Thadewald, Thorsten; Büning, Herbert

Working Paper
Jarque-Bera test and its competitors for testing normality: A power comparison


Provided in Cooperation with:
Free University Berlin, School of Business & Economics

Suggested Citation: Thadewald, Thorsten; Büning, Herbert (2004) : Jarque-Bera test and its competitors for testing normality: A power comparison, School of Business & Economics Discussion Paper: Economics, No. 2004/9

This Version is available at:
http://hdl.handle.net/10419/49919

Standard-Nutzungsbedingungen:
Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen. Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes. You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Jarque-Bera Test and its Competitors
for Testing Normality - A Power Comparison

Thorsten Thadewald & Herbert Büning
Jarque-Bera Test and its Competitors for Testing Normality - A Power Comparison

Thorsten Thadewald & Herbert Büning *

Institute for Statistics and Econometrics, Free University Berlin,
Boltzmannstr. 20, 14195 Berlin, Germany

Abstract

For testing normality we investigate the power of several tests, first of all, the well known test of Jarque and Bera (1980) and furthermore the tests of Kuiper (1960) and Shapiro and Wilk (1965) as well as tests of Kolmogorov-Smirnov and Cramér-von Mises type. The tests on normality are based, first, on independent random variables (model I) and, second, on the residuals in the classical linear regression (model II). We investigate the exact critical values of the Jarque-Bera test and the Kolmogorov-Smirnov and Cramér-von Mises tests, in the latter case for the original and standardized observations where the unknown parameters \( \mu \) and \( \sigma \) have to be estimated. The power comparison is carried out via Monte Carlo simulation assuming the model of contaminated normal distributions with varying parameters \( \mu \) and \( \sigma \) and different proportions of contamination. It turns out that for the Jarque-Bera test the approximation of critical values by the chi-square distribution does not work very well. The test is superior in power to its competitors for symmetric distributions with medium up to long tails and for slightly skewed distributions with long tails. The power of the Jarque-Bera test is poor for distributions with short tails, especially if the shape is bimodal, sometimes the test is even biased. In this case a modification of the Cramér-von Mises test or the Shapiro-Wilk test may be recommended.

Key words: goodness-of-fit tests, tests of Kolmogorov-Smirnov and Cramér-von Mises type, Shapiro-Wilk test, Kuiper test, skewness, kurtosis, contaminated normal distribution, Monte-Carlo simulation, critical values, power comparison

JEL classification: C12; C15

* Tel.: +49 30 838 52896; fax: +49 30 838 54142
Email address: hbuening@wiwiss.fu-berlin.de (Herbert Büning).
Goodness-of-fit tests play an important role in statistical applications, especially in the case of testing univariate normality, see e.g. D’Agostinno and Stephens (1986). Normality may be the most common assumption in applying statistical procedures as in the classical linear regression model where the (unobserved) disturbance vector $\varepsilon$ is assumed to be normally distributed. It is well known that departures from normality may lead to substantially incorrect statements in the analysis of economic models. Thus, a test on normality based on the (observable) regression residuals is an absolute “must” in any regression analysis. Here, we restrict our attention to the vector $\hat{\varepsilon}$ of OLS residuals which is given by $\hat{\varepsilon} = (I - H) \varepsilon$ with $H = X(X'X)^{-1}X'$, i.e. the vector of residuals $\hat{\varepsilon}$ is a linear transformation of the unobserved disturbance vector $\varepsilon$. Therefore, tests of distributional assumptions on $\varepsilon$ are based on $\hat{\varepsilon}$.

One of the most famous tests for normality of regression residuals is the test of Jarque and Bera (1980, 1987), which has gained great acceptance among econometricians. The test statistic JB is a function of the measures of skewness $S$ and kurtosis $K$ computed from the sample. Under normality, the theoretical values of $S$ and $K$ are 0 and 3, respectively. The purpose of this paper is to compare the Jarque-Bera test with other goodness-of-fit tests like the Shapiro-Wilk test, the Kuiper test as well as with tests of Kolmogorov-Smirnov and Cramér-von Mises type in varying sample situations. As pointed out by several authors, see e.g. Jarque and Bera (1987) and Urzúa (1996), the Jarque-Bera test behaves well in comparison with some other tests for normality discussed in the literature if the alternatives to the normal distribution belong to the Pearson family. For our power study we assume the model of contaminated normal distributions (CN) for the components $\varepsilon_i$ of disturbance vector $\varepsilon$, i.e. $\varepsilon_i \sim F = (1 - p) N(\mu_1, \sigma_1^2) + p N(\mu_2, \sigma_2^2), i = 1, \ldots, n, 0 \leq p \leq 1$. This model covers a broad range of distributions, symmetric and asymmetric ones, and can also be used to describe ”small departures” from normality. The contaminated normal distribution function $F$ can be interpreted as follows: With probability $(1 - p)$, an observation comes from a normal distribution with mean $\mu_1$ and variance $\sigma_1^2$ and with probability $p$ from a normal distribution with mean $\mu_2$ and variance $\sigma_2^2$. Notice, that $F$ is not the cdf of a normal distribution. In section 2 we describe the two models and the hypotheses and give two data examples, one concerning the first model (random sample) and one the second model (linear regression). Section 3 presents all the tests mentioned above and section 4 deals with a power comparison of the tests including a study on the critical values of the Jarque-Bera test and the tests of Kolmogorov-Smirnov and Cramér-von Mises type. The power comparison is carried out via Monte-Carlo simulation. As a result, the Jarque-Bera test is, on the whole, the best one among all tests considered but for special sample situations he doesn’t work very well and other tests should be preferred. Section 5 gives a resumé of the results and some hints for further studies.
2 Model, Hypotheses and Data Examples

We consider two models, a random sample and the classical linear regression.

