# 3 ON COMPOSITE MODELS: WEIBULL-PARETO AND LOGNORMAL-PARETO. - A COMPARATIVE STUDY -

**Vasile PREDA***
**Roxana CIUMARA****

## Abstract

In this paper we make a comparison between two composite models: lognormal-Pareto and Weibull-Pareto. The first one was introduced by Cooray and Ananda in 2005. The second composite distribution was constructed in the same manner as lognormal-Pareto. Here, we prove that these models behave similarly and they could be used in insurance bussiness for modelling actuarial data, especially in the cases where one deals with large loss payments.

**Key words**: composite models; lognormal, Weibull and Pareto distributions; maximum likelihood estimation; smooth empirical estimation of percentils.

**JEL Subject Classification**: C16

## 1. Introduction

In this paper we make a comparison between two composite models. The first one - the composite lognormal-Pareto model - was introduced by Cooray and Ananda (2005). The second one, the composite Weibull-Pareto model was constructed in Ciumara (2006), in the same manner as in Cooray and Ananda (2005). These models could be used in insurance for modeling actuarial data, especially when there are some larger loss payments.

These new distributions, the composite lognormal-Pareto and Weibull-Pareto are combinations between lognormal or Weibull distributions, respectivelly - which cover the behaviour of small losses well - up to a threshold parameter and Pareto for the rest of the domain.

The resulting densities of the composite lognormal-Pareto and Weibull-Pareto have larger tails than the lognormal or Weibull densities, but smaller tails than the Pareto density. Moreover, the shape of these composite densities are similar.

---

* *Professor, Faculty of Mathematics and Informatics, Bucharest University.*
** *Lecturer, Department of Mathematics, Academy of Economic Studies, Bucharest.*

In Section 2 we describe the composite lognormal-Pareto model and the composite Weibull-Pareto model through densities, cumulative distribution functions and the r-th initial moments. Afterwards, the construction of the likelihood function in both cases is presented.

Taking into account the results of Section 2, two algorithms for estimating the parameters of the composite models are described in Sections 3 and 4. The first algorithm is based on maximum likelihood estimation and the second one on smooth empirical estimation of percentiles (Klugmann *et al*.,1998). These algorithms are given in Cooray and Ananda (2005) for the composite lognormal-Pareto model and in Ciumara (2006) for the composite Weibull-Pareto model. Finally, numerical examples based on simulated data sets are presented in Section 5.

## 2. Construction and Characteristics of the Two Composite Models

Following Cooray and Ananda (2005), a composite density is derived from

$$f(x) = \begin{cases} c \cdot f_1(x) & \text{, if } x \in (0,\theta] \\ c \cdot f_2(x) & \text{, if } x \geq \theta \end{cases},$$

imposing the conditions necessary to obtain a continuous and differentiable density, that is $\int_0^\infty f(x)dx = 1$, $f_1(\theta) = f_2(\theta)$ and $f_1^{'}(\theta) = f_2^{'}(\theta)$.

For the lognormal-Pareto model, $f_1(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right)$, where $x > 0$, $\sigma > 0$, $\mu \in \mathbf{R}$ and $f_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}$, $x > \theta$, $\theta > 0$ and $\alpha > 0$. Cooray and Ananda (2005) derived the density of the composite lognormal-Pareto model

$$f_{LP}^c(x) = \begin{cases} \frac{\alpha\theta^\alpha}{(1+\Phi(k))x^{\alpha+1}} \exp\left(-\frac{\alpha^2}{2k^2}\ln^2\left(\frac{x}{\theta}\right)\right) & \text{, if } x \in (0,\theta] \\ \frac{\alpha\theta^\alpha}{(1+\Phi(k))x^{\alpha+1}} & \text{, if } x \geq \theta \end{cases} \quad (2.1)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and $k \approx 0.372238898$. This value of $k$ is obtained imposing the conditions that $f_{LP}^c$ is a continuous differentiable density function on $(0,\infty]$, that is $\int_0^\infty$ …. , $f_1(\theta) = f_2(\theta)$ and $f'_1(\theta) = f'_2(\theta)$.

Initially, $f_{LP}^{c}(x)$ had four parameters, but imposing the conditions specified in the beginning of this section, they reduced to two, $\alpha$ and $\theta$. The other two parameters could be obtained from the previous ones, as $\sigma = \frac{k}{\alpha}$ and $\mu = \ln\theta - \alpha\sigma^2$. Moreover, the constant $c$ that appears in the density expression is, in this case, $c = \frac{1}{1+\Phi(k)}$.

For $\alpha = 0.5$ and $\theta = 50$, the densities of the lognormal (dotted line), Pareto (dashed line) and composite lognormal-Pareto (solid line) are displayed in Figure 2.1. We have to note here that the composite lognormal-Pareto density results from $f_1$ restricted to $(0,\theta]$ and $f_2$ on $(\theta, \infty)$ (in this case $\theta = 50$). Because of continuity and differentiability conditions stated earlier, the graph of this composite density is continuous and therefore $f_1$ restricted to $(0,\theta]$ and $f_2$ on $(\theta, \infty)$ are displayed simultaneously in the solid line. The same remark applies for the other figures (2.3-2.6).

