Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Analyzing insurance data with an exponentiated composite inverse Gamma-Pareto model

Bowen Liu & Malwane M. A. Ananda

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2022.2050399

Analyzing insurance data with an exponentiated composite inverse Gamma-Pareto model

Bowen Liu and Malwane M. A. Ananda

Department of Mathematical Sciences, University of Nevada, Las Vegas, Nevada, USA

ABSTRACT
Exponentiated models have been widely used in modeling various types of data such as survival data and insurance claims data. However, the exponentiated composite distribution models have not been explored yet. In this paper, we introduce an improvement of the one-parameter Inverse Gamma-Pareto composite model by exponentiating the random variable associated with the one-parameter Inverse Gamma-Pareto composite distribution function. The goodness-of-fit of the exponentiated Inverse Gamma-Pareto was assessed using three different insurance data sets. The two-parameter exponentiated Inverse Gamma-Pareto model outperforms the one-parameter Inverse Gamma-Pareto model in terms of goodness-of-fit measures for all datasets. In addition, the proposed exponentiated composite Inverse Gamma-Pareto model provides a very good fit with some well-known insurance datasets.

ARTICLE HISTORY
Received 6 May 2021
Accepted 2 March 2022

KEYWORDS
Composite models; goodness-of-fit; inverse-gamma distribution; Pareto distribution; exponentiated models; insurance data modeling

1. Introduction

Modeling claim size data is one of the major topic in actuarial science. Actuaries often make decisions on financial risk management based on models. Thus, the selection of a proper model for claim sizes is a key task in the actuarial industry. Under normal circumstances, a claim size data set consists of a large number of claims with small sizes and few claims with large sizes. The common distributions in the literature such as exponential, normal, etc. do not have the ability to incorporate all the features of a claim size data set. Hence, the concept of composite distribution was introduced for modeling claim size data. With such concept, many different composite models were developed including lognormal-Pareto (Cooray and Ananda 2005; Scollnik 2007), exponential-Pareto (Teodorescu and Vernic 2006), Weibull-Pareto (Preda and Ciumara 2006), and so on. The idea of general composite model was introduced later (Abu Bakar et al. 2015). With such idea, a large number of possible composite models been explored (Grün and Miljkovic 2019). In general, Pareto distribution is considered good for modeling claims with large size. However, for modeling claims with small size, there are many variations in the literature.

Aminzadeh and Deng introduced the Inverse Gamma-Pareto (IG-Pareto) model recently (Aminzadeh and Deng 2019) and it was suggested as a possible model for data

CONTACT Bowen Liu bowen.liu@unlv.edu; Malwane M. A. Ananda malwane.ananda@unlv.edu Department of Mathematical Sciences, University of Nevada, Las Vegas, 89154 NV, USA.

sets with a very heavy tail such as insurance data sets. This is a one-parameter IG-Pareto composite distribution with appealing properties such as continuity and differentiability. However, fitting a one-parameter IG-Pareto model to several insurance data sets does not provide satisfactory performance, as we will show in the Numerical Examples section. Specifically, the mode of fitted IG-Pareto distribution is not large enough to describe the small claims with high frequencies within these insurance data sets. Therefore, we will modify this one-parameter IG-Pareto model by introducing an additional parameter.

Exponentiated distributions were first introduced by Mudholkar and Srivastava (1993). The main idea of exponentiated distributions is to exponentiate the Cumulative Density Function (CDF) of an existing distribution. It adds more flexibility to the traditional models due to the extra parameter. Many modifications of the existing distributions were later introduced following the idea of Mudholkar and Srivastava. For instance, Gupta and Kundu introduced exponentiated exponential (Gupta and Kundu 1999); Nadarajah pioneered exponentiated beta, exponentiated Pareto and exponentiated Gamma (Nadarajah 2005a, 2005b, 2006); Nadarajah and Gupta initiated exponentiated Gamma (Nadarajah and Gupta 2007) and Afify established exponentiated Weibull-Pareto (Afify et al. 2016). However, none of these models were established using CDF of a composite distribution. Moreover, all the exponentiated distributions mentioned above were created by exponentiating the CDF, while the exponentiated Inverse-Gamma model we propose was constructed by exponentiating the random variable associated with the CDF of a composite distribution.

The rest of the paper is organized as follows. Section 2 provides the derivation of exponentiated IG-Pareto model, the description of its behaviors and an algorithm to obtain the maximum likelihood estimators of the model. We briefly summarize the results from simulation studies in Section 3 to assess the accuracy and consistency of the MLE. In Section 4, three numerical examples are presented. Conclusions are provided in Section 5.

2. Methodology

2.1. Introduction of the general composite model in loss data modeling

Let X be a positive real-valued random variable. The general form of a composite model in loss data modeling was formally introduced (Scollnik 2007; Abu Bakar et al. 2015) as follows:

$$f_X(x|\alpha_1,\alpha_2,\theta,\phi)=\begin{cases}\dfrac{1}{1+\phi}f_1^*(x|\alpha_1,\theta) & 0<x\le\theta\\ \dfrac{\phi}{1+\phi}f_2^*(x|\alpha_2,\theta) & \theta<x<\infty\end{cases}$$

along with the continuity and differentiability conditions at the threshold θ:

$$\begin{cases}\lim_{x\to\theta^-}f_X(x|\alpha_1,\alpha_2,\theta,\phi)=\lim_{x\to\theta^+}f_X(x|\alpha_1,\alpha_2,\theta,\phi)\\ \lim_{x\to\theta^-}\dfrac{df_X(x|\alpha_1,\alpha_2,\theta,\phi)}{dx}=\lim_{x\to\theta^+}\dfrac{df_X(x|\alpha_1,\alpha_2,\theta,\phi)}{dx},\end{cases}$$

where f_1^* is the probability density function of random variable X when X takes values between 0 and θ; f_2^* is the probability density function of the random variable X when X takes values that are greater than θ. Here, ϕ is a positive parameter that controls the weights of f_1^* and f_2^*.

