Re-weighting the tails of unimodal distributions with positive support to model insurance losses

Luca Bagnato[a], Antonio Punzo[b,*], Antonello Maruotti[c]

[a] Dipartimento di Discipline Matematiche, Finanza Matematica e Econometria, Università Cattolica del Sacro Cuore
[b] Dipartimento di Economia e Impresa, Università di Catania
[c] DECOPOLI, Libera Università Maria Ss. Assunta, Roma

Abstract

Insurance loss data are positive and their distribution is often unimodal hump-shaped and right-skewed. Many parametric existing models are able to reproduce these peculiarities, but they often fail to cover the behavior of small and/or large losses, being the models' tails thinner than required. To remedy the situation, we propose the following approach. A 2-parameter unimodal hump-shaped model with positive support, parametrized with respect to the mode and to another parameter that is closely related to the distribution variability, is chosen as reference distribution. The parameter related to the variability is then scaled by a convenient distribution (mixing distribution) taking values on all or part of the positive real line and depending on a single parameter governing the tail behavior of the resulting 3-parameter model. The mixing distribution is the key; it gives more flexibility to the tails with respect to the reference distribution. Although the reference distribution can be of any specified form, for illustrative purposes we consider the (unimodal) gamma, the lognormal, and the inverse Gaussian. They are used, by fixing the mode in one, as mixing distributions too; this guarantees that the reference distribution is obtained as special case of the proposed model (nested models). A family of nine models arises by combining these choices. These models are applied on two famous insurance loss datasets and compared with several standard distributions used in the actuarial literature. Comparison is made in terms of goodness-of-fit and through an analysis of the commonly used risk measures computed on the fitted models.

Keywords: Mode, Positive support, Normal scale mixture, Insurance losses, Risk measures, Heavy tailed distributions.

1. Introduction

Insurance loss data are positive and their distribution is often unimodal hump-shaped (Cooray and Ananda, 2005), right-skewed (Lane, 2000), and with heavy tails (Cooray and Ananda, 2005 and Ibragimov et al., 2015). Though many parametric unimodal distributions have been used in the actuarial literature for modeling these data (Klugman et al., 2012), their peculiarities call for more flexible models (Ahn et al., 2012).

Right-skewness may be accommodated by skewed distributions (Lane, 2000). In this class, Vernic (2006), Bolance et al. (2008), Eling (2012), and Kazemi and Noorizadeh (2015) identify the skew-normal as a promising model. However, using the skew-normal distribution is in principle appropriate when the support is the whole real line, while it is not adequate if the support is the positive real line as it causes *boundary bias*, that is, allocation of probability mass outside the theoretical support. A possible solution consists in considering transformations, for example the logarithm, so as to make the support the whole real line and then fitting the skew-normal distribution. Although such a treatment is very simple to use, the transformed variable becomes more difficult to be interpreted (Bagnato and Punzo, 2013). Instead

*Corresponding author. Phone: +39 095 7537640, Fax: +39 095 7537610
Email addresses: luca.bagnato@unicatt.it (Luca Bagnato), antonio.punzo@unict.it (Antonio Punzo), a.maruotti@lumsa.it (Antonello Maruotti)

of applying transformations, there is a growing interest in proposing models having the desired support (see, e.g., Chen, 2000), and a lot of existing parametric unimodal distributions satisfy this requirement.

The remaining task of accurately fitting the tails of insurance losses is the most important in actuarial risk modeling (Ahn et al., 2012 and Abu Bakar et al., 2015). In particular, the losses in the right tail, though rare in frequency, are indeed the ones that have the most impact on the operations of an insurer and could lead to possible bankruptcy of the company. In such circumstances, heavy-tailed distributions have been shown to be reasonably competitive (McNeil, 1997 and Embrechts et al., 2003). Empirical evidence in favor of the skew-t distribution has been provided by Eling (2012). However, such a model suffers of the boundary bias problem. Further traditional parametric models actuaries have employed for heavy-tailed loss data include the Pareto, lognormal, Weibull, and gamma distribution (Hogg and Klugman, 2009 and Klugman et al., 2012). Amongst them, the Pareto distribution, due to the monotonically decreasing shape of the density, does not provide a reasonable fit for many applications, in particular when the density of the data is hump-shaped (Cooray and Ananda, 2005). On the other side, lognormal, Weibull, and gamma distributions cover better the behavior of small losses, but fail to cover the behavior of large losses. Further models have been proposed to cope with this issue (see, e.g., Cooray and Ananda, 2005, Scollnik and Sun, 2012, Nadarajah and Bakar, 2014, and Abu Bakar et al., 2015).

In Section 2, we extend this branch of literature by introducing an approach that allows to account for all the peculiarities of the loss data discussed above. The underlying idea is borrowed by the famous Normal scale mixture model (see, e.g., Watanabe and Yamaguchi, 2004, Chapter 4), and it is roughly described below. A 2-parameter unimodal hump-shaped model, defined on a positive support and parametrized with respect to the mode $\theta > 0$ and to another parameter $\gamma > 0$ that is closely related to the distribution variability, is chosen as benchmark (we use the term "reference distribution"). The γ-parameter is then scaled by a convenient distribution, that we call "mixing distribution", taking values on all or part of the positive real line and depending on a single parameter ν governing the tail behavior of the resulting 3-parameter model. The mixing distribution allows to give more flexibility to the tails of the reference distribution. Advantageously, the way the overall model is defined guarantees unimodality in θ and smoothness.

Various classes of models can be considered to choose for a convenient mixing distribution. We suggest to use mode-parameterized unimodal hump-shaped mixing distributions with mode fixed at $\theta = 1$ and $\gamma = \nu$. This allows to obtain the reference distribution as a special case of the proposed model when ν tends to zero, so to have nested models. A likelihood-ratio test can be so used to determine whether the proposed distribution is a significant improvement over the reference one.

Details about the proposed approach are given by focusing on three reference distributions: the (unimodal) gamma, the lognormal, and the inverse Gaussian (Section **??**). By combining them as reference and mixing distributions, we introduce a novel family of nine different models. All the new models, together with some standard distributions, are fitted to two real benchmark insurance loss data and the results are presented in Section 4. Goodness-of-fit tests, tail risk measures such as value at risk and conditional tail expectation are estimated for the analyzed datasets. Finally, some conclusions, along with future possible extensions, are drawn in Section 5.

