A Family of Loss Distributions with an Application to the Vehicle Insurance Loss Data

Zubair Ahmad
Department of Statistics, Yazd University, P.O. Box 89175-741, Yazd, Iran
zubair@stu.yazd.ac.ir, z.ferry21@gmail.com

Eisa Mahmoudi
Department of Statistics, Yazd University, P.O. Box 89175-741, Yazd, Iran
emahmoudi@yazd.ac.ir

G.G. Hamedani
Department of Mathematics, Statistics and Computer Science, Marquette University, WI 53201-1881, Milwaukee, USA
g.hamedani@mu.edu

Abstract

Actuaries are often in search of finding an adequate loss model in the scenario of actuarial and financial risk management problems. In this work, we propose a new approach to obtain a new class of loss distributions. A special sub-model of the proposed family, called the Weibull-loss model is considered in detail. Some mathematical properties are derived and maximum likelihood estimates of the model parameters are obtained. Certain characterizations of the proposed family are also provided. A simulation study is done to evaluate the performance of the maximum likelihood estimators. Finally, an application of the proposed model to the vehicle insurance loss data set is presented.

Keywords: Family of distributions; Heavy tailed distributions; Weibull distribution; Vehicle insurance losses; Estimation.

1. Introduction

Speaking broadly, modeling insurance loss data with a heavy tail is a prominent research topic. Insurance loss data are positive, for detail see Klugman et al. (2012), and their distribution is often unimodal shaped, for example see Cooray and Ananda (2005), right-skewed, for detail see Lane (2000), Vernic (2006), Bolance et al. (2008), Bernardi et al. (2012), Ahn et al. (2012), Kazemi and Noorizadeh (2015) and Adcock et al. (2015), and with heavy tails, see Ibragimov et al. (2015). Actuaries are often interested in distributions that offer data modeling with heavy tail and provide a good estimate of the associated business risk level. Numerous heavy-tailed models have been proposed in the literature such as Pareto, lognormal, Lomax, Burr, Weibull, and gamma distributions, for a brief discussion, we refer the interested readers to Hogg and Klugman (2009).

Amongst these distributions, the Pareto distribution does not provide a better fit for many applications due to the monotonically decreasing shape of the density function, in particular when the shape of the data is hump-shaped, see, Cooray and Ananda (2005). On the other hand, due to incomplete form of distribution function, the lognormal and gamma distributions causing difficulties in the estimation process and computation of mathematical properties. Whereas, the Weibull model, which is a super-exponential distribution is a prominent heavy-tailed model. Thus Weibull distribution may be the initial choice to use for modeling losses with high outcomes (heavy tail) in finance and insurance, for detail we refer to Benckert and Jung (1974), Hogg and Klugman (1984), Szegö (2005), McNeil et al. (2005) and Klugman et al. (2012).

The Weibull model, however, fails to cover the behavior of large losses. Furthermore, other models have been introduced to handle this issue of fitting the tails of insurance losses, see for example, Scollnik and Sun (2012), Nadarajah and Abu Bakar (2014) and Abu Bakar et al. (2015).

To address some of the former issues with the existing models, several methods have been proposed to extend the existing distributions. These methods including, but not limited to, the following four approaches (i) transformation method, (ii) composition of two or more distributions, (iii) compounding of distributions and (iv) finite mixture of distributions. However, the new distributions introduced through these methods involve two or more extra parameters and the form of the density function becomes more complicated causing difficulties in estimating the model parameters.

To overcome these issues, we propose a new method of constructing new distributions. The proposed method is very flexible adding only one additional parameter to the existing distributions and provide greater distributional flexibility. If a random variable X follows a family of loss (for short F-Loss) distributions, then its cumulative distribution function (cdf) is given by

$$G(x;\sigma,\xi) = 1 - \frac{\sigma\bar{F}(x;\xi)}{\sigma - \log\left(\bar{F}(x;\xi)\right)}, \qquad \sigma,\xi > 0,\ x \in \mathbb{R}, \tag{1}$$

where, $\bar{F}(x;\xi) = 1 - F(x;\xi)$ is the survival function (sf) of the baseline distribution which may depend on the vector parameter $(\xi)^T$. To the best of our knowledge, the proposed method has not been used so far. This is another motivation for using the proposed method. Using our method several new distributions can also be obtained. The probability density function (pdf) corresponding to (1) is given by

$$g(x;\sigma,\xi) = \frac{\sigma f(x;\xi)\left[1 + \sigma - \log\left(\bar{F}(x;\xi)\right)\right]}{\left[\sigma - \log\left(\bar{F}(x;\xi)\right)\right]^2}, \qquad x \in \mathbb{R}. \tag{2}$$

We denote $X \sim F - Loss\,(x;\sigma,\xi)$ a random variable with pdf (2). The sf and hazard rate function (hrf) corresponding to (1) are given respectively, by

$$S(x;\sigma,\xi) = \frac{\sigma\bar{F}(x;\xi)}{\sigma - \log\left(\bar{F}(x;\xi)\right)}, \qquad x \in \mathbb{R},$$

and

$$h(x;\sigma,\xi) = \frac{f(x;\xi)\left[1 + \sigma - \log\left(\bar{F}(x;\xi)\right)\right]}{\bar{F}(x;\xi)\left[\sigma - \log\left(\bar{F}(x;\xi)\right)\right]}, \qquad x \in \mathbb{R}.$$

The goal of this research is to define and study a new family of loss distributions suitable for modeling insurance losses. Another main feature is that, it add greater flexibility to the generated distributions by introducing a single additional parameter rather than two or more parameters. Based on the proposed method, we introduce a three-parameter Weibull-Loss (W-Loss) model and give a comprehensive description of some of its mathematical properties in order to attract wider applications in the insurance sciences and other related areas. In fact, the W-Loss model can provide better fit to the insurance loss data than other well-known competitive models.

