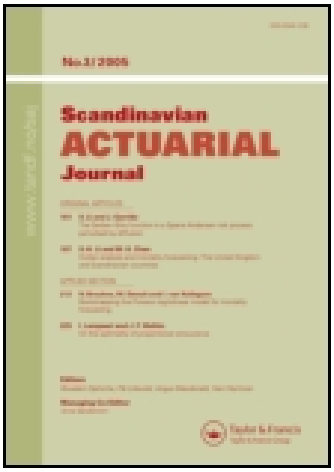

Scandinavian Actuarial Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/sact20

New composite models for the Danish fire insurance data

S. Nadarajah [a] & S.A.A. Bakar [a]
[a] School of Mathematics , University of Manchester , Manchester , UK
Published online: 15 Aug 2012.

To cite this article: S. Nadarajah & S.A.A. Bakar (2014) New composite models for the Danish fire insurance data, Scandinavian Actuarial Journal, 2014:2, 180-187, DOI: 10.1080/03461238.2012.695748

To link to this article: http://dx.doi.org/10.1080/03461238.2012.695748

Scandinavian Actuarial Journal, 2014
Vol. 2014, No. 2, 180–187, http://dx.doi.org/10.1080/03461238.2012.695748

Taylor & Francis
Taylor & Francis Group

Original Article

New composite models for the Danish fire insurance data

S. NADARAJAH* and S.A.A. BAKAR

School of Mathematics, University of Manchester, Manchester, UK

(*Accepted May 2012*)

In recent years, several composite models based on the lognormal distribution have been developed for the Danish fire insurance data. In this note, we propose new composite models based on the lognormal distribution. At least one of the newly proposed models is shown to give a better fit to the Danish fire insurance data.

Keywords: composite models; estimation; lognormal distribution

1. Introduction

Piecewise distributions (sometimes known as spliced distributions) have been introduced in many applications in statistics. Klugman *et al.* (2004) expressed the general form of splicing distributions by the probability density function:

$$f(x) = \begin{cases} a_1 f_1^*(x), & \text{if } c_0 < x < c_1, \\ a_2 f_2^*(x), & \text{if } c_1 < x < c_2, \\ \vdots & \vdots \\ a_k f_k^*(x), & \text{if } c_{k-1} < x < c_k, \end{cases} \tag{1}$$

where:

$$f_i^*(x) = \frac{f_i(x)}{\int_{c_{i-1}}^{c_i} f_i(x)\mathrm{d}x}$$

is a truncated probability density function in which f_i (x), $i = 1,2,\ldots,k$ are standard probability density functions. We shall refer to a_i as mixing weights and c_i as the range limit of the domain.

Composite models that can be defined in terms of two probability density functions are a special case of (1):

$$f(x) = \begin{cases} a_1 f_1^*(x), & \text{if } -\infty < x < \theta, \\ a_2 f_2^*(x), & \text{if } \theta < x < \infty. \end{cases} \tag{2}$$

*Corresponding author. E-mail: mbbsssn2@manchester.ac.uk

Along with the mixing weights, a_i for $i=1$ and 2, (2) serves as a legitimate probability density function. It assumes different weighted truncated distributions for different ranges of the random variable. An early example satisfying such a representation can be found in the resemblance of the symmetric and asymmetric Laplace distributions.

Recent findings by Cooray and Ananda (2005) show that composite models like (2) can give better fits than standard univariate models. They illustrate this fact for the Danish fire insurance data by introducing a model referred to as the *composite lognormal-Pareto model* obtained by piecing together lognormal and Pareto probability density functions. Scollnik (2007) improved the composite lognormal-Pareto model by using mixing weights as coefficients for each piecewise function, replacing the constant weights applied earlier by Cooray and Ananda (2005). Scollnik (2007) also employed the generalized Pareto distribution in replace of the Pareto distribution used by Cooray and Ananda (2005). The later model will be reproduced in Section 3 for the purpose of comparison.

