Contents

List of Tables

List of Figures

On A New Heavy Tailed Pareto-Weibull Distribution and it's Regression Model

Abstract

This article presents a new probability distribution suitable for modeling heavy tailed and right skewed data sets. The proposed distribution is obtained by continuous mixture of scale parameter of Pareto family with Weibull distribution. Analytical expressions for various distributional properties and actuarial risk measures of the proposed model are derived. The applicability of proposed model is assessed by two real life insurance data sets and its performance is compared with existing class of heavy tailed models. This article presents the two mixed Pareto regression models to model various characteristics of the insurance data sets. Pareto distribution is always an indisputable distribution to model the right tail of an income distribution. The monotonically decreasing shape of Pareto model makes him inappropriate to model the hump shaped and positively skewed data set. Through the process of compounding of distribution we reweighed the scale parameter of the Pareto distribution by two parametric distributions which are suitable to model various peculiarities of the insurance data sets like hump shaped and positive skewness. The resultant mixed Pareto distributions can able to handle different features of the insurance data set. Regression specifications are allowed for the parameters of the mixed Pareto models to take care of the heterogeneity of the individual policyholder. EM Algorithm is provided to expedite the process of finding ML estimates of proposed models. The applicability of mixed Pareto regression models is assessed by real life data set and the its performance is compared with existing heavy-tailed regression models.

Keywords: Actuarial risk measures, Continuous mixture distribution, Danish Fire data set, Pareto distribution, usautoBI data set, Weibull distribution.

1 Introduction

Over the several years, modeling heavy tailed data has become a point of attraction to the actuaries where the challenge is to model high frequency of small losses and low frequency of high losses. To handle such type of challenge many traditional distributions are available in the literature. Pareto distribution, Burr distribution, Generalized Pareto distribution, Generalized Inverse Gaussian distribution and many others are being used to analyse heavy tail of the data. In actuarial statistics and finance, the classical Pareto distribution has been considered to be better than other models because it provides a good narration of the random behavior of large claims. Loss data mostly have characteristics such as unimodality, right skewness, and a thick right tail. To accommodate these features in a single model, practitioners have worked on many generalization includes but not limited to this namely (i)transformation method, (ii)composition of two or more distributions, (iii)compounding of distributions and (iv) finite mixture of distributions.

Skewed distributions are popular models since they accommodate right-skewness and high kurtosis present in the data. (see Vernic (2006), Adcock et al. (2015), Kazemi and Noorizadeh (2015) and Eling (2012)). However, insurance losses and financial risks take values on the positive real line and hence these skew class of distributions may not be appropriate as they are real valued. In such instances, variable transformation, particularly exponential transformation, has gained considerable importance. According to Bagnato and Punzo (2013), the transformation is straightforward to use, although inference and computation of several distributional properties can be difficult. Recently Bhati and Ravi (2018) introduced a generalized log-Moyal (GlogM) distribution as a heavy tailed distribution. The two parameter GlogM distribution is obtained by transformation and has uni-modality, right skewness and accompanied with heavier tail than exponential, which are desirable properties. Density function of the GlogM distribution is available in the closed form. This model also provides a better description of data with possibly heavy tails than the available two parameter distributions prevalent in the actuarial literature

Method of composition of two or more distributions is One more popular method to generate probability distribution. Generally in many discipline because of heterogeneity in the date under consideration, modeling the entire heterogeneous data with single probability distribution may create problem of incorrect inference. In such case overall distribution can be approximate by concatenating several well known model at a threshold point from where the heterogeneous data can be partitioned into several non overlapping groups of homogeneous data. Cooray and Ananda (2005) proposed a composite model considering Lognormal density up to a certain threshold and Pareto density there after. Over the several years various composite models are available in the literature (see Ciumara,2006, Scollnik, 2007, Nadarajah and Bakar, 2014, Calderin Odeja and Kwok, 2016). Modeling loss data using composite distributions is quite cumbersome task. There is a no well established rule is available with the choice of the density for the head and tail part of the data set.

Punzo et al. (2017) proposed a three-parameter compound distribution to account for features like uni-modality, hump-shaped, right-skewed, and heavy tails. Authors introduced family of nine different 3-parameter compounding models. which allow to give more flexibility to the tails of the conditional distribution. As Punzo et al. (2017) pointed out, the final density derived using this method may not always contain closed form expressions, making estimation more difficult. Furthermore, not all moments are provided in closed form, and no skew parameter is given.

The k-component finite mixture models from parametric, non-Gaussian families of distributions are another way for modelling multi-modal insurance loss data sets. To model losses, Bernardi et al. (2012) used the skew normal mixture. Lee and Lin (2010) and Verbelen et al. (2015) looked at finite mixes of Erlang distributions later on. Miljkovic and Grün (2016) recently extended this technique to include finite combinations of Burr, Gamma, inverse Burr, inverse Gaussian, log-normal, and Weibull distributions.

The present work is motivated by Colombi(1990), where the author developed three parameter income distribution called as Pareto-Lognormal distribution by considering the scale parameter of Pareto distribution follows Lognormal distribution. The model is developed by assuming the income to be the product of two components having respectively a Paretian and a lognormal distribution. The author presented some statistical properties of the Pareto-Lognormal distribution and also compare its fit with that of Singh-Maddala's and Dagum's Type I models. The aim of this paper is to introduce a Pareto-Weibull distribution to model heavier tail of the data sets. The Pareto-Weibull distribution is generated by assuming the scale parameter of the Pareto distribution follows Weibull distribution. In the actuarial literature, Weibull distribution also proved to be a suitable and flexible to model small losses. The Pareto distribution due to its monotonically decreasing shape of the density, does not provide good fit for many application. Utilizing this fact, we obtain Pareto-Weibull distribution as a re-scaled version of Pareto distribution by scaling the variability-related parameter of a Pareto distribution by a suitable Weibull distribution. Due to this, the shape of resultant density will not be monotonically decreasing for certain values of the parameters. The Pareto-Weibull distribution is then allows to handle small, moderate and large insurance losses by giving more flexibility to the tail of Pareto distribution.

Claims in an insurance portfolio often comprise small size claims that occur with high frequency and large size claims that occur with low frequency. These characteristics in claims size data happen due to a heterogeneous group of individuals/claimants in the portfolio.

In this study, we propose a mixed Pareto regression model to handle the unimodal, right skewed and heavy-tailed data set which often arises in the field of insurance and finance. Due to the monotonically decreasing nature of the Pareto distribution, modeling unimodal and heavy-tailed data with the Pareto distribution may not provide an adequate fit. In order to overcome this difficulty, the mixed Pareto model is formulated by considering that the scale parameter of the Pareto distribution follows a continuous and at least twice differentiable parametric distribution with unit mean to make the resultant mixed Pareto regression model identifiable. Furthermore, we introduce covariate information in both the mean and dispersion parameters of the mixed Pareto model, which enables us to model heavy-tailed data in a more efficient way due to fact that the skewness of the response variable is usually influenced by the mean and the dispersion parameters

The Pareto-Weibull distribution has a complicated density which is not straightforward to estimate via traditional maximum-likelihood estimation (ML) schemes. Moreover, computational complexity is increased by introducing regression structures on every parameter of the model. To overcome this issue, an Expectation-Maximization (EM) type algorithm is used to facilitate the ML estimation procedure for the proposed model. The applicability of the Pareto-Inverse Gaussian regression model is assessed by application to a real world motor insurance claim size data set.

The performance of the Pareto-Weibull regression model having regression to both the parameters involved in the mean and dispersion is compared with the Pareto-Weibull regression model having regression to the parameter involved in the dispersion only.

Regression modelling involving heavy-tailed response distributions, which have heavier tails than the exponential distribution, has become increasingly popular in many insurance settings including non-life insurance. Mixed Exponential models can be considered as a natural choice for the distribution of heavy-tailed claim sizes since their tails are not exponentially bounded. The author is concerned with introducing a general family of mixed Exponential regression models with varying dispersion which can efficiently capture the tail behaviour of losses. Our main achievement is that the author present an Expectation-Maximization (EM) type algorithm which can facilitate maximum likelihood (ML) estimation for our class of mixed Exponential models which allows for regression specications for both the mean and dispersion parameters. Finally, a real data application based on motor insurance data is given to illustrate the versatility of the proposed EM type algorithm

The rest of the paper is organized as follows: In section 2 we discuss the genesis of proposed Pareto-Weibull distribution, section 3 consist of some statistical properties of Pareto-weibull distribution. section 4 presents some actuarial measures for the Pareto-Weibull distribution. Numerical application is given in section 5. Finally concluding remarks can be found in section 6.

