DOI: 10.2478/auom-2022-0032
An. Şt. Univ. Ovidius Constanţa

sciendo
Vol. 30(3),2022, 21–35

On the three-spliced Exponential-Lognormal-Pareto distribution

Adrian Bâcă and Raluca Vernic

Dedicated to the memory of Professor Constantin Popa

Abstract

To model statistical data coming from two different distributions, Cooray and Ananda [1] introduced a composite (two-spliced) Lognormal-Pareto model, that was further extended by Scollnik [9] and fitted to insurance data. Inspired by these studies, more general three-spliced composite models are considered in this work, built by joining three different distributions. In particular, the study is focused on the three-spliced Exponential-Lognormal-Pareto distribution. The main characteristics of this model, as well as statistical inference are discussed. The parameters estimation is illustrated on random generated data.

1 Introduction

A spliced distribution is built of different distributions in subdivided intervals, providing hence more flexibility in capturing the behavior within distinct regions. Therefore, spliced distributions have the potential to better capture tails of loss distributions, from where they have a wide range of applications in general insurance, health insurance, life insurance etc. Klugman et al. [5] introduced splicing as a method for creating new distributions. Starting with

Key Words: Spliced distribution, Exponential, Lognormal, Pareto, parameters estimation.
2010 Mathematics Subject Classification: Primary 60E05, 62F10; Secondary 62-08.
Received: 08.12.2021
Accepted: 08.04.2022

[1], two-component spliced distributions, also called composite distributions, are most studied in the literature in connection with skewed loss data. See, for example, the comprehensive analysis of composite models on a real insurance data set (the Danish fire losses) provided in [4].

However, spliced distributions with more than two components are not so studied, possibly because the thresholds where a spliced distribution changes shape are difficult to estimate. Particular three-spliced regression models were considered by [2] (describing fractional response variables with unignorable zeros and ones) and [3] (with a first component containing zeros). In this paper, a more general three-spliced model is considered, having as first component an exponential distribution, as second component a lognormal distribution, while the third component is Pareto. The aim of such distributions is to better capture some behavior specific to e.g., actuarial data (but not only), where many small and medium claims are recorded, but also some very large claims consistent with a heavy-tailed distribution.

The structure of the paper is as follows: Section 2 first recalls the two-component spliced (composite) distribution, then defines the three-component spliced model, presents some properties, discusses parameters estimation and associated challenges. Subsection 2.2 concentrates on the particular case of the three-spliced Exponential-Lognormal-Pareto distribution. A numerical illustration of its parameters estimation is provided in Section 3. Conclusions and some future research plans are presented in Section 4.

2 Three-component spliced distributions

2.1 General model

Recall the form of a general two-component spliced (composite) probability density function (pdf)

$$f(x;\theta) = \begin{cases} r\frac{f_1(x)}{F_1(\theta)}, & x \le \theta \\ (1-r)\frac{f_2(x)}{1-F_2(\theta)}, & x > \theta \end{cases}, \tag{1}$$

where f_1, f_2 are two pdfs, F_1, F_2 are the corresponding cumulative distribution functions (cdfs), θ is the threshold and $r \in [0,1]$ is a normalizing constant.

In a similar way, the three-component spliced pdf is defined by

$$f(x) = \begin{cases} r_1\frac{f_1(x)}{F_1(\theta_1)}, & x \le \theta_1 \\ r_2\frac{f_2(x)}{F_2(\theta_2)-F_2(\theta_1)}, & \theta_1 < x \le \theta_2 \\ r_3\frac{f_3(x)}{1-F_3(\theta_2)}, & x > \theta_2 \end{cases}, \tag{2}$$

where $f_i, i = 1, 2, 3$ are three pdfs, $F_i, i = 1, 2, 3$ are the corresponding cdfs, $\theta_1 < \theta_2$ are the thresholds and $r_i \in [0, 1], i = 1, 2, 3$ are normalizing constants such that $r_1 + r_2 + r_3 = 1$. Note that (2) can be rewritten as

$$f(x) = \begin{cases} r_1 f_1^*(x), & x \leq \theta_1 \\ r_2 f_2^*(x), & \theta_1 < x \leq \theta_2 \\ r_3 f_3^*(x), & x > \theta_2 \end{cases}, \tag{3}$$

where f_1^*, f_2^*, f_3^* are, respectively, the right truncation of f_1, the left-right truncation of f_2 and the left truncation of f_3.