This article is organized as follows: In Section 2, we define the W-Loss distribution and provide some plots for its pdf. We provide in Section 3 some general mathematical properties of the F-Loss distributions. The maximum likelihood estimates (MLEs) of the unknown parameters and simulation study are presented in Section 4. In Section 5, certain characterizations of the proposed model are provided. The proposed W-Loss distribution is applied to the vehicle insurance loss data in Section 6. Finally, the article is concluded in Section 7.

2. A Special Sub-Model

In this section, we introduce a three-parameter special sub-model of the proposed family. Consider the cdf and pdf of the two-parameter Weibull distribution given by $F(x;\xi) = 1 - e^{-\gamma x^{\alpha}}$, $x \geq 0, \xi > 0$, and $f(x;\xi) = \alpha\gamma x^{\alpha-1} e^{-\gamma x^{\alpha}}$, respectively, where $\xi = (\alpha, \gamma)$. Then, the cdf of W-Loss distribution is given by

$$G(x;\sigma,\xi) = 1 - \frac{\sigma e^{-\gamma x^{\alpha}}}{\sigma + \gamma x^{\alpha}}, \qquad x \geq 0,\ \sigma,\xi > 0, \tag{3}$$

with pdf

$$g(x;\sigma,\xi) = \frac{\alpha\sigma\gamma x^{\alpha-1} e^{-\gamma x^{\alpha}} \left(1 + \sigma + \gamma x^{\alpha}\right)}{\left(\sigma + \gamma x^{\alpha}\right)^{2}}, \qquad x > 0. \tag{4}$$

For different combination of the parameters values, the plots of the density function are sketched in Figure 1.

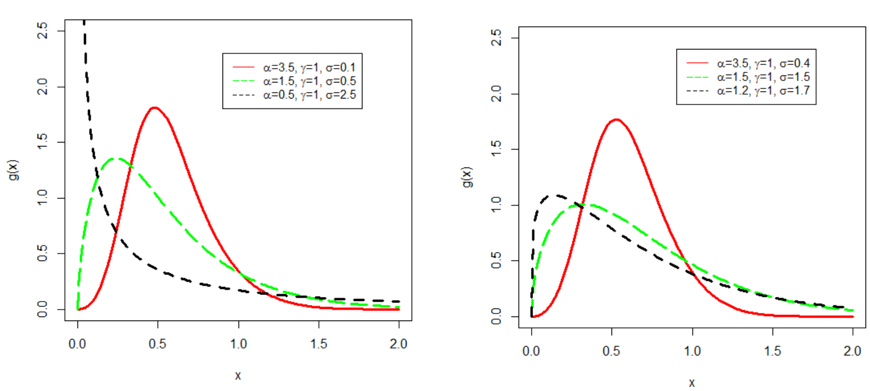

Figure 1: Plots of density function for selected parameter values.

3. Mathematical Properties

Here, statistical properties of the F-Loss distributions including quantile function, moments and moment generating function are discussed.

3.1 Quantile Function

The quantile function of the F-Loss distributions is derived as

$$x = Q\left(u\right) = G^{-1}\left(u\right) = F^{-1}\left(t\right), \tag{5}$$

where t is the solution of the equation$\sigma\left(1-t\right) - \left(1-u\right)\left(\sigma - \log\left(1-t\right)\right) = 0$, and u has the uniform distribution on the interval (0, 1). The nonlinear equation (3) can be used to obtain the random numbers for the F-Loss family of distributions.

3.2 Moments

In this sub-section, we derive the *rth* moment of a random variable X following the expression (2) as follows

$$\mu_r^{/} = \int_{-\infty}^{\infty} x^r \frac{f\left(x;\xi\right)\left[1+\sigma-\log\left(\bar{F}\left(x;\xi\right)\right)\right]}{\sigma\left[1-\frac{\log\left(\bar{F}\left(x;\xi\right)\right)}{\sigma}\right]^2} dx, \tag{6}$$

Using the following series (see, https://en.wikipedia.org/wiki/Taylor_series) $\frac{1}{(1+x)^2} = \sum_{n=1}^{\infty}\left(-1\right)^n n x^{n-1}$, and letting $x = \frac{-\log\left(\bar{F}(x;\xi)\right)}{\sigma}$, we have

$$\frac{1}{1+\frac{1}{\sigma}\{-\log(\bar{F}(x,\xi))\}} = \sum_{n=1}^{\infty}\left(-1\right)^n \frac{n}{\sigma^{n-1}}\left[-\log\left(\bar{F}\left(x;\xi\right)\right)\right]^{n-1}. \tag{7}$$

