STATISTICS IN TRANSITION new series, December 2020
Vol. 21, No. 5, pp. 151–178, DOI 10.21307/stattrans-2020-060
Received – 17.09.2020 ; accepted – 07.11.2020

Comparison of selected tests for univariate normality based on measures of moments

Czesław Domański[1], **Piotr Szczepocki**[2]

ABSTRACT

Univariate normality tests are typically classified into tests based on empirical distribution, moments, regression and correlation, and other. In this paper, power comparisons of nine normality tests based on measures of moments via the Monte Carlo simulations is extensively examined. The effects on power of the sample size, significance level, and on a number of alternative distributions are investigated. None of the considered tests proved uniformly most powerful for all types of alternative distributions. However, the most powerful tests for different shape departures from normality (symmetric short-tailed, symmetric long-tailed or asymmetric) are indicated.

Key words: normality tests, Monte Carlo simulation, power of test.

1. Introduction

The normality of the data assumption is one of the most commonly found in statistical studies, especially in econometric models and generally in research on applied economics. It is well known that departures from normality may lead to substantial inaccuracy of estimation procedures and incorrect inference. Popular graphical methods (Q–Q plot, histogram or box plot) are unable to provide formal conclusive evidence that the normal assumption holds. Therefore, formal statistical tests are required to conclude the normality of the data.

The problem of testing normality has gained considerable importance and has led to the development of a large number of goodness-of-fit tests to detect departures from normality. Comprehensive descriptions and power comparisons of such tests have been the focus of attention of many previous works (for the newest research see: Thadewald and Büning, 2007, Romão, Delgado and Costa, 2010, Yap and Sim, 2011, Wijekularathna, Manage and Scariano, 2019). Although the referred comparison studies have been appearing over the years, there are fewer works that compare only normality tests based on the measures of the moments. The more recent ones, Domański (2010) and Domański and Jędrzejczak (2016), do not include several interesting and more recently developed tests. This class of tests is very broad, and among other encompasses one of the most popular econometric normality test (the Jarque–Bera test) and a number of new tests based on robust estimates of the moments. Furthermore, these tests offer also clear interpretation of results, which may be very useful for users: when normality is rejected, one also obtains information on

[1]Department of Statistical Methods, The Faculty of Economics and Sociology, University of Lodz, Poland. E-mail: czedoman@uni.lodz.pl. ORCID: https://orcid.org/0000-0001-6144-6231
[2]Department of Statistical Methods, The Faculty of Economics and Sociology, University of Lodz, Poland. E-mail: piotr.szczepocki@uni.lodz.pl.ORCID: https://orcid.org/0000-0001-8377-3831.

the sample: the distribution may be skewed to the left/right and/or long (or short) tailed. A further comparison of such normality tests can, therefore, be considered to be of foremost interest.

In Section 2, we present the procedures for normality tests considered in this study. The Monte Carlo simulation methodology for comparisons of the power of the normality tests and results are discussed in Section 3. Finally, a conclusion is given in Section 4.

2. Tests for Normality

In this article, we assume that we have a random sample $X_1, X_2, ..., X_n$ of independently and identically distributed random variables from a continuous univariate distribution with an unknown probability density function $f(x,\theta)$, where $\theta = (\theta_1, \theta_2, ..., \theta_k)$ is a vector of real-valued parameters. We test normality of this sample by verifying a composite null hypothesis:

$$H_0 : f(x;\theta) \in N(x;\mu,\sigma)$$

against the alternative:

$$H_1 : f(x;\theta) \notin N(x;\mu,\sigma)$$

where $N(x;\mu,\sigma)$ is a class of normal distributions with mean μ standard deviation σ, and probability density function given by

$$g(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Let a random variable X be distributed with mean μ and standard deviation σ. Then, the third (skewness) and fourth (kurtosis) standardized moment central moments (provided they exist) are respectively given by:

$$\sqrt{\beta_1} = \frac{E(X-\mu)^3}{\left[E(X-\mu)^2\right]^{3/2}} = \frac{E(X-\mu)^3}{\sigma^3} \tag{1}$$

and

$$\beta_2 = \frac{E(X-\mu)^4}{\left[E(X-\mu)^2\right]^2} = \frac{E(X-\mu)^3}{\sigma^4}. \tag{2}$$

These measures of probabilistic distribution are sometimes referred to as Pearson's moment coefficient of skewness and kurtosis.

Skewness is a measure of symmetry about the mean of a probability density. Kurtosis is a measure of the peakness of a probability density. For the normal distribution $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$. However, there are also non-normal distributions that are symmetric (e.g. t-Student) or have kurtosis equal to three (e.g. the Tukey distribution with parameter $\lambda = 0.135$). Furthermore, testing only skewness, when kurtosis is uncontrolled, may lead to incorrect conclusions. This is often the case of testing skewness in financial returns, for which kurtosis is significantly higher than in the case of normal distribution (Piontek, 2007).

Therefore, usually for the normality testing both skewness and kurtosis are involved. Such normality tests are often referred to as 'omnibus', because they are able to detect deviations from normality due to either skewness or kurtosis. In this study we only compare omnibus tests due to their convenience for practitioners: clear interpretation of results.

Empirical counterparts of the skewness and kurtosis are respectively given by

$$\sqrt{b_1} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S}\right)^3, \tag{3}$$

and

$$b_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S}\right)^4, \tag{4}$$

where $\bar{X} = 1/n\sum_{i=1}^{n} X_i$ is mean and $S = \sqrt{1/n\sum_{i=1}^{n}(X_i - \bar{X})^2}$ is standard devation. A number of transformations and alternative measures of skewness and kurtosis are the basis for the considered univariate normality tests presented below.

2.1. The D'Agostino–Pearson K^2 test

D'Agostino and Pearson (1973) proposed the test statistic K^2 that combines normalizing transformations of sample skewness and kurtosis.

The transformation of sample skewness $\sqrt{b_1}$ is based on Johnson's S_U transformation (Johanson, 1949) and is given by

$$Z(\sqrt{b_1}) = \frac{\ln\left(Y/c + \sqrt{(Y/c+)^2+1}\right)}{\sqrt{\ln(w)}}, \tag{5}$$

where

$$Y = \sqrt{b_1}\sqrt{\frac{(n+1)(n+3)}{6(n-2)}}, \qquad w^2 = -1 + \sqrt{2\gamma_2 - 1},$$

$$\gamma_2 = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \qquad c = \sqrt{\frac{2}{(w^2-1)}}.$$

D'Agostino and Pearson (1973) gave only percentage points of the distribution of transformation of b_2 under normal distribution. Anscombe and Glynn (1983) proposed similar transformation for sample kurtosis b_2 by fitting a linear function of the reciprocal of a chi-squared variable and then using the Wilson-Hilferty transformation (Wilson and Hilferty, 1931). The transformed sample kurtosis from Anscombe and Glynn (1983) is given by

$$Z(\sqrt{b_2}) = \left[\left(1 - \frac{2}{9A}\right) - \sqrt{\frac{1-2/A}{1+y\sqrt{2/(A-4)}}}\right]\sqrt{\frac{9A}{2}}, \tag{6}$$

where

$$y = \frac{b_2 - 3(n-1)/(n+1)}{24n(n-2)(n-3)/\left[(n+1)^2(n+3)(n+5)\right]},$$

$$A = 6 + \frac{8}{\sqrt{\gamma_1}}\left(\frac{2}{\sqrt{\gamma_1}} + \sqrt{1 + \frac{4}{\gamma_1}}\right),$$

$$\sqrt{\gamma_1} = \frac{6(n^2 - 5n + 2)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}.$$

The test statistic $K^2 = \left[Z(\sqrt{b_1})\right]^2 + [Z(b_2)]^2$ that combines D'Agostino and Pearson's transformation of sample skewness (5) and Anscombe and Glynn's transformation of sample kurtosis (6) follows approximately chi-squared distributed with two degrees of freedom as the sum of squares of two asymptotically independent standardized normals (D'Agostino, Belanger and D'Agostino, 1990).

