

Comparison of some multivariate normality tests: A simulation study



Ozlem Alpu *, Demet Yuksek

Department of Statistics, Eskisehir Osmangazi University, Eskisehir 26480, Turkey

ARTICLE INFO

Article history:

Received 5 September 2016

Received in revised form

10 November 2016

Accepted 7 December 2016

Keywords:

Tests of multivariate normality

Power

Type I error

Multivariate normal distribution
generalized Shapiro-Wilk test

ABSTRACT

Many classical multivariate statistical methods are mostly based on the assumption of multivariate normality. Departures from normality, called non-normality, render those statistical methods inaccurate, so it is important to know if datasets are normal or non-normal. Especially in medical and life sciences, most statistical tests required the assumption of multivariate normality have been extensively used. In this study, after summarizing the properties of several most widely used multivariate normality tests, we aim to compare the power and type I error rates of these tests, which have been developed in recent years by many researchers. So, the reader will elucidate the differences and the similarities/superiorities and weaknesses of the tests in order to make the appropriate choice in their practical applications. For this purpose we carried a Monte Carlo simulation study with nominal α level, small, medium and large sample size, different dimension and multivariate distributions which includes different skewness and kurtosis. In conclusion, the results obtained from the comparative study are given.

© 2016 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Many statistical methods for continuous variables assume an underlying multivariate normal distribution. This methodologies have been widely applied in researches. Thus, several formal tests were developed for assessing the multivariate normality of a set of random variables. Let $\mathbf{X}=(X_1, X_2, \dots, X_p)$ be a vector of p random variables distributed as a multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, that is null hypothesis to be tested can be written as $H_0: \mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

A considerable number of ways for assessing multivariate normality can be found in the literature. Some recent references are Thulin (2014), Hanusz and Tarasińska (2012), Delmail et al. (2011), Tenreiro (2011), Cardoso De Oliveira and Ferreira (2010), Okamoto and Seo (2010), Villasenor Alva and Estrada (2009). It can also be seen reviews by Liang et al. (2009), Koizumi et al. (2009), Hanusz and Tarasińska (2008), Doornik and Hansen (2008), Farrell et al. (2007), Arcones (2007), Sürücü (2006), Holgersson (2006), Székely and Rizzo (2005), Pudielko (2005), Mecklin and

Mundfrom (2005), Dufour et al. (2003), Burdinski (2000), Henze and Wagner (1997), Henze (1997), Looney (1995), Romeu and Ozturk (1993), and Henze and Zirkler (1990).

Various multivariate normality tests, which have been included in the literature in decades, will be discussed in the second part of this study, while in the third section the Monte Carlo simulation study results, which compares the type I error rates and their powers of various multivariate normality tests in respect of the different numbers of variables, the sizes of samples and significance levels (p, n, α), will be revealed. Some results and discussions obtained from the comparative study in section 3 are given in section 4. Finally, conclusions are given in section 5.

2. Methods

Various multivariate normality tests which have been included in the literature in decades will be summarized in this section of the study.

2.1. Villasenor-Alva and Gonzalez-Estrada's generalized Shapiro-Wilk (GSW) test

Villasenor Alva and Estrada (2009) proposed a goodness of fit test for multivariate normality which is based on Shapiro and Wilk (1965) statistic for univariate normality and on an empirical standardization of the observations.

* Corresponding Author.

Email Address: aalpu@ogu.edu.tr (O. Alpu)

<https://doi.org/10.21833/ijaas.2016.12.011>

2313-626X/© 2016 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Assume $\mathbf{X}_1, \dots, \mathbf{X}_n$, are independent identically distributed random vector in \mathbf{R}^p , ($p \geq 1$). Let $N^p(\mu, \Sigma)$ denote the p -variate normal density with mean μ and covariance matrix Σ . Let $\mathbf{0}$ be the null vector of order p and let \mathbf{I} be the identity matrix of order $p \times p$. To test the null hypothesis H_0 : $\mathbf{X}_1, \dots, \mathbf{X}_n$ is a sample from $N^p(\mu, \Sigma)$, where μ and Σ are unknown, they proposed the test statistic as

$$W^* = \frac{1}{p} \sum_{i=1}^p W_{Z_i} \quad (1)$$

where W_{Z_i} is Shapiro-Wilk's statistic evaluated on the i th coordinate of the transformed observations Z_{i1}, \dots, Z_{in} , $i=1, \dots, p$. The test based on W^* rejects H_0 at a test size α if $W^* < c_{\alpha; n, p}$, where $c_{\alpha; n, p}$ satisfies the equation $\alpha = P\{W^* < c_{\alpha; n, p} | H_0 \text{ holds}\}$.

2.2. Kankainen-Taskinen-Oja's skewness (b1, new) test

Kankainen et al. (2007) obtained generalizations of classical Mardia's measures of skewness and kurtosis by using special choices of location and scatter estimators.

If the multivariate measure of skewness is constructed so that T_1 and C are the sample mean vector and sample covariance matrix and T_2 is the one-step M-estimator that uses T_1 and C as initial estimators and weight function $v_1(r) = r^2$, then the resulting skewness measure is easily seen to be

$$b_{1, \text{new}} = \text{ave}_i \{r_i^2\}^{-2} \text{ave}_{i,j} \{r_i^2 r_j^2 r_{ij}\}. \quad (2)$$

Note that this measure is equivalent to that introduced in Móri et al. (1993). The limiting distribution of $b_{1, \text{new}}$ is that of $\eta_1 U_1$,

$$b_{1, \text{new}} \sim \frac{1}{n} \left(\frac{2(p+2)}{p^2} \chi_p^2 \right) \quad (3)$$

where $U_1 \sim \chi_p^2$ and $\eta_1 = 2(p+2)/p^2$. See also Henze (1997).

2.3. Kankainen - Taskinen - Oja' kurtosis (b2,new) test

If the new kurtosis measure is such that it uses the sample covariance matrix as C_1 and one-step M-estimator based on the sample mean vector and sample covariance matrix with weight function $v_2(r) = (p+2)^{-1} r^2$ as C_2 , then they have obtained the generalization of the classical Mardia's measure of kurtosis, that is,

$$b_{2, \text{new}} = (p+2)^{-2} \text{ave}_{i,j} \{r_i^2 r_j^2 r_{ij}^2\} - 2(p+2)^{-1} \text{ave}_i \{r_i^4\} + p \quad (4)$$

and the limiting distribution of $b_{2, \text{new}}$ is that of $\eta_2 W_1 + \eta_3 W_2$,

$$b_{2, \text{new}} \sim \frac{1}{n} \left(\frac{4(p+4)}{(p+2)^2} \chi_{p(p+2)/2-1}^2 + \frac{8}{p+2} \chi_1^2 \right) \quad (5)$$

where $W_1 \sim \chi_{p(p+1)/2-1}^2$ and $W_2 \sim \chi_1^2$ (Kankainen et al., 2007).

2.4. Székely - Rizzo's ε -statistic (Energy Statistic) for testing multivariate normality

The ε -test of multivariate normality was proposed and implemented by Székely and Rizzo (2005). The test statistic for d -variate normality is given by

$$\varepsilon_{n,p} = n \left(\frac{2}{n} \sum_{j=1}^n E \|x_j - X\| - E \|X - X'\| - \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \|x_j - x_k\| \right) \quad (6)$$

where x_1, \dots, x_n is the standardized sample, X, X' are iid standard d -variate normal, and $\|\cdot\|$ denotes Euclidean norm. This new nonparametric test for equality of two or more multivariate distributions is based on Euclidean distance between sample units. Their proposed test was universally consistent against all fixed alternatives (not necessarily continuous) with finite second moments. The test rejects H_0 for large values of $\varepsilon_{n,p}$ (Székely and Rizzo, 2005).

