Extending Composite Loss Models Using a General Framework of Advanced Computational Tools

Bettina Grün*

Department of Applied Statistics,
Johannes Kepler University Linz, 4040 Linz, Austria

Tatjana Miljkovic †

Department of Statistics,
Miami University, Oxford, OH 45056, USA

Abstract

Composite models have a long history in actuarial science because they provide a flexible method of curve-fitting for heavy-tailed insurance losses. The ongoing research in this area continuously suggests methodological improvements for existing composite models and considers new composite models. A number of different composite models have been previously proposed in the literature to fit the popular data set related to Danish fire losses. This paper provides the most comprehensive analysis of composite loss models on the Danish fire losses data set to date by evaluating 256 composite models derived from 16 parametric distributions that are commonly used in actuarial science. If not suitably addressed, inevitable computational challenges are encountered when estimating these composite models that may lead to sub-optimal solutions. General implementation strategies are developed for parameter estimation in order to arrive at an automatic way to reach a viable solution, regardless of the specific head and/or tail distributions specified. The results lead to an identification of new well-fitting composite models and provide valuable insights into the selection of certain composite models for which the tail-evaluation measures can be useful in making risk management decisions.

KEY WORDS: composite models; Danish fire insurance data; goodness-of-fit; heavy tailed distributions; risk measures

*Electronic address: `Bettina.Gruen@jku.at`; Corresponding author
†Electronic address: `miljkot@miamioh.edu`

1 Introduction

Fitting distributions to insurance losses is an integral component of pricing and risk management for many insurance products. Generally, the claims analyzed by insurers cover several orders of magnitude and therefore are not amenable to standard parametric models. However, the distribution can be approximated by combining several standard parametric models. Composite parametric loss models assume that the observed losses can be split into two groups, where one group consists of small and moderate losses for the head of the distribution, while the other group contains the large losses in the tail of the distribution. For each group of observations, a standard truncated parametric distribution is used to approximate its distribution (Klugman et al., 2012). Different composite models emerge depending on the distributions used for the head and the tail as well as the conditions imposed at the threshold, where these two parametric distributions are switched. The performance of the proposed composite models has often been assessed based on the popular Danish fire losses data set (Davison, 2013).

The first generation of composite models was based on the idea of using the Log-normal distribution to model the moderate and small claims that occur with high frequency in the head and the Pareto distribution to model the large claims that occur with low frequency in the tail. Cooray and Ananda (2005) proposed such a model with the Log-normal density used up to a specified threshold and the Pareto density used beyond this threshold value. They used the Log-normal and the Pareto density directly for the two components with a common normalization constant. They also imposed additional constraints in order to ensure continuity and differentiability of the composite density at the threshold value, and thus obtained a smooth density curve. This model was criticized by Scollnik (2007) because it implies that the mixing weights of the two components are fixed a-priori. Fixing the mixing weights in advance implies that the proportion of observations from the Log-normal distribution as well as from the Pareto distribution cannot be varied when fitting the composite distribution density, and thus cannot be adapted to fit the available data. In order to remedy this issue, Scollnik (2007) proposed a composite Log-normal-Pareto model with a-priori unrestricted mixing weights. This implies that the fitted distribution can accommodate different proportions of observations from either of the two distributions. Furthermore, the Log-normal-Pareto model proposed by Scollnik (2007) uses the Generalized Pareto Distribution (GPD), instead of the two-parameter Pareto distribution for modeling losses in the tail of the distribution, in order to increase the flexibility of the model. Pigeon and Denuit (2011) extended this model by assuming that the threshold is not the same for all observations, but rather allowed the threshold to vary among observations. This corresponds to introducing a distribution for the threshold and assuming that each observation has its own threshold as a realization of a random variable from the threshold distribution. The authors considered the Gamma and the Log-normal distributions as threshold distributions when modeling Danish fire losses. Nadarajah and Bakar (2014) further extended the methodology of Scollnik (2007) by introducing the Log-normal-Burr composite model.

In the spirit of Log-normal-Pareto models, another class of composite models has emerged based on fitting the Weibull distribution up to an unknown threshold, and then using the

Pareto distribution beyond the threshold value. These models include the composite Weibull-Pareto models proposed by Ciumara (2006), Cooray (2009), and Scollnik and Sun (2012). Bakar et al. (2015) further extended this class of Weibull composite models by considering a family of Transformed Beta distributions for modeling the tail losses, beyond the unknown threshold. These distributions include the following: Burr, Log-logistic, Paralogistic, Generalized Pareto, Pareto, Inverse Burr, Inverse Pareto, and Inverse Paralogistic. In these composite models the mixing weights, and the unknown threshold are both in principle allowed to vary. However, due to the continuity and differentiability conditions imposed on the composite density at the threshold, these parameters are already implicitly defined by the other parameters of the composite density.

Recently, Calderín-Ojeda and Kwok (2016) introduced composite Log-normal-Stoppa and Weibull-Stoppa models by matching the modes of the two distributions of the composite model. The truncated Log-normal or Weibull component was used up to its modal value. At this point it is switched to the appropriate truncated Stoppa distribution which also has its modal value at this point. This "mode-matching" approach replaced the previously used continuity and differentiability conditions. The authors argued that it is easier to match modal values than to deal with the second derivatives of the corresponding density functions. In fact, when matching modes, the parameter estimates can be derived in closed form.

Extensions of composite models to left-truncated data were studied by Brazauskas and Kleefeld (2016). These authors employed Log-normal-Pareto, Log-normal-GPD (Generalized Pareto), Weibull-Pareto, and Weibull-GPD composite models in modeling the distribution and measuring the tail risk of Norwegian fire claims. The authors showed through an extensive model validation process that the composite models with GPD in the tail "do not consistently yield a statistically closer fit when compared to that of the simpler models" (p. 15). However, these composite models provided substantially different results related to the risk evaluations; therefore, using multiple models is helpful when testing the sensitivity of risk measures in the tail.

Punzo et al. (2018) also studied curve-fitting of insurance losses. The authors proposed nine compound models which have greater tail flexibility than if two-parameter unimodal distributions are used such as the Log-normal, Gamma, and Inverse Gaussian distributions. The greater tail flexibility is achieved by adding one parameter, which leads to the final three-parameter compound models. While these models are different from composite models, they have been tested on the same real data sets and therefore they add to the mix of models used in the general area of loss modeling and curve-fitting.

Despite the progress made in composite modeling so far, it remains unclear if there are other composite models yet to be discovered that can provide a superior fit compared to the aforementioned models for modeling Danish fire losses. Also, the computational challenges associated with fitting the composite models have not been sufficiently discussed in the literature so far. This makes the estimation problems and how they can be addressed less transparent within the research community, and therefore estimation more challenging for practitioners who aim to use these models in practice on different data sets. Previous applications obviously faced estimation problems. Specifically, Bakar et al. (2015) indicated the use of different optimizers to obtain the best solutions for different composite models, where the results for the different composite models are reported for different general purpose optimizers available in the R environment for statistical computing and graphics (R Core Team, 2018).

This paper is motivated by the recent work of Miljkovic and Grün (2016) where the best model for fitting Danish fire losses was found to be a finite mixture based on either Burr or Inverse Burr components. The fact that these two distributions improve the fit of the previously published composite models by Bakar et al. (2015) prompted the exploration of additional composite models which have either Burr or Inverse Burr as head distribution and the various other distributions in the tail. We undertake a large-scale comparison of 256 different composite models for curve-fitting which emerge from combining 16 commonly used parametric distributions as head and tail distributions. This comprehensive evaluation of composite models has the following research objectives: 1) to uncover any potential composite models that fit well which have not been previously studied; 2) to identify suitable computational strategies that can be used to obtain maximum likelihood estimates in a reliable way without manual tuning; 3) to assess the implications of different composite models in risk management applications with a special focus on risk measures, such as Value-at-Risk (VaR) and Conditional-Tail-Expectation (CTE); 4) to enable and indicate the advantages of automatically fitting a large number of composite models rather than only fitting a subset of models based on some preconceived ideas.

This paper is organized as follows. In Section 2, we discuss the methodology which includes model specification, derivation of the risk measures, numerical and computational methods for parameter estimation and risk measure determination, and model selection strategies. In Section 3, we present the empirical analysis, which contains the results of all composite models fitted to the Danish fire losses data set. Further, goodness-of-fit measures and tail-risk measures are evaluated for the best subset of models selected using a suitable model selection criterion. Section 4 concludes.

2 Methodology

2.1 Model specification

The general composite model, as introduced by Bakar et al. (2015) for modeling loss data takes the following form (using slightly different notation than used in Bakar et al., 2015, to improve clarity):

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

together with the two conditions on continuity and continuous differentiability at the threshold θ

$$\lim_{x\to\theta-} f(x|\vartheta_1,\vartheta_2,\theta,\phi) = \lim_{x\to\theta+} f(x|\vartheta_1,\vartheta_2,\theta,\phi),$$
$$\lim_{x\to\theta-} f'(x|\vartheta_1,\vartheta_2,\theta,\phi) = \lim_{x\to\theta+} f'(x|\vartheta_1,\vartheta_2,\theta,\phi).$$

ϑ_1 and ϑ_2 are the unknown parameter sets associated with the probability density function up to the threshold θ and the one beyond the threshold θ, respectively. The two conditions imply that the weight parameter ϕ and the threshold parameter θ are already implicitly determined given ϑ_1 and ϑ_2 and the set of free parameters can be considered to only consist of ϑ_1 and ϑ_2.