2 Genesis

2.1 Genesis of the Proposed Model for Pareto-Weibull

In order to proposed the model, for the sake of completeness, we define the following: Let random variable X follow Weibull(ϕ, τ) having density

$$g_X(x) = \frac{\tau}{\phi} e^{-\left(\frac{x}{\phi}\right)^{\tau}} \left(\frac{x}{\phi}\right)^{\tau-1}, \quad x > 0, \phi > 0, \tau > 0, \tag{1}$$

and the conditional distribution of $Y|X = x$ follows Pareto(x, α) with density

$$f_{Y|X}(y|x) = \frac{\alpha x^{\alpha}}{y^{\alpha+1}}, \quad y > x, \alpha > 0. \tag{2}$$

We say that the random variable Y follows a Pareto-Weibull (PW) distribution if it admits the stochastic representation:

$$Y|X \sim Pareto(x, \alpha), \quad X \sim Weibull(\phi, \tau),$$

The unconditional distribution of Y will be denoted by $Y \sim PW(\alpha, \phi, \tau)$ and its density is given as

$$f_Y(y; \alpha, \phi, \tau) = \int_0^y \frac{\alpha x^{\alpha-1} \left(\tau e^{-\left(\frac{x}{\phi}\right)^{\tau}} \left(\frac{x}{\phi}\right)^{\tau}\right)}{y^{\alpha+1}} dx.$$

which reduced to

$$f_Y(y; \alpha, \phi, \tau) = \frac{\alpha \phi^{\alpha}}{y^{\alpha+1}}. \left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau}, \left(\frac{y}{\phi}\right)^{\tau}\right)\right), \tag{3}$$

where $\Gamma(a, k) = \int_k^{\infty} u^{a-1} e^{-u} du$ is the upper incomplete gamma function.
The cumulative distribution function (cdf) of $PW(\alpha, \phi, \tau)$ is given by

$$F_Y(y; \alpha, \phi, \tau) = 1 - \left(\left(\frac{\phi}{y}\right)^{\alpha} \left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau}, \left(\frac{y}{\phi}\right)^{\tau}\right)\right) + e^{-\left(\frac{y}{\phi}\right)^{\tau}}\right). \tag{4}$$

Remark 1: The proposed Pareto-Weibull rv Y follows the relation $Y =^d XZ$, where r v X and Z are two stochastically independent rvs following Pareto and Weibull distribution respectively.

2.1.1 Shape of the Pareto-Weibull distribution

A Pareto-Weibull distribution is uni-modal and its mode M is the solution of following equation

$$-\phi^{(\tau+\alpha)}\left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right)-\Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{y}{\phi}\right)^{\tau}\right)\right).(\alpha+1)-\tau y^{\alpha+\tau}e^{-(\frac{y}{\phi})^{\tau}}=0 \tag{5}$$

Proof: The probability density function of Pareto-Weibull distribution is given by

$$_{PW}f_y(y;\alpha,\phi,\tau)=\frac{\alpha\phi^{\alpha}}{y^{\alpha+1}}\left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right)-\Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{y}{\phi}\right)^{\tau}\right)\right)$$

Let $g(y)=\log(f(y))$ then the $g(y)$ can be written as

$$_{PW}g_y(y;\alpha,\phi,\tau)=\log\left(\frac{\alpha\phi^{\alpha}}{y^{\alpha+1}}\left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right)-\Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{y}{\phi}\right)^{\tau}\right)\right)\right)$$

After equating the first derivative of the above equation to zero, we have following necessary condition for the existing of the mode M as

$$-\phi^{(\tau+\alpha)}\left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right)-\Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{y}{\phi}\right)^{\tau}\right)\right)(\alpha+1)-\tau y^{\alpha+\tau}e^{-(\frac{y}{\phi})^{\tau}}=0$$

It is easy to verify that, $f(0)=0$ and $f(+\infty)=0$, the Pareto-Weibull distribution is always uni-modal. ■

2.2 Genesis for the Pareto-Weibull Regression Model

In this study, we consider two mixed Pareto regression models with non constant variance, can be given as follows. Let us assume that Z_i are independent and identically distributed (i.i.d) random variables and $Y_i|Z_i$ be the claim from a i^{th} policyholder, $i=1,2,....,n$, are i.i.d random variables follows a one-parameter Pareto distribution having probability density function (pdf) as

$$f_{Y|Z}(y_i|z_i)=\frac{\alpha_i z_i^{\alpha_i}}{y_i^{\alpha_i+1}}, \tag{6}$$

where $y_i\geq z_i$, $\alpha_i>1$
The expected value and variance of the $Y_i|z_i$ are given by

$$\mathbb{E}(Y_i|z_i)=\frac{\alpha_i z_i}{(\alpha_i-1)}$$

$$\mathbb{V}ar(Y_i|z_i) = \frac{\alpha_i z_i^2}{(\alpha_i - 2)(\alpha_i - 1)^2}$$

Here we consider that the Z_i are random variables follows certain distribution having pdf $g(z_i; \tau_i)$ and df $G(z_i; \tau_i)$, we consider $\mathbb{E}(Z_i) = 1$ for the model to become identifiable and $\tau_i > 0$ is the dispersion parameter. It is easy to compute the unconditional distribution of Y_i having mixed Pareto distribution with pdf

$$f(y_i) = \int_0^y f(y_i|z_i) g(z_i; \tau_i)\, dz_i \tag{7}$$

Here, several covariates($\mathbf{X}$) pertaining to various attributes of insurance claimants were combined in the linear predictor of the parameter involved in mean, $\alpha_i = \exp(\beta_1^\top \mathbf{x}_{1,i})$ and dispersion parameter $\tau_i = \exp(\beta_2^\top \mathbf{x}_{2,i})$, and were chosen because of their anticipated effect on claim size, where $\beta_1, \beta_2 \in R^p$ are the vectors of regression coefficients and $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ are the vectors of explanatory variables having dimension $p_1 \times 1$ and $p_2 \times 1$ respectively . The mean and variance of Y_i can be computed using the following equations

$$\mathbb{E}(Y_i) = \mathbb{E}_{Z_i}[\mathbb{E}(Y_i|Z_i = z_i)] = \frac{\alpha_i}{(\alpha_i - 1)} \mathbb{E}_{Z_i}[Z_i] = \frac{\alpha_i}{(\alpha_i - 1)} \tag{8}$$

$$\begin{aligned} \mathbb{V}ar(Y_i) &= \mathbb{E}_{Z_i}(\mathbb{V}ar(Y_i|Z_i = z_i)) + \mathbb{V}ar_{Z_i}(\mathbb{E}(Y_i|Z_i = z_i)) \\ &= \frac{\alpha_i}{(\alpha_i - 1)^2(\alpha_i - 2)} \mathbb{E}_{Z_i}(Z_i^2) + \frac{\alpha_i}{(\alpha_i - 1)} \mathbb{V}ar_{Z_i}(Z_i) \end{aligned} \tag{9}$$

In this paper, we proposed regression structure on the both parameters α_i and ϕ_i which are involved in the mean and dispersion of the mixed Pareto distribution namely Pareto-Weibull to assess heavy tailed behaviour of insurance claims.

2.2.1 Mixed Pareto regression model-1 : Pareto-Weibull regression model

Pareto-Weibull regression model is generated by considering scale parameter of the Pareto distribution follows unit mean Weibull distribution having parameter ϕ_i with pdf given by

$$g(z_i; \phi_i) = \phi_i. \left[\Gamma\left(1 + \frac{1}{\phi_i}\right)\right] e^{-\left(z_i\left[\Gamma(1+\frac{1}{\phi_i})\right]\right)^{\phi_i}} \left(z_i\left[\Gamma\left(1 + \frac{1}{\phi_i}\right)\right]\right)^{\phi_i - 1} \tag{10}$$

$z_i > 0$ and $\mathbb{E}(Z_i) = 1$ and $\mathbb{V}ar(Z_i) = \left(\frac{1}{\left[\Gamma\left(1+\frac{1}{\phi_i}\right)\right]}\right)^2 \left[\Gamma\left(1 + \frac{2}{\phi_i}\right) - \Gamma^2\left(1 + \frac{1}{\phi_i}\right)\right]$, where $i = 1, 2, \ldots.., n$.

Using equation (7), the pdf of two-parameter Pareto-Weibull distribution can be given by

$$PWf_y(y_i;\phi_i,\alpha_i) = \frac{\alpha_i}{y_i^{\alpha_i+1}.\left[\Gamma(1+\frac{1}{\phi_i})\right]^{\alpha_i}}.\left[\Gamma\left(\frac{\alpha_i+\phi_i}{\phi_i}\right) - \Gamma\left(\frac{\alpha_i+\phi_i}{\phi_i},\left(y_i\left[\Gamma(1+\frac{1}{\phi_i})\right]\right)^{\phi_i}\right)\right] \tag{11}$$

where $y_i > 0$, $\alpha_i > 0$ and $\phi_i > 0$. The parameter α_i and ϕ_i contains the covariate information as $\alpha_i = \exp(\beta_1^\top \mathbf{x}_{1,i})$ $\phi_i = \exp(\beta_2^\top \mathbf{x}_{2,i})$ The mean and variance of Pareto-Weibull regression model can be computed using equation (8) and (9) respectively as

$$\mathbb{E}(Y_i) = \frac{\alpha_i}{(\alpha_i - 1)}, \quad \alpha_i > 1 \tag{12}$$

$$\mathbb{V}ar(Y_i) = \left(\frac{\alpha_i}{\alpha_i - 1}\right)^2.\left[2.\left(\frac{1}{\left[\Gamma\left(1+\frac{1}{\tau_i}\right)\right]}\right)^2\left[\Gamma\left(1+\frac{2}{\tau_i}\right) - \Gamma^2\left(1+\frac{1}{\tau_i}\right)\right]+1\right], \quad \alpha_i > 2 \tag{13}$$

