Asymptotic distributions and performance of empirical skewness measures

Andreas Eberl, Bernhard Klar*

Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Germany.

Abstract

Over the years, a number of skewness measures have been proposed in the literature to be applied to theoretical distributions. However, the corresponding empirical counterparts have been analyzed only rarely, especially with respect to their asymptotic properties and limit distributions. In this paper, we consider six of these empirical measures. After discussing some general properties, we derive the limiting distribution for each measure under weak assumptions. The performance of these estimators is analyzed in simulations using tests and the coverage probabilities of confidence intervals. In doing so, a particular focus is put on the standardized central third moment as the most popular measure of skewness. Since it turns out to behave poorly, especially when sample sizes are small, we recommend the use of alternative and more suitable skewness measures. A real data application illustrates some of our findings.

Keywords: Asymmetry, Skewness, Skewness estimator, Asymptotic Normality.

1. Introduction

The skewness of a probability distribution is an important characteristic in many applications; typical examples are (right-skewed) income distributions or the (left-skewed) age at death for a population. Over the years, many different measures have been proposed to quantify and compare distributions with regard to their skewness. Additionally, some properties have been established, which should be satisfied by a skewness measure in order to be considered adequate. However, very little attention has been given to the empirical counterparts of these measures and their performance as estimators for small or moderate sample sizes, be it in simulations or real data applications.

It is the purpose of this work to systematically introduce these empirical measures and analyze their properties, in particular their asymptotic distributions. Based on these results, we conduct a simulation study, comparing the various skewness estimators on samples of specific distributions.

*Corresponding author

Email addresses: andreas.eberl@kit.edu (Andreas Eberl), bernhard.klar@kit.edu (Bernhard Klar)

First, let us introduce six skewness measures considered in this work. Generally, we only assume that the underlying distribution function F or the corresponding random variable $X \sim F$ is non-degenerate. If, additionally, $\mathbb{E}|X|^3 < \infty$, the moment skewness [see 25] is defined by

$$\gamma_M(X) = \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^3\right],$$

where $\mu_X = \mathbb{E}X$ and $\sigma_X^2 = \mathbb{V}[X]$. Denoting the quantile function of X by q_X, we define the quantile skewness [see 3, 14] by

$$\gamma_Q^{(\alpha)}(X) = \frac{q_X(1-\alpha) + q_X(\alpha) - 2q_X(1/2)}{q_X(1-\alpha) - q_X(\alpha)} \tag{1}$$

for $\alpha \in (0, 1/2)$ if $q_X(1-\alpha) > q_X(\alpha)$. Integration of numerator and denominator with respect to α yields the integrated quantile skewness [see 12], given by

$$\gamma_{IQ}(X) = \frac{\mu_X - q_X(1/2)}{\mathbb{E}|X - q_X(1/2)|},$$

if $\mathbb{E}|X| < \infty$. If we strengthen this assumption to $\mathbb{E}X^2 < \infty$, Pearson's skewness measure [see 25, 30] is similarly defined by

$$\gamma_P(X) = \frac{\mu_X - q_X(1/2)}{\sigma_X}.$$

Denoting the expectile function of a random variable X with finite mean by e_X, Eberl and Klar [7] proposed the expectile skewness

$$\gamma_E^{(\alpha)}(X) = \frac{1}{1-2\alpha}\frac{e_X(1-\alpha) + e_X(\alpha) - 2e_X(1/2)}{e_X(1-\alpha) - e_X(\alpha)}$$

for $\alpha \in (0, 1/2)$. Unsurprisingly, this definition resembles the definition of the quantile skewness in (1), since expectiles can be seen as a smoothed version of quantiles, and also measure non-central location. As limiting value of the expectile skewness for $\alpha \to 1/2$, we obtain Tajuddin's skewness measure [see 29, 7]

$$\gamma_T(X) = 2F(\mu_X) - 1$$

as our last candidate.

The properties S1. to S3., gathered e.g. by Groeneveld and Meeden [12] or Oja [24], appear to be appropriate for a skewness measure γ:

S1. For $c > 0$ and $d \in \mathbb{R}$, $\gamma(cX + d) = \gamma(X)$.
S2. The measure γ satisfies $\gamma(X) = -\gamma(X)$.
S3. If F and G, the cdf's of X and Y, are continuous, and F is smaller than G in convex transformation order (i.e. $G^{-1}(F(x))$ is convex, written $F \leq_c G$), then $\gamma(X) \leq \gamma(Y)$.

Additionally, normalized skewness measures are preferable for better comparability and interpretability. The convex transformation order mentioned in property S3 is only valid for a subset of all distributions on $\mathbb{R}$, most notably for the subset of all distributions with continuous cdf's. For arbitrary cdf's, $F \leq_c G$ holds if $G^{-1}(F(x))$ is convex on $D_F = \mathbb{R} \setminus F^{-1}(\{0,1\})$ and $\inf G(D_G) \geq \inf F(D_F)$ holds [see 8]. If and under which conditions the considered skewness measures satisfy properties S1 - S3 is summarized in Table 1. The depicted results are gathered from Groeneveld and Meeden [12], Eberl and Klar [7, 8], Tajuddin [29]. These results are only valid if the so called central quantiles are used for the calculation of $\gamma_Q^{(\alpha)}$, γ_{IQ} and γ_P. For $p \in (0,1)$, they are defined as the mean of the right quantile $\sup\{t \in \mathbb{R} : F(t) \leq p\}$ and the left quantile $\inf\{t \in \mathbb{R} : F(t) \geq p\}$. If left quantiles are used instead, weak assumptions are needed for $\gamma_Q^{(\alpha)}$, γ_{IQ} and γ_P to satisfy S2; furthermore, γ_P is then normalized to $(-1,1]$.

Measure	Standardization	S1	S2	S3
γ_M	No	Yes	Yes	Yes if $\sup G(D_G) \leq \sup F(D_F)$
$\gamma_Q^{(\alpha)}$	Yes to $[-1,1]$	Yes	Yes	Yes generally
γ_{IQ}	Yes to $[-1,1]$	Yes	Yes	Yes generally
γ_P	Yes to $(-1,1)$	Yes	Yes	No
$\gamma_E^{(\alpha)}$	Yes to $(-1,1)$	Yes	Yes	Unknown
γ_T	Yes to $(-1,1)$	Yes	Yes if $\mathbb{P}(X = \mu_X) = 0$	Yes if F is continuous

Table 1: Summary of properties of the six skewness measures when considering central quantiles.

This paper is organized as follows. In section 2, we recall the definitions and some properties of quantiles and expectiles. In section 3, empirical counterparts of the six skewness measures are introduced, and their limit distributions and further asymptotic properties are discussed. In Section 4, we compare the different measures for specific families of distributions. Theoretical skewness values are compared in subsection 4.1, whereas the behavior of the pertaining estimators is analyzed in subsection 4.2, using tests and confidence intervals. The behavior of the different skewness measures and their empirical counterparts is illustrated with a real data example in section 5.

2. Quantiles and expectiles

For a cdf F, $X \sim F$ and $p \in (0,1)$, we define the p-quantile of X as

$$q_p = F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) \geq p\},$$

thereby using the so-called left quantile throughout this paper. For independent and identically distributed (iid) random variables $X_1, \ldots, X_n, \ldots$ with empirical cdf $\hat{F}_n(t) =$

$1/n \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}}$, $t \in \mathbb{R}$, the empirical p-quantile is defined as

$$\hat{q}_p = \hat{F}_n^{-1}(p) = X_{(\lceil np \rceil)},$$

where $X_{()} = (X_{(1)}, ..., X_{(n)})$ denotes the order statistic of $X_1, ..., X_n$ and $\lceil \cdot \rceil$ denotes the ceiling function. A connection between the theoretical and the empirical p-quantile is given by the Bahadur representation as follows.

Proposition 1 (Ghosh [11, Theorem 1]). *Let F be a cdf with existing and non-negative derivative f at the point q_p for some $p \in (0,1)$. Then*

$$\hat{q}_p = q_p + \frac{(1 - \hat{F}_n(q_p)) - (1-p)}{f(q_p)} + R_n^{(p)},$$

where $\sqrt{n} R_n^{(p)} \xrightarrow{\mathbb{P}} 0$.

Based on this representation, one can derive the following multivariate central limit theorem for quantiles.

Theorem 2 (Serfling [27, section 2.3.3, Theorem B]). *Let $k \in \mathbb{N}$ and $p_1, ..., p_k \in (0,1)$ with $p_1 < p_2 < ... < p_k$. Furthermore, let F be a cdf with existing density f in neighbourhoods of the quantiles $q_{p_1}, ..., q_{p_k}$, satisfying $f(q_{p_1}), ..., f(q_{p_k}) > 0$. Then, for iid $X_1, ..., X_n \sim F$,*

$$\sqrt{n} \left(\begin{pmatrix} \hat{q}_{p_1} \\ \vdots \\ \hat{q}_{p_k} \end{pmatrix} - \begin{pmatrix} q_{p_1} \\ \vdots \\ q_{p_k} \end{pmatrix} \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_Q'\right),$$

where

$$\left(\Sigma_Q'\right)_{i,j} = \frac{\min\{p_i, p_j\}\left(1 - \max\{p_i, p_j\}\right)}{f(q_{p_i}) f(q_{p_j})}$$

for $i, j \in \{1, ..., k\}$.

Additionally, the strong consistency of the empirical quantiles can be shown.