$f_i^*(x|\vartheta_i,\theta)$, $i = 1, 2$ are the truncated probability density functions (PDFs) of the composite model with parameters ϑ_i. The truncated distributions are obtained from the PDFs f_i and the cumulative distribution functions (CDFs) F_i as

$$f_1^*(x|\vartheta_1,\theta) = \frac{f_1(x|\vartheta_1)}{F_1(\theta|\vartheta_1)}, \qquad f_2^*(x|\vartheta_2,\theta) = \frac{f_2(x|\vartheta_2)}{1-F_2(\theta|\vartheta_2)}.$$

ϕ and θ can be obtained given ϑ_1 and ϑ_2 in the following way:

- Using the continuity constraint at the threshold θ the parameter ϕ is determined in closed form as a function of the other parameters ϑ_1, ϑ_2 and θ:

$$\phi = -\frac{\dfrac{d\ln F_1(\theta|\vartheta_1)}{d\theta}}{\dfrac{d\ln(1-F_2(\theta|\vartheta_2))}{d\theta}} = \frac{\dfrac{f_1(\theta|\vartheta_1)}{F_1(\theta|\vartheta_1)}}{\dfrac{f_2(\theta|\vartheta_2)}{1-F_2(\theta|\vartheta_2)}}. \tag{2.1}$$

- Inserting the expression for ϕ in Equation (2.1) into the continuous differentiability condition yields the following condition for θ

$$\frac{d}{d\theta}\ln\left[\frac{f_1(\theta|\vartheta_1)}{f_2(\theta|\vartheta_2)}\right] = 0,$$

  which is equivalent to

$$\frac{f_1'(\theta|\vartheta_1)}{f_1(\theta|\vartheta_1)} - \frac{f_2'(\theta|\vartheta_2)}{f_2(\theta|\vartheta_2)} = 0. \tag{2.2}$$

2.2 Risk measures

In order to evaluate exposure to risk, two well-known risk measures are VaR and CTE. Following the notation of Klugman et al. (2012), at the $100p\%$ security level, $\text{VaR}_p(X)$ or π_p denotes the $100p\%$ quantile of the continuous distribution of X satisfying

$$P(X < \pi_p) = p, \qquad F^{-1}(p) = \pi_p,$$

where $F^{-1}(\cdot)$ denotes the inverse of the CDF of X. In the case of a composite model, the CDF is defined based on the parameters ϑ_1, ϑ_2, the threshold θ, and the weight parameter ϕ as follows

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

The $\text{VaR}_p(X)$ function is defined for the composite model as follows

$$\text{VaR}_p(X)=\begin{cases}F_1^{-1}\big(p(1+\phi)F_1(\theta|\vartheta_1)|\vartheta_1\big), & \text{if } 0<p\leq\frac{1}{1+\phi},\\ F_2^{-1}\big(F_2(\theta|\vartheta_2)+\frac{1-F_2(\theta|\vartheta_2)}{\phi}(p(1+\phi)-1)|\vartheta_2\big), & \text{if } \frac{1}{1+\phi}<p\leq 1.\end{cases}$$

The $\text{CTE}_p(X)$, at the $100p\%$ security level, represents the expected loss conditional on the loss exceeding the $100p\%$ quantile of the continuous distribution of X, and it is defined as

$$\text{CTE}_p(X)=\text{E}(X|X>\pi_p)=\frac{\int_{\pi_p}^{\infty}xf(x)dx}{1-F(\pi_p)}.$$

If we consider the two possible locations of π_p relative to the threshold θ separately, the $\text{CTE}_p(X)$ is computed as follows

$$\text{CTE}_p(X)=\begin{cases}\frac{1}{1-p}\left[\frac{\int_{\pi_p}^{\theta}xf_1(x|\vartheta_1)dx}{F_1(\theta|\vartheta_1)}+\frac{\int_{\theta}^{\infty}xf_2(x|\vartheta_2)dx}{1-F_2(\theta|\vartheta_2)}\right], & \text{if } 0<p\leq\frac{1}{1+\phi},\\ \frac{1}{1-p}\frac{1}{1-F_2(\theta|\vartheta_2)}\left[\int_{\pi_p}^{\infty}xf_2(x|\vartheta_2)dx\right], & \text{if } \frac{1}{1+\phi}<p\leq 1.\end{cases}$$

Computation of the CTE_p only gives finite values if the first moment of the tail distribution exists.

Additional actuarial risk measures based on higher order moments were investigated by Maria Sarabia and Calderín-Ojeda (2018). These measures are also referred to as k conditional tail moments and only exist if the higher order moments of the tail distributions exist. Maria Sarabia and Calderín-Ojeda (2018) developed analytical expressions for these measures based on the kth moment of the composite distribution given by

$$\text{E}(X^k)=\frac{1}{1+\phi}\text{E}(X_1^k)\frac{F_1^{(k)}(\theta|\vartheta_1)}{F_1(\theta|\vartheta_1)}+\frac{\phi}{1+\phi}\text{E}(X_2^k)\frac{1-F_2^{(k)}(\theta|\vartheta_2)}{1-F_2(\theta|\vartheta_2)},$$

where $F_i^{(k)}$ for $i=1,2$ represents the kth incomplete moment distribution associated with the ith component distribution and X_i denotes the random variable with PDF $f_i(\cdot)$. More specifically, the k conditional tail moment, defined as $\text{E}(X^k|X>\pi_p)$, is given by

$$\begin{cases}\frac{1}{1-p}\left[\frac{1}{1+\phi}\text{E}(X_1^k)\frac{F_1^{(k)}(\theta|\vartheta_1)-F_1^{(k)}(\pi_p|\vartheta_1)}{F_1(\theta|\vartheta_1)}+\frac{\phi}{1+\phi}\text{E}(X_2^k)\frac{1-F_2^{(k)}(\theta|\vartheta_2)}{1-F_2(\theta|\vartheta_2)}\right], & \text{if } 0<\pi_p<\theta,\\ \text{E}(X_2^k)\frac{1-F_2^{(k)}(\pi_p|\vartheta_2)}{1-F_2(\pi_p|\vartheta_2)}, & \text{if } \theta<\pi_p<\infty.\end{cases}$$

The kth incomplete moment distributions can be further used to compute the kth moment of the stop-loss and limited-loss variables. Additional details about these distributional quantities can be found in Klugman et al. (2012).

2.3 Parameter estimation and risk measure determination

Given a data set $\boldsymbol{x} = (x_1, \ldots, x_n)$, maximum likelihood estimation aims at determining the values for ϑ_1 and ϑ_2 which maximize

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

The parameters θ and ϕ are not included in the log-likelihood function stated above. Due to the continuity and smoothness conditions imposed in composite models two degrees of freedom are lost and these parameters are already determined given values for the remaining parameters ϑ_1 and ϑ_2. In fact θ and ϕ can be calculated based on Equations (2.1) and (2.2). Thus, for estimating composite models the marginal likelihood is maximized where these two parameters are implicitly determined.

No closed form solutions are available for determining the parameter values ϑ_1 and ϑ_2 that maximize the log-likelihood which could be used regardless of the parametric distributions specified for the head and tail. As a result, a general procedure is developed which determines the parameter values of a composite distribution regardless of the parametric distributions used for the head and tail. Numeric optimization, derivative calculation, as well as root finding methods are employed to determine the maximum likelihood estimates when the functions for evaluating the PDFs and CDFs of the parametric tail and head distributions are provided.

The steps required for numerical maximum likelihood estimation of a composite distribution are summarized in Algorithm 1 and explained in the following in more detail. Given the data, select two parametric distributions f_1 and f_2 to be used in the composite model. Define the functions which can be used to evaluate the probability density functions (PDFs) as well as the cumulative distribution functions (CDFs) for the two parametric distributions, given the parameter values ϑ_1 and ϑ_2. Define a function h_1 that determines θ and ϕ based on Equations (2.1) and (2.2) when given ϑ_1 and ϑ_2. Function h_1 involves calculating the derivatives of the PDFs using numerical methods as well as employing a root finding method. This step is complicated by the fact that the solutions are not necessarily unique (see the example given below).

[Algorithm 1 about here.]

Define a function h_2 that evaluates the log-likelihood given ϑ_1, ϑ_2, and $\boldsymbol{x}$ by using h_1 internally to determine suitable values for θ and ϕ. If θ and ϕ are not uniquely determined, the values that provide a higher likelihood value are retained in this step, thus resolving the ambiguity associated with multiple suitable values for θ and ϕ given ϑ_1 and ϑ_2 by also taking $\boldsymbol{x}$ into account.

Select the initial values for ϑ_1 and ϑ_2. The maximum likelihood estimates can then be determined based on a general purpose optimizer. This general purpose optimizer requires initial values for ϑ_1 and ϑ_2 and the function h_2 (using h_1 internally) as input in order to iteratively find a (local) maximum of the log-likelihood.

For different composite models, only the parametric distributions used for the head and tail are varied in this numeric procedure. Thus, the algorithm only needs to be changed by inserting different PDFs and CDFs for the parametric distributions. All other steps are generic.

