Modeling with Weibull-Pareto Models

Article *in* North American Actuarial Journal · April 2012

DOI: 10.1080/10920277.2012.10590640

David Scollnik
The University of Calgary

Modeling with Weibull-Pareto Models

David P. M. Scollnik* and Chenchen Sun†

Abstract

In this paper we develop several composite Weibull-Pareto models and suggest their use to model loss payments and other forms of actuarial data. These models all comprise a Weibull distribution up to a threshold point, and some form of Pareto distribution thereafter. They are similar in spirit to some composite lognormal-Pareto models that have previously been considered in the literature. All of these models are applied, and their performance compared, in the context of a real-world fire insurance data set.

1. Introduction

In a series of recent papers, various authors have directed attention toward a variety of composite distributional models featuring the Pareto distribution as one of the component distributions. Cooray and Ananda (2005) initiated this current line of research by proposing a composite lognormal-Pareto model for use with loss payments data of the sort arising in the actuarial and insurance industries. Their model is based on a lognormal density up to an unknown threshold value and a two-parameter Pareto density thereafter. They suggested that this model may be appropriate for use by actuaries when faced with smaller data with higher frequencies as well as occasional larger data with lower frequencies. A key feature of this model is that it allows the data to select the most appropriate threshold value. These authors imposed conditions on the model parameters to ensure continuity and differentiability at the threshold point.

Ciumara (2006) developed a composite Weibull-Pareto model for use with actuarial data using the same design approach as in Cooray and Ananda (2005). This composite Weibull-Pareto model is defined and constructed as follows. Let X be a random variable with the probability density function

$$f(x) = \begin{cases} cf_1(x) & \text{if } 0 < x \le \theta \\ cf_2(x) & \text{if } \theta \le x < \infty \end{cases}, \tag{1.1}$$

where c is the normalizing constant, $f_1(x)$ has the form of the Weibull density, and $f_2(x)$ has the form of the basic two-parameter Pareto density. More specifically,

$$f_1(x) = \left(\frac{\tau}{x}\right)\left(\frac{x}{\phi}\right)^{\tau} \exp\left\{-\left(\frac{x}{\phi}\right)^{\tau}\right\}, \qquad x > 0, \tag{1.2}$$

$$f_2(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \qquad x > \theta, \tag{1.3}$$

where θ, α, τ, and ϕ are unknown parameters such that $\theta > 0$, $\alpha > 0$, $\tau > 0$, and $\phi > 0$. To complete this definition of a composite Weibull-Pareto model, Ciumara (2006) imposed continuity and differentiability conditions at θ; that is,

$$f_1(\theta) = f_2(\theta) \quad \text{and} \quad f_1'(\theta) = f_2'(\theta), \tag{1.4}$$

* David P. M. Scollnik, ASA, PhD, is Full Professor in the Department of Mathematics and Statistics at University of Calgary, 2500 University Drive, N.W., Calgary, Alberta, Canada, scollnik@ucalgary.ca.

† Chenchen Sun, MSc, is Statistical Analyst in the Screening Program for Health Promotion, Disease and Injury Prevention at Alberta Health Services, 2210-2nd Street SW, Calgary, Alberta, Canada.

where $f_1'(\theta)$ and $f_2'(\theta)$ are the first derivatives of $f_1(x)$ and $f_2(x)$ evaluated at θ, respectively.

Ciumara (2006) and Cooray (2009) showed that the composite model identified above and satisfying conditions (1.4) does, in fact, exist. Its density function can be written as

$$f(x) = \begin{cases} \dfrac{(k+1)^2\alpha}{(2k+1)x}\left(\dfrac{x}{\theta}\right)^{\alpha k} \exp\left\{-\left(\dfrac{k+1}{k}\right)\left(\dfrac{x}{\theta}\right)^{\alpha k}\right\} & \text{if } 0 < x \leq \theta \\ \left(\dfrac{k+1}{2k+1}\right)\left(\dfrac{\alpha}{x}\right)\left(\dfrac{\theta}{x}\right)^{\alpha} & \text{if } \theta \leq x < \infty \end{cases}. \tag{1.5}$$

The value of k is a known constant given by the positive solution of the equation

$$\exp\left(1 + \frac{1}{k}\right) = k + 1. \tag{1.6}$$

The numerical value of k is approximately equal to 2.8573348. The construction of this composite Weibull-Pareto model also requires that

$$1 + \frac{\tau}{\alpha} = \exp\left(1 + \frac{\alpha}{\tau}\right) = \exp\left\{\left(\frac{\theta}{\phi}\right)^{\tau}\right\} \tag{1.7}$$

so that

$$k = \frac{\tau}{\alpha} \tag{1.8}$$

and

$$c = \frac{k+1}{2k+1} \approx 0.5744638. \tag{1.9}$$

This makes it clear that the model really has only two unknown parameters, $\theta > 0$ and $\alpha > 0$.