The Danish fire insurance data have been a popular data-set in the statistics literature. It has been used as the main example data for many newly developed statistical models and methods. Some recent uses of the data are with respect to: Bayesian estimation of finite time ruin probabilities (Ausin *et al.*, 2009); hybrid Pareto models (Carreau and Bengio, 2009); beta kernel quantile estimation (Charpentier and Oulidi, 2010); bivariate compound Poisson process (Esmaeili and Kluppelberg, 2010); and non-parametric Bayesian inference on bivariate extremes (Guillotte *et al.*, 2011). See also Burnecki and Weron (2004) and Drees and Muller (2008).

The aim of this note is to improve the fittings for the Danish fire insurance data using various new composite models. The newly developed models include the composite lognormal-Burr model. Prior to any attempt to construct the models, some characteristics of composite distributions are introduced in Section 2. The composite lognormal-Pareto model due to Scollnik (2007) and the newly developed models are given in Section 3. Finally, results of the fits to the Danish fire insurance data are discussed in Section 4. Note that Section 3 gives the methodological, technical contribution.

The models considered in this note are parametric. One could also consider non-parametric models, a possible future work. But parametric models have several advantages: (1) parametric models are simple to fit while non-parametric models can be computationally more expensive; (2) parametric models can be easily used for extrapolation and interval estimation with more power; (3) the lognormal distribution is historically associated with modeling actuarial data; (4) parametric approaches are better understood; (5) most models fitted for the Danish fire insurance data are parametric; and, so on.

2. Some criteria for the models

The generalized composite model proposed by Klugman *et al.* (2004) in (1) is not generally continuous. Therefore, several criteria are imposed to ensure that the resulting composite model is continuous and smooth. Below are some considerations taken into account.

2.1. Continuity

Although (1) asserts a legitimate probability density function it is not continuous in general. For these criteria to hold, we impose the following condition:

$$f(\theta-) = f(\theta+). \tag{3}$$

Using this, the left piecewise function is joined to the tail of the distribution at a threshold, θ. We shall regard θ as a model parameter.

2.2. Differentiability

Another important criterion is smoothness of the probability density function. The following differential condition is necessary to ensure that the resulting probability density function is smooth:

$$f'(\theta-) = f'(\theta+). \tag{4}$$

In order to apply this, the initial probability density function, $f(x)$, must be continuous at every point x.

2.3. Mixing weights

In model (1), it is necessary that the summation of all the coefficients, a_i, $i = 1,2,\ldots,n$, equals 1. In a composite model which involves a two-piecewise distribution, this can be simplified to $a_1 = r$ and $a_2 = 1 - r$, where $0 < r < 1$.

However, there are several other ways of representing the mixing weights in forms so that they add to 1. Since:

$$r = \frac{f_2(\theta)F_1(\theta)}{f_2(\theta)F_1(\theta) + f_1(\theta)[1 - F_2(\theta)]},$$

it is more convenient to choose the mixing weights in a logistic form, that is:

$$a_1 = \frac{1}{1+\phi}$$

and:

$$a_2 = \frac{\phi}{1+\phi},$$

where:

$$\phi = \frac{f_1(\theta)[1 - F_2(\theta)]}{f_2(\theta)F_1(\theta)}.$$

Note that $\phi > 0$.

Scollnik (2007) suggests that having mixing weights depending on the distribution parameters gives a better fit as compared to constant coefficients. So, we take mixing

weights as functions of the distributional parameters involved. This function might not always have an explicit form.

3. A new composite model

In this section, new composite models are constructed with threshold value, θ, and flexible mixing weights, a_1 and a_2, whereby $a_1 + a_2 = 1$. We derive the mixing weights as a function of the model parameters. Also, we reduce the parameters estimated by expressing the location parameter, μ, as a function of the remaining parameters.

