Modeling the Frequency and Severity of Auto Insurance Claims Using Statistical Distributions

Cyprian Ondieki Omari, Shalyne Gathoni Nyambura, Joan Martha Wairimu Mwangi

Department of Statistics and Actuarial Science, Dedan Kimathi University of Technology, Nyeri, Kenya
Email: cyomari@dkut.ac.ke, nshalyne@gmail.com, jmjoanmwangi@gmail.com

Abstract

Claims experience in non-life insurance is contingent on random eventualities of claim frequency and claim severity. By design, a single policy may possibly incur more than one claim such that the total number of claims as well as the total size of claims due on any given portfolio is unpredictable. For insurers to be able to settle claims that may occur from existing portfolios of policies at some future time periods, it is imperative that they adequately model historical and current data on claims experience; this can be used to project the expected future claims experience and setting sufficient reserves. Non-life insurance companies are often faced with two challenges when modeling claims data; selecting appropriate statistical distributions for claims data and establishing how well the selected statistical distributions fit the claims data. Accurate evaluation of claim frequency and claim severity plays a critical role in determining: An adequate premium loading factor, required reserve levels, product profitability and the impact of policy modifications. Whilst the assessment of insurers' actuarial risks in respect of their solvency status is a complex process, the first step toward the solution is the modeling of individual claims frequency and severity. This paper presents a methodical framework for choosing a suitable probability model that best describes automobile claim frequency and loss severity as well as their application in risk management. Selected statistical distributions are fitted to historical automobile claims data and parameters estimated using the maximum likelihood method. The Chi-square test is used to check the goodness-of-fit for claim frequency distributions whereas the Kolmogorov-Smirnov and Anderson-Darling tests are applied to claim severity distributions. The Akaike information criterion (AIC) is used to choose between competing distributions. Empirical results indicate that claim severity data is better modeled using heavy-tailed and skewed distributions. The lognormal distribution is selected as the best distribution to model the claim size while negative binomial and geometric dis-

tributions are selected as the best distributions for fitting the claim frequency data in comparison to other standard distributions.

Keywords

Actuarial Modeling, Claim Frequency, Claim Severity, Goodness-of-Fit Tests, Maximum Likelihood Estimation (MLE), Loss Distributions

1. Introduction

The development of insurance business is driven by general demands of the society for protection against various types of risks of undesirable random events with a significant economic impact. Insurance is a process that entails the provision of an equitable method of offsetting the risk of a likely future loss with a payment of a premium. The underlying concept is to create a fund to which the insured members contribute predetermined amounts of the premium for given levels of loss. When the random events that policyholders are protected against occur giving rise to claims then claims are settled from the fund. The characteristic feature of such an arrangement is that the insured members are faced with a homogeneous set of risks. The positive aspect of forming such communities is the pooling together of risks which enables members to benefit from the weak law of large numbers.

In the non-life insurance industry, there is increased interest in the automobile insurance because it requires the management of a large number of risk events. These include cases of theft and damage to vehicles due to accidents or other causes as well as the extent of damage and parties involved [1]. General insurance companies deal with large volumes of data and need to handle this data in the most efficient way possible. According to Klugman *et al.* [2], actuarial models assist insurance companies to handle these vast amounts of data. The models are intended to represent the real world of insurance problems and to depict the uncertain behavior of future claims payments. This uncertainty necessitates the use of probability distributions to model the occurrence of claims, the timing of the settlement and the severity of the payment, that is, the amount paid.

One of the main challenges that non-life insurance companies face is to accurately forecast the expected future claims experience and consequently determining an appropriate reserve level and premium loading. The erratic results often reported by non-life insurance companies lead to the pre-conclusion that it is virtually impossible to conduct the business on a sound basis. Most non-life insurance companies base their estimations of claim frequency and severity on their own historical claims data. This is sometimes complemented with data from external sources and is used as a base for managerial decisions [3]. The management, with the guidance of actuaries, ensures that the insurance compa-

ny is run on a sound financial basis by charging the appropriate premium corresponding to the specific risks and also maintaining a suitable reserve level.

A general approach to claims data modeling is to consider separately claim count experience from the claim size experience. Both claim frequency and severity random variables; hence there is a risk that future claims experience will deviate from past eventualities. It is therefore imperative that appropriate statistical distributions are applied in modeling of claims data variables. Although the empirical distribution functions are useful tools in analyzing claims processes, it is always convenient to model claims data by fitting a probability distribution with mathematically tractable features.

Statistical modeling of claims data has gained popularity in the past recent years, particularly in the actuarial literature all in an attempt to address the problem of premium setting and reserves calculations. Hogg and Klugman [4] provide an extensive description of the subject of insurance loss models. Loss models are useful in revealing claim-sensitive information to insurance companies and giving insight into decisions on required premium levels, premium loading reserves and assessing the profitability of insurance products. The main components of aggregate claims of a non-life insurance company are frequency/count and size/severity. Claim frequency refers to the total count of claims that occur on a block of policies, while claim severity or claim size is the monetary amount of loss on each policy or on the portfolio as a whole. The claims experience random variables are often assumed to follow certain probability distributions, referred to as loss distributions [5]. Loss distributions are typically skewed to the right with relatively high right-tail probabilities. Various studies available literature describes these probability distributions as being long-tailed or heavy-tailed.

When assessing portfolio claim frequency, it is often found that certain policies did not incur any claims since the insured loss event did not occur to the insured. Such cases result in many zero claim counts such that the claim frequency random variable takes the value zero with high probability. Antonio *et al.* [6] present the Poisson distribution as the modeling archetype of claim frequency. In available economic literature, considerations have been made for models of count data on claims frequency that allow for excess zeros. For instance, Yip and Yau [7] fit a zero-inflated count model to their insurance data. Boucher *et al.* [8] considered zero-inflated and hurdles models with application to real data from a Spanish insurance company. Mouatassim and Ezzahid [9] analyzed the zero-inflated models with an application to private health insurance data.