3 Distributional properties of Pareto-Weibull distribution

3.1 Moments:

Using the laws of total expectation and total variance and the moments of the Weibull-Pareto distribution is given by

The Pareto-Weibull random variable has finite moment of order k only if k is less than α

Proof: By using the postulation 2.1, we can write the k^{th} order moment of Pareto-Weibull distribution as

$$\mathtt{E}(Y^k) = \mathtt{E}(X^k).\mathtt{E}(Z^k) = \frac{\alpha}{\alpha - k}\phi^k\Gamma\left(1+\frac{k}{\tau}\right) \quad \text{where} \quad \alpha > k \tag{14}$$

From the equation (14), following expressions of the expected value, second raw moment and the variance of a Pareto-Weibull random variable can be derive

$$\mathtt{E}(Y) = \frac{\alpha}{\alpha - 1}\phi\Gamma\left(1+\frac{1}{\tau}\right), \quad \alpha > 1 \tag{15}$$

$$\mathtt{E}(Y^2) = \frac{\alpha}{\alpha - 2}\phi^2\Gamma\left(1+\frac{2}{\tau}\right), \quad \alpha > 2, \tag{16}$$

$$\mathtt{Var}(Y) = \alpha\phi^2\left(\frac{\Gamma(1+\frac{2}{\tau})}{(\alpha-2)} - \alpha\frac{\left(\Gamma(1+\frac{1}{\tau})\right)^2}{(\alpha-1)^2}\right) \quad \alpha > 2 \tag{17}$$

$\blacksquare$

4 Extremal Properties and Regularly varying tail behavior of the Pareto-Weibull distribution

In general insurance, the application of heavy right tailed distributions is of prime importance. The choice of a heavy-tailed distribution implies that we will see more extreme observations or that the model will be able to represent this type of behavior. Following postulation establishes that the Pareto-Weibull have survival function with regularly varying tails. It is already known that an important class of heavy-tailed distributions is the class of regularly varying distribution functions. Which means that a distribution function belongs to regularly varying class of distribution is a heavy tailed distribution.

Postulation 4.1: A Pareto-Weibull density belongs to regularly varying tailed distribution.

Proof: A distribution function F is said to belong to the regularly varying class if

$$\lim_{t\to\infty}\frac{\bar{F}(ty)}{\bar{F}(t)} = y^{-\alpha} \quad \forall \quad y>0 \tag{18}$$

Note that the survival function of Pareto-Weibull (α,ϕ,τ) is

$$\bar{F}(y) = \left(\left(\frac{\phi}{y}\right)^{\alpha}\left[\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{y}{\phi}\right)^{\tau}\right)\right) + e^{-\left(\frac{y}{\phi}\right)^{\tau}}\right)$$

Using equation (13), we have

$$\begin{aligned}\lim_{t\to\infty}\frac{\bar{F}(ty)}{\bar{F}(t)} &= \lim_{t\to\infty}\frac{\left(\left(\frac{\phi}{ty}\right)^{\alpha}\left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{ty}{\phi}\right)^{\tau}\right)\right) + e^{-\left(\frac{ty}{\phi}\right)^{\tau}}\right)}{\left(\left(\frac{\phi}{t}\right)^{\alpha}\left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{t}{\phi}\right)^{\tau}\right)\right) + e^{-\left(\frac{t}{\phi}\right)^{\tau}}\right)} \\ &= \lim_{t\to\infty}\frac{\left(\left(\frac{\phi}{ty}\right)^{\alpha}\left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{ty}{\phi}\right)^{\tau}\right) + e^{-\left(\frac{ty}{\phi}\right)^{\tau}}\right)\right)}{\left(\left(\frac{\phi}{t}\right)^{\alpha}\left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau},\left(\frac{t}{\phi}\right)^{\tau}\right) + e^{-\left(\frac{t}{\phi}\right)^{\tau}}\right)\right)} = y^{-\alpha}\end{aligned} \tag{19}$$

The above equation shows that the Pareto-Weibull belongs to regularly varying tailed class of distribution.This suggest that the Pareto-Weibull distribution is heavy tailed. α is the heavy-tailed parameter of the Pareto-Weibull distribution which control the tail of the distribution. The larger the α, less heavy the tail becomes. The smaller the α, more heavy the tail becomes.

Moreover, by Theorem, F belongs to maximum domain of attraction(MDA) if and only if $\bar{F} = 1 - F$ is regularly varying and by Theorem 3.3.7, Embrechts et al. (2003), the distribution function F belongs to the Fréchet maximum domain of attraction(MDA), which means that $\frac{M_n - a_n}{b_n} \xrightarrow{d} F$ where a_n and b_n are normalizing sequences such that a_n=0 and b_n= $F^{\leftarrow}(1-\frac{1}{n})$, and $Y_1, ..., Y_n$ are independent and identically distributed (i.i.d.) rvs having common distribution function F , with rv F having the Fréchet distribution with parameter α ■

5 Actuarial Measures

5.1 Limited Expected Value (LEV)

Reinsurance is used by the majority of insurance firms to shift financial risk above a certain level. The expected value on or below the threshold u, known as the limited expected value, is one of the most important factors in determining the reinsurance premium..In an insurance claim, u is a policy limit that establishes a ceiling on the benefit payable.

If the rv Y is the claim size with probability density function, then

$$\texttt{LEV}_u(Y) = \mathbb{E}(Y \wedge u) = \int_0^u yf(y)dy + u(1 - F(u))$$

The Limited Expected Value for the Pareto-Weibull distribution is given by

$$\begin{aligned}\texttt{LEV}_u(Y) =& \mathbb{E}(Y \wedge u) = \int_0^u y \frac{\alpha\phi^\alpha}{y^{\alpha+1}} \left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau}, \left(\frac{y}{\phi}\right)^\tau\right) \right) dy \\ &+ u \left(\left(\frac{\phi}{u}\right)^\alpha \left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau}, \left(\frac{u}{\phi}\right)^\tau\right) \right) + e^{-\left(\frac{u}{\phi}\right)^\tau} \right) \end{aligned} \tag{20}$$

Unfortunately, the above integral can not be simplified but it can be computed by making use of numerical integration.

5.2 Value-at-Risk

To ensure the insolvency of the insurer with a specified degree of certainty, the actuarial measure Value-at-Risk (`VaR`) is widely used by practitioners. `VaR` of a rv Y is the q-th quantile of its distribution function (Artzner, 1999).
If random variable Y follows Pareto-Weibull (α, ϕ, τ), q-th quantile can computed using the following equation

$$1 - \left(\left(\frac{\phi}{y}\right)^\alpha \left(\Gamma\left(\frac{\alpha+\tau}{\tau}\right) - \Gamma\left(\frac{\alpha+\tau}{\tau}, \left(\frac{y}{\phi}\right)^\tau\right) \right) + e^{-\left(\frac{y}{\phi}\right)^\tau} \right) = q$$

A closed form expression of the Value-at-Risk for the Pareto-Weibull distribution can not be given due to complex structure of the distribution function. q-th quantile can be found by solving the above equation numerically for y.

6 Estimation

The Pareto-Weibull distribution is generated by assuming the scale parameter of the Pareto distribution follows Weibull distribution, due to this operation missing data were produced. This missing data also called as latent variables. In our case (y_i, x_i) is the complete data vector and x_i is the latent information generated during the mixing procedure.

Let $Y_1, Y_2, ..., Y_n$ be a sample of size n from Pareto-Weibull random variable Y. For each y_i we take missing data the Weibull parameter x_i. $\theta = (\alpha, \phi, \tau)$ be the vector of parameters of the Pareto-Weibull distribution. Then the the log-likelihood function for the complete data is

$$l_c(\theta; y_i, x_i) = \sum_{i=1}^{n} \log(f(y_i|x_i)) + \sum_{i=1}^{n} \log(f(x_i|\phi, \tau)) \tag{21}$$

$$\begin{aligned} l_c(\theta; y_i, x_i) = &\sum_{i=1}^{n} [\log(\alpha) + \alpha \log(x_i) - (\alpha + 1)\log(y_i)] \\ &+ \sum_{i=1}^{n} \left[\log(\tau) - \log(\phi) - \left(\frac{x_i}{\phi}\right)^{\tau} + (\tau - 1)(\log(x_i) - \phi)\right] \end{aligned} \tag{22}$$

Due to presence of missing data, the complete data log-likelihood is not quadratic in parameters which makes usual maximum likelihood estimation procedure cumbersome. In a variety of situations, the Expectation-Maximization (EM) algorithm is a attractive alternative. It is now a widely used method for iterative maximum likelihood estimation in a range of challenges including missing data or inadequate information.

