Hacettepe Journal of Mathematics and Statistics
Volume 43 (5) (2014), 843 – 860

A Suitable Alternative to the Pareto Distribution

Emilio Gómez–Déniz* and Enrique Calderín–Ojeda†

Received 26 : 07 : 2013 : Accepted 21 : 12 : 2013

Abstract

Undoubtedly, the single parameter Pareto distribution is one of the most attractive distribution 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 behaviour of large losses. In this paper, we introduce a distribution which can be considered as alternative to the single parameter Pareto distribution. A comprehensive treatment of its mathematical properties is considered with relevant emphasis on results concerning insurance. Additionally, estimation by the method of moments and maximum likelihood is discussed. Then, an analysis of estimation performance is carried out. Finally, the performance of the model is examined by using two examples of real claims data.

2000 AMS Classification: 62E10, 62F10, 62F15, 62P05.

Keywords: Credibility; excess-of-loss reinsurance; Pareto Distribution; Gamma Distribution; Ruin Function; Threshold.

1. Introduction

The single parameter Pareto distributions ([2] and [1], among others) has been proved to be useful as predicting tools in different socioeconomic contexts. In this regard, data related with income and city size (see [12] and references therein) has been modeled with Pareto-positive stable distribution. In actuarial settings, the single parameter Pareto distribution has been largely considered over other probability distributions, apart from its nice properties, for its appropriateness to model the claim size. When modeling losses, there is widely concern on the frequencies and sizes of large claims, in particular, the study of the right tail of the distribution. On this subject, the single parameter Pareto distribution gives a good description of the random behaviour of large losses.

*Department of Quantitative Methods in Economics. University of Las Palmas de Gran Canaria, Spain,
Email: `emilio.gomez-deniz@ulpgc.es` Corresponding author.

†Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Australia.

Perhaps, one of the most important characteristics of the Pareto distribution is that it produces a better extrapolation from the observed data when pricing high excess layers, in situations where there is little or no experience. In this regard, its efficacy dealing with inflation in claims and with the effect of deductibles and excess-of-loss levels for reinsurance has been demonstrated. It has also been used to handle risky insurance (see [16] and [14], among other authors for applications of the Pareto model in rating property excess-of-loss reinsurance).

In the last decades, a lot of attempts have been made to define new families of probability distributions (discrete or continuous) as alternative to other well–known models which could provide more flexibility when modelling data in practice. For the Pareto case, the Stoppa's generalized Pareto distribution (see [22]) and the Pareto positive stable distribution (see [20]) are typical examples of generalizations of the classical Pareto distribution. Nevertheless they has the disadvantage of adding a parameter to the parent distribution (the Pareto distribution), therefore, complicating the parameter estimation. In this paper, a distribution that can be considered as alternative to the single parameter Pareto distribution is introduced. Besides, a comprehensive treatment of its mathematical properties is provided. In this sense, expressions for the moments, variance, cumulative distribution function, asymptotic ruin function among other properties are derived. Estimation is carried out by maximum likelihood and moments methods. Closed-form expressions has been achieved for both estimators. Moreover, an analysis of the estimator performance is completed, being possible, if desired, to obtain unbiased modifications of these estimators. Finally, the performance of the model is examined by using two examples of real claims data. As it can be seen later, the new distribution introduced in this manuscript produces better fit to data than the Pareto distribution in the examples considered.

The distribution proposed here can be used as a basis for excess of loss quotations. Furthermore, it can be a good description of the random behaviour of large losses, in a similar way as the Pareto distribution does. See for example [18]. In addition to this, another interesting application of this distribution in the setting of credibility theory is that it is suitable for premiums calculation. In this sense, new credibility expressions are obtained when the premium charged to a policyholder is computed on the basis of his/her own past claims and the accumulated past claims of the corresponding portfolio of policyholders under weighted squared-error loss function.

The remainder of the paper is organized as follows. In Section 2, the new proposed distribution is shown, including some of its more relevant properties. In the last part of this section, results connected with insurance are provided. Next, Section 3 deals with parameter estimation. Two estimation methods are considered in this paper: moments method and maximum likelihood estimation. Later in this section, an analysis of estimation performance is carried out for the distribution proposed in this manuscript. In Section 4 numerical applications by using real insurance data are considered. Then, the obtained results are compared with those obtained for the classical Pareto distribution. Finally, some conclusions are given in the last section.

2. The proposed distribution

It is not difficult to observe that the expression

$$f(x|\alpha,\theta) = \frac{\theta^2 \log(x+\bar{\alpha})}{(x+\bar{\alpha})^{\theta+1}}, \tag{2.1}$$

where $\bar{\alpha} = 1-\alpha$, for $x \geq \alpha$, $\alpha \geq 0$ and $\theta \geq 0$ is a proper probability density function (pdf). As it can be easily seen, the parameter α marks a lower bound on the possible values that (2.1) can take on. This distribution can be considered a particular case of the log-gamma distribution (see [7] and [11])

$$\bar{f}(x|\beta,\lambda,\alpha) = \frac{\beta^{-\lambda}}{\Gamma(\lambda)}(x+\bar{\alpha})^{-\frac{1+\beta}{\beta}} \log(x+\bar{\alpha})^{\lambda-1} \tag{2.2}$$

when $\beta = \frac{1}{\theta}$ and $\lambda = 2$, assuming $x \geq \alpha$, $\beta > 0$ and $\lambda > 0$.

The particular case $\alpha = 0$ appears in [15]. The first derivative of (2.1) is

$$\frac{d}{dx}f(x) = \frac{\theta^2[1-(1+\theta)\log(x+\bar{\alpha})]}{(x+\bar{\alpha})^{2+\theta}}$$

from which it is simple verified that $f(x)$ increases in $(0, x_m)$ and decreases in (x_m, ∞), where $x_m = \exp((1+\theta)^{-1}) - \bar{\alpha}$ is the unique mode of the distribution. Note that the mode of the simple Pareto distribution with pdf $f(x) = \kappa\beta^\kappa/x^{\kappa+1}$, $x \geq \beta$, is located always at β.

Figure 1 shows the pdf of (2.1) for selected values of θ and $\alpha = 0$. As it can be inferred from this chart that the greater is the value of θ, the narrower is the peak and with thinner tails. On the other hand, for lower values of θ a more rounded peak and fatter tails are obtained.