2.5. Henze-Zirkler (HZ) test

Henze and Zirkler (1990) test which is another multivariate normal known for good power, is based on the empirical characteristic function. As an appealing property of this test have been stated that it was a consistent test. The test statistic is based on a nonnegative functional $D(\cdot, \cdot)$ that using characteristic functions to measure the distance between the hypothesized function and the empirical function. In order to consistent test statistic, the functional $D(\cdot, \cdot)$ must equal zero, providing that the data is multivariate normal distribution. The non-negative functional is given as

$$D_\beta(P, Q) = \int_{\mathbb{R}^p} |\hat{P}(t) - \hat{Q}(t)|^2 \varphi_\beta(t) dt \quad (7)$$

where $\hat{P}(t)$ is the characteristic function of the proposed distribution and $\hat{Q}(t)$ is the empirical characteristic function. The functional consists of a weighting function $\varphi_\beta(t)$, where $\beta = \frac{1}{\sqrt{2}} \left\{ \frac{n(2p+1)}{4} \right\}^{p+4}$ is a smoothing parameter. The test statistic was proposed as

$$T_\beta(p) = \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \exp \left(-\frac{\beta^2}{2} D_{ik}^2 \right) - 2(1 + \beta^2)^{\frac{-p}{2}} \frac{1}{n} \sum_{i=1}^n \exp \left(-\frac{\beta^2}{2(1+\beta^2)} D_{ii}^2 \right) + (1 + 2\beta^2)^{\frac{-p}{2}} \quad (8)$$

where D_{ik}^2 is the squared Mahalanobis distance between two given observations and D_{ii}^2 is the squared Mahalanobis distance from a given

observation to the centroid. The test statistic is approximately lognormally distributed and this distribution is used to find the critical values of the test. The test rejects H_0 for large values of $T_\beta(p)$.

2.6. Royston's H multivariate normality test (1992)

Let X_1, X_2, \dots, X_n be a multivariate random sample of size n , where $X_j \in R^p$ and if $X_{(1)k}, X_{(2)k}, \dots, X_{(n)k}$ denotes an ordered univariate sample for the k th variate, the Shapiro-Wilk's test statistic is

$$W_k = \frac{[\sum_{j=1}^n \tilde{a}_j X_{(j)k}]^2}{\sum_{j=1}^n (X_{jk} - \bar{X}_{.k})^2} \quad (9)$$

where $\bar{X}_{.k} = \sum_{j=1}^n X_{jk}/n$ is the sample mean, and \tilde{a}_j the estimator of the normalized best linear unbiased coefficients for $j = 1, 2, \dots, n$. Royston (1992) suggested a normalizing transformation of W_k , obtaining a standard normal score Z_k , for $k = 1, 2, \dots, p$. Next, the statistic is calculated,

$$\kappa_k = \left\{ \Phi^{-1} \left[\frac{1}{2} \Phi(-Z_k) \right] \right\}^2, \quad \text{for } k = 1, 2, \dots, p \quad (10)$$

where $\Phi(x) = 1/\sqrt{2\pi} \int_{-\infty}^x \exp -t^2/2 dt$ denotes the standard normal cumulative distribution function. Finally, the test statistic proposed by Royston (1992) was defined by

$$H = v \sum_{k=1}^p \kappa_k \quad (11)$$

that is approximately χ_v^2 , where v is referred to as the equivalent degrees of freedom, since κ_k are not independent. See also an estimate for v based on the method of moments (Royston, 1983; Cardoso De Oliveira and Ferreira, 2010).

3. Simulation study

A simulation study for the comparison of the tests introduced in the previous section is discussed in this section. This simulation study is comprised of two parts. In the first part of the study, the Generalized Shapiro-Wilk test (GSW), which has been proposed by Villasenor-Alva and Gonzalez-Estrada, the Kankaiken-Taskinen-Oja's skewness test ($b_{1,new}$), the Kankaiken-Taskinen-Oja's kurtosis test ($b_{2,new}$), the Energy test, the Henze-Zirkler (HZ) test and the Royston (1992) test statistics are compared in terms of type I errors. The second part is comprised of the power comparisons of these tests. In the simulation study, the iteration number has been taken to be 10000, and an R code programme has been prepared for each test procedure. The results which have been obtained have been displayed using tables and figures. For the comparison of type I errors and the

comparisons in respect of the power of the tests, the number of variables has been taken to be $p = 2, 3, 5, 10$, and the sizes of the samples have been taken to be $n=20, 50, 100$ and 200. Comparisons in terms of type I errors and power of tests have been performed for nominal $\alpha = 0.05$.

In the Monte Carlo simulation study the selection of the alternatives to be used against a multivariate normal distribution is quite important. The purpose here is to define the multivariate normality tests which possess the best performance in many situations. The first distribution which needs to be paid attention to is the multivariate normal distribution. There are two reasons for the comparison of tests according to normal distribution. The first is to check to be sure that the algorithms used in the calculation of the test statistics has been programmed correctly, while the second is used to be sure that the tests reject normality only at approximate nominal α levels (thus making a type I error).

The study will consider several normal mixture models in order to perform the simulation of the sample units derived from two different populations as a second distribution. The general form of the normal mixture model is $\tilde{p}N_p(\mu_1, \Sigma_1) + (1 - \tilde{p})N_p(\mu_2, \Sigma_2)$. The \tilde{p} here is the contamination parameter, and specifies the rate of the sample obtained from one population. Three contamination levels will be reviewed. The first level is $\tilde{p} = 0.9$ (90%, 10%), and contains mild contamination. Therefore, it is skewed and leptokurtic. The second levels is $\tilde{p} = 0.788675$ (78.8675%, 21.1325%), and contains moderate contamination. It is skewed and mesokurtic. The third level is $\tilde{p} = 0.5$ (50%, 50%), and contains severe contamination, and is symmetrical and platykurtic. The second normal mixture model has quite an interesting characteristic. The kurtosis of this model is normal, but the model has a distribution which is not normal. Therefore, it is expected that the kurtosis tests of our data set will be of low power. The total number of normal mixture models considered in the study is 6. In these models, the symbols are defined as follows:

μ_1 : Mean vector whose components are zero;

μ_2 : Mean vector whose components are one;

I : Unit matrix with a dimension of $p \times p$;

Σ_1 : Quadratic matrix with a dimension of $p \times p$, whose diagonal components are 0.2;

Σ_2 : Quadratic matrix with a dimension of $p \times p$, whose diagonal components are 0.5;

Σ_3 : Quadratic matrix with a dimension of $p \times p$, whose diagonal components are 1, and non-diagonal components are 0.5 (Mecklin and Mundfrom, 2000).

Thus, the normal mixture models have been obtained as below:

1. $0.9N_p(0, \Sigma_1) + 0.1N_p(1, \Sigma_2)$
2. $0.788675N_p(0, \Sigma_1) + 0.211325N_p(1, \Sigma_2)$
3. $0.5N_p(0, \Sigma_1) + 0.5N_p(1, \Sigma_2)$
4. $0.9N_p(0, \Sigma_3) + 0.1N_p(1, I)$
5. $0.788675N_p(0, \Sigma_3) + 0.211325N_p(1, I)$
6. $0.5N_p(0, \Sigma_3) + 0.5N_p(1, I)$

The elliptical contoured distributions considered in the simulation study are symmetrical distributions whose contours possess equal intensity. These distributions have an elliptical shape. The form of the density function of the elliptical distribution is $f(x) = k_p |\Sigma|^{-1/2} g[(x - \mu)' \Sigma^{-1} (x - \mu)]$. Here, x is a random vector, k_p is a fixed scalar, and $g(\cdot)$ is a non-increasing positive function. The elliptical contoured distributions are shown with $EC_p(\mu, \Sigma, g)$. Multivariate normal distribution with $k_p = (2\pi)^{-p/2}$ and $g(t) = \exp(-\frac{t}{2})$ values are a special case of the elliptical contoured distribution. Elliptical distributions are symmetrical distributions which have a close relationship with normal ones. Therefore, mild departures occur from normality. The type II Pearson distribution and type VII Pearson distribution have been used as an example of elliptical distributions. Johnson (1987) has stated that type II and type VII Pearson distributions are quite suitable for Monte Carlo studies. That is because these distributions are easy to generate, and they cover many of the elliptical distribution family. In our study, the Pearson type II distribution has been generated for different shape parameters of $m = 2, 4$, and 10 . For $m = 10$ this distribution is quite close to multivariate normality. Multivariate t and Cauchy distributions are a special case of the Pearson type VII family. A multivariate Cauchy distribution with $\nu = 1$ degrees of freedom and a multivariate t distribution with $\nu = 10$ degrees of freedom are obtained. In our study, the type VII Pearson distribution with $\nu = 1$ and 10 degrees of freedom has been generated. Other than these, a multivariate t distribution and a multivariate Cauchy distribution with 2 degrees of freedom has also been generated.