In Figure 2.2 we set $\alpha = 1$ and $\theta = 50$ for the same three densities and in Figure 2.3 $\alpha = 1.5$ and $\theta = 50$.

**Figure 2.1**

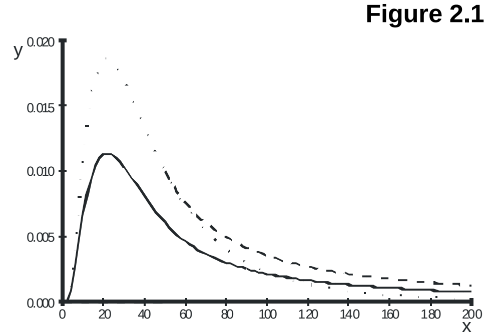



**Figure 2.2**

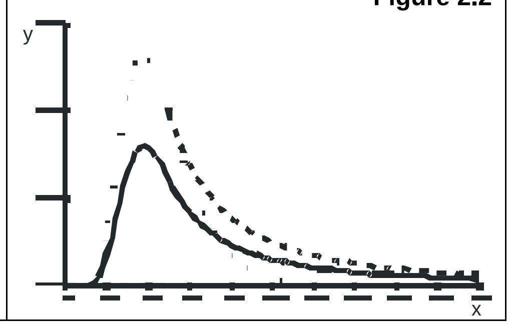
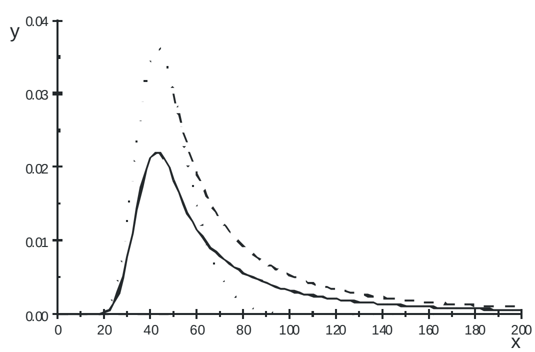



**Figure 2.3**



It is obvious from these graphs that the composite lognormal-Pareto fade away to zero more quickly than Pareto (it has a smaller tail than Pareto), but, in the same time, it has a larger tail than the lognormal. As $\alpha$ increases, the three densities approaches zero faster.

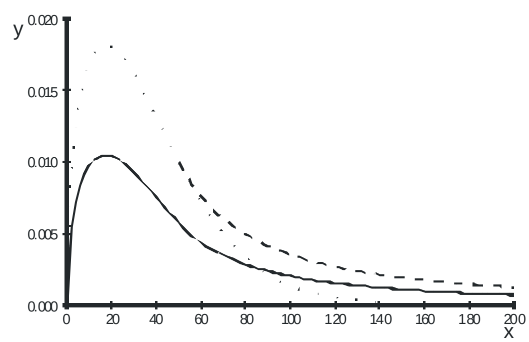
We discuss now the composite Weibull-Pareto model, where $f_1(x)=\frac{\beta}{\gamma^{\beta}}x^{\beta-1}\exp\left(-\left(\frac{x}{\gamma}\right)^{\beta}\right)$ $x>0$, $\gamma>0$, $\beta>1$ and $f_2(x)=\frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}$, $x>\theta$, $\theta>0$ and $\alpha>0$. From Ciumara (2006), the density of this composite distribution is

$$f_{WP}^{c}(x)=\begin{cases}\frac{(t_0+1)^2}{(t_0+2)}\frac{\beta}{x}\left(\frac{x}{\theta}\right)^{\beta}\exp\left(-(t_0+1)\left(\frac{x}{\theta}\right)^{\beta}\right), & \text{if } x\in(0,\theta] \\ \frac{t_0(t_0+1)}{t_0+2}\frac{\beta}{x}\left(\frac{\theta}{x}\right)^{\beta t_0}, & \text{if } x\geq\theta\end{cases} \qquad (2.2)$$

where $t_0 \approx 0.3499764854$. Again, this value of $t_0$ is obtained imposing the conditions that $f^c_{WP}$ is a continuous differentiable density function on $(0, \infty]$.

As in the case of the composite lognormal-Pareto model, we have again the number of parameters reduced from four to two, $\beta$ and $\theta$. For the other two, the relations are $\alpha=\beta t_0$ and $\gamma=\theta(t_0+1)^{-\frac{1}{\beta}}$. Again, the constant from the density expression from the beginning is now $c=\frac{t_0+1}{t_0+2}$.

**Figure 2.4**




In the following three graphs we compare the densities of the composite Weibull-Pareto (solid line), Weibull (dotted line) and Pareto (dashed line) for different values of $\beta=\frac{\alpha}{t_0}$

and $\theta$: $\alpha = 0.5$ and $\theta = 50$ in Figure 2.4, $\alpha = 1$ and $\theta = 50$ in Figure 2.5 and $\alpha = 1.5$ and $\theta = 50$ in Figure 2.6.

