



HAL
open science

A modified skewness measure for testing asymmetry

Thomas H. E. Holgersson

► **To cite this version:**

Thomas H. E. Holgersson. A modified skewness measure for testing asymmetry. Communications in Statistics - Simulation and Computation, 2010, 39 (02), pp.335-346. 10.1080/03610910903453419 . hal-00569406

HAL Id: hal-00569406

<https://hal.science/hal-00569406>

Submitted on 25 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A modified skewness measure for testing asymmetry

| | |
|-------------------------------|---|
| Journal: | <i>Communications in Statistics - Simulation and Computation</i> |
| Manuscript ID: | LSSP-2008-0183.R1 |
| Manuscript Type: | Original Paper |
| Date Submitted by the Author: | 11-Mar-2009 |
| Complete List of Authors: | Holgersson, Thomas; Dept. of Economics and Statistics |
| Keywords: | Skewnes, Median, Bootstrap, graphics |
| Abstract: | Statistical practitioners frequently wish to know whether a variable is symmetrically distributed. There are a number of different tests available but the most commonly used one is perhaps that based on the standardised third central moment, as defined by Pearson and Fisher in the early 1900:s. While this traditional skewness measure uniquely determines the symmetry of a variable within the Pearson family, it does not uniquely determine symmetry for a general distribution. In this paper we propose a modified version of the classical skewness test which is easy to conduct and consistent against a wide family of asymmetric distributions. |



A modified skewness measure for testing asymmetry

By

H.E.T. Holgersson

Dept of Economics and Statistics

Jönköping International Business School

P. O. Box 1026

SE-551 11 Jönköping

Sweden

e-mail: thomas.holgersson@ihh.hj.se

Statistical practitioners frequently wish to know whether a variable is symmetrically distributed. There are a number of different tests available but the most commonly used one is perhaps that based on the standardised third central moment, as defined by Pearson and Fisher in the early 1900:s. While this traditional skewness measure uniquely determines the symmetry of a variable within the Pearson family, it does not uniquely determine symmetry for a general distribution. In this paper we propose a modified version of the classical skewness test which is easy to conduct and consistent against a wide family of asymmetric distributions.

Keywords: *Skewness, median, bootstrap, graphics.*

1. Introduction

1
2
3
4
5 Statistical analysis frequently involves assessing the shape of the distribution. In
6 particular, it is often of great importance to know if a variable is symmetric about its
7 centre or not. Examples include studies conducted of income distributions, stochastic
8 variation in quality control and pharmaceutical analyses, which requires appropriate
9 statistical methods. According to the literature it appears that in the early 19th century,
10 skewness and asymmetry were almost considered as identical concepts (for example,
11 Yule, 1917, Pearson, 1905). Later on the standardised third order central moment came
12 to dominate the literature, and it is sometimes considered to be identical to the term
13 “skewness statistic”, even though it is formally only one of several possible measures.
14 Fisher (1929) derived sampling distributions of the third central moment and therefore, in
15 a sense, put this statistic into a context of hypothesis testing. The third central moment is
16 well known to uniquely determine the (a)symmetry of distributions within the Pearson
17 family (Ord, 1972) and is nowadays a standard option in most computer packages. In the
18 1940s and later, it was recognised that tests based on this statistic cannot be expected to
19 have power against asymmetric distributions with zero third order central moment
20 (Churchill, 1946). Hence a number of alternative tests have been developed which all
21 have their own pros and cons. To mention some important contributions, Antille et al
22 (1982) considered a family of asymmetry tests for unimodal distributions based on
23 symmetric differences of gaps, while Doksum et al (1977) proposed plots with
24 confidence bounds to assess asymmetry. Csörgö and Heathcote (1987) proposed an
25 asymmetry test based on the fact that a random variable is symmetrically distributed
26 about its mean if, and only if, its characteristic function is real. In general, no test can be
27 said to be uniformly superior; issues of robustness to irrelevant distributional properties,
28 levels of complication and size/power properties vary strongly between families of tests.
29 In this paper we introduce a modification of the standardized third central moment to test
30 whether a random variable is symmetrically distributed. More specifically, the test
31 statistic proposed is shown to be a function of the standardized third central moment plus
32 a function of the difference between the mean and median. Hence it has the potential to
33 detect asymmetric distributions that tests based only on the third central moment cannot.
34 Moreover, bootstrap consistency of the test statistic is proved and hence an operational

test is available. A Monte Carlo simulation is included to investigate some size and power properties of the proposed test relative to tests based on only the third central moment. It is also demonstrated that the bootstrap consistency of the proposed statistic allows for graphical investigation of the (a)symmetry of the investigated variable. In the following sections we will pay more attention to these matters.

2. Testing symmetry versus asymmetry

A random variable X is said to be symmetric if and only if $F_X(a-x) + F_X(a+x) = 1$ for all x in \mathbb{R} . The definition itself does not lead to a feasible test of symmetry and a number of approaches have been proposed in the literature. In this section we will present a modified skewness coefficient for testing asymmetry, restricting ourselves to properties of purely continuous variables. One of the most commonly applied measures is perhaps the classical skewness coefficient defined by

$$\gamma = \mu_3 / \mu_2^{3/2} \quad (2.1)$$

where $\mu_r = E[X - E[X]]^r$ is the r :th central moment. The skewness coefficient plays an important role as a characteristic of a random variable and is frequently used as a measure of shape. The null hypothesis $H_0 : \gamma = 0$ (or equivalently $H_0 : \mu_3 = 0$) may be tested by the sample skewness defined by $\hat{\gamma} = m_3 / m_2^{3/2}$, where $m_j = (1/n) \sum_{i=1}^n (X_i - \bar{X})^j$. If X is distributed independently with kurtosis $\beta = \mu_4 / \mu_2^2 = 3$ and skewness $\gamma = 0$, then we have the well-known result

