Goodness-of-fit tests based on a robust measure of skewness

G. Brys * M. Hubert† A. Struyf‡

August 16, 2004

Abstract

In this paper we propose several goodness-of-fit tests based on robust measures of skewness and tail weight. They can be seen as generalisations of the Jarque-Bera test (Bera and Jarque, 1981) based on the classical skewness and kurtosis, and as an alternative to the approach of Moors et al. (1996) using quantiles. The power values and the robustness properties of the different tests are investigated by means of simulations and applications on real data. We conclude that MC-LR, one of our proposed tests, shows the best overall power and that it is moderately influenced by outlying values.

1 Introduction

The third and fourth moments of a distribution are called the skewness and kurtosis. For any distribution F with finite central moments μ_k up to $k = 3$, the *skewness* is defined as

$$\gamma_1(F) = \frac{\mu_3(F)}{\mu_2(F)^{3/2}}.$$

*Faculty of Applied Economics, University of Antwerp (UA), Prinsstraat 13, B-2000 Antwerp, Belgium, *Guy.Brys@ua.ac.be*

†Department of Mathematics, Katholieke Universiteit Leuven (KULeuven), W. de Croylaan 54, B-3001 Leuven, Belgium, Mia.Hubert@wis.kuleuven.ac.be

‡Postdoctoral Fellow of the Fund for Scientific Research - Flanders (Belgium), Department of Mathematics and Computer Science (UIA), Universiteitsplein 1, B-2610 Wilrijk, Belgium, Anja.Struyf@ua.ac.be

Skewness describes the asymmetry of a distribution. A symmetric distribution has zero skewness, an asymmetric distribution with the largest tail to the right has positive skewness, and a distribution with a longer left tail has negative skewness.

For any distribution F with finite central moments μ_k up to $k = 4$, the *kurtosis* is defined as

$$\gamma_2(F) = \frac{\mu_4(F)}{\mu_2(F)^2}.$$

There is no agreement on what it really measures. Strictly speaking, kurtosis measures both peakedness and tail heaviness of a distribution relative to that of the normal distribution. Consequently, its use is restricted to symmetric distributions. Finite-sample versions of γ_1 and γ_2 will be denoted by b_1 and b_2.

The classical skewness and kurtosis coefficient have some common disadvantages. They both have a zero breakdown value and an unbounded influence function, and so they are very sensitive to outlying values. One single outlier can make the estimate become very large or small, making it hard to interpret. Another disadvantage is that they are only defined on distributions having finite moments.

In Section 2 we propose several measures of skewness and of left and right tail weight for univariate continuous distributions. Their interpretation is clear and they are robust against outlying values. Contrary to the kurtosis coefficient, the tail weight measures can be applied to symmetric as well as asymmetric distributions. In Section 3 we introduce some robust goodness-of-fit tests. Section 4 and 5 include simulation results while Section 6 applies the tests on real data. Finally, Section 7 concludes.

2 Robust measures of skewness and tail weight

Assume we have independently sampled n observations $X_n = \{x_1, x_2, ..., x_n\}$ from a continuous univariate distribution F. We will consider the *medcouple* (MC), a robust skewness measure, proposed in Brys et al. (2003) and extensively discussed in Brys et al. (2004a). It is defined as

$$MC(F) = \underset{x_1 < m_F < x_2}{\text{med}} h(x_1, x_2)$$

with x_1 and x_2 sampled from F, $m_F = F^{-1}(0.5)$ and the kernel function h given by

$$h(x_i, x_j) = \frac{(x_j - m_F) - (m_F - x_i)}{x_j - x_i}.$$

This estimator has a breakdown value of 25% and a bounded influence function.

Furthermore, we consider the *left medcouple* (LMC) and *right medcouple* (RMC), respectively the left and right tail weight measure, as defined in Brys et al. (2004c). To construct these measures we have applied the medcouple to respectively the left and right half of the samples:

$$\mathrm{LMC}(F) = -\mathrm{MC}(x < m_F) \quad \text{and} \quad \mathrm{RMC}(F) = \mathrm{MC}(x > m_F),$$

yielding a breakdown value of 12.5%.

Finite sample versions will be denoted by MC_n, LMC_n and RMC_n. These measures can be computed at any distribution, even when finite moments do not exist. Their computation can be performed in $\mathrm{O}(n \log n)$ time due to the fast algorithm described in Brys et al. (2004a). They satify all natural requirements of skewness or tail weight measures including location and scale invariance. More details can be found in the cited references.

3 Description of the tests

In this section we discuss goodness-of-fit tests for the following null and alternative hypothesis:

$$\begin{cases} H_0 : \text{The sample is drawn from a distribution } F \\ H_1 : \text{The sample is not drawn from a distribution } F \end{cases}$$

In this paper we will investigate the performance of the tests at F taken to be the χ^2_2 distribution, the Student t_3 distribution and the Tukey's class of gh-distributions (Hoaglin et al., 1985). When a random variable Z is standard gaussian distributed, then

$$Y_{g,h} = \begin{cases} \frac{(e^{gZ}-1)}{g} e^{\frac{hZ^2}{2}} & g \neq 0 \\ Z e^{\frac{hZ^2}{2}} & g = 0 \end{cases}$$

is said to follow a gh-distribution $G_{g,h}$ with parameters $g \in I\!R$ and $h \geq 0$. The parameter g controls the skewness of the distribution, whereas h effects the tail weight.