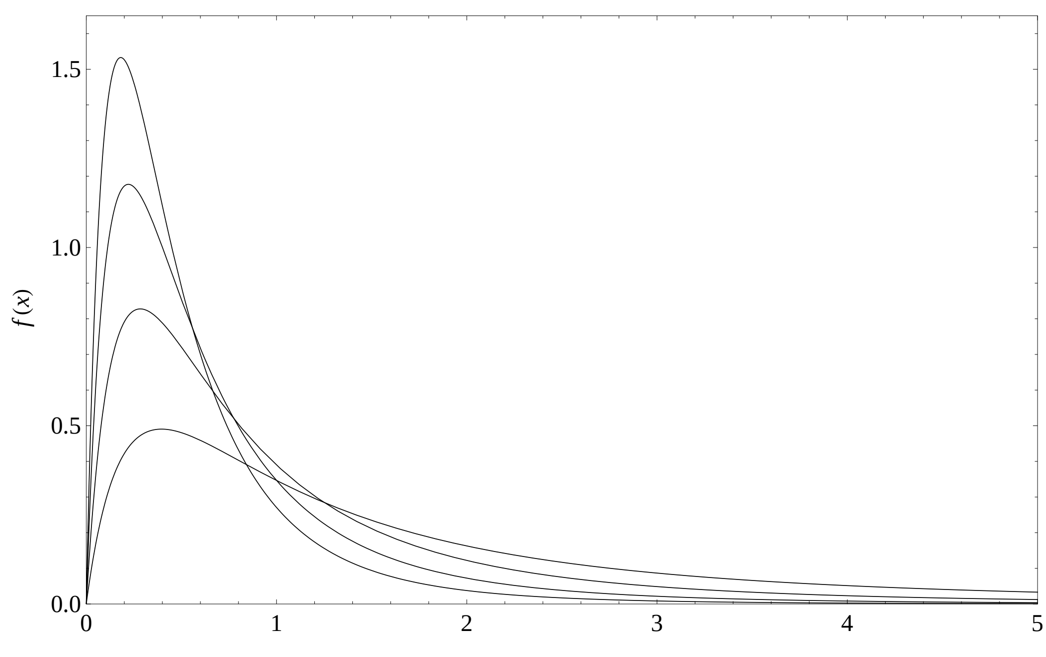

Figure 1. Graphs of the probability density function (2.1) for different values of parameter θ and $\alpha = 0$. From top to down $\theta = 5$, $\theta = 4$, $\theta = 3$ and $\theta = 2$

The distribution with pdf given in (2.1) can be obtained from a monotonic transformation of the $Gamma(2, \theta)$ distribution, as it can be seen in the next result.

2.1. Theorem. *Let us assume that Y follows a $Gamma(2, \theta)$ distribution with pdf*

$$g(y) = \theta^2\, y\, e^{-\theta\, y}, \quad y > 0,$$

where $\theta > 0$. Then the random variable

$$X = e^Y - \bar{\alpha} \tag{2.3}$$

is distributed according to (2.1) for $x \geq \alpha$, $\alpha \geq 0$ and $\theta \geq 0$.

Proof. The proof follows after a simple change of variable. Note that (2.3) is a monotonic function of Y. □

Simple computations provide the cumulative distribution function

$$F(x) = 1 - \frac{1 + \theta\, \log(x + \bar{\alpha})}{(x + \bar{\alpha})^\theta}, \quad x \geq \alpha. \tag{2.4}$$

Consequently, from (2.4) the survival function is easily derived

$$\bar{F}(x) = \frac{1 + \theta\, \log(x + \bar{\alpha})}{(x + \bar{\alpha})^\theta}, \quad x \geq \alpha, \tag{2.5}$$

and using (2.5) together with (2.1), the failure rate (hazard rate) function $r(x)$ is given by

$$r(x) = \frac{\theta^2 \log(x + \bar{\alpha})}{(x + \bar{\alpha})(1 + \theta \log(x + \bar{\alpha}))}, \quad x \geq \alpha. \tag{2.6}$$

Figure 2 displays the hazard rate function of (2.1) for selected values of θ and $\alpha = 0$.

From this chart it can be observed that the higher is the value of θ the higher is the value of $r(x)$ for a given $x \geq 0$.

Now, by using representation (2.3) and the moments of the Gamma distribution, the expression for the r-th moment about zero of distribution (2.1) is easily obtained,

$$\mu'_r = \sum_{j=0}^{r} \binom{r}{j} (-\bar{\alpha})^j \left(\frac{\theta}{\theta - r + j} \right)^2, \quad r = 1, 2, \ldots .$$

In particular we have

$$E(X) = \left(\frac{\theta}{1 - \theta} \right)^2 - \bar{\alpha}, \quad \theta \neq 1, \tag{2.7}$$

$$E(X^2) = \theta^2 \left[\frac{1}{(2 - \theta)^2} - \frac{2\bar{\alpha}}{(1 - \theta)^2} \right] + \bar{\alpha}^2, \quad \theta \neq 1, \theta \neq 2. \tag{2.8}$$

Therefore, from (2.7) and (2.8) the variance is

$$var(X) = \theta^2 \left[\frac{1}{(2 - \theta)^2} - \frac{\theta^2}{(1 - \theta)^4} \right], \quad \theta \neq 1, \theta \neq 2.$$

From (2.7) it can be seen that, for a fixed value of α, the mean increases when $\theta < 1$ and decreases when $\theta > 1$; similarly it always increases with α. It is also important to point out that for $\theta = 1$ the mean does not exist. In addition to this, second order moment around the origin and variance do not exist for $\theta = 2$. This is a practical limitation of this distribution. A similar drawback is shown by the single parameter Pareto distribution.

Furthermore, by using the representation given by (2.3) the following result is obtained

$$E(\log^r(X + \bar{\alpha})) = \theta^{-r}\, \Gamma(r + 2), \quad r = 1, 2, \ldots .$$