Model I (random sample):
Let be $X_1, \ldots, X_n$ independent random variables with absolutely continuous distributions function $F$. We wish to test

$$H_0 : F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) \text{ for all } x \in \mathbb{R}$$

vs. the two sided alternative $H_1 : F(x) \neq \Phi\left(\frac{x - \mu}{\sigma}\right)$ for at least one $x \in \mathbb{R}$,

where $\Phi$ is the cdf of the standard normal distribution and $\mu (-\infty < \mu < \infty)$ as well as $\sigma (\sigma > 0)$ may be known or unknown. In the case of known $\mu$ and $\sigma$ we assume with any loss of generality $\mu = 0$ and $\sigma = 1$. In the case of unknown $\mu$ and $\sigma$ the parameters are estimated by the sample mean $\bar{x}$ and sample standard deviation $s$, respectively.

Model II (classical linear regression):
Let be $y = X\beta + \varepsilon$, where $y$ is a $(n, 1)$-vector, $X$ is a nonstochastic $(n, k)$-matrix of rank $k$, $\beta$ is the $(k, 1)$-vector of unknown parameters and $\varepsilon$ is the disturbance $(n, 1)$-vector whose components are assumed to be uncorrelated and distributed with expectation zero and constant variance $\sigma^2$.

Since the vector $\varepsilon$ is unobservable, a test for normality generally is based on sample residuals such as OLS residuals $\hat{\varepsilon}_i$ which are given by $\hat{\varepsilon} = (I - H)\varepsilon$ with the so-called hat matrix $H = X(X'X)^{-1}X'$. Note, that the components $\hat{\varepsilon}_i$ of $\hat{\varepsilon}$ are not uncorrelated and do not have equal variances in general.

Now, let us present two real data examples to which the selected tests are applied later on, example 1 for model I and example 2 for model II.

Example 1:
The first example refers to returns of the share index DAX. The DAX measures the performance of the Prime Standards 30 largest German companies. The data set contains 200 daily DAX returns (in percent) from June 18, 2002, until March 31, 2003.

Example 2:
The dependency of corporate revenues on ad spending of leading U.S. advertisers is subject of the second example. The following data set shows the total U.S. ad spending and U.S. corporate revenue (in million dollars for 2001) of 50 large companies e.g. General Motors, Ford, Daimler-Chrysler, AT&T, IBM, Hewlett-Packard, Sony, Walt Disney, Coca-Cola, PepsiCo and Unilever.
Fig. 1. Histogram of daily DAX returns

Table 1
Data set U.S. advertisers 2001

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>430</td>
<td>2354</td>
<td>289</td>
<td>5528</td>
<td>449</td>
<td>3414</td>
</tr>
<tr>
<td>444</td>
<td>17522</td>
<td>664</td>
<td>74476</td>
<td>1137</td>
<td>39900</td>
</tr>
<tr>
<td>656</td>
<td>12356</td>
<td>884</td>
<td>6745</td>
<td>385</td>
<td>16726</td>
</tr>
<tr>
<td>1372</td>
<td>52550</td>
<td>3374</td>
<td>132399</td>
<td>552</td>
<td>7716</td>
</tr>
<tr>
<td>421</td>
<td>5804</td>
<td>881</td>
<td>14525</td>
<td>290</td>
<td>9452</td>
</tr>
<tr>
<td>339</td>
<td>37668</td>
<td>366</td>
<td>4557</td>
<td>2210</td>
<td>18215</td>
</tr>
<tr>
<td>543</td>
<td>19597</td>
<td>899</td>
<td>31725</td>
<td>479</td>
<td>3534</td>
</tr>
<tr>
<td>974</td>
<td>13154</td>
<td>778</td>
<td>53553</td>
<td>812</td>
<td>10131</td>
</tr>
<tr>
<td>397</td>
<td>5021</td>
<td>1103</td>
<td>34673</td>
<td>1310</td>
<td>21127</td>
</tr>
<tr>
<td>527</td>
<td>7842</td>
<td>994</td>
<td>35215</td>
<td>926</td>
<td>39888</td>
</tr>
<tr>
<td>410</td>
<td>12791</td>
<td>426</td>
<td>9382</td>
<td>1484</td>
<td>11315</td>
</tr>
<tr>
<td>903</td>
<td>7526</td>
<td>1086</td>
<td>32004</td>
<td>1462</td>
<td>64649</td>
</tr>
<tr>
<td>1985</td>
<td>72708</td>
<td>1618</td>
<td>20204</td>
<td>1283</td>
<td>19466</td>
</tr>
<tr>
<td>302</td>
<td>21760</td>
<td>422</td>
<td>6129</td>
<td>596</td>
<td>15980</td>
</tr>
<tr>
<td>289</td>
<td>3997</td>
<td>304</td>
<td>9328</td>
<td>1757</td>
<td>20970</td>
</tr>
<tr>
<td>746</td>
<td>15651</td>
<td>512</td>
<td>50098</td>
<td>312</td>
<td>1724</td>
</tr>
<tr>
<td>2408</td>
<td>108296</td>
<td>428</td>
<td>9363</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3 Goodness-of-fit tests

