KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

LOSS SEVERITY OF CLAIMS AMOUNTS IN MOTOR INSURANCE. A CASE STUDY OF STATE INSURANCE COMPANY

By

ACQUAH DAVID

(BSc. Management Studies)

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS,

KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD

OF MSc. ACTUARIAL SCIENCE.

April 14, 2016

Declaration

I hereby declare that this submission is my own work towards the award of the MSc. degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

Acquah David	....	....
Student	Signature	Date

Certified by:

Dr. A. Y. Omari-Sasu	....	....
Supervisor	Signature	Date

Certified by:

Prof. S. K. Amponsah	....	....
Head of Department	Signature	Date

Dedication

This project work is dedicated to my lovely wife, Sara Efuah Ackon for her support towards my education.

Abstract

General Insurance companies in Ghana are confronted with problems when they want to use past or present claims amounts in forecasting future claim severity. This study seeks to determine an appropriate statistical distribution for the claims amounts of SIC, to determine the posterior distribution of the claims amounts of SIC, and to estimate the expected future claims amounts of SIC using Bayesian methodology. Secondary data from motor policy was obtained from State Insurance Company (SIC) over a period of one year (from January 2015 to December 2015) was used. Claims data that were above GHS 4,000 were considered by the researcher. The analysis of the data was done using Excel, Statistical Package for Social Science (SPSS) and EasyFit software. The research work revealed that the claims data and the posterior distribution of the claim amount followed log-normal distribution and the expected future claim amount is GHS 21,525.27.

Acknowledgements

My humble gratitude goes to God Almighty for his guidance and confidence that has enabled me to complete this work successfully.

I wish to express my most heartily gratitude and thanks to my supervisor, Dr. A.Y. Omari-Sasu for the valuable and useful discussion which invaluably improved this thesis.

I am also very grateful to Nana Kena Frimpong for his valuable suggestions and corrections for this thesis.

My profound thanks go to my wife Mrs. Sarah Efua Acquah and my children Petrout Papa Kwesi Acquah, Euodia Efua Acquah and Nana Kwamena Acquah for their encouragement and commitment towards the course of my career development.

However, I am completely answerable for any limitation that may be detected in this work.

Contents

List of Tables

List of Figures

List of Acronyms

SIC - State Insurance Company Limited

AIC - Akaike's Information Criteria

COTTOR - Committee On Theory Of Risk

P-P Plot - Probability - Probability Plot

ZAIG - Zero - Adjusted Inverse Gaussian

EIW - Exponentiated Inverse Weibull

Chapter 1

Introduction

1.1 Background to the Study

Over three quarters of a million people are killed and tens of millions injured on the roads in low income countries each year. Many, if not most, will come from poor households, particularly vulnerable to the risk of road trauma and its economic consequences. While road safety is traditionally focused on prevention activities, fair and timely compensation systems will help bereaved families and injured victims recover from the shock of a road crash (Murray, 2013).

Claims amounts from insurance companies are sometimes with large data, which are relatively heavy tails. To model such data, heavy tail statistical distributions like gamma, exponential, weibull and lognormal are used (Boland, 2006).

Owing to the rapid growth of the number of motor vehicles operating in developing countries and the obvious economic and social ramifications, there have been various research in the actuarial literature to model such insurance claims reported by insured drivers (Denuit et al, 2007).

1.2 Problem Statement of the Study

In actuarial profession, the ability to model claim depends on the understanding and interpretation of loss distribution as it is vital in making insurance decisions in estimating premiums, expected profits and reserves. Knowledge on distribution of insurance claims can also help in advising insurance companies to consider reinsurance (Boland, 2006).

State Insurance Company has two major motor policies, namely, comprehensive and third party only. The researcher focused on comprehensive motor policies that have large claims amounts. Most insurance fund managers place much emphasis on large claims amounts and in view of this selecting the best model to model the large claims data is vital to them.

A clearer understanding of probability and loss distribution in general insurance is important because it does not only help to summarize and model large amounts of claim but also help to give timely outcomes (Raz & Shaw, 2000).

However, various models were employed to determine the goodness of fit of the models that fit the claims amounts. The aim was to select the best method out of the sample methods that will provide accurate and consistent data for planning.

1.3 Objective of the Study

The objectives of this research work were:

- to determine an appropriate statistical distribution for the claims amounts of State Insurance Company (SIC).
- to determine the posterior distribution of the claims amounts of SIC.
- to estimate the expected future claims amounts of SIC.

1.4 Methodology

This study focused on secondary data from State Insurance Company. Descriptive statistics were employed to ascertain the nature of the data with respect to its mean, median, mode, standard deviation, variance, skewness, kurtosis and sum. Loss distribution such as Pareto, exponential, gamma, log normal and Weibull were used to determine the claims amounts. Bayesian methodology was employed. With this method the likelihood is multiplied with the prior to arrive at the posterior distribution. Continuous uniform distribution was used as a prior for the data. The posterior provides the current distribution.

1.5 Justification of the Study

The significant of the study was to highlight the essence of understanding probability and loss distribution that is used by general insurance. This will enable the insurance company to make decisions such as estimating premiums, expected profits and reserves. Knowledge about the distribution of insurance claims can also help in advising insurance companies to consider reinsurance.

Insurance companies often receive large amounts of claim during certain periods. The design of a suitable loss distribution that will model the severity of claims will enable insurance fund managers to have a better understanding of claims data.

1.6 Limitation of the Study

The research could have considered claims amounts beyond one year but due to time constrains the study was limited to one year claims data.

As a result of time constrain, this research considered five loss distributions. This limited the distributions of the research.

There was difficulty in getting access to data from insurance companies but due to further explanation of the research work to the insurance company the data was finally made available.

1.7 Organization of the Study

The work is organized into five main chapters that is chapters one to five. Chapter One covers introduction, which is made up of background of the study, statement of the problem, objectives, methodology, justification of the study, limitation of the study, and chapter organization. Chapter two deals with the review of related literature, while Chapter three describes the methodology, including the scope of the data, actuarial modeling process, data processing and Bayesian methodology.

Chapter four contains results and discussions of the study and chapter five is summary, conclusions and recommendations of the study.

Chapter 2

Literature Review

2.1 Introduction

This chapter was developed with the motive of reviewing pass works so as to get theories and empirical evidence to support this research.

2.2 Relevant Literature Related to the Research

Finger & Robertson (1976) postulated that an insurance excess layer has inflationary effect over that attributed to overall growth of claim cost. They also claimed that total claim cost increases at an annual rate and the overall rate is less than the trend of the basic limit cost. Their paper discussed the trend that exists between the claim cost and the basic limit cost. They derived a model to estimate the basic limit trend out of the overall trend.

Heckman et al. (1983) discussed aggregate loss distributions from the prospective of collective risk theory. Their aim was to come out with an accurate, reliable and practical method to determine the cumulative probabilities and excess pure premium ratios for loss distributions when claim severity and claim count distributions are considered. They identified the disadvantages of the collective risk model to be the uncertainty of its parameters. They acknowledged that further research need to be carried out in the area of collective risk. They emphasized the relevance of testing the predictions of the collective risk model against the actual total claim loss. Secondly, they stress the need to test the sensitivity of the collective risk model to violations of the assumptions underlying it.

