Trial Modelling For MTPL Claims In Albania

Oriana Zacaj
Department of Mathematics, Faculty of Mathematics and Physics Engineering, Polytechnic University Of Tirana
Tirana,Albania
[1] ana_25al@yahoo.com

Eralda Dhamo (Gjika)
Departments of Mathematics, Faculty of Natural Science, University of Tirana
Tirana,Albania
[2] eralda.dhamo@unitir.edu.al

Abstract— **MTPL is the largest portfolio in Albania with 65% of the market, and for the insurers is a concern knowing about the nature and the expectation of their claims. Claims are composed with two elements which are the frequency and the severity. Finding a fitted model for claim severity and claim frequency, and then estimating that model for appropriateness becomes the main issue for the insurers. In this paper we'll try to use several non negative distributions to model each of the claim components with the data of an Albanian insurance company and then use some of the diagnostic tests to estimate which fits better to these data using R software.**

Keywords— claim; severity; frequency; distribution; insurance contracts; estimation

I. INTRODUCTION

Knowing about the behaviour of their claims is very important for the insurance companies. Modelling claims helps in estimating current and future liabilities deriving from insurance contracts. In this context loss models are very important in pricing new products and reviewing the actual prices.

Modelling losses in realistic basis helps in estimating different ratios regarding the solvency of the insurer and also helps in modelling the planning of the future operations of the insurer.

The largest and the most important insurance portfolio in Albanian market is the MTPL portfolio.

In this paper we'll try to present different models of discrete non negative distributions for claim severity and nonnegative continuous distributions for claim severity, and then we'll try to fit these distributions to the data of an Albanian insurance company

In this paper are discussed actuarial models for claims deriving from the MTPL portfolio of an insurance company. The subject is divided in three parts In the first part we'll try to present different models of discrete non negative distributions for claim severity and nonnegative continuous distributions for claim severity.

In the second part we discuss the model construction and evaluation which are two important aspects of the empirical implementation of loss models. To construct a parametric model of loss distributions, the parameters of the distribution have to be estimated based on observed data. Alternatively, we may consider the estimation of the distribution function or density function without specifying their functional forms, in which case nonparametric methods are used. Competing models are selected and evaluated based on model selection criteria, including goodness-of-fit tests. Computer simulation using random numbers is an important tool in analyzing complex problems for which analytical answers are difficult to obtain. We discuss methods of generating random numbers suitable for various continuous and discrete distributions. We also consider the use of simulation for the estimation of the mean squared error of an estimator and the p-value of a hypothesis test.

And in the third part we'll, try to fit the selected distributions to the data of an Albanian insurance company and then use the diagnostic tests to find the most appropriate of them

II. TYPES OF DISTRIBUTIONS

A. *Claim frequency distribution*

An important measure of claim losses is the claim number and especially the claim frequency, which is the number of claims in a block of insurance policies over a period of time. Though claim frequency does not directly show the monetary losses of insurance claims, it is an important variable in modelling the losses. The claim number is modelled as a discrete non-negative random variable, and the claim frequency as the parameter of the distribution Some commonly used discrete distributions in modelling claim-number and claim frequency may be Binomial distribution, Geometric distribution, Negative Binomial distribution, and Poisson distribution. Also mixture distributions like Poisson distribution with mixing Gamma distributions can be used.

For the data in consideration we'll use N/λ Poisson (λ) with mixing distribution for the value of claim frequency and finding the distribution which better fits to the claim frequency.

B. *Claim severity distribution*

The aggregate claims for losses of the block of policies, is the sum of the monetary losses of all the claims. Unlike claim frequency, which is a nonnegative integer-valued random variable, claim severity is usually modelled as a nonnegative continuous random variable. Some standard continuous distributions for modelling claim severity 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-parameterWeibull 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 $\mathrm{Lognormal}(\mu, \sigma)$ $\mu > 0, \sigma > 0$ 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 $X \sim \mathrm{Lognormal}(\mu, \sigma)$ 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 claim severity and claim frequency which are compared with different nonnegative continuous distributions. The data used consist on monthly claim frequency an monthly claim severity during years 2005 – 2014 for the MTPL portfolio of an Albanian insurance company.

