A Wald's Type Goodness-of-fit Test for Binormality

Yevgeniy Voinov [1]
Institute for Mathematics and Mathematical Modeling, Ministry of Education and Science of the Republic of Kazakhstan, 125 Pushkin Str., 050010 Almaty, Kazakhstan

Abstract
A new Wald's type chi-squared invariant goodness-of-fit test for binormality is introduced. The test is based on a linear transformation of a two-dimensional sample from a population that diagonalizes the sample covariance matrix, and a modification of Moore and Stubblebine technique for construction chi-squared type tests proposed in 1981. More precise formulation of the well known Moore's 1977 theorem given by Voinov in 2013 permitted to get this new result. A comparison of simulated power of the test with respect to numerous alternatives is presented. The simulated power of the proposed modified McCulloch's test with respect to nine different alternatives is comparable with the power of the well-known Anderson-Darling, Cramer-von Mises, Henze and Zirkler , Doornik and Hansen, and, modified by Royston in 1992, the Shapiro-Wilk's W tests. The overall conclusion of this research is that all seven tests considered can be used in practice.

Key Words: Binormality, chi-squared goodness-of-fit, invariant tests, power of tests

Introduction

The assumption of multivariate normality (MVN), in particular of binormality, is of great importance for applied multivariate statistics, e.g., for analysis of variance, discriminant analysis, canonical correlation and factor analysis, analysis of regression residuals and residuals in time-series models, etc. To check that assumption the most powerful tests should be used, because tests with low power cannot discriminate for sure between the null hypothesis of binormality and supposed alternatives. Several examples of such a situation in the univariate case can be mentioned (see, e.g., Voinov et al. [1], Sec. 3.10.1, and Sec. 3.9). Henze [2, p.468] stated that those tests of MVN, and binormality must be invariant with respect to affine transformations of sample data, and consistent. Consistency means that the probability to fall into rejection region under alternative should tend to 1 if sample size $n \to \infty$.

Two nice surveys of the state of art are known: Henze [2], and Mecklin and Mundfrom [3]. Those surveys list tenth of publications devoted to testing of MVN , and binormality as a particular case. It is almost impossible to find out and cite all existing papers on tests of binormality. We would like to mention here the papers of Szkutnik [4], Versluis [5], Bogdan [6], Székely and Rizzo [7], Sürücü [8], Doornik and Hansen [9], Villasenor Alva and Estrada [10], and Hanusz and Tarasińska [11].

In this article we introduce new invariant and consistent chi-squared goodness-of-fit tests for binormality. Section 1 is devoted to the theoretical background of the proposal. Closed form expressions for the tests proposed are derived in Section 2. In Section 3 a

---

[1] Email: v.yevgeniy@yahoo.com
This research has been supported by the grant 0760/GF of the Ministry of Education and Science of the Republic of Kazakhstan.

Monte-Carlo study of power comparison of different tests is presented. A conclusion and recommendations are given in Section 4.

1. Theoretical background

The theory of modified chi-squared statistics for tests of fit (Moore and Spruill [12], and Moore [13]) can be briefly described as follows. Let $X_1,...,X_n$ be independent identically distributed (iid) univariate random variables, and we intend to test the composite null hypothesis $\boldsymbol{H}_0$ that the distribution function of X_i belongs to a parametric family $F(x,\boldsymbol{\theta})$ of continuous distribution functions, where $\boldsymbol{\theta}=(\theta_1,...,\theta_s)^T \in \boldsymbol{\Theta} \subset R^s$, $\boldsymbol{\Theta}$ being an open set. Chi-squared tests of fit for $\boldsymbol{H}_0$ are based on the observed frequencies $N_j^{(n)}$, the number of $X_1,...,X_n$ that fall into intervals $\boldsymbol{I}_j$ such that $\boldsymbol{I}_i \cap \boldsymbol{I}_j = \varnothing$ for $i \neq j$, and $\boldsymbol{I}_1 \cup \cdots \cup \boldsymbol{I}_r = R^1$, $i, j = 1,...,r$. Since probabilities $p_i(\boldsymbol{\theta}) = \int_{\boldsymbol{I}_i} dF(x,\boldsymbol{\theta})$ depend on unknown parameter $\boldsymbol{\theta}$, the parameter should be estimated from data by any $\sqrt{n}$-consistent estimator $\boldsymbol{\theta}_n = \boldsymbol{\theta}_n(X_1,...,X_n)$. Denote $\mathbf{V}^{(n)}(\boldsymbol{\theta}_n)$ the *r*-vector of standardized frequencies with components $[N_i^{(n)} - np_i(\boldsymbol{\theta}_n)]/[np_i(\boldsymbol{\theta}_n)]^{1/2}$, $i=1,...,r$. Moore and Spruill [12] showed that the vector $\mathbf{V}^{(n)}(\boldsymbol{\theta}_n)$ asymptotically follows a multivariate normal distribution $N_r(\mathbf{0},\boldsymbol{\Sigma})$, where $\mathbf{0}$ is a zero *r*-vector of means, and $\boldsymbol{\Sigma}$ is a nonsingular covariance matrix. The covariance matrix $\boldsymbol{\Sigma}$ essentially depends on the way of the parameter $\boldsymbol{\theta}$ estimation. Wald's idea was as follows. Let a random vector $\boldsymbol{X}$ possess the $N_r(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution and $\boldsymbol{\Sigma}$ is nonsingular of rank *r*, then the quadratic form $(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$ will be distributed in the limit as chi-squared with *r* degrees of freedom (d.f.), χ_r^2. Nikulin [14] was first who used this idea and MLE $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ for testing univariate normality when the matrix $\boldsymbol{\Sigma}$ is singular of rank $r-1$. Moore [13] generalized Wald's approach and developed the method of the construction of chi-squared tests of fit valid for any $\sqrt{n}$-consistent estimator $\boldsymbol{\theta}_n$. The Moore's method is simple. One has to find the limit covariance matrix $\boldsymbol{\Sigma}$ of the vector $\mathbf{V}^{(n)}(\boldsymbol{\theta}_n)$ of standardized frequencies for a particular $\sqrt{n}$-consistent estimator $\boldsymbol{\theta}_n$, derive any generalized matrix inverse $\boldsymbol{\Sigma}^-$, and, for testing $\boldsymbol{H}_0$, use the statistic $Y_n(\boldsymbol{\theta}_n) = \mathbf{V}^{(n)T}(\boldsymbol{\theta}_n)\boldsymbol{\Sigma}_n^{-1}\mathbf{V}^{(n)}(\boldsymbol{\theta}_n)$, where $\boldsymbol{\Sigma}_n^-$ is an estimate of $\boldsymbol{\Sigma}^-$. Voinov [15], Theorem 2, noted that the statistic $Y_n(\boldsymbol{\theta}_n)$ will follow the chi-squared limit distribution if and only if the limit covariance matrix $\boldsymbol{\Sigma}$ of the *r*-vector $\mathbf{V}^{(n)}(\boldsymbol{\theta}_n)$ of standardized frequencies does not depend on unknown parameters, and, hence, the statistic to be used is