Proposition 3 (Serfling [27, section 2.3.1, pp.74-75]). *Let $p \in (0,1)$. If $F(x) > p \ \forall x > q_p$, then $\hat{q}_p \xrightarrow{a.s.} q_p$, i.e. $\hat{q}_p$ is a strongly consistent estimator of q_p.*

As an alternative to the p-quantile, which minimizes the loss function

$$p\mathbb{E}[(X-t)^+] + (1-p)\mathbb{E}[(X-t)^-],$$

one can also consider the minimizer of the loss function

$$p\mathbb{E}[((X-t)^+)^2] + (1-p)\mathbb{E}[((X-t)^-)^2],$$

the so-called p-expectile $e_X(p)$ [see 22]. This minimizer is always unique and can be more formally defined for any integrable X as

$$e_p = e_X(p) = \arg\min_{t \in \mathbb{R}} \{\mathbb{E}[\ell_p(X - t) - \ell_p(X)]\},$$

where

$$\ell_p(t) = t^2(p\mathbb{1}_{[0,\infty)}(t) + (1-p)\mathbb{1}_{(-\infty,0)}(t)).$$

The p-expectile is therefore uniquely characterized by the corresponding first-order condition

$$p\mathbb{E}[(X-e_p)^+] = (1-p)\mathbb{E}[(X-e_p)^-],$$

which is equivalent to $\mathbb{E}[I_p(e_p, X)] = 0$, where

$$I_p(x,y) = p(y-x)\mathbb{1}_{\{y\geq x\}} - (1-p)(x-y)\mathbb{1}_{\{y<x\}}, \quad x, y \in \mathbb{R},$$

denotes the so-called identification function. Some basic properties of expectiles are summarized in the following proposition, collected from Newey and Powell [22] and Bellini et al. [2].

Proposition 4. *Let $X \in L^1$ with cumulative distribution function (cdf) F and $p \in (0,1)$. Then*

a) $e_{X+h}(p) = e_X(p) + h$, for each $h \in \mathbb{R}$,
b) $e_{\lambda X}(p) = \lambda e_X(p)$, for each $\lambda > 0$,
c) $e_X(p)$ is strictly increasing with respect to p,
d) $e_X(p)$ is continuous with respect to p,
e) $e_{-X}(p) = -e_X(1-p)$,
f) for continuous cdf F, e_X has derivative

$$e_X'(p) = \frac{E\,|X - e_X(p)|}{(1-p)F(e_X(p)) + p\,(1-F(e_X(p)))}.$$

The empirical p-expectile $\hat{e}_p = \hat{e}_n(p)$ of a sample $X_1, ..., X_n$ is defined as solution of the empirical analogue of the identification function condition

$$I_p(t, \hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n} I_p(t, X_i) = 0, \quad t \in \mathbb{R}. \tag{2}$$

As with quantiles, a multivariate central limit theorem as well as strong consistency can be proved for empirical expectiles [see 15].

Theorem 5. *Let $k \in \mathbb{N}$ and $p_1, ..., p_k \in (0,1)$ with $p_1 < p_2 < ... < p_k$. Furthermore, let F be a cdf with existing first two moments and without a point mass at any of the points $e_{p_1}, ..., e_{p_k}$. Then, for iid $X_1, ..., X_n \sim F$,*

$$\sqrt{n}\left(\begin{pmatrix} \hat{e}_{p_1} \\ \vdots \\ \hat{e}_{p_k} \end{pmatrix} - \begin{pmatrix} e_{p_1} \\ \vdots \\ e_{p_k} \end{pmatrix}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_E'\right),$$

where

$$(\Sigma_E')_{i,j} = \frac{\mathbb{E}[I_{p_i}(e_{p_i}, X_1) I_{p_j}(e_{p_j}, X_1)]}{(p_i + F(e_{p_i})(1-2p_i))\left(p_j + F(e_{p_j})(1-2p_j)\right)}$$

for $i, j \in \{1, ..., k\}$.

Proposition 6. *Let $p \in (0,1)$. If the first moment of the cdf F exists, then $\hat{e}_p \xrightarrow{a.s.} e_p$, i.e. $\hat{e}_p$ is a strongly consistent estimator of e_p.*

3. Empirical skewness measures and their asymptotics

As general setting, we consider iid random variables $X, X_1, X_2, \ldots \sim F$ for some non-degenerate distribution function F on $\mathbb{R}$. In each subsection, we only consider cdf's F for which the corresponding theoretical skewness measure is well-defined.

3.1. Moment skewness

The empirical moment skewness is obtained as plug-in estimator, i.e.

$$\hat{\gamma}_M(X_1, ..., X_n) = \frac{1}{n} \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^3}{\hat{\sigma}_n^3} = \frac{\overline{X_n^3} - 3\overline{X_n^2}\bar{X}_n + 2\bar{X}_n^3}{(\overline{X_n^2} - \bar{X}_n^2)^{3/2}},$$

where $\bar{X}_n = 1/n \sum_{i=1}^n X_i$ denotes the first empirical moment and $\overline{X_n^k} = 1/n \sum_{i=1}^n X_i^k$ the k-th empirical moment for $k = 2, 3, \ldots$. Throughout this section, we use the factor $1/n$ in any mean.

Since $\hat{\gamma}(X_1, ..., X_n)$ is a differentiable function of the first three empirical moments, the multivariate central limit theorem together with the δ-method yields part a) of the following theorem, whereas the strong law of large numbers together with Slutzky's Theorem yields part b). For this, we introduce the the short-hand notation $\mu = \mu_X$ as well as $\mu_k = \mathbb{E}X^k$ for the k-th theoretical moment of X, $k = 2, 3, \ldots$.

Theorem 7. *a) If $\mathbb{E}X^6 < \infty$, then*

$$\sqrt{n}(\hat{\gamma}_M(X_1, ..., X_n) - \gamma_M(X)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_M^2(F)),$$

where

$$\sigma_M^2(F) = \frac{\mu_6 - \mu_3^2}{\sigma^6} + \frac{3\left[(\mu_2^2 - \mu_3\mu)(3\mu_2^2 - 2\mu_4 - \mu_3\mu) - (\mu_5 - \mu_3\mu_2)(\mu_3 - \mu_2\mu)\right]}{\sigma^8} + \frac{9(\mu_3 - \mu_2\mu)^2(\mu_4 + 3\mu_2^2 - 4\mu_3\mu)}{4\sigma^{10}}$$

[see, e.g. 13].

b) $\hat{\gamma}_M(X_1, ..., X_n)$ is a strongly consistent estimator of $\gamma_M(X)$, i.e. $\hat{\gamma}_M(X_1, ..., X_n) \xrightarrow{a.s.} \gamma_M(X)$.

There exist a number of further distributional results for the moment skewness under normality, particularly for small samples (see, e.g. [10], [6], [5] and [21]).

Since the asymptotic variance is a function of the first six theoretical moments, a variance estimator $\hat{\sigma}_M^2(F)$ can easily be obtained as plug-in estimator, replacing the theoretical moment by the corresponding empirical moments. It follows directly that, if $\mathbb{E}X^6 < \infty$, $\hat{\sigma}_M^2(F)$ is a strongly consistent estimator of $\sigma_M^2(F)$. However, since higher empirical moments converge very slowly, large sample sizes are required in order that these asymptotic results manifest themselves in simulations or real data applications (see sections 4.2 and 5).

3.2. Quantile skewness

We define the empirical quantile skewness once again as plug-in estimator, i.e.

$$\hat{\gamma}_Q^{(\alpha)}(X_1, ..., X_n) = \frac{\hat{q}_{1-\alpha} + \hat{q}_\alpha - 2\hat{q}_{1/2}}{\hat{q}_{1-\alpha} - \hat{q}_\alpha}, \quad \alpha \in (0, 1/2).$$

As with the empirical moment skewness, the central limit theorem for quantiles (Theorem 2) and the δ-method yield a central limit theorem for $\hat{\gamma}_Q^{(\alpha)}$. The strong consistency in part b) is a consequence of Proposition 3.

Theorem 8. *a) Let $\alpha \in (0, 1/2)$ and let F be a cdf with existing density f in a neighbourhood of q_p with $f(q_p) > 0$ for $p \in \{\alpha, 1/2, 1-\alpha\}$. Then*

$$\sqrt{n}\left(\hat{\gamma}_Q^{(\alpha)}(X_1, ..., X_n) - \gamma_Q^{(\alpha)}(X)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_Q^2(F, \alpha))$$

with

$$\begin{aligned}\sigma_Q^2(F,\alpha) = &- \frac{4\alpha^2}{(q_{1-\alpha} - q_\alpha)^4}\left[\frac{(q_{1-\alpha} - q_{1/2})}{f(q_\alpha)} - \frac{(q_{1/2} - q_\alpha)}{f(q_{1-\alpha})}\right]^2 \\ &+ \frac{4\alpha}{(q_{1-\alpha} - q_\alpha)^4}\left[\left(\frac{(q_{1-\alpha} - q_{1/2})}{f(q_\alpha)}\right)^2 + \left(\frac{(q_{1/2} - q_\alpha)}{f(q_{1-\alpha})}\right)^2\right] \\ &- \frac{4\alpha}{(q_{1-\alpha} - q_\alpha)^3 f(q_{1/2})}\left[\frac{(q_{1-\alpha} - q_{1/2})}{f(q_\alpha)} + \frac{(q_{1/2} - q_\alpha)}{f(q_{1-\alpha})}\right] \\ &+ \frac{1}{(q_{1-\alpha} - q_\alpha)^2 f(q_{1/2})^2}.\end{aligned}$$

b) Let $\alpha \in (0, 1/2)$ and, for $p \in \{\alpha, 1/2, 1-\alpha\}$, let F be a cdf satisfying $F(t) > p$ for all $t > q_p$ and . Then

$$\hat{\gamma}_Q^{(\alpha)}(X_1, ..., X_n) \xrightarrow{a.s.} \gamma_Q^{(\alpha)}(X),$$

i.e. $\hat{\gamma}_Q^{(\alpha)}(X_1, ..., X_n)$ is a consistent estimator for $\gamma_Q^{(\alpha)}(X)$.

A related test for the nullity of $\gamma_Q^{(\alpha)}$ along with corresponding asymptotic results is considered in [23].

