



Modeling loss data using composite models

S.A. Abu Bakar^{a,*}, N.A. Hamzah^a, M. Maghsoudi^a, S. Nadarajah^b

^a Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia

^b School of Mathematics, University of Manchester, Manchester M13 9PL, UK

ARTICLE INFO

Article history:

Received September 2013

Received in revised form

August 2014

Accepted 24 August 2014

Available online 4 September 2014

Keywords:

Allocated loss adjustment expenses data

Composite Weibull models

Heavy tailed distributions

Danish fire insurance data

Risk measures

ABSTRACT

We develop several new composite models based on the Weibull distribution for heavy tailed insurance loss data. The composite model assumes different weighted distributions for the head and tail of the distribution and several such models have been introduced in the literature for modeling insurance loss data. For each model proposed in this paper, we specify two parameters as a function of the remaining parameters. These models are fitted to two real insurance loss data sets and their goodness-of-fit is tested. We also present an application to risk measurements and compare the suitability of the models to empirical results.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Modeling insurance loss data of a unimodal type with a heavy tail has been an interesting topic for actuaries. Distributions that can mimic the heavy tail of the insurance loss data are crucial to sufficiently provide a good estimate of the associated business risk level. Several heavy tailed models have been discussed in the literature including the Pareto, lognormal, Weibull and gamma models. Besides modeling claim losses in insurance, the application of these models is diverse, for instance, to model financial returns and file sizes on network servers, see Resnick (2007). The recent focus for modeling such data has been directed towards composite models following a paper by Cooray and Ananda (2005). These models are made up by piecing together two weighted distributions at a specified threshold. For simplicity, we refer to the distribution up to the threshold and the one beyond it as the head and tail of the distribution, respectively.

In general, the composite model discussed above takes the following form

$$f(x) = \begin{cases} a_1 f_1^*(x), & \text{if } -\infty < x \leq \theta, \\ a_2 f_2^*(x), & \text{if } \theta < x < \infty, \end{cases} \quad (1)$$

where a_i , $i = 1, 2$ are the weights and $f_i^*(x)$, $i = 1, 2$ are the truncated probability density functions (pdfs) of the composite model.

Note, however, that the form in (1) is not continuous and smooth in general.

Cooray and Ananda (2005) models the head and tail by weighted lognormal and Pareto distributions, respectively, and showed better fit to a real skewed loss data than several standard models. They also employ the continuity and differentiability conditions at the threshold, θ , to ensure that the resulting models are both continuous and smooth. Scollnik (2007) improved the composite lognormal–Pareto model by allowing flexible mixing weights, replacing a constant weight applied by Cooray and Ananda (2005), resulting in a better fit to the loss data. He reiterates this fact using the composite Weibull–Pareto model in Scollnik and Sun (2012). The mixing weights for the models developed in Cooray and Ananda (2005) can be restrictive, so we shall not pursue such an approach. The most recent findings by Nadarajah and Bakar (2014) suggest that loss data can be better modeled by a composite lognormal model with the tail of a Burr distribution than one with the tail of a Pareto distribution. All these authors illustrate their findings using the well known Danish fire insurance loss data.

The model proposed by Scollnik and Sun (2012) allowing a flexible mixing weight, r , can be expressed as follows

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

where $f_1^*(x) = \frac{f_1(x)}{F_1(\theta)}$ and $f_2^*(x) = \frac{f_2(x)}{1-F_2(\theta)}$ are the truncated Weibull and truncated Pareto of the second kind pdfs, respectively, where $f_1(x)$ and $f_2(x)$ are the Weibull and Pareto of the second kind pdfs (the latter also known as Lomax pdf) specified by

$$f_1(x) = \frac{\alpha}{\lambda} \left(\frac{x}{\lambda} \right)^{\alpha-1} \exp \left[- \left(\frac{x}{\lambda} \right)^{\alpha} \right], \quad x > 0$$

* Corresponding author.

E-mail address: saab@um.edu.my (S.A. Abu Bakar).

and

$$f_2(x) = \frac{\beta\sigma^\beta}{(\sigma+x)^{\beta+1}}, \quad x > 0,$$

respectively, and $F_i(x)$, $i = 1, 2$ are the corresponding cumulative distribution functions (cdfs). Applying the continuity condition, $f(\theta-) = f(\theta+)$, r has the following general form as given in Nadarajah and Bakar (2014)

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

where $0 \leq r \leq 1$. The mixing weight of the composite Weibull–Pareto model proposed by Scollnik and Sun (2012) is given by

$$r = \frac{\frac{\beta}{\alpha}}{\frac{\sigma+\theta}{\theta} \frac{\frac{\theta}{\lambda}}{\exp(\frac{\theta}{\lambda})-1} + \frac{\beta}{\alpha}}.$$

Applying the differentiability condition at θ , $f'(\theta-) = f'(\theta+)$, Scollnik and Sun (2012) showed that the parameters of the model can be reduced, that is, by expressing one of the parameters in terms of the others

$$\lambda = \frac{\theta}{\left[\frac{\beta\theta-\sigma}{\alpha(\sigma+\theta)} - 1 \right]^{\frac{1}{\alpha}}}.$$

The resulting composite model is therefore continuous and smooth on the space $x > 0$. In brief, the composite Weibull–Pareto model is defined by specifying the mixing weight, r , and the scale parameter of the Weibull distribution, λ .