Inserting (7) in (6), we arrive at

$$\begin{aligned} \mu_r^{/} = &\sum_{n=1}^{\infty}\left(-1\right)^n \tfrac{n(1+\sigma)}{\sigma^n}\int_{-\infty}^{\infty} x^r f\left(x;\xi\right)\left[-\log\left(\bar{F}\left(x;\xi\right)\right)\right]^{n-1} dx \\ &+\sum_{n=1}^{\infty}\left(-1\right)^n \tfrac{n}{\sigma^n}\int_{-\infty}^{\infty} x^r f\left(x;\xi\right)\left[-\log\left(\bar{F}\left(x;\xi\right)\right)\right]^{n} dx. \end{aligned} \tag{8}$$

A useful expansion of the expression $\left[-\log\left(\bar{F}\left(x;\xi\right)\right)\right]^n$ and $\left[-\log\left(\bar{F}\left(x;\xi\right)\right)\right]^{n-1}$can be derived using the formula (http://functions.wolfram.com/Elementary Functions/ Log/06/01/04/03/):

$$\left(-\log\left(1-z\right)\right)^a = a\sum_{i=0}^{\infty}\left(\begin{array}{c} i-a \\ i \end{array}\right)\sum_{j=0}^{i}\frac{\left(-1\right)^{i+j}}{a-j}\left(\begin{array}{c} i \\ j \end{array}\right) p_{j,i} z^{a+i}, \tag{9}$$

where $a > 0$ is any real value. The constants $p_{j,i}$ can be calculated, recursively, via

$$p_{j,i} = \frac{1}{i}\sum_{m=1}^{i}\frac{\left(jm+m-i\right)\left(-1\right)^m}{m+1} p_{j,i-m}, \quad \text{for } i = 1,\ 2,\ 3, ...,\ \text{and } p_{j,0} = 1. \tag{10}$$

Taking $a = n$ and $b = n-1$, in (9), respectively, then from (8), we have

$$\begin{aligned}\mu_r^{/} &= \sum_{n=1}^{\infty} (-1)^n \frac{n(1+\sigma)K_{b,j,i}}{\sigma^n} \int_{-\infty}^{\infty} x^r f(x;\xi) F(x;\xi)^{b+i} dx \\ &+ \sum_{n=1}^{\infty} (-1)^n \frac{nK_{a,j,i}}{\sigma^n} \int_{-\infty}^{\infty} x^r f(x;\xi) F(x;\xi)^{a+i} dx,\end{aligned} \quad (11)$$

where $K_{b,j,i} = b\sum_{i=0}^{\infty} \binom{i-b}{i} \sum_{j=0}^{i} \frac{(-1)^{i+j}}{b-j} \binom{i}{j} p_{j,i}$ and

$K_{a,j,i} = a\sum_{i=0}^{\infty} \binom{i-a}{i} \sum_{j=0}^{i} \frac{(-1)^{i+j}}{a-j} \binom{i}{j} p_{j,i}$.

The moment generating function (mgf), say $M_X(t)$, of the F-Loss distributions can be obtained as follows

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r^{/}. \quad (12)$$

Using (11) in (12), we get the mgf of the F-Loss distributions.

4. Estimation and Simulation

In this section, we derive the maximum likelihood estimators of the F-Loss distributions using the maximum likelihood estimation method and provide the simulation study evaluating the behavior of these estimators.

4.1 Maximum likelihood estimation

Let $x_1, x_2, ..., x_n$ be the observed values from the F-Loss distributions with parameters (σ, ξ). The total log-likelihood function corresponding to (2) is

$$\begin{aligned}L(x_i;\sigma,\xi) &= n\log\sigma + \sum_{i=1}^{n} \log f(x_i;\xi) + \sum_{i=1}^{n} \log\left[1+\sigma-\log\left(1-F(x_i;\xi)\right)\right] \\ &- 2\sum_{i=1}^{n} \log\left[\sigma-\log\left(1-F(x_i;\xi)\right)\right].\end{aligned} \quad (13)$$

The log-likelihood function can be maximized either directly or by solving the non-linear likelihood equation obtained by differentiating (13). We used the goodness of fit function in R with "L-BFGS-B" algorithm to obtain the MLEs. The partial derivatives of (13) with respective to the parameters are given, respectively by

$$\frac{\partial}{\partial\sigma} L(x_i;\sigma,\xi) = \frac{n}{\sigma} + \sum_{i=1}^{n} \frac{1}{1+\sigma-\log\left(\bar{F}(x_i;\xi)\right)} - \sum_{i=1}^{n} \frac{2}{\sigma-\log\left(\bar{F}(x_i;\xi)\right)}. \quad (14)$$

and

$$\begin{aligned}\frac{\partial}{\partial\xi} L(x_i;\sigma,\xi) &= \sum_{i=1}^{n} \frac{\partial f(x_i;\xi)/\partial\xi}{f(x_i;\xi)} + \sum_{i=1}^{n} \frac{\left(\bar{F}(x_i;\xi)\right)^{-1} \partial F(x_i;\xi)/\partial\xi}{\left[1+\sigma-\log\left(\bar{F}(x_i;\xi)\right)\right]} \\ &- 2\sum_{i=1}^{n} \frac{\left(\bar{F}(x_i;\xi)\right)^{-1} \partial F(x_i;\xi)/\partial\xi}{\left[\sigma-\log\left(\bar{F}(x_i;\xi)\right)\right]}.\end{aligned} \quad (15)$$

