ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 29, Number 1, June 2025
Available online at https://ojs.utlib.ee/index.php/ACUTM

Lambert W normal distribution: a viable extension for skew-normal?

Meelis Käärik and Anne Selart

Abstract. Several distributions, and their families, such as skew-normal and related distributions, are proposed to model skewed data. Lambert W transformation offers a flexible approach to introduce skewness to any random variable. In this paper, we compare the Lambert W normal distribution with the widely known skew-normal distribution and point out the similarities and differences, focusing mainly on risk-related characteristics and data fitting. While mathematically a different construction, one can think of Lambert W normal approach as an extension of the skew-normal to a more skewed situation.

1. Introduction

In statistical analysis, there are often cases where the data is almost normally distributed but with a light skewness. Such situations arise very often in financial and actuarial mathematics, for example, in the case of asset yield or insurance loss data. One widely used choice to model such data is the skew-normal distribution, where the shape of the normal distribution is deformed by a certain skewness parameter (Azzalini [1]). Analogously, other asymmetric distributions, like skew t-distribution by Azzalini and Capitanio [4], have been developed. Lambert W random variables can be seen as a generalization because the input distribution can be arbitrary (Goerg [12]). Still, because of the wide use of skew-normal distribution, we are particularly interested in Lambert W normal approach and its possible advantages over the skew-normal.

The paper is organized as follows. We first briefly recall the construction of Lambert W random variables and skew-normal distribution. Next, we compare some risk-related characteristics, such as skewness, kurtosis,

Received November 22, 2024.
2020 *Mathematics Subject Classification.* 60E05, 62P05.
Key words and phrases. Probability distributions, skewness, loss distributions, model fitting, actuarial science, risk measures.
https://doi.org/10.12697/ACUTM.2025.29.04

quantiles (value at risk) and expected shortfall of skew-normal and Lambert W normal distributions. In the third section, we address the question of fitting the Lambert W normal and skew-normal distributions to data, focusing especially on the similarities of the approaches in the range of skewness available for the skew-normal distribution. An important practical question of which method to use for parameter estimation is also discussed in this section. The findings of the paper are concluded in the final section.

2. Definitions

2.1. Lambert W random variables. We start with defining the Lambert W function and then give a brief overview of its properties. We refer to [6, 9, 7] for more details on the topic.

Lambert W function is a set of inverse functions for the following function: $f(x') = x'e^{x'}$ ($x' \in \mathbb{R}$). In other words,

$$x' = f^{-1}(x'e^{x'}) = W(x'e^{x'}).$$

Substituting $x = x'e^{x'}$ leads us to the definition of the Lambert W function.

Definition 1. Lambert W function $W(x)$ is defined by the following equality:

$$W(x)e^{W(x)} = x, \quad x \in \left[-\frac{1}{e}, \infty\right). \tag{1}$$

Equation (1) has infinitely many solutions, most of which are complex. Following the notation from [7], we denote the different branches of the function by $W_k(x)$, where k is the branch index, $k \in \{0, \pm 1, \pm 2, \dots\}$, and $x \in \mathbb{C}$. For real x, all branches besides $W_0(x)$ and $W_{-1}(x)$ are complex. We denote the branch corresponding to $W(x) \geq -1$ by $W_0(x)$ and call it the principal branch, and the branch corresponding to $W(x) \leq -1$ by $W_{-1}(x)$ and call it the non-principal branch.

Based on [12], Lambert W random variables are defined for three different cases: the plain (noncentral, nonscaled) version, the scaled version and the location-scale version. We will introduce only the most general version, the location-scale Lambert W random variables. For a more thorough discussion, including the forms of cumulative distribution function (cdf) and probability density function (pdf), we refer to [12] and [14].

Definition 2. Let X be a continuous random variable from a location-scale family with cdf $F_X(x|\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is the corresponding row vector of parameters. Denote the mean and standard deviation of X as μ and σ, then for standardized variable $U = \dfrac{X-\mu}{\sigma}$, E$U = 0$, Var$U = 1$ and

$$Y := \{U \exp(\gamma U)\}\sigma + \mu, \quad \gamma \in \mathbb{R}, \ \ \sigma_X > 0, \tag{2}$$

is location-scale Lambert random variable with parameters $\boldsymbol{\beta}$ and γ.

If $\gamma > 0$, the location-scale Lambert random variable Y takes values in the interval $(\mu - \frac{\sigma}{\gamma e}, \infty)$. If $\gamma < 0$, Y has an upper bound and takes values in the interval $(-\infty, \mu - \frac{\sigma}{\gamma e})$.

2.2. Lambert W normal distribution. In the following, we apply the location-scale transformation (2) to a normal random variable $X \sim N(\mu, \sigma)$. The resulting random variable

$$Y = \frac{X-\mu}{\sigma} \exp\left(\gamma \frac{X-\mu}{\sigma}\right)\sigma + \mu$$

has Lambert W normal distribution with parameter vector $(\mu, \sigma, \gamma)^T$ and we denote it as $Y \sim WN(\mu, \sigma, \gamma)$. If $\mu = 0$ and $\sigma = 1$, the distribution reduces to the standard version of the Lambert W normal distribution, $Y \sim WN(0, 1, \gamma)$.

Without the loss of generality, we assume that the skewness parameter γ is positive; the situation is mirrored for negative γ (left skew instead of right skew). Now, if $\gamma > 0$, the cdf for the $WN(\mu, \sigma, \gamma)$ distribution has the following form:

$$F_Y(y|\mu,\sigma,\gamma) = \begin{cases} 0, & \text{if } y \leq \mu - \frac{\sigma}{\gamma e}, \\ \Phi\left(\frac{W_0(\gamma z)}{\gamma}\right) - \Phi\left(\frac{W_{-1}(\gamma z)}{\gamma}\right), & \text{if } \mu - \frac{\sigma}{\gamma e} < y < \mu, \\ \Phi\left(\frac{W_0(\gamma z)}{\gamma}\right), & \text{if } y \geq \mu, \end{cases}$$

where $z = \frac{y-\mu}{\sigma}$ and $\Phi(\cdot)$ is the cdf of a standard normal random variable. Similarly, the pdf for $\gamma > 0$ can be written as

$$f_Y(y|\mu,\sigma,\gamma) = \begin{cases} 0, & \text{if } y \leq \mu - \frac{\sigma}{\gamma e}, \\ \frac{1}{\sigma} f_0\left(\frac{y-\mu}{\sigma}\right) - \frac{1}{\sigma} f_{-1}\left(\frac{y-\mu}{\sigma}\right), & \text{if } \mu - \frac{\sigma}{\gamma e} < y < \mu, \\ \frac{1}{\sigma} f_0\left(\frac{y-\mu}{\sigma}\right), & \text{if } y \geq \mu, \end{cases}$$

where $f_0(z)$ and $f_{-1}(z)$ denote the components of the pdf corresponding to the principal and non-principal branch, respectively:

$$f_0(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(W_0(\gamma z))^2}{2\gamma^2}\right) \frac{\exp(-W_0(\gamma z))}{1 + W_0(\gamma z)},$$

$$f_{-1}(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(W_{-1}(\gamma z))^2}{2\gamma^2}\right) \frac{\exp(-W_{-1}(\gamma z))}{1 + W_{-1}(\gamma z)}.$$

Depending on the value of skewness parameter γ, we can distinguish three main shapes of the pdf of Lambert W normal random variables (see [14]). First, if $\gamma \in (0, \sqrt{2}-1)$, the pdf has two local extrema (the left panel of Figure 1). Second, if $\gamma \in [\sqrt{2}-1, \sqrt{2}+1]$, the pdf is a strictly decreasing function of y (the middle panel). Third, for $\gamma > \sqrt{2}+1$, the pdf again has two local extrema, but compared to the first case, the overall shape of the pdf is different, as can be seen in the right panel.

In the comparison between Lambert W normal and skew-normal distributions, where we limit the skewness to the available range for the skew-normal, Lambert W normal distribution has the first density shape.

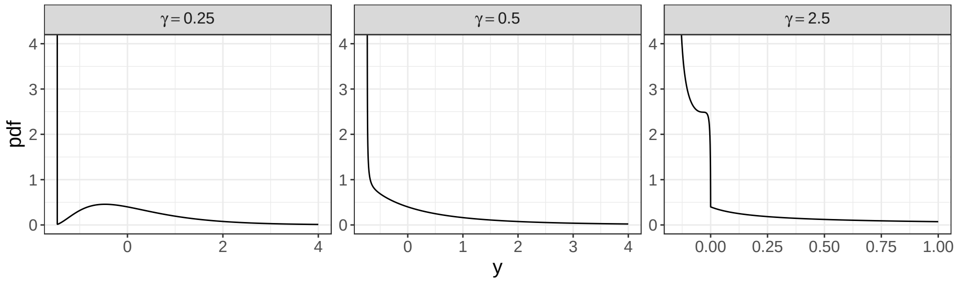

FIGURE 1. Three main shapes of Lambert W normal densities for $WN(0, 1, \gamma)$.