$$Y_n(\boldsymbol{\theta}_n) = \mathbf{V}^{(n)T}(\boldsymbol{\theta}_n)\boldsymbol{\Sigma}^{-1}\mathbf{V}^{(n)}(\boldsymbol{\theta}_n)\,. \tag{1}$$

Let $\mathbf{X}_1,...,\mathbf{X}_n$ be independent identically distributed two-dimensional normal random vectors with the following joint probability density function

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

where $\boldsymbol{\mu}$ is a two-dimensional vector of means and $\boldsymbol{\Sigma}$ is a non-singular 2×2 covariance matrix. Let a given vector of unknown parameters be $\boldsymbol{\theta}=\left(\mu_1,\mu_2,\sigma_{11},\sigma_{12},\sigma_{22}\right)^{T}$. Given $0=c_0<c_1<\cdots<c_r=\infty$, the r grouping cells can be defined as (Moore and Stubblebine [16])

$$E_{in}\left(\hat{\boldsymbol{\theta}}_n\right)=\left\{\mathbf{X}\text{ in }\mathrm{R}^2:\ c_{i-1}\le\left(\mathbf{X}-\bar{\mathbf{X}}\right)^{T}\mathbf{S}^{-1}\left(\mathbf{X}-\bar{\mathbf{X}}\right)<c_i\right\},\ i=1,...,r,$$

where $\bar{\mathbf{X}}$ and $\mathbf{S}$ are maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ correspondingly. The estimated probability to fall into each cell is

$$p_{in}\left(\hat{\boldsymbol{\theta}}_n\right)=\int_{E_{in}\left(\hat{\boldsymbol{\theta}}_n\right)}f(\mathbf{x}\,|\,\hat{\boldsymbol{\theta}}_n)d\mathbf{x}.$$

If c_i is the i/r point of the central chi-squared distribution with 2 degrees of freedom, then the cells are equiprobable under the estimated parameter value, $\hat{\boldsymbol{\theta}}_n$, and $p_{in}\left(\hat{\boldsymbol{\theta}}_n\right)=1/r,\ i=1,...,r.$ Denote $\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n)$ a vector of standardized cell frequencies with the components $V_{in}\left(\hat{\boldsymbol{\theta}}_n\right)=\left(N_i^{(n)}-n/r\right)/\sqrt{n/r},\ i=1,...,r,$ where $N_i^{(n)}$ is the number of random vectors $\mathbf{X}_1,...,\mathbf{X}_n$ falling in $E_{in}\left(\hat{\boldsymbol{\theta}}_n\right)$. The 5×5 Fisher information matrix $\mathbf{J}\left(\boldsymbol{\theta}\right)$ for one observation can be presented as (Moore and Stubblebine [16])

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

where $\mathbf{Q}$ is the 3×3 covariance matrix of $\mathbf{r}=\left(s_{11},s_{12},s_{22}\right)^{T}$, a vector of the entries of $\sqrt{n}\mathbf{S}$. The elements of $\mathbf{Q}$ can be written as (Press [17])

$$\mathbf{Var}\left(s_{ij}\right)=\sigma_{ij}^2+\sigma_{ii}\sigma_{jj},\quad i,j=1,2,\ i\le j,$$

$$\mathbf{Cov}\left(s_{ij},s_{kl}\right)=\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk},\quad i,j,k,l=1,2,\ i\le j,k\le l,$$

where $\sigma_{ij},\ i,j=1,2,$ are elements of $\boldsymbol{\Sigma}$. In our case the matrix $\mathbf{Q}$ will be

$$\mathbf{Q}=\begin{pmatrix}\mathbf{Var}\left(s_{11}\right) & \mathbf{Cov}\left(s_{11},s_{12}\right) & \mathbf{Cov}\left(s_{11},s_{22}\right)\\ \mathbf{Cov}\left(s_{12},s_{11}\right) & \mathbf{Var}\left(s_{12}\right) & \mathbf{Cov}\left(s_{12},s_{22}\right)\\ \mathbf{Cov}\left(s_{22},s_{11}\right) & \mathbf{Cov}\left(s_{22},s_{12}\right) & \mathbf{Var}\left(s_{22}\right)\end{pmatrix}=\begin{pmatrix}2\sigma_{11}^2 & 2\sigma_{11}\sigma_{12} & 2\sigma_{12}^2\\ 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22}+\sigma_{12}^2 & 2\sigma_{12}\sigma_{22}\\ 2\sigma_{12}^2 & 2\sigma_{12}\sigma_{22} & 2\sigma_{22}^2\end{pmatrix}. \quad (2)$$