Bera and Jarque (1981) proposed a normality test using the classical skewness and kurtosis coefficient. As been stated in Moors et al. (1996), under the normality assumption ($\gamma_1 = 0$ and $\gamma_2 = 3$) we can write:

$$\sqrt{n} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \rightarrow_{\mathcal{D}} N_2 \left(\begin{pmatrix} 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 24 \end{pmatrix} \right)$$

which leads to the Jarque-Bera test statistic:

$$T = n\left(\frac{b_1^2}{6} + \frac{(b_2-3)^2}{24}\right) \approx \chi_2^2.$$

This test can be viewed as a special case of the following generalization. Let $w = (w_1, w_2, ..., w_k)^t$ be estimators of $\omega = (\omega_1, \omega_2, ..., \omega_k)^t$, such that

$$\sqrt{n}\begin{pmatrix} w_1 & ... & w_k \end{pmatrix}^t \rightarrow_{\mathcal{D}} N_k(\omega, \Sigma_k)$$

then, under H_0, the generalized test statistic T

$$T = n(w-\omega)^t \Sigma_k^{-1}(w-\omega) \approx \chi_k^2.$$

We can thus easily construct new goodness-of-fit tests, analogous to Brys et al. (2004b). Taking $k = 2$, $w_1 = b_1$ and $w_2 = b_2$ leads to the generalized Jarque-Bera test (JB) with $\omega_1 = \gamma_1$ and $\omega_2 = \gamma_2$. A test based on the medcouple (MC) given in Brys et al. (2004a) has $k = 1$ and $w_1 = \text{MC}$. Secondly, Brys et al. (2004b) propose to use the test of LMC or RMC with $k = 1$ and $w_1 = \text{LMC}$ or $w_1 = \text{RMC}$. Combining the skewness and respectivily the left and right tail weight of a distribution leads to MC-L and MC-R with $k = 2$, $w_1 = \text{MC}$, and respectivily $w_2 = \text{LMC}$ and $w_2 = \text{RMC}$. Next, we can define a test based on the left and right medcouple (LR) with $k = 2$, $w_1 = \text{LMC}$ and $w_2 = \text{RMC}$. Finally, we propose a goodness-of-fit test MC-LR where $k = 3$, $w_1 = \text{MC}$, $w_2 = \text{LMC}$ and $w_3 = \text{RMC}$.

We will also include a test proposed in Moors et al. (1996). This test (MOORS) fits in the framework of the above generalisation by using $w_1 = \frac{F^{-1}(0.75)+F^{-1}(0.25)-2F^{-1}(0.5)}{F^{-1}(0.75)-F^{-1}(0.25)}$ as a robust measure of skewness and $w_2 = \frac{F^{-1}(0.875)-F^{-1}(0.625)+F^{-1}(0.375)-F^{-1}(0.125)}{F^{-1}(0.75)-F^{-1}(0.25)}$ as a robust measure of kurtosis. By using only quantiles of the data this test is resistant to 12.5% outliers in the data. This is the same as for all tests based on LMC and/or RMC.

Table 1 shows the values of ω and Σ_k for several distributions F for the generalized Jarque-Bera test, the MOORS test and the MC-LR test. By using the latter it is possible to write down ω and Σ_k for the other goodness-of-fit tests based on MC. Table 1 is derived from the influence function of the estimators, as described in Brys et al. (2004a) and in Brys et al. (2004c).

4 Simulation study at uncontaminated distributions

We investigate the seven proposed tests by generating $m = 1000$ samples of size $n = 100$ and $n = 1000$ from a wide range of distributions. Note that in all Figures in this Section

	ω(JB)	Σ_k(JB)	ω(MOORS)	Σ_k(MOORS)	ω(MC-LR)	Σ_k(MC-LR)
$G_{0,0}$	$\begin{pmatrix} 0 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 6 & 0 \\ 0 & 24 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1.23 \end{pmatrix}$	$\begin{pmatrix} 1.84 & 0 \\ 0 & 3.14 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0.199 \\ 0.199 \end{pmatrix}$	$\begin{pmatrix} 1.25 & 0.323 & -0.323 \\ 0.323 & 2.62 & -0.0123 \\ -0.323 & -0.0123 & 2.62 \end{pmatrix}$
χ^2_2	$\begin{pmatrix} 2 \\ 9 \end{pmatrix}$	$\begin{pmatrix} 72 & 720 \\ 720 & 8.06e(3) \end{pmatrix}$	$\begin{pmatrix} 0.262 \\ 1.31 \end{pmatrix}$	$\begin{pmatrix} 1.78 & -0.152 \\ -0.152 & 5.09 \end{pmatrix}$	$\begin{pmatrix} 0.338 \\ -0.109 \\ 0.333 \end{pmatrix}$	$\begin{pmatrix} 1.27 & 0.360 & -0.310 \\ 0.360 & 2.75 & -1.87e(-5) \\ -0.310 & -1.87e(-5) & 2.54 \end{pmatrix}$
t_3	$\begin{pmatrix} - \\ - \end{pmatrix}$	$\begin{pmatrix} - & - \\ - & - \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1.40 \end{pmatrix}$	$\begin{pmatrix} 1.87 & 0 \\ 0 & 4.62 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0.297 \\ 0.297 \end{pmatrix}$	$\begin{pmatrix} 1.36 & 0.221 & -0.221 \\ 0.221 & 2.58 & -0.0231 \\ -0.221 & -0.0231 & 2.58 \end{pmatrix}$

Table 1: Asymptotic mean ω and covariance matrix Σ_k of the (joint) distribution of several measures of skewness and tail weight, used in the JB test, the MOORS test and the MC-LR test. (Note that $1e(3)$ stands for 1000.)

and in Section 5 the upper panel represents the results for $n = 100$ and the lower panel shows them for $n = 1000$. Under the null hypothesis of $G_{0,0}$, the alternatives are taken to be $G_{g,h}$ with $(g, h) = (0, 0.1), (0.1, 0), (0.1, 0.1)$, while under the null of χ^2_2, the alternatives are χ^2_k with $k = 1, 3, 4$ and under the null of t_3, they are t_k with $k = 1, 2, 4$. The results were summarised by looking at p-value plots and size-power curves proposed by Wilk and Gnanadesikan (1968) and recently reviewed by Davidson and MacKinnon (1998).