In this paper, we propose a new approach to develop a composite model for any two pieces of distributions. Several new composite models with the head based on the Weibull distribution and the tail belonging to a family of transformed beta distributions are proposed. In Section 2, we describe the method of constructing the composite model by specifying the mixing weights and the threshold in terms of other parameters of the model. All the new models together with some standard distributions are fitted to two real data and the results are presented in Section 3. In addition, some application of the models to risk measurements are presented in Section 4. Finally, some conclusions are drawn in Section 5.

2. Overview of statistical methodology

We now describe the methodological procedures of constructing a composite model with any two pieces of distributions. Since the loss model assumes positive values, the range of (1) is altered accordingly. The pdf (1) with $a_i = \frac{1}{1+\phi}$ is given by:

$$f(x) = \begin{cases} \frac{1}{1+\phi} f_1^*(x), & \text{if } 0 < x \leq \theta, \\ \frac{\phi}{1+\phi} f_2^*(x), & \text{if } \theta < x < \infty. \end{cases} \quad (2)$$

Although (2) is a legitimate pdf, it is not continuous and smooth in general. The followings account for these criteria:

- (i) Continuity condition: The head is joined to the tail of the distribution at a threshold θ . This is done as follows:

$$\lim_{x \rightarrow \theta^-} f(x) = \lim_{x \rightarrow \theta^+} f(x),$$

where θ is regarded as the parameter of the model.

- (ii) Differentiability condition: The following ensures that the resulting distribution is smooth:

$$\lim_{x \rightarrow \theta^-} \frac{df(x)}{dx} = \lim_{x \rightarrow \theta^+} \frac{df(x)}{dx}.$$

The initial pdfs $f_i(x)$, $i = 1, 2$ must be continuous at the threshold, θ , in order for this condition to hold.

- (iii) Mixing weights: Note that the mixing weights of (1) introduced earlier, that is, a_i , $i = 1, 2$ are parameters of the model. However, applying the continuity condition, they can be expressed in terms of other parameters of the model. In order to serve as a legitimate pdf, it is necessary that the summation of the mixing weights equals 1. For convenience, we choose $a_1 = \frac{1}{1+\phi}$ and $a_2 = \frac{\phi}{1+\phi}$ for $\phi > 0$, as presented in (2). Applying the continuity conditions to (2), we can see that ϕ has the following general form:

$$\phi = -\frac{\frac{d}{d\theta} \ln F_1(\theta)}{\frac{d}{d\theta} \ln \bar{F}_2(\theta)}, \quad (3)$$

where $\bar{F}_i(\cdot) = 1 - F_i(\cdot)$, $i = 1, 2$.

- (iv) Threshold: The threshold, θ , is also regarded as one of the model parameters. However, in this new approach, it will be expressed as a function of other parameters of the composite model. Its value can be obtained by solving the following equation

$$\frac{d}{d\theta} \ln \left[\frac{f_1(\theta)}{f_2(\theta)} \right] = 0.$$

Although a closed form of θ may not exist, the equation can be solved for using any numerical method.

2.1. Several new composite models

The aforementioned statistical procedures provide a general rule to develop a composite model with any two pieces of distributions. We will now specify several new composite models using these procedures, paying special attention to the composite Weibull–Burr model. The other composite models proposed can be constructed similarly. The models proposed are restricted to those with a weighted Weibull distribution forming the head and the tail belonging to a family of transformed beta distributions. The properties of the transformed beta family distributions can be found in Klugman et al. (2012). We consider eight of them here consisting of two or three parameters: the Burr, the loglogistic, the paralogistic, the generalized Pareto, the Pareto, the inverse Burr, the inverse Pareto and the inverse paralogistic distributions. They are popular in modeling loss severity and have been considered in Klugman and Parsa (1999).

The pdf of a transformed beta distribution with three shape parameters, β , γ and τ , and a scale parameter, σ , is given by

$$f(x) = \frac{1}{B(\beta, \tau)} \frac{\gamma(x/\sigma)^{\gamma\tau}}{x [1 + (x/\sigma)^{\gamma}]^{\beta+\tau}}, \quad x > 0, \quad (4)$$

where $B(\cdot)$ is the beta function defined by

$$B(\beta, \tau) = \int_0^1 s^{\beta-1} (1-s)^{\tau-1} ds.$$

All the parameters are greater than zero. The tail of the composite Weibull–Pareto model proposed by Scollnik and Sun (2012) belongs to the transformed beta family for $\gamma = \tau = 1$. In the next section, we will refit the composite Weibull–Pareto model for comparisons.