A. *Claim severity*

Table 1 shows a summary of the basics characteristics of the monthly claim severity values. This statistics is necessary as a preliminary step to find the dispersal of the data and then fitting an appropriate known distribution.

TABLE I. Descritpive Statistics For Claim Severity (values are in ALL)

N	Min	Max	Mean	Median	1st Qu.	3rd Qu.	Skewness	Kurtosis
120	54930	363600	138800	124600	91840	165500	1.2	4.248

An histogram of the data (Figure 1) shows a skewness in the left of the data. The proposals for the probability model of the claim severity are: gamma, weibull, lognormal, pareto etc.

Based on the previous assumptions on the claim severity 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.

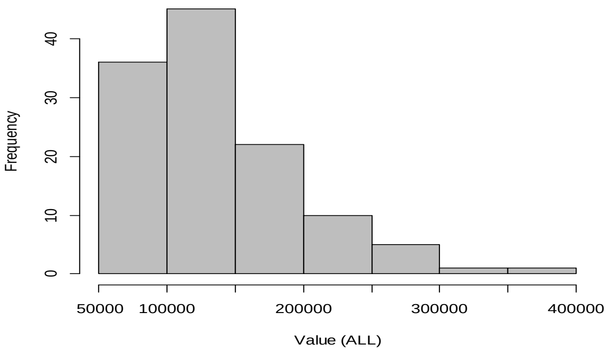

Fig. 1. *Histogram of the claim severity*

TABLE II. SUMMARY OF ESTIMATION RESULTS FOR CLAIM SEVERITY

Distribution	Shape	Scale	AIC	BIC	Loglikelihood
Weibull	2.413	1.57	2974.721	2980.296	-1485.36
Lognormal	11.755	0.407	2950.179	2955.754	-1473.09

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 lognormal model has the smallest value of AIC. This is a singn that this distribution can serve better to model the claim severity. 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 (Figure 3)

Four graphics tests show that the most appropriate probability model for the claim severity 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 the significance level 0.05. Those tests are available only for maximum likelihood estimations.

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 lognormal distribution.

So , at the end of many tests we agree that between weibull distribution and lognormal distribution the one that best fit claim severity is the lognormal distribution.

