

# Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?

Martin Eling\*

*Institute of Insurance Economics, University of St. Gallen, Kirchlistrasse 2, 9010 St. Gallen, Switzerland*

## ARTICLE INFO

### Article history:

Received June 2011

Received in revised form

February 2012

Accepted 3 April 2012

### Keywords:

Goodness of fit

Risk measurement

Skew-normal

Skew-student

## ABSTRACT

This paper analyzes whether the skew-normal and skew-student distributions recently discussed in the finance literature are reasonable models for describing claims in property-liability insurance. We consider two well-known datasets from actuarial science and fit a number of parametric distributions to these data. Also the non-parametric transformation kernel approach is considered as a benchmark model. We find that the skew-normal and skew-student are reasonably competitive compared to other models in the literature when describing insurance data. In addition to goodness-of-fit tests, tail risk measures such as value at risk and tail value at risk are estimated for the datasets under consideration.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

The normal distribution is the most popular distribution used for modeling in economics and finance. In general, however, insurance risks have skewed distributions, which is why in many cases the normal distribution is not an appropriate model for insurance risks or losses (see, e.g., Lane, 2000; Vernic, 2006). Besides skewness, some insurance risks (especially those exposed to catastrophes) also exhibit extreme tails (see Embrechts et al., 2002). The skew-normal distribution as well as other distributions from the skew-elliptical class thus might be promising alternatives to the normal distribution since they preserve advantages of the normal one with the additional benefit of flexibility with regard to skewness (e.g., with the skew-normal) and kurtosis (e.g., with the skew-student).

In this paper, we analyze whether these skewed distributions are reasonably good models for describing insurance claims. We consider two datasets widely used in literature and fit the skew-normal and skew-student to these data. A number of benchmark models are involved in the model comparison, as well as a goodness-of-fit procedure, in order to compare the performance of the skewed distributions in describing the insurance claims data. The motivation for consideration of the skew-normal and the skew-student is that these are popular in recent finance literature (Adcock, 2007, 2010; De Luca et al., 2006), easy to interpret, and easy to implement.

This work is related to Bolancé et al. (2008), who fit the skew-normal and log skew-normal to a set of bivariate claims data

from the Spanish motor insurance industry. To our knowledge, Bolancé et al. (2008) is the only paper to date that uses the skew-normal distribution to fit insurance claims. We build on and extend their results by considering the skew-student distribution and by using different datasets. Furthermore, our analysis is broader than Bolancé et al. (2008) in that we compare our results to a large number of – 18 – alternative distributions, whereas Bolancé et al. (2008) restrict their presentation to the normal, the skew-normal, and a transformation kernel approach. We also include the transformation kernel approach as a non-parametric alternative in our discussion.

To preview our main results, we find that the skew-student and skew-normal are reasonably good models compared to other models presented in literature (see, e.g., Kaas et al., 2009). Given that this paper presents only some first tests of claims modeling in actuarial science using two well-known datasets, we call for more applications of these distributions in the field of insurance in order to more closely analyze whether the skew-normal and skew-student are promising distributions for claims modeling.

The remainder of the paper is organized as follows. In Section 2, we describe the methods we use in our estimation, especially the models involved. Section 3 presents the data. The estimation results are set out in Section 4, both those pertaining to goodness of fit and risk measurement. Conclusions are drawn in Section 5.

## 2. Skew-elliptical distributions

We briefly describe the distributions to be investigated in this paper, along with a short description of the benchmark models we use in the goodness-of-fit context. More detail on skewed distributions can be found in Genton (2004) and a fuller description

\* Tel.: +41 71 2247980; fax: +41 71 2247990.

E-mail address: [martin.eling@unisg.ch](mailto:martin.eling@unisg.ch).

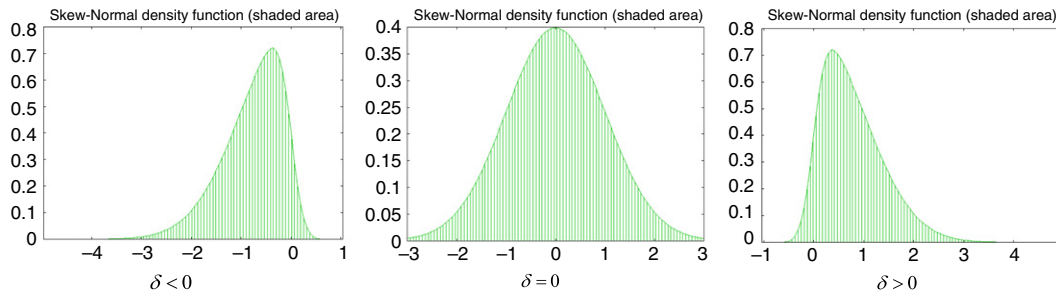


Fig. 1. Skew-normal distribution for three different values of delta.

of the benchmark models can be found in actuarial textbooks, for example, Mack (2002) or Panjer (2007).

### 2.1. Skew-normal

A continuous random variable  $X$  has a skew-normal distribution if its probability density function (pdf) has the form:

$$f(x) = 2\phi(x)\Phi(\alpha x). \quad (1)$$

$\alpha$  is a real number,  $\phi(\cdot)$  denotes the standard normal density function, and  $\Phi(\cdot)$  its distribution function (see Azzalini, 1985). The distribution shown by Eq. (1) is called the skew-normal distribution with shape parameter  $\alpha$ , i.e.,  $X \sim SN(0, 1, \alpha)$ . The skew-normal distribution reduces to the standard normal distribution when  $\alpha = 0$  and to the half-normal when  $\alpha \rightarrow \pm\infty$ . In both empirical and theoretical work, location and scale parameters are necessary. These can be included via the linear transformation  $Y = \xi + \omega X$ , which is said to have the skew-normal distribution  $Y \sim SN(\xi, \omega^2, \alpha)$ , with  $\omega > 0$ . The parameters  $\xi$ ,  $\omega$ , and  $\alpha$  are called location, scale, and shape, respectively. When  $\alpha = 0$ , the random variable  $Y$  is distributed as  $N(\xi, \omega^2)$ .

An alternative representation of the skew-normal that is especially popular in financial modeling is the characterization of skew-normality given by Pourahmadi (2007). A continuous random variable  $Y \sim SN(\xi, \omega^2, \alpha)$  can be written as a special weighted average of a standard normal variable and a half-normal one.  $Y$  is said to have a skew-normal distribution if and only if the following representation holds:

$$Y = \xi + \omega X = \xi + \omega \left( \delta |Z_1| + \sqrt{1 - \delta^2} Z_2 \right), \quad (2)$$

with  $\delta = \alpha / \sqrt{1 + \alpha^2} \in [-1, 1]$ .  $Z_1$  and  $Z_2$  are independent  $N(0, 1)$  random variables.  $Y$  collapses into  $N(\xi, \omega^2)$  if  $\delta = 0$ . Eq. (2) offers a direct financial interpretation, i.e., besides the location parameter  $\xi$ , the return  $Y$  is driven by two components:

- a half-Gaussian driver  $|Z_1|$  modulated by  $\omega\delta$ , and
- a Gaussian driver  $Z_2$  modulated by  $\omega\sqrt{1 - \delta^2}$ .