Guiahi (2000) provided better statistical method for fitting parametric distributions to loss data involving deductibles, policy limits and rating variables. The presence of deductibles and policy limits did not give him a good fit and he resorted into using Akaike's Information Criteria (A.I.C). The criteria employed by Guiahi for estimating which probability distribution best suit the claims data was the value of Akaike's Information Criteria, A.I.C. With this criterion the model with the smaller AIC value is considered to be the appropriate one. The paper also discussed statistical tests of hypotheses to know the effect of rating variables on loss distribution. In his view lognormal was the best fit of the claims data.. A statistical package, S-Plus, were used to compute the maximum likelihood estimate (MLE) of the model parameters. Finally, he concluded that a lot of statistical models could be used to fit distributions to claims data.

Renshaw (2004) focused on modeling both the claim frequency and claim severity components of the claims process in general insurance in the presence of rating factors using the quasi-likelihood. The maximum likelihood estimates was used by him to estimate the parameters. In his view the quasi-likelihood parameter estimates exhibits similar asymptotic properties as the maximum likelihood parameter estimates. His research also indicated how selected parameterized variance functions can be used to model heterogeneity in the claim frequency process and to provide a parameterized family of claim response variables. He introduced parameterized power link function which includes the log-link as a special case.

Fiete (2005) developed model to fit into 490 claims amounts drawn from 7 various years and used maximum likelihood estimate to estimate parameter values. He used P-P plots to know the goodness of fit of all possible outcome to avoid depending on a single number. He used several statistical distributions such as gamma, lognormal, weibull, inverse weibull and Pareto to fit the insurance claim severity. He also assumed no zero claims. According to him Pareto distribution had the best of fit and hence assumed that claims within each year are generated by a Pareto distribution.

Wright (2005) also presented actuarial modeling on his COTOR answer using inverse Pareto, Pareto, burr, Paralogistic, inverse Paralogistic, log logistic, Pearson VI, inverse burr, log-normal and restricted benktander families. He used actuarial procedure to fit models to 490 claims of 7 different years and also used maximum likelihood method to estimate parameters. In his research the P-P plots and Kolmogorov-smirnov test (K-S test) was used to determine the effectiveness of fit. The restricted benktander family is a one-shape-parameter sub-family of the benktander II family. It has the property (like Pareto and exponential families) that left-truncation gives another distribution in the same family. The research was salient on this particular distribution because it is useful for modeling excess loss amounts but the study focused on claims.

Meyers (2005) fitted distributions to 250 claims and the distributions used was lognormal, weibull and gamma. Meyers based his methodology on Bayesian solution and used maximum likelihood estimate to determine the parameters. This means that Meyers multiplied the likelihood with the prior to derive the posterior distribution. In this particular study the likelihood was used to calculate the A.I.C with the aim of selecting the appropriate probabilistic model for the claims amounts.

Shevchenko (2010) researched into calculation of aggregate loss distribution, estimated the operational risk capital under the loss distribution approach requires evaluation of aggregate (compound) loss distributions which is one of the classic problems in risk theory. According to him distributions that are used in operational risk, closed-form solutions are not available. However, modern computers use numerical methods to calculate this distribution. His research addressed numerical algorithms that can be used to compute aggregate loss distributions. Monte Carlo, Panjer recursion and Fourier transformation methods were compared.

Chapter 3

Methodology

3.1 Introduction

The main aim of this chapter was to discuss the various methods adopted for this study. This study focused on the loss severity of claims amounts using the State Insurance Company as case study. This chapter discusses the scope of data, actuarial modeling process, data processing and Bayesian methodology.

3.2 Scope of Data

Secondary data from motor policy was obtained from State Insurance Company. A period of one year (January, 2015 - December, 2015) data was used.The researcher wanted to consider higher claims and upon consultation with the claims manager at SIC, higher claims in their insurance company are GHS 4000 and beyond. Upon studying the data from January, 2015 - December, 2015, the higher claims were 135. The data used in the study follows the following assumptions:

1. All motor vehicles registered under the policy must make a claim (No zero claims in a year).

2. The claims are independent and identically distributed

3. All future claims should be derived from the same distribution

3.3 Descriptive Statistics

Creswell (2009) in his research explains that information on current data is mainly concerned with describing the nature of the present situation of the data. In this study the descriptive statistics was used so as to know the expected claims above GHS 4,000.00 at SIC. The mean, median and mode shows skewness of the individual claim amount. This helped the researcher to fit the positively skewed distributions to the State Insurance Company data.

3.4 Actuarial Modeling Process of Claims Amounts

In general insurance the frequency and severity of claim analysis is important because it helps in estimating future expected claims for pricing. The modeling process of general insurance on claims amounts are done by continuous probability distributions such as Pareto, exponential, gamma, Weibull, lognormal distributions, etc. The log-likelihood function was used as a tool to select the distribution that fit the claims data. That is the distribution with the highest log-likelihood estimate was selected to be the best fit. The expectation was used to determine the claims amounts that an individual policyholder could claim. The definitions of the above distributions are given below:

3.4.1 Pareto Distribution, $P(\alpha,\lambda)$

The Pareto distribution function is denoted as $P(\alpha,\lambda)$. The random variable X has Pareto distribution denoted as $X \sim P(\alpha,\lambda)$, if its probability function (PDF)

is:

$$Parameters : \alpha,\lambda(\alpha > 0,\lambda > 0)$$

$$PDF : f(x) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}, x > 0 \quad (3.1)$$

let $x_i, i = 1,2,...n$ be the claims amounts, then the likelihood function is given as:

$$L = \prod_{i=1}^{n} f(x/\alpha,\lambda) \quad (3.2)$$

$$L = \prod_{i=1}^{n} \frac{\alpha\lambda^{\alpha}}{(\lambda + x_i)^{\alpha+1}} \quad (3.3)$$

$$L = \alpha\lambda^{\alpha}(\lambda + x_1)^{-(\alpha+1)} \times \alpha\lambda^{\alpha}(\lambda + x_2)^{-(\alpha+1)} \times \cdots \times \alpha\lambda^{\alpha}(\lambda + x_n)^{-(\alpha+1)} \quad (3.4)$$

∴ The likelihood function is:

$$L = \alpha^{n}\lambda^{n\alpha} \prod_{i=1}^{n} (\lambda + x_i)^{-(\alpha+1)} \quad (3.5)$$

Take logs. This will usually simplify the algebra.

$$\log l = n\log\alpha + n\alpha\log\lambda - (\alpha + 1)\sum_{i=1}^{n}\log x_i \quad (3.6)$$

Expectation of Pareto Distribution

Write the first x in the integral as $(\lambda + x) - \lambda$ and make the integral "look like"Pareto probabilities

$$\int_0^{\infty} x \cdot \alpha\lambda^{\alpha}(\lambda + x)^{-\alpha-1}dx = \int_0^{\infty} [(\lambda + x) - \lambda]\alpha\lambda^{\alpha}(\lambda + x)^{-\alpha-1}dx \quad (3.7)$$

Splitting up the integral into two bits and inserting appropriate constants so that both integrals look like Pareto PDF's:

$$\int_0^\infty x \cdot \alpha\lambda^\alpha(\lambda+x)^{-\alpha-1}dx = \frac{\alpha\lambda}{\alpha-1}\int_0^\infty(\alpha-1)\lambda^{\alpha-1}(\lambda+x)^{-\alpha}dx - \lambda\int_0^\infty \alpha\lambda^\alpha(\lambda+x)^{-\alpha-1}dx \tag{3.8}$$

$$= \frac{\alpha\lambda}{\alpha-1}P[0 < Pa(\alpha-1,\alpha) < \infty] - \lambda P[0 < Pa(\alpha,\lambda) < \infty] \tag{3.9}$$

$$= \frac{\alpha\lambda}{\alpha-1} - \lambda \tag{3.10}$$

$$= \frac{\lambda}{\alpha-1} \tag{3.11}$$

3.4.2 Exponential Distribution, *Exp*(λ)

The exponential distribution is denoted as *Exp.*(λ). The random variable X has a exponential distribution denoted as $X \sim Exp.(\lambda)$, if its probability density function (PDF) is:

$$Parameter : \lambda(\lambda > 0)$$

$$PDF : f(x) = \lambda e^{-\lambda x}, x > 0 \tag{3.12}$$

Let $x_i, i = 1,2,...,n$ be the claims amounts, then the likelihood function is given as:

$$L = \prod_{i=1}^{n} f(x|\lambda) \tag{3.13}$$

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

$$\lambda e^{-\lambda x_1} \times \lambda e^{-\lambda x_2} \times \cdots \times \lambda e^{-\lambda x_n} \tag{3.14}$$