2.1.1 Some model properties

To obtain a smooth pdf, continuity and differentiability conditions are imposed at both θ_1 and θ_2, leading to the following result:

Proposition 1. *a) By imposing continuity conditions at $\theta_i, i = 1, 2$, it results that:*

$$\frac{r_1}{r_2} = \frac{f_2(\theta_1)}{f_1(\theta_1)} \cdot \frac{F_1(\theta_1)}{F_2(\theta_2) - F_2(\theta_1)}, \tag{4}$$

$$\frac{r_2}{r_3} = \frac{f_3(\theta_2)}{f_2(\theta_2)} \cdot \frac{F_2(\theta_2) - F_2(\theta_1)}{1 - F_3(\theta_2)}. \tag{5}$$

b) If, moreover, differentiability conditions are imposed at $\theta_i, i = 1, 2$, the following restrictions must hold:

$$\frac{f_1'(\theta_1)}{f_1(\theta_1)} = \frac{f_2'(\theta_1)}{f_2(\theta_1)}, \tag{6}$$

$$\frac{f_3'(\theta_2)}{f_3(\theta_2)} = \frac{f_2'(\theta_2)}{f_2(\theta_2)}. \tag{7}$$

Proof. The proof of (a) is immediate, hence omitted.
b) The first relation is proved below, the second resulting in a similar way. The differentiability condition at θ_1 yields:

$$r_1 \frac{f_1'(\theta_1)}{F_1(\theta_1)} = r_2 \frac{f_2'(\theta_1)}{F_2(\theta_2) - F_2(\theta_1)} \Rightarrow \frac{r_1}{r_2} = \frac{f_2'(\theta_1)}{f_1'(\theta_1)} \cdot \frac{F_1(\theta_1)}{F_2(\theta_2) - F_2(\theta_1)},$$

which combined with relation (4) gives formula (6). This completes the proof.□

Remark 1. *From relations (4) and (5), the following formulas of* r_1, r_2, r_3 *are obtained:*

$$\begin{aligned} r_2 &= \left(1+\frac{f_2(\theta_1)}{f_1(\theta_1)}\cdot\frac{F_1(\theta_1)}{F_2(\theta_2)-F_2(\theta_1)}+\frac{f_2(\theta_2)}{f_3(\theta_2)}\cdot\frac{1-F_3(\theta_2)}{F_2(\theta_2)-F_2(\theta_1)}\right)^{-1}, \\ r_1 &= r_2\cdot\frac{f_2(\theta_1)}{f_1(\theta_1)}\cdot\frac{F_1(\theta_1)}{F_2(\theta_2)-F_2(\theta_1)}, \\ r_3 &= r_2\cdot\frac{f_2(\theta_2)}{f_3(\theta_2)}\cdot\frac{1-F_3(\theta_2)}{F_2(\theta_2)-F_2(\theta_1)}. \end{aligned}$$

In Figure 1, several three-component spliced pdfs satisfying all continuity and differentiability conditions are plotted, while Figure 2 displays similar pdfs which do not satisfy the differentiability condition in θ_1. Note the variety of shapes.

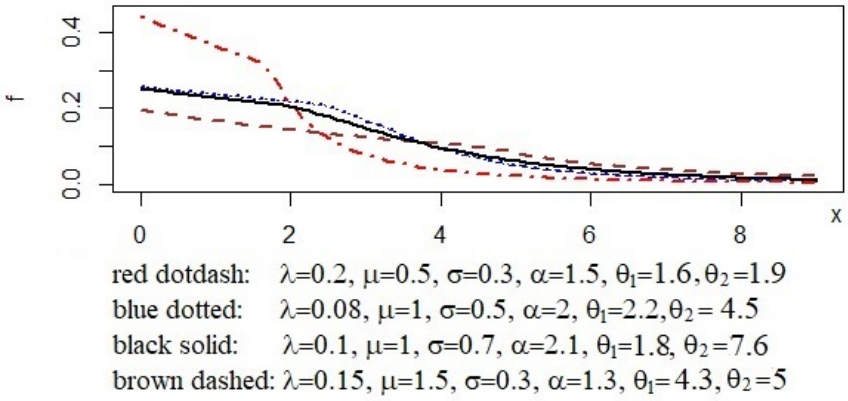

Figure 1: Exponential-Lognormal-Pareto three-component spliced pdfs (with differentiability conditions)

Let $M_f(t) = E(e^{tX})$ denote the moment generating function (mgf) of the random variable (rv) X and let $E_n(f) = E(X^n)$ denote its initial moment of order n. Some properties of the three-component spliced distribution are given in the following:

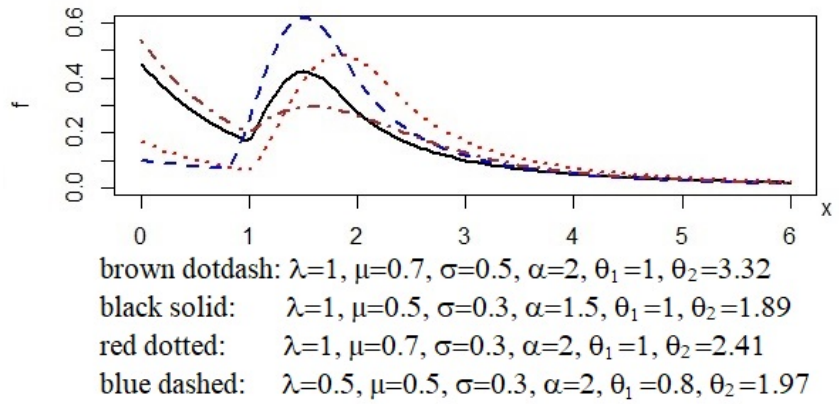

Figure 2: Exponential-Lognormal-Pareto three-component spliced pdfs without differentiability condition in θ_1

Proposition 2. *a) The cdf of (2) is*

$$F(x) = \begin{cases} r_1 \frac{F_1(x)}{F_1(\theta_1)}, & x \leq \theta_1 \\ r_1 + r_2 \frac{F_2(x)-F_2(\theta_1)}{F_2(\theta_2)-F_2(\theta_1)}, & \theta_1 < x \leq \theta_2 \\ r_1 + r_2 + r_3 \frac{F_3(x)-F_3(\theta_2)}{1-F_3(\theta_2)}, & x > \theta_2 \end{cases}. \tag{8}$$

b) Its mgf is

$$M_f(t) = r_1 M_{f_1^*}(t) + r_2 M_{f_2^*}(t) + r_3 M_{f_3^*}(t).$$

c) Its initial moment of order n is

$$E_n(f) = r_1 E_n(f_1^*) + r_2 E_n(f_2^*) + r_3 E_n(f_3^*).$$

Proof. a) Three cases are identified:
- If $x \leq \theta_1$ then $F(x) = r_1 \int_{-\infty}^{x} \frac{f_1(y)}{F_1(\theta_1)} dy = \frac{F_1(x)}{F_1(\theta_1)}$;
- If $\theta_1 < x \leq \theta_2$ then

$$F(x) = r_1 \int_{-\infty}^{\theta_1} \frac{f_1(y)}{F_1(\theta_1)} dy + r_2 \int_{\theta_1}^{x} \frac{f_2(y)}{F_2(\theta_2) - F_2(\theta_1)} dy,$$

which yields the second formula of F;
- If $x > \theta_2$ then

$$F(x) = r_1 \int_{-\infty}^{\theta_1} \frac{f_1(y)}{F_1(\theta_1)} dy + r_2 \int_{\theta_1}^{\theta_2} \frac{f_2(y)}{F_2(\theta_2) - F_2(\theta_1)} dy + r_3 \int_{\theta_2}^{x} \frac{f_3(y)}{1 - F_3(\theta_2)} dy,$$

and the third formula of F is obtained.
b) Using the spliced pdf formula (3) gives

$$\begin{aligned} M_f(t) &= r_1 \int_{-\infty}^{\theta_1} e^{tx} f_1^*(x)dx + r_2 \int_{\theta_1}^{\theta_2} e^{tx} f_2^*(x)dx + r_3 \int_{\theta_2}^{\infty} e^{tx} f_3^*(x)dx \\ &= r_1 M_{f_1^*}(t) + r_2 M_{f_2^*}(t) + r_3 M_{f_3^*}(t). \end{aligned}$$

c) Results similarly to (b). □

Remark 2. *Assuming that F_1, F_2 and F_3 admit inverse functions, the above cdf can be used to generate random values from the three-component spliced pdf (2) by using the inversion method.*

2.1.2 Parameters estimation

Estimating the parameters is already a difficult problem for two-spliced distributions because the threshold where the spliced distribution changes shape is assumed to be a parameter. For three-spliced distributions, the estimation becomes even more challenging due to the necessity to estimate both thresholds in addition to the parameters of f_1^*, f_2^*, f_3^*.