Ciumara (2006) concluded that the composite Weibull-Pareto model is an attractive one. Some of the reasons supporting this are that its parameters are fairly easily estimated, it is similar in shape to the composite lognormal-Pareto and lognormal, and it has a heavier tail than the latter. Therefore, this model can be used in situations in which other models underestimate tail probabilities and consequently underestimate the premium to be paid in the case of large losses. However, the model's performance was tested only against simulated data in Ciumara. Teodorescu and Panaitescu (2009) proposed a truncated version of the composite Weibull-Pareto model examined by Ciumara that could be used to introduce deductibles in many non–life insurance contracts (i.e., when the data are left censored) and applied it to simulated data. Preda and Ciumara (2006) compared the performance of the composite Weibull-Pareto model from Ciumara and the composite lognormal-Pareto model from Cooray and Ananda (2005). They argue that both models are appropriate candidates for modeling actuarial data, especially in cases dealing with large loss payments. Preda and Ciumara illustrated the performance of the two models, but the numerical examples in this paper were again based on simulated data. Cooray (2009) reviewed the construction and properties of the composite Weibull-Pareto model and applied it to three real-world data sets. These data sets were generated from three real-world examples involving survival times and demonstrated that the model was of value for modeling unimodal failure rate data. Teodorescu and Vernic (2009) explored properties of composite models in general, with special attention paid to three versions of a composite exponential-Pareto model. They proposed that these models may be of use to actuaries in the same sort of situation identified by Ciumara, that is, where other models might underestimate the tail probability. As with many of the other papers listed above, the numerical examples in Teodorescu and Vernic were based on simulated data.

Scollnik (2007) explored two extensions of the composite lognormal-Pareto model appearing in Cooray and Ananda (2005), leading to two models that were more flexible and thus capable of providing better fits to real data. This paper will adopt and apply the same approach as in Scollnik to the composite Weibull-Pareto model from Ciumara (2006). This is the subject of Sections 2, 3, and 4. In Section 5 the resulting collection of three composite Weibull-Pareto and three composite lognormal-Pareto models will all be applied to a well-known fire insurance data set. Concluding remarks appear in Section 6.

2. THE FIRST COMPOSITE WEIBULL-PARETO MODEL

Ciumara (2006) and Cooray (2009) do not make this next observation, but it is straightforward to show that the density function of the composite Weibull-Pareto model they discuss can be written as

$$f(x) = \begin{cases} \eta \left(\dfrac{1+k}{k}\right) f_1(x) & \text{if } 0 < x \le \theta \\ (1-\eta) f_2(x) & \text{if } \theta \le x < \infty \end{cases}, \tag{2.1}$$

where $\theta > 0$ and $\alpha > 0$, $f_1(x)$ and $f_2(x)$ are the Weibull and Pareto densities given by (1.2) and (1.3), respectively, the model parameters appearing in $f_1(x)$ and $f_2(x)$ satisfy constraints (1.7), k is the positive solution of (1.6) as before, that is, $k \approx 2.8573348$, and

$$\eta = \frac{k}{2k+1} \approx 0.4255362 \tag{2.2}$$

and

$$1 - \eta = c = \frac{k+1}{2k+1} \approx 0.5744638. \tag{2.3}$$

Now, the Weibull model with density (1.2) has distribution function

$$F_1(x) = 1 - \exp\left\{-\left(\frac{x}{\phi}\right)^{\tau}\right\}, \qquad x > 0. \tag{2.4}$$

Making use of (1.6) to (1.8), we have that

$$F_1(\theta) = 1 - \exp\left\{-\left(\frac{\theta}{\phi}\right)^{\tau}\right\} = \frac{k}{k+1}. \tag{2.5}$$

Hence, the Weibull-Pareto density function can be written as

$$f(x) = \begin{cases} \eta \dfrac{1}{F_1(\theta)} f_1(x) & \text{if } 0 < x \le \theta \\ (1-\eta) f_2(x) & \text{if } \theta \le x < \infty \end{cases}, \tag{2.6}$$

where the parameters satisfy all of the same constraints as before. This allows us to recognize and interpret the composite Weibull-Pareto model as a two-component mixture model with fixed and a priori known mixing weights η and $1 - \eta$. (It could also be interpreted as a two-component spliced model as in Klugman et al. [2004], but we prefer the mixture model interpretation for the purposes of this paper.) The first component is a Weibull model truncated above at the value of θ; the second is a two-parameter Pareto model with positive support starting at θ. From (2.5), it is apparent that the truncation point θ is *always* the 74.075th percentile of the original underlying Weibull model with density $f_1(x)$. The continuity and differentiability conditions (1.4) imposed on this model are also restrictive enough that the two mixing weights are *forced* to the constant values $\eta \approx 0.4255362$ and $1 - \eta \approx 0.5744638$, respectively.

The theoretical two-component mixture model of Ciumara (2006) and Cooray (2009), with fixed and a priori known mixing weights (or a two-component spliced model with fixed and a priori known coefficients) η and $1 - \eta$, is a very restrictive one. It says that exactly $100\eta\%$ ($\approx 42.55\%$) of the observations are from a lognormal model truncated above at θ (note: always the 74.075th percentile of the underlying Weibull model) and that exactly $100(1 - \eta)\%$ ($\approx 57.45\%$) of the observations are above θ and in accordance with a certain parameter-restricted Pareto model. Of course, the theoretical model may still be fit to any data set, whether or not it is believed that the observations are truly split in accordance with the theoretically correct proportions. But as any predictions based on the fitted model implicitly assume these fixed mixing weights, this model's default application and use for predictive modeling would seem ill-advised without very careful consideration in many practical situations.

In the next two sections, we will consider two new composite Weibull-Pareto models, both of which relax this restriction of fixed and a priori known mixing weights.

3. THE SECOND COMPOSITE WEIBULL-PARETO MODEL

To recap, the Weibull-Pareto model appearing in Ciumara (2006) and Cooray (2009) may be regarded as a mixture model with a truncated Weibull model below a threshold value θ, and a Pareto model above this point. Further, its mixing weights for the two components are fixed and known a priori. In this section a composite Weibull-Pareto model will be designed as a truncated Weibull and Pareto mixture model with threshold value θ, but with a priori unrestricted mixing weights.