3.1 Jarque-Bera-test and its modification

The test statistic JB of Jarque-Bera is defined by

$$JB = \frac{n}{6} \left( S^2 + \frac{(K - 3)^2}{4} \right)$$

where the sample skewness $S = \hat{\mu}_3/\hat{\mu}_2^{3/2}$ is an estimator of $\beta_1 = \mu_3/\mu_2^{3/2}$ and

the sample kurtosis $K = \hat{\mu}_4/\hat{\mu}_2^2$ an estimator of $\beta_2 = \mu_4/\mu_2^2$, $\mu_2$ and $\mu_3$ are the

theoretical second and third central moments, respectively, with its estimates

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^j, \ j = 2, 3, 4.$$ 

JB is asymptotically chi-squared distributed with two degrees of freedom because JB is just the sum of squares of two asymptotically independent standardized normals, see Bowman and Shenton (1975). That means: $H_0$ has to be rejected at level $\alpha$ if $JB \geq \chi^2_{1-\alpha, 2}$.
In the more usual case of linear regression JB is calculated for the regression residuals.

Urzúa (1996) introduced a modification of the Jarque-Bera test - we call it JBU - by standardizing the skewness $S$ and kurtosis $K$ in the formula of JB appropriately in the following way:

$$JBU = \left( \frac{S^2}{v_S} + \frac{(K - e_K)^2}{v_K} \right)$$

with $v_S = \frac{6(n-2)}{(n+1)(n+3)}$, $e_K = \frac{3(n-1)}{n+1}$ and $v_K = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$.

Notice, that JB and JBU are asymptotically equivalent, i.e. $H_0$ has to be rejected at level $\alpha$ if $JBU \geq \chi^2_{1-\alpha,2}$.

3.2 Shapiro-Wilk test

The Shapiro-Wilk test is motivated by a probability plot in which we consider the regression of the ordered observations on the expected values of the order statistics from the hypothesized distribution. The test statistic is defined by

$$SW = \frac{\left( \sum_{i=1}^{n} a_i X_{(i)} \right)^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}.$$ 

$X = (X_1, \ldots, X_n)$ is a vector of random variables and $X_{(i)}$ the corresponding ordered vector. $\overline{X}$ is the usual sample mean. The weights $a_i$, $i = 1, \ldots, n$, are calculated like this. $Y = (Y_1, \ldots, Y_n)$ is a vector of random variables from a normal distribution and $Y_{(i)}$ again the corresponding ordered vector. The determination of $a_i$ requires the calculation of the vector of expectation values and the covariance matrix of $Y_{(i)}$: $m' = (m_1, \ldots, m_n)$ where $m_i = E(Y_{(i)})$ and $V$ where $v_{ij} = \text{Cov}(Y_{(i)}, Y_{(j)})$. The vector $a$ of the weights $a_i$ yields as follows: $a' = m'V^{-1}(m'V^{-1}m)^{-1/2}$.

$H_0$ has to be rejected, if $SW \leq w_\alpha$.

For the components of the vector $a$ we have $a_i = -a_{n-i+1}$, they are tabulated by Shapiro and Wilk (1965) for $n \leq 50$, where critical values $w_\alpha$ of SW are given, too, see also Shapiro et al. (1968) and Shapiro and Francia (1972).
3.3 Tests of Kolmogorov-Smirnov type

Now, let us consider nonparametric goodness-of-fit tests which are based on the empirical distribution function.

(1) Kolmogorov-Smirnov test KS

As above, let be $X_{(1)}, \ldots, X_{(n)}$ the order statistics of $X_1, \ldots, X_n$ and let $F_n$ be the usual empirical distribution functions for the $X$-sample. Then the Kolmogorov-Smirnov statistic $KS$ is defined by

$$KS = \sup_x |F_n(x) - F_0(x)|$$

In our case of testing normality we have $F_0(x) = \Phi \left( \frac{x - \mu}{\sigma} \right)$. The corresponding test rejects $H_0$ if $KS \geq k_{1-\alpha}$. Exact critical values $k_{1-\alpha}$ are reported by Buning and Trenkler (1994), where the asymptotic null distribution of $KS$ is given, too.

(2) Modified KS-test

The test statistic $KS$ can be modified by introducing appropriate non-negative weight functions $W(F_0(x))$ in order to give different weights to the difference $|F_n(x) - F_0(x)|$, see Bünning (2001). As a special case we choose $W(F_0(x)) = \sqrt{F_0(x)(1 - F_0(x))}/n$ which is symmetric about the center $F_0(x) = 0.5$. This weight function was introduced by Anderson and Darling (1952). Then we can define a modification of $KS$ by

$$KSW = \sup_x \frac{\sqrt{n} |F_n(x) - F_0(x)|}{\sqrt{F_0(x)(1 - F_0(x))}}$$

Obviously, the denominator $\sqrt{F_0(x)(1 - F_0(x))}$ place high weight on the upper and lower part of the underlying distribution. The corresponding test rejects $H_0$ if $KSW \geq k_{1-\alpha}^{(w)}$. Critical values $k_{1-\alpha}^{(w)}$ are reported in table 2, page 9.