The composite IG-Pareto model was established by Aminzadeh and Deng (2019) by utilizing the theory introduced above. Suppose a random variable X is known to follow a composite Inverse Gamma-Pareto distribution such that the pdf of X is as follows:

$$f_X(x|\theta) = \begin{cases} \dfrac{c(k\theta)^{\alpha} x^{-\alpha-1} e^{-\frac{k\theta}{x}}}{\Gamma(\alpha)} & 0 \le x \le \theta \\ \dfrac{c(\alpha-k)\theta^{\alpha-k}}{x^{\alpha-k+1}} & x > \theta, \end{cases} \tag{1}$$

where, $c = 0.711384, k = 0.144351, a = 0.163947, \alpha = 0.308298$. Thus, their proposed IG-Pareto model contains only one parameter θ. In the following subsection, we will discuss the development of exponentiated composite IG-Pareto distribution specifically.

2.2. Development of the exponentiated composite inverse Gamma-Pareto distribution

Now suppose a power transformation is applied to random variable X, say $Y = g(X) = X^{1/\eta}$, where g is monotone increasing for any $\eta > 0$. Also, $X = g^{-1}(Y) = Y^{\eta}$. For any $\eta > 0$, $g^{-1}(y) = y^{\eta}$ has continuous derivative on $(0, \infty)$. Then the probability density function of Y is given by:

$$f_Y(y|\theta,\eta) = \begin{cases} \dfrac{c(k\theta)^{\alpha} (y^{\eta})^{-\alpha-1} e^{-\frac{k\theta}{y^{\eta}}}}{\Gamma(\alpha)} \eta y^{\eta-1} & 0 \le y^{\eta} \le \theta \\ \dfrac{c(\alpha-k)\theta^{\alpha-k}}{(y^{\eta})^{\alpha-k+1}} \eta y^{\eta-1} & y^{\eta} > \theta \end{cases} \tag{2}$$

It can be easily shown that the above density function for exponentiated composite IG-Pareto model is continuous and differentiable on the support $(0, \infty)$.

The motivation for developing exponentiated IG-Pareto model as an improvement of IG-Pareto model for loss data modeling is shown in Figures 1 and 2. Two different values for θ are chosen as 5 (Figure 1) and 10 (Figure 2). For each θ value, three η values 1, 5 and 10 are chosen, where $\eta = 1$ corresponds to the original one-parameter IG-Pareto composite model.

The figures indicate the composite exponentiated IG-Pareto model provides more flexibility to the one-parameter IG-Pareto model due to the introduction of the power parameter η. For fixed value of θ, the mode of the composite exponentiated IG-Pareto increases as η increases.

In an insurance context, one of the important topics is to maximize its benefit. Given a insurance policy limit b and a pdf $f_Y(y)$ associated with a loss random variable Y, the limited loss random variable $Y \wedge b$ is defined as following:

$$Y \wedge b = \begin{cases} Y & y \in (0, b] \\ b & y \in [b, \infty) \end{cases}$$

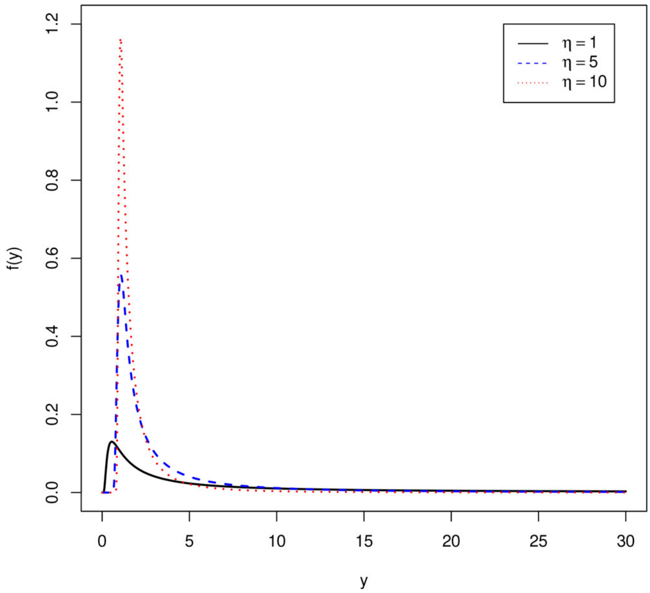

Figure 1. Composite exponentiated IG-Pareto density ($\theta = 5$).

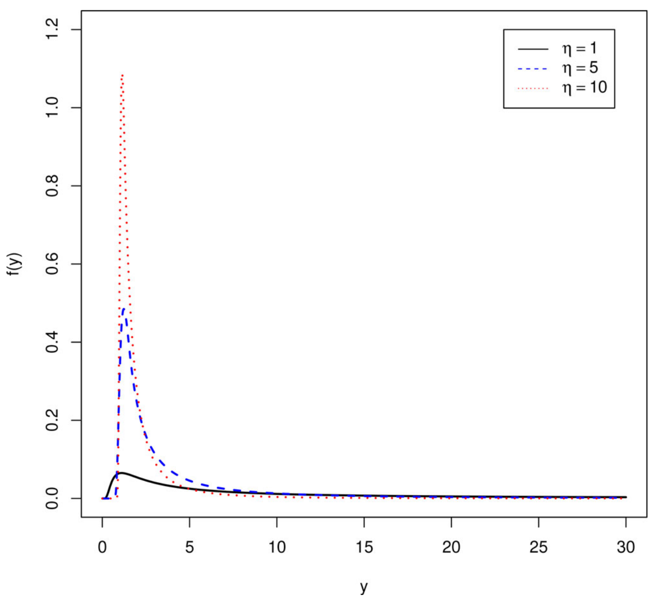

Figure 2. Composite exponentiated IG-Pareto density ($\theta = 10$).