3.3. Integrated quantile skewness

The integrated quantile skewness is estimated by

$$\hat{\gamma}_{IQ}(X_1, ..., X_n) = \frac{\bar{X}_n - \hat{q}_{1/2}}{1/n \sum_{i=1}^n |X_i - \hat{q}_{1/2}|}.$$

Before deriving a central limit theorem for this estimator, we introduce some notation. We denote the mean absolute deviation from the median (median MAD, for short) by $M = M^{(q_{1/2})} = \mathbb{E}|X - q_{1/2}|$, and its empirical version by $\hat{M}_n = \hat{M}_n^{(q_{1/2})} = 1/n \sum_{i=1}^n |X_i - \hat{q}_{1/2}|$. Assuming $F(q_{1/2}) = 1/2$, we have $M = \mu - 2\mu^{(q_{1/2})}$, where $\mu^{(q_{1/2})} = \mathbb{E}[X \mathbb{1}_{\{X \le q_{1/2}\}}]$. Based on this representation of M, we define its higher-order equivalents by $M_k = \mu_k - 2\mu_k^{(q_{1/2})}$ for $k = 2, 3, \ldots$, where $\mu_k^{(q_{1/2})} = \mathbb{E}[X^k \mathbb{1}_{\{X \le q_{1/2}\}}]$, $k = 2, 3, \ldots$. Then, we can prove the following lemma, borrowing some ideas from Lin, Wu and Ahmad [19, Theorem 2.1] and Babu and Rao [1, Theorem 2.5]).

Lemma 9. *Let F be a cdf with existing second moment. If the derivative f of F exists at $q_{1/2}$ with $f(q_{1/2}) > 0$, then*

$$\sqrt{n}\left(\begin{pmatrix} \hat{q}_{1/2} \\ \bar{X}_n \\ \hat{M}_n \end{pmatrix} - \begin{pmatrix} q_{1/2} \\ \mu \\ M \end{pmatrix}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{IQ}),$$

where

$$\Sigma_{IQ} = \begin{pmatrix} \frac{1}{4(f(q_{1/2}))^2} & \frac{M}{2f(q_{1/2})} & \frac{\mu - q_{1/2}}{2f(q_{1/2})} \\ \frac{M}{2f(q_{1/2})} & \mu_2 - \mu^2 & M_2 - M(\mu + q_{1/2}) \\ \frac{\mu - q_{1/2}}{2f(q_{1/2})} & M_2 - M(\mu + q_{1/2}) & \mu_2 - 2\mu q_{1/2} + q_{1/2}^2 - M^2 \end{pmatrix}.$$

Proof. First of all, the assumptions imply that F is continuous at $q_{1/2}$. Hence, $F(q_{1/2}) = 1/2$, which implies $M = \mu - 2\mu^{(q_{1/2})}$.

Now, define for $i \in \{1, \ldots, n\}$ the random vector $Z_i = (\mathbb{1}_{\{X_i > q_{1/2}\}}, X_i, |X_i - q_{1/2}|)^T$. By assumption, these are iid with expectation $\mathbb{E}[Z_1] = (1/2, \mu, M)^T$ and covariance matrix $\mathrm{Cov}(Z_1, Z_2) = S = (s_{i,j})$, where

$$\begin{aligned} s_{1,1} &= \mathbb{V}[\mathbb{1}_{\{X_1 > q_{1/2}\}}] = 1/4, \\ s_{2,2} &= \mathbb{V}[X_1] = \mu_2 - \mu^2, \\ s_{3,3} &= \mathbb{V}[|X_1 - q_{1/2}|] = \mu_2 - 2\mu q_{1/2} + q_{1/2}^2 - M^2, \\ s_{1,2} &= \mathrm{Cov}(\mathbb{1}_{\{X_1 > q_{1/2}\}}, X_1) = \tfrac{1}{2}M, \\ s_{1,3} &= \mathrm{Cov}(\mathbb{1}_{\{X_1 > q_{1/2}\}}, |X_1 - q_{1/2}|) = \tfrac{1}{2}(\mu - q_{1/2}), \\ s_{2,3} &= \mathrm{Cov}(X_1, |X_1 - q_{1/2}|) = M_2 - M(\mu + q_{1/2}). \end{aligned}$$

Applying the Lindeberg-Lévy central limit theorem yields

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \mathbb{E}[Z_1]\right) = \sqrt{n}\begin{pmatrix} (1 - \hat{F}_n(q_{1/2})) - 1/2 \\ \bar{X}_n - \mu \\ \hat{M}_n' - M \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, S), \tag{3}$$

where $\hat{M}_n' = 1/n \sum_{i=1}^{n} |X_i - q_{1/2}|$.

We now turn our attention to the first component of (3). Dividing it by $f(q_{1/2})$ preserves the limit distribution due to Slutzky's Theorem, if the first row and the first column of the asymptotic covariance matrix are also divided by $f(q_{1/2})$. This transforms S into Σ_{IQ}, and we have

$$\sqrt{n}\begin{pmatrix} [(1 - \hat{F}_n(q_{1/2})) - 1/2]/f(q_{1/2}) \\ \bar{X}_n - \mu \\ \hat{M}_n' - M \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{IQ}).$$

The Bahadur representation (Proposition 1) now states that the first component is equal to $\hat{q}_{1/2} - q_{1/2} + R_n$ with $\sqrt{n}R_n \xrightarrow{\mathbb{P}} 0$, so it can be replaced by $\hat{q}_{1/2} - q_{1/2}$ without changing the limit distribution.

It remains to consider the third component of (3); proving $\sqrt{n}(\hat{M}_n' - \hat{M}_n) \xrightarrow{\mathbb{P}} 0$ is sufficient to obtain the asserted result. Using the sign function $\operatorname{sgn}(t) = \mathbb{1}_{(0,\infty)}(t) - \mathbb{1}_{(-\infty,0]}, t \in \mathbb{R}$, we infer that

$$\begin{aligned} |a| - |a-b| &= \int_0^a \operatorname{sgn}(t)\mathrm{d}t - \int_0^{a-b} \operatorname{sgn}(t)\mathrm{d}t = \int_{a-b}^a \operatorname{sgn}(t)\mathrm{d}t = -\int_b^0 \operatorname{sgn}(a-t)\mathrm{d}t \\ &= b\int_0^1 \operatorname{sgn}(a-tb)\mathrm{d}t \end{aligned}$$

for all $a, b \in \mathbb{R}$. By choosing $a = X_i - q_{1/2}$ and $b = \hat{q}_{1/2} - q_{1/2}$, we obtain

$$\begin{aligned} |X_i - q_{1/2}| - |X_i - \hat{q}_{1/2}| &= |X_i - q_{1/2}| - |X_i - q_{1/2} - (\hat{q}_{1/2} - q_{1/2})| \\ &= (\hat{q}_{1/2} - q_{1/2}) \int_0^1 \operatorname{sgn}(X_i - q_{1/2} - t(\hat{q}_{1/2} - q_{1/2}))\mathrm{d}t \\ &= (\hat{q}_{1/2} - q_{1/2}) \int_0^1 \left(1 - 2\mathbb{1}_{\{X_i \le q_{1/2} + t(\hat{q}_{1/2} - q_{1/2})\}}\right) \mathrm{d}t \end{aligned}$$

for all $i \in \{1, ..., n\}$. Subsequent summation over $i = 1, ..., n$ and division by n yields

$$\hat{M}_n' - \hat{M}_n = (\hat{q}_{1/2} - q_{1/2}) \int_0^1 \left(1 - 2\hat{F}_n(q_{1/2} + t(\hat{q}_{1/2} - q_{1/2}))\right) \mathrm{d}t. \tag{4}$$

Since it is known from Theorem 2 that $\sqrt{n}(\hat{q}_{1/2} - q_{1/2})$ converges in distribution to a Gaussian random variable, it remains to be shown that the integral in (4) converges to 0 in probability. To see this, we start by noting

$$\hat{F}_n(q_{1/2}) = 1/n \sum_{i=1}^n \mathbb{1}_{\{X_i \le q_{1/2}\}} \xrightarrow{a.s.} 1/2,$$

since all indicator functions in the sum above are iid symmetric Bernoulli random variables. By assumption, F is not only continuous at $q_{1/2}$, but also in a neighbourhood. Due to Proposition 3, the probability that this neighbourhood contais $\hat{q}_{1/2}$ tends to 1 as n tends to infinity. By definition of the empirical median, $\hat{q}_{1/2}$ is equal $\lceil n/2 \rceil / n$, if there do not exist two indices $i, j \in \{1, ..., n\}$ such that $X_i = \hat{q}_{1/2} = X_j$. This, in turn, is the case if F is continuous at $\hat{q}_{1/2}$. Overall, we obtain

$$\begin{aligned} \mathbb{P}(\hat{F}_n(\hat{q}_{1/2}) \to 1/2) &\ge \mathbb{P}(\hat{F}_n(\hat{q}_{1/2}) = \lceil n/2 \rceil / n) \\ &\ge \mathbb{P}(\nexists i, j \in \{1, ..., n\} : X_i = \hat{q}_{1/2} = X_j) \\ &\ge \mathbb{P}(F \text{ continuous at } \hat{q}_{1/2}) \to 1 \end{aligned}$$

as $n \to \infty$. Since $\hat{F}_n$ is non-decreasing and converges to $1/2$ at both the theoretical and the empirical median, we can infer that $\hat{F}_n(x_n) \xrightarrow{a.s.} 1/2$ for any sequence $(x_n)_{n \in \mathbb{N}}$ with values between $q_{1/2}$ and $\hat{q}_{1/2}$. Specifically, this can be applied to the argument of $\hat{F}_n$ in (4), which lies between $q_{1/2}$ and $\hat{q}_{1/2}$ for all $t \in [0, 1]$. Hence, we have

$$\begin{aligned}
&\lim_{n\to\infty}\int_0^1 \Big(1-2\hat{F}_n(q_{1/2}+t(\hat{q}_{1/2}-q_{1/2}))\Big)\,\mathrm{d}t\\
&=2\int_0^1 \Big(1/2-\lim_{n\to\infty}\hat{F}_n(q_{1/2}+t(\hat{q}_{1/2}-q_{1/2}))\Big)\mathrm{d}t=0 \qquad (5)
\end{aligned}$$

almost surely, where the first identity follows from the dominated convergence theorem. Together with (4), this implies $\sqrt{n}(\hat{M}_n' - \hat{M}_n) \xrightarrow{\mathbb{P}} 0$, thus concluding the proof. $\square$

Based on this lemma, we can now prove the following results for the empirical integrated quantile skewness.

