mathematics

MDPI

Article

Composite and Mixture Distributions for Heavy-Tailed Data—An Application to Insurance Claims

Walena Anesu Marambakuyana and Sandile Charles Shongwe *

Department of Mathematical Statistics and Actuarial Science, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein 9301, South Africa
* Correspondence: shongwesc@ufs.ac.za

Abstract: This research provides a comprehensive analysis of two-component non-Gaussian composite models and mixture models for insurance claims data. These models have gained attraction in actuarial literature because they provide flexible methods for curve-fitting. We consider 256 composite models and 256 mixture models derived from 16 popular parametric distributions. The composite models are developed by piecing together two distributions at a threshold value, while the mixture models are developed as convex combinations of two distributions on the same domain. Two real insurance datasets from different industries are considered. Model selection criteria and risk metrics of the top 20 models in each category (composite/mixture) are provided by using the 'single-best model' approach. Finally, for each of the datasets, composite models seem to provide better risk estimates.

Keywords: claims; composite models; Danish fire loss; heavy-tailed; loss distribution; mixture models; risk measures; single best model approach; skewed

MSC: 62E15

check for updates

Citation: Marambakuyana, W.A.; Shongwe, S.C. Composite and Mixture Distributions for Heavy-Tailed Data—An Application to Insurance Claims. *Mathematics* **2024**, *12*, 335. https://doi.org/10.3390/math12020335

Academic Editors: Arne Johannssen and Nataliya Chukhrova

Received: 21 December 2023
Revised: 18 January 2024
Accepted: 18 January 2024
Published: 19 January 2024

1. Introduction

In the area of loss modelling, basic classical distributions such as the lognormal, Weibull, gamma, Pareto, and Burr distributions are increasingly becoming less popular as composite and mixture models are gaining more attention because of their flexibility. Composite models are developed by piecing together two distributions (which are termed head and tail distributions) at a threshold value so that small and moderate losses are modelled by the head distribution, whereas large losses are modelled by the tail distribution. On the other hand, mixture models are developed as convex combinations of distributions defined on the same overlapping domain, i.e., the positive real line. The different combinations of models that can be constructed provide a large degree of flexibility for modelling heavy-tailed loss data.

The first composite model used to model actuarial data was proposed by Cooray and Ananda [1]. This model has paved the way for more composite model research in the actuarial and risk management curriculum. The idea behind the model was to use the lognormal distribution to model the behaviour of small and moderate losses (high frequency/low severity) and the Pareto distribution to model the behaviour of the large losses (low frequency/high severity). However, this model was criticised by Scollnik [2] as it can be interpreted as a two-component mixture model with fixed and a priori known mixing weights. Scollnik [2] then proposed two models with unrestricted mixing weights. Unlike the model proposed by [1], these models provided more flexibility due to the accommodation of different proportions of the two distributions of the composite model. The models discussed in Scollnik [2] were extended by Pigeon and Denuit [3] for when the threshold is assumed to vary among observations. Pigeon and Denuit [3] proposed two examples of distributions which can be used for the threshold—this resulted in the gamma-distributed threshold and lognormaldistributed threshold. Next, Nadarajah and Bakar [4] introduced the composite lognormal-Burr model, where it was observed that in the case of the Danish fire loss data, this model performed better than the composite lognormal-Pareto family. Parallel to the studies based on the composite lognormal models at that time, Ciumara [5] introduced a model with the Weibull distribution to model the behaviour of small and moderate losses and the Pareto distribution to model the behaviour of large losses. Scollnik and Sun [6] also criticised the restrictive nature of the fixed and a priori known mixing weights of the model discussed in Ciumara [5]. Scollnik and Sun [6] proposed two additional models with unrestricted weights. Abu Bakar et al. [7] extended the class of Weibull composite distributions by proposing seven models with the tail belonging to the family of transformed beta distributions. The new composite models proposed were the composite Weibull-Burr, the composite Weibull-Loglogistic, the composite Weibull-Paralogistic, the composite Weibull-Generalised Pareto, the composite Weibull-Inverse Burr, the composite Weibull-Inverse Pareto and the composite Weibull-Inverse paralogistic models. At the time of their study, [7] found that the composite models with the Weibull as the head distribution performed better for the Danish fire loss data compared to other composite models in the literature. The extension of the framework for composite models was provided by Grün and Miljkovic [8], where they conducted a thorough analysis of 256 distinct composite curve-fitting models which emerged from piecing together two distributions (i.e., head and tail distributions) from the list of 16 widely used parametric distributions—these are provided in Table A1 in Appendix A.

For the Danish fire loss data, Grün and Miljkovic [8] identified the top 20 composite models that fit the data the best and examined the goodness-of-fit characteristics and risk assessments for those 20 models. The composite Weibull-Inverse Weibull, composite Paralogistic-Inverse Weibull, and composite Inverse Burr-Inverse Weibull, respectively, were the top three models based on the Bayesian Information Criterion (BIC). Among the 256 composite models assessed, none of the top 20 best-fitting had the lognormal distribution in the head. Contrarily, using the Weibull, paralogistic, and inverse Burr distributions in the head was proven to be the most practical approach for simulating the small- and moderate-sized claims of Danish fire loss data. The best choices for modelling the long tail of the loss data were the inverse Weibull, inverse paralogistic, loglogistic, Burr, inverse gamma, and paralogistic. Neither the Pareto nor the generalised Pareto distributions were among the top 20 based on the BIC. Calderin-Ojeda and Kwok [9] suggested the use of composite models where the mode is the splice point (or the truncation point). This method is known as the mode matching procedure, and it was used to construct the composite lognormal-Stoppa and the composite Weibull-Stoppa, where the composite Weibull-Stoppa model had the best performance up to date for the Danish data.

Keatinge [10] introduced the use of the mixture of exponentials as a semiparametric approach. Klugman and Rioux [11] stated that a drawback of the mixture of exponentials is its zero mode, and they proposed the augmented mixture of exponentials distribution which consisted of the mixture of exponentials, the gamma or lognormal distribution, and the Pareto distribution, respectively. Lee and Lin [12] stated that a drawback of the augmented mixture of exponentials is that it has a maximum of three modes, and they proposed a mixture of Erlang distributions with the same scale parameter. It is said that the mixture of Erlang distributions is dense in the space of positive, continuous distributions (Tijms, [13]). Lee and Lin [12] also demonstrated that a uniform distribution, a mixture of two gamma distributions, a generalised Pareto distribution, and the lognormal distribution can be approximated by a mixture of Erlang distributions. Finally, Lee and Lin [12] fitted the mixture of Erlang distributions to the US catastrophic loss data. Miljkovic and Grün [14] stated that a drawback of using the mixture of Erlangs with the same scale parameter could be that more components may be required to obtain an adequate fit that could have otherwise been attained without this restriction. Next, Miljkovic and Grün [14] proposed mixtures of non-Gaussian distributions with no restrictions on the parameters. Their best three models for the Danish data based on minimum BIC were the two-component Burr mixture, the three-component inverse Burr mixture, and the five-component lognormal mixture. In addition, Miljkovic and Grün [14] further added that these three models have lower negative log-likelihood (NLL), Akaike Information Criterion (AIC), and BIC in comparison to the composite Weibull-Burr, composite Weibull-Loglogistic, and the Weibull-Inverse paralogistic distributions which were considered to be the best three composite models in [7]. Abu Bakar et al. [15] proposed six two-component mixture models for fitting three real datasets—the Danish, AON Re Belgium, and Norwegian fire loss datasets. The two-component Burr mixture and the two-component lognormal were the first- and second-best models for the three datasets, respectively. However, the two-component exponential mixture was the worst for the Belgian and Danish fire loss datasets, while the two-component Pareto mixture was the worst for the Norwegian fire loss data. Abu Bakar and Nadarajah [16] proposed two-component mixture models based on the inverse transformed gamma and the transformed beta families, where their fit was illustrated using the Danish fire loss data. Furthermore, Abu Bakar and Nadarajah [16] stated that these families are appropriate for modelling loss data because of the high degree of skewness present in the tails of the distributions. This resulted in seventeen two-component mixtures, all with the inverse transformed gamma as the first component distribution and they found that these models have a better fit for the Danish data based on the BIC than all the composite models and mixture models that had been considered for the Danish dataset in the past.

In the spirit of modelling insurance claims data, many other authors have studied loss distributions. Asgharzadeh et al. [17] introduced the generalised inverse Lindley distribution for the Danish data and found it to be better than most of the classical heavy-tailed distributions, but not as good as the composite models. Next, Punzo et al. [18] introduced nine compound models using three real-life datasets (namely, the US indemnity losses, automobile insurance claims, and Norwegian fire claims). These models were said to have more flexibility than the unimodal two-parameter lognormal, inverse Gaussian, and gamma distributions due to the additional parameters. Bhati and Ravi [19] proposed the use of the generalised log-Moyal distribution and fitted it to the Danish and Norwegian fire loss datasets. Motivated by the research work of [18,19], Li et al. [20] proposed the use of the three-parameter gamma mixture of the generalised log-Moyal distribution, and it was shown to be a special case of the four-parameter generalised beta of the second kind. Zhao et al. [21] and Ahmad et al. [22] introduced additional new heavy-tailed distributions for use in insurance data analysis. While the above review is by no means comprehensive, it provides an overall view of the current literature for heavy-tailed (insurance/claims) data analysis.

This paper is motivated by the recent work by: (i) Grün and Miljkovic [8] where 256 composite models were evaluated for the Danish fire loss data (note though, the corresponding 256 mixture models have not all been considered before); and (ii) Maphalla et al. [23] where the standard loss distributions with the best goodness of fit for the South African taxi claims data were found to be the lognormal and the Pareto, and the potential future research idea of modelling the South African taxi claims data using mixture models were suggested. One should note that due to the flexibility of two-component composite and mixture models, overfitting may easily occur. Thus, care needs to be exercised when fitting these models, especially when using mixture models with more than two components, as this may easily lead to overfitting and greatly violate the principle of parsimony.

In this paper, we consider a thorough comparison of 256 composite models and 256 mixture models for curve-fitting which are derived from 16 popular parametric distributions listed in Table A1 in Appendix A. This study focuses on the following objectives: (i) To discover composite models that have not been studied previously for the South African taxi claims data; (ii) To discover mixture models that have not been studied previously for the South African taxi claims data and Danish fire loss data; and (iii) To assess the implications of the different composite models and mixture models using risk measures, such as Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR).

This paper is structured as follows: Section 2 provides the methodology, which includes model specification, risk measures, and model selection criteria. Section 3 provides the analysis, wherein all the results for the top 20 composite models and mixture models that yield the best goodness-of-fit to the Danish fire loss data and the South African taxi claims data are discussed. Different information criteria and risk measures are computed and presented for models studied in this paper, with additional results provided in Appendix A. Finally, Section 4 provides the concluding remarks.

2. Methodology

2.1. The Composite Model

2.1.1. Model Specification

The probability density function (pdf) of a composite model which was introduced in [7] and adapted by [8] is given by

$$f(\vartheta_1, \vartheta_2, \theta, \phi) = \begin{cases} \frac{1}{1+\phi} f_1^*(x|\vartheta_1, \theta), \text{ if } 0 < x \le \theta, \\ \frac{\phi}{1+\phi} f_2^*(x|\vartheta_2, \theta), \text{ if } \theta < x < \infty. \end{cases} \tag{1}$$

The continuity condition and the differentiability conditions are imposed at the threshold θ such that,

$$f(\vartheta_1, \vartheta_2, \theta, \phi) \ = f(\vartheta_1, \vartheta_2, \theta, \phi) \tag{2}$$

$$f'(\vartheta_1, \vartheta_2, \theta, \phi) = f'(\vartheta_1, \vartheta_2, \theta, \phi) \tag{3}$$

where ϑ_1 and ϑ_2 are the parameter sets associated with the pdfs on the disjoint intervals, $(0,\ \theta]$ and $(\theta,\ \infty)$, respectively. The continuity and differentiability conditions ensure that the threshold parameter θ and the weight parameter $\phi > 0$ are defined as functions of the other parameters, ϑ_1 and ϑ_2. In addition, $\frac{1}{1+\phi}$ and $\frac{\phi}{1+\phi}$ are referred to as mixing weights, see [4]. Moreover, the continuity condition at threshold θ ensures that the weight parameter ϕ is expressed as a function of the other parameters $\vartheta_1,\ \vartheta_2,\ \theta$, and cumulative distribution function (cdf) in closed form as,

$$\phi = -\frac{\frac{d\ln F_1(\theta|\vartheta_1)}{d\theta}}{\frac{d\ln[1-F_2(\theta|\vartheta_2)]}{d\theta}} = \frac{\frac{f_1(\theta|\vartheta_1)}{F_1(\theta|\vartheta_1)}}{\frac{f_2(x|\vartheta_2)}{1-F_2(\theta|\vartheta_2)}}. \tag{4}$$