2.3. Skew-normal distribution. Skew-normal distribution is a widespread distribution in the analysis of asymmetric data. It has been studied thoroughly in recent decades and applied in various fields where skewed data is present. In the following, we recall the definition of the skew-normal distribution. More details can be found, e.g., in [1, 2]. Similarly to Lambert W Section 2.2, we only provide the location-scale version of the skew-normal distribution (see also [2]).

Definition 3. If the pdf of a random variable X has the form

$$f(x|\xi,\omega,\alpha) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right), \quad x \in \mathbb{R}, \omega > 0,$$

where ϕ and Φ are the standard normal pdf and cdf, respectively, then we say that X is a location-scale skew-normal random variable with location parameter ξ, scale parameter $\omega > 0$ and skewness parameter α, and write $X \sim SN(\xi, \omega, \alpha)$.

The skewness parameter α can take values from $-\infty$ to ∞, where positive values imply right-skewed distribution and negative values imply left-skewed distribution. If $\alpha = 0$, the distribution reduces to the normal distribution.

If $\xi = 0$ and $\omega = 1$, the distribution reduces to the standard version of the skew-normal distribution, $X \sim SN(0, 1, \alpha)$.

3. Comparison of Lambert W normal and skew-normal distribution: risk-related characteristics

Next, we compare the two distributions of interest based on some properties associated with tail heaviness, which are essential for assessing risk in financial and insurance contexts.

3.1. Skewness coefficient. In the following, we denote the skewness coefficient, i.e. the standardized third moment of a distribution, by γ_1. To distinguish the coefficient for Lambert W normal and skew-normal distributions, we use the corresponding skewness parameter as an argument, so that $\gamma_1(\gamma)$ marks the skewness coefficient for $WN(\mu, \sigma, \gamma)$ and $\gamma_1(\alpha)$ for $SN(\xi, \omega, \alpha)$. It is not necessary to specify the location and scale parameters as the skewness coefficient is location-scale invariant by definition.

For Lambert W normal distribution $WN(\mu, \sigma, \gamma)$, the skewness coefficient $\gamma_1(\gamma)$ has the following form (see Appendix A for a proof):

$$\gamma_1(\gamma) = \frac{\gamma(e^{3\gamma^2}(9+27\gamma^2) - e^{\gamma^2}(3+12\gamma^2) + 2\gamma^2)}{(e^{\gamma^2}(1+4\gamma^2) - \gamma^2)^{\frac{3}{2}}}. \tag{3}$$

Notice that the expression (3) has a minor difference in the denominator compared to the skewness coefficient given in [12].

This skewness coefficient (3) is a monotone function of γ, with the same sign. The value of the function is not bounded, as $\gamma_1(\gamma) \to \pm\infty$ if $\gamma \to \pm\infty$.

For skew-normal distribution $SN(\xi, \omega, \alpha)$, the skewness coefficient $\gamma_1(\alpha)$ has the following form:

$$\gamma_1(\alpha) = \frac{4-\pi}{2} \operatorname{sgn}(\alpha) \left(\frac{\frac{2\delta^2}{\pi}}{1 - \frac{2\delta^2}{\pi}} \right)^{\frac{3}{2}}, \tag{4}$$

with $\operatorname{sgn}(\cdot)$ being the sign function and $\delta = \dfrac{\alpha}{\sqrt{1+\alpha^2}}$.

It has been shown by Azzalini [1] that, for the skewness coefficient of the skew-normal distribution, the following holds:

$$\lim_{\alpha \to \pm\infty} \gamma_1(\alpha) = \pm \frac{4-\pi}{2} \frac{2^{\frac{3}{2}}}{(\pi-2)^{\frac{3}{2}}} \approx \pm 0.9953.$$

In summary, the range of skewness of skew-normal distributions is quite limited, and the Lambert W normal distribution covers a much wider range.

In the following, it makes sense to compare the behaviour of Lambert W normal and skew-normal distributions with both having the skewness coefficient in this range $(-0.9953, 0.9953)$ as both distributions are defined there. For the Lambert W normal distribution, this range is achieved when the skewness parameter γ ranges from -0.1623 to 0.1623 approximately.

3.2. Kurtosis coefficient. For the kurtosis coefficient, we use the notation γ_2. Similarly to the previous subsection, $\gamma_2(\gamma)$ stands for the kurtosis coefficient of Lambert W normal and $\gamma_2(\alpha)$ for skew-normal distribution.

It can be shown, that the kurtosis coefficient of $WN(\mu, \sigma, \gamma)$ has the following form (see Appendix A for a proof):

$$\gamma_2(\gamma) = \frac{e^{6\gamma^2}(256\gamma^4 + 96\gamma^2 + 3) - 36e^{3\gamma^2}\gamma^2(3\gamma^2 + 1) + 6e^{\gamma^2}\gamma^2(4\gamma^2 + 1) - 3\gamma^4}{(e^{\gamma^2}(4\gamma^2 + 1) - \gamma^2)^2}. \tag{5}$$

Again, compared to the result given in [12], our formula (5) has a different denominator, but the resulting numeric values are very similar, especially for small values of γ.

The minimal value of the kurtosis coefficient (5) is 3 when $\gamma = 0$, i.e. when the distribution corresponds to the normal one. As $\gamma \to \pm\infty$, the value of the kurtosis coefficient grows rapidly and has no upper bound.

On the contrary, the kurtosis coefficient of $SN(\xi, \omega, \alpha)$, expressed as (see [1])

$$\gamma_2(\alpha) = 3 + \frac{2(\pi - 3)\left(\delta\sqrt{\frac{2}{\pi}}\right)^4}{\left(1 - \delta^2\frac{2}{\pi}\right)^2},$$

has an upper bound of $3 + \frac{8(\pi-3)}{(\pi-2)^2} \approx 3.8692$.

TABLE 1. Values of kurtosis coefficient γ_2 for skew-normal and Lambert W normal distributions with the same skewness coefficient γ_1.

γ_1	SN: γ_2	WN: γ_2
0.000	3.000	3.000
0.100	**3.041**	3.017
0.200	**3.102**	3.067
0.300	**3.176**	3.150
0.400	3.258	**3.266**
0.500	3.347	**3.416**
0.600	3.443	**3.599**
0.700	3.544	**3.814**
0.800	3.650	**4.063**
0.900	3.760	**4.345**
0.950	3.817	**4.499**
0.990	3.863	**4.628**
0.995	3.869	**4.644**

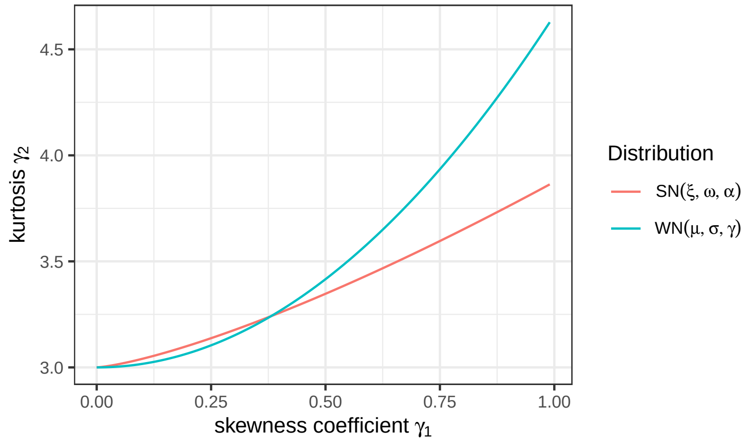

FIGURE 2. Skewness-kurtosis relationship for skew-normal and Lambert W normal distributions with the same skewness coefficient γ_1.

Since for both distributions, the skewness coefficient and the kurtosis coefficient depend only on the skewness parameter, i.e., on γ in the case of Lambert W normal and on α in the case of skew-normal, we can compare the kurtosis coefficients of the distributions having the same value of skewness coefficient γ_1 (see Table 1 and Figure 2). Starting from the skewness coefficient $\gamma_1 = 0$ when both distributions reduce to a normal distribution, we see that the skew-normal kurtosis coefficient is bigger for smaller values of γ_1, up to $\gamma_1 \approx 0.3$, but after that, the Lambert W kurtosis coefficient grows more rapidly and reaches about 4.64 for $\gamma_1 = 0.995$ as the skew-normal kurtosis coefficient stops at around 3.87.

3.3. Quantiles. The p-quantile of a continuous random variable X, denoted by q_p, is the value where $\mathrm{P}(X \leq q_p) = p$, $0 < p < 1$, or, equivalently, $\mathrm{P}(X > q_p) = 1 - p$.

