DOI: 10.15849/IJASCA.230320.05

Int. J. Advance Soft Compu. Appl, Vol. 15, No. 1, March 2023
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Modeling of Loss Severity of Motor Insurance Extreme Claims

V. Selvakumar[1*], Dipak Kumar Satpathi[2], P.T.V. Praveen Kumar[2], V.V. Haragopal[2]

[1] *Bhavan's Vivekananda College of Science, Humanities, and Commerce, Hyderabad, India*
Department of Mathematics, Birla Institute of Technology and Science, Pilani, Hyderabad, India.
[2]*Department of Mathematics, Birla Institute of Technology and Science, Pilani, Hyderabad, India*
* Corresponding author's Email: p20170106@hyderabad.bits-pilani.ac.in, drselva2022@gmail.com

Abstract

There are a wide variety of uses for risk modeling in the insurance sector. A change in strategies for risk control would be beneficial for insurance firms. The risk evaluation of insurance firms in conjunction with their solvency is a complicated and extensive issue, but its solution begins with the mathematical modeling of claims results. This article aims to provide different approaches to achieve an acceptable probability model, allowing for the practical interpretation of insurance risks, which can be used for risk management purposes. Today's computational methods and statistical software enable us to achieve this objective in a wide variety of realistic applications. The article employs procedures focused on the data collected from claim amounts of motor own damage claims, which emphasizes the best fit distribution out of the many distributions using the Kolmogorov –Smirnov one-sample test. Anderson Darling and Cramer-von Mises's test for the real data from insurance companies.

Keywords: ***Density plot, Extreme claims, Parameters, Skewed distributions, Q-Q Plot, Goodness of fit tests***

1 Introduction

Insurance is nothing more than the way specific individuals from a similar pool distribute to the same failure and share their financial losses. Any form of insurance scheme "Non-Life" falls under General Insurance. Insurance policies typically cover the risk of building reserves to satisfy potential claims and profit margins. When the claim number is determined in advance, the premiums will be measured using the life table for the life insurance (that gives the probability of survival). Whereas the size of the settlement is not calculated in advance of non-life insurance, a preliminary estimation of the potential cost of premiums is essential in the measurement of reserves. Compared to life insurance, the distinguishing characteristics of general insurance are: the amount of the claim is not decided in advance, and sometimes without restriction, policies are extended periodically, and under the same scheme, there will be more than one claim. The expense is measured by the frequency and scale of their statements. Statistical modeling suggests that the distribution of assertion amounts is the foundation of the risk mechanism. After the distribution has been established, the likelihood of a claim of a certain size may be estimated. The nature of the distribution of claim amounts and modeling of claims are relevant for deciding premiums and estimating reserves. This Research attempts to build a claim value model by adapting probability distributions from chosen distribution

Received 5 November 2022; Accepted 10 January 2023

families based on historical sample data. In order to do this, we used historical claims data between 1985 and 2020. The study's key contributions are as follows:

- Modeling the claim paid amounts by fitting appropriate probability distributions
- Test if the selected distribution provides a good fitting of the data or not.

2 Related Work

The most important criterion in the auto insurance market is assessing if insurers are deciding high/low claim amount for future reserve planning by modeling the claim amounts based on appropriate distribution. Usually, we find in the literature that the examined data follows either exponential distributions for well-fitted tails, Weibull, Gamma, Lognormal, and Pareto [1], without specifying which distribution for the warped data is the strongest of several distributions. The lognormal distribution will provide a suitable match for the loss data based on the literature on modeling the procedures [2]. Most of the literature suggests that the lognormal distribution is able to yield accurate findings that can be used to draw inferences that are useful for decision-making in the non-life insurance business while checking the goodness of fit, the Q-Q plots, and mathematically utilizing the Akaike Knowledge Requirements meaning. Compared to a class of distributions composed of lognormal, exponential, gamma, and Weibull distributions, the lognormal distribution has been found to be a strong match for the quantities claimed [3]. Based on another report [4], property claim providers offer indexes for damages originating from traumatic accidents in the United States. The lognormal distribution tends to provide a more robust fit than the Paretoian distribution. The generalizations and expansions of the Pareto distribution give fresh insights into tail behavior and risk evaluation of claims [5]. In a motor insurance claim data collection [6], fitted two heavy-tailed distributions, the Weibull and the Burr XII distributions, through simulation using the Pollaczek-Khinchin formula. A statistical test to compare the better fit of a mixed exponential distribution due to the unique structure of insurance claims, where people are often classified into groups [7]. Another study [8] suggested that a mixed copula method was used to construct the joint distribution for the number and size of claims in the first model, based on the conditional probability decomposition model, which includes the Tweedie and two-part generalized linear models when applied to US auto insurance dataset. Using widely-used Danish fire data, a study [9] verified the LogPH model as a viable alternative to GPD. In order to estimate parameters for the claim frequency and average claim severity distributions, a stochastic gradient boosting method is used [10]. Another result also shows that the threshold severity model outperforms the widely used generalized linear model based on the gamma distribution when simulated extreme claim sizes follow a lognormal or Burr Type II distribution [11]. Another Study [12] introduced a novel technique for calculating the extreme quantiles of a large class of heavy-tailed distributions by using High Posterior Density intervals to make Bayesian inferences on extreme quantiles. Among the continuous claim's severity distributions, a study [13] suggests that the lognormal distribution is a relatively effective distribution for modeling claims severity.