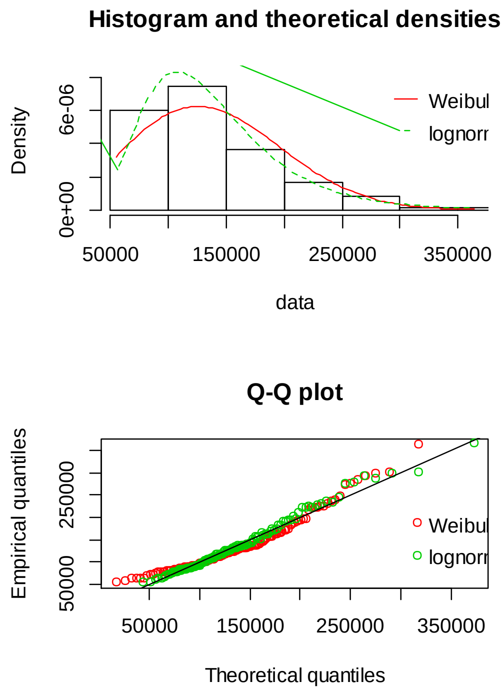

Fig. 2. *Histogram and theoretical densities (Weibull and lognormal)*

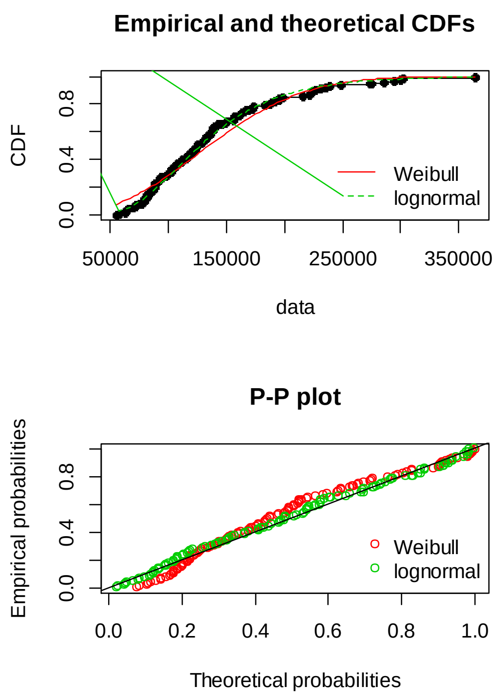

Fig. 3. *Empirical theoretical CDFs and P-P plot*

TABLE III. RESULTS OF STATISTICS FOR SOME TESTS

	Weibull	**Lognormal**
Kolmogorov-Smirnov statistic	0.1142	**0.0548**
Cramer-von Mises statistic	0.3682	**0.0679**
Anderson-Darling statistic	2.3806	**0.4761**

B. *Claim frequency*

Table 4 shows a summary of the basics characteristics of the monthly claim frequency values.

TABLE IV. DESCRITPIVE STATISTICS FOR CLAIM FREQUENCY

N	Min	Max	Mean	Medi an	1st Qu.	3rd Qu.	Skewn ess	Kurtos is
120	0.0066	0.045	0.0251	0.0249	0.0198	0.02987	0.03424	3.0032

A histograme of the data (Figure 4) shows a simetry. On different studies, the proposals for the probability model of the claim frequency are: gamma, weibull, lognormal, pareto distributions etc..

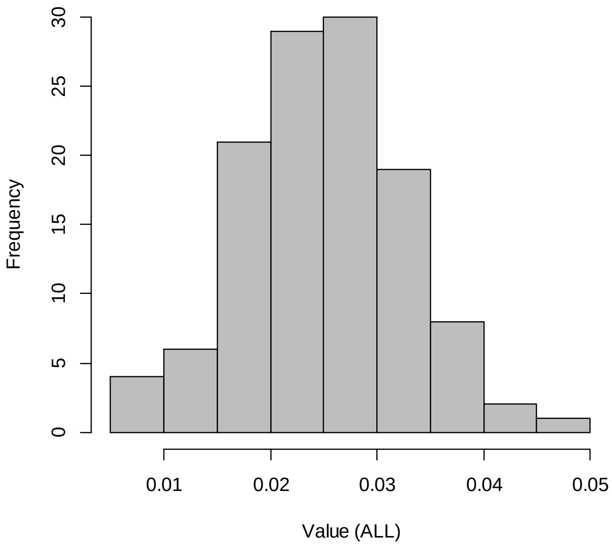

Fig. 4. *Histogram of the claim frequency*

Based on the previous assumptions on the claim frequency 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 5 shows the distribution characteristics for three probability distributions.

TABLE V. SUMMARY OF ESTIMATION RESULTS FOR CLAIM FREQUENCY

Distribution	Shape	Rate	AIC	BIC	Loglikelihood
Gamma	9.27	387.47	-821.22	-815.64	412.61
	Shape	Scale	AIC	BIC	Loglikelihood
Weibull	3.68	0.0277	-829.33	-823.76	416.66
	meanlog	Sdlog	AIC	BIC	Loglikelihood
Lognormal	-3.74	0.342	-809.96	-804.38	406.98