Comparison of the p-quantiles of the distributions provides info about the extremeness of the values corresponding to the same probability level. Also, one of the most commonly used risk measures, value at risk, is defined as a certain quantile of the corresponding risk variable.

In Table 2, the 0.95- and 0.99-quantiles for the Lambert W normal $WN(0, 1, \gamma(\gamma_1))$ and the skew-normal distribution $SN(0, 1, \alpha(\gamma_1))$ for different values of the same skewness coefficient γ_1 are found. As we can see, the quantiles for the Lambert W normal distribution are greater than those of the skew-normal for higher values of the skewness coefficient, implying a heavier right tail for the Lambert W normal distribution in this case.

TABLE 2. Quantiles $q_{0.95}$ and $q_{0.99}$ for Lambert W normal $WN(0,1,\gamma)$ and skew-normal $SN(0,1,\alpha)$ distributions with the same skewness coefficient γ_1.

γ_1	SN: $q_{0.95}$	WN: $q_{0.95}$	SN: $q_{0.99}$	WN: $q_{0.99}$
0.000	1.645	1.645	2.326	2.326
0.100	**1.949**	1.691	**2.573**	2.418
0.200	**1.958**	1.737	**2.576**	2.514
0.300	**1.960**	1.785	2.576	**2.612**
0.400	**1.960**	1.834	2.576	**2.714**
0.500	**1.960**	1.884	2.576	**2.819**
0.600	**1.960**	1.935	2.576	**2.927**
0.700	1.960	**1.986**	2.576	**3.038**
0.800	1.960	**2.039**	2.576	**3.151**
0.900	1.960	**2.091**	2.576	**3.267**
0.950	1.960	**2.118**	2.576	**3.326**
0.990	1.960	**2.139**	2.576	**3.374**
0.995	1.960	**2.142**	2.576	**3.380**

If we denote the p-quantile of the standard normal distribution as z_p, it can be shown that for $p \geq 0.5$ the p-quantile of $WN(0,1,\gamma)$ equals $q_p = z_p e^{\gamma z_p}$, as the p-quantile of Y must satisfy the following equation:

$$p = P(Y \leq q_p) = F_y(q_p; 0, 1, \gamma) = \Phi\left(\frac{W_0(\gamma q_p)}{\gamma}\right) = \Phi(z_p).$$

So, the median of Lambert W normal distribution is always the same as that of the underlying normal distribution, but with $p > 0.5$ and, as γ grows, the Lambert W normal quantile grows exponentially. For right skewed Lambert W normal distribution, no explicit formula can be found for the quantile when $p < 0.5$, as the cdf includes the difference of standard normal cdfs.

For skew-normal quantiles, no explicit formula can be given in relation to normal quantiles, but for positive skewness, skew-normal quantiles are higher than those of underlying normal distribution. When the skewness coefficient γ_1 approaches 0.9953, the skew-normal p-quantile converges to the p-quantile of the half-normal distribution that in turn equals to the $\frac{1+p}{2}$-quantile of the standard normal, $z_{\frac{1+p}{2}}$. As we can see from examples in Table 2, the behaviour for skew-normal quantiles is different from the Lambert W quantiles. The skew-normal p-quantile reaches $z_{\frac{1+p}{2}}$ very quickly, so the right tail of the skew-normal distribution changes very little when γ_1 rises.

The dynamics between p-quantiles of standard skew-normal $SN(0,1,\alpha(\gamma_1))$ and Lambert W normal $WN(0,1,\gamma(\gamma_1))$ distributions is depicted in Figure 3.

For smaller values of p and γ_1, the p-quantile of skew-normal exceeds the p-quantile of Lambert W normal. With the increase of p and γ_1, the quantiles of both distributions also increase, but the growth is faster for the Lambert W normal distribution. Figure 3 shows the borderline when the p-quantiles of skew-normal $SN(0, 1, \alpha(\gamma_1))$ and Lambert W normal $WN(0, 1, \gamma(\gamma_1))$ distributions coincide. Below the line, the skew-normal p-quantiles exceed those of the corresponding Lambert W normal, and above the line, the Lambert W normal p-quantiles exceed those of the corresponding skew-normal. The borderline is found numerically.

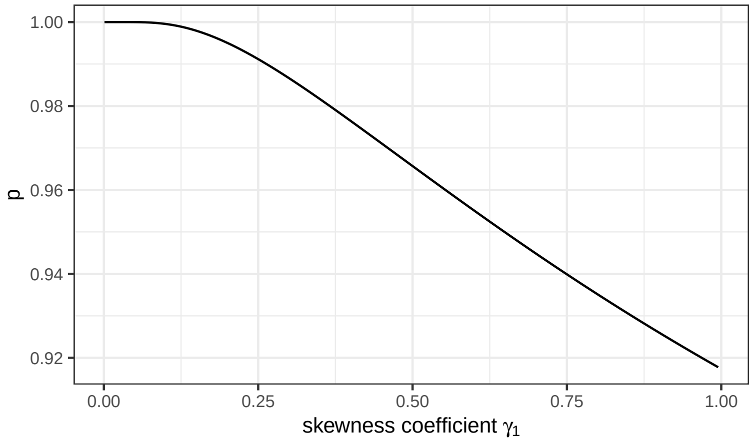

FIGURE 3. The borderline of (γ_1, p) pairs where the $WN(0, 1, \gamma(\gamma_1))$ and $SN(0, 1, \alpha(\gamma_1))$ p-quantiles are equal.

3.4. Expected shortfall. In this section, we recall the concept of expected shortfall and compare the formulas for the expected shortfall for skew-normal and Lambert W normal distributions. Expected shortfall, also called the tail value at risk, is a useful measure in loss models of financial and actuarial applications.

Expected shortfall for a random variable X, $X \sim F$, is defined as

$$\mathrm{ES}_p^F = \mathrm{E}(X|X > q_p),$$

where q_p is the p-quantile of X.

To compare the behaviour of expected shortfall for Lambert W normal and skew-normal distributions, it suffices to handle the expected shortfall for the standard versions $WN(0, 1, \gamma)$ and $SN(0, 1, \alpha)$, as both distributions are in the location-scale family and the expected shortfall generalizes in an obvious way.

For Lambert W standard normal distribution $WN(0,1,\gamma)$, the expected shortfall has the following form:

$$\mathrm{ES}_p^{WN} = \frac{1}{1-p} e^{\frac{\gamma^2}{2}} (\varphi(u) + \gamma(1-\Phi(u))), \tag{6}$$

where $u = \frac{W_0(\gamma q_p)}{\gamma} - \gamma$ and $\varphi(\cdot)$ stands for pdf of standard normal distribution. The proof is straightforward but technical, see Appendix B for details.

We can now estimate the limits of the components in (6) when $\gamma \to \infty$:

$$\lim_{\gamma\to\infty} e^{\frac{\gamma^2}{2}} \varphi(u) = \lim_{\gamma\to\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{W_0^2(\gamma q_p)}{\gamma^2} - W_0(\gamma q_p))} = \infty,$$

and $\gamma(1-\Phi(u)) \to \infty$ since $u \to -\infty$. This assures that the expected shortfall for Lambert normal distribution $WN(0,1,\gamma)$ is unbounded.

Following the ideas given in [5], where the formula for tail conditional expectation $E(X|X < x_p)$ for a skew-normal variable X is derived, we can conclude that the expected shortfall for $X \sim SN(0,1,\alpha)$ is

$$\mathrm{ES}_p^{SN} = \frac{1}{1-p}(f_{SN}(q_p;\alpha) + \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{1+\alpha^2}} (1 - \Phi(q_p\sqrt{1+\alpha^2}))),$$

where $f_{SN}(\cdot;\alpha)$ is the pdf for $SN(0,1,\alpha)$ and $\Phi(\cdot)$ is the standard normal cdf. As the skewness parameter α goes to ($\pm$) infinity, the pdf of skew-normal converges to the pdf of half-normal distribution:

$$\lim_{\alpha\to\pm\infty} f_{SN}(x;\alpha) = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}}, \quad x \cdot \mathrm{sgn}(\alpha) \geq 0. \tag{7}$$

At the same time, the p-quantile of skew-normal distribution approaches the respective half-normal quantile, which, in turn, is equal to the $\frac{1+p}{2}$-quantile of standard normal $z_{\frac{1+p}{2}}$. In addition, the ratio $\frac{\alpha}{\sqrt{1+\alpha^2}} \to \pm 1$ and $\Phi(q_p\sqrt{1+\alpha^2}) \to 1$. As a result, by (7), we have

$$\lim_{\alpha\to\pm\infty} \mathrm{ES}_p^{SN} = \frac{1}{1-p}\sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2} z^2_{\frac{1+p}{2}}\right) = \frac{2}{1-p}\varphi(z_{\frac{1+p}{2}}).$$

This limit corresponds to the expected shortfall of the half-normal distribution.

