Modeling Large Claims with Composite Stoppa Models

Enrique Calderín–Ojeda[a] and Chun Fung Kwok[b]

[a] *Centre for Actuarial Studies, Department of Economics, University of Melbourne, Australia*

[b] *Department of Mathematics and Statistics, University of Melbourne, Australia*

Abstract

In general insurance large losses sometimes arise. The costs faced by insurers often emerge from a mix of moderate and large claims. The payments are typically highly positively skewed and distributed with a thick upper tail. Nevertheless, no standard parametric model seems to provide an acceptable fit to both small and large losses. To overcome this issue, composite parametric models that use Lognormal or Weibull distributions up to the modal value and the three–parameter Stoppa density thereafter are considered in this paper. The derivation of these models is based on a mode–matching procedure. Their performance is compared to the existing composite Pareto families in the context of the well–known Danish fire insurance data.

Key Words: *Stoppa distribution; Lognormal distribution; Weibull distribution; Danish fire losses; Spliced model; Goodness–of–fit; Composite model; Bootstrap.*

1 Introduction

In general insurance large losses sometimes arise. The costs faced by insurers often emerge from a mixture of moderate and large claims. The payments are typically unimodal, highly positively skewed and distributed with a larger upper tail. In this circumstance, the Pareto distribution has been traditionally considered by researchers to model the larger loss data. On the other hand, when losses consist of smaller data with high frequencies and larger losses with low frequencies other continuous parametric families with positive support such as Gamma, Lognormal, Inverse Gaussian and Weibull have been used (see Klugmann et al. (2008)). Nevertheless, no standard parametric model seems to provide an acceptable fit to both small and large losses since probability distributions that provide a good overall fit can be particularly bad at fitting the tail.

To overcome this issue, composite parametric models that use Lognormal (see Cooray and Ananda (2005)) or Weibull (see Ciumara (2006)) up to an unknown single threshold value, estimated from the data, and a two-parameter Pareto density thereafter have been considered. In both approaches, continuity and differentiability conditions at the threshold were imposed to yield a smooth density function and reduce the number of parameters. The resulting models are similar in shape to the Lognormal and Weibull distributions but with a thicker tail. However, these two-component composite models are very restrictive since they have a fixed and a priori known mixing weights. Scollnik (2007) improved the composite Lognormal–Pareto model by using unrestricted mixing weights as coefficients in each component (see also Nadarajah and Bakar (2014)). In a similar fashion, Scollnik and Sun (2012) amended the composite Weibull–Pareto model. Both papers also considered the three–parameter Lomax distribution above the threshold as an extension to the composite Pareto family. Moreover, as both Pareto and Lomax distributions are monotonically decreasing, the resulting composite models are unimodal and their modal values are lower than the scale parameter (threshold) of these distributions.

A new composite models based on the Stoppa distribution is proposed in this work. The Stoppa distribution, not sufficiently well–described in the English language literature, is a generalization of the Pareto distribution (see Stoppa (1990)). This family of distributions is obtained by applying a power transformation to the cumulative distribution function (cdf) of the Pareto distribution. Recently, Burkhauser et al. (2010) used it to model topcoded income values and yielded positive outcomes. One of the main features of the Stoppa distribution is that it presents a heavier tail than the classical Pareto distribution when the additional shape parameter is larger than one. The methodology proposed in this paper assumes a composite-Stoppa model with unrestricted mixing weights via a mode–matching process. Hence, the first component of the spliced model is used up to the modal value (e.g. Lognormal or Weibull), which can be estimated from the data, and the adequate truncation of the Stoppa distribution thereafter.

This methodology presents several advantages over the traditional methods considered in the literature. Firstly, it allows the use of any continuous distribution whose mode can be written in closed form, in the construction of a composite model with the Stoppa distribution. Second, this mode–matching procedure simplifies the derivation of the model significantly, since the mode of the continuous distributions that are commonly considered in the content of actuarial modelling has a much simpler expression than the derivative of the corresponding density function. Third, this approach facilitates the computational implementation of the composite model. In this manuscript two new composite families, the Lognormal–Stoppa and Weibull–Stoppa distributions are derived using this methodology.

The structure of this paper is as follows. In Section 2 a short review on composite Pareto models with unrestricted mixing weight is provided. Next in Section 3 some existing results on the Stoppa distribution are given. Later in Section 4, the genesis of two composite Stoppa models, Lognormal–Stoppa and Weibul–Stoppa, is described. Afterwards, in Section 5 numerical illustrations are provided based upon the well–known the Danish fire insurance data. Here, the composite Stoppa models is compared with existing composite models from two points of view, theoretical plausibility and practical consideration. Finally conclusions are given in the last Section.

2 Composite Pareto Models with unrestricted mixing weights

Scollnik (2007) improved the composite model given in Cooray and Ananda (2005) by incorporating unrestricted mixing weights. The density function of the composite model can be written as

$$f(x) = \begin{cases} r\, f_1^*(x), & 0 < x \leq \theta \\ (1-r)\, f_2^*(x), & \theta < x < \infty \end{cases} \tag{1}$$

with $0 \leq r \leq 1$, $f_1^*(x) = \dfrac{f_1(x)}{F_1(\theta)}$ and $f_2^*(x) = \dfrac{f_2(x)}{1 - F_2(\theta)}$ are adequate truncations of the probability density functions f_1 and f_2 up to and thereafter an unknown threshold value θ where $F_1(\theta)$ and $F_2(\theta)$ denote the cumulative distribution function (cdf) of f_1 and f_2 at θ respectively. Then (1) can be seen as a convex sum of two density functions. After imposing the continuity condition (i.e. $f(\theta^-) = f(\theta^+)$), we have

$$r = \frac{f_2(\theta)\, F_1(\theta)}{f_2(\theta)\, F_1(\theta) + f_1(\theta)\,(1 - F_2(\theta))}. \tag{2}$$

Next, differentiability condition at θ was also imposed in order to make (1) smooth and to reduce the number of parameters.

