International Journal of Innovation in Science and Mathematics
Volume 3, Issue 3, ISSN (Online): 2347–9051

Modeling Individual Claims for Motor Third Party Liability of Insurance Companies in Albania

Oriana Zacaj
Department of Mathematics,
Polytechnic University, Faculty of
Mathematics and Physics Engineering
Email: ana_25al@yahoo.com

Eralda Dhamo (GJIKA)
Departments of Mathematics,
University of Tirana,
Faculty of Natural Science
Email: eralda.dhamo@unitir.edu.al

Shpetim Shehu
Department of Mathematics,
Polytechnic University, Faculty of
Mathematics and Physics Engineering

***Abstract* – When an insurer forecasts its claims, has to deal with two main problems; the first one is to find a model that fits the data, and the second one is to check how well that model fits the data in this paper we try to fit several nonnegative continuous distributions to the data from the claim of the MTPL portfolio of an Albanian insurance company, and then testing how well these models fits to the available data, and judge which among them fits better to our data. The modeling process will establish some probability models that could model the claim amounts, and then we'll perform some diagnostic checks like Kolmogorov – Smirnov test, and Anderson – Darling test and we'll use the Akaike's Information Criterion (A.I.C) and Bayesian information Criterion (B.I.C) and graphically using the Q-Q plots and P – P plots. Finally the study gives a summary, a conclusion and recommendations that can be used by insurance companies to improve their results concerning future claim assumptions**

***Keywords* – Bodily Injury Claims, Claim, Distribution, Insurance Contracts, Estimation, Property Claims, R.**

I. INTRODUCTION

The main concern of the daily work of an actuary in the insurance industry is to asses appropriate values on the cash flows within the insurance company. It is his duty to construct and analyze mathematical models to describe the processes. General insurance is perhaps the most complex sector for actuaries. It includes all types of insurance coverages including health insurance, property insurance and the most complex of all the third party liability insurance. For any country, the insurance industry is of great importance because it is a form of economic remediation. It provides a means of reducing financial loss due to the consequences of risks by spreading or pooling the risk over a large number of people. Insurance being a data-driven industry with the main cash out-flow being claim payments, Insurance companies employ large numbers of analysts, including actuaries, to understand the claims data. Claim actuaries are not interested in the occurrence of the claims themselves but rather in the consequences of its random out-come. That is, they are concern with the amount the insurance company will have to pay than with the particular circumstances which give rise to the claim numbers. The general insurance actuary needs to have an understanding of the various models for the risk consisting of the total or aggregate amount of claims payable by an insurance company over a fixed period of time. Insurance data contains relatively large claim amounts, which may be infrequent, hence there is need to find and use probability models with relatively heavy tails and highly skewed like the exponential, gamma, pareto, weibull and log-normal. These models are informative to the company and they enable it make decisions on amongst other things: solvency, premium loading, expected profits, technical reserves, profitability and the impact of reinsurance and deductibles. In view of the economic importance of motor third party liability insurance in developing countries, it comes a necessity to find a probabilistic model for the distribution of the claim amounts. Although the empirical distribution functions can be useful tools in understanding claims data for motor policy, there is always a necessity to "fit" a probability model with reasonably tractable mathematical properties to the claims data. For that reason this paper involves the steps taken in actuarial modeling to find a suitable probability distribution for the claims data and testing for the goodness of fit of the supposed distribution. Finally, constructing interpretable models for claim amounts can often give one a much added insight into the complexity of the large amounts of claims that may often be hidden in a huge amount of data.

The paper consist in finding an appropriate probability model for the claim data used in actuarial analysis of insurance claim amounts and more specifically in motor policy

In the first part we introduce the behavior of our data: its empirical distribution, its histogram plots, mean, variance, skewness and kurtosis.

Then in the second part we select and try to fit a probability model from: Gamma, Log-normal and Weibull. Parameters are estimated using the maximum likelihood estimation method. After which the log likelihood and the roots estimates are computed. Testing of the goodness of fit is then done using the Q – Q plots, P – P plots Kolmogorov – Smirnov test, Cramer Von Mises test, Anderson Darling test A.I.C and B.I.C.

