

Article

A New Heavy Tailed Class of Distributions Which Includes the Pareto

Deepesh Bhati ¹, Enrique Calderín-Ojeda ^{2,*} and Mareeswaran Meenakshi ¹

¹ Department of Statistics, Central University of Rajasthan, Kishangarh 305817, India; deepesh.bhati@curaj.ac.in (D.B.); mgmari5595@gmail.com (M.M.)

² Department of Economics, Centre for Actuarial Studies, University of Melbourne, Melbourne, VIC 3010, Australia

* Correspondence: ecalderin@unimelb.edu.au

Received: 23 June 2019; Accepted: 10 September 2019; Published: 20 September 2019



Abstract: In this paper, a new heavy-tailed distribution, the mixture Pareto-loggamma distribution, derived through an exponential transformation of the generalized Lindley distribution is introduced. The resulting model is expressed as a convex sum of the classical Pareto and a special case of the loggamma distribution. A comprehensive exploration of its statistical properties and theoretical results related to insurance are provided. Estimation is performed by using the method of log-moments and maximum likelihood. Also, as the modal value of this distribution is expressed in closed-form, composite parametric models are easily obtained by a mode matching procedure. The performance of both the mixture Pareto-loggamma distribution and composite models are tested by employing different claims datasets.

Keywords: pareto distribution; excess-of-loss reinsurance; loggamma distribution; composite models

1. Introduction

In general insurance, pricing is one of the most complex processes and is no easy task but a long-drawn exercise involving the crucial step of modeling past claim data. For modeling insurance claims data, finding a suitable distribution to unearth the information from the data is a crucial work to help the practitioners to calculate fair premiums. In actuarial statistics and finance, the classical Pareto distribution has been deemed better than other models because it provides a good description of the random behaviour of large claims. Loss data mostly have several characteristics such as unimodality, right skewness, and a thick right tail. To accommodate these features in a single model, many probability models have been proposed in the literature.

Over the last decades, different approaches to derive new classes of probability distributions that could provide more flexibility when modelling large losses has been added to the literature. This includes transformation method, the composition of two or more distributions, compounding of distributions or a finite mixture of distributions among other methodologies. In particular for generalizations of the classical Pareto distribution, the reader is referred to the Stoppa distribution (see [Stoppa 1990](#)), the Pareto positive stable distribution (see [Sarabia and Prieto 2009](#)), the Pareto ArcTan distribution (see [Gómez-Déniz and Calderín-Ojeda 2015](#)) and the generalized Pareto distribution proposed in [Ghitany et al. \(2018\)](#). Obviously, adding a parameter to the parent model complicates the parameter estimation.

In this paper, a probabilistic family, the mixture Pareto-loggamma distribution, which belongs to the heavy-tailed class of probabilistic models is introduced. This family is derived by using of an exponential transformation of the generalized Lindley distribution. It is expressed as a convex sum of the classical Pareto and a special case of the loggamma distribution similar to the one presented in

Gómez-Déniz and Calderín-Ojeda (2014). The former model is obtained as a particular case whereas the latter one is a limiting case. We further present expressions of statistical and actuarial measures such as moments, variance, cumulative distribution function, hazard rate function, VaR, TVaR and limited expected values. For many of these distributional characteristics, closed-form expressions are obtained. Estimation of the parameters of this distribution can be easily calculated by the maximum likelihood method by using the numerical search of the maximum and also by the method of log-moments.

In many instances, composite distributions give a reasonably good fit as compared to classical distributions since the former models have the advantage of accommodating both low magnitude values with a high frequency as well as large magnitude figures with low frequency. Cooray and Ananda (2005) introduced a composite lognormal-Pareto distribution with restricted mixing weights. Scollnik (2007) improved the composite lognormal-Pareto model by allowing flexible mixing weights. Bakar et al. (2015) considered several probabilistic models in place of lognormal and Pareto distribution using the approach discussed in Scollnik (2007). Recently, Calderín-Ojeda and Kwok (2016) introduced a new class of composite model using mode-matching procedure. They derived composite models using lognormal and Weibull distributions with a Stoppa model, a generalization of the Pareto distribution. As there exists a closed-form expression for the modal value of this new model, in this article, we use the mode-matching procedure to derive new composite models based on this distribution.

The structure of the paper is organized as follows. In Section 2, we present the genesis of the new distribution, and discuss its relationship with other distributions. Besides, the most relevant distributional properties are studied in the same section. Also, different estimation methods are examined. Finally, composite models based on the proposed distribution are derived. Next, in Section 3, some results related to insurance are displayed. Numerical applications are displayed in Section 4 together with the derivation of some income indices. The last section concludes the paper.

2. The *MPLG* Distribution and Its Properties

It can be easily seen that the following expression

$$f_X(x|\theta, \lambda, x_0) = \frac{\theta^2}{x(\theta + \lambda)} \left(\frac{x}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x}{x_0}\right)\right) \quad x \geq x_0 > 0 \tag{1}$$

with $\lambda \geq 0$ and $\theta > 0$ is a genuine probability density function (pdf). Here θ, λ are shape parameter and x_0 is scale parameter (see the Appendix A). Note that the pdf (1) can be written as a convex sum of the classical Pareto distribution and a special case of the loggamma distribution. The former model is obtained when $\lambda = 0$ and the latter one for $\lambda \rightarrow \infty$. Observe that the pdf of the loggamma distribution considered in this work is

$$f_X(x|\theta, \gamma, x_0) = \frac{\theta\gamma}{x_0 \Gamma(\gamma)} \left(\frac{x}{x_0}\right)^{-(\theta+1)} \log\left(\frac{x}{x_0}\right)^{\gamma-1},$$

with $\theta, \gamma > 0$ and $x \geq x_0$. In our model it is assumed that $\gamma = 2$.

The cumulative distribution function (cdf), survival function (sf) and hazard rate function of a random variable (rv) with pdf (1) are respectively given, for $x \geq x_0$, as

$$F_X(x) = 1 - \frac{\left(\theta + \lambda + \theta\lambda \log\left(\frac{x}{x_0}\right)\right)}{\theta + \lambda} \left(\frac{x}{x_0}\right)^{-\theta}, \tag{2}$$

$$\bar{F}_X(x) = \frac{\left(\theta + \lambda + \theta\lambda \log\left(\frac{x}{x_0}\right)\right)}{\theta + \lambda} \left(\frac{x}{x_0}\right)^{-\theta}, \tag{3}$$

$$h_X(x) = \frac{\theta}{x} \left(1 - \frac{\lambda}{\theta + \lambda + \theta \lambda \log\left(\frac{x}{x_0}\right)} \right). \tag{4}$$

Henceforth, a continuous rv X that follows (1) will be called as mixture Pareto-loggamma distribution and denoted as $\mathcal{MPLG}(\theta, \lambda, x_0)$. The shape of the pdf given in (1) is shown in Figure 1 for different values of the parameters θ and λ and a fixed x_0 . It is observable that the larger is the value of θ , the thinner is the tail. On the contrary, the thickness of the tail decreases when λ drops.

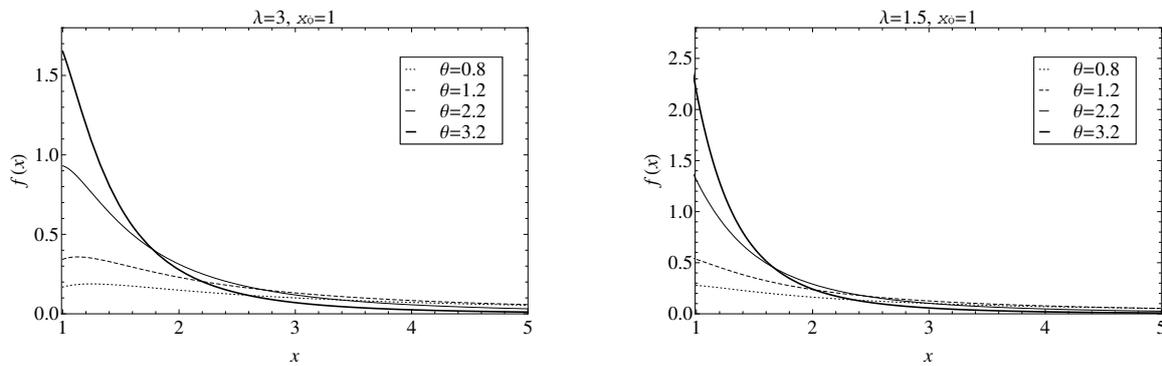


Figure 1. Graphs of the pdf (1) for different values of parameter θ , λ and fixed value of x_0 .

