Folded and log-folded-*t* distributions as models for insurance loss data

Vytaras Brazauskas & Andreas Kleefeld

Published online: 04 Jan 2010.

Submit your article to this journal ↗

Article views: 270

View related articles ↗

Citing articles: 10 View citing articles ↗

Taylor & Francis
Taylor & Francis Group

Original Article

Folded and log-folded-*t* distributions as models for insurance loss data

VYTARAS BRAZAUSKAS and ANDREAS KLEEFELD

Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA

A rich variety of probability distributions has been proposed in the actuarial literature for fitting of insurance loss data. Examples include: lognormal, log-*t*, various versions of Pareto, loglogistic, Weibull, gamma and its variants, and generalized beta of the second kind distributions, among others. In this paper, we supplement the literature by adding the log-folded-normal and log-folded-*t* families. Shapes of the density function and key distributional properties of the 'folded' distributions are presented along with three methods for the estimation of parameters: method of maximum likelihood; method of moments; and method of trimmed moments. Further, large and small-sample properties of these estimators are studied in detail. Finally, we fit the newly proposed distributions to data which represent the total damage done by 827 fires in Norway for the year 1988. The fitted models are then employed in a few quantitative risk management examples, where point and interval estimates for several value-at-risk measures are calculated.

Keywords: Claim severity; Goodness-of-fit; Robust model fitting; Simulations; Value-at-risk

1. Introduction

Fitting of loss models is a necessary first step in many insurance applications such as premium calculations, risk evaluations, and determination of required reserves. A rich variety of probability distributions has been proposed in the actuarial literature for fitting of insurance loss data. Examples include: lognormal, log-*t*, various versions of Pareto, loglogistic, Weibull, gamma and its variants, and transformed beta (generalized beta of the second kind, GB2) distributions, among others (see Klugman *et al.* (2004, Appendix A)). The GB2 family has four parameters, is extremely flexible and includes many of the aforementioned distributions as special or limiting cases. It has been used in non-life insurance (Cummins *et al.* (1990)) and recently in modeling of longitudinal data involving copulas (Sun *et al.* (2008)). While flexibility of a parametric distribution is a desirable feature, it comes at a price. In particular, multi-parameter distributions can present serious computational challenges for parameter estimation, model diagnostics, and for further statistical inference which are necessary in applications. This prompted researchers

---

Corresponding author. E-mail: vytaras@uwm.edu

to pursue simpler distributions for modeling insurance losses (see e.g. composite lognormal-Pareto models of Cooray & Ananda (2005), Scollnik (2007)).

In this paper, we supplement the literature by adding the log-folded-normal and log-folded-t families. The guiding principles for introduction of these new distributions are: mathematical tractability; diagnostic transparency; and practical applicability. The mathematical tractability of these families comes from the fact that they are closely related to two well-understood distributions, normal and t, which in turn implies that they can be transformed into a location–scale family. Model diagnostic tools, such as quantile–quantile type plots, for location–scale families are transparent and especially effective. Further, practical applicability of the log-folded-normal and log-folded-t distributions follows from the observation that virtually all insurance contracts have known lower limit (e.g. deductible or retention level), and we always have a choice of how to treat it for particular data set. One of these choices naturally leads to a folded bell-shaped curve, i.e. the log-folded-normal distribution (see Example 1). Finally, we note that the families we propose are special cases of log-skew-normal and log-skew-t distributions (when the skew parameter α approaches $+\infty$) which have been successfully used for modeling income data in economics (see Azzalini *et al.* (2002)).

EXAMPLE 1.1 *For all numerical and graphical illustrations throughout the paper, we use the Norwegian fire claims data which is taken from Beirlant et al. (1996). The data set has been studied in the actuarial literature, and it represents the total damage done by n = 827 fires in Norway for the year 1988, which exceed 500 thousand Norwegian krones. For this data set, the histogram of the raw observations is not very informative since about 90% of the losses are between 500 and 3000 and the two largest claims (150,597 and 465,365) are much larger than the others. That is, one claim visually suppresses 750 claims into about 5% of the scale on a graph. Therefore, we first take the logarithmic transformation of the data and then make its histogram. Here, however, we have two possibilities. If we treat the lower limit of 500 as a location parameter and subtract it from all losses, then the histogram of the transformed data looks* approximately *bell shaped (see the left panel of Figure 1). This implies that the shifted original losses can be assumed as* roughly *lognormally distributed. Such an approach was taken by Brazauskas (2009) and we will compare his models with the new ones in our numerical illustrations of Section 4. On the other hand, if we divide all losses by 500, take the logarithmic transformation of them and make a histogram, then we observe a half of bell-shaped curve (see the right panel of Figure 1). Subsequently, the rescaled original losses can be assumed as (roughly) log-folded-normally distributed.*

The rest of the article is organized as follows. In Section 2, we provide key distributional properties of the 'folded' distributions and graphically examine shape changes of their density functions. In the next section, we study issues related to model fitting. Specifically, three methods for the estimation of parameters – method of maximum likelihood, method of moments, and method of trimmed moments (MTM) – are presented, and large and small-sample properties of these estimators are investigated in detail. In Section 4, we fit the newly proposed distributions to the Norwegian fire claims data. The fitted models are then employed in a few quantitative risk management examples, where point and interval
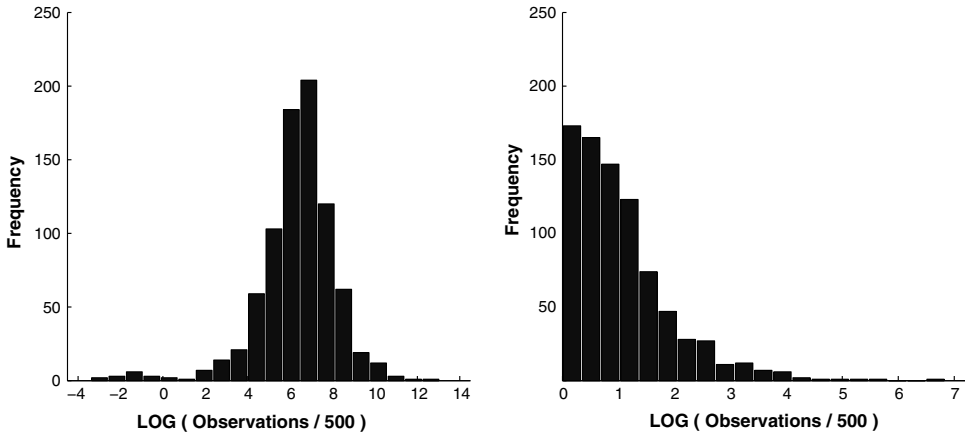
Figure 1. Preliminary diagnostics for the Norwegian fire claims (1988) data.

estimates for several value-at-risk (VaR) measures are calculated. Results are summarized and conclusions are drawn in Section 5.

