University of Nairobi
College of Biological and Physical Sciences
School of Mathematics

A Stochastic Analysis of Claim Reserving in General Insurance Using Bootstrapping Technique

By:Edmond Ochieng' Ochola

Research Report in Actuarial Science, Number 17,2018

A dissertation submitted to the graduate school in partial fulfillment of the requirements for the award of Master of Science in Actuarial Science.
August 2018

Declaration and approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

EDMOND OCHIENG' OCHOLA,

Reg No. I56/79940/2015.

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

Signature Date

Prof Patrick G.O. Weke,
School of Mathematics,
University of Nairobi,
P.O Box 30197, 00100 Nairobi, Kenya.
E-mail: pweke@uonbi.ac.ke

Abstract

Insurance is one of the key sector in ensuring stability and growth in any economy. Reserving actuaries therefore should do the best they can to ensure that the reserves declared are as accurate as possible. In most insurance companies reserves are calculated using the traditional methods which are based on certain algorithms. These include Chain ladder method, Average cost method,Bornhuetter-Ferguson and Standards method among others.

These methods do not take into consideration the actual claim development. Generalised linear models (GLMs) and Bootstrap technique can be used as an alternative as they use the various covariates of the claim process in determining the reserves. In this study I will focus in showing that Bootstrap technique and Generalized Linear Models gives a realistic and structured method of loss reserving. This is because they incorporate more information about the claim process such as Type of claim, Loss, development pattern, Pattern of loss emergence etc.

This makes it possible to determine the predictive distribution of the reserve model and to calculate various measures of risk such as the Value at Risk(VaR) at various levels.

Keywords;Generalized Linear Modes,Chain ladder, Bootstrap,Gamma Model, Over-dispersed Poisson Model

Dedication

This project is dedicated to my mother, Jane Achieng' Ochola. Thank you for your great support, loyalty and love. You are the best.

Contents

1 INTRODUCTION **1**

1.1 BACKGROUND 1

1.2 STATEMENT OF PROBLEM 2

1.3 OBJECTIVES 2

1.3.1 GENERAL OBJECTIVE 2

1.3.2 THE SPECIFIC OBJECTIVES 2

1.4 SIGNIFICANCE OF THE STUDY 2

1.5 THE ORGANIZATION OF THE PROJECT 3

2 LITERATURE REVIEW **4**

3 CHAIN LADDER METHOD AND GENERALIZED LINEAR MODELS **6**

3.1 INTRODUCTORY TO CLAIM RESERVING AND RUN OFF TRIANGLES 6

3.1.1 THE THEORY OF CLAIM RESERVING 6

3.1.2 TYPE OF RESERVES 7

3.1.3 THE STATISTICAL REPRESENTATION OF RUN OFF TRIANGLES 7

3.2 THE CHAIN LADDER RESERVING METHOD 9

3.3 GENERALIZED LINEAR MODELS 10

3.3.1 THE EXPONENTIAL FAMILY OF DISTRIBUTIONS 10

3.3.2 THE LINEAR PREDICTOR 11

3.3.3 THE LINK FUNCTION 12

3.4 MODEL FITTING 12

3.4.1 PARAMETER ESTIMATION 12

3.4.2 ANALYZING THE GOODNESS OF FIT 13

3.5 RESERVE ANALYSIS WITHIN THE FRAMEWORK OF GENERALIZED LINEAR MODELS (GLM) 13

3.5.1 THE GAMMA DISTRIBUTION 14

3.5.2 THE OVER DISPERSED POISSON MODEL (ODP) 15

3.6 BOOTSTRAPPING 15
3.6.1 INTRODUCTION 15
3.6.2 THE PREDICTION ERROR 17
3.6.3 BOOTSTRAPPING THE ODP MODEL 17
3.6.4 BOOTSTRAPPING THE GAMMA MODEL 18