6.1 Maximum likelihood estimation For Pareto-Weibull Model via Expectation-Maximization (EM) Algorithm

The Expectation-Maximization (EM) algorithm (see, Dempster et al., 1977, and McLachlan and Krishnan, 2007) constitutes two steps. In Expectation-step, the information of latent variables can be collected by applying expectation on the conditional distribution of $X|Y$. Once the information of missing data is available, the unknown parameters are estimated in the Maximization-step. The EM type algorithm for the Pareto-Weibull distribution can be described as follows

- E-step: The E-step is used to fill the missing data. It computes the expected value of $l_c(\theta; Y, X)$ given the observed data, Y, and the current parameter estimate, θ^r say. The conditional expectation of complete log-likelihood function given in (22), say Q function can be written as

$$Q(\theta; \theta^{(r)}) = \mathbb{E}_{X_i}(l_c(\theta; Y, X)|Y, \theta^{(r)})$$

$$
\begin{aligned}
Q(\theta;\theta^{(r)}) &= \sum_{i=1}^{n}\left[\log(\alpha^{(r)}) + \alpha^{(r)}.\mathbb{E}(\log(x_i)) - (\alpha^{(r)}+1).\log(y_i)\right] \\
&+ \sum_{i=1}^{n}\left[\log(\tau^{(r)}) - \log(\phi^{(r)}) - \left(\frac{\mathbb{E}(x_i^{\tau^{(r)}})}{\phi^{(r)\tau^{(r)}}}\right) + (\tau^{(r)}-1)(\mathbb{E}(\log(x_i)) - \phi^{(r)})\right]
\end{aligned} \tag{23}
$$

Expectation step consisting computational part of conditional expectation of some functions of unobserved random variable X that are needed for the maximization. Using the current estimates $\hat{\alpha}^{(r)}$, $\hat{\phi}^{(r)}$ and $\hat{\tau}^{(r)}$, we calculate the pseudo-values,

$$
t_i = \mathbb{E}(\log(X_i)|y_i;\hat{\alpha}^{(r)},\hat{\phi}^{(r)},\hat{\tau}^{(r)}) = \frac{\int_0^y \log(x_i)\frac{\alpha^{(r)}x_i^{\alpha^{(r)}}}{y_i^{\alpha^{(r)}+1}}.\frac{\tau^{(r)}}{\phi^{(r)}}e^{-\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}}}\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}-1}dx_i}{\int_0^y \frac{\alpha^{(r)}x_i^{\alpha^{(r)}}}{y_i^{\alpha^{(r)}+1}}.\frac{\tau^{(r)}}{\phi^{(r)}}e^{-\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}}}\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}-1}dx_i}
$$

$$
s_i = \mathbb{E}(X_i^{\hat{\tau}^{(r)}}|y_i;\hat{\alpha}^{(r)},\hat{\phi}^{(r)},\hat{\tau}^{(r)}) = \frac{\int_0^y x_i^{\hat{\tau}^{(r)}}\frac{\alpha^{(r)}x_i^{\alpha^{(r)}}}{y_i^{\alpha^{(r)}+1}}.\frac{\tau^{(r)}}{\phi^{(r)}}e^{-\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}}}\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}-1}dx_i}{\int_0^y \frac{\alpha^{(r)}x_i^{\alpha^{(r)}}}{y_i^{\alpha^{(r)}+1}}.\frac{\tau^{(r)}}{\phi^{(r)}}e^{-\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}}}\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}-1}dx_i}
$$

$$
w_i = \mathbb{E}(X_i^{\hat{\tau}^{(r)}}\log(X_i)|y_i;\hat{\alpha}^{(r)},\hat{\phi}^{(r)},\hat{\tau}^{(r)}) = \frac{\int_0^y x_i^{\hat{\tau}^{(r)}}\log(x_i)\frac{\alpha^{(r)}x_i^{\alpha^{(r)}}}{y_i^{\alpha^{(r)}+1}}.\frac{\tau^{(r)}}{\phi^{(r)}}e^{-\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}}}\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}-1}dx_i}{\int_0^y \frac{\alpha^{(r)}x_i^{\alpha^{(r)}}}{y_i^{\alpha^{(r)}+1}}.\frac{\tau^{(r)}}{\phi^{(r)}}e^{-\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}}}\left(\frac{x_i}{\phi^{(r)}}\right)^{\tau^{(r)}-1}dx_i}
$$

- M-step: The M-step consists of maximizing over parameter vector θ using the conditional expectations computed in E-step. This step generates complete data after the expectation step and updates the parameters of the model. That is, we set

$$
\theta^{(r+1)} = \max Q(\theta;\theta^{(r)})
$$

Where $\theta^{(r)}$ be the estimates of the parameter vector θ at the r^{th} iteration. The updated values of the parameters are

$$
\hat{\alpha}^{(r+1)} = \frac{n}{(\sum_{i=1}^{n}\log(y_i) - \sum_{i=1}^{n}t_i)}
$$

$$\hat{\phi}^{(r+1)} = \left(\frac{\sum_{i=1}^{n} s_i}{n}\right)^{\frac{1}{\hat{\tau}^{(r)}}}$$

$$\frac{n}{\hat{\tau}^{(r+1)}} + \sum_{i=1}^{n} t_i - n\log(\hat{\phi}^{(r)}) - \frac{\sum_{i=1}^{n} w_i}{(\hat{\phi}^{(r)})^{\hat{\tau}^{(r+1)}}} + \frac{\log(\hat{\phi}^{(r)})\sum_{i=1}^{n} s_i}{(\hat{\phi}^{(r)})^{\hat{\tau}^{(r+1)}}} = 0 \qquad (24)$$

Solving (24) for the $\hat{\tau}^{(r+1)}$, we can obtain the improved estimate of $\hat{\tau}^{(r+1)}$.

- Finally, iterate between the E-step and the M-step until some convergence criterion is satisfied, for example the relative change in log-likelihood between two successive iterations is smaller than 10^{-12}.

More specifically, the EM-type algorithm can be written as follows.

Algorithm	**EM Algorithm for the Pareto-Weibull Distribution**
1.	Supply the initial values say $\theta^0 = (\alpha^0, \phi^0, \tau^0)$
2.	(E-step) improve the conditional expectations $t_i = \mathbb{E}(\log(X_i)\vert y_i; \hat{\alpha}^{(r)}, \hat{\phi}^{(r)}, \hat{\tau}^{(r)})$, $s_i = \mathbb{E}(X_i^{\hat{\tau}^{(r)}}\vert y_i; \hat{\alpha}^{(r)}, \hat{\phi}^{(r)}, \hat{\tau}^{(r)})$ and $w_i = \mathbb{E}(X_i^{\hat{\tau}^{(r)}}\log(X_i)\vert y_i; \hat{\alpha}^{(r)}, \hat{\phi}^{(r)}, \hat{\tau}^{(r)})$ using $\theta^{(r)}$ for $i = 1, 2, ..., n$ from the r^{th} iteration
3.	(M-step) Find the maximum global point, $\theta^{(r+1)}$ of the log-likelihood function $Q(\theta; \theta^{(r)})$
4.	If the criterion $\left\vert\frac{l^{(r+1)} - l^{(r)}}{l^{(r)}}\right\vert <$ epsilon is satisfied, the estimate of θ is $\theta^{(r+1)}$. Otherwise, update $\theta^{(r)}$ by $\theta^{(r+1)}$ and return to step 2

6.2 EM algorithm for Maximum Likelihood Estimation of Mixed Pareto Regression Models

Given the random sample $Y_1, Y_2, ..., Y_n$ with $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ as the corresponding vectors of covariates, there are various estimation techniques available in the literature to estimates unknown parameters of the model. Due to heavy-tailed structure of the distributions under consideration and complex structure of the log-likelihood which contains the unsolvable variable, many of the estimation methods fails to give efficient estimates of the model parameters. Hence we consider the maximum likelihood estimation procedure to maximize the log-likelihood function given by

$$l(\theta) = \sum_{i=1}^{n} \log(f(\mathrm{y_i})) \qquad (25)$$

where $\theta = (\beta_1^\top, \beta_2^\top)^\top$ is the parameter vector for the above mixed Pareto regression models with pdf $f(y_i)$ given in (7).
Because of the complex structure of the log-likelihood function of the mixed Pareto models presented in section (2), and when regression structure is applied to the parameters of the mixed Pareto models, direct maximization of the aforesaid function with regard to the

vector of parameters θ is difficult. To make the process of maximization of log-likelihood $l(\theta)$ easier and faster, EM algorithm is presented for the mixed Pareto regression models.