∴ The likelihood function is:

$$L = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \qquad (3.15)$$

Take logs. This will usually simplify the algebra.

$$\log l = n\log\lambda - \lambda\sum_{i=1}^{n} x_i \qquad (3.16)$$

Expectation of Exponential Distribution

In calculating for the expectation of the exponential distribution the log-likelihood was differentiated

$$\frac{d\log l}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \qquad (3.17)$$

$$\frac{d\log l}{d\lambda} = 0 \qquad (3.18)$$

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \qquad (3.19)$$

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i} \qquad (3.20)$$

$$\sum_{i=1}^{n} x_i = \frac{n}{\lambda} \qquad (3.21)$$

$$E(X) = \frac{1}{n} \times \sum_{i=1}^{n} x_i \qquad (3.22)$$

$$E(X) = \frac{1}{n} \times \frac{n}{\lambda} \qquad (3.23)$$

$$E(X) = \frac{1}{\lambda} \qquad (3.24)$$

3.4.3 Gamma Distribution, $Gamma(\alpha,\lambda)$

If X follows a gamma distribution with parameter $\left\{\Gamma\left(\alpha,\frac{1}{\lambda}\right)\right\}$ then its probability density function is:

$$PDF: f(x) = \frac{x^{\alpha-1}e^{\frac{-x}{\lambda}}}{\Gamma(\alpha)\lambda^{\alpha}}, x > 0 \qquad (3.25)$$

Let $x_i, i = 1,2,...,n$ be the claims amounts, then the likelihood function is given

as:

$$L = \prod_{i=1}^{n} f(x|\alpha,\lambda) \qquad (3.26)$$

$$L = \prod_{i=1}^{n} \frac{x^{\alpha-1}e^{\frac{-x}{\lambda}}}{\Gamma(\alpha)\lambda^{\alpha}} \qquad (3.27)$$

$$= \frac{x_1^{\alpha-1}e^{\frac{-x_1}{\lambda}}}{\Gamma(\alpha)\lambda^{\alpha}} \times \frac{x_2^{\alpha-1}e^{\frac{-x_2}{\lambda}}}{\Gamma(\alpha)\lambda^{\alpha}} \times \cdots \times \frac{x_n^{\alpha-1}e^{\frac{-x_n}{\lambda}}}{\Gamma(\alpha)\lambda^{\alpha}} \qquad (3.28)$$

$$L = \frac{\sum_{i=1}^{n} x_i^{\alpha-1} e^{\frac{-\sum_{i=1}^{n} x_i}{\lambda}}}{\Gamma(\alpha)^n \lambda^{\alpha n}} \qquad (3.29)$$

Take logs. This will usually simplify the algebra.

$$\log l = (\alpha-1)\sum_{i=1}^{n}\log x_i - \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n\log\Gamma(\alpha) - n\alpha\log\lambda \qquad (3.30)$$

Expectation of Gamma Distribution

$$\frac{d\log l}{d\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda^2} - \frac{n\alpha}{\lambda} \qquad (3.31)$$

$$\frac{d\log l}{d\lambda}$$

$$\frac{\sum_{i=1}^{n} x_i}{\lambda^2} - \frac{n\alpha}{\lambda}$$

$$\lambda = \frac{n\alpha}{\sum_{i=1}^{n} x_i}$$

$$\sum_{i=1}^{n} x_i = \frac{n\alpha}{\lambda}$$

= 0 (3.32)

= 0 (3.33)

(3.34)

(3.35)

$$E(X) = \frac{1}{n} \times \sum_{i=1}^{n} x_i \quad (3.36)$$

$$E(X) = \frac{1}{n} \times \frac{n\alpha}{\lambda} \quad (3.37)$$

$$E(X) = \frac{\alpha}{\lambda}$$

(3.38)

3.4.4 Weibull Distribution

The weibull distribution is denoted as *weibull*(c,γ). The random variable X has a weibull distribution denoted as $X \sim$ *weibull*(c,γ), if its probability density function (PDF) is:

Parameters : $c,\gamma (c > 0, \gamma > 0)$

$$PDF : f(x) = c\gamma x^{\gamma-1}e^{-cx^{\gamma}}, x > 0 \quad (3.39)$$

Let $x_i, i = 1,2,...,n$ be the claims amounts, then the likelihood function is given as:

$$L = \prod_{i=1}^{n} f(x|c,\gamma) \quad (3.40)$$

$$L = \prod_{i=1}^{n} c\gamma x^{\gamma-1}e^{-cx^{\gamma}} \quad (3.41)$$

$$L = c\gamma x_1^{\gamma-1}e^{-cx_1^{\gamma}} \times c\gamma x_2^{\gamma-1}e^{-cx_2^{\gamma}} \times \cdots \times c\gamma x_n^{\gamma-1}e^{-cx_n^{\gamma}} \quad (3.42)$$

∴ The likelihood function is:

$$L = c^n\gamma^n \sum_{i=1}^{n} x_i^{\gamma-1}e^{-c\sum_{i=1}^{n} x_i^{\gamma}} \quad (3.43)$$

Take logs. This will usually simplify the algebra.

$$\log l = n\log c + n\log\gamma + (\gamma-1)\log\sum_{i=1}^{n} x_i - c\sum_{i=1}^{n} x_i^{\gamma} \quad (3.44)$$

Expectation of Weibull Distribution

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx \quad (3.45)$$

$$= \int_{-\infty}^{\infty} x \cdot c\gamma x^{\gamma-1}e^{-cx^{\gamma}}dx \quad (3.46)$$

Let

$$cx^{\gamma} = t$$

$$x = \left(\frac{t}{c}\right)^{\frac{1}{\gamma}}$$

$$c\gamma x^{\gamma-1}dx = dt$$

$$E(X) = \int_0^{\infty} \left(\frac{t}{c}\right)^{\frac{1}{\gamma}} \cdot e^{-t}dt \quad (3.47)$$

$$= c^{-\frac{1}{\gamma}} \int_0^{\infty} t^{\frac{1}{\gamma}+1-1} \cdot e^{-t}dt$$

(3.48)

The Gammalin function is:

$$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt \quad (3.49)$$

$$\therefore E(X) = c^{-\frac{1}{\gamma}}\Gamma\left(1 + \frac{1}{\gamma}\right)$$

(3.50)

3.4.5 Log-normal Distribution, $LN(\mu,\sigma^2)$

A lognormal distribution function is denoted as $LN(\mu,\sigma^2)$. A random variable X has a lognormal distribution, denoted as $X \sim LN(\mu,\sigma^2)$, if its probability function (PDF) is:

$$Parameters : \mu, \sigma^2(\sigma > 0)$$

$$PDF : \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x}\exp\left(-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right), \sigma > 0, \mu \in \mathbb{R} \quad (3.51)$$

