Scandinavian Journal of Statistics

Expectile-based measures of skewness

Andreas Eberl | Bernhard Klar

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

Correspondence
Bernhard Klar, Institute of Stochastics,
Karlsruhe Institute of Technology (KIT),
Englerstr. 2, 76131, Karlsruhe, Germany.
Email: bernhard.klar@kit.edu

Abstract

In the literature, quite a few measures have been proposed for quantifying the deviation of a probability distribution from symmetry. The most popular of these skewness measures are based on the third centralized moment and on quantiles. However, there are major drawbacks in using these quantities. These include a strong emphasis on the distributional tails and a poor asymptotic behavior for the (empirical) moment-based measure as well as difficult statistical inference and strange behaviour for discrete distributions for quantile-based measures. Therefore, in this paper, we introduce skewness measures based on or connected with expectiles. Since expectiles can be seen as smoothed versions of quantiles, they preserve the advantages over the moment-based measure while not exhibiting most of the disadvantages of quantile-based measures. We introduce corresponding empirical counterparts and derive asymptotic properties. Finally, we conduct a simulation study, comparing the newly introduced measures with established ones, and evaluating the performance of the respective estimators.

KEYWORDS
expectile, omega ratio, skewness measure, stop-loss transform

1 | INTRODUCTION

Symmetry of a distribution or density function is one of the oldest and, at the same time, one of the most important concepts in probability theory and statistics. A real random variable X with cumulative distribution function (cdf) F is symmetric about θ if $X - \theta \sim \theta - X$, or, equivalently, if $F(\theta - x) = 1 - F(\theta + x), x \in \mathbb{R}$. Over time, a sizeable number of asymmetry or skewness measures has been proposed in the literature to quantify the deviation from symmetry.

The best-known skewness measure is certainly the moment-based measure $\gamma_M = E[(X - EX)^3]/\mathrm{Var}(X)^{3/2}$, which is often used synonymously with the notion of skewness itself. However, this measure has a number of disadvantages. First, it is so sensitive to the extreme tails of the distribution that it is difficult to estimate accurately in practice when the distribution is markedly skew (Hosking, 1990). Second, it can not be normalized, which makes the according skewness values less informative and comparable. Third, the asymptotic distribution of its empirical counterpart involves moments up to order 6 of the underlying distribution, implying a very slow convergence especially for heavy-tailed distributions. Finally, $\gamma_M = 0$ does not characterize symmetry. This means that there exist nonsymmetric distributions for which γ_M is equal to zero. For a specific example using the generalized lambda distribution, see Ramberg et al. (1979) and Ramberg et al. (1980). Another example is the gamma difference distribution, also known as generalized asymmetric Laplace distribution. As the latter name indicates, this distribution is asymmetric, but the moment skewness is zero under a specific condition on the parameter values, see Klar (2015). On the plus side, there is its familiarity, and the fact that this measure is also reasonable for discrete distributions with finite third (or sixth) moment.

A distribution is symmetric if and only if $q_X(1 - \alpha) - q_X(1/2) = q_X(1/2) - q_X(\alpha)$ for each $\alpha \in (0, 1/2)$, where $q_X(\alpha)$ denotes the α-quantile of the distribution of X. Hence, one can also provide a quantile-based definition of skewness: A distribution is said to be right-skewed (left-skewed) in the quantile sense, if and only if

$$q_X(1 - \alpha) - q_X(1/2) \geq (\leq) q_X(1/2) - q_X(\alpha), \quad \alpha \in (0, 1/2), \tag{1}$$

and equality does not hold for each $\alpha \in (0, 1/2)$. Corresponding scalar measures of skewness are

$$b_1 = \frac{q_X(3/4) + q_X(1/4) - 2q_X(1/2)}{q_X(3/4) - q_X(1/4)},$$

(see Yule, 1912, Bowley, 1920) or the general version

$$b_2(\alpha) = \frac{q_X(1 - \alpha) + q_X(\alpha) - 2q_X(1/2)}{q_X(1 - \alpha) - q_X(\alpha)},$$

where $\alpha \in (0, 1/2)$, introduced by David and Johnson (1956) (see also Hinkley, 1975, Groeneveld & Meeden, 1984). As a measure of skewness, b_1 can be criticized for being insensitive to the distribution of X any further into the tails than the quartiles (Hosking, 1990). Quantiles are not unique for discrete distributions, in particular for empirical distributions. Furthermore, statistical inference involving the asymptotic distribution of b_1 or b_2 requires the evaluation of the (unknown) density of the underlying distribution, which typically requires bandwidth selection and, hence, leads to a certain arbitrariness. Again, rather large sample sizes are desirable to reliably represent

the population being sampled. On the other hand, b_1 is a robust and quite intuitive measure, and $b_2(\alpha) \equiv 0$ characterizes symmetry.

Both b_1 and b_2 satisfy three properties, introduced by van Zwet (1964), which are often seen appropriate for a skewness coefficient γ (see Oja, 1981, Groeneveld & Meeden, 1984; Tajuddin, 1999):

S1. For $c > 0$ and $d \in \mathbb{R}$, $\gamma(cX + d) = \gamma(X)$.
S2. The measure γ satisfies $\gamma(X) = -\gamma(X)$.
S3. Let F and G, the cdf's of X and Y, be absolutely continuous and strictly increasing on $\{x : 0 < F(x) < 1\}$ and $\{x : 0 < G(x) < 1\}$, respectively. If F is smaller than G in convex transformation order (written $F \leq_2 G$), that is, $G^{-1}(F(x))$ is convex, then $\gamma(X) \leq \gamma(Y)$.

The convex transform order is equivalent to

$$\frac{\left(F^{-1}(w) - F^{-1}(v)\right) - \left(F^{-1}(v) - F^{-1}(u)\right)}{F^{-1}(w) - F^{-1}(u)}$$
$$\leq \frac{\left(G^{-1}(w) - G^{-1}(v)\right) - \left(G^{-1}(v) - G^{-1}(u)\right)}{G^{-1}(w) - G^{-1}(u)} \quad \forall \ 0 < u < v < w < 1. \tag{2}$$

Even though this characterization of the convex transformation order should be known, we were not able to find it in the literature. Thus, we give some details in Appendix. Plugging $v = 1/2, w = 1 - \alpha$ and $u = \alpha$ into (2) shows immediately that b_2 satisfies (S3), that is, it preserves the convex transformation order. A different proof of this fact has been given by Groeneveld and Meeden (1984).

On the other hand, the convex transform order is the strongest of all commonly used skewness orders. Hence, also the requirement $b_{2,X}(\alpha) \leq b_{2,Y}(\alpha)$ for each $\alpha \in (0, 1/2)$ is a very strong one. As Arnold and Groeneveld (1993) argue, one might be willing to forgive minor local violations, rather than to insist on a uniform domination of $b_{2,X}$ by $b_{2,Y}$, and such a proposal is made in the following.

To do so, we consider measures of skewness based on or related to expectiles (which are formally defined in Section 2). Among others, we propose

$$\tilde{s}_2(\alpha) = \frac{e_X(1 - \alpha) + e_X(\alpha) - 2e_X(1/2)}{e_X(1 - \alpha) - e_X(\alpha)}, \quad \alpha \in (0, 1/2),$$

where $e_X(\alpha)$ denotes the α-expectile of X, as a family of expectile-based measures of skewness. Contrary to γ_M, these measures can be normalized, characterize symmetry when equal to zero, exist for any distribution with finite first moment, and have very convenient asymptotic properties. Since expectiles can be seen as smoothed quantiles, $\tilde{s}_2(\alpha)$ has a similar interpretation as $b_2(\alpha)$, but is sensitive to the whole distribution of X.

There are some further skewness measures with structures similar to b_2 and $\tilde{s}_2(\alpha)$. L-skewness, a measure based on L-moments, was introduced by Hosking (1990) and can be written as

$$\tau_3 = \frac{EX_{3:3} - 2EX_{2:3} + EX_{1:3}}{EX_{3:3} - EX_{1:3}},$$

where $EX_{1:3} \leq EX_{2:3} \leq EX_{3:3}$ denotes the order statistics of a random sample of size 3 from the distribution of X. Like $\tilde{s}_2$, it exists whenever $E|X| < \infty$. As for γ_M, $\tau_3 = 0$ does not

characterize symmetry. As an example, consider the generalized extreme value distribution (GEV) with shape parameter k, for which $\tau_3 = 2(1 - 3^{-k})/(1 - 2^{-k}) - 3$ (Hosking, 1990). Hence, $\tau_3 = 0$ for $k^* \approx 0.284$. However, having support $(-\infty, 1/k^*)$, this distribution is asymmetric in the strict mathematical sense. Since the Weibull distribution is related to the GEV by a simple transformation, it also follows that the L-skewness is zero for a Weibull distribution with shape parameter $1/k^* \approx 3.524$. This is very close to the value 3.6 for which γ_M vanishes (MacGillivray, 1986).

Critchley and Jones (2008) defined an asymmetry function as follows. For any $a < b$, they considered the class of all rooted unimodal densities with support (a, b). For any $0 < p < 1$, there are two points $x_L(p)$ and $x_R(p)$, one each side of the mode m, satisfying $f(x_L(p)) = f(x_R(p)) = pf(m)$. Skewness or asymmetry is measured by

$$\gamma^*(p) = \frac{x_R(p) - 2m + x_L(p)}{x_R(p) - x_L(p)}.$$

Even if the idea is interesting, it is difficult to use this measure in practice due to the need of estimating the level points of the density as well as the modal value. Accordingly, not much is known about properties of estimators of this measure. Further, the measure is only defined for a quite restricted class of distributions. Compared with b_2, this is a local measure of asymmetry, whereas $\tilde{s}_2(\alpha)$ can be seen as an integral version of b_2.

Clearly, there exist many other proposals of skewness measures besides the above mentioned. Further examples can be found later in this work; for overviews, see, for example, Benjamini and Krieger (1996), Critchley and Jones (2008), and MacGillivray (1986).

This paper is organized as follows. In Section 2, we recall the definitions of expectiles and some of their properties. Section 3 formally introduces $\tilde{s}_2$ and discusses its properties. We define closely linked skewness measures by normalizing $\tilde{s}_2$ and by considering limiting values with respect to α. In Section 4, we introduce related skewness measures based on Omega ratios and stop-loss transforms. Empirical counterparts of the proposed measures are analyzed in Section 5. We illustrate the behavior of the newly proposed skewness measures and their empirical counterparts for some families of distributions in Section 6. Most proofs are postponed to Appendix.