4 APPLICATION 19
4.1 THE DATA SET 19
4.2 THE STOCHASTIC ANALYSIS OF THE MODELS 23
4.2.1 DETERMINING THE SCALED DEVIANCE STATISTIC AND THE PEARSON χ^2 STATISTIC 23
4.2.2 CLAIM RESERVE CALCULATION USING THE BOOTSTRAPPING TECHNIQUE 23
4.3 FITTING A DISTRIBUTION TO THE IBNR 30

5 CONCLUSIONS AND RECOMMENDATIONS 31
5.1 CONCLUSIONS 31
5.2 RECOMMENDATIONS 32

6 APPENDIX 34
6.1 R Code 34
6.1.1 Importing The Data 34
6.1.2 Chain Ladder Reserving 35
6.1.3 Stochastic Reserving 35

List of Figures

4.1 Graph of cumulative claims against development years 21
4.2 The CDF of reserve for the ODP bootstrap model 25
4.3 The Histogram of reserve for the ODP bootstrap model 26
4.4 The CDF of reserve for the GAMMA bootstrap model 28
4.5 The Histogram of reserve for the GAMMA bootstrap model . . 29

List of Tables

3.1 Table of claim process in general insurance 6
3.2 Claim development triangle 8

4.1 Development triangle of incremental claims 19
4.2 Development triangle of cumulative claims 20
4.3 Table of development factors 22
4.4 The Full Development Triangle of cumulative claims for both the original claims and projected claims 22
4.5 The outstanding reserve per origin year 22

Acknowledgment

First and foremost I would like to express my sincere gratitude to the almighty god for the grace to undertake my Masters studies.

My utmost gratitude to my supervisor, Prof Patrick G.O. Weke ,for his continuous support in my Msc study and project, for his encouragement, support, patience and invaluable guidance. His advise assisted me throughout my Masters studies.

My thanks to the University of Nairobi for the great opportunity to do my Master studies. My gratitude also goes to to my fellow classmates for their cooperation, feedback, and friendship. their support meant a lot to me.

Lastly, my heart felt gratitude my family and friends for the great support throughout my studies and my life in general.

Chapter 1

INTRODUCTION

1.1 BACKGROUND

Insurance is one of the key sector in ensuring stability and growth in any economy. Reserving actuaries therefore should do the best they can to ensure that the reserves declared are as accurate as possible. The accuracy of the reverse is affected by many factors which need to be taken in to consideration in order to ensure proper management of insurance companies. Most companies use the traditional algorithm based methods which leaves out such important factors that affects reserving. Some of the details that need to be factored while reserving include, but not limited to the type of claim, loss development pattern, pattern of loss emergence etc. These can be achieved by doing loss reserving using bootstrapping technique and Generalized Linear Models (GLM). Stochastic models have the following significance;

1. Incorporate more details in loss reserving hence the results is more accurate compared to the other algorithm based methods.

2. Effects of factors affecting loss reserves can be studied separately

3. It makes it easier to deal with changing circumstances as variables are analyzed continuously.

In Kenya, stochastic reserving is one of the claim receiving method that have been included in the insurance act. However, most insurance companies still use the other algorithm bases methods. Considering the random nature of claim occurrence this thesis is expected to encourage and help in the wide spread us of stochastic reserving method. With the introduction of risk based capital model of business in Kenya it is very important for reserves to be determined accurately so as to add value to the various stake holders in the insurance company.

1.2 STATEMENT OF PROBLEM

Most general insurance companies in Kenya have been using algorithm based loss reserving technique such as Chain ladder method among others in determining loss reserves. These methods are quite simple and fast to apply. However algorithm based loss reserving techniques give a point value of loss reserve which do not give more information about the claim development patterns and the distribution of the claim reserves. The availability of this information can greatly help insurance companies to improve their enterprise risk management function by calculating the various risk measures.

In this thesis we seek to examine how Actuaries in general insurance can apply Bootstrapping technique and Generalized Linear Models to improve claim loss reserving.

1.3 OBJECTIVES

1.3.1 GENERAL OBJECTIVE

The general objective of this project is to demonstrate that algorithm based loss reserving models might be overlooking some key factors that need to be incorporates in loss reserving and show that the stochastic loss reserving can provide alternative to the algorithm based methods widely used in loss reserving.