The parameter  $\delta$  plays a key role in determining the skewness since  $\delta$  weights the presence of a half-Gaussian  $|Z_1|$  on  $Y$  (see Eling et al., 2010). The more positive (negative)  $\delta$ , the more pronounced to the right (left) the skewness. Fig. 1 illustrates the impact of  $\delta$  on the skewness of the skew-normal distribution (all figures in this paper were generated using a package available for the software R; see <http://azzalini.stat.unipd.it/SN/>).

An important property of the skew-normal distribution is that all moments exist and are finite. The moment-generating function of  $Y \sim SN(\xi, \omega^2, \alpha)$  is given by:

$$M(t) = E[e^{tY}] = 2 \exp\left(\xi t + \frac{\omega^2 t^2}{2}\right) \Phi(\delta \omega t). \quad (3)$$

Consequently, the moments of  $Y$  are easily derived and we obtain easy-to-read expressions for mean, variance, skewness and kurtosis that highlight the influence of the skewness parameter  $\delta$ :

$$E(Y) = \xi + \omega \sqrt{2/\pi} \delta, \quad (4)$$

$$\text{Var}(Y) = \omega^2 (1 - 2\delta^2/\pi), \quad (5)$$

$$\text{Skewness}(Y) = (4 - \pi)/2 (\delta(2/\pi)^{1/2})^3 / (1 - 2\delta^2/\pi)^{3/2}, \quad (6)$$

$$\text{Excess Kurtosis}(Y) = 2(\pi - 3)(\delta(2/\pi)^{1/2})^4 / (1 - 2\delta^2/\pi)^2. \quad (7)$$

The skew-normal distribution extends the normal distribution in several ways. These can be formalized through a number of properties, such as inclusion (the normal distribution is a skew-normal distribution with shape parameter equal to zero) or affinity (any affine transformation of a skew-normal random vector is skew-normal; for more details, see, e.g., De Luca et al., 2006). Note that the skew-normal distribution can take values of skewness only from  $-1$  to  $1$ . Compared to the normal distribution, it thus extends the range of available skewness, but the range of potential skewness values is still limited.

### 2.2. Skew-student

The skew-student distribution allows regulating both the skewness and kurtosis of a distribution. This attribute is particularly useful in empirical applications where we want to consider distributions with higher kurtosis than the normal, which is often the case in both finance and insurance applications. One limitation of the skew-normal distribution described in Eq. (1) is that it has a kurtosis only slightly higher than the normal distribution (the maximum excess kurtosis is 0.87). An appealing alternative is offered by a skewed version of the Student's  $t$  distribution, introduced by Branco and Dey (2001) and further developed by Azzalini and Capitanio (2003). We define the standardized Student's  $t$  skewed distribution using the transformation:

$$X = \frac{Z}{\sqrt{W/v}}, \quad (8)$$

where  $W \sim \chi^2(\nu)$ , with  $\nu$  degrees of freedom and  $Z$  is an independent  $SN(0, 1, \alpha)$ , instead of  $N(0, 1)$  as used to produce the standard  $t$ . The linear transformation  $Y = \xi + \omega X$  has a skew- $t$  distribution with parameters  $(\xi, \omega, \alpha)$  and we write  $Y \sim ST(\xi, \omega^2, \alpha)$ . Mean and variance of  $Y \sim ST(\xi, \omega^2, \alpha)$  can be computed as follows (see Azzalini and Capitanio, 2003, also for higher moments):

$$E(Y) = \xi + \omega \eta \delta, \quad \text{with } \nu > 1, \quad (9)$$

$$\text{Var}(Y) = \omega^2 \left( \frac{\nu}{\nu - 2} - \eta \delta^2 \right),$$

$$\text{where } \eta = \sqrt{\frac{\nu}{\pi}} \frac{\Gamma(\frac{1}{2}(\nu - 1))}{\Gamma(\frac{1}{2}\nu)}. \quad (10)$$

Comparable to the skew-normal case, Eqs. (9) and (10) highlight the influence of  $\delta$  on the mean and variance of the skew-student

distribution. Again, the former is a linear increasing function of  $\delta$ , whereas the variance is a quadratic function on  $\delta$ . Compared to the skew-normal distribution, the skew-student can take more extreme values, for either skewness or kurtosis.

### 2.3. Benchmark models

The choice of benchmark models is based on their use in actuarial and financial theory. All benchmark models are implemented in the R package *ghyp* and *MASS*. We use both packages to derive maximum likelihood estimators of the best-fitting parameters and to compare these distributions with the above-described skew-normal and skew-student, which are implemented in the R package *sn*. Some of the benchmark distributions will also be involved in the risk measurement procedure where we compare model results with the empirical results in order to evaluate the accuracy of different models when measuring risk.

The R package *ghyp* contains a variety of distributions popular in the fields of both finance and actuarial science. The following distributions can be fitted using the “*stepAIC.ghyp*” command: the normal, the student  $t$  (see Krole et al., 2007), the normal inverse Gaussian (NIG) (Barndorff-Nielsen, 1997), and the hyperbolic (Eberlein et al., 1998). In contrast to the normal distribution, some of the benchmark distributions are able to account for skewness, kurtosis, or even for both (e.g., NIG, hyperbolic). Some of these distributions are related to each other, e.g., the student  $t$ , the normal inverse Gaussian, and the hyperbolic all belong to the class of generalized hyperbolic distributions. The software calculates the value of the log-likelihood function, providing a basis upon which to compare the models using, for example, the Akaike information criteria (AIC).

The R package *MASS* contains 15 distributions that can be considered in a goodness-of-fit context. The benchmark distributions recognized via the command “*fitdistr*” are the beta, Cauchy, chi-squared, exponential,  $f$ , gamma, geometric, log-normal, logistic, negative binomial, normal, Poisson, student  $t$ , and Weibull distribution. Again, to make model comparisons, we derive the log-likelihood and then use it to calculate other criteria such as the AIC or the Bayesian information criterion (BIC).<sup>1</sup>

In the empirical section of this paper (Section 4), we estimate the parameters of the distributions via maximum likelihood estimation. We then compare the skew-normal and skew-student distributions with the various other distributions via the AIC and Kolmogorov–Smirnov goodness-of-fit tests. It should be mentioned that a better AIC value does not necessarily mean that a model is better since there are many other methods and caveats that actuaries need to keep in mind, such as, e.g., the risk of change of the underlying stochastic. To consider some of these aspects we incorporate different methods and datasets (see below for more details). But nevertheless the analysis should be understood as a rather qualitative comparison in order to identify whether the skew-normal and skew-student distribution are reasonable models for actuarial work.

As non-parametric alternative we also implement the transformation kernel approach analyzed in Bolancé et al. (2008) and introduced in Bolancé et al. (2003). The idea of this approach is to first transform the original data and then to apply the classical kernel density estimator to the transformed data. The transformation

kernel approach thus addresses the limitation of classical kernel density estimation which does not perform well when the true density is asymmetric (as is the case in our data). We implement the transformation kernel approach using the R routine described in Pitt et al. (2011) and refer to this work for more details on the modeling approach, e.g., with respect to the target function, the bandwidth selection, and the programming code (we use method 1 described by Pitt et al. 2011).

