Splicing Lindely and some Heavy tailed distributions for Modelling Marine Hull Insurance Claims Data

Abdelrahman Elaraby Mosad Mohammed Saed [1] , Mohamed Tawfik EL-bolkiny[2], Jamaal Abd El Baki Wasif [3] , Mona Elbeshir Elshrbiny [4],

Abstract

Many methods and techniques are used to perform actuarial operations in property insurance companies. Among these methods are probability distributions, which are used to make decisions in Premium Rating, Reservation, Reinsurance Agreements, and Testing for Solvency.

This study aims to develop Probability distributions for Marine Hull Claims data for one Egyptian property insurance company.

This study introduced three mixture probability distributions using the splicing method (Lindely Pareto, Lindely Lomax, Lindely jumble) for Marine Hull insurance claims data.

This study shows that (Lindely Pareto distribution) is the best for Hull insurance claims data.

The previous results were obtained after conducting several tests including the maximized log-likelihood (-ℓ), Anderson-Darling (A), Cramér-Von Mises (W), Kolmogorov Smirnov (KS) statistics (with its p-value), Akaike information criterion (AIC), Bayesian information

[1] *Assistant lecture, Higher Institute for Computer & Information Technology, Elshorouk, Cairo, Egypt.*

[2] *Professor of Actuarial Sciences and Mathematics, Department of Applied Statistics and Insurance, Faculty of Commerce, Mansoura University, Egypt*

[3] *Professor of Actuarial sciences and Mathematics, Department of Applied statistics and Insurance, faculty of commerce, Mansoura University, Egypt*

[4] *Lecturer of Insurance, Department of Applied statistics and Insurance, faculty of commerce, Mansoura University, Egypt*

criterion (BIC), Hannan-Quinn information criterion (HQIC), and consistent Akaike information criterion (CAIC). The study used R statistical package software to calculate the previous measures.

Keywords:

Claims, Marine Hull insurance, Mixture distributions, probability distributions, Actuarial practice in property insurance.

1. Introduction

Data always plays a significant role in the insurance operations; the insurance contract obligates the insurance company to pay the claims arising from contracts. Therefore, the insurance company must maintain capital reserves to meet these future obligations. This means the insurance premium is paid before the actual costs are known. This is what is called the reversal of the production cycle. This means that the process of pricing and calculating reserves Claims are strongly interconnected in actuarial practice. On the other hand, actuaries must determine a fair price for the insurance products they wish to sell based on data generated from analysis models.

Modelling claims losses data is critical in premium rating, reserving, reinsurance agreements and testing for Solvency (Klugman, Panjer et al. 2012). Continuous distributions, such as gamma or lognormal, are usually used to model losses (Bakar, Hamzah et al. 2016). However, these parametric distributions are not always appropriate for actuarial data, which may be Multimodal or heavy-tailed. Loss models must have, on the one hand, flexibility in describing claims and, on the other hand, implement ability in quantitative risk analysis.

When conducting actuarial studies to model general insurance claims and using the resulting models in the pricing operation and calculating reserves, we find that the traditional distributions used (such as gamma, lognormal, Pareto,etc.) are not always appropriate because they do not reflect the nature of the actuarial data, especially with the presence of large fluctuations in claims values (very small and very large). It may also be Multimodal or heavy-tailed.

To clarify the previous problem, this study tried to model Hull Claims data during the period (2014/2015_2020/2021) in one of the Egyptian general insurance market companies using (the Easy Fit Professional 5.2) program.

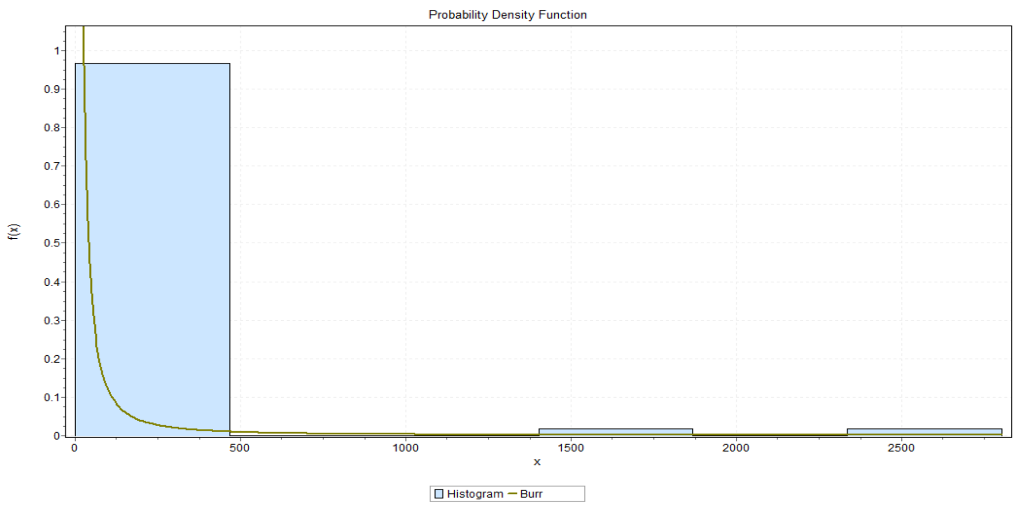

Figure 1: Results of (Easy Fit professional PR. 5.2) Modelling claims data

The above results are:

- The shape of the claims data distribution is skewed to the right with a heavy tail.
- The best distribution is Burr distribution.
- According to Kolmogorov-Smirnov test p.value = 0.178 and statistic = 0.1384688
- Since the p-value is small, the study attempts to find a solution to this problem.