$$\sqrt{n/6} \hat{\gamma} \xrightarrow{\ell} N(0,1), \quad (2.2)$$

(e.g. Bowman and Shenton, 1975). Hence the null hypothesis $H_0 : \gamma = 0$ may be rejected when the squared left hand side of (2.2) exceeds a pre-determined percentile of the $\chi_{(1)}^2$ distribution. In the following we will refer to that test as the Fisher skewness test. An alternative test was proposed by Holgersson (2006) who showed that the distribution of

m_3 may be consistently estimated by bootstrapping, without the assumptions of homoscedasticity and a kurtosis of the value 3. This paper is not so much concerned with robustness aspects but rather the family of asymmetric distributions that the skewness coefficient can detect. For distributions within the Pearson family the γ coefficient is well known to uniquely determining whether a variable is symmetric or not (Ord, 1972). On the other hand, there are asymmetric distributions outside the Pearson family where the third central moment vanishes and the test based on (2.2) thus cannot have power against these distributions (Churchill, 1946, Stuart and Ord, 1993). We will therefore consider a modified version of γ . Let the population mean and variance be defined by $\mu = E[X]$ and $\sigma^2 = E[X - \mu]^2$ respectively and let ψ be an arbitrary measure of centrality of X that coincides with μ for symmetric distributions. We may then obtain a family of skewness measures by $\theta = \sigma^{-3} E[X - \psi]^3$. In order to investigate the relation between this statistic and the regular skewness measure we expand it about μ :

$$\theta = \sigma^{-3} E[X - \mu + \mu - \psi]^3 = \sigma^{-3} E[X - \mu]^3 + \lambda(3 + \lambda^2) \quad (2.3)$$

where $\lambda = (\mu - \psi)/\sigma$. The first term of (2.3) is clearly the γ coefficient. Thus, even though θ cannot be expected to uniquely determine symmetry for all variables, it is evidently more general than the traditional γ skewness measure and has therefore great potential for asymmetry testing. The crucial question is then which ψ to employ. Two obvious candidates are the mode (m) and median (M) respectively, but other possibilities include linear functions of order statistics (L -statistics). Ideally, one would like to have ψ far from μ for asymmetric distributions but there is also a simplicity aspect. Stuart and Ord (1993) gives a definition of unimodality and argue that $(\mu - m) \approx 3(\mu - M)$ for unimodal distributions with only moderate asymmetry, while Hotelling and Solomons (1932) argue that $(\mu - M)/\sigma$ is bounded whereas $(\mu - m)/\sigma$ is not. This speaks in favour of the mode over the median. The mode will not, however, be very practical for multimodal distributions as the modes may then be far from the centre

of the distribution, even for symmetric distributions. Moreover, using the mode will lead to difficulties in finding an operational test statistic. As simplicity is an important issue in this paper we will therefore disregard the mode and restrict ourselves to the median, M . Hence our proposed measure of asymmetry takes the form $\theta = \sigma^{-3} E[X - M]^3$, i.e.,

$$\theta = \sigma^{-3} E[X - \mu]^3 + \lambda(3 + \lambda^2) = \gamma + \nu \quad (2.4)$$

where $\lambda = (\mu - M)/\sigma$, $\gamma = \sigma^{-3} E[X - \mu]^3$ and $\nu = \lambda(3 + \lambda^2)$. A sample version of θ is obtained by

$$\hat{\theta} = (1/n) \sum_{i=1}^n \left(X_i - X_{((n+1)/2)} \right)^3 \hat{\sigma}^{-3} \quad (2.5)$$

where $X_{((n+1)/2)}$ is the sample median defined by the $(n+1)/2$:th ordered value of $\{X_i\}_{i=1}^n$, $\hat{\sigma}^2 = (1/n) \sum_{i=1}^n (X_i - \bar{X})^2$ and the sample size n is taken to be odd, i.e. $n \in (2\mathbb{N}^+ + 1)$. Even sample sizes require a small modification of the sample median (e.g. David and Nagaraja, 2003). Note that (2.5) may be expanded in two equivalent forms:

$$\hat{\theta} = \hat{\gamma} + \hat{\nu} \quad (2.6)$$

and

$$\hat{\theta} = \left(m'_3 - 3m'_2 X_{((n+1)/2)} + 3m'_1 X_{((n+1)/2)}^2 - X_{((n+1)/2)}^3 \right) / \hat{\sigma}^3 \quad (2.7)$$

where $m'_r = (1/n) \sum_{i=1}^n X_i^r$, $\hat{\gamma} = \hat{\sigma}^{-3} n^{-1} \sum_{i=1}^n (X_i - \bar{X})^3$, $\hat{\lambda} = \left(m'_1 - X_{((n+1)/2)} \right) / \hat{\sigma}$ and $\hat{\nu} = \hat{\lambda}(3 + \hat{\lambda}^2)$. Equation (2.6) is convenient in examining the difference between $\hat{\theta}$ and the $\hat{\gamma}$ while (2.7) is more convenient when establishing asymptotical properties. In particular, if $\mu_r = E(X_i^r) < \infty$ and X_i is iid distributed with $E[X_i^{9+\delta}] < \infty$ then $E[(m'_r - \mu_r)^2] = o_p(n^{1/2-\delta})$ $r=1,2,3$, for all $\delta > 0$. Furthermore, if X_i is continuous

and $0 < f(M -) = f(M +) < \infty$ (the left and right derivatives of the densities at M are equal and finite) we have the well known result $n^{1/2} \left\{ X_{((n+1)/2)} - M \right\} \xrightarrow{\ell} N(0, (1/4)/f^2(M))$ (e.g. Serfling 1980). Hence it follows that $E \left[\left(X_{((n+1)/2)} - M \right)^2 \right] = o_p(n^{1/2-\delta})$ and so $\hat{\theta} \xrightarrow{m.s.} \theta$ and $\hat{\lambda} \xrightarrow{m.s.} \lambda$. In other words, the point estimates of θ and λ are consistent and have limiting distributions. We will now establish bootstrap consistencies of the test statistic.

