2}{1-4r\sum d_i^2}$, and the closed form of the NRR statistic in (3) is

$$Y_n^2 = \sum V_i^2 + \frac{4r\left(\sum V_i d_i\right)^2}{1-4r\sum d_i^2}. \qquad (4)$$

## 2. Proposed chi-squared goodness-of-fit tests for MVN

Consider the case of any symmetrical $\mathbf{\Sigma}$ with the dimensionality $p > 2$.

**Proposition 1.** For any parameter $\boldsymbol{\theta} = (\mu_1, \mu_2, ..., \mu_p, \sigma_{11}, \sigma_{22}, ..., \sigma_{pp})^T$

$$\mathbf{D}_p^{-1} = \frac{4}{1-2pr\sum d_i^2}\begin{pmatrix} \frac{(1-2(p-1)r\sum d_i^2)\sigma_{11}^2}{2} & r\sum d_i^2\sigma_{11}\sigma_{22} & \cdots & r\sum d_i^2\sigma_{11}\sigma_{pp} \\ r\sum d_i^2\sigma_{11}\sigma_{22} & \frac{(1-2(p-1)r\sum d_i^2)\sigma_{22}^2}{2} & \cdots & r\sum d_i^2\sigma_{22}\sigma_{pp} \\ \cdots & \cdots & \cdots & \cdots \\ r\sum d_i^2\sigma_{11}\sigma_{pp} & r\sum d_i^2\sigma_{22}\sigma_{pp} & \cdots & \frac{(1-2(p-1)r\sum d_i^2)\sigma_{pp}^2}{2} \end{pmatrix}. \quad (5)$$

The proof of this is given in Appendix 1. Since in this case

$$\tilde{\mathbf{B}} = \sqrt{r}\begin{pmatrix} d_1/\sigma_{11} & d_1/\sigma_{22} & \cdots & d_1/\sigma_{pp} \\ d_2/\sigma_{11} & d_2/\sigma_{22} & \cdots & d_2/\sigma_{pp} \\ \cdots & \cdots & \cdots & \cdots \\ d_r/\sigma_{11} & d_r/\sigma_{22} & \cdots & d_r/\sigma_{pp} \end{pmatrix},$$

we have

$$\tilde{\mathbf{B}}\mathbf{D}_p^{-1} = \frac{2\sqrt{r}}{1-2pr\sum d_i^2}\begin{pmatrix} d_1\sigma_{11} & d_1\sigma_{22} & \cdots & d_1\sigma_{pp} \\ d_2\sigma_{11} & d_2\sigma_{22} & \cdots & d_2\sigma_{pp} \\ \cdots & \cdots & \cdots & \cdots \\ d_r\sigma_{11} & d_r\sigma_{22} & \cdots & d_r\sigma_{pp} \end{pmatrix},$$

$$\tilde{\mathbf{B}}\mathbf{D}_p^{-1}\tilde{\mathbf{B}}^T = \frac{2pr}{1-2pr\sum d_i^2}\begin{pmatrix} d_1^2 & d_1d_2 & \cdots & d_1d_r \\ d_1d_2 & d_2^2 & \cdots & d_2d_r \\ \cdots & \cdots & \cdots & \cdots \\ d_1d_r & d_2d_r & \cdots & d_r^2 \end{pmatrix}, \quad \mathbf{V}^T\tilde{\mathbf{B}}\mathbf{D}_p^{-1}\tilde{\mathbf{B}}^T\mathbf{V} = \frac{2pr\left(\sum V_id_i\right)^2}{1-2pr\sum d_i^2}.$$

Finally the NRR statistic in (1) for any diagonal covariance matrix of any dimensionality $p$ is

$$Y_n^2 = \sum V_i^2 + \frac{2pr\left(\sum V_id_i\right)^2}{1-2pr\sum d_i^2}. \quad (6)$$

The second term of $Y_n^2$ in (6) recovers information lost due to data grouping. Another useful decomposition of $Y_n^2 = U_n^2 + S_n^2$, where the Dzhaparidze-Nikulin (DN) statistic $U_n^2$ is

$$U_n^2 = \mathbf{V}_n^T(\hat{\theta}_n)[\mathbf{I} - \mathbf{B}_n(\mathbf{B}_n^T\mathbf{B}_n)^{-1}\mathbf{B}_n^T]\mathbf{V}_n(\hat{\theta}_n), \text{ and} \quad (7)$$

$$S_n^2 = Y_n^2 - U_n^2 = \mathbf{V}_n^T(\hat{\theta}_n)\mathbf{B}_n[(\mathbf{J}_n - \mathbf{B}_n^T\mathbf{B}_n)^{-1} + (\mathbf{B}_n^T\mathbf{B}_n)^{-1}]\mathbf{B}_n^T\mathbf{V}_n(\hat{\theta}_n) \quad (8)$$

has been proposed by McCulloch in 1985. McCulloch (1985), Theorem 4.2, proved that if rank of $\mathbf{B}$ is $s$, then $U_n^2$ and $S_n^2$ are asymptotically independent and distributed in the limit as $\chi^2_{r-s-1}$ and $\chi^2_s$ respectively. In the sequel, after a very good idea of McCulloch, we shall call the $S_n^2$ as the McCulloch's (McCull) statistic. Since in our case the first $p$ columns of the matrix $\mathbf{B}$ are columns of zeros and the rest are linearly dependent, the matrix $\mathbf{B}_n$ has rank 1. From the above it follows that $(\mathbf{B}_n^T\mathbf{B}_n)^{-1}$ does not exist, but, using the well known facts from the theory of multivariate normal distribution (see e.g., Moore (1977), p.132), we may replace $\mathbf{A}^{-1} = (\mathbf{B}_n^T\mathbf{B}_n)^{-1}$ by $\mathbf{A}^{-} = (\mathbf{B}_n^T\mathbf{B}_n)^{-}$, where $\mathbf{A}^{-}$ is any generalized matrix inverse of $\mathbf{A}$. So, for testing multivariate normality with random cells $E_{in}(\hat{\boldsymbol{\theta}}_n)$ we may use: the NRR statistic $Y_n^2$ defined by (6), the DN statistic