General purpose optimizers. The parameter space of the parametric distributions usually employed in composite models for insurance loss modeling is restricted, e.g., by imposing a positivity condition on parameters that capture spread. Thus, general purpose optimizers used for parameter estimation need to allow for the inclusion of constraints for the parameters or the parameter vector needs to be suitably transformed to ensure an unrestricted parameter space. Thus after transformation any general purpose optimizer may be used to determine ϑ_1 and ϑ_2 given h_2. In the implementation used to obtain the presented results, the optimization method implemented in the function `nlminb` within the R package **stats** is used after taking the exponential of parameters restricted to be positive.

Numeric tools for determining derivatives. Solving Equation (2.2) requires finding the derivatives of the two PDFs, f_1 and f_2. These can either be determined analytically, or found using numerical methods. Analytic solutions would need to be explicitly specified for the parametric distributions used for head and tail. To obtain an implementation which is generally applicable, regardless of the parametric distributions specified, numeric differentiation methods are employed in the code used to obtain the presented results. Specifically, the R package **numDeriv** (Gilbert and Varadhan, 2016) is used.

Initialization. The general purpose optimizer needs valid initial values in order to start the algorithm. A convergence is reached, at best, only to a local optimum. Values of 1 for all parameters in ϑ_1 and ϑ_2 lie within the feasible region for any of the parametric distributions considered. Therefore, this represents a set of valid initial values that can be applicable for any composite distribution.

Using this single set of initial values alone, the general purpose optimizer failed to detect the global optimum for some of the fitted composite models. This failure was detected by comparing the solutions obtained with those previously reported in the literature with respect to the negative log-likelihood (NLL) values. Lower NLL values in the literature indicated that the specific general purpose optimizer employed, together with this set of initial values, only arrived at a sub-optimal solution.

Thus, more elaborate initialization strategies are required to increase the chances of detecting either the global optimum, or at least a good solution. One possibility is to use random initialization strategies together with local optimizers, and then select the best solution from a set of solutions obtained by using different sets of initial values for the general purpose optimizer. Random initialization strategies need to be able to explore the parameter space sufficiently well, but should also generate sets of initial values that are feasible and target promising areas of the parameter space. To address these issues, a local adaptive strategy is pursued in the following. This random initialization strategy generates feasible initial values in the neighborhood of the currently best solution in order to explore the parameter space.

In particular, given feasible parameter values and the log-likelihood associated with these parameters, new parameter values are proposed using a random walk with independent normal distributions and a pre-specified step size. Several replications of such proposals are generated, by using winsorization to ensure that the parameter values lie within the feasible range. The best proposal is then selected by considering the log-likelihood values. The new proposal is used with probability equal to the exponential value of the log-likelihood of the difference between the old and new parameter values as the new initial value for the general purpose optimizer. This scheme is inspired by random walk Metropolis sampling (e.g., Sherlock et al., 2010).

In this implementation, a step size of 0.5 is employed, together with 100 replications. In addition, this iterative strategy for generating new initial values for the general purpose optimizer is repeated 100 times. This fixed scheme is utilized for any composite model regardless of the head and tail distributions specified. Certainly, tuning of the step size, as well as the number of replications, might be considered in order to increase the acceptance probabilities, and thus the computational efficiency of the scheme.

Root finding algorithms. As noted by Bakar et al. (2015), closed form solutions of Equation (2.2) for θ may not exist, but the equation can be solved by using numerical root finding methods. In fact, efficient methods for finding the root of a univariate function are available, if an interval is provided that contains at least one root, and the function evaluations are of different signs at the end points. To determine θ, an interval can be specified by using the data range. However, the function evaluations may not have different signs at the end points. In addition the interval may include more than one root, or Equation (2.2) might not be finite for all values of θ in this interval. Thus, the interval needs to be split into sub-intervals where the function changes its sign so that standard root finding algorithms can be used for each of the sub-intervals. In the implementation employed first, the function is evaluated on a grid to identify these sub-intervals and then function `uniroot` from the R package **stats** is applied to each sub-interval to determine the roots.

Calculation of standard errors. Numeric tools for determining derivatives can also be employed to determine the Hessian of the log-likelihood function at the maximum likelihood estimate. Thus standard errors of the parameter estimates can be calculated based on this Hessian which is also known a observed Fisher information matrix, $I(\Psi) = -\frac{\partial^2 \ell(\Psi)}{\partial \Psi \partial \Psi'}$ where Ψ represents the parameter vector that consists of both parameter sets $\Psi = (\vartheta_1, \vartheta_2)'$. The inverse of this approximation, $I(\hat{\Psi})^{-1}$ provides a useful estimate for the covariance matrix, $\text{Cov}(\hat{\Psi})$. Taking the square root of the diagonal elements of $\text{Cov}(\hat{\Psi})$ gives estimates of the standard errors of the corresponding parameter estimates. For transformed parameters the variances are determined on the original scale using the Delta method.

Risk measures. For determining the VaR, the inverse at the specific values can again be calculated using a root finding procedure such as the function `uniroot()` from the R package **stats**. The numerical integration for some parts of the $\text{CTE}_p(X)$ formula can be done using `quadinf()` in the R package **pracma** developed by Borchers (2017). The other actuarial risk measures based on higher order moments may be calculated analogously.

Example illustrating non-unique θ. A composite model with the Inverse Gaussian distribution in the head and the Log-logistic distribution in the tail, is fitted to the Danish fire data set. One possible model has the estimated parameters $\hat{\mu} = 3.06$ and $\hat{\tau} = 3.44$ for the Inverse Gaussian distribution and $\hat{\alpha} = 2.62$ and $\hat{\sigma} = 1.76$ for the Log-logistic distribution. For these parameter values, there are five non-unique θ values possible, which all produce a continuous and continuously differentiable density. These solutions are summarized in Table 1.

[Table 1 about here.]

Solutions 1 and 2 are special cases of a composite model where either f_1 is the empty component and f_2 covers the entire data range, resulting in a single component Log-logistic model, or f_1 covers the entire data range and f_2 is the empty component, yielding a single component Inverse Gaussian model. Solutions 3, 4 and 5 represent genuine composite model distributions, with Solution 3 providing the best model fit according to the NLL.

The densities of the five solutions are visualized in Figure 1. The solid, dashed, and narrow dashed lines represent composite models where the head and tail distributions are the same with identical parameter values ϑ_1 and ϑ_2 but different threshold values. The vertical lines indicate the thresholds of Solutions 3 (solid line) and 4 (dashed line). The threshold of Solution 5 is at 12.15 and hence not visible in the plot. The composite model which fits the Danish fire losses data set best is represented by the solid line. The dotted line corresponds to the Log-logistic distribution with the same parameter value ϑ_2 as in the composite models and the dashed-dotted line to an Inverse Gaussian distribution with the same parameter value ϑ_1 as in the composite models.

[Figure 1 about here.]

2.4 Model selection

In the area of loss modeling and curve-fitting, two commonly used model selection criteria appear in the literature: Akaike Information Criterion (Akaike, 1974) and Bayesian Information Criterion (Schwarz, 1978), known as AIC and BIC, respectively. The values of AIC and BIC are computed for each composite model using the following equations:

$$\text{AIC} = 2\text{NLL} + 2p,$$
$$\text{BIC} = 2\text{NLL} + p\ln(n),$$

where NLL represents the value of the negative log-likelihood function, p represents the number of estimated parameters in the model, and n is the sample size. The best model is selected by finding the lowest value of AIC or BIC. We compute both AIC and BIC for each of the composite models considered. However, we analyze the results with a special focus on BIC only.

3 Empirical analysis and results

In this section, we illustrate the proposed methodology using the well-known Danish fire losses data set, and then compare our findings to the previously published results for the same data set. The Danish fire losses data set contains 2,492 claims, reported in millions of 1985 Danish Krones (DKK), for the time period 1980 to 1990 inclusive. These losses are highly skewed with a skewness coefficient of 19.9. Standard insurance data sets share many common characteristics similar to those observed in the Danish fire losses; thus, the Danish fire data set has been very attractive for demonstrating the suitability of different proposed models in the area of curve-fitting and loss modeling. For the purpose of our analysis, we obtained the Danish fire losses data through the **SMPracticals** (Davison, 2013) package in R.

3.1 Goodness-of-fit

The parametric distributions considered for the head and/or the tail of the composite model are listed in Table 2. In total, sixteen parametric distributions are used, leading to a total of 256 different composite models. Table 2 also gives the parameters of the distributions and the probability density functions in order to unambiguously define the distributions, given the parameters.

[Table 2 about here.]

The composite models determined are first evaluated by comparing them to previously reported fitted models (see Appendix A). Table 3 contains the NLL values as well as the degrees of freedom, AIC, and BIC values for the composite models previously reported. The Log-normal-Pareto model with the same NLL value is also reported in Scollnik and Sun (2012) as well as Nadarajah and Bakar (2014). The same results for the Log-normal-Burr composite model are also reported in Nadarajah and Bakar (2014). In Bakar et al. (2015), the composite models with head distribution Weibull and all listed tail distributions are also considered. In addition, Scollnik and Sun (2012) also considered the Weibull-Pareto composite model. The same NLL values are reported in all cases, except for slight differences when the generalized Pareto and inverse Burr distribution are in the tail. A comparison of the estimated parameter values indicates that the slight differences might be due to the maximization algorithm being stopped prematurely for the previously reported results.