2.1 Lognormal–Pareto Models

Let

$$f_1(x) = \frac{1}{\sqrt{2\,\pi}\, x\, \sigma} \exp\left(-\frac{1}{2} \left(\frac{\ln x - \mu}{\sigma} \right)^2 \right), \quad x > 0 \tag{3}$$

be the probability density function (pdf) of a two–parameter Lognormal distribution and

$$f_2(x) = \frac{\alpha\, \theta^\alpha}{x^{\alpha+1}}, \quad x > \theta, \tag{4}$$

be the pdf of a two–parameter Pareto distribution.
The density function of this composite model is given by

$$f(x) = \begin{cases} r\, \dfrac{f_1(x)}{\Phi(\alpha\,\sigma)}, & 0 < x \leq \theta \\ (1-r)\, f_2(x), & \theta < x < \infty \end{cases} \tag{5}$$

with $0 \leq r \leq 1$ and $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. By allowing for continuity and differentiability at θ, we have that

$$
\begin{aligned}
r &= \frac{\sqrt{2\,\pi}\,\alpha\,\sigma\,\Phi(\alpha\,\sigma)\,\exp\left(\frac{1}{2}(\alpha\,\sigma)^2\right)}{\sqrt{2\,\pi}\,\alpha\,\sigma\,\Phi(\alpha\,\sigma)\,\exp\left(\frac{1}{2}(\alpha\,\sigma)^2\right)+1} \quad \text{and} \\
\alpha\,\sigma &= \frac{\ln\theta-\mu}{\sigma}.
\end{aligned}
$$

Then (20) is defined by means of the threshold θ, a tail index α and a small loss parameter σ.

Scollnik (2007) also considered another composite model, the Lognormal-Type II Pareto (Lomax) with pdf

$$
f(x) = \begin{cases} r\,\dfrac{f_1(x)}{\Phi\left(\frac{\ln\theta-\mu}{\sigma}\right)}, & 0 < x \leq \theta \\ (1-r)\,f_2(x), & \theta < x < \infty \end{cases} \tag{6}
$$

where $f_2(x)$ is the pdf of the Type II Pareto which is given by

$$
f_2(x) = \frac{\alpha\,(\lambda+\theta)^{\alpha}}{(\lambda+x)^{\alpha+1}}, \quad x > \theta,\ \theta > 0,\ \alpha > 0, \text{ and } \lambda > -\theta.
$$

After imposing the continuity requirement at θ, r is provided by

$$
r = \frac{\sqrt{2\,\pi}\,\alpha\,\theta\,\sigma\,\Phi\left(\dfrac{\ln\theta-\mu}{\sigma}\right)\,\exp\left(\dfrac{1}{2}\left(\dfrac{\ln\theta-\mu}{\sigma}\right)^2\right)}{\sqrt{2\,\pi}\,\alpha\,\theta\,\sigma\,\Phi\left(\dfrac{\ln\theta-\mu}{\sigma}\right)\,\exp\left(\dfrac{1}{2}\left(\dfrac{\ln\theta-\mu}{\sigma}\right)^2\right)+\lambda+\theta}. \tag{7}
$$

Similarly, imposing the differentiability condition at θ gives a smooth density function, at the same time reducing the number of parameters of the model to four. In this case

$$
\frac{\ln\theta-\mu}{\sigma} = \left(\frac{\alpha\,\theta-\lambda}{\lambda+\theta}\right)\,\sigma.
$$

Note that this model nests the composite Lognormal–Pareto model if $\lambda = 0$.

2.2 Weibull–Pareto Models

Let

$$f_1(x) = \frac{\tau}{x}\left(\frac{x}{\phi}\right)^{\tau}\exp\left(-\left(\frac{x}{\phi}\right)^{\tau}\right), \quad x > 0 \tag{8}$$

be the pdf of a two–parameter Weibull distribution and

$$f_2(x) = \frac{\alpha\,\theta^{\alpha}}{x^{\alpha+1}}, \quad x > \theta, \tag{9}$$

be the pdf of a two–parameter Pareto distribution, where $\theta > 0$, $\alpha > 0$, $\tau > 0$ and $\phi > 0$.

Scollnik and Sun (2012) constructed a composite model with pdf

$$f(x) = \begin{cases} r\,\dfrac{f_1(x)}{F_1(\theta)}, & 0 < x \leq \theta \\ (1-r)\,f_2(x), & \theta < x < \infty \end{cases} \tag{10}$$

with $0 \leq r \leq 1$ and $F_1(\theta)$ is the cdf of the Weibull distribution at θ. By allowing for continuity at θ, r is given by

$$r = \frac{\dfrac{\alpha}{\tau}}{\dfrac{\left(\dfrac{\theta}{\phi}\right)^{\tau}}{\exp\left(\left(\frac{\theta}{\phi}\right)^{\tau}\right)-1}+\dfrac{\alpha}{\tau}}. \tag{11}$$

The differentiability condition leads to $\left(\dfrac{\theta}{\phi}\right)^{\tau} = \dfrac{\alpha}{\tau}+1$, simplifying r to

$$r = \frac{\alpha\,\exp\left(\dfrac{\alpha}{\tau}+1\right)-\alpha}{\alpha\,\exp\left(\dfrac{\alpha}{\tau}+1\right)+\tau}, \tag{12}$$

leading to a three-parameter composite model.

In a similar fashion as in the Lognormal–Pareto composite family, Scollnik and Sun (2012) also considered the composite Weibull-Type II Pareto (Lomax) with pdf

$$f(x) = \begin{cases} r\,\dfrac{f_1(x)}{F_1(\theta)}, & 0 < x \leq \theta \\ (1-r)\,f_2(x), & \theta < x < \infty \end{cases} \tag{13}$$

where $f_2(x)$ is the pdf of the Type II Pareto and it is given by

$$f_2(x) = \frac{\alpha\,(\lambda+\theta)^{\alpha}}{(\lambda+x)^{\alpha+1}}, \quad x>\theta,\ \theta>0,\ \alpha>0,\ \text{and}\ \lambda>-\theta.$$

Then by imposing the continuity condition at θ, r is provided by

$$r = \frac{\dfrac{\alpha}{\tau}}{\dfrac{\lambda+\theta}{\theta}\dfrac{\left(\frac{\theta}{\phi}\right)^{\tau}}{\exp\left(\left(\frac{\theta}{\phi}\right)^{\tau}\right)-1}+\dfrac{\alpha}{\tau}}. \tag{14}$$

Analogously, the resulting density function is smooth by imposing the differentiability condition at θ, and the number of parameters of the model is reduced to four. In this case the following restriction is included

$$\left(\frac{\theta}{\phi}\right)^{\tau} = \frac{\alpha\,\theta-\lambda}{(\lambda+\theta)\,\tau}+1,$$

clearly, this model nests the composite Weibull–Pareto model if $\lambda = 0$.