$$U_n^2 = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)[\mathbf{I} - \mathbf{B}_n(\mathbf{B}_n^T\mathbf{B}_n)^{-}\mathbf{B}_n^T]\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n) = \sum V_i^2 - \frac{1}{\sum d_i^2}\left(\sum V_i d_i\right)^2, \text{ and} \qquad (9)$$

$$S_n^2 = Y_n^2 - U_n^2 = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)\mathbf{B}_n[(\mathbf{J}_n - \mathbf{B}_n^T\mathbf{B}_n)^{-1} + (\mathbf{B}_n^T\mathbf{B}_n)^{-}]\mathbf{B}_n^T\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n) = \frac{\left(\sum V_i d_i\right)^2}{\left(1 - 2pr\sum d_i^2\right)\sum d_i^2} \qquad (10)$$

that have asymptotically $\chi_{r-1}^2,\ \chi_{r-2}^2,$ and $\chi_1^2$ distributions correspondingly.

**Proposition 2**. In our multivariate case statistics $U_n^2$ in (9) and $S_n^2$ in (10) are asymptotically independent, and, hence, $U_n^2$ and $S_n^2$ can be used as test statistics independently from each other (for a proof see Appendix 2).

**Proposition 3.** The right-hand-side rejection region tests of multivariate normality based on $Y_n^2$ and $S_n^2$ are consistent against all alternatives that approximately satisfy the condition $P(\|X\|^2 \le c_i) \ne i/r$ for at least one $i \in \{1,...,r-1\}$.
This proposition can be easily proved using arguments of Henze (2002), p. 484.

The above results for testing the MVN null hypothesis with a symmetrical covariance matrix $\boldsymbol{\Sigma}$ suggest the following procedure: 1) produce the Karhunen-Loéve transformation of a sample data that will diagonalize a sample covariance matrix, and 2) use the statistics $Y_n^2,\ U_n^2,$ and $S_n^2$ as defined in formulas (6), (9), and (10). Let $\boldsymbol{\Phi} = [\mathbf{e}_1\ \mathbf{e}_2 \cdots \mathbf{e}_p]$ be a matrix whose columns $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_p$ are orthogonal normalized eigen-vectors of $\mathbf{S}$, then the Karhunen-Loéve transformation will be $\mathbf{Y}_i = \boldsymbol{\Phi}^T\mathbf{X}_i,\ i = 1,...,n.$ From this it follows that tests in (6), (9), and (10) with frequencies $N_i^{(n)},\ i = 1,...,n,$ defined by the number of vectors $\mathbf{Y}_1,...,\mathbf{Y}_n$ that will fall into intervals $\tilde{E}_{in}\left(\hat{\boldsymbol{\theta}}_n\right) = \left\{\mathbf{Y} \text{ in } \mathrm{R}^2:\ c_{i-1} \le \left(\mathbf{Y} - \bar{\mathbf{Y}}\right)^T \mathbf{S}_y^{-1}\left(\mathbf{Y} - \bar{\mathbf{Y}}\right) < c_i\right\},\ i = 1,...,r,$ where $\mathbf{S}_y$ is the sample covariance matrix of $\mathbf{Y}_1,...,\mathbf{Y}_n$, can be used. In the sequel we shall analyze tests (6), (9), and (10) based on the Karhunen-Loéve transformation of the original data.

## 3. Simulation study

A simulation study was conducted to compare the power of the Doornik and Hansen (2008), Henze and Zirkler (1990), Royston (1992) tests (DH, HZ, and R92 as in Farrell et al (2007)), Nikulin-Rao-Robson (NRR) test (6), McCulloch's (McCull) test (10), Anderson-Darling (AD) and Cramer -von Mises (CM) tests as they were defined in Henze (2002), p. 483. Throughout this study only simulated critical values for the right-tailed rejection region of size $\alpha = 0.05$ for all test statistics were used.

In the sequel we shall not consider the Dzhaparidze-Nikulin (DN) and Chernoff-Lehmann (ChLeh) (see Moore and Stubblebine (1981), p. 718) tests because of their low power (see, e.g., Figure 1).

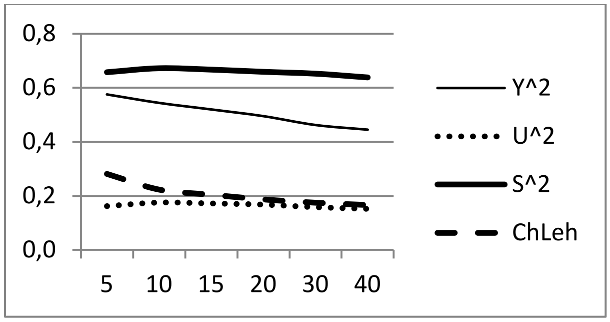
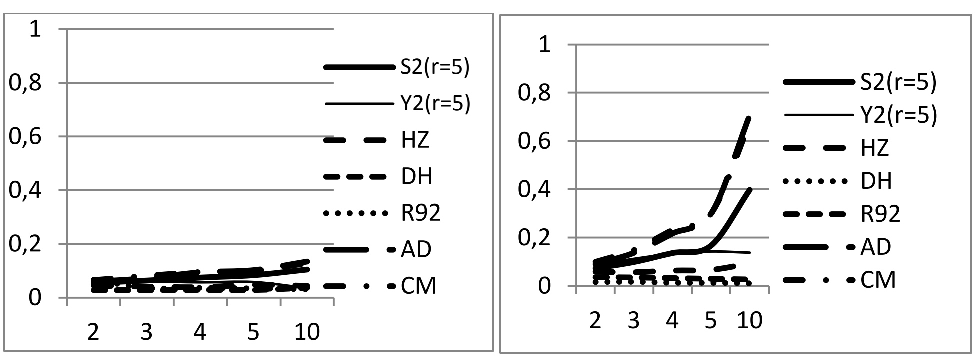