2. Folded-*t* and related distributions

As it is well-known, the probability density function (pdf) of a scaled *t*-distribution is given by

$$f_{T(v)}(x|\sigma) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sigma\sqrt{v\pi}} \frac{1}{(1 + \frac{1}{v}(x/\sigma)^2)^{(v+1)/2}}, \qquad -\infty < x < \infty, \qquad (2.1)$$

where $\sigma > 0$ is the scale parameter and $v = 1, 2, 3, \ldots$ represents the degrees of freedom. Notice two special cases: for $v = 1$, expression (2.1) reduces to the pdf of Cauchy(0, σ), and for $v \to \infty$, it converges to the pdf of normal(0, σ). Moreover, after the transformation $Y = |X|$ one easily obtains the folded-*t* distribution with scale parameter $\sigma > 0$, degrees of freedom $v = 1, 2, 3, \ldots$, and the pdf

$$f_{\text{FT}(v)}(y|\sigma) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{2}{\sigma\sqrt{v\pi}} \frac{1}{(1 + \frac{1}{v}(y/\sigma)^2)^{(v+1)/2}}, \qquad y > 0. \qquad (2.2)$$

Similar to the *t*-distribution case, for $v = 1$, expression (2.2) reduces to the pdf of folded Cauchy, and for $v \to \infty$, it converges to the pdf of folded normal. In Figure 2, we illustrate shape changes of $f_{\text{FT}(v)}(y|\sigma)$ for various combinations of σ and v.

REMARK 2.1 *In most statistical problems involving hypothesis testing and estimation, the scale parameter σ and the degrees of freedom v are known. In modeling insurance losses this is no longer the case – both parameters have to be estimated from the data and the degrees of freedom v do not have to be an integer. In order to facilitate comparison of the numerical examples of Section 4.2 with the most relevant existing literature, we will estimate σ and*
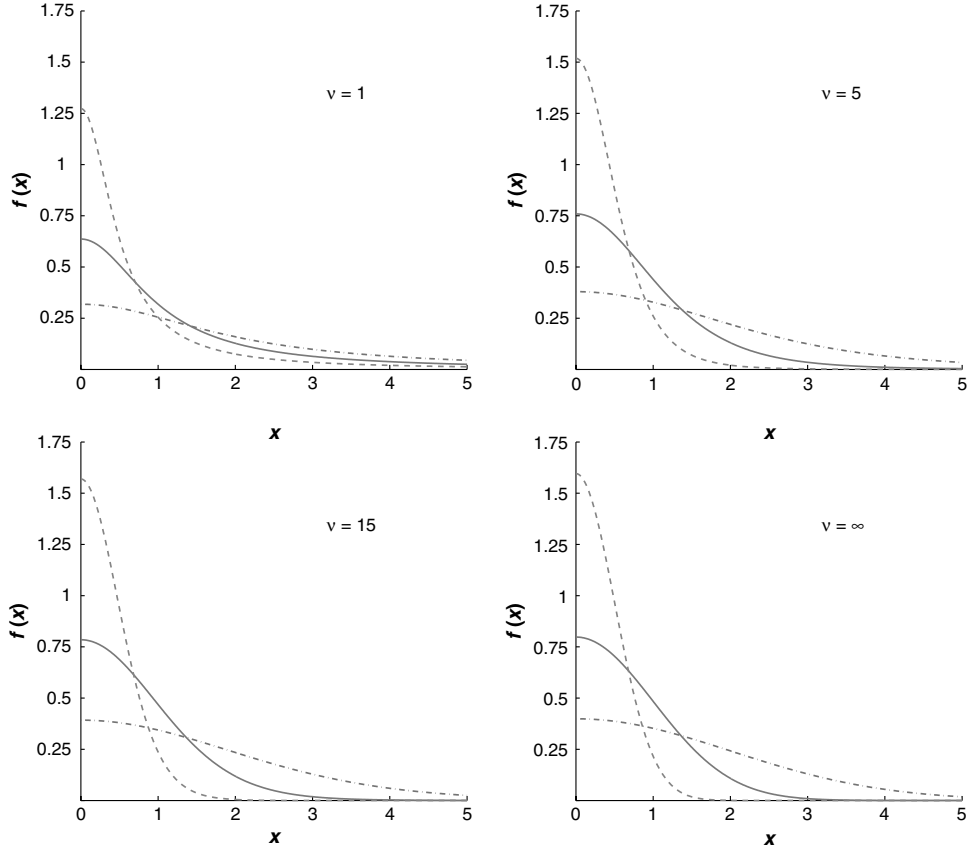
Figure 2.   Shapes of the pdf of folded-*t* distributions for σ =0.5 (dashed line), σ =1.0 (solid line), σ =2.0 (dash-dotted line), and ν =1, 5, 15, ∞.

assume that ν *is a* known integer. *A similar approach was taken by Brazauskas (2009) for fitting a* t_8 *model to the logarithm of Norwegian claims, where each claim had been shifted – not rescaled – by the deductible of 500 (see the left panel of Figure 1). For a general treatment of this problem, i.e. joint estimation of* σ *and* ν, *the reader can be referred to Johnson et al. (1995, Section 28.6).*

Further, due to the close relation between the *t*-distribution and the folded-*t*, the following properties for pdf, cumulative distribution, and quantile functions (cdf and qf, respectively) of the two distributions can be easily established:

$$\text{pdf:}\quad f_{\text{FT}(\nu)}(y|\sigma) = (1/\sigma)\, f^*_{\text{FT}(\nu)}(y/\sigma) = (2/\sigma)\, f^*_{T(\nu)}(y/\sigma), \qquad y > 0, \qquad (2.3)$$

$$\text{cdf:}\quad F_{\text{FT}(\nu)}(y|\sigma) = F^*_{\text{FT}(\nu)}(y/\sigma) = 2\,[F^*_{T(\nu)}(y/\sigma) - 0.5], \qquad y > 0, \qquad (2.4)$$

$$\text{qf:}\quad Q_{\text{FT}(\nu)}(u|\sigma) = \sigma\, Q^*_{\text{FT}(\nu)}(u) = \sigma Q^*_{T(\nu)}((u+1)/2), \qquad 0 < u < 1, \qquad (2.5)$$

where f^*, F^*, and Q^* denote the standard (i.e. with θ =0 and/or σ =1) pdf, cdf, and qf, respectively, of the underlying location–scale family. Also, the mean and variance of the folded-*t* distribution are given by

$$\mathrm{E}(Y) = \sigma\sqrt{\nu/\pi}\,\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} =: \sigma\,c_0, \qquad \nu = 2, 3, 4, \ldots, \tag{2.6}$$

$$\mathrm{Var}(Y) = \sigma^2\left(\frac{\nu}{\nu-2} - c_0^2\right), \qquad \nu = 3, 4, 5, \ldots. \tag{2.7}$$