4.2 Simulation Study

In this sub-section, we evaluate the performance of the maximum likelihood estimators presented in sub-section 4.1 for W-Loss distribution with respect to the sample size n. The mean square errors (MSEs), biases and absolute biases of the model parameter estimates are calculated by means of R software. The evaluation procedure is based on a simulation study as follow:

1. We generate $N = 1000$ samples of sizes $n = 25, 50, \ldots, 1000$ from the W-Loss distribution.

2. Calculate the maximum likelihood estimates for the model parameters.

3. Compute the MSEs and biases given by $MSE\,(n) = \frac{1}{1000}\sum_{i=1}^{1000}(\hat{w} - w)^2$ and $Bias\,(n) = \frac{1}{1000}\sum_{i=1}^{1000}(\hat{w} - w)$ for $w = (\alpha, \sigma, \gamma)$, respectively.

Figures 2 and 3, sketch the numerical results for the α with green line, σ with blue line and γ with red line. From Figures 2 and 3, we can easily observe that when the sample size increases, the empirical means approach the true parameter value. Whereas the biases and absolute biases decreases as the sample size n increases. This fact reveals the accuracy property of the MLEs. Furthermore, the estimated MSEs also decay toward zero as n increases. This fact reveals the consistency property of the MLEs.

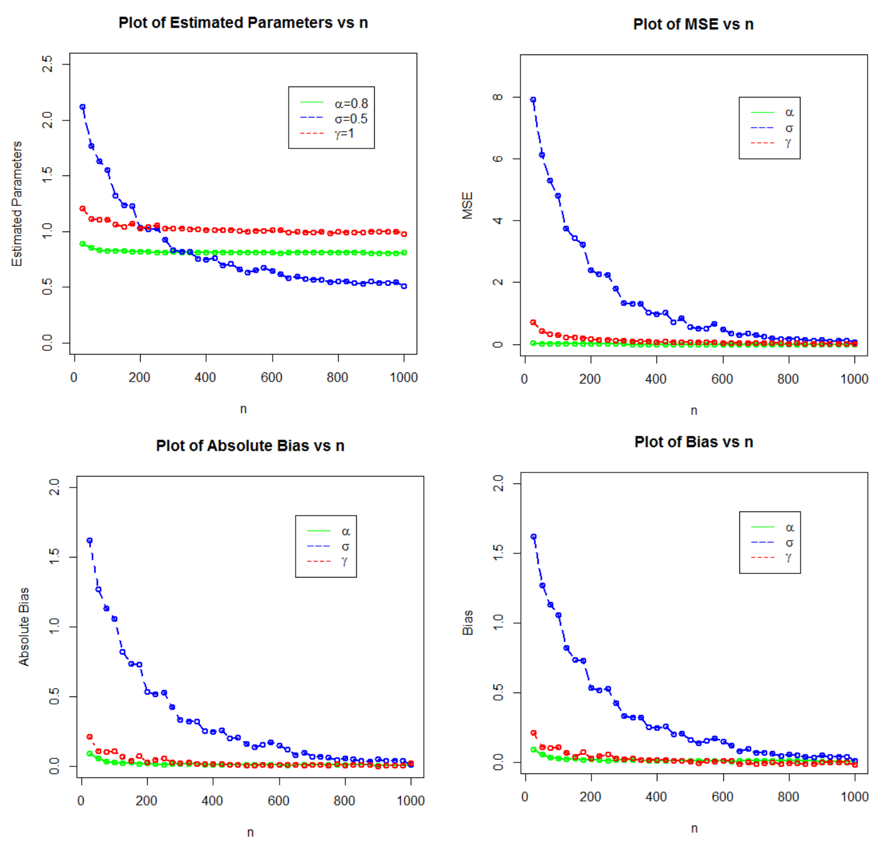

Figure 2: Plots of Estimated parameters, MSEs, Absolute Biases and Biases for α=0.8, σ=0.5 and γ=1.