Theorem 10. *Let F be a cdf with $\mathbb{E}|X| < \infty$.*

a) If $\mathbb{E}X^2 < \infty$ and the derivative f of F exists at $q_{1/2}$ with $f(q_{1/2}) > 0$, then

$$\sqrt{n}\,(\hat{\gamma}_{IQ}(X_1,...,X_n) - \gamma_{IQ}(X)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{IQ}(F)),$$

with

$$\begin{aligned}
\sigma^2_{IQ}(F) =& \frac{(\mu-q_{1/2})^2(\mu_2-2\mu q_{1/2}+q_{1/2}^2)}{M^4} + \frac{\mu-q_{1/2}}{M^3}\left(\frac{\mu-q_{1/2}}{f(q_{1/2})} - 2M_2\right)\\
&+\frac{1}{M^2}\left(\frac{1}{4(f(q_{1/2}))^2} + \mu_2 + 2\mu q_{1/2} - 3q_{1/2}^2\right) - \frac{1}{Mf(q_{1/2})}.
\end{aligned}$$

b) If F is continuous at $q_{1/2}$ and $F(t) > 1/2$ holds for all $t > q_{1/2}$, then

$$\hat{\gamma}_{IQ}(X_1,...,X_n) \xrightarrow{a.s.} \gamma_{IQ}(X),$$

i.e. $\hat{\gamma}_{IQ}(X_1,...,X_n)$ is a strongly consistent estimator of $\gamma_{IQ}(X)$.

Proof. Part a) follows directly by applying Lemma 9 and the δ-method.

For part b) it suffices to show that the three components $\hat{q}_{1/2}, \bar{X}_n, \hat{M}_n$ in Lemma 9 almost surely converge to $q_{1/2}, \mu, M$, respectively. For the empirical median and the arithmetic mean, this is ensured by Proposition 3 and by the strong law of large numbers, respectively. It remains to consider the empirical median MAD.

Similar to the proof of Lemma 9, we do this in two steps. First, we invoke the strong law of large numbers again to obtain $\hat{M}_n' = 1/n\sum_{i=1}^n |X_i - q_{1/2}| \xrightarrow{a.s.} M$. Due to (4), the remaining difference $\hat{M}_n' - \hat{M}_n$ is the product of the difference between empirical and theoretical median and an integral. The almost sure convergence of the integral to 0 is given in (5), for the median difference it is once again given by Proposition 3. Thus, $\hat{M}_n \xrightarrow{a.s.} M$, and the assertion follows. $\square$

3.4. Peason's skewness measure

We define Pearson's empirical skewness measure as

$$\hat{\gamma}_P(X_1,...,X_n) = \frac{\bar{X}_n - \hat{q}_{1/2}}{\hat{\sigma}_n},$$

where $\hat{\sigma}_n = \sqrt{1/n\sum_{i=1}^n (X_i - \bar{X}_n)^2}$. The following result can be shown similarly to Lemma 9, replacing the median by an arbitrary p-quantile.

Lemma 11. *Let $p \in (0,1)$ and let F be a cdf with finite first four moments. If the derivative f of F exists at the point q_p with $f(q_p) > 0$, then*

$$\sqrt{n}\left(\begin{pmatrix} \hat{q}_p \\ \bar{X}_n \\ \overline{X_n^2} \end{pmatrix} - \begin{pmatrix} q_p \\ \mu \\ \mu_2 \end{pmatrix}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma'_P),$$

where

$$\Sigma'_P = \begin{pmatrix} \frac{p(1-p)}{(f(q_p))^2} & \frac{p\mu-\mu^{(p)}}{f(q_p)} & \frac{p\mu_2-\mu_2^{(p)}}{f(q_p)} \\ \frac{p\mu-\mu^{(p)}}{f(q_p)} & \mu_2 - \mu^2 & \mu_3 - \mu\mu_2 \\ \frac{p\mu_2-\mu_2^{(p)}}{f(q_p)} & \mu_3 - \mu\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix}.$$

Applying this lemma for $p = 1/2$ and the δ-method yields the first part of the following theorem. The strong consistency in part b) follows again by Proposition 3 and the strong law of large numbers.

Theorem 12. *Let F be a cdf with existing first two moments.*

a) If $\mathbb{E}X^4 < \infty$ and the derivative f of F exists at the point $q_{1/2}$ with $f(q_{1/2}) > 0$, then

$$\sqrt{n}\left(\hat{\gamma}_P(X_1, ..., X_n) - \gamma_P(X)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_P^2(F))$$

with

$$\begin{aligned} \sigma_P^2(F) = {} & \frac{(q_{1/2} - \mu)}{\sigma^6}\left[q_{1/2}\left(\frac{\mu_4}{4} - \frac{\mu_2^2}{4} - \mu_3\mu + \mu_2\mu^2\right) - \frac{\mu_4\mu}{4} + \mu_3\mu_2 - \frac{3\mu_2^2\mu}{4}\right] \\ & + \frac{1}{f(q_{1/2})\sigma^4}\left[q_{1/2}\left(\mu M - \frac{M_2}{2}\right) + \frac{\mu M_2}{2} - \mu_2 M\right] + \frac{(\mu_2 - q_{1/2}\mu)^2}{\sigma^4} \\ & + \frac{1}{4(f(q_{1/2}))^2\sigma^2}. \end{aligned}$$

b) If $F(t) > 1/2$ holds for all $t > q_{1/2}$, then

$$\hat{\gamma}_P(X_1, ..., X_n) \xrightarrow{a.s.} \gamma_P(X),$$

i.e. $\hat{\gamma}_P(X_1, ..., X_n)$ is a strongly consistent estimator of $\gamma_P(X)$.

As for the quantile skewness, additional asymptotic results and subsequent tests for the nullity of this measure are given by [23].

3.5. Expectile skewness

The empirical expectile skewness is obtained by replacing the theoretical expectiles by their empirical counterparts as defined in (2). Since the empirical 1/2-expectile equals the arithmetic mean, we obtain

$$\hat{\gamma}_E^{(\alpha)}(X_1, ..., X_n) = \frac{1}{1-2\alpha}\frac{\hat{e}_{1-\alpha} + \hat{e}_\alpha - 2\bar{X}_n}{\hat{e}_{1-\alpha} - \hat{e}_\alpha}, \quad \alpha \in (0, 1/2).$$

A central limit theorem and strong consistency of this measure can be derived analogously to the quantile skewness, now based on Theorem 5 and Proposition 6. Both results can also be found in [7]. Using the notations $\eta(\tau_1,\tau_2) = E[I_{\tau_1}(e_{\tau_1},X)I_{\tau_2}(e_{\tau_2},X)]$ for $\tau_1,\tau_2 \in (0,1)$ and

$$A(\tau) = (2\mathbb{1}_{\{\tau<1/2\}} - 1)\frac{e_{1-\tau} - \mu}{\tau + F(e_\tau)(1-2\tau)}$$

for $\tau \in \{\alpha, 1-\alpha\}$, the following holds true.

Theorem 13. *Let F be a cdf with $\mathbb{E}|X| < \infty$.*

a) If $\mathbb{E}X^2 < \infty$ and F does not have a point mass at e_α, μ or $e_{1-\alpha}$, then

$$\sqrt{n}\left(\hat{\gamma}_E(X_1,...,X_n) - \gamma_E(X)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_E^2(F,\alpha))$$

with

$$\begin{aligned}\sigma_E^2(F,\alpha) = \frac{4}{(1-2\alpha)^2}&\left[\frac{4\eta(1/2,1/2)}{(e_{1-\alpha}-e_\alpha)^2} - \frac{4[A(\alpha)\eta(\alpha,1/2) + A(1-\alpha)\eta(1/2,1-\alpha)]}{(e_{1-\alpha}-e_\alpha)^3}\right.\\ &\left. + \frac{(A(\alpha))^2\eta(\alpha,\alpha) + A(\alpha)A(1-\alpha)\eta(\alpha,1-\alpha) + (A(1-\alpha))^2\eta(1-\alpha,1-\alpha)}{(e_{1-\alpha}-e_\alpha)^4}\right].\end{aligned}$$

b) $\hat{\gamma}_E^{(\alpha)}(X_1,...,X_n)$ is a strongly consistent estimator of $\gamma_E^{(\alpha)}(X)$, i.e.

$$\hat{\gamma}_E^{(\alpha)}(X_1,...,X_n) \xrightarrow{a.s.} \gamma_E^{(\alpha)}(X).$$

3.6. Tajuddin's skewness measure

An obvious estimator for Tajuddin's skewness measure is given by

$$\hat{\gamma}_T(X_1,...,X_n) = 2\hat{F}_n(\bar{X}_n) - 1.$$

Theorem 14. *Let F be a cdf with $\mathbb{E}X^2 < \infty$ and with existing and positive derivative f at μ. Then*

a)

$$\sqrt{n}\left(\hat{\gamma}_T(X_1,...,X_n) - \gamma_T(X)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_T^2(F)),$$

where

$$\sigma_T^2(F) = 4\left[F(\mu)(1-F(\mu)) + f^2(\mu)\sigma^2 - f(\mu)\mathbb{E}|X-\mu|\right].$$

b)

$$\hat{\gamma}_T(X_1,...,X_n) \xrightarrow{a.s.} \gamma_T(X),$$

i.e. $\hat{\gamma}_T(X_1,...,X_n)$ is a strongly consistent estimator of $\gamma_T(X)$.