Modeling claim frequency and claim severity of an insurance company is an essential part of insurance pricing and forecasting future claims. Being able to forecast future claim experience enables the insurance company to make appropriate prior arrangements to reduce the chances of making a loss. Such arrangements include setting a suitable premium for the policies and setting aside

money required to settle future claims (reserves). A suitable premium level for an individual policy, with proper investment, should be enough to at least cover the cost of paying a claim on the said policy. Adequate reserves set aside should enable the insurance company to remain solvent, such that it can adequately settle claims when they are due, and in a timely manner. Bahnemann [10] argued that the number of claims arising in a portfolio is a discrete quantity which makes discrete standard distributions ideal since their probabilities are defined only on non-negative integers. Further, they showed claims severity has support on the positive real line and tends to be skewed to the right. They modeled claim severity as a non-negative continuous random variable. Several actuarial models for claims severity are based on continuous distributions while claim frequencies are modeled using discrete probability distributions. The log-normal and gamma distributions are among the most commonly applied distributions for modeling claim severity. Other distributions for claim size are the exponential, Weibull, and Pareto distributions. Achieng [2] modeled the claim amounts from First Assurance Company limited, Kenya for motor comprehensive policy. The log-normal distribution was chosen as the appropriate model that would provide a good fit to the claims data. The Pareto distribution has been shown to adequately mimic the tail-behavior of claim amounts and thus provides a good fit.

The objective of this paper is to present a hypothetical procedure for selecting probability models that approximately describe auto-insurance frequency and severity losses. The method employs fitting of standard statistical distributions that are relatively straightforward to implement. However, the commonly used distributions for modeling claims frequency and severity may not appropriately describe the actual claims data distributions and therefore may require modifications of standard distributions. Most insurance companies rely on existing models from developed markets that are normally customized to model their claims frequency and severity distributions. In practice, specific models from different markets cannot be generalized to appropriately model the claims data of all insurance companies as different markets have unique characteristics.

This paper gives a general methodology in modeling the claims frequency and severity using the standard statistical distributions that may be used as starting point in modeling claims frequency and severity. The actuarial modeling techniques are utilized in an attempt to fit an appropriate statistical probability distribution to the general insurance claims data and select the best fitting probability distribution. In this paper, a sample of the automobile portfolio datasets obtained from the insurance Data package in R with variables; Auto Collision, data Car, and data Ohlsson are used. These data are chosen since they are complete, freely and easily available in R statistical software package. However, any other appropriate dataset especially company-specific data may also be used. The parameter estimates for fitted models are obtained using the Maximum Likelihood Estimation procedure. The Chi-square test is used to check the goodness-of-fit for claim frequency models whereas the Kolmogorov-Smirnov and

Anderson-Darling tests are used for the claim severity models. The AIC and BIC criteria are used to choose between competing distributions.

The remainder of the paper is organized as follows: Section 2 discusses the statistical modeling procedure for claims data. Section 3 presents the empirical results of the study. Finally, Section 4 concludes the study.

2. Statistical Modeling of Claims Data

The statistical modeling of claims data involves the fitting of standard probability distributions to the observed claims data. Kaishev and Krachunov [11] suggested the following four-step statistical procedure for fitting an appropriate statistical distribution to claims data:

1) Selection of the claims distribution family

2) Estimation of parameters of the chosen fitted distributions

3) Specification of the criteria to select the appropriate distribution from the family of distributions.

4) Testing the goodness of fit of the approximate distributions.

2.1. Selection of Claims Distributions

The initial selection of the models is based on prior knowledge on the nature and form of claim's data. Claim frequency is usually modeled using non-negative discrete probability distributions since the number of claims is discrete and non-zero. Claim severity is known to be best modeled using non-zero continuous distributions which are skewed to the right and have heavy tails. Kaas, *et al.* [12] attributes this to the fact that extremely large claim values often occur in the upper-right tails of the distribution. Prior knowledge of claims experiences in non-life insurance coupled with descriptive analysis of prominent features of the claims data and graphical techniques a used to guide the selection of the initial approximate probability distributions of claim size and frequency. Merz and Wüthrich [13] noted that majority of claims data arising from the general insurance industry are positively skewed and have heavy tails. They argued that statistical distributions which exhibit these characteristics may be appropriate for modeling such claims.

This study proposes a number of standard probability distributions that could be used to approximate the distributions for claim amount and claim count random variables. The binomial, geometric, negative-binomial and Poisson distributions are considered for modeling claim frequency as they are discrete. On the other hand, five standard continuous probability distributions are proposed for modeling claim severity. These are the exponential, gamma, Weibull, Pareto and the lognormal distributions. The probability distribution functions along with their parameters estimation and their respective properties are discussed subsequently.

2.2. Estimation of Parameters

In this paper, the parameters of the selected claim distributions are estimated

using the maximum likelihood estimation (MLE) technique. The MLE is a commonly applied method of estimation in a variety of problems. It often yields better estimates compared to other methods like; least-squares estimation (LSE), the method of quantile and method of moments especially when the sample size is large. Boucher *et al.* [14] argued that the MLE approach fully utilizes all the information about the parameters contained in the data and yields highly flexible estimator with better asymptotic properties.

Suppose $X_1, X_2, \cdots, X_n$ is a random sample of independent and identically distributed (iid) observations drawn from an unknown population. Let $X = x$ denote a realization of a random variable or vector X with probability mass or density function $f(x;\theta)$; where θ is a scalar or a vector of unknown parameters to be estimated. The objective of statistical inference is to infer θ from the observed data. The MLE involves obtaining the likelihood function of a random variable. The likelihood function $L(\theta)$ is the probability mass or density function of the observed data x, expressed as a function of the unknown parameter θ. Given that $X_1, X_2, \cdots, X_n$ have a joint density function $f(X_1, X_2, \cdots, X_n | \theta)$ for every observed sample of independent observations $\{x_i\}, i = 1, 2, \cdots, n$, the likelihood function of θ is defined by

$$L(\theta) = L(\theta \mid x_1, x_2, \cdots, x_n) = f(x_1, x_2, \cdots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta) \qquad (1)$$

The principle of maximum likelihood provides a means of choosing an asymptotically efficient estimator for a parameter or a set of parameters $\hat{\theta}$ as the value for the unknown parameter that makes the observed data "most probable". The maximum likelihood estimate (MLE) $\hat{\theta}$ of a parameter θ is obtained through maximizing the likelihood function $L(\theta)$.

$$\hat{\theta}(x) = \arg \max_{\theta} L(\theta) \qquad (2)$$

Since the logarithm of the likelihood function is a monotonically non-decreasing function of X, maximizing $L(\theta)$ is equivalent to maximizing $\log L(\theta)$. Therefore, the log of the likelihood function denoted as $l(\theta)$ is defined as

$$l(\theta) = \ln L(\theta) = \ln \prod_{i=1}^{n} f(x_i | \theta) = \sum_{i=1}^{n} \ln f(x_i | \theta) \qquad (3)$$

The probability distribution functions of the selected claim distributions along with their maximum likelihood estimates for the parameters are given as follows:

2.2.1. 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 on the portfolio and p is the probabili-

ty that there is at least one claim on an individual policy. The probability distribution function of X is defined as:

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for} \quad x = 0, 1, \cdots, n \quad \text{and} \quad 0 \le p \le 1 \tag{4}$$

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

$$L(p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{5}$$

Therefore the log-likelihood function is

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

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

$$\frac{\partial l(p)}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = 0 \tag{6}$$

Solving the Equation (6) gives the MLE. Thus, the MLE is $\hat{p} = x/n$.