A study [14] explores the stability of maximum likelihood estimates derived from the Weibull count model by simulating claim frequency. According to another work [15], statistical modeling utilizing generalized multivariate Pareto (GP) distributions is the multivariate analog of the commonly used univariate peaks over threshold modeling in finance and engineering. The Kumaraswamy Marshal-olkin Log-Logistic model is suitable for fitting Iranian Auto Insurance Claim data, as determined by a study [16]. Another study [17] found that the skew-student distribution and skew-normal in property liability are competitive as contrasted to other parametric distributions dependent on insurance data. Another research recommends that the lognormal distribution best fits lower arguments, while the Pareto distribution gives the best fit in the case of a large claim [18]. Generalized linear models are used as a systematic modeling technique for evaluating claims processes in the presence of covariates. The negative binomial regression model predicts the occurrence of a claim based on the factors, i.e., driver age, gender, a form of car, vehicle age, and no claim discount, based on another analysis, a hierarchical model on three elements, i.e., frequency, policy type, and seriousness of claims. The multinomial logit model often predicts the form of insurance claim dependent on injuries to third parties, insured own harm, damage to third property, and type of car [19]. Generalized beta (2), the third model for the severity variable, predicts claim amounts [20]. Another study stated that claim size follows a Poisson distribution, and the claim amount will follow a lognormal distribution [21]. The mixed copula method makes it possible to rely on the number of claimants [22] and the average size of claims by utilizing the Gaussian copula applied to the automobile insurance scheme portfolio, suggesting its supremacy over the classical Poisson compound model [23]. The spatial results significantly boost the models for claim frequency and claim size [24] by incorporating spatial dependence trends in a Poisson model for claim frequency and gamma distribution for average claim size models in non-life insurance.

Another study identified that the GLM-Gamma distribution model's findings are more contentious, indicating that female drivers are likely to result in greater compensation for car damages [25]. Another analysis found that the Generalized Pareto Distribution provides the appropriate fit for the data collection, which contradicts the common procedure of any insurance data fitting into the Lognormal Life Data Distribution [26, 27]. In order to forecast model parameters, the maximum likelihood method is used to predict the unknown parameters for data on healthcare expenditure [28]. According to another study [29], increasing the number of components(k) results in Finite mixture lognormal fitting is best for Motor insurance claims. Also, the conditional mean square error of prediction (MSEP) for the one-year CDR uncertainty is the critical uncertainty perspective under Solvency II in the claim [30]. A three-parameter compound model includes the unimodal gamma, the lognormal, and the inverse Gaussian are applied to three well-known insurance loss datasets and compared to various standard distributions often employed in the actuarial field [31]. A class of logit-weighted reduced mixture of experts (LRMoE) models for multivariate frequency or severity distributions of claims is proposed to predict the claims [32]. Generalized linear models are fitted to the marginal frequency and conditional severity components of the Canadian Motor Insurance Data Set [33]. Based on another study [34] recommend using explanatory variables in a three-component parametric individual model using the Canadian Insurance data set. Another research study [35] contributes to further spreading the very effective statistical extreme value techniques in the engineering sciences.

Many of the above research studies shows that a skewed and heavy-tailed distribution best represents claim severity. Also, the lognormal distribution was used to describe claims severity across numerous continuous distributions. This analysis focuses on this path and determines the effective distribution of India's motor insurance data for the extreme value modeling of data for a claim and check whether the Pareto distribution fits well with extreme claims data by using Kolmogorov Smirnov (KS-test), Anderson

Darling (AD-test), Cramer-von Mises statistic, AIC, and BIC Criteria. These results may benefit vehicle insurers that can predict excessive claims.

3 Research Design and Methodology

3.1 Data used for Research

The secondary data collected from different India's insurance companies is collected from 1985 to 2020 for 36 years. We have studied the own damage claim amount variable for fitting an appropriate statistical distribution for 40940 data claims using R Programming. We now discuss the techniques of model building as follows.