Distributions where severe departures are seen from normality possess skewness, and are outside the elliptical contoured distribution family. The multivariate normality tests, which are being examined, are expected to display high levels of power for these types of distributions. Multivariate chi-square and multivariate lognormal distributions belonging to this category have been considered in our simulation study. Both of these distributions exhibit a heavily skewed and non-normal kurtosis. Therefore, they are expected to easily determine that all of the tests have nonnormality and thus to

possess high levels of power. The multivariate chi-square distribution is the extended form of the well-known univariate chi-square distribution. The chi-square distribution with 1 degrees of freedom displays positive skewness. Therefore, the multivariate chi-square distribution with $1, 2$ and 4 degrees of freedom has been generated in our study. The values of the lognormal distribution are inevitably positive and the distribution is positively skewed. Therefore, it is expected to display a good performance in terms of the determination of deviations from multivariate normality of all tests. A simple logarithmic transformation will bring the data from lognormal distribution to normality.

Distributions, where the univariate marginal distributions are normal but joint distributions are not normal, are cases where the departure from multivariate normality is theoretically interesting. It is not possible to detect this situation using only the univariate method, and it is problematic even for multivariate methods. One of the multivariate distributions which fits this definition is the Khintchine's family of distribution. This distribution is expected to clarify the real power of the tests. It is expected that the tests do not behave well when detecting that the data does not possess multivariate normality.

Another case which is theoretically interesting is the multivariate nonnormal distributions which possess multivariate normal skewness and kurtosis. The family of generalized exponential power distribution possesses this feature. Horswell (1990) has used two members from this distribution family in his own study.

Apart from these distributions, the symmetrical and multivariate Laplace distribution has also been used in our simulation study. The multivariate Laplace distribution is the multi-dimensionally extended form of the univariate, symmetrical Laplace distribution. This distribution is leptokurtic. Therefore, it is one of the alternative distributions which can be used against multivariate normal distribution (Farrell et al., 2007; Székely and Rizzo, 2005).

The purpose of this study is to examine the power of the 6 multivariate normality tests, which have been developed in the recent years, in order to evaluate multivariate normality via the simulation. With this purpose in mind, the Monte Carlo simulation has been used. The sizes of the samples within the study have been determined as $n = 20, 50, 100$, and 200 , while the numbers of the variables have been determined as $p = 2, 3, 5$, and 10 . The data sets have been generated from 4 different sizes of samples and 4 different numbers of variables, from 21 different multivariate distributions. As a result, the simulations of $21 \times 4 \times 4 = 336$ cases have been tested with significance level of $\alpha = 0.05$. 10000 simulations of each combination have been performed. It has been chosen to restrict the sample sizes and numbers of variables in order to keep the required calculation

time for the simulation at reasonable levels and to be able to evaluate the small sample sizes for multivariate analyses. According to researchers, these sample sizes define the most critical cases of the multivariate normality assumption (in terms of type I and type II errors). The multivariate normality tests examined in the study are the generalized Shapiro-Wilk test (GSW), which has been proposed by Villasenor-Alva and Gonzalez-Estrada, the Kankainen-Taskinen-Oja's skewness test ($b_{1,new}$), the Kankainen-Taskinen-Oja's kurtosis test ($b_{2,new}$), the Energy test, the Henze-Zirkler (HZ) test and the Royston (1992) test. The Monte Carlo simulation method has been used to determine the power of the tests. In this context, several multivariate distributions have been considered. These distributions are multivariate normal distribution, multivariate normal mixture distributions (which contain various contamination levels, means and variances), the Pearson type II and Pearson type VII distributions from the elliptical contoured distribution family, the multivariate t distribution, the multivariate Cauchy distribution and multivariate Laplace distribution, from among the symmetrical distributions, the multivariate chi-square distribution and multivariate lognormal distribution from among the heavily skewed distributions, the Khintchine distribution and the generalized exponential power distribution, from among distributions which, according to their features belong to multivariate normal distributions, but which are not themselves multivariate normal distributions.

When making a comparison in terms of type I errors, empirical type I error values have been obtained for the different p, n at nominal α value. For each test, the convergence level of the empirical type I error to the nominal α value has been examined. The rejection rates of the null hypothesis for multivariate normality have been taken as

empirical type I errors, and have been obtained for every case and every distribution from every test, and presented in tables. Thus, the empirical power of each test statistic has been determined.

4. Results and Discussion

The empirical type I error rates and powers of the generalized Shapiro-Wilk test (GSW), which has been proposed by Villasenor-Alva and Gonzalez-Estrada, the Kankainen-Taskinen-Oja's skewness test ($b_{1,new}$), the Kankainen-Taskinen-Oja's kurtosis test ($b_{2,new}$), the Energy test, the Henze-Zirkler (HZ) test and the Royston (1992) test, which have been selected for the application part of this study in the multivariate normality tests, have been compared using the Monte Carlo simulation method. The first distribution on which the simulation was carried out was the multivariate normal distribution. In this case, the null hypothesis that the data set has been generated from a multivariate normal distribution will be tested. It is expected that the empirical rejection rates of data sets generated from multivariate normal distributions are close to the nominal α significance level. A rejection rate considerably higher than the α level will mean that there is a problem in the type I error rate. The other distributions considered in this study will display deviations from multivariate normality, which vary between mild to severe. Under these circumstances it is necessary to assess the null hypothesis as being incorrect, and to reject it. Low rejection rates will mean, in particular, that there is a problem in the type II error rate in the comparisons with the other tests, as well as a problem in the power of the tests.

The rates of rejection of the H_0 hypothesis in connection with tests of the multivariate normal distributions have been given as a percentage, as in Table 1.

Table 1: Type I error rates of the tests for multivariate normal distribution(MVN) $\alpha=5\%$

		GSW	$b_{1,new}$	$b_{2,new}$	Energy	H-Z	Royston
p=2	n=20	4.89	1.67	0.77	4.74	3.66	4.79
	n=50	4.96	3.73	2.37	4.78	4.16	4.91
	n=100	4.55	4.04	3.12	4.44	4.40	4.51
	n=200	5.01	4.52	3.94	4.92	5.15	4.78
p=3	n=20	4.76	0.94	0.86	4.97	3.37	4.71
	n=50	4.78	3.06	3.00	4.70	4.19	4.89
	n=100	4.70	3.90	4.17	4.65	4.60	4.53
	n=200	4.70	4.45	4.49	4.60	5.19	4.52
p=5	n=20	4.67	0.12	1.00	4.90	2.79	4.71
	n=50	4.88	2.02	4.16	4.87	4.46	4.81
	n=100	4.68	2.99	5.46	4.61	4.78	4.61
	n=200	4.78	3.99	6.20	4.73	4.91	4.77
p=10	n=20	4.71	0.00	1.76	4.82	2.81	4.35
	n=50	4.89	5.50	5.02	5.17	4.10	4.64
	n=100	4.73	3.48	8.65	5.03	4.79	4.94
	n=200	4.73	3.48	8.65	5.03	4.79	4.94

When Table 1 is examined, it is observed that the rejection rates of the H_0 hypothesis in certain tests is much lower than the nominal α level, and in

some it is higher than the nominal α level. The rates of rejection of the H_0 hypothesis in the Kankainen-Taskinen-Oja ($b_{1,new}$) skewness test show the

changes between 0.00 to 5.50 for $\alpha=0.05$. The rates of rejection of the H_0 hypothesis in the Kankainen-Taskinen-Oja ($b_{2,new}$) kurtosis test show the changes between 0.77 to 8.65 for $\alpha=0.05$. The $b_{1,new}$ and $b_{2,new}$ tests show especially severe deviations from the nominal α significance level at the situation where $n = 20$. The other tests possess rates of rejection at nominal α level for the multivariate normal data. According to the Henze-Zirkler (HZ) test; generalized Shapiro-Wilk (GSW) test, and the Energy and Royston tests, even though the nominal α level is a little further away from the significance level at $n = 20$, in general all four tests possess consistent rejection rates. The results for $n=20$ and $p= 2, 3, 5$, and 10 related to the multivariate distributions are shown in Table 2. It is expected that the kurtosis test will be the least sensitive to the lognormal distribution. As can be seen from Tables 2 and 3, in most cases the test with the lowest empirical power is the kurtosis ($b_{2,new}$) test. In addition, the $b_{2,new}$ test displays a lower power than the other four tests at $n = 20$ and 50 , and its power is seen to rise as the sample size increases. The Royston test is more powerful for $n = 20, p=2$ and 3 , and the GSW test is more powerful for $n = 50, p = 2$. At $n = 100$ and $p=5$ and 10 the empirical power of all tests are 100%. The Energy test has the lowest empirical power at $n = 100, p = 2, 3$ and $n = 200$. When $n = 200$, the empirical power of the Energy test decreases.