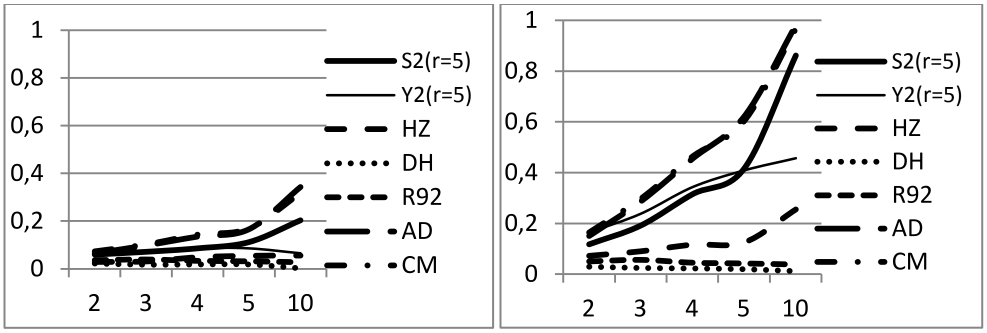

**Figure 1**: Simulated power of the NRR ( $Y^\wedge 2$ ), DN ( $U^\wedge 2$ ), McCull ( $S^\wedge 2$ ), and ChLeh tests as the function of the number *r* of equiprobable cells. The alternative is the two-dimensional Student *t* distribution with 10 degrees of freedom (d.f.) with dependent components (formula (5) of Farrell et al (2007) with $\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$ was used for simulation) , $p$ = 2, $n$ = 250, N = 5000.

From this figure we see that the power of McCull test ( $S^\wedge 2$ ) in this particular case is the highest. Another important feature of the McCull test is that its power almost does not depend on *r*. The same behavior was observed for all alternatives considered in this research. This very nice property of the McCull test distinguishes it from all other chi-squared type tests, the power of which essentially depends on *r* (see e.g. Henze (2002), p. 484). Since in all our simulations the power of McCull test attains the smooth maximum at *r* = 5, in the sequel we shall use *r* = 5 as a universal recommended number of grouping cells for both McCull and NRR tests (at least for alternatives that are investigated in this research).

Some results of our simulation study are presented in Figures 2-8 below.



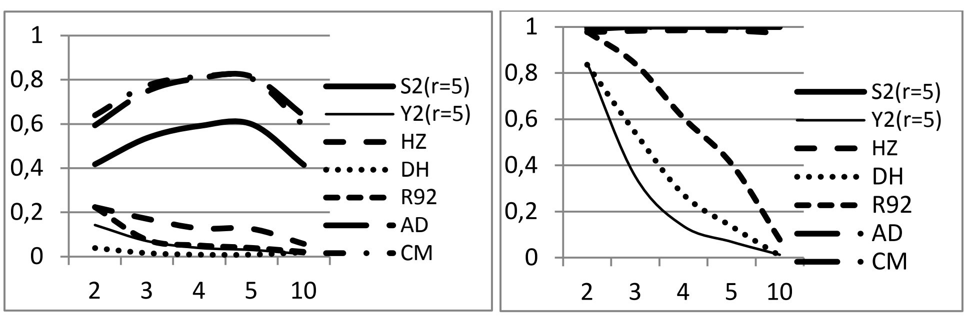

(a) n=25 (c) n=100



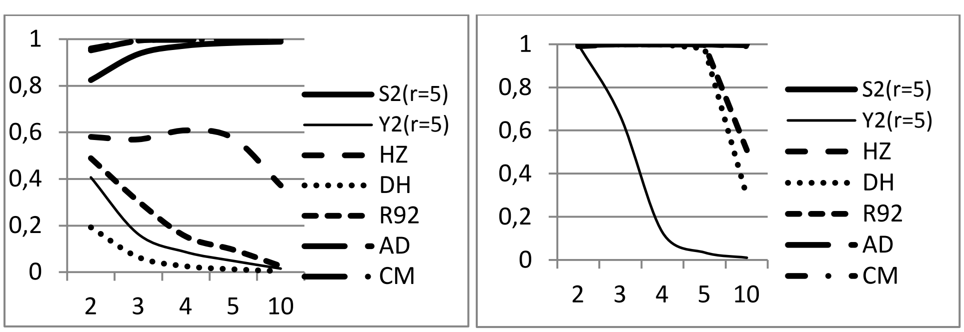

(b) n=50 (d) n=250

**Figure 2**: Empirical powers for the DH, HZ, R92, NRR (Y2), McCull (S2), AD, and CM test statistics for the Pearson Type II alternative distribution with shape parameter *m* = 10. Based on 20,000 samples of sizes *n* = 25, 50, 100, and 250, for dimensionalities *p* = 2,3,4,5, and 10.

From Fig. 2 we see that w.r.t. the Pearson Type II distribution with shape parameter *m* = 10 the power of AD and CM tests is the largest for all sample sizes and dimensionalities considered. McCull test $S2$ possesses lower power that is still comparable with that of AD and CM statistics. The NRR test, we proposed, possesses almost the same power as $S2$ test for $n \geq 50$ and *p* = 2,3,4,5. One sees also that HZ, R92, and DH tests are not competitive with the rest four tests.



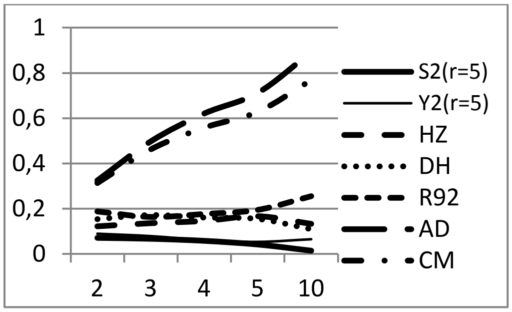

(a) n=25 (c) n=100



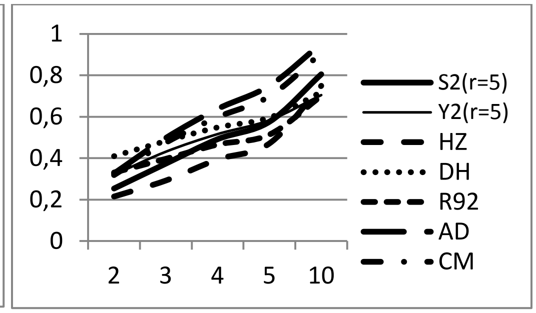

(b) n=50 (d) n=250

**Figure 3**: Empirical powers for the DH, HZ, R92, NRR (Y2), McCull (S2), AD, and CM test statistics w.r.t. the multivariate uniform distribution (Pearson Type II, m=0). Based on 20,000 samples of sizes *n* = 25, 50, 100, and 250, for *p* = 2,3,4,5, and 10.