3.2 Methodology

The methodological portion focuses on modeling the number of claims with various likelihood distributions from chosen families to specify the number of claims observed. A suitable behavioral recognition pattern may be diagnosed through careful data collection. Also, by systematically analyzing these motor insurance datasets, we can determine the distribution and its ideal parameter for the dataset.

Modeling Process:

- ***Identification of a model family of distributions*:**

The assumption emerging from a particular distribution family is identified based on frequency distribution, graphical methods, knowledge about nature, and process decisions.

- ***Parameter Selection*:**

Determine the parameters based on the claim amount data and distributions using the maximum likelihood and method of moments. Often, choose the parameters that decide a particular case of a distribution family.

- ***The validity of the distribution*:**

The chosen distributions and their parameters are a good fit for the data. They were tested using statistical measures such as the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Cramer-von Mises's test.

- ***Check the Model fit:***

Quantile-Quantile (Q-Q) plots are analyzed after validating the sample distribution to verify how well a theoretical distribution model suits. More effectively, claims data is usually skewed to the extreme right, so these claims can be modeled using a skewed distribution. Prior expertise and practice with curve fitting, computer tools, and exploratory data analysis were the deciding factors for choosing the appropriate distribution. This meant finding the descriptive analysis such as the average, the medians, the standard deviation, the range, the coefficient of variation, the skewness, and Kurtosis. The statistical software tool "R Programming" is used here to build the model.

The following are some of the probability density functions (pdf) of the different probability distributions:

Table 1: Probability density functions of Distributions

S.No	Probability Distributions	Probability Density function
1	Gamma Distribution	$f(x)=\frac{\lambda^{n}e^{-\lambda x}x^{\alpha-1}}{\Gamma n}, x>0$ with parameters: shape α > 0 and scale λ > 0
2	Lognormal Distribution	$f(x)=\frac{1}{\sigma x\sqrt{2\pi}}e^{-\frac{1}{2\sigma^{2}}(ln(x)-\mu)}, x>o$ with parameters: location μ and scale λ > 0
3	Weibull Distribution	$f(x)=\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, x>0$ with parameters: shape α > 0 and scale β > 0
4	Pareto Distribution	$f(x)=\left(\frac{\alpha\beta^{\alpha}}{x^{\alpha+1}}\right), x\geq\beta$ with parameters: shape α > 0 and scale β > 0
5	Loglogistic Distribution	$f(x)=\left(\frac{\beta}{\alpha}\right)\frac{\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)^{2}}, x>0$ with parameters α < 0 and β > 0
6	FatigueLife (Birnbaum-Saunders) Distribution (3P)	$f(x)=\frac{\sqrt{\frac{x-\mu}{\beta}}+\sqrt{\frac{\beta}{x-\mu}}}{2\gamma(x-\mu)}\Phi\left(\frac{\sqrt{\frac{x-\mu}{\beta}}-\sqrt{\frac{\beta}{x-\mu}}}{\gamma}\right), x>\mu$ where μ is the location parameter, β > 0 is the Scale parameter, and γ > 0 is the shape parameter

We have fitted above mentioned skewed distribution and also estimated the parameters using R Programming.

3.3 Estimation

There are several methods to estimate the unknown population parameter the Method of Maximum Likelihood, the Method of Moments, the Method of Minimum Chi-Square, and the Least Square Method. We have used the maximum likelihood estimator method to estimate the selected family distributions' unknown parameters among the four methods mentioned above.

3.4 Validation of the Model

The validity of various skewed distributions and their parameter derived is checked whether the fitted model is appropriate using various statistical tests. For all tests, the null hypotheses tested are

H_0: The claim amount of data follows a specified distribution.

H_1: The claim amount of data does not follow a specified distribution.

3.4.1 Kolmogorov-Smirnov test (K-S test):

Under this test, the decision about Ho is taken based on the empirical cumulative distributions of the claim amount data for the fitted cumulative distribution. If Fn(x) and F(x) are the empirical distributions, respectively, the K-S statistics is computed as

$$D_n = \max_{overall\ x} \left| F_n(x) - F_n \right| \quad (1)$$

3.4.2 Anderson Darling (AD) test:

The Anderson-Darling test is used to test how well the claim amount data fits a specified distribution. It also gives more weight to the skewed distribution's tails than the Kolmogorov-Smirnov one-sample test. Anderson Darling test is commonly used as a test for normality. The AD test statistics is given by

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(x_i) + \ln F(x_{n-i+1})] \quad (2)$$

3.4.3 Cramer-Von Mises's test:

Cramer-von Mises's test is an aggregate empirical distribution function test for normality. The test statistic is the total of the squared differences between the actual and predicted cumulative proportions.

$$T = \frac{1}{12n}\sum_{i=1}^{n}\left(\left(\frac{2i-1}{2n}\right) - F(x_i)\right) \quad (3)$$

3.4.4 Performance Criteria:

We shall calculate four performance metrics AIC, BIC and HQIC consider such analytical measures to determine the goodness of fit among the applied distributions.