Correspondingly, the limited t^{th} moment of a loss random variable Y, denoted by $E[(Y \wedge b)^t]$ is defined as:

$$E\left[(Y \wedge b)^t\right] = \int_0^b y^t f_Y(y)dy + \int_b^\infty b^t f_Y(y)dy \tag{3}$$

Suppose Y follows a exponentiated IG-Pareto distribution with parameters θ and η. It is easy to show that the t^{th} limited moment of Y is given by:

$$E\left[(Y\wedge b)^t\right]=\begin{cases} c\left[\dfrac{\Gamma\left(\alpha-\dfrac{t}{\eta},\dfrac{k\theta}{b^\eta}\right)(k\theta)^{\frac{t}{\eta}}+b^t\Gamma(\alpha,k)-b^t\Gamma\left(\alpha,\dfrac{k\theta}{b^\eta}\right)}{\Gamma(\alpha)}+b^t\right] & b\in(0,\theta^{1/\eta})\\ c\left[\dfrac{\Gamma\left(\alpha-\dfrac{t}{\eta},k\right)(k\theta)^{\frac{t}{\eta}}}{\Gamma(\alpha)}+b^t\right] & b=\theta^{1/\eta}\\ c\left\{\dfrac{\Gamma\left(\alpha-\dfrac{t}{\eta},k\right)(k\theta)^{\frac{t}{\eta}}}{\Gamma(\alpha)}+\dfrac{(\alpha-k)\left[b^{t-\eta(\alpha-k)}\theta^{\alpha-k}-\theta^{\frac{t}{\eta}}\right]}{k-\alpha+\dfrac{t}{\eta}}+b^{t-(\alpha-k)\eta}\theta^{\alpha-k}\right\} & b\in(\theta^{1/\eta},\infty), \end{cases}$$

where $\Gamma(.,.)$ stands for an upper incomplete gamma function, $\Gamma(\alpha,x)=\int_x^\infty t^{\alpha-1}e^{-t}dt$.

2.3. Parameter estimation

Let $y_1,\dots,y_n$ be a random sample from the exponentiated composite pdf given in (2). Without loss of generality, assume that $y_1<y_2<\dots<y_n$ is an ordered random sample generated from the pdf. The likelihood function can be written as follows:

$$\begin{aligned} L(\mathbf{y}|\theta,\eta)&=\prod_{i=1}^m\frac{c(k\theta)^\alpha(y_i^\eta)^{-\alpha-1}e^{-\frac{k\theta}{y_i^\eta}}}{\Gamma(\alpha)}\eta y_i^{\eta-1}\prod_{j=m+1}^n\frac{c(\alpha-k)\theta^{\alpha-k}}{(y_j^\eta)^{\alpha-k+1}}\eta y_j^{\eta-1}\\ &=\frac{c^n\eta^nk^{\alpha m}(\alpha-k)^{n-m}\left(\prod_{i=1}^m y_i\right)^{-\alpha\eta-1}}{(\Gamma(\alpha))^m\left(\prod_{j=m+1}^n y_j\right)^{(\alpha-k)\eta+1}}\theta^{\alpha m+(\alpha-k)(n-m)}e^{-k\theta\sum_{i=1}^m\frac{1}{y_i^\eta}}\\ &=Q\theta^{\alpha m+(\alpha-k)(n-m)}e^{-k\theta\sum_{i=1}^m\frac{1}{y_i^\eta}}, \end{aligned}\tag{4}$$

where

$$Q=\frac{c^n\eta^nk^{\alpha m}(\alpha-k)^{n-m}\left(\prod_{i=1}^m y_i\right)^{-\alpha\eta-1}}{(\Gamma(\alpha))^m\left(\prod_{j=m+1}^n y_j\right)^{(\alpha-k)\eta+1}}.$$

The above likelihood assumes that there exist an value m such that $y_m^\eta<\theta<y_{m+1}^\eta$. The MLE of θ and η can be obtained by solving the following equations:

$$\begin{cases}\dfrac{\partial L(\mathbf{y}|\theta,\eta)}{\partial\theta}=0\\ \dfrac{\partial L(\mathbf{y}|\theta,\eta)}{\partial\eta}=0\end{cases}$$

Closed-form expressions for MLE of θ and η cannot be obtained. In addition, m needs to be determined before finding the solution of the above equations. However, for the given values of η and m, the closed-form solution of θ can be written as follows:

$$\hat{\theta}|_{\eta,m}=\frac{\alpha m+(\alpha-k)(n-m)}{k\sum_{i=1}^m\frac{1}{y_i^\eta}}\tag{5}$$

Table 1. Simulation results for $\theta = 1$ and $\eta = 0.8$.

n	$\hat{\eta}_{\text{mean}}$	$\hat{\theta}_{\text{mean}}$	$\hat{\eta}_{\text{SD}}$	$\hat{\theta}_{\text{SD}}$
20	0.876	1.304	0.204	1.437
50	0.828	1.094	0.117	0.474
100	0.816	1.040	0.084	0.315
500	0.804	1.006	0.037	0.135

Thus, we designed a simple search algorithm to find the MLE of θ and η by utilizing Equation (5). The description of the search algorithm is as follows:

I. Obtain the sorted observations of a sample as $y_1 \le y_2 \le \ldots \le y_n$
II. Determine the range of η, the parameter search will be done within the pre-defined range. Note that $\eta > 0$. Hence, the search needs to be done within an left-open interval with 0 as the left endpoint. The right endpoint of this interval is data-specific.
III. For a known η in the range, we start with m = 1 and calculate the MLE of θ given η based on (5). If $y_1^\eta \le \hat{\theta}|_{\eta,m} \le y_2^\eta$, then m = 1. Otherwise jump to step (IV)
IV. Let m = 2. If $y_2^\eta \le \hat{\theta}|_{\eta,m} \le y_3^\eta$, then m = 2. We shall continue the above steps until m is identified. Once m is identified, keep $\hat{\theta}|_{\eta,m}$ as the MLE of θ for the known η.
V. Search for the optimal η that maximizes $L(\boldsymbol{y}|\theta,\eta)$. Find the corresponding $\hat{\theta}$ using equation (5). These are the MLEs for η and θ.

3. Simulation

We conducted a limited simulation study to check the accuracy for the estimates of $\hat{\theta}$ and $\hat{\eta}$. For the selected sample size n, θ and η values, $N = 5000$ samples were generated from the composite density given in (2).

Tables 1–6 present the results of all simulations under different scenarios. $\hat{\eta}_{\text{mean}}$, $\hat{\theta}_{\text{mean}}$ stand for the average of $\hat{\eta}$ and $\hat{\theta}$; $\hat{\eta}_{\text{SD}}$ and $\hat{\theta}_{\text{SD}}$ denote the standard deviation of $\hat{\eta}$ and $\hat{\theta}$ values, respectively.

We observed that when sample size n increases, the mean of the estimates of θ gets closer to the underlying true $\hat{\theta}$ under all simulation scenarios. Similarly, the mean of $\hat{\eta}$ gets closer to the underlying true η. In addition, the standard deviation of both $\hat{\theta}$ and $\hat{\eta}$ decreases as the sample size increases for different settings of the simulation parameters. Thus, the MLE of θ and η become more accurate as the sample size increases, which is a property of maximum likelihood estimation.

4. Numerical examples

In this section, we presented the performance of the exponentiated IG-Pareto model with three different insurance data sets, namely Danish Fire Insurance Data, Norwegian Fire Insurance Data, and Society of Actuaries Group Medical Insurance Large Claims Data. In our comparisons, allong with the proposed exponentiated IG-Pareto model, we used Weibull, Inverse Gamma and IG-Pareto models. Danish Fire Insurance Data Set is

Table 2. Simulation results for $\theta = 1$ and $\eta = 1$.

n	$\hat{\eta}_{\text{mean}}$	$\hat{\theta}_{\text{mean}}$	$\hat{\eta}_{\text{SD}}$	$\hat{\theta}_{\text{SD}}$
20	1.093	1.262	0.248	1.025
50	1.036	1.092	0.145	0.478
100	1.017	1.039	0.102	0.307
500	1.005	1.005	0.049	0.137

Table 3. Simulation results for $\theta = 1$ and $\eta = 1.2$.

n	$\hat{\eta}_{\text{mean}}$	$\hat{\theta}_{\text{mean}}$	$\hat{\eta}_{\text{SD}}$	$\hat{\theta}_{\text{SD}}$
20	1.322	1.263	0.312	1.094
50	1.240	1.091	0.174	0.463
100	1.220	1.048	0.120	0.314
500	1.206	1.005	0.0582	0.140

Table 4. Simulation results for $\theta = 5$ and $\eta = 0.8$.

n	$\hat{\eta}_{\text{mean}}$	$\hat{\theta}_{\text{mean}}$	$\hat{\eta}_{\text{SD}}$	$\hat{\theta}_{\text{SD}}$
20	0.877	7.464	0.203	9.776
50	0.829	5.555	0.117	1.992
100	0.813	5.276	0.082	1.263
500	0.805	5.049	0.037	0.512

Table 5. Simulation results for $\theta = 5$ and $\eta = 1.0$.

n	$\hat{\eta}_{\text{mean}}$	$\hat{\theta}_{\text{mean}}$	$\hat{\eta}_{\text{SD}}$	$\hat{\theta}_{\text{SD}}$
20	1.098	7.256	0.258	6.480
50	1.036	5.626	0.146	2.070
100	1.017	5.269	0.101	1.232
500	1.003	5.048	0.049	0.511

Table 6. Simulation results for $\theta = 5$ and $\eta = 1.2$.

n	$\hat{\eta}_{\text{mean}}$	$\hat{\theta}_{\text{mean}}$	$\hat{\eta}_{\text{SD}}$	$\hat{\theta}_{\text{SD}}$
20	1.317	7.245	0.305	7.213
50	1.244	5.566	0.173	2.025
100	1.224	5.283	0.121	1.261
500	1.206	5.059	0.0579	0.511

well analyzed data set (Grün and Miljkovic 2019). Grün and Miljkovic used 256 models to analyze the Danish Fire Insurance Data. Utilizing all those 256 models are beyond the scope of this paper. However, in their study, Weibull-Inverse Weibull came out as the best fitting model to describe the Danish Fire Insurance Data. Therefore, in addition to the abovementioned models, we included the Weibull-Inverse Weibull model in our study. Furthermore, we include the Weibull-Pareto model since we want to pick another composite model with the Pareto tail for comparison purposes. In fact, according to Grün and Miljkovic, Weibull-Pareto composite model came out as the best model among all the composite distributions with the Pareto tails.

