Strathmore UNIVERSITY

Loss Distributions for Motor Insurance Claim Severity in Kenya

Nyairo Frankline Okindo - 101428

Submitted in partial fulfillment of the requirements for the Degree of Bachelor of Business Science in Actuarial Science at Strathmore University

Strathmore Institute of Mathematical Sciences

Strathmore University

Nairobi, Kenya

February 2021

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Project contains no material previously published or written by another person except where due reference is made in the Research Project itself.

Nyairo Frankline Okindo

.... [*Signature*]

..................09-02-2021............... [*Date*]

This Research Project has been submitted for examination with my approval as the Supervisor.

Kevin Otieno

.... [*Signature*]

..................09/02/21........................ [*Date*]

Strathmore Institute of Mathematical Sciences

Strathmore University

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

This chapter begins with the background information of the study in section 1.1 by explaining the key concepts, main developments, and conceptualization of the study. In section 1.2, the problem statement of the study is elaborated at length. Section 1.3 contains the research objectives and questions. This chapter concludes with the significance of the research in section 1.4.

1.1 Background Information

1.1.1 Key Concepts

Claim: A legal application made by a policyholder to an insurer for indemnity covered under the policy agreement.

Distribution: A function that shows the possible values for a variable and how often they occur.

Loss: The basis of a claim for damages under the terms of an insurance policy.

Severity: The cost of a claim.

1.1.2 Main Developments and Conceptualization of Study

The insurance industry is one of the oldest industries. Insurance companies exist to provide indemnity and make profits since insurance is a business like any other. The advancement of the insurance market is compelled by the prevailing interest of the public for cover against different forms of risks of unacceptable arbitrary incidents with a considerable financial effect (Omari et al., 2018). A policyholder is supposed to pay a premium and make a claim when a certain event occurs within a given period as per the policy. The insurer is then obliged to settle the claim, and this is referred to as loss. Insurers are keen with the results of the random outcome of claims instead of the existence of the claims. They are concerned with the loss rather than the circumstances that give rise to the loss (Achieng, 2010). The aggregate amount of claims in a given duration is a measure that is vital to the operations of an insurance company.

A general insurance actuary of a company needs to understand different risk models comprising of the aggregate claim amount overdue in a certain period. These models enlighten a company and allow it to decide on things such as anticipated profits, premiums to be charged, required reserves that will guarantee profitability with a high likelihood, and the effect of reinsurance and policy excess (Boland, 2006).

Actuaries are tasked with the responsibility of developing cashflow models for insurance companies. These models are crucial since they are used to assist in the day-to-day work of insurers and to provide checks and controls on the business. An actuary may decide to develop a stochastic model and estimate one of the parameters e.g. claim size by assigning it a probability distribution. If the assigned probability distribution is not appropriate for the claim size parameter in the stochastic cash flow model, it will lead to adverse future experience by the insurance company due to parameter error which results in a model error. An actuary will, therefore, be concerned with objectively assigning an appropriate probability distribution to the value of the claim size parameter.

In the general insurance business, there is heightened concern in motor insurance because it involves the control of many risk events. These include fire, theft, third party bodily injury and accidental damage to the vehicle. In most countries, the motor insurance industry is growing rapidly due to legislations that make motor insurance compulsory for all vehicles.

The insurance industry is driven by data, and insurers engage a lot of analysts to comprehend claims data (Boland, 2006). The claims data contains among other things, the frequency and size of claims that a company has received within a given period. Based on the claims data, mathematical methods can be applied to model individual claims. The mathematical models are known as loss distributions.

Loss distributions are vital in the insurance industry since they are used for many purposes which include: premium rating (deciding the premium rates to be paid by policyholders), reserving (determining the required amount of funds to be retained to offset the cost of claims), reviewing reinsurance arrangements and testing for solvency (evaluating the insurer's financial health). This explicitly highlights the importance of having a good estimate of an insurance company's loss distributions.

A loss distribution is the associated probability distribution of a claim-size variable. The claim-size is a non-negative continuous random variable since the claim arising from a covered incident can be measured in the lowest unit of currency e.g. cents. Loss distributions are usually positively skewed and long-tailed. To model the size of insurance claims, it is prudent to fit a continuous parametric claim-size distribution to a discrete sample of claim data. This involves employing a variety of parametric families of continuous distributions which include: Gamma, Lognormal, Exponential and Pareto distributions, among others.

The gamma and lognormal distributions are among the common distributions that have been applied for modelling claim severity. The exponential, Pareto, Weibull, and Burr distributions are also used to model claim severity. This is primarily because all these distributions are positively skewed. Omari et al. (2018) suggested that a lognormal distribution is suitable to model claim severity based on a sample of the automobile portfolio datasets obtained from the insurance Data package in R. Achieng (2010) concluded that the lognormal distribution was a suitable model for the motor comprehensive policy claim severity of First Assurance Company Limited, Kenya. Nduwayezu (2016) found out that the exponential distribution is suitable to model insurance data; though he left the research open for further studies to be carried out to determine the distributions that are most suitable for each class of insurance.

Most research papers have provided a good framework to use when modelling loss distributions for motor insurance claims severity. However, it can be noted that the statistical distributions suggested for modelling claims severity are general and not exhaustive since they are based on the sample data that was being used by the various researchers. In the Kenyan environment, there are no specific statistical distributions that have been recommended to be used for modelling motor insurance claim severity based on the motor insurance claims data in Kenya. This creates a need for research on suitable statistical distributions that can be used within the Kenyan environment as this will enhance the motor insurance industry.

1.2 Problem Statement

One of the major challenges that general insurers face is to precisely estimate the likely prospective claims experience and therefore charging suitable premiums and setting aside sufficient reserves (Omari et al., 2018). For these companies to overcome the challenge of accurately forecasting future claims experience, they need to have a good estimate of loss distributions. A good estimate of a loss distribution entails selecting a suitable statistical distribution that fits the claims data.

Determining motor insurance claims distributions often comprises associating the value of claims with two elements: the occurrence of an accident and the claim amount in case of an accident (Frees & Valdez, 2008). Records from insurance databases show the following claim types: third-party liability claims, and damage claims to the policyholder, comprising property damage, injury, theft, and fire. This, therefore, implies that for every accident, it is probable for multiple types of claims to be incurred; thus, increasing the claim severity of an insurance company for every single accident. This creates a need for having good models of loss distributions that will enable an insurer to plan accordingly to lower the probability of incurring such a loss and reducing the claim severity incurred.

Motor insurance is the biggest segment out of the 14 distinctive classes of the non-life insurance market in Kenya accounting for over KES 46 billion gross written premiums representing close to 35.8% of the entire non-life insurance market in Kenya in 2018. It is not surprising that motor insurance is a huge business in Kenya given that more than 7000 cars are imported monthly. This can be credited to the fact that motor insurance is compulsory in Kenya, and thus, for every new vehicle purchase in the country, a motor insurance policy is added to the existing motor insurance policies for the vehicles on the Kenyan roads. To this end, the huge role played by motor insurance to the Kenyan insurance industry and the economy at large cannot be overlooked. This implies that motor insurance companies need to have good models that will enable them to accurately forecast future claims experience and thus be able to set aside enough reserves.

According to Insurance Outlook Report 2019/2020, East Africa by Deloitte, motor insurance in Kenya is one of the largest general insurance classes together with medical insurance. However, they are also among the top loss-making businesses. This could be partly attributed to the fact that motor insurance companies in Kenya do not have good models for loss distributions, and thus cannot be able to correctly forecast future claims experience. This leads them to undergo huge losses because of failing to plan accordingly to lower the probability of incurring such losses.

This research paper will, therefore, seek to address this gap in the Kenya motor insurance industry by providing a good model of loss distribution for claim severity. This will, in turn, help insurers to precisely estimate prospective claims experience and thus plan accordingly to reduce their huge losses and the chances of them making such losses.

1.3 Research Objectives and Questions

1.3.1 Research Objectives

The objective of this research is to determine the appropriate statistical distribution that fits claim severity data of motor insurance in Kenya and can be used to accurately forecast future claims experience.

1.3.2 Research Questions

Throughout this research, the paper will aim to answer the following questions:

i. What is the most appropriate statistical distribution for claim severity?
ii. How well does this loss distribution fit the claims data?

1.4 Significance of the Research

This research will provide motor insurance companies in Kenya with the most suitable loss distribution for claim severity. This will enable them to accurately forecast future claims experience and thus be able to correctly: rate premiums, reserve, review reinsurance arrangements and test for solvency. Given that risk associated with claim severity would have been minimised, motor insurance policyholders will consequently pay reduced premiums. The research will enhance understanding of the complexity of the extensive volume of claims that is usually concealed in a large amount of data.

CHAPTER TWO: LITERATURE REVIEW

This chapter discusses the theoretical and empirical framework. It begins with a brief introduction of various studies that have been previously carried out on claims data modelling in section 2.1. In section 2.2, the Maximum Likelihood Estimation is then presented as a method of estimating parameters. Subsequently in section 2.3, various continuous distributions are discussed. This chapter closes by presenting Kolmogorov-Smirnov and Anderson-Darling as goodness-of-fit tests in section 2.4, and Akaike Information Criterion and Bayesian Information Criterion as model selection criteria in section 2.5.

2.1 Introduction

A loss distribution is a mathematical method of modelling individual claims. It involves fitting statistical distributions to observed claims data and then testing for the goodness of fit. The fitted loss distributions can then be used to estimate probabilities. The main assumption in all the distributions in this study is that the amount of a claim and its occurrence can be considered independently. Thus, a claim arises according to some elementary model for incidents ensuing in time, then the claim amount is selected from a distribution representing the claim amount.