A p-value plot represents the empirical distribution function $F(p)$ of the p-values obtained by the simulation of the goodness-of-fit test at the null distribution. As in this case significance values are supposed to be uniformly distributed, we expect this line to be as close to the 45 degree line as possible. A confidence bound is plotted across this bissectrice to take into account of sampling errors:

$$\left[p - 1.96\sqrt{\frac{p(1-p)}{m}}; p + 1.96\sqrt{\frac{p(1-p)}{m}}\right]$$

A size-power curve plots the empirical distribution function of the p-values at the null distribution against its counterpart at an alternative distribution. In this way we are able to compare tests with different size values as could be detected by a p-value plot. A powerful test will have a size-power curve converging very rapidly towards one.

Let us first discuss the given tests concerning their p-value plots of Figure 1 in which the left, middle and right panel respectively show the p-value plot with $G_{0,0}$, t_3 and χ^2_2 as null distribution. As all curves on the p-value plot are close to the 45 degree line, we may accept our tests to be well defined. Only the JB tests aberrates from the confidence bound. This is due to the slow rate of convergence of b_1 and b_2 to the bivariate normal limiting distribution

of γ_1 and γ_2, a remark already made in Moors et al. (1996). Note that the JB test is omitted at the t_3 distribution because of the inexistence of the third and fourth moment in this case.

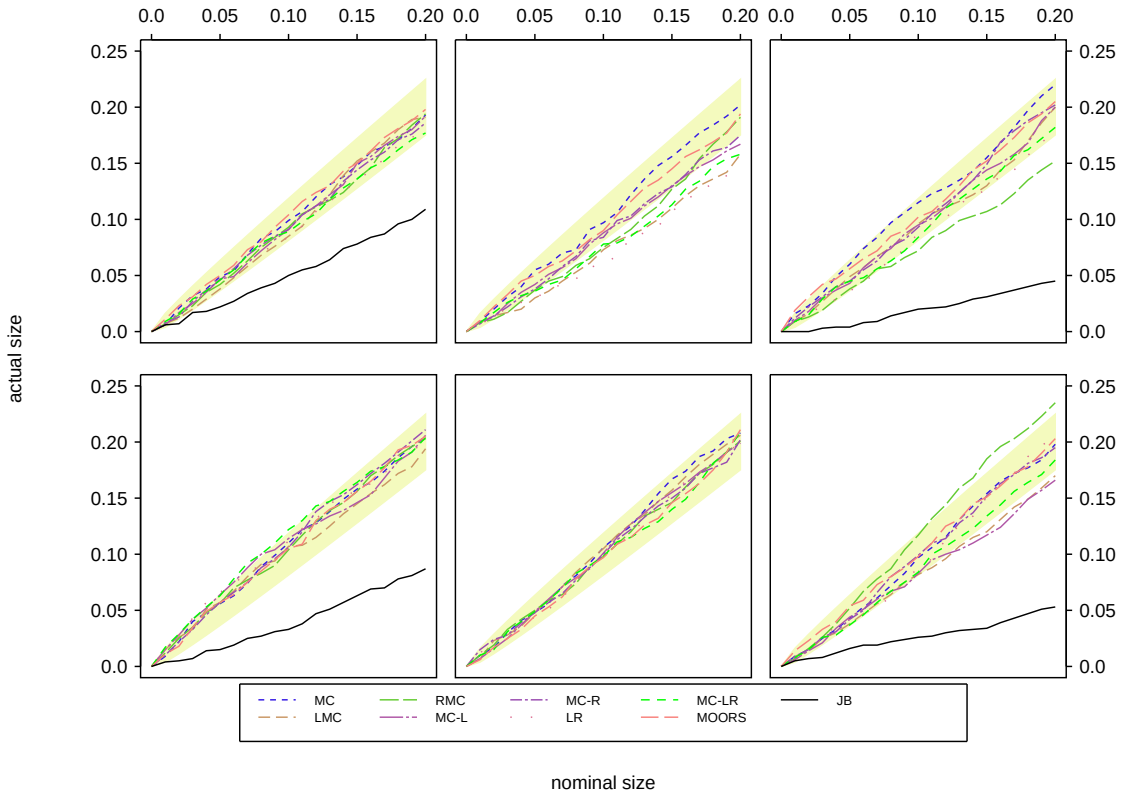


Figure 1: The p-value plots with $G_{0,0}$ (left panel), t_3 (middle panel) and χ^2_2 (right panel) as null distribution. The upper panel represents $n = 100$ and the lower panel $n = 1000$.

In Figures 2-4 size-power plots are drawn. In Figure 2 the JB test clearly outperforms all other tests. It can easily be seen that a slightly right skewed alternative (left panel) is not very well detected by LMC, RMC and LR as they only use measures of tail weight. The middle panel illustrates the incapability of MC to detect a heavier tailed distribution. A right skewed and fat tailed distribution (right panel) again is difficult to detect using only LMC. Of our proposed measures MC-LR presents the best power values. The MOORS test only has higher power values on the middle panel of Figure 2.