Akaike Information Criteria:

$$AIC = -2L_i + (2 * p) \quad (4)$$

Bayesian Information Criteria:

$$BIC = -2L_i + (p * log(n)) \quad (5)$$

Where L_i denotes a log-likelihood function of the distribution, p – number of model parameters of the distribution, n – total number of claims

3.4.5 Diagnosis of Probability Models:

Quantile-Quantile plot is a useful tool for evaluating distributional fit. Suppose we need to compare two distributions visually. In that case, we use the quantile plot, which visually compares two distributions by graphing the quantiles of one versus the quantiles of the other distribution.

Probability modeling in non-life insurance is challenging, and an excellent fit to the tail probability is crucial. In this study, we have fitted five skewed distributions based on 40940 data claims from different Indian insurance companies. Also, we have considered the extreme claim amount, included other model's distribution families, and checked the model's validity to test whether the extreme claim amount follows a specified distribution.

4 Results and Discussions

The secondary data were analyzed using descriptive statistics, Kolmogorov Smirnov, Cramer Von Mises, and Anderson Darling tests. The descriptive data analyses of the claimants paid are summarized in Table 2 below.

Table.2 Descriptive Statistics of Own Damage Extreme Claim Amount

Sample Size	40940
Mean	143410
SD	170370
Minimum	600001
Maximum	5216000
Skewness	9.7984
Kurtosis	212.53

From table 2 above, we can observe the summary statistics of the own damage extreme claim amount. The average claim paid was Rs. 143410, the standard deviation was Rs. 170370, and the skewness coefficient is 9.7984, indicating that the data is asymmetric. It suggests that the claim amount was positively skewed. Kurtosis has a value of 212.53; this measure's coefficient value indicates that the curve is Leptokurtic. Furthermore, the minimum number of extreme claims paid to the policyholders is Rs. 500001, while the maximum amount of claims paid is Rs. 52,16,000. This means that the insurer compensated a policyholder the most extreme claim sum during the era. In general, from the analytics, the claims' details have fewer claim amounts that occurred consistently. Whereas significant claim amounts infrequently happen, implying that a smaller claim sum has a higher relative frequency and a more significant claim amount has a lower relative frequency.

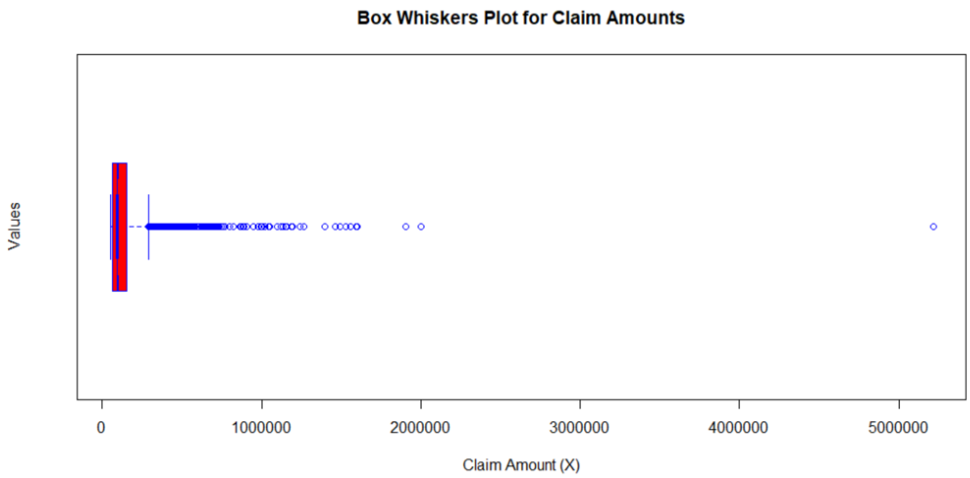


Figure. 1 Box Whisker Plot;

From Fig. 1, Box-Whisker's plot [36] and density curve confirm the skewed existence of the paid claims. The long or thick tail is often seen. Once we have observed the above details, we have placed together the distributions' results to see which distribution of probabilities is best for the data. These findings are in Table 4.