Ignatov et al. (2001) provided a statistical process of fitting a suitable model to claims data. The process begins with the selection of a family of distributions for the claims model and then estimating the parameters for the model. A selection criterion to determine the appropriate distribution from the family of distributions should be specified. Finally, a goodness-of-fit test should be carried out on the selected appropriate distribution.

Achieng (2010) modelled the claim severity motor comprehensive data of First Assurance Company Limited, Kenya (June 2006 – June 2007). The estimates for the parameters were obtained using the Maximum Likelihood Estimation method. The Akaike Information Criteria and Quantile-Quantile plots were further utilised to carry out a goodness-of-fit test. The finding of the study was that the lognormal distribution was a suitable model for the claims data.

Mazviona and Chiduza (2013) used four distributions (Gamma, Pareto, Exponential and Lognormal) to model the claims for a motor portfolio in Zimbabwe. Their study used Maximum Likelihood Estimation and Method of Moments Estimation to estimate parameters for the models. The Chi-Square and Kolmogorov-Smirnov tests were used as goodness-of-fit tests for the models. The finding of their study was that the Lognormal distribution is suitable for smaller claims while the Pareto distribution is appropriate for larger claims.

Packová and Brebera (2015) used Gamma, Weibull, Lognormal and Pareto distributions to model data obtained from a Czech insurance company for compulsory motor third-party liability insurance. The Maximum Likelihood Estimator method was used to estimate the parameters of the selected parametric distributions. They further used the Anderson-Darling, Chi-Square and Kolmogorov-Smirnov tests to determine whether the chosen distribution provides a good fit to the data. The finding of the study was that the Pareto distribution can be assumed to be a good model for the losses.

Omari et al. (2018) modelled a sample of the automobile portfolio datasets obtained from the insurance Data package in R with variables; Auto Collision, data Car, and data Ohlsson used. They used the Maximum Likelihood Estimation method to obtain parameter estimates for the fitted models. The Anderson-Darling and Kolmogorov-Smirnov tests were then used as goodness-of-fit tests for the claim severity models. The Akaike Information Criterion and Bayesian Information Criterion were further applied to choose between competing distributions. The finding of their study was that the lognormal distribution provides a good model for claims severity on a short-term basis. The study recommended that for a long-term basis, insurers should adjust the distributions accordingly based on insurer-specific claims experience.

To this end, this study seeks to apply the theoretical and empirical frameworks of past studies and extend it in the Kenyan economy. The aim is to fit a suitable loss distribution to the claim severity data for motor insurance companies using the company-specific data set.

In line with the research objectives of this study, it is prudent to explain the theoretical framework that will be applied throughout this paper. This will include parameter estimation, standard continuous distributions, goodness-of-fit test, and model selection criteria. These will then form the basis of the subsequent parts of the study that will eventually yield an appropriate loss distribution model.

2.2 Maximum Likelihood Estimation

The method of maximum likelihood is generally considered as the best general method of finding estimators. Maximum likelihood estimators have excellent and normally simply determined asymptotic properties and so are especially good in the large-sample situation. The likelihood function of a random variable, X, is the probability (or probability density function) of observing what was observed given a hypothetical value of the parameter, θ. The maximum likelihood estimate (MLE) is the one that provides the highest probability (or probability density function), i.e. that maximises the likelihood function.

2.3 Standard Continuous Distributions

The claim-size is a non-negative continuous random variable since the claim arising from a covered incident can be measured in the lowest unit of currency e.g. cents. To model the size of insurance claims, it is prudent to fit a continuous parametric claim-size distribution to a discrete sample of claim data. Due to the infrequency of relatively large claims which are of concern, Boland (2006) suggested the use of relatively fat-tailed distributions. Kaas et al. (2008) stated that claim-size is best modelled using continuous distributions that are positively skewed and long-tailed. This is because very large claims occur at the upper-right tails of the distribution. The following parametric families of continuous distributions will be considered: Exponential, Gamma, Lognormal, Pareto, Weibull, and Burr. In this section, the distribution functions and probability density functions will be given.

2.3.1 Exponential Distribution

The exponential distribution is one of the elementary models for claim severity. A random variable X has the exponential distribution with parameter $\lambda > 0$ if it has distribution function:

$$F(X) = 1 - e^{-\lambda X}, X > 0$$

In that case, we write X ~ Exp (λ).

The probability density function is:

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

Achieng (2010) used the exponential distribution due to its heavy-tailed and highly skewed nature to model claim severity motor comprehensive data of First Assurance Company Limited, Kenya (June 2006 – June 2007). The distribution was not a good fit due to its low log-likelihood value and low-density value of its probability density function graphical plot.

Mazviona and Chiduza (2013) used the exponential distribution to model a motor dataset. Their study found out that this distribution failed to fit the data very closely based on the critical value for the chi-square test and thus rejecting the null hypothesis.

Omari et al. (2018) used exponential distribution to model an automobile dataset. The study rejected the null hypothesis for this distribution because it had the largest values among the distributions used for the Kolmogorov-Smirnov and Anderson-Darling tests.

2.3.2 Gamma Distribution

The random variable X has a gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$ if it has probability density function:

$$f(X) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} X^{\alpha-1} e^{-\lambda X}, \qquad x > 0$$

The parameter α changes the shape of the graph of the probability density function, and the parameter λ changes the x-scale. In that case, we write X ~ Ga (α, λ).

Achieng (2010) used the gamma distribution due to its heavy-tailed and highly skewed nature to model claim severity motor comprehensive data of First Assurance Company Limited, Kenya (June 2006 – June 2007). The study concluded that gamma distribution was not a suitable model for the claims data based on its Q-Q plot.

Mazviona and Chiduza (2013) used the gamma distribution to model a motor dataset. Their study found out that this distribution failed to fit the data very closely based on the critical value for the chi-square test and thus rejecting the null hypothesis.

Packová and Brebera (2015) used gamma distribution to model data obtained from a Czech insurance company for compulsory motor third-party liability insurance. The reason for using this distribution was because it is specifically applicable for modelling of claim severity. The study found out that the gamma distribution failed to be a suitable model for the losses based on the Anderson-Darling test value.

Omari et al. (2018) used gamma distribution to model an automobile dataset. The study found out that the gamma distribution is a better model among the others based on the log-likelihood value.

2.3.3 Lognormal Distribution

X has a lognormal distribution if log X has a normal distribution. If X represents, for example, claim size and Y = log X has a normal distribution, then X is said to have a lognormal distribution.

When log X ~ N (μ, σ^2), X ~ LN (μ, σ^2).

The probability density function of the lognormal distribution is defined by:

$$f(X) = \frac{1}{X\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log X-\mu}{\sigma}\right)^2} \; for\; 0 < X < \infty$$

If X has a lognormal distribution with parameters μ and σ , then we can write X ~ log N (μ , σ^2).

Achieng (2010) used the lognormal distribution due to its heavy-tailed and highly skewed nature to model claim severity motor comprehensive data of First Assurance Company

Limited, Kenya (June 2006 – June 2007). The finding of the study was that the lognormal distribution would be the best statistical distribution to model the claim amounts of First Assurance Company Limited at a 99% level of confidence. This was because the lognormal distribution had the smallest Akaike Information Criterion value.

Burnecki et al. (2010) used the lognormal distribution to model Danish fire losses dataset, which concerns major fire losses that occurred between 1980 and 1990 and were recorded by Copenhagen Re. This distribution was chosen since it is a typical candidate for claim size distributions considered in application. The study concluded by suggesting the lognormal distribution as a model for the Danish fire loss amounts because it was the only distribution that passed all the applied tests therein.

Mazviona and Chiduza (2013) used the lognormal distribution to model a motor dataset. Their study found out that this distribution fit the data very closely and based on the graphical plot of its probability density function, concluded that it produced the best fit for lower claims.

Packová and Brebera (2015) used lognormal distribution to model data obtained from a Czech insurance company for compulsory motor third-party liability insurance. The reason for using this distribution was because it is specifically applicable for modelling of claim severity. The study found out that the lognormal distribution failed to be a good model for larger losses based on the chi-square test value.

Omari et al. (2018) used lognormal distribution to model an automobile dataset. The study found out that the lognormal distribution was the most suitable model since it had the lowest Akaike Information Criterion and Bayesian Information Criterion values.

2.3.4 Pareto Distribution

A random variable X has the Pareto distribution with parameters $\alpha > 0$ and $\lambda > 0$ if it has distribution function:

$$F(X) = 1 - \left(\frac{\lambda}{\lambda + X}\right)^{\alpha}, X > 0$$

In that case, we write X ~ Pa (α, λ).

The probability density function is given by:

$$f(X) = \frac{\alpha\lambda^{\alpha}}{(\lambda + X)^{\alpha+1}}, X > 0$$

Mazviona and Chiduza (2013) used the Pareto distribution to model a motor dataset. Their study found out that this distribution fit the data very closely and based on the graphical plot of its probability density function, concluded that it provided the best fit for larger claims. The study further recommended the Pareto distribution because it does not undervalue the probabilities for larger claims.

Packová and Brebera (2015) used Pareto distribution to model data obtained from a Czech insurance company for compulsory motor third-party liability insurance. This was because this distribution is frequently used as a model for insurance losses required to obtain well-fitted tails. The study concluded that the Pareto distribution is a good model for large claims based on the tests carried out.

Omari et al. (2018) used Pareto distribution to model an automobile dataset. This distribution was used since it has been shown to sufficiently mimic the tail-behaviour of claims amount thereby providing a good fit. However, the study discarded the Pareto distribution since its values were extremely out of range based on the tests carried out.