Now a log-folded-t distribution emerges in a very natural way. That is, we will say a random variable Z is log-folded-t distributed if $\log Z$ follows a folded-t distribution. This implies that, in conjunction with Eqs. (2.3)–(2.5), the pdf, cdf, and qf of Z satisfy the following relationships:

$$\text{pdf:} \quad f_{\mathrm{LFT}(\nu)}(z|\sigma) = (2/\sigma)z^{-1}f_{T(\nu)}^*(\log(z)/\sigma), \qquad z > 1, \tag{2.8}$$

$$\text{cdf:} \quad F_{\mathrm{LFT}(\nu)}(z|\sigma) = 2[F_{T(\nu)}^*(\log(z)/\sigma) - 0.5], \qquad z > 1, \tag{2.9}$$

$$\text{qf:} \quad Q_{\mathrm{LFT}(\nu)}(u|\sigma) = \exp\{\sigma Q_{T(\nu)}^*((u+1)/2)\}, \qquad 0 < u < 1. \tag{2.10}$$

As before, it is worth noting two special cases: for $\nu = 1$ and $\nu \to \infty$, log-folded-t becomes a log-folded-Cauchy and log-folded-normal variable, respectively. Also, unlike the 'folded' families, the log-folded distributions possess no moments.

3. Parameter estimation

In this section, we study issues related to model fitting. Three methods for estimation of the scale parameter of a folded-t distribution are provided. Specifically, in Subsection 3.1, standard estimators, based on the maximum likelihood and method-of-moments approaches, are presented and their large-sample properties are examined. Then, in Subsection 3.2, we consider a recently introduced robust estimation technique, the MTM, and study asymptotic behavior of estimators based on it. Finally, Subsection 3.3 is devoted to small-sample properties of the estimators, which are investigated using simulations.

Also, throughout this section, we will consider a sample of n independent and identically distributed random variables, $Y_1, \ldots, Y_n$, from a folded-t family with its pdf, cdf, and qf given by Eqs. (2.3)–(2.5), and denote $Y_{1:n} \leq \cdots \leq Y_{n:n}$ the order statistics of $Y_1, \ldots, Y_n$.

3.1. Standard methods

A method-of-moments estimator for σ is found by matching the population mean, given by Eq. (2.6), and the sample mean $\bar{Y}$, and then solving the equation with respect to σ. This leads to

$$\hat{\sigma}_{\mathrm{MM}} = \sqrt{\pi/\nu}\,\frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu-1}{2})}\,\bar{Y} = \bar{Y}/c_0.$$

As follows from, e.g. Serfling (1980, Section 2.2), the estimator $\hat{\sigma}_{MM}$ is asymptotically normal with mean σ and variance $n^{-1}\sigma^2\,(\frac{\nu}{\nu-2} - c_0^2)/c_0^2$. To summarize this result, we shall write:

$$\hat{\sigma}_{MM} \sim \mathscr{A}\mathscr{N}\left(\sigma, \frac{\sigma^2}{n}\,\Delta_0\right), \tag{3.1}$$

where $\Delta_0 = (\frac{\nu}{\nu-2} - c_0^2)/c_0^2$ and $\mathscr{A}\mathscr{N}$ stands for 'asymptotically normal'.

The maximum likelihood estimator $\hat{\sigma}_{MLE}$ is obtained by maximizing the log-likelihood function, which is equivalent to solving

$$\sum_{i=1}^{n} \frac{\sigma^2(\nu + 1)}{Y_i^2 + \sigma^2\nu} - n = 0$$

for σ with a root finding algorithm. Using standard asymptotic results for MLEs (see e.g. Serfling (1980, Section 4.2)), one can show that

$$\hat{\sigma}_{MLE} \sim \mathscr{A}\mathscr{N}\left(\sigma, \frac{\sigma^2}{n}\,\frac{\nu + 3}{2\nu}\right). \tag{3.2}$$

In this case, the asymptotic variance of MLE is optimal, i.e. it attains the Cramér–Rao lower bound. Therefore, since both estimators are consistent and asymptotically normal, we are interested in comparing their asymptotic variances. In other words, we would like to know how much efficiency is lost due to using $\hat{\sigma}_{MM}$ instead of $\hat{\sigma}_{MLE}$. Clearly, a more efficient estimator is preferred because that has a direct impact on the accuracy of pricing and risk measuring models. In practice, however, other criteria, such as computational simplicity and estimator's robustness to model misspecification and/or data contamination by outliers, have to be considered as well.

As follows from Eqs. (3.1) and (3.2), the *asymptotic relative efficiency* (ARE) of the MM estimator with respect to the MLE, defined as the ratio of their asymptotic variances, is given by

$$\text{ARE}(\hat{\sigma}_{MM},\ \hat{\sigma}_{MLE}) = \frac{\nu + 3}{2\nu}\,\frac{1}{\Delta_0}. \tag{3.3}$$

In Table 1, we provide numerical illustrations of expression (3.3) for selected values of the degrees of freedom ν. In view of MM's minimal sacrifice of efficiency for almost all values of ν (ARE is at least 87% for $\nu \geq 4$) along with its explicit, computationally simple formula, one can argue that $\hat{\sigma}_{MM}$ is indeed a competitive alternative to the MLE. However, note that for $\nu = 1, 2$, the population variance, given by Eq. (2.7), is infinite and hence the ARE is 0.

Table 1. $\text{ARE}(\hat{\sigma}_{MM},\ \hat{\sigma}_{MLE})$ for selected values of ν.