Secondly, we consider Figure 3 which shows the size-power plot of the given tests at the fat-tailed null distribution t_3. Again the MC test cannot detect deviations from the

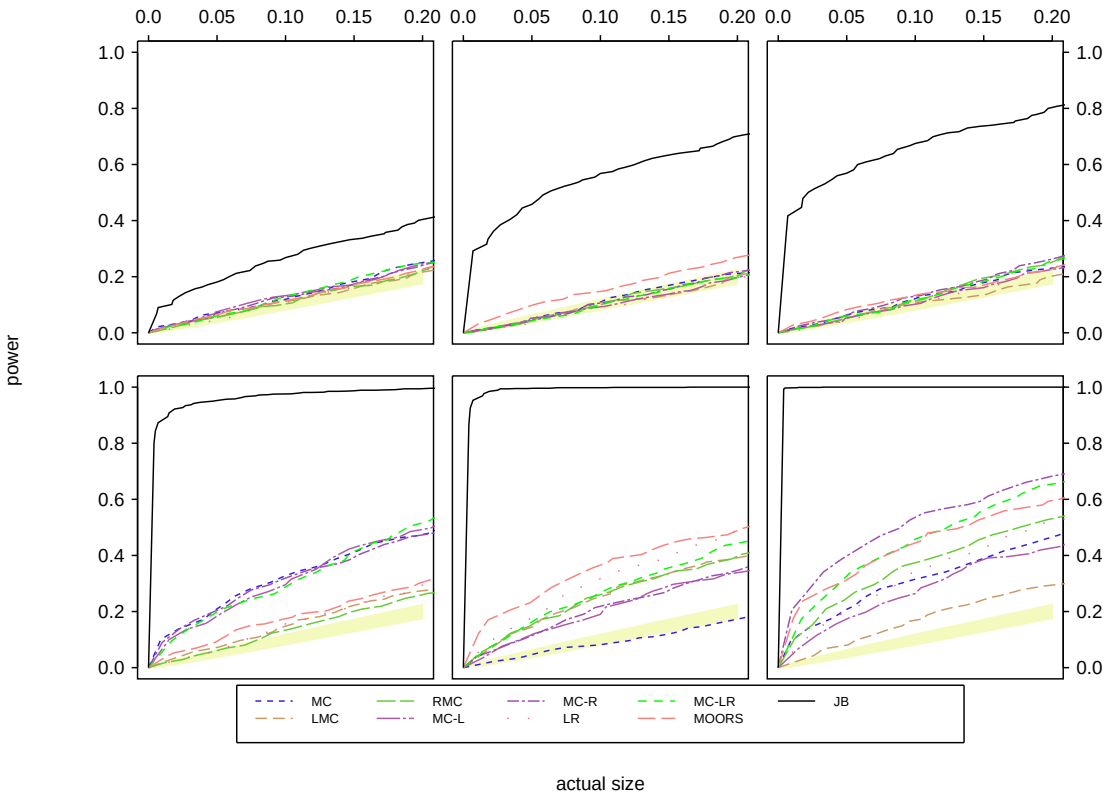


Figure 2: The size-power curve at the null distribution of $G_{0,0}$ and at the alternative distributions $G_{0.1,0}$ (left panel), $G_{0,0.1}$ (middle panel) and $G_{0.1,0.1}$ (right panel).

null distribution as all alternatives also have zero skewness. Here, the MOORS test is most optimal, followed by LR and MC-LR. The JB test is omitted because of the inexistence of the third and fourth moment at the t_3 distribution.

In case χ^2_2 is taken as the null distribution, we obtain in Figure 4 the resulting size-power plot. Here, the MC-LR test and the MC-R test appear to be the best one, although at $n = 100$ they are sometimes outperformed by the JB test.

5 Simulation study at contaminated distributions

In this section we want to compare the proposed tests with respect to their robustness. To this end, we generated here contaminated $m = 1000$ samples of size $n = 100$ and $n = 1000$ of a distribution F by taking a sample of size $n(1-\varepsilon)$ of that distribution F and

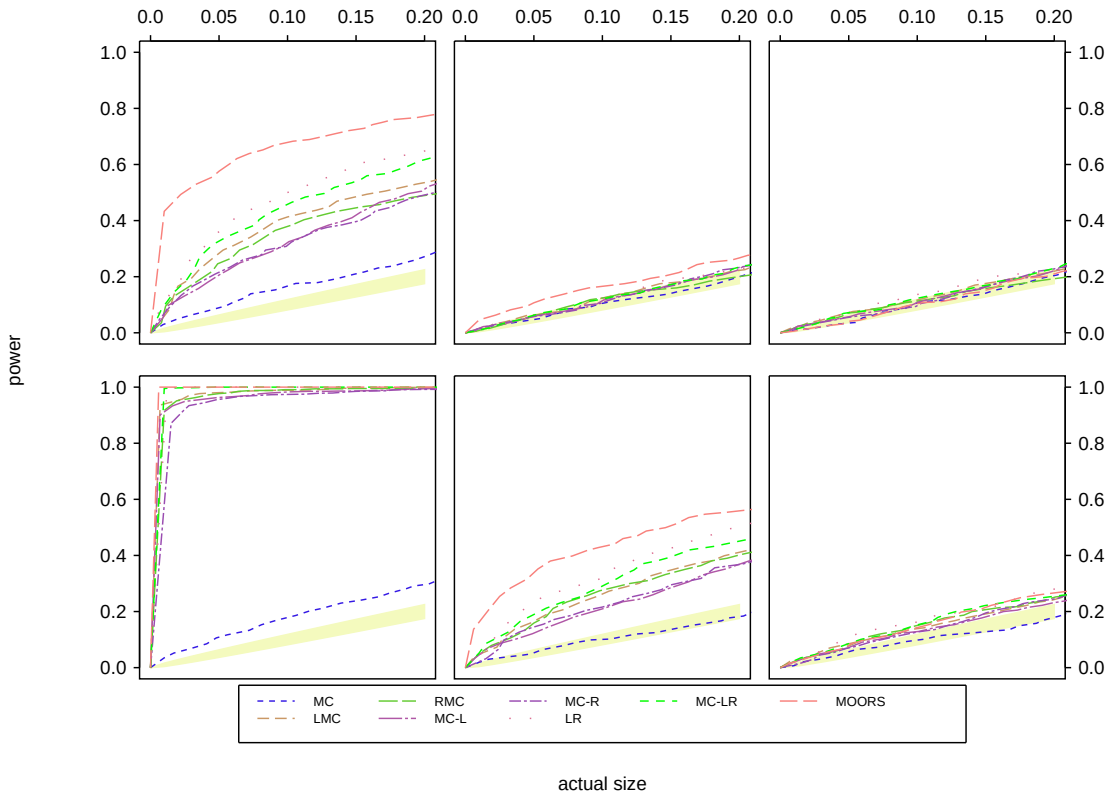


Figure 3: The size-power curve at the null distribution of t_3 and at the alternative distributions t_1 (left panel), t_2 (middle panel) and t_4 (right panel).