### 3. Data

We consider two datasets widely used in the actuarial literature. The first is a dataset of US indemnity losses and the second is comprised of Danish fire losses. Both datasets are freely available and on request we provide documentation of the R code used; thus others should be able to easily replicate and expand the results presented here.

The first dataset is comprised of US indemnity losses used in Frees and Valdez (1998). The data consist of 1500 general liability claims, giving for each the indemnity payment (denoted in the data as “loss”) and the allocated loss adjustment expense (denoted in the data as “*alae*”), both in USD. The latter is the additional expense associated with settlement of the claim (e.g., claims investigation expenses and legal fees). We focus here on the pure loss data and do not consider the expenses, but results taking these expenses into consideration are available upon request. The dataset can be found in the R packages *copula* and *evd*. It is used in other work, including Klugman and Parsa (1999) and Dupuis and Jones (2006). For scaling purposes, we divide the data by 1000; we thus consider TUSD instead of USD.

The second dataset is comprised of Danish fire losses analyzed in McNeil (1997). These data represent Danish fire losses in million Danish Kroner and were collected by a Danish reinsurance company. The dataset contains individual losses above 1 million Danish Kroner, a total of 2167 individual losses, covering the period from January 3, 1980 to December 31, 1990. The data is adjusted for inflation to reflect 1985 values. The dataset can be found in the R packages *fEcofin* and *fExtremes*. These data are considered in Resnick (1997), Cooray and Ananda (2005), and Dell'Aquila and Embrechts (2006), among others.

Fig. 2 presents two histograms for the datasets considered here, as well as the corresponding normal Q–Q plots. The two left diagrams show the indemnity loss data from Frees and Valdez (1998) and the two diagrams on the right represent the Danish fire losses from McNeil (1997). Both histograms reveal a very typical feature of insurance claims data: a large number of small losses and a lower number of very large losses. The absolute values for the indemnity losses presented in the left histogram are higher than the values presented in the right histogram, which is simply due to scaling (TUSD on the left, million Danish Kroner on the right).<sup>2</sup>

Table 1 presents descriptive statistics for the two datasets. In addition to the number of observations, indicators for the first four moments (mean, standard deviation, skewness, excess kurtosis), and minimum and maximum, we also present the 99% quantile and the mean loss, if the loss is above 99%. The 99% quantile is the value at risk (at 99% confidence level) and the mean loss exceeding the 99% quantile the tail value at risk.

<sup>1</sup> We also implemented the Pareto distribution which is an important model in catastrophe insurance (especially for large losses) and which is not implemented in the above mentioned R packages. To integrate this distribution in our analysis we used the distribution function  $a^*(x_{\min}/x)^{a+1}$ , with  $a > 0$  and  $x_{\min} > 0$ . It only produced reasonable results in one case (the original Danish fire data), while in all other cases it does not fit the data well. The results are available upon request.

<sup>2</sup> In both cases, individual losses are considered. The individual losses could be aggregated to a portfolio of losses, which would then allow considering number of claims, individual claim size, and aggregate loss amount. In this paper, we restrict the analysis to the distribution of the individual claim sizes. Note also the kink in the Q–Q plot for the US indemnity losses that occurs around 500 million Euros. The kink arouses interest in subjecting the data to other modeling approaches, such as peak over threshold models from extreme value theory (see McNeil and Saladin, 1997; McNeil, 1997; Embrechts et al., 1997). We do not consider extreme value theory explicitly in this paper, but we do discuss its implications in the paper's Conclusion section.

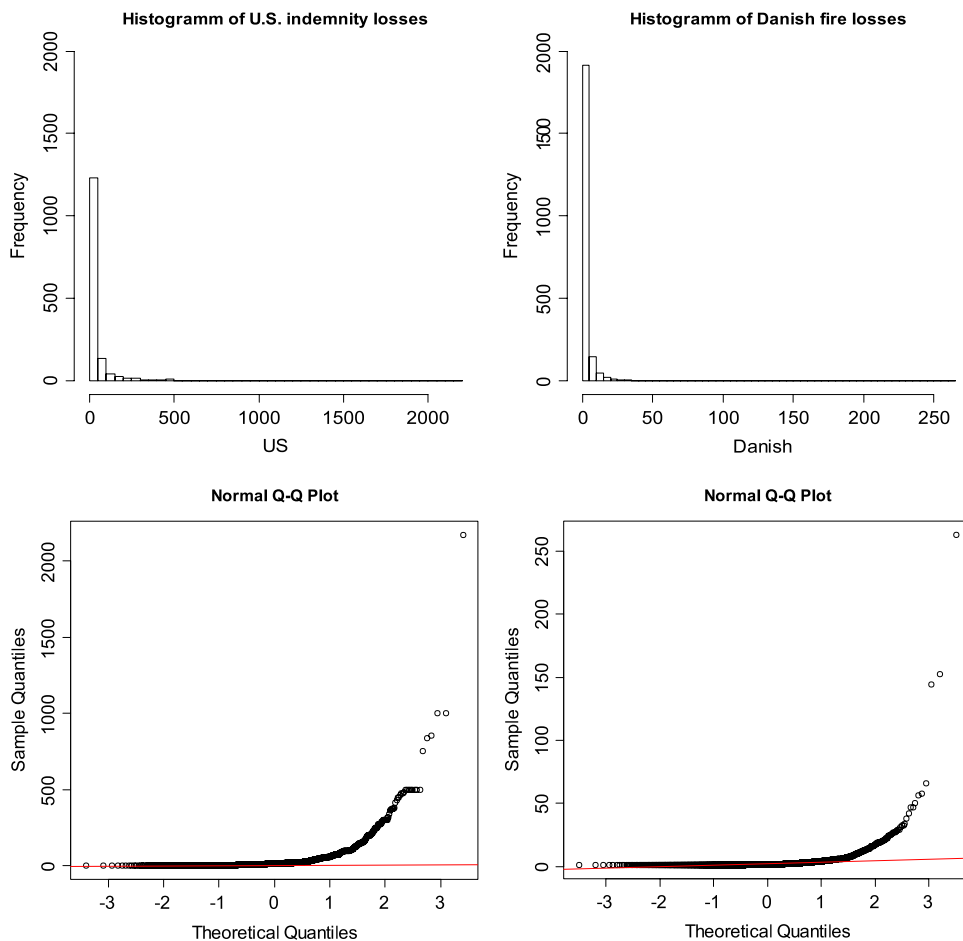


Fig. 2. US indemnity losses (left) and Danish fire losses (right).

**Table 1**  
Descriptive statistics for original data.

	US indemnity losses	Danish fire losses
No. of observations	1500	2167
$E(X)$	41.21	3.39
St. Dev ( $X$ )	10.27	8.51
Skewness ( $X$ )	9.15	18.74
Kurtosis ( $X$ )	141.98	482.20
Minimum	0.01	1.00
Maximum	2173.60	263.25
99% Quantile (value at risk)	475.06	26.04
$E(X X \geq \text{value at risk})$	739.62	58.59

The descriptive statistics show that the Danish fire losses are more extreme with respect to skewness and kurtosis than the US indemnity losses. Values for skewness and kurtosis are around 9.15 and 141.98 for the indemnity losses; the corresponding values for the fire losses are 18.74 and 482.12. Both the indemnity and the fire data are thus significantly skewed to the right and exhibit high kurtosis.<sup>3</sup> These characteristics are also reflected in the relatively high values for value at risk and tail value at risk. These findings make the skew-student look like an especially promising distribution to be fitted to this type of data since it accounts for both skewness as well as kurtosis.