Substituting the expression for ϕ obtained in Equation (4) into the differentiability condition in Equation (3) gives the following condition for the threshold θ, which simplifies to,

$$\begin{gathered} \frac{d}{d\theta}\ln\left[\frac{f_1(\theta|\vartheta_1)}{f_2(\theta|\vartheta_2)}\right] = 0 \\ \frac{f_1'(\theta|\vartheta_1)}{f_1(\theta|\vartheta_1)} = \frac{f_2'(\theta|\vartheta_2)}{f_2(\theta|\vartheta_2)}. \end{gathered} \tag{5}$$

Lastly, $f_1^*(x|\vartheta_1,\ \theta)$ and $f_2^*(x|\vartheta_2,\ \theta)$ are truncated pdfs which are defined in terms of their corresponding pdfs and cdfs are

$$f_1^*(x|\vartheta_1, \theta) = \frac{f_1(x|\vartheta_1)}{F_1(\theta|\vartheta_1)}, \tag{6}$$

$$f_2^*(x|\vartheta_2, \theta) = \frac{f_2(x|\vartheta_2)}{1 - F_2(\theta|\vartheta_2)}, \tag{7}$$

and also,

$$F(\vartheta_1,\ \vartheta_2, \theta, \phi) = \begin{cases} \frac{1}{1+\phi}\frac{F_1(x|\vartheta_1)}{F_1(\theta|\vartheta_1)}, & \text{if } 0 < x \le \theta, \\ \frac{1}{1+\phi}\left[1 + \phi\frac{F_2(x|\vartheta_2) - F_2(\theta|\vartheta_2)}{1 - F_2(x|\vartheta_2)}\right], & \text{if } \theta < x < \infty. \end{cases} \tag{8}$$

The kth raw moment of the composite model is given in Grün and Miljkovic [8] as

$$\mathbb{E}\left[X^k\right] = \frac{1}{1+\phi}\mathbb{E}\left[X_1^k\right]\frac{F_1^{(k)}(\theta|\vartheta_1)}{F_1(\theta|\vartheta_1)} + \frac{\phi}{1+\phi}\mathbb{E}\left[X_2^k\right]\frac{1-F_2^{(k)}(\theta|\vartheta_2)}{1-F_2(\theta|\vartheta_2)}, \tag{9}$$

where X_i is the random variable associated with the ith component and $F_i^{(k)}$ is the kth incomplete moment distribution of the ith component distribution. For a random sample $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$, the log-likelihood function which was introduced in Grün and Miljkovic [8] is given by

$$\updownarrow(\vartheta_1, \vartheta_2|\boldsymbol{x}) = \sum_{i=1}^{n} \ln(f(x_i|\vartheta_1, \vartheta_2)). \tag{10}$$

2.1.2. Risk Measures

Abu Bakar et al. [7] and Grün and Miljkovic [8] defined the theoretical estimate for the VaR of X as

$$\text{VaR}_p(X) = \begin{cases} F_1^{-1}(p(1+\phi)F_1(\theta)), & \text{if } 0 < p \le \frac{1}{1+\phi}, \\ F_2^{-1}(F_2(\theta) + (p(1+\phi)-1)(1-F_2(\theta))/\phi), & \text{if } \frac{1}{1+\phi} < p < 1. \end{cases} \tag{11}$$

The theoretical estimates for the TVaR of X are defined in [8] as

$$\text{TVaR}_p(X) = \begin{cases} \frac{1}{1-p}\left[\frac{\int_{\pi_p}^{\theta} x f_1(x)dx}{F_1(\theta)} + \frac{\int_{\theta}^{\infty} x f_2(x)dx}{1-F_2(\theta)}\right], & \text{if } 0 < p \le \frac{1}{1+\phi}, \\ \frac{1}{1-p}\frac{1}{1-F_2(\theta)}\left[\int_{\pi_p}^{\infty} x f_2(x)dx\right], & \text{if } \frac{1}{1+\phi} < p < 1. \end{cases} \tag{12}$$

Finite values of Equation (12) can only be obtained if the first moment of the tail distribution exists (Grün and Miljkovic, [8]).

2.2. The Mixture Model

2.2.1. Model Specification

The pdf of a two-component mixture model is given by

$$f(x|\vartheta_1, \vartheta_2, \phi) = \frac{1}{1+\phi}f_1(x|\vartheta_1) + \frac{\phi}{1+\phi}f_2(x|\vartheta_2), \qquad x > 0, \tag{13}$$

where $\phi > 0$ is the weight parameter, and ϑ_1 and ϑ_2 are the parameter sets associated with the first and second component distributions, respectively, where f_1 and f_2 are the corresponding pdfs. The component distributions are both defined on R^+. Therefore, the set of parameters of the mixture model is $\{\phi, \vartheta_1, \vartheta_2\}$. Unlike the composite model, the weight parameter, ϕ, is not a function of the other parameters. Rather, the weight parameter is also a model parameter, which is estimated by the maximum likelihood method in a similar fashion as the other model parameters. The coefficients of f_1 and f_2 are called mixing weights and for $\phi > 0$, it is clear that $\frac{1}{1+\phi} + \frac{\phi}{1+\phi} = 1$. For $\phi = 1$, the component distributions have equal mixing weights of 0.5, i.e., $\frac{1}{1+\phi} = \frac{\phi}{1+\phi} = 0.5$. For $\phi < 1$, the first component distribution has a greater weight to the mixture model than the second component distribution. For $\phi > 1$, the second component distribution has a greater weight to the mixture model than the first component distribution.

The corresponding cdf is given by

$$F(x|\vartheta_1, \vartheta_2, \phi) = \frac{1}{1+\phi}F_1(x|\vartheta_1) + \frac{\phi}{1+\phi}F_2(x|\vartheta_2), x > 0, \tag{14}$$

where F_1 and F_2 are the cdfs of the first and second components, respectively. The kth raw moment of a two-component mixture model is given by

$$E\left[X^k\right] = \frac{1}{1+\phi}\mathbb{E}\left[X_1^k\right] + \frac{\phi}{1+\phi}\mathbb{E}\left[X_2^k\right], x > 0, \quad (15)$$

where $\mathbb{E}\left[X_1^k\right]$ and $\mathbb{E}\left[X_2^k\right]$ are the kth raw moments of the first and second components, respectively, given that they exist. The moment-generating function (mgf) of a two-component mixture model is given by

$$M_X(t) = \frac{1}{1+\phi}\mathbb{M}_{X_1}(t) + \frac{\phi}{1+\phi}\mathbb{M}_{X_2}(t), \quad (16)$$

where $\mathbb{M}_{X_1}(t)$ and $\mathbb{M}_{X_2}(t)$ are the mgfs of the first and second components, respectively, given that they exist.

For a random sample $\boldsymbol{x} = \{x_1,\ x_2, \ldots,\ x_n\}$, the log-likelihood function was introduced in Abu Bakar and Nadarajah [16] as

$$l(\vartheta_1,\ \vartheta_2, \phi|\boldsymbol{x}) = -n\ln(1+\phi) + \sum_{i=1}^{n}\{\ln[f_1(x_i|\vartheta_1) + \phi f_2(x_i|\vartheta_2)]\}. \quad (17)$$

2.2.2. Flexibility for Unimodal and Multimodal Data

Abu Bakar and Nadarajah [16] illustrated the flexibility of two-component mixture models by their adaptability to unimodal and bimodal density functions. In this section, we extend the demonstration of flexibility (adaptability to unimodality and bimodality) by illustrating graphically with two additional two-component mixture models. The first model has different parametric distributions (i.e., inverse transformed gamma and transformed beta distributions) and the second has the same parametric distribution (i.e., Burr distributions) with different parameters. Varying the parameter estimates as indicated in Table 1 leads to different shapes of the pdfs in Figure 1 to illustrate how the mixture of inverse transformed gamma and transformed beta tends to account for unimodality and bimodality.

Table 1. Parameter estimates corresponding to the models in Figure 1.

	$\boldsymbol{\phi}$	$\boldsymbol{\vartheta_1}$	$\boldsymbol{\vartheta_2}$
Model A	$\phi = 1$	$\alpha = 0.5,\ \tau = 2,\ \theta = 1$	$\alpha = 0.1,\ \gamma = 0.5,\ \tau = 2,\ \theta = 1$
Model B	$\phi = 2.5$	$\alpha = 0.1,\ \tau = 0.5,\ \theta = 1$	$\alpha = 0.9,\ \gamma = 1.5,\ \tau = 5,\ \theta = 1$
Model C	$\phi = 5$	$\alpha = 10,\ \tau = 3,\ \theta = 10$	$\alpha = 0.2,\ \gamma = 5,\ \tau = 2\ , \theta = 0.5$

Similarly, varying the parameter estimates as indicated in Table 2 leads to different shapes of the pdfs in Figure 2 to illustrate how the mixture of two Burr distributions tends to account for unimodality and bimodality. Note that similar patterns can be illustrated for other types of mixture distributions.

Table 2. Parameter estimates corresponding to the models in Figure 2.

	$\boldsymbol{\phi}$	$\boldsymbol{\vartheta_1}$	$\boldsymbol{\vartheta_2}$
Model A	$\phi = 1$	$\alpha = 0.1,\ \gamma = 2, \theta = 1$	$\alpha = 2,\ \gamma = 5, \theta = 1$
Model B	$\phi = 2$	$\alpha = 2,\ \gamma = 0.6, \theta = 1$	$\alpha = 3,\ \gamma = 0.2, \theta = 1$
Model C	$\phi = 1$	$\alpha = 10,\ \gamma = 10, \theta = 3$	$\alpha = 2.5,\ \gamma = 8, \theta = 8$

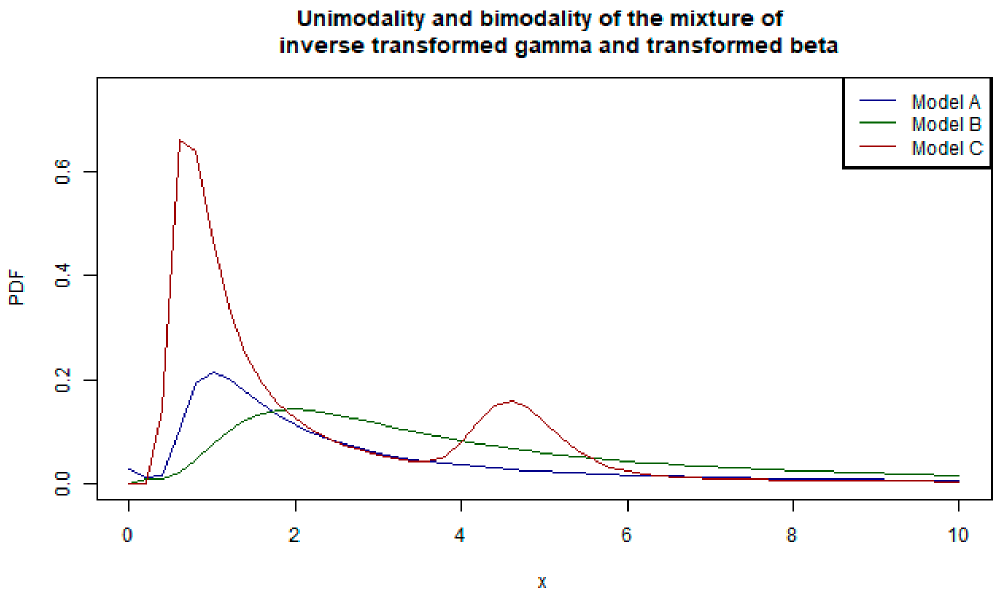

Figure 1. The pdfs of the two-component inverse transformed gamma and transformed beta mixture for different parameters.

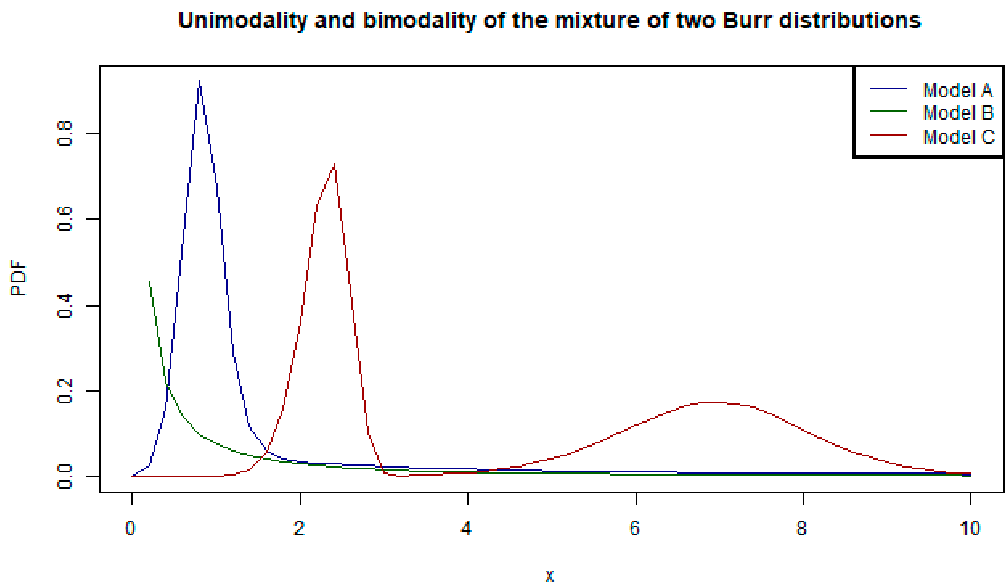

Figure 2. The pdfs of the two-component Burr mixture for different parameters.