Undoubtedly, the single parameter Pareto distribution is one of the most attractive distributions in statistics; a power-law probability distribution found in a large number of real-world situations inside and outside the field of economics. Furthermore, it is usually used as a basis for Excess of Loss quotations as it gives a pretty good description of the random behavior of large losses. In this sense, many probability distributions can be used for modelling single loss amounts. Then, if the loss is assumed to follow the pdf (1), we have that θ defines the tail behavior of the distribution.

Below, the hazard rate function has been plotted in Figure 2 for different values of the parameters θ and λ for fixed x_0 . It is observable that the hazard rate function increases when the parameter θ grows and the parameter λ declines.

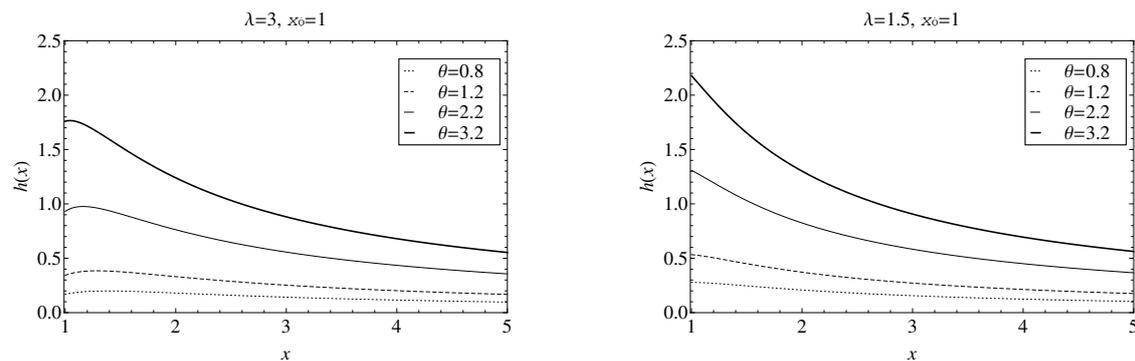


Figure 2. Graphs of the hazard rate function (4) for different values of parameter θ , λ and for fixed value of x_0 .

The next result shows the relationship between the \mathcal{MPLG} and the Generalized Lindley distribution, which is extensively used in lifetime analysis and reliability.

Remark 1. The $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution can be also obtained by taking $X = x_0 \exp\{Y\}$, where Y follows a generalized Lindley distribution with density

$$g_Y(y) = \frac{\theta^2}{(\theta + \lambda)} (1 + \lambda y) \exp\{-\theta y\}, \quad y > 0, \lambda, \theta > 0. \tag{5}$$

From this result, it is straightforward to generate random variates from the $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution by observing that the pdf given in (5), can be rewritten as

$$w(y) = p a_1(y) + (1 - p)a_2(y),$$

where $p = \frac{\theta}{\theta + \lambda}$, $a_1(\cdot)$ is the pdf of the exponential distribution with mean $1/\theta$ and $a_2(\cdot)$ is the pdf of the Erlang distribution with shape parameter 2 and rate parameter θ . Then, a random variate from the \mathcal{MPLG} distribution, x , can be generated following a modification of the algorithm presented in Ghitany et al. (2008) as shown below.

- Generate two random numbers u_1 and u_2 from the standard uniform distribution, $U(0, 1)$.
- Generate random variates \tilde{a}_1 from the exponential distribution with mean $1/\theta$ and \tilde{a}_2 from the Erlang distribution with shape parameter 2 and rate parameter θ by using u_1 .
- If $u_2 \leq p$, then set $y = \tilde{a}_1$; otherwise, set $y = \tilde{a}_2$.
- Generate $x = x_0 \exp\{y\}$.

Observe that an analogous algorithm could be implemented by using the fact that (1) is a convex sum of the densities of Pareto and loggamma distributions.

2.1. Moments and Log-Moments

The r th raw moment of $\mathcal{MPLG}(\theta, \lambda, x_0)$ is given by

$$\mathbb{E}(X^r) = \frac{\theta^2 x_0^r (\theta + \lambda - r)}{(\theta + \lambda)(r - \theta)^2} \quad \theta \neq r.$$

In particular, we have that

$$\begin{aligned} \mu = \mathbb{E}(X) &= \frac{\theta^2 x_0 (\theta + \lambda - 1)}{(\theta - 1)^2 (\theta + \lambda)}, & \theta \neq 1 \text{ and} \\ \mathbb{E}(X^2) &= \frac{\theta^2 x_0^2 (\theta + \lambda - 2)}{(\theta + \lambda)(2 - \theta)^2}, & \theta \neq 2. \end{aligned}$$

The variance is given by

$$\mathbb{V}(X) = \frac{\theta^2 x_0^2 (\theta + \lambda - 2)}{(\theta + \lambda)(2 - \theta)^2} - \left(\frac{\theta^2 x_0 (\theta + \lambda - 1)}{(\theta + \lambda)(1 - \theta)^2} \right)^2, \quad \theta \neq 1, 2. \tag{6}$$

Remark 1 can be used to obtain the r th raw log-moment of the density (1)

$$l_s = \mathbb{E} \left(\log \left(\frac{X}{x_0} \right) \right)^s = \frac{s! (\theta + \lambda + \lambda s)}{\theta^s (\theta + \lambda)}. \tag{7}$$

The moment estimator of θ and λ with known x_0 can be simply derived from (7).

2.2. Unimodality

The modal value of the \mathcal{MPLG} distribution can be found by taking the first derivative of the pdf (1),

$$f'(x) = \frac{\theta^2 x_0^\theta x^{-\theta-2}}{\theta + \lambda} \left(\lambda - (1 + \theta) \left(1 + \lambda \log \left(\frac{x}{x_0} \right) \right) \right), \tag{8}$$

then, by setting (8) equal to zero and solving for x , the mode is $x_m = x_0 \exp \left\{ \frac{\lambda - \theta - 1}{\lambda(\theta + 1)} \right\}$. If $\lambda \leq \theta + 1$, then the mode is at $x_m = x_0$.

2.3. Stochastic Ordering

Many parametric families of distributions can be ordered by some stochastic orders according to the value of their parameters. A rv X_1 is said to be stochastically smaller ($X_1 \leq_{st} X_2$) than X_2 if $F_{X_1}(x) \geq F_{X_2}(x)$ for all x . The formal definition of stochastic ordering is

Definition 1. Let X_1 and X_2 be continuous random variables with densities f_1 and f_2 , respectively, such that $\frac{f_2(x)}{f_1(x)}$ is non-decreasing over the union of the supports of X_1 and X_2 . Then X_1 is said to be smaller than X_2 in the likelihood ratio order. ($X_1 \leq_{lr} X_2$) (see Section 1.C of [Shaked and Shanthikumar \(2007\)](#)).

Similarly, a rv can be also ordered for stochastic (distribution function), hazard rate and mean excess orders when the following results hold:

- (i) Stochastic order ($X_1 \leq_{st} X_2$) if $F_{X_1}(x) \geq F_{X_2}(x)$ for all x .
- (ii) Hazard rate order ($X_1 \leq_{hr} X_2$) if $h_{X_1}(x) \geq h_{X_2}(x)$ for all x .
- (iii) Mean excess order ($X_1 \leq_{me} X_2$) if $e_{X_1}(x) \geq e_{X_2}(x)$ for all x , where $e_X(\cdot)$ is mean excess function given in expression (17).

Using Theorem 1.C.1 and Theorem 2.A.1 of [Shaked and Shanthikumar \(2007\)](#), the above stochastic orders hold following implications result

$$\begin{aligned} (X_1 \leq_{lr} X_2) &\Rightarrow (X_1 \leq_{hr} X_2) \Rightarrow (X_1 \leq_{me} X_2) \\ &\Downarrow \\ &(X_1 \leq_{st} X_2). \end{aligned} \tag{9}$$

This gives rise to Proposition 1.