Proof. Theorem 2 in [11] states that

$$\sqrt{n}\left(\hat{F}_n(\bar{X}_n) - F(\mu)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\mathbb{1}_{\{X_i\le\mu\}} - F(\mu) + (X_i-\mu)f(\mu)\right] + o_{\mathbb{P}}(1). \qquad (6)$$

Part a) now follows directly by the central limit theorem, noting that the variance of the asymptotic normal distribution is given by

$$\begin{aligned} 4\mathbb{V}\left(\mathbb{1}_{\{X\leq\mu\}} + Xf(\mu)\right) &= 4\left(\mathbb{V}[\mathbb{1}_{\{X\leq\mu\}}] + f^2(\mu)\mathbb{V}[X] + 2f(\mu)\mathrm{Cov}(\mathbb{1}_{\{X\leq\mu\}}, X)\right) \\ &= \sigma_T^2(F). \end{aligned}$$

Part b) follows from (6) and the strong law of large numbers. □

4. Comparison of the skewness measures

In this section, we compare both the values of the theoretical skewness measures and the performance of the estimators for several specific families of distributions. For the latter, $H_0 : \gamma = 0$ is tested against $H_1 : \gamma \neq 0$, where γ stands for one of the skewness measures under consideration, and the powers of these tests will be examined. However, it should be noted that none of these tests is consistent against the alternative that the distribution in question is asymmetric.

All computations and simulations were carried out with the R software [26] with an archived version of the R-package expectreg [28] being used for the calculation of theoretical and empirical expectiles. Additionally, the R package evmix [17] was utilized for density estimation of distributions with bounded support. Throughout this section, we fix the parameter α concerning the quantile and expectile skewness at $1/4$. An assessment of the mean squared error of the estimators can be found in [7] for most of the considered measures and varying values of the parameter α.

4.1. Values of the theoretical skewness measures

In this subsection, we examine whether the six skewness measures behave similarly. Therefore, we compare the skewness values for several families of distributions, namely the gamma, lognormal, beta, skew-normal and skew-t distributions. Functions for calculating the theoretical expectiles of the latter two distributions were written emulating the expectreg-methodology, using the explicit formulas for the truncated first moments given in [9] and [18]. Since all skewness measures are invariant under affine transformations (see Table 1), location and scale parameters are always fixed. Instead of the skewness values itself, their quotients with the corresponding moment skewness are considered in order to increase comparability between the measures and among the distributions.

In Figure 1, these relative skewness values are plotted for the gamma and the lognormal distribution. For the gamma distribution with varying shape parameter k, all relative skewness values are fairly constant with slight deviation for small values of k and therefore high skewness. In contrast, for the lognormal distribution with varying log-standard-error τ exists a clear trend. Compared to the moment skewness, all other measures are decreasing as the log-variance, and thereby the skewness, of the distribution increases. This indicates that the moment skewness is more sensitive to heavy-tailed distributions than the rest. The limit of the relative skewness values for $k \to \infty$ for the gamma and for $\tau \to 0$ for the lognormal distribution seem to coincide. This is most likely due to the fact that, in both cases, the underlying limit distribution is an affine transformation of a Gaussian distribution.

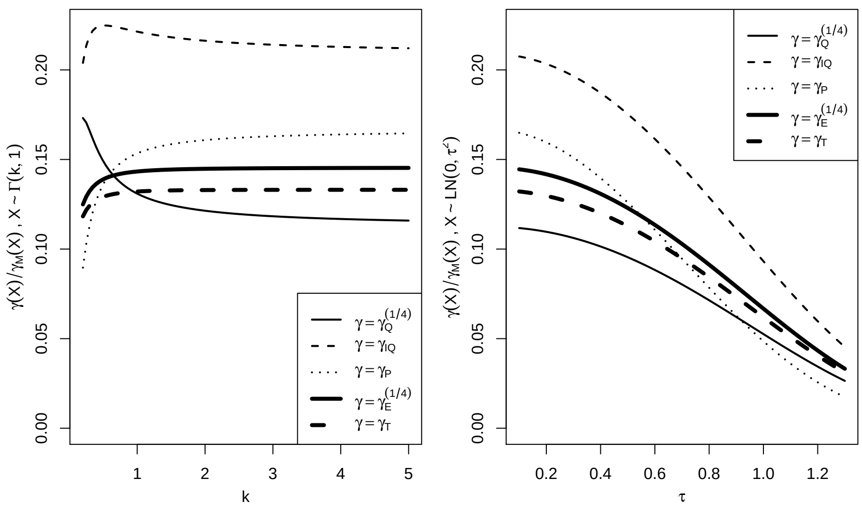

Figure 1: Theoretical skewness measures relative to the moment skewness for the Gamma distribution with varying shape parameter $k \in [0.2, 5]$ (left panel) and for the lognormal distribution with log-mean 0 and varying log-standard-error $\tau \in [0.1, 1.3]$ (right panel).

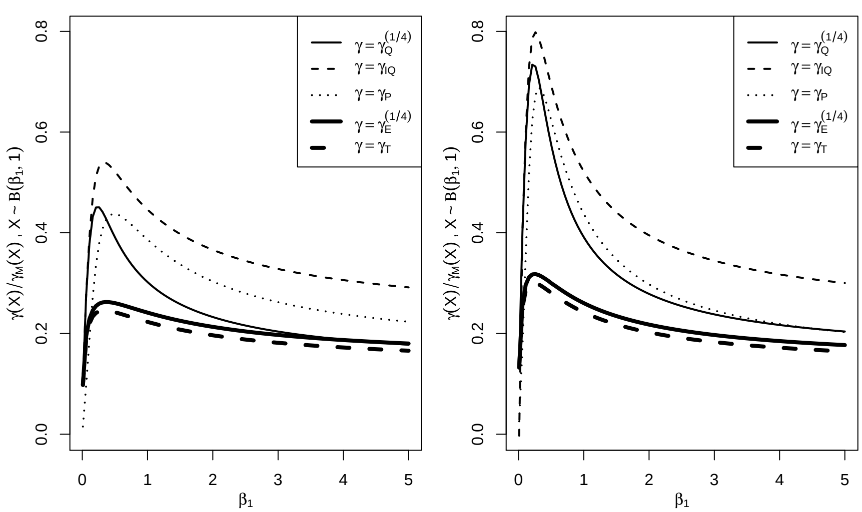

Figure 2: Theoretical skewness measures relative to the moment skewness for the Beta distribution with varying first shape parameter $\beta_1 \in [0.01, 5]$ and second shape parameter β_2 fixed at 1 (left panel) and 0.5 (right panel).

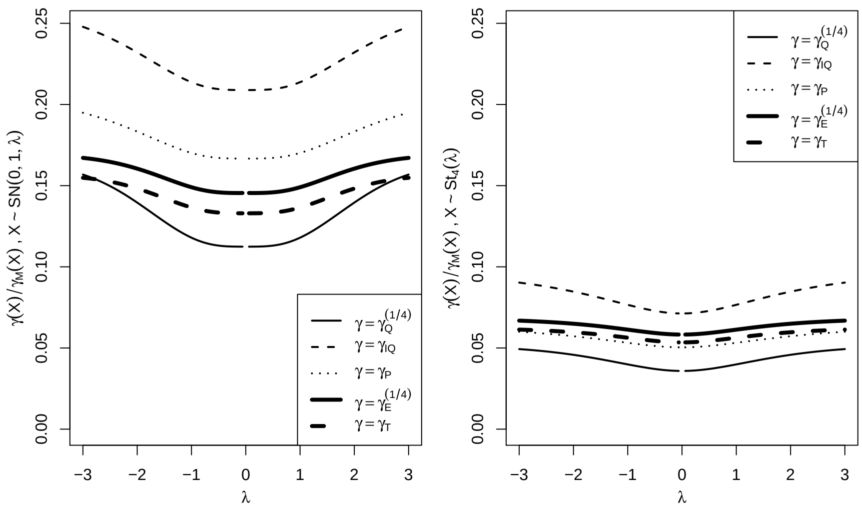

Figure 3: Theoretical skewness measures relative to the moment skewness for the standard skew-normal distribution with varying skewness parameter $\lambda \in [-3, 3]$ (left panel) and for the skew-t distribution with 4 degrees of freedom and varying skewness parameter $\lambda \in [-3, 3]$ (right panel).

In Figure 2, the beta distribution is considered. Since it is symmetric in its two shape parameters, it is sufficient to fix the second one and only to look at varying values of the first parameter β_1. Except for very low values of β_1, the values are consistently higher than for the gamma and the lognormal distribution. Since beta distributions have comparably light tails, this strengthens the aforementioned hypothesis that the moment skewness puts more emphasis on heavy tailed distributions. Additionally, all curves have a peak between $\beta_1 = 0.25$ and 0.5 before flattening out for increasing parameter values. For $\beta_1 = 1$ in the left panel and for $\beta_1 = 0.5$ in the right panel, respectively, we obtain symmetric distributions. However, the relative skewness values at these points are different from the limiting symmetric cases in Figure 1.

Finally, in Figure 3, the relative skewness values of the standard skew-normal distribution as well as the skew-t distribution with 4 degrees of freedom are plotted as a function of the skewness parameter. The shapes of the curves in the two panels are fairly similar. The main difference is that the relative skewness values for the skew-t distribution are less than half of those for the skew-normal. This is a further affirmation for the strong influence of the distributional tails on the moment skewness. Both distributions become symmetric for $\lambda = 0$. The values for the skew-normal distribution as $\lambda \to 0$ corresponds to the limit distributions in Figure 1.