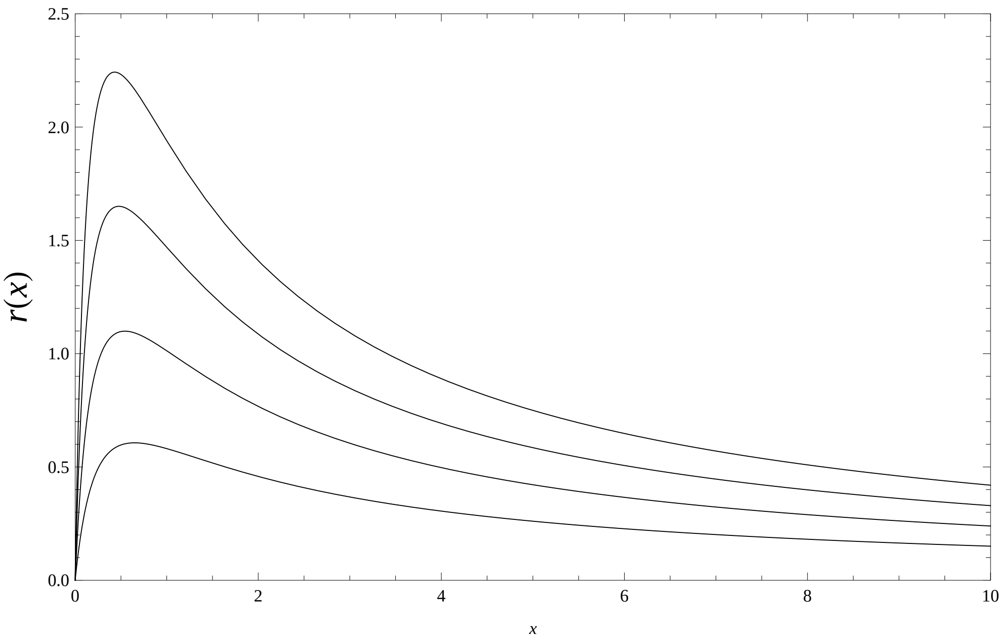

Figure 2. Graphs of the hazard rate function (2.6) for different values of parameter θ and $\alpha = 0$. From top to down $\theta = 5$, $\theta = 4$, $\theta = 3$ and $\theta = 2$.

In particular, we have that

$$E(\log(X + \bar{\alpha})) = \frac{2}{\theta}. \tag{2.9}$$

In the following, we will assume that X_1 and X_2 are two continuous random variables distributed according to (2.1) with pdfs given by $f(\cdot|\alpha, \theta_1)$ and $f(\cdot|\alpha, \theta_2)$, respectively. Now, we show that the distribution given in(2.1) can be ordered in terms of the likelihood ratio order when X_1 and X_2 have the same support. The likelihood ratio order is defined as follows (see Section 1.C of [21]).

2.2. Definition. Let X_1 and X_2 be continuous random variables with densities f_1 and f_2, respectively, such that

$$\frac{f_2(x)}{f_1(x)} \text{ is non-decreasing over the union of the supports of } X_1 \text{ and } X_2.$$

Then X_1 is said to be smaller than X_2 in the likelihood ratio order (denoted by $X_1 \leq_{LR} X_2$).

The likelihood ratio order is stronger than the hazard rate order that is defined as follows (see Section 1.B of [21]).

2.3. Definition. Let X_1 and X_2 be two random variables with hazard rates r_1 and r_2, respectively. Then X_1 is said to be smaller than X_2 in the hazard rate order, denoted by $X_1 \leq_{HR} X_2$, if $r_1(x) \geq r_2(x)$ for all x.

The distribution given in (2.1) can be ordered in the following way

2.4. Theorem. *Let X_1 and X_2 be two continuous random variables distributed according to (2.1) with pdfs given by $f(\cdot|\alpha,\theta_1)$ and $f(\cdot|\alpha,\theta_2)$, respectively. If $\theta_1 \geq \theta_2$ then $X_1 \leq_{LR} X_2$.*

Proof. Firstly, let us observe that the ratio

$$h(x) = \frac{f\left(x\,|\alpha,\theta_2\,\right)}{f\left(x\,|\alpha,\theta_1\,\right)} = \frac{\theta_2^2}{\theta_1^2}\left(x+\bar{\alpha}\right)^{\theta_1-\theta_2}$$

is non-decreasing if and only if $h'\left(x\right) \geq 0$ for $x \geq \alpha$.

Simple calculations show that

$$h'\left(x\right) = \frac{\theta_2^2}{\theta_1^2}\left(\theta_2-\theta_1\right)\left(x+\bar{\alpha}\right)^{\theta_1-\theta_2-1}.$$

Now, by having into account that $\theta_1 \geq \theta_2$ then $h'\left(x\right) \geq 0$ for $x \geq \alpha$ the result holds. □

As a consequence of this result, we have

2.5. Corollary. *Let X_1 and X_2 be two continuous random variables distributed according to (2.1) with pdf given by $f(\cdot|\alpha,\theta_1)$ and $f(\cdot|\alpha,\theta_2)$, respectively. If $\theta_1 \geq \theta_2$ then $r_1\left(x\right) \geq r_2\left(x\right)$ for all $x \geq \alpha$.*

Proof. It is well-known (see [21]) that

$$X_1 \leq_{LR} X_2 \Longrightarrow X_1 \leq_{HR} X_2. \tag{2.10}$$

Then, the result follows by combining (2.10) and Definition 2.3. □

2.1. Insurance applications. The integrated tail distribution (also known as equilibrium distribution) is an important probability model that often appears in insurance and many other applied fields. Let $\bar{F}(x)$ be the survival function given in (2.5), then the integrated tail distribution of $F(x)$ (see for example [24]) is given by $F_I(x) = \frac{1}{E(X)}\int_0^x \bar{F}(y)\,dy$. For the distribution proposed in this work, as it is proven in the following result, the integrated tail distribution can be obtained as a closed-form expression.

2.6. Proposition. *The integrated tail distribution of the cumulative distribution function (2.4) is given by*

$$\begin{aligned} F_I(x) &= \frac{1}{\theta^2-\bar{\alpha}(1-\theta)^2}\Big[\bar{\alpha}^{1-\theta}(2\theta-1-\theta(1-\theta)\log\bar{\alpha}) \\ &- (x+\bar{\alpha})^{1-\theta}(2\theta-1-\theta(1-\theta)\log(x+\bar{\alpha}))\Big]. \end{aligned} \tag{2.11}$$

Proof. By using (2.5) and (2.7) and after some algebra, the result follows. □

The failure rate of the integrated tail distribution, which is given by $\gamma_I(x) = \bar{F}(x)/\int_x^\infty \bar{F}(y)\,dy$ is also obtained in a closed-form. Furthermore, the reciprocal of $\gamma_I(x)$ is the mean residual life, which is given by

$$\begin{aligned} e(x) &= E(X-x|X>x) = \frac{1}{\bar{F}(x)}\int_x^\infty \bar{F}(y)\,dy \\ &= \frac{x+\bar{\alpha}}{(1-\theta)^2}\,\frac{\theta-(1-\theta)(1+\theta\log(x+\bar{\alpha}))}{1+\theta\log(x+\bar{\alpha})}. \end{aligned} \tag{2.12}$$

It is noted that, unlike the classical Pareto distribution this expression is not a linear function of x.

Under the classical model (see [19] and [24]) and assuming a positive security loading, ρ, for the claim size distributions with regularly varying tails we have that, by using

(2.11), it is possible to obtain an approximation of the probability of ruin, $\Psi(u)$, when $u \to \infty$. In this case the asymptotic approximations of the ruin function is given by

$$\Psi(u) \quad \sim \quad \frac{1}{\rho}\bar{F}_I(u), \quad u \to \infty,$$

where $\bar{F}_I(u) = 1 - F_I(u)$.