The composite Weibull–Burr model has the tail specified by (4) with $\tau = 1$. That is,

$$f_2(x) = \frac{\beta\gamma(x/\sigma)^\gamma}{x [1 + (x/\sigma)^\gamma]^{\beta+1}}, \quad x > 0.$$

The mixing weights are obtained using (3) as follows:

$$\phi = \frac{\alpha \left(\frac{\theta}{\lambda} \right)^\alpha \left[\left(\frac{\theta}{\sigma} \right)^\gamma + 1 \right] \left(\frac{\sigma}{\theta} \right)^\gamma}{\beta\gamma \left[e^{\left(\frac{\theta}{\lambda} \right)^\alpha} - 1 \right]}.$$

The threshold, θ , although does not have an explicit expression, can be solved numerically. Thus, the composite Weibull–Burr model consists of five unknown parameters, α , λ , β , γ and σ , all greater than zero.

The other new composite models (namely, the composite Weibull–Loglogistic model, the composite Weibull–Paralogistic model, the composite Weibull–Generalized Pareto model, the composite Weibull–Inverse Burr model and the composite Weibull–Inverse Paralogistic model) can be developed in a similar manner. The family of beta distributions proposed for the tail Klugman et al. (2012) is implemented in many statistical softwares. In the R statistical environment, the *actuar* package provides all the functions related to the beta family of distributions and the *stats* package provides all the functions related to the Weibull distribution. Hence, the new composite models can be easily set up in the R environment.

3. Model selection

In this section, several goodness-of-fit measures are used to measure the appropriateness of the fitted models. These measures also assess the validity of the models considered. A lower value of the measures gives an indication of the preferred model. The measures we consider are the Negative Log-Likelihood (NLL), Akaike Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC).

The Negative Log-Likelihood is the minimized value of the Negative Log-Likelihood function. Let $L(\theta)$ denote the maximized log-likelihood function of a model, then the Negative Log-Likelihood is defined by

$$NLL = -L(\theta).$$

NLL can be used to compare models with the same number of parameters. A lower value indicates a better fit of the model to the data. Comparison of models with differing number of parameters is discussed next.

The Akaike Information Criterion measures the relative goodness-of-fit of a statistical model. It was developed by Hirotugu Akaike, under the name of “an information criterion”, and was first published by Akaike (1974). It has the advantages of comparing both the nested and non-nested models as well as models with varying parameters and thus avoids multiple testing issues. AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. It is defined by

$$AIC = 2k - 2L(\theta),$$

where k is the number of parameters of the fitted model. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value.

The fit of all the models are also compared using the Schwarz's Bayesian Criterion (SBC), also known as the Bayesian Information Criterion. It was introduced by Schwarz (1978). It penalizes model complexity more than does the AIC. SBC is defined by:

$$SBC = k \ln(n) - 2L(\theta),$$

where n is the number of observations. Similar to the two measures above, a smaller value of the SBC indicates a better fit across all the models compared.

3.1. Insurance loss data

Two well-known data sets related to insurance are considered to illustrate the goodness-of-fit of our new models. The data sets are the allocated loss adjustment expenses data and the Danish fire insurance data, widely used by many authors for model comparisons. We compare the seven newly proposed composite models

with several standard distributions as well as with the composite Weibull–Pareto model proposed by Scollnik and Sun (2012). All the new models have their heads described by a weighted Weibull distribution and tails described by a weighted transformed beta distribution. The transformed beta distributions Klugman et al. (2012) considered here consist of two and three parameters. The seven new models are the composite Weibull–Burr model, the Weibull–Loglogistic model, the Weibull–Paralogistic model, the Weibull–Generalized Pareto model, the Weibull–Inverse Burr model, the Weibull–Inverse Pareto model and the Weibull–Inverse Paralogistic model. The Weibull–Burr and the Weibull–Inverse Burr models have five parameters each. All others have four parameters each.

The allocated loss adjustment expense recorded in thousands of the United States dollar are the additional expenses associated with the settlement of a claim (e.g., claims investigation expenses and legal fees). The data set can be found in the R package *evd*. The sample size is 1500. It has been considered in numerous actuarial studies: see Frees and Valdez (1998) for example. The following are some summary statistics for the data: minimum equals 0.015, first quantile equals 2.333, third quantile equals 12.590, maximum equals 501.900, and mean equals 12.590.

The second data set consists of 2492 losses arising from fire claims in Copenhagen. It was recorded in millions of Danish Krone (DKK) for a period from 1980 to 1990, inclusive and the data have been adjusted to reflect inflation. McNeil (1997) analyzed part of the data which is over one million DKK. The data is highly skewed and have been applied in many papers, including Cooray and Ananda (2005), Scollnik (2007), Scollnik and Sun (2012) and Nadarajah and Bakar (2014) to model insurance losses using composite models. Eling (2012) fitted the skew-normal and skew- t distributions to the data while Bernardi et al. (2012) analyzed the data using a Bayesian approach. The data set can be found in the R package *SMPRACTICALS*. The following are some summary statistics for the data: minimum equals 0.3134, first quantile equals 1.1570, third quantile equals 2.6450, maximum equals 263.3, and mean equals 3.0630.