2 | EXPECTILES AND EXPECTILE LOCATION ORDER

Throughout the paper we assume that all mentioned random variables X are nondegenerate, have a finite mean (denoted as $X \in L^1$) and are defined on a common probability space $(\Omega, \mathcal{A}, P)$ unless stated otherwise. Recall that the expectiles $e_X(\alpha)$ of a random variable $X \in L^2$ have been defined by Newey and Powell (1986) as the minimizers of an asymmetric quadratic loss:

$$e_X(\alpha) = \arg \min_{t \in \mathbb{R}} \{E\ell_\alpha(X - t)\}, \tag{3}$$

where

$$\ell_\alpha(x) = \begin{cases} \alpha x^2, & \text{if } x \geq 0, \\ (1 - \alpha)x^2, & \text{if } x < 0, \end{cases}$$

and $\alpha \in (0, 1)$. For $X \in L^1$, equation (3) has to be modified (Newey & Powell, 1986) to

$$e_X(\alpha) = \arg \min_{t \in \mathbb{R}} \{E[\ell_\alpha(X - t) - \ell_\alpha(X)]\}. \tag{4}$$

The minimizer in (3) or (4) is always unique and is identified by the first order condition

$$\alpha E(X - e_X(\alpha))_+ = (1 - \alpha)E(X - e_X(\alpha))_-, \tag{5}$$

where $x_+ = \max\{x, 0\}$, $x_- = \max\{-x, 0\}$. This is equivalent to characterizing expectiles via an identification function, which, for any $\alpha \in (0, 1)$ is defined by

$$I_\alpha(x, y) = \alpha(y - x)\mathbb{1}_{\{y \geq x\}} - (1 - \alpha)(x - y)\mathbb{1}_{\{y < x\}}$$

for $x, y \in \mathbb{R}$. The α-expectile of a random variable $X \in L^1$ is then the unique solution of

$$EI_\alpha(t, X) = 0, \quad t \in \mathbb{R}.$$

Similarly, the *empirical α-expectile* $\hat{e}_n(\alpha)$ of a sample $X_1, \ldots, X_n$ is defined as the solution of

$$\frac{1}{n}\sum_{i=1}^{n} I_\alpha(t, X_i) = 0, \quad t \in \mathbb{R}.$$

As quantiles, expectiles are measures of non-central location; we collect some of their properties from Newey and Powell (1986) and Bellini et al. (2014) in the following proposition.

Proposition 1. *Let $X \in L^1$ with distribution function F and $\alpha \in (0, 1)$. Then:*

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

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

Clearly, expectiles depend only on the distribution of the random variable X; they can be seen as statistical functionals defined on the set of distribution functions with finite mean on $\mathbb{R}$.

Quantiles and expectiles are closely connected as measures of non-central location. Bellini et al. (2018) introduced the expectile (location) order between two random variables: Two random variables $X, Y \in L^1$ are ordered in expectile order (written $X \leq_e Y$) if $e_X(\alpha) \leq e_Y(\alpha)$ for all $\alpha \in (0, 1)$.

It is well known that the usual stochastic order $\leq_{st}$ is equivalent to the pointwise ordering of the quantiles. In view of this, the preceding definition seems quite natural, since quantiles are just replaced by expectiles.

The next theorem shows that the usual stochastic order implies the expectile order, that is, the ordering of the quantiles implies the ordering of the expectiles.

Theorem 1. *Let $X, Y \in L^1$. Then, $X \leq_{st} Y$ implies $X \leq_e Y$.*

This has been proved by Bellini (2012) by an order-theoretic comparative static approach. Since this proof may be not easily accessible for all readers, we give a rather elementary proof in the appendix.

3 | AN EXPECTILE-BASED ORDERING WITH RESPECT TO SKEWNESS

The distribution of a random variable X is symmetric around θ if $X - \theta \sim \theta - X$. Since a distribution is uniquely determined by the expectile function, and since the mean $\mu = e_X(1/2)$ of X coincides with the center of symmetry for a symmetric distribution, it follows that a distribution is symmetric if and only if $e_{X-\mu}(\alpha) = e_{\mu-X}(\alpha)$ for each $\alpha \in (0,1)$. Using properties (a) and (e) in Proposition 1, this is equivalent to

$$e_X(1-\alpha) - \mu = \mu - e_X(\alpha) \quad \text{for each} \quad \alpha \in (0,1/2).$$

In analogy to (1), this leads to the following definition of expectile-based skewness.

Definition 1. A distribution is called right-skewed (left-skewed) in the expectile sense, if and only if

$$e_X(1-\alpha) - \mu \ \geq \ (\leq) \ \mu - e_X(\alpha), \quad \alpha \in (0,1/2), \tag{6}$$

and equality does not hold for each $\alpha \in (0,1/2)$.

Corresponding scalar measures of skewness are

$$\tilde{s}_1 = \frac{e_X(3/4) + e_X(1/4) - 2\mu}{e_X(3/4) - e_X(1/4)}, \tag{7}$$

or the general version

$$\tilde{s}_2(\alpha) = \frac{e_X(1-\alpha) + e_X(\alpha) - 2\mu}{e_X(1-\alpha) - e_X(\alpha)}, \quad \alpha \in (0,1/2). \tag{8}$$

The numerator of (8) is the difference of two positive numbers $e_X(1-\alpha) - \mu$ and $\mu - e_X(\alpha)$, while the denominator is the sum of these numbers. Hence, $-1 \leq \tilde{s}_2(\alpha) \leq 1$.

Note that the following decomposition holds for the expectile as a measure of noncentral tendency:

$$\begin{aligned} e_X(1-\alpha) = e_X(1/2) + 1/2 \left\{ e_X(1-\alpha) - e_X(\alpha) \right\} \\ + 1/2 \left\{ e_X(1-\alpha) + e_X(\alpha) - 2e_X(1/2) \right\}. \end{aligned} \tag{9}$$

This is the counterpart to the decomposition given in Benjamini and Krieger (1996) for the quantile. The first term in (9), the mean, is a measure of central location; the second term, half of the expectile distance $e_X(1-\alpha) - e_X(\alpha)$, is a measure of variability for $\alpha < 1/2$. Finally, the third term, which is essentially the numerator in (8), is zero for symmetric distributions and, hence, quantifies the deviations from symmetry.

The actual range of $\tilde{s}_2(\alpha)$ depends on α and can be considerably smaller than the interval $[-1, 1]$. This is specified in the following result whose proof can be found in Appendix.

Proposition 2. *Let $\alpha \in (0, 1/2)$. Then, $-1 + 2\alpha < \tilde{s}_2(\alpha) < 1 - 2\alpha$ for all random variables, for which $\tilde{s}_2(\alpha)$ is defined, and both bounds cannot be improved.*

Based on this result we redefine our expectile-based skewness measure as

$$s_2(\alpha) = \frac{1}{1 - 2\alpha} \tilde{s}_2(\alpha), \quad \alpha \in (0, 1/2).$$

Then, $-1 < s_2(\alpha) < 1$, and both inequalities are sharp for any $\alpha \in (0, 1/2)$. Further, we define $s_1 = s_2(1/4) = 2\tilde{s}_1$.

Clearly, for the comparison of the skewness of two random variables, it does not matter if we use $\tilde{s}_2$ or s_2. Hence, we say that Y is more skewed to the right than X in the expectile sense if $\tilde{s}_{2,X}(\alpha) \leq \tilde{s}_{2,Y}(\alpha)$ for each $\alpha \in (0, 1/2)$. Using the properties of expectiles, it is clear that $\tilde{s}_2(\alpha)$ satisfies the skewness properties S1 and S2. The validity of S3 is an open question.

An even more general definition of an expectile-based skewness order would be the analogue to display (2); however, this seems to be rather unmanageable in applications.

For the quantile quotient $b_2(\alpha)$, MacGillivray (1992) proposed $\sup_{p < \alpha \leq 1/2} |b_2(\alpha)|$ as "a measure of overall asymmetry in the central $100 \cdot (1 - 2p)\%$ of the distribution." Because of the high similarity between b_2 and s_2 (or $\tilde{s}_2$), it suggests itself to propose $\sup_{p < \alpha \leq 1/2} |s_2(\alpha)|$ as a measure for the same asymmetry. However, $s_2(\alpha)$ and its empirical counterpart are much more stable as a function of α than $b_2(\alpha)$ and its empirical counterpart (see Eberl & Klar, 2020). Hence, there is no necessity for this kind of cumulative version of s_2.

Further indices of skewness could also be constructed by replacing expectiles in (8) by more general measures of noncentral location such as M-quantiles (Jones, 1994).

3.1 | The limiting case $\alpha \to 1/2$

In this section, we examine the behavior of $s_2(\alpha)$ if α approaches its upper bound $1/2$. For this, we assume that the cdf F of X is differentiable with density f; then, $e_X(\alpha)$ is twice differentiable by Theorem 1(f). Note that we can rewrite s_2 as a ratio of first- and second-order difference quotients

$$s_2(\alpha) = 1/2 \frac{(e_X(1/2 + h) - 2e_X(1/2) + e_X(1/2 - h))/h^2}{(e_X(1/2 + h) - e_X(1/2 - h))/h},$$

where $h = 1/2 - \alpha$. Splitting the central difference in the denominator into a forward and a backward difference, and taking the left limit yields

$$\lim_{\alpha \to 1/2-} s_2(\alpha) = \frac{e_X''(1/2)}{4e_X'(1/2)}.$$

By Proposition 1(f), $e_X'(1/2) = 2\delta_X$, where $\delta_X = E|X - \mu|$ denotes the mean absolute deviation (MAD) from the mean. For the calculation of $e_X''(1/2)$, we denote numerator and denominator of

$e'_X(\alpha)$ in Theorem 1(f) by $u(\alpha)$ and $v(\alpha)$, respectively. Then, $\lim_{\alpha\to 1/2} u(\alpha) = \delta_X$ and $\lim_{\alpha\to 1/2} v(\alpha) = 1/2$ as well as

$$u'(\alpha) = e'_X(\alpha)(2F(e_X(\alpha)) - 1) \to e'_X(1/2)(2F(\mu) - 1),$$
$$v'(\alpha) = (1 - 2F(e_X(\alpha))) + (1 - 2\alpha)f(e_X(\alpha))e'_X(\alpha) \to 1 - 2F(\mu),$$

for $\alpha \to 1/2$. By combining these results, it follows

$$e''_X(1/2) = \lim_{\alpha\to 1/2} \frac{u'(\alpha)v(\alpha) - u(\alpha)v'(\alpha)}{(v(\alpha))^2} = 8\delta_X(2F(\mu) - 1),$$

which overall yields

$$s_3 = \lim_{\alpha\to 1/2-} s_2(\alpha) = 2F(\mu) - 1. \tag{10}$$

Apparently, s_3 can also be used as a measure of skewness; in fact, it has already been introduced as such by Tajuddin (1999). Besides, the quantity $F(\mu)$ is the theoretical counterpart of the test statistic of the sign test for symmetry with estimated center (Gastwirth, 1971). The measure s_3 exploits the idea that the difference between mean μ and median $q_{1/2}$ indicates the skewness of the underlying distribution, which is also prevalent in other popular skewness measures like $(\mu - q_{1/2})/\sigma$ or $(\mu - q_{1/2})/E|X - q_{1/2}|$. Since a substitution of the mean by the median in s_3 always results in the value 0 for continuous distributions, a positive value of $\mu - q_{1/2}$ yields a positive value of s_3 and thus right-skewness and vice versa.