1.3.2 THE SPECIFIC OBJECTIVES

The specific objectives of the project are as below;

1. To show how Bootstrap Technique and Generalized Linear Models (GLMs) can provide a more elaborate and a comprehensive framework for loss reserving.

2. To obtain the predictive distribution of the claim reserves.

3. to calculate the Value at Risk of the reserves at various levels.

1.4 SIGNIFICANCE OF THE STUDY

It is very important for insurance companies to ensure that the reserve set aside is adequate in meeting future claim obligations. This put actuaries in a position where they need to keep improving the technique they us in loss reserving.

The significance of the study is to show that the Bootstrapping technique and Generalized Linear Models can be applied to get the loss reserve distribution. This can then be used to calculate the various measures of risk like Value at Risk at a given confidence interval. This can significantly help insurance companies to improve loss reserving and risk management.

1.5 THE ORGANIZATION OF THE PROJECT

The second chapter is the literature review of the various research that have been done in the area of chain ladder reserving technique, generalized Loss Reserving and bootstrapping . Chapter three discusses Chain Ladder reserving technique, Generalized Linear Models as well as the Bootstrapping technique. Chapter four looks at data analysis , calculating claim reserves using Chain Ladder technique, bootstrap Gamma model and the bootstrap Over-dispersed Poisson Model. Conclusion and recommendations for future research are given in chapter five. Lastly a list of reference is finally given.

Chapter 2

LITERATURE REVIEW

Various scholars and practising actuaries have done various studies on the area of loss reserving. In this chapter we will look at some of the studies on loss reserving.

Schmidt (2006) in his work, Methods and Models of Loss Reserving Based on Run-Off Triangles: A Unifying Survey, gives detailed discussion of various loss reserving models that uses the run-off triangles. He notes that the success of the various methods that are based on the run off triangles are greatly assumes that the loss development of each origin year follows the same pattern of development.

S Haberman and A E Renshaw (1999) in their excellent and comprehensive paper describe the various applications of the Generalized Linear Models (GLM) in actuarial work. These included the analysis of the various loss distribution in the general insurance, analysis of the lapse rate with policy characteristics in the life insurance and analyzing the changes in force of mortality in the underwriting of life assurance. Various suitable models are used in all these cases. In conclusion it is noted that there is a great scope if applying GLM in various other actuarial areas such as group life risk premium and marine insurance risk premium.

Hoedemakers et al. (2005) constructs bound for the discounted loss reserves within the framework of GLMs. In particular the upper and lower bound are determined followed by some numerical examples to illustrate the importance of the constructed bounds. Generalised Linear Models(GLMs) were used to model the los reserve while considering stochastic discount factors. The Brownian motion was used in modeling the stochastic discounting factors. it is shown that GLM offers an opportunity to model the claim according to some particular distribution of exponential family such like Gamma Poisson,independent normal etc. Due to the difficulty of presenting the distribution function in explicit form, the approximations to the distribution are used.

In their paper titled “Practitioner’s Guide to Generalized Linear Models” Duncan et al(2007) third edition, discuss ways in which practicing actuaries can use GLMs to analyze insurance date. The paper comprehensively discuses GLM in three sections. The first section discuss the statistical theory of GLM, giving various examples. This basically entails how GLMs are formulated and solved. A comprehensive background,upon which GMLs are built is described including the linear models and the minimum bias procedures. The various methods of solving GLMs ie the maximum likelihood estimates and numerical techniques are well explained. The second part provide the model output for each part of GLM analysis. These are data preparation,model selection, model refinement and model interpretation. A plan of undertaking GLM analysis is well set out as well as how to interpret the various results. The third section deals with topics that applies GLM such as retention modeling and scoring algorithms. The paper also describes how GLMs can be used in credit-based insurance scores. The paper demonstrates that GLM has a lot of advantages and these can be used by insurance companies to achieve competitive advantage and improve profitability. GLMs can be used, among many other things, to analyze the effect of various factors on the experience and provide information about certainty of model results.It is clear,from the paper, that GLM can be applied to areas of insurance industry including but not limited to pricing, underwriting, marketing and reserving as well as in other industries.