The quantiles show that both data are highly skewed. The first, second and third quantiles are very far apart from the maximum value. This is further verified by the histograms and the boxplots in Fig. 1. Both data have the same features: a large number of small losses and a small number of extreme losses.

3.2. Goodness-of-fit to the allocated loss adjustment expense data

It is clear from Table 1 that the composite Weibull–Inverse Paralogistic model gives the lowest AIC and SBC values and thus stands as the best model when fitted to the allocated loss adjustment expense data.

The AIC and SBC values for all the standard distributions are substantially larger than those for the composite Weibull–Inverse Paralogistic model. A difference of greater than 10 in SBC is a strong evidence of a better fit as opposed to models with larger SBC values, see Kass and Raftery (1995).

Among the composite models, the Weibull–Inverse Burr model has the second smallest values for the AIC and SBC. The composite Weibull–Loglogistic model has the third smallest values for the AIC and SBC. Also, the composite Weibull–Inverse Paralogistic model is preferred to the Weibull–Pareto model introduced by Scollnik and Sun (2012).

The composite Weibull–Pareto and the composite Weibull–Burr models are nested, i.e., the composite Weibull–Burr model contains the composite Weibull–Pareto model as a particular case. The likelihood ratio test can be used to determine whether the composite Weibull–Burr model is a significant improvement over the

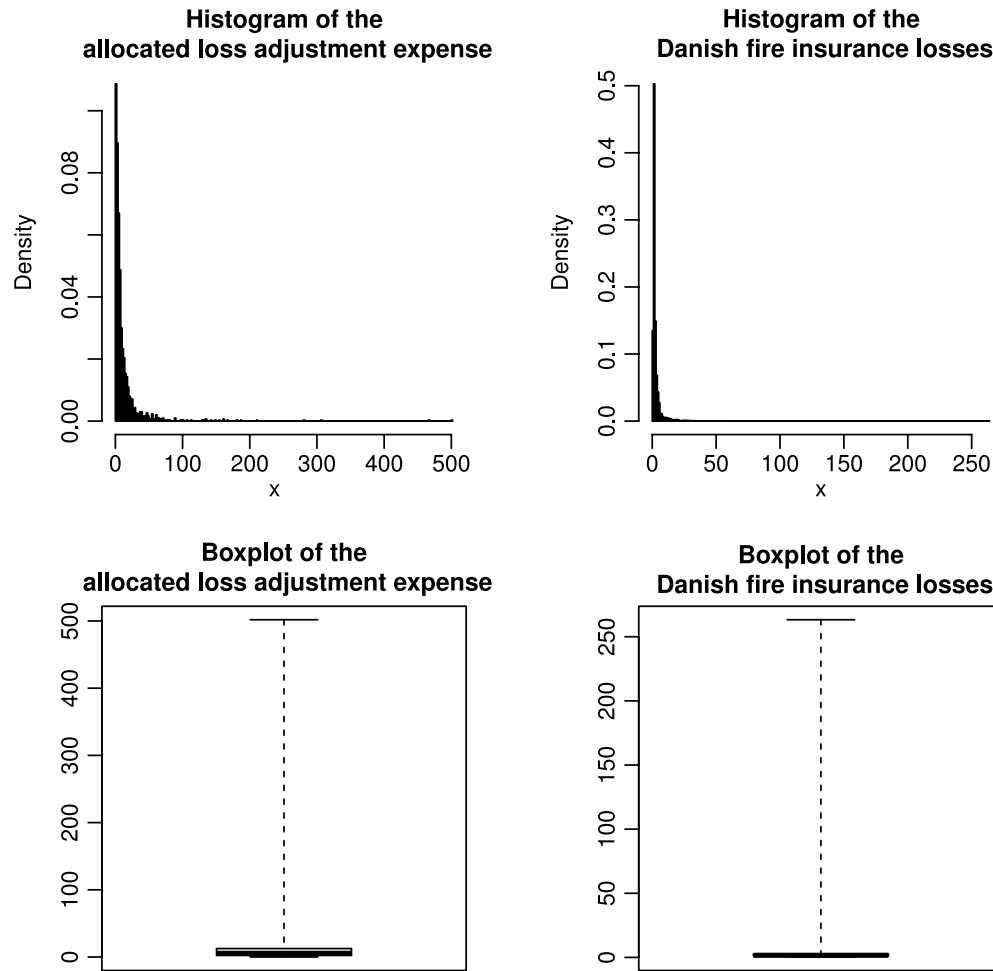


Fig. 1. Histograms and boxplots of the allocated loss adjustment expense and the Danish fire insurance loss.

composite Weibull–Pareto model. Under the null hypothesis of no improvement, the test statistic is

$$D = -2[L(\theta_0) - L(\theta_1)],$$

where $L(\theta_0)$ and $L(\theta_1)$ are the maximized log-likelihood values under the composite Weibull–Pareto and composite Weibull–Burr models, respectively. Using Wilks' theorem, the D can be approximated by a chi-square random variable with one degree of freedom. The statistic $D = 2(1861.724 - 1861.202)$, so that the p -value equals 0.307. This is a strong evidence that the composite Weibull–Burr model is not a significant improvement over the composite Weibull–Pareto model.