4.1. Goodness-of-fit of the exponentiated IG-Pareto model

To compare the performance of the different models when fitting the insurance datasets, NLL, AIC, BIC, AICc and CAIC were used. The description of the measures are listed as follows:

- **NLL:** Negative Log-Likelihood is defined as the additive inverse of the loglikelihood function as follows:

$$NLL = -logL(\hat{\boldsymbol{\theta}}|\boldsymbol{y})$$

- **AIC:** Akaike's Information Criterion (Burnham and Anderson 2002) is defined as follows:

$$AIC = -2logL(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) + 2k,$$

where k is the number of free parameters.

- **BIC:** Bayesian Information Criterion (Burnham and Anderson 2002) is provided as follows:

$$BIC = -2logL(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) + klog(n),$$

where k is the number of parameters and n is the sample size of the data set.

- **AICc:** Hurvich and Tsai's Criterion (Hurvich and Tsai 1989) is provided as follows:

$$AICc = -2logL(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) + \frac{2nk}{(n-k-1)},$$

- **CAIC:** Bozdogan's criterion (Bozdogan 1987) is provided as follows:

$$CAIC = -2logL(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) + k(log(n) + 1),$$

R software was used to compute the MLEs of the parameters in different models as well as NLL, AIC, BIC, AICc, and CAIC values.

4.2. Value-at-Risk

The estimation of extreme quantiles $Q(p)$ with p being large is an important topic in insurance data modeling. These extreme quantiles are named as Value-at-Risk (VaR). For a loss random variable, VaR at the level of p is defined as:

$$P(X < VaR_p(X)) = p,$$

In the context of insurance industry, VaR represents the amount of capital that an insurance company needs to have to protect the company against bankruptcy due to extreme claims.

4.2.1. Case 1: Danish fire insurance data

Danish fire insurance data was widely used by many researchers to check the performance of different composite models. The data set contains 2492 claims in millions of Danish Krones (DKK) from the years 1980 to 1990. From the *SMPracticals* package in R (Davison 2019), we were able to obtain the data and complete the analysis. The histogram of this data set is presented as Figure 3.

Table 7 provides the results from this data set. Exponentiated IG-Pareto model outperforms the original one-parameter IG-Pareto model in terms of *NLL*, *AIC*, *BIC*, *AICc*, and *CAIC*. This is consistent with Figure 4. Figure 4 presents the comparison of IG-

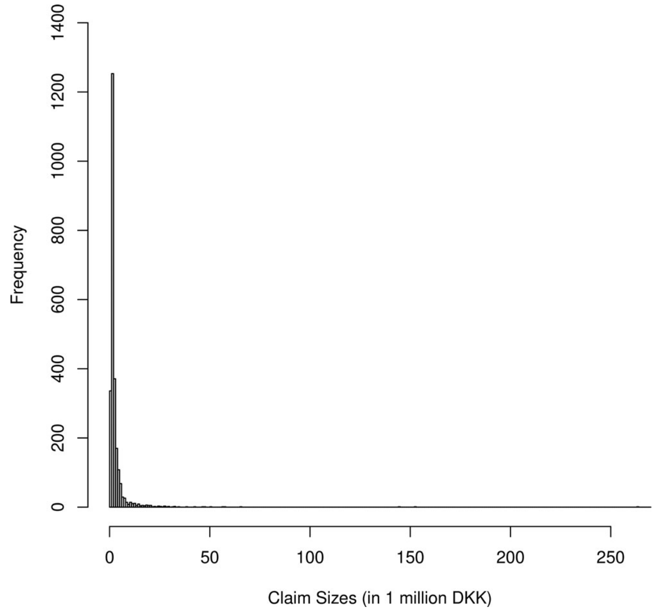

Figure 3. Histogram of Danish Fire Insurance Data Set.

Pareto model, exponentiated IG-Pareto model and the Gaussian kernel density estimate of the Danish Fire Insurance Dataset. Exponentiated IG-Pareto model provides a satisfactory fit to the Danish Fire Insurance Data while the original one-parameter IG-Pareto model does not fit the data well. Among the three two-parameter models we chose, Inverse-Gamma model performed slightly better compared to exponentiated IG-Pareto model. However, in terms of *NLL*, *AIC*, and *BIC*, the exponentiated IG-Pareto model gives a better performance compared to the two-parameter Weibull model. Both Weibull-Inverse Weibull and Weibull-Pareto composite models perform better than the proposed exponentiated IG-Pareto model in terms of all the goodness-of-fit measures.

4.2.2. Case 2: Norwegian fire insurance data

Similar to the Danish fire insurance loss data set, the Norwegian fire insurance data was used by several researchers to investigate the performance of various loss models. The data set consists of 9181 claims in 1000s of Norwegian Krones (NKK) from the years 1972 to 1992 for a Norwegian insurance company. We obtained the data set through R package *ReIns* (Reynkens and Verbelen 2020). Note the claims with size less than 500,000 NKK are forced to be 500,000 NKK. However, none of the claim values from the year 1972 are truncated, and therefore we selected the data from the year 1972 to assess the performance of the proposed model. Dealing with the truncated data is beyond the scope of this article.

Figure 5 shows the histogram of the data set. The claim data from the year 1972 consists of 97 values and the claim values in millions of Norwegian Krones (NKK) are as follows:

0.520, 0.529, 0.530, 0.530, 0.544, 0.545, 0.546, 0.549, 0.553, 0.555, 0.562, 0.565, 0.565, 0.568, 0.579, 0.586, 0.600, 0.600, 0.604, 0.605, 0.621, 0.627, 0.633, 0.636, 0.667, 0.670,

Table 7. Goodness-of-fit of different models to the Danish fire data based on MLEs.