From these three probability distributions we'll chose the one tha better fits the real data. Observing the value of AIC for both distribution (Table 4) we coan find that the lognormal distribution has the smallest value of AIC and BIC. This is a sign that this probability distributeon can serve better to model the claim frequency distribution.. Histogram of theoretical densities and Quantile-Quantile plot where also used to compare the two fitted distribution. Graphical results are shown in Figure 5 In addition the empirical theoretical CDFs and P-P plot were used to compare the goodness of each evaluation (Figure 6)

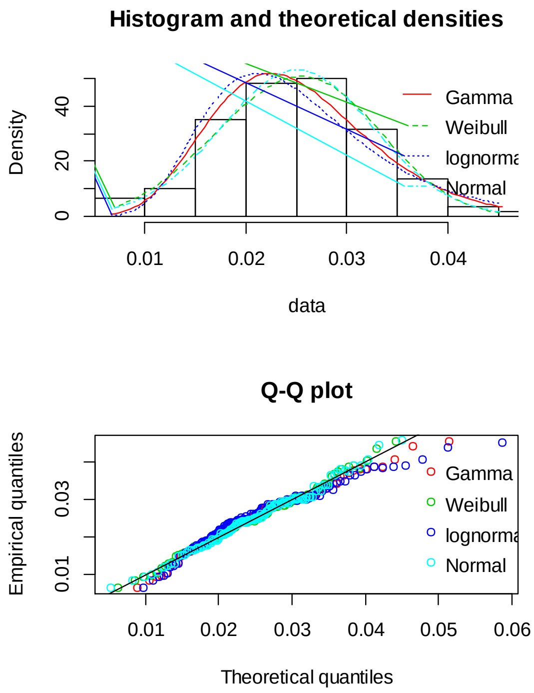

Fig. 5. *Histogram and theoretical densities*

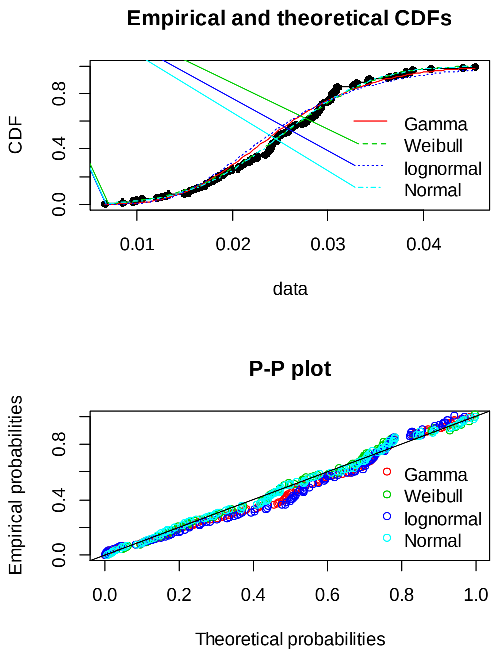

Fig. 6. *Empirical theoretical CDFs and P-P plot*

As seen from Figure 5 and Figure 6 the lognormal distribution is the one that fits better to our data. So, further tests should be performed in order to define the most appropriate model.

TABLE VI. RESULTS OF STATISTICS FOR SOME TESTS

	Gamma	Weibull	Lognormal
Kolmogorov-Smirnov statistic	0.097	0.077	0.1197
Cramer-von Mises statistic	0.1892	0.0597	0.3316
Anderson-Darling statistic	1.0848	0.3684	1.9524

As seen from the results in Table 6, the Weibull distribution model has smaller values for the three statistics. So, it results that it is the Weibull distribution that fits better to our data.

V. CONCLUSIONS

In this study we discussed actuarial models for claim losses. We discussed the component of the losses (claim severity and claim frequency) separately. We discussed modeling of these two components by introducing some techniques for modeling nonnegative integer-valued random variables and 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 claim frequency and claim severity for the portfolio of motor third party liability of an Albanian insurance company, applying the methods discussed in this study

VI. REFERENCES

[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