Toward this end, let X be a random variable with the probability density function

$$f(x) = \begin{cases} r\,\dfrac{1}{F_1(\theta)} f_1(x) & \text{if } 0 < x \le \theta \\ (1 - r) f_2(x) & \text{if } \theta \le x < \infty \end{cases}, \tag{3.1}$$

where r is a mixing weight, $0 \le r \le 1$, and $f_1(x)$ and $f_2(x)$ are the Weibull and Pareto densities given by (1.2) and (1.3), respectively. As before, the parameters appearing in $f_1(x)$ and $f_2(x)$ are θ, α, τ, and ϕ, such that $\theta > 0$, $\alpha > 0$, $\tau > 0$, and $\phi > 0$.

If we impose a continuity requirement at θ so that $f(\theta-) = f(\theta+)$, then

$$r = \frac{\dfrac{\alpha}{\tau}}{\dfrac{\left(\dfrac{\theta}{\phi}\right)^{\tau}}{\exp\left\{\left(\dfrac{\theta}{\phi}\right)^{\tau}\right\} - 1} + \dfrac{\alpha}{\tau}}. \tag{3.2}$$

Observe that the mixing weight r is not a fixed and known value as was the mixing weight η in the previous section; rather, r is free to vary, $0 \le r \le 1$, its precise value depending on the particular values of the parameters θ, α, τ, and ϕ.

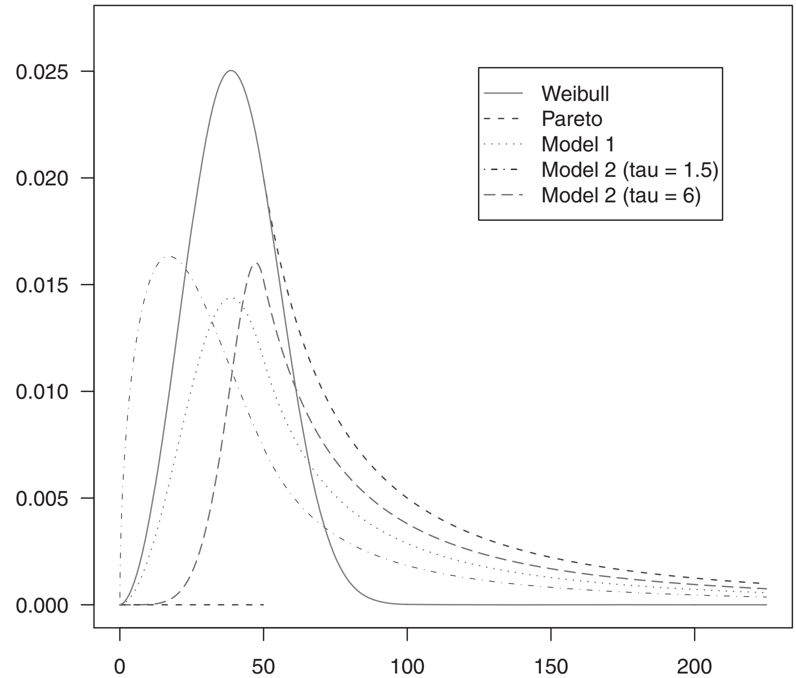
We may ensure that the resulting density function is smooth if we also impose a differentiability condition at θ such that $f'(\theta+) = f'(\theta-)$. This leads to

$$\left(\frac{\theta}{\phi}\right)^{\tau} = \frac{\alpha}{\tau} + 1. \tag{3.3}$$

Then the expression for r simplifies to

$$r = \frac{\alpha \exp\left(\dfrac{\alpha}{\tau} + 1\right) - \alpha}{\alpha \exp\left(\dfrac{\alpha}{\tau} + 1\right) + \tau}, \tag{3.4}$$

Figure 1
Illustrative Weibull, Pareto, and Composite Weibull-Pareto Density Curves



and r still varies in the interval of $0 \leq r \leq 1$. We refer to the model in (3.1), subject to the continuity and differentiability conditions leading to Equations (3.2) and (3.3), as the second composite Weibull-Pareto model to differentiate it from the original (or first) Weibull-Pareto model appearing in Ciumara (2006) and Preda and Ciumara (2006). It reduces to the original composite Weibull-Pareto distribution in the rare instance that $\tau/\alpha = k = 2.857334$.

The second composite Weibull-Pareto model is a model in three unknown parameters, for example, $\theta > 0$, $\alpha > 0$, and $\tau > 0$, with ϕ and r each determined as functions of these. It may be tempting to impose a second derivative condition at θ, that is, $f''(\theta+) = f''(\theta-)$, to further reduce the number of free parameters. However, this leads to

$$\left(\frac{\theta}{\phi}\right)^{\tau} = \left(\frac{\alpha}{\tau} + 1\right)\left(\frac{1}{1 - \tau}\right), \tag{3.5}$$

which can be seen to be inconsistent with (3.3) since $\tau > 0$.

The models discussed in this paper so far are illustrated in Figure 1. The Weibull, Pareto, and first composite Weibull-Pareto model density curves in this figure assume $\theta = 50$ and $\alpha = 1$ (with τ and ϕ satisfying [1.7] and [1.8]). The second composite Weibull-Pareto model also has its density curves plotted for these same values of α and θ, but with $\tau = 1.5$ and $\tau = 6$. Among these models, it is apparent that the Weibull has the thinnest tail and that the second composite model has its peakedness and tail behavior determined by τ, for fixed values of α and θ.

4. THE THIRD COMPOSITE WEIBULL-PARETO MODEL

The third and final composite Weibull-Pareto model to be discussed in this paper will also be developed in terms of a mixture model. As before, a truncated Weibull model is used below some threshold value θ. However, to complete the model this time we will adopt a version of the generalized Pareto model

(GPD) above the threshold value. Theoretical grounds supporting the use of the GPD to model data that exceed high thresholds may be found in McNeil (1997).