The use of heavy right-tailed distribution is of vital importance in general insurance. In this regard, Pareto and log-normal distributions have been employed to model losses in motor third liability insurance, fire insurance or catastrophe insurance. It is already known that any probability distribution, that is specified through its cumulative distribution function $F(x)$ on the real line, is heavy right-tailed if and only if for every $t > 0$, $e^{tx}\bar{F}(x)$ has an infinite limit as x tends to infinity. On this particular subject, (2.1) decays to zero slower than any exponential distribution and it is long-tailed since for any fixed $t > 0$ (see [18]) it is verified that

$$\bar{F}(x+t) \sim \bar{F}(x), \quad x \to \infty.$$

Therefore, as a long-tailed distribution is also heavy right-tailed, the distribution introduced in this manuscript is also heavy right-tailed.

Another important issue in extreme value theory is the regular variation (see [3] and [18]). This concept is formalized in the following definition.

2.7. Definition. A distribution function is called regular varying at infinity with index $-\beta$ if

$$\lim_{x\to\infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{-\beta},$$

where the parameter $\beta \geq 0$ is called the tail index.

Next theorem establishes that the cumulative distribution function given in (2.4) is a regular variation Lebesgue measure.

2.8. Theorem. *The pdf given in (2.1) is a distribution with regularly varying tails.*

Proof. Let us firstly consider the survival function given in (2.5). Then, after simple computations it is obtained that

$$\lim_{x\to\infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{-\theta},$$

and having into account that $\theta \geq 0$ the theorem hence. □

The pdf given in (2.1) can also be applied in rating excess-of-loss reinsurance as it can be seen in the next result.

2.9. Proposition. *Let X be a random variable denoting the individual claim size taking values only for individual claims greater than d. Let us also assumed that X follows the pdf (2.1), then the expected cost per claim to the reinsurance layer when the losses excess of m subject to a maximum of l is given by*

$$\begin{aligned} E[\min(l, \max(0, X-m))] &= \frac{1}{(1-\theta)^2}\Big[(m+\bar{d})^{1-\theta}(2\theta - 1 \\ &- \theta(1-\theta)\log(m+\bar{d})) - (l+m+\bar{d})^{1-\theta} \\ &\times (2\theta - 1 - \theta(1-\theta)\log(l+m+\bar{d}))\Big]. \end{aligned}$$

Proof. The result follows by having into account that

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

from which we get the result after some algebra. □

The following results show that both, the inverse Gaussian distribution and the gamma distribution are conjugate with respect to the distribution proposed in this work.

2.10. Proposition. *Let X_i, $i=1,2,\ldots,N$ independent and identically distributed random variables following the pdf (2.1). Let us suppose that θ follows a prior inverse Gaussian distribution $\pi(\theta)$ with parameters μ and ϕ, i.e. $\pi(\theta)\propto\theta^{-3/2}\exp\left[-\frac{1}{2}\left(\frac{\phi}{\mu^2}\theta+\frac{\phi}{\theta}\right)\right]$. Then the posterior distribution of θ given the sample information $(X_1,\ldots,X_n)$ is a generalized inverse Gaussian distribution $GIG(\lambda^*,\mu^*,\phi^*)$, where*

$$\begin{aligned}\lambda^* &= 2N-\frac{1}{2},\\ \mu^* &= \mu\left(\frac{\phi}{\phi+2\mu^2\sum_{i=1}^N\log(X_i+\bar{\alpha})}\right)^{1/2},\\ \phi^* &= \phi.\end{aligned}$$

Proof. The result follows after some computations by applying Bayes' Theorem and arranging parameters. □

2.11. Proposition. *Let X_i, $i=1,2,\ldots,N$ independent and identically distributed random variables following the pdf (2.1). Let us suppose that θ follows a prior gamma distribution $\pi(\theta)$ with a shape parameter λ and a scale parameter σ, i.e. $\pi(\theta)\propto\theta^{\lambda-1}\exp(-\sigma\theta)$. Then the posterior distribution of θ given the sample information $(X_1,\ldots,X_n)$ is again a gamma distribution with shape parameter $\lambda+2N$ and scale parameter $\sigma+\sum_{i=1}^N\log(X_i+\bar{\alpha})$.*

Proof. Again, the result follows after some algebra by using Bayes' Theorem and arranging parameters. □

This result can be used in credibility theory, where the premium charged to a policyholder is computed on the basis of his/her own past claims and the accumulated past claims of the corresponding portfolio of policyholders. In order to obtain an appropriate expression for this quantity, different approaches within Bayesian statistical decision theory have been proposed in the actuarial literature. In this sense, the following loss function is used to calculate an appropriate credibility for the pdf (2.1)

$$L(x,a)=\left[\log(x+\bar{\alpha})-a\right]^2. \tag{2.13}$$

This is a special case of the weighted squared-error loss function $L(x,a)=g(x)(h(x)-a)^2$ considered by [9]. For a comprehensive study of credibility theory see also [9] and [10]. Next result shows the risk, collective and Bayes premium under the loss given in (2.13).

2.12. Proposition. *Let us assumed that the claim size follows the pdf (2.1) and that θ follows a prior gamma distribution $\pi(\theta)$ with a shape parameter λ and a scale parameter σ. Let us also suppose that the practitioner compute the risk premium by using the loss* *function (2.13) and the collective and Bayes premium by using the net premium principle, i.e.* $L(x,a)=(x-a)^2$*, then the risk, collective and Bayes premium are given by*

$$P_R = \frac{2}{\theta},$$

$$P_C = \frac{2\sigma}{\lambda-1}, \quad \lambda>1, \tag{2.14}$$

$$P_B = \frac{2\left[\sigma+\sum_{i=1}^{N}\log(X_i+\bar{\alpha})\right]}{\lambda+2N-1}, \tag{2.15}$$

respectively.

Proof. The risk premium is obtained in the following way

$$P_R = \arg\min_a \int_\alpha^\infty \left[\log(x+\bar{\alpha})-a\right]^2 f(x|\alpha,\theta)\,dx,$$

from which it is not difficult to derive that $a=P_R=E[\log(x+\bar{\alpha})]=\frac{2}{\theta}$.

Now, the collective premium is computed as

$$P_C = \int_0^\infty \frac{2}{\theta}\pi(\theta)\,d\theta,$$

where $\pi(\theta)\propto\theta^{\lambda-1}\exp(-\sigma\theta)$, then it is directly obtained (2.14). Next, the Bayes premium is computed by simply replacing the parameter λ and σ by the updated parameters $\lambda+2N$ and $\sigma+\sum_{i=1}^{N}\log(X_i+\bar{\alpha})$. □

Now, it is proven that the net Bayes premium in (2.15) can be written as a credibility expression.