2.3.5 Weibull Distribution

This is a very flexible distribution which can be used to model claim severity. It is a modification of the Pareto and exponential distributions usually with $\gamma < 1$. A random variable X has a Weibull distribution with parameters c > 0 and $\gamma > 0$ if it has distribution function:

$$F(x) = 1 - \exp(-cx^{\gamma}), \qquad x > 0$$

In that case, we write X ~ W (c, γ). The probability density function of the W (c, γ) distribution is:

$$f(x) = c\gamma x^{\gamma-1}\exp(-cx^{\gamma}), \qquad x > 0$$

Achieng (2010) used the Weibull distribution due to its heavy-tailed and highly skewed nature to model claim severity motor comprehensive data of First Assurance Company Limited, Kenya (June 2006 – June 2007). The study established that the Weibull distribution is not a suitable model for the claims data based on the tests carried out.

Packová and Brebera (2015) used Weibull distribution to model data obtained from a Czech insurance company for compulsory motor third-party liability insurance. The reason for using this distribution was because it is specifically applicable for modelling of claim severity. The study found out that the Weibull distribution failed to be a good model for the losses based on the Anderson-Darling test value.

Omari et al. (2018) used Weibull distribution to model an automobile dataset. The study discarded this distribution as the appropriate distribution since it failed to meet the selection criteria based on the tests carried out.

2.3.6 Burr Distribution

The distribution function of the Pareto Distribution Pa (α, γ) is:

$$F(x) = 1 - \frac{\lambda^{\alpha}}{(\lambda + x)^{\alpha}}, x > 0$$

A further parameter $\gamma > 0$ can be introduced by setting:

$$F(x) = 1 - \frac{\lambda^{\alpha}}{(\lambda + x^{\gamma})^{\alpha}}, x > 0$$

This is the distribution function of the transformed Pareto or Burr distribution. The extra parameter provides additional flexibility when a fit to data is needed. The probability density function is given by:

$$f(x) = \frac{\alpha\gamma\lambda^{\alpha}x^{\gamma-1}}{(\lambda + x^{\gamma})^{\alpha+1}}, x > 0$$

Burnecki et al. (2010) used the Burr distribution to model fire losses dataset. This distribution was chosen since it is a typical candidate for claim size distributions considered in application. The study failed to suggest the Burr distribution as a good model since it failed to pass the applied tests therein.

It is worth to mention that while this distribution can be used to model claim size distributions, most studies fail to utilize it among their chosen distributions. It would be interesting to conduct a study around the same to determine the rationale behind the exclusion of this distribution in major studies.

2.4 Goodness-of-fit Test

This refers to verifying whether a particular loss distribution provides a good model for the observed claim amounts i.e. whether a model provides a "good fit" to the data. This involves determining the quantitative "compatibility between the estimated theoretical distributions against the empirical distributions of the sample data" (Omari et al., 2018). This enables one "to determine whether the observed sample was drawn from a population that follows a particular probability distribution" (Dodge, 2008).

Myung (2003) argued that even though a good fit is required, it is not an adequate requirement for one to conclude that one model provides a closer approximation to data than does another model simply because the former model fits the data better than the latter. A better fit (larger value of the maximized log-likelihood) simply places the model in a series of competing models for additional considerations such as goodness-of-fit tests. The Kolmogorov-Smirnov and Anderson-Darling tests are used because they are suitable for performing an exact test on continuous distributions.

2.4.1 Kolmogorov-Smirnov Test

The Kolmogorov–Smirnov test is a nonparametric goodness-of-fit test and is used to determine whether an underlying probability distribution differs from a hypothesized distribution.

Consider an independent random sample $(x_1, x_2, \ldots, x_n)$, a sample of size n with unknown distribution function $F(x)$ coming from a population with a specific and known distribution function $F_0(x)$. The hypothesis to test is as follows:

$H_0: F(x) = F_0(x)$

H_1: $F(x) \neq F_0(x)$

If $F(x)$ is the empirical distribution function of the random sample, then the statistical test T_n is defined as the greatest vertical distance between $F_0(x)$ and $F(x)$:

$$T_n = \sup_x |F_0(x) - F(x)|$$

The decision rule is to reject H_0 at the significance level α if T_n is greater than the value of the Kolmogorov table having for the parameters n and 1-α, which is denoted by $t_{n,1-\alpha}$, i.e., if:

$$T_n > t_{n,1-\alpha}$$

2.4.2 Anderson-Darling Test

The Anderson–Darling test is a goodness-of-fit test which allows controlling the hypothesis that the distribution of a random variable observed in a sample follows a certain theoretical distribution.

Consider a random variable X, which follows a particular distribution, and has a distribution function $F_0(x;\ \theta)$, where θ is a parameter (or a set of parameters) that determine, F_0. We further assume θ to be known. An observation of a sample of size n issued from the variable X gives a distribution function $F(x)$. The Anderson–Darling statistic, denoted by A^2, is then given by the weighted sum of the squared deviations $F_0(x;\ \theta) - F(x)$:

$$A^2 = \frac{1}{n}\left(\sum_{i=1}^{n}\left(F_0(x;\ \theta) - F(x)\right)^2\right)$$

Starting from the fact that A^2 is a random variable that follows a certain distribution over the interval [0; +∞], it is possible to test, for a significance level that is fixed a priori, whether $F(x)$ is the realization of the random variable $F_0(X;\ \theta)$; that is, whether X follows the probability distribution with the distribution function $F_0(x;\ \theta)$.

The computation of A^2 Statistic is as follows: Arrange the observations $x_1, x_2, \dots, x_n$ in the sample obtained from X in ascending order i.e., $x_1 < x_2 < \dots < x_n$. The A^2 is then computed as:

$$A^2 = -\frac{1}{n}\left(\sum_{i=1}^{n}(2i-1)(\ln(z_i) + \ln(1 - z_{n+1-i}))\right) - n$$

where $z_i = F_0(x_i;\ \theta),\ (i = 1, 2, \dots, n)$

The null hypothesis is rejected beyond the limiting values of A^2 depending on the significance level based on the Anderson-Darling Test Table.

2.5 Model Selection Criteria

An information criterion measures the quality of a model by analysing how well the model fits the data and the simplicity of the model. Information criteria are used to analyse different models that are fitted to the same data set. Ceteris paribus, a model with a smaller value is preferable to one with a larger value.

2.5.1 Akaike Information Criterion

The Akaike Information Criterion (AIC) was formulated by and named after Akaike (1974). The AIC is an estimator of out-of-sample prediction error that yields a technique for model selection. AIC estimates the relative amount of information lost by a given model: the fewer information that is lost by a model, the better the model's quality.

In a series of models, the most appropriate model that is selected is the one with the least AIC value. The AIC value for a model is calculated as follows:

$$AIC = 2k - 2\ln(\hat{L})$$

where k is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function of the model.

2.5.2 Bayesian Information Criterion

The Bayesian Information Criterion (BIC) is also known as the Schwarz Information Criterion (SIC). It is a criterion used to select models among a definite series of models. BIC was formulated by Schwarz (1978) and given a Bayesian argument for its adoption. It has been widely used for model selection and can be applied to any set of maximum-likelihood based models. The appropriate model that is preferred is one that has the lowest BIC value since it implies a lower penalty term. It is similar to the Akaike Information Criterion and it is based to some extent on the likelihood function. BIC introduces a larger penalty term than AIC.

The BIC value for a model is calculated as follows:

$$BIC = k\ln(n) - 2\ln(\hat{L})$$

where k is the number of estimated parameters in the model, n is the number of observations and $\hat{L}$ is the maximized value of the likelihood function of the model.

CHAPTER THREE: METHODOLOGY

This chapter outlines the methodological framework that will be used in the study. Section 3.1 discusses the design of the research; section 3.2 discusses the population and sample of the study while section 3.3 discusses the collection of the study's data. Finally, section 3.4 explains the data analysis processes that will be carried out by the study.

3.1 Research Design

This study adopts a quantitative method which is focusing on modelling an appropriate loss distribution for motor insurance claim severity in Kenya. The selected appropriate loss distribution will then be tested to check on its goodness-of-fit before being recommended as an appropriate model. The variable of interest is the claim size in the motor insurance industry. This study will use data for Kenya motor insurance companies from 2014 to 2018. The choice of this period is to make the study to be appropriate, relevant, reliable, and applicable as possible.

3.2 Population and Sampling

The focus of the study is motor insurance companies in Kenya. As of 2018, there were 36 licensed insurance companies in Kenya which were providing motor insurance services in Kenya. These companies will form the population of the study consequently. This study will focus on the whole population since the data is readily available to obtain and the population size is relatively small. It is worth to note that this study will not focus on reinsurance companies since they need to be studied independently.

3.3 Data Collection

This study seeks to apply the theoretical and empirical frameworks of past studies and extend it in the Kenyan economy. The aim is to fit a suitable loss distribution to the claim severity data for motor insurance companies using the company-specific data set. This study is focused on claim sizes for motor insurance companies in Kenya from 2014 to

2018. This implies the type of data to be used within the study to be quantitative continuous ratio panel data. The population that is being focused on by the study is made up of 36 licensed motor insurance companies that are regulated by the Insurance Regulatory Authority (IRA).

IRA produces annual reports highlighting activities within the insurance industry by various insurance companies. This study will use the data that is contained within these annual reports that have been published by the regulator. The study will thus use secondary data provided by IRA that includes among other things the claim sizes of various licensed insurance companies providing motor insurance services.

The annual reports provided by IRA are available in Microsoft Excel Binary File Format. This will make it easier for the study to extract the relevant data from the report using programs such as Microsoft Excel and R. These programs will then be further used to analyse the extracted data according to the research objectives.

3.4 Data Analysis

In compliance with the research objectives, this study seeks to find an appropriate model that fits claims size of motor insurance companies. Ignatov et al. (2001) provided the following steps to be followed when fitting a suitable model to claims data:

a. Select a family of distributions for the claims model.

b. Estimate the parameters for the model.

c. Specify a selection criterion to determine the appropriate distribution from the family of distributions.

d. Carry out a goodness-of-fit test on the selected appropriate distribution.