2.2. The Jarque–Bera test

The Jarque–Bera test is one of the most popular goodness-of-fit test in the field of econometrics. Although it was first proposed by Bowman and Shenton (1975), it is mostly known from the work of Jarque and Bera (1987). The test Statistic *JB* is based on sample skewness and kurtosis and is defined as

$$JB = n\left(\frac{(b_1^{1/2})^2}{6} + \frac{(b_2 - 3)^2}{24}\right). \tag{7}$$

This test statistic is derived from the fact that, under normality, the asymptotic means of $b_1^{1/2}$ and b_2 are 0 and 3, and the asymptotic variances are $6/n$ and $24/n$, and finally the asymptotic covariance is zero. Thus, *JB* statistic is the sum of squares of two asymptotically independent standardized normals and has approximately chi-squared distribution with two degrees of freedom. However, the statistics $b_1^{1/2}$ and b_2 are not independently distributed and the sample kurtosis approaches normality very slowly. Thus, asymptotic critical values are strongly not recommended.

Jarque and Bera (1987) also proved that if the alternative distributions are in the Pearson family, *JB* statistic is the corresponding Lagrange multiplier test (also known as Rao's score test) for normality.

2.3. The Urzùa test

Urzùa (1996) proposed a modification of the Jarque–Bera test called the adjusted Lagrange multiplier test by standardizing the sample skewness and kurtosis in the formula of *JB* statistics in the following way

$$ALM = n\left(\frac{(b_1^{1/2})^2}{c_1} + \frac{(b_2 - c_2)^2}{c_3}\right), \tag{8}$$

where

$$c_1 = \frac{6(n-2)}{(n+1)(n+3)}, \qquad c_2 = \frac{3(n-1)}{(n+1)}, \qquad c_3 = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}.$$

The idea of this modification is to use, instead of the asymptotic means and variances of the standardized third and fourth moments, their exact counterparts. On the basis of Fisher (1930) $k-$statistics, Urzùa showed that under normality, the exact mean and variance of $b_1^{1/2}$ are 0 and c_1, and the exact mean and variance of b_2 are c_2 and c_3.

On the basis of asymptotical distributions of *ALM* statistic, the hypothesis of normality is rejected at some significance level if the value of statistic exceeds critical value of a chi-squared distribution with two degrees of freedom. This modification of *JB* statistic behaves much better for small- and medium-size samples, than the original statistic when one uses asymptotical tables of critical values (Urzùa, 1996). However, in the case of Monte Carlo simulated critical values, Thadewald and Büning (2007) reported no improvement of power to the classical *JB* test.

2.4. The Doornik–Hansen test

Doornik and Hansen (2008) introduced another modification of the Jarque–Bera test for which the transformation creates statistics that are much closer to standard normal than in original *JB* statistic. Statistic of Doornik-Hansen test is given by

$$DH = \left[Z(\sqrt{b_1})\right]^2 + [z_2]^2, \tag{9}$$

in which they proposed to use the transformed sample skewness $Z(\sqrt{b_1})$ according to equation (5) and sample kurtosis is transformed to a chi-squared distribution with non-integer degrees of freedom, which is then translated into standard normal using the Wilson–Hilferty transformation

$$z_2 = \left[\left(\frac{\xi}{2a}\right)^{\frac{1}{3}} - 1 + \frac{1}{9a}\right]\sqrt{9a}, \tag{10}$$

where

$$\xi = (b_2 - 1 - b_1)2k, \qquad k = \frac{(n+5)(n+7)(n^3+37n^2+11n-313)}{12(n-3)(n+1)(n^2+15n-4)},$$

$$a = \frac{(n+5)(n+7)\left[(n^2+27n-70)+b_1(n-7)(n^2+2n-5)\right]}{6(n-3)(n+1)(n^2+15n-4)}.$$

The formulae (10) break down for $n \leqslant 7$. The DH statistic is also approximately chi-squared distribution with two degrees of freedom. However, because of its fast coverage, *DH* statistic does not require simulated quantiles of distribution under null hypothesis (Doornik and Hansen, 2008).

2.5. The Gel–Gastwirth test

Gel and Gastwirth (2008) proposed a modification of *JB* that uses a robust estimate of the dispersion, the average absolute deviation from the sample median (MAAD), instead of the second order central moment m_2. MAAD is defined by

$$\mathrm{MAAD} = \frac{\sqrt{\pi/2}}{n}\sum_{i=1}^{n}|X_i - \mathrm{med}_F|, \tag{11}$$

where med_F is the sample median. Robust dispersion measure is used, due to the fact that sample moments are known to be sensitive to outliers, and the sample variance is even more affected by outliers than the mean (Gel and Gastwirth, 2008). Thus, *RJB* statistic performs better than *JB* statistics in the case of long-tailed distributions (Gel and Gastwirth, 2008). However, in the case of short-tailed distribution robust measures of the dispersion may not be necessary.

The test statistic is given by

$$RJB = \frac{n}{6}\left(\frac{m_3}{\mathrm{MAAD}^3}\right)^2 + \frac{n}{64}\left(\frac{m_4}{\mathrm{MAAD}^4} - 3\right)^2. \tag{12}$$

Gel and Gastwirth (2008) also proved that under the null hypothesis of normality, the *RJB* test statistic asymptotically follows the chi-square distribution with two degrees of freedom. However, similarly to *JB*, for small and moderate samples the Monte Carlo simulated critical values are more preferable than asymptotic chi-squared distribution values.

2.6. The Bontemps-Meddahi tests

Bontemps and Meddahi (2005) proposed a family of normality tests developed on the basis of generalized method of moments approach and Hermite polynomials. The family of test statistics is given by

$$BM_{3-\rho} = \sum_{k=3}^{\rho}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_k\left(\frac{x_i - \bar{x}}{s}\right)\right)^2, \tag{13}$$

where H_k is the $k-$th order normalized Hermite polynomial. The considered moment conditions in the Bontemps-Meddahi tests are based on the Stein equation (Stein, 1972). The important property of the Stein equation is that, the expectation of the considered function is zero by construction. Bontemps and Meddahi (2005) showed that special examples of this equation correspond to the zero mean of any Hermite polynomial. The family of the Bontemps-Meddahi tests asymptotically follows the chi-square distribution with $\rho - 2$ degrees of freedom. The *JB* statistic almost coincides with BM_{3-4}. The only difference is that in *JB* test, the variance is estimated by $S = 1/n\sum_{i=1}^{n}(X_i - \bar{X})^2$ while in the Hermite case it is estimated by $1/(n-1)\sum_{i=1}^{n}(X_i - \bar{X})^2$. In the presented study we use the Bontemps-Meddahi test termed BM_{3-6}, because tests based on Hermite polynomials of degree at least seven do not provide gain in power (Bontemps and Meddahi, 2005).

2.7. The Hosking test

Hosking (1990) proposed to use L-moments, linear combinations of the order statistics, instead of classic central moments in order to obtain more powerful test in case of long-tailed distributions. L-moments are less affected by sample variability, and thus more robust to outliers.

Based on the second, third and fourth sample L-moments, which correspond to the second, third and fourth central moments, Hosking (1990) introduced new measures of skewness and kurtosis, termed L-skewness τ_3 and L-kurtosis τ_4 defined as

$$\tau_3 = \frac{l_3}{l_2}, \qquad \tau_4 = \frac{l_4}{l_2} \tag{14}$$

where l_r are order sample L-moment that can be estimated by

$$l_r = \sum_{k=0}^{r-1} p^*_{r-1,k} b_k, \tag{15}$$

where

$$p^*_{r-1,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}, \qquad b_k = \frac{1}{n} \sum_{i=1}^{n} \frac{(i-1)(i-2)\cdots(i-k)}{(n-1)(n-2)\cdots(n-k)}.$$

Hosking (1990) proposed to test normality by the following statistic

$$T_{Lmom} = \frac{\tau_3 - \mu_{\tau_3}}{\mathrm{var}(\tau_3)} + \frac{\tau_4 - \mu_{\tau_4}}{\mathrm{var}(\tau_4)}, \tag{16}$$

where values of means ($\mu_{\tau_3}, \mu_{\tau_4}$) and variances ($\mathrm{var}(\tau_3), \mathrm{var}(\tau_4)$) of L-skewness τ_3 and L-kurtosis τ_4 may be obtained by simulation. The T_{Lmom} is approximately chi-squared distribution with two degrees of freedom.