(3) Kuiper test

Let be $D^+ = \sup_x (F_n(x) - F_0(x))$ and $D^- = \sup_x (F_0(x) - F_n(x))$. Then the statistic of Kuiper (1960) is defined by

$$KUI = D^+ + D^-.$$ 

It should be noted that the Kolmogorov-Smirnov statistic $KS$ can be written as $KS = \max(D^+, D^-)$. $H_0$ has to be rejected if $KUI \geq k_{1-\alpha}^{(u)}$. Critical values $k_{1-\alpha}^{(u)}$ are given in table 2.
3.4 Tests of Cramér-von Mises type

(1) Cramér-von Mises test
The Cramér-von Mises statistic CM is defined as follows:
\[ CM = n \cdot \int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 f_0(x) \, dx, \]
which can be written as
\[ CM = \frac{1}{12n} + \sum_{i=1}^{n} \left[ F_0(x_{(i)}) - \frac{2i - 1}{2n} \right]^2. \]
The test rejects \( H_0 \) if \( CM \geq c_{1-\alpha} \). Approximate critical values \( c_{1-\alpha} \) can be found in Anderson and Darling (1952).

(2) Modified CM-test
In the same manner as for the KS-test we can modify the CM-test by introducing an appropriate weight function. Here, we choose the weight function \( W(F_0(x)) = F_0(x)(1 - F_0(x)) \), i.e.
\[ CM_{W} = n \cdot \int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} f_0(x) \, dx \text{ for } F_0(x) \neq 0, 1. \]
For computational simplification \( CM_{W} \) can be written as
\[ CM_{W} = n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)(\ln p_{(i)} + \ln(1 - p_{(n+1-i)})) \text{ with } p_{(i)} = F_0(x_{(i)}). \]
The corresponding test rejects \( H_0 \) if \( CM_{W} \geq c_{1-\alpha}^{(w)} \). Critical values \( c_{1-\alpha}^{(w)} \) are reported in table 2.

In the following table critical values of the test statistics KSW, CMW and KUI are presented for selected sample sizes and \( \alpha = 0.05 \). The critical values were obtained by Monte Carlo simulation (100 000 replications). The suffix "S" indicates a test based on standardized observations. For example, KSS is the abbreviation for the corresponding KS test. For further information see subsection 4.1.2, page 12.

Critical values of the KSW-test can also be found e.g. in Canner (1975).

For a comprehensive study of tests based on the empirical distribution function, see Stephens (1974).
Table 2
Critical values of tests for various sample sizes \( n, \alpha = 0.05 \)

<table>
<thead>
<tr>
<th>Tests</th>
<th>( n = 10 )</th>
<th>( n = 20 )</th>
<th>( n = 50 )</th>
<th>( n = 100 )</th>
<th>( n = 200 )</th>
<th>( n = 500 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>JB</td>
<td>2.5347</td>
<td>3.7677</td>
<td>5.0037</td>
<td>5.4479</td>
<td>5.7275</td>
<td>5.8246</td>
</tr>
<tr>
<td>KS</td>
<td>0.4101</td>
<td>0.2936</td>
<td>0.1886</td>
<td>0.1338</td>
<td>0.0951</td>
<td>0.0603</td>
</tr>
<tr>
<td>KSS</td>
<td>0.2615</td>
<td>0.1920</td>
<td>0.1243</td>
<td>0.0890</td>
<td>0.0635</td>
<td>0.0403</td>
</tr>
<tr>
<td>KSW</td>
<td>2.0236</td>
<td>1.4369</td>
<td>0.9165</td>
<td>0.6451</td>
<td>0.4544</td>
<td>0.2869</td>
</tr>
<tr>
<td>KSW2</td>
<td>0.8535</td>
<td>0.8049</td>
<td>0.6637</td>
<td>0.5266</td>
<td>0.4007</td>
<td>0.2697</td>
</tr>
<tr>
<td>KUI</td>
<td>0.5142</td>
<td>0.3719</td>
<td>0.2401</td>
<td>0.1715</td>
<td>0.1217</td>
<td>0.0774</td>
</tr>
<tr>
<td>KUIS</td>
<td>0.4318</td>
<td>0.3166</td>
<td>0.2049</td>
<td>0.1469</td>
<td>0.1046</td>
<td>0.0665</td>
</tr>
<tr>
<td>CM</td>
<td>0.4584</td>
<td>0.4534</td>
<td>0.4601</td>
<td>0.4567</td>
<td>0.4627</td>
<td>0.4598</td>
</tr>
<tr>
<td>CM2</td>
<td>0.1193</td>
<td>0.1230</td>
<td>0.1241</td>
<td>0.1253</td>
<td>0.1258</td>
<td>0.1265</td>
</tr>
<tr>
<td>CM2W</td>
<td>2.5265</td>
<td>2.4737</td>
<td>2.4868</td>
<td>2.4695</td>
<td>2.5280</td>
<td>2.4814</td>
</tr>
<tr>
<td>CM2WS</td>
<td>0.6905</td>
<td>0.7176</td>
<td>0.7380</td>
<td>0.7460</td>
<td>0.7479</td>
<td>0.7506</td>
</tr>
<tr>
<td>SW</td>
<td>0.8451</td>
<td>0.9040</td>
<td>0.9486</td>
<td>0.9642</td>
<td>0.9727</td>
<td>0.9795</td>
</tr>
</tbody>
</table>

3.5 Applications

Now, let us test the hypothesis of normality for the data in example 1 and example 2 by applying all the tests above. We get the following results:

Table 3
Testing normality for DAX data (short tails)

