Synopsis of

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

M.Sc Student:

Buddepu Pavan (2020MSTA004)

Supervisor:

Dr. Deepesh Bhati

A major project submitted in fulfillment of the requirements

for the degree of

Masters

in

Statistics

Department of Statistics

School of Mathematics, Statistics and Computational Science,

Central University of Rajasthan, Kishangarh, Ajmer

May 2022

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.

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

2 Literature Review

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.

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.

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

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.

George Tzougas et al. (2020) proposed a 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.

3 Methodology

Pareto-Weibull distribution

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

$$g_X(x) = \frac{\tau}{\phi} e^{-\left(\frac{x}{\phi}\right)^{\tau}} \left(\frac{x}{\phi}\right)^{\tau-1}, \quad x > 0, \phi > 0, \tau > 0, \tag{3.1}$$

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

$$f_{Y|X}(y|x) = \frac{\alpha x^{\alpha}}{y^{\alpha+1}}, \quad y > x, \alpha > 0. \tag{3.2}$$

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

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

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

$$f_{Y|Z}(y_i|z_i) = \frac{\alpha_i z_i^{\alpha_i}}{y_i^{\alpha_i+1}}, \tag{3.3}$$

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

$$f(y_i) = \int_0^y f(y_i|z_i) g(z_i; \tau_i)\, dz_i \tag{3.4}$$

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

$$\mathbb{E}(Y_i) = \mathbb{E}_{Z_i}[\mathbb{E}(Y_i|Z_i = z_i)] = \frac{\alpha_i}{(\alpha_i - 1)} \mathbb{E}_{Z_i}[Z_i] = \frac{\alpha_i}{(\alpha_i - 1)} \tag{3.5}$$

$$
\begin{aligned}
\mathbb{V}ar(Y_i) &= \mathbb{E}_{Z_i}(\mathbb{V}ar(Y_i|Z_i = z_i)) + \mathbb{V}ar_{Z_i}(\mathbb{E}(Y_i|Z_i = z_i)) \\
&= \frac{\alpha_i}{(\alpha_i - 1)^2(\alpha_i - 2)}\mathbb{E}_{Z_i}(Z_i^2) + \frac{\alpha_i}{(\alpha_i - 1)}\mathbb{V}ar_{Z_i}(Z_i)
\end{aligned} \tag{3.6}
$$

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.

4 Proposed Problems

- To develop a new heavy tailed mixed model using the methodology given in Colombi (1990).
- To provide an Expectation-Maximization (EM) algorithm to expdiate the process of finding maximum likelihood estimates of the model parameter
- To assess the applicability pf proposed EM type algorithm for new heavy tailed mixed model using a real world insurance data sets.
- To develop new heavy tailed regression model to accomodate available covariates to increase the predictive power of the model.
- To study the EM type algorithm with the regression sturucture to the parameters of the proposed mixed model.
- To examine the versatility of the proposed EM type algorithm using a real world insurance data set.

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.