In a second step, we analyze the logarithm of the data. Fig. 3 presents the histograms and normal Q – Q plots for the indemnity

loss data (left) and the Danish fire losses (right) after taking the natural logarithm of all data values. Consideration of logarithm data is a widespread practice in statistics and actuarial science in order to decrease extreme values of skewness and kurtosis for modeling purposes (see, e.g., Bolancé et al., 2008). In our context, consideration of the log data can also be interpreted as a kind of robustness test for analyzing model performance.

After taking the natural logarithm, the indemnity loss data look much more like the normal distribution, whereas the fire losses distribution is still skewed to the right.<sup>4</sup> Table 2 presents the descriptive statistics for the log data. Again, the number of observations, mean, standard deviation, skewness, excess kurtosis, minimum and maximum, the 99% quantile, and the mean loss, if the loss is above 99%, are presented.

We still see deviations from normality with the indemnity loss data, but the tails are not very extreme. In this case, the skew-normal might be a reasonably good model. The Danish fire data are now also less extreme. Given the kurtosis value of 4, the skew-student might be a reasonably good model for describing this type of data. Note that in the empirical tests we shift the distribution of the log data by  $-\min$  and add a very small number ( $1E-10$ ) to the data, since some of the distributions we analyze are defined only for values  $> 0$  (the skew-normal and skew-student are not among these distributions, but, e.g., the log-normal is).

<sup>3</sup> Tests for normality, such as, e.g., the Jarque–Bera test, are rejected at very high confidence levels. See Jarque and Bera (1987) for the test.

<sup>4</sup> One reason why the log of the Danish data is skewed is that the original data are truncated at 1, resulting in a minimum value of 0 for the log data. The US data are not truncated at 1 so that negative values for the log data are possible. The data in Fig. 3 is shifted which is the reason why there are no negative values in Fig. 3 (see next paragraph for an explanation).

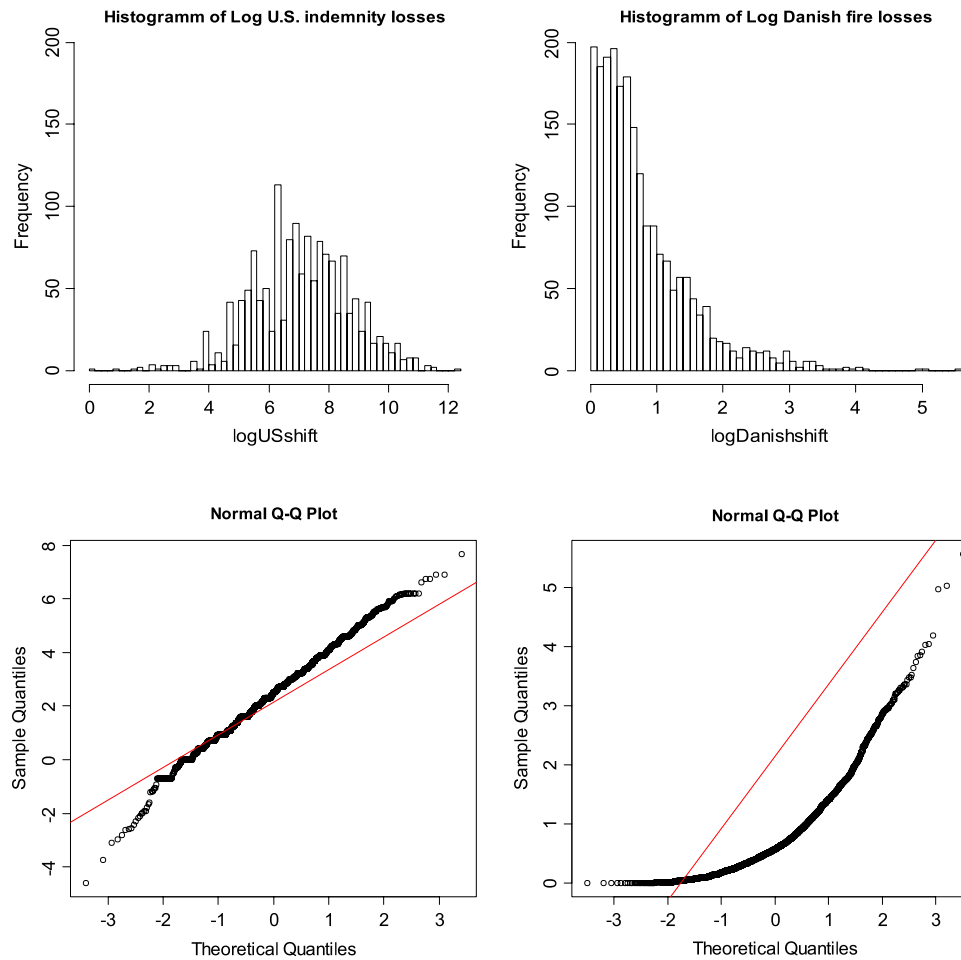


Fig. 3. Logarithm of US indemnity losses (left) and Danish fire losses (right).

**Table 2**  
Descriptive statistics for log data.

	US indemnity losses	Danish fire losses
No. of observations	1500	2167
$E(X)$	2.47	0.79
St. Dev ( $X$ )	1.64	0.72
Skewness ( $X$ )	−0.15	1.76
Kurtosis ( $X$ )	0.32	4.18
Minimum	−4.61	0.00
Maximum	7.68	5.57
99% Quantile (value at risk)	6.16	3.26
$E(X X \geq \text{value at risk})$	6.50	3.82

#### 4. Results

In this section we first estimate the parameters of the skew-normal and skew-student distributions and analyze their properties for the two empirical datasets introduced in Section 3. All parameters of the distributions are estimated based on maximum likelihood estimation.<sup>5</sup> A comparison of the models (i.e., distributions) is made based on the Akaike information criteria (AIC).<sup>6</sup> The preferred model is the one with the lowest AIC value.

<sup>5</sup> For some of the distributions, e.g., the skew-normal, an expectation maximization algorithm is also implemented in R, but we rely on the more widespread standard maximum likelihood algorithm that is available for all distributions considered in this paper.

<sup>6</sup> The AIC is calculated as  $-2 \cdot \log \text{likelihood} + 2 \cdot K$ , with  $K$  as the number of estimated parameters. Results for other criteria, such as the BIC, are available upon

request. While the AIC might be used to compare models, it could be the case that all the models are very bad at describing the data; to test for this possibility, we use a Kolmogorov–Smirnov goodness-of-fit test.<sup>7</sup> This test analyzes whether the theoretical distributions fit the empirical data reasonably well. Moreover, the results of the Kolmogorov–Smirnov test can also be used to compare the skew-normal and skew-student models with the benchmark models. Our result is that both the skew-normal and skew-student are competitive compared to other distributions in widespread use. Finally, we calculate value at risk and tail value at risk using the estimated parameters and compare the estimation results with the empirical values for value at risk and tail value at risk. All tests presented in this section were conducted with the R packages *sn* (for skew-normal and skew-student) or *ghyp* and *MASS* (for the benchmark distributions considered in the goodness-of-fit context). For the transformation kernel approach we refer to Pitt et al. (2011).