Greg Taylor and Grainne McGuire(2004) in their paper, Loss reserving with GLMs: a case study, discuses an application of generalized linear models,GLM to loss reserving. It is noted that although chain ladder method is widely used by actuaries it is subject to a number of restrictions. It is noted that some loss/claim data may present some features that may violate the conditions of chain ladder model. Generalized linear models are presented as an alternative form of data analysis that can be used in claim reserving. This enables the modeling and analysis of the various features data that violates the conditions of chain ladder models.The models are compared using different diagnostics. Greg Taylor and Grainne McGuire shows that stochastic modeling of claim reserves provides an opportunity of choosing the form of distribution,assumed to be followed by the claim data. The paper concludes that there is need to incorporate stochastic modeling and data analysis in claim reserving.

England and Verrall (2002) in their paper; Stochastic claim reserving in general insurance, Emphasises the application of GLM framework in the analysis of claim reserving. Various GLM models including Gamma model, log-normal models etc are discussed. It is noted predictive distribution claim reserves can be obtained using bootstraping. The Bornhuetter-Ferguson reserving method and Bayesian technique are used. This provides a way to incorporate experts' opinion to give prior estimate of ultimate claims. England and Verrall concludes that general insurance companies should incorporate stochastic claim reserving for a full and more comprehensive analysis of their financial well being.

Chapter 3

CHAIN LADDER METHOD AND GENERALIZED LINEAR MODELS

3.1 INTRODUCTORY TO CLAIM RESERVING AND RUN OFF TRIANGLES

Here we will discus theory of claim reserving in non-life insurance and also give an introduction to the run off triangle. We will look at the frame work of runoff triangles, the statistical models for run off triangles and also introduce the various notations used.

3.1.1 THE THEORY OF CLAIM RESERVING

At the end of each financial year, an insurance company would want to calculate the amount of surplus that it can declare to its shareholders. To do this the company has to determine that amount money it need to set aside in order to pay any outstanding claim amount that it received premium during that particular financial year.Claim settlement stages in non-life insurance can be expressed as shown below;

Claim event occurred	====>	Claim reported	====>	Claim payment(s) made	====>	Claim file closed

Table 3.1: Table of claim process in general insurance

Normally, it might take some time from the occurrence of loss to the time where the full the total claim amount which has to be settled is determined. Some of the reasons for such delays are;

1. Delay in reporting claim to an insurance company
2. A claim amount payable may need to be determined by a court process which may take some times
3. An insurance company may need to carry of investigation in to the reported cases to determine its viability
4. A closed claim can be reopened after some years.

Due to the delays an insurance company has to determine the amount of money it need to set aside, out of the premiums it has received in order to take care of any future claim.

3.1.2 TYPE OF RESERVES

Typically reserves can be categorized in two depending on the settlement stage that claim has reached.

1. Claims reserve required in respect of claims that have been incurred but not reported these are referred to as the Incurred But Not Reported (IBNR),
2. Claims reserve is needed for claims that have been reported, but not yet been closed. These are known as to as Reported But Not Settled reserves.

3.1.3 THE STATISTICAL REPRESENTATION OF RUN OFF TRIANGLES

The reported claim data is usually represented in triangular form called the run off triangles. The vertical axis represents the origin or accident year i of the claim and the horizontal axis represent the development year j of the claim. "I" represent the last year of claim occurrence and "J" represent the maximum number of claim development years.

In most cases, though not always, I=J.Suppose that for accident year i and development year j, the incremental claim amount is represented by C_{ij}.

Then, the upper triangle can be denoted as;

$$D_I^U = \{C_{ij} : i + j \leq I; 0 \leq j \leq J\} \tag{3.1}$$

and the lower triangle can be denoted as;

$$D_I^L = \{C_{ij} : i + j \geq I; 0 \leq j \leq J\} \tag{3.2}$$

Each entry, C_{ij} , in the run-off triangle represents the incremental claims and can be expressed in general terms; Each C_{ij} can be denoted as;

$$C_{ij} = r_j * s_i * x_{i+j} + e_{ij} \tag{3.3}$$

Where:

1. r_j is the development factor for year j, representing the proportion of claim payments in Development Year j. Each r_j is independent of the Origin Year i.