[Table 3 about here.]

[Table 4 about here.]

An overview of the model fit for all composite models considered is provided in Figure 2. The composite models are organized using the head distribution on the x-axis and the tail distribution on the y-axis. Each bubble on this plot represents the size of the individual model weight, w_i. The model weights are computed based on the approximate calculation of the posterior probability of each model, assuming equal prior model weights. Specifically, the formula $w_i = \frac{e^{-\frac{1}{2}\cdot \text{BIC}_i}}{\sum_{j=1}^{256} e^{-\frac{1}{2}\cdot \text{BIC}_j}}$ is used to determine the posterior model weights. More information about how to determine posterior model weights can be found in Burnham and Anderson (2003). The top 20 best models, based on the minimum value of BIC, have the largest weights calculated based on the posterior model probability. These 20 models are enumerated, and can be easily located on the grid by their corresponding tail and head distributions. More details on these top 20 best models are reported in Table 4 and in Appendix B.

[Figure 2 about here.]

Figure 2 reveals several interesting observations. Previous applications of composite models evolved around fitting either a Weibull or Log-normal distribution in the head. However, the results indicate that this approach is only partially valid. There are four distributions in the head that repeatedly produce the best results: Weibull, Paralogistic, Inverse Burr, and Log-logistic. The Log-normal distribution in the head does not show up in the top 20 best models. The top three models have a Weibull, Paralogistic or Inverse Burr distribution in the head. While the Weibull distribution was already widely considered in previous research, Inverse Burr received no consideration until the paper by Miljkovic and Grün (2016), which indicated that the Burr and Inverse Burr distributions could be used to obtain a superior fit to the Danish fire losses data. None of the previous applications considered a Paralogistic distribution in the head of a composite model fitted to the Danish fire losses. It is worth noting that some of the composite models that utilized the Paralogistic distribution in the head perform significantly better than all the models proposed by Cooray and Ananda (2005); Scollnik (2007); Pigeon and Denuit (2011); Scollnik and Sun (2012); Nadarajah and Bakar (2014). In addition, these composite models perform better than the models that used the Log-normal distribution in the head considered by Calderín-Ojeda and Kwok (2016). The only model that outperforms these top three models, based on the BIC, is the Weibull-Stoppa model proposed by Calderín-Ojeda and Kwok (2016).

The previous research on composite modeling had a strong focus on fitting the Pareto and Generalized Pareto distributions in the tail. However, our findings show that several distributions can be considered in the tail: Inverse Weibull, Inverse Paralogistic, Log-logistic, Burr, Inverse Gamma, and Paralogistic. The top three distributions based on BIC all have the Inverse Weibull distribution in the tail. Interestingly, none of the top 20 models include the Pareto or Generalized Pareto distributions in the tail. While there is a general perception that Pareto and Generalized Pareto distributions are good options for curve-fitting of heavy tail insurance losses, other distributions should not be ignored because they too exhibit excellent goodness-of-fit properties.

The top three models based on the BIC are Weibull-Inverse Weibull, Paralogistic-Inverse Weibull and Inverse Burr-Inverse Weibull. The fourth best model according to the BIC is the Weibull-Inverse Paralogistic composite model, which was already reported by Bakar et al. (2015). However, the authors did not pair Weibull with Inverse Weibull in order to achieve an even better fit to the Danish fire losses data. Miljkovic and Grün (2016) suggested that using the Burr or Inverse Burr distributions as the component distribution of a finite mixture model can provide a density with improved goodness-of-fit when estimated to the Danish fire loss data. The three top models that involve the Weibull, the Paralogistic and the Inverse Burr distributions in the head have comparable values of BIC (7671.30 vs. 7671.56 vs. 7671.79), which indicates that these three models are indistinguishable with respect to this criterion (the difference in BIC is less than 1).

In addition to NLL, AIC, and BIC which are used for model selection and determining overall goodness-of-fit, specific goodness-of-fit tests have been proposed in the literature related to curve-fitting for insurance losses in order to compare different models. In the following the results for the Kolmogorov-Smirnov, Anderson-Darling, and Chi-square goodness-of-fit tests are considered as in Lee and Lin (2010) and Miljkovic and Grün (2016). Table 5 gives these results for the composite models previously considered in the literature, and Table 6 those for the 20 best fitting models based on the BIC. For most of the previously considered composite models, as well as for all 20 of the best composite models, the Kolmogorov-Smirnov and the Anderson-Darling goodness-of-fit tests would retain the null hypothesis at a significance level of 5%, which indicates a suitable fit. With respect to the Chi-square test, the null hypothesis would be rejected at a significance level of 5%, except for the three composite models where the tail distribution is the Burr distribution, and for the Weibull-Inverse Weibull composite model. In these cases the p-value obtained is slightly above 5%.

The results for the Chi-square test need to be critically viewed. The Chi-square test requires the data to be binned before application and the results strongly depend on how the data is binned. Scollnik and Sun (2012) and Scollnik (2007) point out that the Chi-square test is not a good tool to compare model fit across multiple models under consideration in case of highly skewed data such as the Danish fire losses due to the issue of determining adequate class limits. For the presented results the data was grouped into 10 classes using the sample quantiles for the probabilities from 0 to 1 with increments of 0.10 as thresholds for the binning. The bins thus have unequal length while containing approximately the same number of observations. Chen and Miljkovic (2019) show through a simulation study that the Chi-square test is more sensitive for skewed than symmetric data in curve-fitting applications if quantiles and a small number of classes are used. Thus, it is not surprising that only four models passed the Chi-square test at the 5% significance level.

[Table 5 about here.]

[Table 6 about here.]

3.2 Risk measures

The most common application of curve-fitting models in risk management is to determine risk measures. Table 7 shows the summary of two risk measures, VaR and CTE at the 0.95 and 0.99 security levels for the composite models where results for the Danish fire losses were already reported in the literature. Table 8 provides a summary of the same risk measures for the 20 best fitting models reported in Table 4.

[Table 7 about here.]

[Table 8 about here.]

Figure 3 shows a panel plot of the VaR and CTE at the security level 0.99 for the 20 best models displayed in Table 8. The labels on the x-axis show the distributions used in the head and each of the panels correspond to a different distribution in the tail. The size of each bubble corresponds to the model weight, w_i. A few interesting observations can be drawn from this panel plot. First, there is an obvious pattern of clustering in the VaR and CTE values. The highest VaR and CTE values are observed in the three of the 20 models that have the Burr distribution in the tail: Weibull-Burr, Paralogistic-Burr, and Inverse Burr-Burr. This cluster with the three aforementioned models is well-separated from the remaining 17 models. Second, Weibull-Inverse Weibull, Paralogistic-Inverse Weibull, and Inverse Burr-Inverse Weibull do not produce substantially different values of VaR and CTE when compared to the other composite models, although they do have the smallest BIC values. Further, the minimum values of both VaR and CTE are observed for the Inverse Burr-Inverse Gamma composite model.

[Figure 3 about here.]

Based on AIC and BIC, the Weibull-Pareto and Weibull-Generalized Pareto models were not ranked among the top 20 best models in Table 8 even though their VaR and CTE values are among the closest to the empirical counterparts in comparison to the other models reported. The models in Table 8 provide the best overall fit to the data, not necessary the best fit in the tail area. In fact Blostein and Miljkovic (2019) argue that risk measures and model selection criteria should be simultaneously evaluated and considered on a grid across the entire space of models under consideration. This gives insights into the variability of risk measures and model selection criteria and may guide risk assessment. This also confirms Embrechts et al. (2014) who point out that model uncertainty is a crucial aspect when aiming at reliable risk measures.

4 Conclusion

Composite models have been embraced as a versatile model class for curve-fitting in actuarial science and statistics despite the computational issues faced in their estimation. In this paper, a comprehensive set of composite models obtained by combining 16 parametric distributions has been tested on the Danish fire losses data set. The computational tools developed are generally applicable once the probability density and cumulative distribution functions are available for the parametric distributions considered for the head and the tail. In the early phase of research on composite models, special attention was given to composite models that used the Log-normal distribution in the head and the Pareto distribution in the tail. Later on, the use of a Weibull distribution in the head received more attention. The most recent research used the Weibull distribution in the head, with various other distributions considered for the tail. The presented results expand this research by using a larger set of parametric distributions for the head as well as the tail.

Rather than taking a small incremental step in the search to discover a new best fitting model, we take a more comprehensive approach. We evaluated a very large set of composite models, consisting of 256 different models that involve 16 commonly used parametric distributions in actuarial science. Using the BIC model selection criterion, we identified the top 20 best fitting composite models for the Danish fire losses data set. For these models, the goodness-of-fit properties and risk measures were analyzed.

The findings of this study do not support previous approaches regarding the selection of distributions for the head and the tail in order to obtain the best fit for the Danish fire losses data set. Previous research focused mostly on the Log-normal and Weibull distributions for modeling the head of the composite model. Based on analyzing the BIC results, none of the top 20 best fitting models among the considered 256 composite models contain the Log-normal distribution in the head. On the other hand, using distributions such as Weibull, Paralogistic, and Inverse Burr in the head is found to be ideal for modeling the small and moderate size claims of Danish fire losses. The tail distributions such as Inverse Weibull, Inverse Paralogistic, Log-logistic, Burr, Inverse Gamma, and Paralogistic seem to be the best choices for modeling the long tail of Danish fire losses. Interestingly, the Pareto and Generalized Pareto distributions are not included among the top 20 best models based on the BIC.