Similarly, the composite Weibull–Inverse Pareto and the composite Weibull–Inverse Burr models are nested. The same test shows strong evidence that the composite Weibull–Inverse Burr model is a significant improvement over the composite Weibull–Inverse Pareto model.

3.3. Goodness-of-fit to the Danish fire insurance data

Table 2 shows that the composite Weibull–Burr model has the smallest AIC value, the composite Weibull–Inverse Paralogistic model has the second smallest AIC value and the composite Weibull–Loglogistic model has the third smallest AIC value. The table also shows that the composite Weibull–Inverse Paralogistic model has the smallest SBC value, the composite Weibull–Loglogistic model has the second smallest SBC value and the composite Weibull–Burr model has the third smallest SBC value. So, these three models (the composite Weibull–Burr model, the

composite Weibull–Inverse Paralogistic model and the composite Weibull–Loglogistic model) give the best fits to the Danish fire insurance data. The AIC and SBC values for all the standard distributions are substantially larger.

Comparing the nested composite Weibull–Pareto and the composite Weibull–Burr models by the likelihood ratio test, we see that the latter provides a significantly better fit. Comparing the nested composite Weibull–Inverse Pareto and the composite Weibull–Inverse Burr models by the likelihood ratio test, we see again that the latter provides a significantly better fit.

4. Numerical illustration of risk measures

In this section, we discuss two well known risk measures for describing the appropriateness of the fitted models to empirical measures. Table 3 reports the empirical as well as the theoretical estimates of Value-at-Risk (VaR) and Conditional Tail Expectation (CTE) at the 99% confidence level. The parameter estimates in Tables 1 and 2 were used to compute the theoretical estimates. The results are presented only for the composite Weibull models. We noted earlier that the composite Weibull models provided significantly better fits than the standard distributions for both data sets, so the results are not provided for the latter.

We also investigate the validity of the two aforementioned risk measures calculated from the fitted composite models using backtesting. The backtesting examines whether the proportion of violations and the empirical mean of violations obtained using the theoretical estimates of VaR and CTE are compatible with

Table 1

Parameter estimates of the composite models for the allocated loss adjustment expense data.

Model	R optimization function	Estimated parameters	NLL	AIC	SBC
Weibull	nlm	$\alpha = 0.742$ $\lambda = 9.983$	5133.528	10271.06	10281.68
Burr	nlm	$\beta = 1.705$ $\gamma = 1.099$ $\sigma = 10.334$	5048.334	10102.75	10118.69
Loglogistic	nlm	$\gamma = 1.275$ $\sigma = 5.318$	5061.091	10126.18	10136.81
Paralogistic	nlm	$\beta = 1.216$ $\sigma = 6.707$	5053.235	10110.47	10121.1
Generalized Pareto	nlm	$\beta = 2.025$ $\tau = 1.124$ $\sigma = 11.505$	5049.057	10104.11	10120.05
Pareto	nlm	$\beta = 2.223$ $\sigma = 15.133$	5051.816	10107.63	10118.26
Inverse Burr	nlm	$\gamma = 0.640$ $\tau = 1.539$ $\sigma = 8.602$	5047.755	10101.51	10117.45
Inverse Pareto	nlm	$\tau = 1.395$ $\sigma = 3.364$	5099.633	10203.27	10213.89
Inverse Paralogistic	nlm	$\gamma = 1.171$ $\sigma = 4.367$	5071.857	10147.71	10158.34
Weibull–Burr	nlm	$\alpha = 71.516$ $\lambda = 750.529$ $\beta = 3.184$ $\gamma = 0.999$ $\sigma = 37.412$	1861.202	3732.405	3758.971
Weibull–Loglogistic	nlm	$\alpha = 72.015$ $\lambda = 698.964$ $\gamma = 1.198$ $\sigma = 3.849$	1777.565	3563.131	3584.384
Weibull–Paralogistic	nlm	$\alpha = 66.957$ $\lambda = 1537.192$ $\beta = 1.134$ $\sigma = 5.378$	1976.765	3961.53	3982.783
Weibull–Generalized Pareto	nlm	$\alpha = 77.020$ $\lambda = 353.738$ $\beta = 8.676$ $\tau = 1.538$ $\sigma = 105.256$	2226.836	4463.672	4490.238
Weibull–Pareto	nlm	$\alpha = 71.516$ $\lambda = 750.529$ $\beta = 3.184$ $\sigma = 37.412$	1861.724	3731.448	3752.701
Weibull–Inverse Burr	nlm	$\alpha = 85.940$ $\lambda = 129.075$ $\gamma = 0.207$ $\tau = 0.492$ $\sigma = 93.152$	1767.376	3544.752	3571.318
Weibull–Inverse Pareto	nlm	$\alpha = 28.470$ $\lambda = 179.068$ $\tau = 0.000$ $\sigma = 57.689$	4083.808	8175.617	8196.869
Weibull–Inverse Paralogistic	nlm	$\alpha = 76.883$ $\lambda = 360.051$ $\gamma = 1.174$ $\sigma = 2.823$	1601.291	3210.583	3231.836

Table 2

Parameter estimates of the composite models for the Danish fire insurance loss data.