Following Moore and Stubblebine [16] for a specified $\boldsymbol{\theta}_0$ define

$$p_i\left(\boldsymbol{\theta},\boldsymbol{\theta}_0\right)=\int_{E_i\left(\boldsymbol{\theta}_0\right)}f(\mathbf{x}\,|\,\boldsymbol{\theta})d\mathbf{x}\text{, where}$$

$$E_i\left(\boldsymbol{\theta}_0\right)=\left\{\mathbf{X}\text{ in }\mathrm{R}^2:\ c_{i-1}\le\left(\mathbf{X}-\boldsymbol{\mu}\right)^{T}\boldsymbol{\Sigma}_0^{-1}\left(\mathbf{X}-\boldsymbol{\mu}\right)<c_i\right\},\ i=1,...,r.$$

Define $r\times 3$ matrix $\mathbf{B}=\mathbf{B}\left(\boldsymbol{\theta},\boldsymbol{\theta}_0\right)$ with its elements being

$$B_{ij} = \frac{1}{\sqrt{p_i(\boldsymbol{\theta},\boldsymbol{\theta}_0)}} \frac{\partial p_i(\boldsymbol{\theta},\boldsymbol{\theta}_0)}{\partial \theta_j},\ i=1,...,r,\ j=1,2,3.$$

From Lemma 1 of Moore and Stubblebine [16] it follows that for any c_i and $\boldsymbol{\theta}_0$

$$\left.\frac{\partial p_i(\boldsymbol{\theta},\boldsymbol{\theta}_0)}{\partial \mu_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0,\ \ 1\le i\le r,\ 1\le j\le 2,$$

$$\left.\frac{\partial p_i(\boldsymbol{\theta},\boldsymbol{\theta}_0)}{\partial \sigma_{jk}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = d_i\sigma^{jk},\ \ 1\le i\le r,\ 1\le j\le k\le 2,$$

where σ^{jk} are the elements of $\boldsymbol{\Sigma}^{-1}$ and

$$d_i = \frac{1}{4}\left(c_{i-1}e^{-c_{i-1}/2} - c_i e^{-c_i/2}\right),\ i=1,...,r.$$

As per Moore and Stubblebine [16] the Wald's type statistic based on the MLEs can be presented as

$$Y_n^2(\hat{\boldsymbol{\theta}}_n) = \mathbf{V}^{(n)T}(\hat{\boldsymbol{\theta}}_n)\boldsymbol{\Sigma}_n^{-}\mathbf{V}^{(n)}(\hat{\boldsymbol{\theta}}_n), \tag{3}$$

where $\boldsymbol{\Sigma}_n$ is the estimate of the limiting covariance matrix $\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$ of standardized frequencies $\mathbf{V_n}(\hat{\boldsymbol{\theta}}_n)$, $\mathbf{q}$ being the r-vector with its entries as $1/\sqrt{r}$. The statistic (3) can be presented as (Moore and Spruill [12, p. 610])

$$Y_n^2(\hat{\boldsymbol{\theta}}_n) = \mathbf{V}^{(n)T}(\hat{\boldsymbol{\theta}}_n)(\mathbf{I} - \mathbf{B}_n\mathbf{J}_n^{-1}\mathbf{B}_n^{T})^{-1}\mathbf{V}^{(n)}(\hat{\boldsymbol{\theta}}_n), \tag{4}$$

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

The Fisher's information matrix for one observation is $\mathbf{J} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1} \end{pmatrix}$, and $\mathbf{J}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}$, where $\mathbf{Q}$ is defined by formula (2). The matrix $\mathbf{B}$ can be blocked as $\begin{pmatrix} \mathbf{0} & \tilde{\mathbf{B}} \end{pmatrix}$, where

$$\tilde{\mathbf{B}} = \begin{pmatrix} d_1\frac{\sqrt{r}\sigma_{22}}{\delta} & -d_1\frac{\sqrt{r}\sigma_{12}}{\delta} & d_1\frac{\sqrt{r}\sigma_{11}}{\delta} \\ d_2\frac{\sqrt{r}\sigma_{22}}{\delta} & -d_2\frac{\sqrt{r}\sigma_{12}}{\delta} & d_2\frac{\sqrt{r}\sigma_{11}}{\delta} \\ \cdots & \cdots & \cdots \\ d_r\frac{\sqrt{r}\sigma_{22}}{\delta} & -d_r\frac{\sqrt{r}\sigma_{12}}{\delta} & d_r\frac{\sqrt{r}\sigma_{11}}{\delta} \end{pmatrix},$$

$\sigma_{ij},\ i,j=1,2,$ being elements of $\boldsymbol{\Sigma}$, and $\delta = \sigma_{11}\sigma_{22} - \sigma_{12}^2$. Evidently that $\mathbf{B}^T = \begin{pmatrix} \mathbf{0} \\ \tilde{\mathbf{B}}^T \end{pmatrix}$.

After simple matrix algebra one gets $\mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T = \tilde{\mathbf{B}}\mathbf{Q}\tilde{\mathbf{B}}^T$, where

$$\tilde{\mathbf{B}}\mathbf{Q}\tilde{\mathbf{B}}^T = \frac{r[4\sigma_{11}^2\sigma_{22}^2 - 3\sigma_{11}\sigma_{12}^2\sigma_{22} + \sigma_{12}^4]}{\delta^2}\begin{pmatrix} d_1^2 & d_1 d_2 \cdots d_1 d_r \\ d_1 d_2 & d_2^2 \cdots d_2 d_r \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ d_M d_1 & d_M d_2 \cdots d_r^2 \end{pmatrix}, \tag{5}$$

and the statistic (4) becomes

$$Y_n^2(\hat{\boldsymbol{\theta}}_n) = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)(\mathbf{I}_r - \tilde{\mathbf{B}}_n\mathbf{Q}_n\tilde{\mathbf{B}}_n^T)^{-1}\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n). \tag{6}$$

In our notations the limit covariance matrix $\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$ of the standardized frequencies $\mathbf{V_n}(\hat{\boldsymbol{\theta}}_n)$ can be written down as $\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \tilde{\mathbf{B}}\mathbf{Q}\tilde{\mathbf{B}}^T$. Because of (5) the matrix $\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$ depends on unknown parameters $\sigma_{11}, \sigma_{12}, \sigma_{22}$, and, in accordance with the Theorem 2 of Voinov [15], the limit distribution of the statistic $Y_n^2(\hat{\boldsymbol{\theta}}_n)$ cannot be chi-squared and will depend on unknown parameters. From (5) we see that the limit covariance matrix $\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \tilde{\mathbf{B}}\mathbf{Q}\tilde{\mathbf{B}}^T$ will not depend on unknown parameters if $\sigma_{12} = 0$, i.e., if $\boldsymbol{\Sigma}$ is a diagonal matrix.