Overall, especially when considering the limiting symmetric cases, it is obvious that γ_M is the most sensitive measure to heavy-tailed distributions, followed by the expectile based measures $\gamma_E^{(1/4)}$ and γ_T. On the other end of the spectrum, the quantile based skewness measures $\gamma_Q^{(1/4)}$, γ_{IQ} and γ_P seem to be most sensitive to distributions with light tails.

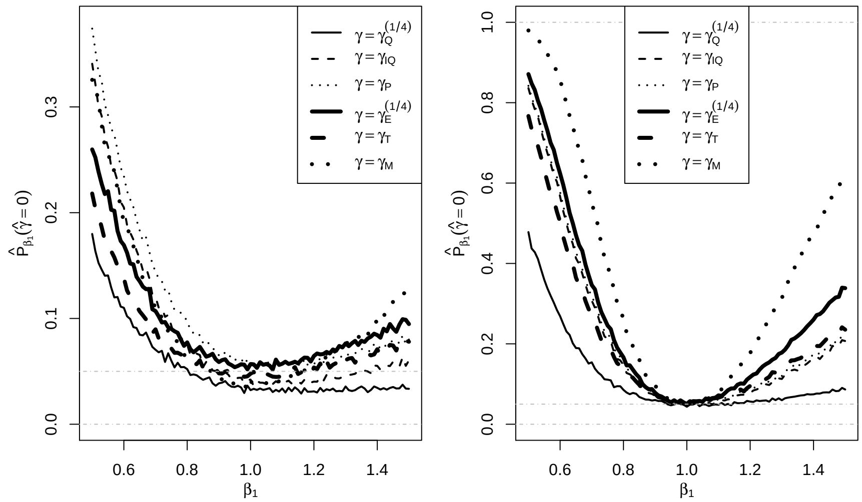

Figure 4: Estimated powers of tests on nullity of each skewness measure with underlying beta distribution $B(\beta_1, 1), \beta_1 \in [0.5, 1.5]$, based on 10000 repetitions. Left panel: sample size $n = 20$. Right panel: sample size $n = 100$.

4.2. Performance of the empirical skewness measures

For all considered skewness measures, it is straightforward to construct a test for nullity of the measure based on the corresponding central limit theorem. Let γ be a theoretical skewness measure with empirical version $\hat{\gamma}$ and asymptotic variance σ^2. Then, the null hypothesis $H_0 : \gamma = 0$ is rejected at the significance level of $\beta = 5\%$ in favour of the alternative $H_1 : \gamma \neq 0$ if $|\hat{\gamma}\hat{\sigma}/\sqrt{n}| > z_{1-\beta/2}$. Here, $z_{1-\beta/2}$ denotes the $(1 - \beta/2)$-quantile of the standard normal distribution, and $\hat{\sigma}$ denotes the estimated standard error of $\hat{\gamma}$. For the latter, we always used plug-in estimator, whose consistency can easily be shown. This assures that all tests are asymptotically valid, i.e. they maintain the significance level $\beta = 5\%$ for increasing sample size.

Since we only consider families of distributions which contain a symmetric one, we can examine whether the actual significance level of the test corresponds to the asymptotic one and also analyze the power of the test as a function of the parameter values. All empirical simulation results are based on 10000 repetitions.

First, we consider the beta distribution with the second shape parameter fixed at one and with variable first shape parameter $\beta_1 > 0$. For $\beta_1 = 1$, the distribution is the uniform distribution on $[0, 1]$ and therefore symmetric. In Figure 4, the results are depicted for sample sizes $n = 20$ and $n = 100$. For $n = 20$, the actual significance levels vary between 3.2% and 6.1% and are therefore already fairly close to the asymptotic one. Nevertheless, such differences between empirical levels have a considerable effect on the power of the tests. The plot is quite asymmetric with considerably higher power for $\beta_1 < 1$ than for $\beta_1 > 1$. The lowest power is attained by the test involving the quantile skewness for all values of β_1. The maximum power is attained for $\beta_1 = 0.5$ at 37% for γ_P with γ_{IQ} close behind, while the power of γ_{IQ} barely exceeds the significance level at $\beta_1 = 1.5$. For $n = 100$, all tests maintain the asymptotic significance level very closely.

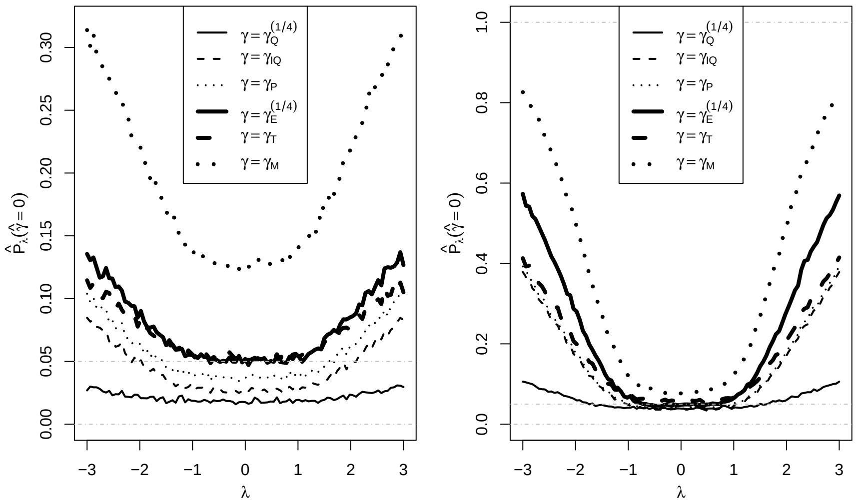

Figure 5: Estimated powers of tests for nullity of each skewness measure with underlying skew-normal distribution $SN(0, 1, \lambda), \lambda \in [-3, 3]$, based on 10000 repetitions. Left panel: sample size $n = 20$. Right panel: sample size $n = 100$.

Once again, $\gamma_Q^{(1/4)}$ clearly performs worst in terms of power, while γ_M has the highest power throughout, even reaching 1 for $\beta_1 = 0.5$. The expectile skewness performs second best, slightly better than γ_T, γ_{IQ} and γ_P, which behave quite similar. For $n = 1000$ (not shown), the steepness when moving away from $\beta_1 = 1$ increases so that the powers of all tests reach 1 at some point. The ranking of the skewness measures is the same as for $n = 100$ with the curves becoming more symmetric.

For the skew-normal distribution (see Figure 5) with sample size $n = 20$, the differences between the powers of the different skewness tests are considerably bigger than for the beta distribution. The two expectile based skewness measures $\gamma_E^{(1/4)}$ and γ_T are the only ones that maintain the 5% significance level exactly. The power for γ_M is far too high with 12.3%, while the powers for the quantile based measures are too low, between 1.6% and 3.4%. The steepness of the incline of the curves when moving away from $\lambda = 0$ seems to be correlated with the significance level, since it is highest for the moment skewness and lowest for the quantile skewness, not exceeding 5% at all. For $n = 100$, the actual significance levels are all close to 5% with the exception of γ_M, where the actual level still exceeds the theoretical one. Apart from that, the plot bears similarities to the corresponding one for the beta distribution. The γ_M-test has the highest power, followed by $\gamma_E^{(1/4)}$, and $\gamma_Q^{(1/4)}$ is once again far behind with a maximum power of 10.6%. The low powers for $\lambda \in [-1, 1]$ are due to the fact that the skew-normal distribution is nearly symmetric for these parameter values. When we increase the sample size to $n = 1000$, all tests maintain the asymptotic significance level closely.

The results for the skew-t distribution with 8 degrees of freedom (see Figure 6) are similar. For $n = 20$, the actual significance level of the γ_M-test is even higher at 22.1%, while the level of the $\gamma_Q^{(1/4)}$-test is once again very low at 1.4%. The significance levels

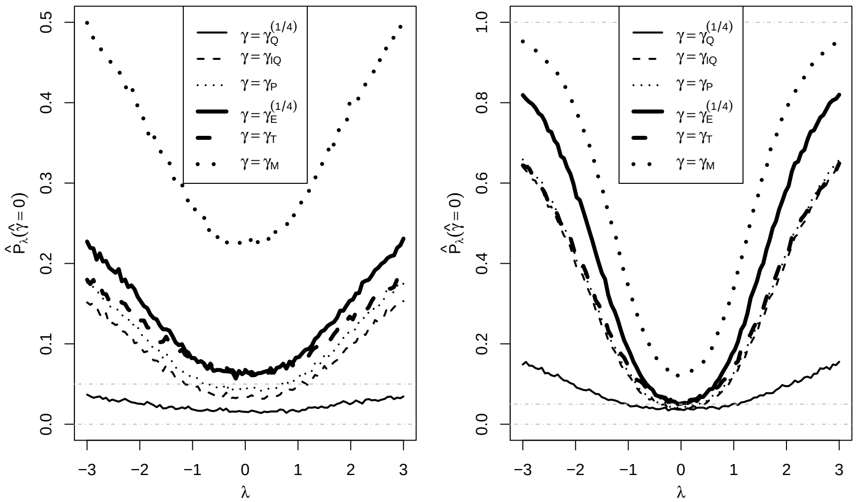

Figure 6: Estimated powers of tests for nullity of each skewness measure with underlying skew-t distribution $St_8(0, 1, \lambda), \lambda \in [-3, 3]$ with 8 degrees of freedom, based on 10000 repetitions. Left panel: sample size $n = 20$. Right panel: sample size $n = 100$.

of the other tests are all relatively close to 5% with $\gamma_E^{(1/4)}$ reaching the highest power of 23.1%. The increase in power when shifting to sample size $n = 100$ is larger than for the skew-normal distribution; however, the significance level of the γ_M-test at 12.1% is still markedly too high. The ranking of the other tests is the same as for the other classes of distributions. For the sample size $n = 1000$, the power of all tests increase rapidly for deviations of λ from 0. For $|\lambda| > 2$, the powers are equal to 1 for all skewness measures except for $\gamma_Q^{(1/4)}$. All other tests are very close to each other in terms of power with the expectile skewness being slightly superior to the moment skewness in this case. Additionally, all tests attain the asymptotic significance level of 5% for this large sample size.

