Modeling Frequency and Severity of Insurance Claims in an Insurance Portfolio

James Kiprotich Ng’elechei1,*, Joel Cheruiyot Chelule1, Herbert Imboga Orango2, Ayubu Okango Anapapa2

1Jomo Kenyatta University of Agriculture and Technology, Department of Statistics and Actuarial Science, Nairobi, Kenya
2University of Eldoret, Department of Mathematics and Computer Science, Eldoret, Kenya
*Corresponding author: jemongelechei@gmail.com

Received October 04, 2020; Revised November 05, 2020; Accepted November 11, 2020

Abstract

Premium pricing is always a challenging task in general insurance. Furthermore, frequency of the insurance claims plays a major role in the pricing of the premiums. Severity in insurance on the other hand, can either be the amount paid due to a loss or the size of the loss event. For insurer’s to be in a position to settle claims that occur from existing portfolios of policies in future, it is necessary that they adequately model past and current data on claim experience then use the models to project the expected future experience in claim amounts. In addition, non-life insurance companies are faced with problems when modeling claim data i.e selecting appropriate statistical distribution and establishing how well it fits the claimed data. Therefore, the study presents a framework for choosing the most suitable probability distribution and fitting it to the past motor claims data and the parameters are estimated using maximum likelihood method (MLE). The goodness of fit of frequency distributions was checked using the chi-square test and Anderson-Darling tests was applied to severity claim distributions. Best chosen models from frequency models and severity models were used to estimate the expected claim amount per risk in the following year. The study employed AIC to choose between competing models. Pareto and Negative Binomial model best fit severity claims, and frequency claims respectively. The two models were used for projection.

Keywords: claim frequency, claim severity, goodness-of-fit tests, Maximum Likelihood Estimation (MLE)


1. Introduction

In the classical Bonus-Malus System (BMS), the transition rule is governed by frequency; it ignores the severity information, which implies that the future loss of a policyholder can be appropriately modeled by predicting the frequency only, [1] and [2]. In reality, this frequency-driven BMS transition rule is the standard practice in many jurisdictions. It carries an implicit assumption that the frequency and severity are independent, such that the premium can be simply computed as a product of the mean frequency and mean severity. Theoretically, the structure of the traditional BMS is directly related to the classical theory of the collective risk model, which assumes independence between frequency and severity for mathematical tractability and convenience, [3]. However, a series of recent empirical studies have shown that the dependence between frequency and severity in auto insurance is statistically significant [4,5]. This phenomenon invalidates the practice of using frequency-driven BMS and highlights the need to extend the classical collective risk model by allowing some dependence structure between frequency and severity. Existing studies on dependent frequency-severity models and the associated insurance premiums include copula-based models [4,6], two-step frequency-severity models [5,7,8,9], and bivariate random effect-based models [10,11,12]. The random effect model is especially popular in insurance ratemaking because of the mathematical tractability in its prediction. The bivariate random effect model consists of two random effect components. The first random effect induces the dependence among frequencies and the second induces the dependence among individual severities. These two random effects are then jointly modeled to induce the dependence between frequency and severities at the distribution level. Statistical methods for prediction using the random effect model are well developed in the statistical and insurance literature, such as in; [10,11,12]. Statistical modeling and analysis for accommodating the various dependent structures in the collective risk models have been actively conducted. For the above reasons, modeling frequency and severity of insurance claims are being done simultaneously globally.
2. Literature Review

2.1. Review of Previous Studies on Severity of Insurance Claims

[13], did a comparison of risk classification methods for claims severity. They compared several risk classification methods for claim severity data by using weighted equation which was written as a weighted difference between the observed and fitted values. The weighted equation was applied to estimate claim severities which was equivalent to the total claim cost divided by the number of claims. From their data, they also observed that, the classical and regression fitting procedures gave equal values for parameter estimates however, the regression procedure provided a faster convergence. The multiplicative and additive models gave similar parameter estimates. The smallest chi-squares were given by the minimum chi-squares model except for the exponential model. All models provided similar values for absolute difference. This study therefore was meant to come up with best model among the chosen.

2.2. Review of Previous Studies on Insurance Claim Frequency

The Poisson regression model which is a GLM model was introduced by [14] and deeply looked into by [15].[16] Did put forward the Poisson distribution as the model to be used in modeling frequency of claims. Although it’s statistically conducive and favorable, it emphasized that the Poisson distribution had some shortcomings which limited its usage. The Poisson model had equidispersion property which is the equality of mean and variance. This study has considered wide range of models to come up with best model that can be used in insurance analysis of frequency claims.