6.2.1 EM algorithm for ML estimation of Pareto-Weibull regression model

Due to the complex structure of density function of the Pareto-Weibull model, direct maximization of the log-likelihood through usual way will not result into efficient estimates of the parameters of the model. An EM algorithm can be used to efficiently compute the estimates of the various parameter of the Pareto-Weibull model as follows
The complete data log-likelihood takes the form

$$\begin{aligned} l_c(\theta) = & \sum_{i=1}^{n} [\log(\alpha_i) + \alpha_i . \log(z_i) - (\alpha_i + 1). \log(y_i)] \\ & + \sum_{i=1}^{n} \left[\log(\phi_i) + \log\left(\Gamma\left(1+\frac{1}{\phi_i}\right)\right) + (\phi_i - 1). \left(\log(z_i) + \log\left(\Gamma\left(1+\frac{1}{\phi_i}\right)\right)\right) - z_i . \Gamma\left(1+\frac{1}{\phi_i}\right)\right] \end{aligned} \tag{26}$$

The expectation of $\log(z_i)$ and z_i are needed for the process of M-step. The EM algorithm can be written as

- E-step: The required expectations for $i = 1, 2...n$ can be computed as

$$t_i = \mathbb{E}_{Z_i}\left[\log(Z_i)|y_i;\theta^{(r)}\right]$$

$$t_i = \frac{\int_0^y \log(z_i) \frac{\alpha_i^{(r)} z_i^{\alpha_i^{(r)}}}{y_i^{\alpha_i^{(r)}+1}} . \phi_i^{(r)} . \left[\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right] e^{-\left(z_i\left[\Gamma(1+\frac{1}{\phi_i^{(r)}})\right]\right)^{\phi_i^{(r)}}} \left(z_i\left[\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right]\right)^{\phi_i^{(r)}-1} dz_i}{\int_0^y \frac{\alpha_i^{(r)} z_i^{\alpha_i^{(r)}}}{y_i^{\alpha_i^{(r)}+1}} . \phi_i^{(r)} . \left[\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right] e^{-\left(z_i\left[\Gamma(1+\frac{1}{\phi_i^{(r)}})\right]\right)^{\phi_i^{(r)}}} \left(z_i\left[\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right]\right)^{\phi_i^{(r)}-1} dz_i} \tag{27}$$

and

$$w_i = \mathbb{E}_{Z_i}\left[Z_i|y_i;\theta^{(r)}\right]$$

$$w_i = \frac{\int_0^y (z_i) \frac{\alpha_i^{(r)} z_i^{\alpha_i^{(r)}}}{y_i^{\alpha_i^{(r)}+1}} . \phi_i^{(r)} . \left[\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right] e^{-\left(z_i\left[\Gamma(1+\frac{1}{\phi_i^{(r)}})\right]\right)^{\phi_i^{(r)}}} \left(z_i\left[\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right]\right)^{\phi_i^{(r)}-1} dz_i}{\int_0^y \frac{\alpha_i^{(r)} z_i^{\alpha_i^{(r)}}}{y_i^{\alpha_i^{(r)}+1}} . \phi_i^{(r)} . \left[\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right] e^{-\left(z_i\left[\Gamma(1+\frac{1}{\phi_i^{(r)}})\right]\right)^{\phi_i^{(r)}}} \left(z_i\left[\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right]\right)^{\phi_i^{(r)}-1} dz_i} \tag{28}$$

where $\phi_i^{(r)} = \exp(\mathrm{x}_2^\top \beta_2^{(r)})$.
The closed form expressions for the above expectations are not easily available hence numerical approximations are required to compute above mentioned quantities.

- M-step: Using the numerical approximate value of t_i, update the regression parameters β_1 using Newton-Raphson method.

$$h_2(\beta_2) = \left[1 + \left(\left(1 - \Gamma\left(1 + \frac{1}{\phi_i^{(r)}}\right)^{\phi_i^{(r)}} .w_i^{\phi_i^{(r)}}\right)\left(\phi_i^{(r)}.\log\left(\Gamma\left(1 + \frac{1}{\phi_i^{(r)}}\right)\right) - \digamma\left(1 + \frac{1}{\phi_i^{(r)}}\right)\right.\right.\right.$$
$$\left.\left.\left. + \phi_i^{(r)}.t_i\right)\right)\right] \mathbf{x}_{2,ij} \tag{29}$$

$$H_2(\beta_2) = [A_1 + A_2 + A_3 + A_4 + A_5]\, \mathbf{x}_{2,ij}\mathbf{x}_{2,ij}^\top = X_2^\top W_2 X_2 \tag{30}$$

where $i = 1, 2, 3, ...n$ and $j = 1, 2, 3....p_2$. The matrix W_2 can be written as $W_2 = diag\{A_1 + A_2 + A_3 + A_4 + A_5\}$.

Where $A_1 = \phi_i^{(r)}\log(\Gamma(1+\frac{1}{\phi_i^{(r)}})) - \Psi^{(0)}(1+\frac{1}{\phi_i^{(r)}})$, $A_2 = \frac{\Psi^{(1)}\left(1+\frac{1}{\phi_i^{(r)}}\right)}{\phi_i^{(r)}}$, $A_3 = \phi_i^{(r)}\log(z_i)$,

$A_4 = \log(z_i)\left[1 + \phi_i^{(r)}.\log(z_i) + \phi_i^{(r)}.\log(\Gamma(1+\frac{1}{\phi_i^{(r)}})) - \Psi^{(0)}(1+\frac{1}{\phi_i^{(r)}})\right].z_i^{\phi_i^{(r)}}\left(\Gamma(1+\frac{1}{\phi_i^{(r)}})\right).\phi_i^{(r)}$

and $A_5 = z_i^{\phi_i^{(r)}}.\left(\Gamma(1+\frac{1}{\phi_i^{(r)}})\right)\left[\left(\phi_i^{(r)}\right)^2.\log\left(\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right)\right)\left(1 + \phi_i^{(r)}.\log(z_i) + \phi_i^{(r)}.\log\left(\Gamma\left(1+\frac{1}{\phi_i^{(r)}}\right.\right.\right.\right.$

The improved estimates of $\beta_2^{(r)}$ can be obtained using Newton Raphson method.

- Finally, iterate between the E-step and the M-step until some convergence criterion is satisfied, for example the relative change in log-likelihood between two successive iterations is smaller than 10^{-12}.

7 Numerical Application

7.1 Numerical Application for Pareto-Weibull

In this section, we illustrate the applicability of Pareto-Weibull distribution using two real world insurance losses data sets. Both the data sets exhibits right skewness, uni-modality and consists of large losses.

7.1.1 Danish fire insurance data set

First, we consider the well-known Danish fire insurance dataset that consists of 2492 fire insurance losses in millions of Danish kroner (DKr) from the years 1980 to 1990 (both inclusive), adjusted to reflect 1985 values. This dataset may be found in the 'SMPrcacticals' add-on package for `R`, available from the CRAN website http://cran.r-project.org/.

Table 1 provides the descriptive statistics for the Danish fire insurance data set and the visualization of data given in Figure 1. There are large losses present in the data along with right-skewness. The significant difference between Q_3 and Maximum of the

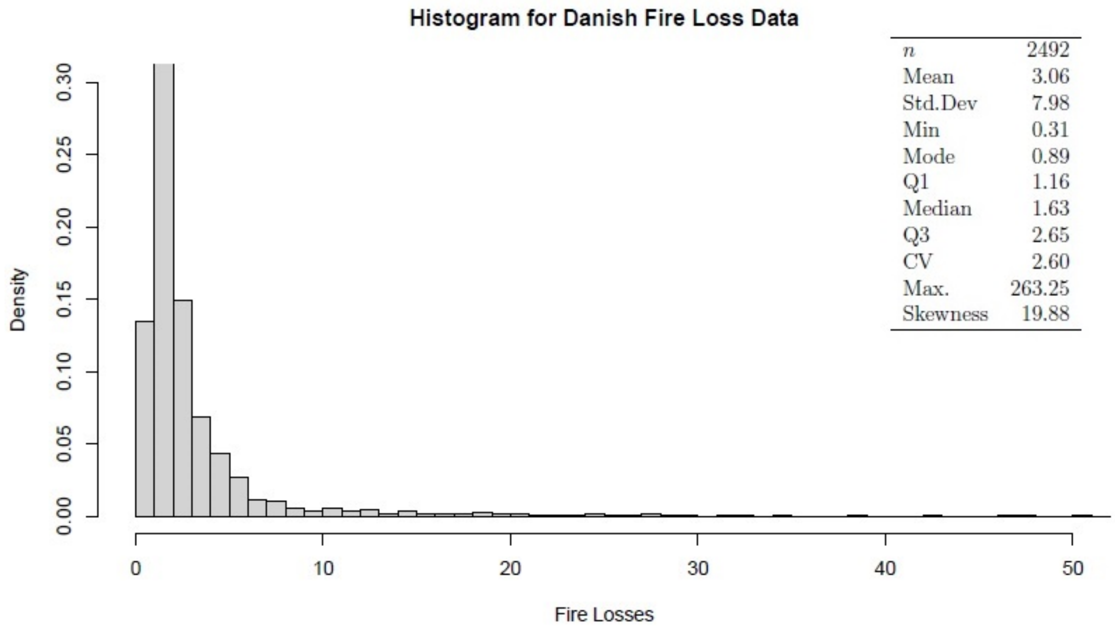

Figure 1: Histrogram for the claims

losses, represents the heavy tail nature of the data set. Due to presence of small number of large losses and large number of small losses Danish fire insurance data becomes the path breaker for researchers working in insurance.