2. A general framework to re-weight the tails of unimodal densities with positive support

Let X be a positive random variable. Requiring that the density function $p(x)$ of X should be unimodal hump-shaped and positively skewed, a general (3-parameter) unimodal density function could assume the form

$$\begin{aligned} p(x;\theta,\gamma,\nu) &= \int_0^\infty f(x;\theta,\gamma/w)\, dH(w;\nu) \\ &= \int_0^\infty f(x;\theta,\gamma/w)\, h(w;\nu)\, dw, \quad x>0, \end{aligned} \tag{1}$$

where

- $f(x;\theta,\gamma)$ is the unimodal density function chosen as reference distribution for X, with $\theta > 0$ denoting the mode and $\gamma > 0$ governing the concentration of f around the mode;
- $H(w;\nu)$ is the cumulative distribution function, depending on the parameter ν, of the mixing positive variable W;
- $h(w;\nu)$ is the probability density (or mass) function, depending on the parameter ν, of the mixing positive variable W.

It is straightforward to show that model (1) is unimodal, with mode in θ; moreover, $p(x;\theta,\gamma,\nu)$ guarantees a different tail behavior, governed by ν, with respect to the reference distribution $f(x;\theta,\gamma)$. Some cases of practical interest can be obtained under a convenient choice of the mixing variable. If W is a degenerate random variable taking value 1 ($W \equiv 1$), then the reference distribution $f(x;\theta,\gamma)$ is obtained. If W takes values on $(0,1)$, then the tails of $p(x;\theta,\gamma,\nu)$ are heavier than the tails of the reference distribution $f(x;\theta,\gamma)$, while, if W takes values in $(1,\infty)$, then the tails of $p(x;\theta,\gamma,\nu)$ are lighter than those of $f(x;\theta,\gamma)$.

As commonly required to models that aim to make more flexible the behavior of the tails of a reference distribution (as, e.g., the well-known normal scale mixture model), it would be better for model (1) to embed the reference distribution as a special case (under a particular choice of ν), so that the two models are nested. This allows to use inferential procedures such as the likelihood-ratio test (cf. Section 4). With this in mind, mode-parameterized densities could be also useful to define convenient mixing density functions $h(w;\nu)$. In particular, by using

$$h(w;\nu) = g(w;\theta = 1,\gamma = \nu), \quad w > 0, \tag{2}$$

where g is a density function with the same characteristics of f, the reference distribution $f(x;\theta,\gamma)$ is obtained as a limiting case of $p(x;\theta,\gamma,\nu)$ when ν tends to zero.

3. Considered mode-parameterized distributions

Among the existing 2-parameter models that can be used for the reference distribution f, as well as for the mixing density function h in (2), we have chosen to adopt unimodal gamma (UG), lognormal (LN), and inverse Gaussian (IG) densities parameterized with respect to the mode. By combining these three density functions with respect to the reference distribution f and to the mixing distribution h, we introduce a novel family of nine different models. For notation purposes, a model with a UG as reference distribution and a LN as mixing distribution, will be denoted as a UG-LN model. However, other 2-parameter distributions defined on a positive support may be used if they can be parameterized according to the mode and to a further parameter governing the concentration around the mode; an example could be the Weibull distribution (Bartels and van Metelen, 1975).

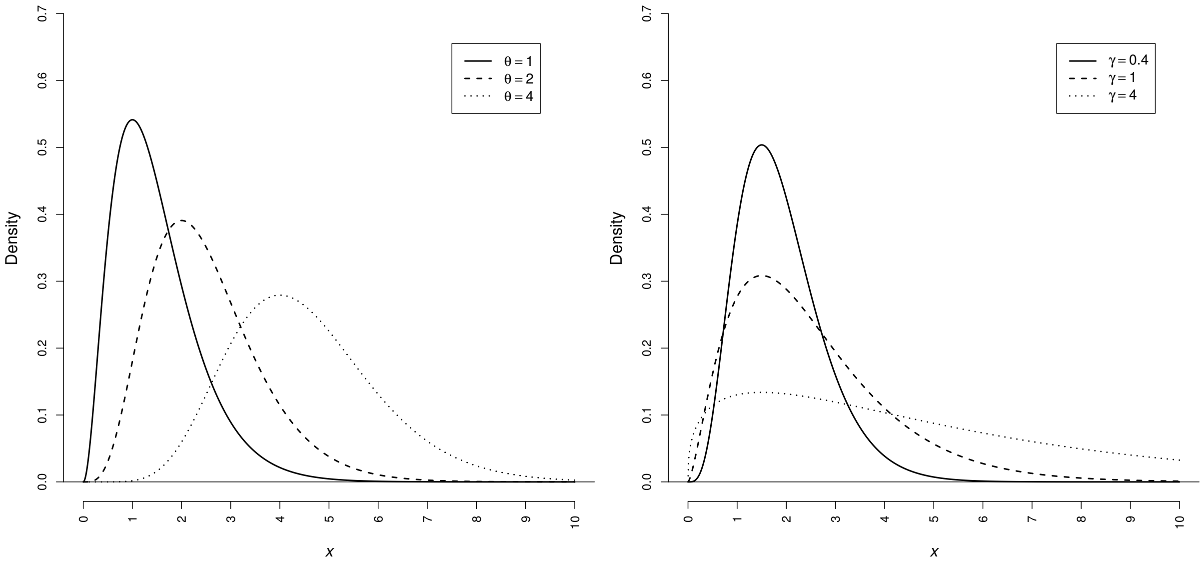
In the following, formulation and properties of the adopted mode-parameterized density functions are outlined.

3.1. Mode-parameterized unimodal gamma distribution

We consider the following mode-parameterized unimodal gamma (UG) density

$$f(x;\theta,\gamma) = \frac{x^{\frac{\theta}{\gamma}} e^{-\frac{x}{\gamma}}}{\gamma^{\frac{\theta}{\gamma}+1}\Gamma\left(\frac{\theta}{\gamma}+1\right)}, \quad x > 0, \tag{3}$$

with $\theta > 0$ and $\gamma > 0$. Note that, in (3), we are focusing only on the subclass of unimodal hump-shaped gamma densities, omitting all the (unlimited) reverse J-shaped cases that have a vertical asymptote in $x = 0$. This parameterization of the gamma distribution has been successfully considered in Chen (2000) and Bagnato and Punzo (2013).