Model	R optimization function	Estimated parameters	NLL	AIC	SBC
Weibull	nlm	$\alpha = 0.948$ $\lambda = 2.952$	5270.471	10544.942	10556.58
Burr	nlm	$\beta = 0.088$ $\gamma = 14.927$ $\sigma = 0.921$	3835.119	7676.238	7693.701
Loglogistic	nlm	$\gamma = 2.653$ $\sigma = 1.770$	4280.587	8565.174	8576.816
Paralogistic	nlm	$\beta = 1.846$ $\sigma = 2.807$	4514.882	9033.764	9045.406
Generalized Pareto	optim	$\beta = 1.180$ $\tau = 1.081$ $\sigma = 13.497$	4100.322	8206.644	8224.107
Pareto	nlm	$\beta = 5.169$ $\sigma = 11.900$	5051.907	10107.814	10119.46
Inverse Burr	optim	$\gamma = 1.022$ $\tau = 1.067$ $\sigma = 11.497$	3967.904	7941.808	7959.271
Inverse Pareto	optim	$\tau = 1.179$ $\sigma = 9.311$	4647.717	9299.434	9311.076
Inverse Paralogistic	nlm	$\gamma = 2.413$ $\sigma = 1.101$	4093.318	8190.636	8202.278
Weibull–Burr	nlm	$\alpha = 16.202$ $\lambda = 0.949$ $\beta = 0.395$ $\gamma = 3.646$ $\sigma = 0.846$	3817.57	7645.14	7674.244
Weibull–Loglogistic	nlm	$\alpha = 15.652$ $\lambda = 0.962$ $\gamma = 1.568$ $\sigma = 0.680$	3821.229	7650.458	7673.741
Weibull–Paralogistic	nlm	$\alpha = 15.512$ $\lambda = 0.965$ $\beta = 1.267$ $\sigma = 0.622$	3822.441	7652.882	7676.165
Weibull–Generalized Pareto	optim	$\alpha = 15.846$ $\lambda = 0.963$ $\beta = 1.644$ $\tau = 4.949$ $\sigma = 0.191$	3822.752	7655.504	7684.608
Weibull–Pareto	nlm	$\alpha = 15.343$ $\lambda = 0.969$ $\beta = 1.653$ $\sigma = 0.560$	3823.698	7655.396	7678.679
Weibull–Inverse Burr	optim	$\alpha = 15.555$ $\lambda = 0.963$ $\gamma = 1.045$ $\tau = 1.568$ $\sigma = 0.668$	3821.202	7652.404	7681.508
Weibull–Inverse Pareto	optim	$\alpha = 15.629$ $\theta = 0.963$ $\beta = 27.383$ $\lambda = 0.000$	3894.884	7797.768	7821.051
Weibull–Inverse Paralogistic	nlm	$\alpha = 15.806$ $\lambda = 0.960$ $\gamma = 1.567$ $\sigma = 0.563$	3820.935	7649.87	7673.153

Table 3

VaR and CTE at 99% confidence level for the ALAE and Danish loss data.

Composite models	VaR		CTE	
	ALAE data	Danish data	ALAE data	Danish data
Empirical	131.71	24.61	222.68	54.60
Weibull–Burr	446.83	25.18	1273.04	82.59
Weibull–Loglogistic	180.21	22.70	633.52	62.80
Weibull–Paralogistic	191.49	22.60	597.01	60.35
Weibull–Generalized Pareto	95.21	22.49	121.56	58.15
Weibull–Pareto	121.95	22.65	189.41	58.21
Weibull–Inverse Burr	170235	22.68	33	62.74
Weibull–Inverse Pareto	873.63	86.78	4651.21	NA
Weibull–Inverse Paralogistic	163.82	22.64	598.45	62.65

Table 4

Backtesting of VaR for the ALEA and Danish fire loss data.

	VaR	N Viol.	Prop. Viol.	conf. int	p-value
Backtesting for the ALEA data					
Weibull–Burr	446.832	2	0.001	0.000 0.005	0.000
Weibull–Loglogistic	180.207	7	0.005	0.002 0.010	0.037
Weibull–Paralogistic	191.489	5	0.003	0.001 0.008	0.006
Weibull–Generalized Pareto	95.212	22	0.015	0.009 0.022	0.089
Weibull–Pareto	121.947	17	0.011	0.007 0.018	0.602
Weibull–Inverse Burr	170235.013	0	0.000	0.000 0.002	0.000
Weibull–Inverse Pareto	873.633	0	0.000	0.000 0.002	0.000
Weibull–Inverse Paralogistic	163.817	8	0.005	0.002 0.011	0.069
Backtesting for the Danish fire loss data					
Weibull–Burr	25.178	24	0.010	0.006 0.014	1.000
Weibull–Loglogistic	22.698	28	0.011	0.007 0.016	0.544
Weibull–Paralogistic	22.605	28	0.011	0.007 0.016	0.544
Weibull–Generalized Pareto	22.489	28	0.011	0.007 0.016	0.544
Weibull–Pareto	22.648	28	0.011	0.007 0.016	0.544
Weibull–Inverse Burr	22.680	28	0.011	0.007 0.016	0.544
Weibull–Inverse Pareto	86.778	3	0.001	0.000 0.004	0.000
Weibull–Inverse Paralogistic	22.639	28	0.011	0.007 0.016	0.544

the expected nominal levels of the two risk measures. This can be verified through a binomial test comparing the number of violations observed with nominal probability of VaR (1%) and through a t -test comparing the CTE of each model with the average of the values that empirically exceeded the VaR calculated using each model.

