ARTICLE IN PRESS

Insurance: Mathematics and Economics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Compound unimodal distributions for insurance losses

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

[a] Dipartimento di Economia e Impresa, Università di Catania, Catania, Italy
[b] Dipartimento di Discipline Matematiche, Finanza Matematica e Econometria, Università Cattolica del Sacro Cuore, Milano, Italy
[c] Dipartimento di Scienze Economiche, Politiche e delle Lingue Moderne, Libera Università Maria Ss. Assunta, Roma, Italy

ARTICLE INFO

Article history:
Received April 2017
Received in revised form October 2017
Accepted 22 October 2017
Available online xxxx

JEL classification:
C46
C51
C52

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

ABSTRACT

The distribution of insurance losses has a positive support and is often unimodal hump-shaped, right-skewed and with heavy tails. In this work, we introduce a 3-parameter compound model to account for all these peculiarities. As conditional distribution, we consider a 2-parameter unimodal hump-shaped distribution with positive support, parameterized with respect to the mode and to another variability-related parameter. The compound is performed by scaling the latter parameter by a convenient 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 compound distribution. Although any 2-parameter distribution can be considered to derive its compound version in our framework, for illustrative purposes we consider the unimodal gamma, the lognormal, and the inverse Gaussian. They are also used as mixing distributions; this guarantees that the un-compound distribution is nested in the compound model. A family of nine models arises by combining these choices. These models are applied on three 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.

1. Introduction

Insurance loss data are positive (Klugman et al., 2012) and their distribution is often unimodal hump-shaped (Cooray and Ananda, 2005), right-skewed (Lane, 2000), and with heavy tails (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; Bernardi et al., 2012). In this class, Vernic (2006), Bolancé et al. (2008), Adcock et al. (2015), 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 may become more difficult to be interpreted (Bagnato and Punzo, 2013). Instead 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 (Abu Bakar et al., 2015; Ahn et al., 2012). 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; Embrechts et al., 2003). Empirical evidence in favor of the skew-t distribution has been provided by Eling (2012, 2014). However, such a model suffers of the boundary bias problem. Further traditional parametric models that actuaries have been employed for heavy-tailed loss data include the Pareto, lognormal, Weibull, and gamma distribution (Hogg and Klugman, 2009; 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

* Correspondence to: Dipartimento di Scienze Economiche, Politiche e delle Lingue Moderne, Libera Università Maria Ss. Assunta, Via Pompeo Magno 22, 00192, Roma, Italy.

E-mail addresses: antonio.punzo@unict.it (A. Punzo), luca.bagnato@unicatt.it (L. Bagnato), a.maruotti@lumsa.it (A. Maruotti).

https://doi.org/10.1016/j.insmatheco.2017.10.007

better the behavior of small losses, but fail to cover the behavior of large losses. Further models have been proposed to cope with the issue of fitting the tails of insurance losses (see, e.g., Cooray and Ananda, 2005; Scollnik and Sun, 2012; Nadarajah and Abu Bakar, 2014; and Abu Bakar et al., 2015).

We extend this branch of literature by introducing a novel compound approach that allows to account for all the peculiarities of the loss data as discussed above (Section 2). Compound distributions can be dated back to Greenwood and Yule (1920) and have been widely used in the actuarial literature (see e.g. Tahir and Cordeiro, 2016). Most of the existing compound distributions are, however, discrete Poisson-based models (see e.g. Zhang et al., 2014 and references therein). Here, instead, we consider continuous compounding aiming at improving the tail behavior of any unimodal distribution with positive support. The underlying idea is borrowed from one of the most famous compound models, the normal scale mixture (see, e.g., Watanabe and Yamaguchi, 2004 Chapter 4), in which the variability-related parameter is scaled by a convenient random variable. In particular, we consider a 2-parameter unimodal hump-shaped model, defined on a positive support and parameterized with respect to the mode $\theta > 0$ and to another parameter $\gamma > 0$ that is closely related to the distribution variability. The γ-parameter is then scaled by some parameterized mixing distribution taking values on all or part of the positive real line and depending on a single parameter ν governing the tail behavior. As a result, we obtain a 3-parameter compound distribution which allows to give more flexibility to the tails of the conditional distribution. Advantageously, our compound model 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 distributions for the mixing distribution too, with mode at $\theta = 1$ and $\gamma = \nu$. This allows to obtain the conditional distribution as a special case of the proposed model when ν tends to zero, so to have nested models. Thus, a likelihood-ratio test can be used to determine whether the proposed compound distribution is a significant improvement over the conditional one (Section 5).

Details about the proposed approach are given by focusing on three unimodal hump-shaped distributions: the unimodal gamma, the lognormal, and the inverse Gaussian (Section 3). By combining them as conditional and mixing distributions, we introduce a novel family of nine different models; some of their properties are illustrated in Section 4, in addition to some considerations about inference via the maximum likelihood approach. All the new models, together with some standard distributions, are fitted to three real benchmark insurance loss data and the results are presented in Section 6. Goodness-of-fit is evaluated and tail risk measures, such as value at risk and conditional tail expectation, are estimated for the analyzed datasets (Section 5). Finally, some conclusions, along with future possible extensions, are drawn in Section 7.

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 unconditional probability density function (pdf) $p(x)$ of X should be unimodal hump-shaped and positively skewed, a general (3-parameter) compound unimodal pdf could assume the form

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

where $f(x;\theta,\gamma)$ is the unimodal conditional pdf, with $\theta > 0$ denoting the mode and $\gamma > 0$ governing the concentration of f around the mode, and $h(w;\nu)$ is the mixing probability density (or mass) function depending on the parameter ν.

The compound distribution p resembles the original distribution f, but it guarantees a different tail behavior, governed by ν, with respect to the conditional distribution f (cf. Section 4.1). Some cases of practical interest can be obtained under a convenient choice of the mixing random variable W (random counterpart of w). If W is degenerate in 1 ($W \equiv 1$), then the conditional distribution f is obtained. If W takes values on $(0, 1)$, then the tails of p are heavier than the tails of the conditional distribution f, while, if W takes values in $(1,\infty)$, then the tails of p are lighter than those of f.

As commonly required to models that aim to make more flexible the behavior of the tails of a conditional distribution (as, e.g., the well-known normal scale mixture model), it would be better for model (1) to embed the conditional 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 5). With this in mind, mode-parameterized pdfs could be also useful to define convenient mixing pdfs $h(w;\nu)$. In particular, by using

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

where g is a pdf with the same characteristics of f, the conditional pdf $f(x;\theta,\gamma)$ 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 conditional pdf f, as well as for the mixing pdf h in (2), we have chosen to adopt unimodal gamma (UG), lognormal (LN), and inverse Gaussian (IG) pdfs parameterized with respect to the mode. In the following, formulation and properties of the adopted mode-parameterized pdfs are outlined.

3.1. Mode-parameterized unimodal gamma distribution

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

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

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

The effect of varying the mode θ, the other parameter kept fixed, is shown by a set of UG pdfs displayed in Fig. 1(a).