The previous studies used traditional distributions such as, (Abdelhamid , N. A. 2019) which proposed a quantitative model for pricing agricultural crop insurance from all factors that affect the degree of risk. The study assumed some important probability distributions (negative exponential distribution, Pareto distribution, log-normal distribution, and gamma distribution) to try to modelling the loss data and goodness of fit tests (Kolmogorov-Smirnov) were performed to choose the appropriate distribution for pricing agricultural-crops.

Other study (EL-bolkiny, wasif et al. 2018) tried to build a quantitative model used in pricing non-life insurance. That paper uses the Easy Fit Professional program to try to modelling the frequency and severity of claims data. The study found that the best distribution for the frequency of claims data was negative binomial distribution and for the severity of claims data was gamma distribution.

(Agwa , Abdelhamid .2017) presented a study for actuarial model for engineering insurance claims using heavy-tailed probability distributions. The paper used three heavy-tailed probability distributions to model engineering insurance claims (Pareto distribution Log-normal distribution Generalized Pareto distribution).

Foreign studies have attempted to develop the probability distributions used, such as (Ahmad, Mahmoudi et al. 2022) the study propose a new family of distributions called the "Beta Power Transformed" (BPT) family, specifically designed to model insurance losses. Other study (Arif, Khan et al. 2021) propose a new family of heavy-tailed distributions explicitly designed to model insurance loss data. They call it the "New Exponentiated Heavy-Tailed" (NEHT) family. (Okhli, Nooghabi .2021) The authors introduce the contaminated exponential (CE) distribution as a more robust alternative for modelling positive-valued insurance claim data containing outliers. (Ahmad, Mahmoudi et al. 2020) the study introduces a new distribution called the Weighted T-X Weibull (WT-XW) distribution. This distribution is part of the T-X family and is designed to create more flexible models.

(Raschke, 2020) also explores several alternative approaches for modelling and inferring claim size distributions in insurance. These methods provide more flexibility and better results in different scenarios. One of these approaches is Finite Mixture Models which combine multiple simple distributions to create a more complex distribution, allowing greater flexibility in fitting different data patterns.(Punzo, Bagnato et al. 2018) also present a specific methodology for constructing compound unimodal distributions tailored explicitly for insurance loss modelling. (Leinwander , Aziz . 2018) The study proposes using two techniques to improve the modelling of insurance claims:

Skewed Distributions**:** These distributions allow for asymmetry, providing a better fit to many types of insurance loss data.

Mixture Distributions**:** These combine multiple underlying distributions to represent situations where claims come from different populations (e.g., small and large claims).

(Omari, Nyambura et al. 2018) This study proposes using various statistical distributions to model the frequency and severity of auto insurance claims separately. This tailored approach aims to provide a more accurate representation of the data.

We find that the studies that were applied to Hull insurance in the Egyptian Market dos not concern with developing the probability distributions used in claims data modelling and were limited to conducting goodness of fit tests for ready-made distributions, and this is what is addressed in this study.
Our problem in this research is to use a method to find mixed distribution to fit the data of marine hull insurance claims data.

2. The Method of Splicing

Actuaries have introduced several methods for making new probability distributions to fit claims data (Klugman, Panjer et al. ,2012),(Reynkens , Verbelen et al., 2017)**,** such as (Multiplication by a Constant, Raising to a Power, Exponentiation, Mixing, and Splicing). Each method is considered a solution to a specific problem in the data modelling process. The study found that the Splicing method is suitable for solving the problem of Marine hull insurance claims data, as a heavy tail characterizes this data. The Splicing method can be explained as follows:
The idea of Splicing Model Method is based on trying to divide claims data into two parts by taking a **cut point** (Which is tested more than once to reach the most suitable one) : the first includes small and medium claims, whose frequency is often very high, and

the second represents large claims, whose frequency is often small, as the following figure ٢ shows:

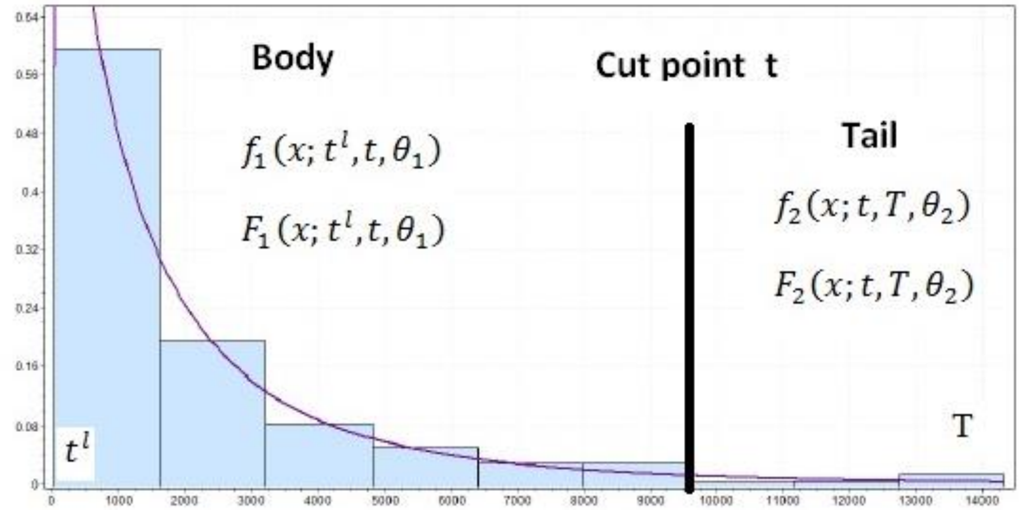

Figure 2: **An illustration of Splicing Method**

To reach the mixture probability distribution, the following steps are followed:

<u>First: finding a probability density function (pdf)</u>

1) To find a probability density function for the body $f_1(x; t^l, t, \theta_1)$ and tail $f_2(x; t, T, \theta_2)$.