3 Generalizing the Pareto distribution

Although not sufficiently well–described in the English language literature, a generalization of the Pareto distribution was proposed by Stoppa (1990). The methodology to derive this family of distributions involves applying a power transformation to the Pareto cdf. The cdf of the Stoppa distribution is given by

$$F(x) = \left[1-\left(\frac{x}{x_0}\right)^{-\delta}\right]^{\gamma}, \quad 0<x_0\leq x,$$

where δ, $\gamma>0$ specify the shape of the distribution and x_0 is the minimum possible value. The classical Pareto distribution is obtained when $\gamma = 1$. The pdf of the Stoppa distribution is provided by

$$f(x) = \gamma\,\delta\,x_0^{\delta}\,x^{-(\delta+1)}\left[1-\left(\frac{x}{x_0}\right)^{-\delta}\right]^{\gamma-1}, \quad 0<x_0\leq x.$$

Some properties of this distribution can be found in Kleiber and Kotz (2003). In this regard, the k^{th} order moment exists for $k < \delta$ is given by

$$E(X^k) = \gamma\, x_0^k\, Be\left(1 - \frac{k}{\delta}, \gamma\right),$$

where $Be(\cdot,\cdot)$ represents the beta function defined by

$$Be(a,b) = \int_0^1 z^{a-1}\,(1-z)^{b-1}\,dz \quad \text{with} \quad a,b > 0.$$

Additionally, the quantile function can easily derived,

$$F^{-1}(u) = x_0\,\left(1 - u^{1/\gamma}\right)^{-1/\delta}, \quad 0 < u < 1.$$

As compared with the Pareto distribution, the Stoppa distribution is more flexible since it includes an additional shape parameter γ that allows for unimodality for $\gamma > 1$ and zeromodality when $\gamma \leq 1$. For the former case the mode is located at

$$x_{Mode} = x_0\,\left(\frac{1+\gamma\,\delta}{1+\delta}\right)^{1/\delta},\ \gamma > 1, \tag{15}$$

whereas for the latter one at x_0. Figure 3 shows the effect of increasing the shape parameter γ on probability density function of the Stoppa distribution while holding x_0 and δ fixed. As $\gamma \geq 1$ increases, the mode moves to the right producing a thicker tail.

4 The Composite Stoppa Models

The Composite Pareto models described use the Pareto distribution above the threshold value θ. However, as both classical and Type II Pareto distributions are monotonically decreasing, θ is always larger than the modal value of the composite model. On the other hand, as discussed in the previous section, the Stoppa distribution has a heavier tail than the classical Pareto distribution when $\gamma > 1$. Then, it would be natural to assume composite model with unrestricted mixing weights based on the Stoppa distribution via a mode–matching process. The first component of the spliced model is used up to the modal value (that can be estimated from the data) and the adequate truncation of the Stoppa distribution thereafter.

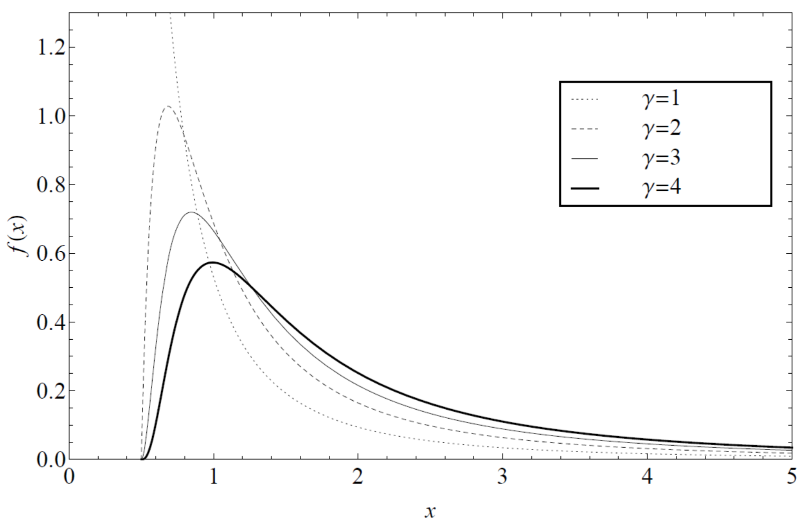


Figure 1: Stoppa densities for $x_0 = 0.5$, $\delta = 1.5$, and $\gamma = 1$ (dotted), $\gamma = 2$ (dashed), $\gamma = 3$ (solid thin) and $\gamma = 4$ (solid thick).

Then, the density function of the Composite Stoppa model can be written as

$$f(x) = \begin{cases} r\, f_1^*(x), & 0 < x \leq x_m \\ (1-r)\, f_2^*(x), & x_m < x < \infty \end{cases} \tag{16}$$

with $0 \leq r \leq 1$ and $f_1^*(x) = \dfrac{f_1(x)}{F_1(x_m)}$ is an adequate truncation of the probability density functions f_1 up to the modal value where $F_1(x_m)$ is the cdf of f_1 evaluated at x_m and $f_2^*(x) = \dfrac{f_2(x)}{1-F_2(x_m)}$ is also a suitable truncation of the Stoppa thereafter, where $1 - F_2(x_m)$ the survival function evaluated at x_m. Again, (16) can be seen as a convex sum of two density functions.

In place of the usual continuity and differentiability conditions, a mode–matching procedure is used. This procedure ensures the continuity and differentiability conditions are satisfied. In addition, it gives a much simpler derivation of the model and allows for an easy implementation of any distributions with a mode that has a closed form expression. The mode–matching conditions are given as follows.

Denote the modes of the distributions used by the first and second components of the composite model by $x_m^{first}, x_m^{second}$ respectively. Then the mode–matching conditions are:

$$x_m^{first} = x_m^{second} \tag{17}$$

$$f_1^*(x_m^{first}) = f_2^*(x_m^{second}) \tag{18}$$

Clearly, (18) implies the continuity condition is satisfied, and since the equality in (17) allows us to drop the 'first' and 'second' labels, the simple equation of the mixing weight, as seen in the other existing composite models, is preserved and is given by

$$r = \frac{f_2(x_m)\, F_1(x_m)}{f_2(x_m)\, F_1(x_m) + f_1(x_m)\, (1 - F_2(x_m))}. \tag{19}$$

Next, note that the derivative at the mode is 0, so it is clear the differentiability condition is also satisfied.