**Figure 2.5**

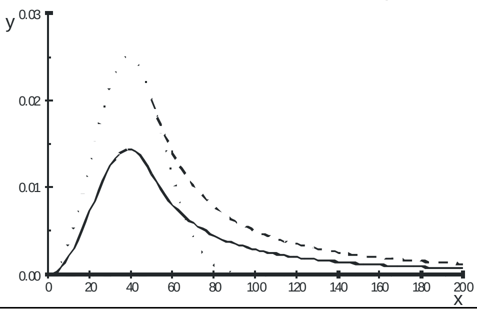
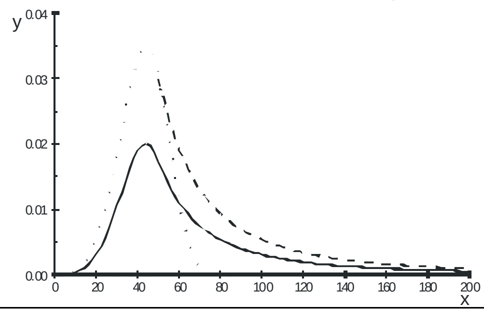



**Figure 2.6**

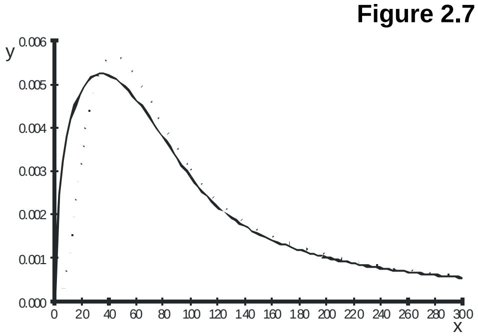



It might be noticed again that the composite Weibull-Pareto density has a smaller tail than Pareto, but longer than Weibull. This observation is useful when one has to choose a proper model for loss payments, especially when some few losses are a lot larger than the others.

We turn now to the comparison between the two composite models. Therefore, in the following graphs the composite Weibull-Pareto density (solid line) and the composite lognormal-Pareto density (dotted line), respectively, are illustrated for different $\theta$ and $\beta = \dfrac{\alpha}{t_0}$ (or, equivalently, $\alpha$ ).

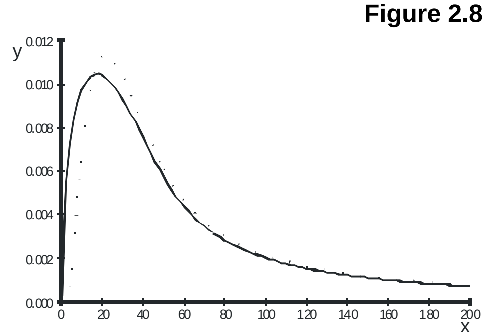
From Figures 2.7-2.9 one may see that if $\alpha < 1$ then, eventhough the shapes of the two composite densities are similar, the mode of the composite Weibull-Pareto is less than the mode of the composite lognormal-Pareto.

**Figure 2.7**




$\alpha = 0.5$, $\theta = 100$

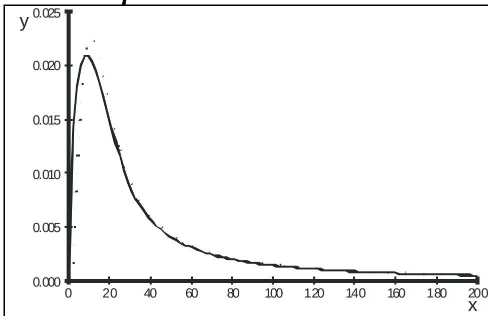
**Figure 2.8**



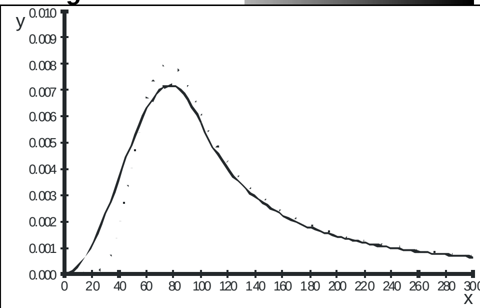

$\alpha = 0.5$, $\theta = 50$

**Figure 2.9**

**Figure 2.10**



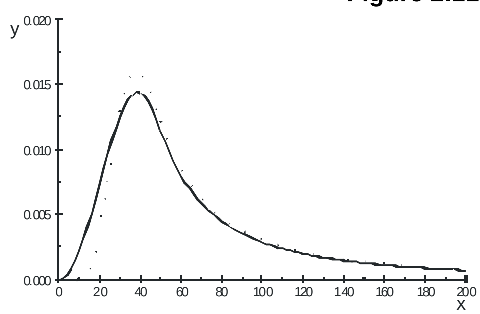

$\alpha = 0.5$, $\theta = 25$




$\alpha = 1$, $\theta = 100$

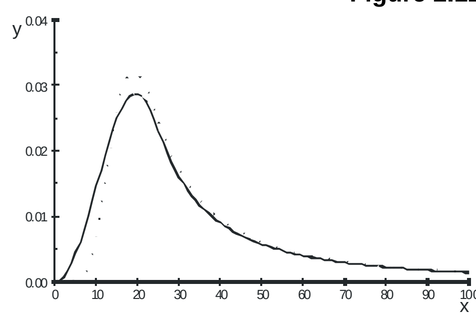
As $\alpha$ increases, the shapes of the two composite densities are even more similar, as Figures 2.11-2.14 reveal.