The form of generalized Pareto model we adopt has its distribution function given by

$$G(x) = 1 - \left(1 + \frac{x - \theta}{\alpha\beta}\right)^{-\alpha}, \tag{4.1}$$

where $-\infty < \theta < \infty$, $\alpha > 0$, and $\beta > 0$, and the support is $x > \theta$ (see McNeil 1997). The composite model being developed is intended for use with positive loss payments data, and so we will further restrict this model so that $\theta > 0$. The density function for this model can be written as

$$g(x) = \frac{\alpha(\alpha\beta)^{\alpha}}{(\alpha\beta - \theta + x)^{\alpha+1}}, \qquad x > \theta, \tag{4.2}$$

where $\theta > 0$, $\alpha > 0$, and $\beta > 0$. Letting $\lambda = \alpha\beta - \theta$, this can be written as

$$g(x) = \frac{\alpha(\lambda + \theta)^{\alpha}}{(\lambda + x)^{\alpha+1}}, \qquad x > \theta, \tag{4.3}$$

where $\theta > 0$, $\alpha > 0$, and $\lambda > -\theta$. The form of the density function given by (4.3) is popular in the actuarial literature. However, the form given by (4.2) may be a simpler one to work with in some applications, because of its slightly simpler parameter space formulation.

Now, let X be a random variable with the probability density function

$$f(x) = \begin{cases} r\dfrac{1}{F_1(\theta)} f_1(x) & \text{if } 0 < x \le \theta \\ (1 - r)g(x) & \text{if } \theta \le x < \infty \end{cases}, \tag{4.4}$$

where r is a mixing weight, $0 \le r \le 1$, $f_1(x)$ is the Weibull density given by (1.2), and $g(x)$ is the generalized Pareto density given by (4.3). The parameters appearing in $f_1(x)$ and $g(x)$ are θ, α, τ, ϕ, and λ, with $\theta > 0$, $\alpha > 0$, $\tau > 0$, $\phi > 0$, and $\lambda > -\theta$.

If we impose a continuity requirement at θ so that $f(\theta+) = f(\theta-)$, then

$$r = \frac{\dfrac{\alpha}{\tau}}{\dfrac{\lambda + \theta}{\theta}\,\dfrac{\left(\dfrac{\theta}{\phi}\right)^{\tau}}{\exp\left\{\left(\dfrac{\theta}{\phi}\right)^{\tau}\right\} - 1} + \dfrac{\alpha}{\tau}}. \tag{4.5}$$

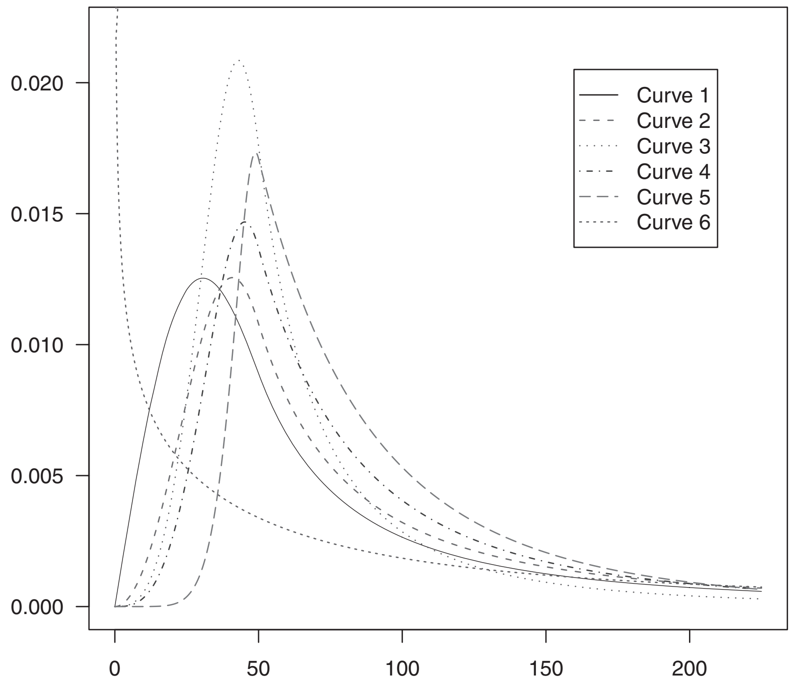
Observe that the mixing weight r is such that $0 \le r \le 1$, with its precise value depending on the values of the other five parameters: θ, α, τ, ϕ, and λ. A differentiability condition at θ can be imposed such that $f'(\theta+) = f'(\theta-)$ to ensure a smooth density curve. This results in the restriction

$$\left(\frac{\theta}{\phi}\right)^{\tau} = \frac{\alpha\theta - \lambda}{(\lambda + \theta)\tau} + 1. \tag{4.6}$$

We will refer to the model in Equation (4.4), subject to the continuity and differentiability conditions leading to (4.5) and (4.6), as the third composite Weibull-Pareto model. It is a model with four unknown parameters, $\theta > 0$, $\alpha > 0$, $\tau > 0$, and $\lambda > -\theta$, with ϕ and r determined as a function of these. It reduces to the second composite Weibull-Pareto model when $\lambda = 0$.