Theorem 1: Let X_i be a strictly continuous iid random variable such that $0 < f(M -) = f(M +) < \infty$ and $E[X_i]^{9+\delta} < \infty$, $\delta > 0$. Also, let $\phi = E[X - M]^3$, $\hat{\phi} = (1/n) \sum_{i=1}^n \left((X_i - M) + (M - X_{((n+1)/2)}) \right)^3$ and $\hat{\phi}^*$ be the bootstrap version of $\hat{\phi}$.

Then $\rho_\infty \left(P_* \left(\sqrt{n} (\hat{\phi}^* - \hat{\phi}) \leq x \right) - P \left(\sqrt{n} (\hat{\phi} - \phi) \leq x \right) \right) \xrightarrow{p} 0$, where ρ_∞ is the sup-norm distance and P_* denotes bootstrap probability.

Proof:

Note that $\hat{\phi} = (1/n) \sum_{i=1}^n \left((X_i - M) + (M - X_{((n+1)/2)}) \right)^3 = \tilde{\mu}_3 - 3\tilde{\delta}\tilde{\mu}_2 + 3\tilde{\delta}^2\tilde{\mu}_1 - \tilde{\delta}^3$ where $\tilde{\mu}_k = (1/n) \sum_{i=1}^n (X_i - M)^k$ and $\tilde{\delta} = (X_{((n+1)/2)} - M)$. The finite ninth order moment is sufficient for $\rho_\infty \left(P_* \left(\sqrt{n} (\tilde{\mu}_k^* - \tilde{\mu}_k) \leq x \right) - P \left(\sqrt{n} (\tilde{\mu}_k - \mu_k) \leq x \right) \right) \xrightarrow{a.s.} 0$, $k = 1, 2, 3$, (see Holgersson, 2006) while the smoothness of f is sufficient for

$\rho_\infty \left(P_* \left(\sqrt{n} \left(X_{((n+1)/2)}^* - X_{((n+1)/2)} \right) \leq x \right) - P \left(\sqrt{n} \left(X_{((n+1)/2)} - M \right) \leq x \right) \right) \xrightarrow{a.s.} 0$ (Shao and Tu 1995). Finally, since $\hat{\phi}$ is a measurable function of $\tilde{\mu}_k$ and $\tilde{\delta}^k$ ($k = 1, 2, 3$) it follows that $\hat{\phi}^*$ converges in bootstrap probability to $\hat{\phi}$.

Corollary 1: Let $\theta := \phi/\sigma^3$, $\hat{\theta} := \hat{\phi}/\hat{\sigma}^3$ and $\hat{\theta}^* := \hat{\phi}^*/\hat{\sigma}^{3*}$ be the bootstrapped version of $\hat{\theta}$. Then $\rho_\infty \left(P_* \left(\sqrt{n} (\hat{\theta}^* - \hat{\theta}) \leq x \right) - P \left(\sqrt{n} (\hat{\theta} - \theta) \leq x \right) \right) \xrightarrow{p} 0$. This follows from Theorem 1 since $\rho_\infty \left(P_* \left(\sqrt{n} (\hat{\sigma}^{2*} - \hat{\sigma}^2) \leq x \right) - P \left(\sqrt{n} (\hat{\sigma}^2 - \sigma^2) \leq x \right) \right) \xrightarrow{a.s.} 0$ (Shao and Tu 1995) and $\hat{\theta}$ is a smooth, measurable function of $\hat{\phi}$ and $\hat{\sigma}^2$.

Corollary 2:

Let $\nu := \lambda \{3 + \lambda^2\}$, $\hat{\nu} := \hat{\lambda} \{3 + \hat{\lambda}^2\}$, $\hat{\lambda} = (\bar{X} - X_{(n/2)})/\hat{\sigma}^2$ and let $\hat{\nu}^* := \hat{\lambda}^* \{3 + \hat{\lambda}^{*2}\}$ be the bootstrapped version of $\hat{\nu}$. Also, let $\gamma = \mu_3/\sigma^3$ where $\hat{\gamma} = m_3/\hat{\sigma}^3$ where $m_3 = (1/n) \sum_{i=1}^n (X_i - \bar{X})^3$ and let $\hat{\gamma}^* = m_3^*/\hat{\sigma}^{3*}$ be the bootstrapped version of $\hat{\gamma}$. Then

$$\rho_\infty \left(P_* \left(\sqrt{n} (\hat{\gamma}^* - \hat{\gamma}) \leq x, \sqrt{n} (\hat{\nu}^* - \hat{\nu}) \leq y \right) - P \left(\sqrt{n} (\hat{\gamma} - \gamma) \leq x, \sqrt{n} (\hat{\nu} - \nu) \leq y \right) \right) \xrightarrow{p} 0,$$

that is, $\hat{\gamma}^*$ and $\hat{\nu}^*$ converge jointly in \square^2 . This follows since $\hat{\nu}^*$ converges according to Theorem 1 and Corollary 1. The convergence of $\hat{\gamma}^*$ is established in Holgersson (2006). Finally, when $(\hat{\gamma}^*, \hat{\nu}^*)$ are obtained jointly from the same resample, the joint distribution follows.