3. Methodology

3.1. Scope of the Research

Research used secondary data from APA insurance company, from January 2017 to July 2019, regarding motor insurance.

Three assumptions were made on the data before use, these included:

i. All the claims came from the same distribution
ii. Zero claims are assumed to be non-inflated
iii. There exists no catastrophic claim

3.2. Procedures for Processing Data

This section describes the steps that were followed in fitting a statistical distribution to the claim data. There are basic standards that were followed in fitting the various probability distributions under consideration.

These steps were:

i. Choosing the model of distribution.
iii. Check goodness of fit of the model

iv. Specification of the criteria to choose best model from the selected distributions

3.3. Choosing Model of Distribution

There are made Considerations on the number of parametric probability models as potential candidates for the data generating mechanism of the claim amounts. Frequency distributions selected were Poisson, Binomial and Negative binomial distributions. They are best in modeling claims that occurs at an interval of time hence giving the best projection to insurance companies. The distribution selected for modeling Severity were Weibull, Log normal, Pareto and Gamma distributions. They are continuous distributions that best fit modeling of severity of insurance claims. Nevertheless, the list of potential probability distributions is large and it is worth to note that the choice of distributions is to some extend subjective. For this study the choice of the sample distributions was with regard to:

i. considering time constraint
ii. Accessibility of computer soft-ware to facilitate the study
iii. The volume and quality of data

3.4. Estimation of Probability Distribution Parameters

In the sense of estimating the parameters, the likelihood is usually observed as a function of the parameters to be measured. MLE was used because it provided a number of appropriate properties, including efficiency, impartial, asymptotic consistency and normality of invariance. The merit of considering the MLE is that it makes full use of all available information on the known parameters contained in the data and that it is highly flexible and that the method is statistically well understood.

3.5. Review of Severity Models

3.5.1. Weibull Distribution

The Weibull distribution is a continuous distribution that is commonly used to model the lifetimes of components. Weibull probability density function has two parameters, both positive constants that determine its location and shape. The probability density function of the Weibull distribution is

\[
f_x(x; \gamma, \beta) = \frac{\gamma}{\beta} x^{\gamma-1} \exp\left(-\left(\frac{x}{\beta}\right)^\gamma\right);
\]

\[x > 0, \gamma > 0, \beta > 0\]  \hspace{1cm} (3.1)

where $\gamma$ is the shape parameter and $\beta$ the scale parameter. When $\gamma =1$, the Weibull distribution is reduced to the exponential distribution with parameter $\lambda = \beta$.

The likelihood function for the Weibull distribution is given by:

\[
L(\gamma, \beta) = \prod_{i=1}^{n} \left(\frac{\gamma}{\beta}\right) x_i^{\gamma-1} \exp\left(-\left(\frac{x_i}{\beta}\right)^\gamma\right)
\]

\[\hspace{1cm} (3.2)\]

The log-likelihood function is therefore given by
Thus, to determine the parameter estimates, we equate the derivatives of the log-likelihood function to zero and solve the following equations

\[
\frac{\partial l(\gamma, \beta)}{\partial \gamma} = -\frac{n}{\gamma} - n \ln \beta + \sum_{i=1}^{n} \ln x_i - \frac{1}{\beta} \sum_{i=1}^{n} x_i^\gamma \ln x_i = 0 \quad (3.4)
\]

\[
\frac{\partial l(\gamma, \beta)}{\partial \beta} = \frac{n \beta \gamma}{\beta} - \gamma \sum_{i=1}^{n} x_i^\gamma = 0 \quad (3.5)
\]

By eliminating $\beta$ and simplifying, we obtained the following non-linear equation

\[
\sum_{i=1}^{n} x_i^\gamma \ln x_i - \frac{1}{\gamma} \sum_{i=1}^{n} x_i^\gamma = 0 \quad (3.6)
\]

This can be solved numerically to obtain the estimate of $\gamma$ by using the Newton-Raphson method. The MLE estimate for $\beta$ is given by

\[
\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} x_i^\gamma. \quad (3.12)
\]

### 3.5.2. Gamma Distribution

The Gamma distribution is a two-parameter family of continuous probability distributions and has a right skewed distribution. It is very useful for risk analysis modeling, particularly, for claims size modeling. It is an extension of the exponential distribution. It is expressed in terms of the gamma function which is defined by;

\[
\Gamma(x) = \int_{0}^{\infty} t^{x-1} e^{-t} dt, x > 0 \quad (3.7)
\]

The probability distribution function (PDF) with two parameters is given by;

\[
f_x(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x \geq 0, \alpha, \beta > 0 \quad (3.8)
\]

Where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter.