2.13. Corollary. *The net Bayes premium given in (2.15) is a credibility expression and can be rewritten as*

$$Z_N\, g(X_1,\ldots,X_N)+(1-Z_N)P_C,$$

where $Z_N=N/(N+K)$ *is the credibility factor,* $K=(\lambda-1)/2$ *and* $g(X_1,\ldots,X_N)=(1/N)\sum_{i=1}^{N}\log(X_i+\bar{\alpha})$.

Proof. After straightforward computations the result is obtained. □

Furthermore, the credibility factor in Corollary 2.13 is similar to the classical formula given by [6] that has been calculated for different pairs of likelihood-prior distributions. See for example [8] and [9].

2.14. Corollary. *It is verified that* $K=E[var(\log(x+\bar{\alpha})|\theta)]/var[E(\log(x+\bar{\alpha})|\theta)]$*, i.e. the credibility factor is of the same form as the Bühlmann credibility factor.*

Proof. The result follows again after some algebra. □

3. Estimation

In the following, we will outline two classical parameter estimation schemes, the moments method and the maximum likelihood method (ML). For that reason, let us assume that $\{x_1,x_2,\ldots,x_n\}$ is a random sample obtained from distribution (2.1). Without loss of generality $\widehat{\alpha}=\min\{x_i\},\ i=1,2,\ldots,n$, can be assumed to maximize the likelihood function as $x_i\geq\alpha$ for each $i=1,2,\ldots,n$. In other words, the parameter α is chosen to be equal to the smallest value of $x_i,\ i=1,2,\ldots,n$.

3.1. Moments method. The moment estimator of θ can be obtained by setting equal the theoretical mean obtained from (2.7) with the corresponding sample mean $\bar{x} = 1/n \sum_{i=1}^{n} x_i$. This simple equation is obtained

$$\left(\frac{\theta}{1-\theta}\right)^2 - \widehat{\bar{\alpha}} = \bar{x}, \tag{3.1}$$

where $\widehat{\bar{\alpha}}$ is the estimator of α given by $\widehat{\alpha} = \min\{x_i\},\ i = 1, 2, \ldots, n$. From (3.1) two possible solutions are found corresponding to the values given by

$$\hat{\theta} = \frac{\sqrt{\bar{x} + \widehat{\bar{\alpha}}}}{\sqrt{\bar{x} + \widehat{\bar{\alpha}}} \pm 1}$$

The moment estimator of θ can be also obtained directly from Table 1, where the exact value of the mean is displayed for different values of the parameter θ, given that α is known. Then, a numerical interpolation method can be used to derive, in an exact form, the estimator of θ.

Table 1. Mean of the distribution for different values of θ

0.6	0.8	2	4	6	8	10
$1.25 + \alpha$	$15 + \alpha$	$3.00 + \alpha$	$0.77 + \alpha$	$0.44 + \alpha$	$0.30 + \alpha$	$0.23 + \alpha$

Note also that it could be more convenient to use equation (2.9) to obtain the estimator of θ given by

$$\widehat{\theta} = \frac{2}{\bar{\bar{x}}}, \quad \bar{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} \log(x_i + \widehat{\alpha}), \tag{3.2}$$

since expression (3.2) is non-negative.

3.2. Maximum likelihood estimation. Parameter estimation by the moments method is simple; nevertheless, as not the whole set of observations is used, it tends to behave poorly. Anyhow, the moment estimate can be used as starting value to calculate maximum likelihood estimator. For that reason, let us assume that $\{x_1, x_2, \ldots, x_n\}$ is a random sample selected from the distribution (2.1). Then, the likelihood function is given by

$$L(\alpha, \theta | x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{\theta^2 \log(x_i + \bar{\alpha})}{(x_i + \bar{\alpha})^{\theta+1}}, \quad 0 < \alpha \leq \min\{x_i\}, \quad \theta > 0,$$

from which the log-likelihood function results

$$\begin{aligned} \log L(\alpha, \theta | x_1, x_2, \ldots, x_n) &= \frac{2n}{\theta} - \sum_{i=1}^{n} \log(\log(x_i + \bar{\alpha})) \\ &- (\theta + 1)\log(x_i + \bar{\alpha}). \end{aligned} \tag{3.3}$$

By setting the derivative of (3.3) equal to zero and, after some algebra, the maximum likelihood estimator (unique solution) of the parameter θ is given by

$$\widehat{\theta} = \frac{2n}{\sum_{i=1}^{n} \log(x_i + \widehat{\bar{\alpha}})}. \tag{3.4}$$

Besides, the second partial derivative is $\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=\widehat{\theta}} = -\frac{2n}{\widehat{\theta}^2} < 0$; therefore, the solution given by (3.4) corresponds to a maximum. From this expression, the asymptotic variance of the maximum likelihood estimator $\widehat{\theta}$, given by $\widehat{\theta}/\sqrt{2n}$, can be obtained.

3.3. Estimator performance. In this subsection, a simple performance analysis of the estimator obtained is developed. The following result shows that the estimator of θ is positively biased.

3.1. Theorem. *The maximum likelihood estimator $\widehat{\theta}$ of θ is positively biased.*

Proof. Let us assume that $\widehat{\theta} = g(T)$ where $g(T) = 2n/T$, being $T = \sum_{i=1}^{n} \log(\bar{\alpha} + T_i)$. Now, observe that $T > 0$ and $g''(T) = 4n/T^3 > 0$. Therefore, $g(T)$ is strictly convex. Thus, by Jensen's inequality we have that $E(g(T)) > g(E(T))$. Now

$$g(E(T)) = g(E(\sum_{i=1}^{n} \log(\bar{\alpha} + T))) = g\left(\frac{2n}{\theta}\right) = \theta.$$

Therefore we have that $E(\widehat{\theta}) > \theta$ and hence the theorem. □

3.2. Theorem. *The unique maximum likelihood estimator $\widehat{\theta}$ of θ is consistent and asymptotically normal and therefore*

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \widehat{\theta}/\sqrt{2n}).$$

Proof. The pdf (2.1) satisfies the regularity conditions (see [13], p. 449) under which the unique maximum likelihood estimator $\widehat{\theta}$ of θ is consistent and asymptotically normal. These conditions are simply verified in the following way. Firstly, the parameter space $(0, \infty)$ is a subset of the real line and the range of x is independent of θ. Then, it is not difficult to observe that $E(\frac{\partial \log f(x)}{\partial \theta}) = 0$. Now, since $\frac{\partial^2 \ell}{\partial \theta^2}\Big|_{\theta=\widehat{\theta}} < 0$, the Fisher's information is positive. Finally, by taking $M(x) = 4n^2/\theta^3$, we have that $\left|\frac{\partial^3 \log p_x}{\partial \lambda^3}\right| \leq M(x)$, being obviously $E(M(x)) < \infty$. □

Moreover, by using Corollary 3.11 in [13], p. 450, it may be concluded that the maximum likelihood estimate $\widehat{\theta}$ of θ is asymptotically efficient.