2.2.3. Risk Measures

The theoretical estimate for the VaR of the mixture model does not have a closed-form solution and requires a numerical solution of

$$F_X\left(\mathrm{VaR}_p(X)\right) = p \tag{18}$$

which can be evaluated using software as stated in [14]. However, for the mixture model, the theoretical estimate for the TVaR of X can be simplified by the "linearity" property stated in [14] to give the weighted sum of the $\mathrm{TVaR}_p(X)$ of each of the component distributions.

2.3. Model Selection Criteria

This section discusses some commonly used model selection criteria that appear in the area of loss distributions. Three information criteria are considered: NLL, AIC, and BIC; see Abu Bakar et al. [7]. The BIC is also known as the Schwarz's Bayesian Criterion (SBC).

For all three criteria, a lower value implies that the theoretical model provides a better fit to the data. The NLL is appropriate only when comparing models with the same number of parameters; however, the AIC and BIC are more appropriate for comparing models with a different number of parameters. Let $l(\theta)$ denote the maximised log-likelihood function of a model, then the NLL is defined as

$$\text{NLL} = -l(\theta). \tag{19}$$

The AIC was introduced by Akaike [24] and is defined as

$$\text{AIC} = 2\text{NLL} + 2\text{p}, \tag{20}$$

where p is the number of free parameters or degrees of freedom. The BIC was introduced by Schwarz [25] and it is defined as

$$\text{BIC} = 2\text{NLL} + \text{plog}(n), \tag{21}$$

where n is the number of observations. An analysis of the results is given with an emphasis on the BIC.

3. Empirical Analysis

In this section, the statistical computations were performed in R (R Core Team, [26]). Two real-life datasets are considered—the South African taxi claims data and the Danish fire insurance loss data. The taxi (or minibus) industry in South Africa, well known for its taxi turf wars, provides the most commonly used mode of public transport, especially for lower-income communities (which account for a larger proportion of the population due to South Africa's high level of inequality in income levels and high unemployment rates). The types of disasters that this industry faces include and are not limited to road accidents (due to potholes, tyre bursts, improper road infrastructure, vehicle malfunctioning, and drunk driving), hijacking, theft of taxi parts, and damage or fires due to public protests because of poor service delivery by elected officials. However, the Danish fire loss data, which was collected by Copenhagen Reinsurance, covers losses from fire due to buildings, contents, and profits. The Danish data is from Denmark, which is in Europe (Northern Hemisphere), a developed first-world country, whereas the South African taxi claims data is from South Africa, which is in the southernmost part of Africa in the Southern Hemisphere, a developing third-world country. Considering that Denmark is a well-developed country, in the case of a fire hazard, the fire can be extinguished quickly because of Denmark's well-developed social service delivery. On the other hand, with the many hazards that can occur in the taxi industry in South Africa, they may not all be avoidable because of the many underdevelopments. When it comes to economic development, a large portion of the South African population is impoverished and burdened with unemployment, whereas only a small portion of the population in Denmark is lacking.

The South African taxi claims data, which was kindly provided for our study by [23] (this data has been made available in the Supplementary Materials of this paper), consists of 48,043 observations and was divided by 100 for computational ease. The Danish fire loss data, however, is very popular and has a long history of applications. It consists of 2492 observations which were adjusted for inflation to reflect 1985 values. Most of the composite models in actuarial literature have used the Danish data as an application. The Danish dataset is available in the *SMPracticals* package Version 1.4-3 in R, Davidson [27]. The full R code used for the analysis in this paper has been made available in the Supplementary Materials of this paper. Tables 3 and 4 provide the summary of the descriptive statistics for the South African taxi claims data and Danish fire loss data, respectively.

Table 3. Descriptive statistics of the South African taxi claims data.

Minimum	Quantiles	Mean	Maximum	Standard Deviation	Coefficient of Variation	Skewness	Kurtosis
0.1	(20.8, 45, 120.8)	132.3	4803.3	284.1563	2.15	6.474	63.64

Table 4. Descriptive statistics of the Danish fire loss data.

Minimum	Quantiles	Mean	Maximum	Standard Deviation	Coefficient of Variation	Skewness	Kurtosis
0.3134	(1.1572, 1.6339, 2.6455)	3.0627	263.2504	7.976703	2.60	19.896	549.5736

Figure 3 provides the boxplots for the South African taxi claims data and the Danish fire loss data, respectively. The dotted vertical line represents the mean value for the datasets. For both datasets, it is clear that the data are skewed to the right.

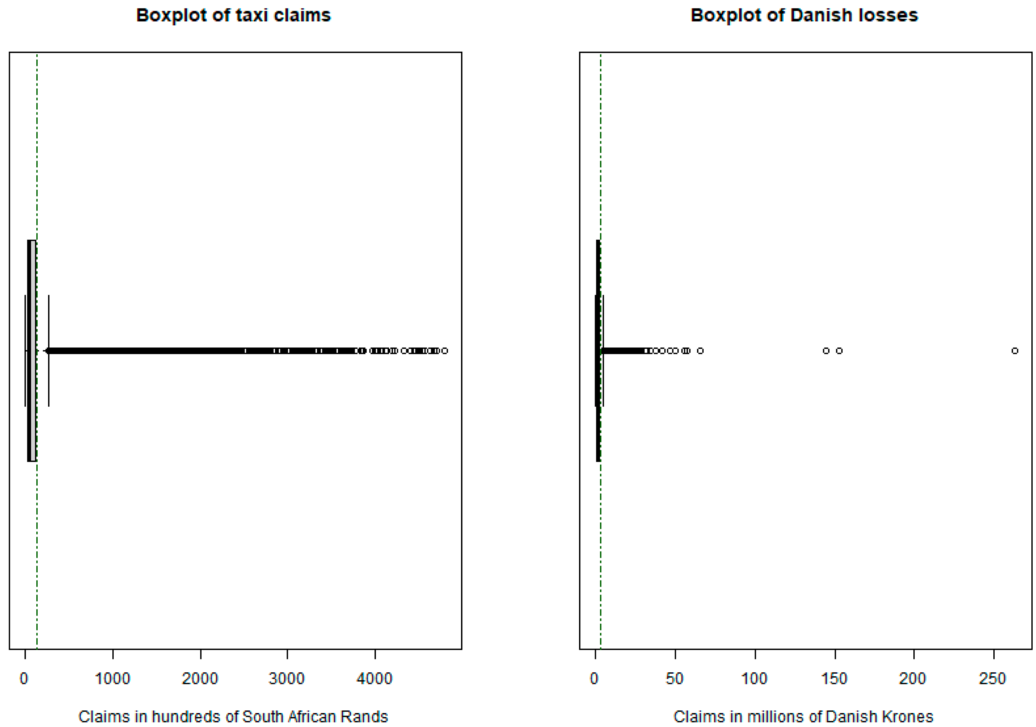

Figure 3. Boxplots of the taxi claims and the Danish loss data.

Figure 4 displays the histograms of the taxi claims and the Danish fire loss data, respectively. By visual inspection of the histograms, the claims data are positive (or at least nonnegative), unimodal and hump-shaped, skewed to the right with long upper tails, and the smaller claims occur with more frequency whereas the larger claims are less frequent.

Figure 5 displays the mean excess plots of the taxi claims on the left and the Danish fire losses data on the right. The mean excess plot for the taxi claims data is initially ultimately increasing, then ultimately constant, and then ultimately decreasing for the remainder of the plot. Therefore, the underlying distribution of the taxi claims data can be said to be heavy-tailed for the lower (left) tail and light-tailed for the upper (right) tail. The mean excess plot for the Danish losses is ultimately increasing (with two observations being the exception). Therefore, the underlying distribution of the Danish data can be said to be heavy-tailed throughout, apart from two observations.

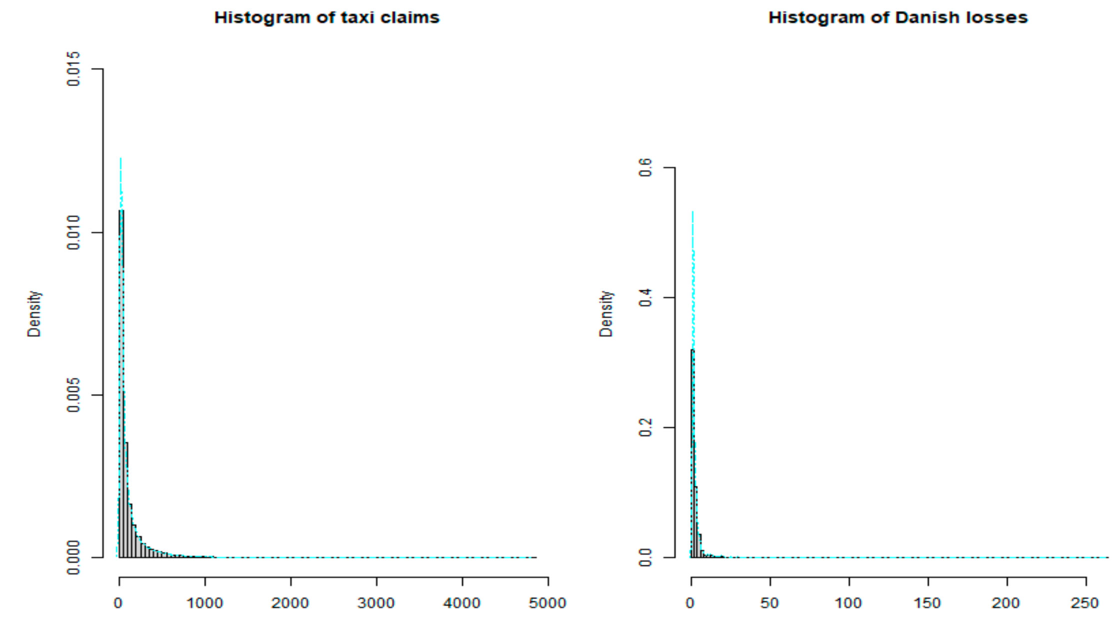

Figure 4. Histograms of the taxi claims and the Danish loss data.

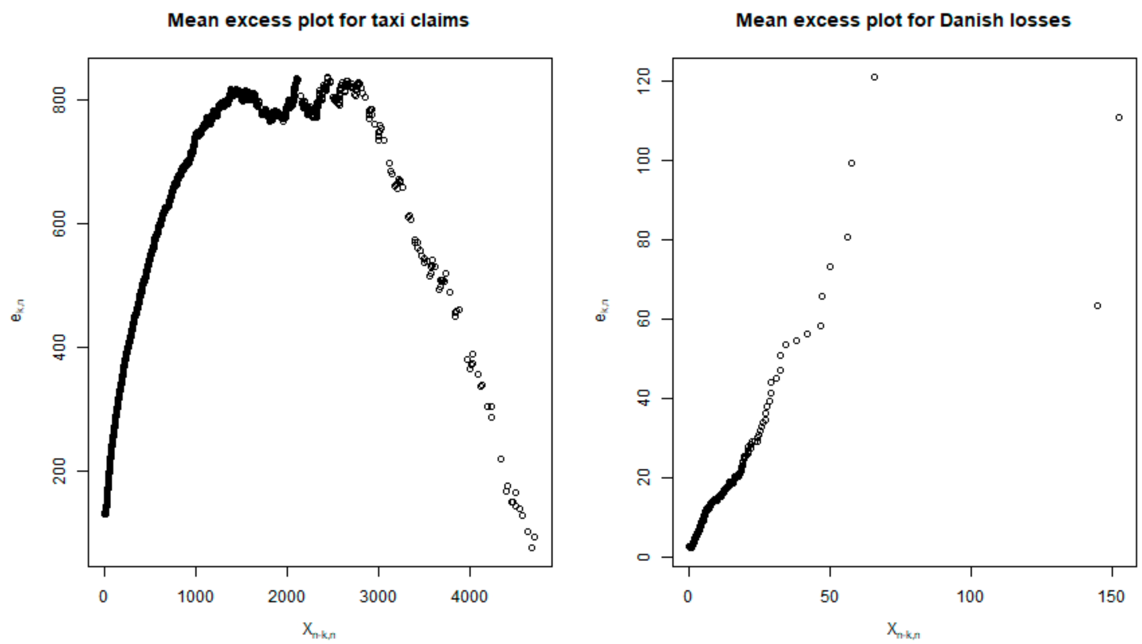

Figure 5. Mean excess plots of the taxi claims and the Danish loss data.