The dynamics between expected shortfalls of standard skew-normal $SN(0,1,\alpha(\gamma_1))$ and Lambert W normal $WN(0,1,\gamma(\gamma_1))$ distributions is similar to the dynamics between the quantiles, see also Table 3. For smaller values of p and γ_1, the expected shortfall of skew-normal exceeds the expected shortfall of Lambert W normal, and with the increase of p and γ_1, the relation reverses. The corresponding borderline where the expected shortfalls are equal is shown in Figure 4. Below the line, the skew-normal expected shortfall exceeds that of the corresponding Lambert W normal, and above the line, vice versa.

TABLE 3. Expected shortfalls $ES_{0.95}$ and $ES_{0.99}$ for Lambert W normal $WN(0,1,\gamma)$ and skew-normal $SN(0,1,\alpha)$ distributions with the same skewness coefficient γ_1.

γ_1	$ES^{SN}_{0.95}$	$ES^{WN}_{0.95}$	$ES^{SN}_{0.99}$	$ES^{WN}_{0.99}$
0.000	2.063	2.063	2.665	2.665
0.100	**2.332**	2.137	**2.891**	2.788
0.200	**2.337**	2.214	2.892	**2.916**
0.300	**2.338**	2.294	2.892	**3.050**
0.400	**2.338**	2.376	2.892	**3.189**
0.500	2.338	**2.461**	2.892	**3.333**
0.600	2.338	**2.548**	2.892	**3.482**
0.700	2.338	**2.636**	2.892	**3.636**
0.800	2.338	**2.727**	2.892	**3.795**
0.900	2.338	**2.820**	2.892	**3.959**
0.950	2.338	**2.867**	2.892	**4.043**
0.990	2.338	**2.906**	2.892	**4.110**
0.995	2.338	**2.910**	2.892	**4.119**

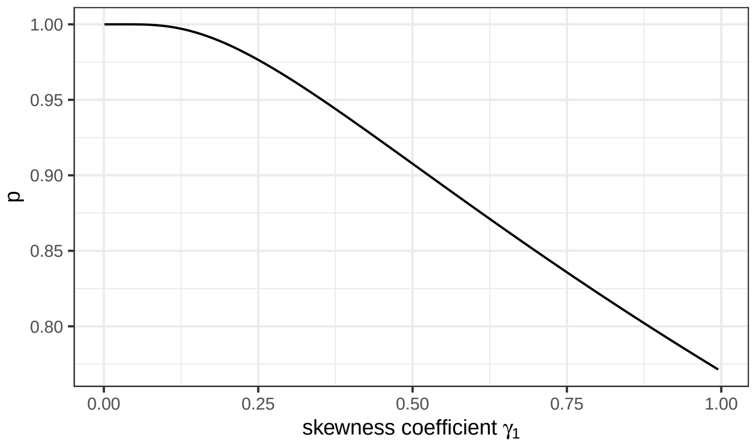

FIGURE 4. The borderline of (γ_1, p) pairs where the expected shortfalls of $WN(0,1,\gamma(\gamma_1))$ and $SN(0,1,\alpha(\gamma_1))$ are equal.

4. Fitting Lambert W normal and skew-normal distributions to data

In [12] the Lambert W normal and skew-normal distributions are fitted to simulated $Y \sim WN(\mu,\sigma,\gamma)$, where $\gamma \in \{0, -0.05, 0.3\}$, and estimates of $\mathrm{E}Y$, Var$Y$ and γ are studied based on bias and RMSE. The results show that for lower values of γ, both approaches give accurate and precise estimates for considered characteristics, but for heavily skewed data, the skew-normal framework fails.

We continue with the comparison of the Lambert W normal and skew-normal approaches by assessing their fit on simulated datasets. Goerg [12] compared the Lambert W and skew-normal model fit on Lambert W normal data; we look at the vice-versa case as well and fit the Lambert W and skew-normal model to skew-normal data. We illustrate the fit with parametric density estimates, goodness of fit tests and log-likelihood values.

The results for some of the most illustrative situations are given in this section; additional information is given in Appendix C.

As the key point for skew-normal and Lambert W normal distributions is to introduce skewness, we will take the skewness as the starting point for our simulation study. We first fix the value of the skewness coefficient γ_1 and generate samples of size 1000 from $WN(0,1,\gamma)$ and $SN(0,1,\alpha)$ with γ and α chosen to produce $\gamma_1(\gamma) = \gamma_1(\alpha) = \gamma_1$. For the skew-normal distribution, the expression (4) can be inverted to get α, but for the Lambert W normal distribution, the parameter γ value for given γ_1 must be found numerically.

For every sample, we fit the distributions $WN(\mu,\sigma,\gamma)$ and $SN(\xi,\omega,\alpha)$ using maximum likelihood estimation (MLE) and maximum spacing estimation (MSP, Ranneby [17]). The reason to use two estimation methods is the shape of Lambert W normal density, as the unbounded likelihood is known to produce issues with the MLE. When presenting the results, we use the MSP estimates for Lambert W normal distribution and the MLE estimates for skew-normal as these combinations (SN/MLE and WN/MSP) were the most stable. We found non-convergence issues with estimating the Lambert W parameters with the MLE and problems with biased estimates for heavily skewed skew-normal with the MSP.

The range of values considered for the skewness coefficient is chosen according to the skew-normal distribution (with positive skewness) with a grid $0.1, 0.2, \ldots, 0.9, 0.95, 0.99$. For each γ_1 value, the data generation and estimation steps were repeated as long as the aim of 1000 samples with successful estimates for both models and methods was reached. The extra samples were mostly needed in cases where the data was skew-normal with extreme skewness, i.e., with skewness coefficient γ_1 over 0.9 and the $WN(\mu,\sigma,\gamma)$ model was fitted using MLE.

In summary, in the simulations, we consider 4 scenarios for each γ_1:

- data from $WN(0,1,\gamma)$ and fitted $SN(\hat{\xi},\hat{\omega},\hat{\alpha})$,
- data from $WN(0,1,\gamma)$ and fitted $WN(\hat{\mu},\hat{\sigma},\hat{\gamma})$,
- data from $SN(0,1,\alpha)$ and fitted $SN(\hat{\xi},\hat{\omega},\hat{\alpha})$,
- data from $SN(0,1,\alpha)$ and fitted $WN(\hat{\mu},\hat{\sigma},\hat{\gamma})$.

We use R (see [16]) to conduct the simulation experiment. The packages *sn* [3] and *LambertW* [13] are used for data simulation and *fitdistrplus* [8] for estimation.

At the end of the section, we also provide a comparison of performance between Lambert W normal and skew-normal distributions on a real-life dataset.

4.1. Parameter estimation based on simulated data. First, we study the situation when data is simulated from $WN(0,1,\gamma)$ and compare the fit of two models, namely $WN(\mu,\sigma,\gamma)$ and $SN(\xi,\omega,\alpha)$.

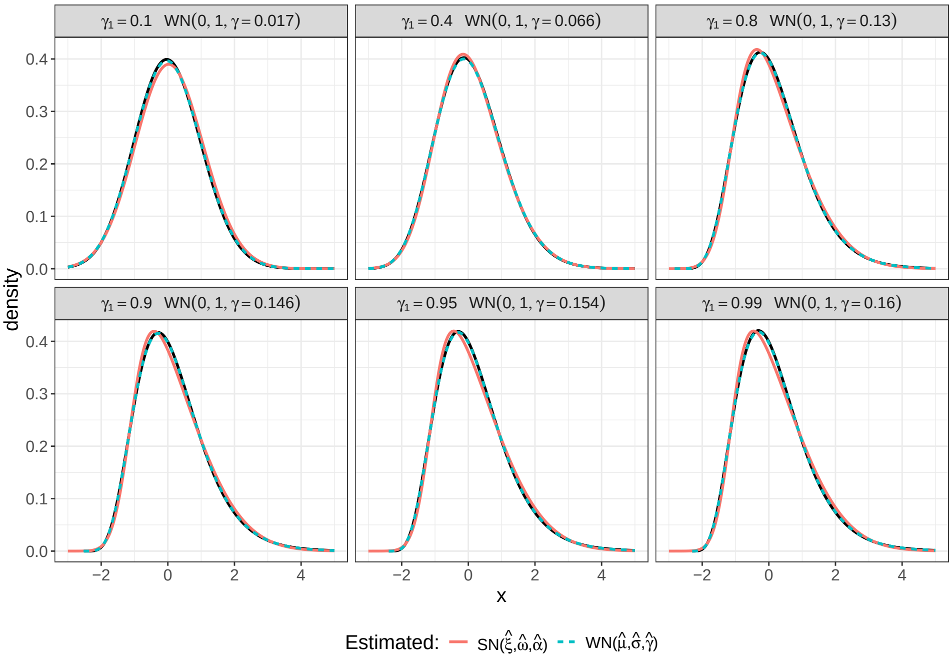

FIGURE 5. Examples of $WN(0,1,\gamma)$ density (black solid line) with fitted $SN(\hat{\xi},\hat{\omega},\hat{\alpha})$ and $WN(\hat{\mu},\hat{\sigma},\hat{\gamma})$ models, where parameters are averaged over 1000 sample estimates.