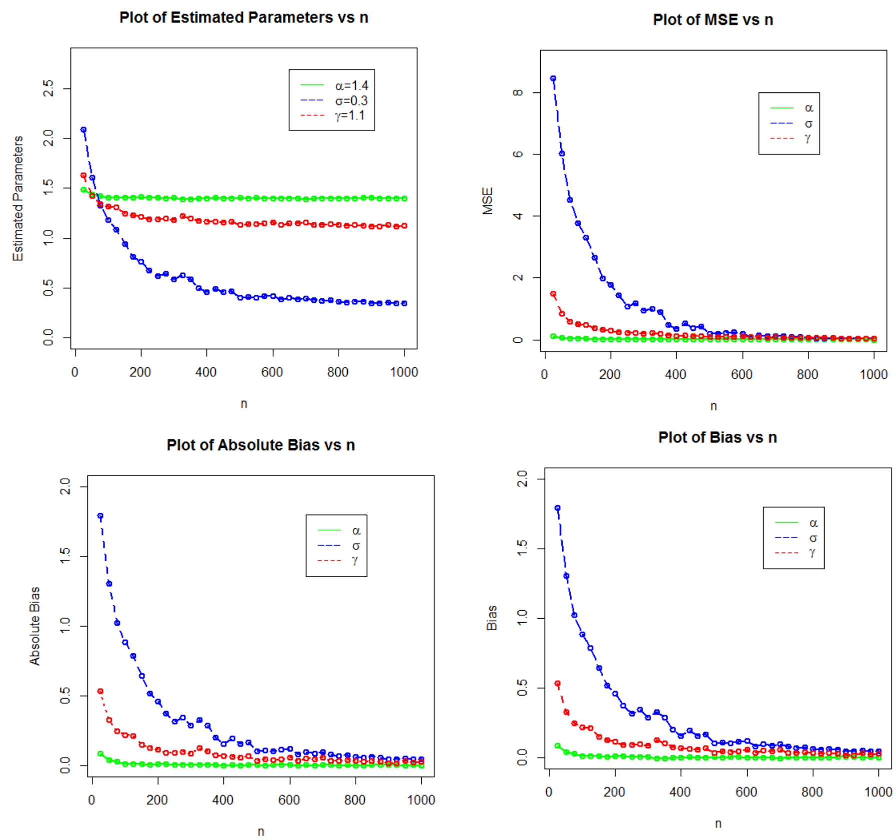

Figure 3: Plots of Estimated parameters, MSEs, Absolute Biases and Biases for α=1.4, σ=0.3 and γ=1.1.

5. Characterizations

This section deals with the characterizations of the F-Loss distribution based on: (i) a simple relation between two truncated moments; (ii) hazard function and (iii) the conditional expectation of a function of the random variable. Welike to mention that the characterization (i) can be employed when the cdf does not have a closed form. We present our characterizations (i)-(iii) in three subsections.

5.1 Characterizations based on two truncated moments

In this subsection we present characterizations of F-Loss distribution in terms of a simple relationship between two truncated moments. The first characterization result employs a theorem due to Glänzel (1987); see Theorem 1 below. Note that the result holds also when the interval H is not closed. Moreover, as mentioned above, it could be also applied when the cdf G does not have a closed form. As shown in Glänzel (1990), this characterization is stable in the sense of weak convergence.

Theorem 1. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a given probability space and let $H = [d, e]$ be an interval for some $d < e$ ($d = -\infty$; $e = \infty$ might as well be allowed). Let X: $\Omega \rightarrow H$ be a continuous random variable with the distribution function G and let q_1 and q_2 be two real functions defined on H such that

$$E\left(q_2\left(X\right)|X\geq x\right)=E\left(q_1\left(X\right)|X\geq x\right)\eta\left(x\right), \qquad x\in H,$$

is defined with some real function η. Assume that $q_1, q_2 \in C^1(H)$, $\eta \in C^2(H)$ and G is twice continuously differentiable and strictly monotone function on the set H. Finally, assume that the equation $\eta q_1 = q_2$ has no real solution in the interior of H. Then G is uniquely determined by the functions q_1, q_2 and η particularly

$$G\left(x\right)=\int_a^x C\left|\frac{\eta^{/}\left(u\right)}{\eta\left(u\right)q_1\left(u\right)-q_2\left(u\right)}\right|\exp\left(-s\left(u\right)\right)du,$$

where the function s is a solution of the differential equation $s^{/}=\frac{\eta^{/}q_1}{\eta q_1-q_2}$ and C is the normalization constant, such that $\int_H dG=1$.

Proposition 5.1. Let X:$\Omega\rightarrow\mathbb{R}$ be a continuous random variable and let $q_1(x)=\frac{\left[\sigma-\log\left(\bar{F}(x;\xi)\right)\right]^2}{1+\sigma-\log\left(\bar{F}(x;\xi)\right)}$ and $q_2(x)=q_1(x)F(x;\xi)$ for $x\in\mathbb{R}$. The random variable X has pdf (2) if and only if the function η defined in Theorem 1 is of the form

$$\xi\left(x\right)=\frac{1}{2}\left(1+F\left(x;\xi\right)\right), \qquad x\in\mathbb{R}.$$

Proof. Suppose the random variable X has pdf (2), then

$$\left(1-G\left(x\right)\right)E\left(q_1\left(X\right)|X\geq x\right)=\sigma\left(1-F\left(x;\xi\right)\right), \qquad x\in\mathbb{R},$$

and

$$\left(1-G\left(x\right)\right)E\left(q_2\left(X\right)|X\geq x\right)=\frac{\sigma}{2}\left(1-F\left(x;\xi\right)^2\right), \qquad x\in\mathbb{R},$$

and finally

$$\eta\left(x\right)q_1\left(x\right)-q_2\left(x\right)=\frac{q_1\left(x\right)}{2}\left(1-F\left(x;\xi\right)\right)>0, \qquad \text{for} \quad x\in\mathbb{R}.$$

Conversely, if η is of the above form, then

$$s^{/}\left(x\right)=\frac{\eta^{/}\left(x\right)q_1\left(x\right)}{\eta\left(x\right)q_1\left(x\right)-q_2\left(x\right)}=\frac{f\left(x;\xi\right)}{1-F\left(x;\xi\right)}, \qquad x\in\mathbb{R},$$

and hence

$$s\left(x\right)=-\log\left(1-F\left(x;\xi\right)\right), \qquad x\in\mathbb{R}.$$

Now, in view of Theorem 1, X has density (2).