2. s_i is a parameter varying by Origin Year, i, representing the exposure, for example the number of claims (or claim amount) incurred in the Origin Year i.

3. x_{i+j} is a parameter varying by calendar year, for example representing inflation.

4. e_{ij} is an error term.

The general statistical form of runoff triangles can be presented as;

	Development Years						
Accident years	1	2	...	j	...	J-1	J
1	$C_{1,1}$	$C_{1,2}$	...	$C_{1,j}$	...	$C_{1,J-1}$	$C_{1,J}$
2	$C_{2,1}$	$C_{2,2}$	...	$C_{2,j}$	...	$C_{2,J-1}$	$C_{2,J}$
...	...	...	i+j=<I	...	...	...	...
i	$C_{i,1}$	$C_{i,2}$	...	$C_{i,j}$	...	$C_{i,J-1}$	$C_{i,J}$
...	...	...	...	...	$i+j>=I$	...	...
I-1	$C_{I-1,1}$	$C_{I-1,2}$	...	$C_{I-1,j}$	...	$C_{I-1,j-1}$	C_{I-1J}
I	$C_{I,1}$	$C_{I,2}$	...	$C_{I,j}$	...	$C_{I,J-1}$	$C_{I,J}$

Table 3.2: Claim development triangle

C_{ij} can represent either the incremental claims numbers or the incremental claim amount.

The cumulative claim amount or number can be expressed as D_{ij}, where i is the accident year and j is the development year. Then the can get D_{ij} as;

$$D_{ij} = \sum_{k=1}^{j} C_{ik} \tag{3.4}$$

The value C_{ij} is only known for $i+j<I$ these are the observed claims and represents the upper triangle. The lower triangle represents the claim amounts,C_{ij} where $i+j>I$, to be predicted. We can define R_i, the claim reserve outstanding for accident year I as;

$$R_i = D_{ij} - D_{i,n-i+1}, 1 \leq j \leq J \tag{3.5}$$

The total claim reserve outstanding, R, will therefore be define as;

$$R = \sum_{i=1}^{I} R_i \tag{3.6}$$

This is what the insurance company will as reserve. In the next section we will briefly look at chain ladder reserving. Most of the other claim reserving techniques are an improvement of chain ladder method.

3.2 THE CHAIN LADDER RESERVING METHOD

The chain ladder method assumes the following;

1. Each accident year has the same pattern of claim development
2. Weighted average if inflation in the past years will be experienced in the future since inflation is one for the factors that is carried on in to the future by the development factors.

The chain ladder reserving is used within run off triangles' framework as introduced in the earlier section. The algorithm of basic chain ladder method is as follows;

1. The exists development factors, f_j,defined as;

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j+1} C_{ij}}{\sum_{i=1}^{I-j+1} C_{ij-1}} \tag{3.7}$$

2. The factor are then used to forecast the future cumulative claim reserves by applying them to the cumulative claims on each row as follows;

$$\hat{C}_{ij-i+2} = C_{ij-i+1} * \hat{f}_{j-i+2} \quad \text{for some} \quad 2 \leq j \leq J \tag{3.8}$$

and for the K^{th} row we have

$$\hat{C}_{ik} = \hat{C}_{ik-i} * \hat{f}_j \quad \text{for some} \quad 2 \leq j \leq J \quad \text{and} \quad 3-i+n \leq k \leq n. \tag{3.9}$$

However this loss reserving technique is usually sensitive to claim variability and outlying values. In the following chapter we will introduce the concept of Generalized Linear Models. Then in the section that will follows we will discuss claim reserving within the frame work of Generalized Linear Models.