2. New tests suggested

Let us construct a chi-squared Wald's type goodness of fit test for the two-dimensional normal distribution if $\boldsymbol{\Sigma}$ is a diagonal matrix. In the two-dimensional case for $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$ the Fisher's information matrix $\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1} \end{bmatrix}$ for one observation is

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

$$\tilde{\mathbf{B}} = \begin{pmatrix} d_1/\sigma_{11} & d_1/\sigma_{22} \\ d_2/\sigma_{11} & d_2/\sigma_{22} \\ \cdots\cdots\cdots\cdots\cdots & \\ d_M/\sigma_{11} & d_M/\sigma_{22} \end{pmatrix}, \text{ and } \tilde{\mathbf{B}}^T\tilde{\mathbf{B}} = r\sum d_i^2 \begin{pmatrix} 1/\sigma_{11}^2 & 1/(\sigma_{11}\sigma_{22}) \\ 1/(\sigma_{11}\sigma_{22}) & 1/\sigma_{22}^2 \end{pmatrix}. \tag{7}$$

The matrix $\mathbf{D} = \mathbf{Q}^{-1} - \tilde{\mathbf{B}}^T\tilde{\mathbf{B}}$ is:

$$\mathbf{D} = \begin{pmatrix} \dfrac{1 - 2r\sum d_i^2}{2\sigma_{11}^2} & -\dfrac{r\sum d_i^2}{\sigma_{11}\sigma_{22}} \\ -\dfrac{r\sum d_i^2}{\sigma_{11}\sigma_{22}} & \dfrac{1 - 2r\sum d_i^2}{2\sigma_{22}^2} \end{pmatrix}.$$ It is easy to show that

$$\mathbf{D}^{-1} = \frac{4}{1-4r\sum d_i^2}\begin{pmatrix} \frac{(1-2r\sum d_i^2)\sigma_{11}^2}{2} & \mathrm{r}\sum d_i^2\sigma_{11}\sigma_{22} \\ \mathrm{r}\sum d_i^2\sigma_{11}\sigma_{22} & \frac{(1-2r\sum d_i^2)\sigma_{22}^2}{2} \end{pmatrix}. \tag{8}$$

Now

$$\tilde{\mathbf{B}}\mathbf{D}^{-1} = \frac{2\sqrt{r}}{1-4r\sum d_i^2}\begin{pmatrix} d_1\sigma_{11} & d_1\sigma_{22} \\ d_2\sigma_{11} & d_2\sigma_{22} \\ \cdots\cdots & \cdots\cdots \\ d_r\sigma_{11} & d_r\sigma_{22} \end{pmatrix}, \text{ hence,}$$

$$\tilde{\mathbf{B}}\mathbf{D}^{-1}\tilde{\mathbf{B}}^T = \frac{4r}{1-4r\sum d_i^2}\begin{pmatrix} d_1^2 & d_1d_2 & \cdots & d_1d_r \\ \cdots & \cdots & \cdots & \cdots \\ d_rd_1 & d_rd_2 & \cdots & d_r^2 \end{pmatrix}.$$

Denoting $V_i = \dfrac{N_i^{(n)} - n/r}{\sqrt{n/r}}$, $i = 1,...,r$, from the above it follows that $\mathbf{V}^T(\hat{\boldsymbol{\theta}}_n)\tilde{\mathbf{B}}\mathbf{D}^{-1}\tilde{\mathbf{B}}^T\mathbf{V}(\hat{\boldsymbol{\theta}}_n) = 4r\left(\sum V_i d_i\right)^2 / \left(1-4r\sum d_i^2\right)$, and the closed form of the NRR statistic (4) becomes

$$Y_n^2(\hat{\boldsymbol{\theta}}_n) = \sum V_i^2 + \frac{4r\left(\sum V_i d_i\right)^2}{1-4r\sum d_i^2}. \tag{9}$$

In this case the limiting covariance matrix $\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q}\mathbf{q}^T - \mathbf{B}\mathbf{J}^{-1}\mathbf{B}^T$ of the standardized frequencies does not depend on parameters, and the NRR statistic follows in the limit the chi-squared distribution with *r* - 1 d.f.

McCulloch [18] proposed a very useful decomposition of $Y_n^2(\hat{\boldsymbol{\theta}}_n)$, namely $Y_n^2(\hat{\boldsymbol{\theta}}_n) = U_n^2(\hat{\boldsymbol{\theta}}_n) + S_n^2(\hat{\boldsymbol{\theta}}_n),$ where Dzhaparidze-Nikulin (DN) statistic $U_n^2(\hat{\boldsymbol{\theta}}_n)$ (Dzhaparidze and Nikulin [19]) is

$$U_n^2(\hat{\boldsymbol{\theta}}_n) = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)[\mathbf{I} - \mathbf{B}_n(\mathbf{B}_n^T\mathbf{B}_n)^{-1}\mathbf{B}_n^T]\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n), \tag{10}$$

and

$$S_n^2(\hat{\boldsymbol{\theta}}_n) = Y_n^2(\hat{\boldsymbol{\theta}}_n) - U_n^2(\hat{\boldsymbol{\theta}}_n) = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)\mathbf{B}_n[(\mathbf{J}_n - \mathbf{B}_n^T\mathbf{B}_n)^{-1} + (\mathbf{B}_n^T\mathbf{B}_n)^{-1}]\mathbf{B}_n^T\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n). \tag{11}$$

McCulloch [18, Theorem 4.2] showed that if the rank of $\mathbf{B}$ is s, then $U_n^2(\hat{\boldsymbol{\theta}}_n)$ and $S_n^2(\hat{\boldsymbol{\theta}}_n)$ are asymptotically independent and distributed in the limit as χ^2_{r-s-1} and χ^2_s respectively.