Let $\mathbf{x} = (x_1, ..., x_n)$ be a random data sample and let $\delta_1, ...\delta_s, \theta_1, \theta_2$ denote the parameters of pdf (2), with $s \in \mathbb{N}$. Also, assume that the data sample is ordered, i.e., $x_1 \leq ... \leq x_n$. If the unknowns parameters $\theta_i \in [x_{m_i}, x_{m_i+1}], i = 1, 2$, then the corresponding likelihood function is

$$L(\mathbf{x}; \delta_1, ...\delta_s, \theta_1, \theta_2) = \prod_{i=1}^{m_1} r_1 f_1^*(x_i) \prod_{i=m_1+1}^{m_2} r_2 f_2^*(x_i) \prod_{i=m_2+1}^{n} r_3 f_3^*(x_i). \quad (9)$$

Similarly to the usual maximum likelihood estimation (MLE) based algorithm used for a two-spliced distribution, which consists in sorting the data set and looking for the MLE solution with the threshold in-between each two consecutive data, the following algorithm is proposed:

Step 1. For $i = 1$ to $n - 2$

For $j = i + 1$ to $n - 1$

Evaluate $\delta_1, ...\delta_s, \theta_1, \theta_2$ as solutions of the optimization problem:

$$\max \log L(\mathbf{x}; \delta_1, ...\delta_s, \theta_1, \theta_2),$$

under the constraints: $\theta_1 \in [x_i, x_{i+1}], \theta_2 \in [x_j, x_{j+1}]$, continuity and differentiability.

Step 2. Among the solutions obtained at Step 1 choose the one that maximizes the log-likelihood function.

Unfortunately, this algorithm proved to be very time consuming. For example, for only $n = 100$ data, this algorithm implemented in MATLAB using the `fmincon` function took about 30 minutes. Therefore, other methods can be considered to reduce the searching intervals of i and j at Step 1. For example, some empirical quantiles or a combination with the method of moments (e.g., with the expected value) can be used, provided that the resulting systems of equations are tractable (depending on the distributions involved).

Another alternative is, as described by [8], to use the mean excess plot to detect different parts of the distribution by viewing where a transition from one part of the distribution to another part is suitable. For example, a Pareto tail could be detected if a point t beyond which the mean excess plot is linearly increasing can be found.

The mean excess plot consists of estimates for the mean excess values

$$e(u) = E(X - u \,|X > u\,) = \frac{1}{1 - F(u)} \int_u^{\infty} (1 - F(x))dx,$$

where $u = X_{n-k,n} = \hat{Q}\left(\frac{n-k}{n+1}\right), k = 1, ..., n-1$ are order statistics, the cdf F is estimated by the empirical cdf and $\hat{Q}$ is the corresponding empirical quantile function.

2.2 Particular case: Exponential-Lognormal-Pareto spliced distribution

The first well studied two-component spliced distribution was the composite Lognormal-Pareto distribution, see [1], [9], [7] or [6]. Starting from this distribution, in the following, the three-component spliced Exponential-Lognormal-Pareto distribution is proposed, obtained as a particular case of (2), where f_1 is the exponential $\text{Exp}(\lambda)$ pdf, f_2 is the Lognormal $\text{LogN}(\mu, \sigma^2)$ pdf, while f_3 is the Pareto type I $\text{Pa}(\alpha, \theta_2)$ pdf, $\lambda, \sigma, \alpha > 0, \mu \in \mathbb{R}$. Therefore, pdf (2) becomes in this particular case

$$f(x) = \begin{cases} r_1 \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda \theta_1}}, & x \leq \theta_1 \\ r_2 \frac{\varphi(\ln x; \mu, \sigma)}{x\left(\Phi\left(\frac{\ln \theta_2 - \mu}{\sigma}\right) - \Phi\left(\frac{\ln \theta_1 - \mu}{\sigma}\right)\right)}, & \theta_1 < x \leq \theta_2 \\ r_3 \frac{\alpha \theta_2^{\alpha}}{x^{\alpha+1}}, & x > \theta_2 \end{cases}, \tag{10}$$

where $\varphi(\cdot; \mu, \sigma)$ denotes the pdf of the normal distribution $\text{N}(\mu, \sigma^2)$ and Φ the cdf of the standard normal distribution N(0,1).