For comparative purposes, we reproduce the composite lognormal-Pareto model introduced by Scollnik (2007). This model gave the most outstanding fit among all models that Scollnik (2007) examined. We will comfortably use the form discussed in Section 2 to express the mixing weights. Let X be a random variable having the probability density function:

$$f(x) = \begin{cases} \frac{1}{1+\phi} f_1^*(x), & \text{for } 0 < x \leq \theta, \\ \frac{\phi}{1+\phi} f_2^*(x), & \text{for } \theta \leq x < \infty, \end{cases} \tag{5}$$

where $f_1^*(x)$ and $f_2^*(x)$ are the truncated probability density functions. Scollnik (2007) took the truncated distributions to be lognormal and Pareto of the second kind (also known as Lomax distribution). That is, $f_1(x)$ and $f_2(x)$ are given by:

$$f_1(x) = \frac{1}{x\sigma} \psi\left(\frac{\ln x - \mu}{\sigma}\right) \tag{6}$$

and:

$$f_2(x) = \frac{\alpha \lambda^\alpha}{(x+\lambda)^{\alpha+1}},$$

respectively, for $x > 0$, $\sigma > 0$, $\alpha > 0$, $\lambda > 0$, and $-\infty < \mu < \infty$, where $\psi(\cdot)$ denotes the standard normal probability density function. Applying the continuity condition at θ gives:

$$\phi = \frac{(\theta + \lambda)\psi[(\ln \theta - \mu)/\sigma]}{\alpha\theta\sigma\Phi[(\ln \theta - \mu)/\sigma]},$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. It is also a concern that the resulting probability density function may not be smooth. Imposing the condition (4) at θ not only makes the function smooth but it also reduces the number of model parameters. Applying the condition leads to:

$$\mu = \ln(\theta) - \left(\frac{\alpha\theta - \lambda}{\theta + \lambda}\right)\sigma^2.$$

Note that, although μ can be expressed as a function of the remaining parameters, the method of maximum likelihood does not give explicit solutions to the parameters. So, the maximum likelihood estimates must be obtained numerically using some quasi-Newton algorithm.

We propose new composite models by taking $f_1(\cdot)$ as the lognormal probability density function in (6) but taking a variety of different forms for $f_2(\cdot)$. The parametric distributions considered for $f_2(\cdot)$ are those implemented in the R contributed package actuar (Dutang *et al.*, 2008; R Development Core Team, 2011): Burr, paralogistic, inverse Burr, inverse paralogistic, transformed beta, *F*, Fréchet, Weibull, generalized Pareto, inverse Pareto, loglogistic, exponential, inverse exponential, gamma, inverse gamma, transformed gamma, and inverse transformed gamma. The model for each $f_2(\cdot)$ is constructed in the same way as the composite lognormal-Pareto model.

One important issue is estimation. In the real data application described in Section 4, we estimate parameters by the method of maximum likelihood. As we shall see now, this method does not give explicit solutions to the parameters.

Suppose we have a random sample $x_1,x_2,\ldots,x_n$ from (5). Let $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\ldots,\lambda_q)$ be the parameters of $f_2(\cdot)$. Suppose also that μ and ϕ can be expressed as $\mu=\mu(\sigma,\theta,\boldsymbol{\lambda})$ and $\phi=\phi(\sigma,\theta,\boldsymbol{\lambda})$, respectively. It is not difficult to see that the maximum likelihood estimates of σ and $\boldsymbol{\lambda}$ are the simultaneous solutions of:

$$0=-\frac{n\partial\phi/\partial\sigma}{1+\phi}+\sigma^{-3}\sum_{x_i\le\theta}(\ln x_i-\mu)^2+\frac{M(\ln\theta-\mu)\psi[(\ln\theta-\mu)/\sigma]}{\sigma^2\Phi[(\ln\theta-\mu)/\sigma]}+\frac{m}{\phi}\frac{\partial\phi}{\partial\sigma}-\frac{M}{\sigma}$$