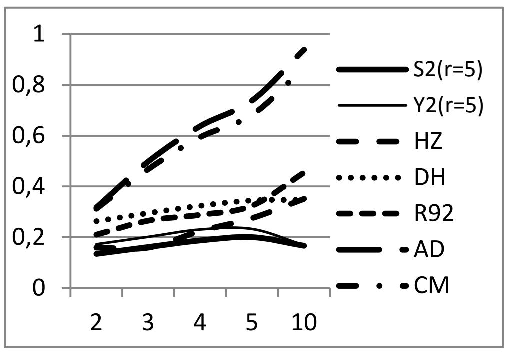
From Fig. 3 we see that for samples of size $n$ = 25 and $n$ = 50 under short-tailed symmetrical alternatives the HZ, R92, NRR, and DH tests cannot be recommended for applications. For samples of size $n$ = 100 and $n$ = 250 only McCull, AD, CM, and HZ tests are the best.



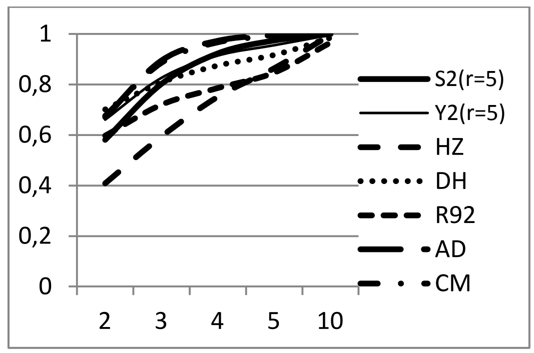

(a) n=25




(c) n=100



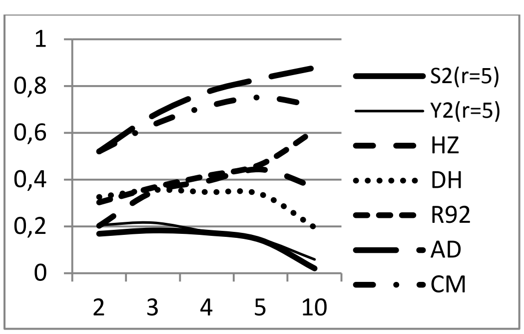

(b) n=50



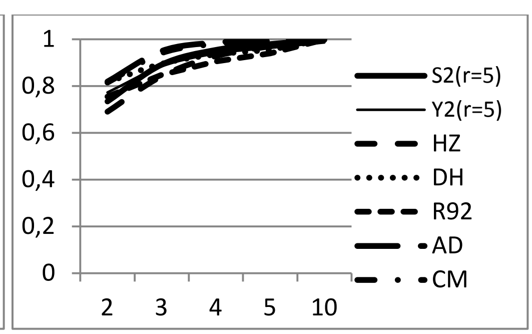

(d) n=250

**Figure 4**: Empirical powers for the DH, HZ, R92, NRR (Y2), McCull (S2), AD, and CM test statistics against the multivariate Student *t* distribution with 10 degrees of freedom. Based on 20,000 samples of sizes $n$ =25, 50, 100, and 250 for $p$ =2, 3, 4, 5, and 10.

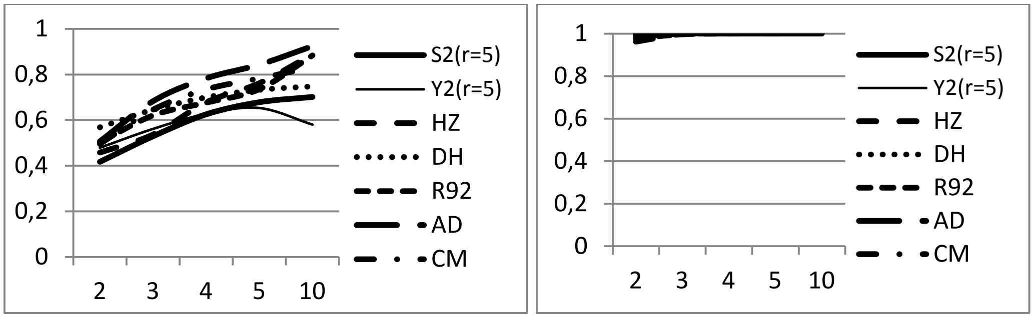
From Fig. 4 one sees that against long-tailed Student *t* distribution with 10 d.f. for $n$ = 25 and $n$ = 50 only AD and CM tests can be recommended. For $n$ = 100 and $n$ = 250 all tests are approximately equivalent with the exception of HZ test that is more conservative for dimensionalities $p$ = 2,3,4, and 5.