When imposing the continuity conditions, the normalizing constants r_1, r_2, r_3 are obtained from Remark 1. Moreover, the differentiability conditions lead to the following result:

Proposition 3. *By imposing differentiability conditions to the Exponential-Lognormal-Pareto pdf (10) the following restrictions must hold:*

$$\begin{aligned} i)\, \lambda\theta_1 &= 1 + \frac{\ln\theta_1 - \mu}{\sigma^2}, \\ ii)\, \alpha &= \frac{\ln\theta_2 - \mu}{\sigma^2}. \end{aligned}$$

Proof. It holds that:

$$\begin{aligned} f_1(x) &= \lambda e^{-\lambda x} \Rightarrow f_1'(x) = -\lambda^2 e^{-\lambda x} = -\lambda f_1(x), \\ f_2(x) &= \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \Rightarrow f_2'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{\sigma^2}} \left[-\frac{1}{x^2} - \frac{1}{x^2}\frac{2(\ln x-\mu)}{2\sigma^2}\right] \\ &= -\frac{1}{x}\left(1 + \frac{\ln x - \mu}{\sigma^2}\right) f_2(x), \\ f_3(x) &= \frac{\alpha\theta_2^\alpha}{x^{\alpha+1}} \Rightarrow f_3'(x) = -\alpha(\alpha+1)\frac{\theta_2^\alpha}{x^{\alpha+2}} = -\frac{\alpha+1}{x} f_3(x). \end{aligned}$$

Therefore,

$$\frac{f_1'(\theta_1)}{f_1(\theta_1)} = -\lambda, \frac{f_2'(\theta_1)}{f_2(\theta_1)} = -\frac{1}{\theta_1}\left(1 + \frac{\ln\theta_1 - \mu}{\sigma^2}\right), \frac{f_3'(\theta_2)}{f_3(\theta_2)} = -\frac{\alpha+1}{\theta_2},$$

which inserted in (6) yields formula (i) and inserted in (7) yields (ii).□

In Figures 1 and 2, several Exponential-Lognormal-Pareto pdfs satisfying all continuity and differentiability conditions and, respectively, not satisfying the differentiability condition in θ_1 are plotted.

In next proposition, formulas for the cdf and the expected value of the Exponential-Lognormal-Pareto distribution are presented.

Proposition 4. *The cdf and expected value of the Exponential-Lognormal-Pareto distribution are, respectively, given by:*

$$i)\, F(x) = \begin{cases} r_1 \frac{1-e^{-\lambda x}}{1-e^{-\lambda\theta_1}}, & x \le \theta_1 \\ r_1 + r_2 \frac{\Phi\left(\frac{\ln x-\mu}{\sigma}\right) - \Phi\left(\frac{\ln\theta_1-\mu}{\sigma}\right)}{\Phi\left(\frac{\ln\theta_2-\mu}{\sigma}\right) - \Phi\left(\frac{\ln\theta_1-\mu}{\sigma}\right)}, & \theta_1 < x \le \theta_2 \\ r_1 + r_2 + r_3\left(1 - \left(\frac{\theta_2}{x}\right)^\alpha\right), & x > \theta_2 \end{cases};$$

$$ii)\, E_1(f) = \frac{r_1}{\lambda}\left(1-\frac{\lambda\theta_1 e^{-\lambda\theta_1}}{1-e^{-\lambda\theta_1}}\right) + r_2 e^{\mu+\frac{\sigma^2}{2}} \frac{\Phi\left(\frac{\ln\theta_2-\mu-\sigma^2}{\sigma}\right)-\Phi\left(\frac{\ln\theta_1-\mu-\sigma^2}{\sigma}\right)}{\Phi\left(\frac{\ln\theta_2-\mu}{\sigma}\right)-\Phi\left(\frac{\ln\theta_1-\mu}{\sigma}\right)} + \frac{r_3\alpha\theta_2}{\alpha-1}, \alpha>1.$$

Proof. Formula (i) results by inserting the corresponding cdfs (exponential, lognormal and Pareto) into formula (8).
ii) From (c) in Proposition 2, it is known that

$$E_1(f) = r_1E_1(f_1^*) + r_2E_1(f_2^*) + r_3E_1(f_3^*),$$

hence each part must be calculated separately. For the truncated exponential it holds that:

$$\begin{aligned} E_1(f_1^*) &= \int_0^{\theta_1} x\frac{\lambda e^{-\lambda x}}{1-e^{-\lambda\theta_1}}dx = \frac{1}{1-e^{-\lambda\theta_1}}\left[-xe^{-\lambda x}\Big|_0^{\theta_1} - \int_0^{\theta_1}(-e^{-\lambda x})dx\right] \\ &= \frac{1}{1-e^{-\lambda\theta_1}}\left[-\theta_1 e^{-\lambda\theta_1} - \frac{e^{-\lambda x}}{\lambda}\Big|_0^{\theta_1}\right] \\ &= \frac{1}{\lambda\left(1-e^{-\lambda\theta_1}\right)}\left(-\lambda\theta_1 e^{-\lambda\theta_1}+1-e^{-\lambda\theta_1}\right) \\ &= \frac{1}{\lambda}\left(1-\frac{\lambda\theta_1 e^{-\lambda\theta_1}}{1-e^{-\lambda\theta_1}}\right). \end{aligned}$$