The likelihood function for the gamma distribution function is given by;

\[
L(\alpha, \beta) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i}. \quad (3.9)
\]

The log-likelihood function is

\[
l(\alpha, \beta) = n(\alpha \ln \beta - \ln \Gamma(\alpha)) + (\alpha - 1) \sum_{i=1}^{n} \ln x_i - \beta \sum_{i=1}^{n} x_i \quad (3.10)
\]

Thus, to determine the parameter estimates, we equate the derivatives of the log-likelihood function to zero and solve the following equations

\[
\frac{\partial l(\alpha, \beta)}{\partial \alpha} = n \left\{ \ln \beta - \frac{d}{d \alpha} \ln \Gamma(\hat{\alpha}) \right\} + \sum_{i=1}^{n} \ln x_i = 0 \quad (3.11)
\]

Thus

\[
\frac{\partial l(\alpha, \beta)}{\partial \beta} = n \frac{\alpha}{\beta} - \sum_{i=1}^{n} x_i = 0 \quad (3.12)
\]

Substituting $\beta = \frac{\hat{\alpha}}{X}$ in the equation results in the following relationship for $\hat{\alpha}$

\[
n \left\{ \ln \beta - \ln X - \frac{d}{d \alpha} \ln \Gamma(\hat{\alpha}) \right\} + \ln x \right\} \quad (3.13)
\]

This result is a non-linear equation in $\alpha$ that cannot be solved in a closed form. This can be solved numerically using the root-finding methods.

### 3.5.3. Log-normal Distribution

The lognormal distribution is applicable to random variables that are constrained by zero but have a few very large values. The resulting distribution is asymmetrical and positively skewed. The application of a logarithmic transformation to the data can allow the data to be approximated by the symmetrical normal distribution, although the absence of negative values may limit the validity of this procedure.

Log normal distribution ($log \sim N(\mu, \sigma^2)$). Parameter; $-\infty < \mu < \infty, \sigma > 0$. The probability distribution is given by;

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{\ln x - \mu}{\sigma} \right)^2 \right\}, x > 0 \quad (3.14)
\]

With parameters: location $\mu$, scale $\sigma > 0$.

The likelihood function for the lognormal distribution is

\[
L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ -\frac{1}{2} \left( \frac{\ln x_i - \mu}{\sigma} \right)^2 \right\} \quad (3.15)
\]

Therefore the log-likelihood function is given by

\[
\ln L(\mu, \sigma) = -n \ln \sigma - \frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^{n} \left( \ln x_i - \frac{\left( \ln x_i - \mu \right)^2}{2\sigma^2} \right) \quad (3.16)
\]

The parameter estimator’s $\hat{\mu}$ and $\hat{\sigma}$ for the parameters $\mu$ and $\sigma$ can be determined by equating the derivatives of the log-likelihood function to zero and solve the following two equations

\[
\frac{\partial \ln L(\mu, \sigma)}{\partial \mu} = 0 \quad (3.17)
\]

And

\[
\frac{\partial \ln L(\mu, \sigma)}{\partial \sigma} = 0 \quad (3.18)
\]

The resulting estimates are $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \ln x_i$ and $\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} \left( \ln x_i - \mu \right)^2$ respectively.
3.5.4. Pareto Distribution

With parameter $\alpha > 0$ and $\gamma > 0$, is given by the distribution

$$f_x(x, \alpha, \gamma) = \frac{\alpha \gamma^\alpha}{(x + \gamma)^{\alpha+1}}, \ x \geq 0 \quad (3.19)$$

Where $\alpha > 0$ represent the shape, and $\gamma > 0$ the scale parameter.

