UNIVERSITY OF NAIROBI

Loss Distributions For Motor Insurance Claim Severity

Case Study: Kenya

Group Members:

	Name	Registration Number
1.	Lillian Ayoo	I07/0817/2018
2.	Nelvine Anyango	I07/0811/2018
3.	Joy Kanyi	I07/132677/2018
4.	Kennedy Mwavu	I07/0807/2018
5.	Rachael Kanini	I07/0878/2018

Data Analysis

Introduction

The motor insurance claim severity data used was obtained from the annual reports of Insurance Regulatory Authority. The data was from 2016 – 2020 and contained 37 insurance companies, all licensed and regulated by `IRA`. Incurred claims for both motor commercial and motor private classes were analyzed side by side to obtain suitable models for each category. All the analysis was done using the R Programming Language.

Descriptive Statistics

Getting a summarised overview of the data was critical in discovering patterns, especially on skewness.

Table 1: Descriptive statistics for incurred claims (2016-2020)

Stat	Motor Commercial	Motor Private
No. Of Observations	185	185
Mean	296940.7	408137.7
Standard Error	23745.16	30003.67
Median	173695	271419
Standard Deviation	322969.1	408094.0
Kurtosis	5.937997	5.777819
Skewness	1.741602	1.683955
Minimum	0	-24861
Maximum	1470770	1992246
Sum	54934035	75505472

It is clearly evident that incurred claims data for both classes of motor insurance are positively skewed, with motor private having a higher mean than motor commercial.

The next step was to get a visual representation of the distribution of the data.

From the histograms we can affirm that the data from both classes is not only skewed to the right, but also long-tailed. This gives a good hint of the kind of distributions most appropriate to model the data.

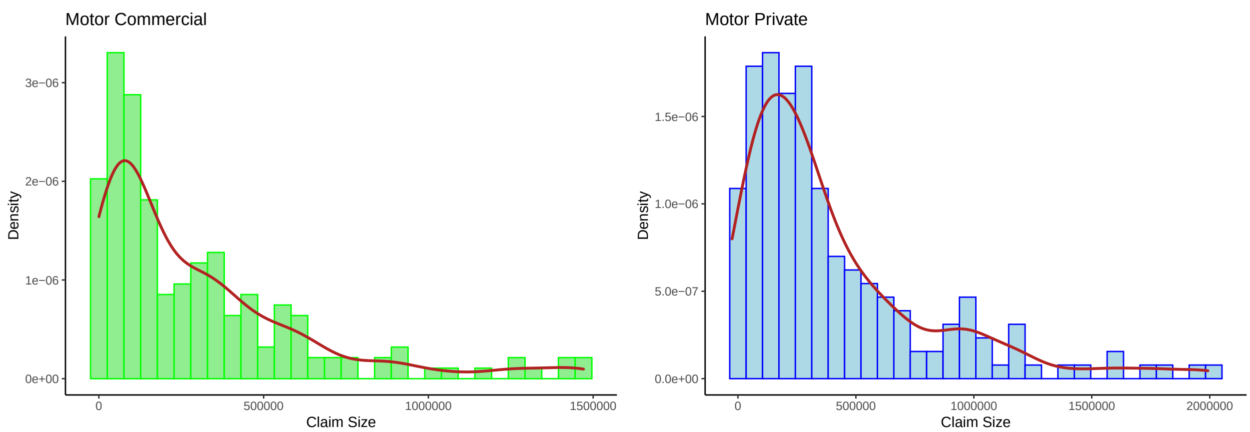

Figure 1: Histograms of original datasets

The Normal QQ-plots in Figure 2 show that the datasets don't match the normal standard distribution.