(a) $\gamma = 0.5$ (b) $\theta = 1.5$

Figure 1: Mode-parameterized unimodal gamma densities in (3).

The effect of varying the mode θ, the other parameter kept fixed, is shown by a set of UG densities displayed in Figure 1(a). The variance of a random variable with density function (3) is

$$\gamma^2 + \theta\gamma. \tag{4}$$

Fixing θ in (4), the variance increases if γ increases, confirming that γ governs the spread of the distribution. The effect of varying γ, for fixed θ, is illustrated in Figure 1(b).

3.2. Mode-parameterized lognormal distribution

We consider the following mode-parameterized lognormal (LN) density

$$f(x;\theta,\gamma) = \frac{e^{-\frac{(\ln x - \ln\theta - \gamma)^2}{2\gamma}}}{\sqrt{2\pi\gamma}x}, \quad x > 0, \tag{5}$$

with $\theta > 0$ and $\gamma > 0$, as reference distribution in (2).

The effect of varying the mode θ, the other parameter kept fixed, is shown in Figure 2(a). The variance of a random variable with density function (5) is

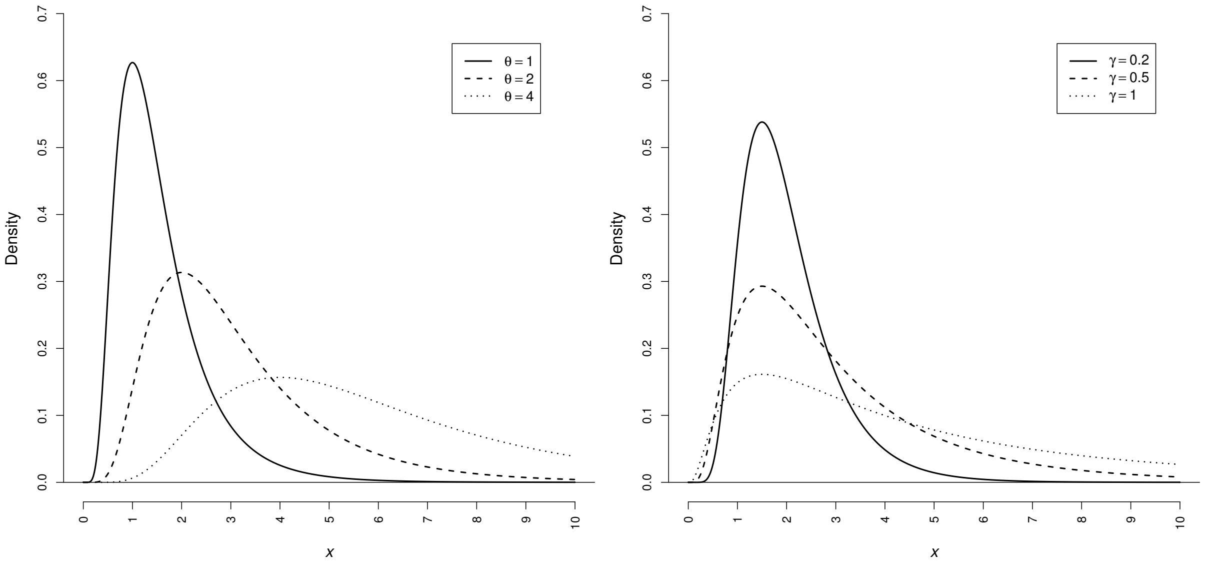
$$\left(e^{\gamma} - 1\right)\theta^2 e^{3\gamma}. \tag{6}$$

Fixing θ in (6), the variance increases if γ increases, confirming that γ governs the spread of the distribution. The effect of varying γ, for fixed θ, is illustrated in Figure 2(b).

3.3. Mode-parameterized inverse Gaussian distribution

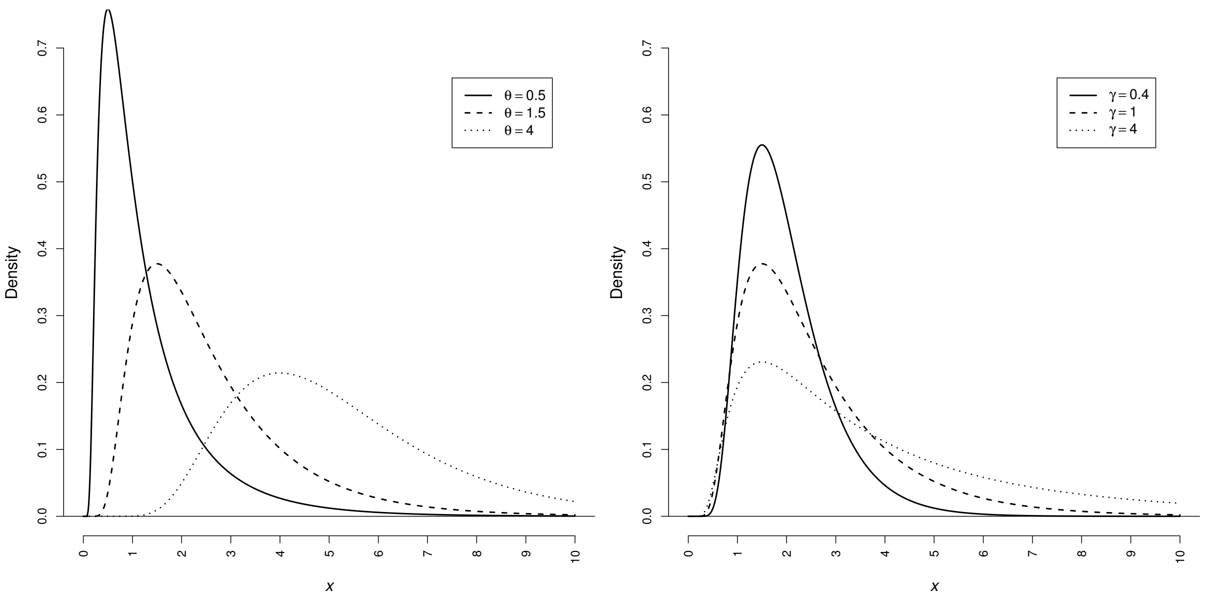
The mode-parameterized inverse Gaussian (IG) distribution we use has density function

$$f(x;\theta,\gamma) = \sqrt{\frac{\theta(3\gamma+\theta)}{2\pi\gamma x^3}} \exp\left\{ -\frac{\left[x - \sqrt{\theta(3\gamma+\theta)}\right]^2}{2\gamma x} \right\}, \quad x > 0, \tag{7}$$



(a) $\gamma = 0.3$ (b) $\theta = 1.5$

Figure 2: Mode-parameterized lognormal densities in (5).