This study will, therefore, follow the above steps in line with the research objectives while analysing the data. Microsoft Excel and R computer programs will be used together for data analysis within the study. The first step that will be carried out on the data is to find its descriptive statistics such as mean, variance and skewness. These values will come in handy when comparing them with the results obtained from the various models to select an appropriate loss distribution.

In line with the research objectives of this study, it is prudent to explain the framework that will be applied in the process of analysing data. This will include parameter estimation, standard continuous distributions, goodness-of-fit test, and model selection criteria. These will then form the basis of data analysis that will eventually yield an appropriate loss distribution model. It is important to note that sections 3.4.1 to 3.4.4 below are outlining the theoretical framework that will be applied in analysing the data using Microsoft Excel and R computer programs.

3.4.1 Maximum Likelihood Estimation

The parameters of the chosen loss distribution in this study will be estimated using the Maximum Likelihood method. The most important stage in applying the method is that of writing down the likelihood:

$$L(\theta) = \prod_{1}^{n} f(x_i; \theta)$$

for a random sample $x_1, x_2, \dots, x_n$ from a population with density or probability function $f(x; \Theta)$.

In most cases taking logs greatly simplifies the determination of the maximum likelihood estimator (MLE) $\hat{\theta}$.

The following steps are used when determining a maximum likelihood estimate (MLE):

1. Specify the likelihood function for the available data.

$$L(\theta) = \prod_{1}^{n} f(x_i; \theta)$$

2. Simplify the algebra using natural logs.

$$I\ (\theta) = \log L\ (\theta) = \sum_{i=1}^{n} \log f\ (x_i\ |\theta)$$

3. Maximise the log-likelihood function by differentiating the log-likelihood function with respect to each of the unknown parameters and equating the resulting expression(s) to zero.

$$\frac{d}{d\theta} l(\hat{\theta}) = 0$$

4. The MLEs of the parameters are obtained by solving the resulting equation(s). To ensure that the obtained values maximise the likelihood function, differentiate a second time.

3.4.2 Standard Continuous Distributions

This study will employ the following parametric families of continuous distributions: Exponential, Gamma, Lognormal, Pareto, and Weibull. In this section, the mean, variance, and the maximum likelihood estimates for the parameters will be given.

A. Exponential Distribution

The distribution function of an exponential distribution is given by:

$$F(X) = 1 - e^{-\lambda X}, X > 0$$

The probability density function is:

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

The mean and variance of X are:

$$E(X) = \frac{1}{\lambda} \qquad var(X) = \frac{1}{\lambda^2}$$

The likelihood function is:

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i} = \lambda^n e^{-\lambda n \bar{x}}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

The log-likelihood function becomes:

$$\log L = n \log \lambda - \lambda n\bar{x}$$

Determine stationary points by differentiating:

$$\frac{\partial}{\partial \lambda} \log L = \frac{n}{\lambda} - n\bar{x}$$

Setting this to zero gives:

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

B. Gamma Distribution

The probability density function of the gamma distribution is given by:

$$f(X) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} X^{\alpha-1} e^{-\lambda X}, \qquad x > 0$$

The mean and variance of X are:

$$E(X) = \frac{\alpha}{\lambda} \qquad var(X) = \frac{\alpha}{\lambda^2}$$

The moment estimators are used as initial estimators for the MLEs since they cannot be obtained in closed form (i.e. in terms of elementary functions).

C. Lognormal Distribution

The probability density function of the lognormal distribution is defined by:

$$f(X) = \frac{1}{X\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log X - \mu}{\sigma}\right)^2} \; for \; 0 < X < \infty$$

The mean and variance of X are:

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2} \qquad var(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$$

Estimating the MLEs is simple since μ and σ^2 may be estimated using the log-transformed data. Let $x_1, x_2, \ldots, x_n$ be the observed values and let $y_i = \log x_i$.

The MLEs will be given by:

$$\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

$$\hat{\sigma}^2 = s_y^2$$

where the subscript y signifies a sample variance computed on the y values

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2$$

D. Pareto Distribution

The distribution function of the Pareto distribution is defined as:

$$F(X) = 1 - \left(\frac{\lambda}{\lambda + X}\right)^{\alpha}, X > 0$$

The probability density function is given by:

$$f(X) = \frac{\alpha\lambda^{\alpha}}{(\lambda + X)^{\alpha+1}}, X > 0$$

The mean and variance of X are:

$$E(X) = \frac{\lambda}{\alpha - 1}(\alpha > 1) \qquad var(X) = \frac{\alpha\lambda^2}{(\alpha - 1)^2(\alpha - 2)}(\alpha > 2)$$

The likelihood function is:

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

The log-likelihood function becomes:

$$\log L = n\log(\alpha) + \alpha n\log(\lambda) - (\alpha+1)\sum_{i=1}^{n}\log(x_i)$$

To maximize the log-likelihood function, set $\hat{\lambda} = \min(x_i)$, such that λ is less than the least x_i. Differentiating and setting to zero:

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + n\log(\lambda) - \sum_{i=1}^{n}\log(x_i) = 0$$

This will result in:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n}\log(\frac{x_i}{\hat{\lambda}})}$$

E. Weibull Distribution

The distribution function of the Weibull distribution is:

$$F(x) = 1 - \exp(-cx^{\gamma}), \qquad x > 0$$

The probability density function is:

$$f(x) = c\gamma x^{\gamma-1}\exp(-cx^{\gamma}), \qquad x > 0$$

The method of maximum likelihood is not simple to apply if both c and γ are unknown. Nevertheless, the equations are elementary when a computer is used. In the case where γ has the known value γ^*, maximum likelihood is easy enough. We use the data transformation $y_i = {x_i}^{\gamma}$. The y values will now have an exponential distribution. The MLE analysis can now be done easily.

3.4.3 Goodness-of-fit Test

It is important to test whether a particular loss distribution is a suitable model for the observed claim amounts i.e. whether a model provides a "good fit" to the data. This enables one "to determine whether the observed sample was drawn from a population that follows a particular probability distribution" (Dodge, 2008). In this paper, both the Kolmogorov-Smirnov and Anderson-Darling tests will be applied because they are suitable for performing an exact test on continuous distributions. For all the goodness-of-fit tests, the hypotheses will be formulated as follows:

H_0: The claim severity data follows a particular distribution $[F(x) = F_0(x)]$

H_1: The claim severity data does not follow the particular distribution $[F(x) \neq F_0(x)]$

where $F(x)$ is the unknown distribution function of the claim severity data (sample) while $F_0(x)$ is a specific and known distribution function (population).

3.4.4 Information Criteria

For all the selected claim severity distributions that pass the goodness-of-fit test, both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) will be used to select the best model for the claim severity data. It is possible to increase the likelihood by adding parameters when fitting models. However, this may lead to overfitting. Both AIC and BIC introduce a penalty term for the number of parameters in the model in a bid to resolve the problem of overfitting. BIC introduces a larger penalty term than AIC. Even though BIC value is always higher than AIC value, the lower the value of these two criteria the better a model is.

3.4.5 Model Selection Criteria

The research objective is to determine an appropriate loss distribution that provides a good fit to the claim severity data of motor insurance companies in Kenya. Based on the data analysis procedures that have been elaborately outlined in sections 3.4.1 to 3.4.4 above, the loss distribution that will be selected as being an appropriate model is one that has:

1. The maximum MLE value subject to passing the goodness-of-fit tests, and
2. The minimum AIC and BIC value

The study will focus on finding the loss distribution that meets the above requirements.

CHAPTER FOUR: DATA ANALYSIS

4.1 Introduction

The data required for this study was motor insurance claim severity for Kenya. This data was readily available from publications of the Insurance Regulatory Authority (IRA) in form of annual reports. The annual reports were obtained in Microsoft Excel format and the required data was extracted therein. The data contained claim severity for 36 insurance companies that were licensed and regulated by IRA from 2014-2018. The data for Kenya Motor Insurance incurred claims was provided in terms of motor commercial and motor private. These two sets of data for the period 2014-2018 were analysed separately to obtain suitable models for each category of motor insurance. The data was analysed using R software.

4.2 Descriptive Statistics

The first step of data analysis was to determine the descriptive statistics of the data to get a general overview and allow a simpler interpretation of the data. Table 1 shows the descriptive statistics of the motor insurance incurred claims. The descriptive statistics in Table 1 confirm the positive skewness of the claims data and as such, positively skewed and long-tailed distributions being appropriate to model this data.

Table 1. Kenya Motor Insurance Incurred Claims Descriptive Statistics for 2014-2018

	Motor Commercial	Motor Private
No. of observations (n)	170	174
Mean	281,703,882.35	386,762,402.30
Standard Error	23,826,723.55	30,253,278.83
Median	178,457,000.00	216,641,000.00
Standard Deviation	310,662,467.00	399,068,156.03
Sample Variance	9.65E+16	1.59E+17
Kurtosis	3.7716	2.6583
Skewness	1.9078	1.6815
Range	1,470,681,000.00	1,991,252,000.00
Minimum	89,000.00	994,000.00
Maximum	1,470,770,000.00	1,992,246,000.00
Sum	47,889,660,000.00	67,296,658,000.00

Figure 1 shows the histograms of the original data of motor commercial and motor private claim sizes. The normal curves superimposed on the histograms show that the claim sizes are skewed to the right. The purpose of carrying out the normality test was to determine whether to use either parametric tests or non-parametric tests on the data after fitting distributions. Given that the data was not following the normal curve, this was a confirmation that non-parametric tests would be applied to the distributions that would be fitted on the data.