Model	p	NLL	AIC	BIC	AICc	CAIC
Weibull	2	5270.5	10545.0	10556.6	10545.0	10558.6
Inverse Gamma	2	4097.9	8199.8	8211.4	8199.8	8213.4
Inverse Gamma-Pareto (One-Parameter)	1	6983.8	13969.6	13975.5	13969.6	13976.5
Exponentiated Inverse Gamma-Pareto	2	4287.7	8591.0	8590.0	8579.4	8593.0
Weibull-Pareto	4	3823.7	7655.4	7678.6	7655.4	7682.5
Weibull-Inverse Weibull	4	3820.0	7648.0	7671.3	7648.0	7675.3

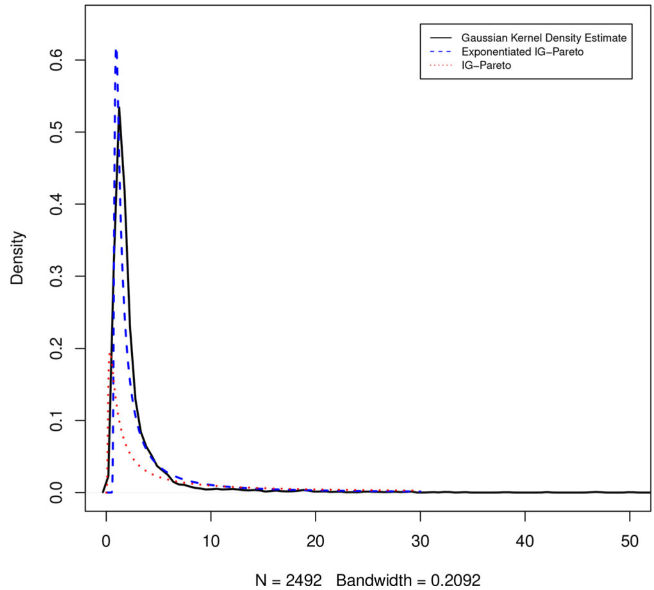

Figure 4. Density Plot of Danish Fire Insurance Data with corresponding exponentiated IG-Pareto and IG-Pareto model fit.

0.671, 0.676, 0.681, 0.682, 0.699, 0.706, 0.725, 0.729, 0.736, 0.741, 0.744, 0.750, 0.758, 0.764, 0.767, 0.778, 0.797, 0.810, 0.849, 0.856, 0.878, 0.900, 0.916, 0.919, 0.922, 0.930, 0.942, 0.943, 0.982, 0.991, 1.051, 1.059, 1.074, 1.130, 1.148, 1.150, 1.181, 1.189, 1.218, 1.271, 1.302, 1.428, 1.438, 1.442, 1.445, 1.450, 1.498, 1.503, 1.578, 1.895, 1.912, 1.920, 2.090, 2.370, 2.470, 2.522, 2.590, 2.722, 2.737, 2.924, 3.293, 3.544, 3.961, 5.412, 5.856, 6.032, 6.493, 8.648, 8.876, 13.911, 28.055

Table 8 presents the goodness-of-fit results from this data set. Similar to what we observed for the Danish Fire Insurance Data, the Exponentiated IG-Pareto model performed better than the original one-parameter IG-Pareto model in terms of all goodness-of-fit measures: *NLL*, *AIC*, *BIC*, *AICc*, and *CAIC*. This is consistent with Figure 6, where exponentiated IG-Pareto model fits with the Norwegian Fire Insurance Data satisfactorily while the original one-parameter IG-Pareto model does not fit this data set well. Among the three two-parameter models we chose, exponentiated IG-Pareto model performed the best in terms of all these goodness-of-fit criteria including *NLL, AIC* and *BIC*. In terms of *BIC* and *CAIC*, the exponentiated IG-Pareto model also demonstrated comparable performance against the Weibull-Inverse Weibull model and the Weibull-Pareto model. However, both of these models still performed better in terms of *AIC* and *AICc*.

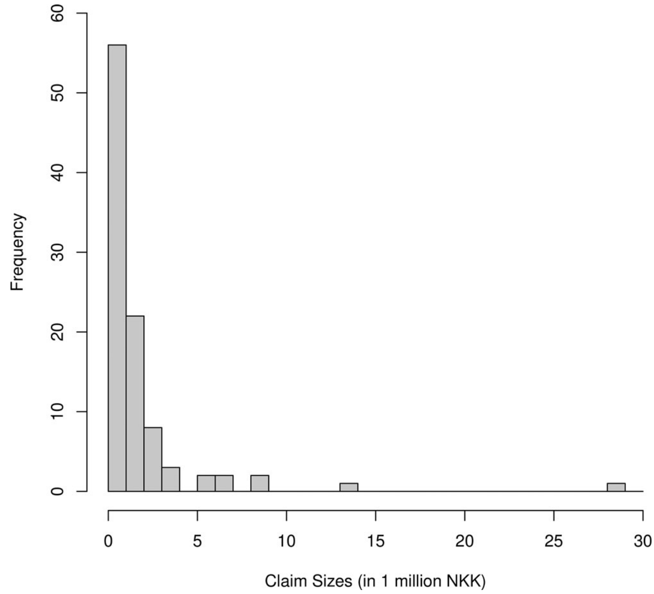

Figure 5. Histogram of Norwegian Fire Insurance Data Set from Year 1972.

Table 8. Goodness-of-fit of different models to the Norwegian fire insurance data from year 1972 based on MLEs.