Figure 5 compares the density curves of theoretical $WN(0,1,\gamma)$ distribution (black solid line) and fitted distributions for selected values of skewness coefficient γ_1. The parameter estimates for fitted $WN(\mu,\sigma,\gamma)$ and $SN(\xi,\omega,\alpha)$ distributions are averaged over 1000 estimates. Both fitted densities follow the true densities very well in these examples, with only minor differences in the middle part of the density curve for skew-normal that become more pronounced when γ_1 grows.

The averaged parameter estimates over the whole grid of γ_1 values are given in Table 4. For Lambert W normal, the averages of parameter estimates are close to their true values. There are no true counterparts to skew-normal parameter estimates, but we can compare the fixed skewness coefficients γ_1 and calculated skewness coefficients $\gamma_1(\hat{\alpha})$ also given in Table 4. As one can see, these values are close for γ_1 values 0.3 and 0.4, but the estimates are lower for other skewness coefficient values, especially for higher γ_1.

TABLE 4. Data from $WN(0,1,\gamma)$. Averages of parameter estimates for fitted $SN(\hat{\xi},\hat{\omega},\hat{\alpha})$ and $WN(\hat{\mu},\hat{\sigma},\hat{\gamma})$ models. The last column corresponds to the parametric estimate of the skewness coefficient γ_1 for the skew-normal variable.

γ_1	γ	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\omega}$	$\hat{\alpha}$	$\gamma_1(\hat{\alpha})$
0.100	0.017	0.001	1.007	0.017	-0.501	1.171	0.756	0.071
0.200	0.033	0.001	1.005	0.032	-0.716	1.256	1.165	0.189
0.300	0.050	0.000	1.004	0.050	-0.837	1.341	1.506	0.302
0.400	0.066	0.001	1.006	0.066	-0.916	1.411	1.818	0.401
0.500	0.083	0.001	1.005	0.082	-0.971	1.464	2.115	0.485
0.600	0.099	0.002	1.004	0.098	-1.019	1.516	2.473	0.570
0.700	0.115	0.001	1.004	0.114	-1.061	1.565	2.870	0.646
0.800	0.130	0.000	1.003	0.130	-1.092	1.605	3.295	0.709
0.900	0.146	0.001	1.004	0.146	-1.117	1.645	3.782	0.764
0.950	0.154	0.002	1.004	0.154	-1.127	1.665	4.069	0.790
0.990	0.160	0.002	1.005	0.160	-1.136	1.680	4.302	0.808

Next, we look at the case when data is generated from a skew-normal distribution, and again we examine the fit of two competing models $SN(\xi,\omega,\alpha)$ and $WN(\mu,\sigma,\gamma)$.

Figure 6 compares the density curves of theoretical $SN(0,1,\alpha)$ distribution (black solid line) and fitted distributions for selected values of skewness coefficient γ_1. The parameter estimates for fitted $WN(\hat{\mu},\hat{\sigma},\hat{\gamma})$ and $SN(\hat{\xi},\hat{\omega},\hat{\alpha})$ distributions are averaged over 1000 estimates. As one can see, the fit of the skew-normal model is very good, as is the fit of Lambert W normal model for smaller γ_1 values. For higher skewness coefficients, the discrepancy between $SN(0,1,\alpha)$ and Lambert W normal is more noticeable.

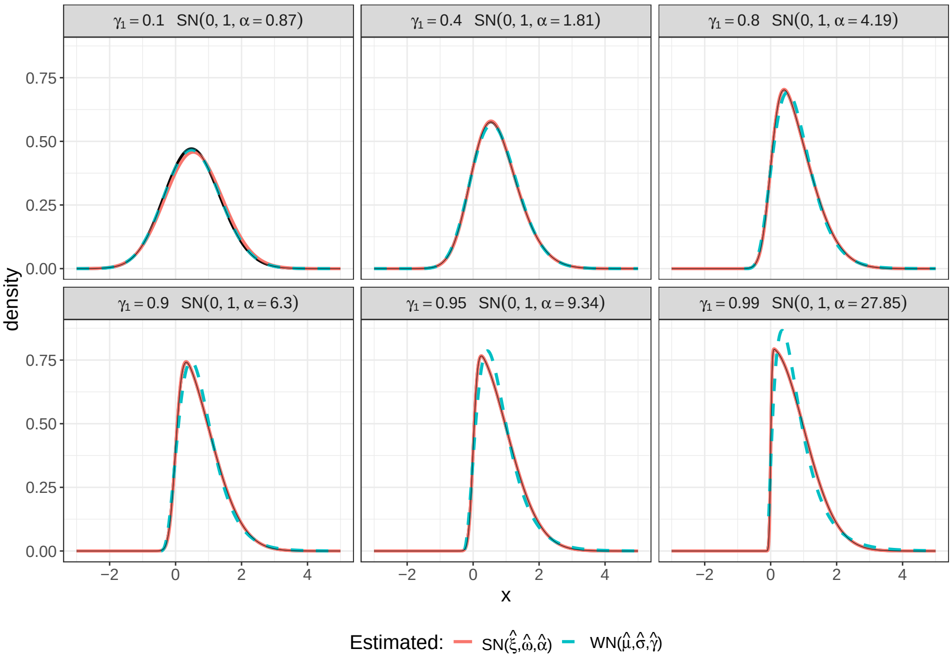

FIGURE 6. Examples of $SN(0, 1, \alpha)$ density (black solid line) with fitted $SN(\hat{\xi}, \hat{\omega}, \hat{\alpha})$ and $WN(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ models, where parameters are averaged over 1000 sample estimates.

The averaged parameter estimates over the whole grid of γ_1 values are given in Table 5. For the skew-normal parameter estimates, we know the true values and the averages of estimates are close to them, except for the last row where $\gamma_1 = 0.99$ corresponds to $\alpha = 27.855$, but the average of estimates is 30.274; the estimates for location and scale parameters are still close to real values.

For the Lambert W normal distribution, we cannot make an equivalent comparison, but considering the skewness parameter γ, we can notice that the estimates $\hat{\gamma}$ tend to produce higher skewness coefficient values than we fixed for the data generation, especially for higher γ_1 values, see the last column of Table 5. The differences of γ_1 and $\gamma_1(\hat{\gamma})$ are less than 0.02 up to $\gamma_1 = 0.6$. As $\gamma_1 \geq 0.7$, the difference starts to grow. For data originating from a skew-normal distribution with a skewness coefficient of 0.99, the estimated Lambert W normal model already has a skewness coefficient equal to 1.881.

TABLE 5. Data from $SN(0,1,\alpha)$. Averages of parameter estimates (from 1000 simulations) for fitted $SN(\hat{\xi},\hat{\omega},\hat{\alpha})$ and $WN(\hat{\mu},\hat{\sigma},\hat{\gamma})$ models. The last column corresponds to the parametric estimate of the skewness coefficient γ_1 for Lambert W normal variable.

γ_1	α	$\hat{\xi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\gamma_1(\hat{\gamma})$
0.10	0.871	0.086	0.998	0.757	0.511	0.855	0.017	0.100
0.20	1.199	0.019	0.993	1.181	0.587	0.791	0.032	0.194
0.30	1.499	0.006	0.996	1.501	0.628	0.747	0.048	0.289
0.40	1.814	0.006	0.995	1.816	0.654	0.710	0.064	0.386
0.50	2.174	0.004	0.998	2.186	0.671	0.680	0.081	0.492
0.60	2.619	0.001	0.999	2.658	0.680	0.652	0.101	0.615
0.70	3.226	0.001	1.000	3.273	0.687	0.627	0.122	0.749
0.80	4.190	0.002	0.999	4.235	0.687	0.602	0.149	0.919
0.90	6.298	0.001	0.999	6.442	0.679	0.576	0.189	1.182
0.95	9.343	0.002	0.999	9.444	0.670	0.562	0.220	1.400
0.99	27.855	0.001	0.999	30.274	0.642	0.552	0.286	1.881

Based on simulations in this section, skew-normal and Lambert W approaches offered very similar distributions and gave close fits to the datasets. The distribution the data was generated from gave a better fit on both occasions, but the alternative was not far behind both visually and by the characteristics. Taking into account the fact that the Lambert W normal distribution covers a much wider range of skewness, one can say that, in summary, the Lambert W normal approach offers a flexible alternative for the more traditional skew-normal distribution, being quite similar in less skewed situations, but having the possibility to describe more skewed data as well. Still, as pointed out before, one has to be careful with extremely skewed data, where the pdf of a Lambert W normal random variable has a specific distorted form.