In addition to this, as a consequence of Theorem 3.2, we have that the large-sample $100(1-\alpha)\%$ confidence interval for θ is given by $\widehat{\theta} \pm z_{\alpha/2}\widehat{\theta}/\sqrt{2n}$, where $z_{\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution.

We also observe that T follows a gamma distribution with parameters $2n$ and θ. Therefore, it is obtained that

$$\begin{aligned} E(\widehat{\theta}) &= \frac{2n\theta}{2n-1}, \\ var(\widehat{\theta}) &= \frac{4n^2\theta^2}{(2n-1)^2(2n-2)}. \end{aligned}$$

Then, the bias of the estimator $\widehat{\theta}$ is given by $\theta/(2n-1)$ as it can be easily seen. Furthermore, the estimate $\widehat{\widehat{\theta}} = \frac{2n-1}{T} = \frac{2n-1}{2n}\widehat{\theta}$ is unbiased and it satisfies that $var(\widehat{\widehat{\theta}}) = \frac{\theta^2}{2n-2} < var(\widehat{\theta})$. Thus, $\widehat{\widehat{\theta}}$ is a better estimator of the parameter θ than $\widehat{\theta}$.

In Table 2, the variance of the two estimators $\widehat{\theta}$ and $\widehat{\widehat{\theta}}$ for different values of θ and sample size n is compared.

Table 2. Comparison of the variance of $\widehat{\theta}$ and $\widehat{\widehat{\theta}}$ and bias of $\widehat{\theta}$

	$\theta = 1$				$\theta = 2$		
n	$var(\widehat{\theta})$	Bias	$var(\widehat{\widehat{\theta}})$	n	$var(\widehat{\theta})$	Bias	$var(\widehat{\widehat{\theta}})$
5	0.1543	0.1111	0.1250	5	0.6170	0.2222	0.5000
10	0.0616	0.0526	0.0555	10	0.2460	0.1053	0.2222
15	0.0382	0.0344	0.0357	15	0.1529	0.0690	0.1429
20	0.0277	0.0256	0.0263	20	0.1107	0.0513	0.1053
25	0.0217	0.0204	0.0208	25	0.0868	0.0408	0.0833
50	0.0104	0.0101	0.0102	50	0.0416	0.0202	0.0408
100	0.0051	0.0050	0.0050	100	0.0204	0.0100	0.0202
	$\theta = 3$				$\theta = 4$		
n	$var(\widehat{\theta})$	Bias	$var(\widehat{\widehat{\theta}})$	n	$var(\widehat{\theta})$	Bias	$var(\widehat{\widehat{\theta}})$
5	1.3890	0.3333	1.1250	5	2.4700	0.4444	2.0000
10	0.5540	0.1579	0.5000	10	0.9850	0.2105	0.8890
15	0.3440	0.1034	0.3214	15	0.6120	0.1379	0.5710
20	0.2490	0.0769	0.2368	20	0.4430	0.1026	0.4211
25	0.1952	0.0612	0.1875	25	0.3470	0.0816	0.3333
50	0.0937	0.0303	0.0918	50	0.1666	0.0404	0.1633
100	0.0459	0.0150	0.0454	100	0.0816	0.0201	0.0808

4. Numerical experiments

In this section, the versatility of (2.1), as compared with the classical Pareto distribution, is tested by analyzing two different sets of continuous actuarial data. Later, the distribution introduced in this article is used together with other eight continuous probability distributions with positive support to fit a third set of loss data. At the view of the obtained results, the distribution introduced in this manuscript outperforms the performance of the classical Pareto distribution in the three examples considered.

The first set of data describes losses due to wind-related catastrophes that occurred in 1977. The losses were recorded to the nearest \$1,000,000; data also include 40 rounded loss amounts of \$2,000,000 or more. The data set appears in Table 6 in the Appendix section. These data have been taken from [11] (p.64), where they are discussed in detail. They have also been considered by [4], [5] and [17].

As [5] and [17] have pointed out, although these data are assumed to arise from a continuous distribution, they were discretized by grouping. Then, by using the method proposed by [5] and [17], they are de-grouped in the following way

$$x_k = \left(1 - \frac{k}{m+1}\right) a + \frac{k}{m+1}\, b, \quad k = 1, 2, \ldots, m,$$

where (a, b) contains m groups sample observations.

The second set of data analyzed in this section comes from the statistics of the French insurance union (see [25]). It displays all business interruption claims due to fire that are greater than \$1,600. Data comprise the year 1988 and claims are located in France. All the claims are valued in French francs. This set of data is shown in Table 7.

We have fitted the new distribution given in (2.1) and the classical Pareto distribution to the two data sets. Parameters have been estimated by the method of maximum likelihood. Table 3 provides parameter estimates together with standard errors (in brackets) computed by inverting the approximations of the observed information matrices. The maximum of the log-likelihood (ℓ_{max}) and Akaike's Information Criteria (AIC) have been firstly considered as measures of goodness-of-fit. From the results obtained, for the first set of data (Data set 1) the new distribution (New) gives a better fit to data based on these measures. The threshold considered has been 1.5, i.e. $\beta = 1.5$ for the Pareto distribution and $\alpha = 1.5$ for the new distribution. The estimator $\widehat{\widehat{\theta}}$ described above is 1.259 with a corresponding AIC value of 252.534, higher than the one obtained via the maximum likelihood estimator. Again, for the second set of data (Data set 2) the new model yields better fit with a lower AIC value.

Table 3. Estimates, standard errors, maximum of the likelihood ℓ_{max} and AIC for Data Set 1 and Data Set 2

	Data set 1		Data set 2	
	Pareto	New	Pareto	New
$\widehat{\kappa}$	0.764	–	0.465	–
	(0.121)		(0.003)	
$\widehat{\theta}$	–	1.275	–	0.783
		(0.142)		(0.142)
ℓ_{max}	–119.333	–118.779	–96782.70	–92141.70
AIC	240.665	239.559	193567	184285

In the following three goodness-of-fit measures for individual data based on the empirical distribution function such as the Kolmogorov-Smirnov (KS) test, Crámer von Mises test and the Anderson-Darling (AD) test (see [5] and [17] for details) have been applied to both Pareto and New. For that reason, only Data set 1 has been considered. As it can be observed in Table 6, the three tests support the suitability of the new distribution when $\alpha = 1.5$ to explain this set of data with goodness-of-fit values 0.1044 (KS), 0.0933 (CvM) and 0.7908 (AD). The first two values slightly improved the ones obtained for the Pareto distribution (with $\beta = 1.5$), 0.1072 (KS), 0.1107 (CvM) and 0.7336 (AD) respectively. These results are confirmed by their corresponding p-values. They have been calculated via Monte Carlo methods using a simulation size of 100000 repetitions.