Let $x_i, i = 1,2,...,n$ be the claims amounts, then the likelihood function is given

as:

$$L = \prod_{i=1}^{n} f(x_i/\mu,\sigma^2) \quad (3.52)$$

$$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{1}{2} \left(\frac{\log x - \mu}{\sigma} \right)^2 \right) \tag{3.53}$$

$$L = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x_1} \exp\left(-\frac{1}{2} \left(\frac{\log x_1 - \mu}{\sigma} \right)^2 \right) \times \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x_2} \exp\left(-\frac{1}{2} \left(\frac{\log x_2 - \mu}{\sigma} \right)^2 \right) \times \cdots \times \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x_n} \exp\left(-\frac{1}{2} \left(\frac{\log x_n - \mu}{\sigma} \right)^2 \right) \tag{3.54}$$

∴ The likelihood function is:

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \frac{1}{\prod_{i=1}^{n} x_i} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{\log x_i - \mu}{\sigma} \right)^2 \right) \tag{3.55}$$

Take logs. This will usually simplify the algebra.

$$\log l = -n \log \sigma\sqrt{2\pi} - \log \prod_{i=1}^{n} x_i - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{\log x_i - \mu}{\sigma} \right)^2 \tag{3.56}$$

Expectation of Log-normal Distribution

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \tag{3.57}$$

$$= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2} dx \tag{3.58}$$

Let

$$t = \frac{\log x - \mu}{\sigma} - \sigma$$

$$\frac{dt}{dx} = \frac{1}{x\sigma}$$

$$dx = x\sigma dt$$

However,

$$t + \sigma = \frac{\log x - \mu}{\sigma}$$

$$x = e^{\sigma t + \sigma^2 + \mu}$$

$$E(X) = \int_{-\infty}^{\infty} e^{\sigma t + \sigma^2 + \mu} \cdot \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(t+\sigma)^2} x\sigma dt \qquad (3.59)$$

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\mu + \frac{1}{2}\sigma^2 - \frac{1}{2}t^2} dt$$

$$= e^{\mu + \frac{1}{2}\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

$$= e^{\mu + \frac{1}{2}\sigma^2} [\phi(\infty) - \phi(-\infty$$

(3.60)

(3.61)

)] (3.62)

$$= e^{\mu + \frac{1}{2}\sigma^2} [1 - 0] \qquad (3.63)$$

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2} \qquad (3.64)$$

3.5 Checking model fit

It was assumed that none of the sampled models considered was true and because of that the aim was to select the best approximating model (Anderson & Burnham, 2004). The mere fact that out of the sampled distributions only one had the highest log-likelihood estimate does not necessarily means that it could provide a good fit to the claim data. A further test had to be conducted to ascertain whether that selected distribution actually fit the data by using P-P Plots and

Akaike's Information Criteria (A.I.C.).

3.5.1 The Probability-Probability (P-P) plot

The probability-probability (p-p) plot is a probability plot for assessing how closely two data sets agree, which plots the two cumulative distribution functions against each other. It is used to evaluate the skewness of a distribution. The plot will be approximately linear if the specific theoretical distribution is the correct model (Wilk & Gnanadesikan, 1968).

3.5.2 The Akaike's Information Criteria (AIC)

The AIC is a type of criteria used in selecting the best model for making inference from a sample of models. With this criteria, the model with the smallest AIC value is considered to be the desirable one. This is because the model is estimated to be the closest to the unknown truth among the models considered (Anderson & Burnham, 2004). The AIC is defined by:

AIC = -2 (maximized log-likelihood) + 2 (no. of parameters estimated) In

this research, the AIC was used to check the goodness of fit.

3.6 Data processing

The researcher used computer statistical packages to perform tests and graphical presentations. EasyFit software was used for fitting distribution, the parameter estimation and computation of likelihood values. SPSS was used to performed descriptive analysis of the data and plotting of the histograms.

3.7 Determination of posterior distribution of claim amount

After the goodness of fit test, the posterior distribution was determined by multiplying the likelihood function of the claims amounts by the prior distribution under the Bayesian methodology.

$$f(\mu/x_i) \quad \propto \quad \prod_{i=1}^{n} f(x_i/\mu) \times f(\mu)$$

3.7.1 Likelihood function

The researcher determined the likelihood functions of the claims amounts after the goodness of fit of the distribution of claims amounts is known. If the likelihood is based on a set of known values $x_1,x_2,...,x_n$, then the likelihood function will take the form:

$$f(x_1/\mu)f(x_2/\mu)\cdots f(x_n/\mu)$$

where $f(x)$ is the PDF of the likelihood to be fitted.

$$L(\mu) = \prod_{i=1}^{n} f(x_i/\mu)$$

3.7.2 Prior distribution

Frees et al (2014) stated that, there may not be any prior information about a particular parameter. Alternatively, with all the inherent concerns about overconfidence, a more objective approach may be preferred even when expert opinion is available. The researcher used continuous uniform distribution to be the prior because of non-availability of expert opinion. In view of this the researcher assumed that each claims amounts have equal chance of occurring. With a finite sample space, a natural choice is to divide the prior density evenly throughout the space.

Also, if the parameter is continuous over a bounded interval, $[a,b]$, then

$$\pi(\mu) = \frac{1}{b-a}$$

3.8 Determining expected future claims amounts from the posterior distribution

1. If the likelihood is based on a set of known values $x_1, x_2, \ldots, x_n$, then the likelihood function will take the form $f(x_1/\mu)f(x_2/\mu)\cdots f(x_n/\mu)$ where f(x) is the PDF of the distribution that is to be fitted.

$$L(\mu) = \prod_{i=1}^{n} f(x_i/\mu)$$

2. Take logs. This will usually simplify the algebra

$$l(\mu) = \log L(\mu) = \sum_{i=1}^{n} \log f(x_i/\mu)$$

3. This usually involves differentiating the log-likelihood function with respect to the parameter(s), and setting the resulting expression(s) equal to zero.

$$\frac{d}{d\mu} l(\tilde{\mu}) = 0$$

4. Solve the resulting equation(s) to find the parameters.

Chapter 4

Analysis and Findings

4.1 Introduction

The purpose of this chapter is to present the results and discuss the empirical findings of the study. The chapter is divided into six sections; we start with the summary statistics, followed by interpretation of histograms, maximum likelihood estimates, the log-likelihoods, goodness of fit test and Bayesian methodology. The actuarial modeling process that was used by the researcher and the results arrived at would be shown numerically and graphically. Conclusions will be drawn on the findings in order to identify which loss distribution best fit the claims amounts.

4.2 Summary Statistics

The descriptive statistics of the claims amounts of SIC under motor insurance were computed in order to show the vital features of the data. A sample size of 135 claims was analyzed. The descriptive statistics of the amount of claims were computed using SPSS.

The table 4.1 below shows the descriptive statistics.

Table 4.1: Descriptive Statistics of claim amount in SIC

n	135
Mean	2.9537E4
Median	1.8000E4
Mode	1.60E4
Standard deviation	3.2502E4

Variance	1.056E9
Skewness	3.248
Kurtosis	11.852
Sum	3.99E6

Source: Computed from SIC data

Explanation to Table 4.1

The data is heavily skewed and therefore most distributions that show such features can be used to model the claims. From the table 4.1, the descriptive statistics of the claims data were not normally distributed because the mean, median and mode with values of GHS 29,537.00, GHS 18,000.00 and GHS 16,000.00 respectively are not equal. The skewness is 3.248 which shows that the data is positively skewed. The standard deviation which shows the variation from the mean is 32,502. This means that most of the data lies between ±32,502 away from the mean. The kurtosis of 11.852 shows that the data is having high peak.