**Figure 2.11**



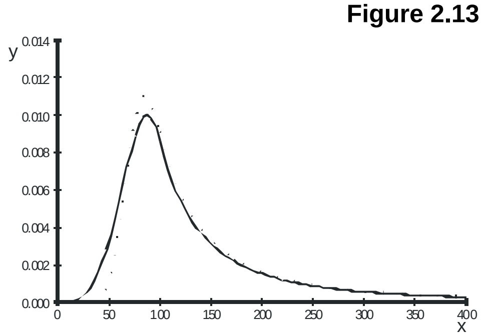

$\alpha = 1$, $\theta = 50$

**Figure 2.12**



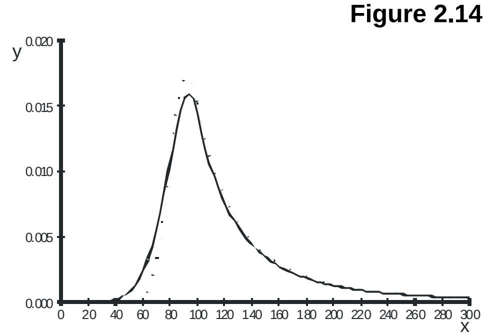

$\alpha = 1$, $\theta = 25$

**Figure 2.13**




$\alpha = 1.5$, $\theta = 100$

**Figure 2.14**

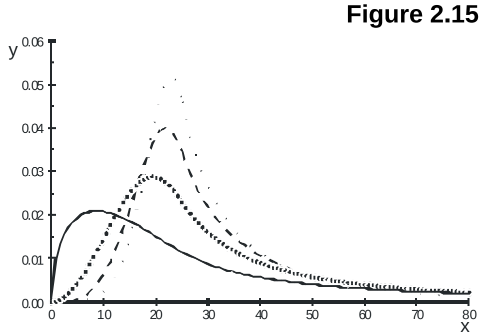



$\alpha = 2.5$, $\theta = 100$

The next two graphs show how the shape of densities changes – in the same manner – when keeping $\theta = 25$ and $\alpha = 0.5$ (solid line), $\alpha = 1$ (dotdashed line), $\alpha = 1.5$ (dashed line), $\alpha = 2$ (dotted line) for the composite Weibull-Pareto (Figure 2.15) and composite lognormal-Pareto (Figure 2.16), respectively.

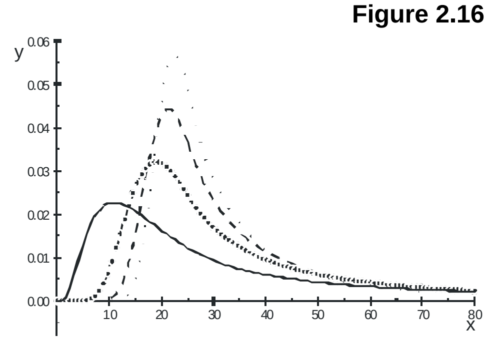
Finally, we keep $\alpha = 1$ and let $\theta = 100$ (solid line), $\theta = 75$ (dotdashed line), $\theta = 50$ (dashed line), $\theta = 25$ (dotted line) in the cases of the composite Weibull-Pareto (Figure 2.17) and the composite lognormal-Pareto (Figure 2.18). The variation of the densities' shape is similar for these two last cases as well.

**Figure 2.15**



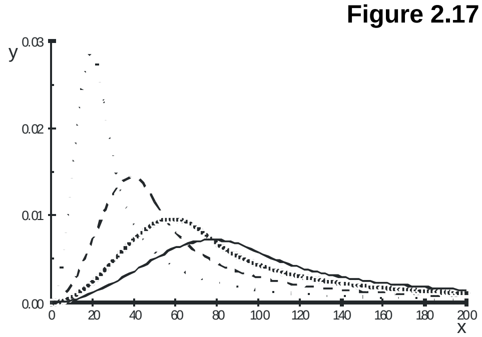

Weibull-Pareto

**Figure 2.16**



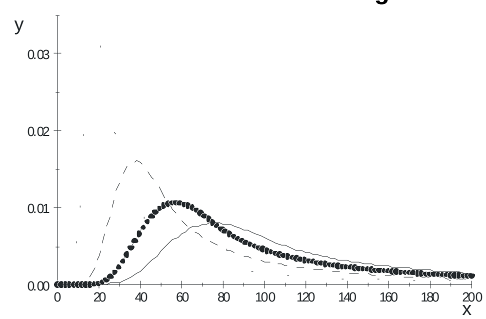

Lognormal-Pareto

**Figure 2.17**




Weibull-Pareto

**Figure 2.18**




Lognormal-Pareto

In what concerns other characteristics of these two composite models, the cumulative distribution function of the composite Weibull-Pareto model is given in Ciumara (2006) by