2.2.2. 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 λ 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

$$f_X(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \text{for} \quad x = 1, 2, 3, \cdots \tag{7}$$

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 λ.

The Poisson likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-\lambda}}{x_1! x_2! \cdots x_n!} \tag{8}$$

Therefore, the log-likelihood function is

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

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

$$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0 \tag{9}$$

Solving the Equation (9) gives the MLE. Thus, the MLE is $\hat{\lambda} = \sum_{i=1}^{n} x_i / n$.

2.2.3. Geometric Distribution

The geometric distribution is a discrete distribution that generates the probability that the first occurrence of an event of interest requires x independent trials, where each trial results in either success or failure, and the probability of success in any individual trial is constant. The probability distribution function of the geometric distribution is

$$f_X(x; p) = p(1-p)^x; \; x = 0, 1, 2, \cdots \tag{10}$$

where p is the probability of success, and x is the number of failures before the first success. The geometric distribution is the only discrete memoryless random distribution analogous to the exponential distribution.

The likelihood function for n independent and identically distributed trials is

$$L(p) = \prod_{i=1}^{n} p(1-p)^{x_i} = p^n \prod_{i=1}^{n} (1-p)^{x_i} \tag{11}$$

The log-likelihood function is:

$$l(p) = n \ln p + \ln(1-p) \sum_{i=1}^{n} x_i$$

Thus, to determine the parameter estimate, equate the derivative of the log-likelihood function to zero.

$$\frac{\partial l(p)}{\partial p} = \frac{n}{p} - \frac{1}{1-p} \sum_{i=1}^{n} x_i = 0 \tag{12}$$

Therefore, the MLE is $\hat{p} = \dfrac{1}{1+\bar{x}}$.

2.2.4. Negative Binomial Distribution

The negative binomial distribution is a generality 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

$$f_X(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x; \tag{13}$$
$$x = 0, 1, 2, \cdots$$

The negative binomial distribution has an advantage over the Poisson 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

$$L(p;x,r) = \prod_{i=1}^{n}\binom{x_i+r-1}{r-1}p^r(1-p)^{x_i} = \prod_{i=1}^{n}\binom{x_i+r-1}{r-1}p^{nr}(1-p)^{\sum x_i} \quad (14)$$

The log-likelihood function is obtained by taking logarithms

$$l(p;x,r) = \sum_{i=1}^{n}\ln\binom{x_i+r-1}{r-1} + nr\ln p + \sum_{i=1}^{n}x_i\ln(1-p)$$

Taking the derivative with respect to p and equating to zero.

$$\frac{\partial l}{\partial p} = \frac{nr}{p} - \frac{\sum x_i}{1-p} = 0 \quad (15)$$

The resulting MLE estimate is $\hat{p} = \dfrac{nr}{nr+\overline{x}}$. Therefore, the negative binomial random variable can be expressed as a sum of r independent, identically distributed (geometric) random variables.

2.2.5. Exponential Distribution

The exponential distribution is a continuous distribution that is usually used to model the time until the occurrence of an event of interest in the process. A continuous random variable X is said to have an exponential distribution if its probability density function (pdf) is given by:

$$f_X(x) = \lambda\exp(-\lambda x),\, x > 0 \quad (16)$$

where $\lambda > 0$ is the rate parameter of the distribution. The exponential distribution is a special case for both the gamma and Weibull distribution.

The likelihood function is

$$L(\lambda) = \prod_{i=1}^{n}\lambda\exp(-\lambda x_i) = \lambda^n\exp\left(-\lambda\sum_{i=1}^{n}x_i\right) \quad (17)$$

The log-likelihood functions is therefore

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

The parameter estimator $\hat{\lambda}$ is be obtained by setting the derivative of the log-likelihood function to zero and solving for λ

$$\frac{d\ln l(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n}x_i = 0 \quad (18)$$

The resulting MLE is given by $\hat{\lambda} = 1/\overline{x}$ where $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n}x_i$ is the mean observed time.

2.2.6. Gamma Distribution

The gamma distribution and Weibull distributions are extensions of the exponential distribution. Both of them are expressed in terms of the gamma function which is defined by

$$\Gamma(x) = \int_0^{\infty}t^{x-1}e^{-t}dt,\, x > 0$$

The two-parameter gamma distribution function is given by

$$f_X(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x > 0 \tag{19}$$

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

$$L(\alpha, \beta) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} \tag{20}$$

The log-likelihood functions is

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

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

$$\frac{\partial l(\alpha, \beta)}{\partial \alpha} = n\left(\ln \hat{\beta} - \frac{d}{d\alpha} \ln \Gamma(\hat{\alpha})\right) + \sum_{i=1}^{n} \ln x_i = 0 \tag{21}$$

Hence

$$\frac{\partial l(\alpha, \beta)}{\partial \beta} = n\frac{\hat{\alpha}}{\hat{\beta}} - \sum_{i=1}^{n} x_i = 0 \quad \text{or} \quad \overline{x} = \frac{\hat{\alpha}}{\hat{\beta}}.$$

Substituting $\hat{\beta} = \hat{\alpha}/\overline{x}$ in Equation (21) results in the following relationship for $\hat{\alpha}$,

$$n\left(\ln \hat{\alpha} - \ln \overline{x} - \frac{d}{dx} \ln \Gamma(\hat{\alpha}) + \overline{\ln x}\right) = 0. \tag{22}$$

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

2.2.7. Weibull Distribution

The Weibull distribution is a continuous distribution that is commonly used to model the lifetimes of components. The 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

$$f_X(x; \gamma, \beta) = \frac{\gamma}{\beta^\gamma} x^{\gamma-1} \exp\left(-\left(\frac{x}{\beta}\right)^\gamma\right); \quad x > 0, \gamma > 0, \beta > 0 \tag{23}$$

where γ is the shape parameter and β 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:

$$L(\gamma, \beta) = \prod_{i=1}^{n} \left(\frac{\gamma}{\beta}\right)\left(\frac{x_i}{\beta}\right)^{\gamma-1} \exp\left(-\left(\frac{x_i}{\beta}\right)^\gamma\right) \tag{24}$$

