Truncated Gamma-Truncated Lomax Distribution in Modelling Data Claims

Irmatul Hasanah, Wahri Irawan and Ikin Ainul Yakin

***Abstract*—One of the methods to analyze the risk of loss on insurance companies based on the historical data on claim payments is modelling the data into severity distribution. This research deals with severity distribution by connecting two distribution Truncated Gamma and Truncated Lomax in modelling claim payments. The Kolmogorov-Smirnov test is used to test the fit of model. The result shows that Truncated Gamma-Truncated Lomax distribution is the best model to analyze the risk of loss based on data claim payments. The AIC value of 1533,915 and the BIC value of 1550,132.**

Keywords: Severity Distribution, Heavy Tail Distribution, Truncated Distribution, Modelling Claims.

I. INTRODUCTION

AN insurance companies as insurer is responsible for the risk of losses that occur to the customer as policyholder based on the insurance policy. As the loss occurs, the claim will be filled by customers from the insurer. The amount of losses experienced by customers is not known for certain. This becomes uncertainty regarding the risks experienced by an insurance company. Therefore the bigger the claim, the more losses the company will experience. So it is necessary to have a loss analysis based on historical data on claim payments. One way that can be used to analyze risk is by carrying out an analysis using the distribution function approach from historical data on claim payments [1].

The severity distribution is a distribution that is commonly used to model the magnitude of claim data. There are various types of severity distribution. There is a severity distribution that has two parameters, three parameters, or more than three parameters [2]. The Lomax distribution is a severity distribution that has two parameters with thick tail characteristics [3].

Several previous studies used the Gamma distribution in modeling the size of claims data because this distribution has right-skewed characteristics [4]. In modeling claim data, it can be done in various ways, one of which is the splicing distribution [5]. Risna, et al. 2021 uses the Gamma distribution connection and the Weibull distribution in modeling the magnitude of the Phoenix City claim data using a threshold value obtained from the largest likelihood value [6]. Furthermore, this study will discuss other distribution connections, namely the Gamma distribution connection and the Lomax distribution and will provide comparisons with other distributions.

I. Hasanah is with the Department of Islamic Banking, Faculty of Islamic Economic and Business, Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Jl. Syech Al-Bantani Serang-Banten, Indonesia, e-mail*: lrmatul.hasanah@uinbanten.ac.id

W. Irawan, I. A. Yakin are with the Department of Islamic Insurance, Faculty of Islamic Economic and Business, Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Jl. Syech Al-Bantani Serang-Banten, Indonesia

Manuscript received July 18, 2023; accepted September 14, 2023.

II. RESEARCH METHOD

This research use to several stages to find the best model to analyze the risk of loss. The following are the stages:

1) Descriptive Analysis
   Descriptive analysis in this study aims to look at the distributive characteristics of a data in order to obtain the right type of distribution for the data used in the study. Some of the characteristics seen are the shape of the distribution and the type of tail of the distribution. A distribution is said to be right-skewed if the skewness value is greater than zero. Meanwhile, a distribution is said to be thick tailed if the excess mean value increases linearly.
2) Choose The Type of Distribution
   Choosing a distribution that matches the characteristics of the data size needs to be done as a basis for modeling the connection function.
3) Modelling the Splicing Function
   After selecting the type of distribution according to the characteristics of the data, then modeling the Truncated Gamma-Truncated Lomax distribution function with a known threshold value. The threshold value used in this study is the threshold value that has been found by Risna, u = 0.969869.
4) Determine the estimated value of the parameters using MLE.
5) Testing Model Fit
   Model fit test was performed on the distribution of truncated joints using the Kolmogorov-Smirnov test.

III. RESULTS AND DISCUSSION

In this research, the data used in the illustration of the Truncated Gamma-Truncated Lomax distribution model is Phoenix City claim data for the period July 2019-June 2020 obtained from https://phoenixopendata.com/dataset/risk-management-claims. Then standardize the data by dividing the data by 1000. The following is a statistical description of the data presented in the following Table I:

TABLE I: Data Descriptive Statistics Klaim *Phoenix City*

N	Mean	Sd	Median	Min	Max	Range	Skew	Kurtosis
426	2.84	8.12	1.07	0.03	143.81	143.78	13.01	213.34

Based on Table I. The skewness value of the Phoenix City claim data is greater than zero, which is 13.01, which means that the data is skewed to the right. So that the distribution that has a right skewed characteristic is suitable for modeling the phoenix city claim data. One distribution that is suitable for modeling data that has right-swinging characteristics is the gamma distribution. The mean and median values of the Phoenix City data have different values, this indicates that the data is not symmetrical. The following is a histogram plot and mean excess from the phoenix city claim data.