Proposition 1. Let X_1 and X_2 be two rv's having \mathcal{MPLG} distribution with parameters $(\theta_i, \lambda_i, x_{0i})$ $i = 1, 2$. Then following results hold

- i. If $x_{01} = x_{02}$, $\theta_1 \geq \theta_2$ and $\lambda_1 \leq \lambda_2$, then $X_1 \leq_{lr} X_2$, $X_1 \leq_{hr} X_2$ and $X_1 \leq_{st} X_2$
- ii. If $x_{01} \leq x_{02}$, $\theta_1 = \theta_2$ and $\lambda_1 = \lambda_2$, then $X_1 \leq_{lr} X_2$, $X_1 \leq_{hr} X_2$ and $X_1 \leq_{st} X_2$.

Proof. See the Appendix A. \square

2.4. Integrated Tail Distribution and Equilibrium Hazard Rate

As we have already seen previously, the proposed distribution can be viewed as a mixture of two thick-tailed distributions, the Pareto and loggamma models, thus it can be used to describe events that include heavy-tailed behaviour. In this particular, we define another distribution function called Integrated tail distribution (ITD) (also known as equilibrium distribution) $F_e(x)$ which often appears in insurance fields (see [Yang 2004](#)). The ITD has many interesting applications, e.g., approximation of the ruin function (see [Yang 2004](#)) or characterisation of the tail of the distribution (see [Su and Tang 2003](#)) just to name a few. Hence, for $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution bluewith $\theta > 1$, ITD is obtained as

$$\begin{aligned} F_e(x) &= \frac{1}{\mathbb{E}(X)} \int_{x_0}^x \bar{F}(y) dy \\ &= \frac{(\theta - 1)^2(\theta + \lambda)}{\theta^2 x_0 (\theta + \lambda - 1)} \left(\frac{\theta \lambda \left(x_0 - x \left(\frac{x}{x_0} \right)^{-\theta} \left(1 - (1 - \theta) \log \left(\frac{x}{x_0} \right) \right) \right)}{(1 - \theta)^2 (\theta + \lambda)} + \frac{x \left(\frac{x}{x_0} \right)^{-\theta} - x_0}{1 - \theta} \right) \\ &= \frac{\left(\frac{x}{x_0} \right)^{-\theta} \left((\theta(1 - \theta - 2\lambda) + \lambda) \left(x - x_0 \left(\frac{x}{x_0} \right)^\theta \right) + (1 + \theta)\theta \lambda x \log \left(\frac{x}{x_0} \right) \right)}{\theta^2 x_0 (\theta + \lambda - 1)}. \end{aligned} \tag{10}$$

The associated equilibrium hazard rate $r_e(x) = (-\log(1 - F_e(x)))'$, for $\theta > 1$ is given by

$$r_e(x) = \frac{(\theta - 1)^2 \left(\theta + \lambda + \theta \lambda \log \left(\frac{x}{x_0} \right) \right)}{x \left(\theta(\theta + 2\lambda - 1) - \lambda + (\theta - 1)\theta \lambda \log \left(\frac{x}{x_0} \right) \right)}. \quad (11)$$

Moreover, it can be easily verified that $\lim_{x \rightarrow \infty} r_e(x) = 0$. Hence by using Theorem 2.1 of Su and Tang (2003) we can conclude that $\mathcal{MPLG}(\theta, \lambda, x_0)$ is a heavy-tailed distribution. Heavy-tailed distributions are important in non-life insurance when modeling losses related to motor third-party liability insurance, fire insurance or catastrophe insurance.

2.5. Estimation

Let x_1, \dots, x_n be an independent and identically distributed (iid) random sample of size n drawn from a population which follows a $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution. Estimate of parameter x_0 can easily be obtained, as the support of the $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution is greater than $\hat{x}_0 = \min(x_1, \dots, x_n)$. Therefore for x_0 known, we estimate the parameters θ and λ by (i) method of log-moments and (ii) maximum likelihood method.

2.5.1. Method of Log-Moments

The first and second log-moments of $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution obtained by using (7) are given by $l_1 = \frac{(\theta + 2\lambda)}{\theta(\theta + \lambda)}$ and $l_2 = \frac{2(\theta + 3\lambda)}{\theta^2(\theta + \lambda)}$. Solving these equations for θ and λ , we have that these estimates are provided by

$$\tilde{\theta} = \frac{\sqrt{l_1^2 - l_2} + l_1}{l_2} \quad \text{and} \quad \tilde{\lambda} = \frac{3l_1 l_2 - 3l_1^3 - 3l_1^2 \sqrt{l_1^2 - l_2} + 2l_2 \sqrt{l_1^2 - l_2}}{l_2(3l_1^2 - 4l_2)}.$$

2.5.2. Maximum Likelihood Estimation

The log-likelihood function given the iid random sample $\underline{x} := x_1, \dots, x_n$ of size n is

$$\begin{aligned} \ell_n(\theta, \lambda | \underline{x}) &= 2n \log \theta - n \log(\theta + \lambda) + \sum_{i=1}^n \log \left(1 + \lambda \log \left(\frac{x_i}{x_0} \right) \right) \\ &\quad - \theta \sum_{i=1}^n \log \left(\frac{x_i}{x_0} \right) - \sum_{i=1}^n \log x_i, \end{aligned} \quad (12)$$

and the normal equations obtained by differentiating (12) with respect to θ and λ are

$$\frac{\partial \ell_n(\theta, \lambda | \underline{x})}{\partial \theta} = \frac{2n}{\theta} - \frac{n}{\theta + \lambda} - \sum_{i=1}^n \log \left(\frac{x_i}{x_0} \right), \quad \frac{\partial \ell_n(\theta, \lambda | \underline{x})}{\partial \lambda} = -\frac{n}{\theta + \lambda} - \sum_{i=1}^n \frac{\log \left(\frac{x_i}{x_0} \right)}{1 + \lambda \log \left(\frac{x_i}{x_0} \right)}.$$

The above equations cannot be solved analytically and hence maximum likelihood estimates of θ and λ can be computed numerically using in-built R-function such as `nlm()`, `maxlik()` or `optim()`. Moreover, in all these functions to initialize the program we use initial values obtained from the method of log-moments. The second partial derivative of log-likelihood function with respect to parameter θ and λ are

$$\begin{aligned} \frac{\partial^2 \ell_n(\theta, \lambda | \underline{x})}{\partial \theta^2} &= -\frac{2n}{\theta^2} + \frac{n}{(\theta + \lambda)^2}, \quad \frac{\partial^2 \ell_n(\theta, \lambda | \underline{x})}{\partial \theta \partial \lambda} = \frac{n}{(\theta + \lambda)^2}, \\ \frac{\partial^2 \ell_n(\theta, \lambda | \underline{x})}{\partial \lambda^2} &= \frac{n}{(\theta + \lambda)^2} - \sum_{i=1}^n \frac{\log^2\left(\frac{x_i}{x_0}\right)}{\left(1 + \lambda \log\left(\frac{x_i}{x_0}\right)\right)^2}. \end{aligned} \tag{13}$$

Furthermore,

$$\mathbb{E} \left(\frac{\log^2\left(\frac{X}{x_0}\right)}{\left(1 + \lambda \log\left(\frac{X}{x_0}\right)\right)^2} \right) = \frac{\theta^2}{(\theta + \lambda)} \int_{x_0}^{\infty} \frac{\log^2\left(\frac{x}{x_0}\right)}{\left(1 + \lambda \log\left(\frac{x}{x_0}\right)\right)^2} \left(\frac{x}{x_0}\right)^{-\theta} \frac{1}{x} \left(1 + \lambda \log\left(\frac{x}{x_0}\right)\right) dx,$$

substituting $t = \lambda \log\left(\frac{x}{x_0}\right)$, and using Tricomi confluent hypergeometric function $\mathcal{U}(a, b, z) = \frac{1}{\Gamma(a)} \int_0^{\infty} t^{a-1} (1+t)^{-a+b-1} e^{-zt} dt$, we get

$$\mathbb{E} \left(\frac{\log^2\left(\frac{X}{x_0}\right)}{\left(1 + \lambda \log\left(\frac{X}{x_0}\right)\right)^2} \right) = \frac{2\theta^2}{(\theta + \lambda)\lambda^3} \mathcal{U}(3, 3, \theta/\lambda).$$