The likelihood function is

$$L(a, \gamma) = \prod_{i=1}^{n} \frac{\alpha \gamma^a}{(x_i + \gamma)^{\alpha+1}}, \ 0 < \gamma \leq \min \{x_i\}, \ a > 0 \quad (3.20)$$

The log-likelihood function

$$l(a, \gamma) = n \ln(a) + a \ln(\gamma) - (a + 1) \sum_{i=1}^{n} \ln(x_i) \quad (3.21)$$

It is important to note that the best way to maximize the log-likelihood function is by adjusting as follows, $\gamma = \min \{x_i\}$ such that $\gamma$ cannot be larger than the smallest value of $x$ in the data. Thus, to determine the parameter estimate $\alpha$ equates the derivative of the log-likelihood function to zero we solve the following equations

$$\frac{\partial \ln(\gamma, a)}{\partial a} = \frac{n}{a} \ln(\gamma) - \sum_{i=1}^{n} \ln(x_i) = 0 \quad (3.22)$$

Hence, we obtain $\hat{a} = \frac{n}{\sum_{i=1}^{n} \ln \left( \frac{x_i}{\gamma} \right)}$.

3.6. Review Frequency Distributions

3.6.1. Binomial Distribution

The binomial distribution is a popular discrete distribution for modeling count data. Given a portfolio of $n$ independent insurance policies, let $X$ denote a binomially distributed random variable that represents the number of policies in the portfolio that result in a claim. The claim count variable $X$ can be said to follow a binomial distribution with parameters $n$ and $p$, where $n$ is a known positive integer representing the number of policies in the portfolio and $p$ is the probability that there is at least one claim on an individual policy. The probability distribution function of $X$ is defined as:

$$f_x(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad (3.23)$$

for $x = 0, 1, ..., n$ and $0 \leq p \leq 1$

The expected value of the binomial distribution is $np$ and variance $np(1-p)$. Hence, the variance of the binomial distribution is smaller than the mean. The corresponding binomial likelihood function is

$$L(p) = \binom{n}{x} p^x (1 - p)^{n-x} \quad (3.24)$$

Therefore the log-likelihood function is

$$l(p) = k + x \ln(p) + (n - x) \ln(1 - p) \quad (3.25)$$

where $k$ is a constant that does not involve the parameter $p$. Therefore, to determine the parameter estimate $\hat{p}$ we obtain the derivative of the log-likelihood function with respect to the parameter and equate to zero.

$$\frac{\partial l(p)}{\partial p} = \frac{x}{p} - \frac{n - x}{1 - p} = 0 \quad (3.26)$$

Solving the above equation gives the MLE. Thus, the MLE is $\hat{p} = \frac{x}{n}$.

3.6.2. Poisson Distribution

The Poisson distribution is a discrete distribution for modeling the count of randomly occurring events in a given time interval. Let $X$ be the number of claim events in a given interval of time and $\lambda$ is the parameter of the Poisson distribution representing the mean number of claim events per interval. The probability of recording $x$ claim events in a given interval is given by

$$f_x(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \text{ for } x = 1, 2, 3, ..., \quad (3.27)$$

A Poisson random variable can take on any positive integer value. Unlike the Binomial distribution which always has a finite upper limit. In general, the expected value and variance functions of the Poisson distribution are both equal to $\lambda$. The Poisson likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! ... x_n!} \quad (3.28)$$

Therefore, the log-likelihood function is

$$l(\lambda) = \sum_{i=1}^{n} x_i \ln \lambda - n \lambda \quad (3.29)$$

Differentiating the log-likelihood function with respect to $\lambda$, ignoring the constant term that does not depend on $\lambda$ and equating to zero

$$\frac{\partial l(\lambda)}{\partial \lambda} = 1 - \lambda n = 0 \quad (3.30)$$

Solving the above Equation gives the MLE. Thus, the MLE is $\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$.

3.6.3. Negative Binomial

The negative binomial distribution is a generalization of the geometric distribution. Consider a sequence of independent Bernoulli trials, with common probability $p$, until we get $r$ successes. If $X$ denotes the number of failures which occur before the $r$th success, then $X$ has a negative binomial distribution given by

$$f_x(x, r, p) = \binom{x + r - 1}{r - 1} p^x (1 - p)^r; \ x = 0, 1, 2, ..., \quad (3.31)$$
The negative binomial distribution has an advantage over the Poisson binomial distribution in modeling because it has two positive parameters \( p > 0 \) and \( r > 0 \). The most important feature of this distribution is that the variance is bigger than the expected value. A further significant feature in comparing these three discrete distributions is that the binomial distribution has a fixed range but the negative binomial and Poisson distributions both have infinite ranges. The likelihood function is

\[
L(p, x, r) = \prod_{i=1}^{n} \left( \frac{x_i + r - 1}{r - 1} \right)^p \left( 1 - p \right)^{x_i}
\]

\[
= \prod_{i=1}^{n} \left( \frac{x_i + r - 1}{r - 1} \right)^{nr} \left( 1 - p \right)^{\sum x_i}
\]