For multivariate Cauchy distribution, the differences between the tests can be seen when $n = 20$. There is quite an interesting situation for the Energy test.

While the Energy test possesses the highest empirical power when $n = 20$ and 50 , when $n = 100$, $p=2, 3$ and $n = 200$, it has the lowest empirical power. A consistent increase is seen in the power of the $b_{1,new}$ test based on the sample size. Therefore it can be said that the $b_{1,new}$ test is sensitive to sample size. However, it still displays a lower power than the other tests for $n = 20$ and 50 . The GSW test possesses the lowest power (44.04 %) at 0.05 significance levels for $n = 20$ and $p = 10$. Something else which stands out here is that the $b_{1,new}$ test also displays quite low power (56.65%) at 0.05 significance levels for $n = 20$ and $p = 10$. All tests possess the maximum empirical power (100%) at $n=50, 100$ and $p = 10, n = 100$ and $p = 5$.

It is expected that the skewness tests in the multivariate Laplace distribution display low power. Where the sample size is 20, it is observed that the empirical power for all tests are low. However, when the samples sizes rise, the empirical power of the tests increases. Even still, there is no situation where all of the tests reach the maximum (100%) empirical power levels (in connection with n and p). The test with the highest empirical power in all cases with this distribution is the Energy test.

For multivariate $t_{(2)}$ distribution as with the multivariate Laplace distribution, the test with the lowest power is the $b_{1,new}$ test, and the empirical

power at the 5% significance level for $p = 10, n=20$ is 5.08%. The empirical power of the Energy test increases when the sample size changes from $n = 20$ to $n = 50$ and 100 , but when the sample size is $n = 200$, this has a contrasting effect. Despite the fact that the $b_{1,new}$ test has low empirical power, it is sensitive to sample size. In the same way, when the sample size for the GSW, HZ, Royston and $b_{2,new}$ tests increases, an increase is also seen in their power. All tests reach 100% power at $n=100$ and $p = 10$.

The skewness and kurtosis values of the generalized exponential power distributions are equal to the multivariate normal distribution values, but the distribution is not a multivariate normal distribution. Most of the tests display low power for the generalized exponential power distribution. In particular, the $b_{2,new}$ test displays a poor power performance. The $b_{2,new}$ test can be said to be the worst test for this distribution. Apart from this, it is also seen that the $b_{1,new}$ test has the lowest power (0.00%) of all the tests for $n = 20$. Therefore, it possesses the worst power performance for $n = 20$ and $p = 10$. However, as the sample size grows, an increase in the power is also seen. When $n = 20$ and 50 for the $b_{1,new}$ test, a decrease in power is seen together with an increase in the number of variables. The Royston test and the GSW test can be said to display good levels of power for the generalized exponential power distribution.

When Tables 2-5 is examined in order to the Khintchine distribution, it can be seen that in particular the GSW and Royston tests display quite low power. While the empirical power values for the GSW test show the changes between 3.06 and 4.89, the empirical power values of the Royston test changes between 4.85 and 6.19. The $b_{1,new}$ test possesses the lowest empirical power (0.00%) for $n = 20$ and $p = 10$. The Energy test can be said to display better empirical power than the other tests for this distribution.

As it is known that the multivariate chi-square distribution with 1 degrees of freedom is a severely skewed distribution with heavy tailed, it is expected that the values related to the skewness test will be high. When Tables 2-5 is examined, it can be seen that the $b_{1,new}$ test is sensitive to sample size. However, as the number of variables increases at $n = 20$, the power of the $b_{1,new}$ test decreases. An interesting situation seen with the $b_{1,new}$ test is that it displays quite low power (4.01 for $\alpha=0.05$) at $p = 10$ and $n = 20$. The $b_{2,new}$ test displays low power at $n = 20$, while the remaining four tests display quite a good power performance.

All tests reach maximum power (100%) at $n = 100, p = 5, 10$. The Royston test reaches 100% power in almost every case. When the multivariate chi-square distribution with 2 degrees of freedom in Tables 2-5 is examined, it can be seen that all tests are sensitive to sample size. However, when $n = 20, p = 10$, the $b_{1,new}$ test has the lowest power (0.00 for $\alpha=0.05$).

Table 2: Empirical power (in percent): ($\alpha=5\%$, $n=20$, $p=2, 3, 5$ and 10)

Alternative Distr.	p=2						p=3					
	GSW	b1,new	b2,new	Energy	H-Z	Royston	GSW	b1,new	b2,new	Energy	H-Z	Royston
MV Lognormal	99.24	88.05	60.28	97.18	96.18	99.53	99.9	89.58	67.42	98.70	97.46	100
MV Cauchy	92.58	88.02	88.64	96.84	96.32	94.78	93.01	93.06	93.6	98.86	98.37	97.12
MV Laplace	32.23	26.07	19.88	45.85	40.49	36.44	34.30	30.27	25.84	62.52	54.61	44.47
MV $t_{(2)}$	61.99	55.07	51.61	65.32	61.92	62.29	62.68	62.13	60.97	76.79	70.77	71.61
Generalized Exp power	73.08	16.78	2.98	50.02	44.97	77.83	82.17	11.39	3.03	47.84	41.44	89.53
Khintchine	4.04	4.93	2.97	10.44	7.78	5.12	3.55	6.46	5.55	19.21	13.23	5.15
MV $\chi^2_{(1)}$	99.91	88.14	50.99	99.15	98.5	99.99	99.98	87.85	55.87	99.59	98.98	100
MV $\chi^2_{(2)}$	95.35	62.74	28.63	85.81	82.24	96.68	98.59	60.74	31.46	88.95	83.85	99.50
MV $\chi^2_{(4)}$	71.26	36.51	15.22	55.43	50.10	100	82.06	31.60	15.91	58.62	48.62	100
Pearson II (2)	5.85	0.21	0.00	4.25	3.54	5.37	5.72	0.12	0.18	2.81	2.77	5.55
Pearson II (4)	3.51	0.56	0.19	3.46	2.84	3.19	2.93	0.23	0.32	2.95	2.43	2.81
Pearson II (10)	3.66	1.02	0.55	4.25	3.09	3.36	3.50	0.50	0.54	3.73	2.58	3.34
Pearson VII (1)	97.33	87.9	87.85	97.09	96.42	97.79	99.49	93.02	92.68	98.96	98.55	99.73
Pearson VII (10)	11.55	5.81	3.83	10.02	7.46	12.04	12.62	4.14	3.80	10.52	6.40	13.59
MVN Mix I (0.9)	19.75	23.13	13.30	28.02	23.64	31.14	16.05	21.46	16.53	34.20	24.07	37.07
MVN Mix I (0.788675)	23.34	23.14	9.07	36.72	32.97	36.94	21.18	22.80	12.40	44.50	35.16	46.03
MVN Mix I (0.5)	14.90	5.59	2.64	19.31	16.71	16.94	15.23	5.84	5.00	26.09	21.29	20.55
MVN Mix II (0.9)	5.73	2.02	2.03	5.37	4.20	4.97	6.02	1.73	4.15	5.97	3.76	5.30
MVN Mix II (0.788675)	5.59	1.93	2.16	5.47	4.10	4.65	6.35	1.79	4.30	6.59	4.16	5.00
MVN Mix II (0.5)	4.73	1.72	1.35	5.42	4.01	4.29	5.65	1.37	2.97	6.51	4.12	4.42
Alternative Distr.	p=5						p=10					
	GSW	b1,new	b2,new	Energy	H-Z	Royston	GSW	b1,new	b2,new	Energy	H-Z	Royston
MV Lognormal	99.99	85.11	72.68	99.29	97.4	100	100	9.46	59.9	97.75	80.97	100
MV Cauchy	89.58	95.61	96.49	99.74	99.36	99.06	44.04	56.65	93.94	99.91	97.83	99.71
MV Laplace	33.99	29.13	30.22	82.33	71.98	56.45	19.63	3.25	27.77	93.81	77.47	73.72
MV $t_{(2)}$	55.08	62.03	69.14	88.33	77.39	82.78	21.93	5.08	56.45	93.58	62.65	92.45
Generalized Exp power	89.37	2.62	2.60	37.13	28.50	98.06	89.26	0.00	2.38	16.30	9.69	99.99
Khintchine	3.06	5.82	8.23	42.40	27.95	4.85	3.19	0.15	12.63	72.00	40.69	5.31
MV $\chi^2_{(1)}$	100	79.04	56.54	99.44	98.36	100	100	4.01	39.19	96.11	98.87	100
MV $\chi^2_{(2)}$	99.67	41.99	29.01	88.78	78.05	99.98	99.88	0.00	15.49	72.38	37.68	100
MV $\chi^2_{(4)}$	90.72	14.22	13.34	57.29	38.92	100	95.23	0.00	6.86	38.16	14.33	100
Pearson II (2)	5.78	0.00	0.33	1.79	2.02	6.63	4.68	0.00	1.07	1.82	2.41	7.60
Pearson II (4)	3.24	0.00	0.34	2.54	1.75	2.85	2.76	0.00	1.25	3.06	2.61	2.28
Pearson II (10)	3.57	0.00	0.68	3.86	2.30	3.36	3.11	0.00	1.51	3.75	2.69	2.69
Pearson VII (1)	99.93	95.30	96.31	99.78	99.34	99.98	100	54.93	96.30	99.88	97.40	100
Pearson VII (10)	14.56	0.79	2.76	10.46	4.81	16.54	16.65	0.00	2.69	10.05	3.68	21.64
MVN Mix I (0.9)	10.91	9.55	19.29	37.15	18.84	48.99	5.37	0.00	12.04	26.41	6.84	66.81
MVN Mix I (0.788675)	14.73	12.75	16.41	49.05	30.31	60.82	7.35	0.00	20.24	38.90	11.58	82.26
MVN Mix I (0.5)	14.25	3.94	9.41	34.77	25.74	27.58	8.52	0.00	25.90	39.00	17.33	41.71
MVN Mix II (0.9)	5.90	0.31	9.47	7.26	3.99	5.58	5.38	0.00	25.63	7.82	3.42	6.30
MVN Mix II (0.788675)	6.75	0.47	9.24	8.32	4.52	5.09	5.80	0.00	25.79	10.55	3.88	5.64
MVN Mix II (0.5)	6.34	0.30	5.44	8.54	4.59	4.51	5.68	0.00	17.34	11.89	4.68	4.82