(a) n=25



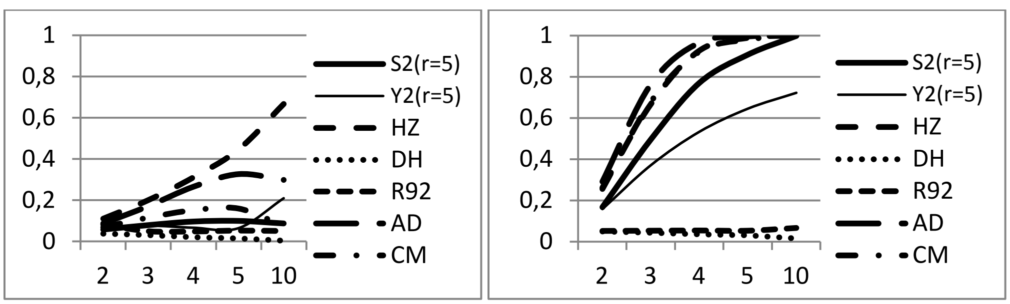

(c) n=100



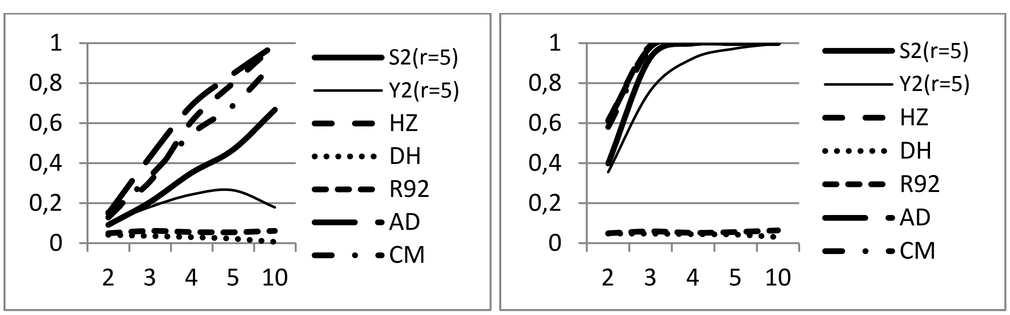

(b) n=50 (d) n=250

**Figure 5**: Empirical powers for the DH, HZ, R92, NRR (Y2), McCull (S2), AD, and CM test statistics against the multivariate Student *t* distribution with 5 degrees of freedom. Based on 20,000 samples of sizes $n$ = 25, 50, 100, and 250, for $p$ = 2, 3, 4, 5, and 10.

From Fig. 5. we see that AD and CM tests are perfect for samples of size $n$ = 25 and that all other tests under consideration are not good except, possibly, R92 for $p$ = 10. For samples of size $n > 25$ all tests seem to be equivalent.



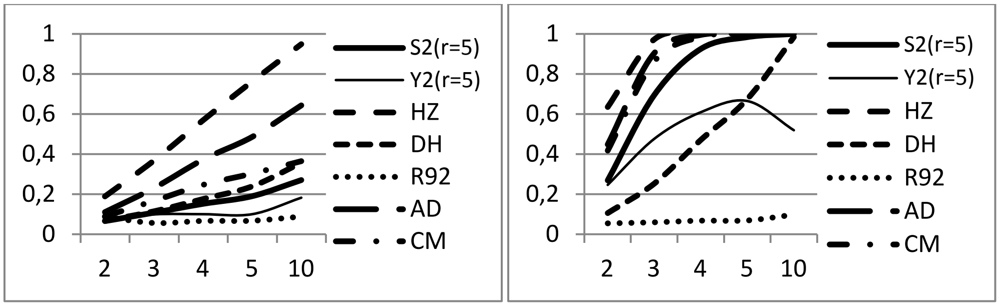

(a) n=25 (c) n=100



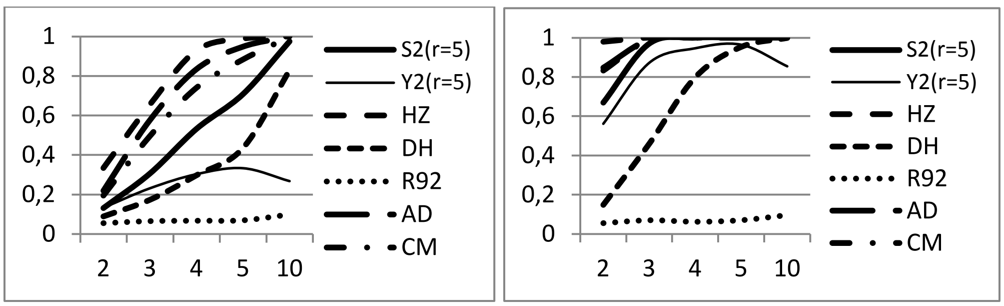

(b) n=50 (d) n=250

**Figure 6**: Empirical powers for the DH, HZ, R92, NRR (Y2), McCull (S2), AD, and CM test statistics against the alternative multivariate Khinchine (KHN) distribution with normal marginals . Based on 20,000 samples of sizes $n$ = 25, 50, 100, and 250 for $p$ = 2, 3, 4, 5, and 10.

From Fig. 6. we see that against KHN alternative the HZ test possesses an evident advantage for $n$ = 25. For $n$ = 50 and more the best are AD, CM, and HZ tests. The McCull test is comparable with them for $n \geq 100$. It is of importance to note that for all sample sizes considered the R92 and DH tests possess no power and are biased w.r.t. right-tailed rejection region. It is quite possible that these tests are powerful and unbiased for left-tailed rejection region, but this needs a separate thorough investigation.