(a) $\gamma = 1$ (b) $\theta = 1.5$

Figure 3: Mode-parameterized inverse Gaussian densities in (7).

where $\theta > 0$ and $\gamma > 0$.

The effect of varying the mode θ, the other parameter kept fixed, is shown in Figure 3(a). The variance of the random variable X with density function (7) is

$$\gamma\sqrt{\theta}\sqrt{3\gamma + \theta}.$$

The last expression, analyzed as a function of γ, is monotone increasing; consequently, fixed θ, the variability increases in line with the value of γ, confirming that γ governs the spread of X. The effect of varying γ, the mode θ kept fixed, is illustrated in Figure 3(b).

4. Analyses on real insurance loss data

In this section, we investigate the behavior of our nine models through real insurance loss data widely used by many authors. The proposed models are compared with several standard distributions used in the actuarial literature. The competitors are the unimodal gamma (UG) in (3), the lognormal (LN) in (5), the inverse Gaussian (IG) in (7), the Weibull, the logistic, the skew-normal, and the skew-t. Parameters are estimated based on maximum likelihood (ML) and the whole analysis is conducted in the R statistical software (R Core Team, 2016). We implemented a convenient code to find ML estimates for the three parameters of our models, and for the two parameters of UG, LN, and IG; this code is available at `http://www.olddei.unict.it/punzo/Rcode.htm`. ML estimates for the parameters of Weibull and logistic are obtained by the `fitdist()` function of the **fitdistrplus** package (Delignette-Muller and Dutang, 2015 and Delignette-Muller et al., 2017), while for skew-normal and skew-t by the `snormFit()` and `sstdFit()` functions, respectively, of the **fGarch** package (Wuertz and Chalabi, 2016). The comparison is made in terms of goodness-of-fit and through an analysis of the commonly used risk measures arising from the fitted models.

To compare models with the same number of parameters, in terms of goodness-of-fit, we use the log-likelihood (in addition to the criteria described below). Comparison of models with differing number of parameters is accomplished, as usual, via the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978) that, in our formulation, need to be maximized. Moreover, the likelihood-ratio (LR) test is used to compare nested models. To clarify the application of the LR test in our context we can consider, as an example, the following nested models (cf. Section 2): the proposed LN-IG and the classical LN. These models are nested because the LN-IG (alternative model) contains the LN (null model) as a particular case. The LR test can be used to determine whether the LN-IG model is a significant improvement over the LN model. In particular, under the null hypothesis of no improvement, the test statistic is

$$LR = -2\left[l(\widehat{\theta},\widehat{\gamma}) - l(\widehat{\theta},\widehat{\gamma},\widehat{\nu})\right],$$

where $\widehat{\theta}$, $\widehat{\gamma}$ and $\widehat{\nu}$ are the maximum likelihood estimates of θ, γ and ν, respectively, and where $l(\widehat{\theta},\widehat{\gamma})$ and $l(\widehat{\theta},\widehat{\gamma},\widehat{\nu})$ are the maximized log-likelihood values under the LN and LN-IG models, respectively. Using Wilks' theorem, LR can be approximated by a χ^2 random variable with one degree of freedom (given by the difference in the number of estimated parameters between the alternative and the null model), and this allows us to compute a p-value.

To describe the appropriateness of the fitted models in reproducing empirical risk measures, we adopt the value-at-risk (VaR) and the conditional tail expectation (CTE); they are computed, using the estimated parameters, at the 95% and 99% confidence levels. Note that we compute VaR numerically while CTE is obtained through simulations. More specifically, for simulations we consider one million random numbers from the fitted model and, in the fashion of Eling (2012), we checked for the stability of simulated data by looking at the convergence of the simulated means, standard deviations and risk measures. Finally, we use the backtesting procedure to test when models provide reasonable estimates of the VaR. The backtesting examines whether the proportion of violations obtained using the estimates of the VaR are compatible with the expected nominal level. This is verified through a binomial test comparing the number of violations observed with nominal probabilities of VaR (5% and 1%). The test is performed via the `VaRTest()` function of the **rugarch** package (Ghalanos, 2015).

4.1. U.S. indemnity losses

The first dataset consists of 1500 U.S. indemnity losses, general liability claims giving, for each, the indemnity payment measured in thousands of U.S. dollars. These data were first analyzed by Frees and

Valdez (1998) and are available in the R packages **copula** (Kojadinovic et al., 2010 and Hofert et al., 2017) and **evd** (Stephenson, 2015).

Table 1 reports some descriptive statistics while Figure 4 shows the histogram of the data. The losses are right-skewed and leptokurtic, with a long tail and few points laying quite far from the bulk of the data. The histogram reveals a further very typical feature of insurance loss data: a large number of small losses and a lower number of very large losses.

No. Observations	1,500
Mean	41.21
St. Dev.	102.75
Skewness	9.16
Kurtosis	145.17
Minimum	0.01
Maximum	2,173.60

Table 1: U.S. indemnity losses: descriptive statistics.

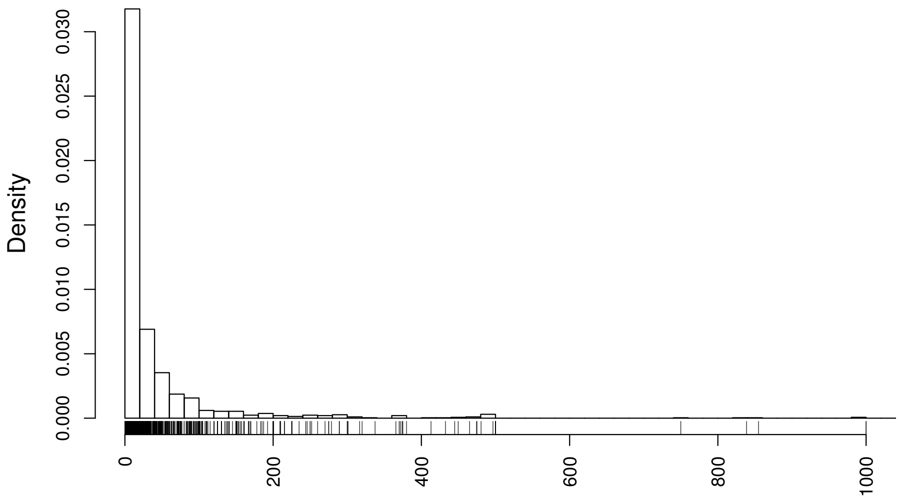


Figure 4: U.S. indemnity losses: histogram.