Corollary 5.1. Let X: $\Omega\rightarrow\mathbb{R}$ be a continuous random variable and let $q_1(x)$ be as in Proposition 5.1. The random variable X has pdf (2) if and only if there exist functions q_2 and η defined in Theorem 1 satisfying the following differential equation

$$\frac{\eta^{/}\left(x\right)q_1\left(x\right)}{\eta\left(x\right)q_1\left(x\right)-q_2\left(x\right)}=\frac{f\left(x;\xi\right)}{1-F\left(x;\xi\right)}, \qquad x\in\mathbb{R}.$$

The general solution of the differential equation in Corollary 5.1 is

$$\eta(x) = (1 - F(x;\xi))^{-1} \left[- \int f(x;\xi) (q_1(x))^{-1} q_2(x) \, dx + D \right],$$

Note that a set of functions satisfying the above differential equation is given in Proposition 5.1 with D=1/2. However, it should be also noted that there are other triplets $(q_1(x), q_2(x), \eta(x))$ satisfying the conditions of Theorem 1.

5.2 Characterization in terms of hazard function

It is known that the hazard function, $h_G(x)$, of a twice differentiable distribution function, G, satisfies the following first order differential equation

$$\frac{g^{/}(x)}{g(x)} = \frac{h_G^{/}(x)}{h_G(x)} - h_G(x).$$

For many univariate continuous distributions, this is the only characterization available in terms of the hazard function. The following characterization establish a non-trivial characterization of F-Loss distribution in terms of the hazard function, which is not of the above trivial form.

Proposition 5.2. Let $X\colon \Omega \to \mathbb{R}$ be a continuous random variable. The random variable X has pdf (2) if and only if its hazard function $h_G(x)$ satisfies the following differential equation

$$h_G^{/}(x) = \frac{f^{/}(x;\xi)}{f(x;\xi)} h_G(x) = f(x;\xi) \frac{d}{dx} \left\{ \frac{1 + \sigma - \log\left(\bar{F}(x;\xi)\right)}{\bar{F}(x;\xi)\left[\sigma - \log\left(\bar{F}(x;\xi)\right)\right]} \right\} \qquad x \in \mathbb{R}.$$

Proof. Is straightforward and hence omitted.

5.3 Characterization based on the conditional expectation of certain function of the random variable

The following proposition has already appeared in Hamedani's previous work (2013), so we will just state it here which can be used to characterize the F-Loss distribution.

Proposition 5.3. Let X: $\Omega \to (a,b)$ be a continuous random variable with cdf F. Let $\psi(x)$ be a differentiable function on (a,b) with $\lim_{x\to a^+} \psi(x) = 1.$. Then for $\delta \neq 1$,

$$E[\psi(X) | X \geq x] = \delta \psi(x), \qquad x \in (a,b),$$

if and only if

$$\psi(x) = \{1 - F(x)\}^{\frac{1}{\delta} - 1}, \qquad x \in (a,b),$$

Remark 5.1. For $(a,b) = \mathbb{R}$, $\psi(x) = \frac{\sigma \bar{F}(x;\xi)}{\sigma - \log\left(\bar{F}(x;\xi)\right)}and\delta = \frac{1}{2}$, Proposition 5.3 provides a characterization of F-Loss distribution. Of course there are other suitable functions than the one we mentioned above, which is chosen for simplicity.

6. Application to the Vehicle Insurance Loss Data

The main applications of the heavy tail models are the so-called extreme value theory or insurance loss phenomena. We consider a data set from insurance losses. The data set representing the vehicle insurance losses available at:
http://www.businessandeconomics.mq.edu.au. We fitted the proposed model in comparison with the other heavy tailed distributions including the two parameters Weibull, Marshall-Olkin Weibull (MOW), modified Weibull (MW), exponentiated Weibull (EW), new Weibull burr X-II (NWBX-II), Kumaraswamy Weibull (Ku-W), Lomax, exponentiated Lomax (EL) and Burr X-II (BX-II) distributions. The distribution functions of the competitive models are:

1. Weibull

$$G(x) = 1 - e^{-\gamma x^{\alpha}}, \qquad x \geq 0, \alpha, \gamma > 0.$$

2. Marshall Olkin Weibull

$$G(x) = \frac{1 - e^{-\gamma x^{\alpha}}}{1 - (1-\sigma) e^{-\gamma x^{\alpha}}}, \qquad x \geq 0, \alpha, \gamma, \sigma > 0.$$