The sum of the claim amount is GHS 3,990,000.00.

4.3 Interpretation of Histograms

4.3.1 Histogram of the Claims Amounts of SIC

The histogram of the claims amounts of SIC in 2015 is shown in figure 4.1.

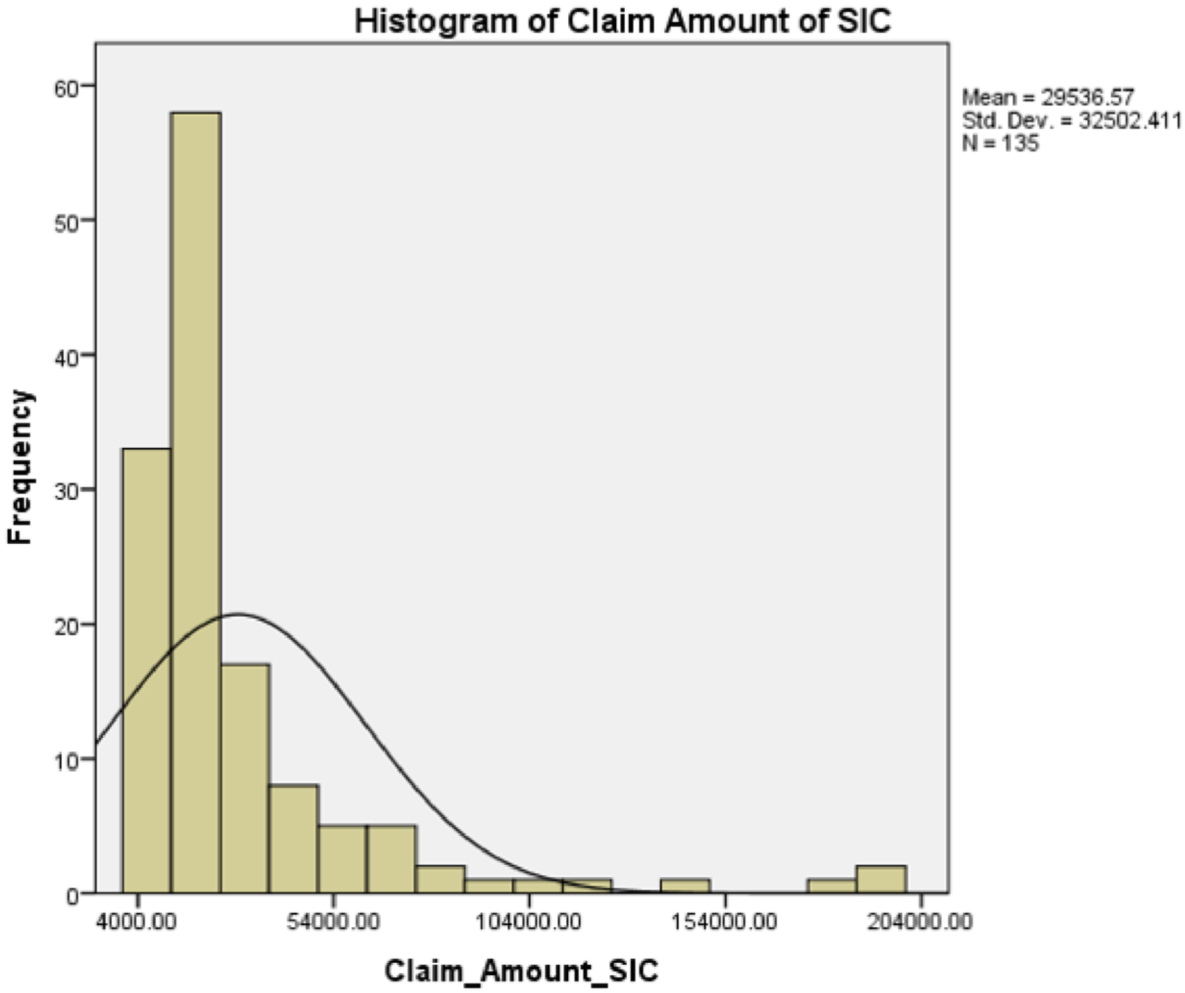

Figure 4.1: Histogram showing claims amounts of SIC

There were gaps in the histogram such that it may not have been practical to begin fitting a distribution in its original form. The histogram has a normal curve superimposed in it. The curve shows the skewness of the claims data. It can be observed from the diagram that, the original claims data has a heavy right-hand tail. This means that very few of the claims amounts were of high values whiles most of the claims amounts were of low values.

4.4 Descriptive Statistics of log of claims amounts in SIC

Table 4.2 below shows the descriptive statistics of log claims amounts in SIC:

Table 4.2: Descriptive Statistics of log of claims amounts in SIC

n	135
Mean	9.9775
Median	9.7981
Mode	9.68
Standard deviation	0.70821
Variance	0.502
Skewness	1.106
Kurtosis	0.910
Sum	1346.96

Explanation to Table 4.2

From table 4.2, the descriptive statistics of the log of claims amounts shows the mean, median and mode with values of 9.9775, 9.7981 and 9.68 respectively are equal. The skewness which is still positively skewed has reduced to 1.106. The standard deviation which shows the variation from the mean has also reduced to 0.70821. This indicates that majority of the data lies between ±0.70821 away from the mean. The peakness of the data has reduced to 0.910. The sum of the claim amount is now 1346.96.

4.4.1 Histogram of the log of claims amounts

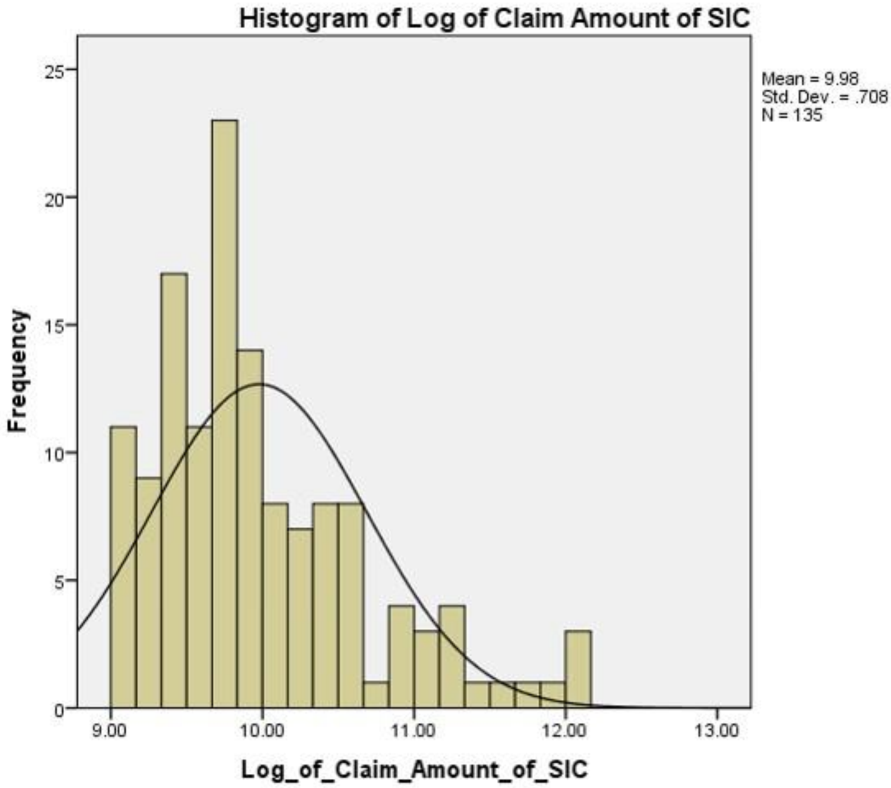

Figure 4.2: Histogram showing log of claims amounts in SIC

In an attempt to reduce the skewness and increasing the shape of the bars of the original data log transformation was employed. The log of the claims data was computed using Excel. From figure 4.2 the bars of the histogram are well pronounced and the normal curve is less skewed. The log transformation was used to model the claims data. The normal curve that was superimposed on the histogram is relatively normally distributed. Therefore, the log of the claims amounts was chosen and the maximum likelihood was used to fit the five distributions. A closer observation of figure 4.2 revealed that the log of the claims data could be used to fit the five distributions that were selected. At this stage the P-P plot, parameter estimation and the test statistics at 0.01 significance level for the five distributions were computed.