Hence, according to Corollary 1 the proposed test statistic $\hat{\theta}$ may be consistently bootstrapped and the hypothesis $H_0 : \theta = \theta_0$ against $H_A : \theta \neq \theta_0$, where $\theta_0 = 0$, may be tested on the α -level by rejecting H_0 when $\left(\sqrt{n} (\hat{\theta} - \theta_0) \right)^2$ exceeds the $(1-\alpha)B$:th ordered value from $\left\{ \left(\sqrt{n} (\hat{\theta}_b^* - \hat{\theta}) \right)^2 \right\}_{b=1}^B$.

Before the properties of the proposed test is examined further some things should be noted. Firstly, the proposed statistic θ is here mainly used to provide a simple and (in some sense) useful test of symmetry versus asymmetry. In case the analysis primarily

1
2
3 involves the problem of comparing skewness between two distributions, e.g.
4 $G(x)$ vs. $F(x)$, then other measures may be more adequate. For example, just because
5 $G(x)$ is more skewed than $F(x)$ with respect to γ , it does not imply that the same
6 skewness ordering is preserved with respect to θ . More specifically, the λ statistic (and
7 hence the θ statistic) does not preserve the skewness ordering in the sense of van Zwet
8 (1964). An in-depth discussion of orderings of various skewness/asymmetry measures,
9 including γ and λ , is given in MacGillivray (1986). Secondly, apart from the power to
10 detect “unexpected” asymmetric distributions (e.g. with zero third central moment) there
11 is also an issue of the power to detect “expected” deviations from asymmetry. In
12 particular, since the γ coefficient does uniquely determine (a)symmetry of Pearson
13 distributions, and such distributions arise naturally within many sciences, it is tractable to
14 have a test with high power to detect these distributions. Since the γ coefficient is an
15 ingredient within the proposed measure it, the test based on it may indeed be expected to
16 have relatively high power against such distributions. Thirdly, as there are two terms of
17 the θ statistic there is a possibility that they are of opposite sign and hence cancel out
18 each other. Specifically, since $(3 + \lambda^2) > 0$ and $\sigma > 0$, the sign of λ will play an
19 important role in the comparison of θ and γ . Sufficient conditions for determining the
20 sign of $(\mu - M)$ are given in van Zweet (1979) and Groenveld and Meeden (1984). A
21 particularly important contribution is that if x is unimodal and $f(\mu + x) - f(\mu - x)$
22 change sign once in x , then λ and γ will bear the same sign (MacGillivray, 1981) so
23 that the proposed test should have relatively high power against these distributions.
24 However, there does exist distributions such that γ and ν will have opposite signs, even
25 though the set of such distributions seem rather small (empirically). Fortunately this
26 problem need not be that serious. According to Corollary 2, there is an option of
27 analysing the distribution of $\hat{\gamma}$ and $\hat{\nu}$ separately or even jointly. For example, bootstrap
28 histograms of $\hat{\gamma}^*$ and $\hat{\nu}^*$ will be helpful to empirically assess their individual signs.
29 Cases when one of them is significantly smaller than zero and the other significantly
30 larger than zero would indicate a peculiar distribution of X , and further investigation such
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 as plots of ordered observations (e.g. Doksum, 1977, Wilk and Gnanadesikan, 1968)
4 should be conducted. In the next section we will investigate some basic properties of the
5 proposed test, including the option of separate analysis of γ and ν .
6
7
8
9

10 11 12 **3. Empirical investigation** 13

14
15
16 In this section we will investigate some of the properties of the proposed test and also
17 compare it with the bootstrapped γ coefficient and the traditional Fisher skewness test
18 based on the asymptotic $\chi_{(1)}^2$ distribution. We will investigate the size and power
19 properties by calculating the rejection frequencies of $r = 10\ 000$ replications of different
20 symmetric and asymmetric distributions respectively, using $B = 99$ bootstrap resamples
21 in each replicate. The distributions used in the simulation are specified in Table 1 below.
22 Details of the distributions may be found in e.g. Johnson et al, (1995) and Ramberg et. al
23 (1979). Moreover, to demonstrate the possibility of investigating the signs of γ vs. ν ,
24 bootstrap histograms of $\{\hat{\gamma}_b^*\}_{b=1}^B$ and $\{\hat{\nu}_b^*\}_{b=1}^B$ respectively have been obtained for a single
25 drawing of sample size $n = 10000$ using $B = 399$ bootstrap resamples (Fig. 1-3).
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Insert Table 1 about here

Power simulations are usually conducted by letting the parameter of relevance increase away from the null hypothesis while keeping the sample size fixed. This is not quite appropriate when investigating tests for distributional properties as there is no unique way to quantify asymmetry. We will instead investigate the proposed test for a fixed distribution and let the sample size increase from 30 pseudo observations up to 10 000. All simulations involve testing at the nominal level of 5%. In Table 2 below we display the rejection frequencies of the three tests when operating on the normal distribution and $t_{(4)}$ distribution respectively. All tests are seen to behave well for the normal distribution. For the $t_{(4)}$ distribution the tests of the bootstrapped γ and θ statistic over reject with up