<table>
<thead>
<tr>
<th>Test</th>
<th>JB</th>
<th>JBU</th>
<th>KS</th>
<th>KSW</th>
<th>KUI</th>
<th>CM</th>
<th>CMW</th>
<th>SW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistik</td>
<td>5.1218*</td>
<td>5.1388</td>
<td>0.0460</td>
<td>0.2050</td>
<td>0.0751</td>
<td>0.0801</td>
<td>0.6071</td>
<td>0.9827</td>
</tr>
<tr>
<td>p-Wert</td>
<td>0.062</td>
<td>0.072</td>
<td>0.38</td>
<td>0.22</td>
<td>0.49</td>
<td>0.21</td>
<td>0.22</td>
<td>0.015</td>
</tr>
<tr>
<td>asymp. p</td>
<td>0.077</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* \( S = 0.3433 \) and \( K = 2.6216 \)

Table 4
Testing normality for advertising data (residuals)

<table>
<thead>
<tr>
<th>Test</th>
<th>JB</th>
<th>JBU</th>
<th>KS</th>
<th>KSW</th>
<th>KUI</th>
<th>CM</th>
<th>CMW</th>
<th>SW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistik</td>
<td>3.9716*</td>
<td>5.6718</td>
<td>0.1386</td>
<td>0.4726</td>
<td>0.2257</td>
<td>0.1666</td>
<td>0.9099</td>
<td>0.9594</td>
</tr>
<tr>
<td>p-Wert</td>
<td>0.071</td>
<td>0.062</td>
<td>0.017</td>
<td>0.115</td>
<td>0.015</td>
<td>0.014</td>
<td>0.019</td>
<td>0.084</td>
</tr>
<tr>
<td>asymp. p</td>
<td>0.137</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* \( S = 0.4600 \) and \( K = 4.0296 \)
Obviously, the results for the eight tests are extremely different. In example 1 the Kuiper test has the greatest p-value of 49% and the Shapiro-Wilk test the smallest one with 1.5%, it is the only test which rejects at the 5% level. The differences of the p-values in example 2 are not so considerable as in the first example, but now, half of the tests reject at the 5% level. Thus, a power comparison of all the tests becomes important, it is carried out in section 4.2.

4 Power study

4.1 Critical values of some tests

4.1.1 Exact critical values of JB for models I and II

For model I and II let us compare the asymptotic critical values of the Jarque-Bera test JB with the exact ones calculated by simulation for sample sizes $n = 10, 20, 50, 100, 200$ and $500$ and varying levels $0.01, 0.02, 0.05, 0.1$ and $0.2$. In model II we consider the cases of one, three and six independent regression variables assuming different distribution functions such as uniform, normal and exponential. Table 5 displays the critical values of JB for model I.

Table 5
Critical values of JB-test (random sample)

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$n = 10$</th>
<th>$n = 20$</th>
<th>$n = 50$</th>
<th>$n = 100$</th>
<th>$n = 200$</th>
<th>$n = 500$</th>
<th>$n \to \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>5.738</td>
<td>9.458</td>
<td>12.331</td>
<td>12.296</td>
<td>11.750</td>
<td>10.601</td>
<td>9.210</td>
</tr>
<tr>
<td>0.02</td>
<td>4.274</td>
<td>6.583</td>
<td>8.721</td>
<td>9.089</td>
<td>8.788</td>
<td>8.349</td>
<td>7.824</td>
</tr>
<tr>
<td>0.05</td>
<td>2.535</td>
<td>3.768</td>
<td>5.004</td>
<td>5.448</td>
<td>5.728</td>
<td>5.825</td>
<td>5.991</td>
</tr>
<tr>
<td>0.10</td>
<td>1.618</td>
<td>2.335</td>
<td>3.192</td>
<td>3.643</td>
<td>4.081</td>
<td>4.324</td>
<td>4.605</td>
</tr>
<tr>
<td>0.20</td>
<td>1.126</td>
<td>1.556</td>
<td>2.122</td>
<td>2.474</td>
<td>2.748</td>
<td>2.985</td>
<td>3.219</td>
</tr>
</tbody>
</table>

From table 5 we can state the JB-test based on asymptotical critical values is conservative for $\alpha \geq 0.05$, and that considerably for small sample sizes; for $\alpha < 0.05$, the pattern is not so clear. As an example: For $\alpha = 0.01$, the approximation by the Chi-square distribution is better for $n = 20$ than for $n = 200$, of course, a surprising result. In general, the approximation by the Chi-squared distribution does not work well even not for large sample sizes, the speed of convergence is very slow. Thus, approximate critical values $\chi^2_{1-\alpha}$ for the JB-test should be used cautiously in empirical studies and for a meaningful power comparison exact critical values have to be used.
Figure 3 presents the graphs of the exact and asymptotic distribution of the JB-statistic. Obviously, the approximation becomes more precise with increasing sample sizes.

Fig. 3. Simulated and asymptotic distribution of the JB-test in comparison

Now, let us consider model II. Table 6 lists critical values of JB with three independent regression variables from a standard normal distribution. We restrict our presentation to this case because our simulation has shown that the pattern of critical values is nearly the same for one, three and six variables as well as for other distributions mentioned above.

Table 6
Critical values of JB-test (residuals)
From table 5 and 6 we see that for all levels the differences between the critical values in model I and II are negligible for \( n \geq 50 \), the pattern in model II is nearly the same as in model I.

### 4.1.2 Exact critical values of KS and CM for original and standardized data

We study critical values of the KS-test and CM-test for the original data and for the standardized observations where the unknown parameters \( \mu \) and \( \sigma \) of the normal distribution function are estimated by the arithmetic mean \( \bar{x} \) and the empirical standard deviation \( s_x \), i.e. \( z_i = (x_i - \bar{x})/s_x \). We again consider sample sizes \( n = 10, 20, 50, 100, 200 \) and 500, but here only the usual levels \( \alpha = 0.01, 0.05 \) and 0.10.