The variance of a random variable with pdf (3) is

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

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

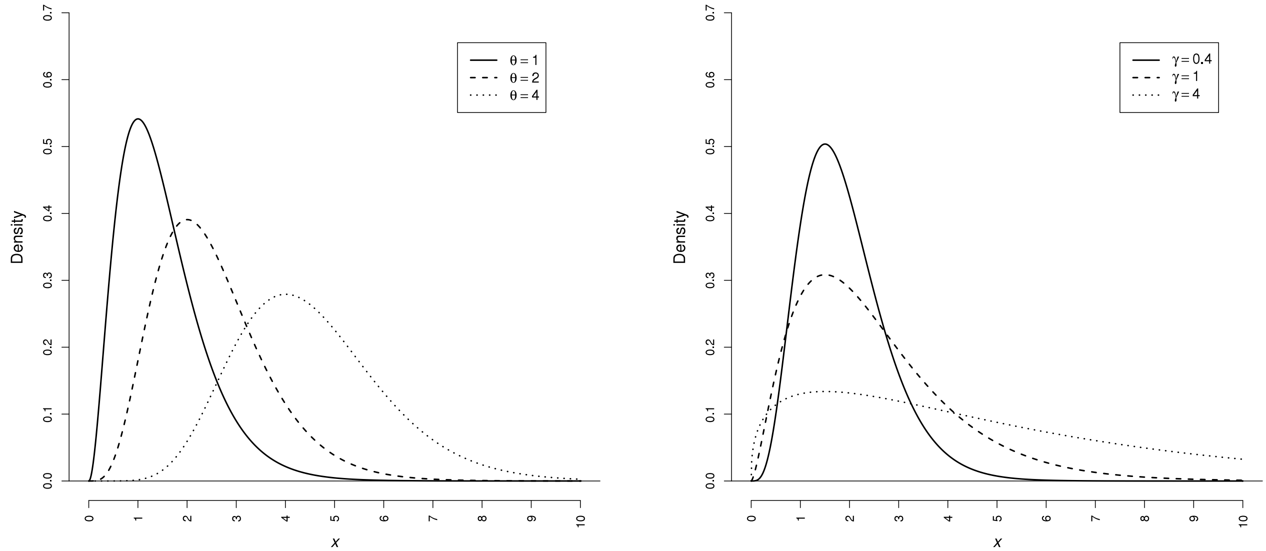
3.2. Mode-parameterized lognormal distribution

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

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

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

Please cite this article in press as: Punzo A., et al., Compound unimodal distributions for insurance losses. Insurance: Mathematics and Economics (2017), https://doi.org/10.1016/j.insmatheco.2017.10.007.



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

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

The effect of varying the mode θ, the other parameter kept fixed, is shown in Fig. 2(a).

The variance of a random variable with pdf (5) is

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

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

3.3. Mode-parameterized inverse Gaussian distribution

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

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

where $\theta > 0$ and $\gamma > 0$. This parameterization of the IG distribution has been recently proposed by Punzo (2017).

The effect of varying the mode θ, the other parameter kept fixed, is shown in Fig. 3(a). The variance of the random variable X with pdf (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 variability of X. The effect of varying γ, the mode θ kept fixed, is illustrated in Fig. 3(b).

4. The proposed family of models: properties and inference

By combining the three pdfs illustrated in Section 3, with respect to the conditional 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 conditional 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). For other compound models involving these three distributions, see Johnson and Kotz (1970, Section 8.2) and McDonald and Butler (1987).

4.1. The effect of ν

The parameter ν gives further flexibility to the compound distribution $p(x;\theta,\gamma,\nu)$ with respect to the simple conditional pdf $f(x;\theta,\gamma)$. The effect of varying ν, the other parameters θ and γ kept fixed, is shown by a set of UG-UG pdfs displayed in Fig. 4. Similar results are obtained for the others members of our family, but the corresponding plots are not displayed for the sake of space. As we can see by Figs. 4(a) and 4(c), the greater the value of ν, the heavier are the tails of the compound distribution with respect to the conditional UG pdf. At the same time, the greater the value of ν, the higher is the peakedness of the resulting pdf around the mode θ.

4.2. Closed form expression for the pdf

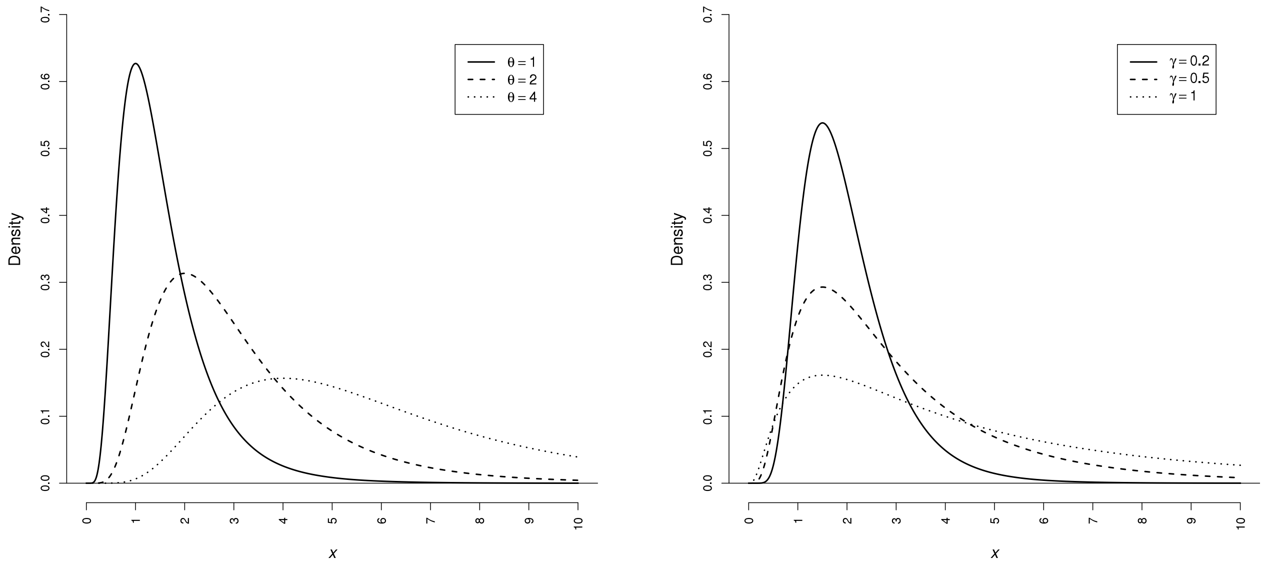
As emphasized by Shevchenko (2010), simple closed form expressions for the pdf of compound models are often not available. In our case, two exceptions are the LN-UG model, whose pdf can be written as

$$\begin{aligned} p(x;\theta,\gamma,\nu) = {} & \frac{\gamma}{\theta\Gamma\left(\frac{1}{\nu}\right)}\sqrt{\frac{2}{\pi}}\left(\frac{\gamma}{\nu}\right)^{\frac{1}{\nu}} K_{\frac{3}{2}+\frac{1}{\nu}} \\ & \times \left(\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right) \\ & \times \left[\frac{\nu}{2\gamma+\nu(\ln x-\ln\theta)^2}\right]^{\frac{3}{4}+\frac{1}{2\nu}}, \end{aligned} \tag{8}$$