The log-likelihood function is therefore given by

$$l(\gamma, \beta) = n \ln \gamma - n\gamma \ln \beta + (\gamma - 1) \sum_{i=1}^{n} \ln x_i - \sum_{i=1}^{n} \left(\frac{x_i}{\beta} \right)^{\gamma}$$

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

$$\frac{\partial l(\gamma, \beta)}{\partial \gamma} = \frac{n}{\gamma} - n \ln \beta + \sum_{i=1}^{n} \ln x_i - \frac{1}{\beta^{\gamma}} \sum_{i=1}^{n} x_i^{\gamma} \ln x_i = 0 \tag{25}$$

$$\frac{\partial l(\gamma, \beta)}{\partial \beta} = -\frac{n\gamma}{\beta} + \frac{\gamma}{\beta^{\gamma-1}} \sum_{i=1}^{n} x_i^{\gamma} = 0 \tag{26}$$

By eliminating β and simplifying, we obtained the following non-linear equation

$$\frac{\sum_{i=1}^{n} x_i^{\gamma} \ln x_i}{\sum_{i=1}^{n} x_i^{\gamma}} - \frac{1}{\gamma} - \frac{1}{n} \sum_{i=1}^{n} \ln x_i = 0 \tag{27}$$

This can be solved numerically to obtain the estimate of γ by using the Newton-Raphson method. The MLE estimate for β is given by $\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} x_i^{\gamma}$

2.2.8. Log-Normal Distribution

A positive random variable X is log-normally distributed if the logarithm of the random variable is normally distributed. Hence X follows a lognormal distribution if its probability density function is given by

$$f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi} x \sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right), \ x > 0 \tag{28}$$

with parameters: location μ, scale $\sigma > 0$.

The likelihood function for the lognormal distribution is

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} x_i \sigma} \exp\left(-\frac{(\ln x_i - \mu)^2}{2\sigma^2} \right) \tag{29}$$

Therefore the log-likelihood function is given by

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

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

$$\frac{\partial \ln L(\mu, \sigma)}{\partial \mu} = 0 \ \text{ and } \ \frac{\partial \ln L(\mu, \sigma)}{\partial \sigma} = 0 \tag{30}$$

The resulting estimates are $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \ln x_i$ and $\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2}$

respectively.

2.2.9. Pareto Distribution

The Pareto distribution with parameter $\alpha > 0$ and $\gamma > 0$, is given by

$$f_X(x;\alpha,\gamma) = \frac{\alpha\gamma^{\alpha}}{(x+\gamma)^{\alpha+1}}, \; x \geq 0 \tag{31}$$

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

The likelihood function is

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

The log-likelihood function

$$l(\gamma,\alpha) = n\ln(\alpha) + \alpha n \ln(\gamma) - (\alpha+1)\sum_{i=1}^{n}\ln(x_i) \tag{32}$$

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

$$\frac{\partial \ln L(\gamma,\alpha)}{\partial \alpha} = \frac{n}{\alpha} + n\ln(\gamma) - \sum_{i=1}^{n}\ln(x_i) = 0 \tag{33}$$

Thus, we obtain $\hat{\alpha} = n \Big/ \sum_{i=1}^{n}\ln\left(\frac{x_i}{\hat{\gamma}}\right)$

After estimating the parameters of the selected distributions, the next step is typically the determination of the best fitting distribution, *i.e.*, the distribution that provides the best fit to the data set at hand. Establishing the underlying distribution of a data set is fundamental for the correct implementation of claims modeling procedures. This is an important step that must be implemented before employing the selected distributions for forecasting insurance claims or pricing insurance contacts. The consequences of misspecification of the underlying distribution may prove very costly. One way of dealing with this problem is to assess the goodness of fit of the selected distributions.

2.3. Goodness of Fit Tests

A goodness-of-fit (GoF) test is a statistical procedure that describes how well a distribution fits a set of observations by measuring the quantifiable compatibility between the estimated theoretical distributions against the empirical distributions of the sample data. The GoF tests are effectively based on either of the two distribution functions: the probability density function (PDF) or cumulative distribution function (CDF) in order to test the null hypothesis that the unknown distribution function is, in fact, a known specified function. The tests considered for testing the suitability of the fitted distributions to claims data include; the Chi-Square goodness of fit test, Kolmogorov-Smirnov (K-S) test, and the Anderson-Darling (A-D) test. The three GoF tests were selected for two reasons.

First, they are among the most common statistical test for small samples (and they can as well be used for large samples). Secondly, the tests are implemented in most statistical packages and are therefore widely used in practice. The Chi-Square test is based on the PDF while both, the K-S and A-D GoF tests use the CDF approach and therefore they are generally considered to be more powerful than the Chi-Square goodness of fit test.

For all the GoF tests, the hypotheses of interest are:

H_0: The claims data sample follows a particular distribution,

H_1: The claims data samples do not follow the particular distribution.

2.3.1. The 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:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(O_j - E_j\right)^2}{E_j} \tag{34}$$

where O_j is the observed number of cases in interval j, 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.

2.3.2. The Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (K-S) test compares a hypothetical or fitted cumulative distribution function (CDF) $\hat{F}(x)$ with an empirical $F_n(x)$ CDF in order to assess the goodness-of-fit of a given data set to a theoretical distribution. The CDF uniquely characterizes a probability distribution. The empirical $F_n(x)$ CDF is expressed as the proportion of the observed values that are less than or equal to x and is defined as

$$F_n(x) = \frac{I(x)}{n} \tag{35}$$

where n is the size of the random sample, $I(x)$ is the number of xi's less than or equal to x. To test the null hypothesis, the maximum absolute distance between empirical $F_n(x)$ CDF and fitted CDF $\hat{F}(x)$ and is computed. The K-S test statistic D_n is the largest vertical distance between $F_n(x)$ and $\hat{F}(x)$ for all values of x; *i.e.*

$$D_n = \sup_x \left| F_n(x) - \hat{F}(x) \right| \tag{36}$$

where $F(x)$ is the theoretical cumulative distribution of the distribution being tested which must be a continuous distribution, and it must be fully specified.

2.3.3. 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. The test statistic is non-directional, and is calculated using the following formula:

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln\left(x_{(i)}\right) + \ln\left(1 - \left(x_{(n+1-i)}\right)\right)\right] \qquad (37)$$

where $\left\{x_{(1)} < \cdots < x_{(n)}\right\}$ is the ordered sample of size n and $F_n(x)$ is the underlying theoretical distribution to which the sample is compared. The Chi-square goodness-of-fit test is applied to select the appropriate discrete claim frequency distribution while the K-S and A-D tests are used to select the continuous claim severity distributions. Therefore, for the GoF tests considered the null hypothesis that the sample data are taken from a population with the underlying distribution $F_n(x)$ is rejected if the p-value is small than the criterion value at a given significance level α, such as 1% or 5%.