Model	p	NLL	AIC	BIC	AICc	CAIC
Weibull	2	158.7	321.4	326.6	321.5	328.6
Inverse Gamma	2	167.2	338.4	343.6	338.6	345.6
Inverse Gamma-Pareto (One-Parameter)	1	221.8	445.7	448.2	445.7	449.2
Exponentiated Inverse Gamma-Pareto	2	96.1	196.2	201.3	196.3	203.3
Weibull-Pareto	4	91.2	190.4	200.7	190.8	204.7
Weibull-Inverse Weibull	4	90.5	189.0	199.3	189.4	203.3

4.2.3. Case3: Society of actuaries (SOA) group medical insurance large claims data

The SOA Group Medical Insurance Claims is a publicly available data set that was published in year 1997. This data set contains 75,789 claims from year 1991 in US Dollars (USD). This data set is available in R package ReIns (Reynkens and Verbelen 2020). For the analysis concern, we rescaled the claim sizes so all the claim sizes had a unit of 10,000 USD in our analysis. The histogram of this data set is shown in Figure 7.

Table 9 illustrates the goodness-of-fit results for this medical insurance data set. With the medical insurance data set, exponentiated IG-Pareto model outperforms all the other methods with respect to all the goodness-of-fit measures. Figure 8 also demonstrates the exponentiated IG-Pareto model provides a very good fit to this data. Furthermore, it is clear that the one-parameter IG-Pareto model is not a good fit for this data set.

Table 10 presents VaR estimates for different models at the level of $0.90, 0.95$ and 0.99. Our model provided the closest fit compared to the empirical estimates of VaR at the level of 0.90 and 0.95. However, Weibull-Pareto composite model provided the

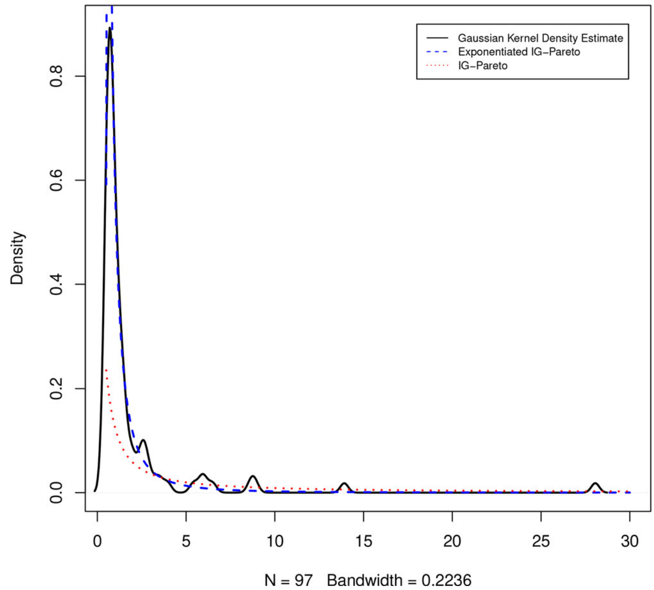

Figure 6. Density Plot of Norwegian Fire Insurance Data from year 1972 with corresponding exponentiated IG-Pareto and IG-Pareto model fit.

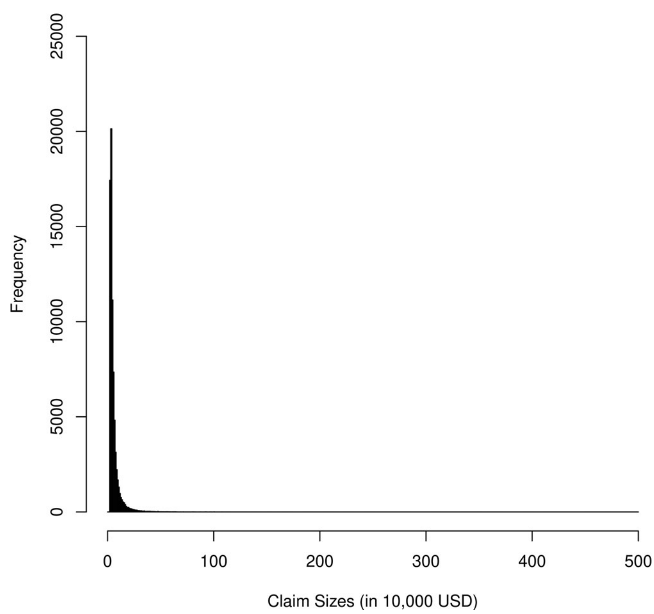

Figure 7. Histogram of SOA Group Medical Insurance Large Claims.

closest estimate among all models at the 0.99 level, compared to its empirical counterparts. Notice in terms of all the goodness-of-fit measures, Weibull-Pareto is not the best model. This indicates a model that provides the best fit to the whole data set does not necessarily provides the best performance at the extreme upper tail area.

Table 9. Goodness-of-fit of different models to the SOA data based on MLEs.

Model	p	NLL	AIC	BIC	AICc	CAIC
Weibull	2	204223.9	408451.8	408470.3	408451.8	408472.3
Inverse Gamma	2	173619.6	554883.8	554893.0	554883.8	554894.0
Inverse Gamma-Pareto (One-Parameter)	1	277440.9	347243.2	347261.7	347243.2	347263.7
Exponentiated Inverse Gamma-Pareto	2	160836.6	321677.2	321695.7	321677.2	321697.7
Weibull-Pareto	4	209464.3	418936.6	418973.5	418936.6	418977.5
Weibull-Inverse Weibull	4	167745.8	335499.6	335536.5	335499.6	335540.5