It is important to note that the approach used in this paper is based on the 'single best model' as performed in well-known studies like [7,8,14,16]. That is, using the BIC (as the main model selection criteria), we extract the top 20 best goodness-of-fit models, calculate the corresponding risk metrics, and select the model with the VaR and/or TVaR closer to the corresponding empirical values. Note though that Blostein and Miljkovic [28] as well as Miljkovic and Grün [29] suggested two alternative methods to select the optimal model, which consider both the goodness-of-fit measures and the risk metrics. It is worth mentioning that [29] stated that the 'single best model' approach is still the most used one because although "*model averaging has been recognized in the actuarial field, it has not yet been embraced as a standard practice neither in risk management nor the regulatory capital environment*", but researchers and practitioners are starting to realise the importance of applying the 'model averaging approach'.

Fitting composite models to the taxi claims data

Using the 16 loss distributions outlined in Appendix A's Table A1, it is observed in Table 5 that using distributions such as the gamma, loglogistic, paralogistic, and inverse paralogistic in the head is found to be ideal for modelling the small and moderate size claims of the taxi claims data. However, the tail distributions such as the Weibull, inverse Gaussian, Burr, Pareto, or generalised Pareto and lognormal seem to be the best choices for modelling the upper tail of taxi claims. In an effort to conserve writing space, the corresponding parameter estimates of the top 20 models in Table 5 are provided in Table A2 in Appendix A.

Table 5. Summary of the information criteria of the top 20 composite models for taxi claims (based on the BIC).

Head	Tail	p	NLL	AIC	BIC
Gamma	Weibull	4	**270,197.8**	**540,403.6**	**540,438.8**
Paralogistic	Inverse Gaussian	4	270,198.1	540,404.2	540,439.3
Loglogistic	Inverse Gaussian	4	270,200.7	540,409.5	540,444.6
Paralogistic	Weibull	4	270,201.2	540,410.5	540,445.6
Inverse paralogistic	Inverse Gaussian	4	270,217.0	540,442.1	540,447.2
Weibull	Weibull	4	270,202.4	540,412.8	540,447.9
Gamma	Burr	5	270,197.8	540,405.6	540,449.5
Loglogistic	Weibull	4	270,204.4	540,416.8	540,452.0
Paralogistic	Burr	5	270,201.1	540,412.5	540,456.4
Weibull	Burr	5	270,202.4	540,414.8	540,458.7
Inverse Burr	Weibull	5	270,202.5	540,415.1	540,459.0
Loglogistic	Burr	5	270,204.4	540,418.8	540,462.7
Inverse Burr	Burr	6	270,201.2	540,417.1	540,469.8
Inverse paralogistic	Weibull	4	270,223.4	540,454.7	540,489.9
Inverse paralogistic	Burr	5	270,223.4	540,456.8	540,500.7
Burr	Pareto	5	270,246.1	540,502.3	540,546.2
Weibull	Lognormal	4	270,259.9	540,527.8	540,562.9
Gamma	Lognormal	4	270,260	540,528.8	540,563.9
Gamma	Generalised Pareto	5	270,257.0	540,524.0	540,567.9
Paralogistic	Lognormal	4	270,262.9	540,533.9	540,569.0

Table 6 reports the empirical risk estimates, the estimated risk measures for the top 20 composite models for taxi claims, and the percentage deviation in parenthesis of each estimated risk measure with respect to the empirical risk estimates. The risk estimates obtained from using the top 20 composite distributions closely match the empirical risk estimates. However, using the lognormal or the generalised Pareto distributions as a tail distribution leads to much higher estimates for the TVaR than when using the Burr distribution, the Pareto distribution, or the Weibull distribution. Using the inverse Gaussian distribution as the tail distribution leads to TVaR estimates that are much lower than the empirical estimates.

Table 6. Summary of the empirical risk estimates, risk measures of the top 20 composite models for taxi claims data, and the percentage deviation with respect to the empirical risk estimates in parenthesis.

		$VaR_{0.95}$	$VaR_{0.99}$	$TVaR_{0.95}$	$TVaR_{0.99}$
Empirical Estimates		**525.1509**	**1396.901**	**1085.583**	**2206.203**
Parametric					
Head	**Tail**				
Gamma	Weibull	521.76 (−0.6%)	1396.27 (0.0%)	1125.77 (3.7%)	2422.04 (9.8%)
Paralogistic	Inverse Gaussian	547.96 (4.3%)	1361.21 (−2.6%)	1068.6 (−1.6%)	2055.10 (−6.8%)
Loglogistic	Inverse Gaussian	547.84 (4.3%)	1361.62 (−2.5%)	1068.80 (−1.5%)	2056.09 (−6.8%)
Paralogistic	Weibull	521.92 (−0.6%)	1394.70 (−0.2%)	1124.46 (3.6%)	2416.23 (9.5%)
Inverse paralogistic	Inverse Gaussian	547.38 (4.2%)	1363.79 (−2.4%)	1070.18 (−1.4%)	2061.91 (−6.5%)
Weibull	Weibull	522.21 (−0.6%)	1389.41 (−0.5%)	1119.77 (3.1%)	2396.8 (8.6%)
Gamma	Burr	521.78 (−0.6%)	1396.5 (0.0%)	1126.1 (3.7%)	2423.17 (9.8%)
Loglogistic	Weibull	521.62 (−0.7%)	1393.86 (−0.2%)	1123.85 (3.5%)	2415.2 (9.5%)
Paralogistic	Burr	521.92 (−0.6%)	1394.71 (−0.2%)	1124.46 (3.6%)	2416.26 (9.5%)
Weibull	Burr	-	-	-	-
Inverse Burr	Weibull	522.15 (−0.6%)	1395.0 (−0.1%)	1124.66 (3.6%)	2416.12 (9.5%)
Loglogistic	Burr	-	-	-	-
Inverse Burr	Burr	521.91 (−0.6%)	1394.08 (−0.2%)	1123.92 (3.5%)	2414.26 (9.4%)
Inverse paralogistic	Weibull	520.81 (−0.8%)	1393.196 (−0.3%)	1123.73 (3.5%)	2418.417 (9.6%)
Inverse paralogistic	Burr	521.06 (−0.8%)	1394.12 (−0.2%)	1124.48 (3.6%)	2420.34 (9.7%)
Burr	Pareto	532.21 (1.3%)	1334.65 (−4.5%)	1112.35 (2.5%)	2391.68 (8.4%)
Weibull	Lognormal	516.02 (−1.7%)	1513.17 (8.3%)	1270.78 (17.1%)	3067.21 (39.0%)
Gamma	Lognormal	516.54 (−1.6%)	1524.43 (9.1%)	1286.50 (18.5%)	3133.02 (42.0%)
Gamma	Generalised Pareto	585.11 (11.4%)	1587.63 (13.7%)	1396.13 (28.6%)	3404.88 (54.3%)
Paralogistic	Lognormal	513.46 (−2.2%)	1511.45 (8.2%)	1274.91 (17.4%)	3098.64 (40.5%)

Fitting mixture models to the taxi claims data

Using the 16 loss distributions outlined in Appendix A's Table A1, it is observed from the results in Table 7 that the lognormal distribution seems to be an ideal component distribution for most of the best-fitting mixture models. It seems that the conclusion by Maphalla et al. [23] that the lognormal distribution is the best for taxi claims data is supported by the top mixture models with a lognormal distribution component. In an effort to conserve writing space, the corresponding parameter estimates of the top 20 models in Table 7 are provided in Table A3 in Appendix A.

Table 7. Summary of the information criteria of the top 20 mixture models for taxi claims data (based on the BIC).

First Component	Second Component	p	NLL	AIC	BIC
Inverse gamma	Lognormal	5	270,142.25	**540,294.5**	**540,338.4**
Inverse Gaussian	Lognormal	5	270,142.45	540,294.91	540,338.81
Generalised Pareto	Lognormal	6	**270,142.17**	540,296.35	540,349.03
Inverse paralogistic	Lognormal	5	270,148.17	540,306.35	540,350.25
Inverse Weibull	Lognormal	5	270,148.29	540,306.58	540,350.48
Inverse Burr	Lognormal	6	270,146.59	540,305.17	540,357.85
Loglogistic	Lognormal	5	270,158.197	540,326.39	540,370.29
Burr	Lognormal	6	270,155.86	540,323.72	540,376.4
Gamma	Lognormal	5	270,164.59	540,339.19	540,383.09
Paralogistic	Lognormal	5	270,167.79	540,345.59	540,389.49
Lognormal	Weibull	5	270,186.3	540,382.6	540,426.5
Loglogistic	Generalised Pareto	6	270,225.71	540,463.43	540,516.11
Generalised Pareto	Paralogistic	6	270,227.97	540,467.94	540,520.62
Loglogistic	Paralogistic	5	270,239.59	540,489.17	540,533.07
Burr	Loglogistic	6	270,236.85	540,485.70	540,538.38
Paralogistic	Paralogistic	5	270,247.06	540,504.11	540,548.01
Burr	Burr	7	270,236.71	540,487.43	540,548.89
Inverse gamma	Paralogistic	5	270,248.70	540,507.41	540,551.31
Inverse gamma	Generalised Pareto	6	270,243.88	540,499.75	540,552.43
Paralogistic	Burr	6	270,246.83	540,505.67	540,558.35

For the taxi claims data, the two-component Burr mixture also performs better than the two-component gamma mixture, the two-component Pareto mixture, the two-component Weibull mixture, and the two-component exponential mixture—this is similar to the results observed in Abu Bakar et al. [15] for the Danish, Belgian, and Norwegian loss datasets. Additionally, for the taxi claims data, the two-component paralogistic mixture performs better than the two-component Burr based on the BIC. In fact, the two-component gamma mixture, the two-component exponential mixture, and the two-component Weibull mixture did not converge for the taxi claims data. Other components such as the paralogistic distribution, the Burr distribution, the generalised Pareto distribution, the loglogistic distribution, and the inverse gamma distribution also seem to be optimal component distributions for the mixture models for the taxi claims data.

The mixture models considered provide fair estimates for the VaR at both 95% and 99% security levels, although the VaR at a 95% security level is underestimated by all the models (see Table 8). The TVaR at both 95% and 99% security levels is not underestimated for any of the models, which provides a bit of comfort since the TvaR is a coherent risk measure and more attractive than the VaR.

Table 8. Summary of the empirical risk estimates, risk measures of the top 20 mixture models for the taxi claims data, and the percentage deviation with respect to the empirical risk estimates in parenthesis.

		$\text{VaR}_{0.95}$	$\text{VaR}_{0.99}$	$\text{TVaR}_{0.95}$	$\text{TVaR}_{0.99}$
Empirical Estimates		**525.1509**	**1396.901**	**1085.583**	**2206.203**
Parametric					
First Component	**Second Component**				
Inverse gamma	Lognormal	513.98 (−2.1%)	1382.65 (−1.0%)	1142.3 (5.2%)	2558.96 (16.0%)
Inverse Gaussian	Lognormal	514.55 (−2.0%)	1373.99 (−1.6%)	1134.85 (4.5%)	2529.36 (14.6%)
Generalised Pareto	Lognormal	514.63 (−2.0%)	1383.4 (−1.0%)	1142.86 (5.3%)	2558.81 (16.0%)
Inverse paralogistic	Lognormal	515.07 (−1.9%)	1390.59 (−0.5%)	1149.02 (5.8%)	2580.52 (17.0%)
Inverse Weibull	Lognormal	510.55 (−2.8%)	1378.04 (−1.4%)	1139.31 (4.9%)	2560.72 (16.1%)
Inverse Burr	Lognormal	513.78 (−2.2%)	1389.68 (−0.5%)	1148.59 (5.8%)	2583.77 (17.1%)
Loglogistic	Lognormal	517.67 (−1.4%)	1391.51 (−0.4%)	1149.05 (5.8%)	2570.78 (16.5%)
Burr	Lognormal	516.78 (−1.6%)	1400.85 (0.3%)	1157.8 (6.7%)	2607.82 (18.2%)
Gamma	Lognormal	513.25 (−2.3%)	1360.23 (−2.6%)	1123.05 (3.5%)	2489.42 (12.8%)
Paralogistic	Lognormal	516.78 (−1.6%)	1373.68 (−1.7%)	1133.63 (4.4%)	2515.69 (14.0%)
Lognormal	Weibull	512.52 (−2.4%)	1342.93 (−3.9%)	1107.41 (2.0%)	2431.57 (10.2%)
Loglogistic	Generalised Pareto	518.2 (−1.3%)	1349.01 (−3.4%)	1157.27 (6.6%)	2664.54 (20.8%)
Generalised Pareto	Paralogistic	511.48 (−2.6%)	1363.65 (−2.4%)	1190.45 (9.7%)	2846.36 (29.0%)
Loglogistic	Paralogistic	513.22 (−2.3%)	1372.25 (−1.8%)	1223.63 (12.7%)	3002.93 (36.1%)
Burr	Loglogistic	515.45 (−1.8%)	1351.42 (−3.3%)	1181.42 (8.8%)	2799.21 (26.9%)
Paralogistic	Paralogistic	507.18 (−3.4%)	1384.18 (−0.9%)	1255.56 (15.7%)	3176.84 (44.0%)
Burr	Burr	516.17 (−1.7%)	1351.72 (−3.2%)	1181.84 (8.9%)	2798.08 (26.8%)
Inverse gamma	Paralogistic	504.42 (−3.9%)	1436.60 (2.8%)	1327.5 (22.3%)	3496.71 (58.5%)
Inverse gamma	Generalised Pareto	507.28 (−3.4%)	1402.21 (0.4%)	1268.23 (16.8%)	3220.81 (46.0%)
Paralogistic	Burr	507.44 (−3.4%)	1376.18 (−1.5%)	1240.94 (14.3%)	3109.60 (40.9%)

Fitting composite models to the Danish data

Using the 16 loss distributions outlined in Appendix A's Table A1, it is observed that having distributions such as the Weibull, paralogistic, and inverse Burr in the head is found to be ideal for modelling the small and moderate size claims of Danish fire losses (Grün and Miljkovic, [8]). The tail distributions such as inverse Weibull, inverse paralogistic, loglogistic, Burr, inverse gamma, and paralogistic seem to be the best choices for modelling the long tail of Danish fire losses (Grün and Miljkovic, [8]). In an effort to conserve writing space, the corresponding parameter estimates of the top 20 models in Table 9 are provided in Table A4 in Appendix A.