ν	3	4	5	6	7	8	9	10	15	25	50	100	∞
ARE	0.681	0.875	0.941	0.964	0.972	0.973	0.971	0.967	0.949	0.935	0.903	0.890	0.876

3.2. Robust estimation

As presented by Brazauskas *et al.* (2009, Section 2.1), the MTM procedure is operationally equivalent to the method-of-moments approach. The difference is that for the MTM we match 'trimmed' moments rather than 'simple' moments. Thus, in order to obtain an MTM estimator of σ, we first compute a sample trimmed moment

$$\hat{\mu} = \frac{1}{n - m_n - m_n^*} \sum_{i = m_n + 1}^{n - m_n^*} Y_{i:n},$$

where m_n and m_n^* are integers $0 \le m_n < n - m_n^* \le n$ such that $m_n/n \to a$ and $m_n^*/n \to b$ when $n \to \infty$, where the trimming proportions a and b $(0 \le a + b < 1)$ are chosen by the researcher. Then, we derive the corresponding population trimmed moment

$$\mu := \mu(\sigma) = \frac{1}{1 - a - b} \int_a^{1-b} Q_{\mathrm{FT}(v)}(u | \sigma) \mathrm{d}u$$

$$= \frac{\sigma}{1 - a - b} \int_a^{1-b} Q_{T(v)}^*((u+1)/2) \mathrm{d}u =: \sigma c(a, b).$$

Equating $\hat{\mu}$ to μ and solving the equation with respect to σ yields the MTM estimator

$$\hat{\sigma}_{\mathrm{MTM}} = \hat{\mu}/c(a, b).$$

Asymptotic properties of MTM estimators are extensively studied by Brazauskas *et al.* (2009, Section 2.2 and Appendix A). Adaptation of their formulas to our case implies that

$$\hat{\sigma}_{\mathrm{MTM}} \sim \mathscr{AN}\left(\sigma, \frac{\sigma^2}{n} \Delta(a, b)\right), \tag{3.4}$$

where $\Delta(a, b) = C(a, b)/c^2(a, b)$ with

$$C(a, b) = \frac{1}{(1 - a - b)^2} \{a(1-a)[Q_{T(v)}^*((a+1)/2)]^2 + b(1-b)[Q_{T(v)}^*(1-b/2)]^2$$

$$- 2ab \, Q_{T(v)}^*((a+1)/2) \, Q_{T(v)}^*(1-b/2) - (1-a-b)^2 \, c^2(a, b) + (1-a-b)d(a,b)$$

$$- 2(1-a-b)[a \, Q_{T(v)}^*((a+1)/2) + b Q_{T(v)}^*(1-b/2)]c(a, b)\},$$

where $c(a, b)$ is defined above and $d(a, b) = (1 - a - b)^{-1} \int_a^{1-b} [Q_{T(v)}^*((u+1)/2)]^2 \mathrm{d}u$.

REMARK 3.1 *When $m_n = m_n^* = 0$, then $\hat{\mu} = \bar{Y}$ and $c(0,0) = c_0$; consequently, the MTM estimator $\hat{\sigma}_{\mathrm{MTM}}$ becomes $\hat{\sigma}_{\mathrm{MM}}$. Also note that since $\Delta(a, b) \to \Delta_0$ when $a = b \to 0$, the MM's asymptotic distribution follows from Eq. (3.4). Hence, for the folded-t distribution, the MTM can be viewed as a robustified version of MM.*

Now let us turn to the efficiency investigations. As follows from Eqs. (3.2) and (3.4), the ARE of an MTM estimator with respect to the MLE is given by

$$\mathrm{ARE}(\hat{\sigma}_{\mathrm{MTM}}, \, \hat{\sigma}_{\mathrm{MLE}}) = \frac{v + 3}{2v} \frac{1}{\Delta(a, b)}. \tag{3.5}$$

In Table 2, we provide numerical illustrations of expression (3.5) for selected values of a and b and for $v = 1, 5, 15, \infty$. Several conclusions emerge from the table. First, the MTM

procedures with $a > 0$ and $b > 0$ are valid for all values of ν, thus they expand the range of applicability of the MM estimator. Second, for a fixed ν, there is always at least one MTM estimator which is more efficient than the MM. Third, for very heavy-tailed folded-t distributions (i.e. when ν is small), it is beneficial to trim data even when there are no outliers because that improves efficiency and accuracy of estimation. In summary, while from the computational point of view MTMs are a bit more complex than MMs, they are still simpler than the MLE. They also offer various degrees of robustness against outliers. In real-data examples of Section 4, we will illustrate how to choose the trimming proportions a and b based on the data at hand.

3.3. Simulations

Here we supplement the large-sample results of Subsections 3.1 and 3.2 with finite-sample investigations. The objective is to see how large the sample size n is needed for the MLE and MTM estimators, including the MM as a special case of MTMs, to achieve

Table 2.   ARE($\hat{\sigma}_{MTM}$, $\hat{\sigma}_{MLE}$) for selected a, b, and $\nu = 1$, 5, 15, ∞, with the boxed numbers highlighting the case $a = b$.

ν	a	b								
		0	0.01	0.03	0.05	0.10	0.15	0.25	0.49	0.70
1	0	0	0.202	0.391	0.519	0.730	0.858	0.974	0.843	0.512
	0.05	0	0.202	0.390	0.518	0.729	0.857	0.974	0.847	0.522
	0.10	0	0.201	0.389	0.515	0.725	0.854	0.972	0.855	0.542
	0.15	0	0.199	0.386	0.511	0.720	0.847	0.967	0.863	0.566
	0.25	0	0.195	0.376	0.497	0.699	0.825	0.947	0.874	0.613
	0.49	0	0.175	0.330	0.432	0.602	0.709	0.824	0.821	–
	0.70	0	0.141	0.256	0.329	0.448	0.524	0.609	–	–
5	0	0.941	0.985	0.979	0.961	0.900	0.834	0.703	0.423	0.220
	0.05	0.941	0.986	0.981	0.962	0.901	0.836	0.706	0.427	0.226
	0.10	0.942	0.988	0.983	0.965	0.906	0.841	0.713	0.436	0.237
	0.15	0.942	0.990	0.986	0.969	0.912	0.848	0.721	0.448	0.251
	0.25	0.938	0.991	0.991	0.977	0.923	0.864	0.742	0.476	0.284
	0.49	0.881	0.952	0.967	0.962	0.927	0.880	0.778	0.540	–
	0.70	0.725	0.819	0.852	0.860	0.849	0.821	0.748	–	–
15	0	0.949	0.926	0.885	0.848	0.765	0.693	0.566	0.326	0.167
	0.05	0.951	0.928	0.887	0.850	0.768	0.695	0.568	0.329	0.171
	0.10	0.955	0.933	0.892	0.855	0.773	0.701	0.575	0.337	0.180
	0.15	0.962	0.939	0.899	0.862	0.781	0.709	0.584	0.347	0.191
	0.25	0.976	0.955	0.916	0.880	0.800	0.729	0.606	0.371	0.217
	0.49	0.990	0.975	0.942	0.911	0.839	0.773	0.657	0.433	–
	0.70	0.919	0.918	0.898	0.876	0.820	0.766	0.666	–	–
∞	0	0.876	0.844	0.797	0.757	0.674	0.604	0.487	0.275	0.140
	0.05	0.878	0.846	0.799	0.760	0.676	0.606	0.489	0.278	0.143
	0.10	0.884	0.852	0.805	0.765	0.682	0.612	0.495	0.284	0.150
	0.15	0.893	0.861	0.813	0.774	0.691	0.621	0.503	0.293	0.160
	0.25	0.914	0.881	0.834	0.794	0.711	0.641	0.525	0.315	0.182
	0.49	0.961	0.929	0.882	0.844	0.762	0.694	0.579	0.372	–
	0.70	0.947	0.918	0.877	0.843	0.770	0.708	0.602	–	–