Table 3. Empirical power (in percent): ($\alpha=5\%$, $n=50$, $p=2, 3, 5$, and 10)

Alternative Distr.	p=2						p=3					
	GSW	b1,new	b2,new	Energy	H-Z	Royston	GSW	b1,new	b2,new	Energy	H-Z	Royston
MV Lognormal	100	100	97.75	100	100	99.96	100	100	99.3	100	100	100
MV Cauchy	99.97	98.46	99.99	100	100	99.98	99.99	99.77	100	100	100	99.99
MV Laplace	69.89	50.05	70.48	84.52	82.47	71.28	78.18	63.14	86.21	96.18	95.51	82.25
MV $t_{(2)}$	95.54	85.08	94.64	94.87	93.54	93.03	97.85	92.85	98.50	98.69	97.97	97.23
Generalized Exp power	99.99	76.68	7.11	97.33	95.99	100	100	74.03	8.05	98.00	96.71	100
Khintchine	4.53	10.91	11.85	15.84	14.06	4.95	4.14	20.53	27.35	39.34	34.06	4.93
MV $\chi^2_{(1)}$	100	99.99	94.93	100	100	100	100	100	97.89	100	100	100
MV $\chi^2_{(2)}$	100	99.74	77.41	100	99.96	100	100	99.84	84.69	100	99.99	100
MV $\chi^2_{(4)}$	99.64	93.52	51.41	96.47	93.87	100	100	94.67	57.83	98.57	95.71	100
Pearson II (2)	22.55	0.00	0.75	10.77	14.56	21.21	28.58	0.11	0.13	7.57	13.17	26.73
Pearson II (4)	5.42	0.41	0.12	4.59	5.42	4.99	5.34	0.45	0.18	3.54	5.09	4.71
Pearson II (10)	3.68	1.51	0.73	4.47	4.48	3.66	3.66	1.46	1.21	3.85	4.20	3.32
Pearson VII (1)	99.99	98.33	99.97	99.99	99.99	99.99	100	99.65	100	100	100	100
Pearson VII (10)	20.76	15.32	17.26	14.52	11.25	20.91	24.39	15.12	20.53	16.21	10.34	25.33
MVN Mix I (0.9)	50.07	71.38	52.22	61.71	52.25	65.12	48.90	82.73	69.19	77.32	61.72	77.20
MVN Mix I (0.788675)	59.76	80.08	33.55	82.01	76.55	79.69	57.97	89.08	46.99	92.82	86.52	90.11
MVN Mix I (0.5)	38.45	26.51	8.83	53.91	53.20	45.8	39.79	37.17	17.95	70.22	68.61	56.58
MVN Mix II (0.9)	6.82	5.13	6.44	5.86	5.17	5.61	8.19	6.35	13.08	6.87	5.14	5.60
MVN Mix II (0.788675)	7.32	5.40	6.69	6.69	5.77	5.73	9.28	6.39	13.68	8.16	6.36	5.20
MVN Mix II (0.5)	5.13	5.40	4.64	6.80	6.06	4.40	7.46	6.39	8.81	9.42	7.85	4.41
Alternative Distr.	p=5						p=10					
	GSW	b1,new	b2,new	Energy	H-Z	Royston	GSW	b1,new	b2,new	Energy	H-Z	Royston
MV Lognormal	100	100	99.87	100	100	100	100	100	99.98	100	100	100
MV Cauchy	100	100	100	100	100	100	100	100	100	100	100	100
MV Laplace	86.56	80.92	96.99	99.82	99.71	92.79	90.53	95.88	99.92	100	100	98.61
MV $t_{(2)}$	98.93	98.40	99.84	99.92	99.76	99.31	98.99	99.92	100	100	99.98	99.95
Generalized Exp power	100	63.46	9.52	97.44	93.60	100	100	26.23	10.37	85.47	65.24	100
Khintchine	3.58	39.97	60.31	83.81	78.28	5.45	3.46	74.81	95.33	99.79	99.03	6.09
MV $\chi^2_{(1)}$	100	100	99.21	100	100	100	100	100	99.74	100	100	100
MV $\chi^2_{(2)}$	100	99.87	90.53	100	100	100	100	99.17	94.07	100	99.71	100
MV $\chi^2_{(4)}$	100	94.19	65.59	99.33	94.92	100	100	84.46	71.46	99.12	81.04	100
Pearson II (2)	38.52	0.00	0.27	3.14	11.00	39.18	50.09	0.00	0.54	0.82	7.53	64.4
Pearson II (4)	5.53	0.33	0.45	2.43	5.31	4.84	5.45	0.00	1.18	1.10	4.80	5.38
Pearson II (10)	3.1	0.98	1.75	3.29	4.25	2.79	3.03	0.23	2.47	2.68	4.26	2.46
Pearson VII (1)	100	99.98	100	100	100	100	100	100	100	100	100	100
Pearson VII (10)	30.40	13.54	23.19	18.81	10.02	32.7	41.23	6.23	25.1	21.21	7.47	47.33
MVN Mix I (0.9)	42.15	90.85	85.39	89.17	64.37	88.44	25.17	90.17	95.68	93.86	37.02	96.4
MVN Mix I (0.788675)	51.90	94.79	68.78	97.72	89.92	97.68	36.99	97.13	94.23	99.27	70.42	99.9
MVN Mix I (0.5)	41.28	58.88	35.99	87.33	83.82	71.89	34.91	84.95	74.02	98.21	88.19	88.98
MVN Mix II (0.9)	9.48	7.53	29.07	10.55	6.18	6.50	8.31	5.31	65.25	17.15	5.69	7.22
MVN Mix II (0.788675)	11.54	9.27	29.85	14.47	8.08	6.24	11.91	10.49	70.15	29.93	8.34	7.09
MVN Mix II (0.5)	9.63	9.27	18.37	15.67	10.52	4.64	11.84	10.49	47.29	35.67	13.07	4.27

The $b_{2,new}$ test has less power than the other tests. All tests reach maximum power (100%) at $n = 200$, $p = 5$. The test with the best performance for the distribution is the Royston test. According to the results related to the multivariate chi-square distribution with 4 degrees of freedom, there is no situation where all of the tests which have been considered reach maximum. Increases are seen in the empirical power of all tests as the sample size increases. The $b_{2,new}$ test is more powerful than the other tests. When $n = 20$, $p = 10$, the $b_{1,new}$ test has the lowest power performance (0.00%). As the number of variables increases at $n = 20$, an increase is seen in the power of the GSW and Royston tests. The best test for this distribution is the Royston test.