Let f_1^*, f_2^* are PDF with corresponding CDF F_1^*, F_2^* on the parameter vectors $\theta_{1,}\, \theta_2$ define:

f_1: The probability density of the body on $[t^l, t]$. (Lower truncated at t^l and upper truncated at t)

$$f_1(x; t^l, t, \theta_1) = \begin{cases} \dfrac{\mathrm{f}_1^*(x, \theta_1)}{\mathrm{F}_1^*(t, \theta_1) - \mathrm{F}_1^*(t^l, \theta_1)} & , \quad t^l \le x \le t \\ 0 & , other\ wise \end{cases}$$

f_2 : The density of the tail on $[t, T]$. (Lower truncated at t and upper truncated at T)

$$f_2(x; t, T, \theta_2) = \left\{ \begin{array}{cc} \dfrac{\mathrm{f}_2^*(x, \theta_2)}{\mathrm{F}_2^*(T, \theta_2) - \mathrm{F}_2^*(t, \theta_2)} & , \quad t \leq x \leq T \\ 0 & , otherwise \end{array} \right\}$$

Where $0 \leq t^l \leq \mathrm{t} \leq \mathrm{T}$ are fixed points

2) Arriving to the probability density function by substituting into the following relationship:

Consider splicing weight $\pi \in [0,1]$, and then the splicing density is

$$f(x; t^l, t, T, \theta) = \left\{ \begin{array}{cc} 0 & if\ x \leq t^l \\ \pi\ f_1(x, t^l, t, \theta_1) & if t^l \leq x \leq t \\ (1 - \pi)\ f_2(x, t, T, \theta_2) & if t \leq x \leq T \\ 0 & if\ x \geq T \end{array} \right\}$$

Second: Finding a cumulative distribution function (cdf)

1) To find a cumulative density function for the body $F_1(x; t^l, t, \theta_1)$ and tail $F_2(x; t, T, \theta_2)$.

F_1: The CDF of the body on $[t^l, t]$

$$F_1(x; t^l, t, \theta_1) = \left\{ \begin{array}{cc} 0 & , \quad x \leq t^l \\ \dfrac{\mathrm{F}_1^*(x, \theta_1) - \mathrm{F}_1^*(t^l, \theta_1)}{\mathrm{F}_1^*(t, \theta_1) - \mathrm{F}_1^*(t^l, \theta_1)} & , \quad t^l \leq x \leq t \\ 1 & , \quad x \geq T \end{array} \right\}$$

F_2: The CDF of the tail on $[t, T]$

$$F_2(x; t, T, \theta_2) = \left\{ \begin{array}{cc} 0 & , \quad x \leq t \\ \dfrac{\mathrm{F}_2^*(x, \theta_2) - \mathrm{F}_2^*(t, \theta_2)}{\mathrm{F}_2^*(T, \theta_2) - \mathrm{F}_2^*(t, \theta_2)} & , \quad t^l \leq x \leq T \\ 1 & , \quad x \geq T \end{array} \right\}$$

2) Arriving to the cumulative density function by substituting into the following relationship:

Consider splicing weight $\pi \in [0,1]$, and then the splicing CDF is:

$$F(x; t^l, t, T, \theta) = \left\{ \begin{array}{ll} 0 & if\ x \le t^l \\ \pi \quad F_1(x, t^l, t, \theta_1) & if t^l \le x \le t \\ \pi + (1-\pi)\, F_2(x, t, T, \theta_2) & if t \le x \le T \\ 1 & if\ x \ge T \end{array} \right\}$$

3. Derivation of new Mixture Distribution for Hull insurance claims Data

In this study, we apply Lindley distribution (light-tailed) for the body, which is lower truncated with $t^l$and and upper truncated with t. Other extreme value distributions (heavy-tailed) for the tail, which is lower truncated with t and upper truncated with T, are also applied.

3.1 Splicing Lindely and Pareto distributions

3.1.1 Truncated Lindley distribution

Consider the following PDF and CDF from Lindley(Ghitany, Atieh et al. 2008):

$$f\ (x) = \frac{\theta^2}{1+\theta}(1+x)\ e^{-\theta x}$$

$$F\ (x) = 1 - \frac{1+\theta+\theta x}{1+\theta}\ \mathrm{e}^{-\theta x}$$

After truncation with limits $\mathbf{t^l}$ and $\mathbf{t}$, the PDF is given by:

$$let \quad f^t_{lindly}(x; t^l, t, \theta) = \left\{ \begin{array}{ll} \dfrac{f_{lindly}(x,\theta)}{F_{lindly}(t,\theta) - F_{lindly}(t^l,\theta)} \quad , & t^l \le x \le t \\ 0 & , other\ wise \end{array} \right\}$$

<u>Where</u> $\qquad f_{lindly}(x,\theta) = \frac{\theta^2}{1+\theta}(1+x)\ e^{-\theta x}$

$$F_{lindly}(t,\theta) = 1 - \frac{1+\theta+\theta t}{1+\theta}\,\mathrm{e}^{-\theta t} \quad , \quad F_{lindly}(t^l,\theta) = 1 - \frac{1+\theta+\theta t^l}{1+\theta}\,\mathrm{e}^{-\theta t^l}$$

$$\therefore F_{lindly}(t,\theta) - F_{lindly}(t^l,\theta)$$

$$= \frac{(1+\theta+\theta t^l)\mathrm{e}^{-\theta t^l} - (1+\theta+\theta t)\mathrm{e}^{-\theta t}}{1+\theta}$$

$$f_{lindly}^{t}(x;t^l,t,\theta)$$

$$= \frac{\theta^2}{1+\theta}(1+x)\,e^{-\theta x} \,.\, \frac{1+\theta}{(1+\theta+\theta t^l)\mathrm{e}^{-\theta t^l} - (1+\theta+\theta t)\mathrm{e}^{-\theta t}}$$

$$= \frac{\theta^2(1+x)\,e^{-\theta x}}{(1+\theta+\theta t^l)\mathrm{e}^{-\theta t^l} - (1+\theta+\theta t)\mathrm{e}^{-\theta t}}$$

$$\boxed{\therefore f_1(x;t^l,t,\underline{\theta}) = \frac{{\theta_j}^2(1+x)\,e^{-\theta_j x}}{(1+\theta_j+\theta_j t^l)\mathrm{e}^{-\theta_j t^l} - (1+\theta_j+\theta_j t)\mathrm{e}^{-\theta_j t}}}$$