This modeling process is performed using R soft ware

II. PROBLEM STATEMENT

For claim actuaries, claim modeling is very crucial since a good understanding and Interpretation of loss distribution is the back-bone of all the decisions made in the Insurance industry regarding, premium loading, expected profits, reserves necessary to Ensure profitability and the impact of re-insurance. MTPL portfolio and MTPL claims takes the majority of the business in Albania. For this reason it is crucial to have a good knowledge on the behavior of these claims. Choosing the

most suitable loss distribution therefore is important especially for the large amounts of claims settled by insurance companies. An understanding of the probability and statistical distribution is vital for the general Insurance actuary to be able to summarize and model the large amounts of claims data and give timely outcomes. These distributions are "must-have" tools for any actuarial assessment, that's why the study went further to get their specific descriptions to emphasize their different properties and how they are useful in insurance claims Data. There is need to compare the different methods used to test for the goodness of fit of the selected probability distribution model chosen for the claims data. This is vital because it is only by choosing the best method amongst the sampled methods that the actuaries attain accuracy and consistency in making financial sense of the future.

The general objective was to test for an appropriate probability distribution model for the MTPL claim amounts and to test how well the chosen probability distribution model fits the claims data

Appropriate probability distribution models for claim distributions

The values of claims within a block of insurance contracts are is usually modelled as a nonnegative continuous random variable. Some standard continuous distributions for modelling claim values may be Exponential distribution, Gamma distribution, Weibull distribution, Lognormal distribution, and Pareto distribution. Also we may build other non-negative continuous distributions by methods of transformation, splicing, and mixture distribution:

For the data in consideration we'll use Gamma distribution, Weibull distribution, Lognormal distribution:

X is said to have a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, denoted by $\mathcal{G}(\alpha, \beta)$, if its pdf is

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} \text{ for } x \geq 0 \text{ and } \Gamma(\alpha)$$

$$= \int_0^{\infty} y^{\alpha-1} e^{-y} dy$$

The mean and the variance of X are:

$$E(X) = \alpha\beta, Var(X) = \alpha\beta^2$$

The moment generating function of X is $M_X(t) = \frac{1}{(1-\beta t)^{\alpha}}$

A random variable X has a 2-parameter Weibull distribution $\mathcal{W}(\alpha, \lambda)$ $\alpha > 0, \lambda > 0$ if its pdf is

$$f_X(x) = \left(\frac{\alpha}{\lambda}\right)\left(\frac{x}{\lambda}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\lambda}\right)^{\alpha}\right], \quad x \geq 0$$

where α is the shape parameter and λ is the scale parameter.

The distribution function of X is

$$F_X(x) = 1 - \exp\left[-\left(\frac{x}{\lambda}\right)^{\alpha}\right], \quad x \geq 0,$$

The mean and the variance of X are:

$$E(X) = \lambda\Gamma\left(1 + \frac{1}{\alpha}\right), (2.47)\ Var(X) = \lambda^2\Gamma\left(1 + \frac{2}{\alpha}\right) - \mu^2$$

A random variable X has a Lognormal distribution Lognormal(μ, σ) $\mu > 0, \sigma > 0$ if its pdf is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left[-\frac{1}{2}\left(\frac{\log x - x}{\sigma}\right)\right], \quad x \geq 0$$

We have the relationship if X~Lognormal(μ, σ) then $Y = \log X \sim N(\mu, \sigma^2)$

III. Model Construction and Evaluation

In this part of the paper we are concerned in modeling claim numbers and claim sizes; that is, fitting probability distributions from selected families to sets of data consisting of observed claim numbers or claim sizes. The family may be chosen after an exploratory analysis of the data set – looking at numerical summaries such as mean, median, mode, standard deviation (or variance), skewness, kurtosis and plots such as the empirical distribution function. Of course, also we will try to fit a distribution from each of several families to provide comparisons among the fitted models, comparisons with previous work and choice.

Various criteria are available, including the method of moments, the method of maximum likelihood, the method of percentiles and the method of minimum distance.

After a model has been estimated, we have to evaluate it to ascertain that the assumptions applied are acceptable and supported by the data. This should be done prior to using the model for prediction and pricing. Model evaluation can be done using graphical methods, as well as formal misspecification tests and diagnostic checks.

Nonparametric methods have the advantage of using minimal assumptions and allowing the data to determine the model. However, they are more difficult to analyze theoretically. On the other hand, parametric methods are able to summarize the model in a small number of parameters, although with the danger of imposing the wrong structure and oversimplification. Using graphical comparison of the estimated df and pdf, we can often detect if the estimated parametric model has any abnormal deviation from the data.