$$F_{WP}^{c}(x) = \begin{cases} \dfrac{t_0+1}{t_0+2}\left[1-\exp\left(-(t_0+1)\left(\dfrac{x}{\theta}\right)^{\beta}\right)\right], & \text{if } x \in (0,\theta] \\ 1-\dfrac{t_0+1}{t_0+2}\left(\dfrac{\theta}{x}\right)^{\beta t_0}, & \text{if } x \geq \theta \end{cases} \tag{2.3}$$

and

$$F_{LP}^{c}(x)=\begin{cases}\frac{1}{1+\Phi(k)}\Phi\left(\frac{\alpha}{k}\ln\left(\frac{x}{\theta}\right)+k\right), & \text{if } x\in(0,\theta]\\ 1-\frac{1}{1+\Phi(k)}\left(\frac{\theta}{x}\right)^{\alpha}, & \text{if } x\geq\theta\end{cases} \quad (2.4)$$

is the cumulative distribution function for the composite lognormal-Pareto, given in Cooray and Ananda (2005).

The composite Weibull-Pareto and the lognormal-Pareto are unimodal, with the mode for the first distribution

$$x_{\text{mode}}^{cWP}=\theta\left(\frac{\beta-1}{\beta(t_0+1)}\right)^{\frac{1}{\beta}} \quad (2.5)$$

given in Ciumara (2006) and the mode for the second one,

$$x_{\text{mode}}^{cLP}=\exp\left(\ln\theta-\frac{\alpha+1}{\alpha^2}k^2\right). \quad (2.6)$$

Another characteristic of the models we discuss in this article concerns the initial moments. Thus, the initial $r$-th moment of the composite Weibull-Pareto distribution is given in Ciumara (2006) by

$$E(X^r)=\frac{t_0+1}{t_0+2}\theta^r\left[(t_0+1)^{-\frac{r}{\beta}}\cdot\Gamma\left(\frac{r}{\beta}+1;t_0+1\right)+\frac{\beta t_0}{\beta t_0-r}\right] \quad (2.7)$$

for $r<\beta t_0$, where $\Gamma(s;z)=\int_0^z y^{s-1}e^{-y}dy$ is the incomplete Gamma function.

Similarly, the $r$-th initial moment of the composite lognormal-Pareto distribution is given in Cooray and Ananda (2005) by

$$E(X^r)=\frac{1}{1+\Phi(k)}\theta^r\left[\Phi\left(k-\frac{kr}{\alpha}\right)\cdot\exp\left(\frac{1}{2}\left(\frac{k}{\alpha}\right)^2\left(r^2-2\alpha r\right)\right)+\frac{\alpha}{\alpha-r}\right] \quad (2.8)$$

for $r<\alpha$.

The likelihood function of the composite Weibull-Pareto model is given in Ciumara (2006). Moreover, it is proved that in certain conditions the maximum likelihood estimates of the two parameters of this model exists and they can be computed as the solution of some equations. In what it follows we state the result mentioned previously.

Let $x_1, x_2, \dots, x_n$ be a random sample from the two-parameter composite Weibull-Pareto model. Assuming that $x_1\leq x_2\leq\ldots\leq x_n$ and $x_m\leq\theta\leq x_{m+1}$, the likelihood function is given by

$$L(x_1, x_2, ..., x_n; \beta, \theta) = C^{WP} \cdot \beta^n \cdot \theta^{\beta((n-m)t_0 - m)} \frac{\prod_{i=1}^{m} x_i^{\beta}}{\prod_{i=m+1}^{n} x_i^{\beta t_0}} \cdot \exp\left(-\frac{t_0+1}{\theta^{\beta}} \sum_{i=1}^{m} x_i^{\beta}\right) \qquad (2.9)$$

where $C^{WP} = \left(\frac{t_0+1}{t_0+2}\right)^n (t_0+1)^m t_0^{n-m} \prod_{i=1}^{n} x_i^{-1}$.

Moreover, if

$$m > n\frac{t_0}{t_0+1} \qquad (2.10)$$

and

$$n - t_0 \sum_{i=m+1}^{n} \ln x_i + \sum_{i=1}^{m} \ln x_i - (m - t_0(n-m)) \frac{\sum_{i=1}^{m} x_i \ln x_i}{\sum_{i=1}^{m} x_i} > 0 \qquad (2.11)$$

then the maximum likelihood estimate of $\beta$, denoted by $\hat{\beta}_{\mathrm{ML}}$, is the (unique) solution of the equation

$$\frac{n}{\beta} - t_0 \sum_{i=m+1}^{n} \ln x_i + \sum_{i=1}^{m} \ln x_i - (m - t_0(n-m)) \frac{\sum_{i=1}^{m} x_i^{\beta} \ln x_i}{\sum_{i=1}^{m} x_i^{\beta}} = 0 \qquad (2.12)$$

and the maximum likelihood estimate of $\theta$ is

$$\hat{\theta}_{\mathrm{ML}} = \left(\frac{t_0+1}{m - t_0(n-m)} \sum_{i=1}^{m} x_i^{\hat{\beta}_{\mathrm{ML}}}\right)^{\frac{1}{\hat{\beta}_{\mathrm{ML}}}} \qquad (2.13)$$

Some remarks could be made in order to clarify the above result. Condition $(2.10)$ is equivalent to $m > 0.26n$. In other words, if at least one quarter of the data set consists of "small" losses one may think that the model could be Weibull-Pareto. Otherwise, the Pareto model seems more appropriate. Of course, if $m$ approaches $n$, then a simple Weibull or lognormal model should be chosen.