Table 3 presents the estimated parameters for the skew-normal and skew-student distributions. The parameters estimated for other distributions considered below (see Table 4) are not set out here, but the results are available from the authors upon request.

request. Consideration of other criteria that vary, e.g., in the penalty imposed for the number of parameters involved in the estimation, is important since using these criteria can lead to different results. In our case, however, the relative evaluation is not affected by the choice of criteria.

<sup>7</sup> Results for other tests, such as Andersen–Darling, are available upon request. Again, other tests generally give the same results.



**Table 3**

Estimated parameters for the skew-normal and skew-student distributions.

Model		Original data		Log data	
		US indemnity	Danish fire	US indemnity	Danish fire
Skew-normal	Location	−0.1872299	0.9721663	8.133202	−0.001315211
	Scale	110.6819405	8.8584110	1.951962	1.126483
	Shape	1533.7683374	1533.7683374	−0.933825	1533.768
Skew-student	Location	0.009959833	0.9999994	7.8820586	−0.0000001
	Scale	10.74798	0.8240933	1.7715133	0.8191495
	Shape	3291725	71916290	−0.6769532	5567005
	Degrees of freedom	0.8616305	1.100162	33.7823242	4.600740

**Table 4**

Log-likelihood and AIC for 19 distributions and the transformation kernel.

Model	Symmetric # of parameters		R package	Log-likelihood				AIC			
				Original data		Log data		Original data		Log data	
				US indemnity	Danish fire	US indemnity	Danish fire	US indemnity	Danish fire	US indemnity	Danish fire
Skew-normal	False	3	sn	−8149.49	−6295.59	−2865.90	−1716.75	16304.98	12597.18	<b>5737.79</b>	3439.49
Skew-student	False	4	sn	−6594.89	<b>−3337.51</b>	<b>−2865.03</b>	<b>−1633.11</b>	13197.78	<b>6683.03</b>	5738.06	<b>3274.22</b>
Exponential	False	1	MASS	−7077.96	−4809.40	−4433.98	−1647.81	14157.93	9620.79	8869.95	3297.61
Geometric	False	1	MASS	−7078.39	−4876.11	−4446.54	−2212.29	14158.78	9754.21	8895.08	4426.57
Chi-square	False	1	MASS	−25814.4	−4802.55	−3551.81	−1909.12	51630.80	9607.10	7105.62	3820.24
Cauchy	True	2	MASS	−7257.03	−4118.09	−3130.22	−2292.69	14518.07	8240.17	6264.44	4589.38
Gamma	False	2	MASS	−6766.61	−4767.10	−3219.11	−1647.81	13537.22	9538.19	6442.22	3299.61
Gauss	True	2	ghyp	−9076.32	−7713.76	−2868.22	−2352.58	18156.65	15431.52	5740.44	4709.15
Logistic	True	2	MASS	−8270.46	−5737.86	−2874.96	−2208.59	16544.91	11479.71	5753.92	4421.17
Log-normal	False	2	MASS	<b>−6566.77</b>	−4057.90	−4445.56	−2750.31	<b>13137.53</b>	8119.79	8895.12	5504.62
<i>t</i> (non-central)	True	2	MASS	−7243.32	−4115.93	−2866.06	−2146.95	14492.64	8237.85	5738.12	4299.90
Weibull	False	2	MASS	−6658.85	−4803.62	−2959.98	−1645.13	13321.70	9611.24	5923.95	3294.27
Hyperbolic	True	3	ghyp	−7939.19	−5213.58	−2866.21	−2178.95	15884.38	10433.17	5738.41	4363.90
Normal inverse Gaussian	True	3	ghyp	−7254.88	−4115.81	−2866.19	−2148.97	14515.76	8237.61	5738.38	4303.93
Variance gamma	True	3	ghyp	−7127.77	−4541.84	−2866.33	−2184.59	14261.53	9089.69	5738.65	4375.17
Generalized hyperbolic	True	4	ghyp	−7243.32	−4108.33	−2866.22	−2145.10	14494.64	8224.65	5740.43	4298.21
Hyperbolic	False	4	ghyp	−7077.62	−4050.64	−2865.08	−1649.92	14163.24	8109.27	5738.16	3307.83
Normal inverse Gaussian	False	4	ghyp	−6570.33	−3399.40	−2865.06	−1685.07	13148.66	6806.79	5738.12	3378.14
Variance gamma	False	4	ghyp	−7084.73	−3698.04	−2865.30	−1636.53	14177.46	7404.07	5738.61	3281.06
Generalized hyperbolic	False	5	ghyp	−6567.96	−3382.93	−2865.30	−1636.53	13145.91	6775.85	5740.61	3283.06
Transformation kernel	False	/	/	<b>−6564.90</b>	−3362.60	<b>−2854.60</b>	−1653.60	/	/	/	/

The estimation results for the skew-normal lead to a skewness value close to 1 for the US indemnity losses as well as for the Danish fire claims. As mentioned, the skew-normal model can take values of skewness from  $-1$  to  $1$ . The model's skewness values thus confirm the right skew of the empirical data, but the skewness values the model can take are less extreme. This might be seen as a limitation of the skew-normal model compared to other skewed distributions.

Fitting the skew-student model to the original data is not without difficulty, since the values for the degrees of freedom are very low. We observe a value of 0.86 for the US data and a value of 1.10 for the Danish data. Both values imply that the moments of these distributions do not exist, e.g., at least four degrees of freedom are needed so that the first four moments exist. This implies that when these fitted models are involved in a simulation study, the resulting random numbers will not lead to stable results for mean, standard deviation, skewness, or kurtosis. Moreover, the tails of the simulated random numbers are not stable and thus no reasonable values can be calculated for value at risk or tail value at risk.

There are two options for solving this problem. The first is to reconsider the maximum likelihood estimation with a fixed degree of freedom parameter of, e.g., four. This would solve the computational problems, but lead to a decrease in goodness of fit. The second option is to use the logarithm of the data in the maximum likelihood procedure (as shown in Columns 3 and 4 of Table 3, there are more degrees of freedom when considering the logarithm of the data, i.e., 33.8 for the US data and 4.6

for the Danish data). Both approaches are taken in this paper, i.e., in calculating the value at risk, we use a modified maximum likelihood estimation with four degrees of freedom, and we also undertake an application employing the logarithm of the data.