The type II Pearson distribution, for which the shape parameter is $m = 2$, is an elliptical contoured distribution with a short tailed. The power levels of all tests other than the skewness test ($b_{1,new}$) display a change according to the sample size. The $b_{1,new}$ test has the worst power performance. Apart from this when $p = 10$ the power at all sample sizes is 0.00. The $b_{2,new}$ test does not display good performance either, while the power levels of the GSW and Royston tests show a rapid increase based on the sample size. Both the GSW and Royston tests have a power of 100% at $p=5$, 10 and $n = 200$. When the results related to the type II Pearson distribution, for which the shape parameter is $m = 4$, are examined in Tables 2-5, as with the previous distribution, the test with the worst power performance is the $b_{1,new}$ test. The power levels of the GSW and Royston tests display a lower increase for the type II Pearson distribution with the $m = 4$ shape parameter. The HZ and Energy tests display lower power than these two tests. None of the tests are able to reach maximum power (100%). When the results related to the type II Pearson distribution, for which the shape parameter is $m = 10$, are considered in the tables, the power levels of all tests decrease based on the increase in the shape parameter m . The test with the highest power when $m = 10$ is the HZ test, but this power level only has the low value of 6.55%. For $n = 20$ and 50, the power levels of all tests remain beneath the nominal α significance level.

For the results related to the type VII Pearson distribution with 1 degrees freedom in Tables 2-5, it can be said that the $b_{1,new}$ test is less powerful than the other tests when $n = 20$ and 50. An increase is seen in the power levels of the GSW, $b_{2,new}$, HZ and Royston tests based on the sample size and number of variables. While the Energy test displays consistent power levels at $n=20$, 50 and $n = 100$, $p = 5$, 10, it has quite low power at $n = 100$, $p = 2$, 3 and $n=200$. An important decrease is seen in the power levels of the Energy test, in particular when $n = 200$. All tests reach maximum power at $n = 50$, $p = 10$ and $n = 100$, $p = 5$, 10. The GSW and Royston tests have the best power performance for this distribution. When the results related to the

type VII Pearson distribution with 10 degrees of freedom are examined, a decrease in the power levels of the tests has been observed together with the increase in degrees of freedom. There is no situation when all of the tests reach maximum power. The $b_{1,new}$ test has the lowest empirical power (0.00%) at $n=20$, $p=10$. The power of all tests display increases based on sample size. As the number of variables rises for all sample sizes, the power of the GSW and Royston tests increases. However, when $n = 20$, the power of the other four tests decreases as the number of variables increases.

According to the results related to the type I multivariate normal mixture distribution with three different contamination levels (0.9-0.1, 0.788675-0.211325, 0.5-0.5) in Tables 2-5, while the power of the tests increases when the contamination parameter falls from 0.9 to 0.788675, when the contamination parameter falls from 0.788675 to 0.5, the power of the tests decreases. At all three contamination levels, the $b_{1,new}$ test has the lowest power (0.00%) at $n=20$ and $p = 10$. It can be seen that the power levels of the tests increase at different sample sizes and numbers of variables, as the sample size increases, at the first contamination level (0.9-0.1). When $n = 20$, the power of the $b_{1,new}$ test decreases based on an increase in the number of variables. When $p = 5$ and 10 lower power values are obtained for the HZ test. The power of the Royston test increases together with an increase in the number of variables, while the power of the GSW test decreases. Therefore, the test with the best power performance at the first contamination level is the Royston test. It can be seen that the power of the tests increases at different sample sizes and numbers of variables, as the sample size increases, for the second contamination level (0.788675-0.211325). As with the previous distribution, when $n = 20$, the power of the $b_{1,new}$ test decreases based on an increase in the number of variables. When the number of variables increases, an increase is seen in the power of the $b_{1,new}$, $b_{2,new}$, Energy and Royston tests, while the power of the GSW test decreases. Additionally, when $p = 5$ and 10 lower power values are obtained for the HZ test. As a result, it can be said that the Royston test displays a more powerful performance with the multivariate normal mixture distribution at the second contamination level. As the sample size increases, it can be seen that there is also an increase in the power of the tests. When the number of variables increases, the power levels of the $b_{2,new}$, Energy, HZ and Royston tests also increase. In general, it can be said that the Royston test displays a more powerful performance with the multivariate normal mixture distribution at the third contamination level.

According to the results related to the type II multivariate normal mixture distribution with three different contamination levels (0.9-0.1, 0.788675-0.211325, 0.5-0.5) are examined, it can be seen that

all tests possess quite low power levels for all three contamination levels.

Table 4. Empirical power (in percent): ($\alpha=5\%$, $n=100$, $p=2, 3, 5$, and 10)