3.3 GENERALIZED LINEAR MODELS

In this section we will introduce thestatistical frame work of Generalized Linear Models, its assumption and characteristics. GLM allow us to assume that the data does not come from a normal distribution. This assumption is importation because most actuarial data are usually not normarmally distributed e.g. poisson distibustion is used to model claim frequency, while claim severity cam be modeled using expornential, gamma or lognormal distributons. GLM relate the relates the response variables to the independent variable, the variable that we have information about. GLM can be divided into 3 components as described in McCullagh and Nelder (1989);

1. **The distribution** of the data, eg. Exponential, Normal Poisson etc.
2. **A linear predictor**, η.
3. **The link function**, $g(.)$.

Below is a brief description of each of these components;

3.3.1 THE EXPONENTIAL FAMILY OF DISTRIBUTIONS

The GLMs assumes that the observed random variable are drawn from the exponential family of distribution. As discussed in McCullagh and Nelder (1989), a probability distribution function, $f(y;\theta,\phi)$, is a member of exponential family if it can be written as;

$$f(y;\theta,\phi) = \exp(\frac{(y\theta - b(\theta))}{a(\phi)} - c(y,\phi)) \tag{3.10}$$

Where;
y is the random variable of an observation Y
θ; the location parameter also referred to as the canonical parameter,
ϕ; the dispersion parameter also referred to as the scale parameter.
$b(.)$; the cumulant parameter, which determines the distribution shape.
$c(y,\phi)$ produces the total unit ma of the distribution, is the normalizing factor.

The mean and variance of exponential family of distribution are expressed as below;

$$E[Y] = \mu = \frac{d}{dt}m_y(t)|_{t=0} = b^{'}(\theta) \tag{3.11}$$

$$Var[Y] = \frac{d^2}{dt^2}m_y(t)) - [\frac{d}{dt}m_y(t)]^2|_{t=0} = a(\phi) * b^{''}(\theta) \tag{3.12}$$

We use prime to express differentiation with respect to θ.

From the above,it is clear that the variance of a random variable Y is a product of $b^{''}(\theta)$, referred to as the variance function, dependent only on the canonical parameter and the function $a(\phi)$, which does not depend on θ . The variance function can be denoted as $V(\mu)$ because the dependence on the canonical parameter implies its dependency on μ. Normal distribution is the only one where the variance does not depend on the mean since $V(\mu) = 1$.

3.3.2 THE LINEAR PREDICTOR

It is a function of covariates and is linear to the parameter. Suppose we have y, a vector of observation such that;

$$y = (y_1, \ldots y_n) \tag{3.13}$$

Suppose we have a nxp $(p \leq n)$ matrix X, of independent (response) variables and a vector of un known parameters β, of dimension p, such that

$$\beta = (\beta_i \ldots \beta_p)^T \tag{3.14}$$

Then we can express the mean of vector Y denoted as vector μ as follows;

$$\mu = X\beta \tag{3.15}$$

The assumption is that the error(random) part of the model are normally and independently distributed. We can represent the linear predictor as below;

$$\eta = \sum_{i=1}^{p} X_i \beta_i \tag{3.16}$$

where;

1. x_i are a vector of the covariate matrix, X.

2. β_i is a vector, of size p, of unknown parameters.

The covariate in can interact in various ways depending on whether they are dependent or independent.

3.3.3 THE LINK FUNCTION

The linear predictor, $\eta = g(\mu)$ and the response variables are linked together by the link function. Normally each distribution has an accepted link function that gives μ which fits that particular distribution. This is known as canonical link function.

3.4 MODEL FITTING

In the previous sections we introduce the Generalized Linear Models and discussed its various components. In this section we will look at the various ways of fitting the model ie parameter determination, deviance analysis to test the suitability of the model and testing the parameter significance.

3.4.1 PARAMETER ESTIMATION

Using the Maximum Likelihood Estimation

The Maximum Likelihood Estimation is applied in the estimation of parameters. The log-likelihood functions are determined depending on the various link functions. To achieve this the log-likelihood is with respect to the respective parameters. McCullagh and Nelder (1989) gives more information on this method.