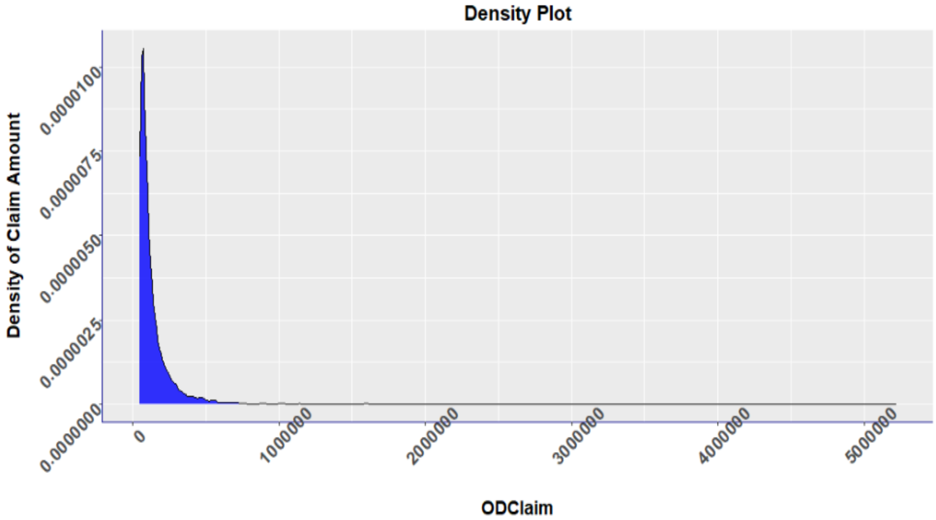


Figure. 2 Density trace of the overall Claim amount

In Fig. 2, the estimated probability density function of a claim paid data indicates the positively skewed distribution with a long and heavy tail to the right. Before analyzing the claims data, we must first identify a parameter of the specific statistical distribution that applies to those claims. The most probable probability for estimating the model parameters is determined by using the Maximum Likelihood Estimation and the Method of Moments [37-39]. Table 3 shows the parameters of five distributions that were fitted to the claims data. The estimated distribution parameters were used to calculate the log-likelihoods, AIC, and BIC values for each of the five distributions.

Table 3. Parameter Estimation of OD Claim

Distribution	**Estimated Parameters**	
Gamma	α = 0.7085	β = 202410
Weibull	α = 1.6981	β = 153600
Lognormal	μ = 11.604	σ = 0.64252
Pareto	α = 0.398	β = 57571
Fatigue Life	α = 0.69494	β = 116350
log logistic	α = 2.6723	β = 109390

Using the Akaike Information Criterion, Bayesian Information Criterion, log-like hood, and Goodness-of-fit tests, the selected distributions will be evaluated to see whether they correctly represent the data or not.

Table.4 Performance Criterion of OD Claim Amount

Distribution	Gamma	Weibull	Lognormal	Pareto	log logistic	Fatigue Life
Kolmogorov-Smirnov Statistics	0.428	0.168	0.111	0.019	0.11	0.137
Anderson Darling Statistics	710.4	283.8	102.5	3.4	100.6	152.4
Cramer-von Mises statistics	3015.4	N/A	328.11	17.6	312.8	414.8
log-likelihood	-52225.7	-52534.1	-51498.8	-50602.1	-51423.3	-51103.1
AIC	104455.4	105072.2	103001.6	101208.2	102850.6	102210.2
BIC	104468	105084.8	103014.2	101220.8	102863.2	102222.8

After evaluating the parameters and adapting different probability distributions, it is validated by the Kolmogorov Smirnov test, Anderson Darling test, Cramer-Von Mises test, AIC, and BIC values are listed, and the results are given in Table 3. Based on the findings, the distribution with minimum AIC, BIC, and log-likelihood values may be chosen for modeling the whole OD claim data. In this case, the Pareto distribution had shown minimum performance criteria values. As a result, the Pareto distribution was found to be the best-fitting distribution for the claims data, as it had the lowest AIC and BIC values. It is advisable to consider the 'goodness of fit' until concluding how well the lognormal distribution matches the claims results. In this sequence, its Q-Q plots and P-P plots were plotted and interpreted as follows