2.8. The Brys-Hubert-Struyf & Bonett-Seier test

The Brys-Hubert-Struyf & Bonett-Seier test $T_{MC-LR} - T_w$ is omnibus test for normality proposed in (Romão, Delgado and Costa, 2010) as combination of two tests: the Bonett-Seier test (Bonett and Seier, 2002) and the Brys–Hubert–Struyf $MC\text{˘}LR$ test (Brys, Hubert and Struyf, 2007). The former is a kurtosis associated test, the latter is a skewness-based test. The statistic of the Bonett-Seier test is defined as

$$T_w = \frac{(\hat{\omega} - 3)\sqrt{n+2}}{3.54}, \tag{17}$$

where

$$\hat{\omega} = 13.29 \left[\ln \sqrt{m_2} - \ln \left(\frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}| \right) \right],$$

in which m_2 is a sample second central moment. Statistics T_w approximately follows a standard normal distribution, and consequently null hypothesis is rejected for both small and large values of T_w. The Bonett-Seier statistic is a simple transformation of Geary's measure of kurtosis (Geary, 1936) , which is defined as τ/σ, where $\tau = E(|X-\mu|)$. After transformation (17), like its Pearson's counterpart (given by equation (4)), Geary's measure of kurtosis equals 3 under normality and increases without bound with increasing leptokurtosis.

The Brys–Hubert–Struyf MC–LR T_{MC-LR} test is given by

$$T_{MC-LR} = n(\nu - \mathrm{v})'V^{-1}(\nu - \mathrm{v}), \tag{18}$$

where v is the vector of robust measures of skewness $[MC, LMC, RMC]'$, and ν, V are estimates based on the distribution under null hypothesis. In the case of normal distribution ν, V are given by

$$\nu = [0, 0.199, 0.199]', \qquad V = \begin{bmatrix} 1.25 & 0.323 & -0.323 \\ 0.323 & 2.62 & -0.0123 \\ -0.323 & -0.0123 & 2.62 \end{bmatrix}.$$

T_{MC-LR} statistic approximately follows the chi-square distribution with three degrees of freedom.

The first element of vector of v is medcouple, proposed in Brys et al. (2004), defined as

$$MC = \underset{X_{(i)} \le \mathrm{med}_F \le X_{(j)}}{\mathrm{med}} h\left(X_{(i)}, X_{(j)}\right), \tag{19}$$

where med is the median, h is the kernel function given by

$$h\left(X_{(i)}, X_{(j)}\right) = \frac{\left(X_{(j)} - \mathrm{med}_F\right) - \left(\mathrm{med}_F - X_{(i)}\right)}{X_{(i)} - X_{(j)}}.$$

Medcouple is a robust skewness measure bounded by $[-1, 1]$.

The two other elements of vector v are the left medcouple (LMC) and the right medcouple (RMC), the left and right tail weight measure, proposed in Brys et al. (2006). LMC and RMC are respectively defined as

$$LMC = -MC(x < \mathrm{med}_F), \qquad RMC = MC(x > \mathrm{med}_F). \tag{20}$$

Like medcouple, they are robust against outlying values. These three measures have great advantage that can be computed at any distribution, even when finite moments do not exist (Brys, Hubert and Struyf, 2007).

The joint test $T_{MC-LR} - T_w$ proposed by Romão, Delgado and Costa (2010) is based on the assumption that individual tests can be considered independent. This assumption was positively verified in simulation study of 200,000 samples of size 100 drawn from a standard normal distribution. In order to control the overall type I error at the nominal level α, the normality hypothesis of the data is rejected for the joint test when rejection is obtained for either one of the two individual tests for a significance level of $\alpha/2$ (Romão, Delgado and Costa, 2010).

2.9. Desgagnéa and Lafaye de Micheaux test

Desgagnéa and Lafaye de Micheaux (2018) has recently proposed new alternatives to the classical Pearson's measures of skewness and kurtosis, which they termed 2nd-power skewness and kurtosis. They used them to build two tests of normality. First test X^a_{APD} can be derived as the Lagrange multiplier test on the asymmetric power distribution (APD) class, introduced by Komunjer (2007). This class of distribution is a generalization of the generalized power distribution (GPD) (also known as the generalized error distribution (GED)), which is symmetric, to a broader class that includes asymmetric distributions. The APD class encompasses all GPD distributions (i.e. the Laplace distribution, normal distribution) and asymmetric distributions (i.e. asymmetric Laplace distribution, split normal distribution).

The basis of this test are 2nd-power skewness B_2 and 2nd-power kurtosis K_2, which are defined as

$$B_2 = \frac{1}{n}\sum_{i=1}^{n} Z_i^2 \text{sign}(Z_i), \qquad \text{and} \qquad K_2 = \frac{1}{n}\sum_{i=1}^{n} Z_i^2 \ln(|Z_i|), \tag{21}$$

where $Z_i = (X_i - \bar{X})/S$. This sample statistics are analogous to 2nd-power skewness and kurtosis for a random variable X, which are defined as $E(Z^2\text{sign}(Z))$ and $E(Z^2\ln(Z))$, respectively.

The X^a_{APD} statistics is defined as

$$X^a_{APD} = \frac{nB_2^2}{3-8/\pi} + \frac{n\left(K_2 - (2-\ln 2-\gamma)/2\right)^2}{(3\pi^2-28)/8}, \tag{22}$$

where γ is the Euler–Mascheroni constant. The X^a_{APD} is approximately chi-squared distribution with two degrees of freedom as a sum of squares of two independent standard normals. However, X^a_{APD} has rather poor small sample properties (just as JB statistic). Thus, Desgagnéa and Lafaye de Micheaux (2018) proposed the second statistic X_{APD} defined as

$$X_{APD} = Z^2(B_2) + Z^2(K_2 - B_2), \tag{23}$$

where

$$Z(B_2) = \sqrt{\frac{nB_2^2}{(3-8/\pi)(1-1.9/n)}}$$

is transformed 2nd-power skewness, and

$$Z(K_2 - B_2) = \frac{\sqrt{n}\left[(K_2 - B_2^2)^{1/3} - ((2-\ln 2-\gamma)/2)^{1/3}(1-1.026/n)\right]}{\sqrt{((2-\ln 2-\gamma)/2)^{-4/3}(3\pi^2-28)(1-2.25/n^{0.8})/72}}$$

is transformed 2nd-power net kurtosis. Under the null hypothesis X_{APD} is, with high numerical precision, approximately distributed as chi-squared distribution with two degrees of freedom, for all sample sizes with at least 10 observations. This is a rare and desirable characteristic for normality test statistic based on measures of the moments. In the simulation study we use only X_{APD} statistics.

3. Simulation study

In our simulation study we considered three levels of significance: $\alpha = 0.01$, 0.05 and 0.10, and five different sample sizes: n = 10, 20, 50, 100, 500. First, appropriate critical values were obtained for each test based on 100,000 simulated samples from a standard normal distribution. We decided to use empirical rather than approximated limit distributions, because many preavious studies emphasized that in the case of Jarque-Berra test and their modifications chi-squared distribution approximation of the limit distribution did not work well, even for large sample sizes (Thadewald and Büning, 2007 and Romão, Delgado and Costa, 2010).

In order to investigate the power of the various tests a total of 10,000 samples of the appropriate size were drawn from each of 15 different non-normal distributions. These distributions are categorized as symmetric short-tailed, symmetric long-tailed and asymmetric in shape (the same categories were considered by Farrell and Rogers-Stewart, 2006). The choice of shape category is based on the values of Pearson's measures of the skewness and kurtosis of the distribution given by the formulas (1) and (2). Specifically, asymmetric distributions have $\sqrt{\beta_1} \neq 0$, symmetric short-tailed $\sqrt{\beta_1} = 0$ and $\beta_2 < 3$ and symmetric long-tailed $\sqrt{\beta_1} = 0$ and $\beta_2 > 3$.

Tables 1, 2, 3 presents results for the first category of alternative distributions, namely symmetrical short-tailed distributions, respectively for three levels of significance: $\alpha = 0.01$, 0.05 and 0.10. Distributions are ordered from the distribution with the lowest kurtosis (the most distinct from normal), to the distribution with the highest kurtosis (the closest to normal). The average power across all short-tailed distributions is presented in Table 4. Firstly, power of normality test for this group of distributions is not sufficient. Especially, at significance level $\alpha = 0.01$ and small samples sizes (below 50), all considered tests perform very poorly. When the significance level and/or sample size increase tests become more powerful. For the smallest sample sizes X_{APD} statistics seems to perform best for most alternative distributions. For moderate and big samples K^2 achieves good power for alternative distributions with low kurtosis, but for distributions with kurtosis more close to normal $T_{MC-LR} - T_w$ statistics performs even better. On the basis of average results, T_{Lmom} tests perform fairly well for moderate and big samples, too. The results show that for symmetrical short-tailed distributions the popular JB statistic performs poorly. From modifications of this statistics DH seems to be the best. It performs quite well for all sample sizes.