and:

$$0=-\frac{n\partial\phi/\partial\lambda_j}{1+\phi}+\frac{1}{\sigma^2}\frac{\partial\mu}{\partial\lambda_j}\sum_{x_i\le\theta}(\ln x_i-\mu)-\sum_{x_i\le\theta}\frac{1}{\theta}\frac{\partial\theta}{\partial\lambda_j}-\frac{M\psi[(\ln\theta-\mu)/\sigma]}{\sigma\Phi[(\ln\theta-\mu)/\sigma]}\left(\frac{1}{\theta}\frac{\partial\theta}{\partial\lambda_j}-\frac{\partial\mu}{\partial\lambda_j}\right)$$
$$+\sum_{x_i>\theta}\frac{\partial f_2(x_i)/\partial\lambda_j}{f_2(x_i)}+m\frac{\partial F_2(\theta)/\partial\lambda_j}{1-F_2(\theta)}+\frac{m}{\phi}\frac{\partial\phi}{\partial\lambda_j}$$

for $j=1,2,\ldots,q$, where $M=\sum_{i=1}^{n}I(x_i\le\theta)$ and $m=\sum_{i=1}^{n}I(x_i>\theta)$. The log-likelihood function is not continuous with respect to the parameter θ. So, the maximum likelihood estimate of θ must be obtained by segment-wise maximization (e.g., Chapter 6 of Seber and Lee, 2003).

It is clear that the maximum likelihood estimates of $(\sigma,\theta,\boldsymbol{\lambda})$ cannot be obtained in closed form. So, they must be obtained numerically using some quasi-Newton algorithm.

Because of space restrictions, we describe the new composite models for only one of the choices for $f_2(\cdot)$. The details for others including statistical properties and programs in R (R Development Core Team, 2011) used to fit all composite models can be obtained from the corresponding author. Here, we take $f_2(\cdot)$ to correspond to the Burr distribution, resulting in the composite lognormal-Burr model, a piecewise continuous function made up of a truncated lognormal probability density function and a truncated Burr probability

density function joined by threshold, θ. Accordingly, the composite lognormal-Burr model has a form as in (5) with $f_1(x)$ as in (6) and $f_2(x)$ given by:

$$f_2(x) = \frac{\alpha\beta(x/s)^{\beta}}{x\left[1+(x/s)^{\beta}\right]^{\alpha+1}} \tag{7}$$

for $x>0$, $\alpha>0$, $\beta>0$ and $s>0$. Note that this is a scaled Burr probability density function. Note also that (7) reduces to a Pareto probability density function when $\beta=1$.

Applying (3) at θ leads to the following expression for φ:

$$\phi = \frac{(\theta^{\beta}+s^{\beta})\psi[(\ln\theta-\mu)/\sigma]}{\sigma\alpha\beta\theta^{\beta}\Phi[(\ln\theta-\mu)/\sigma]}.$$

Applying the differentiability condition (4) leads to the following simplification:

$$\mu = \ln\theta - \sigma^2\left\{\frac{(\alpha+1)\beta\theta^{\beta}}{\theta^{\beta}+s^{\beta}} - \beta\right\}.$$

Hence, the composite lognormal-Burr model has five unknown parameters $\sigma>0$, $\theta>0$, $\alpha>0$, $\beta>0$ and $s>0$.

4. Results and discussion

Here, we illustrate an application to the Danish fire insurance data (in millions DKK) of 2492 measurements. These data have been formerly applied by Cooray and Ananda (2005) for their composite lognormal-Pareto model. Scollnik (2007) improved Cooray and Ananda's (2005) model by using the Pareto distribution of the second kind to replace the classical Pareto distribution in the second piece of the model. Scollnik (2007) also allowed for flexible mixing weights.

Some summary statistics of the data are: the minimum value is 0.3134, the first quartile is 1.1570, the median is 1.6340, the mean is 3.0630, the third quartile is 2.6450, the maximum value is 263.3000, and the standard deviation is 7.976703.