It is easy to see that s_3 satisfies the skewness properties S1 generally and S2 under the assumption $P(X = \mu) = 0$. For continuous distributions, the crucial property S3 follows from Jensen's inequality.

Besides its simplicity, an argument for the use of s_3 is that $s'_2(\alpha)$ converges to zero as α tends to 1/2, so $s_2(\alpha)$ flattens out toward s_3. This means that, at least for values of α close to 1/2, s_3 is close to and thereby representative for a range of values of $s_2(\alpha)$ without the need of a specific choice of the parameter α. This result on the gradient of s_2 can be proved under the assumption of $e_X \in C^4((0, 1))$ (which is equivalent to the assumption that the density of X is twice differentiable) as follows.

First, we differentiate s_2 with respect to α, which yields

$$s'_2(\alpha) = 2\frac{e'_X(\alpha)(e_X(1 - \alpha) - e_X(1/2)) - e'_X(1 - \alpha)(e_X(1/2) - e_X(\alpha))}{(1 - 2\alpha)(e_X(1 - \alpha) - e_X(\alpha))^2}$$
$$+ 2\frac{e_X(1 - \alpha) - 2e_X(1/2) + e_X(\alpha)}{(1 - 2\alpha)^2(e_X(1 - \alpha) - e_X(\alpha))},$$

for $\alpha \in (0, 1/2)$. Using the notation $h = 1/2 - \alpha$, this can once again be rewritten as a composition of difference quotients. Now, using Taylor expansions for each of them such that the remainders are of order $O(h^3)$ in the numerators as well as in the denominators yields after some computations

$$s'_2(\alpha) = 1/4\frac{-e'_X(1/2)e''_X(1/2)h + O(h^3)}{(e'_X(1/2))^2h^2 + O(h^3)} + 1/4\frac{e''_X(1/2) + 1/12e^{(4)}_X(1/2)h^2 + O(h^3)}{e'_X(1/2)h + O(h^3)}$$
$$= 1/4\frac{O(h^3)}{e'_X(1/2)h^2 + O(h^3)},$$

where we used for the second equality that $e'_X(1/2) > 0$ by Proposition 1c). Then, taking the limit $h \to 0$ yields the asserted result $\lim_{\alpha \to 1/2} s'_2(\alpha) = \lim_{h \to 0} s'_2(\alpha) = 0$.

4 | RELATION TO OMEGA RATIOS AND STOP-LOSS TRANSFORMS

In this section, we give conditions based on Omega ratios and stop-loss transforms which are equivalent to Definition 1.

Expectiles are related to the Omega ratio, which has been introduced in the financial literature by Keating and Shadwick (2002) as

$$\Omega_X(t) = \frac{E(X - t)_+}{E(X - t)_-}.$$

Then, Equation (5) can be written as

$$\Omega_X(e_X(\alpha)) = \frac{1 - \alpha}{\alpha}, \tag{11}$$

(Remillard, 2013), which gives the following one-to-one relation between expectiles and Omega ratios:

$$e_X(\alpha) = \Omega_X^{-1}\left(\frac{1 - \alpha}{\alpha}\right), \quad \Omega_X(t) = \frac{1 - e_X^{-1}(t)}{e_X^{-1}(t)}.$$

Arguing as at the beginning of Section 3, condition (6) for right-skewness is equivalent to

$$e_{X-\mu}(\alpha) \geq e_{-(X-\mu)}(\alpha), \quad \alpha \in (1/2, 1).$$

For $\alpha \in (1/2, 1)$, put $\beta = (1 - \alpha)/\alpha \in (0, 1)$. From Equation (11), the condition $e_{-(X-\mu)}(\alpha) \leq e_{X-\mu}(\alpha)$ is equivalent to

$$\Omega_{-(X-\mu)}(x) = \beta, \Omega_{X-\mu}(y) = \beta \Rightarrow x \leq y. \tag{12}$$

Since $\Omega_{-(X-\mu)}(0) = \Omega_{X-\mu}(0) = 1$, since $\Omega_{-(X-\mu)}$ and $\Omega_{X-\mu}$ are strictly decreasing and since (12) holds for each $\beta \in (0, 1)$, it follows that

$$\Omega_{-(X-\mu)}(t) \leq \Omega_{X-\mu}(t) \text{ for all } t \in (0, \infty). \tag{13}$$

Using $\Omega_{X-\mu}(t) = \Omega_X(\mu + t)$ and $\Omega_{-(X-\mu)}(t) = 1/\Omega_X(\mu - t)$, we finally obtain that (13), and, hence, condition (6) for right-skewness, is equivalent to

$$\Omega_X(\mu + t) \cdot \Omega_X(\mu - t) \geq 1 \quad \text{for all } t > 0. \tag{14}$$

Further note that a distribution is symmetric if and only if equality holds in (14) for every $t > 0$.

The Omega ratio, in turn, is closely related to the *stop-loss transform* $\pi_X(t) = E(X - t)_+$, which is well known in the actuarial literature (see e.g., Müller, 1996), since it describes

the expected cost of a stop-loss insurance contract with deductible t for a risk X. From

$$E(X - t)_- = t - EX + E(X - t)_+$$

we immediately get

$$\Omega_X(t) = \frac{\pi_X(t)}{t - EX + \pi_X(t)}. \tag{15}$$

Plugging (15) into condition (14), we obtain the following result:

Theorem 2.

(a) *The distribution of a random variable X with cdf F and finite mean $\mu = EX$ is symmetric around μ if and only if*

$$S_X(t) = \frac{1}{t}\left\{\pi_X(\mu + t) - \pi_X(\mu - t)\right\} + 1 = 0 \quad \text{for each } t > 0.$$

(b) *A distribution is right-skewed (left-skewed) in the sense of Definition 1 if and only if*

$$S_X(t) \geq (\leq) 0 \quad \text{for each } t > 0.$$

The following proposition collects some properties of S_X.

Proposition 3. *Let X be random variable with cdf F and finite mean μ. Then:*

(a) $\lim_{t\to\infty} S_X(t) = 0$.
(b) $S_X(t) = \frac{1}{t}\int_{\mu-t}^{\mu+t} F_X(z)dz - 1$.
(c) $-1 \leq S_X(t) \leq 1$.

Proof. Since $\lim_{t\to\infty}\pi_X(\mu + t) = 0$, and since the monotone convergence theorem implies

$$\lim_{t\to\infty}\left\{\pi_X(\mu - t) - t\right\} = \lim_{t\to\infty} E\left[\max\{X - \mu, -t\}\right] = 0,$$

we obtain $\lim_{t\to\infty} S_X(t) = 0$. Further,

$$S_X(t) = \frac{1}{t}\left\{\int_{\mu+t}^{\infty}\overline{F}_X(z)dz - \int_{\mu-t}^{\infty}\overline{F}_X(z)dz\right\} + 1$$

$$= 1 - \frac{1}{t}\int_{\mu-t}^{\mu+t}\overline{F}_X(z)dz = \frac{1}{t}\int_{\mu-t}^{\mu+t} F_X(z)dz - 1,$$

where $\overline{F}_X(z) = 1 - F_X(z)$ denotes the survivor function. Part (c) follows directly from (b). ∎

Figure 1 illustrates the area below $F_X(z)$ for $z \in [\mu - t, \mu + t]$ for a (symmetric) normal distribution $N(2, 4)$ (left panel, $t = 2.5$) and a right-skewed exponential distribution with mean 5 (right
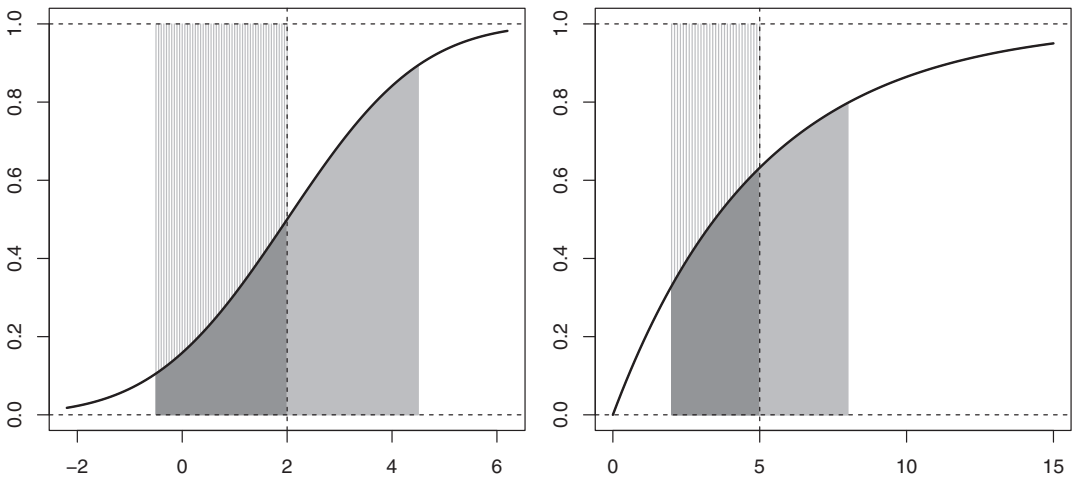
FIGURE 1 Area below $F_X(z)$ for $z \in [\mu - t, \mu + t]$ for a (symmetric) normal distribution $N(2, 4)$ (left panel) and a right-skewed exponential distribution with mean 5 (right panel)

panel, $t = 3$). For the normal distribution, the gray areas below F_X sum up to $t = 2.5$, whereas the sum is larger than $t = 3$ in case of the exponential distribution.

Remark 1. (i) The representation of S_X in Proposition 3(c) bears some similarity to skewness functionals defined in Arnold and Groeneveld (1993). In particular, they proposed

$$\lambda_X(u) = \int_0^u F^{-1}(1/2 + v) + F^{-1}(1/2 - v) dv, \ 0 < u < 1/2,$$

as skewness function. Note, however, that this is a skewness measure with respect to the median, whereas S_X is a measure with respect to the mean (cf. MacGillivray, 1986).

(ii) Note that $S_X(t) = 2\int_{\mathbb{R}} F_X(z) dH(z) - 1$, where H is the cdf of the uniform distribution on $(\mu - t, \mu + t)$. Replacing H by the Dirac measure in μ results in s_3 in (10). Another reasonable choice for H would be any cdf with unimodal density, that is symmetric around μ, for example a normal distribution with mean μ.

$S_X(t)$ is location invariant, but not scale invariant. This is not an issue if one analyzes the skewness of a single distribution. However, scale invariance is essential for a meaningful comparison between several distributions. As a scale invariant modification, we propose

$$\tilde{S}_X(t) = S_X(t\delta_X), \tag{16}$$

where $\delta_X = E|X - EX|$, as before, denotes the MAD from the mean. This dispersion measure is strongly related to the stop-loss transform, since $\pi_X(EX) = \delta_X/2$ is just the absolute semideviation. In principle, one could use any other dispersion measure σ_X satisfying $\sigma_{cX} = c\sigma_X$ for $c > 0$ instead of δ_X, but the latter is particularly suitable for our purpose.