Unlike the situation with the second composite model in the previous section, it is possible to impose a second derivative condition that $f''(\theta+) = f''(\theta-)$ to reduce the number of free parameters without

Figure 2
Illustrative Density Curves for the Third Composite Weibull-Pareto Model



running into an inconsistent set of conditions. This second derivative condition leads to the additional requirement

$$\left(\frac{\theta}{\phi}\right)^{\tau} = \frac{\theta^2(\alpha + 1)}{(\lambda + \theta)^2(1 - \tau)\tau} - \frac{1}{\tau}. \tag{4.7}$$

This does not conflict with (4.6), and it does reduce the number of free parameters by one by allowing, for example, α to be determined as a function of the others. In this case, and by making use of (4.6), we have

$$\alpha = \frac{\theta^2 - (\tau(\lambda + \theta) + \theta)(1 - \tau)(\lambda + \theta)}{\theta(1 - \tau)(\lambda + \theta) - \theta^2}. \tag{4.8}$$

However, this comes at the cost of significantly restricting the shape of the model's density curve—so much so, that we have not found it to be a particularly useful model thus far. Note that α determined on the basis of (4.8) must still satisfy both (4.6) and (4.7), with $\alpha > 0$ and $\phi > 0$.

Six illustrative density curves for the third composite Weibull-Pareto model are presented in Figure 2. The parameters used for these curves are $\alpha = 1$, $\theta = 50$, $\tau = 2$, $\lambda = 10$ (curve 1); $\alpha = 1$, $\theta = 50$, $\tau = 3$, $\lambda = 10$ (curve 2); $\alpha = 2$, $\theta = 50$, $\tau = 4$, $\lambda = 10$ (curve 3); $\alpha = 2$, $\theta = 50$, $\tau = 4$, $\lambda = 50$ (curve 4); $\alpha = 4$, $\theta = 50$, $\tau = 8$, $\lambda = 140$ (curve 5); $\alpha = 0.5$, $\theta = 50$, $\tau = 0.75$, $\lambda = 50$ (curve 6). The sixth of these curves was constrained by the second derivative condition (4.7).

5. Illustrative Example

To illustrate the usefulness and value of the composite Weibull-Pareto models presented in this paper, we apply them to a classic insurance data set. This is the set of Danish data on 2,492 fire insurance losses in Danish kroner (DKr) from the years 1980 to 1990, inclusive, adjusted to reflect 1985 values.

The adjusted values range from (in millions of DKr) 0.3134041 to 263.2503660. The data set may be found in the ''SMPracticals'' add-on package for R, available from the CRAN website cran.r-project.org.

This Danish fire insurance data set was previously considered by McNeil (1997), who analyzed the 2,156 losses exceeding 1 (i.e., 1 million DKr). McNeil found the tail of the lognormal model to be a little too thin to capture the behavior of the largest losses, and the ordinary Pareto overestimated probabilities of large losses and led to unrealistically high answers. McNeil found that the GPD was somewhere between the lognormal and Pareto models in the tail area and was a good explanatory model for the highest losses.

More recently Cooray and Ananda (2005) applied six different parametric models fit using the method of maximum likelihood (ML) to the complete Danish fire insurance data set. These models were the lognormal, ordinary Pareto, inverse Gaussian, gamma, Weibull, and original (first) composite lognormal-Pareto. Scollnik (2007) fit a second and third version of composite lognormal-Pareto model to a data set consisting of the uppermost 2,168 values (i.e., values above 1) in the original Danish fire insurance data set combined with a proxy set of 324 simulated values below 1. See Cooray and Ananda and Scollnik for the definitions and details of these composite lognormal-Pareto models. In the present example, we apply the composite Weibull-Pareto models developed in this paper to the complete Danish fire insurance data set. At the same time, we take this opportunity to update the analysis in Scollnik by applying its composite lognormal-Pareto models to the complete Danish data set (rather than the combined data set used in that paper). For the sake of comparison, results for the five other two parameter models considered in Cooray and Ananda will also be presented.

Parameter estimation for all the models will proceed using ML (implemented using the ''nlm'' function in R). Goodness-of-fit will be measured and compared in various different ways, including on the basis of the value of the negative log-likelihood (NLL) at the values of the ML estimates, and the Akaike information criterion (AIC) equal to twice the NLL at the values of the ML estimates plus twice the number of parameters. The AIC thus adjusts for the number of parameters in a model and allows models with differing numbers of parameters to be more fairly compared. For both goodness-of-fit measures, smaller values are good when comparing across models.

The ML parameter estimates (along with their estimated asymptotic variances in brackets) for the fitted models, along with the corresponding values of the NLL and AIC and additional information discussed below, are given in Table 1.

The NLL and AIC results indicate that the composite models provide a better fit than do the simpler lognormal, Pareto, inverse Gaussian, gamma, or Weibull models, even when allowances are made for the larger number of parameters in some of the composite models. These results for the composite lognormal-Pareto models are consistent with what was reported in Scollnik (2007). It is apparent that the second and third composite Weibull-Pareto models provide appreciably better fits than any of the other models, with the third Weibull-Pareto model providing a better fit than the second. The main portions of the density curves for the fitted composite models are presented in Figures 3 and 4, with an empirical histogram of the Danish data overlaid on top.

Cooray and Ananda (2005) employed the χ^2 goodness-of-fit test to compare their model against alternatives. The use of this test statistic was examined in Scollnik (2007). The discussion of the illustrative example in Scollnik (pp. 29–30) made clear that the χ^2 goodness-of-fit statistic was not a good basis to use for comparing models in the context of the Danish data set (i.e., ''with such a large and highly skewed data set as the one under consideration, the chi-square goodness of fit test is not a particularly good tool to use for comparing models unless there is some objective and meaningful way to determine the class limits, that is also relevant to the problem at hand''). For this reason we do not consider it further in this paper.