4.1 Lognormal–Stoppa Model

The composite Lognormal–Stoppa model will be derived in terms of the mixture model (16). Its density function is given by

$$f(x) = \begin{cases} r \, \dfrac{\dfrac{1}{\sqrt{2\,\pi}\,x\,\sigma}\,\exp\left(-\dfrac{1}{2}\left(\dfrac{\ln x-\mu}{\sigma}\right)^2\right)}{\Phi\left(\frac{\ln x_m-\mu}{\sigma}\right)}, & 0 < x \leq x_m \\ (1-r)\,\dfrac{\gamma\,\delta\,x_0^{\delta}\,x^{-(\delta+1)}\left[1-\left(\frac{x}{x_0}\right)^{-\delta}\right]^{\gamma-1}}{1-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}, & x_m < x < \infty \end{cases} \tag{20}$$

with $\mu \in \mathbb{R}$, $\sigma > 0$, $\gamma > 1$, $\delta > 0$, $0 \leq r \leq 1$ and $\Phi(\cdot)$ denotes the cdf of the standard normal distribution.

Using the mode–matching procedure, (17) gives

$$\sigma = \sqrt{\mu - \ln\left[x_0\left(\frac{1+\gamma\delta}{1+\delta}\right)^{1/\delta}\right]}, \tag{21}$$

note that this result implies that $\mu > \ln\left[x_0\left(\dfrac{1+\gamma\delta}{1+\delta}\right)^{1/\delta}\right]$. Then, substituting the corresponding densities and distribution functions into (19) gives

$$\begin{aligned} r &= \gamma\,\delta\,x_0^{\delta}\,x_m^{-(\delta+1)}\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma-1}\Phi\left(\frac{\ln x_m-\mu}{\sigma}\right) \\ &\times \left\{\gamma\,\delta\,x_0^{\delta}\,x_m^{-(\delta+1)}\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma-1}\Phi\left(\frac{\ln x_m-\mu}{\sigma}\right)\right. \\ &+ \left.\frac{1}{\sqrt{2\,\pi}\,x_m\,\sigma}\exp\left(-\frac{1}{2}\left(\frac{\ln x_m-\mu}{\sigma}\right)^2\right)\left(1-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}\right)\right\}^{-1}. \end{aligned}$$

These guarantee (20) is continuous and smooth. Note that the number of parameters are reduced to four.

The cdf of the composite Lognormal–Stoppa distribution is provided by

$$F(x) = \begin{cases} r \dfrac{\Phi\left(\dfrac{\ln x - \mu}{\sigma}\right)}{\Phi\left(\dfrac{\ln x_m - \mu}{\sigma}\right)} \\ r + (1-r)\, \dfrac{\left[1 - \left(\frac{x}{x_0}\right)^{-\delta}\right]^{\gamma} - \left[1 - \left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}{1 - \left[1 - \left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}, \quad x_m < x < \infty \end{cases} \tag{22}$$

Furthermore, the moment of order k^{th} of the composite Lognormal–Stoppa distribution exists when $\delta > k$. Its analytical expression is given by

$$\begin{aligned} E(X^k) &= r \frac{\Phi\left(\dfrac{\ln x_m - \mu - k\sigma^2}{\sigma}\right)}{\Phi\left(\dfrac{\ln x_m - \mu}{\sigma}\right)} e^{k\mu + \frac{k^2\sigma^2}{2}} \\ &+ (1-r)\, \frac{1}{1 - \left[1 - \left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}} \gamma\, x_0^k\, Be\left(\left(\frac{x_m}{x_0}\right)^{-\delta}; 1 - \frac{k}{\delta}, \gamma\right), \end{aligned}$$

where $Be(\cdot;\cdot,\cdot)$ represents the incomplete beta function defined by

$$Be(x; a, b) = \int_0^x z^{a-1}\,(1-z)^{b-1}\,dz \quad \text{with} \quad a, b > 0.$$

In preparation for the numerical applications section, a procedure for generating random variates from the composite Lognormal–Stoppa distribution is presented. As the cdf of the Lognormal and Stoppa distributions can be inverted, the inverse transformation method of simulation can be used for this composite family. If u is a value generated from the uniform distribution $U(0,1)$, then a value generated from (20) is obtained as

- If $u \leq r$ then solve

$$x = \exp\left\{\mu + \sigma \cdot \Phi^{-1}\left(\frac{u}{r}\,\Phi\left(\frac{\ln x_m - \mu}{\sigma}\right)\right)\right\}$$

- If $u > r$ then

$$x = x_0\left\{1 - \left(\frac{u-r}{1-r}\left[1 - \left(1 - \left(\frac{x_m}{x_0}\right)^{-\delta}\right)^{\gamma}\right] + \left(1 - \left(\frac{x_m}{x_0}\right)^{-\delta}\right)^{\gamma}\right)^{1/\gamma}\right\}^{-1/\delta}$$

4.2 Weibull–Stoppa Model

The composite Weibull–Stoppa model will be also obtained in terms of the mixture model (16). Its density function is given by

$$f(x) = \begin{cases} r\,\dfrac{1}{1-\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)}\left(\dfrac{\tau}{x}\right)\left(\dfrac{x}{\phi}\right)^{\tau}\exp\left(-\left(\dfrac{x}{\phi}\right)^{\tau}\right), & 0 < x \le x_m \\ (1-r)\,\dfrac{\gamma\,\delta\,x_0^{\delta}\,x^{-(\delta+1)}\left[1-\left(\frac{x}{x_0}\right)^{-\delta}\right]^{\gamma-1}}{1-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}, & x_m < x < \infty \end{cases} \tag{23}$$

with $\phi > 0$, $\gamma > 1$, $\delta > 0$, $0 \le r \le 1$ and $\tau > 1$ to define a positive modal value.