Table 1: Empirical power results for symmetrical short-tailed distributions ($\alpha = 0.01$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR}-T_w$	X_{APD}
Uniform (a=0,b=1)	10	0.24	0.24	0.3	1.72	0.21	0.26	0.35	0.67	**3.44**
$\sqrt{\beta_1}=0, \beta_2=1.8$	20	1.32	0	0	1.44	0	0.03	2.81	1.02	**5.62**
	50	**47.5**	0	0	8.8	0	0	41.96	16.57	35.26
	100	**96.72**	0	0	62.14	0	1	90.25	55.82	84.82
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Tukey (l=0.8)	10	0.33	0.33	0.3	1.97	0.29	0.33	0.35	0.74	**3.46**
$\sqrt{\beta_1}=0, \beta_2=1.86$	20	1.44	0.01	0	1.52	0	0	2.81	0.97	**6.1**
	50	**46.94**	0	0	8.41	0	0	41.01	16.25	33.71
	100	**97.03**	0	0	61.76	0	1.29	90.47	56.12	84.77
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Tukey (l=0.3)	10	0.39	0.4	0.51	0.73	0.37	0.39	0.47	0.92	**0.99**
$\sqrt{\beta_1}=0, \beta_2=2.41$	20	0.13	0.08	0	0.3	0.07	0.09	0.21	0.7	**0.58**
	50	1.07	0	0	0.13	0	0	0.97	**1.28**	1.14
	100	**4.46**	0	0	0.33	0	0	3.05	2.25	2.86
	500	**78.89**	10.99	8.8	43.61	2.33	0.52	0.34	28.93	57.35
Beta (a=5,b=5)	10	0.51	0.51	0.5	0.72	0.51	0.5	0.6	0.88	**0.91**
$\sqrt{\beta_1}=0, \beta_2=2.53$	20	0.14	0.13	0.4	0.29	0.13	0.16	0.28	**0.79**	0.42
	50	0.37	0	0	0.11	0.01	0	0.59	**0.9**	0.64
	100	**1.97**	0	0	0.26	0	0	1.72	1.41	1.54
	500	**34.41**	1.25	0.9	10.86	0.18	0.02	100	10.96	22.58
Beta (a=10,b=10)	10	0.7	0.68	0.7	0.87	0.72	0.72	0.7	**0.97**	0.85
$\sqrt{\beta_1}=0, \beta_2=2.73$	20	0.29	0.3	1	0.47	0.34	0.3	0.49	**0.71**	0.49
	50	0.4	0.18	0	0.39	0.2	0.19	0.62	**0.75**	0.52
	100	0.65	0.06	0	0.25	0.08	0.09	0.71	**0.88**	0.61
	500	**3.66**	0.12	0.1	1	0.03	0	100	2.39	3.14

Table 2: Empirical power results for symmetrical short-tailed distributions ($\alpha = 0.05$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR} - T_w$	X_{APD}
Uniform (a=0,b=1)	10	2.43	1.86	1.6	6.58	1.67	2.33	5.7	4.48	**9.75**
$\sqrt{\beta_1} = 0, \beta_2 = 1.8$	20	13.3	0.21	0.1	10.08	0.18	3.65	**20.18**	8.45	19.58
	50	**77.79**	1.16	0	45.96	0.02	44.53	70.41	38.82	62.68
	100	**99.68**	75.63	53.6	95.32	1.14	93.76	97.76	80.01	96.52
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Tukey (l=0.8)	10	1.86	1.56	1.6	6.45	1.56	2.27	5.45	4.45	**9.53**
$\sqrt{\beta_1} = 0, \beta_2 = 1.86$	20	12.85	0.25	0.1	9.33	0.24	3.43	**19.11**	7.95	18.44
	50	**78.32**	1.06	0	45.74	0.03	44.21	70.17	38.43	62.45
	100	**99.76**	75.16	53.6	95.49	1.04	93.89	97.75	79.62	96.53
	500	99.99	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Tukey (l=0.3)	10	2.68	2.73	2.4	3.46	2.74	2.97	3.32	**4.39**	4.26
$\sqrt{\beta_1} = 0, \beta_2 = 2.41$	20	2.01	1.01	1	1.77	1.12	1.36	2.9	**4.21**	3.6
	50	**6.4**	0.17	0.2	1.94	0.18	1.44	6.11	6.19	6.25
	100	**17.34**	0.46	0	5.17	0.03	3.87	13.23	10.24	13.1
	500	**94.85**	80.93	76.7	84.53	59.18	62.26	22.64	54.74	82.1
Beta (a=5,b=5)	10	3.48	3.39	3.2	3.52	3.26	3.3	3.8	**4.68**	4.31
$\sqrt{\beta_1} = 0, \beta_2 = 2.53$	20	2.18	1.55	1.9	2.04	1.63	1.97	3.3	**4.35**	3.41
	50	3.8	0.48	0.2	1.54	0.5	1.16	4.01	**4.9**	3.71
	100	**8.77**	0.32	0.36	2.87	0.18	1.66	7.86	7.18	7.69
	500	65.12	38.12	33.6	44.66	21.69	22.26	**100**	28.03	48.96
Beta (a=10,b=10)	10	3.96	3.95	3.9	4.56	4.23	4.07	4.45	**4.77**	4.72
$\sqrt{\beta_1} = 0, \beta_2 = 2.73$	20	3.08	2.84	3	3.39	2.82	3.23	3.84	**4.71**	3.93
	50	3.37	1.73	1.7	2.62	1.56	2.33	4.22	**4.99**	3.87
	100	3.82	0.91	0.9	2.33	0.88	1.57	4.52	**5.36**	4.16
	500	15.67	6	4.3	8.68	2.99	2.6	**100**	8.44	12.76

Table 3: Empirical power results for symmetrical short-tailed distributions ($\alpha = 0.1$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR}-T_w$	X_{APD}
Uniform (a=0,b=1)	10	8.02	4.71	4	11.45	3.61	8.05	13.71	8.91	**15.01**
$\sqrt{\beta_1}=0, \beta_2=1.8$	20	27.79	2.97	0.4	20.8	0.76	22.29	**34.08**	16.87	30.53
	50	**88.96**	50.18	26.6	68.86	0.1	75.85	82.45	54.33	76.84
	100	**99.93**	98.66	97.1	98.96	74.52	99.11	99.15	88.59	98.67
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Tukey (l=0.8)	10	8.36	4.98	4	11.88	3.94	8.5	14.13	9.59	**15.8**
$\sqrt{\beta_1}=0, \beta_2=1.86$	20	27.67	3.45	0.4	21.35	0.85	22.75	**34.44**	17.02	30.69
	50	**88.64**	50.27	56.11	68.72	0.17	74.76	81.96	53.54	76.19
	100	**99.88**	98.46	97.1	98.77	73.63	98.85	98.87	87.71	98.4
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Tukey (l=0.3)	10	6.46	5.88	6.1	7.07	5.95	6.64	7.81	**9.61**	8.26
$\sqrt{\beta_1}=0, \beta_2=2.41$	20	6.06	3.41	3.2	5.08	3.02	5.97	8.09	**9.31**	8
	50	**13.07**	2.15	1.4	6.34	0.83	7.92	12.62	12.24	11.92
	100	30.21	10.37	5.7	14.86	1.27	16.36	**24.41**	17.56	24.39
	500	**98.02**	93.97	92.4	93.83	83.65	85.09	86.61	68.58	90.22
Beta (a=5,b=5)	10	7.21	6.81	7.3	8.02	6.91	7.24	8.17	**8.87**	8.82
$\sqrt{\beta_1}=0, \beta_2=2.53$	20	6.29	4.73	3.9	5.81	4.29	6.66	7.8	**9.26**	7.78
	50	8.43	2.84	1.5	5.13	1.7	6.01	9.82	**10.72**	8.95
	100	**16.64**	5.7	2.6	8.52	1.37	9.01	15.22	13.03	14.44
	500	78.06	63.62	60.5	63.89	48.5	48.3	**100**	40.19	63.25
Beta (a=10,b=10)	10	8.44	8.54	7.8	8.94	8.4	8.94	8.96	**9.99**	9.21
$\sqrt{\beta_1}=0, \beta_2=2.73$	20	7.86	7.28	6.7	7.55	6.98	7.92	8.45	**9.46**	8.66
	50	7.3	4.97	4.7	5.84	4.79	6.23	8.67	**10**	7.76
	100	9.18	4.93	3.6	6.13	3.35	6.19	10.16	**10.6**	9.33
	500	**26.61**	17.69	14.2	19.18	12.41	11.39	11.0	16.08	22.92

Table 4: Average Power for symmetrical short-tailed distributions.