The log-likelihood function is obtained by taking logarithms

\[
\ln \left( \frac{x_i + r - 1}{r - 1} \right) + n r \ln p + \sum x_i \ln (1 - p) = \sum_{i=1}^{n} \ln \left( \frac{x_i + r - 1}{r - 1} \right) + n r \ln p + \sum x_i \ln (1 - p)
\]

Taking the derivative with respect to \( p \) and equating to zero.

\[
\frac{\partial \ln}{\partial p} = \frac{nr}{p} \sum \frac{x_i}{1 - p} = 0
\]

The resulting MLE estimate is \( \hat{p} = \frac{nr}{nr - X} \) therefore, the negative binomial random variable can be expressed as a sum of \( r \) independent, identically distributed (geometric) random variables.

3.7. Checking the Goodness of Fit

3.7.1. Chi-Square Goodness-of-Fit Test

The Chi-Square goodness of fit test is used to test the hypothesis that the distribution of a set of observed data follows a particular distribution. The Chi-square statistic measures how well the expected frequency of the fitted distribution compares with the observed frequency of a histogram of the observed data. The Chi-square test statistic is:

\[
\chi^2 = \sum_{j} \frac{(O_j - E_j)^2}{E_j}
\]

Where \( O_j \) the observed number of cases in interval \( j \) is, \( E_j \) is the expected number of cases in interval \( j \) of the specified distribution and \( k \) is the number of intervals the sample data is divided into. The test statistic approximately follows a Chi-Squared distribution with \( k - p - 1 \) degrees of freedom where \( p \) is the number of parameters estimated from the (sample) data used to generate the hypothesized distribution. A G-o-F test considers the null hypothesis that the sample data are taken from a population with the underlying distribution \( F(x) \) is rejected if the p-value is smaller than the criterion value at a given significance level \( \alpha \), at 5%.

3.7.2. The Anderson-Darling Test

The Anderson-Darling (A-D) test is an alternative to other statistical tests for testing whether a given sample of data is drawn from a specified probability distribution. This test gives more weight to the tails than Kolmogorov-Smirnov test which tends to be more sensitive near the center of the distribution than at the tails. Anderson-Darling (A-D) is given by the formula;

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

Where \( x_{(1)} < \cdots < x_{(n)} \) is the ordered sample of size \( n \) and \( F(x) \) is the underlying theoretical distribution to which the sample is compared. An A-D test is used to select the continuous claims severity distributions. Therefore, G-o-F tests consider null hypothesis that the sample data are taken from a population with the underlying distribution \( F(x) \) is rejected if the p-value is smaller than the criterion value at a given significant level \( \alpha \), at 5%.

3.8. Criteria for Choosing One Distribution for the Claims’ Data

The values of Akaike’s Information Criterion AIC were used to select the best model. Bayesian Information Criteria BIC can also be used to select best model but it has some limitation compared to AIC. Unlike AIC, BIC penalizes free additional parameters more strongly. AIC is good for making asymptotically equivalent to cross validation. On the contrary, BIC is good for consistent estimation. AIC tries to find unknown model that has high dimensional reality while BIC comes across only true models. For the above reasons AIC was the best for making inference from a sampled group of models. Therefore, the model with the smallest value of AIC was selected because this model is estimated to be closest to the unknown truth among models considered.

The AIC criterion is defined by:

\[
AIC = -2\ln(likelihood) + 2K
\]

where \( K \) is the parameter.

4. Data Analysis and Results

4.1. Frequency Claims Models

Frequency distribution was performed based on three models including binomial, Poisson and negative binomial distribution models. The results as shown in Table 1 below suggest that the p-values for the z-statistic in all the three models are less than 0.05 indicating that the parameter estimates for the three variables including the intercept are highly significant based on these models. Therefore, a significant relationship exists between the predictor variables and the response. The claims based on binomial, negative binomial and Poisson distributions models.

Moreover, the distributions of the residuals appear to be symmetrical in all the three models hence models
predict all points closer to the actual observed points. The findings show that the null deviance for Poisson, Binomial and Negative Binomial distributions were found to be 1778.3, 1225.5 and 1202.08 respectively hence Negative Binomial distribution was considered to fit well as compared to Poisson distribution and Binomial distribution since it has a smaller value of the null deviance.