<u>Also, define</u>

Let $$F_{lindly}^{t}(x;t^l,t,\theta) = \begin{Bmatrix} \frac{F_{lindly}(x,\theta) - F_{lindly}(t^l,\theta)}{F_{lindly}(t,\theta) - F_{lindly}(t^l,\theta)} & , \quad t^l \le x \le t \\ 0 & , other\ wise \end{Bmatrix}$$

<u>Where</u>

$$F_{lindly}(x,\theta) - F_{lindly}(t^l,\theta)$$

$$= \frac{(1+\theta+\theta t^l)\mathrm{e}^{-\theta t^l} - (1+\theta+\theta x)\mathrm{e}^{-\theta x}}{1+\theta}$$

$$F_{lindly}(t,\theta) - F_{lindly}(t^l,\theta) = \frac{(1+\theta+\theta t^l)\mathrm{e}^{-\theta t^l} - (1+\theta+\theta t)\mathrm{e}^{-\theta t}}{1+\theta}$$

<u>Therefore</u>

$$F_{lindly}^{t}(x;t^l,t,\theta) = \frac{(1+\theta+\theta t^l)e^{-\theta t^l} - (1+\theta+\theta x)e^{-\theta x}}{(1+\theta+\theta t^l)e^{-\theta t^l} - (1+\theta+\theta t)e^{-\theta t}}$$

$$\boxed{\therefore F_1(x;t^l,t,\underline{\theta})=\frac{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l}-(1+\theta_j+\theta_j x)\mathrm{e}^{-\theta_j x}}{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l}-(1+\theta_j+\theta_j t)\mathrm{e}^{-\theta_j t}}}$$

3.1.2 Truncated Pareto distribution

The truncated Pareto distribution can be obtained as follows **(**Reynkens, T. et al, 2017**)**

<u>PDF</u> $$\mathrm{f}_2(\mathrm{x};\mathrm{t},\mathrm{T},\gamma)=\frac{\frac{1}{\gamma t}(\frac{x}{t})^{-\frac{1}{\gamma}-1}}{1-(\frac{T}{t})^{-\frac{1}{\gamma}}}\quad t<x<T$$

<u>With CDF</u>

$$\mathrm{F}_2(x;t,T,\gamma)=\frac{1-\left(\frac{x}{t}\right)^{-\frac{1}{\gamma}}}{1-\left(\frac{T}{t}\right)^{-\frac{1}{\gamma}}}\quad t<x<T$$

3.1.3 Lindely - Pareto distributions

Splicing model with $\pi\in(0,1)$ the pdf and CDF is given by:

$$f(x;t^l,t,T,\theta)=\left\{\begin{array}{lll}\pi & \frac{{\theta_j}^2(1+x)\,e^{-\theta_j x}}{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l}-(1+\theta_j+\theta_j t)\mathrm{e}^{-\theta_j t}} & t^l<x<t\\ (1-\pi) & \frac{\frac{1}{\gamma t}(\frac{x}{t})^{-\frac{1}{\gamma}-1}}{1-(\frac{T}{t})^{-\frac{1}{\gamma}}} & t<x<T\end{array}\right\}$$

$$F(x;t^l,t,T,\theta) = \begin{cases} 0 & if\ x \le t^l \\ \pi \dfrac{\left(1+\theta_j+\theta_j t^l\right)e^{-\theta_j t^l} - (1+\theta_j+\theta_j x)e^{-\theta_j x}}{\left(1+\theta_j+\theta_j t^l\right)e^{-\theta_j t^l} - (1+\theta_j+\theta_j t)e^{-\theta_j t}} & if t^l \le x \le t \\ \pi + (1-\pi)\dfrac{1-\left(\frac{x}{t}\right)^{-\frac{1}{\gamma}}}{1-\left(\frac{T}{t}\right)^{-\frac{1}{\gamma}}} & if t \le x \le T \\ 1 & if\ x \ge T \end{cases}$$

3.2 Splicing Lindely and Lomax distributions

3.2.1Truncated Lomax Distribution

Consider the following PDF and CDF from Lomax distribution (Oguntunde, Khaleel , et al., 2017)

f_2^*(x) = $\alpha\beta(1+\beta x)^{-(\alpha+1)}$ (x,$\alpha,\beta) > 0$

F_2^*(x) = $1-(1+\beta x)^{-\alpha}$ (x,$\alpha,\beta) > 0$

$After$ Truncation with limits t, T the PDF and CDF become:

$$\therefore f_2(x;t,T,\theta_2) = \frac{f_2^*(x,\theta_2)}{F_2^*(T,\theta_2) - F_2^*(t,\theta_2)} \quad , \quad t \le x \le T$$