adding a contaminated sample of size $n\varepsilon$. The latter can be $N(F^{-1}(0.5) + 2 * F^{-1}(0.999) - F^{-1}(0.001), 0.1)$ (right contamination, RC), $N(F^{-1}(0.5), F^{-1}(0.999) - F^{-1}(0.001))$ (symmetric contamination, SC), $N(F^{-1}(0.5) + 2 * F^{-1}(0.001) - F^{-1}(0.999), 0.1)$ (left contamination, LC) or $N(F^{-1}(0.5), F^{-1}(0.51) - F^{-1}(0.49))$ (central contamination, CC). Here we have taken $\varepsilon = 0.01$ and $\varepsilon = 0.02$.

We will restrict ourselves to two tests, namely the MOORS test and MC-LR. The other proposed robust alternatives are omitted due to their lower power values. Furthermore, it is straightforward to see that the JB goodness-of-fit test is not able to handle outlying values in a robust way. Indeed, as this test is based on moments of the data, it cannot cope right, symmetric or left contamination. An exception is made at central contamination, because in that case the third and fourth moment is only slightly changed, and the JB test remains well defined here.

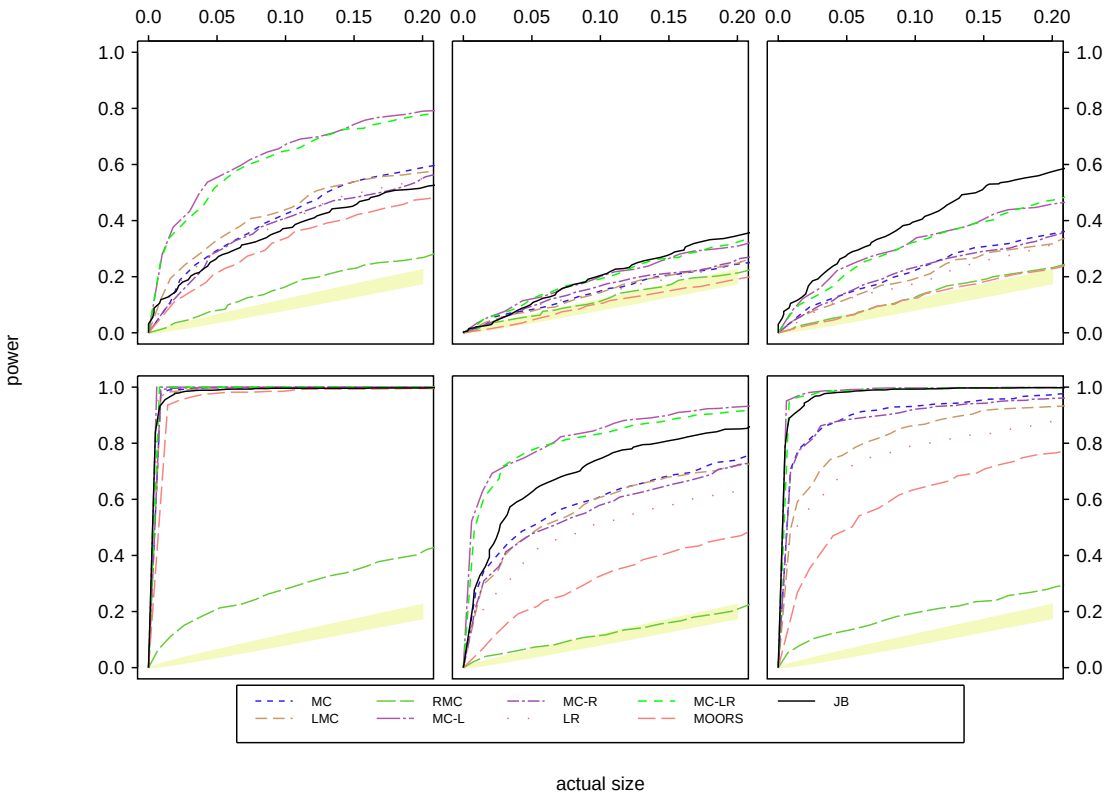


Figure 4: The size-power curve at the null distribution of χ^2_2 and at the alternative distributions χ^2_1 (left panel), χ^2_3 (middle panel) and χ^2_4 (right panel).

From Figures 5 and 6 it is straightforward to see that the MOORS test and the MC-LR test behave fairly correct in presence of outliers. When ε increases the MC-LR test deviates more strongly than the MOORS test from the confidence bound, especially at $n = 1000$ (lower panel of Figure 6). Nevertheless, compared to the generalized Jarque-Bera test, the MC-LR test is extremely better able to handle outlying values.

6 Applications

In this section we analyse four data sets which illustrate the robustness of the MOORS and the MC-LR test compared to the JB test.