Table 2 presents a model comparison in terms of goodness-of-fit. To easy the reader in comparing the performance of the considered models in terms of AIC and BIC, Table 2 also gives rankings induced by these criteria. Here AIC and BIC, which provide the same ranking, indicate the UG-LN as the best model. The second best is the LN-LN model. The two models are different in terms of reference distribution (UG and LN) but they share the same mixing density (LN). Interestingly, six out of the nine proposed models occupy the first nine positions of this ranking. Of particular interest is to note that the UG model has the 14th position if no re-weighting of the tails is applied, while it reaches positions 1, 6, and 7 when our method is used; this is also corroborated by the practically null p-values of the LR test provided in the last column. The last three positions are occupied by the UG, skew-normal, and logistic distributions. Overall, our models appear to be competitive in comparison with the benchmark models presented in Table 2.

By considering the confidence levels of 95% and 99%, Table 3 reports the empirical VaR as well as the estimated VaR from the fitted models; corresponding backtesting results are also provided. Percentage of variation of each estimated VaR with respect to the empirical VaR, and ranking induced by the absolute

Model	# par.	Log-lik.	AIC	Ranking	BIC	Ranking	LR test (null model)
LN-LN	3	-6,561.320	-13,128.639	2	-13,144.579	2	0.001 (LN)
LN-UG	3	-6,566.582	-13,139.163	5	-13,155.103	5	0.053 (LN)
LN-IG	3	-6,566.558	-13,139.116	4	-13,155.056	4	0.518 (LN)
UG-LN	3	-6,558.861	-13,123.722	1	-13,139.662	1	0.000 (UG)
UG-UG	3	-6,571.902	-13,149.805	6	-13,165.745	6	0.000 (UG)
UG-IG	3	-6,585.860	-13,177.719	7	-13,193.659	7	0.000 (UG)
IG-LN	3	-7,017.739	-14,041.477	13	-14,057.417	13	0.534 (IG)
IG-UG	3	-7,017.495	-14,040.989	11	-14,056.929	11	0.350 (IG)
IG-IG	3	-7,017.658	-14,041.316	12	-14,057.256	12	0.460 (IG)
LN	2	-6,566.767	-13,137.534	3	-13,148.160	3	
UG	2	-7,077.964	-14,159.928	14	-14,170.555	14	
IG	2	-7,017.931	-14,039.863	10	-14,050.489	10	
Weibull	2	-6,658.850	-13,321.699	8	-13,332.326	8	
logistic	2	-8,270.456	-16,544.913	16	-16,555.539	16	
skew-normal	3	-8,148.487	-16,302.974	15	-16,318.914	15	
skew-t	4	-6,744.992	-13,497.983	9	-13,511.923	9	

Table 2: U.S. indemnity losses: log-likelihood, AIC, and BIC for the competing models, along with rankings induced by these criteria. In the last column, p-values from the LR tests for the proposed models are given.

value of this measure, are also given to ease performance comparison. When the 95% confidence level is considered, the best model is the IG-LN; it slightly overestimates the empirical VaR by the 0.470%. At the 99% confidence level, the best model is the UG-LN, which overestimates the empirical VaR by the 3.498%. By considering the p-values from the backtesting procedure, it can be noted how the empirical VaR at the 95% confidence level matches up very well with the estimates of VaR provided by the majority of the competing models (apart from UG, logistic, and skew-t). On the contrary, the only model which is able to reproduce the empirical VaR at the 99% confidence level is the UG-LN (p-value equal to 0.793). Then, the UG-LN model, which is the best one in terms of AIC and BIC, is the only one providing a reasonable estimate for the VaR at both the considered confidence levels.

Finally, Table 4 reports the empirical and estimated values of the CTE at confidence levels of 95% and 99%; these levels are chosen in conformity with the VaR. Also in this case, the percentage of variation of each estimated CTE, with respect to the empirical CTE, is provided along with the rankings of the competing models induced by the absolute value of this variation. Regardless from the considered confidence level, the best model is the UG-LN; this choice corroborates the results obtained via AIC and BIC (cf. Table 3). It is interesting to note how the estimates, at confidence levels of 95% and 99%, of the CTE from the best UG-LN model exceeds the empirical value (9.167% and 30.436%, respectively), suggesting sufficient allocations of resources.

Model			VaR				Prop. Viol.		p-value	
	95%	Diff. %	Ranking	99%	Diff. %	Ranking	95%	99%	95%	99%
Empirical	170,400			475,055						
LN-LN	178,805	4,933	9	513,462	8,085	2	0,049	0,004	0,905	0,008
LN-UG	173,646	1,905	6	528,933	11,341	3	0,050	0,004	1,000	0,008
LN-IG	175,783	3,159	8	536,514	12,937	5	0,050	0,004	1,000	0,008
UG-LN	168,412	-1,167	5	491,670	3,498	1	0,051	0,009	0,813	0,793
UG-UG	159,283	-6,524	10	607,856	27,955	6	0,054	0,004	0,483	0,008
UG-IG	151,759	-10,939	11	333,238	-29,853	7	0,057	0,019	0,246	0,001
IG-LN	171,201	0,470	1	714,332	50,368	13	0,050	0,004	1,000	0,008
IG-UG	171,696	0,761	4	712,316	49,944	11	0,050	0,004	1,000	0,008
IG-IG	171,405	0,590	2	713,853	50,267	12	0,050	0,004	1,000	0,008
LN	174,036	2,134	7	531,241	11,827	4	0,050	0,004	1,000	0,008
UG	123,449	-27,553	14	189,772	-60,053	15	0,075	0,047	0,000	0,000
IG	171,661	0,740	3	720,418	51,649	14	0,050	0,004	1,000	0,008
Weibull	151,381	-11,162	12	299,780	-36,896	9	0,057	0,025	0,246	0,000
logistic	108,176	-36,516	15	155,524	-67,262	16	0,081	0,055	0,000	0,000
skew-normal	198,044	16,223	13	243,508	-48,741	10	0,047	0,035	0,549	0,000
skew-*t*	244,132	43,270	16	316,299	-33,419	8	0,035	0,021	0,004	0,000

Table 3: U.S. indemnity losses: VaR, difference (in percentage) with respect to the empirical VaR, ranking induced by the absolute difference, and proportion of violations and p-values from the backtesting of VaR, for the competing models. Confidence levels of 95% and 99% are considered.