Now, again applying the mode-matching conditions, (17) gives

$$\phi = \left[x_0\left(\frac{1+\gamma\delta}{1+\delta}\right)^{1/\delta}\right]\left(\frac{\tau}{\tau-1}\right)^{1/\tau}. \tag{24}$$

Similarly, substituting the corresponding densities and distribution functions into (19) gives

$$\begin{aligned} r &= \gamma\,\delta\,x_0^{\delta}\,x_m^{-(\delta+1)}\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma-1}\left(1-\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)\right) \\ &\times \left\{\gamma\,\delta\,x_0^{\delta}\,x_m^{-(\delta+1)}\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma-1}\left(1-\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)\right)\right. \\ &+ \left.\left(\frac{\tau}{x_m}\right)\left(\frac{x_m}{\phi}\right)^{\tau}\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)\left(1-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}\right)\right\}^{-1}. \end{aligned}$$

The cdf of the composite Weibull–Stoppa distribution is yielded by

$$F(x)=\begin{cases} r\,\dfrac{1-\exp\left(-\left(\frac{x}{\phi}\right)^{\tau}\right)}{1-\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)}, & 0<x\le x_m \\ r+(1-r)\,\dfrac{\left[1-\left(\frac{x}{x_0}\right)^{-\delta}\right]^{\gamma}-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}{1-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}, & x_m<x<\infty \end{cases} \tag{25}$$

Now, k^{th} order moment of the composite Weibull–Stoppa distribution exists again if $\delta > k$. Its analytical expression is provided by

$$\begin{aligned} E(X^k) &= r\,\frac{\phi^k\,\Gamma\left(1+\frac{k}{\tau}\right)-\Gamma\left(1+\frac{k}{\tau};\left(\frac{x_m}{\phi}\right)^{\tau}\right)}{\left(1-\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)\right)} \\ &+ (1-r)\,\frac{1}{1-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}\,\gamma\,x_0^k\,Be\left(\left(\frac{x_m}{x_0}\right)^{-\delta};1-\frac{k}{\delta},\gamma\right), \end{aligned}$$

where $\Gamma(\cdot)$ and $\Gamma(\cdot;\cdot)$ are complete and incomplete gamma functions defined by

$$\Gamma(a)=\int_0^{\infty} z^{a-1}\,e^{-z}\,dz \quad \text{and}\ \Gamma(a;x)=\int_0^{x} z^{a-1}\,e^{-z}\,dz \quad \text{with} \quad a,x>0,$$

respectively and $Be(\cdot;\cdot,\cdot)$ is the incomplete beta function.

The procedure for generating random variates from the Weibull–Stoppa distribution is also presented. Similar to the previous section, the inverse transformation method of simulation can be used as the cdf of the Weibull and Stoppa distributions can be inverted. Then if u is a value generated from the uniform distribution $U(0,1)$, a value generated from (23) is obtained as follows

- If $u \leq r$ then solve

$$x = -\phi\left\{\ln\left[1-\frac{u}{r}\left(1-\exp\left[-\left(\frac{\theta}{\phi}\right)^{\tau}\right]\right)\right]\right\}^{1/\tau}$$

- If $u > r$ then

$$x = x_0\left\{1-\left(\frac{u-r}{1-r}\left[1-\left(1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right)^{\gamma}\right]+\left(1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right)^{\gamma}\right)^{1/\gamma}\right\}^{-1/\delta}$$

5 Numerical applications

In this section, the versatility of the composite Lognormal–Stoppa and Weibull–Stoppa models, as compared with the composite Pareto and composite Lomax families, is tested by analyzing a classic insurance data set. It deals with the set of Danish data on 2,492 fire insurance losses in millions of Danish kroner (DKr) from the years 1980 to 1990 inclusively, 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/`.

Parameter estimation for all the models considered in this paper has been completed by the method of maximum likelihood (ML)(which is implemented using the function "mle"/"mle2" in R). The ML estimates for the different composite models, together with their corresponding standard errors, are reported in Table 1.

5.1 Model assessment

Model assessment is presented from two points of view, theoretical plausibility and practical consideration. For the former point, the theoretical plausibility is justified by means of Kullback-Leibler divergence, suggesting an information-criterion based approach. The following four information criterions are used:

1. Negative log–likelihood (NLL): Calculated by taking the negative of the value of the log–likelihood evaluated at the ML estimates.

2. Akaike information criterion (AIC): Calculated by twice the NLL, evaluated at the ML estimates, plus twice the number of estimated parameters

3. Bayesian information criterion (BIC): Obtained as twice the NLL, evaluated at the ML estimates, plus $k\ln(n)$, where k is the number of estimated parameters and n is the sample size)

4. Consistent Akaike Information Criteria (CAIC): A corrected version of the AIC, proposed by Bozdogan (1987) to overcome the tendency of the AIC overestimating the complexity of the underlying model as it lacks the asymptotic property of consistency. In order to calculate the CAIC, a correction factor based on the sample size is used to compensate for the overestimating nature of AIC. The CAIC is defined as twice the NLL plus $k\,(1+\ln(n))$, again k is the number of free parameters and n refers to the sample size.

Note that for all the information criterion above, smaller values indicate a better fit of the model to the data. The results are shown in Table 1. It can be seen that within each of the Lognormal-composite family and the Weibull-composite family, the model associated with the Stoppa distribution outperforms the ones assocciated with the Pareto or Lomax distributions in all of the goodness-of-fit measures mentioned before. Overall, the Weibull-Stoppa composite model provides the best fit, again consistently across the different measures, to the data. Illustration of the fit of all the composite models is given in Figure 2 and Figure 3.

For the latter point, applications of the composite model, especially in the context of actuarial studies, involve mostly calculations done using the distribution function of the fitted model, for instance, the expected loss above a threshold and the value-at-risk. Hence, it is natural to express the fit of the model to the data in terms of distribution functions. In particular, it is suggested to use the following three empirical distribution function (EDF)

Table 1: Estimated values of different composite models for Danish fire insurance loss data