<u>Where</u>

$$f_2^*(x,\theta_2) = \alpha\beta(1+\beta x)^{-(\alpha+1)}$$

$$F_2^*(T,\theta_2) = 1-(1+\beta T)^{-\alpha}\ , \qquad F_2^*(t,\theta_2) = 1-(1+\beta t)^{-\alpha}$$

$$F_2^*(T,\theta_2) - F_2^*(t,\theta_2) = \left(1-(1+\beta T)^{-\alpha}\right) - \left(1-(1+\beta t)^{-\alpha}\right)$$

$$= (1+\beta t)^{-\alpha} - (1+\beta T)^{-\alpha}$$

$$\boxed{\therefore f_2(x;t,T,\theta_2) = \frac{\alpha\beta(1+\beta x)^{-(\alpha+1)}}{(1+\beta t)^{-\alpha} - (1+\beta T)^{-\alpha}} \quad , \quad t \le x \le T}$$

Also, define

$$F_2(x;t,T,\theta_2) = \frac{\mathrm{F}_2^*(X,\theta_2) - \mathrm{F}_2^*(t,\theta_2)}{\mathrm{F}_2^*(T,\theta_2) - \mathrm{F}_2^*(t,\theta_2)} \quad , \quad t \le x \le T$$

Where

$$\mathrm{F}_2^*(X,\theta_2) - \mathrm{F}_2^*(t,\theta_2) = (1-(1+\beta X)^{-\alpha}) - (1-(1+\beta t)^{-\alpha})$$
$$= (1+\beta t)^{-\alpha} - (1+\beta X)^{-\alpha}$$

$$\boxed{\therefore\ F_2(x;t,T,\theta_2) = \frac{(1+\beta t)^{-\alpha} - (1+\beta X)^{-\alpha}}{(1+\beta t)^{-\alpha} - (1+\beta T)^{-\alpha}} \quad , \quad t \le x \le T}$$

3.2.2 Lindley - Lomax distributions

Splicing model with $\pi \in (0,1)$ the pdf and CDF is given by:

$$f(x;t^l,t,T,\theta) = \left\{ \begin{array}{lll} \pi & \dfrac{\theta_j{}^2(1+x)\, e^{-\theta_j x}}{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l} - (1+\theta_j+\theta_j t)\mathrm{e}^{-\theta_j t}} & t^l < x < t \\ (1-\pi) & \dfrac{\alpha\beta(1+\beta x)^{-(\alpha+1)}}{(1+\beta t)^{-\alpha} - (1+\beta T)^{-\alpha}} & t < x < T \end{array} \right\}$$

$$F(x;t^l,t,T,\theta) = \left\{ \begin{array}{ll} 0 & if\ x \le t^l \\ \pi\, \dfrac{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l} - (1+\theta_j+\theta_j x)\mathrm{e}^{-\theta_j x}}{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l} - (1+\theta_j+\theta_j t)\mathrm{e}^{-\theta_j t}} & if t^l \le x \le t \\ \pi + (1-\pi)\, \dfrac{(1+\beta t)^{-\alpha} - (1+\beta X)^{-\alpha}}{(1+\beta t)^{-\alpha} - (1+\beta T)^{-\alpha}} & if t \le x \le T \\ 1 & if\ x \ge T \end{array} \right\}$$

3.3 Splicing Lindely and jumble (type 2) distributions:

3.3.1 Truncated jumble (type 2) distribution:(Abbas, Hussain et al. 2020)

Consider the following PDF and CDF from jumble (type 2) distribution

$$f_2^*(\mathrm{x}) = \alpha\beta\, x^{-(\alpha+1)} \exp(-\beta x^{-\alpha}) \quad , x,\ \alpha, \beta > 0$$

$$F_2^*(\mathrm{x}) = \exp(-\beta x^{-\alpha}) \quad , x,\ \alpha, \beta > 0$$

$After$ Truncation with limits t, T the PDF and CDF become:

$$\therefore f_2(x; t, T, \theta_2) = \frac{\mathrm{f}_2^*(x, \theta_2)}{\mathrm{F}_2^*(T, \theta_2) - \mathrm{F}_2^*(t, \theta_2)} \quad , \quad t \le x \le T$$

<u>Where</u>

$$\mathrm{f}_2^*(x, \theta_2) = \alpha\beta\, x^{-(\alpha+1)} \exp(-\beta x^{-\alpha})$$

$$\mathrm{F}_2^*(T, \theta_2) = \exp(-\beta T^{-\alpha})$$

$$\mathrm{F}_2^*(t, \theta_2) = \exp(-\beta t^{-\alpha})$$

$$\mathrm{F}_2^*(T, \theta_2) - \mathrm{F}_2^*(t, \theta_2) = \exp(-\beta T^{-\alpha}) - \exp(-\beta t^{-\alpha})$$

$$= e^{-\beta T^{-\alpha}} - e^{-\beta t^{-\alpha}}$$

$$\boxed{\therefore f_2(x; t, T, \theta_2) = \frac{\alpha\beta\, x^{-(\alpha+1)} \exp(-\beta x^{-\alpha})}{e^{-\beta T^{-\alpha}} - e^{-\beta t^{-\alpha}}} \quad , \quad t \le x \le T}$$

<u>Also, define</u>

$$F_2(x; t, T, \theta_2) = \frac{\mathrm{F}_2^*(X, \theta_2) - \mathrm{F}_2^*(t, \theta_2)}{\mathrm{F}_2^*(T, \theta_2) - \mathrm{F}_2^*(t, \theta_2)} \quad , \quad t \le x \le T$$

Where

$$\mathrm{F}_2^*(X, \theta_2) - \mathrm{F}_2^*(t, \theta_2) = \exp(-\beta x^{-\alpha}) - \exp(-\beta t^{-\alpha})$$

$$= e^{-\beta x^{-\alpha}} - e^{-\beta t^{-\alpha}}$$

$$\therefore\ F_2(x;t,T,\theta_2) = \frac{e^{-\beta x^{-\alpha}} - e^{-\beta t^{-\alpha}}}{e^{-\beta T^{-\alpha}} - e^{-\beta t^{-\alpha}}} \quad , \quad t \le x \le T$$