Table 1: Descriptive Statistics of `Danish` fire insurance data

Measures	`Danish`
n	2492
Mean	3.06
Std.Dev	7.98
Min	0.31
Mode	0.89
Q1	1.16
Median	1.63
Q3	2.65
CV	2.60
Max.	263.25
Skewness	19.88

Table 2 gives a comparison of Pareto-Weibull distribution with other competent models in terms of goodness-of-fit. The negative log-likelihood (NLL) value along with AIC, BIC and K-S test statistic are provided for the comparison. Smallest value for all criterion indicates that Pareto-Weibull provides best fit. As compared to the standard parametric models, our model is considered to be competitive.

Table 3 indicates the empirical `VaR` as well as the estimated `VaR` from the fitted model for different security level (γ) ranging from 5% to 99%. In Table 4 , percentage of variation of each estimated `VaR` with respect to the empirical `VaR` is provided to impose ranking to the listed models for the comparison. As far as the ranking is concerned, for lower quantile (10%, 15% and 25%) the best model is Pareto-Weibull except for the security level 5%. For the higher quantile (65%, 75% and 85%) Pareto-Weibull contains the rank 1(65%) and rank 2(75% and 85%)

Finally Table 5 provides the empirical and estimated values of Limited expected value (`LEV`) for the fitted models at different policy limit u.

Table 2: Results for `Danish` fire insurance Data Set

Model	Parameter	NLL	AIC	BIC	K-S	Reference
PW	3	3839.43	7684.86	7689.05	0.049	Proposed Model
CW-P	3	3840.38	7686.76	7690.95	0.052	Calderín and Kwok (2016)
CL-P	3	3865.86	7737.72	7741.90	0.032	Calderín and Kwok (2016)
PL	3	3868.15	7742.30	7746.49	0.106	Colombi (1990)
Pareto	3	5051.91	10109.82	10114.01	0.290	–
E-IG	3	5088.45	10182.91	10187.09	0.283	–

Table 3: Value-at-Risk (`VaR`) for the `Danish` fire insurance data set at security level γ

	Model						
γ	PW	PL	Pareto	E-IG	CL-P	CW-P	Emperical
0.05	0.908	0.99	0.119	0.122	0.869	0.902	0.904
0.1	0.968	1.087	0.245	0.252	0.94	0.957	0.964
0.15	1.017	1.165	0.381	0.391	0.992	1.010	1.02
0.25	1.123	1.309	0.681	0.701	1.080	1.115	1.157
0.35	1.257	1.465	1.035	1.065	1.281	1.249	1.329
0.45	1.434	1.658	1.459	1.504	1.455	1.427	1.516
0.55	1.679	1.918	1.988	2.051	1.696	.673	1.735
0.65	2.047	2.301	2.68	2.764	2.053	2.043	2.049
0.75	2.668	2.937	3.66	3.773	2.653	2.668	2.645
0.85	3.989	4.252	5.276	5.421	3.915	4.002	3.884
0.95	9.475	9.425	9.344	9.446	9.039	9.571	8.406
0.99	33.655	30.254	17.103	16.561	30.793	34.334	24.613

7.1.2 Automobile bodily injury data set (`usautoBI`)

For the second application, we consider the automobile bodily injury claim dataset (`usautoBI`). This dataset contains automobile injury claims collected in 2002 by the Insurance Research Council (part of AICPCU and IIA). There are 1,340 records with demographic information, in addition to the claim amount. The data set may be found in the Package `CASdatasets` in R repository. Table 5 gives a descriptive statistics for `usautoBI` Data sets, which clearly shows that the data sets is heavy tailed and right skewed due to large difference between third quartile and maximum value. Figure 2. gives the histogram for the usautoBI data set. Table 7. provide information regarding the goodness-of-fit criterion applied on the existing models and Pareto-Weibull model. we provide the values of the NLL and Akaike information criterion (AIC) ,Bayesian information criterion (BIC)

Table 4: Ranking of competent models based on absolute value of the percentage of difference between each estimated value of `VaR` and empirical value of `VaR` corresponding to security level γ for `Danish` fire insurance data set

	security level γ											
Model	0.05	0.1	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95	0.99
PW	2	1	1	1	2	4	2	1	2	2	5	5
CW-P	1	2	2	2	3	5	3	3	3	3	6	6
CL-P	3	3	3	3	1	3	1	2	1	1	1	2
PL	4	4	4	4	4	6	4	4	4	4	3	1
Pareto	5	6	6	6	6	2	5	5	5	5	2	3
E-IG	6	5	5	5	5	1	6	6	6	6	4	5

Table 5: `LEV` for the `Danish` fire insurance data set

u	PW	PL	Pareto	E-IG	CL-P	CW-P	Empirical
1	0.957	0.964	0.660	0.678	0.909	0.987	0.935
2	1.387	1.210	0.974	0.992	1.418	1.527	1.551
3	1.594	1.452	1.297	1.315	1.669	1.801	1.839
5	1.755	1.604	1.472	1.517	.942	2.107	2.155
8	1.925	1.782	1.668	1.713	2.158	2.355	2.387
10	1.982	1.835	1.668	1.741	2.250	2.462	2.483
15	2.135	1.976	1.668	1.810	2.401	2.642	2.653
21	2.195	2.035	1.668	1.811	2.513	2.778	2.762
40	2.353	2.165	1.668	1.811	2.697	3.007	2.919
70	2.475	2.259	1.668	1.811	2.829	3.176	2.997
110	2.552	2.305	1.668	1.811	2.920	3.297	3.045
170	2.620	2.350	1.668	1.811	2.997	3.400	3.093
270	2.694	2.408	1.668	1.811	3.067	3.497	3.113

Table 6: Descriptive Statistics of `usautoBI` data set

Measures	`usautoBI`
n	1340
Mean	5.95
Std.Dev	33.14
Min	0.01
Mode	0.25
Q1	0.64
Median	2.33
Q3	4.00
CV	5.57
Max.	1067.70
Skewness	25.66

and K-S test statistic evaluated at the MLEs. After making comparisons across models, results suggests that the Pareto-Weibull model provides the best fit to the `usautoBI` data set, followed by the Exponential-Inverse Gamma model.

Table 9 provides the estimated values of `VaR` for various security level γ and empirical values of `VaR` corresponding to γ. Table 10 focuses on the ranking imposed on the competent models based on the absolute value of the percentage of difference between the each estimated value of `VaR` and empirical `VaR` for security level γ. For the estimation of high quantiles Pareto-Weibull performing better than the some of the existing models.

Estimated values of Limited expected value (`LEV`) for the different policy limit u for the Pareto-Weibull and competent models of `usutoBI` are given in Table 11. Empirical values of different policy limit u are also provided for the comparison.

Table 7: Results for `usautoBI` data set

Model	Parameter	NLL	AIC	BIC	K-S	Reference
PW	3	3128.75	6263.50	6266.88	0.0579	Proposed Model
Pareto	3	3145.92	6297.84	6301.22	0.0678	–
CL-P	3	3155.35	6316.70	6320.08	0.128	Calderin and Kwok(2016)
PL	3	3170.89	6347.78	6351.16	0.0920	Colombi (1990)
E-IG	3	3179.53	6365.06	6368.44	0.0783	–
CW-P	3	3207.95	6421.94	6425.32	0.149	Calderin and Kwok(2016)

Table 8: Value-at-Risk (`VaR`) for the `usautoBI` data set at security level γ

			Model				
γ	PW	PL	Pareto	E-IG	CL-P	CW-P	Emperical
0.05	0.139	0.154	0.119	0.108	0.281	0.174	0.138
0.1	0.284	0.263	0.248	0.227	0.539	0.348	0.238
0.15	0.439	0.377	0.387	0.359	0.791	0.523	0.32
0.25	0.781	0.644	0.709	0.607	1.288	0.872	0.64
0.35	1.184	0.987	1.102	1.071	1.778	1.221	1.318
0.45	1.674	1.449	1.601	1.596	2.265	1.569	1.984
0.55	2.294	2.101	2.261	2.32	2.666	1.918	2.64
0.65	3.133	3.085	3.191	3.382	3.220	2.636	3.305
0.75	4.396	4.731	4.645	5.096	4.146	3.126	3.994
0.85	6.762	8.078	7.403	8.411	6.087	4.778	5.893
0.95	15.438	19.861	16.539	18.88	13.897	11.901	15.508
0.99	50.693	54.402	44.148	43.779	46.572	45.304	67.822

7.2 Numerical Application for the Pareto-Weibull Regression model

To illustrate the applicability of proposed mixed Pareto regression models, we consider the vehicle insurance losses data set based on one-year vehicle insurance policies taken out in 2004 and 2005, The original data set comprises of 67,856 policies and we consider

Table 9: Ranking of competent models based on absolute value of the percentage of difference between each estimated value of VaR and empirical value of VaR corresponding to security level γ for the `usautoBI` data set

	Security Level γ											
Model	0.05	0.1	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95	0.99
PW	1	4	4	4	2	2	3	4	2	2	1	2
CW-P	5	5	5	5	1	5	6	6	5	3	5	4
CL-P	6	6	6	6	6	1	1	2	1	1	3	3
PL	2	3	2	1	5	6	5	5	4	5	6	1
Pareto	3	1	3	3	3	3	4	3	3	4	2	5
E-IG	4	2	1	2	4	4	2	1	6	6	4	6