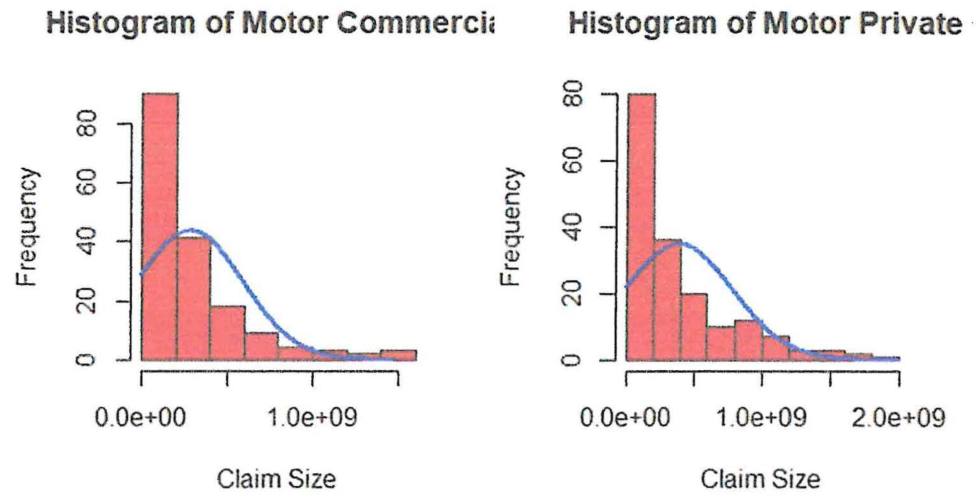


Figure 1

Figure 2 shows the Q-Q plots of the original data of motor commercial and motor private claim sizes. Given that the upper ends of the Q-Q plots are deviating more from the straight line than the lower ends, this confirms that the data is positively skewed. This implied the need to use continuous distributions that are positively skewed to fit the data.

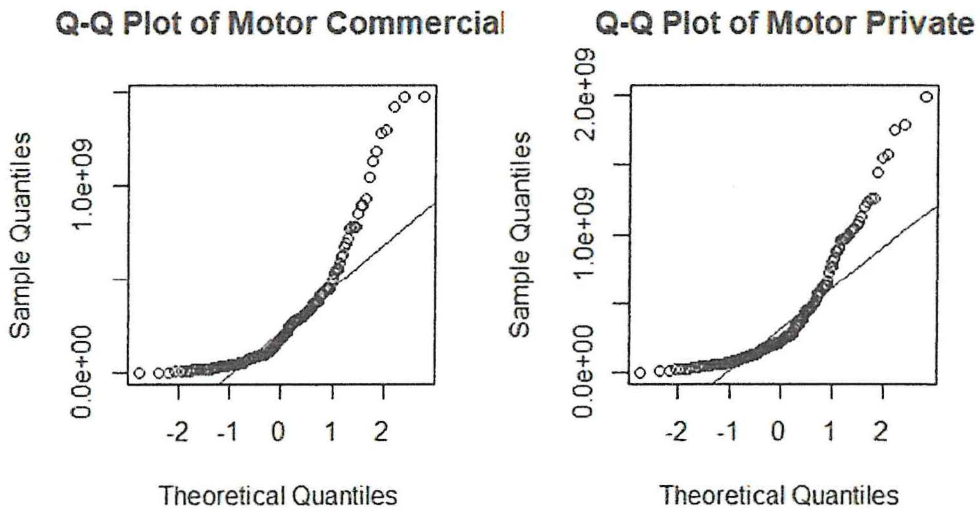


Figure 2

The data was transformed using the cube root function to make it simpler to work with since the original data contained very large values. Figure 3 shows the histograms of the cube-root transformed data of motor commercial and motor private claim sizes.

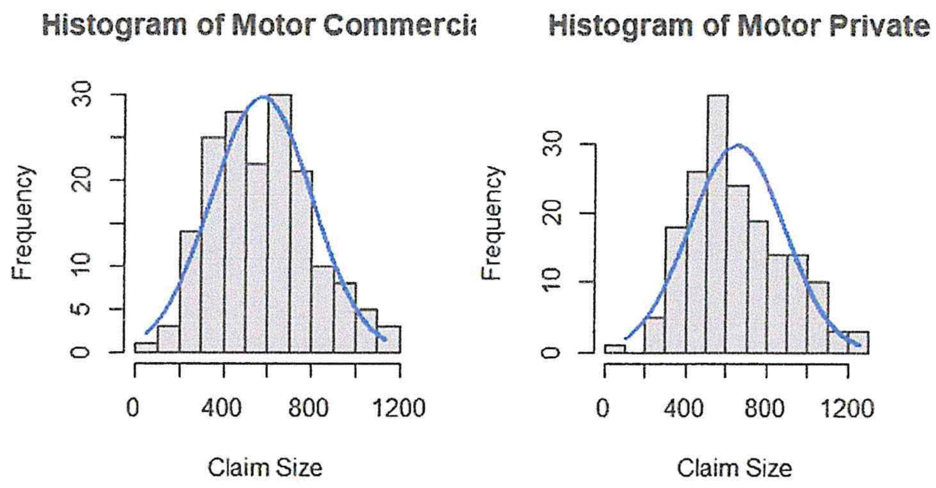


Figure 3

Figure 4 shows the Q-Q plots of the cube-root transformed data of motor commercial and motor private claim sizes.

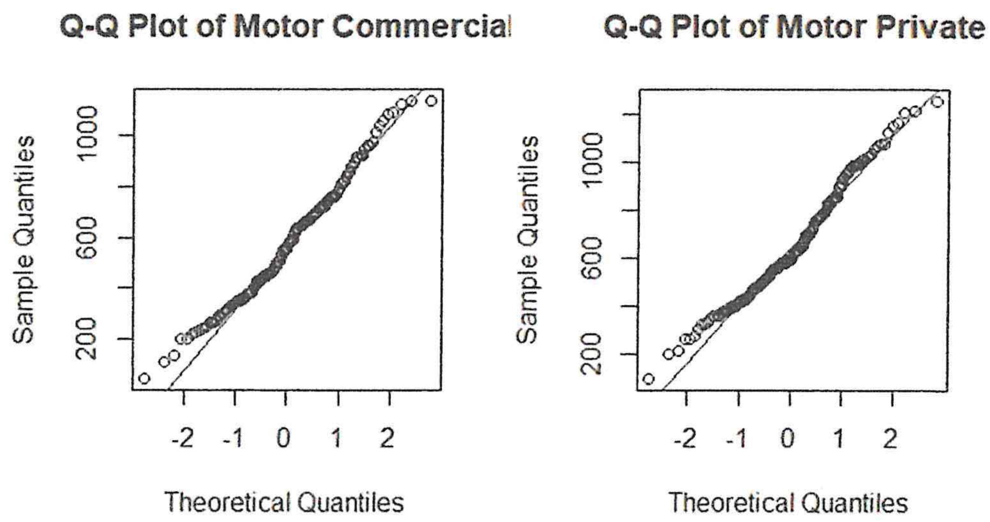


Figure 4

The cube-root transformed data seemed to be more suitable for purposes of fitting the distributions compared to the original data as shown by their respective histograms and Q-Q plots above. With the data being transformed, the next step was fitting distributions to obtain an appropriate model. Subsequently, goodness-of-fit tests were carried out and information criteria applied on the fitted distributions as outlined in the sections below. All these steps were carried out using the R software package "fitdistrplus".

4.3 Parameters Estimation

The parameters for the various fitted distributions were obtained using the MLE method. Table 2 shows the parameter estimates, their corresponding standard errors and the maximum loglikelihood function (LLF). The most appropriate distribution is the one with the highest loglikelihood function.

Table 2. Estimated Parameters for fitted distributions

Distribution	Parameter	Motor Commercial	Motor Private
Exponential	Rate	0.0021	0.0020
	std. error	0.0001	0.0001
	LLF	-1218.773	-1254.143
Gamma	Shape	5.4827	7.3252
	std. error	0.5377	0.7165
	Rate	0.0096	0.0113
	std. error	0.0010	0.0011
	LLF	-1165.534	-1192.602
Lognormal	Meanlog	6.2576	6.4079
	std. error	0.0360	0.0296
	SDlog	0.4696	0.3898
	std. error	0.0255	0.0209
	LLF	-1176.499	-1197.959
Pareto	Shape	5094930	3221289
	(std. error)	-	-
	Scale	2925993718	2095806234
	(std. error)	-	-
	LLF	-1249.768	-1301.128
Weibull	Shape	2.7097	3.0010
	std. error	0.1617	0.1747
	Scale	644.7977	728.9534
	std. error	19.2325	19.4038
	LLF	-1161.697	-1193.441

The parameter values in Table 2 are the ones that maximised the loglikelihood function for each distribution. For motor commercial, Weibull distribution has the maximum loglikelihood function (-1161.697) followed by Gamma (-1165.534), Lognormal (-

1176.499), Exponential (-1218.773) and Pareto (-1249.768) distributions. For motor private, Gamma distribution has the maximum loglikelihood function (-1192.602) followed by Weibull (-1193.441), Lognormal (-1197.959), Exponential (-1254.143) and Pareto (-1301.128) distributions. Based on the LLF values, Weibull and Gamma distributions were the most appropriate for motor commercial and motor private, respectively. The plots for the various fitted distributions are provided in the appendices section.

4.4 Goodness-of-Fit Test

The main aim of carrying out the goodness-of-fit test was to measure the distance between the fitted parametric distribution $F(x)$ and the empirical distribution $F_0(x)$. The Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests were used to determine the appropriateness of the fitted distributions to the claim size data. These two tests helped to determine the most suitable continuous distribution for the claim severity. Table 3 shows the K-S and A-D test statistic values for the distributions fitted on the claim severity data.