Table 9. Summary of the information criteria of the top 20 composite models for Danish fire loss data (based on the BIC)—these results are similar to those reported in Grün and Miljkovic [8].

Head	Tail	p	NLL	AIC	BIC
Weibull	Inverse Weibull	4	3820.01	7648.02	**7671.30**
Paralogistic	Inverse Weibull	4	3820.14	7648.28	7671.56
Inverse Burr	Inverse Weibull	5	3816.34	7642.68	7671.79
Weibull	Inverse paralogistic	4	3820.93	7649.87	7673.15
Inverse Burr	Inverse paralogistic	5	3817.07	7644.14	7673.25
Paralogistic	Inverse paralogistic	4	3821.04	7650.08	7673.36
Weibull	Loglogistic	4	3821.23	7650.46	7673.74
Inverse Burr	Loglogistic	5	3817.37	7644.74	7673.85
Paralogistic	Loglogistic	4	3821.32	7650.65	7673.93
Loglogistic	Inverse Weibull	4	3821.38	7650.76	7674.04
Weibull	Burr	5	3817.57	7645.14	7674.24
Paralogistic	Burr	5	3817.72	7645.43	7674.54
Inverse Burr	Burr	6	**3814.00**	**7639.99**	7674.92
Loglogistic	Inverse paralogistic	4	3822.15	7652.31	7675.59
Inverse Burr	Inverse gamma	5	3818.30	7646.61	7675.71
Paralogistic	Inverse gamma	4	3822.22	7652.43	7675.72
Loglogistic	Loglogistic	4	3822.41	7652.82	7676.10
Weibull	Paralogistic	4	3822.44	7652.88	7676.17
Paralogistic	Paralogistic	4	3822.53	7653.05	7676.34
Inverse Burr	Paralogistic	5	3818.68	7647.37	7676.47

Although the composite inverse Burr-Burr model has the lowest NLL and AIC among the other models in Table 9 (and the 256 considered), there is no strong evidence that it provides a better fit than the other models—its BIC is not at least 10 units less than the BIC of the other models in Table 9 (see Abu Bakar et al. [7]). Additionally, the composite inverse Burr-Burr model has six parameters, and the principle of parsimony does not favour it. Rather, a simpler four-parameter composite model is more favourable here.

Table 10 reports the empirical risk estimates, the estimated risk measures for the top 20 composite models for Danish fire loss data and the percentage deviation in parenthesis of each estimated risk measure with respect to the empirical risk estimates. Most of the risk estimates in Table 10 obtained from using the top 20 composite distributions closely match the empirical risk estimates (except those with the Burr as the tail distribution, in terms of the TVaR). The top 20 composite models provide fair estimates for the VaR at both 95% and 99% security levels, although the VaR at a 95% security level is underestimated by all the models (see Table 10). Using the Burr distribution as a tail distribution leads to much higher estimates for the TVaR than using the inverse Weibull distribution, the inverse paralogistic distribution, or the loglogistic distribution. Using the inverse gamma or the paralogistic distribution as the tail distribution leads to TVaR estimates that are lower than the empirical estimates at a 95% security level.

Table 10. Summary of the empirical risk estimates, the risk measures of the top 20 composite models for Danish fire loss data (reported in Grün and Miljkovic [8]) and the percentage deviation with respect to the empirical risk estimates in parenthesis.

		$VaR_{0.95}$	$VaR_{0.99}$	$TVaR_{0.95}$	$TVaR_{0.99}$
Empirical Estimates		**8.406298**	**24.61378**	**22.15509**	**54.60396**
		Parametric			
Head	**Tail**				
Weibull	Inverse Weibull	8.02 (−4.6%)	22.77 (−7.5%)	22.64 (2.2%)	63.86 (17.0%)
Paralogistic	Inverse Weibull	8.02 (−4.6%)	22.79 (−7.4%)	22.67 (2.3%)	64.00 (17.2%)
Inverse Burr	Inverse Weibull	8.01 (−4.7%)	22.73 (−7.7%)	22.59 (2.0%)	63.67 (16.6%)
Weibull	Inverse paralogistic	8.03 (−4.5%)	22.64 (−8.0%)	22.38 (1.0%)	62.65 (14.7%)
Inverse Burr	Inverse paralogistic	8.03 (−4.5%)	22.65 (−8.0%)	22.39 (1.1%)	62.69 (14.8%)
Paralogistic	Inverse paralogistic	8.03 (−4.5%)	22.68 (−7.9%)	22.44 (1.3%)	62.89 (15.2%)
Weibull	Loglogistic	8.05 (−4.2%)	22.7 (−7.8%)	22.43 (1.2%)	62.8 (15.0%)
Inverse Burr	Loglogistic	8.04 (−4.4%)	22.64 (−8.0%)	22.35 (0.9%)	62.46 (14.4%)
Paralogistic	Loglogistic	8.05 (−4.2%)	22.71 (−7.7%)	22.46 (1.4%)	62.89 (15.2%)
Loglogistic	Inverse Weibull	8.05 (−4.2%)	22.96 (−6.7%)	22.93 (3.5%)	65.02 (19.1%)
Weibull	Burr	8.22 (−2.2%)	25.18 (2.3%)	26.98 (21.8%)	82.59 (51.3%)
Paralogistic	Burr	8.22 (−2.2%)	25.18 (2.3%)	26.98 (21.8%)	82.61 (51.3%)
Inverse Burr	Burr	8.22 (−2.2%)	25.13 (2.1%)	26.88 (21.3%)	82.15 (50.4%)
Loglogistic	Inverse paralogistic	8.05 (−4.2%)	22.79 (−7.4%)	22.6 (2.0%)	63.55 (16.4%)
Inverse Burr	Inverse gamma	8.1 (−3.6%)	22.33 (−9.3%)	21.42 (−3.3%)	57.83 (5.9%)
Paralogistic	Inverse gamma	8.11 (−3.5%)	22.44 (−8.8%)	21.57 (−2.6%)	58.48 (7.1%)
Loglogistic	Loglogistic	8.06 (−4.1%)	22.82 (−7.3%)	22.61 (2.1%)	63.52 (16.3%)
Weibull	Paralogistic	8.11 (−3.5%)	22.6 (−8.2%)	21.98 (−0.8%)	60.35 (10.5%)
Paralogistic	Paralogistic	8.11 (−3.5%)	22.62 (−8.1%)	21.99 (−0.7%)	60.41 (10.6%)
Inverse Burr	Paralogistic	8.1 (−3.6%)	22.47 (−8.7%)	21.78 (−1.7%)	59.52 (9.0%)

Fitting mixture models to the Danish data

Using the 16 loss distributions outlined in Appendix A's Table A1, it is observed from the results in Table 11 that the Burr distribution seems to be an ideal component distribution for most of the best mixture models. For the Danish fire loss data, the two-component Burr mixture performs better than the two-component gamma mixture, the two-component Pareto mixture, the two-component Weibull mixture, and the two-component exponential mixture as also concluded in Abu Bakar et al. [15] for the three fire loss datasets considered (i.e., Danish, Belgian and Norwegian). In an effort to conserve writing space, the corresponding parameter estimates of the top 20 models in Table 11 are provided in Table A5 in Appendix A.

Table 11. Summary of the information criteria of the top 20 mixture models for Danish fire loss data (based on the BIC).

First Component	Second Component	p	NLL	AIC	BIC
Burr	Burr	7	**3786.47**	**7586.95**	**7627.69**
Inverse Weibull	Burr	6	3790.61	7593.22	7628.15
Loglogistic	Burr	6	3791.60	7595.20	7630.13
Inverse paralogistic	Burr	6	3792.02	7596.03	7630.96
Paralogistic	Burr	6	3794.36	7600.72	7635.64
Inverse Burr	Burr	7	3790.73	7595.46	7636.21
Gamma	Burr	6	3798.004	7608.01	7642.93
Lognormal	Burr	6	3799.06	7610.12	7645.05
Generalised Pareto	Burr	7	3797.91	7609.83	7650.57
Inverse Gaussian	Burr	6	3801.97	7615.94	7650.86
Inverse gamma	Burr	6	3803.79	7619.57	7654.5
Inverse exponential	Burr	5	3810.03	7630.06	7659.17
Exponential	Burr	5	3811.32	7632.63	7661.74
Inverse Pareto	Burr	6	3810.06	7632.12	7667.04
Weibull	Burr	6	3810.82	7633.65	7668.57
Pareto	Burr	6	3811.33	7634.65	7669.58
Inverse Weibull	Inverse Burr	6	3833.76	7679.53	7714.45
Inverse paralogistic	Inverse Weibull	5	3840.24	7690.49	7719.59
Inverse Weibull	Inverse gamma	5	3840.53	7691.07	7720.17
Inverse Burr	Inverse Burr	7	3833.79	7681.57	7722.32

Other than the Burr distribution, the inverse Weibull distribution and the inverse gamma distribution also seem like optimal component distributions for the mixture models of the Danish fire loss data.

Table 12 reports the empirical risk estimates, the estimated risk measures for the top 20 mixture models for the Danish fire loss data, and the percentage deviation in parenthesis of each estimated risk measure with respect to the empirical risk estimates. The mixture distributions considered provide fair estimates for the VaR at both 95% and 99% security levels. Most of the mixture models have TVaR estimates much higher than the empirical estimates. For the Danish data, the mixture models proposed, especially the ones with the Burr component, do not adequately capture the area under the tail.

Table 12. Summary of the empirical risk estimates, the risk measures of the top 20 mixture models for Danish fire loss data and the percentage deviation with respect to the empirical risk estimates in parenthesis.

		$\mathbf{VaR_{0.95}}$	$\mathbf{VaR_{0.99}}$	$\mathbf{TVaR_{0.95}}$	$\mathbf{TVaR_{0.99}}$
Empirical Estimates		**8.406298**	**24.61378**	**22.15509**	**54.60396**
		Parametric			
First Component	**Second Component**				
Burr	Burr	8.26 (−1.7%)	27.0 (9.7%)	34.74 (56.8%)	119.72 (119.3%)
Inverse Weibull	Burr	8.19 (−2.6%)	25.23 (2.5%)	27.21 (22.8%)	83.83 (53.5%)
Loglogistic	Burr	9.245 (10.0%)	36.07 (46.5%)	60.00 (170.8%)	234.13 (328.8%)
Inverse paralogistic	Burr	8.2025 (−2.4%)	25.32 (2.9%)	27.37 (23.5%)	84.498 (54.7%)
Inverse Burr	Burr	8.1902 (−2.6%)	25.23 (2.5%)	27.22 (22.9%)	83.87 (53.6%)
Gamma	Burr	9.395 (11.8%)	36.48 (48.2%)	59.80 (169.9%)	232.22 (325.3%)
Inverse Gaussian	Burr	8.59 (2.2%)	31.88 (29.5%)	46.53 (110.0%)	172.91 (216.7%)
Lognormal	Burr	8.71 (3.6%)	32.67 (32.7%)	48.82 (120.4%)	183.18 (235.5%)
Generalised Pareto	Burr	8.81 (4.8%)	33.31 (35.3%)	50.72 (128.9%)	191.77 (251.2%)
Inverse gamma	Burr	9.14 (8.7%)	32.64 (32.6%)	43.71 (97.3%)	156.09 (185.9%)
Inverse exponential	Burr	9.61 (14.3%)	34.555 (40.4%)	-	-
Exponential	Burr	9.51 (13.1%)	33.85 (37.5%)	45.41 (105.0%)	162.22 (197.1%)
Inverse Pareto	Burr	9.61 (14.3%)	34.55 (40.4%)	-	-
Paralogistic	Burr	8.43 (0.3%)	26.85 (9.1%)	30.07 (35.7%)	95.73 (75.3%)
Weibull	Burr	8.44 (0.4%)	26.89 (9.2%)	30.14 (36.0%)	96.04 (75.9%)
Pareto	Burr	9.51 (13.1%)	33.85 (37.5%)	45.39 (104.9%)	162.15 (197.0%)
Inverse Weibull	Inverse Burr	8.15 (−3.0%)	21.51 (−12.6%)	20.11 (−9.2%)	51.88 (−5.0%)
Inverse paralogistic	Inverse Weibull	8.13 (−3.3%)	19.92 (−19.1%)	17.89 (−19.3%)	42.34 (−22.5%)
Inverse Weibull	Inverse gamma	8.53 (1.5%)	22.16 (−10.0%)	19.84 (−10.4%)	48.35 (−11.5%)
Inverse Burr	Inverse Burr	8.15 (−3.0%)	21.49 (−12.7%)	20.096 (−9.3%)	51.83 (−5.1%)

Finally, although the results in Tables 5, 7, 9 and 11 are sorted in terms of the BIC (in the last column), the boldfaced value in each column provides the best goodness of fit with respect to the minimum model selection criterion (NLL, AIC, BIC).