and the LN-IG model, whose pdf can be written as

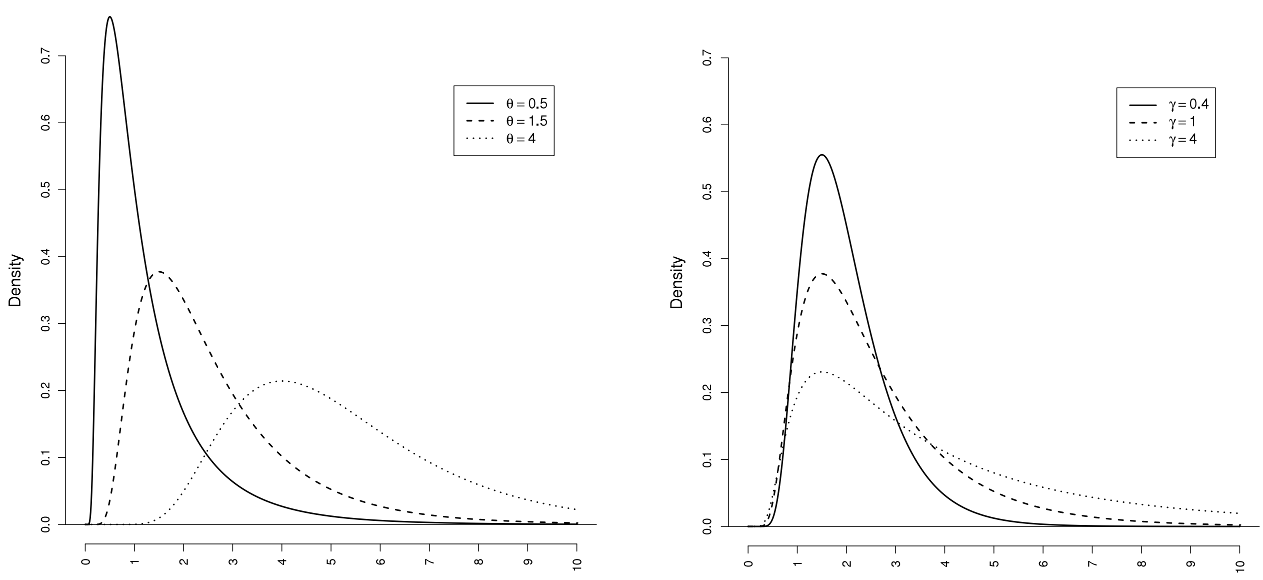
$$\begin{aligned} p(x;\theta,\gamma,\nu) = {} & \frac{1}{\pi\theta}\exp\left(\frac{\sqrt{1+3\nu}}{\nu}\right)\sqrt{\frac{1+3\nu}{\gamma\nu}} K_0 \\ & \times \left(\frac{1}{\nu}\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}\right), \end{aligned} \tag{9}$$

Please cite this article in press as: Punzo A., et al., Compound unimodal distributions for insurance losses. Insurance: Mathematics and Economics (2017), https://doi.org/10.1016/j.insmatheco.2017.10.007.



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

Fig. 2. Mode-parameterized lognormal densities in (5).



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

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

where $K_\alpha(y)$ denotes the modified Bessel function of the second kind.

4.3. The mode

Theorem 1 shows that model (1) is unimodal hump-shaped, with mode in θ.

Theorem 1. *The unconditional pdf $p(x;\theta,\gamma,\nu)$ in (1) is unimodal hump-shaped in θ.*

Proof. The first derivative, with respect to x, of $p(x;\theta,\gamma,\nu)$ is

$$p'(x;\theta,\gamma,\nu) = \int_0^\infty f'(x;\theta,\gamma/w)\, h(w;\nu)\, dw, \quad x > 0. \tag{10}$$

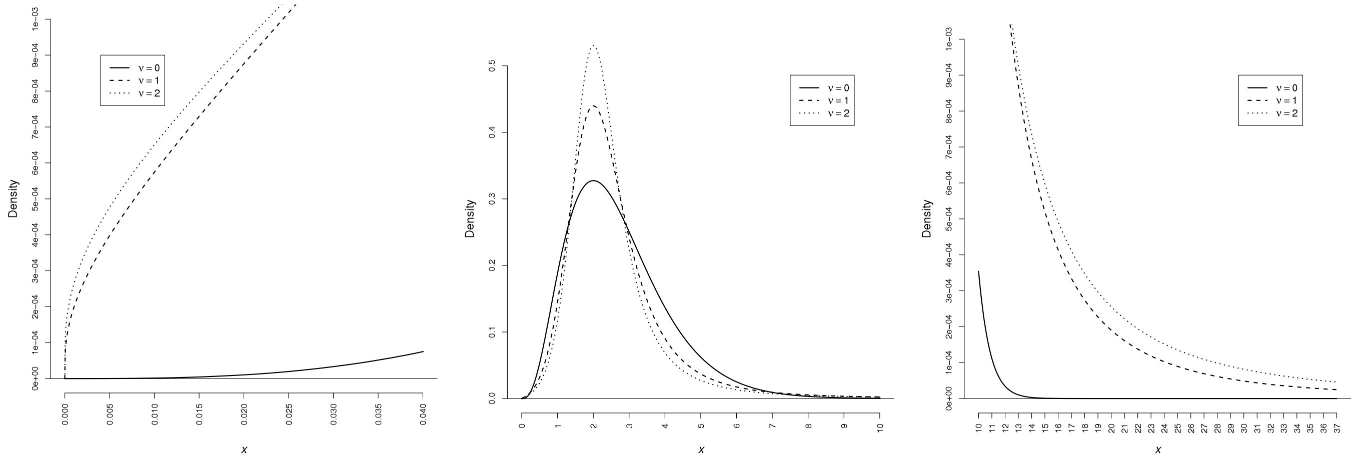
By definition, $f(x;\theta,\gamma/w)$ is unimodal hump-shaped with mode in $x = \theta$. Hence, $f'(x;\theta,\gamma/w) > 0$ for $x < \theta$, $f'(x;\theta,\gamma/w) < 0$ for $x > \theta$, and $f'(x;\theta,\gamma/w) = 0$ for $x = \theta$. Using these results and recalling that $h(w;\nu) \geq 0$ for $w > 0$ and $\nu > 0$, it is straightforward to prove the theorem. □

4.4. Moments

It is easy to show that the rth moment of the compound distribution p in (1) is given by

$$E_p\left(X^r\right) = E_h\left[E_f\left(X^r | W = w\right)\right], \tag{11}$$

where the subscripts of the expectation operators explicitly convey the distribution used to compute the expectation (see Klugman



(a) Zoom on the left tail.

(b) Overall densities.

(c) Zoom on the right tail.

Fig. 4. Effect of ν on a UG-UG distribution ($\theta = 2$ and $\gamma = 0.7$).

Table 1
First moment.

f	h			
	UG	LN	IG	No compound
UG	$\theta+\gamma$	$\theta+\gamma\exp\left(-\frac{\gamma}{2}\right)$	$\theta+\gamma\frac{\nu+\sqrt{1+3\nu}}{1+3\nu}$	$\theta+\gamma$
LN	No closed-form	No closed-form	$\theta\exp\left(\frac{\sqrt{1+3\nu}-\sqrt{1+3\nu(1-\gamma)}}{\nu}\right)\sqrt{\frac{1+3\nu}{1+3\nu(1-\gamma)}}$ [a]	$\theta\exp\left(\frac{3\gamma}{2}\right)$
IG	$-\frac{\theta}{\pi}\Gamma\left(\frac{1}{2}+\frac{1}{\nu}\right)\Gamma\left(-\frac{1}{\nu}\right)U\left(-\frac{1}{2},-\frac{1}{\nu},\frac{3\gamma}{\theta\nu}\right)\sin\left(\frac{\pi}{\nu}\right)$	No closed-form	No closed-form	$\sqrt{\theta\left(3\gamma+\theta\right)}$

[a] if $1+3\nu\left(1-\gamma\right)\geq 0$.