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

Test Statistic	Distribution	Motor Commercial	Motor Private
K-S	Exponential	0.3089418	0.3601346
	Gamma	0.07988259	0.04042086
	Lognormal	0.09186256	0.04602873
	Pareto	0.3084464	0.3600815
	Weibull	0.05444920	0.08136149
A-D	Exponential	27.6395653	33.0098233
	Gamma	0.67351891	0.37752203
	Lognormal	1.68105273	0.59667217
	Pareto	27.5714722	33.0018593
	Weibull	0.41538833	1.08746132

The K-S and A-D test statistic values were used to compare the fit of the distributions to the data 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. For motor commercial, Weibull distribution had the smallest K-S and A-D test statistic values (0.05444920, 0.41538833) followed by Gamma (0.07988259, 0.67351891), Lognormal (0.09186256, 1.68105273), Pareto (0.3084464, 27.5714722) and Exponential (0.3089418, 27.6395653) distributions. For motor private, Gamma distribution had the smallest K-S and A-D test statistic values (0.04042086, 0.37752203) followed by Lognormal (0.04602873, 0.59667217), Weibull (0.08136149, 1.08746132), Pareto (0.3600815, 33.0018593), and Exponential (0.3601346, 33.0098233) distributions. This implied that Weibull and Gamma distributions fitted the data of motor commercial and motor private better, respectively.

4.5 Information Criteria

The Akaike and Bayesian Information Criteria were applied to determine the appropriate distribution among the fitted distributions. Table 4 shows the AIC and BIC values for the fitted distributions. The most appropriate distribution is one which has the minimum AIC and BIC values.

Table 4. AIC and BIC values for fitted distributions

Information Criterion	Distribution	Motor Commercial	Motor Private
AIC	Exponential	2439.547	2510.286
	Gamma	2335.067	2389.204
	Lognormal	2356.999	2399.918
	Pareto	2503.536	2606.255
	Weibull	2327.393	2390.881
BIC	Exponential	2442.682	2513.445
	Gamma	2341.339	2395.522
	Lognormal	2363.270	2406.236
	Pareto	2509.808	2612.573
	Weibull	2333.665	2397.199

For motor commercial, Weibull distribution had the minimum AIC and BIC values (2327.393, 2333.665) followed by Gamma (2335.067, 2341.339), Lognormal (2356.999, 2363.270), Exponential (2439.547, 2442.682) and Pareto (2503.536, 2509.808) distributions. For motor private, Gamma distribution had the minimum AIC and BIC values (2389.204, 2395.522) followed by Weibull (2390.881, 2397.199), Lognormal (2399.918, 2406.236), Exponential (2510.286, 2513.445) and Pareto (2606.255, 2612.573) distributions. This implied that Weibull and Gamma distributions were the most appropriate for motor commercial and motor private, respectively.

4.6 Summary

The most appropriate distribution is one which has the maximum LLF, minimum K-S and A-D test statistic values, and minimum AIC and BIC values. Based on the data analysis carried out, as shown above, the most appropriate distributions to model claim severity data for motor commercial and motor private are Weibull distribution and Gamma distribution, respectively.

Based on past studies that were carried out as discussed in the literature review section, the lognormal distribution is fronted as being the most suitable to model motor insurance claim severity. However, given the findings of this research, Weibull and Gamma distributions are the most appropriate to model motor commercial and motor private claim severity, respectively. The lognormal distribution is the third most suitable model for this purpose based on the findings of the study.

CHAPTER FIVE: CONCLUSION

5.1 Introduction

The objective of this research was to determine the appropriate statistical distribution that fits claim severity data of motor insurance in Kenya and can be used to accurately forecast future claims experience. To achieve this, the study used the steps to be followed when fitting a suitable model to claims data as provided by Ignatov et al. (2001).

The first step entailed selecting a family of distributions for the claims model whereby the Exponential, Gamma, Lognormal, Pareto and Weibull distributions were selected given that they are continuous positively skewed distributions. The parameters for these distributions were estimated using the MLE method. The K-S and A-D tests were applied as goodness-of-fit tests on the fitted distributions. The AIC and BIC information criteria were used to determine the appropriate distribution among the fitted distributions.

5.2 Conclusion

The study carried out an analysis of Kenya motor insurance claim severity data from 2014-2018 based on the above steps. The finding of the study was that the Weibull and Gamma distributions are suitable for modelling motor commercial and motor private data, respectively. This was because they had the maximum LLF, minimum K-S and A-D test statistic values, and minimum AIC and BIC values among the fitted distributions.

Findings of past studies as highlighted in the Literature Review section [Achieng (2010) and Omari et al. (2018)] fronted the Lognormal distribution as being the most suitable for modelling claim severity. Gamma, Weibull, Pareto, and Exponential distributions were also fronted as being appropriate as well in that decreasing order of preference [Mazviona and Chiduza (2013), and Packová and Brebera (2015)]. The findings of this study are different from past studies given that the Weibull and Gamma distributions have been selected as being the most suitable to model claim severity for motor commercial and motor private data respectively. Lognormal, Pareto and Exponential distributions are also appropriate in that decreasing order of preference.

The objectives of this study were achieved given that the Weibull and Gamma distributions have been selected as being the most appropriate to model claim severity of motor commercial and motor private, respectively. These distributions were selected after the data analysis steps used to achieve the research objectives were successfully followed and meeting the selection criteria of the study and thus being selected as the most suitable after emerging the best among the other competing distributions. To this end, Weibull and Gamma distributions fit the claim severity data of motor commercial and motor private respectively and can be used to accurately forecast future claims experience; thus, achieving the objectives of the study.

5.3 Recommendations

The appropriate distributions obtained under this study are suitable for forecasting short-term claims experience given the period covered by the study. It is recommended that insurers use their own claims experience to adjust the distributions accordingly for long-term forecasts. This would allow insurers to consider their specific financial objectives and expected changes in their investment portfolios. These proposed claim severity distributions would also be useful to insurance regulators while testing for solvency and assessing the required levels of reserves for various insurance companies.

5.4 Limitations of the Study

This study focused on statistical distributions only as opposed to considering other approaches such as non-parametric methods, machine learning and deep learning. These approaches are also suitable for modelling claim severity. Non-parametric methods do not assume underlying statistical distributions in the data and thus do not rely on any distribution. This approach may serve well in certain circumstances whereby data fails to follow any statistical distribution and thus be well modelled using this approach. Technological advancement in terms of machine and deep learning has improved efficiency. This can be applied by insurers in the modelling process by predicting patterns of claim volume and augmenting loss analysis using artificial intelligence.

This study did not determine the impact of modelling claim severity on the business of insurance companies. Insurers need to know this impact to make suitable adjustments in their modelling process. Modelling claim severity involves the use of resources and thus, insurers need to know whether it is worth it or not for them to commit resources towards modelling given the value being added to the business in terms of increased or decreased profitability. If modelling has a positive impact on the business in terms of increased profitability, insurers will strive to commit high-quality resources towards the modelling process in a bid to ensure they increase their profitability.

5.5 Suggestions for Further Research

Interested parties may investigate other approaches that may be used to model claim severity such as non-parametric methods, machine learning and deep learning.

Further studies may also be carried out to determine the impact of modelling claim severity on the business of insurance companies.

REFERENCES

1. Omari, C., Nyambura, S. and Mwangi, J. (2018). Modeling the Frequency and Severity of Auto Insurance Claims Using Statistical Distributions. *Journal of Mathematical Finance*, **8**, 137-160. doi: 10.4236/jmf.2018.81012
2. Achieng, O. M. (2010). Actuarial modeling for insurance claim severity in motor comprehensive policy using industrial statistical distributions. International Congress of Actuaries, Cape Town, 7–12 March 2010.
3. Boland, P. J. (2006). Statistical Methods in General Insurance.
4. Frees, E. W., & Valdez, E. A. (2008). Hierarchical Insurance Claims Modeling. Journal of the American Statistical Association, 103(484), 1457–1469. doi: 10.1198/016214508000000823
5. Nduwayezu, F. (2016). Finding appropriate loss distributions to insurance data: Case study of Kenya (2010-2014). Retrieved April 3, 2020, from https://su-plus.strathmore.edu/handle/11071/5361
6. Distributions Used in Actuarial Science. (n.d.). Retrieved April 3, 2020, from https://reference.wolfram.com/language/guide/DistributionsUsedInActuarialScience.html
7. Association of Kenya Insurers (2018). Insurance Industry Annual Report. Retrieved April 23, 2020, from https://www.akinsure.com/images/publications/AKI%20_Ann_Report_2018.pdf
8. Deloitte (2019). Insurance Outlook Report 2019/2020 East Africa. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/ke/Documents/audit/Final%20Insurance%20Outlook%20Report%20EA%202019.pdf
9. Bahnemann, D. (2015). Chapter 2: Claim Size. In DISTRIBUTIONS FOR ACTUARIES (pp. 37–77). Arlington, Virginia 22203: CASUALTY ACTUARIAL SOCIETY.
10. Mazviona, B. W., & Chiduza, T. (2013). THE USE OF STATISTICAL DISTRIBUTIONS TO MODEL CLAIMS IN MOTOR