3.3.2 Lindley – jumble distributions

Splicing model with $\pi \in (0,1)$ the pdf and CDF is given by:

$$f(x;t^l,t,T,\theta) = \begin{cases} \pi \quad \dfrac{\theta_j{}^2(1+x)\, e^{-\theta_j x}}{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l} - (1+\theta_j+\theta_j t)\mathrm{e}^{-\theta_j t}} & t^l < x < t \\ (1-\pi) \quad \dfrac{\alpha\beta\, x^{-(\alpha+1)} \exp(-\beta x^{-\alpha})}{e^{-\beta T^{-\alpha}} - e^{-\beta t^{-\alpha}}} & t < x < T \end{cases}$$

$$F(x;t^l,t,T,\theta) = \begin{cases} 0 & if\ x \le t^l \\ \pi\ \dfrac{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l} - (1+\theta_j+\theta_j x)\mathrm{e}^{-\theta_j x}}{\left(1+\theta_j+\theta_j t^l\right)\mathrm{e}^{-\theta_j t^l} - (1+\theta_j+\theta_j t)\mathrm{e}^{-\theta_j t}} & if t^l \le x \le t \\ \pi + (1-\pi)\ \dfrac{e^{-\beta x^{-\alpha}} - e^{-\beta t^{-\alpha})}}{e^{-\beta T^{-\alpha}} - e^{-\beta t^{-\alpha}}} & if t \le x \le T \\ 1 & if\ x \ge T \end{cases}$$

4. Application

The study apply new three mixture distributions (Lindely Pareto, Lindely Lomax and Lindely jumble distributions) to fitting marine hull insurance claims data during the period (2014/2015-2020/2021) in one of the Egyptian general insurance market companies and compare the results with other probability distributions which previous studies recommend using in modelling claims data.

4.1 goodness-of-fit measures:

(Burnham, 2004),(Pinho, 2017), (Bakar, 2016),(Omari, 2018), (Ahmad, 2020) (Steenkamp, 2014)

In order to compare the fits of the distributions, we consider some measures of goodness-of-fit, including the maximized log-likelihood (-ℓ),

Anderson-Darling (A), Cramér-Von Mises (W), Kolmogorov Smirnov (KS) statistics (with its p-value), Akaike information criterion (AIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), and consistent Akaike information criterion (CAIC). In general, the smaller these statistics are, the better the fit is. We use R statistical package software to calculate all of the previous measures.
We shall compare the fits of the (Lindely_ Pareto, Lindely_ Lomax and Lindely_ jumble) models with other models: Lindely distribution, Lomax distribution, Pareto distribution, jumble distribution, Weibull distribution, and Burr distribution.

Table 1: Results of the R program for goodness of fit tests

Distribution	AIC	CAIC	BIC	HQIC	W	A	$-\ell$	KS	PV
Lindely_ Pareto (π, θ, γ)	**332.55**	**332.9666**	**338.98**	**335.08**	**0.148225 4**	**1.073855**	**163.2799**	**0.09620451**	**0.604337**
Lindely_ Lomax $(\pi, \theta, \alpha, \beta)$	**334.54**	**335.2357**	**343.11**	**337.91**	**0.148228 5**	**1.073874**	**163.273**	**0.09620417**	**0.604331**
Lindely_ jumble $(\pi, \theta, \alpha, \beta)$	**334.25**	**334.9462**	**342.82**	**337.62**	**0.148229 3**	**1.073879**	**163.1283**	**0.09620405**	**0.604333**
Lindely (θ)	814.82	814.8936	816.95	815.66	1.519644	7.897468	406.4135	0.8435338	0.000
Lomax (α, β)	339.32	339.5271	343.61	341.01	0.177982	1.046163	167.6636	0.1407736	0.164576
Pareto (α, β)	339.32	339.5271	343.61	341.01	0.177981 9	1.046163	167.6636	0.1407739	0.164574
jumble (α, β)	344.91	345.1188	349.20	346.60	0.343877 7	1.872642	170.4594	0.153177	0.104000
Weibull (α, β)	376.58	376.7876	380.87	378.27	0.750188 1	4.107314	186.2938	0.2530783	0.000625
Burr $(\alpha, \gamma, \lambda)$	341.28	341.688	347.71	343.80	0.178106 6	1.043555	167.6406	0.1384688	0.178451

It is clear from the previous table(1) that (Lindely Pareto) distribution is the best distribution because P.value is the highest and at the same time all other measures of goodness-of-fit are lowest compared to other probability distributions, which means that the claims data for marine hull insurance may be come from this distribution.

4.2 Parameter Estimation:

To estimate the parameters of the previous distributions, we use the maximum likelihood function method using the R statistical Package program. The results are as follows:

Table 2: Results of the R program for maximum likelihood Parameter Estimation

Distribution	Estimates			
Lindely_ Pareto (π, θ, γ)	0.8412695 0.04603767	2.7545100 0.87105179	0.8361045 0.12228985	
Lindely_ Lomax $(\pi, \theta, \alpha, \beta)$	0.8412683 0.04603793	0.8361070 0.12228996	0.3870372 0.24167395	0.8425606 7.7900596
Lindely_ jumble $(\pi, \theta, \alpha, \beta)$	0.8412681 0.04603799	1.9570473 3.75315746	0.8361078 0.12229000	0.4504020 0.1863843
Weibull (α, β)	0.8589591 0.0427751	1.1581486 0.0982922		
Lomax (α, β)	0.6686694 0.1356751	1.2669953 0.5028843		
Pareto (α, β)	0.6686692 0.1356749	0.7892700 0.3132692		
jumble (α, β)	0.5077085 0.04730831	0.7517543 0.10830587		
Burr $(\alpha, \gamma, \lambda)$	0.6356498 0.1971294	1.0369948 0.1754497	0.7328643 0.3923823	