To calculate $E_1(f_2^*)$, note the fact that f_2^* is the pdf of a r.v., say Y, having a doubly truncated lognormal distribution with truncation limits θ_1, θ_2. Therefore, it is well known that $Z = \ln Y$ follows a doubly truncated normal distribution with the same parameters as Y and truncation limits $\ln\theta_1, \ln\theta_2$. Moreover, $E_1(f_2^*) = E(Y) = E\left(e^Z\right) = \mathcal{L}_Z(-1)$, where $\mathcal{L}_Z$ denotes the Laplace transform of Z, which, for the truncated normal distribution $\mathrm{N}(\mu,\sigma^2;a,b)$ is given by

$$\mathcal{L}_Z(t) = e^{-t\mu+\frac{t^2\sigma^2}{2}} \frac{\Phi\left(\frac{b-\mu+t\sigma^2}{\sigma}\right)-\Phi\left(\frac{a-\mu+t\sigma^2}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}$$

It follows that

$$E_1(f_2^*) = e^{\mu+\frac{\sigma^2}{2}} \frac{\Phi\left(\frac{\ln\theta_2-\mu-\sigma^2}{\sigma}\right)-\Phi\left(\frac{\ln\theta_1-\mu-\sigma^2}{\sigma}\right)}{\Phi\left(\frac{\ln\theta_2-\mu}{\sigma}\right)-\Phi\left(\frac{\ln\theta_1-\mu}{\sigma}\right)}.$$

For the Pareto distribution it is well known that for $\alpha > 1$,

$$E_1(f_3^*) = \frac{\alpha\theta_2}{\alpha - 1}.$$

By inserting the three expectations into the above expression of $E_1(f)$, formula (ii) is obtained.□

Generating random values from the Exponential-Lognormal-Pareto distribution. This can be done by using the inversion method based on the cdf given in the previous theorem as follows: generate u a uniform $\mathrm{U}(0,1)$ value, then

• If $u \leq r_1$ then solve for x the equation $u = r_1 \frac{1-e^{-\lambda x}}{1-e^{-\lambda\theta_1}}$;

• If $r_1 < u \leq r_1 + r_2$ then solve for x the equation $u = r_1 + r_2 \frac{\Phi\left(\frac{\ln x - \mu}{\sigma}\right) - \Phi\left(\frac{\ln \theta_1 - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln \theta_2 - \mu}{\sigma}\right) - \Phi\left(\frac{\ln \theta_1 - \mu}{\sigma}\right)}$;

• If $r_1 + r_2 < u$ then solve for x the equation $u = r_1 + r_2 + r_3 \left(1 - \left(\frac{\theta_2}{x}\right)^{\alpha}\right)$.

The resulting x is an Exponential-Lognormal-Pareto random value.

3 Numerical illustration

In this section, simulated data are used to check the estimation procedure for the Exponential-Lognormal-Pareto distribution. Using the inversion method described above, $n = 1000$ values were simulated from (10) with parameters $\lambda = 0.08, \mu = 1, \sigma = 0.5, \alpha = 2, \theta_1 = 2.213, \theta_2 = 4.482$ and $r_1 = 0.519, r_2 = 0.324, r_3 = 0.157$. Note that these parameters values satisfy the continuity and differentiability conditions given in Proposition 3. The main descriptive statistics of the generated data sample are given in Table 1.

Min	1st Qu.	Median	Mean	3rd Qu.	Max.	St.Dev.	Var.
0.005	1.004	2.163	3.051	3.571	70.247	4.390	19.275

Table 1: Descriptive statistics for the simulated data set

In Figure 3, the histogram of the generated data together with the true Exponential-Lognormal-Pareto density are plotted. Note how well the true density fits the histogram.