4. Conclusions

For the composite models, it seems that the composite Paralogistic-Burr, composite Weibull-Burr, and composite Inverse Burr-Burr are optimal models for both datasets as they both appear in the top 20 composite models. However, for the mixture models, it seems that the two-component Burr mixture, the two-component paralogistic and Burr mixture, and the two-component lognormal and Burr mixture are optimal models for both datasets as they also both appear in the top 20 mixture models. In general, the composite models provide better risk estimates for both of the datasets. The mixture models seem to not adequately capture the area under the tail, especially when using the Burr distribution as a component distribution for the Danish data. Finally, model selection criteria (NLL, AIC, BIC) evaluate the quality of fit of the entire model and not just the tail, so both the model selection criteria and risk estimates are important for deciding which model is optimal.

As it can be observed here, there is no single universal composite or mixture model that is better than the others. Stated differently, the best model depends on the underlying data being used to fit the model and the corresponding risk metrics. Finally, care needs to be taken when interpreting the risk metrics because a model with an excessively large risk metric as compared to the empirical estimate implies that more funds need to be kept in reserve rather than being invested elsewhere, which leads to less profits.

For future research, composite and mixture models (with more than two components) can be fitted to the taxi claims data to evaluate their suitability and other appropriate risk metrics. More importantly, it would be of interest to investigate what would be the best possible back-testing technique that is appropriate for the considered models and risk metrics. While a lot of composite and mixture distributions were considered in this paper, a reader can extend this list by considering the distributions that are discussed in [30]. Next, if data are time-dependent, readers are advised to also investigate analytical methods that involve hidden Markov models. Given that this paper used the 'single best model' approach, it would be interesting to investigate the 'grid map' and 'model averaging' methods discussed in [28,29,31,32] using the datasets and models discussed in this paper. There is also a need for academics to engage with the private sector so that they can be granted access to large datasets and be able to use more advanced and accurate machine learning techniques where data can be split into training, validation, and test sets, as well as for back-testing purposes. However, considering that the data from private companies are usually under many proprietary laws, this is a major limitation when it comes to the analysis of real-life insurance data.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math12020335/s1. Supplementary File: R Code for Mathematics. Table S1. Taxi Claims Data.

Author Contributions: Methodology, W.A.M.; Software, W.A.M.; Validation, W.A.M. and S.C.S.; Formal analysis, W.A.M. and S.C.S.; Investigation, W.A.M. and S.C.S.; Resources, S.C.S.; Data curation, W.A.M. and S.C.S.; Writing—original draft, W.A.M.; Writing—review & editing, S.C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The authors would like to acknowledge the Department of Mathematical Statistics and Actuarial Science, and the Open Access Publication Fund at University of the Free State for assistance with APC.

Data Availability Statement: The data and R codes used in this paper are provided as Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Sixteen distributions that are considered as head and/or tail in the composite model, or first and second components in the mixture model.

Distribution	Parameters	PDF	CDF	$E[X^k]$
Burr	$\alpha > 0,\ \gamma > 0,$ $\theta > 0$	$\frac{\alpha\gamma\left(\frac{x}{\theta}\right)^{\gamma}}{x\left[1+\left(\frac{x}{\theta}\right)^{\gamma}\right]^{\alpha+1}}$	$1 - u^{\alpha}, \quad u = \frac{1}{1+\left(\frac{x}{\theta}\right)^{\gamma}}$	$\frac{\theta^k \Gamma\left(1+\frac{k}{\gamma}\right)\Gamma\left(\alpha-\frac{k}{\gamma}\right)}{\Gamma(\alpha)}, -\gamma < k < \alpha\gamma$
Exponential	$\theta > 0$	$\frac{e^{-\frac{x}{\theta}}}{\theta}$	$1 - e^{-\frac{x}{\theta}}$	$\begin{cases} \theta^k \Gamma(k+1), & k > -1 \\ \theta^k k!, & \text{if } k \text{ is a positive integer} \end{cases}$
Gamma	$\alpha > 0,\ \theta > 0$	$\frac{\left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{x}{\theta}}}{x\Gamma(\alpha)}$	$\Gamma\left(\alpha; \frac{x}{\theta}\right)$	$\begin{cases} \begin{cases} \frac{\theta^k \Gamma(\alpha+k)}{\Gamma(\alpha)}, \end{cases} & k > -\alpha \\ \theta^k (\alpha+k-1)\cdots\alpha, & \text{if } k \text{ is a positive integer} \end{cases}$

Table A1. *Cont.*

Distribution	Parameters	PDF	CDF	$E[X^k]$
Generalised Pareto	$\alpha > 0,\ \tau > 0,$ $\theta > 0$	$\frac{\Gamma(\alpha+\tau)}{\Gamma(\alpha)\Gamma(\tau)} \frac{\theta^{\alpha} x^{\tau-1}}{(x+\theta)^{\alpha+\tau}}$	$\beta(\tau,\alpha;u), \quad u = \frac{x}{x+\theta}$	$\begin{cases} \frac{\theta^k \Gamma(\tau+k)\Gamma(\alpha-k)}{\Gamma(\alpha)\Gamma(\tau)}, & -\tau < k < \alpha \\ \frac{\theta^k \tau(\tau+1)\cdots(\tau+k-1)}{(\alpha-1)\cdots(\alpha-k)}, & \text{if } k \text{ is a positive integer} \end{cases}$
Inverse Burr	$\tau > 0,\ \gamma > 0,$ $\theta > 0$	$\frac{\tau\gamma\left(\frac{x}{\theta}\right)^{\gamma\tau}}{x\left[1+\left(\frac{x}{\theta}\right)^{\gamma}\right]^{\tau+1}}$	$u^{\tau}, \quad u = \frac{\left(\frac{x}{\theta}\right)^{\gamma}}{1+\left(\frac{x}{\theta}\right)^{\gamma}}$	$\frac{\theta^k \Gamma\left(\tau+\frac{k}{\gamma}\right)\Gamma\left(1-\frac{k}{\gamma}\right)}{\Gamma(\tau)}, -\tau\gamma < k < \gamma$
Inverse Exponential	$\theta > 0$	$\frac{\theta e^{-\frac{\theta}{x}}}{x^2}$	$e^{-\frac{\theta}{x}}$	$\theta^k \Gamma(1-k),\ k < 1$
Inverse Gamma	$\alpha > 0,\ \theta > 0$	$\frac{\left(\frac{\theta}{x}\right)^{\alpha} e^{-\frac{\theta}{x}}}{x\Gamma(\alpha)}$	$1 - \Gamma\left(\alpha; \frac{\theta}{x}\right)$	$\begin{cases} \frac{\theta^k \Gamma(\alpha-k)}{\Gamma(\alpha)}, & \text{if } k < \alpha \\ \frac{\theta^k}{(\alpha-1)\cdots(\alpha-k)}, & \text{if } k \text{ is a positive integer} \end{cases}$
Inverse Gaussian	$\mu > 0,\ \theta > 0$	$\left(\frac{\theta}{2\pi x^3}\right)^{\frac{1}{2}} e^{-\frac{\theta z^2}{2x}},$ $z = \frac{x-\mu}{\mu}$	$\Phi\left[z\left(\frac{\theta}{x}\right)^{\frac{1}{2}}\right] + e^{\frac{2\theta}{\mu}} \Phi\left[-y\left(\frac{\theta}{x}\right)^{\frac{1}{2}}\right]$ $z = \frac{x-\mu}{\mu}$	$\sum_{n=0}^{k-1} \frac{(k+n-1)!}{(k-n-1)!n!} \frac{\mu^{n+k}}{(2\theta)^n},\ k = 1, 2, \ldots,$
Inverse Paralogistic	$\tau > 0,\ \theta > 0$	$\frac{\tau^2\left(\frac{x}{\theta}\right)^{\tau^2}}{x\left[1+\left(\frac{x}{\theta}\right)^{\tau}\right]^{\tau+1}}$	$u^{\tau}, \quad u = \frac{\left(\frac{x}{\theta}\right)^{\tau}}{1+\left(\frac{x}{\theta}\right)^{\tau}}$	$\frac{\theta^k \Gamma\left(\tau+\frac{k}{\tau}\right)\Gamma\left(1-\frac{k}{\tau}\right)}{\Gamma(\tau)}, -\tau^2 < k < \tau$
Inverse Pareto	$\tau > 0,\ \theta > 0$	$\frac{\tau\theta x^{\tau-1}}{(x+\theta)^{\tau+1}}$	$\left(\frac{x}{x+\theta}\right)^{\tau}$	$\begin{cases} \frac{\theta^k \Gamma(\tau+k)\Gamma(1-k)}{\Gamma(\tau)}, & -\tau < k < 1 \\ \frac{\theta^k (-k)!}{(\tau-1)\cdots(\tau+k)}, & \text{if } k \text{ is a negative integer} \end{cases}$
Inverse Weibull	$\tau > 0,\ \theta > 0$	$\frac{\tau\left(\frac{\theta}{x}\right)^{\tau} e^{-\left(\frac{\theta}{x}\right)^{\tau}}}{x}$	$e^{-\left(\frac{\theta}{x}\right)^{\tau}}$	$\theta^k \Gamma\left(1-\frac{k}{\tau}\right),\ k < \tau$
Loglogistic	$\gamma > 0,\ \theta > 0$	$\frac{\gamma\left(\frac{x}{\theta}\right)^{\gamma}}{x\left[1+\left(\frac{x}{\theta}\right)^{\gamma}\right]^2}$	$\frac{\left(\frac{x}{\theta}\right)^{\gamma}}{1+\left(\frac{x}{\theta}\right)^{\gamma}}$	$\theta^k \Gamma\left(1+\frac{k}{\gamma}\right)\Gamma\left(1-\frac{k}{\gamma}\right), -\gamma < k < \gamma$
Lognormal	$\mu > 0,\ \sigma > 0$	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{\phi(z)}{\sigma x}$ $z = \frac{lnx-\mu}{\sigma}$	$\Phi(z)$	$e^{k\mu+\frac{1}{2}k^2\sigma^2}$
Paralogistic	$\alpha > 0,\ \theta > 0$	$\frac{\alpha^2\left(\frac{x}{\theta}\right)^{\alpha}}{x\left[1+\left(\frac{x}{\theta}\right)^{\alpha}\right]^{\alpha+1}}$	$1 - u^{\alpha}, \quad u = \frac{1}{1+\left(\frac{x}{\theta}\right)^{\alpha}}$	$\frac{\theta^k \Gamma\left(1+\frac{k}{\alpha}\right)\Gamma\left(\alpha-\frac{k}{\alpha}\right)}{\Gamma(\alpha)}, -\alpha < k < \alpha^2$
Pareto	$\alpha > 0,\ \theta > 0$	$\frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}$	$1 - \left(\frac{\theta}{x+\theta}\right)^{\alpha}$	$\begin{cases} \frac{\theta^k \Gamma(k+1)\Gamma(\alpha-k)}{\Gamma(\alpha)}, & -1 < k < \alpha \\ \frac{\theta^k k!}{(\alpha-1)\cdots(\alpha-k)}, & \text{if } k \text{ is a positive integer} \end{cases}$
Weibull	$\tau > 0,\ \theta > 0$	$\frac{\tau\left(\frac{x}{\theta}\right)^{\tau} e^{-\left(\frac{x}{\theta}\right)^{\tau}}}{x}$	$1 - e^{-\left(\frac{x}{\theta}\right)^{\tau}}$	$\theta^k \Gamma\left(1+\frac{k}{\tau}\right),\ k > -\tau$

Table A2. Parameter estimates and standard errors in parenthesis of the top 20 composite models for the taxi claims data.