Upon the recommendation of an anonymous reviewer, the fits of the different models appearing in Table 1 were also measured and compared on the basis of the Bayesian information criterion (BIC), also known as the Schwarz information criterion (Schwarz 1978). The BIC is calculated as twice the NLL at the values of the ML estimates plus $k\ln(n)$, where k is the number of free parameters to be estimated and n is the sample size. Smaller values of the BIC are good when comparing across models,

Table 1

Estimated Values of Fitted Models for Danish Fire Insurance Loss Data

Distribution	Parameter MLEs (Asymptotic Variances)	NLL	AIC	BIC
Lognormal	$\hat{\mu}$ = 0.671853 (0.000215) $\hat{\sigma}$ = 0.732316 (0.000108)	4,433.891	8,871.782	8,883.424
Pareto	$\hat{\theta}$ = 0.313404 $\hat{\alpha}$ = 0.545816 (0.000120)	5,675.094	11,354.184	11,365.83
Inverse Gaussian	$\hat{\mu}$ = 3.062697 (0.003376) $\hat{\theta}$ = 3.417103 (0.009373)	4,516.307	9,036.614	9,048.256
Gamma	$\hat{\theta}$ = 2.434590 (0.005733) $\hat{\alpha}$ = 1.257994 (0.001025)	5,243.027	10,490.054	10,501.70
Weibull	$\hat{\theta}$ = 2.952494 (0.004411) $\hat{\tau}$ = 0.947586 (0.000127)	5,270.470	10,544.940	10,556.58
Lognormal-Pareto (first model)	$\hat{\theta}$ = 1.385128 (0.000182) $\hat{\alpha}$ = 1.436332 (0.000729)	3,877.844	7,759.688	7,771.33
Lognormal-Pareto (second model)	$\hat{\theta}$ = 1.207430 (0.000892) $\hat{\alpha}$ = 1.328223 (0.000996) $\hat{\sigma}$ = 0.196517 (0.000136)	3,865.864	7,737.728	7,755.19
Lognormal-Pareto (third model)	$\hat{\theta}$ = 1.144585 (0.000886) $\hat{\alpha}$ = 1.563127 (0.007787) $\hat{\sigma}$ = 0.182288 (0.000133) $\hat{\lambda}$ = 0.363363 (0.015606)	3,860.471	7,728.942	7,752.225
Weibull-Pareto (first model)	$\hat{\theta}$ = 1.447231 (0.000221) $\hat{\alpha}$ = 1.564950 (0.000910)	3,959.005	7,922.01	7,933.652
Weibull-Pareto (second model)	$\hat{\theta}$ = 1.002988 (0.000190) $\hat{\alpha}$ = 1.261474 (0.000748) $\hat{\tau}$ = 14.033955 (1.201833)	3,840.376	7,686.752	7,704.215
Weibull-Pareto (third model)	$\hat{\theta}$ = 0.971693 (0.000219) $\hat{\alpha}$ = 1.652557 (0.007998) $\hat{\tau}$ = 15.34259 (1.932738) $\hat{\lambda}$ = 0.560429 (0.014976)	3,823.698	7,655.396	7,678.68

and the BIC tends to favor more parsimonious models than does the AIC. Values of the BIC are reported in Table 1. Once again, it is apparent that the second and third composite Weibull-Pareto models provide better fits than any of the other models, with the third Weibull-Pareto model providing a better fit than the second. Note that the BIC values for the second and third composite Weibull-Pareto models are 7,704.215 and 7,678.68, respectively, yielding a difference between the two of 25.535. According to Kass and Raftery (1995, p. 777), a difference of greater than 10 is very strong evidence against the model with the larger BIC value (note that this comes about as the BIC gives a rough approximation to the logarithm of the Bayes factor for comparing two models). Hence, the observed difference of 25.535 corresponds to very strong support for the use of the third composite Weibull-Pareto model in this illustrative example.

The same reviewer mentioned above also recommended a test of the null hypothesis $\lambda = 0$ versus the alternative in the context of the third composite Weibull-Pareto model. In other words, a test of the nested second versus third composite Weibull-Pareto models. A likelihood ratio test can be used to accomplish this. The test statistic is given by

$$D = -2\ln\left(\frac{\text{likelihood for null model}}{\text{likelihood for alternative model}}\right).$$

Figure 3
Fitted Density Curves for the Composite Lognormal-Pareto and Weibull-Pareto Models

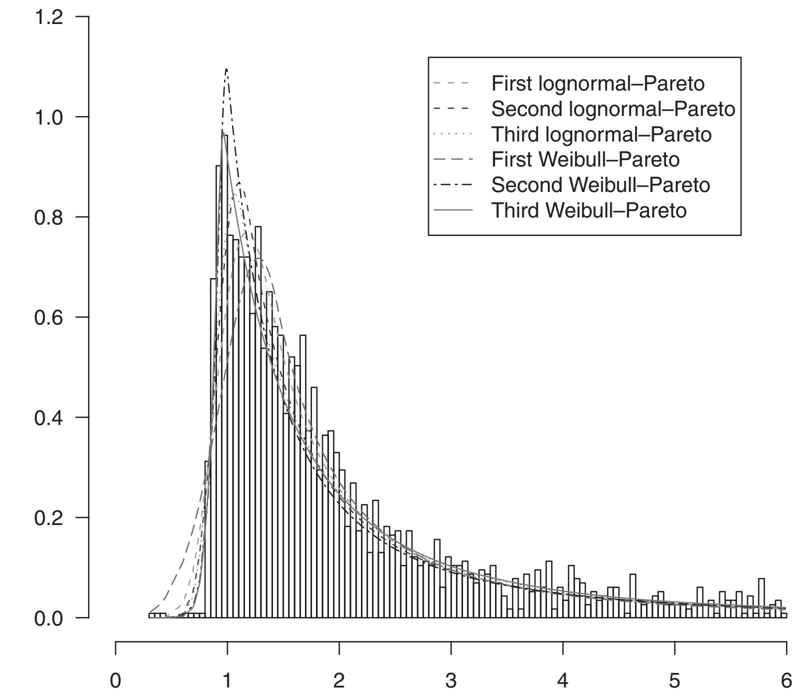


Figure 4
Fitted Density Curves for the Composite Lognormal-Pareto and Weibull-Pareto Models