For the remainder of this section we assume that the cdf's are absolutely continuous and strictly increasing (i.e., F is strictly increasing on $\{x : 0 < F(x) < 1\}$). Oja (1981) showed that F and

G are strongly skewness comparable (i.e., $F \leq_2 G$ or $G \leq_2 F$) if and only if $F(x)$ and $G(ax+b)$ cross each other at most twice for all $a > 0, b \in \mathbb{R}$. He then defined two weakenings of $\leq_2$ in case of finite expectations μ_F and μ_G and finite variances σ_F and σ_G as follows.

- $F \leq_2^* G$ if the standardized distribution functions $F(\sigma_F x + \mu_F)$ and $G(\sigma_G x + \mu_G)$ cross each other exactly once on each side of $x = 0$, with $F(\mu_F) \leq G(\mu_G)$.

- $F \leq_2^{**} G$ if there exist $a > 0, b \in \mathbb{R}$ such that $F(x)$ and $G(ax+b)$ cross each other exactly twice with $F(x) - G(ax+b)$ changing sign from positive to negative to positive.

The following implications hold true (Oja, 1981):

$$F \leq_2 G \quad \Rightarrow \quad F \leq_2^* G \quad \Rightarrow \quad F \leq_2^{**} G.$$

Similarly like $F \leq_2^* G$ (see also definition 2.1 in MacGillivray, 1986), we now define skewness with respect to mean and MAD:

Definition 2. G is more skew with respect to mean and MAD than F ($F <_\mu^\delta G$), if $F(\delta_F x + \mu_F)$ and $G(\delta_G x + \mu_G)$ cross each other exactly once on each side of $x = 0$, with $F(\mu_F) \leq G(\mu_G)$.

More generally, $F \leq_\mu^\delta G$ if $F <_\mu^\delta G$ or $F(\delta_F \cdot + \mu_F)$ and $G(\delta_G \cdot + \mu_G)$ are identical.

The next theorem shows that this new skewness order is weaker than the strong skewness order $\leq_2$; on the other hand, it is stronger than the skewness order implied by $\tilde{S}_X$ given in (16).

Theorem 3. *Let $X \sim F$ and $Y \sim G$ with finite expectations $\mu_F = EX$ and $\mu_G = EY$. Then:*

(a) $F \leq_2 G \quad \Rightarrow \quad F \leq_\mu^\delta G \quad \Rightarrow \quad F \leq_2^{**} G.$
(b) $F \leq_\mu^\delta G \quad \Rightarrow \quad \tilde{S}_X(t) \leq \tilde{S}_Y(t) \ \forall \ t > 0.$
In particular, the skewness measure $\tilde{S}_X(t)$ satisfies skewness property S3 for any $t > 0$.

Remark 2. (i) The proof of Theorem 3 shows that it is reasonable to require exactly two crossings in Definition 2. Exactly one crossing can occur only in specific situations where the standardized distribution functions are identical for all values smaller (larger) than zero; in this case, there is no reasonable comparison between the skewness of the two cdf's.

(ii) From Theorem 3 and the strong connection between $\tilde{S}_X$ and the skewness measures s_2 in (8) we conjecture that s_2 also satisfies property S3. This is reinforced by the validity of S3 for the limiting measure s_3, and by numerical computations for specific examples, see Section 6.

5 | EMPIRICAL EXPECTILE SKEWNESS

By replacing the theoretical expectiles by empirical ones we obtain the plug-in estimator

$$\hat{s}_{2,n}(\alpha) = \frac{1}{1 - 2\alpha} \frac{\hat{e}_n(1 - \alpha) + \hat{e}_n(\alpha) - 2\bar{x}}{\hat{e}_n(1 - \alpha) - \hat{e}_n(\alpha)}, \quad \alpha \in (0, 1/2),$$

and $\hat{s}_{1,n} = \hat{s}_{2,n}(1/4)$. Utilizing the asymptotic normality of a finite number of expectiles (see, e.g., Holzmann & Klar, 2016) and the fact that $s_2(\alpha)$ is a differentiable function of expectiles, the delta method yields the following theorem. Preliminarily, we define

$$\eta(\tau_1, \tau_2) = E[I_{\tau_1}(e_X(\tau_1), X)I_{\tau_2}(e_X(\tau_2), X)],$$

for $\tau_1, \tau_2 \in (0, 1)$ and

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

for $\tau \in \{\alpha, 1 - \alpha\}$.

Theorem 4. *Let $\alpha \in (0, 1/2)$ and let $X, X_1, X_2, \ldots$ be iid with cdf F such that $EX^2 < \infty$. Furthermore, let F not have a point mass in $e_X(p)$ for $p \in \{\alpha, 1/2, 1 - \alpha\}$. Then*

$$\sqrt{n}(\hat{s}_{2,n}(\alpha) - s_2(\alpha)) \xrightarrow{D} N(0, \sigma_\alpha^2),$$

where

$$\sigma_\alpha^2 = \frac{4}{(1 - 2\alpha)^2}\left[\frac{4\eta(1/2, 1/2)}{(e_X(1-\alpha) - e_X(\alpha))^2} - \frac{4[A(\alpha)\eta(\alpha, 1/2) + A(1-\alpha)\eta(1/2, 1-\alpha)]}{(e_X(1-\alpha) - e_X(\alpha))^3}\right.$$

$$\left. + \frac{(A(\alpha))^2\eta(\alpha, \alpha) + A(\alpha)A(1-\alpha)\eta(\alpha, 1-\alpha) + (A(1-\alpha))^2\eta(1-\alpha, 1-\alpha)}{(e_X(1-\alpha) - e_X(\alpha))^4}\right].$$

It can also be shown that $\hat{s}_{2,n}(\alpha)$ is a strongly consistent estimator for $s_2(\alpha)$ for all $\alpha \in (0, 1/2)$. To see this, we consider theorem 2 in Holzmann and Klar (2016), which implies strong consistency of $\hat{e}_n(\tau)$ for $e_X(\tau)$ for all $\tau \in (0, 1)$ under weak assumptions. Using Slutzky's Theorem, the almost sure convergence then also holds for any continuous function of any finite number of expectiles, yielding the following corollary.

Corollary 1. *Let $\alpha \in (0, 1/2)$ and let $X, X_1, X_2, \ldots$ be iid with cdf F such that $E|X|<\infty$. Then, $\hat{s}_{2,n}(\alpha)$ is strongly consistent for $s_2(\alpha)$, that is, $\hat{s}_{2,n}(\alpha) \xrightarrow{a.s.} s_2(\alpha)$.*

In order to obtain the plug-in estimator $\hat{\sigma}_\alpha^2$ of σ_α^2, the expectiles $e_X(\tau)$ are replaced by the empirical expectiles $\hat{e}_n(\tau)$. Moreover, $\eta(\tau_1, \tau_2)$ and $A(\tau)$ are estimated by

$$\hat{\eta}_n(\tau_1, \tau_2) = \frac{1}{n}\sum_{i=1}^{n} I_{\tau_1}(\hat{e}_n(\tau_1), X_i)I_{\tau_2}(\hat{e}_n(\tau_2), X_i),$$

and

$$\hat{A}_n(\tau) = (2\mathbb{1}_{\{\tau<1/2\}} - 1)\frac{\hat{e}_n(1-\tau) - \overline{X}}{\tau + \hat{F}_n(\hat{e}_n(\tau))(1 - 2\tau)},$$

where $\hat{F}_n$ denotes the empirical cdf. It is then easy to see that $\hat{\sigma}_\alpha^2$ is a composition of consistent estimators, hence $\hat{\sigma}_\alpha^2$ itself is a consistent estimator of σ_α^2. Consequently, an asymptotic confidence
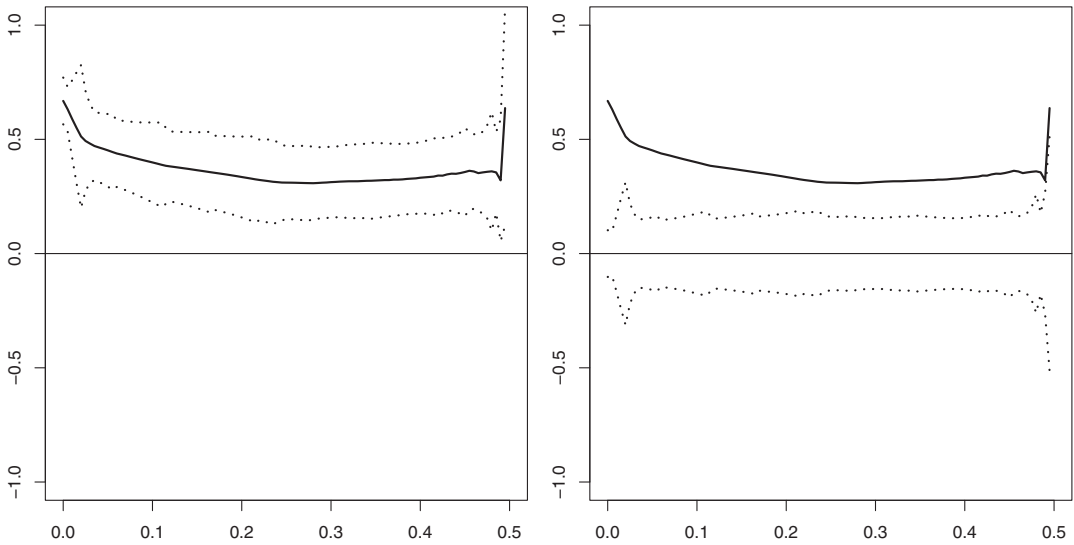
FIGURE 2 Left panel: Plot of $\hat{s}_{2,n}(\alpha)$ with (pointwise) 95%-confidence limits defined in (17) as dotted line. Right panel: Plot of $\hat{s}_{2,n}(\alpha)$ with limits under the assumption of symmetry defined in (18) as dotted line

interval for $s_2(\alpha)$ with confidence level $1 - p$ is given by

$$\hat{s}_{2,n}(\alpha) - \frac{\widehat{\sigma}_\alpha}{\sqrt{n}} z_{1-p/2} \leq s_2(\alpha) \leq \hat{s}_{2,n}(\alpha) + \frac{\widehat{\sigma}_\alpha}{\sqrt{n}} z_{1-p/2}, \tag{17}$$

where z_q denotes the q-quantile of the standard normal distribution. The left panel in Figure 2 shows a plot of $\hat{s}_{2,n}(\alpha)$ with (pointwise) confidence limits for a sample of size $n = 50$ from an exponential distribution with rate 1.