4.5 P-P Plot of the Five Distributions

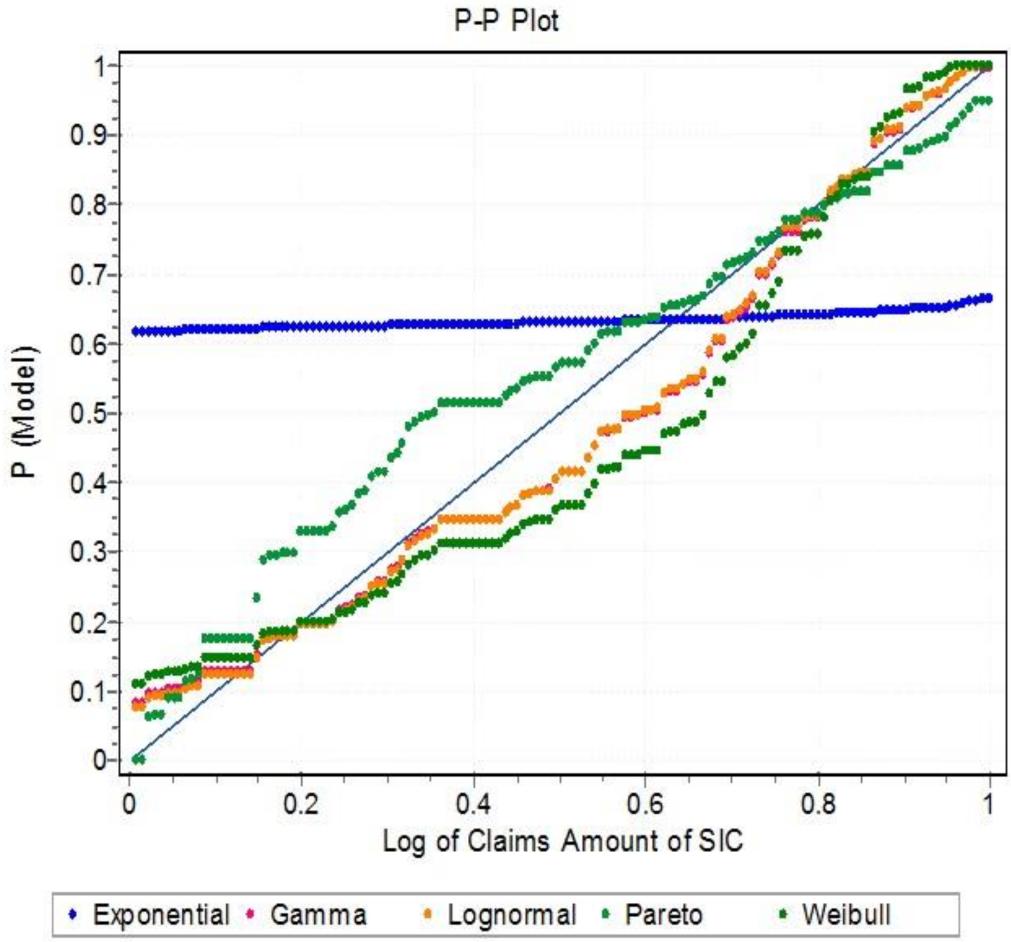

Figure 4.3: P-P plot showing the log of the claim amount of the five distributions

From the graph the exponential distribution depicts the worst fit among the distributions. This is because the points are very far from the reference line. In the case of Pareto and Weibull distributions at the ends of the reference line, some of the points are on it but the point seems to move further away from the middle of the reference line. These two distributions do not provide good fit to the data. However, the gamma and lognormal distributions shows great signs of providing good fit to the data as both were very close to the reference line and it was very difficult to determining the best one. Further computation was done by using AIC to know which one best fit the claims data.

4.6 Test Statistics of Distributions

Test statistics was conducted on State Insurance claims to obtain the best fit.

The test statistics table is shown below.

4.6.1 Test Statistics for Pareto Distribution

Table 4.3: Showing the test statistics for Pareto distribution of claims data

Pareto					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	135 0.15335 0.00308				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.09235	0.10526	0.11688	0.13065	0.1402
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size Statistic	135 5.8009				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Degree of freedom Statistic P-Value	7 21.661 0.00291				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	9.8032	12.017	14.067	16.622	18.475
Reject?	Yes	Yes	Yes	Yes	Yes

Explanation to Table 4.3

The Kolmogorov-Smirnov test, Anderson-Darling and Chi-square goodness of fit test from table 4.3 indicates that at 99% confident level Pareto distribution do not fit the claims amounts of SIC. There is a rejection of all the test statistics as indicated in table 4.3 at 1% significant level showing that the P-value of 0.00308 and 0.00291 of Kolmogorov-Smirnov and Chi-Square tests are less than the 1 % significant level. The statistic of 0.15335, 5.8009 and 21.661 of KolmogorovSmirnov, Anderson-Darling and Chi-Square test are greater than the critical values of 0.1402, 3.9074 and 18.475. State Insurance claims data do not follow Pareto

distribution.

4.6.2 Test Statistic for Exponential distribution

Table 4.4: Showing the test statistics for Exponential distribution of claims data

Exponential					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	135 0.5962 0				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.09235	0.10526	0.11688	0.13065	0.1402
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size Statistic	135 54.139				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Degree of freedom Statistic P-Value	6 1286.8 0				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	8.5581	10.645	12.592	15.033	16.812
Reject?	Yes	Yes	Yes	Yes	Yes

Explanation to Table 4.4

The Kolmogorov-Smirnov, Anderson-Darling and Chi-square goodness of fit test from table 4.4 indicates that at 99% confident level Exponential distribution do not fit the claims amounts of SIC. There is a rejection of all the test statistics as indicated in table 4.4 at 1% significant level showing that the p-value of 0 of

Kolmogorov-Smirnov and Chi-Square test are less than the 1% significant level.

The statistic of 0.5962, 54.139 and 1286.8 of Kolmogorov-Smirnov, AndersonDarling and Chi-Square test are greater than the critical values of 0.1402, 3.9074 and 16.812 respectively. State Insurance claims data do not follow exponential

distribution.