(asymptotic) unbiasedness and for their finite-sample relative efficiency (RE) to reach the corresponding ARE level. The RE of an estimator is defined as the ratio of its estimated (from simulations) mean-squared error and the asymptotic variance of the MLE, which is provided by statement (3.2).

From a specified folded-t distribution we generate 10,000 samples of size n using Monte Carlo. For each sample we estimate the scale parameter σ using MLE and various MTM estimators and then compute the average mean and RE of those 10,000 estimates. This process is repeated 10 times and the 10 average means and the 10 REs are again averaged and their standard deviations are reported. (Such repetitions are useful for assessing standard errors of the estimated means and REs. Hence, our findings are essentially based on 100,000 samples.) The standardized mean that we report is defined as the average of 100,000 estimates divided by the true value of the parameter that we are estimating. The standard error is standardized in a similar fashion. The study was performed for the following choices of simulation parameters:

- Parameters of folded-t: $\sigma = 5$ and $v = 1, 5, 15, \infty$.
- Sample sizes: $n = 50, 100, 250, 500$.
- Estimators of σ:
  - MLE.
  - MM (corresponds to MTM with $a = b = 0$).
  - MTM with: $a = b = 0.05$; $a = b = 0.10$; $a = b = 0.25$; $a = b = 0.49$; $a = 0.10$ and $b = 0.70$; $a = 0.25$ and $b = 0$.

Simulation results are recorded in Tables 3 and 4. Note that the entries of the last columns of these tables are included as target quantities and follow from the theoretical results of Subsections 3.1 and 3.2, not from simulations. First of all, we observe that the most heavy-tailed case (i.e. $v = 1$) differs from the other choices of v. Specifically, as we have seen in Subsection 3.1, the MM estimator (equivalently, MTM with $a = b = 0$) for this distribution does not exist; this fact in simulations manifests itself through uncontrollable bias and RE $= 0$. Moreover, even for the theoretically well-behaved estimators such as MLE and MTM with $a > 0$ and $b > 0$, it still takes $n \geq 500$ to get the bias within 1% of the target. Likewise, convergence of REs to the corresponding AREs is slow. On the other hand, for $v \geq 5$, behavior of all estimators is predictably stable as their bias practically vanishes for $n \geq 100$ and their REs approach large-sample counterparts for $n \geq 50$. In summary, except for very heavy-tailed cases (say, for $v = 1, 2, 3$), it is safe to conclude that the asymptotic results of MLE, MM, and MTM estimators are valid for samples of size 100 or larger.

4. Real-data illustrations

In this section, we fit the log-folded-normal and log-folded-t distributions to the Norwegian fire claims data which were described and preliminary analyzed in Example 1.

Table 3. Standardized mean of MLE and various MTM estimators for ν = 1, 5, 15, ∞.

ν	Estimator	Trimming proportions		Sample size (n)				
		Lower (a)	Upper (b)	50	100	250	500	∞
1	MLE	–	–	1.02 (0.001)	1.01 (0.001)	1.00 (0)	1.00 (0)	1
	MTM	0	0	1.12 (0.074)	1.26 (0.118)	1.30 (0.175)	1.21 (0.037)	–
		0.05	0.05	1.17 (0.001)	1.04 (0.001)	1.03 (0)	1.01 (0)	1
		0.10	0.10	1.05 (0.001)	1.02 (0)	1.01 (0)	1.01 (0)	1
		0.25	0.25	1.04 (0.001)	1.01 (0)	1.01 (0)	1.00 (0)	1
		0.49	0.49	1.03 (0.001)	1.01 (0)	1.01 (0.001)	1.00 (0)	1
		0.10	0.70	1.04 (0)	1.02 (0.001)	1.01 (0)	1.00 (0)	1
		0.25	0	1.41 (0.249)	1.35 (0.123)	1.56 (0.185)	2.12 (0.620)	–
5	MLE	–	–	1.00 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
	MTM	0	0	1.00 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.05	0.05	1.02 (0)	1.01 (0)	1.00 (0)	1.00 (0)	1
		0.10	0.10	1.01 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.25	0.25	1.01 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.49	0.49	1.01 (0.001)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.10	0.70	1.03 (0.001)	1.02 (0)	1.01 (0)	1.00 (0)	1
		0.25	0	0.99 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
15	MLE	–	–	1.00 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
	MTM	0	0	1.00 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.05	0.05	1.01 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.10	0.10	1.01 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.25	0.25	1.01 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.49	0.49	1.01 (0.001)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.10	0.70	1.03 (0.001)	1.02 (0)	1.01 (0)	1.00 (0)	1
		0.25	0	0.99 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
∞	MLE	–	–	1.00 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
	MTM	0	0	1.00 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.05	0.05	1.01 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.10	0.10	1.01 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.25	0.25	1.01 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.49	0.49	1.01 (0.001)	1.00 (0)	1.00 (0)	1.00 (0)	1
		0.10	0.70	1.03 (0.001)	1.02 (0)	1.01 (0)	1.00 (0)	1
		0.25	0	0.99 (0)	1.00 (0)	1.00 (0)	1.00 (0)	1

Note: The entries are mean values (with standard errors in parentheses) based on 100,000 samples.

We also investigate the implications of a model fit on risk evaluations. In particular, we compute point estimates of, and construct confidence intervals for, a number of VaR measures.