To estimate the parameters by the algorithm presented in Section 2.1.2, the running time becomes prohibitive (for only 100 data, it took almost 30 min on a PC i7-8550U CPU, 16 GB RAM, SSD). Therefore, from the plot of the mean excess, it can be noted that there is a significant shape change around $t = 2.1$, see Figure 4. Unfortunately, this plot does not give any visual information

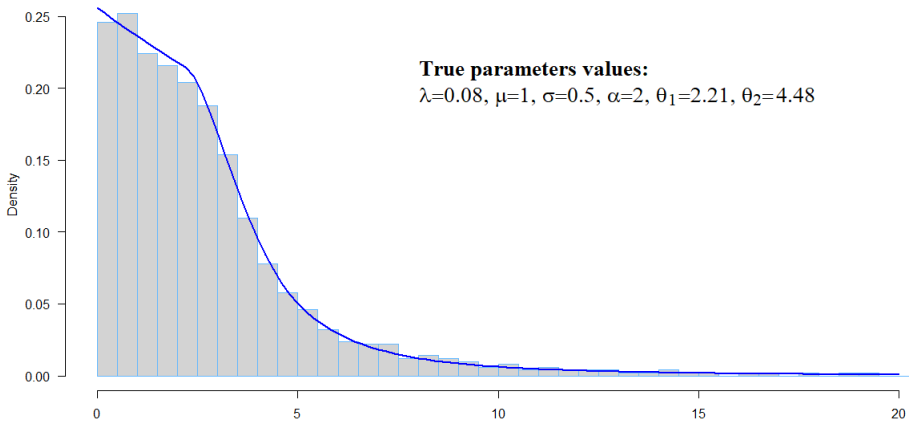

Figure 3: Histogram with true Exponential-Lognormal-Pareto pdf

about a second change of shape. Therefore, in a first step, the search of θ_1 was limited around the value of t; then, in a second step, the search of θ_2 was restricted around the same value. This procedure significantly reduced the computing time, but still, each step lasted a couple of hours.

The estimated parameters resulting by the described procedure are given in the last line of Table 2. They were obtained in the first step (i.e., choice of θ_1 around t) according to the best MLE value from both steps.

	λ	μ	σ	α	θ_1	θ_2	r_1	r_2	r_3
True	0.08	1	0.5	2	2.213	4.482	0.519	0.324	0.157
Estimated	0.093	1.024	0.473	1.903	2.338	4.262	0.538	0.281	0.181

Table 2: True parameters of simulated data (first line) and estimated parameters (second line)

Even more, other models were fitted to the same data, i.e., the exponential distribution, the lognormal distribution and two composite distributions: exponential-lognormal and lognormal-Pareto. The best fist is provided by the Exponential-Lognormal-Pareto model, as it can be seen from Table 3, where several goodness-of-fit measures are displayed (for each one, minimum value means best model). Recall the formula of each such measure, with k the number of free parameters and n the sample size:

- Negative Log-Likelihood: $NLL = -\log L$.

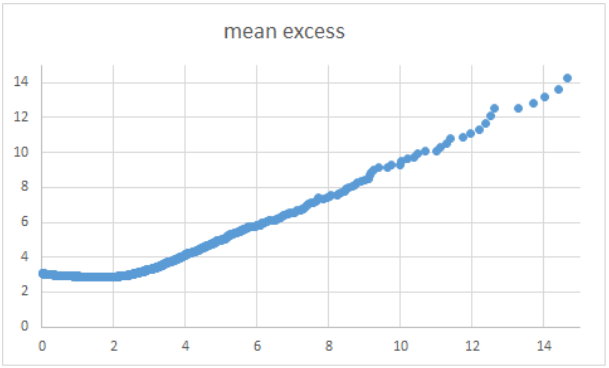

Figure 4: Mean excess plot for simulated data

- Akaike's Information Criterion: $AIC = -2\log L + 2k$.
- Bayesian Information Criterion: $BIC = -2\log L + k\ln n$.
- Hurvich and Tsai's Criterion: $AICc = -2\log L + \frac{2nk}{n-k-1}$.
- Bozdogan's criterion: $CAIC = -2\log L + k(\ln n + 1)$.

Model	NLL	AIC	BIC	AICc	CAIC
Exponential	2115.56	4233.12	4234.12	4233.12	4235.12
Lognormal	2141.65	4287.31	4289.31	4287.32	4291.31
Lognormal-Pareto	2851.25	5708.5	5711.5	5708.52	5714.5
Exponential-Lognormal	2067.06	4140.12	4143.12	4140.15	4146.12
Exponential-Lognormal-Pareto	2050.71	4109.43	4113.43	4109.47	4117.43

Table 3: Goodness-of-fit measures for models comparison

Further on, two goodness-of-fit tests were applied: Kolmogorov-Smirnov and Chi-square. Kolmogorov's distance was calculated using the formula

$$D_n = \max_{i=1,\dots,n} \left| F_n^*(x_i) - F(x_i) \right|,$$

where F_n^* denotes the empirical cdf of the data and F the Exponential-Lognormal-Pareto cdf given in (i) Proposition 4. For the simulated data, the

value $\sqrt{n}D_n = 0.174$ was obtained, which is less than the 95% Kolmogorov quantile $q_{95\%} = 1.358$, hence the hypothesis that the Exponential-Lognormal-Pareto distribution (with the estimated parameters) fits the data is accepted.