Model	Parameter Estimates (S.E.)	NLL	AIC	BIC	CAIC
Lognormal Stoppa	$\mu = 0.0908\,(0.0221)$ $x_0 = 0.9574\,(0.0528)$ $\delta = 1.4543\,(0.0507)$ $\gamma = 1.2704\,(0.1229)$	3858.74	7717.48	7748.76	7752.76
Lognormal Lomax	$\mu = 0.1035\,(0.0196)$ $\sigma = 0.1823\,(0.0112)$ $\lambda = 0.3648\,(0.1234)$ $\theta = 1.1444\,(0.0289)$	3860.47	7728.94	7752.22	7756.22
Lognormal Pareto	$\mu = 0.1372\,(0.0185)$ $\sigma = 0.1966\,(0.0116)$ $\theta = 1.2075\,(0.0297)$	3865.86	7737.72	7755.18	7758.18
Weibull Stoppa	$\tau = 16.1717\,(1.2911)$ $x_0 = 0.7416\,(0.0683)$ $\delta = 1.4952\,(0.0588)$ $\gamma = 1.7307\,(0.3282)$	3818.82	7645.64	7668.92	7672.92
Weibull Lomax	$\tau = 15.3455\,(0.6711)$ $\phi = 0.9690\,(0.0068)$ $\lambda = 0.5613\,(0.1276)$ $\theta = 0.9717\,(0.0069)$	3823.70	7655.40	7676.68	7682.68
Weibull Pareto	$\tau = 14.0483\,(0.5015)$ $\phi = 0.9966\,(0.0075)$ $\theta = 1.0027\,(0.0078)$	3840.38	7686.76	7704.22	7707.22

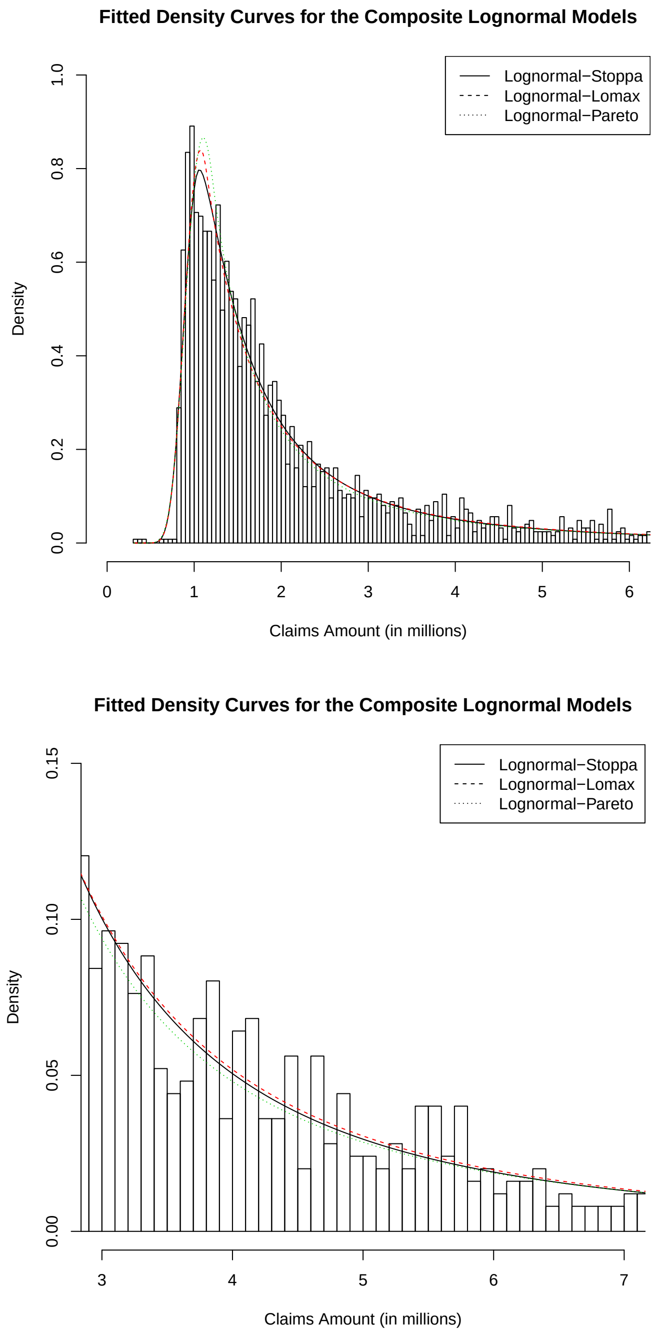
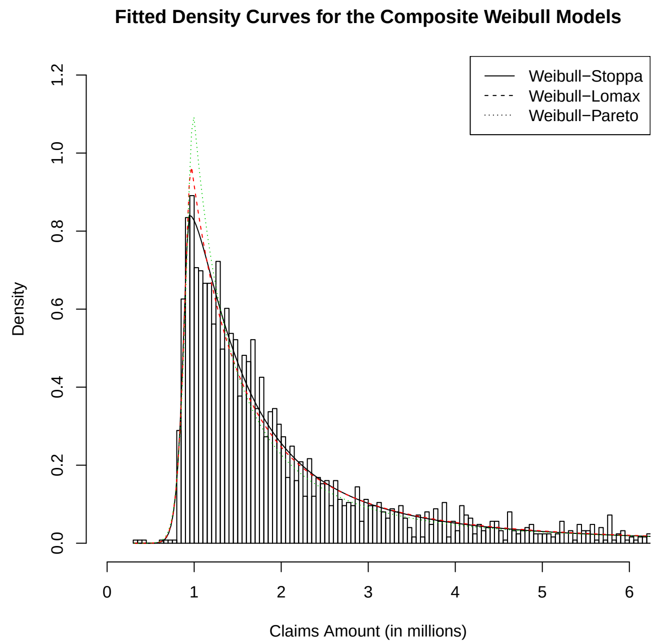


Figure 2: Comparison of empirical histogram for the Danish fire insurance data, Lognormal–Stoppa (solid), Lognormal–Lomax (dashed) and Lognormal–Pareto (dotted).