Under the hypothesis of symmetry, $s_2(\alpha) \equiv 0$. Therefore,

$$\lim_{n \to \infty} P\left(-\frac{\widehat{\sigma}_\alpha}{\sqrt{n}} \cdot z_{1-p/2} \leq \hat{s}_{2,n}(\alpha) \leq \frac{\widehat{\sigma}_\alpha}{\sqrt{n}} \cdot z_{1-p/2} \right) = 1 - \alpha, \tag{18}$$

which define confidence limits under the hypothesis of symmetry. The right panel in Figure 2 shows a plot of $\hat{s}_{2,n}(\alpha)$ together with the limits given in (18) for the same dataset as in the left panel.

With respect to $S(t)$, we again use the plug-in estimator, which is given by

$$S_n(t) = \frac{1}{nt} \sum_{i=1}^{n} \left\{ (X_i - \overline{X} - t)_+ - (X_i - \overline{X} + t)_+ \right\} + 1.$$

Then the following result holds.

Theorem 5. *Let $X_1, X_2, \ldots$ be iid with continuous cdf F and $EX^2 < \infty$. Then,*

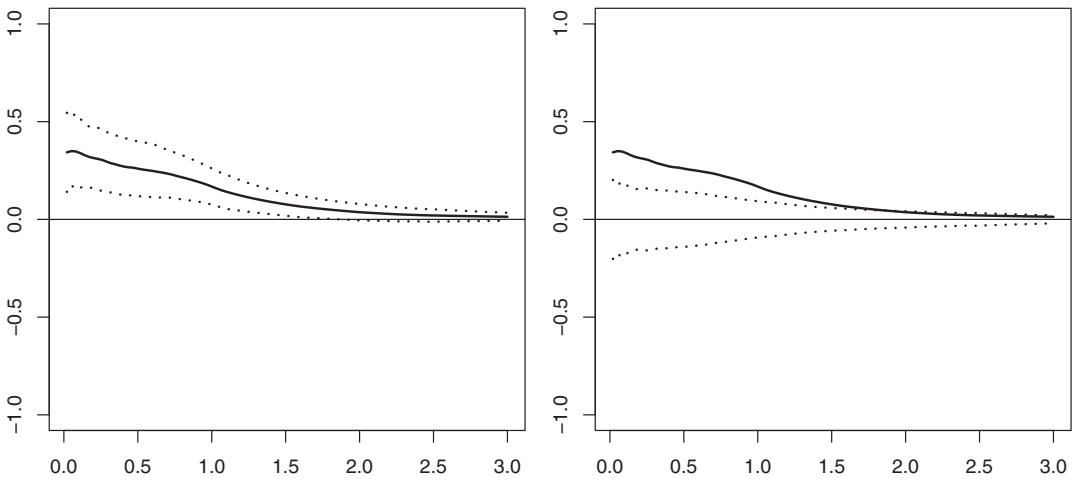$$\sqrt{n}\,(S_n(t) - S(t)) \xrightarrow{D} N(0, \sigma_t^2),$$

FIGURE 3 Left panel: Plot of $S_n(t)$ with (pointwise) 95%-confidence limits defined in (19) as dotted line. Right panel: Plot of $S_n(t)$ with limits under the assumption of symmetry defined in (20) as dotted line

where

$$\sigma_t^2 = \frac{1}{t^2} \cdot \mathrm{Var}\left((X_1 - \mu - t)_+ - (X_1 - \mu + t)_+ + (X_1 - \mu)\left(F(\mu + t) - F(\mu - t)\right)\right).$$

The plug-in estimator for σ_t^2 is

$$\widehat{\sigma}_t^2 = \frac{1}{nt^2} \sum_{i=1}^{n} \left((X_i - \overline{X} - t)_+ - (X_i - \overline{X} + t)_+ + (X_i - \overline{X})\,\hat{p}_t\right)^2$$

$$- \left\{ \frac{1}{nt} \sum_{i=1}^{n} \left((X_i - \overline{X} - t)_+ - (X_i - \overline{X} + t)_+\right) \right\}^2,$$

where $\hat{p}_t = 1/n \sum_{i=1}^{n} \mathbb{1}\{\mu - t < X_i \le \mu + t\}$. Analogous to the expectile skewness $s_2(\alpha)$, an asymptotic confidence interval for $S(t)$ with confidence level $1 - p$ is given by

$$S_n(t) - \frac{\widehat{\sigma}_t}{\sqrt{n}} \cdot z_{1-p/2} \le S(t) \le S_n(t) + \frac{\widehat{\sigma}_t}{\sqrt{n}} \cdot z_{1-p/2}. \tag{19}$$

Exemplary, (pointwise) confidence limits for a sample of size 50 from an exponential distribution with rate 1 are given in the left panel of Figure 3.

As before, the hypothesis of symmetry yields $S(t) \equiv 0$ and, thereby,

$$\lim_{n \to \infty} P\left(-\frac{\widehat{\sigma}_t}{\sqrt{n}} \cdot z_{1-\alpha/2} \le S_n(t) \le \frac{\widehat{\sigma}_t}{\sqrt{n}} \cdot z_{1-\alpha/2}\right) = 1 - \alpha, \tag{20}$$

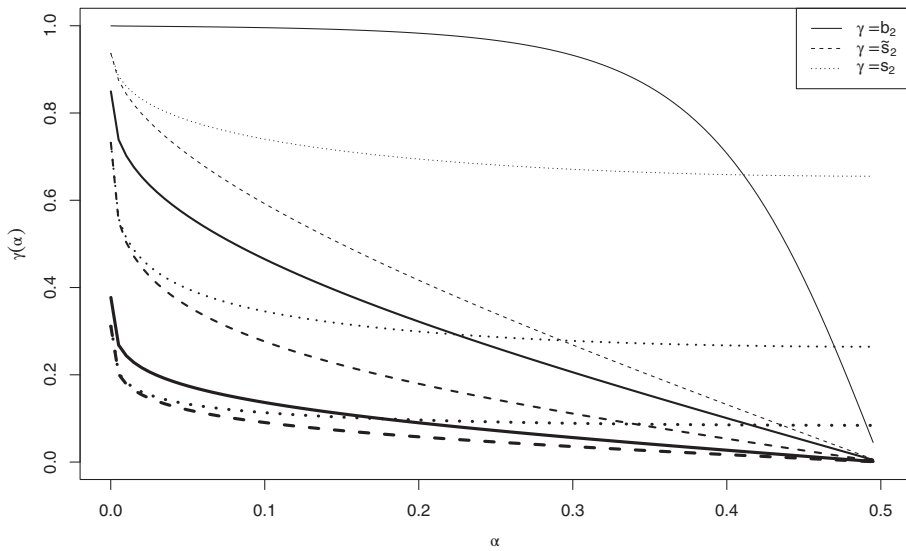the respective confidence limits are given in the right panel of Figure 3.

FIGURE 4 Expectile and quantile skewness as functions of α, for gamma distributions with shape parameters $k = 0.1$, 1 and 10, where the line width increases with the value of k

6 | EXPECTILE AND QUANTILE SKEWNESS FOR SOME FAMILIES OF DISTRIBUTIONS

6.1 | Comparison of theoretical values

In this section we examine how the expectile skewness $s_2(\alpha)$ behaves for specific families of continuous distributions, in particular for the gamma distribution. We analyze how the skewness values depend on α as well as on the distributional parameters. The results are compared with the corresponding values of the quantile skewness $b_2(\alpha)$. Due to property S1, skewness does only depend on the shape parameter of the gamma distribution, but not on the scale parameter. The density of the gamma distribution is given by

$$f_{k,r}(t) = \frac{r^k}{\Gamma(k)} t^{k-1} e^{-rt} \mathbb{1}_{[0,\infty)}(t),$$

with shape parameter $k > 0$ and rate parameter $r > 0$. Here, $\Gamma(\cdot)$ denotes the gamma function. All gamma distributions are right skewed and the degree of their skewness decreases as k increases. Figure 4 depicts skewness as a function of α for gamma distributions with different shape parameters.

First, we only look at the measures b_2 and $\tilde{s}_2$ without the correction term $(1 - 2\alpha)^{-1}$. Both of them tend to 0 as α tends to 1/2. All curves are strictly decreasing in α. While the quantile skewness exceeds the diagonal for highly skewed distributions, the expectile skewness is restricted to values below the diagonal, corresponding to Proposition 2. The curves above the diagonal are concave while the ones underneath are convex.
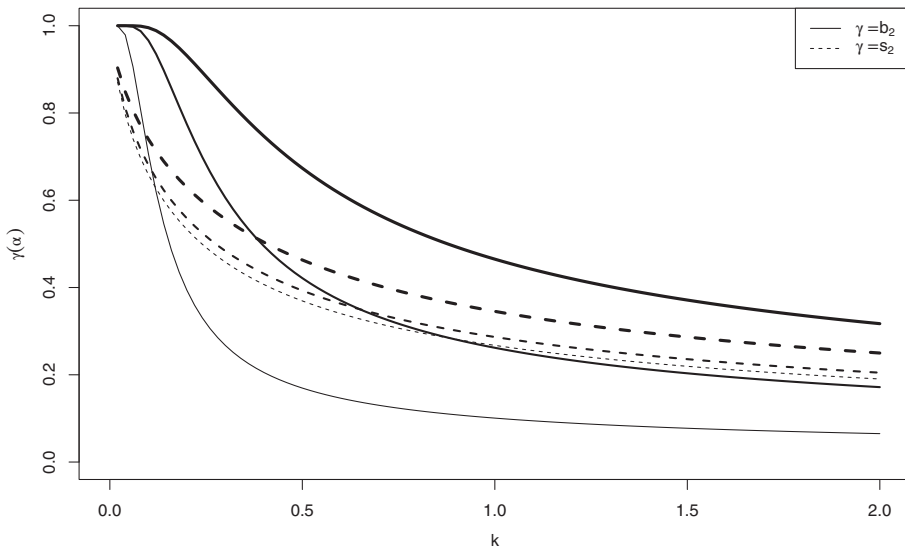
FIGURE 5 Expectile and quantile skewness as functions of the shape parameter k of the underlying gamma distribution for values $\alpha = 0.1, 0.25, 0.4$, where the line width increases with the value of α

If the expectile skewness is normalized to 1, it no longer tends to 0 as α tends to 1/2. Instead, the still convex curves flatten out with increasing α after a steep decline close to 0, illustrating that $s_2'(\alpha)$ converges to zero as α tends to 1/2. Since the curves flatten out rather quickly, this implies that the limiting expectile skewness s_3 is representative of $s_2(\alpha)$ for a considerable part of the range of α.

If the quantile skewness is multiplied with the factor $(1 - 2\alpha)^{-1}$, plots show that it also flattens out toward some limiting value with diminishing gradient as α approaches 1/2. However, these values are then no longer normalized and can be equal to any real number.

The observed behavior is very similar for other popular classes of skewed distributions like the log-normal and the Weibull distribution. If the underlying distribution is skewed to the left, all considered skewness measures increase in α with the expectile skewness curves being concave as long as they stay above the corresponding lower diagonal.

Now we look at the behaviour of the skewness measures b_2 and s_2 as functions of the shape parameter of the underlying distribution. For the shape parameter k of the gamma distribution specifically, this is depicted in Figure 5.