Since the first two columns of the matrix $\mathbf{B}_n$ in our case 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 [13, p.132]), we may replace $\mathbf{A}_n^{-1} = (\mathbf{B}_n^T\mathbf{B}_n)^{-1}$ by $\mathbf{A}_n^- = (\mathbf{B}_n^T\mathbf{B}_n)^-$, where $\mathbf{A}_n^-$ is any generalized matrix inverse of $\mathbf{A}$. So, for testing binormality with cells $E_{in}(\hat{\boldsymbol{\theta}}_n)$ we may use: the NRR statistic defined by (9), the DN statistic

$$U_n^2(\hat{\boldsymbol{\theta}}_n) = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)[\mathbf{I} - \mathbf{B}_n(\mathbf{B}_n^T\mathbf{B}_n)^- \mathbf{B}_n^T]\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n), \tag{12}$$

where $\mathbf{I}$ is the $r \times r$ identity matrix, and

$$S_n^2(\hat{\boldsymbol{\theta}}_n) = \mathbf{V}_n^T(\hat{\boldsymbol{\theta}}_n)\mathbf{B}_n[(\mathbf{J}_n - \mathbf{B}_n^T\mathbf{B}_n)^{-1} + (\mathbf{B}_n^T\mathbf{B}_n)^-]\mathbf{B}_n^T\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n) \tag{13}$$

that will have asymptotically χ_{r-1}^2, χ_{r-2}^2, and χ_1^2 distributions correspondingly.

Using formula (7) for $\tilde{\mathbf{B}}^T\tilde{\mathbf{B}}$, it is easily verified that $(\tilde{\mathbf{B}}^T\tilde{\mathbf{B}})^- = \dfrac{1}{r\sum d_i^2}\begin{pmatrix} \sigma_{11}^2 & 0 \\ 0 & 0 \end{pmatrix}$ is a generalized matrix inverse for $\tilde{\mathbf{B}}^T\tilde{\mathbf{B}}$. From (12) and (13) after simple matrix algebra one gets the following closed form expressions for $U_n^2(\hat{\boldsymbol{\theta}}_n)$ and $S_n^2(\hat{\boldsymbol{\theta}}_n)$:

$$U_n^2(\hat{\boldsymbol{\theta}}_n) = \sum V_i^2 - \frac{\left(\sum V_i d_i\right)^2}{\sum d_i^2}, \tag{14}$$

and

$$S_n^2(\hat{\boldsymbol{\theta}}_n) = \frac{\left(\sum V_i d_i\right)^2}{\left(1 - 4r\sum d_i^2\right)\sum d_i^2}. \tag{15}$$

The above approach suggests the following evident procedure for developing tests for binormality for any nonsingular covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$. Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be a set of two-dimensional random vectors, and **S** is a corresponding sample covariance matrix. Let $\boldsymbol{e}_1$, and $\boldsymbol{e}_2$ be orthogonal normalized eigen-vectors of a sample covariance matrix **S,** then the Karhunen-Loév transform $\mathbf{Y}_i = \left(\boldsymbol{e}_1\ \boldsymbol{e}_2\right)^T \mathbf{X}_i,\ i = 1, ..., n$, will give a set $\mathbf{Y}_1, ..., \mathbf{Y}_n$ of two-dimensional random vectors with the diagonal sample covariance matrix. Hence, tests in (9), (14), and (15) of this Section with frequencies $N_i^{(n)},\ i = 1, ..., n,$ defined by the number of $\mathbf{Y}_1, ..., \mathbf{Y}_n$ that 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.

3. A simulation study of proposed tests

A simulation study was conducted to compare the power of the Doornik and Hansen (DH) [9], Henze and Zirkler (HZ) [20], Royston (R92) [21], Nikulin-Rao-Robson (NRR) test (9), McCulloch's (McCull) test (15), Anderson-Darling (AD), and Cramer von Mises (CM) tests . To simulate DH, HZ, and R92 tests the *R*-scripts given us by Matias Saliban-Barrera were used. AD and CM tests were simulated using corresponding formulas of Henze [2, p.483].

We do not consider the Dzhaparidze-Nikulin (DN) and Chernoff-Lehmann (ChLeh) tests of Moore and Stubblebine [16, p. 718], because of their low power (see Figure 1).

To construct graphs in Fig. 1 and Tables 2-5 we used simulated critical values of level $\alpha = 0.05$ that are given in Table 1.