4.1. Fitting log-folded distributions

Suppose the Norwegian fire claims are a realization of n independent and identically distributed random variables, $X_1, \ldots, X_n$, all defined above the pre-specified deductible $x_0 = 500$. In view of the preliminary diagnostics of Example 1, it is reasonable to assume that the ratios $Z_i = X_i/x_0$, $i = 1, \ldots, n$, follow a log-folded-t distribution, for which the log-folded-normal is a limiting case. That is, the pdf, cdf, and qf of $Z_1, \ldots, Z_n$ are given by expressions (8), (9), and (10), respectively.

Table 4. Relative efficiencies of MLE and various MTM estimators for $v = 1, 5, 15, \infty$.

| v | Estimator | Trimming proportions | | Sample size (n) | | | | |
		Lower (a)	Upper (b)	50	100	250	500	∞
1	MLE	–	–	0.92 (0.004)	0.97 (0.004)	0.99 (0.005)	1.00 (0.004)	1
	MTM	0	0	0 (0)	0 (0)	0 (0)	0 (0)	0
		0.05	0.05	0.19 (0.004)	0.41 (0.003)	0.44 (0.003)	0.49 (0.002)	0.518
		0.10	0.10	0.54 (0.004)	0.63 (0.004)	0.69 (0.003)	0.70 (0.004)	0.725
		0.25	0.25	0.80 (0.006)	0.89 (0.005)	0.92 (0.004)	0.93 (0.004)	0.947
		0.49	0.49	0.75 (0.004)	0.78 (0.003)	0.80 (0.003)	0.81 (0.003)	0.821
		0.10	0.70	0.50 (0.003)	0.52 (0.002)	0.54 (0.003)	0.54 (0.003)	0.542
		0.25	0	0 (0)	0 (0)	0 (0)	0 (0)	0
5	MLE	–	–	1.00 (0.006)	1.00 (0.004)	1.00 (0.006)	1.00 (0.004)	1
	MTM	0	0	0.94 (0.006)	0.94 (0.003)	0.94 (0.007)	0.94 (0.004)	0.941
		0.05	0.05	0.90 (0.003)	0.95 (0.004)	0.96 (0.004)	0.96 (0.003)	0.962
		0.10	0.10	0.90 (0.003)	0.89 (0.003)	0.90 (0.004)	0.90 (0.006)	0.906
		0.25	0.25	0.74 (0.004)	0.74 (0.003)	0.74 (0.004)	0.74 (0.003)	0.742
		0.49	0.49	0.54 (0.003)	0.54 (0.002)	0.55 (0.003)	0.54 (0.003)	0.540
		0.10	0.70	0.23 (0.001)	0.23 (0.001)	0.24 (0.001)	0.24 (0.001)	0.237
		0.25	0	0.94 (0.003)	0.94 (0.004)	0.92 (0.004)	0.94 (0.004)	0.938
15	MLE	–	–	1.00 (0.006)	1.01 (0.004)	0.99 (0.004)	1.00 (0.005)	1
	MTM	0	0	0.95 (0.005)	0.94 (0.003)	0.94 (0.003)	0.95 (0.003)	0.949
		0.05	0.05	0.83 (0.004)	0.85 (0.004)	0.84 (0.005)	0.85 (0.004)	0.850
		0.10	0.10	0.77 (0.003)	0.77 (0.004)	0.77 (0.004)	0.77 (0.004)	0.773
		0.25	0.25	0.61 (0.003)	0.61 (0.004)	0.60 (0.002)	0.61 (0.003)	0.606
		0.49	0.49	0.44 (0.002)	0.44 (0.002)	0.43 (0.002)	0.43 (0.002)	0.433
		0.10	0.70	0.17 (0.001)	0.18 (0.001)	0.18 (0.001)	0.18 (0.001)	0.180
		0.25	0	0.99 (0.003)	0.99 (0.004)	0.97 (0.003)	0.96 (0.003)	0.976
∞	MLE	–	–	1.00 (0.004)	1.01 (0.005)	0.99 (0.004)	1.00 (0.004)	1
	MTM	0	0	0.88 (0.005)	0.88 (0.004)	0.87 (0.006)	0.87 (0.003)	0.876
		0.05	0.05	0.76 (0.004)	0.76 (0.002)	0.76 (0.002)	0.76 (0.003)	0.760
		0.10	0.10	0.69 (0.004)	0.69 (0.003)	0.68 (0.004)	0.68 (0.003)	0.682
		0.25	0.25	0.53 (0.003)	0.53 (0.002)	0.52 (0.002)	0.53 (0.002)	0.525
		0.49	0.49	0.38 (0.001)	0.38 (0.002)	0.37 (0.002)	0.37 (0.002)	0.372
		0.10	0.70	0.15 (0.001)	0.15 (0)	0.15 (0.001)	0.15 (0.001)	0.150
		0.25	0	0.93 (0.004)	0.92 (0.005)	0.92 (0.004)	0.91 (0.003)	0.914

Note: The entries are mean values (with standard errors in parentheses) based on 100,000 samples.

We fit the log-folded-normal model to the data using the MTM method with $a = 0.50$, $b = 0.10$ (MTM1) and the MLE. Using the notation of this section, the estimators of σ are given by

$$\hat{\sigma}_{\mathrm{MLE}} = \left(\frac{1}{n} \sum_{i=1}^{n} \log^2(Z_i) \right)^{1/2} \quad \text{and} \quad \hat{\sigma}_{\mathrm{MTM}} = \hat{\mu}/c(a,b),$$