Table 2
Second moment.

f	h			
	UG	LN	IG	No compound
UG	No closed-form	$\theta^2+3\theta\gamma\exp\left(-\frac{\gamma}{2}\right)+2\gamma^2$	$\theta^2+\frac{1}{(1+3\nu)^{\frac{5}{2}}}\left(3\gamma\theta+6\gamma^2\nu+18\gamma\theta\nu+18\gamma^2\nu^2+27\gamma\theta\nu^2\right)$ $+\frac{1}{(1+3\nu)^2}\left(2\gamma^2+6\gamma^2\nu+3\gamma\theta\nu+6\gamma^2\nu^2+9\gamma\theta\nu^2\right)$	$\left(\theta+\gamma\right)\left(\theta+2\gamma\right)$
LN	No closed-form	No closed-form	$\theta^2\exp\left(\frac{\sqrt{1+3\nu}-\sqrt{1+\nu(3-8\gamma)}}{\nu}\right)\sqrt{\frac{1+3\nu}{1+\nu(3-8\gamma)}}$ [a]	$\theta^2\exp\left(4\gamma\right)$
IG	$\theta\left(3\gamma+\theta\right)+\gamma\sqrt{\theta\left(3\gamma+\theta\right)}$	$\theta\left(3\gamma+\theta\right)+\gamma\sqrt{\theta\left(3\gamma+\theta\right)}$	$\theta\left(3\gamma+\theta\right)+\gamma\sqrt{\theta\left(3\gamma+\theta\right)}$	$\theta\left(3\gamma+\theta\right)+\gamma\sqrt{\theta\left(3\gamma+\theta\right)}$

[a] if $1+\nu\left(3-8\gamma\right)\geq 0$.

et al., 2012 p. 69). By using Eq. (11), Tables 1 and 2 respectively list first and second moments – when available in closed form – for all the members of our family. In the last column of these tables, the corresponding moment for the conditional pdf f is reported too. In Table 1, $U\left(a, b, z\right)$ denotes the confluent hypergeometric function.

From Table 1 it is interesting to note that the mean of the compound models UG-UG and UG-LN do not depend on ν. In addition, the mean of the UG-UG model coincides with the mean of its conditional UG pdf. The same latter finding can be observed in Table 2: the second moments of the compound models having the IG as conditional distribution are equal to the second moment of the IG distribution.

4.5. Maximum likelihood estimates

To find the estimates of the parameters for our models, we use the maximum likelihood (ML) approach. Given a random sample $x_1, \ldots, x_n$ from the pdf in (1), the log-likelihood function related to the compound model is

$$l\left(\theta, \gamma, \nu\right) = \sum_{i=1}^{n} \ln\left[p\left(x_i; \theta, \gamma, \nu\right)\right]. \tag{12}$$

The first order partial derivatives of (12), with respect to $\left(\theta, \gamma, \nu\right)'$, are

$$l'\left(\theta, \gamma, \nu\right) = \sum_{i=1}^{n} \frac{\partial}{\partial\left(\theta, \gamma, \nu\right)'} \ln\left[p\left(x_i; \theta, \gamma, \nu\right)\right]. \tag{13}$$

The values of θ, γ, and ν that maximize $l\left(\theta, \gamma, \nu\right)$ are the ML estimates $\widehat{\theta}$, $\widehat{\gamma}$, and $\widehat{\nu}$ and satisfy the condition

$$l'\left(\widehat{\theta}, \widehat{\gamma}, \widehat{\nu}\right) = \mathbf{0}.$$

As closed-form maximum likelihood estimates are not always available but the (log-)likelihood can easily be evaluated, we simply use numerical methods in an attempt to maximize that likelihood, subject to any constraints that there may be on parameters. As this succeeds, there will be no need to derive and code any sophisticated algorithm (more on this topic can be found in MacDonald, 2014). Operationally (in the analyses of Section 6), we perform maximization of (12) numerically by the general-purpose optimizer `optim()` for R (R Core Team, 2016), included in the **stats** package. In particular, we use the BFGS algorithm for maximization, passed to `optim()` via the argument `method`. However, for the LN-UG and LN-IG models we have a closed-form expression for the pdf (cf. Section 4.2). For these models, details about the tridimensional vector of the first order partial derivatives on the right-hand side of (13) are given in Appendix.

5. Model selection and computational aspects

The proposed models are compared with several standard distributions used in the actuarial literature. The parametric competitors are the UG in (3), the LN in (5), the IG in (7), the Weibull, the logistic, the skew-normal, and the skew-t. The transformation kernel approach (Bolancé et al., 2003, 2008) is also considered as a nonparametric competitor. Parameters are estimated based on ML and the whole analysis is conducted in the R statistical software (R Core Team, 2016). Computation is performed on a Windows 7 Pro PC, with Intel i7 3.40 GHz CPU, 16.0 GB RAM, using R 32 bit, and the elapsed time is computed via the `proc.time()` function of the **base** package. 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/Compound.R. 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 transformation kernel approach is implemented using the R routine given in Pitt et al. (2011), by using what the authors call "method 1". 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 is 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).

Table 3
U.S. indemnity losses: descriptive statistics.

No. of Observations	1500
Mean	41.21
St. Dev.	102.75
Skewness	9.16
Kurtosis	145.17
Minimum	0.01
Maximum	2173.60

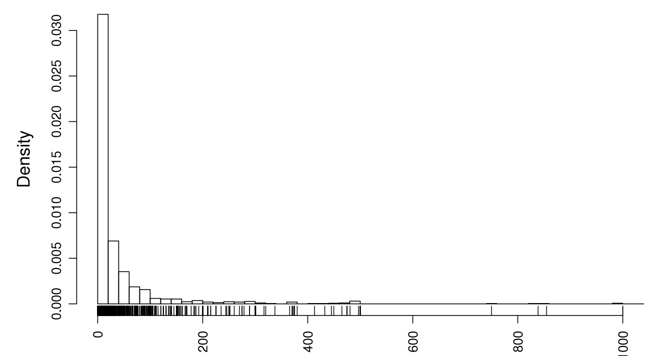


Fig. 5. U.S. indemnity losses: histogram.

6. 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. These datasets cover different levels of tail-weight which may arise when dealing with the empirical distribution of real-world insurance data. In particular the last dataset (Section 6.3), coming from the reinsurance world, presents very large losses. In this situation, it is more challenging to find a model able to accommodate the whole distribution of the data (cf. Klugman et al., 2012).

6.1. 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 and Yan, 2010 and Hofert et al., 2017) and **evd** (Stephenson, 2015).

Table 3 reports some descriptive statistics while Fig. 5 shows the histogram of the data. The losses are right-skewed and leptokurtic, with a long right-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.