4.6.3 Test Statistic for Gamma distribution

Table 4.5: Showing the test statistics for Gamma distribution of claims data

Gamma					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	135 0.1239 0.02896				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.09235	0.10526	0.11688	0.13065	0.1402
Reject?	Yes	Yes	Yes	No	No
Anderson-Darling					
Sample Size Statistic	135 2.9068				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	Yes	No	No
Chi-Squared					
Degree of freedom Statistic P-Value	7 17.536 0.01425				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	9.8032	12.017	14.067	16.622	18.475
Reject?	Yes	Yes	Yes	Yes	No

Explanation to Table 4.5

The Kolmogorov-Smirnov, Anderson-Darling and Chi-square goodness of fit test from table 4.5 indicates that at 99% confident level gamma distribution fit the claims amounts of SIC. There is an acceptance of all the test statistics as indicated in table 4.5 at 1% significant level showing that the P-value of 0.02896 and 0.01425 of Kolmogorov-Smirnov and Chi-Square test are greater than the

1% significant level. The statistic of 0.1239, 2.9068 and 17.536 of KolmogorovSmirnov, Anderson-Darling and Chi-Square test are greater than the critical values of 0.1402, 3.9074 and 18.475 respectively. State insurance claims data follow gamma distribution.

4.6.4 Test Statistic for Weibull distribution

Table 4.6: Showing the test statistics for Weibull distribution of claims data

Weibull					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	135 0.17912 2.9208E-4				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.09235	0.10526	0.11688	0.13065	0.1402
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size Statistic	135 7.6618				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Degree of freedom Statistic P-Value	7 26.278 4.4921E-4				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	9.8032	12.017	14.067	16.622	18.475
Reject?	Yes	Yes	Yes	Yes	Yes

Explanation to Table 4.6

The Kolmogorov-Smirnov, Anderson-Darling and Chi-square goodness of fit test from table 4.6 shows that at 99% confident level Weibull distribution do not fit the claims amounts of SIC. There is a rejection of all the test statistics as indicated in table 4.6 at 1% significant level showing that the P-value of 2.9208E-

4 and 4.4921E-4 of Kolmogorov-Smirnov and Chi-Square test are less than the

1% significant level. The statistic of 0.17912, 7.6618 and 26.278 of KolmogorovSmirnov, Anderson-Darling and Chi-Square test are greater than the critical values of 0.1402, 3.9074 and 18.475 respectively. State insurance claims data do not follow Weibull distribution.

4.6.5 Test Statistic for Log-normal distribution

Table 4.7: Showing the test statistics for lognormal distribution of claims data

Lognormal					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	135 0.11847 0.04152				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.09235	0.10526	0.11688	0.13065	0.1402
Reject?	Yes	Yes	Yes	No	No
Anderson-Darling					
Sample Size Statistic	135 2.703				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	Yes	No	No
Chi-Squared					
Degree of freedom Statistic P-Value	7 16.559 0.02047				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	9.8032	12.017	14.067	16.622	18.475
Reject?	Yes	Yes	Yes	No	No

Explanation to Table 4.7

The Kolmogorov-Smirnov, Anderson-Darling and Chi-square goodness of fit test from table 4.7 indicates that at 99% confident level lognormal distribution fit the claims amounts of SIC. There is an acceptance of all the test statistics as indicated in table 4.7 at 1% significant level showing that the P-value of 0.04152 and 0.02047 of Kolmogorov-Smirnov and Chi-Square test are greater than the

1% significant level. The statistic of 0.11847, 2.703 and 16.559 of KolmogorovSmirnov, Anderson-Darling and Chi-Square test are greater than the critical values of 0.1402, 3.9074 and 18.475 respectively. State Insurance claims data follow lognormal distribution.

The table 4.3 to table 4.7 show that gamma and lognormal has been accepted by the test statistics at 0.01 significance level as the best fit of the log of claims amounts of SIC. It is also noted that Pareto, Weibull and Exponential distributions were rejected by the test statistics at 0.01 significance level and hence cannot be used to fit the claims data. At this stage out of the five sampled distributions only two were accepted at 0.01 significance level. The researcher conducted a further analysis on the two distributions (gamma and lognormal distributions) to identify which one of them fit the claims data best.

4.7 Maximum Likelihood Estimates

The log of the claims data were used to compute the maximum likelihood estimates and the parameter estimates of the two sampled distributions. Table 4.8 below shows the parameters of the two statistical distributions having been fitted to the claims data.

Table 4.8: Showing computed log-likelihoods, A.I.C. and parameters

Distributions	Log-likelihood	A.I.C	Parameters	
Gamma	-141.324814	286.6496279	$\alpha = 198.48$	$\lambda = 0.05027$
Lognormal	-67.50028	139.00056	$\sigma = 0.06845$	$\mu = 2.2979$

The log-likelihood was used as a tool to select the distribution that fits the data. It is also important to note that EasyFit was used to obtain the parameter values of the sampled distributions. From table 4.8, the log-normal and gamma distributions had log-likelihood values of -67.5003 and -141.3248 respectively. This showed that the log-normal distribution had the highest log-likelihood value.

However, the lognormal distribution had the smallest value of AIC(139.0006). It was therefore concluded that the log-normal distribution was the best fitted distribution among the two distributions for the claims data since it had the smallest value of AIC. That is, the log-normal distribution was estimated to be the closest to the unknown true distribution among the two remaining distributions. On the basis of this, the log-normal distribution was the best fit of the sampled distributions.

As compared to Guiahi (2000) researched into issues and methodologies for fitting alternative statistical distributions to samples of insurance data. Lognormal was selected to be the best fit for the data. In his research he used the method of maximum likelihood to estimate model parameters and also his criteria for comparing which probability distribution fits the data set best was based upon the value of akaikes information criteria, AIC. That is, the model with the smaller

AIC value is the more desirable one.

4.8 Determining the Expected Claim amount using the Bayesian Methodology

The researcher then calculated the expected claims amounts of SIC using the Bayesian Methodology. In order to determine the expected claims amounts, the likelihood function of the best fit model (log-normal) of the claims amounts must be combined with the prior distribution(continuous uniform distribution) to obtain the Posterior Distribution. The expectation of the posterior distribution will then be the expected claim for SIC.

4.8.1 Determining Likelihood function

Let $X_i, i = 1,2,\ldots,n$ be the random claims amounts

$$X \sim lognormal(\mu,\sigma^2)$$

Holding the scale parameter constant and varying the shape parameter. Likelihood

$$L = \prod_{i=1}^{n} f(x_i/\mu,\sigma^2) \quad (4.1)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x}\exp\left(-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2\right) \quad (4.2)$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x_1}\exp\left(-\frac{1}{2}\left(\frac{\log x_1-\mu}{\sigma}\right)^2\right) \times \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x_2}\exp\left(-\frac{1}{2}\left(\frac{\log x_2-\mu}{\sigma}\right)^2\right)$$
$$\times\cdots\times \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x_n}\exp\left(-\frac{1}{2}\left(\frac{\log x_n-\mu}{\sigma}\right)^2\right) \quad (4.3)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \frac{1}{\prod_{i=1}^{n} x_i}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{logx_1-\mu}{\sigma}\right)^2\right) \quad (4.4)$$

4.8.2 Determining the Posterior Distribution

The posterior distribution is the product of the likelihood and the prior. It is denoted as Posterior ∝ likelihood function x prior.

$$f(\mu/x_i,\sigma^2) = \prod_{i=1}^{n} f(\mu/x_i,\sigma^2 \times f(\mu) \quad (4.5)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \frac{1}{\prod_{i=1}^{n} x_i}\exp(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{logx_1-\mu}{\sigma}\right)^2 \times \frac{1}{b-a} \quad (4.6)$$

$$\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{logx_1-\mu}{\sigma}\right)^2\right) \quad (4.7)$$

The posterior distribution follows a log-normal distribution.