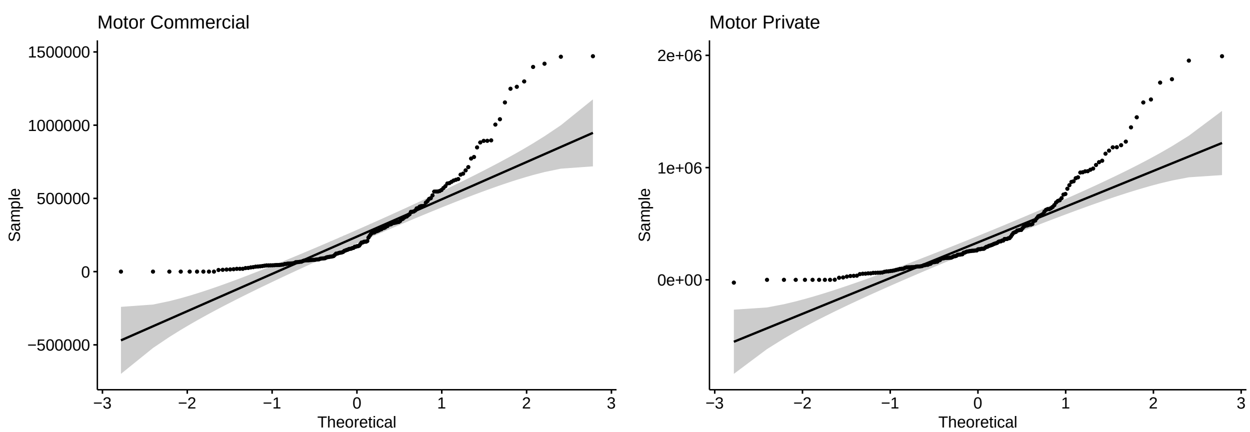

Figure 2: Normal QQ-plots of original data

This is a good indication that after fitting the distributions, non-parametric tests would be applied as opposed to parametric tests.

To make it simpler to work with, we transformed the data using the cube root function. QQ-Plots of the transformed data are shown in Figure 3.

The cube root transformed data seemed to be more suitable for fitting the distributions compared to the original data.

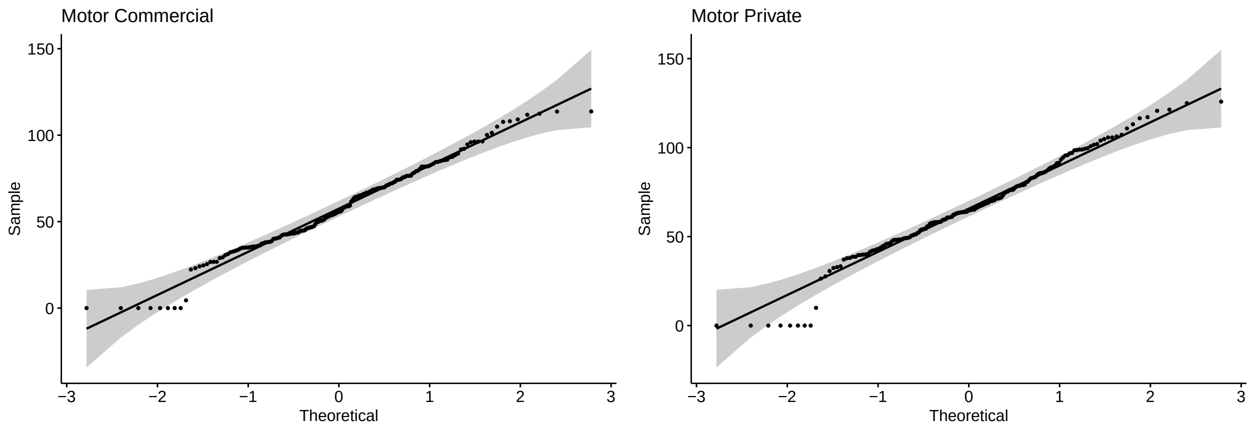

Figure 3: QQ-Plots of transformed data

Parameter Estimation

We used the Maximum Likelihood Estimation (`MLE`) method to obtain the various fitted distributions. Consequently, the most appropriate distribution is the one with the highest log-likelihood function (`LLF`).