Condition $(2.10)$ assures that the system of the likelihood equation has a unique solution, which is the maximum of likelihood function, and condition $(2.11)$ has to be imposed in order to obtain $\beta > 1$.

For the composite lognormal-Pareto distribution, assuming that $x_1 \le x_2 \le \ldots \le x_n$ and $x_m \le \theta \le x_{m+1}$, the likelihood function is given in Cooray and Ananda (2005)

$$L(x_1, x_2, \ldots, x_n; \alpha, \theta) = C^{LP} \alpha^n \theta^{\alpha n} \left( \prod_{i=1}^{n} x_i^{-\alpha} \right) \cdot \exp\left( -\frac{\alpha^2}{2k^2} \sum_{i=1}^{n} \ln^2 \frac{x_i}{\theta} \right) \qquad (2.14)$$

where $C^{LP} = \dfrac{1}{\left( \prod_{i=1}^{n} x_i \right) (1 + \Phi(k))^n}$.

## 3. Algorithms to Estimate the Composite Weibull-Pareto Parameters

The first algorithm is based on the maximum likelihood estimation of both parameters of the composite Weibull-Pareto distribution. It is described in Ciumara (2006) and we shall present it below.

For each $m \in \overline{\left[ n \frac{t_0}{t_0 + 1} \right] + 1, n}$, where $[\cdot]$ indicates the greatest integer function, if condition $(2.11)$ holds, we find $\hat{\beta}_m$ the solution of equation $(2.12)$ and then

$$\hat{\theta}_m = \left( \frac{t_0 + 1}{m - t_0 (n - m)} \sum_{i=1}^{m} x_i^{\hat{\beta}_m} \right)^{\frac{1}{\hat{\beta}_m}}.$$

If $x_m \le \hat{\theta}_m \le x_{m+1}$, then the maximum likelihood estimates of the two parameters are $\hat{\beta}_{\mathrm{ML}} = \hat{\beta}_m$, $\hat{\theta}_{\mathrm{ML}} = \hat{\theta}_m$ and the algorithm stops. Otherwise, we increase the value of $m$ with 1 and repeat the procedure.

If, for $m = n$, $x_n \le \hat{\theta}_n$, then, the maximum likelihood estimate of $\beta$, denoted $\hat{\beta}_{\mathrm{ML}}$, is the solution of the equation

$$\frac{n}{\beta} + \sum_{i=1}^{n} \ln x_i - n \frac{\sum_{i=1}^{n} x_i^{\beta} \ln x_i}{\sum_{i=1}^{n} x_i^{\beta}} = 0$$

and the maximum likelihood estimate of $\theta$ is

$$\hat{\theta}_{\mathrm{ML}} = \left( \frac{t_0 + 1}{n} \sum_{i=1}^{n} x_i^{\hat{\beta}_{\mathrm{ML}}} \right)^{\frac{1}{\hat{\beta}_{\mathrm{ML}}}}.$$

The second algorithm for estimating the parameters of the composite Weibull-Pareto distribution is based on the smooth empirical estimate of percentiles (Klugmann *et al.*, 1998).

We estimate $\theta$ by

$$\tilde{\theta} = (1-h)\cdot x_m + h \cdot x_{m+1},$$

where $m = [(n+1)p]$ or $m = \left[\left(n+\frac{1}{3}\right)p+\frac{1}{3}\right]$, $h = (n+1)p - m$ or $h = \left(n+\frac{1}{3}\right)p+\frac{1}{3} - m$, $p = F_{WP}^{c}(\theta)$ and $[\cdot]$ is the greatest integer function. In this case, $p = \frac{1}{t_0+2}$.

Then, the estimate of $\beta$ is the solution of the equation

$$\tilde{\theta}^{\beta} = \frac{t_0+1}{m - t_0(n-m)}\sum_{i=1}^{m} x_i^{\beta}.$$

Again, we have to impose here the conditions $(2.10)$ and $(2.11)$ in order to obtain a solution in $(1,\infty)$.

If the estimate of $\theta$ is closer to $x_1$ or $x_n$, the Pareto or the Weibull distributions, respectively, would be better models than the composite Weibull-Pareto. Moreover, if condition $(2.10)$ does not hold, then using a simple Pareto model is more appropriate than using the composite one.

## 4. Algorithms to Estimate the Composite Lognormal-Pareto Parameters

In their article, Cooray and Ananda (2006) proposed a simple and straightforward way to compute the maximum likelihood estimates for the two parameters of the composite lognormal-Pareto distribution.