4.3 Visual goodness of fit

The R program was used to perform a Visual goodness-of-fit by displaying propapilty denisty function (pdf) ,cumulative dinsty function (cdf) survival function and the relationshipe between expeceted and opserved claims . This was done for two distributions (Lindely_ Pareto) distribution as the best proposed distribution and Burr distribution as the best traditional distribution:

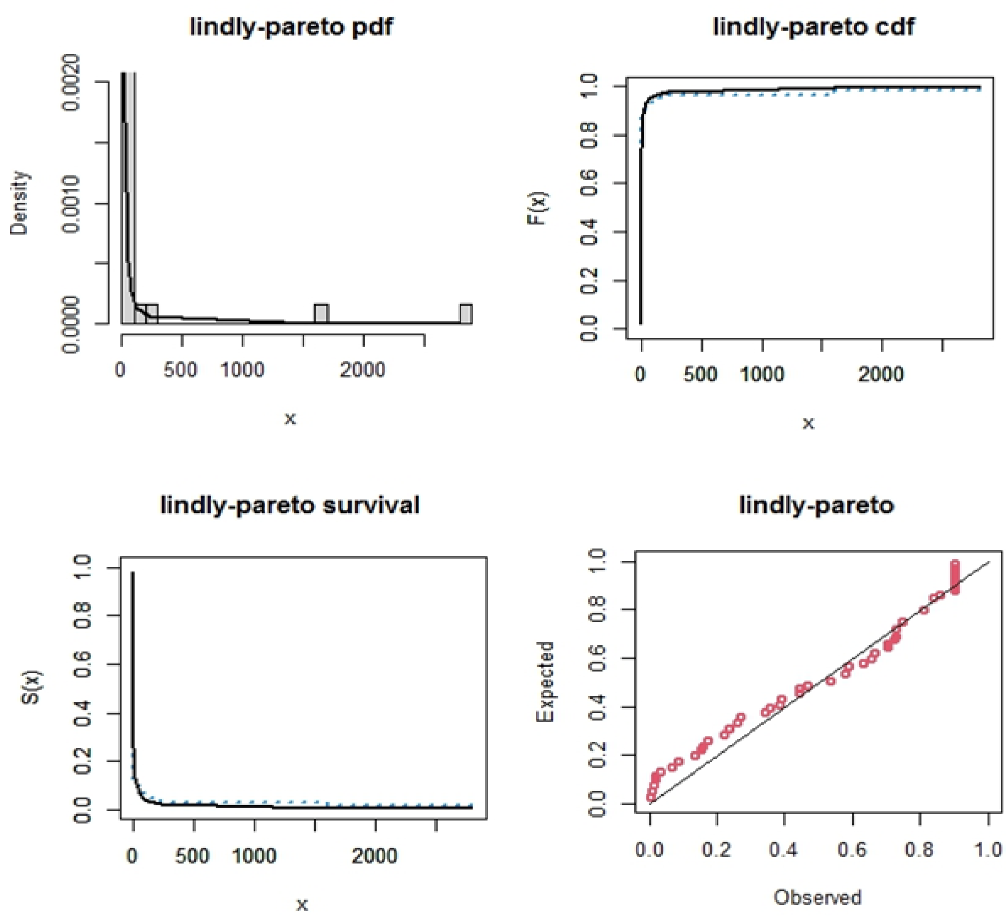

figure٣: The results of R program that attempts to Model claims data using (Lindely Pareto) distribution.