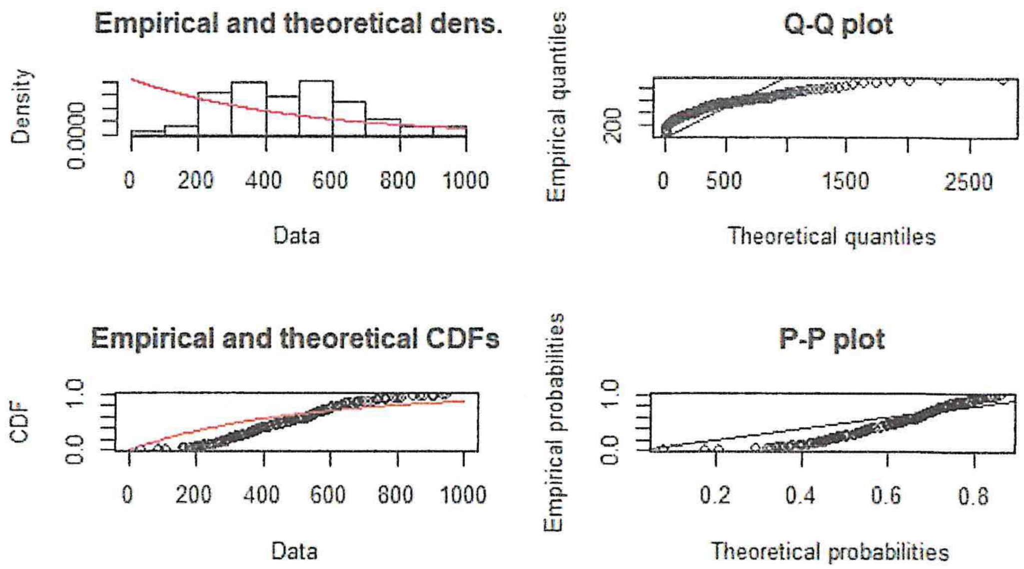
INSURANCE. International Journal of Business, Economics and Law, 3(1), 44–57. Retrieved from https://www.ijbel.com/wp-content/uploads/2014/01/KLB3140-THE-USE-OF-STATISTICAL-DISTRIBUTIONS-TO-MODEL-CLAIMS-IN-MOTOR-INSURANCE-Batsirai.pdf

11. Packová, V., & Brebera, D. (2015). Loss Distributions in Insurance Risk Management. *Recent Advances on Economics and Business Administration: Proceedings of the International Conference on Economics and Statistics (ES 2015). Vienna, Austria, March 15-17, 2015*, 17–22. Retrieved from http://www.inase.org/library/2015/barcelona/ECBAS.pdf
12. Dodge, Y. (2008). The concise encyclopedia of statistics: With 247 tables. New York, NY: Springer.
13. 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
14. Burnecki, K., Misiorek, A., & Weron, R. (2010). Loss Distributions. Statistical Tools for Finance and Insurance, 289-317. doi:10.1007/3-540-27395-6_13
15. Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008). Modern Actuarial Risk Theory Using R. Springer, Berlin, Vol. 53.
16. Ignatov, Z. G., Kaishev, V. K., & Krachunov, R. S. (2001). An improved finite-time ruin probability formula and its Mathematica implementation. Insurance: Mathematics and Economics, 29(3), 375-386. doi:10.1016/s0167-6687(01)00078-6
17. Myung, I. J. (2003). Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology, 47(1), 90-100. doi:10.1016/s0022-2496(02)00028-7
18. Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2), 461-464. doi:10.1214/aos/1176344136
19. Research & Statistics: Annual Reports. Retrieved from https://www.ira.go.ke/index.php/publications/statistical-reports/annual-reports
20. Actuarial Education Company. (2014). Retrieved from https://acted.co.uk/docs/2014/CMP%20Upgrade/CT6-PU-14.pdf?cv=1