Computational advances and numerical methods available in today's world enable us to test a variety of composite models, through pairing different distributions so that we can achieve an optimal fit to a given data set. We have seen that evaluating a subset of models based on some preconceived idea may result in a sub-optimal fit in the case of the Danish fire losses data set. In this instance, the calculation of risk measures will be subject to greater uncertainty, as there is no guarantee that an accurate amount of capital will be reserved for an adverse outcome.

Future work in this area might consider adjustments of the proposed method to accommodate composite modeling of incomplete data, such as the left-truncated loss data (see Brazauskas and Kleefeld, 2016) or the right-censored benefit data (see Miljkovic and Barabanov 2015 and Miljkovic and Orr 2017).

Acknowledgments

We would like to thank the reviewers for their feedback which helped improve the paper. The discussion of additional risk measures based on higher order moments and the calculation of standard errors was added based on their comments. The authors are grateful for the comments received from Ms. Lisa Werwinski that helped improve the readability of the paper.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Bakar, S. A. A., Hamzaha, N. A., Maghsoudia, M., Nadarajah, S., 2015. Modeling loss data using composite models. Insurance: Mathematics and Economics 61, 146–154.

Blostein, M., Miljkovic, T., 2019. On modeling left-truncated loss data using mixtures of distributions. Insurance: Mathematics and Economics, in Press.

Borchers, H. W., 2017. pracma: Practical Numerical Math Functions. R package version 2.1.1.
URL `https://CRAN.R-project.org/package=pracma`

Brazauskas, V., Kleefeld, A., 2016. Modeling severity and measuring tail risk of norwegian fire cliams. North American Actuarial Journal 20 (1), 1–16.

Burnham, K. P., Anderson, D. R., 2003. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 11th Edition. Springer Science & Business Media.

Calderín-Ojeda, E., Kwok, C. F., 2016. Modeling claims data with composite Stoppa models. Scandinavian Actuarial Journal 9, 817–836.

Chen, Y., Miljkovic, T., 2019. From grouped to de-grouped data: a new approach in distribution fitting for grouped data. Journal of Statistical Computation and Simulation 89 (2), 272–291.

Ciumara, R., 2006. An actuarial model based on the composite Weibull-Pareto distribution. Mathematical Reports-Bucharest 8 (4), 401–414.

Cooray, K., 2009. The Weibull-Pareto composite family with applications to the analysis of unimodal failure rate data. Communications in Statistics – Theory and Methods 38 (11), 1901–1915.

Cooray, K., Ananda, M. M., 2005. Modeling actuarial data with a composite lognormal-Pareto model. Scandinavian Actuarial Journal 61 (5), 321–334.

Davison, A., 2013. SMPracticals: Practicals for Use with Davison (2003) "Statistical Models". R package version 1.4-2.
URL http://CRAN.R-project.org/package=SMPracticals

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., Beleraj, A., 2014. An academic response to basel 3.5. Risks 2 (1), 25–48.

Gilbert, P., Varadhan, R., 2016. numDeriv: Accurate Numerical Derivatives. R package version 2016.8-1.
URL https://CRAN.R-project.org/package=numDeriv

Klugman, S. A., Panjer, H. H., Willmot, G. E., 2012. Loss Models: From Data to Decisions, 4th Edition. John Wiley & Sons, Hobuken, NJ.

Lee, S. C. K., Lin, X. S., 2010. Modeling and evaluating insurance losses via mixtures of Erlang distributions. North American Actuarial Journal 14 (1), 107–130.

Maria Sarabia, J., Calderín-Ojeda, E., 2018. Analytical expressions of risk quantities for composite models. JOURNAL OF RISK MODEL VALIDATION 12 (3), 75–101.

Miljkovic, T., Barabanov, N., 2015. Modeling veterans – health benefit grants using the expectation maximization algorithm. Journal of Applied Statistics 42 (6), 1166–1182.

Miljkovic, T., Grün, B., 2016. Modeling loss data using mixtures of distributions. Insurance: Mathematics and Economics 70, 387–396.

Miljkovic, T., Orr, M., 2017. An evaluation of the reconstructed coefficient of determination and potential adjustments. Communications in Statistics – Simulation and Computation 46 (9), 6705–6718.

Nadarajah, S., Bakar, S., 2014. New composite models for the Danish fire insurance data. Scandinavian Actuarial Journal 2, 180–187.

Pigeon, M., Denuit, M., 2011. Composite lognormal-Pareto model with random threshold. Scandinavian Actuarial Journal 3, 177–192.

Punzo, A., Bagnato, L., Maruotti, A., 2018. Compound unimodal distributions for insurance losses. Insurance: Mathematics and Economics 81, 95–107.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6 (2), 461–464.

Scollnik, D., 2007. On composite lognormal-Pareto models. Scandinavian Actuarial Journal 1, 20–33.

Scollnik, D. P., Sun, C., 2012. Modeling with Weibull-Pareto models. North American Actuarial Journal 16, 260–272.

Sherlock, C., Fearnhead, P., Roberts, G. O., 2010. The random walk Metropolis: Linking theory and practice through a case study. Statistical Science 25 (2), 172–190.

A Previous results

Table 9 summarizes the results previously reported for the Log-normal-Pareto and Log-normal-Burr composite models as well as several composite models with the Weibull distribution in the head and eight different parametric distributions in the tail. These models, together with the reported results, are included in Cooray and Ananda (2005); Scollnik (2007); Pigeon and Denuit (2011); Scollnik and Sun (2012); Bakar et al. (2015); Nadarajah and Bakar (2014); Bakar et al. (2015). In addition the Stoppa and Lomax models considered by Calderín-Ojeda and Kwok (2016) are also included.

[Table 9 about here.]

B Detailed results

The parameter estimates together with the standard errors are given in Tables 10 and 11 for the 20 best fitting composite models (according to the BIC).

[Table 10 about here.]

[Table 11 about here.]

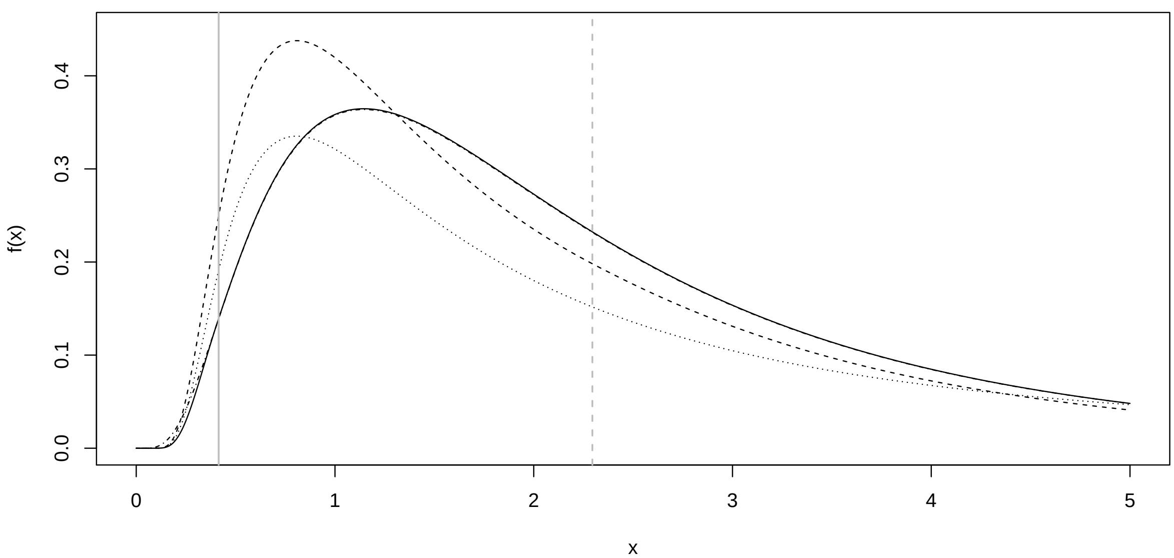

Figure 1: Density plot of the solutions summarized in Table 1. The line style for each density is displayed as follows: Solution 1 (dotted line), Solution 2 (dashed-dotted line), Solution 3 (solid line), Solution 4 (dashed line) and Solution 5 (narrow dashed line). The vertical lines indicate the θ values for Solutions 3 and 4 in the respective line styles.