α	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR}-T_w$	X_{APD}
	10	0.43	0.43	0.46	1.20	0.42	0.44	0.49	0.84	**1.93**
	20	0.66	0.10	0.28	0.80	0.11	0.12	1.32	0.84	**2.64**
0.01	50	**19.26**	0.04	0.00	3.57	0.04	0.04	17.03	7.15	14.25
	100	**40.17**	0.01	0.00	24.95	0.02	0.48	37.24	23.30	34.92
	500	63.39	42.47	41.96	51.09	40.51	40.11	**80.07**	48.46	56.61
	10	2.88	2.70	2.54	4.91	2.69	2.99	4.54	4.55	**6.51**
	20	6.68	1.17	1.22	5.32	1.20	2.73	**9.87**	5.93	9.79
0.05	50	**33.94**	0.92	0.42	19.56	0.46	18.73	30.98	18.67	27.79
	100	**45.87**	30.50	21.69	40.24	0.65	38.95	44.22	36.48	43.60
	500	75.13	65.01	62.92	67.57	56.77	57.42	**84.53**	58.24	68.76
	10	7.70	6.18	5.84	9.47	5.76	7.87	10.56	9.39	**11.42**
	20	15.13	4.37	2.92	12.12	3.18	13.12	**18.57**	12.38	17.13
0.1	50	**41.28**	22.08	18.06	30.98	1.52	34.15	39.10	28.17	36.33
	100	**51.17**	43.62	41.22	45.45	30.83	45.90	49.56	43.50	49.05
	500	**80.54**	75.06	73.42	75.38	68.91	68.96	79.52	64.97	75.28

Results for the second category of alternative distributions, symmetrical long-tailed distributions, are presented in Tables 5, 6, 7. Distributions are ordered from the distributions with the highest kurtosis (the most distinct from normal), to the distribution with the lowest kurtosis (the closest to normal). The average power across all long-tailed distributions is presented in Table 8. For this group of distribution, normality tests perform better than for the short-tailed distributions, but for small sample size results are still not very impressive. On the basis of average results, the *RJB* statistic outperforms other tests for almost all sample sizes. It is not surprisingly, bearing in mind that this test is based on the robust estimate of the dispersion. However, when one takes a closer look at particular alternative distributions, one may see that for the distribution with kurtosis closer to three also D'Agostino–Pearson K^2 test performs well. Contrary to the short-tailed distribution, *JB* statistic has quite good power properties, even better than its modifications (apart from *RJB*).

Table 5: Empirical power results for symmetrical long-tailed distributions ($\alpha = 0.01$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR} - T_w$	X_{APD}
Laplace (mu=0,b=1)	10	6.78	6.75	7.25	6.07	7.63	7.02	**7.89**	3.64	5.17
$\sqrt{\beta_1} = 0, \beta_2 = 6$	20	13.68	14.37	14.59	16.18	**18.09**	16.15	17.31	11.2	15.83
	50	28.58	33.84	30.55	34.76	**43.96**	33.56	43.08	35.39	41.95
	100	48.96	60.25	51.85	62.62	74.35	54.43	**74.61**	69.81	73.89
	500	99.82	99.96	99.06	99.96	**100**	99.76	5.43	**100**	**100**
Logistic (mu=0,sigma=1)	10	2.6	2.6	2.27	2.14	2.66	2.65	**2.78**	1.62	1.93
$\sqrt{\beta_1} = 0, \beta_2 = 4.4$	20	5.19	5.37	5.1	5.27	5.46	**5.51**	4.58	3.01	4.6
	50	10.06	**11.71**	9.18	11.7	12.83	11.38	8.4	5.96	11.4
	100	16.54	21.12	20.87	21.61	**23.49**	18.67	14.17	11.31	20.19
	500	67.32	76.3	63.96	76.75	**80.31**	57.87	2.04	60.4	77.76
Student-t (df=10)	10	2.05	2.02	1.1	1.65	2.07	2.02	**2.1**	1.33	1.39
$\sqrt{\beta_1} = 0, \beta_2 = 4$	20	4.08	4.14	3.25	4.03	**4.25**	4.22	3.38	2.59	3.7
	50	7.92	9.09	6.89	8.8	**9.55**	9.03	5.59	3.88	7.99
	100	13.27	16.22	15.94	15.88	**16.65**	14.63	8.28	6.93	13.79
	500	49.63	58.58	48.09	58.55	**60.22**	43.6	1.85	32.93	53.43
Student-t (df=20)	10	1.26	1.26	1.19	1.07	1.4	1.3	**1.46**	1.09	1.06
$\sqrt{\beta_1} = 0, \beta_2 = 3.375$	20	2.23	2.18	1.84	2.24	**2.41**	2.32	2.11	1.6	2.05
	50	3.4	3.92	2.88	3.41	3.84	**3.95**	2.36	1.68	3.1
	100	4.76	5.48	4.11	5.08	**5.52**	5.3	2.8	1.97	4.21
	500	12.99	16.89	16.13	16.58	**17**	12.51	1.33	5.35	11.93
Student-t (df=30)	10	1.19	1.2	1.1	1.18	1.22	1.19	**1.39**	1.11	1.06
$\sqrt{\beta_1} = 0, \beta_2 = 3.23$	20	1.59	1.56	1.49	**1.68**	1.66	1.57	1.38	1.33	1.5
	50	2.36	2.58	2.02	2.42	2.5	**2.59**	1.86	1.42	2.1
	100	3.09	**3.52**	2.79	3.07	3.49	3.27	1.91	1.58	2.74
	500	5.94	8.01	7.9	7.43	**8.02**	6.3	1.16	2.5	5.45

Table 6: Empirical power results for symmetrical long-tailed distributions ($\alpha = 0.05$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR}-T_w$	X_{APD}
Laplace (mu=0,b=1)	10	17.78	17.41	17.88	18.65	**20.31**	17.84	19.59	10.13	16.81
$\sqrt{\beta_1}=0, \beta_2=6$	20	28.6	30.02	29.81	31.74	**35.6**	32.65	33.11	20.29	31.12
	50	49.1	55.81	52.95	56.81	**65.67**	59.69	61.34	50.92	61.1
	100	72.44	79.84	79.92	80.74	**89.12**	84.12	87.6	82.73	87.09
	500	99.98	99.98	**100**	99.98	**100**	**100**	**100**	**100**	**100**
Logistic (mu=0,sigma=1)	10	9.61	9.51	7.26	8.47	**9.79**	9.45	8.85	6.49	7.81
$\sqrt{\beta_1}=0, \beta_2=4.4$	20	13.77	14.55	12.27	14.02	**14.94**	14.17	12.29	8.17	12.42
	50	22.82	26.48	20.97	25.76	**27.79**	26.36	19.86	13.78	23.38
	100	33.51	39.63	31.98	38.41	**42.39**	39.37	29.41	22.16	36.32
	500	84.56	89.04	83.77	88.88	**91.41**	87.18	79.9	75.51	88.98
Student-t (df=10)	10	**8.84**	8.69	7.34	8.13	**8.84**	8.77	8.21	5.73	7.51
$\sqrt{\beta_1}=0, \beta_2=4$	20	11.67	12.05	9.78	11.74	**12.37**	11.73	9.89	7.14	10.46
	50	18.07	**20.86**	15.14	19.71	21.3	20.4	14.5	10.65	17.35
	100	25.87	30.47	22.96	29.17	**32.09**	29.82	19.94	15.39	25.94
	500	68.91	75.09	67.18	74.33	**76.3**	71.74	71.7	49.77	70.31
Student-t (df=20)	10	**6.79**	6.76	6.21	6.42	6.75	6.6	6.44	5.33	6.27
$\sqrt{\beta_1}=0, \beta_2=3.375$	20	7.53	7.68	8.78	7.69	**7.8**	7.57	6.96	5.45	7.2
	50	9.76	11.03	10.85	10.81	**11.37**	10.84	8.25	6.89	9.46
	100	12.82	15.3	14.34	14.27	**15.69**	14.93	9.84	7.81	12.22
	500	27.59	33.33	33.11	31.91	**33.64**	31.11	6.14	14.3	26.19
Student-t (df=30)	10	**6.06**	6.01	5.52	5.47	5.78	5.89	5.56	5.19	5.47
$\sqrt{\beta_1}=0, \beta_2=3.23$	20	**6.87**	6.84	6.66	6.75	6.64	6.65	6.38	5.89	6.53
	50	8.45	**9.11**	7.45	8.36	8.97	8.81	6.87	5.8	7.78
	100	9.83	**11.22**	10.78	10.52	11.37	11.01	7.55	6.67	9.01
	500	16.56	20.51	19.87	19.05	**20.52**	19.77	5.41	8.76	15.39