For each $m \in \overline{1, n-1}$, calculate $\hat{\alpha}_m$ and $\hat{\theta}_m$ as follows. For $m = 1$,

$$\hat{\alpha}_m = n\left(\sum_{i=1}^{n} \ln\frac{x_i}{x_1}\right)^{-1}$$

and

$$\hat{\theta}_m = x_1 \prod_{i=1}^{n}\left(\frac{x_i}{x_1}\right)^{k^2}.$$

Otherwise,

$$\hat{\alpha}_m = \frac{k^2\left(n\sum_{i=1}^{m}\ln x_i - m\sum_{i=1}^{n}\ln x_i\right)}{2\left(m\sum_{i=1}^{m}\ln^2 x_i - \left(\sum_{i=1}^{m}\ln x_i\right)^2\right)} +$$

$$+ \frac{\sqrt{k^4\left(n\sum_{i=1}^{m}\ln x_i - m\sum_{i=1}^{n}\ln x_i\right)^2 + 4mnk^2\left(m\sum_{i=1}^{m}\ln^2 x_i - \left(\sum_{i=1}^{m}\ln x_i\right)^2\right)}}{2\left(m\sum_{i=1}^{m}\ln^2 x_i - \left(\sum_{i=1}^{m}\ln x_i\right)^2\right)}$$

and

$$\hat{\theta}_m = \left(\prod_{i=1}^{m} x_i\right)^{\frac{1}{m}} \exp\left(\frac{nk^2}{m\hat{\alpha}_m}\right).$$

If $x_m \le \hat{\theta}_m \le x_{m+1}$, then the maximum likelihood estimates of the two parameters are $\hat{\alpha}_{\mathrm{ML}} = \hat{\alpha}_m$, $\hat{\theta}_{\mathrm{ML}} = \hat{\theta}_m$ and the algorithm stops. Otherwise, we increase the value of $m$ with 1 and repeat the procedure.

If, for $m = n$, $x_n \le \hat{\theta}_n$, then, the maximum likelihood estimate of $\alpha$, denoted $\hat{\alpha}_{\mathrm{ML}}$, is

$$\hat{\alpha}_{\mathrm{ML}} = \frac{nk}{\sqrt{n\sum_{i=1}^{n}\ln^2 x_i - \left(\sum_{i=1}^{n}\ln x_i\right)^2}}$$

and

$$\hat{\theta}_{\mathrm{ML}} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \exp\left(\frac{k^2}{\hat{\alpha}_{\mathrm{ML}}}\right).$$

As in section 3, we describe below the second algorithm based on the smooth empirical estimates of percentiles (Cooray and Ananda, 2005). For $x_1 \le x_2 \le \ldots \le x_n$, we estimate $\theta$ by

$$\tilde{\theta} = (1-h)\cdot x_m + h\cdot x_{m+1},$$

where $m=[(n+1)p]$ or $m=\left[\left(n+\frac{1}{3}\right)p+\frac{1}{3}\right]$, $h=(n+1)p-m$ or $h=\left(n+\frac{1}{3}\right)p+\frac{1}{3}-m$, $p=F_{LP}^{c}(\theta)$ and $[\cdot]$ is the greatest integer function. In this case, $p=\frac{\Phi(k)}{1+\Phi(k)}$. Then, the estimate of $\alpha$ is given by

$$\tilde{\alpha}=\frac{\sqrt{k^4\left(\sum_{i=1}^{n}\ln\frac{x_i}{\tilde{\theta}}\right)^2+4nk^2\sum_{i=1}^{m}\ln^2\frac{x_i}{\tilde{\theta}}}-k^2\sum_{i=1}^{n}\ln\frac{x_i}{\tilde{\theta}}}{2\sum_{i=1}^{m}\ln^2\frac{x_i}{\tilde{\theta}}}.$$

## 5. Numerical Examples

**Example 5.1** (Ciumara, 2006). We estimate the two parameters of the Weibull-Pareto composite model using a data set that was simulated from this model with $\beta=\frac{\alpha}{t_0}=\frac{1}{t_0}=2.85$, $\theta=1000$ and $n=50$. The true value of $m$ is, in this case, 20. Furthermore, we compare the estimation results obtained by considering the two composite models, Weibull-Pareto and lognormal-Pareto, and we expect similar results on estimation.

The following table presents the values of the parameters estimates, the value of the log-likelihood function (denoted by L-L), the Kolmogorov-Smirnov test statistic (K-S in the table), and the value of $\aleph^2$ goodness-of-fit test, together with its p-value (Iosifescu, Mihoc, Theodorescu, 1966). In order to estimate the parameters, we apply the two algorithms presented in Section 3 for the composite Weibull-Pareto model, denoted in the table by `W-P alg1' and `W-P alg2', respectively. In what concerns the parameters of the lognormal-Pareto with density given in $(2.1)$, we used again two algorithms presented in Cooray and Ananda (2005) and in Section 4. We have to mention here that the first algorithm is derived from the maximum likelihood estimation (L-P alg1) and the second one is a combination between the smooth empirical estimate of percentiles and the maximum likelihood estimation (L-P alg2).