4.2. Assessing the cross-distributional fitting using Kolmogorov–Smirnov test and log-likelihood method. Next, we use the same datasets and parameter estimates described in Section 4.1. For each dataset generated from $WN(0,1,\gamma)$ or $SN(0,1,\alpha)$, we apply the Kolmogorov–Smirnov (KS) goodness of fit test and log-likelihood values to assess the fit of $WN(\hat{\mu},\hat{\sigma},\hat{\gamma})$ and $SN(\hat{\xi},\hat{\omega},\hat{\alpha})$. As both models have the same number of parameters, the comparison of log-likelihood values is equivalent to the comparison of Akaike information criteria.

In this section, the results for $\gamma_1 = 0.1$ and $\gamma_1 = 0.9$ are presented to give an illustration of the dynamics. The results for several more values of γ_1 values are given in the Appendix C.

The following graphs (Figures 7 and 8) illustrate the p-values from the KS test. The left panel contains the results for datasets from $WN(0,1,\gamma)$, and the right panel from $SN(0,1,\alpha)$. The coordinates are determined by p-values from the KS test: on the horizontal axis for the Lambert W model and on the vertical axis for the skew-normal model.

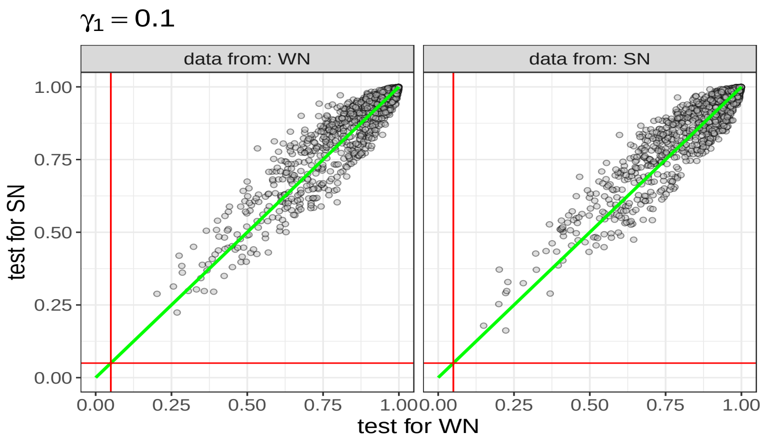

FIGURE 7. Scatterplot of KS test p-values to assess the fit of $WN(\hat{\mu},\hat{\sigma},\hat{\gamma})$ (horizontal axis) and $SN(\hat{\xi},\hat{\omega},\hat{\alpha})$ (vertical axis). Data is generated from $WN(0,1,\gamma(\gamma_1 = 0.1))$ in the left panel, and from $SN(0,1,\alpha(\gamma_1 = 0.1))$ in the right panel.

For $\gamma_1 = 0.1$, we can see that none of the p-values fall below the 0.05 threshold, the red line on the graph. So, no evidence is found against either of the models for moderately skewed Lambert W normal and skew-normal data. In addition, we can see from the graphs that the p-values from the two tests behave similarly: if the p-value is high for one KS test, it most likely is high for the other as well.

The results of KS tests are also given in Tables 6 and 7. The tables include the number of times (out of 1000 samples) we would retain the null hypothesis that a candidate distribution fits the data using a significance level of 0.05. The tables also contain the number of times a candidate distribution had a higher log-likelihood than the competitor.

TABLE 6. The results from KS test (at level 0.05) and log-likelihood comparison for $\gamma_1 = 0.1$. Table cells indicate counts from 1000 samples.

	Kolmogorov–Smirnov test		Log-likelihood	
	H_0 for WN	H_0 for SN	higher for WN	higher for SN
data: WN	1000	1000	395	605
data: SN	1000	1000	348	652

For the $\gamma_1 = 0.1$, the models are equally good based on the KS test, but the log-likelihood comparison gives a different result, as the skew-normal model is preferred in both cases. So, for Lambert W data with low skewness, based on log-likelihood method one would decide in favour of the skew-normal model.

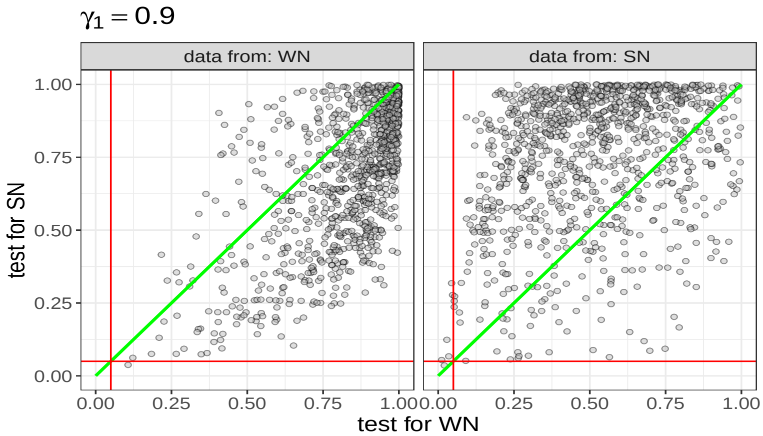

FIGURE 8. Scatterplot of KS test p-values to assess the fit of $WN(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ (horizontal axis) and $SN(\hat{\xi}, \hat{\omega}, \hat{\alpha})$ (vertical axis). Data is generated from $WN(0, 1, \gamma(\gamma_1 = 0.9))$ in the left panel, and from $SN(0, 1, \alpha(\gamma_1 = 0.9))$ in the right panel.

If we consider the datasets with higher skewness coefficient $\gamma_1 = 0.9$, one would make some rejections based on the KS test. If the data is from Lambert W normal distribution, there is one rejection for the skew-normal model. For the skew-normal data, there is also one rejection for the skew-normal model and 6 rejections for the Lambert W model. As the count of rejections is low, we can still say that the KS test cannot distinguish the two models, although the p-values now tend to be higher for the correct model class, see Figure 8. However, the log-likelihood comparison now gives a clear preference for the right model, as shown in Table 7.

TABLE 7. The results from KS test (at level 0.05) and log-likelihood comparison for $\gamma_1 = 0.9$. Table cells indicate counts from 1000 samples.

	Kolmogorov–Smirnov test		Log-likelihood	
	H_0 for WN	H_0 for SN	higher for WN	higher for SN
data: WN	1000	999	868	132
data: SN	994	999	38	962

4.3. Fitting the Lambert W random variables to insurance data. In this section, we revisit an example of fitting Lambert W normal and skew-normal distributions to the real data. The example uses two well-known skewed datasets, US indemnity data [11] and the Danish fire data [15].

These datasets are extensively used in previous research and fitted by various distributions. A comprehensive overview with 19 distributions is given in [10]. Also, as the datasets are skewed and heavy-tailed, both original and log-transformed versions are used, resulting in 4 datasets to be fitted. In [14], the Lambert W normal and Lambert W exponential fit was added to the list, and both models showed a relatively good fit, especially in comparison to skew-normal.

It is also worth looking at the skewness coefficients for these distributions as, by the results of Section 3, the range of values for the skewness coefficient for skew-normal distribution is quite limited. The estimated skewness coefficient is 9.15 for the US indemnity data and 18.74 for the Danish fire loss data. Applying the log-transform alleviates the skewness: for the US indemnity data, the skewness coefficient value reduces to $\gamma_1 = -0.15$ and for the Danish fire loss data $\gamma_1 = 1.76$ (which is still substantial).

The comparison of fit by the Lambert W normal and skew-normal distributions to the four described datasets using AIC values is given in Table 8. As we can see, the Lambert W normal approach wins for the datasets with a larger skewness coefficient. Taking into account that the skewness coefficient for the skew-normal distribution has an upper limit of about 0.9953, the significantly better fit by Lambert W on these three datasets is to be expected. For near-symmetric data (log-transformed US indemnity claims), the two models produce equal results based on AIC.

TABLE 8. AIC values for fitted Lambert W normal and skew-normal distributions.

	US indemnity		Danish fire loss	
	original	log-loss	original	log-loss
Lambert W normal	13397.48	5737.793	6699.82	2978.46
skew-normal	16315.13	5737.790	12608.36	3441.49

In conclusion, the Lambert W normal should clearly be preferred over the skew-normal distribution for highly skewed data. For moderately skewed data, both distributions are suitable.