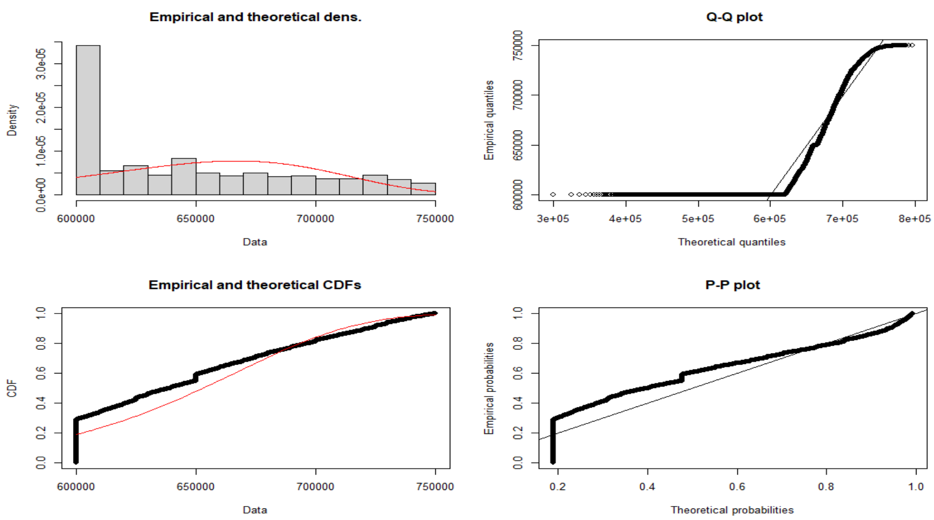


Figure 3: QQ Plot and PP Plot for Pareto Distributions

We have constructed the Q-Q plot and P-P Plot as part of the diagnostics for six skewed distributions. Fig. 3 demonstrates the Q-Q plot and P-P plot for claims paid amount for Pareto distributions. After evaluating the distribution's goodness of fit graphically using Q-Q plot, P-P plot, and Performance criteria values, the measures are taken in the actuarial modeling can produce accurate model fitting that can be used to draw valid inferences for decision-making in the general insurance industry.

The study's main goal was to create a suitable statistical distribution for Own Damage Insurance Extreme claims data and to evaluate how well this distribution matched the claim data used to estimate claim amounts. The analysis attempts to model the data's information to fit it into a certain distribution. As a consequence, an attempt was made to find an optimum statistical distribution that precisely matches the insurance claims data using statistical computation and visual representation in R.

From Table 4, we can observe that the Pareto distribution has the maximum log-likelihood function of -50602.1, implying that it has a greater likelihood of fitting the extreme OD claims data than the other statistical distributions as Gamma, Weibull, log-logistic, Fatigue Life, and gamma distribution. In terms of AIC and BIC, the Pareto distribution has the lowest value of 101208.2 and 101220.8, which implies that it fits the overall TPclaims data well.

As shown in Figure 3, the Q-Q plot and the PP Plot of the Pareto distribution provide an excellent fit to the findings of the extreme claims, with almost all points mapping on the line segment and just a few points falling near the line.

5 Conclusion

Generally, it is essential to use appropriate statistical distributions to extreme model claims in a growing number of motor insurance claims. These mathematical claim amount models play an important role in addressing several complex issues in the motor insurance industry. An accurate insurance payment system that considers all possible uncertainties, costs, losses, and profits is essential for non-life insurance businesses. Insurance claims contribute significantly to the financial outflow of the business. In order to be able to predict future claims, insurance companies must first calculate the expected frequency and severity of claims. Motor insurance Extreme claims are modeled by using both discrete and continuous probability distributions, and those are the distributions utilized in this study. Non-life extreme own damage insurance claims data is modeled to evaluate the probability distributions to discover the best suitable distribution.

Modeling data with a standard distribution is often challenging due to differences in data features such as skewness and extended tail. In comparison to other continuous claims severity distributions, the lognormal distribution was selected by the majority of insurance firms to represent the severity of claims. Using models to simulate extreme claims experience and liability estimates, we have identified that Pareto distribution fits well with extreme claims data. These findings may benefit all motor insurance companies that are effective in forecasting extreme claims behavior.

Also, estimating the extreme claim amount by the application of an appropriate distribution to the motor insurance data results may be advantageous to vehicle insurers. These distributions are useful for anticipating the behaviour of extreme assertions. This would enable companies to prepare for changes in their portfolios and achieve their financial goals, which would be advantageous for insurance companies. The suggested claims distributions would also be useful to insurance regulators, who would use them to evaluate the financial viability of different insurance firms and the needed reserve levels for severe claims.

References

[1] V Packova and D Brebera. (2015). Loss Distribution and Simulations in General Insurance and Reinsurance. *International Journal of Mathematics and Computers in Simulation*, 5, 159-167.

[2] Marlin P. (1984). Fitting the Lognormal Distribution to Loss Data Subject to Multiple Deductibles. *The Journal of Risk and Insurance*, 51(4), 687-701.