Table 4 presents a model comparison in terms of goodness-of-fit. The log-likelihood value for the transformation kernel is given in the last line; it is the third greatest value if compared to the log-likelihoods of the parametric models. To compare the performance of the considered parametric models, Table 4 also gives rankings induced by AIC and BIC. These criteria, which provide the same ranking, indicate the UG-LN as the best model. The second best is the LN-LN model. Interestingly, six out of the nine proposed models occupy the first nine positions of this ranking. Of particular

Please cite this article in press as: Punzo A., et al., Compound unimodal distributions for insurance losses. Insurance: Mathematics and Economics (2017), https://doi.org/10.1016/j.insmatheco.2017.10.007.

Table 4
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.

Model	# par.	Log-lik.	AIC	Ranking	BIC	Ranking	LR test (null model)
LN-LN	3	−6561.320	−13,128.639	2	−13,144.579	2	0.001 (LN)
LN-UG	3	−6566.582	−13,139.163	5	−13,155.103	5	0.053 (LN)
LN-IG	3	−6566.558	−13,139.116	4	−13,155.056	4	0.518 (LN)
UG-LN	3	−6558.861	−13,123.722	1	−13,139.662	1	0.000 (UG)
UG-UG	3	−6571.902	−13,149.805	6	−13,165.745	6	0.000 (UG)
UG-IG	3	−6585.860	−13,177.719	7	−13,193.659	7	0.000 (UG)
IG-LN	3	−7017.739	−14,041.477	13	−14,057.417	13	0.534 (IG)
IG-UG	3	−7017.495	−14,040.989	11	−14,056.929	11	0.350 (IG)
IG-IG	3	−7017.658	−14,041.316	12	−14,057.256	12	0.460 (IG)
LN	2	−6566.767	−13,137.534	3	−13,148.160	3	
UG	2	−7077.964	−14,159.928	14	−14,170.555	14	
IG	2	−7017.931	−14,039.863	10	−14,050.489	10	
Weibull	2	−6658.850	−13,321.699	8	−13,332.326	8	
Logistic	2	−8270.456	−16,544.913	16	−16,555.539	16	
Skew-normal	3	−8148.487	−16,302.974	15	−16,318.914	15	
Skew-t	4	−6744.992	−13,497.983	9	−13,511.923	9	
Transformation kernel		−6564.856					

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 parametric models presented in Table 4. In terms of computational times required by our code to get parameters estimate, the absence of closed-type density formulas (cf. Section 4.2), and the consequent use of the `integrate()` function included in the **stats** package for R, naturally affect the computational burden for ML estimation. For these data, the nine proposed models take 91.79 s, on average, to be estimated ranging from 16.85 for the UG-UG model to 227.42 seconds for the IG-IG model.

By considering the confidence levels of 95% and 99%, Table 5 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 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 the 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 6 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 of the considered confidence level, the best model is the UG-LN; this choice corroborates the results obtained via AIC and BIC (cf. Table 5). It is interesting to note how the estimates, at confidence levels of 95% and 99%, of the CTE from the best UG-LN model exceed the empirical value (9.167% and 30.436%, respectively).

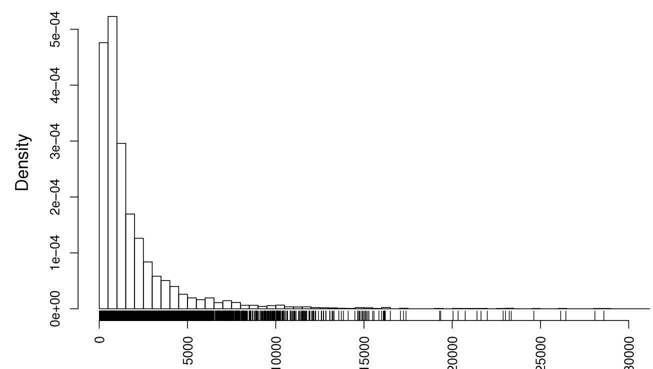


Fig. 6. Automobile insurance claims: histogram.

6.2. 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 **insuranceData** (Wolny-Dominiak and Trzkesiok, 2014).

Table 7 and Fig. 6 present descriptive statistics and the histogram of the data, respectively. As for the previous dataset, we observe right-skewness, leptokurtosis, a large number of small losses and a lower number of very large losses.

In the fashion of Table 4, Table 8 evaluates goodness-of-fit for the competing models. AIC and BIC provide the same ranking apart from an exchange of positions (11th and 12th) 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 conditional distribution, the UG model alone, i.e. without compounding, 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 8). Finally, seven out of the nine proposed models occupy the first nine positions of the ranking. As regards the computational times, the nine proposed models take 164.66 s, on average, to be estimated ranging from 29.25 (UG-UG) to 470.59 (IG-LN) seconds.

Table 5
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	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 6
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.

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 7
Automobile insurance claims: descriptive statistics.

No. of Observations	6773
Mean	1853.03
St. Dev.	2646.91
Skewness	6.24
Kurtosis	87.28
Minimum	9.50
Maximum	60,000.00

Table 9 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 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 level, it can be noted how the empirical VaR at both confidence levels 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 8) works also well in reproducing the empirical VaR.

Finally, Table 10 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 of the considered confidence level, the best model is the UG-IG; this choice corroborates the results obtained via AIC and BIC (cf. Table 8). 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%.

6.3. Norwegian fire claims

The third dataset contains 9181 fire insurance claims, in thousands of Norwegian krones (TNOK), for a Norwegian insurance company for the period 1972–1992. A priority of 500 TNOK was

Table 8
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.

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	
Transformation kernel		−57,125.530					

Table 9
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	VaR						Prop. Viol.		p-value	
	95%	Diff. %	Ranking	99%	Diff. %	Ranking	95%	99%	95%	99%
Empirical	6,356.726			12,052.290						
LN-LN	6,028.198	−5.168	6	11,855.392	−1.634	1	0.057	0.011	0.008	0.606
LN-UG	5,999.926	−5.613	9	11,632.170	−3.486	2	0.058	0.011	0.004	0.383
LN-IG	6,143.101	−3.361	3	12,938.464	7.353	5	0.055	0.008	0.059	0.140
UG-LN	6,253.580	−1.623	2	14,057.200	16.635	10	0.052	0.007	0.395	0.011
UG-UG	6,121.264	−3.704	4	16,117.551	33.730	13	0.056	0.004	0.036	0.000
UG-IG	6,272.222	−1.329	1	12,770.985	5.963	4	0.052	0.009	0.494	0.223
IG-LN	6,703.774	5.460	7	13,051.439	8.290	7	0.046	0.008	0.132	0.140
IG-UG	6,705.112	5.481	8	13,047.294	8.256	6	0.046	0.008	0.132	0.140
IG-IG	6,812.773	7.174	10	13,982.679	16.017	8	0.044	0.007	0.033	0.011
LN	6,106.883	−3.930	5	12,670.840	5.132	3	0.056	0.009	0.031	0.336
UG	5,525.987	−13.069	14	8,478.556	−29.652	12	0.065	0.027	0.000	0.000
IG	6,826.756	7.394	11	14,046.672	16.548	9	0.044	0.007	0.033	0.011
Weibull	5,763.323	−9.335	13	9,115.029	−24.371	11	0.061	0.023	0.000	0.000
Logistic	4,293.018	−32.465	15	5,922.370	−50.861	15	0.096	0.058	0.000	0.000
Skew-normal	5,783.652	−9.015	12	7,107.775	−41.026	14	0.061	0.041	0.000	0.000
Skew-t	14,479.716	127.786	16	21,223.668	76.097	16	0.007	0.002	0.000	0.000

in force; thus, no claims below this limit were recorded. This dataset is available in the R package **ReIns** (Reynkens et al., 2017) and was studied, among others, in Beirlant et al. (1996, 2006) and Brazauskas and Kleefeld (2016). In particular, Brazauskas and Kleefeld (2016) emphasize the need of a highly skewed and heavy-tailed model to capture the distributional characteristics of these claims. In our analysis, data are shifted on the left by 500 TNOK to remove the impact of the priority.