Table 7: Empirical power results for symmetrical long-tailed distributions ($\alpha = 0.1$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR}-T_w$	X_{APD}
Laplace (mu=0,b=1)	10	26.27	26.68	28.4	27.14	**29.13**	26.62	26.77	14.94	25.23
$\sqrt{\beta_1}=0, \beta_2=6$	20	38.91	41.07	42.21	42.32	**47.49**	41.21	42.44	27.4	41.93
	50	60.4	65.4	63.49	66.96	**75.8**	68.95	70.95	59.36	71.01
	100	81.62	86.26	86.07	86.87	**93.13**	89.76	91.72	87.47	91.49
	500	**100**	**100**	**100**	**100**	100	**100**	**100**	**100**	**100**
Logistic (mu=0,sigma=1)	10	15.5	15.7	13.77	15.55	**16.34**	15.15	14.97	10.94	14.32
$\sqrt{\beta_1}=0, \beta_2=4.4$	20	20.99	22.19	23.03	21.88	**23.49**	21.34	19.62	14.46	20.61
	50	31.14	34.57	36.15	34.59	**37.44**	33.58	27.54	20.43	31.17
	100	43.03	48.55	49.44	48.39	**52.66**	49.28	39.6	30.7	45.74
	500	89.97	92.15	88.62	92.39	**94.26**	92.78	14.61	82.51	92.75
Student-t (df=10)	10	14.7	14.85	12.37	13.84	**14.89**	14.3	13.35	10.3	12.6
$\sqrt{\beta_1}=0, \beta_2=4$	20	18.69	19.6	15.94	19.06	**20.39**	18.66	17.3	13.13	17.65
	50	26.42	**28.96**	22.84	28.56	30.8	28.19	22.43	17.11	25.75
	100	35.53	39.91	31.58	39.49	**42.58**	39.6	28.82	22.32	35.33
	500	77.97	81.46	73.47	81.44	**82.82**	81.47	13.64	59.23	78.38
Student-t (df=20)	10	12.43	12.62	11.22	11.95	**12.74**	12.18	11.53	10.55	11.68
$\sqrt{\beta_1}=0, \beta_2=3.375$	20	13.28	13.97	12.26	13.51	**14.34**	13.59	12.77	10.95	12.91
	50	15.82	17.6	15.15	17.46	**18.3**	16.54	13.88	11.88	15.35
	100	19.84	22.25	21.3	21.83	**23.33**	21.78	15.82	13.52	18.98
	500	36.61	40.86	40.29	40.76	**42.11**	41.29	11.9	21.48	34.71
Student-t (df=30)	10	**11.16**	10.88	10.47	10.63	11.06	10.71	10.5	10.47	10.4
$\sqrt{\beta_1}=0, \beta_2=3.23$	20	**12.91**	12.94	11.53	12.6	12.75	12.32	12.03	10.34	12.17
	50	13.97	14.98	14.89	14.58	**15.4**	14.22	12.64	11.03	13.83
	100	16.42	18.04	17.1	17.76	**18.83**	17.88	13.83	11.65	15.84
	500	23.86	26.97	27.43	26.6	28.16	**28.67**	10.73	14.52	21.97

Table 8: Average Power for symmetrical long-tailed distribution.

α	n	Goodness-of-fit tests								
		n	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR}-T_w$	X_{APD}
	10	2.78	2.77	2.58	2.42	3.00	2.84	**3.12**	1.76	2.12
	20	5.35	5.52	5.25	5.88	**6.37**	5.95	5.75	3.95	5.54
0.01	50	10.46	12.23	10.30	12.22	**14.54**	12.10	12.26	9.67	13.31
	100	17.32	21.32	19.11	21.65	**24.70**	19.26	20.35	18.32	22.96
	500	47.14	51.95	47.03	51.85	**53.11**	44.01	2.36	40.24	49.71
	10	9.82	9.68	8.84	9.43	**10.29**	9.71	9.73	6.57	8.77
	20	13.69	14.23	13.46	14.39	**15.47**	14.55	13.73	9.39	13.55
0.05	50	21.64	24.66	21.47	24.29	**27.02**	25.22	22.16	17.61	23.81
	100	30.89	35.29	32.00	34.62	**38.13**	35.85	30.87	26.95	34.12
	500	59.52	63.59	60.79	62.83	**64.37**	61.96	52.63	49.67	60.17
	10	16.01	16.15	15.25	15.82	**16.83**	15.79	15.42	11.44	14.85
	20	20.96	21.95	20.99	21.87	**23.69**	21.42	20.83	15.26	21.05
0.1	50	29.55	32.30	30.50	32.43	**35.55**	32.30	29.49	23.96	31.42
	100	39.29	43.00	41.10	42.87	**46.11**	43.66	37.96	33.13	41.48
	500	65.68	68.29	65.96	68.24	**69.47**	68.84	30.18	55.55	65.56

The results for the last category of alternative distributions, asymmetric distributions, are presented in Tables 9, 10, 11. First three distributions are skewed to the right (ordered from the highest skewness to the most close to zero), and the rest two distributions are left-skewed (one with low skewness and one with close to zero). The average results power across all asymmetric distributions is presented in Table 12. For asymmetric distributions normality tests have much more power than in case of symmetric distributions. The results do not show one particular test that outperforms the rest. The results vary widely depending on the type of asymmetry, sample size and significance level. For lognormal distribution (strongly right-skewed) BM_{3-6} and T_{Lmom} perform the best. From modifications of JB statistic, DH also performs well. However, for big sample sizes (100 and 500) almost all statistics have 100% power. For distributions with weaker right asymmetry BM_{3-6} and DH are the most powerful tests. As far as distributions skewed to the left are concerned, T_{Lmom}, X_{APD} and BM_{3-6} perform the best. Contrary to the left asymmetry, DH test is not better than the standard JB test.

Table 9: Empirical power results for asymmetric distributions ($\alpha = 0.01$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR} - T_w$	X_{APD}
Lognormal (logmean=0,logsd=1)	10	28.03	28.19	29.67	**34.12**	28.42	29.96	29.03	7.86	28.56
$\sqrt{\beta_1} = 6.18, \beta_2 = 5.22$	20	59.66	58.57	51.78	**75.48**	60.17	66.11	74.95	28.88	72.18
	50	96.97	96.58	96.98	99.82	96.44	97.32	**99.9**	81.31	99.77
	100	99.99	99.99	**100**	**100**	99.99	99.99	**100**	99.3	**100**
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Gumbel (mu=1,sigma=1)	10	5.19	5.2	5.23	4.72	5.07	**5.32**	4.56	1.86	3.93
$\sqrt{\beta_1} = 1.14, \beta_2 = 5.4$	20	13.05	12.68	12.12	11.91	11.98	**13.35**	12.18	3.79	11.71
	50	37.17	34.85	39.12	40.46	33.79	31.44	**41.93**	9.31	41.83
	100	71.42	67.65	65.56	**80.97**	66.12	55.62	80.18	17.91	80.55
	500	**100**	**100**	**100**	**100**	**100**	99.99	**100**	88.42	**100**
Chi-squared (df=20)	10	1.92	1.92	1.8	1.67	1.88	**2**	1.7	1.07	1.51
$\sqrt{\beta_1} = 0.63, \beta_2 = 3.6$	20	4.54	4.42	4.25	3.58	3.98	**4.55**	3.79	1.4	3.81
	50	12.14	11.2	12.1	11.3	10.5	9.35	12.45	2.09	**12.57**
	100	26.75	23.85	29.23	31.15	22.42	16.44	30.72	3.09	**31.37**
	500	99.16	98.69	98.8	99.74	98.36	82.91	**100**	18.58	99.56
Weibull (shape=1,scale=1)	10	15.58	15.74	15.79	18.16	14.72	**16.56**	14.51	2.97	14.34
$\sqrt{\beta_1} = -2.85, \beta_2 = 6$	20	36.05	35.01	35.56	**51.54**	34.1	40.76	50.59	9.71	47.35
	50	81.53	79.42	8.33	97.42	77.54	81.33	**98.41**	43.35	96.79
	100	99.59	99.43	**100**	**100**	98.54	98.84	**100**	85.74	**100**
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Beta (a=10,b=5)	10	1.05	**1.07**	1.05	1.05	0.98	**1.07**	0.97	0.93	1.04
$\sqrt{\beta_1} = -0.33, \beta_2 = 2.82$	20	1.12	1.05	1.13	0.94	0.98	1.1	1.11	0.76	**1.21**
	50	1.64	1.25	1.32	1.7	1.15	1.11	**2.78**	1.2	2.4
	100	3.21	1.89	2.59	5.03	1.81	1.16	**7.21**	1.45	6.41
	500	61.14	47.14	48.1	79.87	39.99	5.05	**100**	6.77	75.83