5. Conclusions and discussion

In this study, we analyzed the properties of Lambert W normal random variables and their applicability for modelling skewed data. The results show that Lambert W normal random variables provide a viable alternative to the well-studied skew-normal distribution while covering a wider range of skewness.

The results demonstrate similarities of probability density functions of skew-normal and Lambert W normal distributions in the applicable range of skewness for the skew-normal distribution (i.e. skewness coefficient up to 0.99), although, for higher values of skewness coefficient, the skew-normal distribution clearly deviates and moves towards the half-normal "cut" shape.

It is also shown that the Lambert W normal distributions may have a heavier tail even with the same skewness, which makes it more suitable to model heavy-tailed distributions widely used in finance and insurance. Moreover, the Lambert W normal distribution covers a wider range of skewness and kurtosis than the skew-normal distribution, and the quantiles of the Lambert W normal distribution exceed those of the skew-normal distribution when the skewness increases. The same holds for the expected shortfall.

An obvious downside for the Lambert W distribution is its complexity even with the same number of parameters: besides the Lambert W function not having an explicit form, the density function has an asymptote (although difficult to distinguish in practice), and for some parameter combinations (when skewness coefficient is large) the shape of the distribution can be unsuitable for practical applications. Also, the maximum likelihood estimation of parameters is very sensitive to starting values and may encounter difficulties with the convergence, which complicates using the distribution in practice.

Somewhat surprisingly, we found that the skew-normal distribution also had problems with parameter estimation with both maximum likelihood method and maximum spacing method with big values of skewness parameter α.

Finally, as the Lambert W normal distribution offers more flexibility and is thus applicable for more scenarios, it is definitely a possible alternative to consider for scenarios that are too skewed for the skew-normal distribution. Still, it depends on the particular problem and objective whether the gain in model precision is worth the extra complexity.

References

[1] A. Azzalini, *A class of distributions which includes the normal ones*, Scand. J. Stat. **12** (1985), 171–178. URL

[2] A. Azzalini, *The skew-normal distribution and related multivariate families*, Scand. J. Stat. **32** (2005), 159–188. URL

[3] A. Azzalini, *The R package `sn`: The skew-normal and related distributions such as the skew-t and the SUN (version 2.1.1)*, Università degli Studi di Padova, Italia, 2023. URL

[4] A. Azzalini and A. Capitanio, *Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **65** (2003), 367–389. DOI

[5] M. Bernardi, *Risk measures for skew normal mixtures*, Statist. Probab. Lett. **83** (2013), 1819–1824. DOI

[6] P. Brito, F. Fabião, and A. Staubyn, *Euler, Lambert, and the Lambert w-function today*, The Mathematical Scientist **33** (2008). URL

[7] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, *On the Lambert w function*, Adv. Comput. Math. **5** (1996), 329–359. DOI

[8] M. L. Delignette-Muller and C. Dutang, *fitdistrplus: An R package for fitting distributions*, J. Statistical Software **64** (2015), 1–34. DOI

[9] T. Dence, *A brief look into the Lambert w function*, Applied Mathematics **4** (2013), 887–892. DOI

[10] M. Eling, *Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?* Insurance: Mathematics and Economics **51** (2012), 239–248. DOI

[11] E. W. Frees and E. A. Valdez, *Understanding relationships using copulas*, N. Am. Actuar. J. **2** (1998), 1–25. DOI

[12] G. M. Goerg, *Lambert W random variables – a new family of generalized skewed distributions with applications to risk estimation*, Ann. Appl. Stat. **5** (2011), 2197–2230. DOI

[13] G. M. Goerg, *Lambertw: Probabilistic models to analyze and gaussianize heavy-tailed, skewed data*, R package version 0.6.9-1.

[14] M. Käärik, A. Selart, T. Puhkim, and L. Tee, *Lambert w random variables and their applications in loss modelling*, Symmetry **15** (2023). DOI

[15] A. J. McNeil, *Estimating the tails of loss severity distributions using extreme value theory*, ASTIN Bulletin: The Journal of the IAA **27** (1997), 117–137. DOI.

[16] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023. URL

[17] B. Ranneby, *The maximum spacing method. An estimation method related to the maximum likelihood method*, Scand. J. Stat. **11** (1984), 93–112. URL

Appendix A. Derivation of the formulas for skewness coefficient $\gamma_1(\gamma)$ and kurtosis coefficient $\gamma_2(\gamma)$

As the skewness coefficient and kurtosis coefficient are location-scale invariant it suffices to find these coefficients for $Y \sim WN(0,1,\gamma)$. The moments for Y are found using the relation (see [12])

$$\mathrm{E}Y^k = \frac{1}{k^k}\frac{\partial^k}{\partial\gamma^k}M_{N(0,1)}(\gamma k) = \frac{1}{k^k}\frac{\partial^k}{\partial\gamma^k}\exp\left(\frac{\gamma^2k^2}{2}\right),$$

where $M_{N(0,1)}$ denotes the moment generating function of $N(0,1)$. The first four moments equal to

$$\begin{aligned}
\mathrm{E}Y\ &= \gamma e^{\frac{\gamma^2}{2}},\\
\mathrm{E}Y^2 &= (4\gamma^2+1)e^{2\gamma^2},\\
\mathrm{E}Y^3 &= 9\gamma(3\gamma^2+1)e^{\frac{9\gamma^2}{2}},\\
\mathrm{E}Y^4 &= (16^2\gamma^4+96\gamma^2+3)e^{8\gamma^2}.
\end{aligned}$$

Taking into account the relations between moments and central moments, we get

$$\begin{aligned}
\mathrm{Var}Y &= e^{\gamma^2}((4\gamma^2+1)e^{\gamma^2}-\gamma^2),\\
\mathrm{E}(Y-\mathrm{E}Y)^3 &= 9\gamma(3\gamma^2+1)e^{\frac{9\gamma^2}{2}} - 3\cdot(4\gamma^2+1)e^{2\gamma^2}\cdot\gamma e^{\frac{\gamma^2}{2}} + 2\cdot\gamma^3 e^{\frac{3\gamma^2}{2}}\\
&= \gamma e^{\frac{3\gamma^2}{2}}\big(9(3\gamma^2+1)e^{3\gamma^2} - 3(4\gamma^2+1)e^{\gamma^2} + 2\gamma^2\big),\\
\mathrm{E}(Y-\mathrm{E}Y)^4 &= (16^2\gamma^4+96\gamma^2+3)e^{8\gamma^2} - 4\cdot 9\gamma(3\gamma^2+1)e^{\frac{9\gamma^2}{2}}\cdot\gamma e^{\frac{\gamma^2}{2}}\\
&\quad + 6\cdot(4\gamma^2+1)e^{2\gamma^2}\cdot\gamma^2 e^{\gamma^2} - 3\cdot\gamma^4 e^{2\gamma^2}\\
&= e^{2\gamma^2}\big((16^2\gamma^4+96\gamma^2+3)e^{6\gamma^2} - 36\gamma^2(3\gamma^2+1)e^{3\gamma^2}\\
&\quad + 6\gamma^2(4\gamma^2+1)e^{\gamma^2} - 3\gamma^4\big).
\end{aligned}$$

Consequently, the skewness coefficient can be found as

$$\begin{aligned}
\gamma_1(\gamma) = \frac{\mathrm{E}(Y-\mathrm{E}Y)^3}{\sqrt{(\mathrm{Var}Y)^3}} &= \frac{\gamma e^{\frac{3\gamma^2}{2}}\big(9(3\gamma^2+1)e^{3\gamma^2} - 3(4\gamma^2+1)e^{\gamma^2} + 2\gamma^2\big)}{e^{\frac{3\gamma^2}{2}}((4\gamma^2+1)e^{\gamma^2}-\gamma^2)^{3/2}}\\
&= \frac{\gamma\big(9(3\gamma^2+1)e^{3\gamma^2} - 3(4\gamma^2+1)e^{\gamma^2} + 2\gamma^2\big)}{((4\gamma^2+1)e^{\gamma^2}-\gamma^2)^{3/2}}
\end{aligned}$$

and the kurtosis coefficient can be expressed as

$$\gamma_2(\gamma) = \frac{\mathrm{E}(Y-\mathrm{E}Y)^4}{(\mathrm{Var}Y)^2} =$$

$$= \frac{e^{2\gamma^2}\big((16^2\gamma^4+96\gamma^2+3)e^{6\gamma^2} - 36\gamma^2(3\gamma^2+1)e^{3\gamma^2} + 6\gamma^2(4\gamma^2+1)e^{\gamma^2} - 3\gamma^4\big)}{e^{2\gamma^2}((4\gamma^2+1)e^{\gamma^2}-\gamma^2)^2}$$
$$= \frac{\big((16^2\gamma^4+96\gamma^2+3)e^{6\gamma^2} - 36\gamma^2(3\gamma^2+1)e^{3\gamma^2} + 6\gamma^2(4\gamma^2+1)e^{\gamma^2} - 3\gamma^4\big)}{((4\gamma^2+1)e^{\gamma^2}-\gamma^2)^2}.$$

Appendix B. Derivation of the formula for $\mathrm{E}S_p^{WN}$

Here we derive the formula (6) of the expected shortfall $\mathrm{E}S_p^{WN}$ for the Lambert W standard normal distribution $WN(0,1,\gamma)$.