Both skewness measures decrease in k for all values of α. This was to be expected since van Zwet (1964, pp. 60-62) showed that $F_{k_1}^{-1} \circ F_{k_2}$ is convex for $k_1 \leq k_2$, where F_k denotes the cdf of the gamma distribution with shape parameter k. The qualitative behaviour is analogous for similarly ordered classes of distributions like the Weibull distribution, thus strengthening our conjecture that s_2 satisfies skewness property S3. The curves of the expectile and quantile skewness differ slightly with the former being strictly convex while the latter become concave for k close to 0. However, except for very small values of k, b_2 decreases more rapidly than s_2, especially for small values of α. The plot also further confirms that the range of s_2 for different values of α is substantially smaller than that of b_2.

6.2 | Performance of the empirical skewness measures

In this section, we examine and compare bias and variance of different empirical skewness measures. Here, we include quantile skewness $b_2(\alpha)$, expectile skewness $s_2(\alpha)$, Tajuddin's measure s_3 and the moment skewness $\gamma_M = E((X - \mu)/\sigma)^3$. All corresponding empirical measures are obtained as plug-in estimators. Their behavior is explored for gamma distributions with varying shape parameter $k \in [0.1, 10]$, for the log-normal distribution with fixed log-mean 0 and varying log-variance $\tau^2 \in [0.01, 2.25]$ as well as for Student's t-distribution with varying degrees of freedom $k \in \mathbb{N} \setminus \{1, 2\} \cup \{\infty\}$ (where $k = \infty$ corresponds to the standard normal distribution). For any skewness measure γ with empirical version $\hat{\gamma}_n$, we define the standardized bias, variance, and mean squared error (MSE) by

$$\text{sbias}(\hat{\gamma}_n, \gamma) = \frac{\text{bias}(\hat{\gamma}_n, \gamma)}{\gamma}, \quad \text{sVar}(\hat{\gamma}_n) = \frac{\text{Var}(\hat{\gamma}_n)}{\gamma^2}, \quad \text{sMSE}(\hat{\gamma}_n, \gamma) = \frac{\text{MSE}(\hat{\gamma}_n, \gamma)}{\gamma^2},$$

if $\gamma \neq 0$, and as their nonstandardized versions otherwise. Since the different skewness measures are scaled differently and might also behave differently within the chosen classes of distributions, this makes their biases, variances and MSE's comparable. For the parameter α, the values 0.1, 0.25 and 0.4 have been considered, the sample size n varied from 20 to 10,000, and each simulation is based on 10,000 repetitions.

First, we consider a highly skewed (shape parameter 0.1, see Figure 6) and a mildly skewed (shape parameter 10, see Figure 7) gamma distribution. We observe that the MSE generally seems to decrease with increasing skewness; however, that decrease is slower for γ_M than for the other measures. While b_2 and γ_M have the highest MSE at either end of the skewness spectrum, the expectile skewness is always in an acceptable range and converges fairly quickly toward 0 as n increases. While there is almost no bias for the mildly skewed distribution, all measures are at
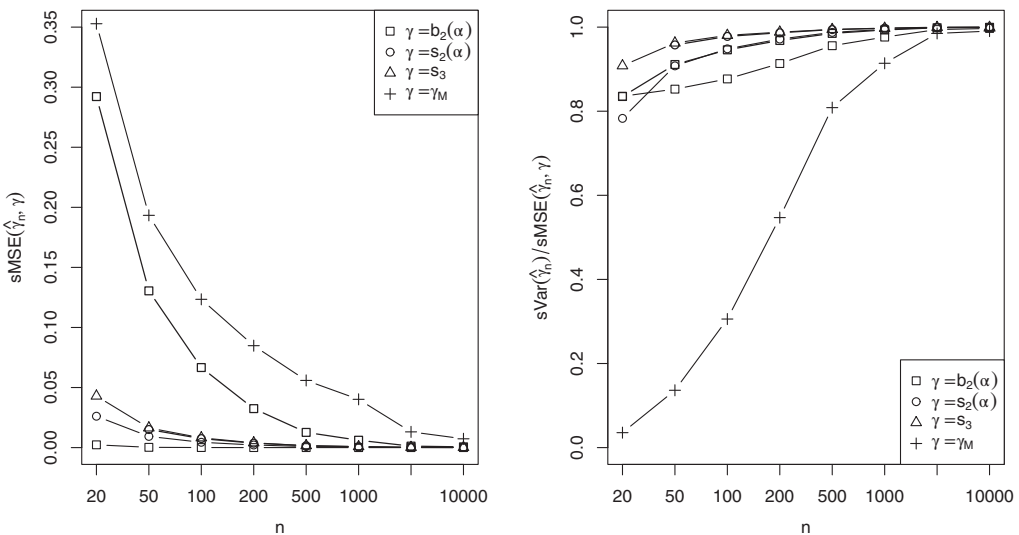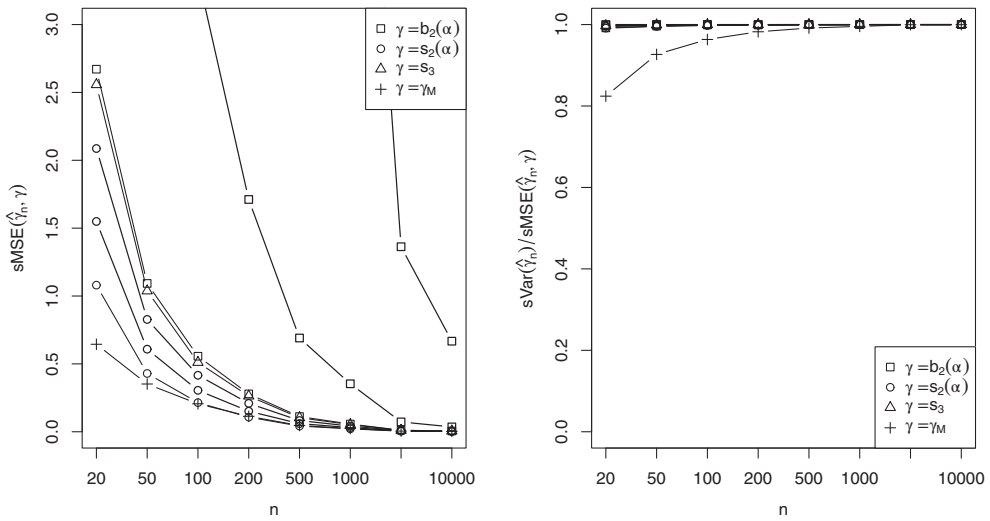

FIGURE 6 Left panel: Standardized MSE's. Right panel: Percentage of the MSE taken up by the variance. Underlying distribution: $\Gamma(0.1, 1)$. Moment skewness: 6.325. For skewness measures depending on a parameter α, increasing line width symbolizes increasing values of $\alpha \in \{0.1, 0.25, 0.4\}$
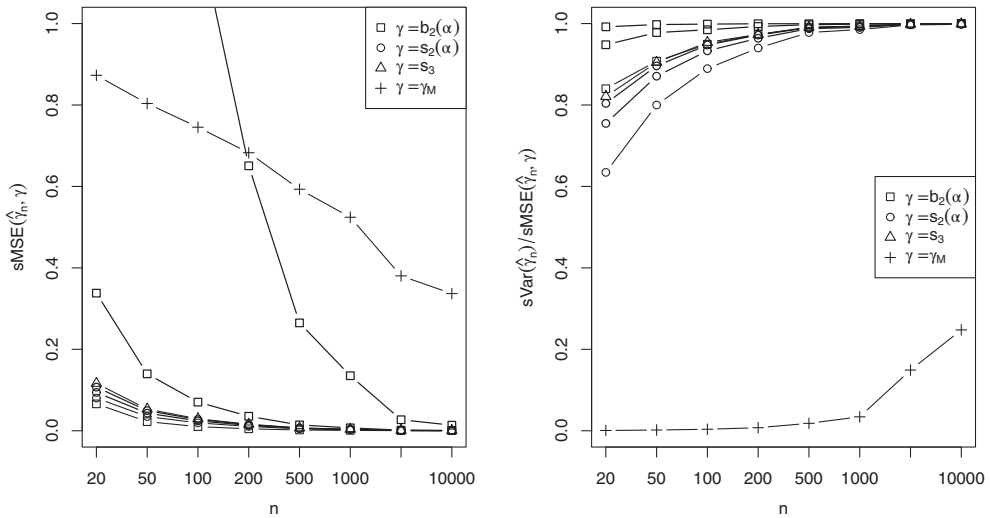
FIGURE 7   Left panel: Standardized MSE's. Right panel: Percentage of the MSE taken up by the variance. Underlying distribution: $\Gamma(10, 1)$. Moment skewness: 0.632. For skewness measures depending on a parameter α, increasing line width symbolizes increasing values of $\alpha \in \{0.1, 0.25, 0.4\}$

FIGURE 8   Left panel: Standardized MSE's. Right panel: Percentage of the MSE taken up by the variance. Underlying distribution: $\mathcal{LN}(0, 2.25)$. Moment skewness: 33.468. For skewness measures depending on a parameter α, increasing line width symbolizes increasing values of $\alpha \in \{0.1, 0.25, 0.4\}$

least slightly biased (relative to their variance) for high skewness. For increasing n, the bias vanishes. Irrespective of the distributional skewness, γ_M is always the most biased measure, for high skewness even to a critical degree.