Similarly, let $q_0 = \Pr(X \leq x_0) = F(x_0)$, $0 \leq q \leq 1$, be the values of q_0 given the corresponding x_0. The values of x_0, given by $x_0 = F^{-1}(q_0)$, can be easily computed by having into account that $x_0 = \exp\left(G^{-1}(q_0; 2, 1/\theta)\right) - \bar{\alpha}$, where $G^{-1}(\cdot)$ represents the inverse of the cumulative distribution function of the gamma distribution. It is not possible to obtain a closed-form expression for these quantiles. Nevertheless, by using the command from the Mathematica package `InverseCDF[GammaDistribution[2,1/θ,x0]`, they can be easily obtained. Then, quantiles for the first example were computed and, next, those values were used to plot quantile-quantile (Q-Q) plots. This chart consists of plots of the observed quantiles against theoretical quantiles for each one of the fitted models. In Figure 3, the Q-Q plots against the new model (righ-hand side) and classical Pareto distribution (left-hand side) are displayed for the first set of data. As usual, estimated quantiles are plotted on the vertical axis and ordered observations on the horizontal axis, where $F^{-1}(q)$ is the estimated q-th quantile and $q = \frac{j}{n+1}$ with $j = 1, \ldots, n$. According to this chart, it can be observed that the new model is a better reasonable choice for the given data.

Table 4. Goodness-of-fist tests and their corresponding p-values for fitted Pareto and New models in Data set 1

	Pareto	New
KS	0.1072	0.1044
p-value	0.7089	0.7388
CvM	0.1107	0.0933
p-value	0.5375	0.6209
AD	0.7336	0.7908
p-value	0.5288	0.4876

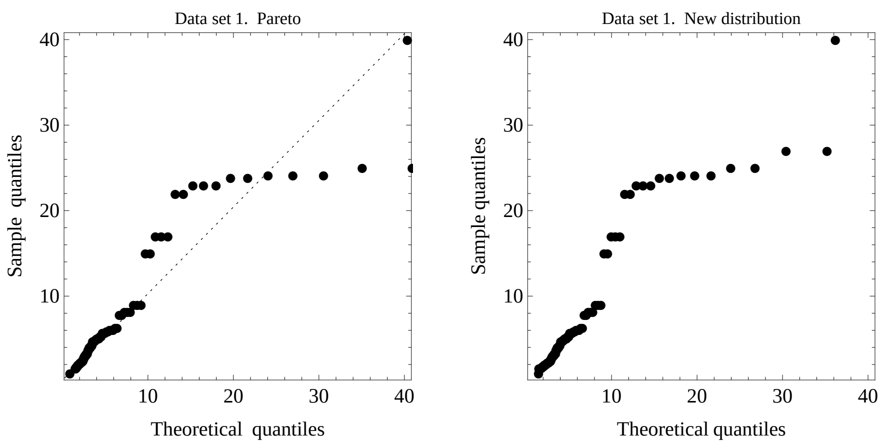

Figure 3. Sample and theoretical quantiles for Pareto and the new distribution for the first set of data.

Finally, the likelihood ratio test proposed by Vuong (see [23]) will be considered as a tool for to model diagnostic. The test statistic is

$$T = \frac{1}{\omega\sqrt{n}}\left(\ell_f(\widehat{\theta}_1) - \ell_g(\widehat{\theta}_2)\right),$$

where

$$\omega^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{f(x_i|\widehat{\theta}_1)}{g(x_i|\widehat{\theta}_2)}\right)\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f(x_i|\widehat{\theta}_1)}{g(x_i|\widehat{\theta}_2)}\right)\right]^2$$

and f and g represent the probability density function of the new distribution and Pareto distributions, respectively. Under the null hypotheses, $H_0 : E[\ell_f(\widehat{\theta}_1) - \ell_g(\widehat{\theta}_2)] = 0$, T is asymptotically normal distributed. It is generally accepted that rejection region for this test in favor of the alternative hypothesis occurs, at the 5% significance level, when $T > 1.96$.

For Data set 1 the value of the statistic T is 0.247, and therefore, no difference between the two model exists, whereas for Data set 2, as $T = 496.951 > 1.96$, the null hypothesis is rejected, and hence the new model is preferable.

Now, the log-gamma distribution given in (2.2) is used to model Data set 1. The likelihood function is maximized at $\hat{\beta} = 0.964$ and $\hat{\lambda} = 1.626$. We are interested, by means of the likelihood ratio test for nested models, in determining whether the new model introduced in this manuscript is suitable for the population that produced this set of data. In this regard, we could formulate the following test of hypothesis $H_0 : \lambda = 2$ vs. $H_1 : \lambda \neq 2$. The test statistic is $T = 2\,(\ell^1_{max} - \ell^2_{max})$ where ℓ^1 and ℓ^2 represent the maximum of the log-likelihood function for the log-gamma and new model respectively. This statistic follows a chi-square distribution with degrees of freedom equal to the number of free parameters under the alternative hypothesis. For the example considered the test statistic is $T = 2\,(-118.234 + 118.776) = 1.084$, it follows a chi-square distribution with 1 degree of freedom. Then, at the 5% significance level, as $1.084 < 3.841$ the null hypothesis is not rejected and the smaller model (New) is preferable at this significance level, the result is confirmed by the p-value (0.2978).

In the following, a final example (Data set 3) is considered to illustrate the value of the new distribution introduced in this manuscript. It deals with the set of Danish data on 2,492 fire insurance losses in Danish kroner (DKr) from the years 1980 to 1990, inclusive and adjusted to reflect 1985 values. This data set may be found in the "SMPracticals" add-on package for R, available from the CRAN website `http://cran.r-project.org/`. In order to explain the claim amount distribution given in this example, several probability laws have been used to model this set of data. They are listed in Table 5. For each distribution it is provided parameter estimates together with their corresponding standard errors (S.E.). Parameter estimation for the models considered was completed by the method of maximum likelihood. Two of the previously considered measures of goodness-of-fit have been used in this example, the maximum of the log-likelihood (ℓ_{max}) and the AIC. It is important to point out that the latter measure adjusts for the number of parameters in the distribution and it allows models with different numbers of parameters to be more fairly compared. The ℓ_{max} and AIC results show that the new model introduced here provides a better fit to data than do the Pareto, log-normal, Inverse Gaussian, Weibull, Gamma and Lomax distribution. However, as expected, the log-gamma distribution outperforms the new distribution. This result is confirmed by the p-value (< 0.0001) associated with the likelihood ratio test.