For the sake of readability, Fig. 7 displays the histogram of the data only up to 40,000 TNOK; this leaves out 36 claims, with the largest one being almost half a million TNOK (cf. Table 11).

Although these 36 claims are only about the 1% of the available claims, they represent about the 20% of the total amount of claims paid. Of particular interest is the very high kurtosis (1,521.25), which further highlights the strong heavy tails of the empirical distribution (cf. Table 11).

AIC and BIC values in Table 12 provide the same ranking. The best three models (UG-UG, UG-LN, and UG-IG) are compound models having the UG as conditional distribution. The simple UG model has the 10th position only, and the p-values from the LR tests (see the last column of Table 12) show how its compound versions are a significant improvement. Finally, six out of the nine proposed models occupy the first eight positions of the ranking. As regards

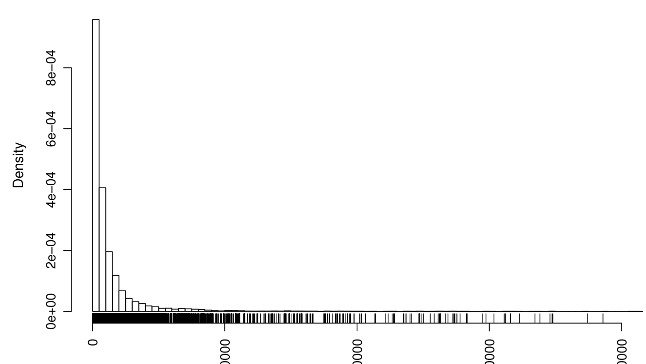


Fig. 7. Norwegian fire claims: histogram.

the computational times, the nine proposed models take 122.18 s, on average, to be estimated, ranging from 13.37 (IG-IG) to 258.99 (UG-IG) seconds.

From Table 13, which reports empirical and estimated VaR values, we observe that, when the 95% confidence level is considered,

Table 10
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.

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 11
Norwegian fire claims: descriptive statistics.

No. of Observations	9181
Mean	1717.22
St. Dev.	7759.97
Skewness	30.85
Kurtosis	1,521.25
Minimum	0.01
Maximum	464,865.01

the best model is the UG-IG (the third model in the rankings of Table 12); the UG-IG model underestimates the empirical VaR by the 3.987%. At the 99% confidence level, the best model is the UG-UG, which underestimates the empirical VaR by the 8.764%. Considering the backtesting procedure, and taking $\alpha = 0.05$, it can be noted how the empirical VaR, at both confidence levels, does not match up very well with the estimates of the VaR provided by the majority of the competing models (apart from the Weibull at the 95% confidence level having a p-value of 0.384). Instead, the compounds of the UG distribution show satisfactory results. One of them, the UG-UG, is the only one working well at the 99%.

Finally, Table 14 reports the model comparisons in terms of CTE. Regardless of the considered confidence level, also here the best model is the UG-UG which overestimates (underestimates) the empirical CTE by the 2.679% (0.334%) at the 95% (99%) confidence level. Summarizing, the UG-UG model can be considered as the best model for these data being the first, in the ranking induced by AIC and BIC (cf. Table 12), and being the only one providing good results for VaR and CTE at the considered confidence levels.

7. 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 compound 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 compound model. The proposed models have been fitted to three well-known and publicly available insurance loss datasets 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/Compound.R 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 compound models are reasonably good models compared to the competitors. However, as pointed out by one of the reviewers, when the logarithm is applied to the considered right-skewed loss data – as sometimes happens in the literature (see, e.g., Eling, 2012, 2014) – the impact of our proposal is slightly reduced, and some classical distributions, such as the logistic, the skew-normal, and the skew-t, work better. The logarithmic transformation tends to symmetrize the distribution of the data (lower skewness) and to reduce the impact of large losses (lower kurtosis). These aspects are in countertrend with the peculiarities of our approach. The results of the analyses on the log-data are not reported in the paper because, in addition to the considerations above, transformed data are more difficult to interpret and the use of any transformation involves the further issue about the choice of the best scale for the available data.

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 conditional model can be parameterized with respect to the mode and to another parameter governing the distribution variability.

Naturally, the proposed models also have some drawbacks. Though different compound models can be obtained by following the approach discussed in Section 2, those discussed in this work have specific features that the data should reflect to outperform other existing distributions. The absence of a tuning skewness parameter makes our models not suitable for symmetric data. Only two of the proposed models possess a closed-form expression for the pdf (cf. Section 4.2). This makes inference based on the maximum likelihood approach more challenging, though the available numerical methods reduce the computational burden in the estimation phase (cf. Section 4.5). Moreover, not all the moments are available in a closed-form (cf. Section 4.4) and this, from an

Table 12
Norwegian fire 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.

Model	# par.	Log-lik.	AIC	Ranking	BIC	Ranking	LR test (null model)
LN-LN	3	−75,336.108	−150,678.217	7	−150,699.592	7	0.000 (LN)
LN-UG	3	−75,336.555	−150,679.109	8	−150,700.484	8	0.001 (LN)
LN-IG	3	−75,336.037	−150,678.074	6	−150,699.448	6	0.000 (LN)
UG-LN	3	−73,951.517	−147,909.035	2	−147,930.409	2	0.000 (UG)
UG-UG	3	−73,857.587	−147,721.173	1	−147,742.548	1	0.000 (UG)
UG-IG	3	−74,114.091	−148,234.183	3	−148,255.558	3	0.000 (UG)
IG-LN	3	−98,922.630	−197,851.259	13	−197,872.634	13	0.000 (IG)
IG-UG	3	−99,012.248	−198,030.495	14	−198,051.870	15	0.005 (IG)
IG-IG	3	−99,014.533	−198,035.066	15	−198,056.441	16	0.077 (IG)
LN	2	−75,341.823	−150,687.645	9	−150,701.895	9	
UG	2	−77,565.327	−155,134.653	10	−155,148.903	10	
IG	2	−99,016.093	−198,036.186	16	−198,050.436	14	
Weibull	2	−74,169.269	−148,342.538	4	−148,356.788	4	
Logistic	2	−85,081.599	−170,167.197	11	−170,181.447	11	
Skew-normal	3	−86,964.836	−173,935.672	12	−173,957.047	12	
Skew-t	4	−75,126.893	−150,261.787	5	−150,281.162	5	
Transformation kernel		−74,010.569					