For all the selected claims frequency and claims severity distributions that pass the GoF tests, information criterion, namely the Akaike's information criterion (AIC) developed by Akaike [15] is utilized for the selection of the most appropriate distribution for the claims data. The AIC is not a test of the distribution in the sense of hypothesis testing; rather it compares between distributions—a tool for distribution selection. Given a data set, several fitted distributions may be ranked according to their AIC. The fitted distribution with the smallest AIC is selected as the most appropriate distribution for modeling the claims data. The AIC is defined by

$$AIC = -2l\left(\hat{\theta}|x\right) + 2k \qquad (38)$$

where $l\left(\hat{\theta}|x\right)$ is the log-likelihood function, $\hat{\theta}$ the estimated parameter vector of the model, x is the empirical data and k the length of the parameter vector. The first part $-2l\left(\hat{\theta}|x\right)$ is a measure of the goodness-of-fit of the selected model and the second part is a penalty term, penalizing the complexity of the model. In contrast to the AIC the Bayesian information criterion (BIC) or Schwarz Criterion (SBC), comprises of the number of observations in the penalty term. Thus in BIC, the penalty for additional parameters is stronger than that of the AIC. Apart from that the BIC is similar to the AIC and is defined as

$$AIC = -2l\left(\hat{\theta}|x\right) + 2\log(n) \qquad (39)$$

where $l\left(\hat{\theta}|x\right)$ is the log-likelihood function, $\hat{\theta}$ the estimated parameter vector of the model with length k and x the empirical data vector of length n. As the AIC, the first term is a measure of the goodness-of-fit and the second part is a penalty term, comprising of the number of parameters as well as the number of observations. Note that for both the AIC and BIC the model specification with the lowest value implies the best fit.

3. Empirical Results

3.1. Data Description

The data set used in modeling the claims frequency and claims severity distributions consists of three data sets: AutoCollision, dataCar and dataOhlsson obtained from the R package insuranceData. Autocollision data set is a sample of 32 observations due to [16] who considered 8942 collision losses from private passenger United Kingdom (UK) automobile insurance policies. The average severity is in Sterling Pounds adjusted for inflation. The dataCar dataset consists of 67,856 one-year motor vehicle insurance policies issued in 2004 and 2005. Finally, dataOhlsson dataset contains 64,548 aggregated claims data for motorcycle insurance policyholders on all insurance policies and claims during 1994-1998.

Table 1 presents the summary statistics for the three claims frequency and claims severity datasets. The summary statistics include the number of observations, minimum, maximum, mean, median, standard deviation, skewness, and kurtosis. The Auto Collision data set has the highest mean for both the claims frequency and claims severity given that the number observations were only 32. For both data Car and data Ohlsson, the mean and the standard deviations are close to zero for the claim frequency. This is because the majority of the observations in the datasets are zeros. More specifically, throughout the analyzed period the descriptive statistics show that the Data Car and Data Ohlsson losses are more extreme with respect to skewness and kurtosis than the Auto Collision losses. All the datasets are thus significantly skewed to the right and exhibit excess kurtosis, especially the claims severity have high kurtosis values compared to claims frequency. This, in addition, confirms that the data sets do not follow the normal distribution. The descriptive statistics confirm the skewed nature of the claims data. As a result of these findings, the skewed-heavy tailed distributions are the most appropriate to fit this type of data since they account for skewness, excess kurtosis as well as heavy tails.

Table 1. Descriptive statistics of the claims frequency and claims severity data.

	AutoCollision		dataCar		dataOhlsson	
	Frequency	Severity	Frequency	Severity	Frequency	Severity
No. Obs	32	32	67,856	67,856	64,548	64,548
Minimum	5.00	153.62	0.00	0.00	0.00	0.00
Maximum	970	797.80	4	55,922	2	365,347
Mean	279.44	276.35	0.073	137.27	0.01	264.02
Median	208.00	250.53	0.00	0.00	0.00	0.00
Std.dev	241.61	110.45	0.2828	1058.30	0.10	4690.42
Skewness	1.19	3.22	4.07	17.50	10.5	30.6
Kurtosis	3.84	15.67	21.50	482.9	121.30	1336.00

Figure 1 displays the histograms of the claims frequency and claims severity data. Most of the claims frequency and claims severity for data Car and data Ohlsson data are zero and it is clear that we have zero-inflation problem. For the Auto Collison dataset for claims frequency, most of the values are less than 200 and the data is slightly positively skewed. The claims severity data values mostly lie between 200 and 300 and also we have very few extremely large values that lie between 700 and 800. Also, the histogram in Figure 1 illustrates that claims data in our case are skewed and exhibits heavy-tailed distributions.

Table 2 reports the summary number of claims occurrences distribution recorded and their respective percentages for the data Car and data Ohlsson datasets. The maximum numbers of reported claims occurrences are four in data Car and two in data Ohlsson respectively. The zero claim account for the significant percentage of the claim occurrences recorded. For data Car, the percentage of total zero claims occurrences are 93.18% and 98.96% for data Ohlsson respectively. This kind of data is defined as zero-inflated because of the zero response of the unaffected policyholders. Therefore it is obvious that the distribution of claim occurrences should be one with a very high probability for a claim occurrence of zero. Such distributions would be discrete distributions with a large parameter (probability of success) since they have the mode at zero and the probability of the other values would become progressively smaller. When data displays over-dispersion, the most likely to be used distribution is the negative
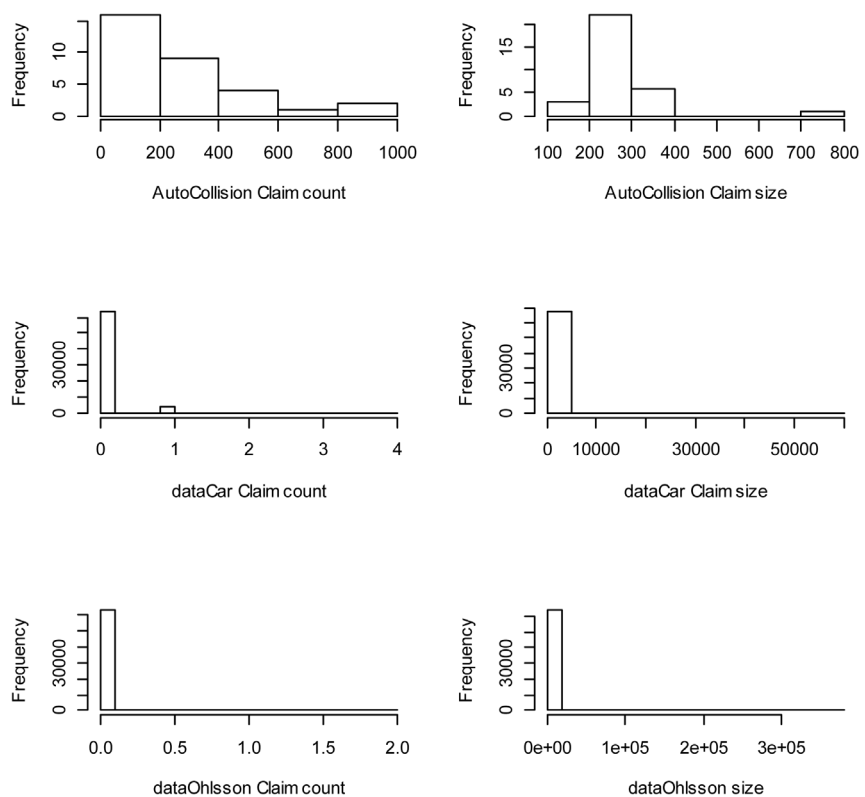