[3] Achieng, O.M. and NO, I. (2010). Actuarial Modeling for Insurance Claims Severity in Motor Comprehensive Policy using Industrial Statistical Distributions. *International Congress of Actuaries*, Vol. 712.

[4] Krzysztof Burnecki, Grzegorz Kukla and RafałWeron (2000). Property insurance loss distributions. *Physica A: Statistical Mechanics and its Applications*, 287(1), 269-278.

[5] Fackler, M. (2013). Reinventing Pareto: Fits for both small and large losses. *Risk Management Electronic Journal*, ASTIN Colloquium, 1-24.

[6] Dilip C. Nath and Jagriti Das. (2016). Modeling of Insurance Data through Two Heavy-Tailed Distributions: Computations of Some of Their Actuarial Quantities through Simulation from Their Equilibrium Distributions and the Use of Their Convolutions. *Journal of Mathematical Finance*, 6(3),378- 400.

[7] David Lee, Wai Keung Li and Tony Siu Tung Wong. (2012). Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. *Insurance: Mathematics and Economics*, 51(3), 538-550.

[8] Peng Shi, Xiaoping Feng and Anastasia Ivantsova. (2015). Dependent frequency–severity modeling of insurance claims. *Insurance: Mathematics and Economics*, 64 (1),417-428.

[9] Soohan Ahn, Joseph H.T. Kim and Vaidyanathan Ramaswami. (2012). A new class of models for heavy tailed distributions in finance and insurance risk. *Insurance: Mathematics and Economics*, 51(1), 43-52.

[10] Su X, Bai M. (2020). Stochastic gradient boosting frequency-severity model of insurance claims. *PLoS ONE*, 15(8).

[11] Christian Laudage, Sascha Desmettre and Jorg Wenzel. (2019). Severity modeling of extreme insurance claims for tariffication. *Insurance: Mathematics and Economics,* 88(1), 77-92.

12] Farias, Rafael & Montoril, Michel & Andrade, J. (2016). Bayesian inference for extreme quantiles of heavy-tailed distributions. *Statistics & Probability Letters*, 113(1), 103-107.

[13] Omari C, Nyambura S and Mwangi J. (2018). Modeling the frequency and severity of Auto Insurance Claims using Statistical Distributions. *Journal of Mathematical Finance*, 8: 137-160.

[14] Kwadwo Asamoah. (2016). On the credibility of insurance claim frequency: Generalized count models and parametric estimators. *Insurance: Mathematics and Economics*, 70, 339-353.

[15] Anna Kiriliouk, Holger Rootzén, Johan Segers & Jennifer L. Wadsworth. (2019). Peaks Over Thresholds Modeling With Multivariate Generalized Pareto Distributions. *Technometrics*, 61(1), 123-135.

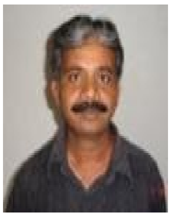
[16] Ramin Kazemi, Abdollah Jalilian and Akram Kohansal. (2017). Fitting Skew Distributions to Iranian Auto Insurance Claim Data. *Applications and Applied Mathematics: An International Journal*, 12(2), 790-802.