where $\hat{\mu} = (n - m_n - m_n^*)^{-1} \Sigma_{i=m_n+1}^{n-m_n^*} \log(Z_{i:n})$ and $c(a,b) = (1-a-b)^{-1} \int_a^{1-b} Q_{T(\infty)}^*((u+1)/2) \mathrm{d}u$. The resulting estimates are $\hat{\sigma}_{\mathrm{MLE}} = 1.37$ and $\hat{\sigma}_{\mathrm{MTM1}} = 1.24$, and the corresponding fits are illustrated in the QQP-plot of Figure 3 (left panel). The QQP stands for 'quantile–quantile–percentile'; the plot is a quantile–quantile plot equipped with an additional vertical axis that shows the percentile levels of empirical quantiles. As discussed by Brazauskas (2009), such plots, besides revealing empirical quantile's relative position within the sample, also provide guidance about the minimal trimming requirements for the MTMs. Therefore, we choose $b = 0.10$; the other trimming proportion is chosen based on
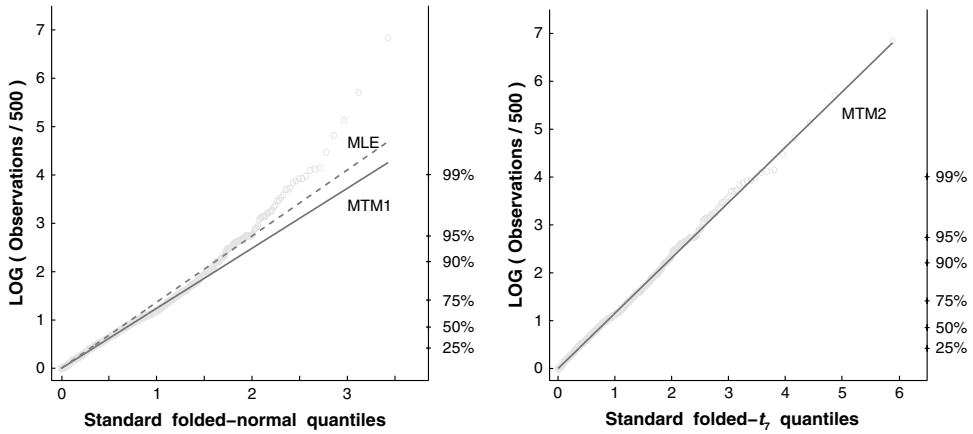
Figure 3.   Log-folded-normal (left panel) and log-folded-t_7 (right panel) QQP plots. The models are fitted using the MLE and MTM methods with $a = 0.50$, $b = 0.10$ (MTM1) and $a = 0.30$, $b = 0.01$ (MTM2). Parameter estimates: $\hat{\sigma}_{MLE} = 1.37$; $\hat{\sigma}_{MTM1} = 1.24$; and $\hat{\sigma}_{MTM2} = 1.16$. (In both graphs, the right vertical axis represents empirical percentile levels.)

the efficiency considerations (ARE of $\hat{\sigma}_{MTM1}$ is 0.764). One can clearly see from the plot that the log-folded-normal model is misspecified, which was done intentionally to demonstrate advantages of the robust MTM fit over the non-robust MLE fit. Indeed, while the MTM line is in close agreement with 80–85% of the data, the MLE line gets attracted by a few largest observations and matches well only 60–65% of the data.

The above analysis suggests that we need to modify our distributional assumption. Since the data deviate from the linear pattern in the upward direction, we have to replace the underlying normal with a fairly heavy-tailed t distribution. From the right panel of Figure 3, we see that the data set forms a nearly perfect straight line. Hence, the log-folded-t_7 distribution is appropriate for the Norwegian fire claims data. We fit the model using the MTM method with $a = 0.30$, $b = 0.01$ (MTM2). In this case, the trimming proportions are selected entirely on the efficiency considerations (ARE of $\hat{\sigma}_{MTM2}$ is 0.995). We see that the chosen MTM estimator is as accurate as MLE which was not included because of its non-explicit formula. The MTM expression is the same as above with $c(a,b)$ computed using the quantile function $Q^*_{T(7)}$ instead of $Q^*_{T(\infty)}$. The resulting estimate is $\hat{\sigma}_{MTM2} = 1.16$.

Finally, to make sure that the proposed model provides a better fit to this data set than some of its closest competitors, we fitted exponential, gamma, generalized Pareto (GPD), and Weibull distributions to the (log) data. We then performed a χ^2 goodness-of-fit test for each distribution. The test results for two specifications of data groupings are summarized in Table 5. Note that the GPD and Weibull models were fitted using the MLE approach which yielded the following estimates of the parameters: $\hat{\sigma} = 1.17$, $\hat{\gamma} = 0.12$ (GPD) and $\hat{\sigma} = 1.06$, $\hat{\tau} = 1.04$ (Weibull). The exponential and gamma model fits were even worse than those of GPD and Weibull, and thus are not included in the table. Also, for all distributions, the p-value computations were based on the χ^2_{m-k-1} statistic, where m is the number of data groups/classes and $k = 2$ is the number of estimated

Table 5. Values of the χ^2 goodness-of-fit statistic (with p values in parentheses) of the folded-t_7, GPD, and Weibull models fitted to $\log(X_1/500), \ldots, \log(X_n/500)$.

	Model		
Edges of classes used for data grouping	Folded-t_7	GPD	Weibull
{0; 0.68; 1.37; 2.05; 2.73; 3.42; 4.10; ∞}	4.767 (0.312)	19.697 (0.001)	26.036 (<0.001)
{0; 0.34; 0.68; 1.03; 1.37; 1.71; 2.05; 2.39; 2.73; 3.08; 3.42; 3.76; ∞}	8.564 (0.479)	26.985 (0.001)	32.666 (<0.001)

parameters. Other choices of data grouping consistently led to the same conclusion: the folded-t_7 fit should be accepted while the other fits should be rejected (at all *typical* significance levels $\alpha = 0.01, 0.05, 0.10$).

4.2. *Quantitative risk management*

To see how the quality of the model fit affects insurance risk evaluations, we will construct confidence intervals for a number of VaR measures. Mathematically, this measure is the $(1-\beta)$-level quantile of the distribution function G, that is, $\text{VaR}(\beta) = G^{-1}(1-\beta)$. For empirical estimation, we replace G with the empirical cdf $\hat{G}_n$; for parametric (MLE) and robust parametric (MTM) estimation, $\hat{G}$ is found by replacing G's parameters with their respective MLE and MTM estimates. In particular, as presented by Kaiser & Brazauskas (2006), the empirical point estimator and the $100(1-\alpha)\%$ *distribution-free* confidence interval of $\text{VaR}(\beta) = G^{-1}(1-\beta)$ are given by

$$\widehat{\text{VaR}}_{\text{EMP}}(\beta) = X_{n-[n\beta]:n} \qquad \text{and} \qquad (X_{k_1:n}, X_{k_2:n}),$$

where $k_1 = [n((1-\beta) - z_{\alpha/2}\sqrt{\beta(1-\beta)/n})]$ and $k_2 = [n((1-\beta) + z_{\alpha/2}\sqrt{\beta(1-\beta)/n})]$. Here $[\cdot]$ denotes 'greatest integer part' and $z_{\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution.