Hence, the expected Fisher’s information matrix associated to the parameters θ and λ is given by $\mathcal{I}_{ij}(\theta, \lambda)$ with $i, j = 1, 2$ where

$$\begin{aligned} \mathcal{I}_{11}(\theta, \lambda) &= \mathbb{E} \left(-\frac{\partial^2 \ell_n(\theta, \lambda | \underline{x})}{\partial \theta^2} \right) = \frac{2n}{\theta^2} - \frac{n}{(\theta + \lambda)^2}, \\ \mathcal{I}_{12}(\theta, \lambda) &= \mathcal{I}_{21}(\theta, \lambda) = \mathbb{E} \left(-\frac{\partial^2 \ell_n(\theta, \lambda | \underline{x})}{\partial \theta \partial \lambda} \right) = -\frac{n}{(\theta + \lambda)^2}, \\ \mathcal{I}_{22}(\theta, \lambda) &= \mathbb{E} \left(-\frac{\partial^2 \ell_n(\theta, \lambda | \underline{x})}{\partial \lambda^2} \right) = -\frac{n}{(\theta + \lambda)^2} + \frac{2\theta^2 n}{(\theta + \lambda)\lambda^3} \mathcal{U}(3, 3, \theta/\lambda). \end{aligned}$$

The standard errors of the estimates $\hat{\theta}$ and $\hat{\lambda}$ can be obtained by inverting the aforementioned matrix and taking the square root of the diagonal entries.

2.6. Composite Models

Composite parametric models are a useful way to describe data that combined loss data of small and moderate size with a high frequency and large observations with a low frequency. They consist of two distributions, the first model is used up to an unknown single threshold value, estimated from the data, and the second distribution beyond this threshold (see [Cooray and Ananda 2005](#)). Another approach to derive composite models is utilizing a mode-matching procedure. In this technique, the two distributions are composed at the common modal value which can be estimated from the claims data. Thus, the composite model uses a truncated version of the first model up to the mode and the rest of the model is based on an appropriate truncation of the second distribution from that modal point onwards. The model after composition is similar in shape to either of the models considered but with a thicker tail. By using this methodology, it is guaranteed that the new density is continuous and smooth. These composite models give a significantly better fit as compared to the standard single models for the same empirical data. Here, we derived composite models of our \mathcal{MPLG} distribution with lognormal, Weibull and paralogistic distributions.

2.6.1. Composite Lognormal- \mathcal{MPLG} Model

Let us consider, for $x > 0$, the two-parameter lognormal distribution with pdf and cdf respectively given by

$$f_1(x) = \frac{1}{\sqrt{2\pi x\sigma}} \exp\left\{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right\} \text{ and } F_1(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right),$$

with $\mu \in \mathbb{R}$, $\sigma > 0$ where $\Phi(\cdot)$ is the cdf of the standard normal distribution. Now by taking expressions (1) and (2) as the pdf, $f_2(x)$, and cdf, $F_2(x)$ of the \mathcal{MPLG} distribution respectively with $x \geq x_0$ and setting equal the modes of lognormal and \mathcal{MPLG} distributions we obtain

$$\sigma = \sqrt{\mu - \log x_0 - \frac{\lambda - \theta - 1}{\lambda(\theta + 1)}}. \tag{14}$$

Note that we must impose the additional constraint $\mu > \log x_0 + \frac{\lambda - \theta - 1}{\lambda(\theta + 1)}$. Now, the unrestricted mixing weight is given by

$$\begin{aligned} r = & \frac{\theta^2}{x_m(\theta + \lambda)} \left(\frac{x_m}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x_m}{x_0}\right)\right) \Phi\left(\frac{\log x_m - \mu}{\sigma}\right) \\ & \times \left\{ \frac{\theta^2}{x_m(\theta + \lambda)} \left(\frac{x_m}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x_m}{x_0}\right)\right) \Phi\left(\frac{\log x_m - \mu}{\sigma}\right) \right. \\ & \left. + \frac{1}{\sqrt{2\pi x_m\sigma}} \exp\left\{-\frac{1}{2}\left(\frac{\log x_m - \mu}{\sigma}\right)^2\right\} \frac{(\theta + \lambda + \theta\lambda \log\left(\frac{x_m}{x_0}\right))}{\theta + \lambda} \left(\frac{x_m}{x_0}\right)^{-\theta} \right\}^{-1}. \end{aligned}$$

Then, the pdf of the four-parameter composite lognormal- \mathcal{MPLG} distribution is given by

$$f(x) = \begin{cases} r F_1(x_m)^{-1} \frac{1}{\sqrt{2\pi x\sigma}} \exp\left\{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right\} & 0 \leq x \leq x_m \\ (1 - r) (1 - F_2(x_m))^{-1} \frac{\theta^2}{x(\theta + \lambda)} \left(\frac{x}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x}{x_0}\right)\right) & x_m \leq x \leq \infty. \end{cases}$$

2.6.2. Composite Weibull- \mathcal{MPLG} Model

Let

$$f_1(x) = \frac{\tau}{x} \left(\frac{x}{\phi}\right)^\tau \exp\left\{-\left(\frac{x}{\phi}\right)^\tau\right\} \text{ and } F_1(x) = 1 - \exp\left\{-\left(\frac{x}{\phi}\right)^\tau\right\}$$

be the pdf and cdf of a two-parameter Weibull distribution with $\phi, \tau > 0$ and $x > 0$. Let us again consider the pdf and cdf of the \mathcal{MPLG} distribution given by (1) and (2) respectively. By equating the modes of the two distributions, we have that

$$\phi = x_0 \exp\left\{\frac{\lambda - \theta - 1}{\lambda(\theta - 1)}\right\} \left(\frac{\tau}{\tau - 1}\right)^{\frac{1}{\tau}}. \tag{15}$$

where the restriction $\tau > 1$ must be imposed to guarantee that the mode of the Weibull distribution is greater than 0. Now, the pdf of the four-parameter composite Weibull- \mathcal{MPLG} distribution is provided by

$$f(x) = \begin{cases} r F_1(x_m)^{-1} \frac{\tau}{x} \left(\frac{x}{\phi}\right)^\tau \exp\left\{-\left(\frac{x}{\phi}\right)^\tau\right\} & 0 \leq x \leq x_m \\ (1 - r) (1 - F_2(x_m))^{-1} \frac{\theta^2}{x(\theta + \lambda)} \left(\frac{x}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x}{x_0}\right)\right) & x_m \leq x \leq \infty, \end{cases}$$

where the unrestricted mixing weight is given by

$$r = \frac{\theta^2}{x_m(\theta + \lambda)} \left(\frac{x_m}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x_m}{x_0}\right)\right) \left(1 - \exp\left\{-\left(\frac{x_m}{\phi}\right)^\tau\right\}\right) \\ \times \left\{ \frac{\theta^2}{x_m(\theta + \lambda)} \left(\frac{x_m}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x_m}{x_0}\right)\right) \left(1 - \exp\left\{-\left(\frac{x_m}{\phi}\right)^\tau\right\}\right) \right. \\ \left. + \frac{\tau}{x_m} \left(\frac{x_m}{\phi}\right)^\tau \exp\left\{-\left(\frac{x_m}{\phi}\right)^\tau\right\} \left(\frac{\theta + \lambda + \theta \log\left(\frac{x_m}{x_0}\right)}{\theta + \lambda} \left(\frac{x_m}{x_0}\right)^{-\theta}\right) \right\}^{-1}.$$