Table 1: Simulated critical values for the right-tailed rejection region of size $\alpha = 0.05$. Simulated relative standard errors were defined as errors of the mean of 20 runs by N = 1,000 each. For all results below that error is not more than 1%.

r	n	McCull	NRR	ChLeh	
3	250	3.882	6.018		
5	250	3.830	9.490	8.033	
10	250	3.730	16.890	15.635	
15	250	3.650	23.640	22.344	
20	250	3.620	30.100	28.924	
25	250	3.600	36.400	35.190	
30	250	3.560	42.550	41.350	
40	250	3.530	54.550	53.140	
3	100	3.868	5.959		
5	100	3.780	9.400		
10	100	3.550	16.750		
15	100	3.430	23.510		
20	100	3.330	30.020		
3	50	3.887	6.133		
5	50	3.690	9.350		
10	50	3.300	16.640		
3	25	4.000	6.100		
5	25	3.570	9.240		

n	HZ	DH	R92	AD	CM
250	1.058	9.594	5.948	1.314	0.221
100	0.957	9.560	5.868	1.293	0.221
50	0.859	9.513	6.018	1.323	0.224
25	0.737	9.411	4.968	1.298	0.223

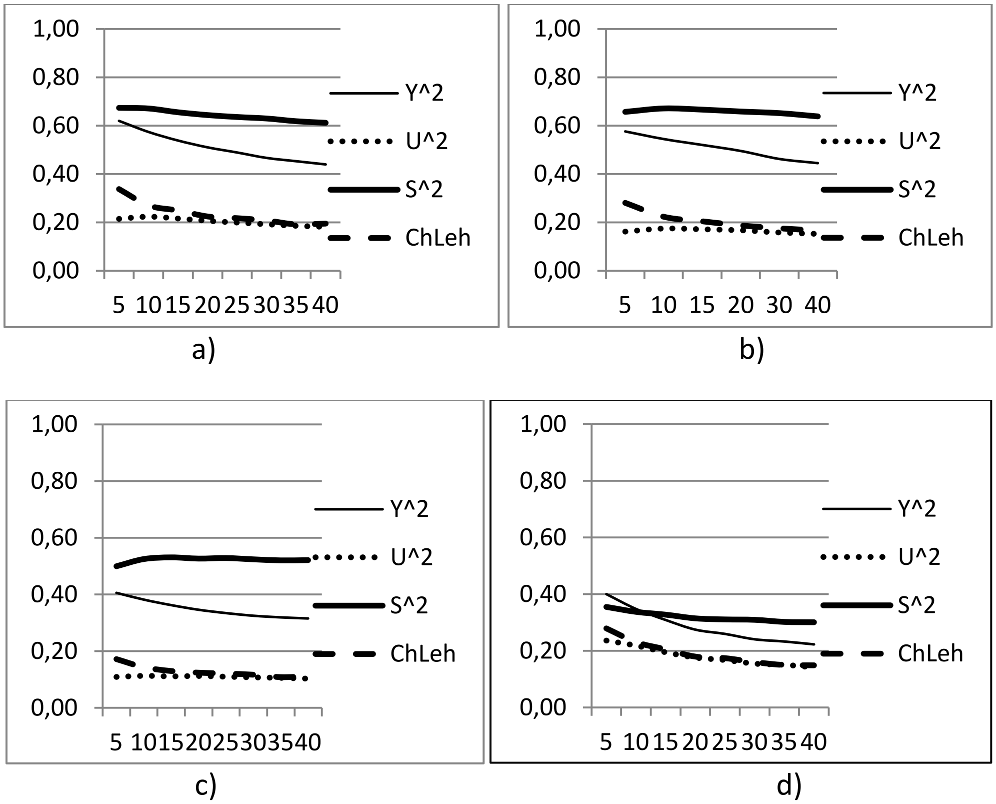

Figure 1: a) Power w.r.t. two-dimensional logistic with independent standard logistic components (n = 250). b) Power w.r.t. two-dimensional Student t(10 d.f.) with independent components (n = 250). c) Power w.r.t. two-dimensional Student t(10 d.f.) with dependent components for n = 250 (formula (5) of Farrell et al. [22] was used). d)) Power w.r.t. two-dimensional Khinchine distribution for n = 250.

From Fig. 1 one sees that the power of McCull test almost does not depend on the number of equiprobable grouping intervals and that power of DN ($\mathrm{U}^\wedge 2$) and ChLeh tests is noticeably less than that of NRR ($\mathrm{Y}^\wedge 2$) and McCull ($\mathrm{S}^\wedge 2$) tests. The same lack of power for the DN and Chernoff-Lehmann tests was observed in the univariate case (see Voinov et al. [23]). Because of this in the sequel we shall not consider the ChLeh and DN tests. Because of weak dependence of the power of McCull test on the number of grouping intervals, the powers of NRR and McCull tests in Tables 2-5 were simulated for r = 5 that seems to be optimal. We considered 9 alternatives. Namely: Pearson Type II (m = 10, and m = 0 (uniform)), two-dimensional Student t (10 and 5 d.f.), two-dimensional Khinchine distribution, and 4 different mixtures of normal distributions.

Table 2: Simulated powers of tests with respect to alternatives considered for n = 25, N = 20,000 replications. In the sequel $\boldsymbol{\mu} = \mathbf{3}$ denotes the mean vector $(3,3)^T$, $\mathbf{\Sigma} = \boldsymbol{B}$ denotes 1 on diagonal and 0.9 off diagonal, **I** is the identity 2×2 matrix.

Alternative	HZ	AD	CM	R92	McCull	NRR	DH
Pearson Type II (m=10)	0.045	0.059	0.066	0.067	0.056	0.055	0.028
Pearson Type II (m=0)	0.223	0.637	0.592	0.223	0.418	0.143	0.038
Multivariate t (10 d.f.)	0.122	0.325	0.313	0.188	0.071	0.090	0.154
Multivariate t (5 d.f.)	0.203	0.520	0.520	0.303	0.168	0.206	0.326
Khinchine	0.111	0.087	0.068	0.082	0.058	0.065	0.038
$0.5N(\mathbf{0},\mathbf{I})+0.5N(\mathbf{3},\mathbf{I})$	0.311	0.274	0.294	0.309	0.173	0.055	0.027
$0.79N(\mathbf{0},\mathbf{I})+0.21N(\mathbf{3},\mathbf{I})$	0.570	0.080	0.075	0.556	0.060	0.057	0.076
$0.5N(\mathbf{0},\boldsymbol{B})+0.5N(\mathbf{0},\mathbf{I})$	0.189	0.110	0.089	0.087	0.066	0.072	0.068
$0.9N(\mathbf{0},\boldsymbol{B})+0.1N(\mathbf{0},\mathbf{I})$	0.221	0.127	0.117	0.065	0.100	0.147	0.128
Average	**0.22**	**0.25**	**0.24**	**0.21**	**0.13**	**0.10**	**0.10**