We fitted the composite lognormal-Pareto and the composite lognormal-Burr models in Section 3 to the data. The method of maximum likelihood was used. Table 1 gives the

Table 1. Parameter estimates for the composite models fitted to the Danish fire insurance data.

Models	Parameter estimates	Log-likelihood	AIC
Lognormal-Pareto	$\hat{\sigma}=0.182(0.012)$	−3860.471	7728.943
	$\hat{\theta}=1.145(0.030)$		
	$\hat{\alpha}=1.563(0.088)$		
	$\hat{s}=0.363(0.125)$		
Lognormal-Burr	$\hat{\sigma}=0.178(0.011)$	−3857.827	7725.654
	$\hat{\theta}=1.093(0.038)$		
	$\hat{\alpha}=0.347(0.161)$		
	$\hat{\beta}=4.111(1.808)$		
	$\hat{s}=0.841(0.107)$		

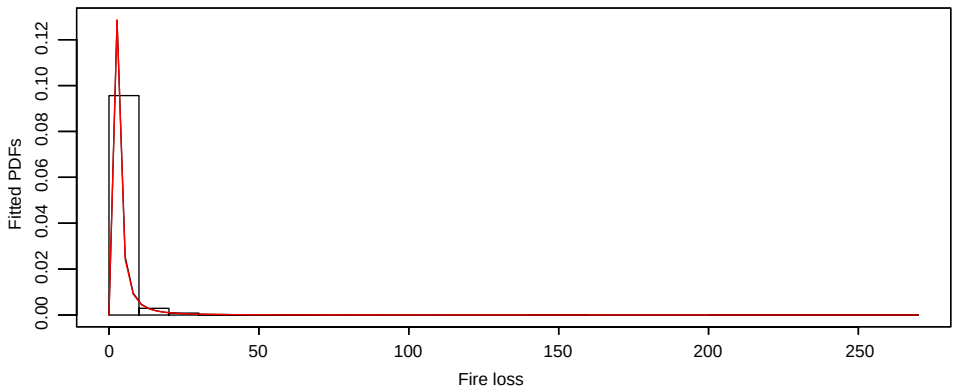

Figure 1. Fitted probability density functions for the Danish fire insurance data: black for the composite lognormal-Pareto model and red for the composite lognormal-Burr model.

parameter estimates, the maximized log-likelihood values and the Akaike information criteria (AIC) values (Akaike, 1974). The numbers within brackets are the standard errors computed by inverting the observed information matrices.

The composite lognormal-Burr model has one more parameter than the composite lognormal-Pareto model. Based on the AIC values, we see that the former gives the better fit.

The conclusion based on the AIC can be verified by means of probability–probability plots, quantile–quantile plots, and density plots. A probability–probability plot consists of plots of the observed probabilities against probabilities predicted by the fitted model. A quantile–quantile plot consists of plots of the observed quantiles against quantiles predicted by the fitted model. A density plot compares the fitted probability density functions of the models with the empirical histogram of the observed data.

The density plot for the two fitted models is shown in Figure 1. The probability–probability and quantile–quantile plots are shown in Figures 2 and 3. The fitted probability density functions appear almost indistinguishable. But they appear to capture the general pattern of the empirical histogram well. The probability–probability plots