1
2
3 to 14%, which may seem high. It should, however, be remembered that the null
4 distribution of these tests assumes all first nine moments to be finite, while the $t_{(4)}$
5 distribution lacks moments above order four. Thus it is curious to see that the bootstrap
6 tests actually do limit their nominal levels asymptotically even though higher order
7 moments do not exist. The null distribution of the Fisher test requires the kurtosis
8 coefficient to be 3, and it is clearly seen that it diverges for the $t_{(4)}$ distribution. We
9 therefore present it in brackets in cases when it is not valid (i.e. for all distributions but
10 the normal distribution). As regards the other two symmetric distributions in Table 3, the
11 proposed test is seen to behave well and rapidly limits its nominal size, as does the test
12 based on the bootstrapped γ skewness coefficient. The Fisher test is again seen to behave
13 peculiarly. In Table 4 some multimodal distributions are investigated. The trimodal
14 distribution is symmetric and the rejection frequencies stays fairly close to the nominal
15 size for this distribution though both the bootstrapped γ test and the proposed test over
16 rejects, but it is revealing that multimodality does not impose any serious distortions.
17 Also, both bootstrapped tests detect the bimodal asymmetric t -distribution but the
18 rejection frequencies does not quite limit 100% within the investigated sample sizes. The
19 bootstrap histograms of the separate components of the proposed statistic ($\hat{\gamma}$ and $\hat{\nu}$) is
20 presented in Figure 1 where it is seen that they have the same sign although the $\hat{\nu}$
21 statistic has a higher absolute value and also a smaller sampling variation. In Table 5
22 some power properties of unimodal distributions are investigated. The gamma
23 distribution with shape parameter equal to 500 yields a distribution that is very close to
24 the normal distribution though not exactly symmetric. The proposed test is clearly
25 consistent against this distribution though slightly inferior to that of the bootstrapped γ
26 coefficient. The power properties against the exponential distribution are clearly higher
27 than those of the near-normal distribution, and again the bootstrapped skewness
28 coefficient is slightly better than the proposed test. The rejection frequencies for the
29 asymmetric distributions with zero third order central moment are shown in Table 6.
30 Indeed, the power of the bootstrapped γ coefficient limits the nominal size (0.05) for the
31 centred gamma distribution while the proposed test is consistent as expected, in the sense
32 that the rejection frequency increases monotonically with the sample size. Moreover,
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 even though the Fisher test simultaneously tests against the kurtosis and skewness, the
4 power of the proposed test is nevertheless of the same magnitude as that of the Fisher
5 test. The same pattern is seen from the Tukey's lambda distribution where the rejection
6 frequencies of the bootstrapped γ coefficient again limits 0.05 while the proposed test
7 fairly quickly limits a rejection frequency of 100%. Moreover, the sampling distributions
8 of the separate components of the proposed statistic are displayed in Figures 2 and 3, and
9 clearly reveal that $\hat{\gamma}$ is centered about zero while $\hat{\nu}$ has no overlap at zero. These figures
10 hence gives some insight to the very different behavior of the rejection frequencies of
11 Table 6, and also demonstrates the use of separate analyses of the two components of the
12 proposed statistic. Finally, it should be stressed that even though the power of the
13 proposed test may seem rather low against some of the distributions, the level of
14 asymmetry (whatever the criteria) vary largely among the distributions which must be
15 taken into consideration. To sum up the discussion, the size properties of the proposed
16 test are more or less the same as those of the test of the bootstrapped γ skewness
17 coefficient but it does have additional power against asymmetric distributions which tests
18 based on γ cannot detect. Hence the proposed test should be a useful supplement to other
19 tests for asymmetry versus symmetry.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

36 **Insert Table 2 about here**

37 **Insert Table 3 about here**

38 **Insert Table 4 about here**

39 **Insert Table 5 about here**

40 **Insert Figure 1 about here**

41 **Insert Figure 2 about here**

42 **Insert Figure 3 about here**
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

4. Conclusive summary

In this paper we propose a modification of the traditional skewness coefficient (defined by the standardized third central moment) to test for symmetry versus (a)symmetry. Instead of the third moment being centred about its mean value it is centred about the median. This adjusted measure is shown to be a function of the traditional skewness coefficient plus a function of the difference between the mean and median. While the proposed measure might not be optimal for the problem of ordering the amount of skewness between two distributions, it is argued that it is more useful than the traditional skewness test for the problem of discriminating between symmetric and asymmetric distributions, specifically for distributions with zero third order central moment. Bootstrap consistency is proved for the sample version of the proposed measure and so an operational test procedure is available. Moreover, a Monte Carlo simulation was conducted to investigate some size and power properties of the proposed test versus two tests based on the third central moment, and it is demonstrated that the new test is indeed consistent against asymmetric distributions which tests based on the classical skewness measure cannot detect. The test only relies on some moment restrictions and that the distribution being analysed is smooth at the median. In particular, it is demonstrated that multimodality imposes no distortions to the test. It is also shown that the two terms within the proposed test statistic may be investigated separately by bootstrap histograms. This graphical option will provide some additional insight into the (a)symmetry of the variable being investigated. Generally, the proposed test is not meant to replace other methods, such as graphical tools, but rather it is seen as a supplementary test. In particular, the proposed test along with separate analysis of the two components of the proposed measure and plots of ordered observations should provide a powerful tool set for investigating whether a variable is symmetrically distributed.

References

Antille A et al. (1982). Testing symmetry. JASA 77:639-646.

Bowman, K. O. and L. A. Shenton (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 . Biometrika, 62(2):243-250.