Table 2: Estimated Parameters For Fitted Distributions

Distribution	Parameter	Motor Commercial	Motor Private
Exponential	Rate	0.02	0.01
	Std. Error	0.00	0.00
	LLF	-901.93	-920.29
Gamma	Shape	6.54	8.90
	Shape Std. Error	0.68	0.93
	Rate	0.11	0.13
	Rate Std. Error	0.01	0.01
	LLF	-800.47	-794.92
Log Normal	Mean Log	4.02	4.17
	Mean Log Std. Error	0.03	0.03
	SD Log	0.42	0.35
	SD Log Std. Error	0.02	0.02
	LLF	-810.28	-801.37
Weibull	Mean Log	2.91	3.33
	Mean Log Std. Error	0.17	0.19
	SD Log	67.40	76.47
	SD Log Std. Error	1.84	1.83
	LLF	-798.55	-795.27

Under the motor commercial class, Weibull distribution has the highest log-likelihood function (-798.55), followed by the Gamma distribution (-800.47), then Log-Normal (-810.28) and finally Exponential distribution (-901.93).

For motor private, Gamma distribution has the highest LLF (-794.92) followed closely by Weibull (-795.27), then Log-Normal (-801.37) and finally the Exponential distribution (-920.29).

From the LLF values, the Weibull distribution is the most appropriate for the motor commercial class and the Gamma distribution for motor private.

Goodness-Of-Fit Test

Typically, measures of goodness-of-fit summarize the discrepancy between observed values and the values expected under the model in question.

Two non-parametric tests were performed:

- Kolmogorov-Smirnov (K-S) test
- Anderson-Darling (A-D) test

Those two were used to determine the appropriateness of the fitted distributions to the incurred claims data.

Table 3: K-S and A-D test statistic values for fitted distributions

Test Statistic	Distribution	Motor Commercial	Motor Private
K-S	Exponential	0.3428	0.3785
	Gamma	0.0693	0.0395
	Log Normal	0.0846	0.0622
	Weibull	0.0655	0.0679
A-D	Exponential	31.8254	37.2804
	Gamma	0.6340	0.2429
	Log Normal	1.1405	0.6198
	Weibull	0.8083	0.9116

To determine the most suitable continuous distribution for the incurred claims, we fish for smaller K-S and A-D test statistic values.

For the motor commercial class, Weibull distribution had the smallest K-S statistic (0.0655), followed very closely by Gamma distribution (0.0693). But under the A-D statistic, Gamma (0.6340) beats Weibull (0.8083) by a considerable margin. The Gamma distribution would be the best fit for this class.

As expected for the motor private class, the Gamma distribution has the least K-S statistic (0.0395) as well as A-D statistic (0.2429) which again makes it the most appropriate fit for this class.

From the K-S and A-D tests, the Gamma distribution seems to be the best fit for both auto-insurance classes.

Information Criteria

Two approaches were used here:

- Akaike's Information Criteria (AIC)
- Bayesian Information Criteria (BIC)

Lower values for both AIC and BIC indicate a more appropriate distribution.

Table 4: AIC and BIC values for fitted distributions

Information Criterion	Distribution	Motor Commercial	Motor Private
AIC	Exponential	1805.85	1842.59
	Gamma	1604.95	1593.83
	Log Normal	1624.55	1606.74
	Weibull	1601.10	1594.54
BIC	Exponential	1809.03	1845.76
	Gamma	1611.30	1600.17
	Log Normal	1630.91	1613.08
	Weibull	1607.45	1600.88

For the motor commercial class, Weibull had the lowest AIC and BIC values (1601.10, 1607.45).

For motor private, Gamma distribution had the minimum AIC and BIC values (1593.83, 1600.17).

From the AIC and BIC values, Weibull was the most appropriate distribution for motor commercial and Gamma for motor private. In both auto-insurance classes, the Exponential distribution seems to have the worst fit.

Conclusion

Our research entailed determining the statistical distribution that best fits claim severity of motor insurance in Kenya. It could then be used to predict future claim incurrence.

After selecting a family of continuous, positively skewed distributions (Exponential, Gamma, Lognormal and Weibull), their parameters were estimated using the MLE method, K-S and A-D tests applied to test their goodness-of-fit, and finally AIC and BIC criterion used to determine the most appropriate distribution among the chosen ones.

The most suitable distribution is the one with:

- Maximum LLF
- Minimum K-S and A-D test statistic values
- Minimum AIC and BIC values

According to our study, the Weibull distribution is the best fit for modelling claim severity in motor commercial class, while the Gamma distribution is best for the motor private class.