Table 4 presents a model comparison based on the log likelihood and AIC. In addition to the skew-normal and the skew-student, 18 other distributions implemented in the R packages ghyp and MASS are presented, including the normal distribution, the classical student *t*, the hyperbolic (hyp), the generalized hyperbolic (ghyp), the normal inverse Gaussian (NIG), and the variance gamma (VG). Implementation in the R package ghyp allows both a symmetric and an asymmetric implementation of all these models, denoted in the table by the column "Symmetric".<sup>8</sup> In Table 4, we first present the results for the skew-normal and the skew-student, as they are the focus of this study. The benchmark models are then sorted, first, according to the

<sup>8</sup> Note that some of the distributions implemented in MASS are not used. For beta and *f* reasonable starting values are needed and the estimation for the discrete Poisson and negative binomial distribution did not lead to any results and are not eligible for this type of analysis. In this analysis, we consider only individual claim sizes. If we aggregated this type of data into a portfolio of losses with claim number and individual claim sizes, the Poisson and binomial distribution could be used to estimate the claim number. There is also an implementation of an asymmetric student *t* in MASS that we do not use since our focus is on the skew-student as defined by Azzalini and Capitanio (2003). Additional tests with the asymmetric student *t* implemented in MASS, however, show, that this skewed distribution is also very promising in a goodness-of-fit context.

**Table 5**  
Kolmogorov–Smirnov goodness-of-fit test for selected distributions.

Model	Original data		Log data	
	US indemnity	Danish fire	US indemnity	Danish fire
Critical value	0.0351	0.0292	0.0351	0.0292
Skew-normal	0.5159	0.5798	<b>0.0281</b>	0.1223
Skew-student	0.0556	<b>0.0200</b>	<b>0.0308</b>	<b>0.0237</b>
Chi-square	0.3112	0.1917	0.2151	0.1442
Gamma	0.1335	0.2019	0.1017	0.0563
Gauss	0.3442	0.3896	<b>0.0265</b>	0.1375
Exponential	0.2648	0.2558	0.4313	0.0565
Log-normal	<b>0.0265</b>	0.1375	0.2502	0.1711
Logistic	0.3043	0.2999	0.0397	0.1345
Weibull	0.0769	0.2733	0.0462	0.0429
Transformation kernel	<b>0.0267</b>	<b>0.0251</b>	<b>0.0303</b>	<b>0.0244</b>

number of parameters involved and, second, in the alphabetic order.<sup>9/10</sup> Other estimation results for all distributions or other estimation criteria, such as BIC, are available upon request. Our findings hold for these other estimation criteria as well. The results for the transformation kernel are presented in the last line.

The results in Table 4 show that the skew-normal and skew-student are reasonably competitive models. Considering AIC and the original data, the skew-student has the fourth best goodness of fit for the US indemnity data (the best is the log-normal) and is actually the best model for the Danish fire data. In the case of the log data, the differences in AIC are not as extreme as they are for the original data. Here, the skew-student again provides very good values for AIC. It ranks as number two for the log of the US indemnity data and, again, as number 1 for the log of the Danish fire data. For the original dataset the skew-normal does not perform especially well, which confirms the empirical results from Bolancé et al. (2008) for auto insurance. For the log data, however, the skew-normal performs better. For example, for the log of the US indemnity data, the skew-normal provides the best AIC value. Other popular distributions such as the gamma and log-normal also perform well, especially for the US indemnity data. The Chi-square does not perform well for the US indemnity data since it cannot well track very high numerical values. For the other data, this distribution fits the data better.

We also compare our results with the transformation kernel approach described in Bolancé et al. (2008). The log likelihood value for the transformation kernel is  $-6564.90$  and  $-3362.60$  for the US indemnity and the Danish fire data. This makes the transformation kernel the best and second best approach when compared with the parametric distributions in terms of the log likelihood value. Also with the log data, the transformation kernel approach performs extremely well which again confirms the results presented by Bolancé et al. (2008) for auto insurance.

Overall, the skew-normal and skew-student thus appear to be competitive with the benchmark models presented in Table 4. It might thus be promising to consider both distributions when modeling insurance claims. However, one question remains unanswered: We compared the log likelihood and AIC, but it could be the case that all the models are very bad at describing the empirical data considered here. To discover whether the models are good at describing the empirical data, we perform a Kolmogorov–Smirnov goodness-of-fit test. The critical value at the 95% confidence level can be calculated by  $1.36/n^{0.5}$ , with  $n$  as the number of observations. If the test statistic of the Kolmogorov–Smirnov goodness-of-fit test remains below that level, we cannot reject the hypothesis that the empirical distribution belongs to this type of distribution. The distributions that meet these criteria are shown in bold in Table 5. In Table 5, we restrict our presentation to distributions that perform reasonably well in Table 4, i.e., have low AIC values. Results for the other distributions are available upon request. The results for the transformation kernel approach are again presented in the last line.

Table 5 reveals that the skew-student is a good model for describing the Danish fire losses, both considering the original data as well as the log data. For the indemnity data in its original version only, the log-normal provides a reasonable goodness of fit. The test value for the skew-student (0.0556) is slightly above the critical value of 0.0350 so that at a 95% confidence level we need to reject the hypothesis that the original data belong to the skew-student distribution. However, looking at the log data, we again see that the skew-student is a good model for describing these data. Moreover, the skew-normal is a very good model in this context. Overall, the test results are thus highly correlated with the AIC results and confirm the ability of the skew-student and skew-normal distributions to describe insurance claims for the data at hand.

Finally, in Table 6 we use the model results to derive estimators for value at risk and tail value at risk and compare them with the empirical data. In Table 6, only values for a confidence level of 99% are presented, but Fig. 4 illustrates the risk measurement results for varying confidence levels. Note that for the original data and the skew-student model, modified maximum likelihood estimators are used with a fixed degree of freedom parameter of four in order to stably determine the risk measures.<sup>11</sup> For the log data, the estimators presented above are used.

The results presented in Table 6 were generated using simulated loss data. Closed-form solutions are available for some of the distributions considered (e.g., the normal and the value at

<sup>9</sup> In general, the higher the number of parameters, the higher the freedom in fitting the empirical data to the theoretical distribution models. While the log-likelihood does not consider this advantage for the distributions that have many parameters, the AIC controls for this aspect because it penalizes a higher number of parameters involved in the estimation. The comparison should thus be based on the AIC criteria. The log-likelihood value is needed to calculate the AIC value, which is why we present it in Table 4.

<sup>10</sup> An alternative systematization popular in actuarial risk theory is by whether the distributions are light tailed or fat tailed. The skew-student, Cauchy, log-normal, student, and Weibull distributions are fat tailed; all other distributions in Table 4 are light tailed. See Mikosch (2009). Note that the term fat tails is used differently in actuarial science than it is in finance. In actuarial theory, the term fat tails refers to the property of a theoretical distribution. In the finance literature, the term is often used when an empirical distribution exhibits a sample kurtosis that is higher than the kurtosis of a normal distribution with sample mean as the expectation parameter and sample variance as the variance parameter since the empirical pdf has then more probability mass in the tail (see, e.g., Kon, 1984).

<sup>11</sup> For the US indemnity data, the estimated values are location = 0.009964385, scale = 28.10352, and shape = 5.342925. The log-likelihood is  $-6987.46$ . For the Danish fire data, the estimated values are location = 0.9999992, scale = 1.550688, and shape = 25.653250. The log-likelihood is  $-3726.81$ .

**Table 6**

Value at risk and tail value at risk at 99% confidence level.