3. Modified Weibull

$$G(x) = 1 - e^{-\gamma x^{\alpha} - \theta x}, \qquad x \geq 0, \alpha, \theta, \gamma > 0.$$

4. Exponentiated Weibull

$$G(x) = \left(1 - e^{-\gamma x^{\alpha}}\right)^{a}, \qquad x \geq 0, \alpha, a, \gamma > 0.$$

5. NWBX-II

$$G(x) = 1 - \exp\left(-\gamma \left(k \log\left(1 + x^{c}\right)\right)^{\alpha}\right), \qquad x \geq 0, c, k, \alpha, \gamma > 0.$$

6. Kumaraswamy Weibull

$$G(x) = 1 - \left(1 - \left(1 - e^{-\gamma x^{\alpha}}\right)^{a}\right)^{b}, \qquad x \geq 0, \alpha, \gamma, a, b > 0.$$

7. Lomax

$$G(x) = 1 - \left(1 + \frac{x}{\gamma}\right)^{-\alpha}, \qquad x \geq 0, \alpha, \gamma > 0.$$

8. Exponentiated Lomax

$$G(x) = \left(1 - \left(1 + \frac{x}{\gamma}\right)^{-\alpha}\right)^{a}, \qquad x \geq 0, \alpha, a, \gamma > 0.$$

9. BX-II

$$G(x) = 1 - (1 + x^c)^{-k}, \qquad x \geq 0, c, k > 0.$$

To decide about the goodness of fit among the applied distributions, we consider certain analytical measures. In this regard, we consider four discrimination measures such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC) and Consistent Akaike Information Criterion (CAIC). These measures are given by

- The AIC is given by

$$AIC = 2k - 2l.$$

- The BIC is given by

$$BIC = k \log(n) - 2l.$$

- The HQIC is given by

$$HQIC = 2k \log(\log(n)) - 2l.$$

- The CAIC is given by

$$CAIC = \frac{2nk}{n-k-1} - 2l,$$

where l denotes the log-likelihood function evaluated at the MLEs, k is the number of model parameters and n is the sample size.
A lower value of these analytical measures is desirable. The maximum likelihood estimates of the model parameters are reported in Table 1. Whereas, the analytical measures are provided in Table 2. These results show that the proposed W-Loss distribution provides better fit than the other considered competitors. In support of Table 2, the estimated pdf and cdf of the W-Loss distribution are sketched in Figure 4. Whereas, the PP and Kaplan Meier survival plots are presented in Figure 5.

Table 1: Estimated values of the proposed and other competitive models for the vehicle insurance loss data.

Dist.	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{k}$
W-Loss	0.760	0.104	8.490				
Weibull	0.759	0.106					
MOW	1.153	0.085	3.098				
MW	0.529	0.001	100.000				
EW	0.176	3.389		9.120			
NWBX-II	0.253	1.883				1.993	0.399
Ku-W	0.4387	2.800		31.00	0.175		
Lomax	1.690	19.139					
EL	0.981	1.508		3.936			
BX-II						0.109	3.984

Table 2: Analytical measures of the proposed and other competitive models for the vehicle insurance loss data.

Dist.	AIC	BIC	CIAC	HQIC
W-Loss	1334.687	1347.750	1334.729	1339.781
Weibull	1432.698	1441.406	1432.719	1436.094
MOW	1410.814	1423.877	1410.856	1415.909
MW	1433.109	1441.056	1431.201	1417.094
EW	1400.419	1413.482	1397.440	1405.514
NWBX-II	1400.026	1417.443	1400.096	1406.819
Ku-W	1397.006	1414.423	1397.076	1403.799
Lomax	1418.450	1427.158	1418.471	1421.846
EL	1403.994	1417.057	1404.036	1409.089
BX-II	1467.001	1475.710	1467.022	1470.397

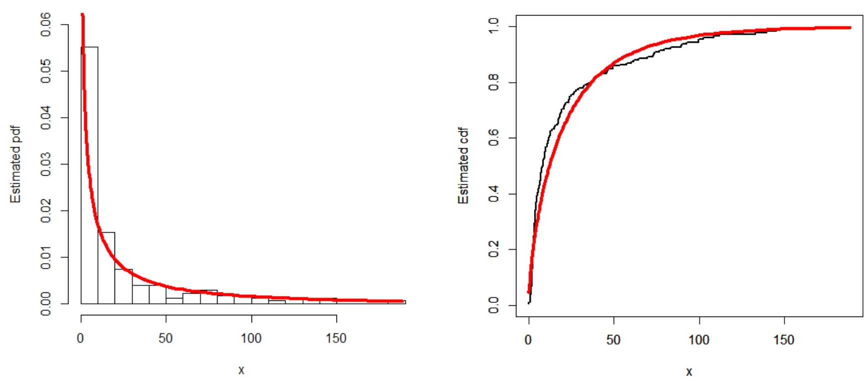

Figure 4: Estimated pdf and cdf of the W-Loss distribution for the vehicle insurance loss data.