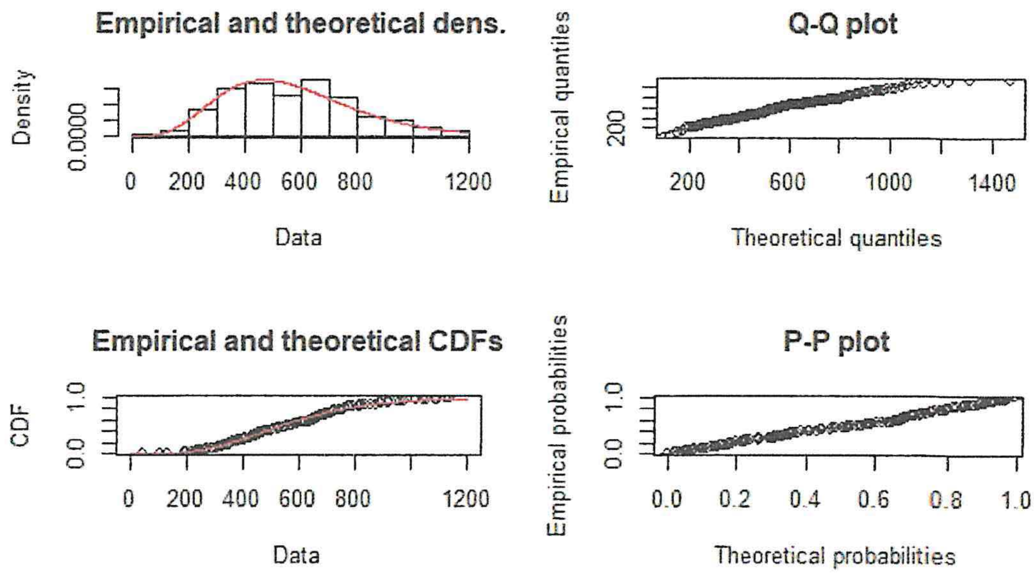
APPENDICES

Appendix A: Fitted Distributions for Motor Commercial

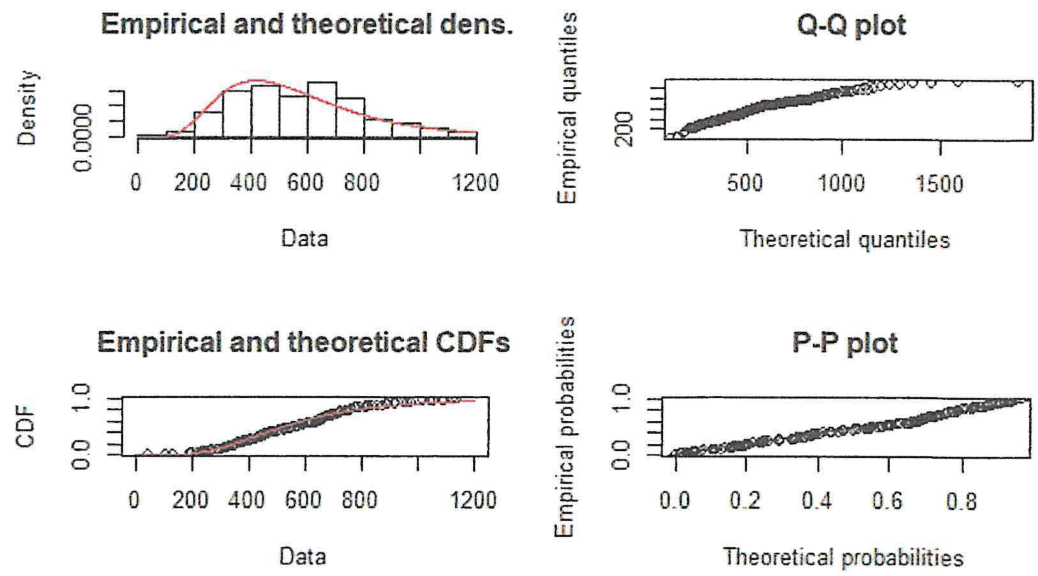
Appendix A1: Exponential Distribution



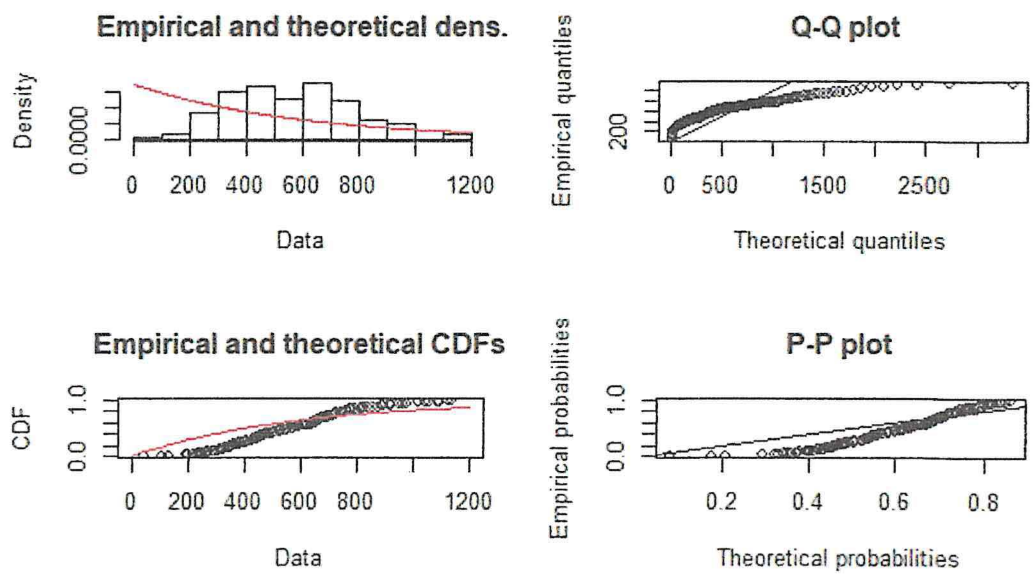
Appendix A2: Gamma Distribution



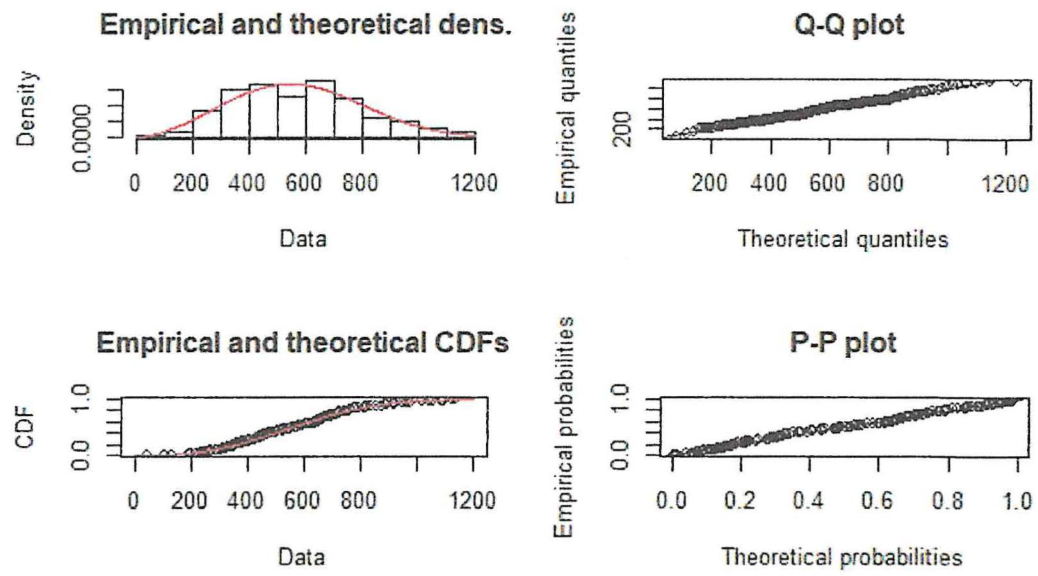
Appendix A3: Lognormal Distribution



Appendix A4: Pareto Distribution

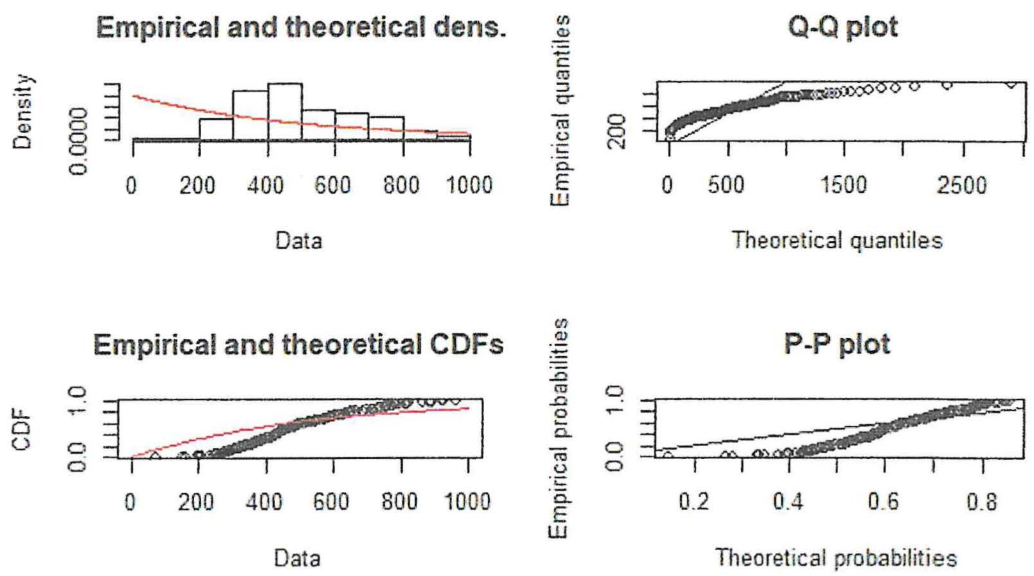


Appendix A5: Weibull Distribution

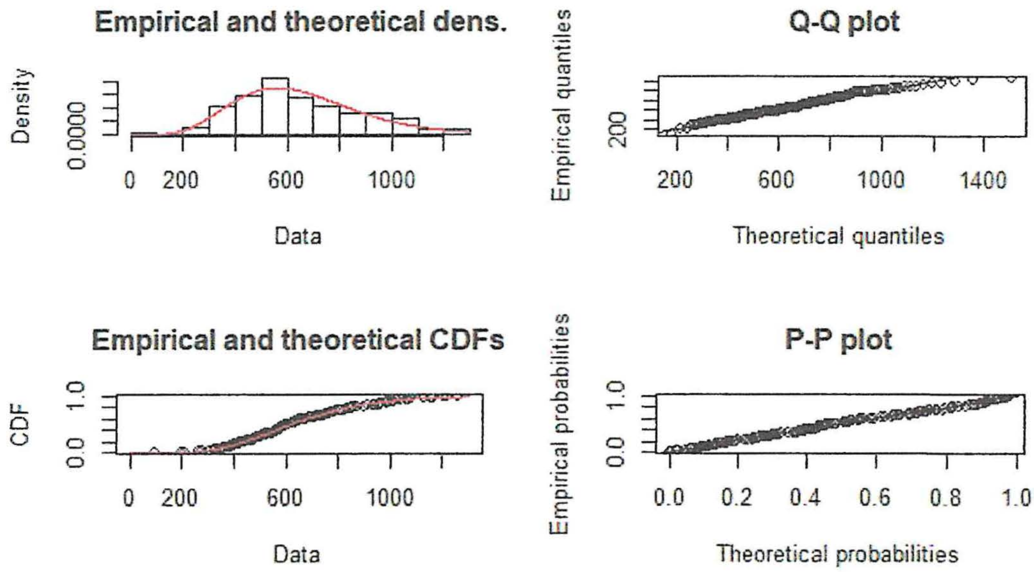


Appendix B: Fitted Distributions for Motor Private

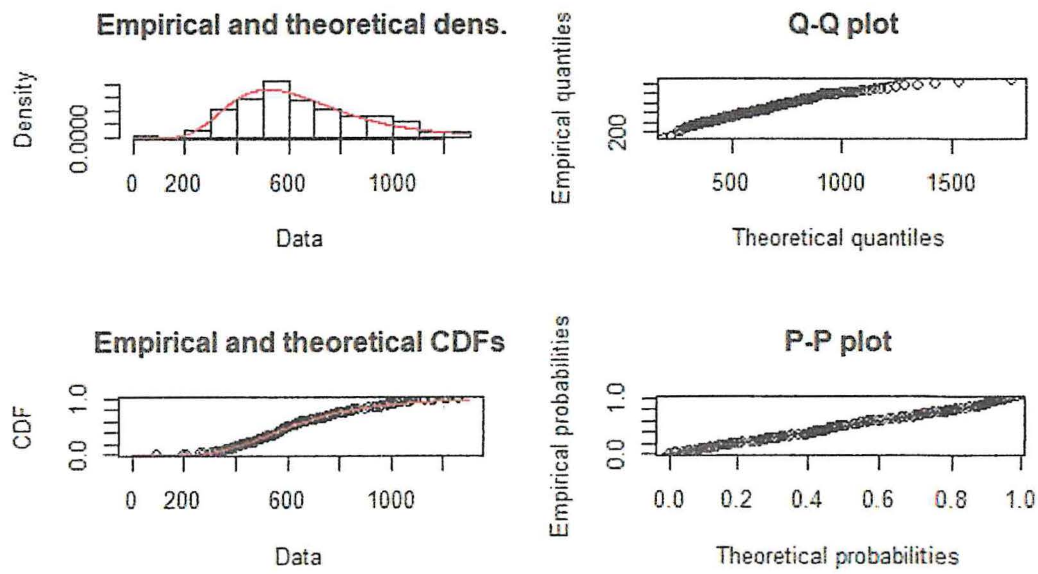
Appendix B1: Exponential Distribution



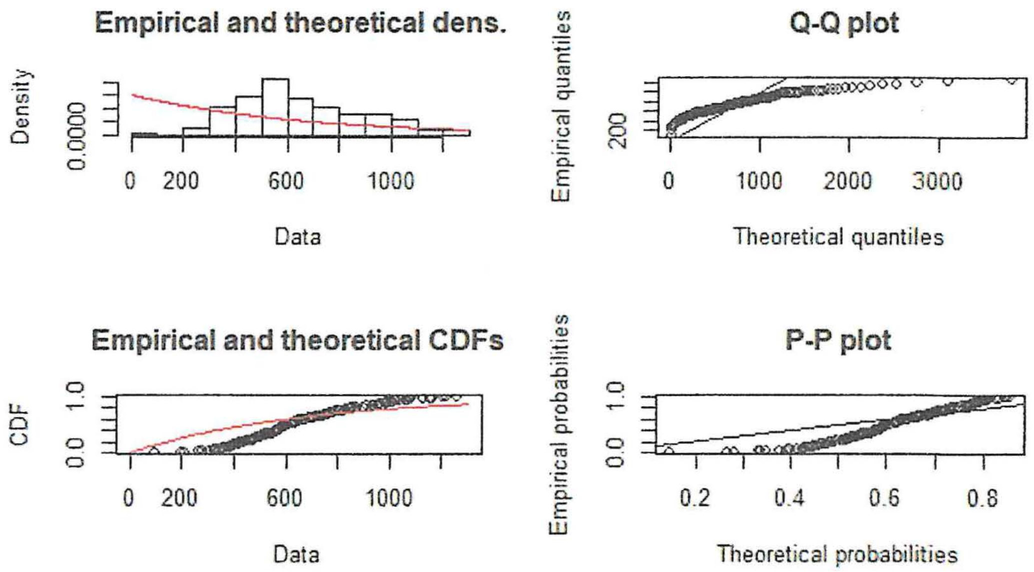
Appendix B2: Gamma Distribution



Appendix B3: Lognormal Distribution



Appendix B4: Pareto Distribution



Appendix B5: Weibull Distribution