Model	Original data		Log data	
	US indemnity	Danish fire	US indemnity	Danish fire
Value at risk				
Skew-normal	285.31	23.78	10.75	2.90
Skew-student	10.83	8.14	10.83	3.45
Chi-square	27.34	11.57	19.65	7.25
Gamma	217.74	13.70	12.98	3.62
Gauss	280.18	23.13	10.88	2.46
Exponential	189.25	15.66	32.52	3.62
Log-normal	530.68	11.60	34.08	40.81
Logistic	155.53	9.67	11.36	2.36
Weibull	297.68	16.19	10.69	3.47
Empirical	475.05	26.04	10.77	3.26
Tail Value at risk				
Skew-normal	320.32	26.62	11.27	3.26
Skew-student	11.44	10.78	11.42	4.55
Chi-square	30.36	13.73	22.25	9.11
Gamma	345.05	16.41	14.13	4.40
Gauss	315.03	25.98	11.43	2.70
Exponential	230.66	19.02	39.66	4.41
Log-normal	1112.03	15.11	44.01	103.32
Logistic	184.50	11.26	12.32	2.73
Weibull	411.86	19.82	11.12	4.19
Empirical	739.61	58.59	11.10	3.82

risk), but not for all of them. It is beyond the scope of this paper to develop formulas for value at risk and tail value at risk for the other distributions so we rely on simulation. In all cases, we consider 1 million random numbers. The results are fairly stable (the convergence of the simulated means, standards deviations, and risk measures was checked). Note also that the empirical values presented in Table 6 correspond to the estimates presented in Tables 1 and 2 (the values for the log of the US indemnity data are 4.61 higher, since the data is shifted; see Section 3).

In general, value at risk and tail value at risk do not perform very well when the original data are considered; the estimators derived using the theoretical distributions are in general much lower than the empirical values. For example, and as expected, the skew-student does not perform very well with the original data, since in this case the degrees of freedom need to be fixed. The other distributions also show large variation in results. The results for value at risk and tail value at risk look better when the log data are considered; the risk estimators derived using the theoretical distributions are very close to the empirical values. For example, with the skew-student, the theoretical estimators for both value at risk and tail value at risk are relatively close to the empirical estimator and thus fit the data quite well. When considering value at risk, the difference between the skew-student model and the empirical estimator is only 0.55% (with the log of the US indemnity data) and 5.9% (with the log of the Danish fire data). With tail value at risk and the log data, the skew-normal is, again, also a good model. The log-normal distribution provides very high numbers for the risk measures, which emphasizes that it is more extreme in the tails than the other distributions considered here.

Fig. 4 presents value at risk and tail value at risk for confidence levels between 90% and 99.9%. The dashed light-colored line presents the empirically observed values for value at risk and tail value at risk, the dashed dark-colored line the corresponding estimators for skew-student, the light-colored continuous line the skew-normal, and the dark-colored continuous line the normal.

As already indicated by the AIC and goodness-of-fit tests, the skew-normal and skew-student, and also the normal distribution approximate the log of the US indemnity data reasonably well; all three distributions are quite close to the empirical distribution (Fig. 4(a)). For the Danish fire data and value at risk (Fig. 4(b)),

the skew-student slightly overestimates the empirical distribution in the right tail, while the skew-normal slightly underestimates the empirical distribution. Both perform better than the normal distribution, which more severely underestimates the empirical risk. Considering tail value at risk (Fig. 4(c)) confirms that the skew-normal is the best distribution for approximating the log US indemnity data. Fig. 4(d) yields the same conclusion as Fig. 4(b): the skew-student slightly overestimates the empirical distribution, the skew-normal slightly underestimates it, and both perform better than the normal distribution, which severely underestimates the empirical risk.

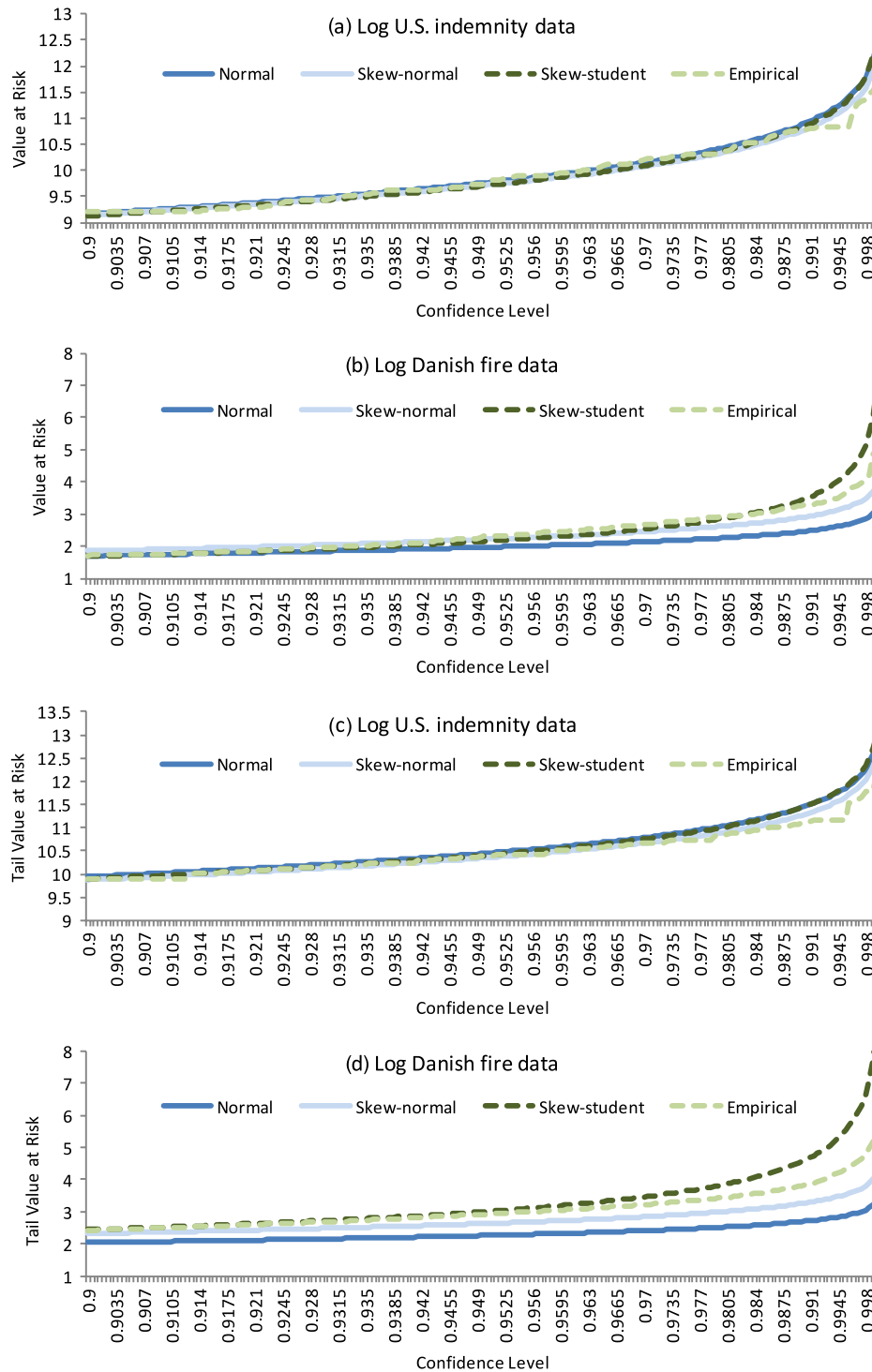
Overall, the skew-student and skew-normal are reasonably good models compared to other models in the literature. Given, however, that this paper presents only some first tests of claims modeling in insurance using two well-known datasets, we call for more applications of these distributions in the field of insurance in order to more closely analyze whether the skew-normal and skew-student are promising distributions for modeling claims.