Churchill, E. (1946). Information given by odd moments. Ann. Math. Statist. 17:244-256.

Csörgö S, Heathcote C R (1987). Testing for symmetry. Biometrika 74(1):177-184.

David, H. A. and H. N. Nagaraja (2003). *Order statistics*. Third ed. Wiley.

Doksum K A et al.(1977). Plots and tests for symmetry. Biometrika 64(3):473-487.

Fisher, R. A. (1929). *Moments and product moments of sampling distributions*. Proceedings of the London mathematical society, 199-238.

Groenveld, R. A. and Meeden, G. (1977). The Mode, median and mean inequality. Am. Stat. 31(3):120-121.

Holgerson, H. E. T. (2006). Robust testing for skewness. Communications in Statistics-Theory and Methods 36(3):485-498.

Hotelling H. and Solomons L. M. (1932). The limits of a measure of skewness. Ann. Math. Statist. 3, 141.

Johnson, N. L., Kotz S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol 1-Vol2. Second ed. Wiley.

1
2
3 MacGillivray, H. L. (1981). The mean, median, mode inequality and skewness for a
4 class of densities. *Austral. J. Statist.* 23(2):247-250.
5
6

7
8 MacGillivray, H. L. (1986). Skewness and assymetry: Measures and orderings, *Ann.*
9 *Stat.* 14(3):994-1011.
10
11

12
13 Ord, J. K. (1972). *Families of frequency distributions*. Griffin's statistical monographs &
14 courses.
15
16

17
18 Pearson, K. (1905). Das fehlergesetz und seine verallgemeinerungen durch Fechner und
19 Pearson“. A rejoinder. *Biometrika.* 4(1/2):169-212.
20
21

22
23 Ramberg, J. S. et. al. (1979). A probability distribution and its uses in fitting data.
24 *Technometrics.* 21(2):201-214.
25
26

27
28 Stuart, A. and Ord J. K. (1993). *Kendall's Advanced Theory of Statistics*. Vol 1. Arnold,
29 London.
30
31

32
33 Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley, New
34 York.
35
36

37
38 Shao, J. and Tu D. (1995). *The Jacknife and Bootstrap*. Springer, New York Berlin
39 Heidelberg.
40
41

42
43 Yule, G. U. (1917). *An introduction to the theory of statistics*. C. Griffen and company.
44
45

46
47 Van Zwet, W. R. (1964). Convex transformations: A new approach to skewness and
48 kurtosis. *Statistika Neerlandica*, 18:433-441.
49
50

51
52 van Zwet, W. R. (1979). Mean, median, mode II. *Statistica Neerlandica*, 33:1-5.
53
54

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Wilk, M. B. and Gnanadesikan R. (1968). Probability plotting methods for analysis of data. *Biometrika*. 55:1-17.