Table 13
Norwegian fire 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	VaR						Prop. Viol.		p-value	
	95%	Diff. %	Ranking	99%	Diff. %	Ranking	95%	99%	95%	99%
Empirical	5,889.010			19,317.410						
LN-LN	13,065.409	121.861	14	55,025.316	184.848	16	0.018	0.003	0.000	0.000
LN-UG	13,091.856	122.310	15	54,765.842	183.505	14	0.018	0.003	0.000	0.000
LN-IG	13,103.524	122.508	16	54,860.269	183.994	15	0.018	0.003	0.000	0.000
UG-LN	5,607.252	−4.784	3	13,786.748	−28.630	3	0.052	0.016	0.420	0.000
UG-UG	5,536.556	−5.985	4	17,624.382	−8.764	1	0.053	0.011	0.218	0.176
UG-IG	5,654.243	−3.987	1	12,136.602	−37.173	5	0.051	0.019	0.537	0.000
IG-LN	142.477	−97.581	12	3,381.696	−82.494	12	0.809	0.092	0.000	0.000
IG-UG	143.834	−97.558	10	3,505.824	−81.851	10	0.809	0.088	0.000	0.000
IG-IG	142.925	−97.573	11	3,403.481	−82.381	11	0.809	0.091	0.000	0.000
LN	13,010.784	120.933	13	53,770.594	178.353	13	0.018	0.003	0.000	0.000
UG	5,127.174	−12.937	6	7,877.257	−59.222	7	0.057	0.033	0.001	0.000
IG	144.291	−97.550	9	3,611.572	−81.304	9	0.000	0.000	0.000	0.000
Weibull	6,135.640	4.188	2	12,419.184	−35.710	4	0.048	0.019	0.384	0.000
Logistic	4,380.748	−25.611	7	6,321.563	−67.275	8	0.069	0.046	0.000	0.000
Skew-normal	11,264.212	91.275	8	13,850.233	−28.302	2	0.020	0.016	0.000	0.000
Skew-t	5,192.543	−11.827	5	9,872.800	−48.892	6	0.057	0.025	0.003	0.000

Table 14
Norwegian fire 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.

Model	95%			99%		
	CTE	Diff. %	Ranking	CTE	Diff. %	Ranking
Empirical	17,890.069			50,793.358		
LN-LN	51,030.498	185.245	14	155,613.754	206.366	11
LN-UG	50,911.830	184.582	13	155,156.082	205.465	10
LN-IG	51,496.096	187.847	15	157,930.490	210.927	12
UG-LN	11,306.695	−36.799	4	23,512.892	−53.709	3
UG-UG	17,410.828	−2.679	1	50,963.141	0.334	1
UG-IG	9,870.214	−44.829	6	17,840.886	−64.876	4
IG-LN	32,516.174	81.755	9	159,784.641	214.578	13
IG-UG	34,064.013	90.407	10	167,464.893	229.698	14
IG-IG	36,696.805	105.124	11	180,687.657	255.731	15
LN	49,660.719	177.588	12	150,179.761	195.668	9
UG	6,833.697	−61.802	7	9,587.069	−81.125	7
IG	4.6e+110	2.6e+108	16	2.3e+111	4.5e+108	16
Weibull	10,191.516	−43.033	5	17,527.699	−65.492	5
Logistic	5,592.754	−68.738	8	7,505.680	−85.223	8
Skew-normal	12,846.310	−28.193	3	15,136.482	−70.200	6
Skew-t	20,482.752	14.492	2	25,300.944	−50.188	2

Please cite this article in press as: Punzo A., et al., Compound unimodal distributions for insurance losses. Insurance: Mathematics and Economics (2017), https://doi.org/10.1016/j.insmatheco.2017.10.007.

$$\frac{\partial \ln\left[p\left(x;\theta,\gamma,\nu\right)\right]}{\partial \nu}=\frac{1}{2\nu^2}\left\{\frac{K_1\left(\frac{\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}}{\nu}\right)\left\{\gamma\left[2+(3+\gamma)\nu\right]+\nu(\ln x-\ln\theta)^2\right\}}{\gamma K_0\left(\frac{\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}}{\nu}\right)\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}}-\frac{2+\nu\left[9\left(1+\nu\right)+\sqrt{1+3\nu}\right]}{\left(1+3\nu\right)^{\frac{3}{2}}}\right\}.$$

Box I.

inferential point of view, does not make possible the use of the method of moments.

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 a compound model introduced herein. Finally, extension of our approach to the case of data defined on a compact or finite support may be considered too by using, for example, the unimodal beta or the discrete unimodal beta, respectively, as conditional distributions (see Bagnato and Punzo, 2013; Punzo, 2010; and Mazza and Punzo, 2011, 2014, 2015).

Acknowledgment

The work of Antonello Maruotti is partially supported by the project on "Environmental processes and human activities: capturing their interactions via statistical methods (EPHASTAT)", funded by MIUR (Italian Ministry of Education, University and Scientific Research) (20154XBK23).

Appendix. First partial derivatives of the log pdf of the LN-UG and LN-IG distributions

The first order partial derivatives with respect to θ, γ, and ν of the logarithm of the pdf of the LN-UG model in (8) are

$$\frac{\partial \ln\left[p\left(x;\theta,\gamma,\nu\right)\right]}{\partial \theta}=\frac{1}{\theta}\left\{\frac{\left(2+3\nu\right)\left(\ln x-\ln\theta\right)}{2\gamma+\nu(\ln x-\ln\theta)^2}\right.$$
$$\left.+\frac{K_{\frac{1}{2}+\frac{1}{\nu}}\left(\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right)\left(\ln x-\ln\theta\right)}{K_{\frac{3}{2}+\frac{1}{\nu}}\left(\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right)\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}}-1\right\},$$

$$\frac{\partial \ln\left[p\left(x;\theta,\gamma,\nu\right)\right]}{\partial \gamma}=\frac{1}{2\gamma+\nu(\ln x-\ln\theta)^2}$$
$$\times\left\{\frac{\left(1+\nu\right)\left(\ln x-\ln\theta\right)^2}{\gamma}\right.$$
$$\left.-\frac{K_{\frac{1}{2}+\frac{1}{\nu}}\left(\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right)\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}}{K_{\frac{3}{2}+\frac{1}{\nu}}\left(\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right)}-1\right\},$$

and

$$\frac{\partial \ln\left[p\left(x;\theta,\gamma,\nu\right)\right]}{\partial \nu}$$
$$=-\frac{1}{2\nu^2}\left\{-\frac{2\gamma K_{\frac{5}{2}+\frac{1}{\nu}}\left[\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right]}{K_{\frac{3}{2}+\frac{1}{\nu}}\left[\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right]\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}}\right.$$
$$+2\left(1+\nu+\ln\frac{\gamma}{\nu}\right)+\ln\left[\frac{\nu}{2\gamma+\nu(\ln x-\ln\theta)^2}\right]$$
$$\left.-2\psi\left(1+\frac{1}{\nu}\right)-\frac{2K^{(1,0)}_{-\frac{3}{2}-\frac{1}{\nu}}\left(\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right)}{K_{\frac{3}{2}+\frac{1}{\nu}}\left(\sqrt{\frac{2\gamma+\nu(\ln x-\ln\theta)^2}{\nu}}\right)}\right\},$$

where $\psi\left(\cdot\right)$ denotes the digamma function and $K_{\alpha}^{(1,0)}\left(y\right)$ denotes the first derivative of the modified Bessel function of the second kind with respect to α.