Formal misspecification tests can be conducted to compare the estimated model (parametric or nonparametric) against a hypothesized model. When the key interest is the comparison of the df, we may use the Kolmogorov–Smirnov test and Anderson–Darling test. The chi-square goodness-of-fit test is an alternative for testing distributional assumptions, by comparing the observed frequencies against the theoretical frequencies. The likelihood ratio test is applicable to testing the validity of restrictions on a model, and can be used to decide if a model can be simplified.

When several estimated models pass most of the diagnostics, the adoption of a particular model may be decided using some information criteria.

IV. Results and Discussions

In this part of the paper we are considering the fitted distribution for our data. The problems discussed are the individual claims for MTPL portfolio of an Albanian

insurance company for the period 2005 – 2014 which are compared with different nonnegative continuous distributions. The value of claims is spitted into two categories which are bodily injury claims and property claims

A. Bodily injury claims

Table 1 shows a summary of the basics characteristics of the individual bodily injury claims. This statistics is necessary as a preliminary step to find the dispersal of the data and then fitting an appropriate known distribution.

Table 1: Descriptive statistics for bodily injury claims (values are in ALL)

Min	Max	Mean	Median	Sd.	Skewness	Kurtosis
1818	6036102	1568580	124600	1453211	0.85	2.794

An histograme of the data (Figure 1) shows a skewness in the left of the data. The proposals for the probability model of the value of the bodily injury claims are: gamma, weibull, lognormal, pareto etc.

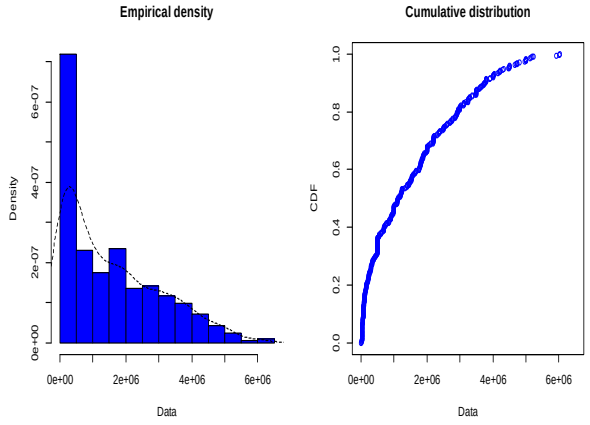

Fig.1. Histogram and cumulative distribution of the bodily injury claims

Based on the previous assumptions on the claim distribution, we have estimated the parameters from: gamma, weibull and lognormal distribution using the MLE (calculations are made with the help of R software). Table 2 shows the distribution characteristics for two probability distributions. We tried to perform an evaluation of the parameters for the gamma distribution but the results were not satisfactory so we decide to go for two main distribution: Weibull and lognormal.

Table 2: Summary of estimation results for bodily injury claims

Distribution	Shape	Scale	AIC	BIC	Loglikelihood
Weibull	8.925	1.49	13002.01	13010.12	-6499.005
	Meanlog	Sdlog	AIC	BIC	Loglikelihood
Lognormal	13.529	0.052	13088.23	13096.34	-6542.117

From the two of the probability distributions we'll chose the one that better fits the real data. Observing the value of AIC for both distribution (Table 2) we can find that the weibull model has the smallest value of AIC. This is a singn that this distribution can serve better to model the bodily injury claims. But this is not enough. Below we've performed some other tests to come to a conclusion of arguing on the appropriateness of the model. Histogram of theoretical densities and Quantile-Quantile plot where used to compare the two fitted distribution. Graphical results are shown in Figure 2. Also empirical theoretical CDFs and P-P plot were used to compare the goodness of each evaluation

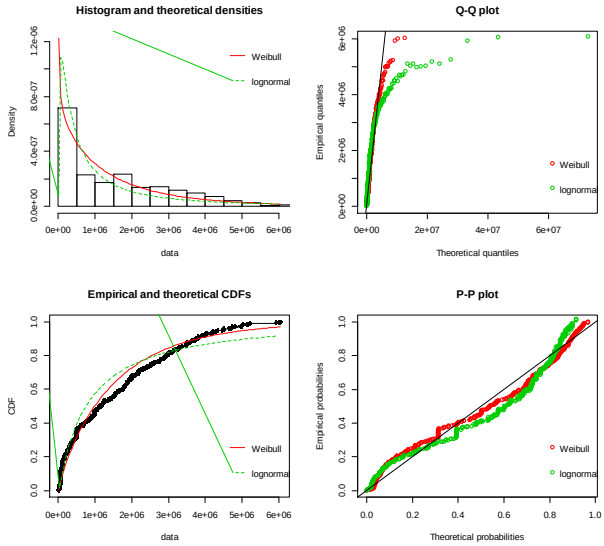

Fig.2. Histogram, theoretical densities, empirical theoretical CDFs, and P-P plot (Weibull and lognormal)