Figure 1. Histogram of the Claims frequency and claims severity data.

Table 2. The number of claims in the datasets.

Occurrence	dataCar		dataOhlsson	
	Frequency	Percentage	Frequency	Percentage
0	63,232	93.18%	63,878	98.96%
1	4333	6.39%	643	1.00%
2	271	0.40%	27	0.04%
3	18	0.03%		
4	2	0.00%		
Total	67,856	100%	64,548	100%

binomial instead of Poisson distribution to model the data.

Table 3 presents the descriptive statistics of the claims data without zero claims. The claims frequency in both data Car and data Ohlsson without zero claims still has a mean and standard deviation close to zero. All the datasets are still significantly skewed to the right and exhibit excess kurtosis. This still confirms the data sets do not follow the normal distribution.

Figure 2 displays the histograms of logarithms of claims frequency and claims severity sets. The histograms look much promising for purposes of fitting the distributions compared to the raw data histograms in Figure 1 especially for the distribution of the claim severity datasets. The Auto Collison data has most of the observations concentrated at the center. The data Car and data Ohisson data sets for claims severity are skewed to the right and left respectively.

3.2. Parameters Estimation

The parameters of the claims frequency and claims severity fitted distributions are estimated using the maximum likelihood estimation method and is implemented via packages in R statistical software. Table 4 represents the parameter estimates and their corresponding (asymptotic) standard errors, the maximum log likelihood function (LLF), AIC and BIC values for the fitted claims frequency distributions to the Auto Collision, data Car and data Ohlsson datasets. The parameter estimates for all the fitted distributions are obtained except for the LLF, AIC, and BIC for the binomial distribution for both data Car and data Ohlsson data.

In order to have better comparisons among the fitted distributions, we us the LLF, AIC and BIC values of the fitted distributions. For the Auto Collision the geometric distribution has the lowest AIC followed by the negative binomial distribution. Both data Car and data Ohlsson have the Poisson distribution with the lowest AIC value followed by the negative binomial distribution. The over-dispersion observed in the data would suggest that Negative binomial or Geometric distributions might be good candidates for the modeling claims frequency data. This will be verified later by the Chi-square goodness-of-fit test.

Table 5 reports the estimated parameters their (asymptotic) standard errors,
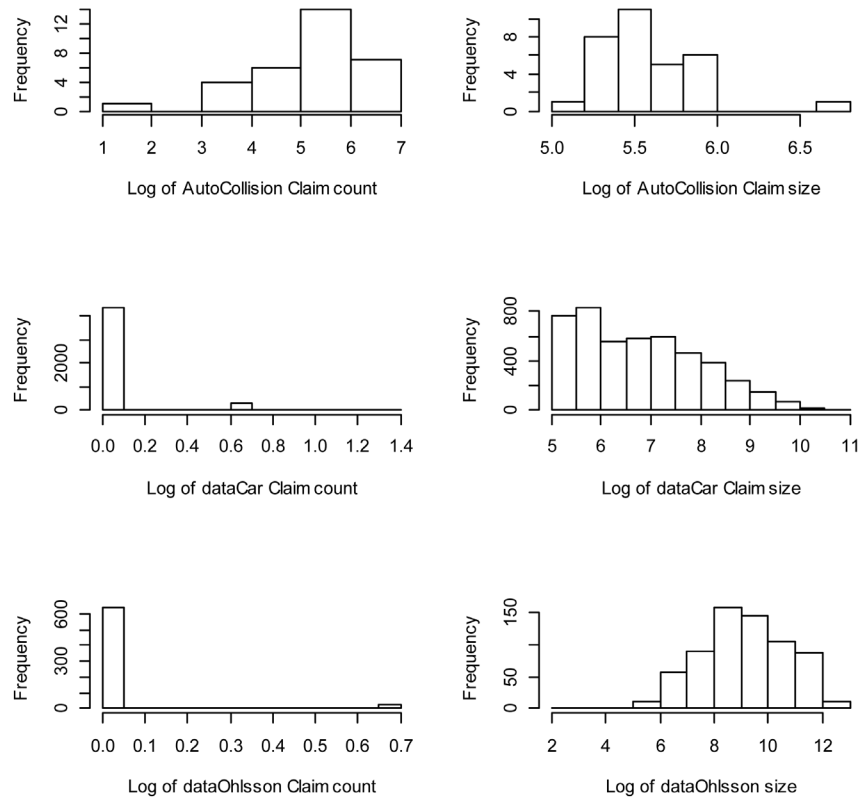
Figure 2. The Histogram of logarithms of claims frequency and claims severity data.

Table 3. Descriptive statistics of the claims data without zero claims.

	dataCar		dataOhlsson	
	Frequency	Severity	Frequency	Severity
Claims Count	67,856	4624	64,548	670
Minimum	0.00	200	0.00	16
Maximum	4	55,922	2	365,000
Mean	0.073	2014	0.01	25,000
Median	0.00	762	0.00	9015
Std.dev	0.2828	3549.65	0.10	38729.83
Skewness	4.07	5.04	10.46	3.10
Kurtosis	21.50	43.20	121.20	17.52

the likelihood function (LLF), AIC and BIC values for the fitted claims severity distributions. The parameter estimates for all the fitted distributions are obtained for the three data sets except for those of the Pareto distribution in the Auto Collision data. The LLF, AIC and BIC criteria are employed for purposes of selecting the appropriate distribution among the fitted distributions. The distribution function with the maximum LLF and lowest AIC or BIC values is

Table 4. Estimated parameters for claims frequency distributions.