2.6.3. Composite Paralogistic-MPLG Model

Finally, let us assume that

$$f_1(x) = \alpha^2 \frac{(x \tau)^\alpha}{x(1 + (x \tau)^\alpha)^{\alpha+1}} \text{ and } F_1(x) = 1 - \frac{1}{(1 + (x \tau)^\alpha)^\alpha}$$

are the pdf and cdf of the two-parameter paralogistic distribution with $\alpha, \tau > 0$, and $x > 0$. Once again, we consider the pdf and cdf of the MPLG model given by (1) and (2). By setting equal the modal values of the two distributions, we have that

$$\tau = \left(x_0 \exp\left(\frac{\lambda - \theta - 1}{\lambda(\theta - 1)}\right)\right)^{-1} \left(\frac{\alpha - 1}{\alpha^2 + 1}\right)^{\frac{1}{\alpha}}. \tag{16}$$

where the restriction $\alpha > 1$ is again established to ensure that the mode of the paralogistic distribution is larger than 0. The unrestricted mixing weight is now given by

$$r = \frac{\theta^2}{x_m(\theta + \lambda)} \left(\frac{x_m}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x_m}{x_0}\right)\right) \left(1 - \left(\frac{1}{1 + (x_m \tau)^\alpha}\right)^\alpha\right) \\ \times \left\{ \frac{\theta^2}{x_m(\theta + \lambda)} \left(\frac{x_m}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x_m}{x_0}\right)\right) \left(1 - \left(\frac{1}{1 + (x_m \tau)^\alpha}\right)^\alpha\right) \right. \\ \left. + \alpha^2 \frac{(x_m \tau)^\alpha}{x_m(1 + (x_m \tau)^\alpha)^{\alpha+1}} \left(\frac{\theta + \lambda + \theta \log\left(\frac{x_m}{x_0}\right)}{\theta + \lambda} \left(\frac{x_m}{x_0}\right)^{-\theta}\right) \right\}^{-1}.$$

The pdf of the four-parameter composite paralogistic-MPLG distribution is provided by

$$f(x) = \begin{cases} r F_1(x_m)^{-1} \alpha^2 \frac{(x \tau)^\alpha}{x(1 + (x \tau)^\alpha)^{\alpha+1}} & 0 \leq x \leq x_m \\ (1 - r) (1 - F_2(x_m))^{-1} \frac{\theta^2}{x(\theta + \lambda)} \left(\frac{x}{x_0}\right)^{-\theta} \left(1 + \lambda \log\left(\frac{x}{x_0}\right)\right) & x_m \leq x \leq \infty. \end{cases}$$

3. Insurance Results

In the following, several theoretical results related to insurance for the $MPLG(\theta, \lambda, x_0)$ distribution and related composite models are derived.

3.1. Mean Excess Function

The mean excess function measures the expected payment per claim on a policy with a fixed amount deductible of x , ignoring the claims with amounts less than or equal to x . It is defined as

$$e(x) = \mathbb{E}(X - x | X > x) = \frac{\int_x^\infty \bar{F}(u) du}{\bar{F}(x)},$$

which can also be obtained by inverting the equilibrium hazard rate (Su and Tang 2003), hence for $\theta > 1$ the mean excess function for the model with pdf (1) is given by

$$e(x) = \frac{x \left(\theta(\theta + 2\lambda - 1) - \lambda + (\theta - 1)\theta\lambda \log\left(\frac{x}{x_0}\right) \right)}{(\theta - 1)^2 \left(\theta + \lambda + \theta\lambda \log\left(\frac{x}{x_0}\right) \right)}. \tag{17}$$

3.2. Excess of Loss Reinsurance

Let X be a rv denoting the individual claim size taking values greater than d . Let us also assume that X follows (1), and the expected cost per claim to the reinsurance layer when the loss in excess of m subject to a maximum of l is given by

$$\mathbb{E}(\min(l, \max(0, X - m))) = \int_m^{m+l} (x - m) f(x) dx + l \bar{F}(m + l).$$

Now by replacing $f(\cdot)$ and $\bar{F}(\cdot)$ by expressions (1) and (3) and solving the integral, we have that

$$\begin{aligned} & \mathbb{E}(\min(l, \max(0, X - m))) \\ &= \frac{\theta^2 x_0^\theta}{\theta + \lambda} \left((l + m)^{-\theta} \left(-\frac{\lambda(\theta l + m) \log\left(\frac{l+m}{x_0}\right)}{(\theta - 1)\theta} - \frac{(\theta + \lambda - 1)(l + m)}{(\theta - 1)^2} + \frac{m(\theta + \lambda)}{\theta^2} \right. \right. \\ &+ \left. \left. \frac{m^{1-\theta} \left(\theta^2 + \theta(2\lambda - 1) - \lambda + (\theta - 1)\theta\lambda \log\left(\frac{m}{x_0}\right) \right)}{(\theta - 1)^2 \theta^2} \right) \right) \\ &+ \frac{l x_0^\theta (l + m)^{-\theta} \left(\theta + \lambda + \theta\lambda \log\left(\frac{l+m}{x_0}\right) \right)}{\theta + \lambda}. \end{aligned}$$

3.3. Value-at-Risk and Tail Value-at-Risk

In this subsection, we firstly discuss the most widely used risk measure, the Value-at-Risk (VaR). It is defined as the minimum value of the distribution such that the probability of the loss larger than this value does not exceed a given probability. In statistical terms, VaR is a quantile of a random variable and the formal definition is as follows.

Definition 2. Let X be a loss rv with a continuous cdf $F_X(\cdot)$, and δ be a probability level such that $0 < \delta < 1$, the Value-at-Risk at probability level δ , denoted by $VaR_\delta(X)$, is the δ -quantile of X . That is

$$VaR_\delta(X) = \inf\{x \in \mathbb{R} : F_X(x) \geq \delta\}.$$

Hence for rv having $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution the Value-at-Risk is given in next Proposition.

Proposition 2. For $\theta, \lambda > 0$, the $VaR_\delta(X)$ of the $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution is

$$VaR_\delta(X) = x_0 \exp \left\{ -\frac{\theta + \lambda}{\theta\lambda} - \frac{1}{\theta} W_{-1} \left(-\frac{(1 - \delta)(\theta + \lambda) \exp \left\{ -\frac{\theta + \lambda}{\lambda} \right\}}{\lambda} \right) \right\} \tag{18}$$

where W_{-1} is the negative branch of Lambert-W function.

Proof. See the Appendix A. \square

The median of $\mathcal{MPLG}(\theta, \lambda, x_0)$ will be obtained by taking $\delta = 0.5$ in $VaR_\delta(X)$. As in general, the loss distributions are typically skewed, the VaR is a non-coherent risk measure due to the lack of subadditivity (see Klugman et al. 2012), for that reason, the Tail Value-at-Risk (TVaR) of X (see Acerbi and Tasche 2002) is usually considered as a more informative and more useful risk measure. The TVaR is given by

$$TVaR_\delta(X) = \frac{1}{1 - \delta} \int_\delta^1 VaR_z(X) dz,$$

which is a coherent risk measure. If X is continuous, $\Pr(X \leq VaR_\delta(X)) = \delta$, and then the TVaR is the conditional tail expectation $TVaR_\delta(X) = \mathbb{E}(X|X > VaR_\delta(X))$.

Proposition 3. For $\theta, \lambda > 0$, the $TVaR_\delta(X)$ for $\mathcal{MPLG}(\theta, \lambda, x_0)$ distribution is given by

$$TVaR_\delta(X) = \frac{1}{(1 - \delta)} \frac{\lambda x_0 \theta \exp \left\{ -\frac{(1 - \theta)(\theta + \lambda)}{\theta\lambda} \right\}}{(1 - \theta)^2(\theta + \lambda)} \times \left((1 + (1 - \theta)\mathcal{K}(\delta; \theta, \lambda)) \exp \left\{ -\frac{(1 - \theta)\mathcal{K}(\delta; \theta, \lambda)}{\theta} \right\} - 1 \right) \tag{19}$$

where $\mathcal{K}(\delta; \theta, \lambda) = W_{-1} \left(-\frac{(1 - \delta) \exp \left\{ -\frac{\theta + \lambda}{\lambda} \right\} (\theta + \lambda)}{\lambda} \right)$.

Proof. See the Appendix A. \square

3.4. Limited Loss Variable for Composite Models

Finally, in this subsection the k th moment of the limited loss variable for composite models based on the \mathcal{MPLG} distribution is presented.

Definition 3. The k th incomplete moment transform of a rv X with cdf $F(x)$ and pdf $f(x)$, given that $\mu_k = \mathbb{E}(X^k) < \infty$, is a rv with pdf given by $f^{(k)}(x) = \frac{x^k f(x)}{\mu_k}$ and cdf given by $F^{(k)}(x) = \int_0^x \frac{t^k f(t) dt}{\mu_k}$, $0 < x < \infty$.

Definition 4. The limited loss variable is defined as, $X \wedge u = \begin{cases} X, & X < u, \\ u, & X \geq u. \end{cases}$ where u is the maximum benefit paid by the insurance policy.