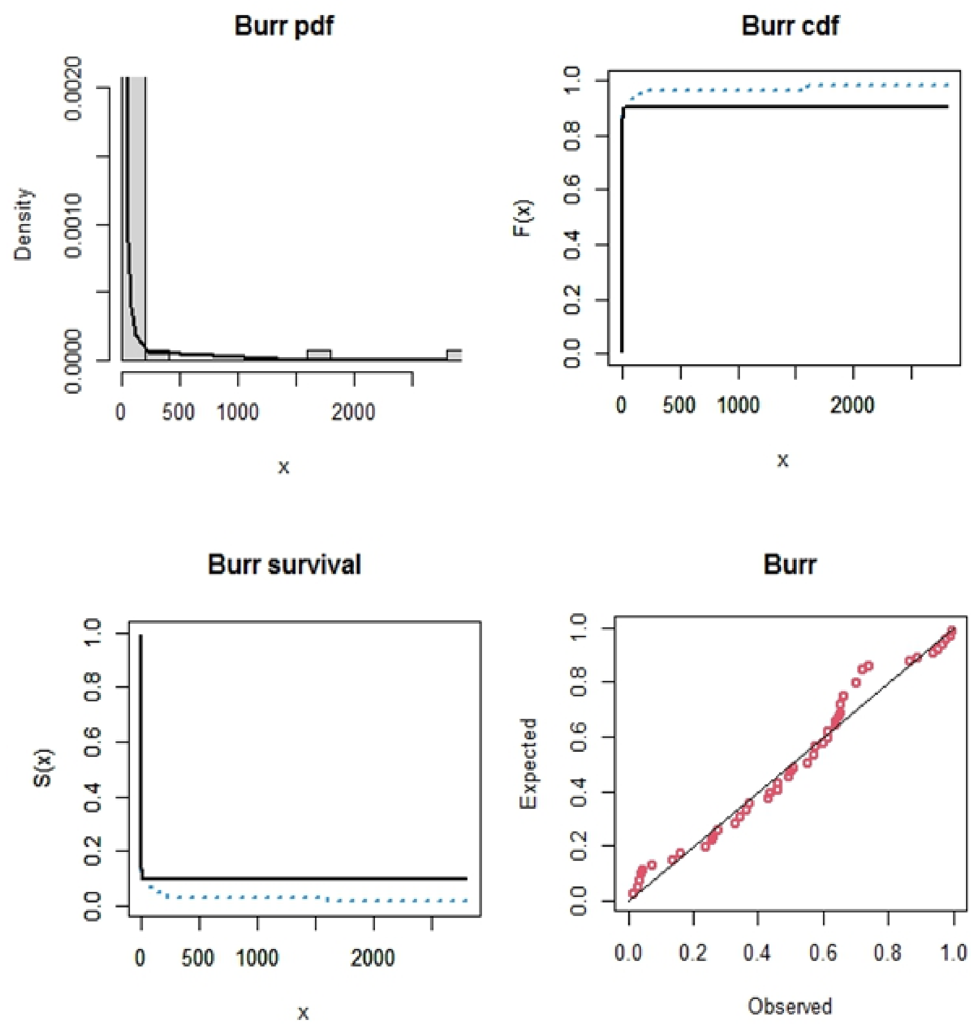

Figure 4: The results of R program that attempts to Model claims data using Burr distribution

It is clear from figures (3,4) that The (Lindely-Pareto) distribution demonstrates a better goodness-of-fit to marine hull claims data compared to the Burr distribution . This finding aligns with the results of the digital tests conducted .

Concluding Remarks

The study aimed to reach a probability distribution consistent with the nature of ship hull insurance claims data, which is skewed to the right and heavy-tailed.

The study concluded that one of the available solutions is to use Splicing Model method to create new Mixture probability distributions that fit the nature of the claims data.

The study proposed three new Micture probability distributions, the best of which was (Lindely Pareto) distribution .

The study recommended property insurance companies to use (Lindely Pareto) distribution while trying to model marine hull insurance data to perform Premium Rating, Reservation, Reinsurance Agreements, and Testing for Solvency.

References

Abbas, K., Hussain, Z., Rashid, N., Ali, A., Taj, M., Khan, S. A., & Khan, D. M. (2020). "Bayesian estimation of gumbel type-II distribution under type-II censoring with medical applicatioNs." Computational and Mathematical Methods in Medicine 2020:1-11.

Abdelhamid , N. A. (2019)"Pricing of Agricultural Crops Insurance Using Probability Distributions: A Case Study on Wheat Crop in Egypt." Journal of Contemporary Commercial Studies, Faculty of Commerce, Kafr El-Sheikh University: pp. 526-550.

Agwa , A. & Abdelhamid , N. A. (2017). "Actuarial Modeling of Engineering Insurance Claims Using Some Heavy-Tailed Probability Distributions." The Egyptian Journal of Commercial Studies, Faculty of Commerce, Mansoura University: pp. 77-103.

Ahmad, Z., Mahmoudi, E., & Alizadeh, M. (2022). "Modelling insurance losses using a new beta power transformed family of distributions." Communications in Statistics-Simulation and Computation **51**(8): 4470-4491.

Ahmad, Z., Mahmoudi, E., Dey, S., & Khosa, S. K. (2020). "Modeling Vehicle Insurance Loss Data Using a New Member of T-X Family of Distributions." Journal of Statistical Theory and Applications **19**(2): 133-147.

Arif, M., Khan, D. M., Khosa, S. K., Aamir, M., Aslam, A., Ahmad, Z., & Gao, W. (2021). "Modelling Insurance Losses with a New Family of Heavy-Tailed Distributions." Computers, Materials & Continua **66**(1).

Bakar, S. A., Hamzah, N. A., Maghsoudi, M., & Nadarajah, S. (2016). Modeling loss data using composite models. Insurance: Mathematics and Economics, 61, 146-154

Burnham, K. P. and D. R. Anderson (2004). "Multimodel inference: understanding AIC and BIC in model selection." Sociological methods & research 33(2): 261-304.

EL-bolkiny, M. T., Wasif, G. A. & El-barkawy , M. A. F (2018). "Pricing Railway Freight Transportation Risks Using Compound Probability Models." The Egyptian Journal of Commercial Studies, Faculty of Commerce, Mansoura University: pp. 174-201.

Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). "Lindley distribution and its application." Mathematics and computers in simulation 78(4): 493-506 .

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

Leinwander, A. J. and M. A. Aziz (2018). "Modeling insurance claims using flexible skewed and mixture probability distributions." Journal of Modern Applied Statistical Methods **17**.

Oguntunde, P. E., Khaleel, M. A., Ahmed, M. T., Adejumo, A. O., & Odetunmibi, O. A. (2017). "A new generalization of the Lomax distribution with increasing, decreasing, and constant failure rate." Modelling and Simulation in Engineering 2017.

Omari, C. O., Nyambura, S. G., & Mwangi, J. M. W. (2018). "Modeling the frequency and severity of auto insurance claims using statistical distributions."

Punzo, A., Bagnato, L., & Maruotti, A. (2018). "Compound unimodal distributions for insurance losses." Insurance: Mathematics and Economics **81**: 95-107.

Pinho, L. G. B. (2017). "Building new probability distributions: the composition method and a computer based method."

Raschke, M. (2020). "Alternative modelling and inference methods for claim size distributions." Annals of Actuarial Science **14**(1): 1-19.

Steenkamp, J. H. H. (2014). Statistical distributions in general insurance stochastic processes (Doctoral dissertation, University of Pretoria).