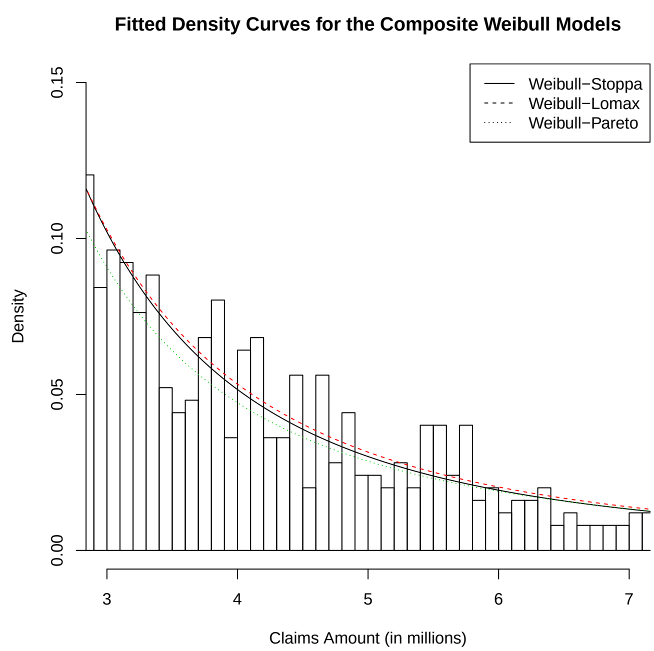


Figure 3: Comparison of empirical histogram for the Danish fire insurance data, Weibull–Stoppa (solid), Weibull–Lomax (dashed) and Weibull–Pareto (dotted).

goodness-of-fit measures to quantify the 'distance' between the empirical distribution function constructed from the data and the cumulative distribution function of the fitted models, namely the Kolmogorov-Smirnov test statistics, the Cramer-von Mises test statistics and the Anderson-Darling test statistics (see Rizzo (2009)). The definition of the test statistics are given as follows: Denote the cumulative distribution function of the fitted model by $\hat{F}$, the original data by $x_1, ..., x_N$ and the ordered data in increasing magnitude by $x_{(1)}, ..., x_{(N)}$, then we have:

1. Kolmogorov-Smirnov (KS) test statistics: $D = max(D^+, D^-)$, where

$$D^+ = \max_{1\leq j\leq N}\left\{\frac{j}{N} - \hat{F}(x_{(j)}))\right\}, D^- = \max_{1\leq j\leq N}\left\{\hat{F}(x_{(j)}) - \frac{j-1}{N}\right\}$$

2. Cramer-von Mises (CvM) test statistics:

$$W^2 = \sum_{j=1}^{N}\left[\hat{F}(x_{(j)}) - \frac{2j-1}{2N}\right]^2 + \frac{1}{12N}$$

3. Anderson-Darling (AD) test statistics:

$$A^2 = -N - \frac{1}{N}\sum_{j=1}^{N}[(2j-1)\log(\hat{F}(x_{(j)})) + (2n+1-2j)\log(1-\hat{F}(x_{(j)}))]$$

For all the EDF goodness-of-fit measures above, smaller values indicate a better fit of the model to the data. The results are summarized in Table 2. It can be seen that using the CvM and AD measure, within each of the Lognormal-composite family and the Weibull-composite family, the composite model associated with the Stoppa distribution gives the best fit. In fact, unlike the case in the information-criterion section, the Lognormal-Stoppa and Weibull-Stoppa models are the best two models considering all the models at once. In the KS case, it is observed that while the Weibull-Stoppa model is still the best overall, the Lognormal-Lomax model outperforms the Lognormal-Stoppa model by a slight margin. The reason is believed to be that Kolmogorov-Smirnov test is relatively insensitive to deviations occuring in the tail (Mason and Schuenemeyer, 1983). And since the improvement brought by using the Stoppa distribution in place of the Lomax distribution

Table 2: EDF goodness-of-fit measures of the composite models

	test statistics		
Model	KS	CvM	AD
Lognormal–Pareto	0.032304	0.47814	3.15964
Lognormal–Lomax	0.019515	0.21406	1.95087
Lognormal–Stoppa	0.019739	0.14493	1.70092
Weibull–Pareto	0.051729	1.51904	7.33822
Weibull–Lomax	0.025506	0.33780	1.90971
Weibull–Stoppa	0.017340	0.12615	0.88225

only occurs in the tail, the KS test statistics may not have fully reflected the improvement, resulting in the slightly superior fit of the Lognormal-Lomax model over the Lognormal-Stoppa model.

Note that the test statistics not only provide a way to measure the fit in terms of distribution functions, but also allow us to perform hypothesis testing for model validation purposes. We remark that to perform the goodness-of-fit tests, it is required that the proposed model is specified completely, i.e. parameters need to be specified too. In the case where parameters are estimated from data, the critical values produced using the standard procedure are no longer valid (Babu and Rao, 2004). To circumvent this problem, we use the bootstrap method. The validity is justified by the work of Babu and Rao(2004), in which the consistency of the bootstrap method estimating the null distribution of the goodness-of-fit test statistics was shown. We present a brief outline of the bootstrap procedure. Denote the data by $x_1, ..., x_N$. For each proposed composite model, fit the model to the data. Then,

1. Compute the goodness-of-fit test statistics, t_{KS}, t_{CvM}, t_{AD}
2. Use the fitted model to perform parametric bootstrapping
   - Generate M sets of resampled data, denote as $\hat{x}_1^{(i)}, ..., \hat{x}_N^{(i)}, i = 1, ..., M$
   - For each set of the resampled data, fit the composite model and compute the test statistics, $t_{KS}^{(i)}, t_{CvM}^{(i)}, t_{AS}^{(i)}, i = 1, ..., M$

3. The p-value of the respective original test statistics are given by

$$\frac{\#\{i : t^{(i)}_{KS} \geq t_{KS}\}}{M}, \frac{\#\{i : t^{(i)}_{CvM} \geq t_{CvM}\}}{M}, \frac{\#\{i : t^{(i)}_{AD} \geq t_{AD}\}}{M}$$

The p-value of the test statistics, computed using $M = 10000$ simulations, are presented in Table 3. We remark that while an extremely small p-value may lead to a confident rejection of the null hypothesis that the data comes from the proposed model and in general a larger p-value is favourable, a p-value being large doesn't serve well as evidence of the model being correct especially when there are other models with a p-value of comparable magnitude. It can be seen that none of the models are rejected, validating that the models are statistically legitimate candidates. In addition, the composite-Stoppa models, together with the composite-Lomax models in this case, have relatively high(hence favourable) p-values across different EDF goodness-of-fit measures.

Table 3: p-values of the EDF goodness-of-fit measures of the composite models, computed with 10000 sets of bootstrap resamples

	p-value of the test statistics		
Model	KS	CvM	AD
Lognormal–Pareto	0.5314	0.5527	0.6968
Lognormal–Lomax	0.9741	0.9219	0.9410
Lognormal–Stoppa	0.9360	0.9530	0.9440
Weibull–Pareto	0.5071	0.5086	0.5324
Weibull–Lomax	0.7949	0.7760	0.8364
Weibull–Stoppa	0.6955	0.7104	0.7571

5.2 Applications and performance

In this section, investigation on two practical concerns, namely the high quantiles and the probable maximal loss, is presented. As both the quantities are related to the distribution function of the model, the performance in these terms is expected to be in line with the results shown in the EDF goodness-of-fit section.