Four graphics tests show that the most appropriate probability model for the bodily injury claims is the Weibull model. Further Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling and statistics are also computed, as defined by Stephens (1986). Using the fitdistrplus package in R an approximate Kolmogorov-Smirnov test, Cramer-von Mises and Anderson-darling tests are performed by assuming the distribution parameters known. The critical value defined by Stephens (1986) for a completely specified distribution is used to reject or not the distribution at the significance level 0.05. Those tests are available only for maximum likelihood estimations.

Table 3: Results of statistics for some tests

	Weibull	Lognormal
Kolmogorov-Smirnov statistic	0.0879	0.1231
Cramer-von Mises statistic	0.8757	2.0034
Anderson-Darling statistic	5.9883	12.1350

As seen from the results in Table 3, again we confirm our suspicion that between Weibull and lognormal probability model, the one that better fits the data is the Weibull distribution.

So, at the end of many tests we agree that between Weibull distribution and lognormal distribution the one that best fit bodily injury claims is the Weibull distribution.

B. Property claims

Table 4 shows a summary of the basics characteristics of the individual property claims. This statistics is

necessary as a preliminary step to find the dispersal of the data and then fitting an appropriate known distribution.

Table 4: Descriptive statistics for property claims (values are in ALL)

Min	Max	Mean	Median	Sd.	3rd Qu.	Skewness	Kurtosis
2600	1982917	84257.03	43000	130225.1	165500	5.5	50.46

An histograme of the data (Figure 3) shows a skewness in the left of the data. The proposals for the probability model of the value of the property claims are: gamma, weibull, lognormal, pareto etc.

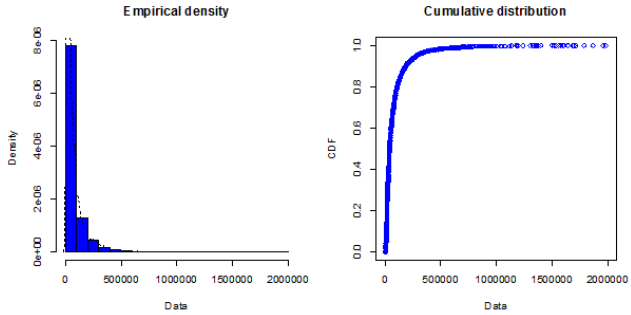

Fig.3. Histogram and cumulative distribution of the property claims

Based on the previous assumptions on the claim distribution, we have estimated the parameters from: gamma, weibull and lognormal distribution using the MLE (calculations are made with the help of R software). Table 4 shows the distribution characteristics for two probability distributions. We tried to perform an evaluation of the parameters for the gamma distribution but the results were not satisfactory so we decide to go for two main distribution: Weibull and lognormal.

Table 5: Summary of estimation results for property claims

Distribution	Shape	Scale	AIC	BIC	Loglikelihood
Weibull	9.310	8.46	335081.2	335096.2	-167538.6
	meanlog	Sdlog	AIC	BIC	Loglikelihood
Lognormal	10.774	1.001	331219.8	331234.8	-165607.9

From the two of the probability distributions we'll chose the one that better fits the real data. Observing the value of AIC for both distributions (Table 5) we can find that the lognormal model has the smallest value of AIC. This is a sign that this distribution can serve better to model the property claims. But this is not enough. Below we've performed some other tests to come to a conclusion of arguing on the appropriateness of the model. Histogram of theoretical densities and Quantile-Quantile plot where used to compare the two fitted distribution. Graphical results are shown in Figure 4. Also empirical theoretical CDFs and P-P plot were used to compare the goodness of each evaluation