Head	Tail	ϑ_1	ϑ_2	θ	ϕ
Gamma	Weibull	$\alpha = 1.8955$ (0.0229), $\frac{1}{\theta} = 0.0599(0.0014)$	$\tau = 0.3325$ (0.0058), $\theta = 7.1857(0.72095)$	35.5261	1.3957
Paralogistic	Inverse Gaussian	$\alpha = 1.6984(101.315)$, $\frac{1}{\theta} = 0.0231(17.4617)$	$\mu = 101.31497$ (3.7430), $\theta = 17.46168$ (1.4671)	37.1193	1.2982
Loglogistic	Inverse Gaussian	$\gamma = 1.7352$ (0.0155), $\theta = 32.5366$ (0.4623)	$\mu = 99.64726$ (4.2422), $\theta = 16.8705$ (1.6193)	40.5153	1.1594
Paralogistic	Weibull	$\alpha = 1.7090$ (0.0154), $\frac{1}{\theta} = 0.02368$ (0.0003)	$\tau = 0.3338$ (0.0058), $\theta = 7.3820$ (0.735)	34.7893	1.4319
Inverse paralogistic	Inverse Gaussian	$\tau = 1.3895$ (0.0072), $\frac{1}{\theta} = 0.0364$ (0.0005)	$\mu = 91.2859$ (6.0897), $\theta = 14.0619$ (2.0543)	51.2797	0.8695
Weibull	Weibull	$\tau = 1.6529$ (0.0166), $\theta = 28.8585$ (0.4634)	$\tau = 0.3376205$ (0.0058), $\theta = 7.9324$ (0.76697)	30.8894	1.6791
Gamma	Burr	$\alpha = 1.8953(0.0236)$, $\frac{1}{\theta} = 0.05999$ (0.0015)	$\alpha = 9290451(2166981)$, $\gamma = 0.3324$ (0.0061), $\frac{1}{\theta} = 1.519089 \times 10^{-22}$ $\left(1.195389 \times 10^{-23}\right)$	35.5336	1.3953

Table A2. *Cont.*

Head	Tail	ϑ_1	ϑ_2	θ	ϕ
Loglogistic	Weibull	$\gamma = 1.7465\ (0.016)$, $\theta = 31.722669$	$\tau = 0.3335(0.0058)$, $\theta = 7.3185(0.7314)$	37.162	1.314
Paralogistic	Burr	$\alpha = 1.7090\ (0.0152\ )$, $\theta = 0.0237(0.0003)$	$\alpha = 132231.5, \gamma = 0.3338$, $\frac{1}{\theta} = 6.186885 \times 10^{-17}$	34.7898	1.4319
Weibull	Burr	$\tau = 1.6528\ (0.0166)$, $\theta = 28.8594\ (0.4602)$	$\alpha = 873910.9$, $\gamma = 0.3376(0.0057)$, $\frac{1}{\theta} = 3.173873 \times 10^{-19}$	30.891	1.6789
Inverse Burr	Weibull	$\tau = 0.8631\ (0.0595)$, $\gamma = 1.9451\ (0.1031)$, $\frac{1}{\theta} = 0.0301(0.0007)$	$\tau = 0.3341\ (0.0058)$, $\theta = 7.4239(0.7363)$	35.0194	1.4186
Loglogistic	Burr	$\gamma = 1.7465\ (0.0156)$, $\theta = 31.7227\ (0.41667)$	$\alpha = 31127.62, \gamma = 0.3335$, $\frac{1}{\theta} = 4.603593 \times 10^{-15}$	37.1616	1.3140
Inverse Burr	Burr	$\tau = 0.8632(0.0599\ )$, $\gamma = 1.9450\ (0.1035\ )$, $\frac{1}{\theta} = 0.03012(0.00066)$	$\alpha = 18358.92$, $\gamma = 0.334237$, $\frac{1}{\theta} = 2.354395 \times 10^{-14}$	35.01608	1.4188
Inverse paralogistic	Weibull	$\tau = 1.3941(0.0077)$, $\frac{1}{\theta} = 0.0372\ (0.0006)$	$\tau = 0.3313\ (0.0062)$, $\theta = 6.9658\ (0.7622)$	44.1095	1.0634
Inverse paralogistic	Burr	$\tau = 1.3941(0.0077)$, $\frac{1}{\theta} = 0.0372\ (0.0006)$	$\alpha = 29397.29$, $\gamma = 0.3313(0.0006)$, $\frac{1}{\theta} = 4.6688 \times 10^{-15}$ $\left(1.0827 \times 10^{-17}\right)$	44.1084	1.0634
Burr	Pareto	$\alpha = 0.3761\ (0.0112)$, $\gamma = 1.8324\ (0.0187)$, $\frac{1}{\theta} = 0.0487\ (0.0010)$	$\alpha = 2.6827\ (0.164)$, $\theta = 443.996\ (61.2433)$	371.0913	0.0883
Weibull	Lognormal	$\tau = 1.68\ (0.0180)$, $\theta = 27.3341\ (0.5253)$	$\mu = 3.2995\ (0.0398)$, $\sigma = 1.6521\ (0.0180)$	27.3694	2.0122
Gamma	Lognormal	$\alpha = 1.9054\ (0.0247)$, $\frac{1}{\theta} = 0.0614\ (0.0017)$	$\mu = 3.2496\ (0.0443)$, $\sigma = 1.6686\ (0.0188)$	32.3335	1.6029
Gamma	Generalised Pareto	$\alpha = 1.9193\ (0.0243)$, $\frac{1}{\theta} = 0.063\ (0.0016)$	$\alpha = 2.00855\ (0.0590)$, $\tau = 0.00000008$ $\left(1.4945 \times 10^{-15}\right)$, $\frac{1}{\theta} = 0.00263\ (0.00015)$	33.0272	1.5616
Paralogistic	Lognormal	$\alpha = 1.7207(0.0264)$, $\frac{1}{\theta} = 0.0243(0.0005)$	$\mu = 3.2653(0.0550)$, $\sigma = 1.6637\ (0.0264)$	30.9552	1.6977

Table A3. Parameter estimates of the top 20 mixture models for the taxi claims data.

First Component	Second Component	ϑ_1	ϑ_2	ϕ
Inverse gamma	Lognormal	$\alpha = 3.8170,\ \theta = 93.6396$	$\mu = 4.074,\ \sigma = 1.4041$	4.4276
Inverse Gaussian	Lognormal	$\mu = 29.8338,\ \theta = 107.9798$	$\mu = 4.0794,\ \sigma = 1.3961$	4.7267
Generalised Pareto	Lognormal	$\alpha = 4.1712,\ \tau = 45.299$, $\theta = 2.2854$	$\mu = 4.0765,\ \sigma = 1.4031$	4.4435
Inverse paralogistic	Lognormal	$\tau = 2.4166,\ \theta = 17.1673$	$\mu = 4.0813,\ \sigma = 1.4069$	4.0799
Inverse Weibull	Lognormal	$\tau = 2.0034\theta = 23.086563$	$\mu = 4.060686,\ \sigma = 1.4098$	4.1863
Inverse Burr	Lognormal	$\tau = 4.5381,\ \gamma = 2.1940$, $\theta = 12.0926$	$\mu = 4.0736,\ \sigma = 1.4101$	4.0411
Loglogistic	Lognormal	$\gamma = 2.9332,\ \theta = 26.4047$	$\mu = 4.1085,\ \sigma = 1.3974$	3.9066

Table A3. *Cont.*

First Component	Second Component	$\boldsymbol{\vartheta_1}$	$\boldsymbol{\vartheta_2}$	$\boldsymbol{\phi}$
Burr	Lognormal	$\alpha = 0.5817,\ \gamma = 3.4263,$ $\theta = 21.4809$	$\mu = 4.0956802, \sigma = 1.4086$	3.6253
Gamma	Lognormal	$\alpha = 4.2428,\ \theta = 6.5837$	$\mu = 4.091793,\ \sigma = 1.3857$	4.769
Paralogistic	Lognormal	$\alpha = 2.403,\ \theta = 40.2599$	$\mu = 4.126997,\ \sigma = 1.3822$	4.002
Lognormal	Weibull	$\mu = 4.1531,\ \sigma = 1.3612$	$\tau = 1.9636,\ \theta = 29.5290$	0.2536
Loglogistic	Generalised Pareto	$\gamma = 1.7759,\ \theta = 29.8996$	$\alpha = 2.2800,\ \ \tau = 1.5141,$ $\theta = 247.543$	0.52375
Generalised Pareto	Paralogistic	$\alpha = 2.0570,\ \tau = 1.614,$ $\theta = 168.3791$	$\alpha = 1.7326,\ \theta = 39.1302$	1.1726
Loglogistic	Paralogistic	$\gamma = 1.7641,\ \theta = 30.5826$	$\alpha = 1.3814,\ \theta = 228.3339$	0.4576
Burr	Loglogistic	$\alpha = 1.6342,\ \gamma = 1.2558,$ $\theta = 256.4782$	$\gamma = 1.7836,\ \theta = 30.326$	1.9518
Paralogistic	Paralogistic	$\alpha = 1.7332,\ \theta = 40.1747$	$\alpha = 1.3464,\ \ \theta = 180.1915$	0.7779
Burr	Burr	$\alpha = 0.95765,\ \gamma = 1.7884,$ $\theta = 29.6828$	$\alpha = 1.663,\ \gamma = 1.2498,$ $\theta = 264.2923$	0.4903
Inverse gamma	Paralogistic	$\alpha = 1.7495,\ \theta = 294.6358$	$\alpha = 1.6836,\ \theta = 43.7961$	2.4313
Inverse gamma	Generalised Pareto	$\alpha = 1.8386,\ \ \theta = 334.996$	$\alpha = 6.2739,\ \tau = 1.9704,$ $\theta = 107.74715$	2.2563
Paralogistic	Burr	$\alpha = 1.7378,\ \theta = 40.1805$	$\alpha = 1.3951,\ \tau = 1.3194,$ $\theta = 184.542$	0.797

Table A4. Parameter estimates and standard errors in parenthesis of the top 20 composite models for the Danish fire loss data.

Head	Tail	$\boldsymbol{\vartheta_1}$	$\boldsymbol{\vartheta_2}$	$\boldsymbol{\theta}$	$\boldsymbol{\phi}$
Weibull	Inverse Weibull	$\tau = 16.094(1.554),$ $\theta = 0.955(0.0149)$	$\tau = 1.555(0.0505),$ $\frac{1}{\theta} = 1.102(0.0994)$	0.955	9.854
Paralogistic	Inverse Weibull	$\alpha = 16.088(1.581),$ $\frac{1}{\theta} = 0.879(0.0232)$	$\tau = 1.554(0.0507),$ $\frac{1}{\theta} = 1.105(0.101)$	0.957	9.688
Inverse Burr	Inverse Weibull	$\tau = 0.204(0.235),$ $\gamma = 68.745(73.872,$ $\frac{1}{\theta} = 1.046(0.0176)$	$\tau = 1.557(0.0493),$ $\frac{1}{\theta} = 1.096(0.0919)$	0.934	12.609
Weibull	Inverse paralogistic	$\tau = 15.806(1.728),$ $\theta = 0.960(0.0176)$	$\tau = 1.567(0.0565),$ $\frac{1}{\theta} = 1.777(0.213)$	0.961	9.256
Inverse Burr	Inverse paralogistic	$\tau = 0.000383(0.00001),$ $\gamma = 35290(1049),$ $\frac{1}{\theta} = 1.078(0.00005)$	$\tau = 1.567(0.053),$ $\frac{1}{\theta} = 1.775(0.181)$	0.928	14.086
Paralogistic	Inverse paralogistic	$\alpha = 15.745(1.968),$ $\frac{1}{\theta} = 0.872(0.0301)$	$\tau = 1.566(0.057),$ $\frac{1}{\theta} = 1.787(0.221)$	0.964	9.054
Weibull	Loglogistic	$\tau = 15.652(1.939),$ $\theta = 0.962(0.0206)$	$\gamma = 1.568(0.0593),$ $\theta = 0.680(0.0979)$	0.964	9.030
Inverse Burr	Loglogistic	$\tau = 0.000395(0.00002),$ $\gamma = 34285(1.517),$ $\frac{1}{\theta} = 1.078(0.00005)$	$\gamma = 1.570(0.0278),$ $\theta = 0.688(0.00275)$	0.928	14.020