For Peer Review Only

Figure 1. Bootstrap histogram of the bimodal $t_{(4)}$ distribution

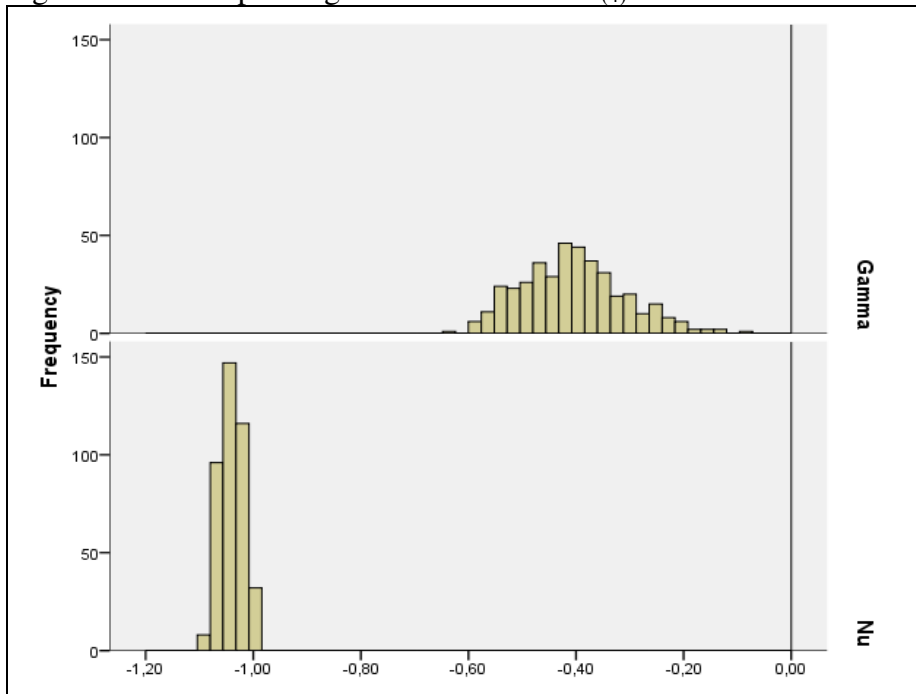


Figure 2. Bootstrap histogram of centred gamma distribution

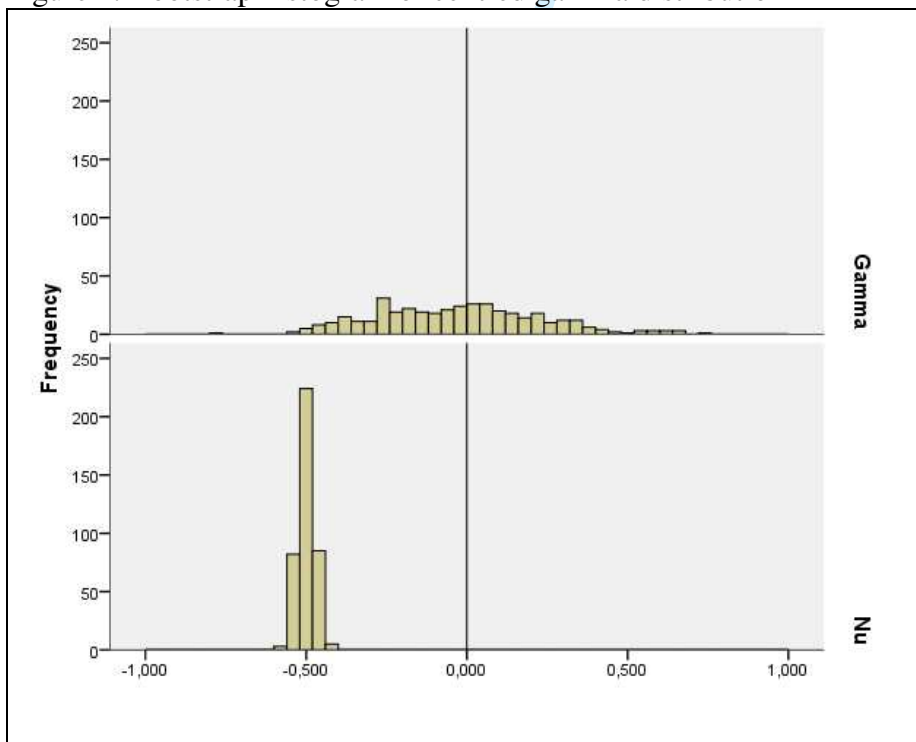
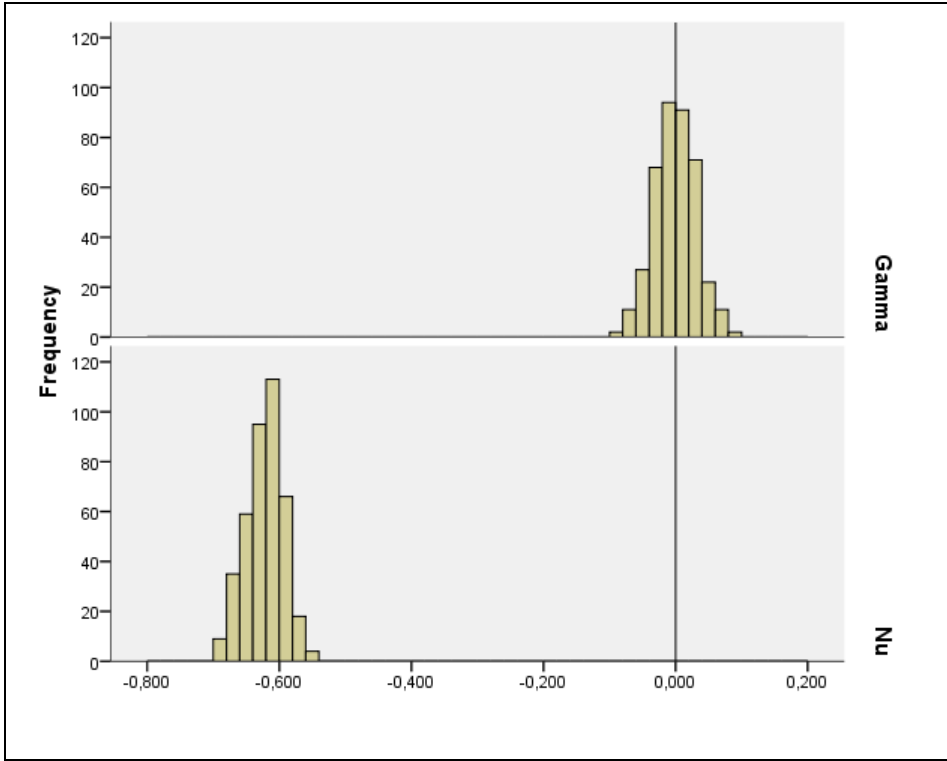


Figure 3. Bootstrap histogram of the Tukey's lambda distribution

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Review Only

Table 1. Variates used in the simulation

| Distribution | Parameter values | Shape | Skewness coefficient |
|-------------------------------------|--|--------------------|----------------------|
| Standard normal | $N(0,1)$ | Bell-shaped | 0 |
| Students t -distribution | $t_{(4)}$ | Bell-shaped | 0 |
| Beta | $B(0.5, 0.5)$ | U-shaped | 0 |
| Uniform | $U(0,1)$ | Box-shaped | 0 |
| *Gamma | $G(\text{rate} = 1, \text{shape} = 500)$ | Asymmetric | 0.004 |
| Exponential | $Exp(1)$ | Asymmetric | 2 |
| **Bimodal $t_{(4)}$ | $0.3(-1 + t_{(4)}) + 0.7(4 + t_{(4)})$ | Bimodal asymmetric | -0.540 |
| **Trimodal normal | $\frac{1}{3}N(-10, 2) + \frac{1}{3}N(0, 4) + \frac{1}{3}N(10, 2)$ | Trimodal symmetric | 0 |
| * $(Y_2 - E[Y_2]) - (X_2 - E[X_2])$ | $Y_2 \sim G(\text{rate} = 1, \text{shape} = 0.125)$ $X_2 \sim G(\text{rate} = 4^{(1/3)}, \text{shape} = 0.5)$ | Asymmetric | 0 |
| Tukey's lambda distribution | $\lambda_3 = 35.498, \lambda_4 = 2.297$ | Asymmetric | 0 |

*The gamma distribution is defined inconsistently in the literature. Here we define the gamma distribution

to be a variable X with characteristic function $E[e^{it}] = (\text{rate}/(\text{rate} + it))^{\text{shape}}$.