Table 3:. Simulated powers of tests with respect to specified alternatives for n = 50, N = 20,000

Alternative	HZ	AD	CM	R92	McCull	NRR	DH
Pearson Type II (m=10)	0.029	0.065	0.073	0.036	0.06	0.067	0.022
Pearson Type II (m=0)	0.581	0.952	0.960	0.490	0.824	0.407	0.192
Multivariate t (10 d.f.)	0.160	0.317	0.312	0.211	0.135	0.173	0.263
Multivariate t (5 d.f.)	0.457	0.507	0.496	0.503	0.417	0.478	0.568
Khinchine	0.153	0.149	0.128	0.049	0.092	0.100	0.042
$0.5N(\mathbf{0},\mathbf{I})+0.5N(\mathbf{3},\mathbf{I})$	0.805	0.585	0.587	0.535	0.369	0.072	0.020
$0.79N(\mathbf{0},\mathbf{I})+0.21N(\mathbf{3},\mathbf{I})$	0.929	0.094	0.089	0.813	0.068	0.069	0.116
$0.5N(\mathbf{0},\boldsymbol{B})+0.5N(\mathbf{0},\mathbf{I})$	0.335	0.219	0.195	0.056	0.131	0.134	0.090
$0.9N(\mathbf{0},\boldsymbol{B})+0.1N(\mathbf{0},\mathbf{I})$	0.326	0.257	0.251	0.044	0.243	0.327	0.266
Average	**0.42**	**0.35**	**0.34**	**0.30**	**0.26**	**0.20**	**0.18**

Table 4: Simulated powers of tests with respect to specified alternatives for n = 100, N = 20,000

Alternative	HZ	AD	CM	R92	McCull	NRR	DH
Pearson Type II (m=10)	0.058	0.090	0.100	0.036	0.073	0.087	0.016
Pearson Type II (m=0)	0.977	1.000	1.000	0.979	0.997	0.843	0.837
Multivariate t (10 d.f.)	0.215	0.327	0.319	0.331	0.254	0.320	0.410

Multivariate t (5 d.f.)	0.690	0.819	0.815	0.755	0.734	0.772	0.816
Khinchine	0.262	0.291	0.256	0.052	0.167	0.158	0.049
*0.5N(**0**,**I**)+0.5N(**3**,**I**)*	0.998	0.924	0.911	0.932	0.728	0.109	0.025
*0.79N(**0**,**I**)+0.21N(**3**,**I**)*	0.999	0.119	0.113	0.992	0.083	0.069	0.243
*0.5N(**0**,**B**)+0.5N(**0**,**I**)*	0.634	0.448	0.418	0.055	0.269	0.246	0.106
*0.9N(**0**,**B**)+0.1N(**0**,**I**)*	0.495	0.494	0.481	0.047	0.465	0.577	0.466
Average	**0.59**	**0.50**	**0.49**	**0.46**	**0.42**	**0.35**	**0.33**

Table 5: Simulated powers of tests with respect to specified alternatives for n = 250, N = 20,000

Alternative	HZ	AD	CM	R92	McCull	NRR	DH
Pearson Type II (m=10)	0.072	0.151	0.164	0.050	0.118	0.164	0.030
Pearson Type II (m=0)	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Multivariate t (10 d.f.)	0.408	0.671	0.676	0.597	0.580	0.660	0.701
Multivariate t (5 d.f.)	0.962	0.995	0.995	0.977	0.986	0.987	0.988
Khinchine	0.614	0.61	0.580	0.050	0.398	0.355	0.048
*0.5N(**0**,**I**)+0.5N(**3**,**I**)*	1.000	1.000	1.000	1.000	0.995	0.219	0.075
*0.79N(**0**,**I**)+0.21N(**3**,**I**)*	1.000	0.154	0.157	1.000	0.105	0.068	0.678
*0.5N(**0**,**B**)+0.5N(**0**,**I**)*	0.980	0.848	0.833	0.055	0.670	0.562	0.148
*0.9N(**0**,**B**)+0.1N(**0**,**I**)*	0.805	0.856	0.854	0.045	0.844	0.914	0.789
Average	**0.76**	**0.70**	**0.70**	**0.53**	**0.63**	**0.55**	**0.50**

It is not easy to compare different tests w.r.t. different alternatives. The simplest way is to compare average (w.r.t. 9 alternatives selected) power of all 7 tests under consideration. The results of our simulations are presented in Figure 2.

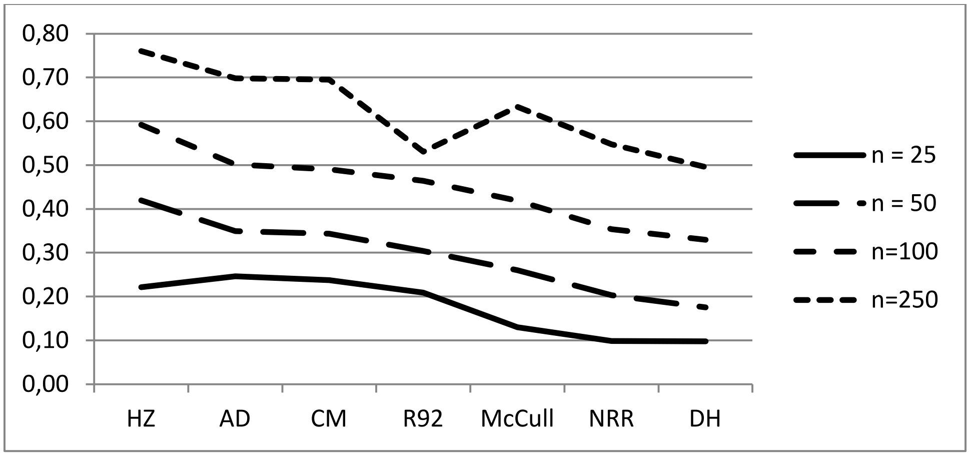

Figure 2: Average powers of HZ, AD, CM, R92, McCull, NRR, and DH tests as functions of the sample size.