The first order partial derivatives with respect to θ, γ, and ν of the logarithm of the pdf of the LN-IG model in (8) are

$$\frac{\partial \ln\left[p\left(x;\theta,\gamma,\nu\right)\right]}{\partial \theta}$$
$$=\frac{\left[1+\left(3+\gamma\right)\nu\right]K_1\left(\frac{\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left(\gamma+\nu(\ln x-\ln\theta)^2\right)}{\gamma}}}{\nu}\right)\left(\ln x-\ln\theta\right)}{\theta\gamma K_0\left(\frac{\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}}{\nu}\right)\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}}$$
$$-\frac{1}{\theta},$$

$$\frac{\partial \ln\left[p\left(x;\theta,\gamma,\nu\right)\right]}{\partial \gamma}$$
$$=\frac{K_1\left(\frac{\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}}{\nu}\right)\left[-\gamma^2+\left(1+3\nu\right)\left(\ln x-\ln\theta\right)^2\right]}{2\gamma^2K_0\left(\frac{\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}}{\nu}\right)\sqrt{\frac{\left[1+(3+\gamma)\nu\right]\left[\gamma+\nu(\ln x-\ln\theta)^2\right]}{\gamma}}}$$
$$-\frac{1}{2\gamma},$$

and $\frac{\partial \ln[p(x;\theta,\gamma,\nu)]}{\partial\nu}$ is given in Box I.

References

Abu Bakar, S.A., Hamzah, N.A., Maghsoudi, M., Nadarajah, S., 2015. Modeling loss data using composite models. Insurance Math. Econom. 61, 146–154.

Adcock, C., Eling, M., Loperfido, N., 2015. Skewed distributions in finance and actuarial science: a review. Eur. J. Finance 21 (13–14), 1253–1281.

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

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

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

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

Bee, M., 2017. Density approximations and VaR computation for compound Poisson-lognormal distributions. Comm. Statist. Simulation Comput. 46 (3), 1825–1841.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2006. Statistics of Extremes: Theory and Applications. John Wiley & Sons.

Beirlant, J., Teugels, J.L., Vynckier, P., 1996. Practical Analysis of Extreme Values. Leuven University Press.

Bernardi, M., Maruotti, A., Petrella, L., 2012. Skew mixture models for loss distributions: a Bayesian approach. Insurance Math. Econom. 51, 617–623.

Bolancé, C., Guillen, M., Nielsen, J.P., 2003. Kernel density estimation of actuarial loss functions. Insurance Math. Econom. 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 Math. Econom. 43, 386–393.

Brazauskas, V., Kleefeld, A., 2016. Modeling severity and measuring tail risk of norwegian fire claims. N. Am. Actuar. J. 20 (1), 1–16.

Chen, S.X., 2000. Probability density function estimation using gamma kernels. Ann. Inst. Statist. Math. 52 (3), 471–480.

Cooray, K., Ananda, M.M.A., 2005. Modeling actuarial data with a composite lognormal-Pareto model. Scand. Actuar. J. 2005 (5), 321–334.

Delignette-Muller, M.L., Dutang, C., 2015. **fitdistrplus**: An R package for fitting distributions. J. Stat. Softw. 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 Math. Econom. 51 (2), 239–248.

Eling, M., 2014. Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models? Insurance Math. Econom. 59, 45–56.

Embrechts, P., Klüppelberg, C., Mikosch, T., 2003. Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin, p. 648.

Frees, E.W., 2010. Regression Modeling with Actuarial and Financial Applications. In: International Series on Actuarial Science, Cambridge University Press, Cambridge.

Frees, E.W., Valdez, E.A., 1998. Understanding relationships using copulas. N. Am. Actuar. J. 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.

Greenwood, M., Yule, G.U., 1920. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. J. Roy. Statist. Soc. 83, 255–279.

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. In: Wiley Series in Probability and Statistics, vol. 249, Wiley, New York.

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

Johnson, N.L., Kotz, S., 1970. Continuous Univariate Distributions. In: Distributions in Statistics, no. v. 2, Hougton Mifflin.

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. In: Wiley Series in Probability and Statistics, vol. 715, Wiley, Hoboken, New Jersey.

Kojadinovic, I., Yan, J., 2010. Modeling multivariate distributions with continuous margins using the **copula** R package. J. Stat. Softw. 34 (9), 1–20.

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

MacDonald, I.L., 2014. Numerical maximisation of likelihood: A neglected alternative to EM? Internat. Statist. Rev. 82 (2), 296–308.

Mazza, A., Punzo, A., 2011. Discrete beta kernel graduation of age-specific demographic indicators. In: Ingrassia, S., Rocci, R., Vichi, M. (Eds.), New Perspectives in Statistical Modeling and Data Analysis. In: Studies in Classification, Data Analysis and Knowledge Organization, Springer-Verlag, Berlin Heidelberg, pp. 127–134.

Mazza, A., Punzo, A., 2014. **DBKGrad**: An R package for mortality rates graduation by discrete beta kernel techniques. J. Stat. Softw. 57 (Code Snippet 2), 1–18.

Mazza, A., Punzo, A., 2015. Bivariate discrete beta kernel graduation of mortality data. Lifetime Data Anal. 21 (3), 419–433.

McDonald, J.B., Butler, R.J., 1987. Some generalized mixture distributions with an application to unemployment duration. Rev. Econ. Stat. 232–240.

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

Nadarajah, S., Abu Bakar, S.A., 2014. New composite models for the Danish fire insurance data. Scand. Actuar. J. 2014 (2), 180–187.

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.

Punzo, A., 2010. Discrete beta-type models. In: Locarek-Junge, H., Weihs, C. (Eds.), Classification As a Tool for Research. In: Studies in Classification, Data Analysis and Knowledge Organization, Springer-Verlag, Berlin Heidelberg, pp. 253–261.

Punzo, A., 2017. A new look at the inverse Gaussian distribution, arXiv.org e-print 1707.04400, available at: https://arxiv.org/abs/1707.04400.

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/.

Reynkens, T., Verbelen, R., Bardoutsos, A., Cornilly, D., Goegebeur, Y., Herrmann, K., 2017. **ReIns**: Functions from "Reinsurance: Actuarial and Statistical Aspects". Version 1.0.4 (2017-06-10). URL https://cran.r-project.org/web/packages/ReIns/index.html.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.

Scollnik, D.P.M., Sun, C., 2012. Modeling with Weibull-Pareto models. N. Am. Actuar. J. 16 (2), 260–272.

Shevchenko, P.V., 2010. Calculation of aggregate loss distributions. J. Oper. Risk 5 (2), 3–40.

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.

Tahir, M.H., Cordeiro, G.M., 2016. Compounding of distributions: a survey and new generalized classes. J. Stat. Distrib. Appl. 3 (1), 13.

Vernic, R., 2006. Multivariate skew-normal distributions with applications in insurance. Insurance Math. Econom. 38 (2), 413–426.

Watanabe, M., Yamaguchi, K., 2004. The EM Algorithm and Related Statistical Models. In: Statistics: A Series of Textbooks and Monographs, Taylor & Francis, New York.

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.

Zhang, H., Liu, Y., Li, B., 2014. Notes on discrete compound Poisson model with applications to risk theory. Insurance Math. Econom. 59, 325–336.

Please cite this article in press as: Punzo A., et al., Compound unimodal distributions for insurance losses. Insurance: Mathematics and Economics (2017), https://doi.org/10.1016/j.insmatheco.2017.10.007.