The first data set comes from the Associated Examining Board in Guilford (Cresswell,

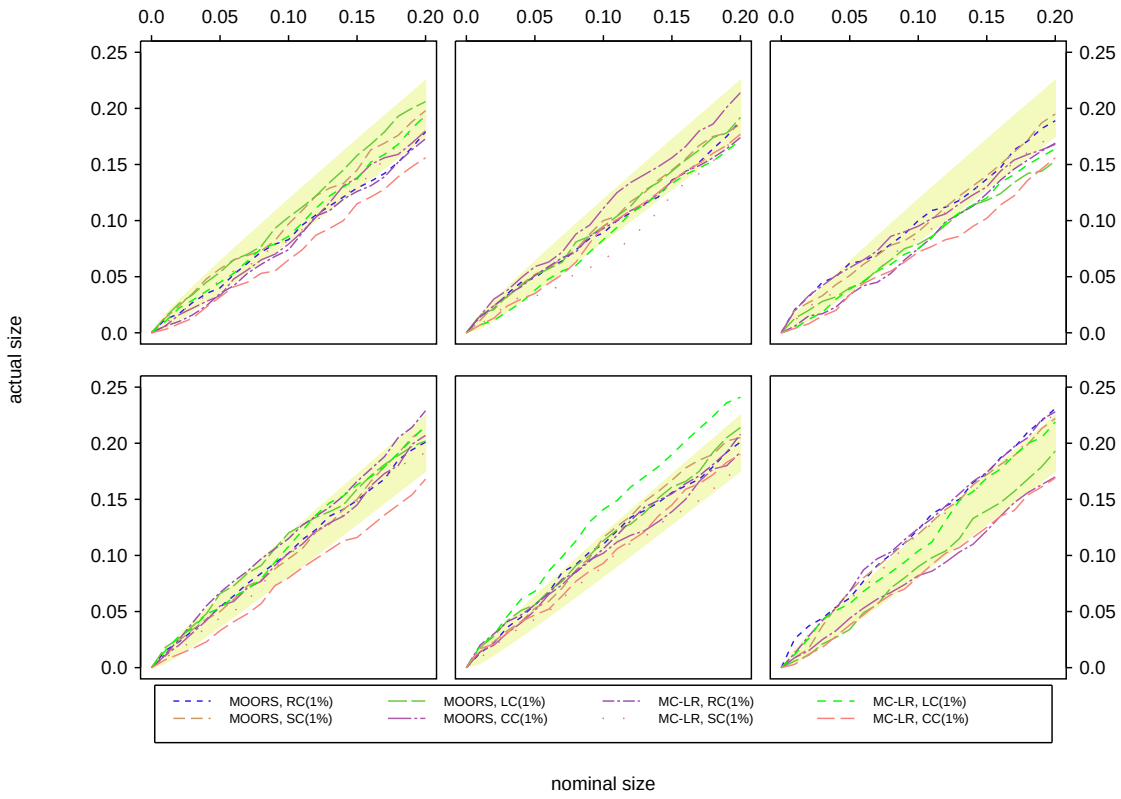


Figure 5: The p-value plot at the null distribution of $G_{0,0}$ (left panel), of t_3 (middle panel), and of χ^2_2 (right panel), contaminated case with $\varepsilon = 0.01$.

1990) and contains a sample of 1000 scores of students on the writing of a paper. From the normal QQ-plot of Figure 7(a) and the boxplot in Figure 7(b) the assumption of normality seems appropriate. Only four minor outliers are visible on the boxplot. In Table 2 the non-robustness of the JB test is illustrated. Normality is rejected at the 5% significance level when the outliers from the boxplot are included, but is accepted when they are excluded. On the contrary, the MOORS test and our proposed MC-LR test is based on the majority of the data and so they behave the same in both situations. As could be expected, they all detect normality in this data set.

The stars data set (Rousseeuw and Leroy, 1987) contains the light intensity and the surface temperature of 47 stars in the direction of Cygnus. A scatter plot of the data and the robust LTS regression line (Rousseeuw, 1984) are shown in Figure 8(a). In regression,

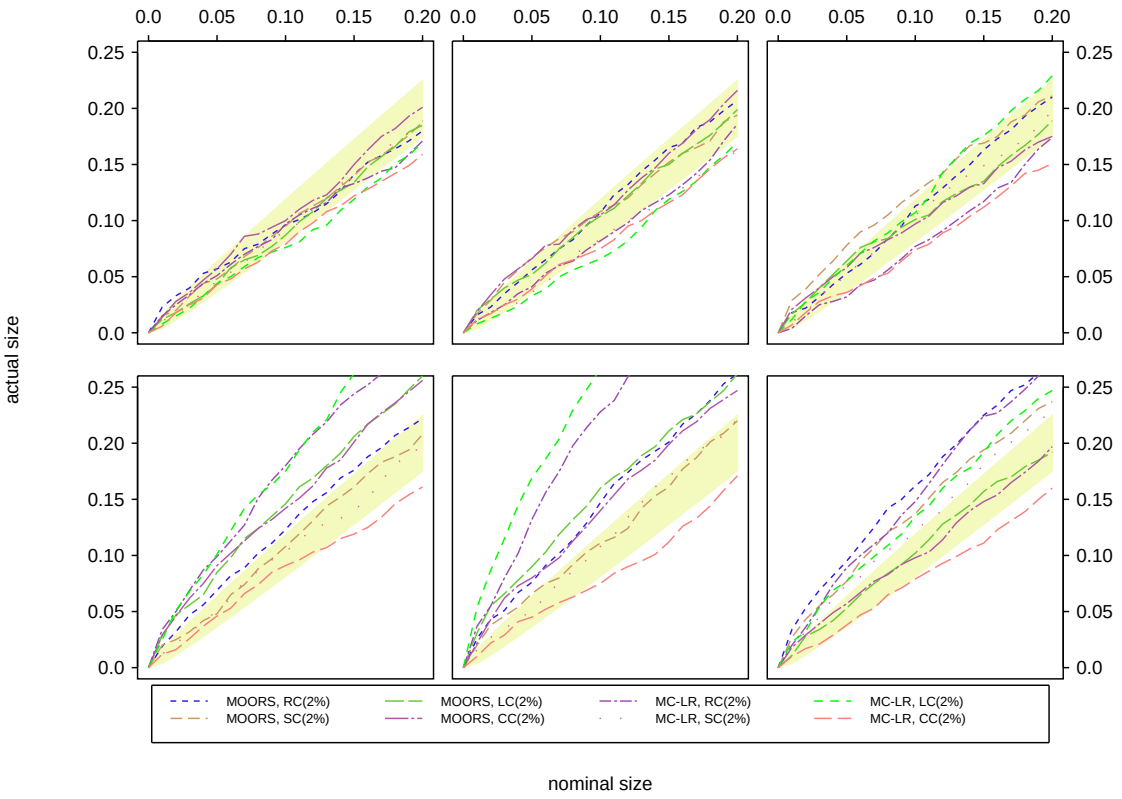
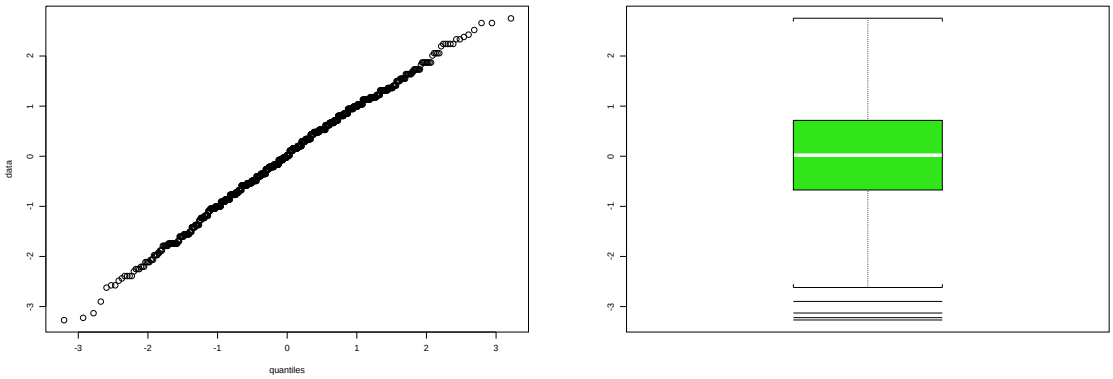