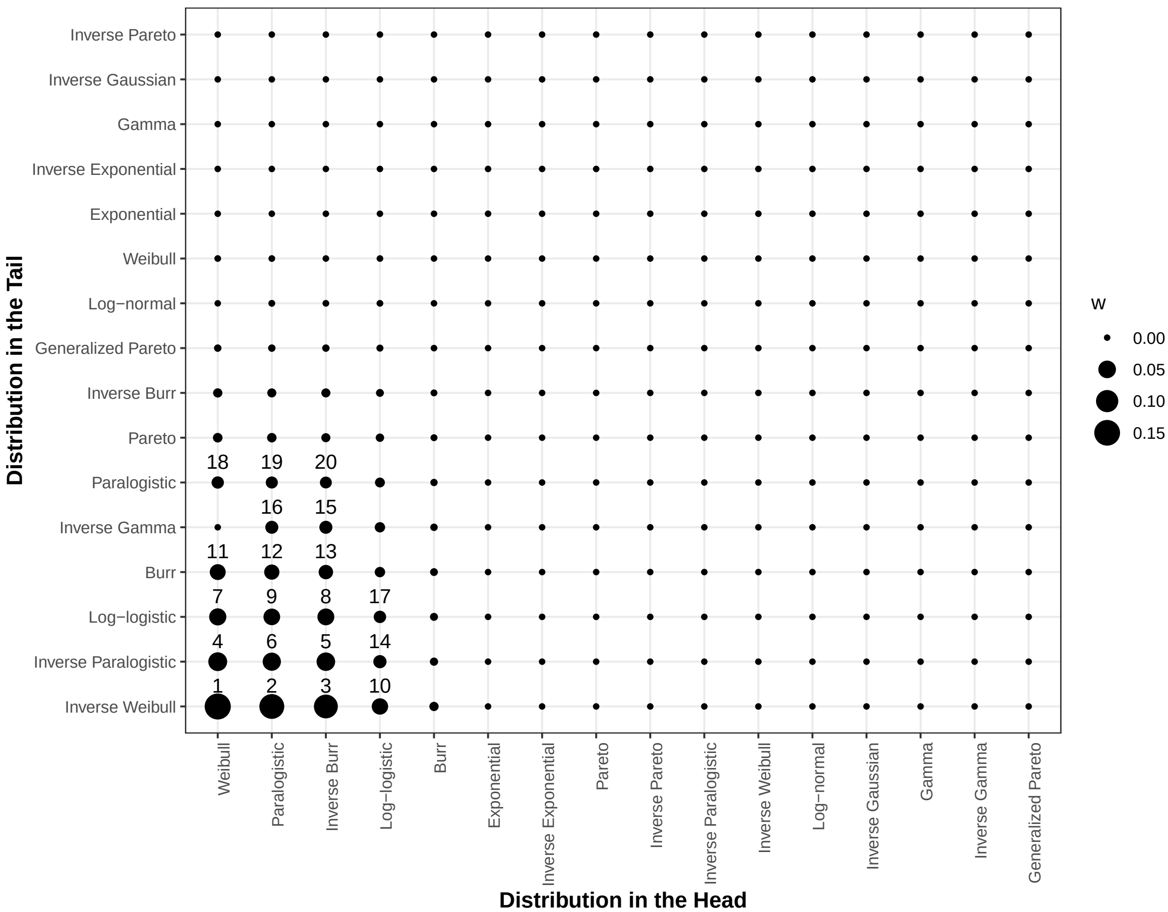

Figure 2: Summary of the composite models. The head distribution is shown on the x-axis, while the tail distribution is shown on the y-axis. The size of each bubble represents the model weight, computed as $w_i = \frac{e^{-\frac{1}{2}\cdot \text{BIC}_i}}{\sum_{j=1}^{256} e^{-\frac{1}{2}\cdot \text{BIC}_j}}$. The top 20 best models, based on the minimum value of BIC, have the largest weights based on the estimated posterior model probability. They are enumerated using the numbers 1 through 20.

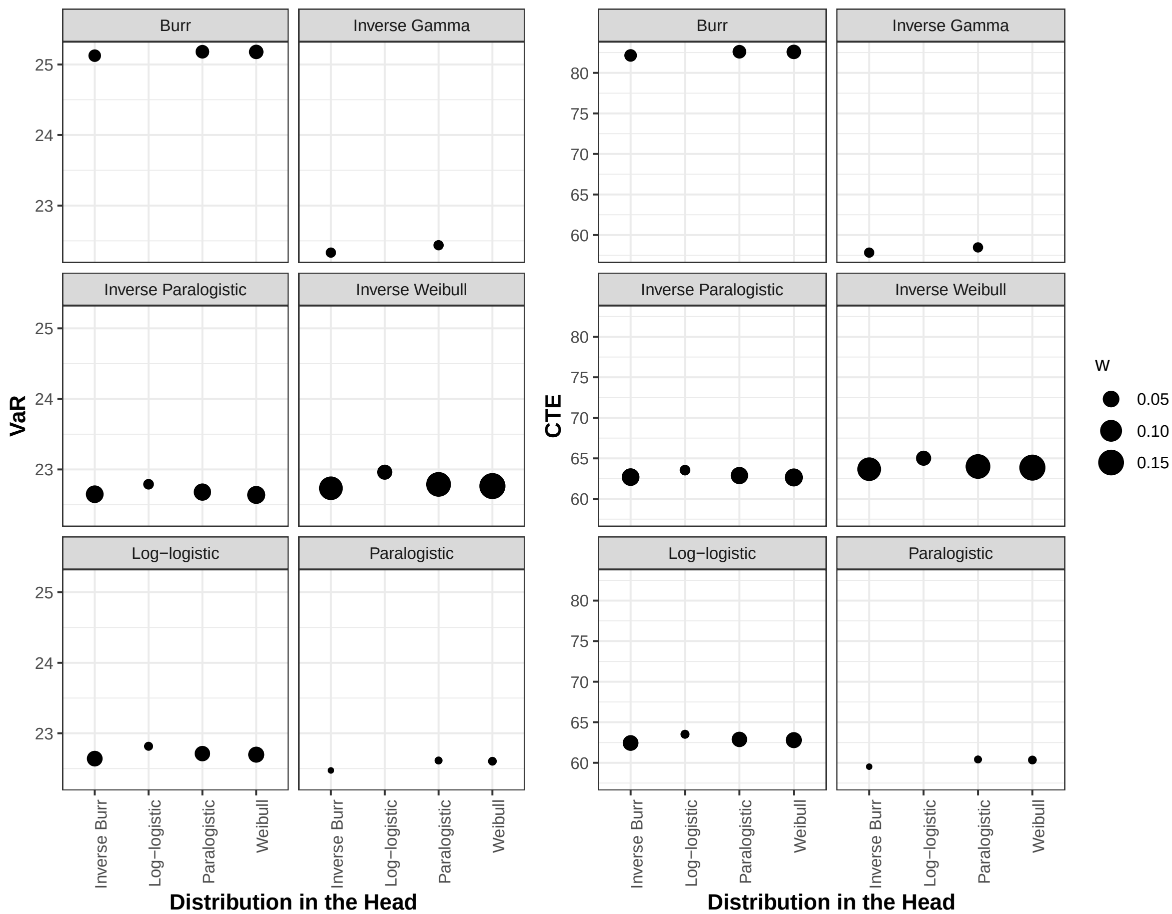

Figure 3: Panel plot for the visual summary of the $\text{VaR}_{0.99}$ (left) and $\text{CTE}_{0.99}$ (right) for the 20 best fitting composite models. The labels on the x-axis show the head distribution. Each panel contains the results for the tail distribution given by the strip name. The size of each bubble represents the model weight, computed as $w_i = \frac{e^{-\frac{1}{2}\cdot \text{BIC}_i}}{\sum_{j=1}^{256} e^{-\frac{1}{2}\cdot \text{BIC}_j}}$.

Possible Solutions	θ	ϕ	NLL	$\text{VaR}_{0.99}$	$\text{CTE}_{0.99}$
Solution 1 (Inverse Gaussian)	∞	0.00	4516.33	14.43	18.44
Solution 2 (Log-logistic)	0.00	∞	4281.05	10.19	16.53
Solution 3 (Inverse Gaussian-Log-logistic)	0.58	23.80	4249.94	10.24	16.61
Solution 4 (Inverse Gaussian-Log-logistic)	1.63	1.04	4313.09	9.92	16.09
Solution 5 (Inverse Gaussian-Log-logistic)	12.15	0.02	4392.69	16.83	27.26

Table 1: Summary of the non-unique solutions for a composite model with respect to θ and ϕ given ϑ_1 and ϑ_2. Both VaR and CTE are computed at the 0.99 security level.

Distribution	Parameters	Density
Burr	α, β, $\tau = 1/\sigma$	$f(x) = \frac{\alpha\beta(x/\sigma)^{\beta}}{x(1+(x/\sigma)^{\beta})^{\alpha+1}}$
Exponential	λ	$f(x) = \lambda e^{-\lambda x}$
Gamma	α, $\tau = 1/\sigma$	$f(x) = \frac{x^{\alpha-1}}{\sigma^{\alpha}\Gamma(\alpha)} e^{-(x/\sigma)}$
Generalized Pareto	α, β, $\tau = 1/\sigma$	$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\sigma^{\alpha} x^{\beta-1}}{(x+\sigma)^{\alpha+\beta}}$
Inverse Burr	α, β, $\tau = 1/\sigma$	$f(x) = \alpha\beta \frac{(x/\sigma)^{\alpha\beta}}{x(1+(x/\sigma)^{\beta})^{\alpha+1}}$
Inverse Exponential	λ	$f(x) = \lambda e^{-\lambda/x} \frac{1}{x^2}$
Inverse Gamma	α, $\tau = 1/\sigma$	$f(x) = \frac{1}{\sigma^{\alpha}\Gamma(\alpha)} x^{-(\alpha+1)} e^{-(1/(\sigma x))}$
Inverse Gaussian	μ, $\tau = 1/\sigma$	$f(x) = \sqrt{\frac{1}{2\pi\sigma x^3}} e^{-\frac{(x-\mu)^2}{2\mu^2\sigma x}}$
Inverse Paralogistic	α, $\tau = 1/\sigma$	$f(x) = \alpha^2 \frac{(x/\sigma)^{\alpha^2}}{x(1+(x/\sigma)^{\alpha})^{(\alpha+1)}}$
Inverse Pareto	α, σ	$f(x) = \alpha\sigma \frac{x^{\alpha-1}}{(x+\sigma)^{(\alpha+1)}}$
Inverse Weibull	α, $\tau = 1/\sigma$	$f(x) = \alpha(\sigma/x)^{\alpha} e^{-(\sigma/x)^{\alpha}} \frac{1}{x}$
Log-logistic	α, σ	$f(x) = \alpha \frac{(x/\sigma)^{\alpha}}{x(1+(x/\sigma)^{\alpha})^2}$
Log-normal	μ, σ	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-((\log(x)-\mu)^2/(2\sigma^2))}$
Paralogistic	α, $\tau = 1/\sigma$	$f(x) = \alpha^2 \frac{(x/\sigma)^{\alpha}}{x(1+(x/\sigma)^{\alpha})^{(\alpha+1)}}$
Pareto	α, σ	$f(x) = \frac{\alpha\sigma^{\alpha}}{(x+\sigma)^{(\alpha+1)}}$
Weibull	α, β	$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{(\alpha-1)} e^{-(x/\beta)^{\alpha}}$

Table 2: Sixteen parametric distributions considered as the tail and/or head distributions in the composite model.