Overall, the moment skewness performs seem to perform best in terms of power; however, this is due to the fact that the test is very liberal with an actual significance level far in excess of the nominal level, especially for heavy-tailed distributions and small sample sizes. A better alternative is the expectile skewness, which performs well in terms of power, and consistently has a significance level very close to the theoretical one. Among the six skewness tests, the quantile skewness has by far the lowest power.

Given empirical data, we can also construct asymptotic confidence intervals for any skewness measure γ considered in this paper. Since $\sqrt{n}(\hat{\gamma} - \gamma)$ has a limiting normal distribution, an approximate confidence interval at level $1 - \beta$ is given by

$$\left[\hat{\gamma} - \frac{\hat{\sigma} z_{1-\beta/2}}{\sqrt{n}}, \hat{\gamma} + \frac{\hat{\sigma} z_{1-\beta/2}}{\sqrt{n}}\right],$$

where $\hat{\sigma}$ denotes a consistent estimator of the standard error. In the following, we keep $1 - \beta$ fixed at 95%, and we consider a sample size of $n = 50$.

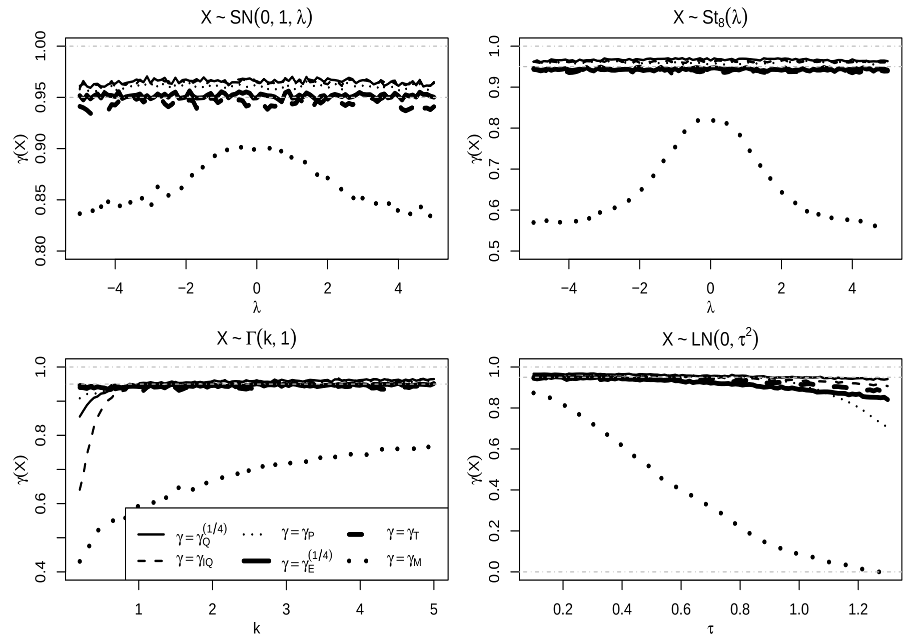

Figure 7: Empirical probabilities of the confidence intervals based on each skewness measure covering the theoretical skewness value, based on 10000 repetitions and sample size $n = 50$.

We now analyze the empirical coverage probability, i.e. the proportion of the time that the interval contains the true skewness value, based on 10000 repetitons. In Figure 7, these proportions are shown as a function of the parameter for four different distributions. For all moderately skewed distributions, these coverage probabilities are fairly close to 95%, except for the moment skewness. Whereas the coverage probabilities for the quantile based measures $\gamma_Q^{(1/4)}$, γ_{IQ} and γ_P tend to be a little higher than 95%, they are exactly at 95% or slightly below for the expectile based measures $\gamma_E^{(1/4)}$ and γ_T. The moment skewness γ_M, however, consistently falls short of the nominal coverage probability with a maximal value of 90.3% over all families of distributions. The probability gets even lower, the more skewed the underlying distribution is. Additionally, it tends to be lower for heavy-tailed distributions, namely for the skew-t and the log-normal distributions.

For heavily skewed distributions (e.g. $\Gamma(k, 1)$ for $k < 1$ and $LN(0, \tau^2)$ for $\tau > 1$), the coverage probability of the moment skewness decreases even further, and some of the other curves also fall significantly below 95%. For the more light-tailed gamma distribution, the expectile based measures maintain the asymptotic probability, while the quantile based measures decline to 64.1% for γ_{IQ}. For the heavy-tailed log-normal distribution, all probabilities decline with a smaller gradient. Here, γ_P attains the lowest value with 69.5%, followed by the expectile based measures with 84.1% for $\gamma_E^{(1/4)}$ and 87.6% for γ_T.

Overall, this reinforces the strong sensitivity of γ_M to heavy-tailed distributions, making it unusable under some scenarios. To a much lesser extent, this also holds for the quantile based measures for light-tailed distributions with high skewness.

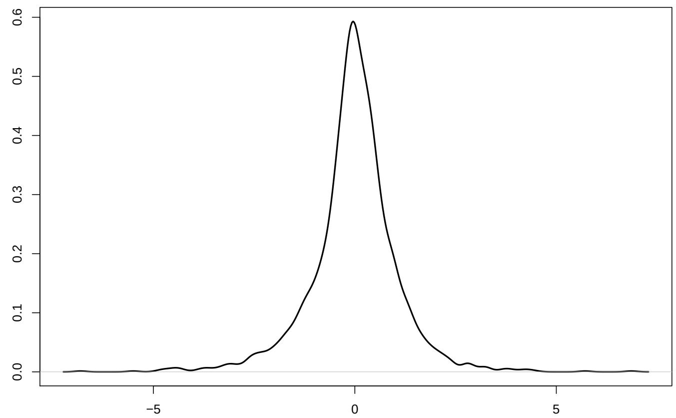

Figure 8: Estimated density for the data set of residuals, sample size $n = 2013$.

	γ_M	$\gamma_Q^{(1/4)}$	γ_{IQ}	γ_P	$\gamma_E^{(1/4)}$	γ_T
Estimates	-0.4445	0.0993	-0.0060	-0.0042	-0.0275	-0.0159
p-value	0.191	0.183	0.296	0.295	0.018	0.164

Table 2: Estimates and p-values of tests on nullity for the different skewness measure.

5. Real data applications

We apply the different skewness measures to the residuals of a time series model fitted to the daily log returns $x_t = \ln(P_t/P_{t-1})$, where P_t is the closing price of the S&P 500 on day t, adjusted for dividends and splits. The sample period runs from January 1, 2009 until December 31, 2016, giving a total of 2013 observations. The data has been downloaded from `http://finance.yahoo.com` and was also analyzed in a forecasting context in Holzmann and Klar [16]. The underlying model is the following GARCH(1,1) process,

$$x_t = \mu + \sigma_t z_t, \qquad \sigma_t^2 = \omega + \alpha_1 (x_{t-1} - \mu)^2 + \beta_1 \sigma_{t-1}^2.$$

The residuals are then given by $e_t = (x_t - \hat{\mu})/\hat{\sigma}_t$, where $\hat{\mu}$ is the estimated mean, and $\hat{\sigma}_t$ denotes the fitted volatility process. A typical finding in empirical applications of GARCH models to stock returns is that the residuals are leptokurtic compared to a normal distribution, but they are often approximately symmetric.

In the present case, the data seems to be slightly left-skewed; most prominently, there seems to be more probability mass around the value -1 than around 1 with the left tail seeming a bit thicker overall (see Figure 8). The p-values of the tests on nullity for the six considered skewness measures based on this data set are given in Table 5. The expectile skewness is the only measure that deems the data to be significantly skewed. All skewness measures judge the data to be left-skewed except for the quantile skewness.

It would be desirable for quantile and expectile skewness that their test statistics are reasonably stable with respect to the parameter α, which was fixed at $1/4$ up to now. The p-values of these tests are plotted as a function of α in Figure 9. Obviously, the expectile

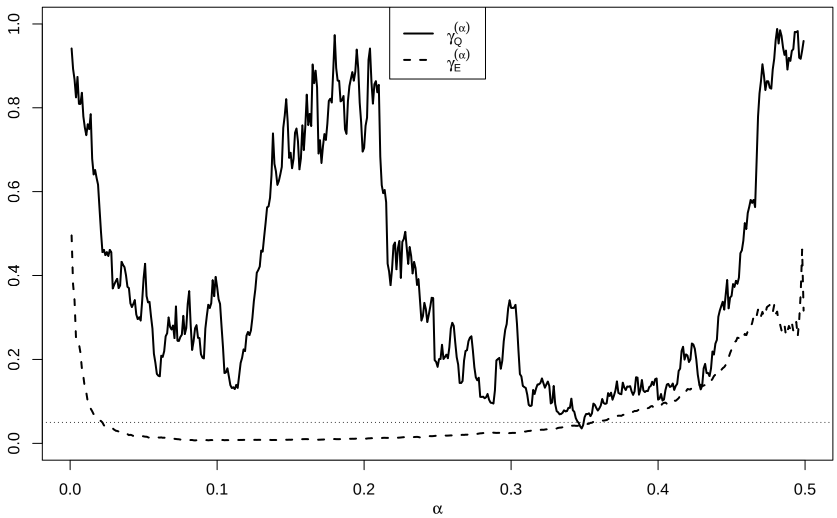

Figure 9: p-values of tests on nullity of the quantile and expectile skewness as a function of the parameter α.

skewness curve is a lot smoother and more stable, excluding α-values close to 0 or 1/2. It deems the data significantly left-skewed for the relatively wide range of $\alpha \in [0.022, 0.355]$. On contrary, the quantile skewness is very unstable with respect to α: it comes close to deeming the data significantly left-skewed at $\alpha = 0.112$ (p-value 0.129), then changes sides at $\alpha = 0.18$ and deems the data significantly right-skewed at $\alpha = 0.348$ (p-value 0.036). This echoes the findings of section 4.2 that the quantile skewness is not really suitable for testing for skewness, especially compared to the expectile skewness.