## 5. Conclusion

The aim of this work is to fit two standard datasets of insurance claims to skewed distributions used in recent finance literature (Adcock, 2007, 2010; De Luca et al., 2006). The motivation for conducting this study is to discover whether these models are also appropriate for describing insurance claims data. Claims data in non-life insurance are very skewed and exhibit high kurtosis. For this reason, the skew-normal and skew-student might be promising candidates for both theoretical and empirical work in actuarial science.

The main finding from the empirical section of the paper is that the skew-normal and skew-student are reasonably good models compared to 18 benchmark distributions. However, as we consider only two datasets, much work remains to be done. Other than the work by Bolancé et al. (2008), we are not aware of any studies that use the skew-normal and related distributions to describe claims data. Given the flexibility, interpretability, and tractability of the skew-normal and skew-student models, we believe that these models should be closer considered in actuarial science. More applications of the skewed models are thus needed.





**Fig. 4.** Value at risk and tail value at risk for varying confidence levels.

Another aspect worth emphasizing is that financial literature shows that the skew-normal and skew-student models lead to important theoretical results, especially in the field of portfolio selection and asset pricing. For example, Stein's Lemma, which is very useful in portfolio selection, has been extended from the normal distribution to the skew-normal and the skew-student (see [Adcock, 2007](#), [Adcock, 2010](#)). It thus might be that the skew-normal model is also a promising candidate for theoretical work in actuarial science, e.g., in analyzing a portfolio of insurance contracts, i.e., the individual and collective risk model widespread in actuarial science.

This paper presents only a first analysis of the use of the skew-normal and skew-student in insurance claims modeling, leaving open many possibilities for extension in various research directions. First, more applications are needed in order to justify the use of the skew-normal and skew-student in insurance claims modeling. In this paper, we restrict the empirical analysis to two well-known datasets that are publicly available so that all results can be easily verified and replicated. Other applications could use company-specific data such as, for example, the data used in [Bolancé et al. \(2008\)](#). In this context, it might be interesting to compare the parametric estimators considered in this work with

other type of non-parametric estimators such as kernel estimators. It might be interesting to compare different lines of insurance, for example, those with and without catastrophe losses. Furthermore, other modeling approaches could be used. A modeling approach very popular in actuarial science, especially for extreme data, is the extreme value theory. In this type of analysis, different models are used to evaluate the high number of small claims and the low number of extremely high claims (see, e.g., McNeil et al., 2005). This approach allows more emphasis on the goodness of fit in the tail of the distribution, which is of high importance in actuarial practice since these determine the most costly events.

## References

- Adcock, C.J., 2007. Extensions of Stein's lemma for the skew-normal distribution. *Communications in Statistics—Theory and Methods* 36, 1661–1671.
- Adcock, C.J., 2010. Asset pricing and portfolio selection based on the multivariate extended skew-student- $t$  distribution. *Annals of Operations Research* 176, 221–234.
- Azzalini, A., 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171–178.
- Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew- $t$  distribution. *Journal of the Royal Statistical Society: Series B* 65, 367–389.
- Barndorff-Nielsen, O.E., 1997. Normal inverse Gaussian distributions and the modelling of stock returns. *Scandinavian Journal of Statistics* 24, 1–13.
- Bolancé, C., Guillen, M., Nielsen, J.P., 2003. Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics* 32 (1), 19–36.
- Bolancé, C., Guillen, M., Pelican, E., Vernic, R., 2008. Skewed bivariate models and nonparametric estimation for the CTE risk measure. *Insurance: Mathematics and Economics* 43 (3), 386–393.
- Branco, M.D., Dey, D.K., 2001. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* 79 (1), 99–113.
- Cooray, K., Ananda, M.M.A., 2005. Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal* 2005 (5), 321–334.
- Dell'Aquila, R., Embrechts, P., 2006. Extremes and robustness: a contradiction? *Financial Markets and Portfolio Management* 20 (1), 103–118.
- De Luca, G., Genton, M.G., Loperfido, N., 2006. A multivariate skew-GARCH model. *Advances in Econometrics* 20, 33–57.
- Dupuis, D.J., Jones, B.L., 2006. Multivariate extreme value theory and its usefulness in understanding risk. *North American Actuarial Journal* 10 (4), 1–27.
- Eberlein, E., Keller, U., Prause, K., 1998. New insights into smile, mispricing and value at risk: the hyperbolic model. *Journal of Business* 71, 371–405.
- Eling, M., Farinelli, S., Rossello, D., Tibiletti, L., 2010. Tail risk in hedge funds: classical skewness coefficients vs Azzalini's skewness parameter. *International Journal of Managerial Finance* 6, 290–304.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events for Insurance and Finance*. Springer.
- Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. In: Dempster, M.A.H. (Ed.), *Risk Management: Value at Risk and Beyond*. Cambridge University Press, Cambridge, pp. 176–223.
- Frees, E., Valdez, E., 1998. Understanding relationships using copulas. *North American Actuarial Journal* 2, 1–25.
- Genton, M.G., 2004. *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality*. Chapman & Hall, CRC.
- Jarque, C.M., Bera, A.K., 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55, 163–172.
- Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M., 2009. *Modern Actuarial Risk Theory*. Springer.
- Klugman, S.A., Parsa, R., 1999. Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics* 24, 139–148.
- Kole, E., Koedijk, K., Verbeek, M., 2007. Selecting copulas for risk management. *Journal of Banking & Finance* 31, 2405–2423.
- Kon, S.J., 1984. Models of stock returns—a comparison. *Journal of Finance* 39, 147–165.
- Lane, M.N., 2000. Pricing risk transfer transactions. *ASTIN Bulletin* 30 (2), 259–293.
- Mack, T., 2002. *Schadenversicherungsmathematik*. VVW-Verlag.
- McNeil, A., 1997. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* 27, 117–137.
- McNeil, A., Frey, R., Embrechts, P., 2005. *Quantitative Risk Management*. Princeton University Press.
- McNeil, A., Saladin, T., 1997. The peaks over thresholds method for estimating high quantiles of loss distributions. In: *Proceedings of 28th International ASTIN Colloquium*, pp. 23–43.
- Mikosch, T., 2009. *Non-Life Insurance Mathematics*. Springer.
- Panjer, H.H., 2007. *Operational Risk*. John Wiley & Sons.
- Pitt, D., Guillen, M., Bolancé, C., 2011. Estimation of parametric and nonparametric models for univariate claim severity distributions—an approach using R. Working Paper. Available at: <http://ssrn.com/abstract=1856982>.
- Pourahmadi, M., 2007. Skew-normal ARMA models with nonlinear heteroscedastic predictors. *Communications in Statistics: Theory and Methods* 36, 1803–1819.
- Resnick, S.I., 1997. Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin* 27 (1), 139–151.
- Vernic, R., 2006. Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics* 38, 413–426.