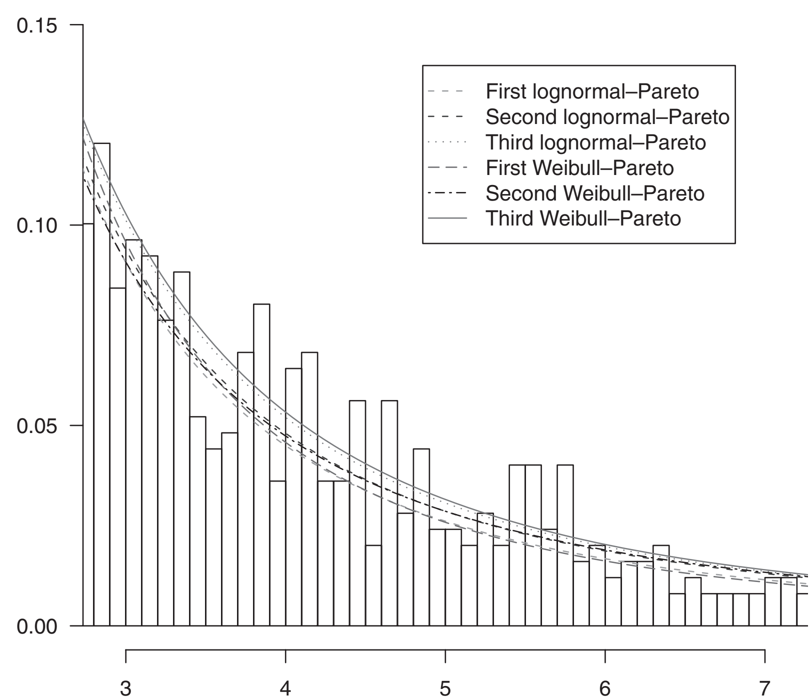


Table 2

Empirical and Fitted Lognormal-Pareto Model Quantiles

	Empirical		Fitted Lognormal-Pareto Models		
Quantile	Type 7	Type 8	First Model	Second Model	Third Model
50%	1.634	1.634	1.587	1.572	1.611
75%	2.645	2.645	2.571	2.650	2.712
90%	5.080	5.086	4.866	5.282	5.164
95%	8.406	8.458	7.884	8.902	8.249
97.5%	14.395	14.496	12.775	15.001	13.054
99%	24.614	24.870	24.177	29.903	23.750
99.5%	32.431	32.811	39.173	50.391	37.207
99.9%	105.893	146.010	120.121	169.277	104.835
99.95%	150.509	199.020	194.626	285.259	163.540
99.99%	235.641	263.250	596.811	958.261	458.572

The exact distribution of this test statistic is not available. However, Wilks's theorem suggests that as the sample size n approaches ∞, this test statistic will be asymptotically χ^2 distributed with degrees of freedom ν equal to the difference in the dimensionality of the two models under consideration. In this case, $D = 2\ (3{,}840.376 - 3{,}823.698) = 33.356$, $\nu = 4 - 3 = 1$, and $n = 2{,}492$, so the p value of the test is nearly 0. This is very strong support to reject the second composite Weibull-Pareto model in favor of the third.

Given in Tables 2 and 3 are a variety of empirical quantiles for the data along with the corresponding theoretical quantiles from the six fitted composite models. It is instructive to examine how closely the theoretical tail quantiles from the fitted models match the empirical quantiles associated with the data set. Of course, a number of different and competing methods are available with which to calculate empirical quantiles, and the empirical quantiles, particularly in the extreme tails, can be very dependent on the method used. Hyndman and Fan (1996) describe nine quantile algorithms. For the sake of this illustration, Tables 2 and 3 present the empirical quantiles calculated using two of these algorithms. The Type 7 empirical quantiles are calculated using the default method used in R. The Type 8 empirical quantiles are calculated using the method recommended by Hyndman and Fan (1996). Overall, however, the third composite Weibull-Pareto model does provide a very good fit regardless of which of these empirical quantile algorithms is used. We remark that not too much should be read into the precise values of the 99.9%, 99.95%, and 99.99% empirical quantiles. The empirical quantiles in this extreme portion of the tail describe events that occur 1 in 1,000 times or less and are based on a highly skewed data set of only 2,492 observations. Estimates of the 99.99 % empirical quantile, for example, range at least between approximately 235 and 263, depending upon the particular definition of empirical

Table 3

Empirical and Fitted Weibull-Pareto Model Quantiles

	Empirical		Fitted Weibull-Pareto Models		
Quantile	Type 7	Type 8	First Model	Second Model	Third Model
50%	1.634	1.634	1.581	1.542	1.615
75%	2.645	2.645	2.463	2.671	2.749
90%	5.080	5.086	4.423	5.522	5.201
95%	8.406	8.458	6.888	9.566	8.203
97.5%	14.395	14.496	10.726	16.571	12.770
99%	24.614	24.870	19.262	34.262	22.648
99.5%	32.431	32.811	29.996	59.353	34.742
99.9%	105.893	146.010	83.888	212.586	92.931
99.95%	150.509	199.020	130.635	368.271	141.649
99.99%	235.641	263.250	365.342	1,319.032	376.050

quantile that is used in the calculation (as seen in Table 3). Note that the five largest observations in the Danish fire insurance loss data set are 57.411, 65.707, 144.658, 152.413, and 263.250.