Table 7
Critical values of KS- and CM-test

<table>
<thead>
<tr>
<th>Critical values of KS</th>
<th>(original data)</th>
<th>Critical values of KS</th>
<th>(standardized data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>( n = 10 )</td>
<td>( n = 20 )</td>
<td>( n = 50 )</td>
</tr>
<tr>
<td>0.01</td>
<td>0.4898</td>
<td>0.3494</td>
<td>0.2248</td>
</tr>
<tr>
<td>0.05</td>
<td>0.4101</td>
<td>0.2936</td>
<td>0.1886</td>
</tr>
<tr>
<td>0.10</td>
<td>0.3701</td>
<td>0.2640</td>
<td>0.1699</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Critical values of CM</th>
<th>(original data)</th>
<th>Critical values of CM</th>
<th>(standardized data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>( n = 10 )</td>
<td>( n = 20 )</td>
<td>( n = 50 )</td>
</tr>
<tr>
<td>0.01</td>
<td>0.7143</td>
<td>0.7147</td>
<td>0.7341</td>
</tr>
<tr>
<td>0.05</td>
<td>0.4584</td>
<td>0.4534</td>
<td>0.4601</td>
</tr>
<tr>
<td>0.10</td>
<td>0.3469</td>
<td>0.3439</td>
<td>0.3487</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Critical values of CM</th>
<th>(standardized data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>( n = 10 )</td>
</tr>
<tr>
<td>0.01</td>
<td>0.1687</td>
</tr>
<tr>
<td>0.05</td>
<td>0.1193</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0989</td>
</tr>
</tbody>
</table>

From table 7 we can state that for the tests KS and CM there is a great difference between the critical values of original and standardized data. If we have to estimate the unknown parameters \( \mu \) and \( \sigma \) and apply, however, both tests by using critical values of KS and CM based on original data then the tests are extremely conservative. The same is true for the modifications KSW.
and CMW of KS and CM, respectively. Figure 4 illustrates the great difference between the distributions of KS, CM (original data) and KSS, CMS (standardized data).

Fig. 4. Distributions of KS, KSS, CM, CMS (n = 50)

4.2 Power comparison

We investigate via Monte Carlo simulation (10 000 replications) the power of all the tests presented in section 3. To conduct the simulation study we select the model of contaminated normal CN for $H_1$, the distribution function $F$ of which can be given in the following form, see section 1:

$$F = (1 - p) N(\mu_1, \sigma_1^2) + p N(\mu_2, \sigma_2^2)$$ with $0 \leq p \leq 1$.

Without any loss of generality we assume $\mu_1 = 0$ and $\sigma_1^2 = 1$. Furthermore, we choose $\mu_2 = 0, 1, 2, 3, 4, \sigma_2 = 1, 2, 3, 4, 6$ and $p = 0.01, 0.05, 0.20, 0.35, 0.50, 0.65, 0.80, 0.90, 0.95, 0.99$. The choice of such parameters guarantees a broad range of distributions, unimodal and bimodal ones, symmetric distributions with short up to very long tails and asymmetric ones with different strength of skewness. For all the tests we use exact critical values. In the simulation the tests of Kolmogorov-Smirnov and Cramér-von Mises type are carried out on a basis of standardized observations.
At first, comparing the JB-test and JBU-test we can state that the difference in power of the tests is very small, in 93% of the cases the difference is less than 1%, the JB-test seems to be a little bit more efficient than JBU. The power of JBU is never higher than 2% in comparison to JB, but in some cases lower than 10%. Because of the small power differences between JB and JBU we omit JBU from the following power presentations.

Figures 5 to 12 display the power of the seven tests JB, KS, KSW, KUI, CM, CMW and SW from section 3 by selecting different values of $\mu_2$, $\sigma_2$ and $p$. Different sample sizes are chosen in order to produce graphs for a visible power comparison. In the bottom of the figures the corresponding density functions of CN are plotted in comparison to the standard normal density. At first, we consider the case of random sample with original data and then the case of regression variables.

The values of the corresponding parameters of skewness $\beta_1$ and kurtosis $\beta_2$ are given in the following tables 8 and 9.

Table 8
Skewness and kurtosis for various CN distributions, part 1

<table>
<thead>
<tr>
<th>Fig.</th>
<th>$\mu_2$</th>
<th>$p$</th>
<th>$\sigma_2 = 1$</th>
<th>$\sigma_2 = 2$</th>
<th>$\sigma_2 = 3$</th>
<th>$\sigma_2 = 4$</th>
<th>$\sigma_2 = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>0.10</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\beta_2 = 3$</td>
<td>$\beta_2 = 4.44$</td>
<td>$\beta_2 = 8.33$</td>
<td>$\beta_2 = 12.72$</td>
<td>$\beta_2 = 19.33$</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.50</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\beta_2 = 3$</td>
<td>$\beta_2 = 4.08$</td>
<td>$\beta_2 = 4.92$</td>
<td>$\beta_2 = 5.34$</td>
<td>$\beta_2 = 5.68$</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.80</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\beta_2 = 3$</td>
<td>$\beta_2 = 3.37$</td>
<td>$\beta_2 = 3.56$</td>
<td>$\beta_2 = 3.64$</td>
<td>$\beta_2 = 3.7$</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0.50</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0.65$</td>
<td>$\beta_1 = 0.92$</td>
<td>$\beta_1 = 0.96$</td>
<td>$\beta_1 = 0.83$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\beta_2 = 2.04$</td>
<td>$\beta_2 = 2.85$</td>
<td>$\beta_2 = 3.72$</td>
<td>$\beta_2 = 4.37$</td>
<td>$\beta_2 = 5.11$</td>
</tr>
</tbody>
</table>