Table 10: Empirical power results for asymmetric distributions ($\alpha = 0.05$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR} - T_w$	X_{APD}
Lognormal (logmean=0,logsd=1)	10	46.59	50.03	50.09	51.3	47.31	**54.21**	49.18	14.55	50.7
$\sqrt{\beta_1} = 6.18, \beta_2 = 5.22$	20	78.84	82.11	83.13	88.6	78.4	87.96	**90.14**	40.9	88.16
	50	99.58	99.77	99.79	99.95	99.29	99.88	**99.98**	90.4	99.92
	100	**100**	**100**	**100**	**100**	**100**	**100**	**100**	99.78	**100**
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Gumbel (mu=1,sigma=1)	10	15.18	15.67	15.47	12.81	14.23	**16.22**	12.92	5.93	12.74
$\sqrt{\beta_1} = 1.14, \beta_2 = 5.4$	20	27.5	28.7	28.82	26.18	26.07	**29.99**	27.15	8.45	26.68
	50	58.49	62.06	62.55	65	56.83	62.19	63.47	17.87	**64.02**
	100	88.44	90.53	90.44	**93.63**	86.83	89.21	91.8	31.85	92.55
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	95.62	**100**
Chi-squared (df=20)	10	9.03	9.24	9.17	7.82	8.99	**9.44**	8.25	4.83	8.06
$\sqrt{\beta_1} = 0.63, \beta_2 = 3.6$	20	13.38	13.56	13.58	11.93	12.44	**13.82**	12.22	5.43	12.57
	50	26.96	28.42	28.6	**28.79**	25.04	28.16	27.35	7.64	28.16
	100	50.69	53.27	52.47	**58.46**	47.79	49.58	54.21	9.47	56.14
	500	99.95	99.96	**100**	**100**	99.95	99.62	**100**	36.75	99.94
Weibull (shape=1,scale=1)	10	30.27	33.02	34.28	34.36	30.05	**36.53**	31.86	7.11	33.9
$\sqrt{\beta_1} = -2.85, \beta_2 = 6$	20	57.74	62.34	63.85	73.64	55.47	71.49	**77.38**	17.7	72.76
	50	95.36	97.86	97.93	99.63	93.29	99.02	**99.75**	64.34	99.45
	100	**100**	**100**	**100**	**100**	99.98	**100**	**100**	94.65	**100**
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Beta (a=10,b=5)	10	4.92	4.94	5.67	5.18	4.95	5.27	5.17	4.86	**5.73**
$\sqrt{\beta_1} = -0.33, \beta_2 = 2.82$	20	5.76	5.53	5.32	5.17	4.74	5.85	6.1	4.15	**6.37**
	50	8.87	8.04	8.55	9.24	6.15	8.15	**11.01**	5.4	10.35
	100	15.71	13.48	15.19	20.57	10.31	13.54	21.52	6.4	**21.55**
	500	92.18	91.79	92.79	95.47	88.27	77.92	**100**	18.51	92.31

Table 11: Empirical power results for asymmetric distributions ($\alpha = 0.1$).

Alternative	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR}-T_w$	X_{APD}
Lognormal (logmean=0,logsd=1)	10	58.41	64.63	62.95	62.3	58.35	**67.68**	61.86	19.56	62.47
$\sqrt{\beta_1} = 6.18, \beta_2 = 5.22$	20	86.56	92.12	92.15	93.38	86.66	**93.99**	94.34	50.08	93.31
	50	99.94	99.99	**100**	**100**	99.82	**100**	**100**	93.72	99.99
	100	**100**	**100**	**100**	**100**	**100**	**100**	**100**	99.89	**100**
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Gumbel (mu=1,sigma=1)	10	22.7	24.21	23.79	20.66	22.54	**24.39**	20.93	10.51	21.25
$\sqrt{\beta_1} = 1.14, \beta_2 = 5.4$	20	37.46	41.26	41.23	37.63	37.25	**41.77**	37.79	14.32	38.18
	50	69.92	75.73	75.45	**75.8**	69.93	74.47	73.31	25.53	74.18
	100	94.41	96.18	96.24	**96.8**	94.5	95.21	95.56	41.23	95.96
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	97.5	**100**
Chi-squared (df=20)	10	14.43	14.89	14.6	13.46	14.04	**14.95**	13	9.78	13.42
$\sqrt{\beta_1} = 0.63, \beta_2 = 3.6$	20	20.49	22.54	22.53	19.81	20.71	**22.59**	20.17	10.45	20.65
	50	38.17	**42.94**	41.51	41.5	37.87	40.94	39.35	13.29	40.52
	100	65.83	70.92	70.75	**71.66**	65.41	66.67	67.54	16.5	69.47
	500	99.99	99.99	99.99	99.99	99.99	99.96	**100**	50	99.95
Weibull (shape=1,scale=1)	10	41.83	48.58	57.59	45.36	41.31	**52.26**	45.64	11.87	46.29
$\sqrt{\beta_1} = -2.85, \beta_2 = 6$	20	70.42	80.12	80.55	83.32	69.43	84.29	**86.4**	26.37	82.75
	50	99.19	99.79	99.77	99.91	98.12	99.83	**99.94**	73.9	99.89
	100	**100**	**100**	**100**	**100**	**100**	**100**	**100**	96.77	**100**
	500	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Beta (a=10,b=5)	10	9.52	9.78	9.69	9.59	9.12	9.89	9.65	9.46	**10.1**
$\sqrt{\beta_1} = -0.33, \beta_2 = 2.82$	20	11.07	10.82	10.37	10.35	9.86	**12.1**	11.82	9.54	11.81
	50	16.8	16.17	16.25	16.54	12.95	16.99	**19.27**	10.66	18.44
	100	29.52	30.31	31.1	34.1	22.9	28.58	34.2	12.03	**34.25**
	500	97.11	97.41	98.23	98.24	96.41	92.39	**100**	28.01	96.23

Table 12: Average Power for asymmetric distributions.