Proposition 4. The moment of order k of the limited loss variable for the composite models based on the \mathcal{MPLG} distribution is given by

$$\mathbb{E}((X \wedge u)^k) = r\mathbb{E}(X_1^k) \frac{F_1^{(k)}(u)}{F_1(x_m)} + u^k \left(1 - r \frac{F_1(u)}{F_1(x_m)} \right), \tag{20}$$

if $0 < u \leq x_m$ and

$$\begin{aligned} \mathbb{E}((X \wedge u)^k) &= r\mathbb{E}(X_1^k) \frac{F_1^{(k)}(x_m)}{F_1(x_m)} + (1-r)\mathbb{E}(X_2^k) \frac{F_2^{(k)}(u) - F_2^{(k)}(x_m)}{1 - F_2(x_m)} \\ &+ u^k(1-r) \frac{1 - F_2(u)}{1 - F_2(x_m)}, \end{aligned} \tag{21}$$

if $x_m < u < \infty$.

Proof. The k th moment of the limited loss variable can be derived as,

$$E[(X \wedge u)^k] = \int_0^u x^k f_X(x) dx + u^k [1 - F_X(u)].$$

If $0 < u < x_m$,

$$E[(X \wedge u)^k] = \int_0^u r x^k \frac{f_1(x)}{F_1(x_m)} dx + u^k \left(1 - r \frac{F_1(u)}{F_1(x_m)} \right),$$

and we get (20). Now, if $x_m < u < \infty$,

$$\begin{aligned} E[(X \wedge u)^k] &= \int_0^{x_m} r x^k \frac{f_1(x)}{F_1(x_m)} dx + \int_{x_m}^u (1-r) \frac{x^k f_2(x)}{1 - F_2(x_m)} dx \\ &+ u^k \left[1 - (r + (1-r) \frac{F_2(u) - F_2(x_m)}{1 - F_2(x_m)}) \right], \end{aligned}$$

and (21) is obtained. \square

Observe that in expressions (20) and (21), $\mathbb{E}(X_i^k)$ and $F_i^{(k)}(\cdot)$ denotes the k th incomplete moment and the cdf of the k th incomplete moment for lognormal, Weibull or paralogistic distributions ($i = 1$) and $MP\mathcal{L}\mathcal{G}$ distribution ($i = 2$).

4. Case Studies

In this section, we use two claims datasets to assess the performance of $MP\mathcal{L}\mathcal{G}$ distribution and the $MP\mathcal{L}\mathcal{G}$ composite models. Finally, some results related to income indices are given.

4.1. An Application to Automobile Claims Data

First, we examine the performance of $MP\mathcal{L}\mathcal{G}$ distribution as compared to other heavy-tailed distributions available in the literature by employing of a real automobile claims dataset (see De Jong and Heller 2008). This dataset describes a one-year vehicle insurance policies taken out in 2004 or 2005. There are 67,856 policies, of which 4624 (6.8%) had at least one claim. The variable of interest is the size of the claims. The minimum claim amount is \$200 and the maximum value is \$55,922.13. We have fitted to this dataset the $MP\mathcal{L}\mathcal{G}$ distribution by setting the value of x_0 equal to the minimum value to explain the claims amount distribution. In Table 1, parameter estimates and standard errors (S.E.) for the $MP\mathcal{L}\mathcal{G}$ and other two-parameter and three-parameter heavy-tailed probabilistic models with support in \mathbb{R}^+ are illustrated. For a review of these distributions the reader is referred to Hogg and Klugman (2009) or Klugman et al. (2012). For the classical Pareto, PAT the location parameter has been set equal to the minimum value, i.e., $x_0 = 200$. For the loggamma distribution the location parameter can be chosen in the neighborhood of this value. For comparison purposes, three measures of model selection has been included in this Table, the negative of the log-likelihood function (NLL), Akaike’s information criterion (AIC) and Bayesian information criterion (BIC). It is observable that the $MP\mathcal{L}\mathcal{G}$ provides the best fit to data in terms of these measures of model selection. The figures of the measures of model selection reveal that the Pareto distribution is preferable to the loggamma distribution for this dataset.

We have also fitted the three-parameter Burr distribution to this dataset. For this model the value of the NLL is 38,028.13. However, the algorithm used to search for the maximum of the log-likelihood surface to find the estimates stopped at the boundary of the parameter space, and thus these estimates are no longer displayed in Table 1.

Table 1. Estimated values of different heavy-tailed models, corresponding standard errors (S.E.), negative of the log-likelihood function (NLL), Akaike’s information criterion (AIC) and Bayesian information criterion (BIC) for automobile claims dataset (De Jong and Heller 2008).

Distribution	Parameter Estimates (S.E.)	NLL	AIC	BIC
Pareto	$\hat{\theta} = 0.661 (0.010)$	38,024.80	76,051.61	76,058.05
lognormal	$\hat{\mu} = 6.810 (0.017), \hat{\sigma} = 1.189 (0.012)$	38,852.15	77,708.31	77,721.19
loggamma	$\hat{\alpha} = 1.115 (0.021), \hat{\beta} = 5.109 (0.118)$	38,998.18	78,000.36	78,013.23
Fréchet	$\hat{\alpha} = 0.659 (0.019), \hat{s} = 254.0 (11.87)$ $\hat{b} = 157.1 (4.279)$	38,408.30	76,822.60	76,841.92
Weibull	$\hat{\mu} = 0.786 (0.008), \hat{\sigma} = 1690.8 (33.64)$	39,491.60	78,987.19	79,000.07
Lomax	$\hat{\alpha} = 2.047 (0.088), \hat{\lambda} = 2205.1 (133.1)$	39,169.85	78,343.70	78,356.58
PAT	$\hat{\alpha} = 0.895 (0.095), \hat{\theta} = 0.740 (0.017)$	38,006.30	76,016.61	76,029.49
Inverse Weibull	$\hat{\alpha} = 1.053 (0.012), \hat{\sigma} = 518.8 (7.636)$	38,595.61	77,195.22	77,208.09
<i>MPLG</i>	$\hat{\theta} = 0.943 (0.018), \hat{\lambda} = 0.698 (0.073)$	37,965.99	75,935.98	75,948.86

4.2. An Application to Fire Insurance Claims

In our second application, the composite models derived from the *MPLG* distribution are compared to the composite Pareto and composite Lomax families. For that reason, we consider the well-known Danish fire insurance dataset that consists of 2492 fire insurance losses in millions of Danish kroner (DKr) from the years 1980 to 1990 (both inclusive), adjusted to reflect 1985 values. This dataset may be found in the ‘SMPrcacticals’ add-on package for R, available from the CRAN website <http://cran.r-project.org/>. Parameter estimation for all the models considered has been completed by the method of maximum likelihood (which is implemented using the function ‘mle’/‘mle2’ in R). In Table 2, parameter estimates and standard errors (S.E.) for all the composite distributions and the three measures of model validation considered in our first example are exhibited. Among the composite models, we can see that the Weibull-*MPLG* composite model gives the best fit overall for this dataset in terms of NLL, AIC and BIC values.

Finally, to select composite models that provide an acceptable description of the loss process, we must verify that the first-order moment of the limited loss variable and the empirical counterpart, given by $\mathbb{E}_n(u) = \frac{1}{n} \sum_{i=1}^n \min(x_i, u)$ are essentially in agreement. Obviously when u tends to infinity, the former quantity and $\mathbb{E}_n(u)$ converge to $\mathbb{E}(X)$ and the sample mean, respectively. In Table 3 empirical and fitted limited expected value for the seven composite models considered are displayed. As it can be observed, the composite lognormal-Pareto and Weibull-Pareto distributions tend to overestimate the empirical limited expected value when the policy limit u increases. Although, the remainder models stay closer to the empirical limited expected value for different values of u , the composite Weibull-Lomax displays the best behaviour.

Table 2. Estimated values of different heavy-tailed models, corresponding standard errors (S.E.), NLL, AIC, and BIC of different composite models for the Danish fire losses dataset.