Parameter estimation was also done using Maximum Likelihood Estimation MLE method. The results show that LLF for Poisson, Negative Binomial and Binomial were found to be -1314.8528, -598.9715 and -1278.6367 respectively. From the results, it can be shown that the best model is Negative Binomial distribution with the highest value of LLF. This is further confirmed with AIC output where Negative Binomial distribution was found to be the best model since it has a smaller AIC value of 1205.9 as compared to both Poisson and Binomial distributions with 2637.7 and 2567.3 respectively. Hence, Negative Binomial distributions was considered as the best model for forecasting frequency of insurance claims. The results are summarized in Table 1 below.

### 4.2. Severity Claims Models

Severity of the claims was performed based on four models including Gamma, Weibull, Log normal and Pareto distributions. The results as shown in Table 2 below indicates that the p-values for the z-statistic in all the four models are less than 0.05 indicating that the parameter estimates for the five variables including the intercept are highly significant based on these models. Therefore, a significant relationship exists between the predictor variables and the response to severity of the claims based on Gamma, Weibull, log normal and Pareto distributions models.

The distributions of the residuals appear to be strongly symmetrical hence all models predict all points close to the actual observed points. The output of the fitted Null deviance for gamma distribution is 864.19 and residual deviance is 273.52. However, since Weibull, Pareto and Log normal distributions have no null deviance and residual deviance, Gamma distribution may be considered as appropriate model. However, the results of LLF based on MLE for Gamma, Weibull, Lognormal and Pareto distributions were found to be -10842.7846, -10986.1643, -10690.17 and -10346 respectively. From the results of parameter estimation, Pareto distribution was considered as the best model since it has highest value of LLF. This is further confirmed from the output of Akaike’s Information Criterion (AIC), where Gamma, Weibull, Log normal and Pareto distributions were found to have AIC values of 21703, 21988.96, 21396.33 and 20720.01 respectively thus suggesting that Pareto distribution with lower AIC value was considered the best model and for this reason, Pareto model was used in forecasting insurance severity claims. The results are presented in Table 2 below.

| Distribution | Variable | Estimate  | Standard Error | Z value  | Pr(>|z|) |
|--------------|----------|-----------|----------------|----------|----------|
| Intercept:   |          | -0.59985  | 0.06830        | -8.782   | <0.0001  |
| Mvr pts:     |          | 0.19242   | 0.01282        | 5.005    | <0.0001  |
| Car use num: |          | -0.20377  | 0.06922        | -2.944   | 0.00324  |
| Revoked num: |          | 0.38514   | 0.08686        | 4.434    | 9.26e-06 |
| LLF:         |          | -1314.852 |                |          |          |
| Deviance Residuals: | | 1Q | -2.3735 | -1.0478 | -0.9463 | 0.457 | 3.8763 |
| Min           |          |            | Median         | 3Q       | Max      |
| Null Deviance:|          | 1778.3     |                |          |          |
| Residual deviance: | | 1350.3 |            |          |          |
| AIC:          |          | 2637.7     |                |          |          |
| Distribution | Variable | Estimate  | Standard Error | Z value  | Pr(>|z|) |
|--------------|----------|-----------|----------------|----------|----------|
| Intercept:   |          | -1.22948  | 0.13814        | -8.900   | <0.0001  |
| Mvr pts:     |          | 0.54292   | 0.04017        | 13.516   | <0.0001  |
| Car use num: |          | -0.33922  | 0.14705        | -2.307   | 0.02107  |
| Revoked num: |          | 0.6753    | 0.20675        | 3.248    | 0.00116  |
| LLF:         |          | -598.9715 |                |          |          |
| Deviance Residuals: | | 1Q | -2.3812 | -0.7828 | -0.6152 | 0.9343 | 1.8751 |
| Min           |          |            | Median         | 3Q       | Max      |
| Null Deviance:|          | 1202.08    |                |          |          |
| Residual Deviance: | | 1197.9 |            |          |          |
| AIC:          |          | 1205.9     |                |          |          |
| Distribution | Variable | Estimate  | Standard Error | Z value  | Pr(>|z|) |
|--------------|----------|-----------|----------------|----------|----------|
| Intercept:   |          | -0.63786  | 0.08649        | -7.375   | 1.6413   |
| Mvr pts:     |          | 0.21077   | 0.01807        | 11.663   |          |
| Car use num: |          | -0.22377  | 0.08979        | -2.492   | 0.0127   |
| Revoked num: |          | 0.43869   | 0.11800        | 3.718    | 0.0002   |
| LLF:         |          | -1278.636 |                |          |          |
| Deviance Residuals: | | 1Q | -1.6681 | -0.9485 | -0.8605 | 0.225 | 2.9519 |
| Min           |          |            | Median         | 3Q       | Max      |
| Null Deviance:|          | 1225.5     |                |          |          |
| Residual Deviance: | | 1064.5 |            |          |          |
| AIC:          |          | 2567.3     |                |          |          |
4.3. Goodness of Fit of the Models