Head	Tail	NLL	p	AIC	BIC
Log-normal	Pareto	3860.47	4	7728.94	7752.23
Log-normal	Burr	3835.12	5	7680.24	7709.34
Weibull	Burr	3817.57	5	7645.14	7674.24
Weibull	Log-logistic	3821.23	4	7650.46	7673.74
Weibull	Paralogistic	3822.44	4	7652.88	7676.17
Weibull	Generalized Pareto	3822.13	5	7654.25	7683.36
Weibull	Pareto	3823.70	4	7655.40	7678.68
Weibull	Inverse Burr	3820.01	5	7650.02	7679.12
Weibull	Inverse Pareto	3894.88	4	7797.77	7821.05
Weibull	Inverse Paralogistic	3820.93	4	7649.87	7673.15

Table 3: Summary of the results obtained for the previously fitted composite models.

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

Table 4: Summary of results for the 20 best fitting composite models (according to the BIC).

		Kolmogorov-Smirnov		Anderson-Darling		Chi-square	
Head	Tail	T	p-value	T	p-value	T	p-value
Log-normal	Pareto	0.020	0.295	1.950	0.098	23.2	0.006
Log-normal	Burr	0.038	0.001	3.427	0.017	35.6	0.000
Weibull	Burr	0.015	0.655	0.715	0.547	12.3	0.197
Weibull	Log-logistic	0.021	0.214	1.373	0.210	18.4	0.031
Weibull	Paralogistic	0.023	0.151	1.627	0.149	19.5	0.021
Weibull	Generalized Pareto	0.022	0.171	1.565	0.162	19.3	0.023
Weibull	Pareto	0.026	0.078	1.916	0.102	25.4	0.003
Weibull	Inverse Burr	0.021	0.222	1.159	0.284	16.3	0.060
Weibull	Inverse Pareto	0.079	0.000	32.714	0.000	136.9	0.000
Weibull	Inverse Paralogistic	0.021	0.210	1.318	0.227	17.0	0.049

Table 5: Results of goodness-of-fit tests for the models corresponding to previously fitted composite models.

		Kolmogorov-Smirnov		Anderson-Darling		Chi-square	
Head	Tail	T	p-value	T	p-value	T	p-value
Weibull	Inverse Weibull	0.021	0.222	1.159	0.284	16.3	0.060
Paralogistic	Inverse Weibull	0.021	0.226	1.156	0.285	17.1	0.047
Inverse Burr	Inverse Weibull	0.021	0.216	1.160	0.283	17.1	0.047
Weibull	Inverse Paralogistic	0.021	0.210	1.318	0.227	17.0	0.049
Inverse Burr	Inverse Paralogistic	0.021	0.211	1.327	0.224	17.1	0.047
Paralogistic	Inverse Paralogistic	0.021	0.217	1.311	0.229	17.6	0.040
Weibull	Log-logistic	0.021	0.214	1.373	0.210	18.4	0.031
Inverse Burr	Log-logistic	0.021	0.204	1.398	0.203	17.7	0.038
Paralogistic	Log-logistic	0.021	0.216	1.369	0.211	18.9	0.026
Log-logistic	Inverse Weibull	0.020	0.255	1.148	0.288	17.8	0.038
Weibull	Burr	0.015	0.655	0.715	0.547	12.3	0.197
Paralogistic	Burr	0.015	0.656	0.715	0.546	12.3	0.197
Inverse Burr	Burr	0.015	0.636	0.711	0.550	12.4	0.191
Log-logistic	Inverse Paralogistic	0.021	0.235	1.288	0.236	18.9	0.026
Inverse Burr	Inverse Gamma	0.023	0.155	1.612	0.152	18.4	0.031
Paralogistic	Inverse Gamma	0.022	0.172	1.561	0.163	19.3	0.023
Log-logistic	Log-logistic	0.021	0.234	1.340	0.220	19.9	0.018
Weibull	Paralogistic	0.023	0.151	1.627	0.149	19.5	0.021
Paralogistic	Paralogistic	0.023	0.152	1.621	0.150	19.5	0.021
Inverse Burr	Paralogistic	0.023	0.137	1.705	0.134	20.6	0.015

Table 6: Results of goodness-of-fit tests for the 20 best fitting composite models (according to the BIC).

		$\text{VaR}_{0.95}$	$\text{VaR}_{0.99}$	$\text{CTE}_{0.95}$	$\text{CTE}_{0.99}$
Empirical estimates		8.41	24.61	22.16	54.60
Head	Tail				
Log-normal	Pareto	8.25	23.75	23.54	66.57
Log-normal	Burr	9.07	30.98	38.32	130.93
Weibull	Burr	8.22	25.18	26.98	82.59
Weibull	Log-logistic	8.05	22.70	22.43	62.80
Weibull	Paralogistic	8.11	22.60	21.98	60.35
Weibull	Generalized Pareto	8.11	22.43	21.56	58.43
Weibull	Pareto	8.20	22.65	21.63	58.21
Weibull	Inverse Burr	8.02	22.77	22.64	63.86
Weibull	Inverse Pareto	17.35	86.77	–	–
Weibull	Inverse Paralogistic	8.03	22.64	22.38	62.65

Table 7: Summary of the results for VaR and CTE at the 0.95 and 0.99 security level, obtained for previously fitted composite models. Empirical estimates are provided for comparison purposes. The missing value of CTE indicated by "–" is due to the non-existence of the moments for the Inverse Pareto distribution.

		$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.99}$	$\mathrm{CTE}_{0.95}$	$\mathrm{CTE}_{0.99}$
Empirical estimates		8.41	24.61	22.16	54.60
Head	Tail				
Weibull	Inverse Weibull	8.02	22.77	22.64	63.86
Paralogistic	Inverse Weibull	8.02	22.79	22.67	64.00
Inverse Burr	Inverse Weibull	8.01	22.73	22.59	63.67
Weibull	Inverse Paralogistic	8.03	22.64	22.38	62.65
Inverse Burr	Inverse Paralogistic	8.03	22.65	22.39	62.69
Paralogistic	Inverse Paralogistic	8.03	22.68	22.44	62.89
Weibull	Log-logistic	8.05	22.70	22.43	62.80
Inverse Burr	Log-logistic	8.04	22.64	22.35	62.46
Paralogistic	Log-logistic	8.05	22.71	22.46	62.89
Log-logistic	Inverse Weibull	8.05	22.96	22.93	65.02
Weibull	Burr	8.22	25.18	26.98	82.59
Paralogistic	Burr	8.22	25.18	26.98	82.61
Inverse Burr	Burr	8.22	25.13	26.88	82.15
Log-logistic	Inverse Paralogistic	8.05	22.79	22.60	63.55
Inverse Burr	Inverse Gamma	8.10	22.33	21.42	57.83
Paralogistic	Inverse Gamma	8.11	22.44	21.57	58.48
Log-logistic	Log-logistic	8.06	22.82	22.61	63.52
Weibull	Paralogistic	8.11	22.60	21.98	60.35
Paralogistic	Paralogistic	8.11	22.62	21.99	60.41
Inverse Burr	Paralogistic	8.10	22.47	21.78	59.52

Table 8: Summary of the results for VaR and CTE at the 0.95 and 0.99 security level for the 20 best fitting composite models in ascending order based on the BIC value. Empirical estimates are provided for comparison purposes.