Figures 5 to 8 are concerned with power curves as functions of $\sigma_2$ for different values of $p$ and $\mu_2 = 0$ (symmetric case) and $\mu_2 = 3$ (asymmetric case for $\sigma_2 \neq 1$). Of course, with increasing values of $\sigma_2$ the kurtosis $\beta_2$ increases, for $p = 0.1$ much more than for $p = 0.5$ and 0.8. In the symmetric cases (Figures 5 to 7), we see that for small values of $p$ the JB-test is the best one, the power of JB, however, decreases with increasing $p$ where $\beta_2$ increases slower. For great values of $p$ the Kuiper test and CMW-test are superior, the Shapiro-Wilk test is, in general, not a powerful test in comparison to the others. In the asymmetric case with $p = 0.5$, $\mu_2 = 3$ and $\sigma_2 \neq 1$ (Figure 8), where the skewness $\beta_1$ is smaller than 1, JB and KSW have lowest power, both tests are even biased for the symmetric case, where the kurtosis is equal to 2.04, smaller than that of the standard normal distribution ($\beta_2 = 3$).
Fig. 5. Power of the tests (symmetric case 1, random sample)

Fig. 6. Power of the tests (symmetric case 2, random sample)
Fig. 7. Power of the tests (symmetric case 3, random sample)

Fig. 8. Power of the tests (asymmetric case 1, random sample)
Figures 9 and 10 present power curves as functions of $\mu_2$, again, the kurtosis is much greater for $p = 0.05$ than for $p = 0.5$. In Figure 11 the power is plotted as a function of $p$ with decreasing kurtosis.

Table 9
Skewness and kurtosis for various CN distributions, part 2

<table>
<thead>
<tr>
<th>Fig.</th>
<th>$\sigma_2$</th>
<th>$p$</th>
<th>$\mu_2 = 0$</th>
<th>$\mu_2 = 1$</th>
<th>$\mu_2 = 2$</th>
<th>$\mu_2 = 3$</th>
<th>$\mu_2 = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.05</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0.9$</td>
<td>$\beta_1 = 1.71$</td>
<td>$\beta_1 = 2.35$</td>
<td>$\beta_1 = 2.84$</td>
<td>$\beta_1 = 13.47$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\beta_2 = 3$</td>
<td>$\beta_2 = 14.12$</td>
<td>$\beta_2 = 15.75$</td>
<td>$\beta_2 = 17.65$</td>
<td>$\beta_2 = 19.24$</td>
<td>$\beta_2 = 2.92$</td>
</tr>
<tr>
<td>10</td>
<td>0.50</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 2.04$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\beta_2 = 3$</td>
<td>$\beta_2 = 2.92$</td>
<td>$\beta_2 = 2.5$</td>
<td>$\beta_2 = 2.04$</td>
<td>$\beta_2 = 1.72$</td>
<td>$\beta_2 = 1.72$</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
<td>$\beta_1 = 0$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$\beta_2 = 13.47$</td>
<td>$\beta_2 = 9.75$</td>
<td>$\beta_2 = 5.34$</td>
<td>$\beta_2 = 3.64$</td>
<td>$\beta_2 = 3.14$</td>
<td>$\beta_2 = 3.47$</td>
</tr>
</tbody>
</table>

Figure 9 shows that the JB-test is the best one for small $p = 0.05$ even in the asymmetric case but the power loss is dramatic for $p = 0.5$ (Figure 10), a distribution which is symmetric and bimodal with small values of $\beta_2$ (JB is biased). Here, the SW- and CMW-tests are the winner. For fixed $\mu_2 = 0$ and $\sigma_2 = 4$ (Figure 11) the JB-test behaves very well for values of $p$ smaller, say, 0.25, but with increasing values of $p$, the power of JB decreases rapidly, the same is true for the other tests.
Fig. 10. Power of the tests (symmetric case 4, random sample)

Fig. 11. Power of the tests (symmetric case 5, random sample)
Further results on the power of tests on normality can be found in chapter 7 of Thode (2002). At this point it should be noted that the kurtosis $\beta_2$ is not a suitable measure for tailweight, a discussion of this problem "what does $\beta_2$ really measure?" can be found in chapter 3.3 of Buning (1991) where measures of tailweight are given, e.g. the measure

$$T = \frac{x_{0.975} - x_{0.025}}{x_{0.875} - x_{0.125}}$$

where $x_p$ is the $p$-quantile of the distribution function $F$. Table 10 presents values of $\beta_2$ and $T$ for symmetric contaminated normal distributions (CN) with $\mu_2 = 0$ and $\sigma_2 = 3$ varying values of $p$ (see figures 5 to 7).

<table>
<thead>
<tr>
<th>$p$</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2$</td>
<td>3.00</td>
<td>8.33</td>
<td>7.56</td>
<td>6.53</td>
<td>5.64</td>
<td>4.94</td>
<td>3.56</td>
</tr>
<tr>
<td>$T$</td>
<td>1.70</td>
<td>1.99</td>
<td>2.42</td>
<td>2.56</td>
<td>2.47</td>
<td>2.28</td>
<td>1.84</td>
</tr>
</tbody>
</table>

We see that with increasing proportion $p$ of contamination the kurtosis $\beta_2$ decreases whereas the tailweight $T$ is, at first, increasing and then decreasing. That means, CN has, e.g., longer tails for $p = 0.2$ than for $p = 0.1$, the kurtosis, however, is smaller for $p = 0.2$ than for $p = 0.1$.