4.1. Value-at-risk (VaR)

The empirical estimate of VaR can be obtained as:

$$VaR_X(\alpha) = \hat{F}^{-1}(\alpha),$$

where $\hat{F}(\cdot)$ denotes the empirical cdf. The theoretical estimates of VaR can be obtained using the formula:

$$VaR_X(\alpha) = \begin{cases} F_1^{-1}(\alpha(1+\phi)F_1(\theta)), & \text{if } 0 < \alpha \leq \frac{1}{1+\phi}, \\ F_2^{-1}(F_2(\theta) + (\alpha(1+\phi) - 1)(1 - F_2(\theta))/\phi), & \text{if } \frac{1}{1+\phi} < \alpha < 1. \end{cases}$$

The empirical and theoretical estimates of VaR for both data are given in Table 3.

Table 3 shows that the empirical VaR matches up very well with the theoretical estimates of VaR for the Danish data. The differences are less than 3% except for the composite Weibull–Inverse Pareto model. The smallest difference is for the composite Weibull–Burr model.

The empirical VaR does not match up very well with the theoretical estimates of VaR for the ALAE data. The smallest difference of about 7% is for the composite Weibull–Pareto model.

The second smallest difference of about 25% is for the composite Weibull–Inverse Paralogistic model. So, we may say that at least these two models perform reasonably for both data sets.

The corresponding backtesting results are reported in Table 4. Based on the p -values of the binomial test, the composite Weibull–Pareto, the composite Weibull–Generalized Pareto and the composite Weibull–Inverse Paralogistic models show acceptance of the null hypothesis and thus the estimates of VaR under these models are reasonable for the ALAE data. All but the composite Weibull–Inverse Pareto model provide reasonable estimates of VaR for the Danish data.

4.2. Conditional tail expectation (CTE)

The empirical estimate of CTE can be obtained as:

$$CTE_X(\alpha) = \frac{1}{1-\alpha} \int_{\alpha}^1 \hat{F}^{-1}(s) ds,$$

where $\hat{F}(\cdot)$ denotes the empirical cdf. The theoretical estimates of CTE can be obtained using the formula:

$$CTE_X(\alpha) = \frac{1}{1-\alpha} \int_{\alpha}^1 VaR_X(s) ds.$$

The empirical and theoretical estimates of CTE for both data are given in Table 3. The NA in the table shows that a CTE value could not be computed due to convergence issues.

The empirical CTE does not match up very well with the theoretical estimates of CTE for either data set. However, the smallest differences for the Danish data are about 6% for the composite Weibull–Generalized Pareto and composite Weibull–Pareto models and about 10% for the composite Weibull–Paralogistic model.

Table 5
Backtesting of CTE for the ALEA and Danish fire loss data.

	CTE	Mean Viol.	conf. int	p-value
Backtesting for the ALEA data				
Weibull–Burr	1273.036	484.554	265 704	0.014
Weibull–Loglogistic	633.524	304.892	183 427	0.001
Weibull–Paralogistic	597.010	353.391	198 509	0.012
Weibull–Generalized Pareto	121.557	187.212	138 236	0.011
Weibull–Pareto	189.409	211.842	153 271	0.430
Weibull–Inverse Burr	32560897.402	NA	NA	NA
Weibull–Inverse Pareto	4651.210	NA	NA	NA
Weibull–Inverse Paralogistic	598.452	287.642	178 397	0.000
Backtesting for the Danish fire loss data				
Weibull–Burr	82.589	55.839	32.5 79.2	0.027
Weibull–Loglogistic	62.795	51.340	31.1 71.6	0.257
Weibull–Paralogistic	60.347	51.340	31.1 71.6	0.370
Weibull–Generalized Pareto	58.149	51.340	31.1 71.6	0.497
Weibull–Pareto	58.214	51.340	31.1 71.6	0.493
Weibull–Inverse Burr	62.740	51.340	31.1 71.6	0.259
Weibull–Inverse Pareto	NA	186.774	NA	NA
Weibull–Inverse Paralogistic	62.650	51.340	31.1 71.6	0.263

The smallest difference for the ALAE data is about 10% for the composite Weibull–Pareto model. So, we may say that at least the composite Weibull–Pareto model performs reasonably for both data sets.

It is comforting that all theoretical estimates of CTE exceed the empirical estimate, so they provide sufficient allocation in calculating many quantities related to insurance.