Model		95%			99%	
	CTE	Diff. %	Ranking	CTE	Diff. %	Ranking
Empirical	373,811			739,617		
LN-LN	562,246	50,409	13	1656,638	123,986	15
LN-UG	454,346	21,544	3	1141,024	54,272	6
LN-IG	459,780	22,998	5	1154,128	56,044	7
UG-LN	408,079	9,167	1	964,728	30,436	1
UG-UG	697,931	86,707	16	2357,447	218,739	16
UG-IG	269,922	-27,792	6	494,565	-33,132	2
IG-LN	544,900	45,769	10	1383,779	87,094	12
IG-UG	548,352	46,692	12	1397,516	88,951	14
IG-IG	545,255	45,864	11	1384,017	87,126	13
LN	447,317	19,664	2	1104,476	49,331	4
UG	164,657	-55,952	14	230,981	-68,770	9
IG	542,177	45,040	9	1363,533	84,357	11
Weibull	246,191	-34,140	7	415,524	-43,819	3
logistic	137,698	-63,164	15	184,249	-75,089	10
skew-normal	226,094	-39,516	8	266,125	-64,019	8
skew-*t*	288,483	-22,827	4	355,006	-52,001	5

Table 4: U.S. indemnity losses: CTE, difference (in percentage) with respect to the empirical CTE, and ranking induced by the absolute difference, for the competing models. Confidence levels of 95% and 99% are considered.

4.2. Automobile Insurance Claims

The second dataset consists of 6773 amounts (in U.S. dollars) paid, by a large midwestern (U.S.) insurance company, to settle and close claims for private passenger automobile policies. These data were analyzed, among others, by Frees (2010) and Bee (2017) and are available in the R package **insurance-Data** (Wolny-Dominiak and Trzkesiok, 2014).

Table 5 and Figure 5 present descriptive statistics and the histogram of the data, respectively. As for the previous data set, we observe right-skewness, leptokurtosis, a large number of small losses and a lower number of very large losses.

No. Observations	6,773
Mean	1,853.03
St. Dev.	2,646.91
Skewness	6.24
Kurtosis	87.28
Minimum	9.50
Maximum	60,000.00

Table 5: Automobile Insurance Claims: descriptive statistics

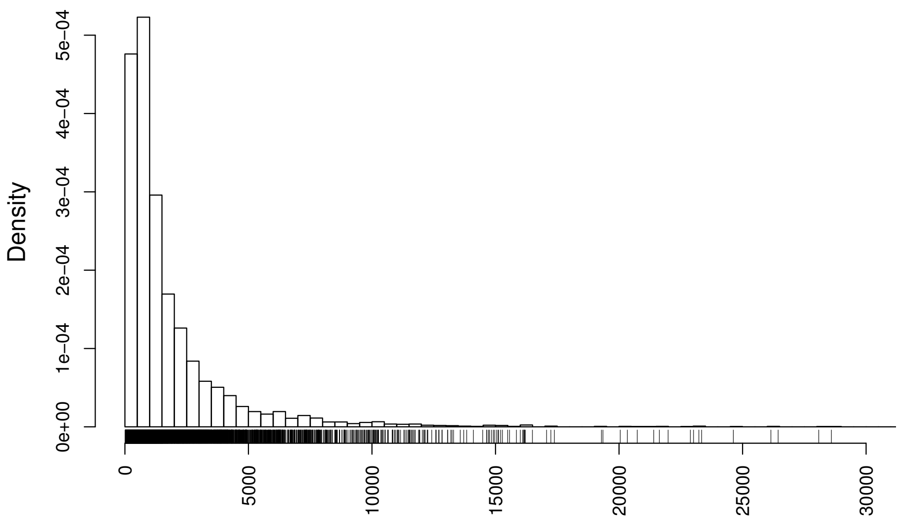


Figure 5: Automobile Insurance Claims: histogram.

In the fashion of Table 2, Table 6 evaluates goodness-of-fit for the competing models. AIC and BIC provide the same ranking apart from an exchange of positions (11st and 12nd) between the IG-IG and IG models. The best model is the UG-IG while the second best is the UG-LN. Although both models have the UG as reference distribution, the UG model alone, i.e. without our adjustment approach, has the 14th position only. The significant improvement in terms of goodness-of-fit of the modified UG models, with respect to the nested UG one, is also confirmed by the practically null p-values from the LR tests (see the last column of Table 6). Finally, seven out of the nine proposed models occupy the first nine positions of the ranking.

Table 7 reports the empirical VaR, the estimated VaR from the fitted models, the percentage of variation of each estimated VaR with respect to the empirical VaR, and ranking induced by the absolute value of this measure. Confidence levels of 95% and 99% are considered. When the 95% confidence level is considered, the best model is the UG-IG; it slightly underestimates the empirical VaR by the 1.329%. At

Model	# par.	Log-lik.	AIC	Ranking	BIC	Ranking	LR test (null model)
LN-LN	3	-57,170.529	-114,347.058	4	-114,367.520	4	0.000 (LN)
LN-UG	3	-57,166.575	-114,339.149	3	-114,359.611	3	0.000 (LN)
LN-IG	3	-57,184.078	-114,374.157	6	-114,394.619	7	0.152 (LN)
UG-LN	3	-57,133.830	-114,273.660	2	-114,294.122	2	0.000 (UG)
UG-UG	3	-57,175.769	-114,357.538	5	-114,378.000	5	0.000 (UG)
UG-IG	3	-57,123.403	-114,252.807	1	-114,273.269	1	0.000 (UG)
IG-LN	3	-57,613.138	-115,232.276	10	-115,252.738	10	0.000 (IG)
IG-UG	3	-57,613.060	-115,232.120	9	-115,252.582	9	0.000 (IG)
IG-IG	3	-57,628.661	-115,263.323	11	-115,283.785	12	0.149 (IG)
LN	2	-57,185.106	-114,374.211	7	-114,387.853	6	
UG	2	-57,736.619	-115,477.239	14	-115,490.880	14	
IG	2	-57,629.705	-115,263.410	12	-115,277.052	11	
Weibull	2	-57,707.938	-115,419.876	13	-115,433.517	13	
logistic	2	-60,882.663	-121,769.325	16	-121,782.967	16	
skew-normal	3	-59,636.402	-119,278.805	15	-119,299.267	15	
skew-t	4	-57,299.125	-114,606.250	8	-114,624.712	8	