The results for the highly skewed (log-variance 2.25, see Figure 8) and the mildly skewed (log-variance 0.01, see Figure 9) log-normal distributions confirm the observations made concerning the gamma distribution. The first log-normal distribution is even more skewed than the first gamma distribution, having the effect that the MSE of γ_M is almost completely dominated by the
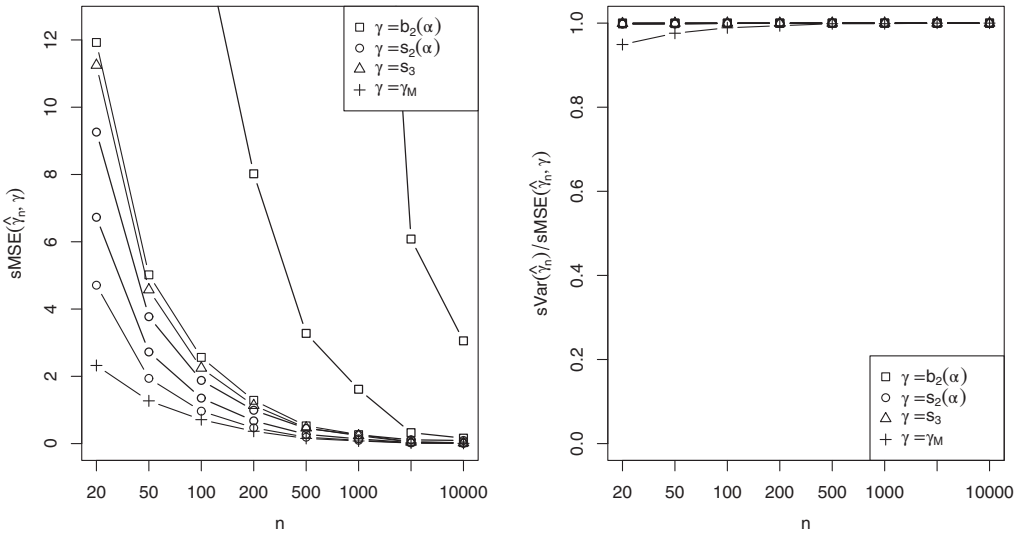
FIGURE 9 Left panel: Standardized MSE's. Right panel: Percentage of the MSE taken up by the variance. Underlying distribution: $\mathcal{LN}(0, 0.01)$. Moment skewness: 0.302. For skewness measures depending on a parameter α, increasing line width symbolizes increasing values of $\alpha \in \{0.1, 0.25, 0.4\}$
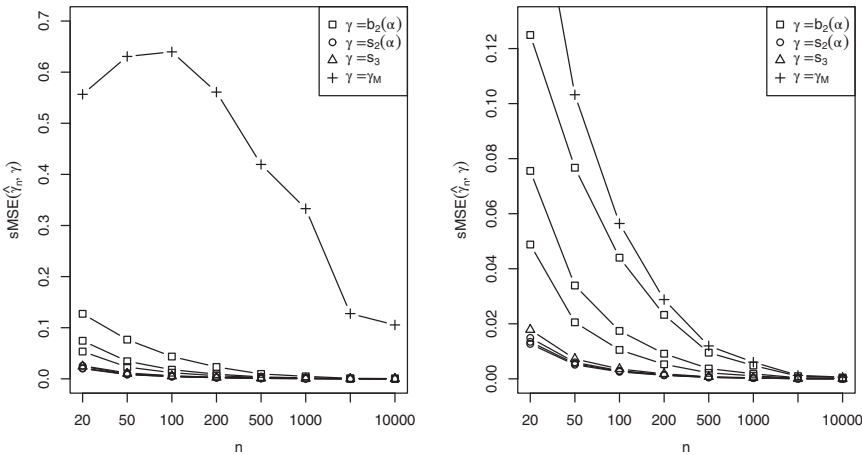

FIGURE 10 Standardized MSE's. Left panel: t_5. Right panel: $t_\infty \sim N(0, 1)$. Moment skewness: 0. For skewness measures depending on a parameter α, increasing line width symbolizes increasing values of $\alpha \in \{0.1, 0.25, 0.4\}$

bias. Additionally, the MSE of γ_M seems to converge very slowly relative to the other measures, possibly suggesting worse behaviour on heavy-tailed distributions.

Finally, we consider Student's t-distribution and the standard normal distribution (as limiting case) as examples of symmetric distributions. As expected for these distributions, the bias is negligible relative to the variance for all skewness measures. While the MSE basically does not change for b_2 and s_2 (with the latter achieving even lower values), γ_M only behaves nicely for higher degrees of freedom. For lower degrees of freedom, it becomes somewhat unstable (see Figure 10), once again showing a poor behaviour for heavy-tailed distributions.

Overall, $s_2(\alpha)$ and s_3 seem to be the most stable skewness measures considered here. While they are outperformed for specific distributions, their MSE never explodes, and their bias is always fairly low compared to their variance.

ORCID

Bernhard Klar https://orcid.org/0000-0002-1419-5473

REFERENCES

Arnold, B. C., & Groeneveld, R. A. (1993). Skewness and kurtosis orderings: An introduction. *Stochastic inequalities, IMS Lecture Notes - Monograph Series*, *22*, 17–24.

Artin, E. (1964). *The gamma function*. Holt, Rinehart and Winston.

Bellini, F. (2012). Isotonicity results for generalized quantiles. *Statistics & Probability Letters*, *82*, 2017–2024.

Bellini, F., Klar, B., & Müller, A. (2018). Expectiles, Omega ratios and stochastic ordering. *Methodology and Computing in Applied Probability*, *20*, 855–873.

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

Benjamini, Y., & Krieger, A. M. (1996). Concepts and measures for skewness with data-analytic implications. *The Canadian Journal of Statistics*, *24*, 131–140.

Bowley, A. L. (1920). *Elements of statistics* (4th ed.). Charles Scribner's Sons.

Critchley, F., & Jones, M. C. (2008). Asymmetry and gradient asymmetry functions: Density-based skewness and kurtosis. *Scandinavian Journal of Statistics*, *35*, 415–437.

David, F. N., & Johnson, N. L. (1956). Some tests of significance with ordered variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *18*, 1–31.

Eberl, A., & Klar, B. (2020). Asymptotic distributions and performance of empirical skewness measures. *Computational Statistics & Data Analysis*, *146*, 106939.

Gastwirth, J. L. (1971). On the sign test for symmetry. *Journal of the American Statistical Association*, *66*, 821–823.

Groeneveld, R. A., & Meeden, G. (1984). Measuring skewness and kurtosis. *Journal of the Royal Statistical Society Series D*, *33*, 391–399.

Hinkley, D. V. (1975). On power transformations to symmetry. *Biometrika*, *62*, 101–111.

Holzmann, H., & Klar, B. (2016). Expectile asymptotics. *Electronic Journal of Statistics*, *10*, 2355–2371.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society Series B*, *52*, 105–124.

Jones, M. C. (1994). Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters*, *20*, 149–153.

Keating, C., & Shadwick, W. F. (2002). *A universal performance measure*. The Finance Development Centre.

Klar, B. (2015). A note on gamma difference distributions. *Journal of Statistical Computation and Simulation*, *85*, 3708–3715.

MacGillivray, H. L. (1986). Skewness and asymmetry: Measures and orderings. *Annals of Statistics*, *14*, 994–1011.

MacGillivray, H. L. (1992). Shape properties of the g-and-h and Johnson families. *Communications in Statistics Theory and Methods*, *21*, 1233–1250.

Müller, A. (1996). Ordering of risks: A comparative study via stop-loss transforms. *Insurance: Mathematics and Economics*, *17*, 215–222.

Müller, A., & Stoyan, D. (2002). *Comparison methods for stochastic models and risks*. John Wiley & Sons Ltd.

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

Oja, H. (1981). On Location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics*, *8*, 154–168.

Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., & Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, *21*, 201–214.

Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., & Mykytka, E. F. (1980). Letters to the editor. *Technometrics*, *22*, 135.

Remillard, B. (2013). *Statistical methods for financial engineering*. Chapman & Hall / CRC Press.

Tajuddin, I. H. (1999). A comparison between two simple measures of skewness. *Journal of Applied Statistics*, *26*, 767–774.

van Zwet, W. R. (1964). *Convex transformations of random variables*. Mathematisch Centrum.

Yule, G. U. (1912). *An introduction to the theory of statistics* (1st ed.). Charles Griffin and Co.

APPENDIX A

Proof of the ⩽ equivalence of $\leq_2$ and (2). Convexity of a function h on an interval I is equivalent to

$$\frac{h(x)(z - y) + h(y)(x - z) + h(z)(y - x)}{(x - y)(y - z)(z - x)} \geq 0, \tag{A1}$$

for all distinct numbers $x, y, z \in I$ (see, e.g., Artin, 1964). Equivalently, h is convex, if and only if the numerator of the left hand side of (A1) is nonnegative for all $x < y < z$.

Let $h = G^{-1} \circ F$, where F and G are continuous and strictly increasing on $\{x : 0 < F(x) < 1\}$ and $\{x : 0 < G(x) < 1\}$, respectively. Putting $F(x) = u, F(y) = v$ and $F(z) = w$ shows that

$$G^{-1}(u)(F^{-1}(w) - F^{-1}(v)) + G^{-1}(v)(F^{-1}(u) - F^{-1}(w)) \\ + G^{-1}(w)(F^{-1}(v) - F^{-1}(u)) \geq 0, \tag{A2}$$

for all $u < v < w$ is equivalent to the convexity of $G^{-1} \circ F$. Now, a direct computation yields that (2) is equivalent to (A2). ∎

Proof of Theorem 1. Let $\pi_X(t) = E(X - t)_+$ and $\pi_Y(t) = E(Y - t)_+$ denote the stop-loss transforms of X and Y. Then,

$$\lim_{t \to \infty} \pi_X(t) = 0 \quad \text{and} \quad \lim_{t \to -\infty} (\pi_X(t) + t) = EX,$$

and analogously for Y (see, e.g., Müller & Stoyan, 2002, theorem 1.5.10). Hence,

$$\lim_{t \to \infty} (\pi_Y(t) - \pi_X(t)) = EY - EX. \tag{A3}$$

Now, assume $X \leq_{st} Y$. This holds if and only if $t \to \pi_Y(t) - \pi_X(t)$ is nonincreasing (Müller & Stoyan, 2002, theorem 1.5.13). In particular,

$$\pi_X(t) \leq \pi_Y(t) \quad \text{for all real } t. \tag{A4}$$

Bellini et al. (2018) proved that $X \leq_e Y$ is equivalent to

$$\pi_X(t)(t - EY) \leq \pi_Y(t)(t - EX), \quad \text{for all real } t. \tag{A5}$$

Since $X \leq_{st} Y$, we have $EX \leq EY$, and (A5) is obviously satisfied for $EX \leq t \leq EY$. For $t > EY$, $t - EX \geq t - EY > 0$, and (A5) holds by (A4). Next, consider the case $t < EX$. Then, (A5) is equivalent to

$$\pi_X(t)(EY - t) \geq \pi_Y(t)(EX - t), \quad \text{for all real } t,$$

or

$$\pi_X(t)(EY - EX) \geq (\pi_Y(t) - \pi_X(t))(EX - t), \quad \text{for all real } t. \tag{A6}$$

Since $t \to \pi_Y(t) - \pi_X(t)$ is nonincreasing, $\pi_Y(t) - \pi_X(t) \leq EY - EX$ by (A3). On the other hand, Jensen's inequality implies

$$(EX - t) \leq (EX - t)_+ \leq E(X - t)_+ = \pi_X(t),$$

and (A6) follows. Hence, (A5) is valid for all real t, which completes the proof. ∎

Proof of Proposition 2. Let X be a random variable with cdf F and finite mean μ. The representation

$$E(X - t)_+ = \int_t^\infty (1 - F(z))dz,$$

shows that $t \mapsto E(X - t)_+$ is strictly decreasing on $\{t \in \mathbb{R} : F(t) < 1\}$. Similarly, $t \mapsto E(X - t)_-$ is strictly increasing on $\{t \in \mathbb{R} : F(t) > 0\}$. The mean $\mu = e_X(1/2)$ always lies within the closures of both of these sets. Since, by Proposition 1c), $e_X(\alpha) < \mu < e_X(1 - \alpha)$, we obtain

$$E(X - e_X(1 - \alpha))_+ < E(X - e_X(\alpha))_+, \tag{A7}$$

$$E(X - e_X(\alpha))_- < E(X - e_X(1 - \alpha))_-. \tag{A8}$$

Furthermore, we can rewrite the first-order condition for expectiles (5) for any $\tau \in (0, 1) \setminus \{1/2\}$ in the following two ways

$$E(X - e_X(\tau))_+ = \frac{1 - \tau}{1 - 2\tau}(\mu - e_X(\tau)), \tag{A9}$$

$$E(X - e_X(\tau))_- = \frac{\tau}{1 - 2\tau}(\mu - e_X(\tau)). \tag{A10}$$

Plugging equations (A9) and (A10) into inequalities (A7) and (A8) yields

$$\alpha(e_X(1 - \alpha) - \mu) < (1 - \alpha)(\mu - e_X(\alpha)),$$
$$\alpha(\mu - e_X(\alpha)) < (1 - \alpha)(e_X(1 - \alpha) - \mu).$$

Transforming the first inequality yields the upper bound for $\tilde{s}_2(\alpha)$, by transforming the second one we obtain the lower bound.