The corresponding backtesting results are reported in Table 5, where NA means either that the theoretical estimate of CTE was missing or that the number of observations needed for testing was not sufficient. Based on the p -values of the t -test, the composite Weibull–Pareto model shows acceptance of the null hypothesis and thus the estimate of CTE under this model is reasonable for the ALAE data. All but the composite Weibull–Burr and composite Weibull–Inverse Pareto models provide reasonable estimates of CTE for the Danish data.

The large differences between the theoretical and empirical estimates may be explained by the extremely skewed tails of both data sets.

5. Conclusions

This paper has proposed seven composite Weibull models with tail belongs to the transformed beta family of distributions. All these models were fitted to two well-known insurance loss data sets and their performance measured. An application of the models to risk measurements was also presented. We observed that all the newly proposed composite models fitted the insurance loss data better than standard models. Among all the composite models, the composite Weibull–Burr model, the composite Weibull–Inverse Paralogistic model and the composite Weibull–Loglogistic model gave the best performance. We found evidence that the composite models based on the Burr (Inverse Burr) distribution gave better fits than those based on the Pareto (Inverse Pareto) distribution.

All the computations for this paper were performed using the R statistical software, see R Core Team (2012). In particular, the parameter estimation and the computation of NLL, AIC and SBC were performed using the `nlm` and `optim` functions. The transformed beta distributions used to model the tails of the composite models are implemented in the `actuar` package, an add-on package to R, see Dutang et al. (2008). Numerical differentiation was performed using the `grad` function in the R package `numDeriv`. The `hist` and `boxplot` functions in R were used to draw the histograms and boxplots. The R software provides many flexible tools to work with the models described in this paper.

The approach described in this paper is the first to provide a general rule to develop a composite model with two pieces. The total number of parameters of the composite model equals to the sum of the parameters of the two pieces. The mixing weight, ϕ , and the threshold, θ , are expressed in terms of other parameters of the model, so their inclusions do not increase the number of parameters. A closed form expression for the threshold may not always exist. But it can always be solved for numerically by some quasi-Newton algorithm.

Recent developments have shown that the Pareto, the lognormal and the gamma distributions are becoming less attractive as models for highly skewed loss data. The Burr distribution is able to capture the tail of a highly skewed distribution better. A comparison of the lognormal–Pareto and lognormal–Burr distributions in Nadarajah and Bakar (2014) showed that the later fitted the loss data better. Some of our results in Sections 3.2 and 3.3 support this conclusion.

This paper has shown the superiority of composite Weibull models in modeling highly skewed data to other common models in the literature to date. The composite Weibull–Burr model, the composite Weibull–Inverse Paralogistic model and the composite Weibull–Loglogistic model have been shown to give best fits. A future work is to apply these models to a variety of other insurance loss data sets.

Acknowledgments

The research for this paper was supported by grants from the University of Malaya, UMRG (RP009B-13AFR) and BKP (BK022-2013). 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 Trans. Automat. Control* 19 (6), 716–723.
- Bernardi, M., Maruotti, A., Petrella, L., 2012. Skew mixture models for loss distributions: a Bayesian approach. *Insurance Math. Econom.* 51 (3), 617–623.
- Cooray, K., Ananda, M.M., 2005. Modeling actuarial data with a composite lognormal-pareto model. *Scand. Actuar. J.* 2005 (5), 321–334.
- Dutang, C., Goulet, V., Pigeon, M., 2008. *actuar: an r package for actuarial science*. *J. Stat. Softw.* 25 (7), 38. URL <http://www.jstatsoft.org/v25/i07>.
- Eling, M., 2012. Fitting insurance claims to skewed distributions: are the skew-normal and skew-student good models? *Insurance Math. Econom.* 51 (2), 239–248.
- Frees, E., Valdez, E., 1998. Understanding relationships using copulas. *N. Am. Actuar. J.* 2, 1–15.

- Kass, R.E., Raftery, A.E., 1995. Bayes factors. *J. Amer. Statist. Assoc.* 90 (430), 773–795.
- Klugman, S.A., Panjer, H.H., Willmot, G.E., 2012. *Loss Models: From Data to Decisions*, Vol. 715. John Wiley and Sons.
- Klugman, S.A., Parsa, R., 1999. Fitting bivariate loss distributions with copulas. *Insurance Math. Econom.* 24 (1), 139–148.
- McNeil, A.J., 1997. Estimating the tails of loss severity distributions using extreme value theory. *Astin Bull.* 27 (1), 117–137.
- Nadarajah, S., Bakar, S., 2014. New composite models for the danish fire insurance data. *Scand. Actuar. J.* 2014 (2), 180–187.
- R Core Team 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0, URL <http://www.R-project.org/>.
- Resnick, S.I., 2007. *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer Science+ Business Media.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Statist.* 6 (2), 461–464.
- Scollnik, D.P., 2007. On composite lognormal-pareto models. *Scand. Actuar. J.* 2007 (1), 20–33.
- Scollnik, D.P., Sun, C., 2012. Modeling with weibull-pareto models. *N. Am. Actuar. J.* 16 (2), 260–272.