Based on Fig. 1 (a) it can be seen that the mean excess value increases linearly, thus indicating that the phoenix city claim data has the characteristic of a thick tail. Some distribution models have thick tail characteristics with two parameter values such as the Weibull, Lomax and Burr (B-XII) distribution. With the description of the data and based on the mean excess plot above from the phoenix city claim data, this study will discuss the Truncated Gamma-Truncated Lomax (TG-TL) distribution model to model the phoenix city claim data. Next, the probability density function of the TG-TL distribution model is given

$$f(x)=\begin{cases} 0 & ,x<0 \\ v\left(\dfrac{(\frac{x}{\theta})^{\alpha-1}e^{-\frac{x}{\theta}}}{\theta[\tau(\alpha)-\tau(\alpha,\frac{u}{\theta})]}\right) & ,0\le x\le u \\ (1-v)\gamma\dfrac{(\beta+\mu)^{\gamma}}{(\beta+x)^{\gamma+1}} & ,x>u \end{cases} \quad (1)$$

The cumulative distribution function of the Truncated Gamma and Truncated Lomax joint distributions is as follows

$$F(x)=\begin{cases} 0 & ,x<0 \\ v\left(\dfrac{[\tau(\alpha)-\tau(\alpha,\frac{x}{\theta})]}{\theta[\tau(\alpha)-\tau(\alpha,\frac{u}{\theta})]}\right) & ,0\le x\le u \\ 1-(1-v)\left(\dfrac{(\beta+\mu)}{(\beta+x)}\right)^{\gamma} & ,x>u \end{cases} \quad (2)$$

Where v probability x is less than the threshold $\mu=0.969869$. Next, the parameter values of the distribution of truncated gamma and truncated lomax joints will be estimated using the maximum likelihood estimator method. The following is the likelihood function of the TG-TL distribution function

$$L(\alpha,\theta,\gamma,\beta;x)=f(x_1,x_2,...,x_n;\alpha,\theta,\gamma,\beta) \quad (3)$$

So that

$$\begin{aligned} \ln L(\alpha,\theta,\gamma,\beta;x) &= \sum_{i=0}^{n} f(x_i;\alpha,\theta,\gamma,\beta) \\ &= \sum_{i=0}^{n}\left\{ z_{i1}\ln\left(v\left(\frac{(\frac{x_i}{\theta})^{a-1}e^{-\frac{x_i}{\theta}}}{\theta[\tau(\alpha)-\tau(\alpha,\frac{u}{\theta})]}\right)\right)\right. \\ &\left. +z_{i2}\ln\left((1-v)\gamma\frac{(\beta+\mu)^{\gamma}}{(\beta+x_i)^{\gamma+1}}\right)\right\} \end{aligned} \quad (4)$$

Where n is the amount of data with z_{i1} having a value of 1 if x_i lies below the u threshold value and a value of 0 if x_i lies above the u threshold value and z_{i2} has a value of 1 if x_i lies above the u threshold value and has a value of 0 if x_i lies below the threshold value u. The parameter values of the TG-TL distribution are obtained by performing a partial differentiation of each parameter whose estimated value you want to know and then equating it to zero. The numerical method is used in calculating the estimated parameter values to obtain the estimated parameter values from the TG-TL distribution. The following parameter estimates are presented in Table II. below,

TABLE II: TG-TL Distribution Parameter Estimation

Distribution	Parameter	Parameter Estimation
TG-TL Distribution	$\hat{\alpha}$	17.037.846
	$\hat{\theta}$	0.3337171
	$\hat{\gamma}$	16.712.036
	$\hat{\beta}$	18.453.996

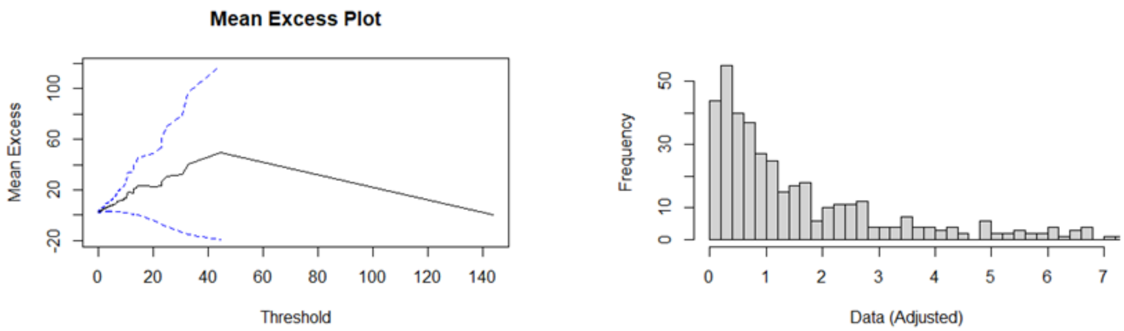
Next, a model suitability test will be carried out using the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test is a Goodness of fit test of a random variable model to see whether a phoenix city claim data distribution is included in a data distribution in the TG-TL distribution model. The following are the steps for testing the fit of the TG-TL distribution model,

1) Hypothesis
   H_0 : Phoenix city claim data is suitable to be modeled with the TG-TL distribution
   H_0 : Phoenix city claims data does not suitable to be modeled with the TG-TL distribution
2) Significance Level $\alpha=0.05$
3) Test Statistic
   The following test statistics are used

$$T=maxt\le x\le u|F_n(x)-F(x)|$$

   The results of the Kolmogorov Smirnov test for hoenix city claim data were obtained $T=0,028169$
4) Decision Rule
   For $\alpha=0.05$, the decision will de reject H_0 if $T>$ the critical value $\frac{1.36}{\sqrt{n}}=\frac{1.36}{\sqrt{426}}=0.065795$
5) Conclusion
   From the test above, the value of the test statistic $T=0.028169$ is obtained which is smaller than the critical value of 0.065795 at a significance level of $\alpha=0.05$. So it can be concluded that the phoenix city claim data is suitable for modeling with the TG-TL distribution with a threshold value of $u=0.96986$.