Distribution	Parameter Estimates (S.E.)		NLL	AIC	BIC
lognormal-Lomax	$\hat{\mu} = 0.104$ (0.020) $\hat{\lambda} = 0.365$ (0.123)	$\hat{\sigma} = 0.182$ (0.011) $\hat{\theta} = 1.144$ (0.029)	3860.47	7728.94	7752.223
lognormal-Pareto	$\hat{\mu} = 0.137$ (0.019) $\hat{\theta} = 1.208$ (0.030)	$\hat{\sigma} = 0.197$ (0.012)	3865.86	7737.72	7755.182
lognormal-MPLG	$\hat{\mu} = 0.045$ (0.017) $\hat{x}_0 = 0.745$ (0.070)	$\hat{\theta} = 2.060$ (0.039) $\hat{\lambda} = 65.804$ (377.425)	3872.40	7752.81	7776.09
Weibull-Lomax	$\hat{\tau} = 15.345$ (0.671) $\hat{\lambda} = 0.561$ (0.127)	$\hat{\phi} = 0.969$ (0.007) $\hat{\theta} = 0.971$ (0.007)	3823.70	7655.40	7678.68
Weibull-Pareto	$\hat{\tau} = 14.048$ (0.502) $\hat{\theta} = 1.003$ (0.008)	$\hat{\psi} = 0.997$ (0.008)	3840.38	7686.76	7704.22
Weibull-MPLG	$\hat{\tau} = 18.763$ (0.986) $\hat{\theta} = 1.938$ (0.034)	$\hat{x}_0 = 0.787$ (0.006) $\hat{\lambda} = 4.614$ (0.159)	3823.30	7654.60	7677.88
paralogistic-MPLG	$\hat{\alpha} = 16.719$ (1.311) $\hat{\theta} = 1.989$ (0.035)	$\hat{x}_0 = 0.901$ (0.020) $\hat{\lambda} = 3.379$ (0.266)	3824.482	7656.96	7680.25

Table 3. Limited expected value for the distributions considered and different values of the policy limit u .

Policy Limit u	Empirical	Composite Models						
		lognormal Pareto	lognormal Lomax	Weibull Pareto	Weibull Lomax	lognormal MPLG	Weibull MPLG	Paralogistic MPLG
1	0.989	0.998	0.985	0.989	0.989	0.987	0.986	0.989
2	1.565	1.726	1.525	1.560	1.548	1.333	1.266	1.467
5	2.138	2.489	2.122	2.155	2.143	2.015	1.943	2.093
10	2.447	2.932	2.362	2.482	2.464	2.341	2.292	2.398
20	2.707	3.285	2.566	2.781	2.674	2.542	2.522	2.591
100	2.958	3.851	2.826	3.299	2.923	2.740	2.778	2.793
200	3.037	4.017	2.884	3.463	2.972	2.770	2.823	2.826
∞	3.063	4.665	3.005	4.289	3.060	2.804	2.885	2.868

4.3. Theil's Income Indices

Among several economic inequality measures the Gini Index is commonly used. However, it does not allow us to vary the sensitivity of the index to redistributive movements at specific income ranges, which can point out different evolutions of disparities when no Lorenz dominance is found (see Sarabia and Castillo 2005). Yitzhaki (1983) further generalized this index which is sensitive to changes in the right tail of the distribution as the parameter increases. The two limiting cases of this family of inequality measures, known as Theil's Indices (T_1 and T_0), corresponding to the Theil entropy index (TEI) and the mean log deviation (MLD), are defined by

$$T_1(X) = -\mathbb{E} \left(\log \frac{X}{\mu} \right) \quad \text{and} \quad T_2(X) = \mathbb{E} \left(\frac{X}{\mu} \log \frac{X}{\mu} \right).$$

Proposition 5. The TEI and MLD indices for a rv X that follows the $MPLG(\theta, \lambda, x_0)$ distribution are provided by

$$T_1(X) = -\frac{\theta + 2\lambda}{\theta(\theta + \lambda)} + \log \left(\frac{\theta^2(\theta + \lambda - 1)}{(\theta - 1)^2(\theta + \lambda)} \right), \quad \theta > 1$$

and

$$T_2(X) = \frac{2}{\theta - 1} - \frac{1}{\theta + \lambda - 1} - \log \left(\frac{\theta^2(\theta + \lambda - 1)}{(\theta - 1)^2(\theta + \lambda)} \right), \quad \theta > 1.$$

Proof. The proof follows directly after some computation. \square

5. Conclusions

In this paper, we have proposed a new heavy-tailed class of distribution, that is obtained by using an exponential transformation of the generalized Lindley distribution. This class generalizes both the classical Pareto model and a special case of the loggamma distribution. Some of its most relevant statistical properties were examined. The model is very flexible since it allows for closed-form expressions for many results related to insurance. Besides, as the mode of this new model can be written in closed-form, composite models based on this distribution can be easily derived. The numerical illustrations reveal that the mixture Pareto-loggamma distribution provides a good fit to loss data, and it is a competitive model with other existing heavy-tailed distributions in the literature.

Author Contributions: The authors contributed equally to this work.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Expression (1) is a genuine pdf:

$$\begin{aligned} f_X(x|\theta, \lambda, x_0) &= \frac{\theta}{\theta + \lambda} \frac{\theta}{x} \left(\frac{x}{x_0}\right)^{-\theta} + \frac{\lambda}{\theta + \lambda} \frac{\theta^2}{x} \left(\frac{x}{x_0}\right)^{-\theta} \log\left(\frac{x}{x_0}\right) \\ &= \frac{\theta}{\theta + \lambda} \theta \frac{x_0^\theta}{x^{\theta+1}} + \frac{\lambda}{\theta + \lambda} \frac{\theta^2}{x_0} \left(\frac{x}{x_0}\right)^{-(\theta+1)} \log\left(\frac{x}{x_0}\right) \\ &= \frac{\theta}{\theta + \lambda} g_1(x) + \frac{\lambda}{\theta + \lambda} g_2(x), \end{aligned}$$

where $g_1(x)$ and $g_2(x)$ are the densities of the Pareto and loggamma distribution respectively.

Proof of Proposition 1. Let X_i be a \mathcal{MPLG} rv's with parameter $(\theta_i, \lambda_i, x_{0i}), i = 1, 2$. Then

$$\frac{d}{dx} \log \frac{f_{X_1}(x)}{f_{X_2}(x)} = \frac{\lambda_1}{x \left(1 + \lambda_1 \log\left(\frac{x}{x_{01}}\right)\right)} - \frac{\lambda_2}{x \left(1 + \lambda_2 \log\left(\frac{x}{x_{02}}\right)\right)} - \frac{\theta_1}{x} + \frac{\theta_2}{x} \tag{A1}$$

it is easy to see that, if $x_{01} = x_{02}$, (A1) is negative when $\theta_1 \geq \theta_2$ and $\lambda_1 \leq \lambda_2$ which implies that $X_1 \leq_{lr} X_2$. Further note that, when $\theta_1 = \theta_2$ and $\lambda_1 = \lambda_2$, (A1) is negative when $x_{01} \leq x_{02}$ and therefore $X_1 \leq_{lr} X_2$. The other results follows from (9). \square

Proof of Proposition 2. By assuming $p = \log\left(\frac{x}{x_0}\right)$, the cdf can be written as

$$F_X(x) = 1 - \frac{(\theta + \lambda + \theta\lambda p)}{\theta + \lambda} \exp\{-\theta p\},$$

for fixed $\theta, \lambda > 0$ and $\delta \in (0, 1)$, the δ th quantile function is obtained by solving $F_X(x) = \delta$. By re-arranging the above, we obtain

$$(1 - \delta)(\theta + \lambda) = (\theta + \lambda + \theta\lambda p) \exp\{-\theta p\} \tag{A2}$$

Now, by multiplying both sides of (A2) by $-\exp\{-(\theta + \lambda)/\lambda\}$, we obtain

$$-\left(\frac{\theta + \lambda}{\lambda} + \theta p\right) \exp\left\{-\left(\frac{\theta + \lambda}{\lambda} + \theta p\right)\right\} = -\frac{(1 - \delta)(\theta + \lambda)}{\lambda} \exp\left\{-\frac{\theta + \lambda}{\lambda}\right\}. \tag{A3}$$