Model	p	NLL	AIC	BIC	Reference
Log-normal-Pareto (1)	2	3877.84	7759.69	7771.33	Cooray and Ananda (2005)
Log-normal-Pareto (2)	2	3878.67	7761.34	7772.98	Cooray and Ananda (2005)
Log-normal-Pareto (1)	2	3877.84	7759.69	7771.33	Scollnik (2007)
Log-normal-Pareto (2)	3	3864.89	7735.79	7753.25	Scollnik (2007)
Log-normal-Pareto (3)	4	3858.83	7725.65	7748.94	Scollnik (2007)
Log-normal-Pareto (4)	3	4422.99	8851.97	8869.43	Scollnik (2007)
Model (1)	2	3878.00	7760.00	7748.76	Pigeon and Denuit (2011)
Model (2)	3	3866.00	7739.00	7755.46^A	Pigeon and Denuit (2011)
Model (3)	4	3860.00	7728.00	7751.28	Pigeon and Denuit (2011)
Model (4)	4	3860.00	7728.00	7751.28	Pigeon and Denuit (2011)
Log-normal-Pareto (2)	3	3865.86	7737.73	7755.19	Scollnik and Sun (2012)
Log-normal-Pareto (3)	4	3860.47	7728.94	7752.23	Scollnik and Sun (2012)
Weibull-Pareto (1)	2	3959.01	7922.01	7933.65	Scollnik and Sun (2012)
Weibull-Pareto (2)	3	3840.38	7686.75	7704.22	Scollnik and Sun (2012)
Weibull-Pareto (3)	4	3823.70	7655.40	7678.68	Scollnik and Sun (2012)
Log-normal-Pareto	4	3860.47	7728.94	7752.22	Nadarajah and Bakar (2014)
Log-normal-Burr	5	3857.83	7725.65	7754.76	Nadarajah and Bakar (2014)
Weibull-Burr	5	3817.57	7645.14	7674.24	Bakar et al. (2015)
Weibull-Loglogistic	4	3821.23	7650.46	7673.74	Bakar et al. (2015)
Weibull-Paralogistic	4	3822.44	7652.88	7676.16	Bakar et al. (2015)
Weibull-Gen. Pareto	5	3822.75	7655.50	7684.61	Bakar et al. (2015)
Weibull-Pareto	4	3823.70	7655.40	7678.68	Bakar et al. (2015)
Weibull-Inv. Burr	5	3821.20	7652.40	7681.51	Bakar et al. (2015)
Weibull-Inv. Pareto	4	3894.88	7797.77	7821.05	Bakar et al. (2015)
Weibull-Inv. Paralogistic	4	3820.93	7649.87	7673.15	Bakar et al. (2015)
Log-normal-Stoppa	4	3858.74	7717.48^B	7748.76	Calderín-Ojeda and Kwok (2016)
Log-normal-Lomax	4	3860.47	7728.94	7752.22	Calderín-Ojeda and Kwok (2016)
Log-normal-Pareto	4	3865.86	7737.72	7755.18	Calderín-Ojeda and Kwok (2016)
Weibull-Stoppa	4	3818.82	7645.64	7668.92	Calderín-Ojeda and Kwok (2016)
Weibull-Lomax	4	3823.70	7655.40	7676.68	Calderín-Ojeda and Kwok (2016)
Weibull-Pareto	3	3840.38	7686.76	7704.22	Calderín-Ojeda and Kwok (2016)

Table 9: Summary of results reported for previously fitted composite models. p denotes the number of estimated parameters.

Exceptions: $^{(A)}$The BIC value is calculated based on the NLL provided in the original paper by Pigeon and Denuit (2011). $^{(B)}$The AIC value is calculated based on NLL and it defers from the value of 7717.48 reported in the original paper by Calderín-Ojeda and Kwok (2016). However, our calculation agrees with the BIC value reported.

Head	Tail	ϑ_1	ϑ_2	θ	ϕ
Weibull	Inverse Weibull	$16.094, 0.955$	$1.555, 1.102$	0.955	9.854
		$(1.553, 0.0149)$	$(0.0505, 0.0994)$		
Paralogistic	Inverse Weibull	$16.088, 0.879$	$1.554, 1.105$	0.957	9.688
		$(1.581, 0.0232)$	$(0.0507, 0.101)$		
Inverse Burr	Inverse Weibull	$0.204, 68.745, 1.046$	$1.557, 1.096$	0.934	12.609
		$(0.235, 73.872, 0.0176)$	$(0.0493, 0.0919)$		
Weibull	Inverse Paralogistic	$15.806, 0.960$	$1.567, 1.777$	0.961	9.256
		$(1.728, 0.0176)$	$(0.0565, 0.213)$		
Inverse Burr	Inverse Paralogistic	$3.83 \cdot 10^{-4}, 35290, 1.078$	$1.567, 1.775$	0.928	14.086
		$(1.14 \cdot 10^{-5}, 1049, 5.27 \cdot 10^{-5})$	$(0.053, 0.181)$		
Paralogistic	Inverse Paralogistic	$15.745, 0.872$	$1.566, 1.787$	0.964	9.054
		$(1.968, 0.0301)$	$(0.057, 0.221)$		
Weibull	Log-logistic	$15.652, 0.962$	$1.568, 0.680$	0.964	9.030
		$(1.939, 0.0206)$	$(0.0593, 0.0979)$		
Inverse Burr	Log-logistic	$3.95 \cdot 10^{-4}, 34285, 1.078$	$1.570, 0.688$	0.928	14.020
		$(2.24 \cdot 10^{-5}, 1.517, 4.69 \cdot 10^{-5})$	$(0.0278, 0.00275)$		
Paralogistic	Log-logistic	$15.683, 0.871$	$1.567, 0.678$	0.965	8.906
		$(1.205, 0.0174)$	$(0.0569, 0.0888)$		
Log-logistic	Inverse Weibull	$16.267, 0.975$	$1.547, 1.130$	0.976	8.216
		$(1.264, 0.0127)$	$(0.0502, 0.105)$		

Table 10: Parameter estimates and standard errors in parenthesis obtained for the 20 best fitting composite models (according to the BIC). Reporting these estimates helps to reproduce the fitted distributions and provides insights into the variability as assessed using a procedure based on the Hessian of the log-likelihood and the Delta method which relies on asymptotics. - Part I.

Head	Tail	ϑ_1	ϑ_2	θ	ϕ
Log-logistic	Inverse Weibull	$16.267, 0.975$	$1.547, 1.130$	0.976	8.216
		$(1.264, 0.0127)$	$(0.0502, 0.105)$		
Weibull	Burr	$16.202, 0.949$	$0.395, 3.646, 1.182$	0.947	11.135
		$(1.263, 0.0107)$	$(0.104, 0.880, 0.0693)$		
Paralogistic	Burr	$16.278, 0.887$	$0.394, 3.649, 1.182$	0.947	11.043
		$(1.257, 0.0166)$	$(0.104, 0.884, 0.0697)$		
Inverse Burr	Burr	$0.262, 53.766, 1.046$	$0.406, 3.549, 1.190$	0.932	13.251
		$(0.108, 19.976, 0.0099)$	$(0.107, 0.853, 0.0713)$		
Log-logistic	Inverse Paralogistic	$16.197, 0.977$	$1.561, 1.819$	0.980	7.876
		$(1.358, 0.0141)$	$(0.0554, 0.216)$		
Inverse Burr	Inverse Gamma	$4.43 \cdot 10^{-4}, 30761, 1.078$	$1.641, 1.148$	0.928	13.945
		$(2.25 \cdot 10^{-8}, 0.885, 3.15 \cdot 10^{-5})$	$(0.0399, 0.0263)$		
Paralogistic	Inverse Gamma	$15.635, 0.869$	$1.635, 1.119$	0.967	8.753
		$(1.285, 0.0186)$	$(0.0733, 0.188)$		
Log-logistic	Log-logistic	$16.153, 0.978$	$1.562, 0.666$	0.981	7.761
		$(1.387, 0.0145)$	$(0.0573, 0.0911)$		
Weibull	Paralogistic	$15.511, 0.965$	$1.267, 1.607$	0.968	8.660
		$(1.314, 0.0129)$	$(0.0273, 0.265)$		
Paralogistic	Paralogistic	$15.557, 0.867$	$1.266, 1.611$	0.969	8.551
		$(1.355, 0.020)$	$(0.0273, 0.267)$		
Inverse Burr	Paralogistic	$9.29 \cdot 10^{-4}, 14718, 1.077$	$1.270, 1.559$	0.928	13.775
		$(1.4 \cdot 10^{-6}, 2.036, 1.52 \cdot 10^{-5})$	$(0.0136, 0.0307)$		

Table 11: Parameter estimates and standard errors in parenthesis obtained for the 20 best fitting composite models (according to the BIC). Reporting these estimates helps to reproduce the fitted distributions and provides insights into the variability as assessed using a procedure based on the Hessian of the log-likelihood and the Delta method which relies on asymptotics. – Part II.

Algorithm 1: Numerical Computation of the Maximum Likelihood Estimates.

Data: $\boldsymbol{x} = (x_1, \ldots, x_n)$.
Select the parametric distributions for the tail and head.
Define their PDFs and CDFs: $f_1(x|\vartheta_1), f_2(x|\vartheta_2), F_1(x|\vartheta_1), F_2(x|\vartheta_2)$.
Provide a function h_1 which given ϑ_1 and ϑ_2 returns the values of θ and ϕ, which are compliant with Equations (2.1) and (2.2).
Provide a function h_2 which evaluates the log-likelihood $h_2 = \ell(\vartheta_1, \vartheta_2|\boldsymbol{x})$ based on h_1.
Initialize the parameters: $\vartheta_1^{(0)}$ and $\vartheta_2^{(0)}$.
Maximize h_2: $\vartheta_1^{(opt)}$, $\vartheta_2^{(opt)} \leftarrow \arg\max_{\vartheta} h_2(\cdot)$.