**The addition here refers to "mixing", with the respective probabilities adding up to one.

Table 2. Standard normal distribution t(4)

| Nobs | γ | θ | Fisher | | γ | θ | Fisher |
|-------|----------|----------|--------|--|----------|----------|---------|
| 30 | 0.048 | 0.039 | 0.033 | | 0.113 | 0.084 | (0.297) |
| 50 | 0.049 | 0.049 | 0.043 | | 0.132 | 0.103 | (0.385) |
| 100 | 0.049 | 0.047 | 0.043 | | 0.137 | 0.112 | (0.501) |
| 300 | 0.051 | 0.050 | 0.053 | | 0.119 | 0.108 | (0.644) |
| 500 | 0.050 | 0.051 | 0.048 | | 0.108 | 0.100 | (0.682) |
| 2000 | 0.049 | 0.046 | 0.051 | | 0.075 | 0.072 | (0.790) |
| 5000 | 0.052 | 0.052 | 0.052 | | 0.061 | 0.062 | (0.846) |
| 10000 | 0.049 | 0.051 | 0.050 | | 0.054 | 0.054 | (0.871) |

Table 3. Beta Uniform

| Nobs | γ | θ | Fisher | | γ | θ | Fisher |
|-------|----------|----------|---------|--|----------|----------|---------|
| 30 | 0.038 | 0.053 | (0.005) | | 0.034 | 0.039 | (0.002) |
| 50 | 0.047 | 0.055 | (0.004) | | 0.046 | 0.040 | (0.001) |
| 100 | 0.044 | 0.050 | (0.003) | | 0.051 | 0.044 | (0.001) |
| 300 | 0.050 | 0.054 | (0.002) | | 0.048 | 0.047 | (0.001) |
| 500 | 0.049 | 0.051 | (0.003) | | 0.048 | 0.048 | (0.001) |
| 2000 | 0.049 | 0.051 | (0.003) | | 0.050 | 0.052 | (0.001) |
| 5000 | 0.050 | 0.049 | (0.004) | | 0.053 | 0.054 | (0.001) |
| 10000 | 0.051 | 0.050 | (0.003) | | 0.051 | 0.049 | (0.001) |

Table 4. Trimodal normal Bimodal t(4)

| Nobs | γ | θ | Fisher | | γ | θ | Fisher |
|-------|----------|----------|---------|--|----------|----------|---------|
| 30 | 0.089 | 0.046 | (0.047) | | 0.495 | 0.496 | (0.209) |
| 50 | 0.085 | 0.038 | (0.043) | | 0.626 | 0.741 | (0.354) |
| 100 | 0.073 | 0.031 | (0.035) | | 0.778 | 0.930 | (0.642) |
| 300 | 0.065 | 0.045 | (0.044) | | 0.883 | 0.974 | (0.938) |
| 500 | 0.065 | 0.051 | (0.052) | | 0.901 | 0.976 | (0.966) |
| 2000 | 0.061 | 0.056 | (0.064) | | 0.946 | 0.990 | (0.994) |
| 5000 | 0.066 | 0.059 | (0.065) | | 0.960 | 0.992 | 1 |
| 10000 | 0.064 | 0.059 | (0.071) | | 0.970 | 0.993 | 1 |

Table 5. Gamma Exponential

| Nobs | γ | θ | Fisher | | γ | θ | Fisher |
|-------|----------|----------|--------|--|----------|----------|---------|
| 30 | 0.063 | 0.022 | (.039) | | 0.779 | 0.437 | (0.867) |
| 50 | 0.073 | 0.036 | (.053) | | 0.832 | 0.673 | (0.987) |
| 100 | 0.083 | 0.054 | (.068) | | 0.884 | 0.873 | (1) |
| 300 | 0.137 | 0.101 | (.126) | | 0.968 | 0.980 | (1) |
| 500 | 0.189 | 0.145 | (.178) | | 0.991 | 0.996 | (1) |
| 2000 | 0.574 | 0.473 | (.559) | | 1 | 1 | (1) |
| 5000 | 0.915 | 0.840 | (.911) | | 1 | 1 | (1) |
| 10000 | 0.997 | 0.986 | (.997) | | 1 | 1 | (1) |

Table 6. Centred Gamma Tukey's lambda

| Nobs | γ | θ | Fisher | γ | θ | Fisher |
|--------------|----------|----------|---------|----------|----------|---------|
| 30 | 0.210 | 0.129 | (0.617) | 0.239 | 0.179 | (0.102) |
| 50 | 0.171 | 0.152 | (0.680) | 0.224 | 0.237 | (0.121) |
| 100 | 0.137 | 0.179 | (0.744) | 0.184 | 0.343 | (0.077) |
| 300 | 0.109 | 0.215 | (0.796) | 0.102 | 0.635 | (0.092) |
| 500 | 0.099 | 0.230 | (0.805) | 0.086 | 0.821 | (0.164) |
| 2000 | 0.084 | 0.322 | (0.827) | 0.069 | 1 | (0.081) |
| 5000 | 0.083 | 0.454 | (0.846) | 0.068 | 1 | (0.105) |
| 10000 | 0.075 | 0.618 | (0.843) | 0.062 | 1 | (0.158) |