6. CONCLUDING REMARKS

The maximum likelihood estimation, statistical calculations, tests, and graphics in this paper were performed using R. The ML estimates and estimated asymptotic variances were determined with the aid of the ''nlm'' function in R. Essentially, use of the nlm function simply requires the user to code the likelihood function and set initial values for the unknown parameters. The R software environment for statistical computing and graphics, along with various add-on packages for it, can be obtained from www.r-project.org and cran.r-project.org.

In this paper we reviewed three composite Weibull-Pareto models for use with the form of very highly positively skewed data that often arise in the actuarial and insurance industries. These three models are similar in spirit to three composite lognormal-Pareto models, as developed in papers by Cooray and Ananda (2005) and Scollnik (2007). All of these composite models, along with five simpler and more traditional loss models, were applied to a classic and well-studied insurance data set. Among all of the models, the second and third composite Weibull-Pareto provided the best fit.

An anonymous reviewer suggested that the composite models provide an indirect way to estimate the Pareto index α and that a comparison with standard approaches used to determine it (e.g., Hill plots) could be interesting so as to check whether the estimated α depends on the non-Pareto component of the composite model. However, it is already clear that the estimated α in any of the composite models will depend on the non-Pareto component. This is apparent from Table 1 as the Pareto, lognormal-Pareto, and Weibull-Pareto models all have different estimated values of α. Note that the Hill plot was used in the case of the Danish data set in Resnick (1997), as were a number of additional methods (e.g., alternative Hill plot, smoothed Hill plot, alternative smoothed Hill plot, and dynamic and static QQ plots). Based on an amalgam of these methods, Resnick (p. 144) settled on an estimate of $\alpha = 1.4$. Significantly, Resnick described the use of the Hill plot as often ''more guesswork than science'' (p. 140) and expressed concern over the sensitivity of this and various other estimation and fitting methods to the choice of threshold or the choice of the number of order statistics used in estimation (p. 150). The use of the composite lognormal-Pareto and Weibull-Pareto models does eliminate this sort of guesswork and subjectivity of threshold selection from the estimation and fitting process. However, it is important to bear in mind that the composite models in this paper were developed for use with data sets containing both small and large values, and to provide a good fit of the model over the entire range of the data (i.e., not just in the tail).

Recall that all of the composite models discussed in this paper have assumed a single threshold value θ applying uniformly to the whole data set. Pigeon and Denuit (2011) consider a situation in which heterogeneity of the threshold parameter in a composite lognormal-Pareto model is allowed in order so that it can vary among observations. Specifically, these authors show how a gamma or a lognormal mixing distribution can be imposed on θ in the context of the second lognormal-Pareto model in Scollnik (2007). They also applied one of the resulting mixture models to the Danish fire insurance loss data set. This model's associated NLL and AIC values were 3,860 and 7,728, respectively. These are about the same as the NLL and AIC values for the third lognormal-Pareto model, although worse than those for the second and third composite Weibull-Pareto models. It might be interesting and worthwhile to explore the use of mixing distributions on θ in the context of the composite Weibull-Pareto models.

ACKNOWLEDGMENTS

The authors thank Dr. Enrique de Alba (coeditor) and anonymous reviewers for their valuable suggestions that have improved this paper. This paper was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

CIUMARA, R. 2006. An Actuarial Model Based on the Composite Weibull-Pareto Distribution. *Mathematical Report–Bucharest* 8(4): 401–414.

COORAY, K. 2009. The Weibull-Pareto Composite Family with Applications to the Analysis of Unimodal Failure Rate Data. *Communications in Statistics–Theory and Methods* 38: 1901–1915.

COORAY, K., AND M. M. A. ANANDA. 2005. Modeling Actuarial Data with a Composite Lognormal-Pareto Model. *Scandinavian Actuarial Journal* 5: 321–334.

HYNDMAN, R. J., AND Y. FAN. 1996. Sample Quantiles in Statistical Packages. *American Statistician* 50: 361–365.

KASS, R., AND A. RAFTERY. 1995. Bayes Factors. *Journal of the American Statistical Association* 90: 773–795.

KLUGMAN, S. A., H. H. PANJER, AND G. E. WILLMOT. 2004. *Loss Models: From Data to Decisions.* 2nd ed. New York: John Wiley & Sons.

MCNEIL, A. J. 1997. Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. *ASTIN Bulletin* 27(1): 117–137.

PIGEON, M., AND M. DENUIT. 2011. Composite Lognormal-Pareto Model with Random Threshold. *Scandinavian Actuarial Journal* 3: 177–192.

PREDA, V., AND R. CIUMARA. 2006. On Composite Models: Weibull-Pareto and Lognormal-Pareto. A Comparative Study. *Romanian Journal of Economic Forecasting* 7(2): 32–46.

RESNICK, S. I. 1997. Discussion of the Danish Data on Large Fire Insurance Losses. *ASTIN Bulletin* 27(1): 139–151.

SCHWARZ, G. 1978. Estimating the Dimension of a Model. *Annals of Statistics* 6: 461–464.

SCOLLNIK, D. P. M. 2007. On Composite Lognormal-Pareto Models. *Scandinavian Actuarial Journal* 1: 20–33.

TEODORESCU, S., AND E. PANAITESCU. 2009. On the Truncated Composite Weibull-Pareto Model. *Mathematical Reports–Bucharest* 11(61): 259–273.

TEODORESCU, S., AND R. VERNIC. 2009. Some Composite Exponential-Pareto Models for Actuarial Work. *Romanian Journal of Economic Forecasting* 12(4): 82–100.

Discussions on this paper can be submitted until October 1, 2012. The authors reserve the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.