To apply the Chi-square test, the generated data were divided into $r = 9$ intervals, as it can be seen from Table 4, where the theoretical (denoted by n_i) and empirical frequencies (np_i) calculated for each interval are displayed. The chi-square distance obtained with the formula

$$X^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i},$$

is $X^2 = 0.9095$. This is smaller than the chi-square quantile $q_{4,95\%} = 9.488$, where the number of the degrees of freedom is calculated as $4 = r - 1 - j$, where j is the number of estimated parameters ($j = 4$ in this case, the other parameters being related as in Proposition 3). Hence, according to the Chi-square test, the Exponential-Lognormal-Pareto distribution also fits the data.

Intervals	Empirical freq., n_i	Theoretical freq., np_i	$\frac{(n_i-np_i)^2}{np_i}$
0.0 - 0.5	123	125.047	0.033
0.5 - 1.0	126	119.372	0.368
1.0 - 1.5	112	113.954	0.033
1.5 - 2.1	125	129.939	0.188
2.1 - 2.7	125	121.803	0.084
2.7 - 3.5	131	129.390	0.020
3.5 - 5.0	123	127.062	0.130
5.0 - 10	100	97.749	0.052
10 - 70.25	35	34.810	0.001
$\sum$	1000	X^2 *dist.* $= 0.9095$	

Table 4: Chi-square test

4 Conclusions and future work

Starting from the composite distributions, with the purpose to better model some real data, three-component spliced distributions were introduced in this work. As seen from the plots, these distributions have flexible shapes that can capture some specific data behavior. The study focused on the particular Exponential-Lognormal-Pareto distribution, for which some close-type formulas were obtained and an estimation procedure was discussed, with accent

on its challenges. The estimation procedure was illustrate on a generated data set. Unfortunately, due to the fact that the thresholds where the spliced distribution changes shape are assumed to be parameters, hence making the estimation more challenging, the proposed estimation algorithm turned out to be very time consuming for the 1000 generated data; therefore, as future work, the intention is to look for a faster estimation method. Also, other particular three-spliced distributions can be studied and fitted on some real data sets.

Acknowledgments

The authors acknowledge the help of the anonymous referee in improving the paper.

References

[1] Cooray, K., Ananda, M. M. (2005). Modeling actuarial data with a composite lognormal-Pareto model. Scandinavian Actuarial Journal, 2005(5), 321-334.

[2] Fang, K., Ma, S. (2013). Three-part model for fractional response variables with application to Chinese household health insurance coverage. Journal of Applied Statistics, 40(5), 925-940.

[3] Gan, G., Valdez, E. A. (2018). Fat-tailed regression modeling with spliced distributions. North American Actuarial Journal, 22(4), 554-573.

[4] Grün, B., Miljkovic, T. (2019). Extending composite loss models using a general framework of advanced computational tools. Scandinavian Actuarial Journal, 2019(8), 642-660.

[5] Klugman, S. A., Panjer, H. H., Willmot, G. E. (2012). Loss models: from data to decisions (Vol. 715). John Wiley & Sons.

[6] Mutali, S., Vernic, R. (2020). On the composite LognormalPareto distribution with uncertain threshold. Communications in Statistics-Simulation and Computation, to appear.

[7] Nadarajah, S., Bakar, S. A. A. (2013). CompLognormal: An R Package for Composite Lognormal Distributions. R J., 5(2), 97.

[8] Reynkens, T., Verbelen, R., Beirlant, J., Antonio, K. (2017). Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions. Insurance: Mathematics and Economics, 77, 65-77.

[9] Scollnik, D. P. (2007). On composite lognormal-Pareto models. Scandinavian Actuarial Journal, 2007(1), 20-33.

Adrian BÂCĂ,
Doctoral School,
Ovidius University of Constanta,
124 Mamaia, 900527 Constanta, Romania.
Email: bacaadi@yahoo.com

Raluca VERNIC,
Faculty of Mathematics and Computer Science,
Ovidius University of Constanta,
124 Mamaia, 900527 Constanta, Romania.
Email: rvernic@univ-ovidius.ro