Let now $X \sim \text{Bin}(1, p)$ for some $p \in (0, 1)$. Some calculations yield

$$\tilde{s}_2(\alpha) = (2\alpha - 1)(2p - 1).$$

Hence, $\tilde{s}_2(\alpha) \to 1 - 2\alpha$ for $p \to 0$ as well as $\tilde{s}_2(\alpha) \to 2\alpha - 1$ for $p \to 1$. Hence, both inequalities are sharp. ∎

Proof of Theorem 3. From

$$EX = \int_0^\infty (1 - F(x)) \, dx - \int_{-\infty}^0 F(x) \, dx,$$

and

$$E|X| = \int_0^\infty (1 - F(x) - F(-x)) \, dx,$$

we obtain

$$EX - EY = \int_{-\infty}^\infty (G - F)(x) \, dx, \tag{A11}$$

$$E|X| - E|Y| = \int_0^\infty (G - F)(x) \, dx - \int_{-\infty}^0 (G - F)(x) \, dx. \tag{A12}$$

Define $\tilde{X} = (X - \mu_F)/\delta_F$ and $\tilde{Y} = (Y - \mu_G)/\delta_G$ with pertaining cdf's $\tilde{F}(\cdot) = F(\delta_F \cdot + \mu_F)$ and $\tilde{G}(\cdot) = G(\delta_G \cdot + \mu_G)$. Set $M_{\tilde{G},\tilde{F}} = \tilde{G}(x) - \tilde{F}(x)$. Since $\mu_{\tilde{F}} = \mu_{\tilde{G}} = 0$ and $\delta_{\tilde{F}} = \delta_{\tilde{G}} = 1$, (A11) and (A12) take the form

$$\int_{-\infty}^\infty M_{\tilde{G},\tilde{F}}(x) \, dx = 0, \tag{A13}$$

$$\int_0^\infty M_{\tilde{G},\tilde{F}}(x) \, dx = \int_{-\infty}^0 M_{\tilde{G},\tilde{F}}(x) \, dx. \tag{A14}$$

(a) Assume $F \leq_2 G$. Hence, $\tilde{F}$ and $\tilde{G}$ cross each other at most twice. Since $\mu_{\tilde{F}} = \mu_{\tilde{G}}$, it follows from (A12) (and is well-known) that $\tilde{F}$ and $\tilde{G}$ are either identical or cross each other at least once. Now, assume that $M_{\tilde{G},\tilde{F}} \not\equiv 0$. Then, $\tilde{G}^{-1}\tilde{F}$ is strictly convex, and Jensen's inequality implies $\tilde{G}^{-1}\tilde{F}(0) < 0$, resulting in $\tilde{F}(0) < \tilde{G}(0)$ (see van Zwet, 1964, p. 10).
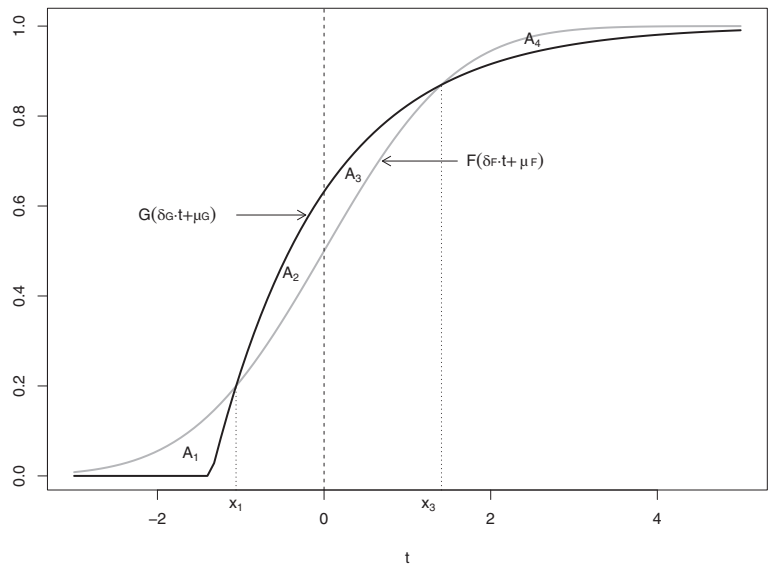
Assume that $M_{\tilde{G},\tilde{F}}$ has exactly one root x_1, where $x_1 \leq 0$. Put $x_0 = -\infty, x_2 = 0, x_3 = \infty$, and

$$A_i = \int_{x_{i-1}}^{x_i} M_{\tilde{G},\tilde{F}}(x) \, dx, \quad i = 1, 2, 3.$$

From (A13) and (A14), we obtain

$$A_1 + A_2 + A_3 = 0, \qquad A_1 + A_2 = A_3.$$

FIGURE A1
Standardized distribution
functions with $F \leq_\mu^\delta G$

Hence, $A_3 = 0$, which implies that $\tilde{F}(x) = \tilde{G}(x)$ for $x \geq 0$, a contradiction to $\tilde{F}(0) < \tilde{G}(0)$. Since an analogous reasoning excludes a single root $x_1 > 0$, it follows that $\tilde{F}$ and $\tilde{G}$ cross each other exactly twice, with $M_{\tilde{G},\tilde{F}}$ changing sign from negative to positive to negative, and $M_{\tilde{G},\tilde{F}}(0) > 0$. The second implication follows from the definitions.

(b) Assume $F \leq_\mu^\delta G$. Denote the two roots of $M_{\tilde{G},\tilde{F}}$ by x_1 and x_3, where $x_1 < 0 < x_3$. Further, put $x_0 = -\infty, x_2 = 0, x_4 = \infty$, and

$$A_i = \int_{x_{i-1}}^{x_i} M_{\tilde{G},\tilde{F}}(x) \, dx, \quad i = 1, \dots, 4.$$

From (A13) and (A14), we obtain

$$A_1 + A_2 + A_3 + A_4 = 0, \qquad A_1 + A_2 = A_3 + A_4$$

(see Figure A1). Hence, $-A_1 = A_2$, and $-A_4 = A_3$, that is, the area $|A_1|$ equals A_2, and $|A_4|$ equals A_3. Consequently,

$$\tilde{S}_Y(t) - \tilde{S}_X(t) = \frac{1}{t} \int_{-t}^{t} M_{\tilde{G},\tilde{F}}(t) \, dt \geq 0.$$

∎

Proof of Theorem 5. First, we obtain

$$\sqrt{n} \, (S_n(t) - S_X(t))$$

$$= \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ (X_i - \overline{X} - t)_+ - (X_i - \overline{X} + t)_+ - (\pi_X(\mu + t) - \pi_X(\mu - t)) \right\}$$

$$= \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ (X_i - \mu - t)_+ - (X_i - \mu + t)_+ - (\pi_X(\mu + t) - \pi_X(\mu - t)) \right\}$$

$$+ \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ (X_i - \overline{X} - t)_+ - (X_i - \mu - t)_+ \right\}$$

$$- \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ (X_i - \overline{X} + t)_+ - (X_i - \mu + t)_+ \right\}$$

$$= T_{n1} + T_{n2} - T_{n3}, \tag{A15}$$

say. Here, T_{n2} can be written as

$$T_{n2} = \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ (\mu - \overline{X})\, \mathbb{1}\{X_i \geq \max(\overline{X} + t, \mu + t)\} \right\}$$

$$+ \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ (X_i - \overline{X} - t)\, \mathbb{1}\{\overline{X} + t \leq X_i < \mu + t\} \right\}$$

$$- \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ (X_i - \mu - t)\, \mathbb{1}\{\mu + t \leq X_i < \overline{X} + t\} \right\}$$

$$= T_{n2}^{(1)} + T_{n2}^{(2)} - T_{n2}^{(3)},$$

where $\mathbb{1}$ denotes the indicator function. Now, we consider $T_{n2}^{(2)}$. If $\overline{X} \leq X_i - t \leq \mu$, then $0 \leq X_i - t - \overline{X} \leq \mu - \overline{X}$, and, hence,

$$\left| T_{n2}^{(2)} \right| \leq \sqrt{n} |\overline{X} - \mu| \frac{1}{tn} \sum_{i=1}^{n} \mathbb{1}\{\overline{X} \leq X_i - t < \mu\}$$

$$= O_p(1)o_p(1) = o_p(1).$$

Likewise, $T_{n2}^{(3)} = o_p(1)$. Similarly,

$$T_{n2}^{(1)} - \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ (\mu - \overline{X})\, \mathbb{1}\{X_i \geq \mu + t\} \right\}$$

$$= \sqrt{n}(\mu - \overline{X}) \frac{1}{tn} \sum_{i=1}^{n} \mathbb{1}\{\overline{X} \leq X_i - t \leq \mu\} = o_p(1).$$

Hence,

$$T_{n2}^{(1)} = \sqrt{n}(\mu - \overline{X}) \frac{1}{tn} \sum_{i=1}^{n} \mathbb{1}\{X_i \geq \mu + t)\} + o_p(1)$$

$$= \sqrt{n}(\mu - \overline{X})\, P(X_1 \geq \mu + t)/t + o_p(1).$$

Using $T_{n2} = T_{n2}^{(1)} + o_p(1)$, we get

$$T_{n2} = \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \left\{ -(X_i - \mu)\,(1 - F(\mu + t)) \right\} + o_p(1). \tag{A16}$$

Completely analogous considerations yield

$$T_{n3} = \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \{-(X_i - \mu)(1 - F(\mu - t))\} + o_p(1). \tag{A17}$$

Plugging (A18) and (A19) in (A17), we finally obtain

$$\sqrt{n}(S_n(t) - S_X(t))$$

$$= \frac{1}{t\sqrt{n}} \sum_{i=1}^{n} \{(X_i - \mu - t)_+ - (X_i - \mu + t)_+ + (X_i - \mu)(F(\mu + t) - (F(\mu - t))$$

$$- (\pi_X(\mu + t) - \pi_X(\mu - t))\} + o_p(1),$$

and the central limit theorem yields the assertion. ∎