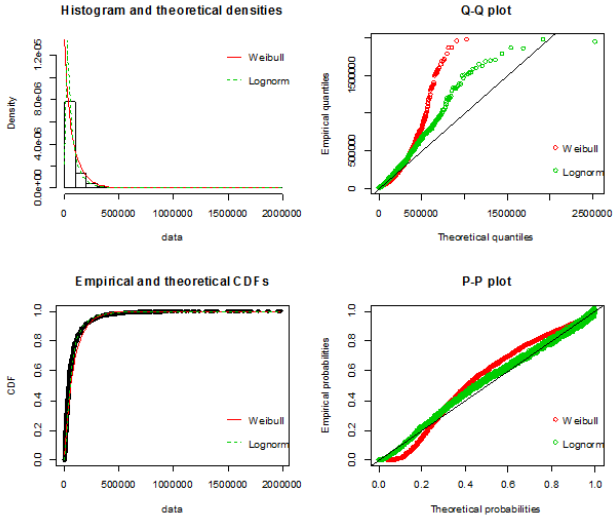

Fig.4. Histogram, theoretical densities, empirical theoretical CDFs, and P-P plot (Weibull and lognormal)

Four graphics tests show that the most appropriate probability model for the property claims is the lognormal model. Further Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling and statistics are also computed, as defined by Stephens (1986). Using the fitdistrplus package in R an approximate Kolmogorov-Smirnov test, Cramer-von Mises and Anderson-darling tests are performed by assuming the distribution parameters known. The critical value defined by Stephens (1986) for a completely specified distribution is used to reject or not the distribution at the significance level 0.05. Those tests are available only for maximum likelihood estimations.

Table 6: Results of statistics for some tests

	Weibull	Lognormal
Kolmogorov-Smirnov statistic	0.1068	0.0509
Cramer-von Mises statistic	65.3660	8.8409
Anderson-Darling statistic	382.7199	54.0210

As seen from the results in Table 6, again we confirm our suspicion that between Weibull and lognormal probability model, the one that better fits the data is the lognormal distribution.

So, at the end of many tests we agree that between Weibull distribution and lognormal distribution the one that best fit property claims is the lognormal distribution.

V. CONCLUSIONS

In this study we discussed actuarial models for claim losses. We discussed the necessity of having a deeper knowledge of te MTPL claims in the Albanian market. We discussed modeling of the individual claims by introducing some techniques for modeling nonnegative continuous random variables. We discussed the model construction and evaluation which are two important aspects of the empirical implementation of loss models.

We also introduced a special case by analyzing the individual claim values for the bodily injury and property claoms for the portfolio of motor third party liability of an Albanian insurance company, applying the methods discussed in this study

REFERENCES

[1] C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995. (8 [1] S. A. Klugman, H. H. Panjer, G E. Willmot,(2004) Loss Models From Data to Decisions,2nd edition
[2] YIU-KUEN TSE, (2009), Nonlife Actuarial Models Theory, Methods and Evaluation
[3] M. H DeGroot,. M. J. Schervish, (2002), Probability and Statistics, 3rd edition
[4] R.V. Hogg, ,A.T. Craig (1995), Introduction to Mathematical Statistics, 5th edition,
[5] N. L. Johnson, S. Kotz, (1969), Distributions in Statistics: Discrete Distributions
[6] N. L. Johnson, S. Kotz, (1970) Distributions in Statistics: Continuous Univariate Distributions-I.
[7] S. Ross, (2006), AFirst Course in Probability, 7th edition,
[8] O.Zacaj, E.Dhamo (2011) Loss Models: Statistical Methods in modelling losses deriving from the insurance contracts, National Conference on advanced studies in the mathematics, chemistry, and physic engineering
[9] D'arcy, Stephen P. (1989) "On becoming an actuary of the third kind" (PDF) Proceedings of the casualty Actuarial Society LXXVI (145)
[10] Roger J. Gray, Susan M. Pitts "Risk modeling in general insurance: From Principles to Practice
[11] Commeau, N., Parent, E., Delignette-Muller, M.-L., and Cornu, M. (2012). Fitting a lognormal distribution to enumeration and absence/presence data. International Journal of Food Microbiology.
[12] D'Agostino, R. and Stephens, M. (1986). Goodness-of-Fit Techniques. Dekker, 1st edition.
[13] Delignette-Muller, M., Pouillot, R., Denis, J., and Dutang, C. (2014). fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data. R package version 1.0-2.
[14] Kohl, M. and Ruckdeschel, P. (2010). R Package distrMod: S4 Classes and Methods for Probability Models. Journal of Statistical Software.
[15] Mal_a, I. (2013). The use of _nite mixtures of lognormal and gamma distributions. Research Journal of Economics, Business and ICT.
[16] Mandl, J., Monteiro, J., Vrisekoop, N., and Germain, R. (2013). T Cell-Positive Selection Uses Self-Ligand Binding Strength to Optimize Repertoire Recognition of Foreign Antigens. Immunity.