Table 10: `LEV` for the `usautoBI` data set

u	PW	PL	Pareto	E-IG	CL-P	CW-P	Empirical
2	1.069	1.016	1.044	1.059	1.350	1.426	1.430
3	1.423	1.363	1.412	1.430	1.298	1.759	1.907
6	1.865	1.895	1.895	1.982	2.012	2.269	2.599
15	2.426	2.644	2.520	2.745	2.736	2.841	3.326
25	2.612	2.903	2.704	2.944	3.054	3.116	3.740
60	2.955	3.281	2.826	3.199	3.488	3.526	4.452
100	3.085	3.316	2.826	3.202	3.689	3.734	4.780
150	3.189	3.316	2.826	3.202	3.826	3.884	5.052
300	3.327	3.316	2.826	3.202	4.021	4.113	5.380
550	3.427	3.316	2.826	3.202	4.159	4.289	5.567
700	3.442	3.316	2.826	3.202	4.207	4.353	5.679
850	3.478	3.316	2.826	3.202	4.242	4.402	5.791
1000	3.494	3.316	2.826	3.202	4.271	4.442	5.902

4,624 policies having at least one claim. The response variable associated with the data set is CLMSIZE. We present the exploratory data analysis of the vehicle insurance data set to analyze the impact of the categorical explanatory variables on the response variables. Table 11 contains the descriptive statistics for the continuous explanatory variable and response variable of the vehicle insurance data set. The histogram of overall claim size is presented in Figure 2, which exhibits the uni-modality, positively skewed and heavy-tailed behavior of the data set. Figure 3 presents the summary of the categorical explanatory variables for the vehicle insurance data set which shows the contribution of the each category of the explanatory variables.

Details of the response variable and other explanatory variables are as follows:

- Claim (CLMSIZE): Total claim (In '000)
- Exposure (EXPSR): Exposure over the range (0-1)
- Gender (GENDR): Gender of the policy holder (0 for Male and 1 for Female)
- Vehicle Age (VEHAGE): Age of vehicle 1,2,3,4 where 1 is for youngest
- age category (AGECAT): Driver's age category 1,2,3,4,5,6 where 1 is for youngest

Table 11: Summary of continuous explanatory variable and response variable of the vehicle insurance loss data set

Statistic	CLMSIZE	EXPSR
Minimum	0.2	0.002
Maximum	55.922	0.999
Q_1	0.354	0.411
Q_3	2.091	0.832
Median	0.761	0.637
Mean	2.014	0.611
Skewness	5.041	-0.346
Kurtosis	43.215	2.128

7.2.1 Modeling Results

This part of the section provides the modeling results of the mixed Pareto regression models containing regression structure to both the parameters which are involved in the mean and dispersion and containig regression structure to only parameter involved in the dispersion. All the computational work is carried out using the R programming language. The models containing regression structure to both the parameters are converged after few iterations of EM algorithm using a strict stopping criterion. In particular, we iterated between the E-step and the M-step until the relative change in log-likelihood, between two successive iterations was smaller than 10^{-12}. The good starting values for the regression

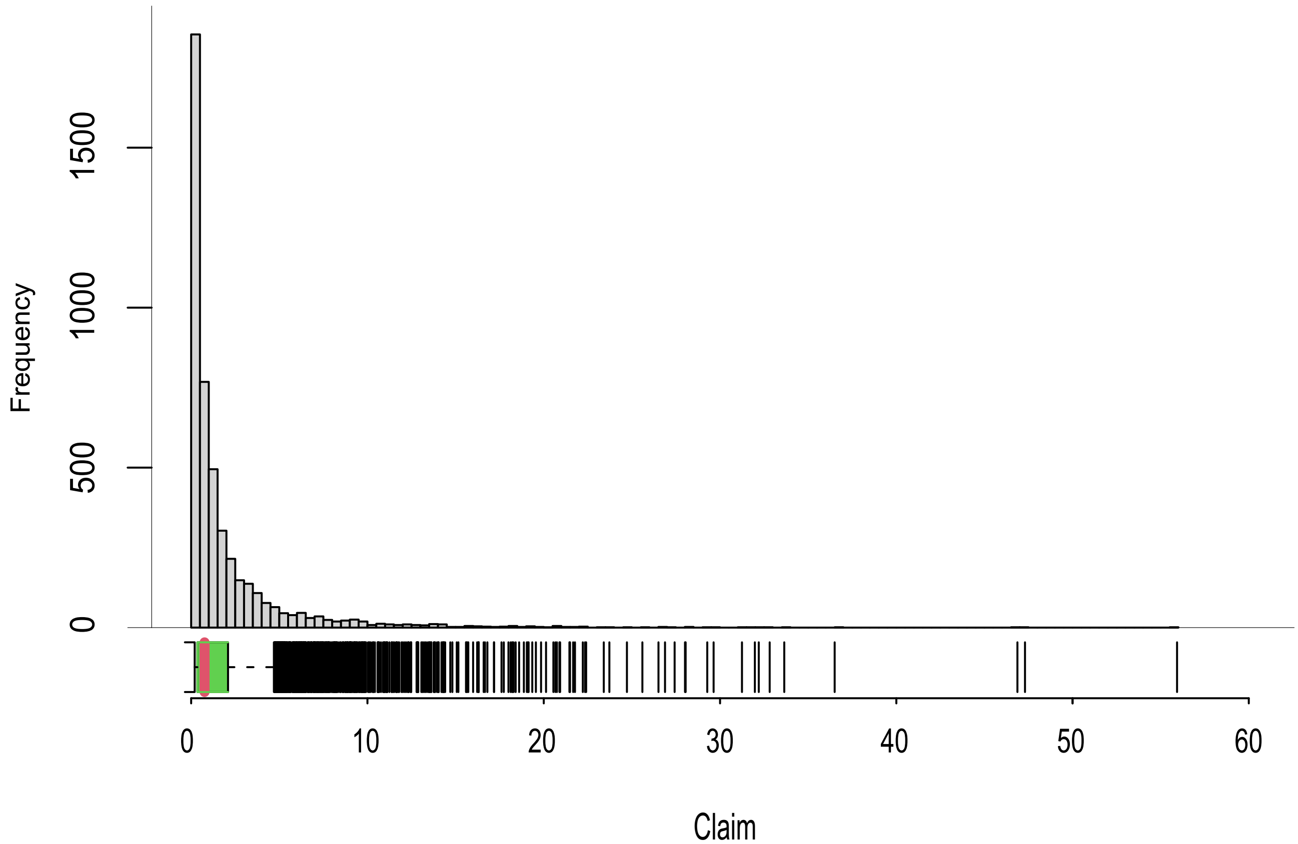

Figure 2: Histogram with Box plot for the overall claim

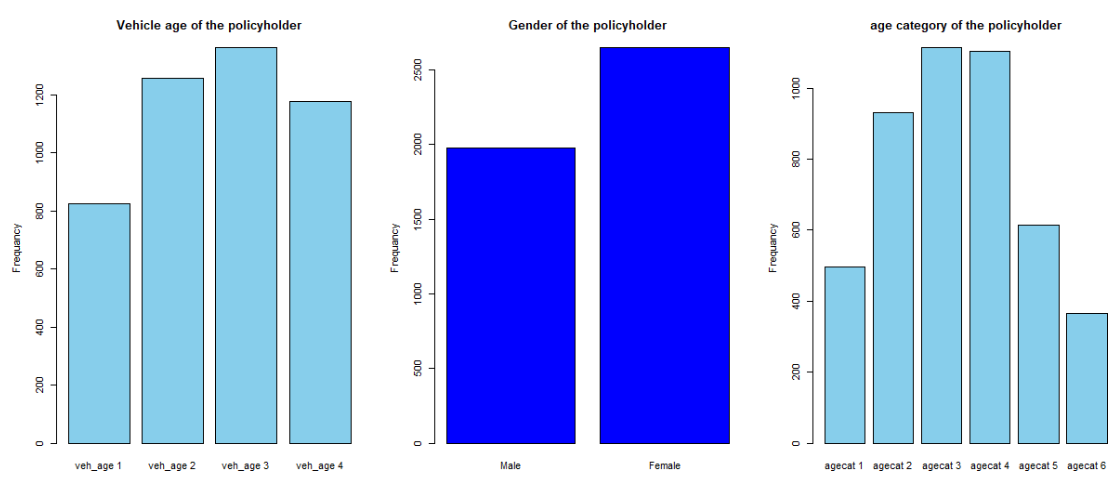

Figure 3: summary of categorical explanatory variable

coefficients β_1 and β_2 plays an importtant role for the EM algorithm to converge without giving inadmissible values of the regression coefficients. Good starting values of β_1 are obtained by the `GAMLSS` package of the `R` programming language by passing `PARETO2` as a `family` in the `GAMLSS` function. For the initial values of regression coefficient vector β_2 we checked many initial vales but for all the time out algorithm is converged to the same solution. To be specific with the selection of initial values of the β_2, we use value `1` for the intercept and `0` for the remaining coefficients.