From expression (A3), we see that $-\left(\frac{\theta+\lambda}{\lambda} + \theta p\right)$ is the Lambert-W function of real argument $-\frac{(1-\delta)(\theta+\lambda)}{\lambda} \exp\left\{-\frac{\theta+\lambda}{\lambda}\right\}$. Thus, we have

$$W\left(-\frac{(1-\delta)(\theta+\lambda)}{\lambda} \exp\left\{-\frac{\theta+\lambda}{\lambda}\right\}\right) = -\left(\frac{\theta+\lambda}{\lambda} + \theta p\right). \tag{A4}$$

Moreover, for any $\theta, \lambda > 0$, it is immediate that $\frac{\theta+\lambda}{\lambda} + \theta p > 1$, and it can also be checked that $-\frac{(1-\delta)(\theta+\lambda)}{\lambda} \exp\left\{-\frac{\theta+\lambda}{\lambda}\right\} \in (-1/e, 0)$ since $\delta \in (0, 1)$. Therefore, by taking into account the properties of the negative branch of the Lambert-W function, we deduce the following

$$W_{-1}\left(-\frac{(1-\delta)(\theta+\lambda)}{\lambda} \exp\left\{-\frac{\theta+\lambda}{\lambda}\right\}\right) = -\left(\frac{\theta+\lambda}{\lambda} + \theta p\right).$$

Again, substituting $p = \log\left(\frac{x}{x_0}\right)$, and solving for x , we obtain

$$\text{VaR}_\delta(X) = x_0 \exp\left\{-\frac{\theta+\lambda}{\theta\lambda} - \frac{1}{\theta} W_{-1}\left(-\frac{(1-\delta)(\theta+\lambda) \cdot \exp\left\{-\frac{\theta+\lambda}{\lambda}\right\}}{\lambda}\right)\right\}.$$

This completes the proof of Proposition 2. \square

Proof of Proposition 3. Let X follows $\mathcal{MPLG}(\theta, \lambda, x_0)$. Then, the $\text{TVaR}_\delta(X)$ is defined as

$$\text{TVaR}_\delta(X) = \frac{1}{1-\delta} \int_\delta^1 \text{VaR}_z(X) dz,$$

substituting the value $\text{VaR}_z(X)$ from (18), we get

$$\begin{aligned} \text{TVaR}_\delta(X) &= \frac{1}{1-\delta} \int_\delta^1 x_0 \exp\left\{-\frac{\theta+\lambda}{\theta\lambda} - \frac{1}{\theta} W_{-1}\left(-\frac{(1-z) \exp\left\{-\frac{\theta+\lambda}{\lambda}\right\} (\theta+\lambda)}{\lambda}\right)\right\} dz \\ &= \frac{x_0}{1-\delta} \exp\left\{-\frac{(\theta+\lambda)(1-\theta)}{\theta\lambda}\right\} \int_\delta^1 \exp\left\{-\frac{1}{\theta} W_{-1}\left(-\frac{(1-z) \exp\left\{-\frac{\theta+\lambda}{\lambda}\right\} (\theta+\lambda)}{\lambda}\right)\right\} dz. \end{aligned}$$

Now, by denoting $-\frac{(1-z)e^{-\frac{\theta}{\lambda}-1}(\theta+\lambda)}{\lambda} = u$, we get

$$\text{TVaR}_\delta(X) = -\frac{x_0\lambda}{1-\delta} \exp\left\{-\frac{(\theta+\lambda)(1-\theta)}{\theta\lambda}\right\} \int_0^{\delta'} \exp\left\{-\frac{1}{\theta} W_{-1}(u)\right\} du,$$

where $\delta' = -\frac{(1-\delta)e^{-\frac{\theta}{\lambda}-1}(\theta+\lambda)}{\lambda}$. By letting $W_{-1}(u) = t$, it gives $t \exp\{t\} = u$, and $du = \exp\{t\} (t+1) dt$, then we have

$$\begin{aligned} \text{TVaR}_\delta(X) &= -\frac{x_0\lambda}{1-\delta} \exp\left\{-\frac{(\theta+\lambda)(1-\theta)}{\theta\lambda}\right\} \int_0^{W_{-1}(\delta')} \exp\left\{-\left(\frac{1}{\theta}-1\right)\right\} (1+t) dt \\ &= \frac{1}{(1-\delta)} \frac{\lambda x_0 \theta \exp\left\{-\frac{(1-\theta)(\theta+\lambda)}{\theta\lambda}\right\}}{(1-\theta)^2(\theta+\lambda)} \\ &\quad \times \left((1+(1-\theta)\mathcal{K}(\delta; \theta, \lambda)) \exp\left\{-\frac{(1-\theta)\mathcal{K}(\delta; \theta, \lambda)}{\theta}\right\} - 1 \right), \end{aligned}$$

where $\mathcal{K}(\delta; \theta, \lambda) = W_{-1}\left(-\frac{(1-\delta) \exp\left\{-\frac{\theta+\lambda}{\lambda}\right\} (\theta+\lambda)}{\lambda}\right)$. Hence the proposition. \square

References

- Acerbi, Carlo, and Dirk Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking & Finance* 26: 1487–503.
- Bakar, Shaiful A., Nor A. Hamzah, Mastoureh Maghsoudi, and Saralees Nadarajah. 2015. Modeling loss data using composite models. *Insurance: Mathematics and Economics* 61: 146–54.
- Calderín-Ojeda, Enrique, and Chun F. Kwok. 2016. Modeling Claims Data with Composite Stoppa Models. *Scandinavian Actuarial Journal* 9: 817–36. [CrossRef]
- Cooray, Kahadawala, and Malwane M. A. Ananda. 2005. Modeling actuarial data with a composite lognormal–Pareto model. *Scandinavian Actuarial Journal* 5: 321–34. [CrossRef]
- De Jong, Piet, and Heller Gillian H. 2008. *Generalized linear models for insurance data*. International series on actuarial science. Cambridge: Cambridge University Press.
- Ghitany, Mohamed E., Barbra Atieh, and Saralees Nadarajah. 2008. Lindley distribution and its application. *Mathematics and Computers in Simulation* 78: 493–506. [CrossRef]
- Ghitany, Mohamed E., Emilio Gómez-Déniz, and Saralees Nadarajah. 2018. A new generalization of the Pareto distribution and its application to insurance data. *Journal of Risk and Financial Management* 11: 10. [CrossRef]
- Gómez-Déniz, Emilio, and Enrique Calderín-Ojeda. 2014. A Suitable Alternative to the Pareto Distribution. *Haceteppe Journal of Mathematics and Statistics* 43: 843–60.
- Gómez-Déniz, Emilio, and Enrique Calderín-Ojeda. 2015. Modeling insurance data with the Pareto ArcTan distribution. *Astin Bulletin* 45: 639–60. [CrossRef]
- Hogg, Robert V., and Stuart A. Klugman. 2009. *Loss Distributions*. Vol. 249 of Wiley Series in Probability and Statistics. New York: Wiley.
- Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2012. *Loss Models: From Data to Decisions*, 4th ed. Wiley Series in Probability and Statistics. Hoboken: John Wiley and Sons, Inc.
- Sarabia, José M., and Enrique Castillo. 2005. About a class of max-stable families with applications to income distributions. *Metron* LXIII: 505–27.
- Sarabia, José M., and Faustino Prieto. 2009. The Pareto-positive stable distribution: A new descriptive model for city size data. *Physica A* 388: 4179–91. [CrossRef]
- Scollnik, David P. M. 2007. On Composite Lognormal-Pareto Models. *Scandinavian Actuarial Journal* 1: 20–33. [CrossRef]
- Shaked, Moshe, and J. George Shanthikumar. 2007. *Stochastic Orders*. New York: Springer Science and Business Media.
- Stoppa, Gabriele. 1990. Proprietà campionarie di un nuovo modello Pareto generalizzato. In *Atti XXXV Riunione Scientifica della Società Italiana di Statistica*. Padova: Cedom, pp. 137–44.
- Su, Chun, and Qihe H. Tang. 2003. Characterizations on heavy-tailed distributions by means of hazard rate. *Acta Mathematicae Applicatae Sinica* 19: 135–42. [CrossRef]
- Yang, Hailiang. 2004. Cramér-Lundberg asymptotics. In *Encyclopedia of Actuarial Science*. New York: Wiley.
- Yitzhaki, Shlomo. 1983. On an Extension of the Gini Inequality Index. *International Economic Review* 24: 617–28.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).