Table A4. *Cont.*

Head	Tail	ϑ_1	ϑ_2	θ	ϕ
Paralogistic	Loglogistic	$\alpha = 15.683(1.205)$, $\frac{1}{\theta} = 0.871(0.0174)$	$\gamma = 1.567(0.0569)$, $\theta = 0.678(0.0888)$	0.965	8.906
Loglogistic	Inverse Weibull	$\gamma = 16.267(1.264)$, $\theta = 0.975(0.0127)$	$\alpha = 1.547(0.0502)$, $\frac{1}{\theta} = 1.130(0.105)$	0.976	8.216
Weibull	Burr	$\tau = 16.267(1.264)$, $\theta = 0.949(0.00107)$	$\alpha = 0.395(0.104)$, $\gamma = 3.646(0.880)$, $\frac{1}{\theta} = 1.182(0.0693)$	0.947	11.13
Paralogistic	Burr	$\alpha = 16.278(1.257)$, $\frac{1}{\theta} = 0.887(0.887)$	$\alpha = 0.394(0.104)$, $\gamma = 3.649(0.884)$, $\frac{1}{\theta} = 1.182(0.0697)$	0.947	11.043
Inverse Burr	Burr	$\tau = 0.262(0.108)$, $\gamma = 53.766(19.976)$, $\frac{1}{\theta} = 1.046(0.0099)$	$\alpha = 0.406(0.107)$, $\gamma = 3.549(0.853)$, $\frac{1}{\theta} = 1.190(0.0713)$	0.932	13.251
Loglogistic	Inverse paralogistic	$\gamma = 16.197(1.358)$, $\theta = 0.977(0.0141)$	$\tau = 1.561(0.0554)$, $\frac{1}{\theta} = 1.819(0.216)$	0.980	7.876
Inverse Burr	Inverse gamma	$\tau = 4.4300(0.000000225)$, $\gamma = 30761(0.885)$, $\frac{1}{\theta} = 1.078(0.0000315)$	$\alpha = 1.641(0.0399)$, $\frac{1}{\theta} = 1.148(0.0263)$	0.928	13.945
Paralogistic	Inverse gamma	$\alpha = 15.635(1.285)$, $\frac{1}{\theta} = 0.869(0.0186)$	$\alpha = 1.635(0.0733)$, $\frac{1}{\theta} = 1.119(0.188)$	0.967	8.753
Loglogistic	Loglogistic	$\gamma = 16.153(1.387)$, $\theta = 0.978(0.0145)$	$\gamma = 1.562(0.0573)$, $\theta = 0.666(0.0911)$	0.981	7.761
Weibull	Paralogistic	$\tau = 15.511(1.314)$, $\theta = 0.965(0.00129)$	$\alpha = 1.267(0.0273)$, $\frac{1}{\theta} = 1.607(0.265)$	0.968	8.660
Paralogistic	Paralogistic	$\alpha = 15.557(1.3555)$, $\frac{1}{\theta} = 0.867(0.020)$	$\alpha = 1.266(0.0273)$, $\frac{1}{\theta} = 1.611(0.267)$	0.969	8.551
Inverse Burr	Paralogistic	$\tau = 0.000929(0.0000014)$, $\gamma = 14718(2.036)$, $\frac{1}{\theta} = 1.077(0.0000152)$	$\alpha = 1.270(0.0136)$, $\frac{1}{\theta} = 1.559(0.0307)$	0.928	13.775

Table A5. Parameter estimates of the top 20 mixture models for the Danish fire loss data.

First Component	Second Component	ϑ_1	ϑ_2	ϕ
Burr	Burr	$\alpha = 0.2706,\ \gamma = 6.6542,$ $\theta = 1.2869$	$\alpha = 0.0257,\ \gamma = 49.3079,$ $\theta = 0.8573$	2.1565
Inverse Weibull	Burr	$\tau = 10.5701,\ \theta = 0.9465$	$\alpha = 0.1577,\ \gamma = 9.0711,$ $\theta = 1.1658$	4.3468
Loglogistic	Burr	$\gamma = 5.1028,\ \theta = 1.7185$	$\alpha = 0.028,\ \gamma = 41.8835,$ $\theta = 0.8645$	4.6562
Inverse paralogistic	Burr	$\gamma = 11.85025,\ \theta = 0.7666$	$\alpha = 0.1541,\ \gamma = 9.2683,$ $\theta = 1.1497$	4.7785
Inverse Burr	Burr	$\tau = 136.2773,\ \gamma = 10.6678,$ $\theta = 0.5971$	$\alpha = 0.1574,\ \gamma = 9.0858,$ $\theta = 1.1646$	4.3782
Gamma	Burr	$\alpha = 8.7777,\ \theta = 0.2029$	$\alpha = 0.0298,\ \gamma = 39.7623,$ $\theta = 0.8677$	5.40768

Table A5. *Cont.*

First Component	Second Component	ϑ_1	ϑ_2	ϕ
Inverse Gaussian	Burr	$\mu = 2.1500,\ \theta = 8.79$	$\alpha = 0.0315,\ \gamma = 38.8700,$ $\theta = 0.8699$	4.7799
Lognormal	Burr	$\mu = 0.6349,\ \sigma = 0.4417$	$\alpha = 0.0308,\ \gamma = 39.48825,$ $\theta = 0.8686$	4.7594
Generalised Pareto	Burr	$\alpha = 12.2499,\ \tau = 10.915,$ $\theta = 2.0900$	$\alpha = 0.0303,\ \gamma = 39.960,$ $\theta = 0.8679$	4.7499
Inverse gamma	Burr	$\alpha = 3.6500,\ \theta = 6.61997$	$\alpha = 0.0404,\ \gamma = 31.330,$ $\theta = 0.877$	4.2999
Inverse exponential	Burr	$\theta = 0.98115$	$\alpha = 0.04097,\ \ \gamma = 25.728,$ $\theta = 0.8903$	134.9374
Exponential	Burr	$\frac{1}{\theta} = 0.3686$	$\alpha = 0.04999,$ $\gamma = 25.2807,\ \ \theta = 0.8908$	119.5614
Inverse Pareto	Burr	$\tau = 66.423,\ \theta = 0.015575$	$\alpha = 0.0491, \gamma = 25.7105,$ $\theta = 0.8903$	134.9823
Paralogistic	Burr	$\alpha = 20.0916,\ \theta = 1.1077$	$\alpha = 0.1294, \gamma = 10.7409,$ $\theta = 1.0445$	10.19554
Weibull	Burr	$\tau = 19.9054,\ \theta = 0.95505$	$\alpha = 0.1291,\ \gamma = 10.7569,$ $\theta = 1.0427$	10.3812
Pareto	Burr	$\alpha = 37.699,\ \theta = 100.9595$	$\alpha = 0.0501,\ \gamma = 25.2500,$ $\theta = 0.8908$	121.669
Inverse Weibull	Inverse Burr	$\tau = 3.9523,\ \theta = 1.1564$	$\tau = 5.8476,\ \gamma = 1.714638,$ $\theta = 0.866$	0.7465
Inverse paralogistic	Inverse Weibull	$\tau = 1.9012,\ \theta = 2.3835$	$\tau = 3.4688,\ \theta = 1.2099$	2.2879
Inverse gamma	Inverse Weibull	$\alpha = 3.6924,\ \theta = 1.183$	$\tau = 1.91129,\ \theta = 4.8491$	0.6334
Inverse Burr	Inverse Burr	$\tau = 5.8468,\ \gamma =$ $1.7148,\ \ \theta = 0.866$	$\tau = 4119.4997,$ $\gamma = 3.9533,\ \theta = 0.1408$	1.33905

References

1. Cooray, K.; Ananda, M.M.A. Modeling actuarial data with a composite lognormal-Pareto model. *Scand. Actuar. J.* **2007**, *2005*, 321–334. [CrossRef]
2. Scollnik, D.P.M. On composite lognormal-Pareto models. *Scand. Actuar. J.* **2007**, *2007*, 20–33. [CrossRef]
3. Pigeon, M.; Denuit, M. Composite Lognormal-Pareto model with random threshold. *Scand. Actuar. J.* **2010**, *2011*, 177–192. [CrossRef]
4. Nadarajah, S.; Bakar, S.A.A. New composite models for the Danish fire insurance data. *Scand. Actuar. J.* **2012**, *2014*, 180–187. [CrossRef]
5. Ciumara, R. An actuarial model based on the composite Weibull-Pareto distribution. *Math. Rep.* **2006**, *8*, 401–414.
6. Scollnik, D.P.M.; Sun, C. Modeling with Weibull-Pareto models. *N. Am. Actuar. J.* **2012**, *16*, 260–272. [CrossRef]
7. Abu Bakar, S.A.; Hamzah, N.A.; Maghsoudi, M.; Nadarajah, S. Modeling loss data using composite models. *Insur. Math. Econ.* **2015**, *61*, 146–154. [CrossRef]
8. Grün, B.; Miljkovic, T. Extending composite loss models using a general framework of advanced computational tools. *Scand. Actuar. J.* **2019**, *2019*, 642–660. [CrossRef]
9. Calderin-Ojeda, E.; Kwok, C.F. Modeling claims data with composite Stoppa models. *Scand. Actuar. J.* **2015**, *2016*, 817–836. [CrossRef]
10. Keatinge, C.L. Modeling losses with the mixed exponential distribution. *Proc. Casualty Actuar. Soc.* **1999**, *LXXXVI*, 654–698.
11. Klugman, S.; Rioux, J. Toward a Unified Approach to Fitting Loss Models. *N. Am. Actuar. J.* **2006**, *10*, 63–83. [CrossRef]
12. Lee, S.C.K.; Lin, S.X. Modeling and Evaluating Insurance Losses Via mixtures of Erlang Distributions. *N. Am. Actuar. J.* **2010**, *14*, 107–130. [CrossRef]
13. Tijms, H. *Stochastic Models: An Algorithm Approach*; Wiley: Hoboken, NJ, USA, 1994; ISBN 0-471-95123-4.
14. Miljkovic, T.; Grün, B. Modeling loss data using mixtures of distributions. *Insur. Math. Econ.* **2016**, *70*, 387–396. [CrossRef]
15. Abu Bakar, S.A.; Nadarajah, S.; Adzhar, Z.A.A.K. Loss modeling using Burr mixtures. *Empir. Econ.* **2017**, *54*, 1503–1516. [CrossRef]

16. Abu Bakar, S.A.; Nadarajah, S. Risk measure estimation under two component mixture models with trimmed data. *J. Appl. Stat.* **2019**, *46*, 835–852. [CrossRef]
17. Asgharzadeh, A.; Nadarajah, S.; Sharafi, F. Generalized inverse Lindley distribution with application to Danish fire insurance data. *Commun. Stat. Theory Methods* **2017**, *46*, 5000–5021. [CrossRef]
18. Punzo, A.; Bagnato, L.; Maruotti, A. Compound unimodal distributions for insurance losses. *Insur. Math. Econ.* **2018**, *81*, 95–107. [CrossRef]
19. Bhati, D.; Ravi, S. On generalized log-Moyal distribution: A new heavy tailed size distribution. *Insur. Math. Econ.* **2018**, *79*, 247–259. [CrossRef]
20. Li, Z.; Beirlant, J.; Meng, S. Generalizing the log-Moyal distribution and regression models for heavy-tailed loss data. *ASTIN Bull. J. IAA* **2021**, *51*, 57–99. [CrossRef]
21. Zhao, J.; Ahmad, Z.; Mahmoudi, E.; Hafez, E.H.; Mohie El-Din, M.M. A new class of heavy-tailed distributions: Modeling and simulating actuarial measures. *Complexity* **2021**, *2021*, 5580228. [CrossRef]
22. Ahmad, Z.; Mahmoudi, E.; Dey, S. A new family of heavy tailed distributions with an application to the heavy tailed insurance loss data. *Commun. Stat.-Simul. Comput.* **2022**, *51*, 4372–4395. [CrossRef]
23. Maphalla, R.; Mokhoabane, M.; Ndou, M.; Shongwe, S.C. Quantifying risk using loss distributions. In *Applied Probability Theory—New Perspectives, Recent Advances and Trends*; Jaoude, A.A., Ed.; IntechOpen: London, UK, 2023; pp. 139–161. [CrossRef]
24. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
25. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
26. R Core Team. R: A Language and Environment for Statistical Computing. 2022. Available online: https://www.R-project.org/ (accessed on 1 April 2023).
27. Davison, A. SMPracticals: Practicals for Use with Davison (2003) Statistical Models. 2019. R Package Version 1.4-3. Available online: https://CRAN.R-project.org/package=SMPracticals (accessed on 1 May 2023).
28. Blostein, M.; Miljkovic, T. On modeling left-truncated loss data using mixtures of distributions. *Insur. Math. Econ.* **2019**, *85*, 35–46. [CrossRef]
29. Miljkovic, T.; Grün, B. Using Model Averaging to Determine Suitable Risk Measure Estimates. *N. Am. Actuar. J.* **2021**, *25*, 562–579. [CrossRef]
30. Rigby, R.A.; Stasinopoulos, M.D.; Heller, G.Z.; De Bastiani, F. *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*; CRC Press: Boca Raton, FL, USA, 2019. [CrossRef]
31. Zou, J.; Wang, W.; Zhang, X.; Zou, G. Optimal model averaging for divergent-dimensional Poisson regressions. *Econom. Rev.* **2022**, *41*, 775–805. [CrossRef]
32. Zou, J.; Yuan, C.; Zhang, X.; Zou, G.; Wan, A.T.K. Model averaging for support vector classifier by cross-validation. *Stat. Comput.* **2023**, *33*, 117. [CrossRef]