Now, we consider each year separately; the corresponding estimated densities are shown in Figure 10, and the p-values of the corresponding tests are given in Table 5. The skewness measures are mostly negative, but not to a significant degree, which corresponds to the slightly left-skewed densities in most years. 2011 and 2014 seem to be the years

Year	γ_M	$\gamma_Q^{(1/4)}$	γ_{IQ}	γ_P	$\gamma_E^{(1/4)}$	γ_T
2009	0.905 (−)	0.353 (−)	0.343 (−)	0.345 (−)	0.183 (−)	0.390 (−)
2010	0.452 (−)	0.933 (−)	0.731 (−)	0.731 (−)	0.283 (−)	0.328 (−)
2011	0.136 (−)	0.362 (−)	0.202 (−)	0.200 (−)	0.028 (−)	0.117 (−)
2012	0.901 (+)	0.775 (+)	0.642 (+)	0.642 (+)	0.402 (+)	0.230 (+)
2013	0.147 (−)	0.654 (−)	0.365 (−)	0.364 (−)	0.190 (−)	0.561 (−)
2014	0.047 (−)	0.674 (+)	0.149 (−)	0.149 (−)	0.015 (−)	0.132 (−)
2015	0.429 (−)	0.174 (+)	0.466 (+)	0.468 (+)	0.644 (+)	0.395 (+)
2016	0.162 (−)	0.279 (+)	0.927 (−)	0.927 (−)	0.617 (−)	0.768 (−)

Table 3: p-values of tests on nullity of the given skewness measure divided up according to the year. The symbol in brackets indicates whether the data was deemed left-skewed (−) or right-skewed (+).

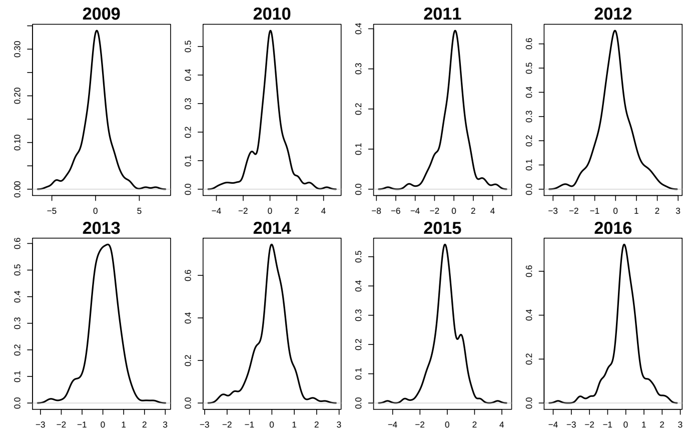

Figure 10: Estimated densities for the data sets of residuals divided up according to the year.

with the most clearly left-skewed data based on the test results of $\gamma_E^{(1/4)}$ and γ_M. Also, there is consensus among the skewness measures that the data of 2012 is rather right-skewed, but not to a significant degree. However, in the years 2014 to 2016, the measures disagree whether the data is rather left- or right-skewed. Considering the densities, the disagreement of $\gamma_Q^{(1/4)}$ in 2014 and 2016 seems to be caused by its disregard of the tails, which contribute significantly to the overall skewness in these data sets. Since the moment skewness puts more emphasis on the tails, the slightly bigger left tail in the 2015 data affects γ_M more than the little bump around 1, contrary to the other measures.

Overall, γ_{IQ}, γ_P, $\gamma_E^{(\alpha)}$ and γ_T seem to be the most reliable measures in this real data application. Of these four, the expectile skewness is the only one to even once judge the data to be significantly skewed at the 5%-level.

6. Conclusion

In this paper, we examined six measures of skewness concerning their theoretical and empirical properties as well as their behaviour in simulations. Supplementary simulation results can be found in Eberl and Klar [7].

While the moment skewness γ_M as the most popular skewness measure seems to have the highest power at first sight, it has several downsides. First, it is not normalizable and therefore difficult to use for comparisons. Second, its empirical version converges very slowly and with a high bias [see 7] due to higher moments having to be estimated. Consequently, the corresponding tests on nullity fail to attain their asymptotic significance level for heavy-tailed distributions and sample sizes below 1000, and the empirical coverage probability of pertaining confidence intervals is often far away from the nominal level.

Based on these simulations, we would recommend the expectile skewness $\gamma_E^{(\alpha)}$ as the best skewness measure. It is neither too sensitive to heavy-tailed nor to light-tailed distributions, and the mean squared error of its empirical version behaves nicely with a relatively small contribution of the bias compared to the variance. Additionally, estimation of its asymptotic variance is very accurate even for small sample sizes, and it detects the asymmetric distributions considered in this paper reliably. Its main downside is that it is an open question if $\gamma_E^{(\alpha)}$ preserves the convex transformation order (property S3). This can be circumvented by using Tajuddin's skewness measure γ_T instead. It performs slightly worse than the expectile skewness in the simulations, but satisfies property S3; moreover, it has a very simple structure.

References

[1] Babu, G. J., Rao, C. R. (1992). Expansions for Statistics involving the Mean Absolute Deviations. *Annals of the Institute of Statistical Mathematics* 44, 387–403.

[2] Bellini, F., Klar, B., Müller, A., Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics* 54, 41–48.

[3] Bowley, A. L. (1920). *Elements of Statistics*, 4th edn. Charles Scribner's Sons, New York.

[4] Cramér, H. (1957). Mathematical Methods of Statistics. *Princeton University Press*, 7th edn.

[5] D'Agostino, R. B., Pearson, E. S. (1973). Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$. *Biometrika* 60, 613–622.

[6] D'Agostino, R. B., Tietjen, G. L. (1973). Approaches to the Null Distribution of $\sqrt{b_1}$. *Biometrika* 60, 169–173.

[7] Eberl, A., Klar, B. (2019a). Expectile based measures of skewness. *ArXiv preprint arXiv:1908.08243.*

[8] Eberl, A., Klar, B. (2019b). On the skewness order of van Zwet and Oja. *Submitted.*

[9] Flecher, C., Allard, D., Naveau, P. (2010). Truncated skew-normal distributions: moments, estimation by weighted moments and application to climatic data. *METRON* 68, 331-–345.

[10] Geary, R. C. (1947). The Frequency Distribution of $\sqrt{b_1}$ for Samples of All Sizes Drawn a Random from a Normal Population. *Biometrika* 34, 68–97.

[11] Ghosh, J. K. (1971). A New Proof of the Bahadur Representation of Quantiles and an Application. *The Annals of Mathematical Statistics*, 42, 1957–1961.

[12] Groeneveld, R. A., Meeden, G. (1984). Measuring Skewness and Kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)* 33, 391–399.

[13] Gupta M. K. (1967). An asymptotically nonparametric test of symmetry. *The Annals of Mathematical Statistics* 38, 849–866.

[14] Hinkley, D. V. (1975). On Power Transformations to Symmetry. *Biometrika*, 62, 101–111.

[15] Holzmann, H., Klar, B. (2016). Expectile Asymptotics. *Electronic Journal of Statistics* 10, 2355–2371.

[16] Holzmann, H., Klar, B. (2017). Focusing on regions of interest in forecast evaluation. *Ann. Appl. Stat.* 11, 2404–2431.

[17] Hu, Y. and Scarrott, C. (2018). evmix: An R package for Extreme Value Mixture Modeling, Threshold Estimation and Boundary Corrected Kernel Density Estimation. *Journal of Statistical Software* 84.5, 1-27. doi: 10.18637/jss.v084.i05 (URL: http://doi.org/10.18637/jss.v084.i05).

[18] Jamalizadeh, A., Pourmousa, R., Balakrishnan, N. (2009). Truncated and Limited Skew-Normal and Skew-t Distributions: Properties and an Illustration. *Communications in Statistics – Theory and Methods* 38, 2653–2668.

[19] Lin, P.-E., Wu, K.-T., Ahmad, I. A. (1980). Asymptotic joint distribution of sample quantiles and sample mean with applications. *Communications in Statistics - Theory and Methods* 9, 51–60.

[20] MacGillivray, H. L. (1986). Skewness and Asymmetry: Measures and Orderings. *The Annals of Statistics*, 14, 994–1011.

[21] Mulholland, H. P. (1977). On the Null Distribution of $\sqrt{b_1}$ for Samples of Size at Most 25, with Tables. *Biometrika* 64, 401–409.

[22] Newey, K., Powell, J. (1986). Asymmetric least squares estimation and testing. *Econometrica* 55, 819–847.

[23] Ngatchou-Wandji, J. (2006). On Testing for the Nullity of Some Skewness Coefficients. *International Statistical Review* 74, 47–65.

[24] Oja, H. (1981). On Location, Scale, Skewness and Kurtosis of Univariate Distributions. *Scandinavian Journal of Statistics*, 8, 154–168.

[25] Pearson, K. (1895). Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London. A*, 186, 343–414.

[26] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/.

[27] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley Series in Probability and Statistics, Wiley.

[28] Sobotka, F., Schnabel, S., Schulze Waltrup, L. with contributions from Eilers, P., Kneib, T., Kauermann, G. (2014). expectreg: Expectile and Quantile Regression. R package version 0.39.

[29] Tajuddin, I. H. (1999). A Comparison between Two Simple Measures of Skewness. *Journal of Applied Statistics* 26, 767–774.

[30] Yule, G. U. (1912). *An Introduction to the Theory of Statistics*, 1st edn. Charles Griffin and Co, London.

[31] Zwet, W. R. van (1964). *Convex Transformations of Random Variables.* Math. Centrum, Amsterdam.