Distribution	Parameter(s)	AutoCollision	dataCar	dataOhlsson
Binomial	n	8942	67856	64548
	p	0.03125122	1.072227e−06	1.672889e−07
	(s.e)	(0.00032496)	(0.000000)	(0.000000)
	LLF	−3249.72	–	–
	AIC	6501.441	–	–
	BIC	6502.907	–	–
Geometric	Probability	0.003565857	0.4836314	0.4901244
	(s.e)	(0.00057441)	(0.00511074)	(0.0135207)
	LLF	−212.3061	−6622.056	−947.2655
	AIC	426.6122	13246.11	1896.531
	BIC	428.0779	13252.55	1901.038
Negative Binomial	size	1.216671	8.392017e+06	1.088848e+07
	(s,e)	(0.2757261)	(1.43127302)	(0.000000)
	mu	279.447627	1.067867	1.040076
	(s.e)	(44.8845448)	(0.01519796)	(0.03939565)
	LLF	−211.9508	−4840.089	−688.1782
	AIC	427.9016	9684.177	1380.356
	BIC	430.8331	9697.055	1389.371
Poisson	Lambda	279.4375	1.06769	1.040299
	(s.e)	(2.955068)	(0.01519544)	(0.03940408)
	LLF	−3144.531	−4840.088	−688.1781
	AIC	6291.062	9682.177	1378.356
	BIC	6292.528	9688.616	1382.863

considered to be the best fit. The lognormal distribution as the highest log-likelihood function and also the smallest AIC and BIC values hence the best fit amongst all the distributions fitted and for all three datasets. The log likelihood of the gamma distribution is reasonably closer to that of the lognormal and it is the second best fitting distribution. The other distributions have values that are extremely out of the range. However, the LLF, AIC and BIC at best shows which distributions are better than the competing distributions but do not necessarily qualify the chosen distributions as the most appropriate. To remedy this problem, the Kolmogorov Smirnov and Anderson-Darling goodness-of-fit tests are used to check for the goodness-of-fit of these claim severity distributions. These goodness-of-fit tests verify whether the proposed theoretical distributions provide a reasonable fit to the empirical data.

3.3. Goodness-of-Fit Tests

The purpose of goodness-of-fit tests is typically to measure the distance between the fitted parametric distribution and the empirical distribution: e.g., the

Table 5. Estimated parameters for claims severity distributions.

Distribution	Parameter(s)	AutoCollision	dataCar	dataOhlsson
Exponential	Rate	0.0036186	0.007284904	0.003787624
	(s.e)	(0.0005856)	(2.742692e−05)	(1.376899e−05)
	LLF	−211.8936	−401839.9	−424468.7
	AIC	425.7873	803681.8	848939.4
	BIC	427.253	803690.9	848948.5
Gamma	Shape	10.14141	0.7500861	0.5951737
	(s.e)	(2.470826)	(0.000000)	(0.000000)
	Scale	0.036695	2686.2118	42728.44
	(s.e)	(0.009158)	(0.000000)	(0.000000)
	LLF	−187.1523	−39662.92	−7392.141
	AIC	378.3046	79329.85	14788.28
	BIC	381.2361	79342.72	14797.3
Pareto	Shape		2.046569	1.482972
	(s.e)		(0.08776018)	(0.211924)
	Scale		2206.086511	16839.7191
	(s.e)	N/A	(132.315102)	(3878.686483)
	LLF		−39169.85	−7377.696
	AIC		78343.7	14759.39
	BIC		78356.58	14768.41
Lognormal	meanlog	5.5715751	6.810081	9.104990
	(s.e)	(0.05141875)	(0.01748793)	(0.06241052)
	sdlog	0.2908684	1.189179	1.615456
	(s.e)	(0.03635662)	(0.01236580)	(0.04413083)
	LLF	−184.1801	−38852.15	−7372.376
	AIC	372.3603	77708.31	14748.75
	BIC	375.2917	77721.19	14757.77
Weibull	Shape	2.459737	0.7857048	0.7026427
	(s.e)	(0.2704631)	(0.0081805)	(0.02048285)
	Scale	309.842046	1690.905575	20437.75
	(s.e)	(23.7470479)	(33.6783355)	(1089.2853)
	LLF	−194.4251	−39491.6	−7377.065
	AIC	392.8502	78987.19	14758.13
	BIC	395.7817	79000.07	14767.14

distance between the fitted cumulative distribution function $\hat{F}(x)$ and the empirical distribution function $F_n(x)$. The Chi-square goodness-of-fit test is used to select the most appropriate claims frequency distribution among the fitted discrete probability distributions. Table 6 summarizes the Chi-square goodness-of-fit test-statistic values and the corresponding *p*-value to further compare fitted distributions.

The p-values are used to determine whether or not to reject the null

Table 6. Chi-squared test for claim frequency distributions.

Distribution	AutoCollision	dataCar	dataOhlsson
	TestStatistic (p-value)	TestStatistic (p-value)	Test Statistic (p-value)
Binomial	1.9155e24 (0.0000)	98.734 (0.0000)	148.01 (0.0000)
Geometric	1.015653 (0.6018)	1.868321 (0.17166)	54.58665 (0.0000)
Negative Binomial	0.514027 (0.4734)	0.255171 (0.61346)	1.493632 (0.22165)
Poisson	1.6768e17 (0.0000)	98.7294 (0.0000)	184.002 (0.0000)

hypotheses. The Negative-Binomial and geometric distributions have p-values greater than $\alpha = 0.01$ in almost all cases except for the geometric distribution for data Ohlsson. The p-values for the binomial and Poisson distributions for all the three data sets are zero; hence the null hypothesis that claims data follows a particular distribution is rejected for the binomial and Poisson distributions with a 99% confidence level. Thus, the Chi-square goodness-of-fit test confirmed that both negative binomial and geometric distributions are appropriate distributions for modeling claims frequency while binomial and Poisson distributions provide the inappropriate fit.