The robust parametric point estimator of $\text{VaR}(\beta)$ is found by transforming $\hat{\sigma}_{\text{MTM}}$ according to Eq. (2.10) and then multiplying the transformation by the deductible $x_0 = 500$; the corresponding $100(1-\alpha)\%$ confidence interval is then derived by applying the delta method to Eq. (3.4). These two steps lead to:

$$\widehat{\text{VaR}}_{\text{MTM}}(\beta) = 500\exp\{\hat{\sigma}_{\text{MTM}} Q^*_{T(v)}(1-\beta/2)\}$$

and

$$\widehat{\text{VaR}}_{\text{MTM}}(\beta) \times (1 \pm z_{\alpha/2}(\hat{\sigma}_{\text{MTM}}/\sqrt{n}) \sqrt{\Delta(a,b)} Q^*_{T(v)}(1-\beta/2)),$$

where $Q^*_{T(v)}(1-\beta/2)$ denotes the $(1-\beta/2)$th quantile of the standard t-distribution with v degrees of freedom, and $z_{\alpha/2}$ is again the $(1-\alpha/2)$th quantile of the standard normal distribution. The MLE point and interval estimators are constructed by following the same two steps. Table 6 presents empirical, parametric, and robust parametric point estimates and 95% interval estimates of $\text{VaR}(\beta)$ for several levels of β and various estimation methodologies. For comparison, we also include the $\text{VaR}(\beta)$ estimates based on the log-t_8 model, which are taken from Brazauskas (2009). In that article, it was found that the log-t_8 distribution provides an excellent fit for the upper 90% of the data.

Table 6. Point estimates and 95% confidence intervals of various value-at-risk measures computed by employing empirical, parametric (MLE), and robust parametric (MTM) methodologies.

		Estimation methodology			
		LOG-folded-normal		log-folded-t_7	log-t_8
Risk measure VaR(β)	Empirical	MLE	MTM1	MTM2	MTM
β = 0.25	2058 (1830; 2268)	2417 (2234; 2600)	2089 (1925; 2254)	2132 (1954; 2310)	2112 (1867; 2357)
β = 0.10	4555 (3758; 5974)	4759 (4243; 5275)	3864 (3428; 4299)	4472 (3906; 5037)	4512 (3821; 5203)
β = 0.05	7731 (6905; 11,339)	7298 (6355; 8242)	5695 (4931; 6460)	7660 (6452; 8869)	7850 (6410; 9290)
β = 0.01	26,791 (20,800; 84,464)	16,774 (13,924; 19,624)	12,119 (9981; 14,258)	27,844 (21,342; 34,346)	28,788 (21,360; 36,217)

Several conclusions emerge from the table. First, for risk evaluations based on the log-folded-normal model, the MTM1 fit is good everywhere except for the upper 15% of the data, which results in an accurate estimation of the empirical risk of moderate significance (e.g. $\beta = 0.25$) but severe underestimation when $\beta \leq 0.10$. The MLE fit, on the other hand, is mostly poor for the upper 35% of the data but (accidentally) matches the data well between the 90th and 95th percentiles. That yields fairly accurate estimates of VaR for $\beta = 0.10$, 0.05 and poor ones for $\beta = 0.25$, 0.01. Second, the point estimates of the risk based on the log-folded-t_7 and log-t_8 models and the empirical approach are very close for all levels of β because both parametric models are in close agreement with the data (see Figure 3). Third, the main advantage of robust parametric methodology over the empirical one is that it produces substantially shorter confidence intervals, especially for extreme significance levels (e.g. $\beta \leq 0.05$). Fourth, notice that the intervals based on the log-folded-t_7 model are slightly shorter than those of the log-t_8. This fact is primarily due to the parsimony of the former model, i.e. it has fewer unknown parameters than the log-t_8.

5. Conclusions

In this article, we have introduced the log-folded-normal and log-folded-t distributions for modeling insurance loss data. The close relationship between these families and the normal and t distributions makes them mathematically tractable and computationally attractive. In insurance context, if one applies the contract deductible for rescaling (instead of shifting) of losses, then the log-folded families emerge very naturally. Another positive feature of these probability distributions is their parsimony.

Further, we have presented and developed two standard (MM and MLE) and a class of robust (MTM) methods for the estimation of the parameters of the log-folded-normal and log-folded-t distributions. Large and small-sample properties of such estimators have been thoroughly investigated. We have concluded that, except for very heavy-tailed cases (e.g. when $\nu \leq 3$), the asymptotic results become valid for samples of size 100 or larger.

Finally, as the real-data example has shown, a log-folded distribution can fit insurance loss data exceptionally well. Subsequently, this translates into correct segmentation and accurate estimation of the empirical (observed) risk. Also, by employing a parametric model, we typically arrive at less variable estimates and shorter confidence intervals for the risk. This is one of the two key advantages of such approach over the empirical methodology; the other one is parametric model's ability to provide more reliable inference beyond the range of the observed data.

References

Azzalini, A., dal Cappello, T. & Kotz, S. (2002). Log-skew-normal and log-skew-*t* distributions as models for family income data. *Journal of Income Distribution* **11** (3–4), 12–20.

Beirlant, J., Teugels, J. L. & Vynckier, P. (1996). *Practical analysis of extreme values*. Leuven, Belgium: Leuven University Press.

Brazauskas, V. (2009). Robust and efficient fitting of loss models: diagnostic tools and insights. *North American Actuarial Journal* **13** (3), 1–14.

Brazauskas, V., Jones, B. & Zitikis, R. (2009). Robust fitting of claim severity distributions and the method of trimmed moments. *Journal of Statistical Planning and Inference* **139** (6), 2028–2043.

Cooray, K. & Ananda, M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal* **2005** (5), 321–334.

Cummins, J. D., Dionne, G., McDonald, J. B. & Pritchett, B. M. (1990). Applications of the GB2 family of distributions in modeling insurance loss processes. *Insurance: Mathematics and Economics* **9** (4), 257–272.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous univariate distributions* (Vol. 2, 2nd ed). New York: Wiley.

Kaiser, T. & Brazauskas, V. (2006). Interval estimation of actuarial risk measures. *North American Actuarial Journal* **10** (4), 249–268.

Klugman, S. A., Panjer, H. H. & Willmot, G. E. (2004). *Loss models: from data to decisions* (2nd ed). New York: Wiley.

Scollnik, D. P. M. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal* **2007** (1), 20–33.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Sun, J., Frees, E. W. & Rosenberg, M. A. (2008). Heavy-tailed longitudinal data modeling using copulas. *Insurance: Mathematics and Economics* **42** (2), 817–830.