[17] Eling, M. (2012). Fitting insurance claims to skewed distributions: are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics,* 51(2), 239-248.
[18] Batsirai Winmore Mazviona and Tafadzwa Chiduza. (2013). The use of Statistical Distributions to Model Claims in Motor Insurance. *International Journal of Business, Economics and Law*, 3(1), 44-57.
[19] Renshaw, A. (1994). Modeling the Claims Process in the Presence of Covariates. *ASTIN Bulletin*, 24(2), 265-285.
[20] Edward.W Frees and Emiliano.A Valdez. (2015). Hierarchical Insurance Claims Modeling. *Journal of the American Statistical Association*, 103, 1457-1469.
[21] Daiane Aparecida Zuanetti, Carlos A.R. Diniz and Jose Galvoa Leite.(2006). A lognormal model for Insurance Claims Data. *RevStat- Statistical Journal*, 4(2), 131-142.
[22] Peter Xue Ken Song. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, 2(1), 305-320.
[23] Claudia Czado, Rainer Kastenmeier, Eike Christian Breachmann, and Aleksey Min. (2011). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, 278-305.
[24] Susanne G and Claudia Czado, (2007). Spatial modeling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 2007(3), 202-225.
[25] Evgenii V Gilenko and Elena A Mironova. (2017). Modern claim frequency and claim severity models: An application to the Russian motor own damage insurance market. *Cogent Economics & Finance*, 5(1), 1311097.
[26] V V Haragopal and K Navatha. (2017). Fitting Distributions to Big Data: Example of Large Claim Insurance Data. *Visleshana, Big Data Analytics*, 1(1).
[27] Zisheng, O., Chi, X. (2006). Generalized pareto distribution fit to medical insurance claims data. *Appl. Math. Chin. Univ. 21*, 21–29.
[28] Antonio Punzo, Angelo Mazza & Antonello Maruotti . (2018). Fitting insurance and economic data with outliers: a flexible approach based on finite mixtures of contaminated gamma distributions. *Journal of Applied Statistics*, 45(14), 2563-2584.
[29] Sattayatham, P. & Talangtam, T. (2012). Fitting of Finite Mixture Distributions to Motor Insurance Claims. *Journal of Mathematics and Statistics*, 8(1), 49-56.
[30] Happ, Sebastian & Merz, Michael & Wüthrich, Mario V. (2012). Claims development result in the paid-incurred chain reserving method. *Insurance: Mathematics and Economics*, 51(1), 66-72.
[31] Antonio Punzo, Luca Bagnato, Antonello Maruotti. (2018). Compound unimodal distributions for insurance losses. *Insurance: Mathematics and Economics*, 81, 95-107.
[32] Chai Fung, T., Badescu, A., & Sheldon Lin, X. (2019). A Class Of Mixture Of Experts Models For General Insurance: Application To Correlated Claim Frequencies. *ASTIN Bulletin*, 49(3), 647-688.
[33] J. Garrido, C. Genest, J. Schulz. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70(1), 205-215.
[34] Juan Sebastian Yanez, Mathieu Pigeon. (2021). Micro-level parametric duration-frequency-severity modeling for outstanding claim payments. *Insurance: Mathematics and Economics*, 98, 106-119.

[35] Gissibl N., Klüppelberg C., Mager J. (2017). Big Data: Progress in Automating Extreme Risk Analysis. In: Pietsch W., Wernecke J., Ott M. (eds) Berechenbarkeit der Welt?. Springer VS, Wiesbaden.
[36] Masoud, M., Jaradat, Y., Rababa, E. and Manasrah A. (2021). Turnover Prediction using Machine Learning: Empirical Study. *International Journal of Advances in Soft Computing & Its Applications*, 13(1), 193-207.
[37] Wang S, Gui W. (2020). Corrected Maximum Likelihood Estimations of the Lognormal Distribution Parameters. *Symmetry*, 12(6), 968.
[38] Afolabi Y. O. and Aliu A. H. (2022). Comparing Methods of Estimation for Two-Parameter Gamma Distribution Using Rainfall Patterns in Nigeria, *International Journal of Mathematics and Statistics Studies*, 10(4), 65-77.
[39] I-Shiang Tzeng & Li-Shya Chen. (2019). Testing the mean of skewed distributions applying the maximum likelihood estimator, *Cogent Mathematics & Statistics*, 6(1), 1588191.

Notes on Contributors

Dr. V SelvaKumar, Assistant Professor in the Department of Mathematics and Statistics, Bhavan's Vivekananda College of Science, Humanities & Commerce. He has published 21 papers in different national and international journals and one book to his credit. Also, presented seven papers at national and international conferences. His areas of interest are Data analytics, Time Series Analysis, and Machine Learning.



Prof. Dipak K Satpathi, is an Associate Professor, Department of Mathematics, Birla Institute of Technology & Science, Pilani, Hyderabad Campus. He authored a book "Mathematical Modeling: Application, Issues, and Analysis" (Published by Ane Books India, New Delhi, India) in the year 2007. He has published 20 papers and presented 16 papers in various national and international conferences. He guided four Students to Ph.D. His areas of interest are Mathematical Modeling of Biological Systems, Fluid Mechanics, and Financial Mathematics.

.

Prof. P.T.V Praveen Kumar, is an Associate Professor in the Department of Mathematics, Birla Institute of Technology & Science, Pilani, Hyderabad Campus. He has published 15 papers in the area of Data Analytics and Biostatistics. He guided one Student for Ph.D . Also, he has presented papers at both international and national conferences. His areas of interest are Multivariate Data Analysis and Biostatistics.

Prof. V V Hara Gopal, Visiting Professor, Dept. of Mathematics, Birla Institute of Technology & Science, Hyderabad Campus. He was a Former Director of the Centre for Quantitative Methods at Osmania University, Hyderabad. He has supervised twenty-Five Ph.D. dissertations and seven MPhil dissertations in the areas of Statistics. He published around 110 research papers in national and International Journals. He has successfully completed many government projects in different states in India