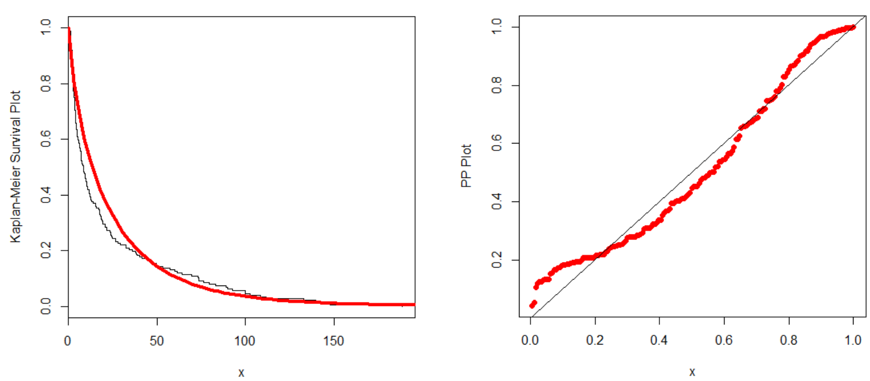

Figure 5: Kaplan Meier Survival and PP-plots of the W-Loss distribution for the vehicle insurance loss data.

7. Concluding Remarks

In this article, a family of loss distributions is proposed. Some of its properties along with certain characterizations of the family are derived. A three-parameter special sub-model of the proposed family called, the Weibull loss distribution capable of modeling heavy tailed data is studied in detail. Maximum likelihood estimators of the model parameters are obtained and simulation study is provided to evaluate the behavior of these estimators. A practical application to the insurance loss data is analyzed and the comparison of the proposed model with nine other well-known competitors is provided. The practical applications shows that the proposed model is a good candidate for modeling insurances losses. We hope that the proposed method will attract the wider applications in the actuarial sciences and other related fields.

Dedication

This article is part of PhD work of the first author. The author would like to dedicate this article to The Allah Almighty, to the memory of his late parents, beloved brothers and sisters.

References

1. Adcock, C., Eling, M. and Loperfido, N. (2015). Skewed distributions in finance and actuarial science: a review. The European Journal of Finance, 21(13-14), 1253-1281.

2. Ahn, S., Kim, J. H. and Ramaswami, V. (2012). A new class of models for heavy tailed distributions in finance and insurance risk. Insurance: Mathematics and Economics, 51(1), 43-52.

3. Bakar, S. A., Hamzah, N. A., Maghsoudi, M. and Nadarajah, S. (2015). Modeling loss data using composite models. Insurance: Mathematics and Economics, 61, 146-154.

4. Benckert, L.G. and Jung, J. (1974). Statistical models of claim distributions in fire insurance. ASTIN Bulletin: The Journal of the IAA, 8(1), 1-25.

5. Bernardi, M., Maruotti, A. and Petrella, L. (2012). Skew mixture models for loss distributions: a Bayesian approach. Insurance: Mathematics and Economics, 51(3), 617-623.

6. Bolance, C., Guillen, M., Pelican, E. and Vernic, R. (2008). Skewed bivariate models and nonparametric estimation for the CTE risk measure. Insurance: Mathematics and Economics, 43(3), 386-393.

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

8. Glänzel, W. (1987). A characterization theorem based on truncated moments and its application to some distribution families. In Mathematical statistics and probability theory (pp. 75-84). Springer, Dordrecht.

9. Glanzel, W. (1990). Some consequences of a characterization theorem based on truncated moments. Statistics, 21(4), 613-618.

10. Hamedani, G. G. (2013). On certain generalized gamma convolution distributions II (No. 484). technical report.

11. Hogg, R. V. and Klugman, S. A. (2009). Loss Distributions, vol. 249.

12. Hogg, R.U. and Klugman, S.A. (1984). Loss Distributions, John Wiley and Sons.

13. Ibragimov, M., Ibragimov, R. and Walden, J. (2015). Heavy-tailed distributions and robustness in economics and finance (Vol. 214). Heidelberg: Springer.

14. Kazemi, R. and Noorizadeh, M. (2015). A Comparison between Skew-logistic and Skew-normal Distributions. Matematika, 31(1), 15-24.

15. Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2012). Loss models: from data to decisions (Vol. 715). John Wiley & Sons.

16. Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2012). Loss models: from data to decisions (Vol. 715). John Wiley & Sons.

17. Lane, M. N. (2000). Pricing Risk Transfer Transactions 1. ASTIN Bulletin: The Journal of the IAA, 30(2), 259-293.

18. McNeil, A. J., Frey, R. and Embrechts, P. (2005). Quantitative risk management: Concepts, techniques and tools (Vol. 3). Princeton: Princeton university press.

19. Nadarajah, S. and Bakar, S. A. (2014). New composite models for the Danish fire insurance data. Scandinavian Actuarial Journal, 2014(2), 180-187.

20. Scollnik, D. P. and Sun, C. (2012). Modeling with weibull-pareto models. North American Actuarial Journal, 16(2), 260-272.

21. Szegö, G. (2005). Measures of risk. European Journal of Operational Research, 163(1), 5-19.

22. Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. Insurance: Mathematics and economics, 38(2), 413-426.