From this Figure one sees that, evidently, HZ, AD, and CM tests seem to be the best. At the same time it can be concluded that all tests have a right to be used in practice. We see also that much depends on an alternative. From Tables 2 and 3, e.g., we see that AD, CM, and McCull tests are the best w.r.t. Pearson Type II (m = 0) alternative, and that HZ test is preferable w.r.t. Khinchine alternative and mixtures of normal distributions for small samples of size n = 25 and n = 50. From Table 5 we see that HZ, R92, AD, CM, and McCull tests are perfect w.r.t. the mixture of normal distributions $0.5N(\mathbf{0},\mathbf{I})+0.5N(\mathbf{3},\mathbf{I})$ if n = 250.

Remark

It is of interest to compare simulated powers of AD, CM, HZ, and McCull tests with those for Z_2, C_2, R_2, $b_{1,2}$, and W_2 tests of Sürücü [8], Table 1, p.1325 for n = 50 (Table 6).

Table6: Simulated powers of tests with respect to specified alternatives for n = 50, N = 20,000

Alternative	Z_2	C_2	R_2	$b_{1,2}$	W_2	AD	CM	HZ	McCull
St. t (2 d.f.)	0.74	0.97	0.99	0.94	0.96	0.98	0.98	0.97	0.96
St. t (6 d.f.)	0.30	0.45	0.58	0.53	0.49	0.39	0.38	0.34	0.31
Uniform	0.07	0.97	0.41	0	0.07	0.95	0.96	0.58	0.83
Average	**0.37**	**0.80**	**0.66**	**0.49**	**0.51**	**0.77**	**0.77**	**0.63**	**0.70**

From this Table we see that as Sürücü [8, p.1324] wrote "as an omnibus goodness-of-fit test the C_2 test is clearly most powerful overall". This is clear indeed, but we have to note that AD, CM, and McCull tests are quite comparable with the C_2 test of Sürücü [8].

4. Conclusion and recommendations

To conclude the research we have to note that no one of seven tests considered can be a panacea when testing for two-dimensional normality. Any of them can be used in practice. But, before selecting a proper test, it is highly desirable to have some imagination about possible alternative.

Acknowledgements

The author is grateful to Matias Saliban-Barrera for giving an access to R-codes used in Farrell et al. [22].

References

[1] Voinov, V.. Nikulin, M.. Balakrishnan, N. (2013). *Chi-Squared Goodness of Fit Tests with Applications*. Academic Press. Elsevier: Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, etc.
[2] Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statistical papers*. Vol.**43,** pp. 467-506.
[3] 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.
[4] Szkutnik, Z. (1988). Most powerful invariant tests for binormality. *The Annals of Statistics*. **16,** pp. 292-301.
[5] Versluis, C. (1996). Comparison of tests for bivariate normality with unknown parameters by transformation to a univariate statistic. *Communications in Statistics - Theory and Methods*. **25,** pp. 647-665.
[6] Bogdan, M. (1999). Data driven smooth tests for bivariate normality. *J. of Multivariate Analysis*. **68,** pp. 26-53.
[7] Székely, G.J. and Rizzo, M.L. (2005). A new test for multivariate normality. *J. of Multivariate Analysis*. **93,** 58-80.
[8] Sürücü, B. (2006). Goodness-of-fit tests for multivariate distributions. *Communications in Statistics - Theory and Methods*. **35,** pp. 1319-1331.
[9] Doornik, J.A.. Hansen, H. (2008). Oxford *Bulletin in Economics and Statistics*. Supplement 0305-9049, pp. 927-939.
[10] Villasenor Alva, J.A., Gonzáles Estrada, E. (2009). A generalization of Shapiro-Wilk's test for multivariate normality. *Communications in Statistics - Theory and Methods*. **38,** pp. 1870-1883.
[11] Hanusz, Z., Tarasińska, J. (2012). New tests for multivariate normality based on Small's and Srivastava's graphical methods. *J. of Statistical Computation and Simulation*. **82**, 1743-1752.
[12] Moore, D.S.. Spruill, M.C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *The Annals of Statistics*. **3,** 599-616.
[13] Moore, D.S. (1977). Generalized inverses. Wald's method. and the construction of chi-squared tests of fit. *J. of the American Statistical Association*. **72,** 131-137.
[14] Nikulin, M.S. (1973). Chi-square test for continuous distributions. *Theory of Probability and its Applications*. **18,** 638-639.
[15] Voinov, V. (2013). A note on Wald's type chi-squared tests of fit. *Communications in Statistics - Theory and Methods*, http://www.tandfonline.com/doi/full/ 10.1080/03610926.2013.768668.
[16] Moore, D.S.. Stubblebine, J.B. (1981). Chi-square tests for multivariate normality with application to common stock prices. *Communications in Statistics - Theory and Methods*. **A10(8),** 713-738.

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

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

[19] Dzhaparidze, K.O.. Nikulin, M.S. (1974). On a modification of the standard statistic of Pearson. *Theory of Probability and Applications*. **19,** 851-853.

[20] Henze, N. and Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications is Statistics- Theory and Methods*. **19,** 3595-3617.

[21] Royston, J.P. (1992). Approximating the Shapiro-Wilk *W*-Test for non-normality. *Statistics and Computing.* **2,** 117-119.

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

[23] Voinov, V., Pya, N., Alloyarova R. (2009). A comparative study of some modified chi-squared tests. *Communications in Statistics - Simulation and Computation*. **38**, 355-367.