Figure 6: The p-value plot at the null distribution of $G_{0,0}$ (left panel), of t_3 (middle panel), and of χ^2_2 (right panel), contaminated case with $\varepsilon = 0.02$.



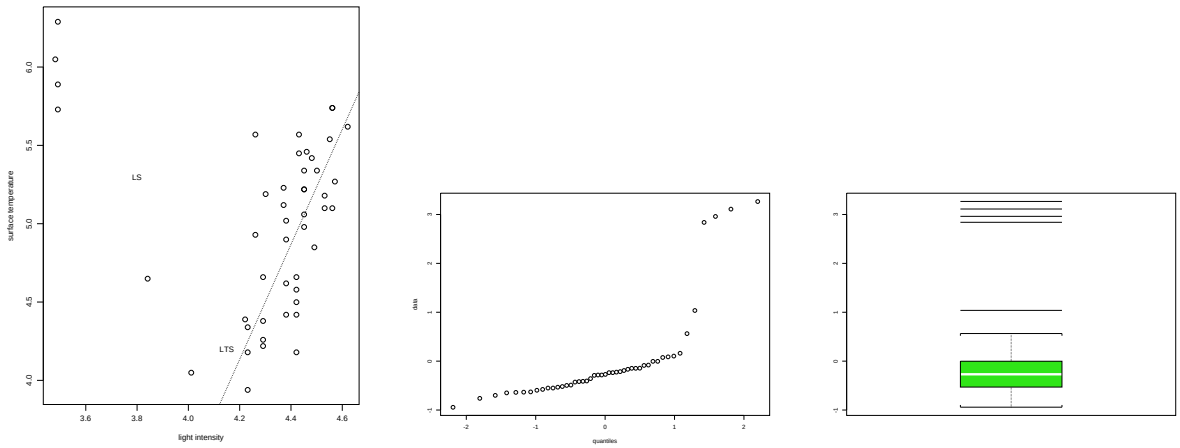
(a) (b)

Figure 7: The Guilford data: (a) normal QQ-plot; (b) boxplot.

	JB	MOORS	MC-LR
Guilford, outliers included	0.039	0.496	0.975
Guilford, outliers excluded	0.087	0.497	0.995
Stars, outliers included	0.000	0.867	0.290
Stars, outliers excluded	0.301	0.320	0.377
Baseball, outliers included	0.000	0.261	0.104
Baseball, outliers excluded	0.919	0.717	0.213
Procter, outliers included	-	0.573	0.652
Procter, outliers excluded	-	0.491	0.606

Table 2: Significance of the goodness-of-fit tests, with outliers included or excluded.

it is important to check normality of the residuals. Figure 8(b) and Figure 8(c) contain the normal QQ-plot and the boxplot of the LTS residuals, from which five clear outliers are visible. Table 2 shows again that the JB test lead to very different conclusions whether or not these outliers are included in the data. Both MOORS and MC-LR are not highly influenced by these outliers and confirm the normality assumption. We should be careful in interpreting these results as this data set is very small and consequently the robust tests are known to be very conservative. But still, this example shows the non-robustness of the JB test.



(a) (b) (c)

Figure 8: The Stars data: (a) Scatter plot with LTS regression line; (b) normal QQ-plot of the residuals; (c) boxplot of the residuals.