5.2.1 Estimation of high quantiles

Frequently, it is convenient for practitioners to obtain reliable information about the tail of the claim size distribution. A measure that yields an acceptable knowledge of the right tail of the model is the estimates of the high quantiles. Empirical and fitted quantiles in the extreme portion of the tail for composite Lognormal and composite Weibull are given in Table 4 and Table 5 respectively. The empirical quantiles have been computed using the Type 8 quantile algorithm suggested by Hyndman and Fan (1996). It is our interest to analyze how much theoretical tail quantiles of each fitted composite model deviate from the empirical quantiles in the extreme portion of the tail. It can be seen that the composite Stoppa models demonstrate a better fit to the data in the high quantiles, likewise suggesting it being a favourable model for the given data. The Lognormal–Pareto and Weibull–Pareto distributions tend to overestimate the exteme tail quantiles whereas Lognormal–Lomax and Weibull–Lomax composite models underestimate them. We remark that interpretation of the results obtained from this table needs to be prudently made given that the extreme-value data is scarce. To illustrate the point, note that the sample size of the dataset is only 2492 while the value of the 99.99% empirical quantile represents an event that occur 1 in 10000 times.

Table 4: Empirical and fitted composite Lognormal model quantiles.

		Composite Lognormal		
Quantiles	Empirical	Pareto	Lomax	Stoppa
0.50	1.634	1.572	1.611	1.590
0.90	5.086	5.282	5.164	5.028
0.95	8.459	8.902	8.249	8.134
0.99	24.870	29.903	23.750	24.683
0.999	146.010	169.227	104.835	120.322
0.9995	199.020	285.259	163.540	193.801
0.9999	263.250	958.261	458.572	586.128

Table 5: Empirical and fitted composite Weibull model quantiles.

		Composite Weibull		
Quantiles	Empirical	Pareto	Lomax	Stoppa
0.50	1.634	1.542	1.615	1.632
0.90	5.086	5.522	5.201	5.129
0.95	8.459	9.566	8.203	8.211
0.99	24.870	34.262	22.648	24.223
0.999	146.010	212.586	92.931	113.126
0.9995	199.020	368.271	141.649	179.885
0.9999	263.250	1319.032	376.050	527.741

5.2.2 Probable maximum loss

We conclude this section by presenting the probable maximum loss (PML) for the family of composite models described in this paper. In general terms, the PML can be defined as the worst loss likely to happen (Cebrián et al (2003)). To be specific, given a random variable N that follows a Poisson distribution with mean ζ and $X_1, \ldots, X_N$ a sequence of independent and identically distributed random variables, define a sequence of maxima $M_N = \max(X_1, \ldots, X_N)$ with cdf F_{M_N}. Then, the PML is the quantile function of the maximum loss M_N and is given by $PML = F_{M_N}^{-1}(q)$. For actual implementation, the value of ζ is needed, we follow Pigeon and Denuit(2011) and use the average annual frequency as an estimate, which is given by $\hat{\zeta} = 226.5455$. Results of PML for values of $q = 0.90$, $q = 0.95$ and $q = 0.99$ are shown in Table 6.

6 Conclusions

In this paper, a new composite model for modelling large claim data has been proposed and its performance is compared with the existing composite models using a classic insurance dataset. The composite model is developed using a mode–matching procedure. The Lognomral or Weibull distribution is used up to its mode as the first component of the spliced model and thereafter, the Stoppa distribution is used as the second component. This model

Table 6: Probable maximal losses for different values of q.

	Probable Maximal Loss q		
Model	0.90	0.95	0.99
Lognormal–Pareto	301.20	517.86	1766.67
Lognormal–Lomax	171.10	271.32	770.04
Lognormal–Stoppa	229.74	376.89	1156.00
Weibull–Pareto	390.10	690.25	2513.00
Weibull–Lomax	147.65	228.48	613.17
Weibull–Stoppa	195.11	315.78	939.37

incorporates an unrestricted mixing weight. The Goodness–of–fit has been analyzed from two perspectives, on one hand considering an information–criterion based approach, motivated by theoretical plausibility, by using four different information criterions and on the other hand a distribution approach, motivated by practical considerations, by quantifying the distance between the empirical and theoretical cumulative distribution function. From the numerical results obtained, the composite Stoppa model outperform the existing composite models under all seven measures of fit. This together with its simple implementation make the composite-Stoppa an appealing tool to model large claims.

References

Babu, G.J. and Rao, C.R. (2004). Goodness–of–fit tests when parameters are estimated *Sankhya: The Indian Journal of Statistics*, 66, 1, 63–74.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, **52**, 3, 345–370.

Burkhauser, R.V., Feng, S. and Larrimore, J. (2010). Improving imputations of top incomes in the public-use current population survey by using both cell-means and variances. *Economics Letters*, 108, 1, 69–72.

Cebrián, A., Denuit, M. and Lambert, Ph. (2003). Generalized Pareto fit to the society of actuaries' large claims data base. *North American Actuarial Journal*, 7, 3, 18–36.

Ciumara, R. (2006). An actuarial model based on the composite Weibull–Pareto distribution. *Mathematical Report–Bucharest*, 8(4), 401–414.

Cooray, K. and Ananda, M.M.A. (2005). Modeling actuarial data with a compostie Lognormal–Pareto model. *Scandinavian Actuarial Journal*, 5, 321–334.

Hyndman, R.J. and Fan, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, 50, 361–365.

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken, NJ: John Wiley & Sons.

Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2008). *Loss Models: From Data to Decisions*. Third Edition. Wiley.

Mason, D. M. and Schuenemeyer, J. H. (1983). A modified Kolmogorov-Smirnov test sensitive to tail alternatives. *Ann. Statist.*, 11, 933–946.

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

Pigeon, M and Denuit, M. (2011). Composite Lognormal–Pareto model with random threshold. *Scandinavian Actuarial Journal*, 3, 177–192.

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

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

Scollnik, D. P. M. and Sun, C. (2012). Modeling with Weibull–Pareto models. *North American Actuarial Journal*, 16, 2, 260–272.

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