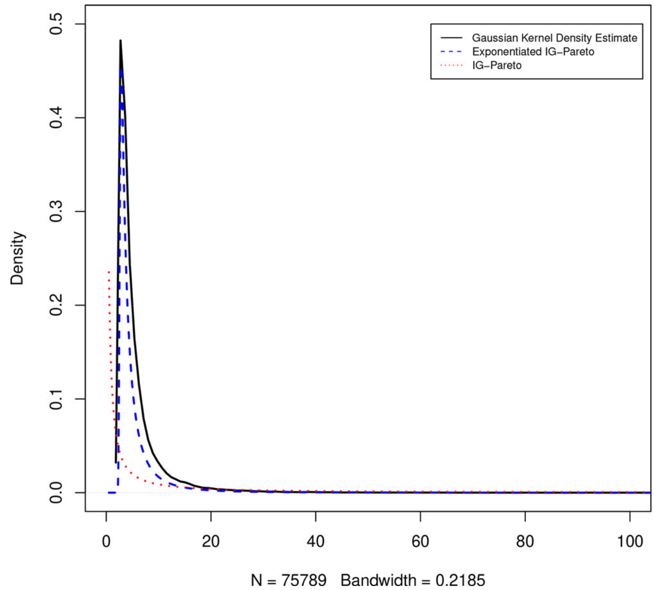

Figure 8. Density Plot of SOA Group Medical Large Claims with corresponding exponentiated IG-Pareto and IG-Pareto model fit.

Table 10. Comparison of VaR of different models for SOA Group Medical Claims Data.

Model	VaR(0.90)	VaR(0.95)	VaR(0.99)
Empirical estimates	10.18	14.76	30.60
Weibull	12.24	15.01	20.93
Inverse Gamma	9.29	11.78	19.25
Inverse Gamma-Pareto (One-Parameter)	1.36×10^6	9.36×10^7	1.72×10^{12}
Exponentiated Inverse Gamma-Pareto	9.77	14.54	36.58
Weibull-Pareto	13.48	17.66	27.60
Weibull-Inverse Weibull	8.67	11.40	21.17

5. Conclusion

In this paper, we proposed a new exponentiated IG-Pareto model to improve the performance of the original one-parameter IG-Pareto model. We provide an algorithm to find the MLE of θ and η in Section 2. Such algorithm presents the ability to identify the MLE as the estimates for both θ and η become more accurate as the sample size gets larger in all simulation scenarios. Three numerical examples are provided and the new exponentiated IG-Pareto model outperforms the original IG-Pareto model for all the examples. For the SOA Group Medical Insurance Large Claims data set, the exponentiated IG-Pareto model provided the best fit to the data among all the models. The development of this model is promising since such exponentiation approach can also be applied to other composite models.

Acknowledgement

We Thank The Reviewers And The Editor For Their Helpful Comments And Suggestions.

References

Abu Bakar, S., N. Hamzah, M. Maghsoudi, and S. Nadarajah. 2015. Modeling loss data using composite models. *Insurance: Mathematics and Economics* 61 (C):146–54.

Afify, A. Z., H. M. Yousof, G. G. Hamedani, and G. R. Aryal. 2016. The exponentiated weibull-pareto distribution with application. *Journal of Statistical Theory and Applications* 15 (4): 326–44. doi:10.2991/jsta.2016.15.4.2.

Aminzadeh, M. S., and M. Deng. 2019. Bayesian predictive modeling for inverse gamma-pareto composite distribution. *Communications in Statistics - Theory and Methods* 48 (8):1938–54. doi:10.1080/03610926.2018.1440595.

Bozdogan, H. 1987. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52 (3):345–70. doi:10.1007/BF02294361.

Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference*, 2nd ed. New York: Springer-Verlag.

Cooray, K., and M. M. A. Ananda. 2005. Modeling actuarial data with a composite lognormal-pareto model. *Scandinavian Actuarial Journal* 2005 (5):321–34. doi:10.1080/03461230510009763.

Davison, A. 2019. *SMPracticals: Practicals for Use with Davison (2003) Statistical Models. R Package Version* 1:3–4.

Grün, B., and T. Miljkovic. 2019. Extending composite loss models using a general framework of advanced computational tools. *Scandinavian Actuarial Journal* 2019 (8):642–60. doi:10.1080/03461238.2019.1596151.

Gupta, R. D., and D. Kundu. 1999. Theory & methods: Generalized exponential distributions. *Australian < Html_Ent Glyph="@Amp;" Ascii="&"/> New Zealand Journal of Statistics* 41 (2): 173–88. doi:10.1111/1467-842X.00072.

Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* 76 (2):297–307. doi:10.1093/biomet/76.2.297.

Mudholkar, G. S., and D. K. Srivastava. 1993. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability* 42 (2):299–302. doi:10.1109/24.229504.

Nadarajah, S. 2005a. Exponentiated beta distributions. *Computers & Mathematics with Applications* 49 (7-8):1029–35. doi:10.1016/j.camwa.2004.11.008.

Nadarajah, S. 2005b. Exponentiated pareto distributions. *Statistics* 39 (3):255–60. doi:10.1080/02331880500065488.

Nadarajah, S. 2006. The exponentiated gumbel distribution with climate application. *Environmetrics* 17 (1):13–23. doi:10.1002/env.739.

Nadarajah, S., and A. K. Gupta. 2007. The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin* 59 (1-2):29–54. doi:10.1177/0008068320070103.

Preda, V., and R. Ciumara. 2006. On composite models: Weibull-pareto and lognormal-pareto. - a comparative study. *Romanian Journal of Economic Forecasting* 3:32–46.

Reynkens, T., and R. Verbelen. 2020. *ReIns: Functions from" Reinsurance: Actuarial and Statistical Aspects"*. R package version 1.0.10.

Scollnik, D. P. M. 2007. On composite lognormal-Pareto models. *Scandinavian Actuarial Journal* 2007 (1):20–33. doi:10.1080/03461230601110447.

Teodorescu, S., and R. Vernic. 2006. A composite exponential-pareto distribution. *The Annals of the "Ovidius" University of Constanta, Mathematics Series* 14 (1):99–108.