Let $Y \sim WN(0,1,\gamma)$ and $\mathrm{P}(Y > q_p) = 1-p$, then

$$(1-p)\mathrm{E}(Y|Y>q_p) = \int_{q_p}^{\infty} y f_{WN}(y)dy$$

giving

$$(1-p)\mathrm{E}S_p^{WN} = \int_{q_p}^{\infty} y\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{W_0^2(\gamma y)}{\gamma^2}}\frac{e^{W_0(\gamma y)}}{1+W_0(\gamma y)}dy.$$

Since

$$y = \frac{1}{\gamma}\gamma y = \frac{1}{\gamma}W_0(\gamma y)e^{W_0(\gamma y)} \quad \text{and} \quad \frac{e^{W_0(\gamma y)}}{1+W_0(\gamma y)}dy = \frac{1}{\gamma}dW_0(\gamma y),$$

the integral above transforms to

$$\frac{1}{\gamma^2}\int_{W_0(\gamma q_p)}^{\infty} W_0(\gamma y)e^{W_0(\gamma y)}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{W_0^2(\gamma y)}{\gamma^2}}dW_0(\gamma y).$$

Next, expressing the product of two exponentials as

$$e^{W_0(\gamma y)}e^{-\frac{1}{2}\frac{W_0^2(\gamma y)}{\gamma^2}} = e^{-\frac{1}{2}\left(\frac{W_0(\gamma y)}{\gamma}-\gamma\right)^2+\frac{\gamma^2}{2}}$$

and using the change of variables with $x = \frac{W_0(\gamma y)}{\gamma} - \gamma$, leads to the integral

$$e^{\frac{\gamma^2}{2}}\int_u^{\infty}(x+\gamma)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx,$$

where $u = \frac{W_0(\gamma q_p)}{\gamma} - \gamma$. This integral can be expressed by standard normal pdf and cdf giving finally

$$(1-p)\mathrm{E}S_p^{WN} = e^{\frac{\gamma^2}{2}}(\varphi(u) + \gamma(1-\Phi(u))).$$

The formula (6) of the expected shortfall $\mathrm{E}S_p^{WN}$ follows.

Appendix C. Comparison by Kolmogorov–Smirnov test and log-likelihood: additional results

The comparison based on KS test and log-likelihood for skewness coefficient γ_1 values 0.1 and 0.9 is discussed in Section 4.2. Here we present results for more values of γ_1 to expand the overview.

As one can see from Table 9, the KS test considers both models suitable for both datasets if the skewness coefficient γ_1 is 0.9 or smaller. For γ_1 equal to 0.95 or 0.99, there are more occasions when Lambert W normal model is rejected for skew-normal data but not vice versa. Although the proportion of H_0/H_1 decisions changes little as γ_1 grows, the scatterplots of KS test p-values in Figure 9 indicate that the p-values tend to point towards the correct model.

Table 9. Comparison by KS test and log-likelihood for different values of γ_1. Results based on 1000 samples

		Kolmogorov–Smirnov test		Log-likelihood	
γ_1	data	H_0 for WN	H_0 for SN	higher for WN	higher for SN
0.1	WN	1000	1000	395	605
	SN	1000	1000	348	652
0.4	WN	999	1000	669	331
	SN	1000	1000	365	635
0.8	WN	1000	1000	812	188
	SN	999	1000	130	870
0.9	WN	1000	999	868	132
	SN	994	999	38	962
0.95	WN	1000	997	867	133
	SN	974	997	15	985
0.99	WN	1000	997	900	100
	SN	714	993	1	999

Also, if we compare the log-likelihood values of the two models (see Table 9), we can see that with increasing skewness, there are more cases when the correct model is preferred and fewer cases when Lambert W model is suitable for skew-normal data.

In summary, we can say that for low and moderate skewness, both models are appropriate, but for the γ_1 value close to 0.99, the skew-normal distribution has a very specific shape that the Lambert W normal cannot describe.

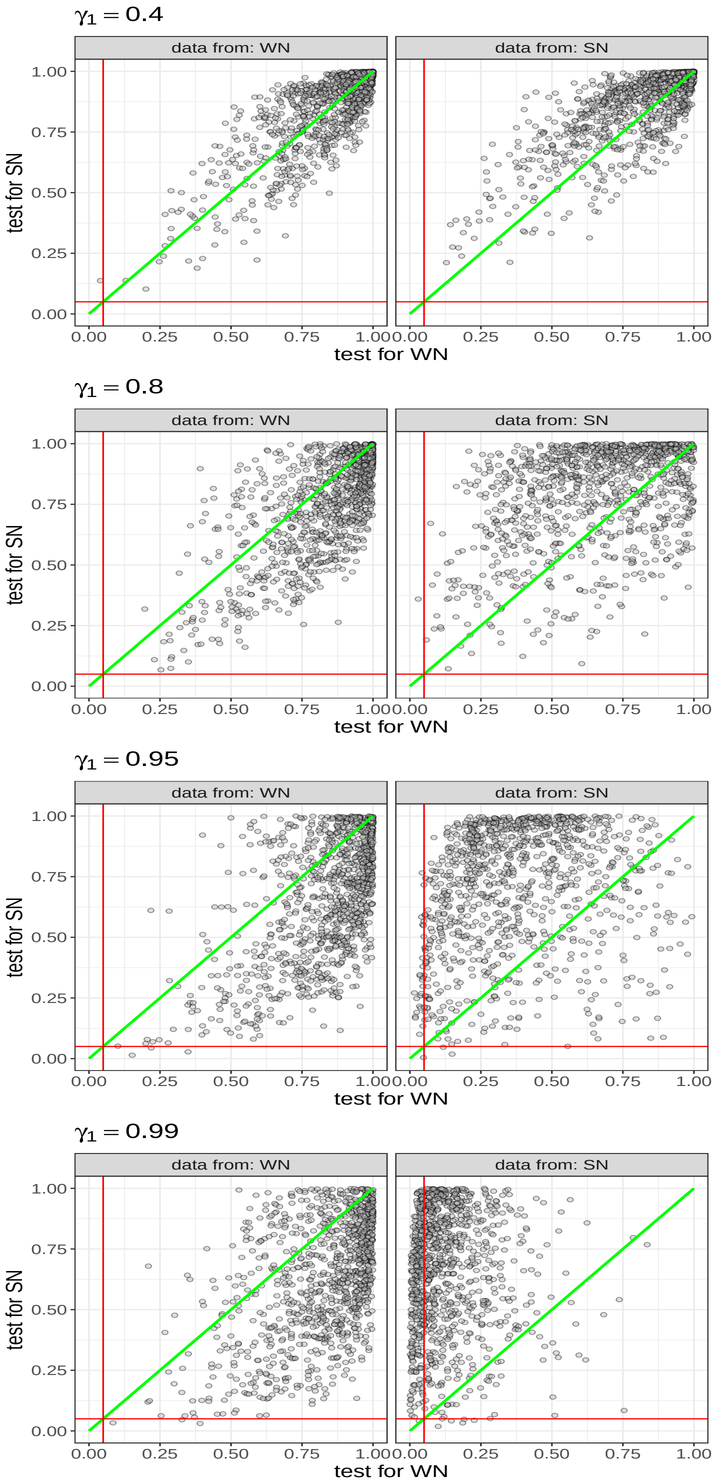

FIGURE 9. Scatterplot of KS test p-values to assess the fit of $WN(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ (horizontal axis) and $SN(\hat{\xi}, \hat{\omega}, \hat{\alpha})$ (vertical axis). Data is generated from $WN(0, 1, \gamma(\gamma_1))$ in the left panel, and from $SN(0, 1, \alpha(\gamma_1))$ in the right panel.

Acknowledgements

The work was supported by the Estonian Research Council grant PRG1197. The authors are thankful to Tuuli Puhkim for input on the earlier draft of the paper and Magnus Käärik for assistance in refining the language and enhancing clarity.

Institute of Mathematics and Statistics, University of Tartu, 18 Narva str., 51009 Tartu, Estonia

E-mail address: meelis.kaarik@ut.ee

E-mail address: anne.selart@ut.ee