Alternative Distr.	p=2						p=3					
	GSW	b1,new	b2,new	Energy	H-Z	Royston	GSW	b1,new	b2,new	Energy	H-Z	Royston
MV Lognormal	100	100	99.97	98.43	100	100	100	100	100	98.61	100	100
MV Cauchy	100	99.6	100	52.91	100	100	100	99.96	100	63.67	100	100
MV Laplace	93.92	59.78	96.16	98.85	98.66	94.26	97.94	75.62	99.56	99.96	99.97	98.28
MV $t_{(2)}$	99.91	94.13	99.93	91.37	99.77	99.71	100	98.27	100	91.17	99.99	99.94
Generalized Exp power	100	99.32	8.54	99.99	99.99	100	100	99.54	10.10	100	100	100
Khinchine	4.89	14.93	22.80	27.11	24.87	5.37	4.52	28.58	55.69	69.65	67.04	5.29
MV $\chi^2_{(1)}$	100	100	99.96	100	100	100	100	100	100	100	100	100
MV $\chi^2_{(2)}$	100	100	96.88	100	100	100	100	100	98.94	100	100	100
MV $\chi^2_{(4)}$	100	99.99	80.57	100	99.95	100	100	99.99	87.31	100	99.98	100
Pearson II (2)	70.58	0.00	29.63	30.41	37.29	68.19	84.75	0.00	19.11	25.44	36.31	82.55
Pearson II (4)	10.75	0.45	2.88	7.78	10.29	9.92	13.01	0.41	1.26	6.37	9.95	11.39
Pearson II (10)	3.46	1.51	0.91	4.73	5.32	3.37	3.54	1.47	1.08	4.39	5.09	3.33
Pearson VII (1)	100	99.59	100	53.79	100	100	100	99.99	100	63.71	100	100
Pearson VII (10)	32.58	22.34	33.36	20.45	15.31	33.04	39.21	23.24	38.66	24.00	15.92	40.29
MVN Mix I (0.9)	81.14	94.77	85.52	88.30	78.16	90.50	81.23	98.56	95.46	96.65	88.18	96.47
MVN Mix I (0.788675)	89.91	99.24	60.73	98.76	97.24	98.33	90.17	99.92	76.90	99.95	99.55	99.77
MVN Mix I (0.5)	70.84	61.84	18.37	89.70	87.99	81.8	74.42	79.43	36.47	97.67	97.31	91.09
MVN Mix II (0.9)	8.57	6.92	10.90	6.20	5.48	6.44	9.71	8.92	20.10	7.86	6.22	6.29
MVN Mix II (0.788675)	9.96	7.02	11.83	7.29	6.37	5.58	13.36	10.32	23.29	11.40	8.05	6.37
MVN Mix II (0.5)	6.44	7.02	7.54	9.26	7.89	4.09	9.59	10.32	14.55	15.73	12.14	4.34
Alternative Distr.	p=5						p=10					
	GSW	b1,new	b2,new	Energy	H-Z	Royston	GSW	b1,new	b2,new	Energy	H-Z	Royston
MV Lognormal	100	100	100	100	100	100	100	100	100	100	100	100
MV Cauchy	100	100	100	100	100	100	100	100	100	100	100	100
MV Laplace	99.72	91.19	100	100	100	99.79	100	99.58	100	100	100	100
MV $t_{(2)}$	100	99.83	100	99.99	100	100	100	100	100	100	100	100
Generalized Exp power	100	99.18	12.39	100	99.99	100	100	95.49	15.33	99.99	99.21	100
Khinchine	4.15	57.55	94.96	99.02	98.86	5.57	3.80	93.56	100	100	99.99	6.19
MV $\chi^2_{(1)}$	100	100	100	100	100	100	100	100	100	100	100	100
MV $\chi^2_{(2)}$	100	100	99.73	100	100	100	100	100	99.93	100	100	100
MV $\chi^2_{(4)}$	100	100	92.62	100	99.99	100	100	99.98	96.36	100	99.9	100
Pearson II (2)	96.33	0.00	2.39	13.16	31.12	96.03	99.77	0.00	0.00	1.79	20.74	99.96
Pearson II (4)	16.43	0.27	0.55	4.02	8.88	14.66	24.11	0.17	0.67	1.46	8.20	22.34
Pearson II (10)	3.38	1.19	1.62	3.76	5.26	2.98	3.23	0.51	2.56	2.49	4.92	2.92
Pearson VII (1)	100	100	100	100	100	100	100	100	100	100	100	100
Pearson VII (10)	50.89	24.33	44.66	28.63	15.31	52.85	69.48	20.64	51.8	34.57	11.54	73.16
MVN Mix I (0.9)	79.38	99.81	99.50	99.47	92.88	98.97	68.90	99.98	100	99.95	83.89	99.86
MVN Mix I (0.788675)	89.20	100	94.01	100	99.92	100	84.08	100	99.94	100	99.27	100
MVN Mix I (0.5)	77.39	96.73	64.86	99.88	99.63	97.75	76.31	100	96.34	100	99.96	99.82
MVN Mix II (0.9)	11.86	14.36	43.96	12.82	7.40	6.72	13.66	23.91	86.84	28.15	8.14	8.05
MVN Mix II (0.788675)	17.64	20.28	48.79	22.45	11.42	6.80	21.53	44.24	91.54	55.80	15.60	7.61
MVN Mix II (0.5)	15.36	20.28	30.05	29.20	17.91	4.27	21.29	44.24	73.50	69.43	30.73	4.08

While the power of the tests increases when the contamination parameter falls from 0.9 to 0.788675, when the contamination parameter falls from 0.788675 to 0.5, the power of the tests decreases. At all three contamination levels, the $b_{1,new}$ test has the lowest power (0.00%) at $n=20$ and $p=10$. It can be seen that the power of the tests increases as the sample size increases. As a result, it can be said that the test with the best power performance at the

first contamination level for the multivariate normal mixture distribution is the GSW test.

It can be seen that the power of the tests increases at different sample sizes and numbers of variables, as the sample size increases, at the second contamination level (0.788675-0.211325). The Energy test has better power than the other tests at $p=3$, and 5. At $p=10$, the $b_{2,new}$ test displays the highest empirical power. As the power of the tests increase at different sample sizes and

numbers of variables, as the sample size increases, at the third contamination level (0.5-0.5), it is

observed that there are increases in the power levels of the GSW, $b_{1,new}$, $b_{2,new}$, Energy and HZ tests.

Table 5. Empirical power (in percent): ($\alpha=5\%$, $n=200$, $p=2, 3, 5$ and 10)

Alternative Distr.	p=2						p=3					
	GSW	$b_{1,new}$	$b_{2,new}$	Energy	H-Z	Royston	GSW	$b_{1,new}$	$b_{2,new}$	Energy	H-Z	Royston
MV Lognormal	100	100	100	85.05	100	100	100	100	100	90.66	100	100
MV Cauchy	100	99.9	100	51.13	100	100	100	99.99	100	34.36	100	100
MV Laplace	99.90	67.38	99.96	99.98	100	99.89	100	81.75	100	99.95	100	99.99
MV $t_{(2)}$	100	98.01	100	78.20	100	100	100	99.64	100	79.02	100	100
Generalized Exp power	100	100	11.25	100	100	100	100	100	12.21	100	100	100
Khintchine	4.75	17.01	41.27	51.24	49.81	4.92	4.67	34.07	87.11	96.17	95.90	5.04
MV $\chi^2_{(1)}$	100	100	100	98.31	100	100	100	100	100	98.85	100	100
MV $\chi^2_{(2)}$	100	100	99.95	99.86	100	100	100	100	100	99.87	100	100
MV $\chi^2_{(4)}$	100	100	97.38	100	100	100	100	100	98.91	99.99	100	100
Pearson II (2)	99.69	0.14	93.39	74.45	76.34	99.49	99.99	0.00	88.86	73.98	77.30	100
Pearson II (4)	33.26	0.39	24.15	16.59	18.73	30.75	44.32	0.35	17.07	14.70	19.16	40.18
Pearson II (10)	5.25	1.53	2.46	5.45	6.20	4.91	5.33	1.53	1.77	5.03	6.55	5.04
Pearson VII (1)	100	99.95	100	50.66	100	100	100	100	100	35.50	100	100
Pearson VII (10)	51.97	26.95	55.52	31.14	22.90	52.03	62.80	29.90	62.73	37.23	25.03	63.09
MVN Mix I (0.9)	98.01	99.85	99.04	99.15	96.45	99.37	98.58	99.99	99.94	99.97	99.21	99.92
MVN Mix I (0.788675)	99.76	100	88.19	100	99.97	100	99.75	100	96.47	100	100	100
MVN Mix I (0.5)	96.35	94.68	38.80	99.87	99.67	98.99	97.57	99.26	68.18	100	99.99	99.91
MVN Mix II (0.9)	10.99	8.59	15.32	7.15	6.28	7.00	14.31	12.29	28.69	10.14	7.30	7.15
MVN Mix II (0.788675)	14.64	8.99	17.59	9.90	7.90	6.85	21.82	15.05	35.31	17.30	11.03	6.95
MVN Mix II (0.5)	8.63	8.99	11.14	15.25	12.19	3.96	16.04	15.05	21.75	31.50	21.13	3.78
Alternative Distr.	p=5						p=10					
	GSW	$b_{1,new}$	$b_{2,new}$	Energy	H-Z	Royston	GSW	$b_{1,new}$	$b_{2,new}$	Energy	H-Z	Royston
MV Lognormal	100	100	100	82.76	100	100	100	100	100	96.43	100	100
MV Cauchy	100	100	100	15.82	100	100	100	100	100	21.08	100	100
MV Laplace	100	95.17	100	100	100	100	100	99.86	100	100	100	100
MV $t_{(2)}$	100	99.99	100	51.80	100	100	100	100	100	79.48	100	100
Generalized Exp power	100	100	14.24	100	100	100	100	100	18.07	100	100	100
Khintchine	4.69	65.96	99.92	100	100	5.63	3.79	97.19	100	100	100	6.19
MV $\chi^2_{(1)}$	100	100	100	99.72	100	100	100	100	100	99.70	100	100
MV $\chi^2_{(2)}$	100	100	100	100	100	100	100	100	100	99.98	100	100
MV $\chi^2_{(4)}$	100	100	99.74	100	100	100	100	100	99.95	99.99	100	100
Pearson II (2)	100	0.00	66.38	60.50	72.12	100	100	0.00	0.90	18.35	54.65	100
Pearson II (4)	62.07	0.21	5.59	10.64	18.22	56.13	86.56	0.18	0.41	3.75	14.63	82.89
Pearson II (10)	5.92	1.37	1.41	3.97	6.05	5.37	6.22	0.77	1.72	2.71	5.77	5.14
Pearson VII (1)	100	100	100	16.42	100	100	100	100	100	21.59	100	100
Pearson VII (10)	77.98	32.38	69.19	44.02	24.21	78.55	93.97	34.09	77.43	53.08	20.10	94.63
MVN Mix I (0.9)	98.62	100	100	100	99.83	100	97.83	100	100	100	99.51	100
MVN Mix I (0.788675)	99.82	100	99.85	100	100	100	99.78	100	100	100	100	100
MVN Mix I (0.5)	98.58	100	93.03	100	100	100	98.98	100	99.99	100	100	100
MVN Mix II (0.9)	18.11	22.41	59.77	18.30	8.70	7.99	23.48	48.32	96.12	45.78	10.93	9.43
MVN Mix II (0.788675)	30.92	34.40	67.92	38.80	17.79	7.74	41.21	80.51	98.78	85.29	30.15	10.03
MVN Mix II (0.5)	27.56	34.40	47.32	61.20	35.50	3.76	42.98	80.51	93.09	97.34	64.49	3.91