Table 6: Automobile Insurance Claims: log-likelihood, AIC, and BIC for the competing models, along with rankings induced by these criteria. In the last column, p-values from the LR tests for the proposed models are given.

the 99% confidence level, the best model is the LN-LN, which underestimates the empirical VaR by the 1.634%. By considering the p-values from the backtesting procedure, and taking $\alpha = 0.05$ as benchmark significance levels, it can be noted how the empirical VaR at both confidence level does not match up very well with the estimates of VaR provided by the majority of the existing models (apart from the LN model with the 99% confidence level which has a p-value of 0.336). On the contrary, some of our models behave well. In particular, the LN-IG, UG-IG, IG-LN, and IG-UG models work well at both the considered confidence levels, the UG-LN model performs well at the 95% confidence level, and the LN-LN and LN-UG models behave well at the 99% confidence level. Therefore, the UG-IG model, selected by AIC and BIC (see Table 6) works well in reproducing the empirical VaR too.

Finally, Table 8 reports the empirical and estimated values of the CTE, the percentage of variation of each estimated CTE, with respect to the empirical CTE, and the rankings of the competing models induced by the absolute value of this variation. Confidence levels of 95% and 99% are considered as for VaR. Regardless from the considered confidence level, the best model is the UG-IG; this choice corroborates the results obtained via AIC and BIC (cf. Table 6). This model overestimates the empirical CTE by the 1.059%, in the case of the 95% confidence level, and by 2.215% at the 99% confidence level.

Model	VaR						Prop. Viol.		p-value	
	95%	Diff. %	Ranking	99%	Diff. %	Ranking	95%	99%	95%	99%
Empirical	6356,726			12052,290						
LN-LN	6028,198	-5,168	6	11855,392	-1,634	1	0,057	0,011	0,008	0,606
LN-UG	5999,926	-5,613	9	11632,170	-3,486	2	0,058	0,011	0,004	0,383
LN-IG	6143,101	-3,361	3	12938,464	7,353	5	0,055	0,008	0,059	0,140
UG-LN	6253,580	-1,623	2	14057,200	16,635	10	0,052	0,007	0,395	0,011
UG-UG	6121,264	-3,704	4	16117,551	33,730	13	0,056	0,004	0,036	0,000
UG-IG	6272,222	-1,329	1	12770,985	5,963	4	0,052	0,009	0,494	0,223
IG-LN	6703,774	5,460	7	13051,439	8,290	7	0,046	0,008	0,132	0,140
IG-UG	6705,112	5,481	8	13047,294	8,256	6	0,046	0,008	0,132	0,140
IG-IG	6812,773	7,174	10	13982,679	16,017	8	0,044	0,007	0,033	0,011
LN	6106,883	-3,930	5	12670,840	5,132	3	0,056	0,009	0,031	0,336
UG	5525,987	-13,069	14	8478,556	-29,652	12	0,065	0,027	0,000	0,000
IG	6826,756	7,394	11	14046,672	16,548	9	0,044	0,007	0,033	0,011
Weibull	5763,323	-9,335	13	9115,029	-24,371	11	0,061	0,023	0,000	0,000
logistic	4293,018	-32,465	15	5922,370	-50,861	15	0,096	0,058	0,000	0,000
skew-normal	5783,652	-9,015	12	7107,775	-41,026	14	0,061	0,041	0,000	0,000
skew-t	14479,716	127,786	16	21223,668	76,097	16	0,007	0,002	0,000	0,000

Table 7: Automobile Insurance Claims: VaR, difference (in percentage) with respect to the empirical VaR, ranking induced by the absolute difference, and proportion of violations and p-values from the backtesting of VaR, for the competing models. Confidence levels of 95% and 99% are considered.

Model	95%			99%		
	CTE	Diff. %	Ranking	CTE	Diff. %	Ranking
Empirical	10,403.811			18,172.931		
LN-LN	10,908.727	4.853	4	22,309.638	22.763	6
LN-UG	10,995.232	5.685	5	23,080.708	27.006	10
LN-IG	10,829.715	4.094	3	20,541.759	13.035	5
UG-LN	11,632.873	11.814	8	22,952.298	26.299	7
UG-UG	14,989.748	44.079	14	38,406.946	111.342	16
UG-IG	10,514.016	1.059	1	18,575.429	2.215	1
IG-LN	11,798.341	13.404	10	22,972.318	26.410	8
IG-UG	11,795.732	13.379	9	22,984.315	26.476	9
IG-IG	11,457.822	10.131	7	20,014.844	10.135	4
LN	10,536.148	1.272	2	19,481.337	7.200	2
UG	7,360.358	-29.253	12	10,311.592	-43.259	12
IG	11,398.068	9.557	6	19,612.610	7.922	3
Weibull	7,859.849	-24.452	11	11,268.956	-37.990	11
logistic	5,306.385	-48.996	15	6,928.844	-61.873	15
skew-normal	6,596.702	-36.593	13	7,773.642	-57.224	14
skew-t	18,797.189	80.676	16	26,407.463	45.312	13

Table 8: Automobile Insurance Claims: CTE, difference (in percentage) with respect to the empirical CTE, and ranking induced by the absolute difference, for the competing models. Confidence levels of 95% and 99% are considered.

5. Conclusions

Gamma, lognormal, and inverse Gaussian are examples of 2-parameter models which are used to represent the distribution of insurance loss data. Although they are able to fit some of the peculiarities of the loss distribution, such as its positive support, hump-shaped unimodality and right-skewness, they often fail to reproduce the empirical behavior of the tails which is typically characterized by smaller losses with higher frequencies as well as occasional larger losses with lower frequencies.