References

[1] Achieng, O.M. (2010). Actuarial Modeling for Insurance Claim Severity in Motor Comprehensive Policy Using Industrial Statistical Distributions. International Congress of Actuaries, Cape Town, Vol. 712. 7-12.

[2] Ahmad, Z., Mahmoudi, E., Sanku, D. and Saima, K. K. (2020). Modeling Vehicle Insurance Loss Data Using a New Member of T-X Family of Distributions. Journal of Statistical Theory and Applications, 19, 133-147.

[3] Akaike, H. (1974). A New Look at Statistical Model Identification. IEEE Transactions on Automatic Control, 19, 716-723.

[4] Bahnemann, D. (2015). Distributions for Actuaries. CAS Monograph Series No. 2, Casualty Actuarial Society, Arlington.

[5] Boland, P. J. (2006). Statistical Methods in General Insurance. Unpublished M. Ed. Dissertation, National University of Ireland, Dublin.

[6] Boucher, J.P., Denuit, M. and Guillén, M. (2007). Risk Classification for Claim Counts: A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models. North American Actuarial Journal, 11, 110-131.

[7] Burnecki, K., Misiorek, A. and Weron, R. (2010). Loss Distributions. Unpublished M. Ed. Dissertation, Wroclaw University of Technology, 50-370 Wroclaw, Poland.

[8] Dodge, Y. (2008). The concise encyclopedia of statistics: With 247 tables. New York, NY: Springer.

[9] Dutta K. and Perry J., (2006). A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital. Federal Reserve Bank of Boston, MA, USA. Working Papers, No. 06-13.

[10] Frees, W. E. and Valdez, E. A. (2012). Hierarchical Insurance Claims Modeling. Journal of the American Statistical Association, 103, 484, 1457-1469.

[11] Hakim, A. R., Fithriani, I. and Novita M. (2021). Properties of Burr distribution and its application to heavy tailed survival time data. Journal of Physics: Conference Series, 1, 1, 6-8.

[12] Hesse, C., Ofosu, J. B. and Nortey, E. N. (2017). Introduction to Nonparametric Statistical Methods. Akrong Publications Ltd, Accra.

[13] Hogg R. V. and Klugman S. A. (1984). Loss Distributions. New York: John Wiley & Sons.

[14] lgnatov, Z. G., Kaishev, V. K. and Krachunov, R. S. (2001). An improved finite time ruin probability formula and its Mathematica implementation. Journal of Insurance: Mathematics and Economics, 29, 3, 375-386.

[15] Mazviona, B. W., & Chiduza, T. (2013). The Use of Statistical Distributions to Model Claims in Motor Insurance. International Journal of Business, Economics and Law, 3, 1, 44-57.

[16] Merz, M. and Wüthrich, M. V. (2008). Modelling the Claims Development Result for Solvency Purposes. Casualty Actuarial Society E-Forum Fall 2008, 542-568.

[17] Mihaela, D. and Jemna Dănuţ-Vasile (2015). Modeling the Frequency of Auto insurance claims by Means of Poisson and Negative Binomial models. Journal of Scientific Annals of Economics and Business of the “Alexandru Ioan Cuza” University of laşi, Romania, 62, 2, 151-168.

[18] Nduwayezu, F. (2016). Finding appropriate Loss Distributions to insurance data: Case study of Kenya (2010-2014) Unpublished M. Ed. Dissertation, Strathmore University, Nairobi, Kenya.

[19] 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.

[20] Packova, V., and Brebera, D. (2015). Loss Distribution in Insurance Risk Management. Proceedings of the International Conference on Economics and Statistics (ES 2015). Vienna, Austria, March 15-17, 2015, 17-22.

[21] Selvakumar, V., Satpathi, D. K., Praveen Kumar, P. T. V. and Haragopal, V. V. (2022). Modeling of Motor Insurance Extreme Claims through Appropriate Statistical Distributions. Unpublished M. Ed. Dissertation, Birla Institute of Technology and Science – Pilani, Hyderabad, India.

[22] Vaughan, E. J. (1997). Risk Management. New York: John Wiley.