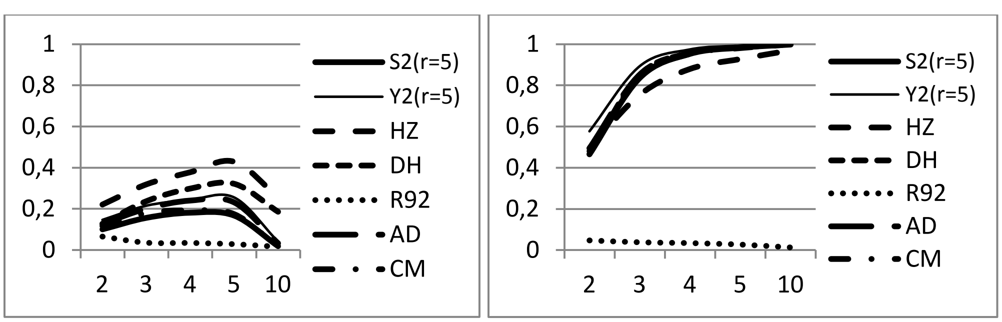

(a) n=25 (c) n=100



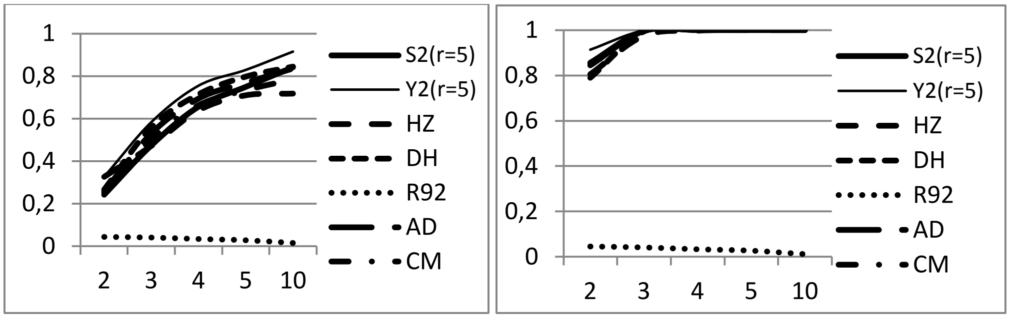

(b) n=50 (d) n=250

**Figure** 7: Empirical powers for the DH, HZ, R92, NRR (Y2), McCull (S2), AD, and CM test statistics against the mixture of normal distributions 0.5N(**0,B**)+0.5N(**0,I**). **B** denotes a matrix with 1 on diagonal and 0.9 off diagonal, **I** is the identity matrix. Based on 20,000 samples of sizes $n$ = 25, 50, 100, and 250 for $p$ = 2, 3, 4, 5, and 10.

From Fig. 7 we see that for $n$ = 25 only HZ and AD tests can be recommended. For $n > 50$ HZ, AD, CM, and McCull tests perform well, and DH test is good only for $p$ = 10. It is clear that R92 test is not powerful and biased (see comment to Fig. 6).




(a) n=25 (c) n=100




(b) n=50 (d) n=250

**Figure 8**: Empirical powers for the DH, HZ, R92, NRR (Y2), McCull (S2), AD, and CM test statistics against the mixture of normal distributions 0.9N(**0**,**B**)+0.1N(**0**,**I**). Based on 20,000 samples of sizes $n$ = 25, 50, 100, and 250 for $p$ = 2, 3, 4, 5, and 10.

From Fig. 8 we see that for n = 25 the HZ tests is the best. For $n \geq 50$ all tests except R92 perform well. Note also that again R92 test is biased and, hence does not work.

From all the above one sees that no one test considered is a panacea when testing for multivariate normality against symmetrical alternatives. Much depends on a prior knowledge about a reasonable alternative.

## 4. Conclusion and recommendations

To summarize we may to conclude that no one of tests considered can be recommended as a universal one. All tests have a right to be used in practice, but, before selecting a proper test, it is highly desirable to have some imagination about possible alternatives. If, on the contrary, one has no idea about an alternative, then the AD, CM, HZ, and McCull tests can be used.

**Acknowledgements**

The authors are grateful to Matias Saliban-Barrera for giving us an access to R-codes used in Farrell et al (2007).

## Appendix 1

**A proof of the Proposition 1.**
Formula (5) holds when $p$ = 2 (see formula (2)). Let it hold for any $p$ >2, then

$$\mathbf{D}_p = \begin{Bmatrix} \frac{1-2r\sum d_i^2}{2\sigma_{11}^2} & -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{22}} & \cdots & -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{p-1,p-1}} & -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{pp}} \\ -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{22}} & \frac{1-2r\sum d_i^2}{2\sigma_{22}^2} & \cdots & -\frac{r\sum d_i^2}{\sigma_{22}\sigma_{p-1,p-1}} & -\frac{r\sum d_i^2}{\sigma_{22}\sigma_{pp}} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{p-1,p-1}} & -\frac{r\sum d_i^2}{\sigma_{22}\sigma_{p-1,p-1}} & \cdots & \frac{1-2r\sum d_i^2}{2\sigma_{p-1,p-1}^2} & -\frac{r\sum d_i^2}{\sigma_{p-1,p-1}\sigma_{pp}} \\ -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{pp}} & -\frac{r\sum d_i^2}{\sigma_{22}\sigma_{pp}} & \cdots & -\frac{r\sum d_i^2}{\sigma_{p-1,p-1}\sigma_{pp}} & \frac{1-2r\sum d_i^2}{2\sigma_{pp}^2} \end{Bmatrix}$$

can be presented as $\mathbf{D}_p = \begin{pmatrix} \mathbf{D}_{p-1} & \mathbf{A}_p \\ \mathbf{A}_p^T & \mathbf{C}_p \end{pmatrix}$, where $\mathbf{D}_{p-1}$ is the upper left $(p-1)\times(p-1)$ submatrix of $\mathbf{D}_p$, $\mathbf{C}_p = \dfrac{1-2r\sum d_i^2}{2\sigma_{pp}^2}$, and $\mathbf{A}_p = \left( -\dfrac{r\sum d_i^2}{\sigma_{11}\sigma_{pp}} \;\; -\dfrac{r\sum d_i^2}{\sigma_{22}\sigma_{pp}} \;\cdots\; -\dfrac{r\sum d_i^2}{\sigma_{p-1,p-1}\sigma_{pp}} \right)^T$.