| | Parameter estimates | L-L | K-S | $\aleph^2$ | p-value |
|---|---|---|---|---|---|
| W-P alg1 | $\hat{m}=21;\hat{\theta}=1143.08;\hat{\beta}=2.10$ | -458.4 | 0.049 | 1.99 | 0.73 |
| W-P alg2 | $\hat{m}=21;\hat{\theta}=1172.12;\hat{\beta}=1.94$ | -458.6 | 0.069 | 2.64 | 0.61 |
| L-P alg1 | $\hat{m}=20;\hat{\theta}=1025.79;\hat{\alpha}=0.56$ | -461.27 | 0.102 | 4.87 | 0.3 |
| L-P alg2 | $\hat{m}=19;\hat{\theta}=997.97;\hat{\alpha}=0.57$ | -461.28 | 0.095 | 4.83 | 0.3 |

At the 95% level of significance, the critical value of the Kolmogorov-Smirnov test is 0.1923 and the one of $\aleph^2$ goodness-of-fit test with 4 degrees of freedom is 9.49. Taking into account the obtained values, we could decide that the Weibull-Pareto is a better fit for this data set, as it was expected. One has to notice, however, the very good estimates obtained when applying the lognormal-Pareto model as well.

**Example 5.2** Now, we simulated a data set starting from the composite lognormal-Pareto distribution with $\alpha = 1$ ($\beta = \frac{\alpha}{t_0} = \frac{1}{t_0} = 2.85$), $\theta = 1000$ and $n = 50$. The true value of $m$ is, in this case, 19. Again, we compare the estimation results obtained by considering the two composite models, the Weibull-Pareto and the lognormal-Pareto.

| | Parameter estimates | L-L | K-S | $\aleph^2$ | p-value |
|---|---|---|---|---|---|
| W-P alg1 | $\hat{m} = 21; \hat{\theta} = 1193.5; \hat{\beta} = 2.398$ | -458.06 | 0.06 | 5.27 | 0.26 |
| W-P alg2 | $\hat{m} = 21; \hat{\theta} = 1240.25; \hat{\beta} = 2.14$ | -458.40 | 0.06 | 6.52 | 0.16 |
| L-P alg1 | $\hat{m} = 20; \hat{\theta} = 1104.37; \hat{\alpha} = 0.73$ | -457.43 | 0.067 | 1.06 | 0.9 |
| L-P alg2 | $\hat{m} = 19; \hat{\theta} = 979.70; \hat{\alpha} = 0.77$ | -457.91 | 0.122 | 2.52 | 0.64 |

Again, we obtained similar results on estimation. The differences between the acuracy of the two models are given by the values of the test statistics employed.

A remark on the estimation results concerns the second algorithm. Here, in the case of the composite Weibull-Pareto model, for a certain $n$ we will obtain the same $m$, no matter what data set is used. The same is true for the composite lognormal-Pareto model as well.

Finally, we generate 50 data sets starting from the composite Weibull-Pareto model with the same parameters as before $\beta = \frac{\alpha}{t_0} = \frac{1}{t_0} = 2.85$, $\theta = 1000$ and $n = 50$. Applying the second algorithm to estimate the parameters of the composite Weibull-Pareto parameters, we find an average of the estimate for the first parameter $\hat{\theta} = 1450.11$ and for the second one $\hat{\beta} = 1.94$, while applying the first algorithm, we get better results (in average) $\hat{\theta} = 1271.53$ and $\hat{\beta} = 2.47$.

## References

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

Beirlant, J., Joossens, E., Segers, J. (2004), "Generalized Pareto fit to the society of actuaries' large claims database", *North-American Actuarial Journal,* Vol. 8, No. 2/2004, 108-111.

Ciumara, R. (2006), "An actuarial model based on composite Weibull-Pareto

distribution", submitted to *Mathematical Reports,* Vol. 8(58), No. 4/2006.
Cooray, K., Ananda, M.M.A. (2005), "Modeling actuarial data with a composite lognormal-Pareto model", *Scandinavian Actuarial Journal,* No. 5/2005, 321-334.
Embrechts, P., Kluppelberg, C., Mikosch, T. (1999), *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.
Hogg, R.V., Klugman, S.A. (1984), *Loss Distribution,* John Wiley & Sons Inc., New York.
Hossack, I.B., Pollard, J.H., Zehnwirth, B. (1983), *Introductory Statistics with Applications in General Insurance,* Cambridge University Press, United Kingdom.
Iosifescu, M., Mihoc, G., Theodorescu, R. (1966), *Teoria probabilităţilor şi statistica Matematică,* Editura Tehnică, Bucureşti.
Klugmann, S., Panjer, H., Willmot, G. (1998), *Loss Models from Data to Decision,* John Wiley & Sons, New York.
McNeil, A. (1997), "Estimating the tails of loss severity distributions using extreme value theory", *ASTIN Bulletin,* Vol. 27, No. 1/1997, 117-137.
Patrik, G. (1980), "Estimating casualty insurance loss amount distributions", *Proceedings* of the Casualty Actuarial Society, Vol. LXVII/1980, 57-109.
Preda, V. (1992), *Statistical Decision Theory,* Editura Academiei Române.
Ramlau-Hansen, H. (1988), "A solvency study in non-life insurance", *Scandinavian* Actuarial Journal, No. 2/1988, 3-34.