It is seen that the power levels of the $b_{1,new}$ test decrease when the number of variables increases for $n = 20$, the power of the $b_{1,new}$ test increases together with the number of variables for $n = 50, 100, 200$. The Energy test provides better power results for the multivariate normal mixture distribution at the third contamination level.

5. Conclusions

In this study, the generalized Shapiro-Wilk test (GSW), which has been proposed by Villasenor-Alva and Gonzalez-Estrada, the Kankaiken-Taskinen-Oja's skewness test ($b_{1,new}$), the Kankaiken-Taskinen-Oja's kurtosis test ($b_{2,new}$), the Energy test, the Henze-

Zirkler (HZ) test and the Royston (1992) test have been introduced for testing of the null hypothesis in connection with multivariate normality, and comparisons have been made concerning the empirical type I errors rates of these tests and their power levels. In the comparisons concerning the type I errors for $n=20$, the $b_{1,new}$ and $b_{2,new}$ tests display quite poor results, while the other four tests display better results, close to the nominal α level. As the sample size increases an improvement in the $b_{1,new}$ and $b_{2,new}$ tests is seen.

In general, in all of the comparisons made concerning the power of the tests, the $b_{1,new}$ test gave the worst results. The $b_{1,new}$ test particularly possesses bad performance for symmetrical and leptokurtic distributions. Therefore, it will be more appropriate to use alternative tests instead of these ones. The $b_{2,new}$ test has given bad results for heavily skewed distributions, but good results for the type II multivariate normal mixture distribution. While the Energy test has poor performance in the symmetrical and platykurtic distributions, it does display better power performance for symmetrical and leptokurtic distributions. The GSW and Royston tests have provided good results for platykurtic tests.

As a result, the Energy and Royston tests have more powerful for type I normal mixture distributions, while the $b_{2,new}$ test has more powerful for type II normal mixture distributions. It has been observed that the GSW and Royston tests are more powerful for elliptical, skewed and generalized exponential power distributions. The Royston test has been found to be more powerful for symmetrical distributions, while the Energy test has been found to be empirically more powerful for the Khintchine distribution.

References

- Arcones MA (2007). Two tests for multivariate normality based on the characteristic function. *Mathematical Methods of Statistics*, 16(3): 177-201.
- Burdenski T (2000). Evaluating Univariate, Bivariate, and Multivariate Normality Using Graphical and Statistical Procedures. *Multiple Linear Regression Viewpoints*, 26(2): 15-28.
- Cardoso De Oliveira IR and Ferreira DF (2010). Multivariate extension of chi-squared univariate normality test. *Journal of Statistical Computation and Simulation*, 80(5): 513-526.
- Delmail D, Labrousse P and Botineau M (2011). The most powerful multivariate normality test for plant genomics and dynamics data sets. *Ecological Informatics*, 6(2): 125-126.
- Doornik JA and Hansen H (2008). An Omnibus Test for Univariate and Multivariate Normality. *Oxford Bulletin of Economics and Statistics*, 70(S1): 927-939.
- Dufour JM, Khalaf L, and Beaulieu MC (2003). Exact Skewness-Kurtosis Tests for Multivariate Normality and Goodness-of-Fit in Multivariate Regressions with Application to Asset Pricing Models. *Oxford Bulletin of Economics and Statistics*, 65(s1): 891-906.
- Farrell P, Salibian-Barrera M, and Naczka K (2007). On tests for multivariate normality and associated simulation studies. *Journal of Statistical Computation and Simulation*, 77(12): 1065-1080.
- Hanusz Z and Tarasińska J (2008). A note on Srivastava and Hui's tests of multivariate normality. *Journal of Multivariate Analysis*, 99(10): 2364-2367.
- Hanusz Z and Tarasińska J (2012). New tests for multivariate normality based on Small's and Srivastava's graphical methods. *Journal of Statistical Computation and Simulation*, 82(12): 1743-1752.
- Henze N (1997). Extreme smoothing and testing for multivariate normality. *Statistics and probability letters*, 35(3): 203-213.
- Henze N and Wagner T (1997). A New Approach to the BHEP Tests for Multivariate Normality. *Journal of Multivariate Analysis*, 62(1): 1-23.
- Henze N and Zirkler B (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10): 3595-3617.
- Holgersson HET (2006). A graphical method for assessing multivariate normality. *Computational Statistics*, 21(1): 141-149.
- Horswell RL (1990). A Monte Carlo comparison of tests of multivariate normality based on multivariate skewness and kurtosis. Ph.D. Dissertation, Louisiana State University, Louisiana, USA.
- Johnson ME (1987). *Multivariate Statistical Simulation*. Wiley, New York, USA.
- Kankainen A, Taskinen S and Oja H (2007). Tests of multinormality based on location vectors and scatter matrices. *Statistical Methods and Applications*, 16(3): 357-379.
- Koizumi K, Okamoto N and Seo T (2009). On Jarque-Bera tests for assessing multivariate normality. *Journal of Statistics: Advances in Theory and Applications*, 1(2): 207-220.
- Liang J, Tang ML, and Chan PS (2009). A generalized Shapiro-Wilk statistic for testing high-dimensional normality. *Computational Statistics and Data Analysis*, 53(11): 3883-3891.
- Looney SW (1995). How to use tests for univariate normality to assess multivariate normality. *The American Statistician*, 49(1): 64-70.

- Mecklin CJ and Mundfrom DJ (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*, 75(2): 93-107.
- Móri TF, Rohatgi VK, and Székely GJ (1993). On multivariate skewness and kurtosis. *Theory of Probability and Its Applications*. 38(3): 547-551.
- Okamoto N and Seo T (2010). On the distributions of multivariate sample skewness. *Journal of Statistical Planning and Inference* 140(10): 2809-2816.
- Pudelko J (2005). On a new affine invariant and consistent test for multivariate normality. *Probability and Mathematical Statistics*, 25(1): 43-54.
- Romeu JL and Ozturk A (1993) A comparative study of goodness of fit tests for multivariate normality. *Journal of Multivariate Analysis*, 46(2): 309-334.
- Royston JP (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. *Applied Statistics*, 32(2): 121-133.
- Royston JP (1992). Approximating the Shapiro-Wilk W-Test for non-normality. *Statistics and Computing*, 2(3): 117-119.
- Shapiro SS and Wilk MB (1965). An analysis of variance test for normality (complete samples), *Biometrika*, 52(3/4): 591-611.
- Sürücü B (2006). Goodness-of-Fit Tests for Multivariate Distributions. *Communications in Statistics - Theory and Methods*, 35:7, 1319-1331.
- Székely GJ and Rizzo ML (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1): 58-80.
- Tenreiro C (2011). An affine invariant multiple test procedure for assessing multivariate normality. *Computational Statistics and Data Analysis*, 55(5): 1980-1992.
- Thulin M (2014). Tests for multivariate normality based on canonical correlations. *Statistical Methods and Applications*, 23(2): 189-208.
- Villasenor Alva JA and Estrada EG (2009). A Generalization of Shapiro-Wilk's Test for Multivariate Normality. *Comm. Statist. Communications in Statistics—Theory and Methods*, 38(11): 1870-1883.