Let us show that (5) holds for $p$ + 1. Due to Rao (1968), p.29, problem 2.7, the matrix $\mathbf{D}_{p+1}^{-1}$ can be evaluated as

$$\mathbf{D}_{p+1}^{-1} = \begin{pmatrix} \mathbf{D}_p & \mathbf{A}_{p+1} \\ \mathbf{A}_{p+1}^T & \mathbf{C}_{p+1} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{D}_p^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}^T & \mathbf{E}^{-1} \end{pmatrix}, \qquad (11)$$

where

$$\mathbf{F} = \mathbf{D}_p^{-1}\mathbf{A}_{p+1} = \frac{4}{1-2pr\sum d_i^2} \begin{pmatrix} \frac{(1-2(p-1)r\sum d_i^2)\sigma_{11}^2}{2} & r\sum d_i^2\sigma_{11}\sigma_{22} & \cdots & r\sum d_i^2\sigma_{11}\sigma_{pp} \\ r\sum d_i^2\sigma_{11}\sigma_{22} & \frac{(1-2(p-1)r\sum d_i^2)\sigma_{22}^2}{2} & \cdots & r\sum d_i^2\sigma_{22}\sigma_{pp} \\ \cdots & \cdots & \cdots & \cdots \\ r\sum d_i^2\sigma_{11}\sigma_{pp} & r\sum d_i^2\sigma_{22}\sigma_{pp} & \cdots & \frac{(1-2(p-1)r\sum d_i^2)\sigma_{pp}^2}{2} \end{pmatrix} \bullet$$

$$\bullet \begin{pmatrix} -\frac{r\sum d_i^2}{\sigma_{11}\sigma_{p+1,p+1}} \\ -\frac{r\sum d_i^2}{\sigma_{22}\sigma_{p+1,p+1}} \\ \cdots \\ -\frac{r\sum d_i^2}{\sigma_{pp}\sigma_{p+1,p+1}} \end{pmatrix} = -\frac{2r\sum d_i^2}{(1-2pr\sum d_i^2)\sigma_{p+1,p+1}} \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \cdots \\ \sigma_{pp} \end{pmatrix},$$

$$\mathbf{E} = \mathbf{C}_{p+1} - \mathbf{A}_{p+1}^T\mathbf{F} = \frac{1-2(p+1)r\sum d_i^2}{2(1-2pr\sum d_i^2)\sigma_{p+1,p+1}^2},$$

$$-\mathbf{F}\mathbf{E}^{-1} = \frac{4M\sum d_i^2\sigma_{p+1,p+1}}{1-2(p+1)M\sum d_i^2} \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \cdots \\ \sigma_{pp} \end{pmatrix},$$

$$\mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T = \frac{8(r\sum d_i^2)^2}{(1-2pr\sum d_i^2)(1-2(p+1)r\sum d_i^2)} \begin{pmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{22} & \cdots & \sigma_{11}\sigma_{pp} \\ \sigma_{11}\sigma_{22} & \sigma_{22}^2 & \cdots & \sigma_{22}\sigma_{pp} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{11}\sigma_{pp} & \sigma_{22}\sigma_{pp} & \cdots & \sigma_{pp}^2 \end{pmatrix},$$

$$\mathbf{E}^{-1} = \frac{4}{1-2(p+1)r\sum d_i^2} \cdot \frac{(1-2pr\sum d_i^2)\sigma_{p+1,p+1}^2}{2}.$$

Substituting the above expressions into (11) we see that the expression (5) holds for $p$ + *1*, and, hence, by induction is valid for any $p \geq 2$.

## Appendix 2

From the theory of quadratic forms it is known that if $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive definite matrix, then the two quadratic forms $\mathbf{Y}^T\mathbf{AY}$ and $\mathbf{Y}^T\mathbf{BY}$ are distributed independently if $\mathbf{A\Sigma B} = \mathbf{0}$. Since asymptotically $\mathbf{V}_n(\hat{\theta}_n) \sim N(\mathbf{0}, \mathbf{I} - \mathbf{BJ}^{-1}\mathbf{B}^T)$, then to prove that $U_n^2$ and $S_n^2$ are independent we have to show that

$$\left\{\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-}\mathbf{B}^T\right\}\left\{\mathbf{I} - \mathbf{BJ}^{-1}\mathbf{B}^T\right\}\left\{\mathbf{B}\left[(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1} + (\mathbf{B}^T\mathbf{B})^{-}\right]\mathbf{B}^T\right\} = \mathbf{0}. \qquad (12)$$

Let $(\mathbf{B}^T\mathbf{B})^{-}$ be the Moore-Penrose matrix inverse for $\mathbf{B}^T\mathbf{B}$ (it is well known that under normality quadratic forms do not depend on the way of inversion). From the properties of the Moore-Penrose matrix inverse we have $(\mathbf{B}^T)^{-}\mathbf{B}^T = (\mathbf{B}^{-})^T\mathbf{B}^T = (\mathbf{BB}^{-})^T = \mathbf{BB}^{-}$, so

$$\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-}\mathbf{B}^T = \mathbf{BB}^{-}(\mathbf{B}^T)^{-}\mathbf{B}^T = \mathbf{BB}^{-}(\mathbf{B}^{-})^T\mathbf{B}^T = \mathbf{BB}^{-}\mathbf{BB}^{-} = \mathbf{BB}^{-}.$$

Using matrix multiplication, from (12) we get

$$\begin{aligned}
&\left\{\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-}\mathbf{B}^T\right\}\left\{\mathbf{I} - \mathbf{BJ}^{-1}\mathbf{B}^T\right\}\left\{\mathbf{B}\left[(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1} + (\mathbf{B}^T\mathbf{B})^{-}\right]\mathbf{B}^T\right\} = \\
&\left\{\mathbf{I} - \mathbf{BB}^{-}\right\}\left\{\mathbf{I} - \mathbf{BJ}^{-1}\mathbf{B}^T\right\}\left\{\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-}\mathbf{B}^T\right\} = \qquad (13)\\
&\left\{\mathbf{I} - \mathbf{BB}^{-}\right\}\left\{\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{BB}^{-} - \mathbf{BJ}^{-1}\mathbf{B}^T\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{BJ}^{-1}\mathbf{B}^T\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-}\mathbf{B}^T\right\}.
\end{aligned}$$

The factor of the last term of the above expression is simplified as

$$\begin{aligned}
&\mathbf{B}^T\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-}\mathbf{B}^T = \mathbf{B}^T\mathbf{BB}^{-}(\mathbf{B}^T)^{-}\mathbf{B}^T = \mathbf{B}^T\mathbf{BB}^{-}\mathbf{BB}^{-} = \\
&\mathbf{B}^T\mathbf{BB}^{-} = \mathbf{B}^T(\mathbf{BB}^{-})^T = \mathbf{B}^T(\mathbf{B}^T)^{-}\mathbf{B}^T = \mathbf{B}^T.
\end{aligned}$$