The baseball data (Reichler, 1991) consists of 162 major league baseball players who achieved true free agency. This means that the player could sell his services to the highest

bidding team. A player is expected to handle in two possible directions. Or he plays badly in the year of his free agency, because he is unhappy with his current team and he will play much better in the next year. Or he pushes his performance in his free agency year in order to get to a better team, but then he will play less well the next year. Here, we wanted to test whether the batting average (hits per at bat) at the free agency year and at the next year is bivariate normally distributed. Therefore we calculated the robust distances given by

$$(x - \hat{\mu})^t \hat{\Sigma}^{-1} (x - \hat{\mu})$$

in which $\hat{\mu}$ and $\hat{\Sigma}$ are the Minimum Covariance Estimator (MCD) estimates of location and scale (Rousseeuw, 1984). If the data follow a bivariate normal distribution, these robust distances are approximately χ^2_2 distributed. On the χ^2_2 based QQ-plot of Figure 9 we notice two prominent outliers. With these outliers included, the generalized Jarque-Bera test rejects the null hypothesis, but the robust tests accept that the majority of the distances are χ^2_2 distributed. When excluding these two extreme values, both the JB and the robust tests accept the null hypothesis, which implies that the original data are bivariate normally distributed. We thus see that the JB test rejects the null hypothesis only in the presence of the two outliers.

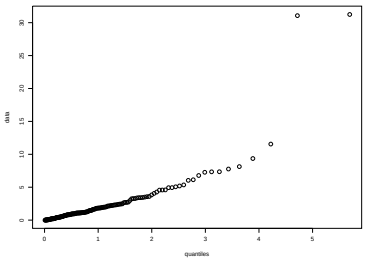


Figure 9: The Baseball data: χ^2_2 based QQ-plot of the robust distances.

From Datastream we collected the daily logaritmic returns of the Procter & Gamble stock from Januari 2000 to December 2003, leading to a univariate data set consisting of 1004 values. From the t_3 based QQ-plot of Figure 10, we could believe these data to be likewise distributed. Indeed, both the MOORS and the MC-LR test do not reject the null hypothesis, which is probably due to the majority of points which follow closely the imaginary line on the QQ-plot. Excluding the extreme value didn't change the results.

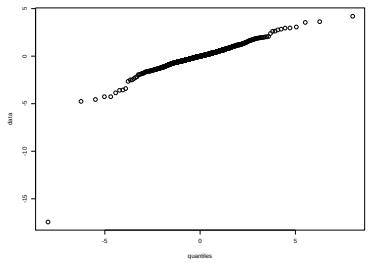


Figure 10: The Procter & Gamble data: t_3 based QQ-plot.

7 Conclusion

In this paper we discussed several goodness-of-fit tests in terms of robustness. The commonly used Jarque-Bera test of normality was extended to become a goodness-of-fit test. Main advantage of the generalized Jarque-Bera test is that its power values are reasonably high. But, by means of p-value plots and size-power curves we noted that this test often fails to lead to a correct actual size, due to the slow rate of convergence towards the limiting distribution. Moreover, the test cannot be performed at distributions without finite moments, and as it is based on moments of the data it is strongly influenced by the presence of outlying values.

Therefore Moors et al. (1996) proposed to replace the classical skewness and kurtosis coefficient by robust alternatives, leading to the MOORS test. We conducted a similar approach by using the measures proposed in Brys et al. (2004a) and in Brys et al. (2004c). Combining the medcouple (MC), a robust skewness measure, with left and right tail weight measures (LMC and RMC), we constructed the MC-LR test, which came out to be the best of our proposed robust goodness-of-fit tests. Indeed, it appeared to be well defined at the null distribution and it also appeared to be quite powerful. Compared to the MOORS test the MC-LR test has often higher power values, and comparable sensitivity towards outliers.

In practice, we recommend to perform both the JB test and the robust MC-LR test. If they give contradictory answers, this can either be due to the failure of the JB test in the presence of outliers, or due to the conservative behaviour of the MC-LR test. In that case, a further investigation of the data is required.

References

[1] Bera, A. and Jarque, C., 1981. Efficient tests for normality, heteroskedasticity and serial independence of regression residuals: Monte Carlo evidence. Economics Letter, 7, 313–318.

[2] Brys, G., Hubert, M., and Struyf, A., 2003. A Comparison of Some New Measures of Skewness. In: R. Dutter, P. Filzmoser, U. Gather and P.J. Rousseeuw (Ed.), Developments in Robust Statistics, ICORS 2001. Springer-Verlag, Heidelberg, 98–113.

[3] Brys, G., Hubert, M., and Struyf, A., 2004a. A Robust Measure of Skewness. Journal of Computational and Graphical Statistics (JCGS), to appear.

[4] Brys, G., Hubert, M., and Struyf, A., 2004b. A robustification of the Jarque-Bera test of normality. In: J. Antoch (Ed.), COMPSTAT 2004 Proceedings, Springer, Physica Verlag, to appear.

[5] Brys, G., Hubert, M., and Struyf, A., 2004c. Robust Measures of Tail Weight. Submitted.

[6] Cresswell, M.J., 1990. Gendar Effects in GCSE, some initial analyses. Research Report, Associated Examining Board, Guilford, 517.

[7] Davidson, R. and MacKinnon, J.G., 1998. Graphical methods for investigating the size and power of test statistics. The Manchester School, 66, 1–26.

[8] Hoaglin, D.C., Mosteller, F., and Tukey, J.W., 1985. Exploring Data Tables, Trends and Shapes. John Wiley and Sons, New York.

[9] Moors, J.J.A., Wagemakers, R.T.A., Coenen, V.M.J., Heuts, R.M.J., and Janssens, M.J.B.T., 1996. Characterizing systems of distributions by quantile measures. Statistica Neerlandica, 50, 417–430.

[10] Reichler, J.L., 1991. The Baseball Encyclopedia. Macmillan, New York.

[11] Rousseeuw, P.J., 1984. Least Median of Squares Regression. Journal of the American Statistical Association, 79, 871–881.

[12] Rousseeuw, P.J. and Leroy, A.M., 1987. Robust Regression and Outlier Detection. Wiley, New York.

[13] Smirnov, N. V., 1948. Table for estimating the goodness of fit of empirical distributions. Annals of Mathematical Statistics, 19, 279–281.

[14] Wilk, M.B. and Gnanadesikan, R., 1968. Probability Plotting Methods for the Analysis of Data. Biometrika, 33 (No 1), 1–17.