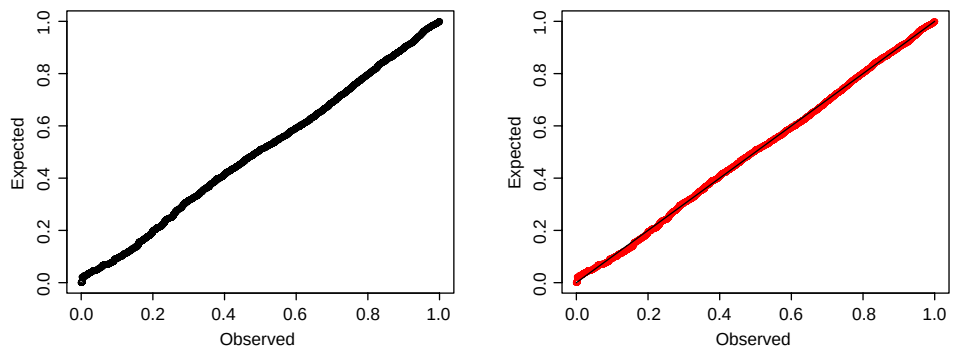

Figure 2. Probability–probability plots for the Danish fire insurance data: black for the composite lognormal-Pareto model and red for the composite lognormal-Burr model.

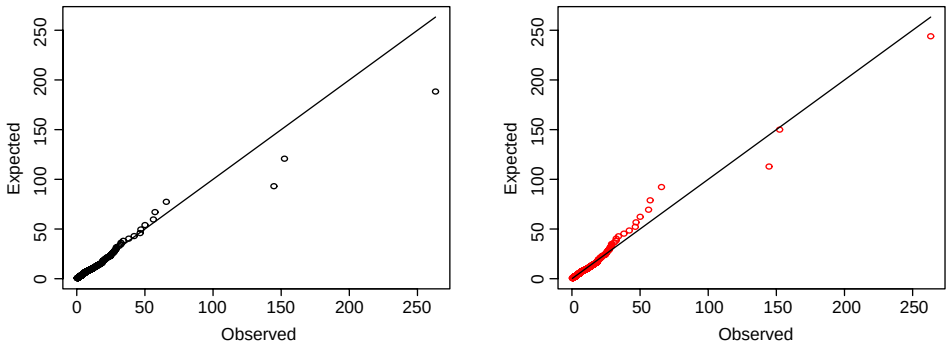

Figure 3. Quantile–quantile plots for the Danish fire insurance data: black for the composite lognormal-Pareto model and red for the composite lognormal-Burr model.

suggest that both models provide an adequate fit to the data. The quantile–quantile plots suggest that the composite lognormal-Burr model has points closer to the diagonal line in the upper tail.

Acknowledgements

The authors would like to thank the Editor and the referee for careful reading and for their comments which greatly improved the paper.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723.

Ausin, M. C., Wiper, M. P. & Lillo, R. E. (2009). Bayesian estimation of finite time ruin probabilities. *Applied Stochastic Models in Business and Industry* **25**, 787–805.

Burnecki, K. & Weron, R. (2004). Modeling the risk process in the XploRe computing environment. *Lecture Notes in Computer Science* **3039**, 868–875.

Carreau, J. & Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes* **12**, 53–76.

Charpentier, A. & Oulidi, A. (2010). Beta kernel quantile estimators of heavy-tailed loss distributions. *Statistics and Computing* **20**, 35–55.

Cooray, K. & Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal* **5**, 321–334.

Drees, H. & Muller, P. (2008). Fitting and validation of a bivariate model for large claims. *Insurance: Mathematics and Economics* **42**, 638–650.

Dutang, C., Goulet, V. & Pigeon, M. (2008). actuar: An R package for actuarial science. *Journal of Statistical Software* **25**, 1–37.

Esmaeili, H. & Kluppelberg, C. (2010). Parameter estimation of a bivariate compound Poisson process. *Insurance: Mathematics and Economics* **47**, 224–233.

Guillotte, S., Perron, F. & Segers, J. (2011). Non-parametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society,* B **73**, 377–406.

Klugman, S. A., Panjer, H. H. & Willmot, G. E. (2004). *Loss models: from data to decisions* (second edition). New York: John Wiley and Sons.

R Development Core Team. (2011). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

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

Seber, G. A. F. & Lee, A. J. (2003). *Linear regression analysis* (second edition). Hoboken, NJ: John Wiley and Sons.