To select the most appropriate continuous distributions for the claims severity, two goodness-of-fit statistics are typically considered: K-S and A-D statistics [17] are used to test the suitability of fitted claims severity distributions to the data sets. Table 7 reports the K-S and A-D test statistic values for the claim severity distributions that are used to model the claims severity data. The K-S and A-D test statistic values are used to compare the fit of competing distributions as opposed to an absolute measure of how a particular distribution fits the data. Smaller K-S and A-D test statistic values indicate that the distribution fits the data better. However, these statistics should be used cautiously especially when comparing fits of various distributions. The Kolmogorov-Smirnov and Anderson-Darling statistics computed for several claims severity distributions fitted on the same data set like in our case are hypothetically complicated to compare. Moreover, such a statistic, as K-S one, doesn't take into account the complexity of the distribution (*i.e.* the number of parameters). It is less problematic where the compared distributions are characterized by the same number of parameters, but again it could methodically encourage the selection of the more complex distributions. Both the K-S and A-D goodness-of-fit statistics based on the CDF distance choose the lognormal distribution, which is also the preference for the AIC and BIC values for all the datasets as the most appropriate claims severity distribution.

Table 7. K-S and A-D test statistic values for claims severity distributions.

Distribution	K-S test statistic	A-D test statistic
Auto Collision		
Exponential	0.4695586	8.1269415
Gamma	0.1605827	1.2760080
Weibull	0.2339492	2.8469195
Pareto	-	-
Lognormal	0.1410449	0.8257456
DataCar		
Exponential	0.1870179	341.7652236
Gamma	0.1502237	191.3327388
Weibull	0.1704634	139.5430172
Pareto	0.1627262	87.9184259
Lognormal	0.1021038	72.4949308
DataOhlsson		
Exponential	0.21022191	59.0055091
Gamma	0.09457255	7.97122388
Weibull	0.07629782	4.48316734
Pareto	0.05641147	3.01466283
Lognormal	0.04244645	1.82690162

4. Conclusions

Non-life insurance companies require an accurate insurance pricing system that makes adequate provision for contingencies, expenses, losses, and profits. Claims incurred by an insurance company form a large part of the cash outgo of the company. An insurance company is required to model its claims frequency and severity in order to forecast future claims experience in order to prepare adequately for claims when they fall due. In this paper, selected discrete and continuous probability distributions are used as approximate distributions for modeling both the frequency and severity of claims made on automobile insurance policies. The probability distributions are fitted to three datasets of insurance claims in an attempt to determine the most appropriate distributions for describing non-life insurance claims data.

The findings from empirical analysis indicate that claims severity distribution is more accurately modeled by a skewed and heavy-tailed distribution. Amongst the continuous claims severity distributions fitted, the lognormal distribution is selected as a reasonably good distribution for modeling claims severity. On the other hand, negative binomial and geometric distributions are selected as the most appropriate distributions for the claims frequency compared to other standard discrete distributions. These results may be considered as informative

assumed models for automobile insurance companies while choosing models for their claims experience and making liability forecasts. The forecasts obtained under such distributions are however suitable for projecting the short-term claims behavior. For long-term usage, it is recommended that the company uses its own claims experience to make necessary adjustments to the distributions. This would allow for anticipated changes in the portfolios and for company specific financial objectives. These proposed claims distributions would also be useful to insurance regulators in their own assessment of required reserve levels for various companies and in checking for solvency.

A suggestion for further research, interested parties may look into extensions of the standard probability distributions covered here such as the zero-inflated models. Due to a large number of zeros in claims frequency data, there is need to consider other distributions such as the zero-truncated Poisson or zero-truncated negative binomial and zero-modified distributions from the class $(a, b, 1)$ to model this unique phenomenon.

References

[1] David, M. and Jemna, D.V. (2015) Modeling the Frequency of Auto Insurance Claims by Means of Poisson and Negative Binomial Models. *Analele Stiintifice Ale Universitatii Al I Cuza Din Iasi—Sectiunea Stiinte Economice*, **62**, 151-168. https://doi.org/10.1515/aicue-2015-0011

[2] Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2012) Loss Models: From Data to Decisions. 3rd Edition, Society of Actuaries.

[3] Ohlsson, E. and Johansson, B. (2010) Non-Life Insurance Pricing with Generalized Linear Models. EAA Lecture Notes, Springer, Berlin, 71-99. https://doi.org/10.1007/978-3-642-10791-7

[4] H. R. W. (1984) Loss Distributions by Hogg Robert V. and Klugman Stuart A. *Transactions of the Faculty of Actuaries*, **39**, 421-423. https://doi.org/10.1017/S0071368600009046

[5] Jindrová, P. and Pacáková, V. (2016) Modelling of Extreme Losses in Natural Disasters. *International Journal of Mathematical Models and Methods in Applied Sciences*, **10**, 171-178.

[6] Antonio, K., Frees, E.W. and Valdez, E.A. (2010) A Multilevel Analysis of Intercompany Claim Counts. *ASTIN Bulletin*, **40**, 151-177. https://doi.org/10.2143/AST.40.1.2049223

[7] Yip, K.C.H. and Yau, K.K.W. (2005) On Modeling Claim Frequency Data in General Insurance with Extra Zeros. *Insurance: Mathematics and Economics*, **36**, 153-163. https://doi.org/10.1016/j.insmatheco.2004.11.002

[8] 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. https://doi.org/10.1080/10920277.2007.10597487

[9] Mouatassim, Y. and Ezzahid, E.H. (2012) Poisson Regression and Zero-Inflated Poisson Regression: Application to Private Health Insurance Data. *European Actuarial Journal*, **2**, 187-204. https://doi.org/10.1007/s13385-012-0056-2

[10] Bahnemann, D. (2015) Distributions for Actuaries. CAS Monograph Series No. 2,

Casualty Actuarial Society, Arlington.

[11] Achieng, O.M. and NO, I. (2010) Actuarial Modeling for Insurance Claims Severity in Motor Comprehensive Policy using Industrial Statistical Distributions. *International Congress of Actuaries*, Cape Town, 7-12 March 2010, Vol. 712.

[12] Ignatov, Z.G., Kaishev, V.K. and Krachunov, R.S. (2001) An Improved Finite-Time Ruin Probability Formula and Its Mathematica Implementation. *Insurance: Mathematics and Economics*, **29**, 375-386.
https://doi.org/10.1016/S0167-6687(01)00078-6

[13] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008) Modern Actuarial Risk Theory: Using R. Springer, Berlin, Vol. 53.

[14] Merz, M. and Wüthrich, M.V. (2008) Modelling the Claims Development Result for Solvency Purposes. CAS E-Forum, 542-568.

[15] Akaike, H. (1974) A New Look at Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.
https://doi.org/10.1109/TAC.1974.1100705

[16] Mildenhall, S. (1999) A Systematic Relationship between Minimum Bias and Generalized Linear Models. *Proceedings of the Casualty Actuarial Society*, **86**, 393-487.

[17] Koziol, J.A., D'Agostino, R.B. and Stephens, M.A. (1987) Goodness-of-Fit Techniques. *Journal of Educational Statistics*, **12**, 412.