α	n	Goodness-of-fit tests								
		K^2	JB	ALM	DH	RJB	BM_{3-6}	T_{Lmom}	$T_{MC-LR}-T_w$	X_{APD}
	10	10.35	10.42	10.71	**11.94**	10.21	10.98	10.15	2.94	9.88
	20	22.88	22.35	20.97	**28.69**	22.24	25.17	28.52	8.91	27.25
0.01	50	45.89	44.66	31.57	50.14	43.88	44.11	**51.09**	27.45	50.67
	100	60.19	58.56	59.48	63.43	57.78	54.41	63.62	41.50	**63.67**
	500	92.06	89.17	89.38	95.92	87.67	77.59	**100.00**	62.75	95.08
	10	21.20	22.58	22.94	22.29	21.11	**24.33**	21.48	7.46	22.23
	20	36.64	38.45	38.94	41.10	35.42	41.82	**42.60**	15.33	41.31
0.05	50	57.85	59.23	59.48	**60.52**	56.12	59.48	60.31	37.13	60.38
	100	70.97	71.46	71.62	**74.53**	68.98	70.47	73.51	48.43	74.05
	500	98.43	98.35	98.56	99.09	97.64	95.51	**100.00**	70.18	98.45
	10	29.38	32.42	33.72	30.27	29.07	**33.83**	30.22	12.24	30.71
	20	45.20	49.37	49.37	48.90	44.78	**50.95**	50.10	22.15	49.34
0.1	50	64.80	**66.92**	66.60	66.75	63.74	66.45	66.37	43.42	66.60
	100	77.95	79.48	79.62	**80.51**	76.56	78.09	79.46	53.28	79.94
	500	99.42	99.48	99.64	99.65	99.28	98.47	**100.00**	75.10	99.24

4. Conclusions

In this study, we performed a comprehensive investigation of nine tests for normality based on measures of the moments. In a simulation study we focused on a different forms of shape departure from normality such as symmetric short-tailed, symmetric long-tailed, or asymmetric. None of the tests considered in this study is uniformly most powerful for all types of alternative distributions, sample sizes and significance levels considered. If the distribution is symmetric and short-tailed two test are the most powerful, Desgagnéa and Lafaye de Micheaux's X_{APD} test and D'Agostino and Pearson's K^2 . Gel and Gastwirth's RJB test is one of the most powerful tests for normality based on measures of the moments across a wide array of symmetrical and long-tailed alternative distributions. For the last category of alternative distributions, asymmetric distributions, it is difficult to distinguish one test. Bontemps-Meddahi's BM_{3-6} for right-skewed distributions and Hosking's T_{Lmom} for left-skewed perform fairly well.

The JB test performs well for symmetric distributions with long tails and for slightly skewed distributions with long tails. However, the power of the JB test is very poor for distributions with short tails. As Thadewald and Büning (2007) reported the Urzùa test has no improvement of power to the classical JB test in the case of Monte Carlo simulated critical values. Gel and Gastwirth modification of JB that uses a robust estimate of the dispersion seems to be the best modification in the case distributions with long tails and Doornik–Hansen modification in the case of short-tailed distributions.

Finally, the authors would like to indicate two tests that have quite reasonable power for all alternative distributions and have advantage of being very closely approximated by chi-squared distribution with two degrees of freedom. These two test are the Doornik–Hansen test and the Desgagnéa and Lafaye de Micheaux test. As a concluding remark, practitioners should carefully act when graphical techniques such as histogram or moment statistics suggest that the sample comes from symmetric distribution. In this case, normality tests do not perform well for small sample sizes (below 50), especially when symmetry is accompanied by short tail of distribution.

Acknowledgement

This paper was presented on the MSA 2019 conference, which financed its publication. Organization of the international conference "Multivariate Statistical Analysis 2019" (MSA 2019) was supported from resources for popularization of scientific activities of the Minister of Science and Higher Education in the framework of agreement No. 712/P-DUN/202019.

References

BONETT, D., SEIER, E., (2002). A test of normality with high uniform power. *Computational Statistics & Data Analysis*, 40(3), pp. 435–445.

BONTEMPS, C., MEDDAHI, N., (2005). Testing Normality: A GMM Approach. *Journal of Econometrics*, 124(1), pp. 149–186. doi:10.1016/j.jeconom.2004.02.014.

BOWMAN, K. O., SHENTON, L. R., (1975). Omnibus test contours for departures from normality based on b1 and b2. *Biometrika*, 62(2), pp. 243–250.

BRYS, G., HUBERT, M. and STRUYF, A., (2006). Robust measures of tail weight. *Computational Statistics & Data Analysis*, 50(3), pp. 733–759. doi:10.1016/j.csda.2004.09.012.

BRYS, G., HUBERT, M. and STRUYF, A., (2004). A Robust Measure of Skewness. *Journal of Computational and Graphical Statistics*, 13(4), pp. 996–1017. doi:10.1198/106186004x12632.

BRYS, G., HUBERT, M. and STRUYF, A., (2007). Goodness-of-fit tests based on a robust measure of skewness. *Computational Statistics*, 23(3), pp. 429–442. doi:10.1007/s00180-007-0083-7.

D'AGOSTINO, R. B., PEARSON, E. S., (1973). Tests for Departure From Normality. Empirical Results for the Distributions of b2 and p b1. *Biometrika*, 60, pp. 613–622. doi:10.1093/biomet/60. 3.613.

D'AGOSTINO, R., BELANGER, A. and D'AGOSTINO, R. JR., (1990). A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician*, 44(4), p. 316. doi:10.2307/2684359.

DOMAŃSKI, C. K., (2010). Uwagi o testach Jarque-Bera, *Przegląd Statystyczny*, 57(4), pp. 19–26.

DOMAŃSKI, C. K., JĘDRZEJCZAK, A., (2017). Consistency Tests Based on Moments. *Annales Universitatis Mariae Curie-Skłodowska, sectio H, Oeconomia*, 50(4), pp. 89–99. doi:dx.doi.org/10.17951/h.2016.50.4.89.

DOORNIK, J. A., HANSEN, H., (2008). An Omnibus Test for Univariate and Multivariate Normality. *Oxford Bulletin of Economics and Statistics*, 70, pp. 927–939. doi:10.1111/j.1468-0084. 2008.00537.x.

FARRELL, P. J., ROGERS-STEWART, K., (2006). Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. *Journal of Statistical Computation and Simulation*, 76(9), pp. 803–816. doi:10.1080/10629360500109023.

FISHER, R. A., (1930), The moments of the distribution for normal samples of measures of departure from normality, *Proceedings of the Royal Statistical Society A*, 130, pp. 16–28.

GEARY, R., (1936). Moments of the Ratio of the Mean Deviation to the Standard Deviation for Normal Samples. *Biometrika*, 28(3/4), p. 295. doi:10.2307/2333953.

GEL, Y. R., GASTWIRTH, J. L., (2008). A Robust Modification of the Jarque-Bera Test of Normality. *Economics Letters*, 99(1), pp. 30–32. doi:10.1016/j.econlet.2007.05.022.

HOSKING, J. R. M., (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1), pp. 105–124.

JARQUE, C. M., BERA, A. K., (1987). A Test for Normality of Observations and Regression Residuals. International Statistical Review. *Revue International de Statistique*, 55(2), pp. 163–172. doi:10.2307/1403192.

JOHNSON, N. L., (1949). Bivariate Distributions Based on Simple Translation Systems . *Biometrika*. 36 (3/4), pp. 297–304. doi:10.1093/biomet/36.3-4.297.

KOMUNJER, I., (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, 22(5), pp. 891–921. doi:10.1002/jae.961.

PIONTEK, K., (2007). Pomiar i testowanie skośności rozkładów stóp zwrotu instrumentów finansowych. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu. Taksonomia*, 14, pp. 122–130.(in Polish).

ROMÃO, X., DELGADO, R. and COSTA, A., (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation*, 80(5), pp. 545–591. doi:10.1080/00949650902740824.

THADEWALD, T., BÜNING, H., (2007). Jarque–Bera Test and its Competitors for Testing Normality – A Power Comparison. *Journal of Applied Statistics*, 34(1), pp. 87–105. doi:10.1080/02664760600994539.

STEIN, C., (1972). A bound for the error in the normal approximation to the distribution of a sum of dependant random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematics, Statistics and Probability*, Vol. 2, pp. 583–602.

URZÙA, C., (1996). On the correct use of omnibus tests for normality. *Economics Letters* 53, pp. 247–251.doi:10.1016/S0165-1765(96)00923-8.

WIJEKULARATHNA, D. K., MANAGE, A. B. W. and SCARIANO, S. M. (2019). Power analysis of several normality tests: A Monte Carlo simulation study. *Communications in Statistics - Simulation and Computation*, pp. 1–17. doi:10.1080/03610918.2019.1658780.

WILSON E. B., HILFERTY M. M., (1931). The Distribution of Chi-square. *Proceedings of the National Academy of Sciences of the United States of America* Vol. 17, No. 12, pp. 684–688. doi:10.1073/pnas.17.12.684.

YAP, B. W., SIM, C. H., (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), pp. 2141–2155.

ZGHOUL, A. A., (2010). A Goodness of Fit Test for Normality Based on the Empirical Moment Generating Function. *Communications in Statistics – Simulation and Computation*, 39(6), pp. 1292–1304.