5. Concluding remarks

In this paper, a continuous probability distribution function with positive support suitable for fitting insurance data has been introduced. The distribution, that arises from a monotonic transformation of the classical Gamma distribution, can be considered a particular case of the log-gamma distribution. This new development, which has a promising approach for data modeling in actuarial field, may be very useful for practitioners who handles large claims. For that reason, it can be deemed as alternative to the classical Pareto distribution. Besides, an extensive analysis of its mathematical properties has been provided. In this regard, this new approach allows to compute new credibility expressions when the premium charged to a policyholder is computed on the basis of past claims and the accumulated past claims of the corresponding portfolio of policyholders under weighted squared-error loss function.

Acknowledgements. EGD was partially supported by ECO2009–14152 (MICINN, Spain). The authors are indebted to the two anonymous referees for helpful comments and suggestions which, without doubt, helped to improve an earlier version of the paper.

Table 5. Esimated values of fitted values, standard errors, maximum of the log-likelihood and AIC for the Danish Fire Claims Data (Data set 3).

Distribution	Parameter MLEs	S.E.	ℓ_{max}	AIC
Pareto	$\widehat{\kappa} = 0.546$	0.011	–5675.11	11352.2
New	$\widehat{\theta} = 1.971$	0.028	–4425.78	8853.57
Log-normal	$\widehat{\mu} = 0.672$	0.015	-4433.89	8871.78
	$\widehat{\sigma} = 0.732$	0.010		
Inverse Gaussian	$\widehat{\mu} = 3.063$	0.058	–4516.31	9036.61
	$\widehat{\sigma} = 3.417$	0.097		
Weibull	$\widehat{\mu} = 0.948$	0.011	–5270.47	10544.9
	$\widehat{\sigma} = 2.953$	0.066		
Gamma	$\widehat{\mu} = 1.258$	0.032	–5243.03	10490.1
	$\widehat{\sigma} = 2.435$	0.076		
Lomax	$\widehat{\alpha} = 5.169$	0.423	–5051.91	10107.8
	$\widehat{\lambda} = 11.900$	1.127		
Log-gamma	$\widehat{\lambda} = 3.917$	0.177	–4173.79	8351.57
	$\widehat{\beta} = 0.259$	0.006		

References

[1] Albrecher, H. and Kortchak, D. (2009). On ruin probability and aggregate claim representations for Pareto claim size distribution. *Insurance: Mathematics and Economics*, 45, 362–373.

[2] Arnold, B.C. (1983). *Pareto Distributions*, International Cooperative Publishing House, Silver Spring, MD.

[3] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. Cambridge University Press, Cambridge.

[4] Boyd, A.V. (1988). Fitting the truncated Pareto distribution to loss distributions. *JSS*, 31, 151–158.

[5] Brazauskas, V. and Serfling, R. (2003). Favorable estimator for fitting Pareto models: a study using goodness-of-fit measures with actual data. *Astin Bulletin*, 33, 2, 365–381.

[6] Bühlmann, H. (1967). Experience rating and credibility. *Astin Bulletin*, 4, 3, 199–207

[7] Dutang, C., Goulet, V. and Pigeon, M. (2008). actuar: An R Package for Actuarial Science. *Journal of Statistical Software*, 25, 7, 1–37.

[8] Gómez-Déniz, E. (2009). Some Bayesian Credibility Premiums Obtained by Using Posterior Regret Γ–Minimax Methodology. *Bayesian Analysis*, 4, 2, 223-242.

[9] Heilmann, W. (1989) Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics* 8, 1, 77–95.

[10] Herzog, T.N. (1996) *Introduction to Credibility Theory*, 2nd ed. ACTEX Publications, Winsted.

[11] Hogg, R. and Klugman, S.A. (1984). *Loss Distributions*. John Wiley and Sons, New York.

[12] Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economic and Actuarial Sciences*. Wiley.

[13] Lehmann, E.L. and Casella, G. (1998). Theory of Point Estimation. Springer, New York.

[14] Mata, A. (2000). Princing excess of loss reinsurance with reinstatements. *Astin Bulletin*, 30, 2, 349–368.

[15] Patel, J. (1973). A catalogue of failure distributions. *Communications in Statistics–Theory and Methods*, 1, 3, 281–284.

[16] Pinto, E. and Gogol, D.F. (1987). An analysis of excess loss development. *Proceedings of the Casualty Actuarial Society*, LXXIV, 2, 142, 227–255.

[17] Rizzo, M.L. (2009). New goodness-of-fit tests for Pareto distributions. *Astin Bulletin*, 39, 2, 691–715.

[18] Rytgaard, M.. (1990). Estimation in the Pareto distribution. *Astin Bulletin*, 20, 2, 201–216.

[19] Rolski, T., Schmidli, H. Schmidt, V. and Teugel, J. (1999). Stochastic processes for insurance and finance. John Wiley & Sons.

[20] Sarabia, J.M. and Prieto, F. (2009). The Pareto-positive stable distribution: a new descriptive model for city size data. *Physica A*, 388, 4179–4191.

[21] Shaked M. and Shanthikumar J.G. (2007). Stochastic orders. Series: Springer Series in Statistics, Springer, New York.

[22] Stoppa, G. (1990). Proprietà campionarie di un nuovo modello Pareto generalizzato. *Atti XXXV Riunione Scientifica della Società Italiana di Statistica, Padova: Cedam*, 137–144.

[23] Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.

[24] Yang, H. (2004). Cramér-Lundberg asymptotics. In *Encyclopedia of Actuarial Science*. Wiley.

[25] Zajdenweber, D. (1996). Extreme values in business interruption insurance. *The Journal of Risk and Insurance*, 63, 1, 95–110.

Appendix

Table 6. Data set 1. Losses due to wind-related catastrophes in the U.S. (millions of dollars) recorded in 1977. (See [11])

2	2	2	2	2	2	2	2	2	2
2	2	3	3	3	3	4	4	4	5
5	5	5	6	6	6	6	8	8	9
15	17	22	23	24	24	25	27	32	43

Table 7. Data set 2. 1988 size distribution data (in millions of francs) for French firms. (See [25])

Size classes	Average sizes	Number of firms insured
Less than 4	n.a.	2612
4–7	5.7	6783
7–12	9.4	5854
12–25	17.3	5063
25–50	34.7	2273
50–100	70.6	1094
100–200	138.2	514
200–500	304.6	311
500–1000	708.3	100
1000–2000	1216.4	49
2000–5000	3085.3	34
Over 5000	10561	16