4.9 Determining expected future claims amounts

The researcher then used the posterior distribution to determine the expected claims. The log-likelihood of the posterior distribution were obtained in order to find the expected claims for SIC.

$$\log f(\mu/x_i,\sigma^2) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\log x_i-\mu}{\sigma}\right)^2 \quad (4.8)$$

$$= -\frac{\sum_{i=1}^{n}\log x^2}{2\sigma^2} + \frac{\mu\sum_{i=1}^{n}\log x_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}$$

$$\frac{d\log f(\mu/x_i,\sigma^2)}{d\mu} = \frac{\sum_{i=1}^{n}\log x_i}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$\frac{d\log f(\mu/x_i,\sigma^2)}{d\mu}$$

$$\frac{\mu\sum_{i=1}^{n}\log x_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}$$

$$\mu = \frac{\sum_{i=1}^{n}\log x_i}{n} = \frac{310.2216337}{135} = 2.$$

$$E(X) = e^{\mu+\frac{1}{2}\sigma^2} = e^{2.297938027} + \frac{1}{2}0.06845^2 = 9.$$

(4.9)

(4.10)

= 0 (4.11)

= 0 (4.12)

297938027 (4.13)

976982883 (4.14)

The antilog of the claims amounts of 9.976982883 is 21525.27. Therefore, the expected claim amount of SIC is GHS 21,525.27. This means that an individual policy holder is likely to make a claim of GHS 21,525.27 from State Insurance

Company based on their previous data.

Chapter 5

Conclusion and Recommendations

This chapter seeks to present a brief summary on the content of the research work undertaken in this thesis and outline the major conclusions that were derived from the empirical results. The next section discusses recommendation and suggested areas for future research.

5.1 Summary of results

The research was carried out at SIC of Ghana. The study adopted the secondary data approach by using claims amounts above GHS 4,000. The study analyzed 135 claims data from SIC. The descriptive statistics had mean, median, mode and skewness to be 2.9537E4, 1.8000E4, 1.60E4 and 3.248 respectively. The mean, median and mode were not equal. The data was positively skewed. There were gaps in the histogram and it would not have been appropriate to start fitting a distribution in its original form. Log transformation was employed to reduce the skewness of the claims data. The claims data was logged once and the skewness of the data was reduced to 1.106. The mean, median and the mode of the log of claims data were almost equal. The log transformation was used to model the claims data.

EasyFit, Excel and SPSS softwares were used in computing the appropriate statistical distributions that best fit the insurance claims data. It was revealed in the analysis in chapter four that the lognormal distribution provided the best model for the claims amounts of comprehensive motor insurance of SIC.

According to table 4.3 to table 4.7 the gamma and lognormal distributions was accepted by the test statistics as the best fit for the log of claims amount of SIC. It was noted that Pareto, Weibull and exponential distribution could not be used to fit the claims amounts of SIC as it was rejected by the test statistics shown in table 4.3, table 4.4 and table 4.6 under 0.01 significant level. However, from table 4.8, the lognormal distribution had the highest log-likelihood function of -67.5003, which implies that among the chosen statistical distributions it had a greater chance in providing a good fit for the claims data. The lognormal distribution was then followed by the gamma distribution which had a log-likelihood of -141.3248. The outcome of the log-likelihood means that the gamma distribution could not provide an appropriate model for the claims data. The lognormal distribution had the highest log-likelihood function. The AIC was computed to test whether the lognormal distribution provided the best fit for the claims data. The AIC confirmed that the lognormal distribution had the smallest value of 139.0006 indicating that it had the best fit for the claims data. The gamma distributions followed in that order with AIC value of 286.6496, showing that it would not be regarded as a good fit to the claims data.

The P-P plots for the statistical distributions was plotted to graphically re-affirm the goodness of fit test computed by the AIC. Figure 4.3 showed that the P-P plot of the lognormal distribution provided the best fit to the claims data as almost all the points are on the reference line.

It was revealed in chapter four that the posterior of the claims amounts follows a log normal distribution. The expected claims amounts was GHS 21,525.27.

5.2 Conclusion

The study revealed that the appropriate statistical distribution for the claims amounts of SIC was lognormal distribution.

The research showed that the posterior distribution of the claims amounts of SIC is a lognormal distribution.

Finally, the expected future claims amounts of SIC was GHS 24,525.27.

5.3 Recommendations

The sample size used for the study could be increased to enhance the reliability of the expected future claims amounts of SIC.

The posterior distribution could be used by researchers as their informative prior for further calculations of their posterior.

Further research could be conducted in this area by introducing more continuous distributions to further improve the accuracy of the distribution that best fits the claims data.

References

Achieng, O. M., & NO, I. (2010). Actuarial Modeling for Insurance Claim Severity in Motor Comprehensive Policy Using Industrial Statistical Distributions. In International Congress of Actuaries (pp. 7-12).

Anderson, D. R. & Burnham, K. P., (2004). Multimodel inference understanding

AIC and BIC in model selection. Sociological methods & research, 33(2), 261-304.

Bartoszewicz, B. (2005). Modelling the Claim Count with Poisson Regression and Negative Binomial Regression. In Innovations in Classification, Data Science, and Information Systems (pp. 103-110). Springer Berlin Heidelberg.

Boland, P. J. (2006). Statistical methods in general insurance.

Claro, D. P. (2009). Estimating claim size and probability in the auto-insurance industry: the zeroadjusted Inverse Gaussian (ZAIG) distribution (No. wpe 159). Insper Working Paper, Insper Instituto de Ensino e Pesquisa.

Creswell ,W. J. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. SAGE Publications, Incorporated.

Denuit, M., MarA˜ c chal, X., Pitrebois, S., & Walhin, J. F. (2007). Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems. John Wiley & Sons.

Dudley, C. (2006). Bayesian analysis of an aggregate claim model using various loss distributions (Doctoral dissertation, Dissertation thesis for Master of

Science in Actuarial Management. Heriot-Watt University, Edinburgh).
Fiete S. (2004) COTOR Challenge Round 3, available at www.casact.org/cotor/ Fiete.doc

Finger, R. J., & ROBERTSON, D. B. J. P. (1976). A Note on Basic Limits Trend Factors. PCAS LXIII, 106.

Frees, W. E, Derrig, A. R and Meyer, G (2014). Predictive Modeling Applications in Actuarial Science. Volume 1 pp. 315-326

Guiahi F. (2000) 'Fitting to loss distributions with emphasis on rating variables'. available at www.casact.org/pubs/forum/01wforum/01wf133.pdf

Heckman, Philip E. Meyers, Glenn G. (1983), 'The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions', Proceedings of the Casualty Actuarial Society Casualty Actuarial Society - Arlington, Virginia 1983: LXX 22 - 61.

Meyers.(2004) COTOR Challenge Round 2, available at www.casact.org/cotor

Mahmoud, O. H. (2014). Construction Actuarial Model for Aggregate Loss under

Exponentiated Inverted Weibull Distribution. Applied Mathematical Sciences, 8(162), 8085-8097.

Murray, R. H. (2013). Robert Muir-Wood, MA Ph. D., Chief Research Officer, Risk Management Solutions (RMS). The Geneva Reports, 91.

Raz .O. and Shaw .M. (2000) 'An Approach to preserving sufficient correctness in open resource coalitions, 10th International Work-shop on software specification & Design (IWSSD-10) IEEE Computer Society, November 2000 .'

Renshaw, A. E. (1994). Modelling the claims process in the presence of covariates. Astin Bulletin, 24(02), 265-285.

Shevchenko P. V., (2010), ’Calculation of aggregate loss distributions’, The Journal of Operational Risk 5 (2), pp. 3 - 40.

Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika, 55(1), 1 - 17

Wright T. (2005) FIA, Deloitte, UK. COTOR Challenge Round 3, available at www.casact.org/cotor/