The Maximum likelihood estimates of the parameters for the mixed Pareto regression models having regression structure for parameters which are involved in the mean and dispersion along with mixed Pareto regression models having regression structure for parameter which is involved in the dispersion regression model are presented in Table (12)

Table 12: Parameter estimates of the fitted models

Models		
Estimates	PW (α_i, ϕ_i)	PW (α, ϕ_i)
	Coefficients $\boldsymbol{\beta_1}$	
Intercept ($\beta_{1,0}$)	-0.0186	–
`EXPSR` ($\beta_{1,1}$)	0.7645	–
`VEHAGE` 2 ($\beta_{1,2}$)	-0.0561	–
`VEHAGE` 3 ($\beta_{1,3}$)	-0.1283	–
`VEHAGE` 4 ($\beta_{1,4}$)	-0.2055	–
`GENDR` 1 ($\beta_{1,5}$)	0.1411	–
`AGECAT` 2 ($\beta_{1,6}$)	0.2474	–
`AGECAT` 3 ($\beta_{1,7}$)	0.2839	–
`AGECAT` 4 ($\beta_{1,8}$)	0.2894	–
`AGECAT` 5 ($\beta_{1,9}$)	0.3996	–
`AGECAT` 6 ($\beta_{1,10}$)	0.3562	–
	Coefficients $\boldsymbol{\beta_2}$	
Intercept ($\beta_{2,0}$)	0.2454	0.1493
`EXPSR` ($\beta_{2,1}$)	-0.0404	0.0524
`VEHAGE` 2 ($\beta_{2,2}$)	0.0242	0.0241
`VEHAGE` 3 ($\beta_{2,3}$)	0.0704	0.0622
`VEHAGE` 4 ($\beta_{2,4}$)	0.1072	0.0918
`GENDR` 1 ($\beta_{2,5}$)	-0.0499	-0.0354
`AGECAT` 2 ($\beta_{2,6}$)	-0.0645	-0.0328
`AGECAT` 3 ($\beta_{2,7}$)	-0.0703	-0.0291
`AGECAT` 4 ($\beta_{2,8}$)	-0.0821	-0.0364
`AGECAT` 5 ($\beta_{2,9}$)	-0.0974	-0.0493
`AGECAT` 6 ($\beta_{2,10}$)	-0.0027	0.0282

7.2.2 Model Comparison

The model fit of the proposed mixed Pareto regression model having regression structure to both the parameters (α_i, ϕ_i) for the vehicle insurance dataset is compared with the mixed Pareto regression model with regression structure to only one parameter (α, ϕ_i). For the selection of best fitted model we have used three model selection criterion namely Deviance, AIC and SBC (also known as BIC). The formulae involved in the computation of above mentioned models selection criterion are

$$DEV = -2l(\hat{\theta})$$

where $l(\hat{\theta})$ is the maximum of the log-likelihood and $\hat{\theta}$ is the vector of the estimated model parameters. The AIC can be computed as

$$AIC = -2l(\hat{\theta}) + 2 \times df$$

and the SBC is given by

$$SBC = -2l(\hat{\theta}) + \log(n) \times df$$

where n is sample size of the data set and df is the number of fitted parameters of the model.

Table (13) represents the values of the log-likelihood, deviance, AIC and SBC of the vehicle insurance data set for different fitted model. It is point to be noted that for all three model comparison criterion smaller value represents the best fit of the model to the data set under consideration. As a common observation is that a model performs better than a competitor model if the difference in their log-likelihood exceeds five, which corresponds to the difference in the AIC exceeds 10 and difference in SBC exceeds 5 (see Burnham and Anderson, 2002). From (13) it is clear that the Pareto-Inverse Gaussian regression model(α_i, ϕ_i) is performing better than Pareto-Inverse Gaussian regression model(α, ϕ_i). Lowest value of NLL and other model comparison criterion of the Pareto-Inverse Gaussian regression model with regression structure to both the parameters involved in the mean and dispersion indicates the improvement in the performance of usual Pareto regression model having regression specification to the parameter involved in the mean.

Table 13: Values of Negative Log-Likelihood Function, Global deviance, AIC and SBC for vehicle insurance dataset

Model	df	NLL	Deviance	AIC	SBC
PW (α_i, ϕ_i)	22	7375.29	14750.61	14794.61	14936.26
PW (α, ϕ_i)	12	7433.33	14866.66	14890.66	14967.93

8 Conclusions

We have proposed the new heavy tailed distribution known as Pareto-Weibull distribution generated by considering the scale parameter of the Pareto density follows Weibull distribution. The important statistical properties of Pareto-Weibull distribution are derived. Some actuarial measures for the proposed distributions are also given to assess the tail behaviour of the Pareto-Weibull distribution. We have illustrated the application of Pareto-Weibull distribution to two real world data sets. By comparing Pareto-Weibull distribution with other existing distribution we conclude that the Pareto-Weibull performs better and can be a considerable heavy tailed distribution. We also proposed the mixed Pareto regression model having regression structure to both the parameters α_i and ϕ_i to account for the heterogeneity associated with the claim amount. A dedicated EM algorithm is employed to efficiently estimates the model parameters. One real life insurance data set namely vehicle insurance data set is provided to emphasize on the applicability of the proposed mixed Pareto regression model and to asses the utility of our proposed estimation methodology. The results obtained for vehicle insurance data set suggest that the Pareto-Inverse Gaussian regression model (α_i, ϕ_i) performs better than the Pareto-inverse Gaussian regression model (α, ϕ_i).

References

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

[2] Artzner, P. (1999). Application of coherent risk measures to capital requirements in insurance. *North American Actuarial Journal*, 3 (2), 11-25.

[3] 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.

[4] Bernardi, M., Maruotti, A., & Petrella M., (2012). Skew mixture models for loss distributions: A Bayesian approach. *Insurance: Mathematics and Economics*, 51 (3), 617-623.

[5] Bettina Grün & Tatjana Miljkovic (2019). Extending composite loss models using a general framework of advanced computational tools. *Scandinavian Actuarial Journal*, 8, 642-660.

[6] Bhati D., Ravi S.(2018). On generalized log-Moyal distribution: A new heavy tailed size distribution. *Insurance: Mathematics and Economics.*

[7] Calderín-Ojéda, E. & Kwok, C. F. (2016).Modeling claims data with composite Stoppa models. *Scandinavian Actuarial Journal*, 2016, 817–836.

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

[9] Colombi, R. (1990). A New Model Of Income Distribution: The Pareto-Lognormal Distribution. *Studies in Contemporary Economics. Income and Wealth Distribution, Inequality and Poverty*, 18-32.

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

[11] George Tzougas & Dimitris Karlis (2020).An EM algorithm for fitting a new class of mixed exponential regression models with varying dispersion. *ASTIN Bulletin: The Journal of the IAA*, 50 (2), 555-583

[12] José María Sarabia, Vanesa Jordá & Lorena Remuzgo (2017). The Theil Indices in Parametric Families of Income Distributions—A Short Review. *The review of income and wealth*, 63 (4), 867-880

[13] 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.

[14] Miljkovic T., & Grün B., (2016). Modeling loss data using mixtures of distributions. *Insurance: Mathematics and Economics*, 70, 387-396.

[15] Nadarajah, S. & Bakar, S. (2014). New composite models for the Danish fire insurance data. *Scandinavian Actuarial Journal*, 2014, 180–187.

[16] Punzo, A., Bagnato, L. & Maruotti, A., (2017). Compound unimodal distributions for insurance losses. *Insurance: Mathematics and Economics*. Accepted Manuscript.

[17] Theil, H. (1967). Economics and Information Theory. *North-Holland Publishing Company*, Amsterdam.

[18] Verbelen, R., Gong, L., Antonio, K., Badescu, A, & Lin, S., (2015). Fitting mixtures of Erlangs to censored and truncated data using the EM algorithm. *ASTIN Bulletin*, 45 (3), 729-758.

[19] Tzougas, G., & Karlis, D. (2020). An EM algorithm for fitting a new class of mixed exponential regression models with varying dispersion. ASTIN Bulletin: The Journal of the IAA, 50(2), 555-583.

Appendix

Table 14: Parameter Estimates for Danish Fire Insurance Data Set

Model	α	ϕ	τ	μ	σ	δ	β
PW	1.2697	0.9273	14.3494	–	–	–	–
CW-P	0.9917	0.9960	14.1014	–	–	1.2602	–
CL-P	1.1130	–	–	0.1457	0.1963	1.313	–
PL	1.3800	–	–	0.0526	0.1700	–	–
Pareto	–	–	–	2.1177	–	5.1694	5.6193
E-IG	–	–	–	2.1795	–	1.3255	5.9078

Table 15: Parameter Estimates for usautoBI Data Set

Model	α	ϕ	τ	μ	σ	δ	β
PW	1.3537	1.4841	1.0150	–	–	–	–
Pareto	–	–	–	1.3800	–	1.9113	3.1588
CL-P	2.7338	–	–	62.26	7.783	1.3300	–
PL	17.6774	–	–	0.5000	1.4775	–	–
E-IG	–	–	–	1.3461	–	3.5954	2.6386
CW-P	1.9055	1905487	1.000001	–	–	1.2040	–