Now, let us study the power of the tests in the case of regression variables. As already mentioned above we considered the cases of one, three and six independent regression variables assuming different distribution functions such as uniform, normal and exponential. It might be a surprising result that the pattern of the power curves of the tests in each sample situation (various numbers of variables and various distribution functions) is always nearly the same. Thus, we restrict the illustration to the case of three standard normally distributed regression variables, the power curves are given in Figure 12 for the symmetric case with $\mu_2 = 0$ and $p = 0.5$ and varying $\sigma_2$. Obviously, the JB-test behaves well, but the CMW-, KUI- and the KS-tests are superior with increasing $\sigma_2$.

For further studies of tests on normality in the linear regression model, see e.g. Huang and Bolch (1974) and White and MacDonald (1980).

We assumed independent regression variables. The question arises how robust are the tests above if the regression variables are correlated, an important problem for testing normality in regression analysis. A first study, assuming an autoregressive process of order one for the error terms with $\rho = -0.5$, shows that the influence on the critical values and therefore on the power of the tests is considerable. This is, however, another topic.
5 Conclusions and Outlook

As an overall result of our power study we can state:

- The asymptotic JB-test is conservative for $\alpha = 0.05, 0.10$, the approximation by the Chi-square distribution does not work well.
- There are no remarkable differences between the critical values of the tests in model I with random samples and in model II with regression variables for $n \geq 50$.
- The nonparametric tests, KS and CM, as well as its modifications KSW and CMW, are very conservative in the case of estimating $\mu$ and $\sigma$ if we use the critical values for the original data.
- The power values for each of the tests are nearly the same for random samples (model I) and for regression variables (model II).
- The JB-test behaves well (it is often even the best one) for symmetric distributions with medium up to long tails and for slightly skewed distributions with long tails.
- The power of the JB-test is poor for distributions with short tails, especially if the shape is bimodal, sometimes JB is even biased. In this case the modification of CM, CMW, or the Shapiro-Wilk test may be recommended.
Two further problems by testing normality in regression analysis are of the-
oretical and practical importance, the case of autocorrelated error terms as
already mentioned above and the case of heteroscedastic error terms. The in-
fluence of autocorrelation on the critical values and the power of the tests
seems to be strong as first studies show. Thus, generalized least square es-
timator (GLS) is more appropriate for the regression parameters instead of
the ordinary least square (OLS). For the case of heteroscedasticity of the error
terms similar studies have to be done, too. For studies of serial correlation and
heteroscedasticity of regression residuals, see e.g. Jarque and Bera (1980).

References

Anderson, T., Darling, D., 1952. Asymptotic theory of goodness of fit criteria
212.
Bowman, K., Shenton, L., 1975. Omnibus test contours for departures from
normality based on $b_1$ and $b_2$. Biometrika 62, 243–250.
Büning, H., 2001. Kolmogorov-smirnov and Cramér-von Mises type two-
sample tests with various weight functions. Communications in Statistics
- Simulation and Computation 30, 847–865.
Walter de Gruyter, Berlin.
Canner, P., 1975. A simulation study of one- and two-sample Kolmogorov-
Smirnov statistics with a particular weight function. Journal of the American
Statistical Association 70, 209–211.
Dekker, New York.
Huang, C., Bolch, B., 1974. On the testing of regression disturbances for nor-
Jarque, C., Bera, A., 1980. Efficient tests for normality homoscedasticity and
Shapiro, S., Francia, R., 1972. An approximate analysis of variance test for
Shapiro, S., Wilk, M., 1965. An analysis of variance test for normality (com-
plete samples). Biometrika 52, 591–611.
Shapiro, S., Wilk, M., Chen, H., 1968. A comparative study of various tests for
Diskussionsbeiträge
des Fachbereichs Wirtschaftswissenschaft
der Freien Universität Berlin

2004

2004/1 MUCHLINSKI, Elke
Kontroversen in der internationalen Währungspolitik.
Retrospektive zu Keynes-White-Boughton & IMF. 32 S.
Volkswirtschaftliche Reihe

2004/2 MUCHLINSKI, Elke
Central Banks - Transparency or Bafflement?
From "monetary mystique" to "creative ambiguity".
Volkswirtschaftliche Reihe

2004/3 STRAUSZ, Roland
Regulating Availability in Bargain Sales. 17 S.
Volkswirtschaftliche Reihe

2004/4 STRAUSZ, Roland
Deterministic Mechanisms are Optimal
in Standard Principal-Agent Models. 22 S.
Volkswirtschaftliche Reihe

2004/5 RENDTEL, Ulrich; KALTENBORN, Ulrich
The Stability of Simulation-based Estimation.
Volkswirtschaftliche Reihe

2004/6 LADWIG, Birgit
Transparenz von Zentralbanken im Kontext der Debatte
um die Europäische Zentralbank.
Volkswirtschaftliche Reihe

2004/7 RIESE, Hajo
Arbeitsmarktpolitik und ökonomische Theorie
- zur Degenerierung einer Forschungspraxis
II: Das erkenntnistheoretische Defizit
Volkswirtschaftliche Reihe

2004/8 EISEND, Martin
Metaanalyse - Einführung und kritische Diskussion
Betriebswirtschaftliche Reihe