4.3.1 Chi-squared Test for Frequency Models

Chi-square goodness-of-fit test was performed in order to select the most appropriate claims frequency distribution among the fitted discrete probability distributions. The results as shown in Table 3 below suggest that the p-values for all the distributions including Poisson, Binomial and Negative Binomial distributions were less than 0.05 hence statistically significant. This confirms the above findings that Poisson, negative binomial and Binomial distributions were appropriate distributions for modeling frequency claims.

| Distribution | Variable         | Estimate | Standard Error | Z value | Pr(>|z|)  |
|--------------|-----------------|----------|----------------|---------|-----------|
| (Intercept):1|                 | 7.529649 | 0.109630       | 68.682  | <0.0001  |
| (Intercept):2|                 | 0.579964 | 0.023509       | 24.670  | <0.0001  |
| Age:         |                 | 0.021415 | 0.001214       | 17.635  | <0.0001  |
| BMI:         |                 | 0.015243 | 0.002919       | 5.222   | 1.77e-07 |
| Children:    |                 | 0.076723 | 0.014193       | 5.406   | 6.45e-08 |
| Car age:     |                 | 0.010673 | 0.003021       | 3.533   | 0.000411 |
| Smoker_num:  |                 | 1.355789 | 0.041258       | 32.861  | <0.0001  |
| Region_num:  |                 | 0.045850 | 0.015633       | 2.932   | 0.003373 |
| LLF:         |                 | -10986.2 |               |         |           |
| Pearson Residuals: |       | Min |              | Median | 3Q | Max       |
| Loglink(scale): |              | -1.20 | -0.5734       | -0.285  | 0.01559  | 11.152    |
| Loglink(shape): |           | -28.23 | 0.0698        | 0.470   | 0.68577  | 0.784     |
| AIC:         |                 | 21988.96 |              |         |           |
|                | Distribution | Variable | Estimate | Standard Error | Z value | Pr(>|z|)  |
|                |            |           |         |                 |         |           |
| (Intercept):1 |            | 7.113482 | 0.13072 | 54.415          | <0.0001 |
| (Intercept):2 |            | 0.026255 | 0.00145 | 19.488          | <0.0001 |
| Age:         |            | 0.047244 | 0.00348 | 4.225           | 2.59e-05|
| BMI:         |            | 0.092466 | 0.01694 | 5.457           | 6.00e-08|
| Children:    |            | 0.008424 | 0.00360 | 2.335           | 0.01970 |
| Car age:     |            | 1.497256 | 0.04926 | 30.395          | <0.0001 |
| Smoker_num:  |            | 0.049702 | 0.01866 | 2.66            | 0.00786 |
| Region_num:  |            | -10842.7 | Min       | 1Q              | Median  | 3Q | Max       |
| LLF:         |            | Min | -1.14523 | -0.34397 | -0.14475 | -0.00009 | 2.65254   |
| Deviance Residuals: |       | Null Deviance: | 864.19 | | |
| Log normal   |            | AIC: | 273.52 |              |         |           |
| Pareto       |            | AIC: | 21703.0 |              |         |           |
|                | Distribution | Variable | Estimate | Standard Error | Z value | Pr(>|z|)  |
|                |            |           |         |                 |         |           |
| (Intercept):1 |            | 6.800027 | 0.0865739 | 78.546          | <0.0001 |
| (Intercept):2 |            | 0.814686 | 0.0213201 | -38.21          | <0.0001 |
| Age:         |            | 0.034299 | 0.0009602 | 35.714          | <0.0001 |
| BMI:         |            | 0.013258 | 0.0023080 | 5.749           | 8.98e-09|
| Children:    |            | 0.103492 | 0.0112224 | 9.222           | <0.0001 |
| Car age:     |            | 0.004926 | 0.0023888 | 2.063           | 0.0392  |
| Smoker_num:  |            | 1.550971 | 0.0326230 | 47.542          | <0.0001 |
| Region_num:  |            | 0.051703 | 0.0123615 | 4.183           | 2.88e-05|
| LLF:         |            | -10690.7 | Min       | 1Q              | Median  | 3Q | Max       |
| Pearson Residuals: |       | Meanlog: | -2.489 | -0.4491 | -0.099 | 0.168 | 4.784 |
| Loglink(sdlg): |           | -0.707 | -0.6945 | -0.637 | -0.177 | 0.1547 |
| LLF:         |            | -10690.7 | Max       |                |         |           |
| AIC:         |            | 21396.33 |            |         |           |
| Pareto       |            | AIC: | 20720.01 |              |         |           |