Substituting this result into expression (13) and opening brackets we obtain

$$\begin{aligned}
&\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{BB}^{-} - \mathbf{BJ}^{-1}\mathbf{B}^T\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{BJ}^{-1}\mathbf{B}^T - \\
&\mathbf{BB}^{-}\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T - \mathbf{BB}^{-}\mathbf{BB}^{-} + \mathbf{BB}^{-}\mathbf{BJ}^{-1}\mathbf{B}^T\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T - \mathbf{BB}^{-}\mathbf{BJ}^{-1}\mathbf{B}^T = \\
&\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{BB}^{-} - \mathbf{BJ}^{-1}\mathbf{B}^T\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{BJ}^{-1}\mathbf{B}^T - \\
&\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T - \mathbf{BB}^{-} + \mathbf{BJ}^{-1}\mathbf{B}^T\mathbf{B}(\mathbf{J} - \mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T - \mathbf{BJ}^{-1}\mathbf{B}^T = \mathbf{0}, \text{ QED.}
\end{aligned}$$

## References

Bartoszyński, R., Pearl, D.K. and Lawrence J. (1997). A multidimensional goodness-of-fit test based on interpoint distances. *JASA*, vol. **92**, pp. 577-586.

Batsidis, A., Martin, N., Pardo, L., Zografos, K. (2013). A necessary power divergence type family tests of multivariate normality, *Commun. Statist.- Simula. Computa.*, **42**, pp. 2253-2271.

Batsidis, A., Zografos, K. (2013). A necessary test of fit of specific elliptical distributions based on an estimator of Song's measure. *J. of Multivariate Analysis*, **113**, pp. 91-105.

Cardoso de Oliveira, I,R., Ferreira, D.F. (2010). Multivariate extension of chi-squared univariate normality test. *Journal of Statistical Computation and Simulation*, Vol.**80**, pp. 513-526.

Doornik, J.A. and Hansen, H. (2008). Practitioners' corner: An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, **70**, supplement, pp. 927-939.

Farrell, P.J., Saliban-Barrera, M. and Naczk, K. (2007). On tests for multivariate normality and associated simulation studies. *Journal of Statistical Computation and Simulation*, Vol.77, pp. 1065-1080.

Follmann, D. (1996). A Simple Multivariate Test For One-Sided Alternatives. *JASA*, Vol. **91**, pp. 854-861.

Hanusz, Z., Tarasińska, J. (2012). New tests for multivariate normality based on Small's and Srivastava's graphical methods. *Journal of Statistical Computation and Simulation*, Vol.**82**, pp. 1743-1752.

Henze, N., Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Commun. Statist.- Theory and Methods*, **19**, pp. 3595-3617.

Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statistical papers*, Vol.**43**, pp. 467-506.

Hwu, T-J., Han, C-P., Rogers, K.J. (2002). The combination test for multivariate normality, *Journal of Statistical Computation and Simulation*, Vol.**72**, pp. 379-390.

Joenssen, D.W., Vogel, J. (2012). A power study of goodness-of-fit tests for multivariate normality implemented in R. *Journal of Statistical Computation and Simulation*, iFirst, 1-24.

Khatri, C.G. (1968). Some results for the singular normal multivariare regression models, *Sankhyā*, **30**, 267-280.

Keselman, H.J. (2005). Multivariate Normality Tests. In: Encyclopedia of Statistics in Behavioral Science, eds: Brian S. Everitt & David C. Howell, Vol.3, pp. 1373-1379, John Wiley and Sons.

McCulloch, C.E. (1985). “Relationships among some chi-square goodness of fit statistics”. *Communications is Statistics- Theory and Methods*. **14**. 593-603.

Mecklin, C.J. and Mundfrom, D.J. (2004). An appraisal and bibliography of tests for multivariate normality. *International Statistical Review*, Vol. **72**, pp. 123-138.

Mecklin, C.J. and Mundfrom, D.J. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation*, Vol. **75**, pp. 93-107.

Moore, D.S. (1977). Generalized inverses. Wald’s method. and the construction of chi-squared tests of fit. *J. of the American Statist. Assoc.*. **72**. 131-137.

Moore, D.S.. Stubblebine, J.B. (1981). Chi-square tests for multivariate normality with application to common stock prices. *Commun. Statist.-Theory and Methods*. **A10(8)**. 713-738.

Press, S.J. (1972). *Applied Multivariate Analysis*. New York: Holt. Rinehart and Winston.

Rao, C.R. (1968). *Linear statistical inference and its applications*, Wiley, New York.

Royston, P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing*, Vol.**2**, pp.117-119.

Small, N.J.H. (1978). Plotting squared radii. *Biometrika*, Vol.**65**, pp. 657-658.

Sürücü, B. (2006). Goodness-of-fit tests for multivariate distributions. *Commun. Statist.- Theory and Methods*. **35**. pp. 1319-1331.

Székely, G.J., Rizzo, M.L. (2005). A new test for multivariate normality. *J. Multivariate Analysis*, Vol.**93**, pp. 58-80.

von Eye, A. (2005). Comparing tests of multinormality- a Monte Carlo study. InterStat. http://interstat. statjournals.net (July, 2012).

Villasenor Alva, J.A., Gonzáles Estrada, E. (2009). A generalization of Shapiro-Wilk's test for multivariate normality. *Commun. Statist.- Theory and Methods*. **38**. pp. 1870-1883.

Voinov, V., Nikulin, M., Balakrishnan, N. (2013). *Chi-squared goodness of fit tests with applications.* Academic Press, Elsevier: Amsterdam, New York, etc.

Voinov, V. (2013a). A note on Wald’s type chi-squared tests of fit. *Commun. Statist.- Theory and Methods.* http://www.tandfonline.com/doi/full/10.1080/03610926.2013.768668.

Voinov, Y. (2013b). A Wald's type goodness-of-fit test for binormality. (Submitted to proceedings of the JSM 2013).