In this paper, nine models have been proposed to improve the flexibility of the tails of the 2-parameter models cited above. The price to pay for this flexibility is the addition of one parameter only; it governs the tails behavior of the resulting 3-parameter model. The proposed models have been fitted to two well-known and publicly available insurance loss data sets and their performance has been measured and compared to some benchmark models. Classical risk measures from the fitted models have been evaluated too. In order to disseminate our approach, at `http://www.olddei.unict.it/punzo/Rcode.htm` we provide the R code to fit our nine models, as well as to compute the classical risk measures from them. The main finding from the empirical study is that the best performing model — both in terms of goodness-of-fit and in the analysis of the commonly used risk measures — is always one of the proposed ones. Moreover, all the proposed models are reasonably good models compared to the competitors.

However, the novelty of the paper is not limited to these nine models. It is worth noticing that our approach to obtain them can be easily extended to any other 2-parameter unimodal distribution, on a positive support, with the aim of making more flexible the behavior of its tails. This can be done provided that the reference model can be parameterized with respect to the mode and to another parameter governing the distribution variability.

There are different possibilities for further work, three of which are worth mentioning. First of all, our approach could be extended to the multivariate setting. Secondly, covariates are often available along with losses in insurance datasets. A straightforward extension would deal with the regression framework, in which the error term is assumed to be distributed according to one of the nine models introduced herein. Finally, extension of our approach to the case of data defined on a compact support may be considered too by using the beta as reference distribution.

References

Abu Bakar, S. A., Hamzah, N. A., Maghsoudi, M., Nadarajah, S., 2015. Modeling loss data using composite models. Insurance: Mathematics and Economics 61, 146–154.

Ahn, S., Kim, J. H. T., Ramaswami, V., 2012. A new class of models for heavy tailed distributions in finance and insurance risk. Insurance: Mathematics and Economics 51, 43–52.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Bagnato, L., Punzo, A., 2013. Finite mixtures of unimodal beta and gamma densities and the k-bumps algorithm. Computational Statistics 28 (4), 1571–1597.

Bartels, C. P. A., van Metelen, H., 1975. Alternative probability density functions of income. Research Memorandum 29, Vrije University Amsterdam.

Bee, M., 2017. Density approximations and var computation for compound poisson-lognormal distributions. Communications in Statistics-Simulation and Computation 46 (3), 1825–1841.

Bolance, C., Guillen, M., Pelican, E., Vernic, R., 2008. Skewed bivariate models and nonparametric estimation for the CTE risk measure. Insurance: Mathematics and Economics 43, 386–393.

Chen, S. X., 2000. Probability density function estimation using gamma kernels. Annals of the Institute of Statistical Mathematics 52 (3), 471–480.

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

Delignette-Muller, M. L., Dutang, C., 2015. **fitdistrplus**: An R package for fitting distributions. Journal of Statistical Software 64 (4), 1–34.

Delignette-Muller, M. L., Dutang, C., Siberchicot, A., 2017. **fitdistrplus**: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data. Version 1.0-8 (2017-02-01).
URL `https://cran.r-project.org/web/packages/fitdistrplus/index.html`

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

Embrechts, P., Klüppelberg, C., Mikosch, T., 2003. Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin.

Frees, E. W., 2010. Regression Modeling with Actuarial and Financial Applications. International Series on Actuar. Cambridge University Press.
URL https://books.google.it/books?id=hhuZtBq5KZoC

Frees, E. W., Valdez, E. A., 1998. Understanding relationships using copulas. North American actuarial journal 2 (1), 1–25.

Ghalanos, A., 2015. **rugarch**: Univariate GARCH Models. Version 1.3-6 (2015-08-16).
URL https://cran.r-project.org/web/packages/rugarch/index.html

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., 2017. **copula**: Multivariate Dependence with Copulas. Version 0.999-16 (2017-01-09).
URL https://cran.r-project.org/web/packages/copula/index.html

Hogg, R. V., Klugman, S. A., 2009. Loss Distributions. Vol. 249 of Wiley Series in Probability and Statistics. Wiley.

Ibragimov, M., Ibragimov, R., Walden, J., 2015. Heavy-Tailed Distributions and Robustness in Economics and Finance. Vol. 214 of Lecture Notes in Statistics. Springer International Publishing, New York.

Kazemi, R., Noorizadeh, M., 2015. A comparison between skew-logistic and skew-normal distributions. Matematika 31 (1), 15–24.

Klugman, S. A., Panjer, H. H., Willmot, G. E., 2012. Loss Models: From Data to Decisions. Vol. 715 of Wiley Series in Probability and Statistics. Wiley.

Kojadinovic, I., Yan, J., et al., 2010. Modeling multivariate distributions with continuous margins using the copula r package. Journal of Statistical Software 34 (9), 1–20.

Lane, M. N., 2000. Pricing risk transfer transactions. Astin Bulletin 30 (2), 259–293.

McNeil, A. J., 1997. Estimating the tails of loss severity distributions using extreme value theory. Astin Bulletin 27, 117–137.

Nadarajah, S., Bakar, S. A. A., 2014. New composite models for the danish fire insurance data. Scandinavian Actuarial Journal 2014 (2), 180–187.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6 (2), 461–464.

Scollnik, D. P. M., Sun, C., 2012. Modeling with weibull-pareto models. North American Actuarial Journal 16 (2), 260–272.

Stephenson, A., 2015. **evd**: Functions for Extreme Value Distributions. Version 2.3-2 (2015-12-25).
URL https://cran.r-project.org/web/packages/evd/index.html

Vernic, R., 2006. Multivariate skew-normal distributions with applications in insurance. Insurance: Mathematics and Economics 38 (2), 413–426.

Watanabe, M., Yamaguchi, K., 2004. The EM Algorithm and Related Statistical Models. Statistics: A Series of Textbooks and Monographs. Taylor & Francis.
URL http://books.google.it/books?id=Q961Osmlv6oC

Wolny-Dominiak, A., Trzkesiok, M., 2014. **insuranceData**: A Collection of Insurance Datasets Useful in Risk Classification in Non-life Insurance. Version 1.0 (2014-09-04).
URL https://cran.r-project.org/web/packages/insuranceData/index.html

Wuertz, D., Chalabi, Y., 2016. **fGarch**: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling. Version 3010.82.1 (2016-08-15).
URL https://cran.r-project.org/web/packages/fGarch/index.html