Table 2. Severity Claims Models
4.4. Expected Amount as Per the Number of Claims Per Risk

This section presents the results for the post model selection fit to affirm the selected models. [17], argued that the central problem in analysis is the kind of model one needs to use for making inferences from the claim dataset which is known as the model selection problem. As already shown and proven in both the selection criterion and the goodness of fit test above, the Pareto model emerged as the best fit for severity claims since it has least AIC and largest LLF. On the other hand, Negative Binomial distribution model emerged as the best fit for frequency claims since it has least AIC and largest LLF. It was also necessary to use goodness of fit criterion in order to select the severity model and frequency model that best fit the claim data. To obtain the expected amount of claims per risk for the following year, it was given by; 

\[ E(X) = e^\mu, \text{ where } \mu \text{ is the mean of the model selected.} \]

The parameter estimates that were used to find expected frequency claim amount per risk is below.

Table 6. Parameter Estimates of Severity Models

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>Estimates</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>( \hat{\alpha} )</td>
<td>163.35309</td>
<td>9.155</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>0.0560441</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>( \beta )</td>
<td>0.8635716</td>
<td>4.610</td>
</tr>
<tr>
<td></td>
<td>( \gamma )</td>
<td>4.6071778</td>
<td></td>
</tr>
<tr>
<td>Log normal</td>
<td>( \mu )</td>
<td>2.1969107</td>
<td>2.197</td>
</tr>
<tr>
<td></td>
<td>( \sigma )</td>
<td>0.8532774</td>
<td></td>
</tr>
<tr>
<td>Pareto</td>
<td>( \alpha )</td>
<td>7.8257900</td>
<td>8.718</td>
</tr>
<tr>
<td></td>
<td>( \gamma )</td>
<td>7.6042440</td>
<td></td>
</tr>
</tbody>
</table>

Since Pareto distribution was the best severity model, with its parameters \( \hat{\alpha} \) and \( \gamma \), mean of Pareto model is 8.7183. Hence, the expected claims amount per risk for the following year was Ksh 6,113.70.

5. Conclusion

The Pareto model was found to be the best fit model for the severity claim data among the four models tested. Anderson Darling test, AIC and LLF were used to measure the fitness of the model. The three tools confirmed, affirmed and reaffirmed that the Pareto model is the best fit model among the four models fitted. The expected claim amount per risk to be paid in the following year was estimated to be Ksh 6,310.70. On the other hand, Negative Binomial distribution model was found to best fit model for frequency claims data among models tested. The Chi-square test, AIC and LLF were used to measure the fitness, the three tools confirmed, affirmed and reaffirmed that Negative Binomial model is the best fit model among the three models fitted. The expected frequency claim amount per risk for the following year was estimated to be 1.07372±1. After a systematic study and going through actuarial modeling process by using comprehensive claim amount paid to policy holder and the frequency claims, the research concluded that Pareto model was the appropriate model for modeling comprehensive insurance claim severity with a heavier tail and Negative Binomial distribution model was the best appropriate model for modeling frequency claims of insurance company.

6. Recommendations

Since motor insurance is the fastest growing industry with overwhelming number of claims which are huge and unexpected, it is recommended that insurance companies use Negative Binomial model and Pareto model in forecasting of claims. This is because it assists insurance companies in their premium loading hence avoiding financial insolvency which threatens to place insurance companies under receivership. For further research, all models should be considered, which is critical. It’s also
recommended that companies should use their own claims experience to make necessary adjustments to the models. This will allow for anticipated changes in the portfolios and for companies’ specific financial objectives. These proposed claims distributions will also be useful to insurance regulators in their own assessment of required reserve levels for various companies and in checking for solvency. It’s also recommended that further analysis should be done using models such as the zero-inflated models. This is because of large number of zeros in claims frequency data, hence there is need to consider other distributions such as the zero-truncated Poisson or zero-truncated negative binomial and zero-modified distributions to model this unique phenomenon.

References


© The Author(s) 2020. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).