The following is a comparison of the estimated parameter values with other distributions such as the Lomax distribution, Burr B-XII, Truncated Gamma and Truncated Weibull, Truncated Gamma-Truncated Burr B-XII.



(a) (b)

Fig. 1: (a) Mean Excess Plot Data Phoenix city; (b) Histogram Data Phoenix City

TABLE III: Parameter Estimation of Other Distribution Models on Phoenix City Claims Data

Distribution	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$
TG-TW	1.7043642	0.3335255	0.320722	0.0971767
TG-TB-XII	1.7043448	0.3335292	1.053261	1.2503972
Lomax			2.793317	2.013779
Burr			1.3516771	0.9299798

From Table II. and Table III. The AIC, BIC and T values obtained from the Kolmogorov-Smirnov test results for each model are modeled with the phoenix city claim data presented in Table IV. below

TABLE IV: Analytical Measures Model for Data Claim Phoenix City

Distribution	AIC	BIC	T
TG-TW	1535.014	1551.232	0.023474
TG-TL	1533.915	1550.132	0.028169
TG-TBXII	1535.424	1551.641	0.028169
Lomax	1557.794	1565.903	0.046948
Burr	1546.527	1554.636	0.035211

Based on Table IV. The smallest AIC and BIC values are obtained for the distribution of Truncated Gamma and Truncated Lomax connections, namely 1533,915 and 1550,132. Therefore, a distribution that has small AIC and BIC values indicates that this distribution model is the best distribution for modeling Phoenix City claim data.

Based on the results of the model fit test using the Kolmogorov-Smirnov test, the T value for each model is less than the critical value of 0,065795 so that each model in the simulation in this study is suitable for modeling the phoenix city claim data. The following presents a histogram of phoenix city claim data and a graph of the estimated distribution function in Table III. based on the parameter estimates

Based on Fig. 2. It can be seen that the graph of the distribution function is suitable in modeling the phoenix city claim data. From the Kolmogorov Smirnov test presented in Table IV. where the statistical value is less than the critical value so that each distribution model in this study is suitable in modeling Phoenix city data.

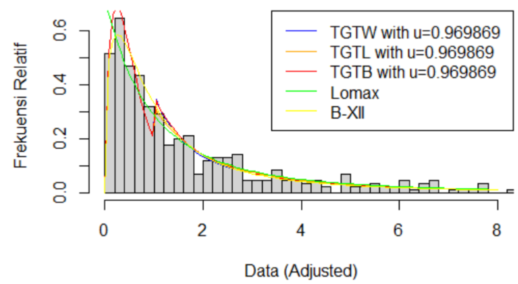


Fig. 2: Histogram graph and Estimated Distribution Function graph

IV. CONCLUSIONS

Based on the simulation results given, the distribution model by connecting the Truncated Gamma and Truncated Lomax distributions with a threshold value of $u = 0.96986$ is the best distribution model. Ideally the connection distribution can be used with a threshold value determined at the highest likelihood value. The selection of the distribution model in modeling the data needs to pay attention to the characteristics possessed by each distribution. In this study the connection distribution is one way to obtain the best distribution in modeling data.

REFERENCES

[1] G. E. Rejda and M. J. McNamara, *Principles of Risk Management and Insurance*. Pearson Education Inc., 2018.

[2] S. Klugman A., H. Panjer H., and G. Willmot E., *Loss Models: Nonlife Actuarial Models*, Fourth Edi. Canada: Wiley, 2012.

[3] J. Zhao, Z. Ahmad, E. Mahmoudi, E. H. Hafez, and M. M. M. El-Din, “A new class of heavytailed distributions: Modeling and simulating actuarial measures,” *Complexity*, vol. 2021, no. iii, 2021, doi: 10.1155/2021/5580228.

[4] J. Garrido, C. Genest, and J. Schulz, “Generalized linear models for dependent frequency and severity of insurance claims,” *Insur. Math. Econ.*, vol. 70, pp. 205–215, 2016, doi: 10.1016/j.insmatheco.2016.06.006.

[5] C. Laudagé, S. Desmettre, and J. Wenzel, “Severity modeling of extreme insurance claims for tariffication,” *Insur. Math. Econ.*, vol. 88, pp. 77–92, 2019, doi: 10.1016/j.insmatheco.2019.06.002.

[6] R. Diandarma, D. Lestari, S. Mardiyati, R. A. Kafi, S. Devila, and L. Safitri, “Truncated gamma-truncated Weibull distribution for modeling claim severity,” *AIP Conf. Proc.*, vol. 2374, no. Iscpms, pp. 1–9, 2021, doi: 10.1063/5.0059259.