Using quasi-likelihood

In practice, it may not be possible to use the maximum likelihood method in parameter estimation since the available information on the random variable may not tell us much about its distribution. We may use the quasi-likelihood in determining the effect of covariates on response variables. The only specification necessary is the relation between the mean and the variance of the random variables. McCullagh and Nelder (1989) gives more information on this method.

3.4.2 ANALYZING THE GOODNESS OF FIT

Analysis of Residuals using scaled deviance

To determine the adequacy of the model in data description we use the scaled deviance. This is achieved by comparing this model to a saturated model. A saturated model is one in which the number of observed data points equals the number of the models' parameters. The saturated model is considered as the base line. Suppose we Ls and ls are the likelihood and logg-likelihood of the saturated model and Lm and lm are the likelihood and log-likelihood of the current model, then we can define the scaled deviance, SD, as follows; SD=$2(l_s - l_m)$ The deviance of the of the current model, Dm, is defined as;

$$Dm = \phi * SD \tag{3.17}$$

Where ϕ is the scaled parameter.

Normally in deciding which model to pick we compare their scaled deviance. those with smaller scaled deviance are better. Normally, models with more parameters have smaller scaled deviance.

Analysis of the Residuals using Pearson χ^2 statistic

The Pearson χ^2 statistic can be defined as

$$\sum r_p = \sum (\frac{y_i - \hat{\mu}}{\sqrt{V(\hat{\mu}_i}}) \tag{3.18}$$

Where the estimated variance of the distribution is $V(\mu_i)$.

Pearson χ^2 statistic follows χ^2 distribution for normal linear models and not for other the distributions. For data which do not have normal distribution,we get a skewed distribution of the Pearson residuals.

3.5 RESERVE ANALYSIS WITHIN THE FRAMEWORK OF GENERALIZED LINEAR MODELS (GLM)

Here we will look at how GLM can be applied to claim reserving by using the various distribution of the exponential family. Normally claims a particular distribution can be used to model the claim count(frequency) or claim amount (severity). Here we will discuss the Gamma and Over Dispersed Poisson models.

3.5.1 THE GAMMA DISTRIBUTION

The gamma distribution has probability density function as below;

$$f(y;\alpha,\lambda) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} \quad \text{for some} \quad y = i,2,3..., \quad \alpha > 0 \quad and \quad \lambda > 0 \tag{3.24}$$

Suppose we take $\lambda = \frac{\alpha}{\mu}$, then we have;

$$f(y;\alpha,\mu) = \frac{\alpha^{\alpha}}{\mu^{\alpha}\Gamma(\alpha)} y^{\alpha-1} e^{-\frac{\alpha}{\mu}y} \quad \text{for some} \quad y = i,2,3..., \quad \alpha > 0 \quad and \quad \lambda > 0 \tag{3.25}$$

Expressing this as exponential family of distributions gives;

$$f(y;\lambda,\mu) = \exp[\alpha(-\frac{y}{\mu} - \log\mu) + (\alpha-1)\log y + \alpha\log\alpha - \log\Gamma(\alpha)] \tag{3.26}$$

$$\text{for some} \quad y = i,2,3..., \quad \alpha > 0 \quad and \quad \lambda > 0$$

We can define the various parameters as below;

$$\phi = \alpha \tag{3.30}$$

$$\theta = -\frac{1}{\mu} \tag{3.29}$$

$$b(\theta) = -\log(-\theta) \tag{3.32}$$

$$c(y,\phi) = (\phi-1)\log y + \phi\log\phi - \log\Gamma(\phi) \tag{3.33}$$

$$a(\phi) = \frac{1}{\alpha} \tag{3.31}$$

The mean $E[Y]$, variance $Var[Y]$ and variance function $V(\mu)$ are given by;

$$E[Y] = b^{'}(\theta) = \frac{-1}{\theta} = \mu \tag{3.34}$$

$$Var[Y] = \frac{\mu^2}{\alpha} \tag{3.36}$$

$$V(\mu) = \frac{1}{\theta^2} = \mu^2 \tag{3.35}$$

The linear predictor η can then be expressed as;

$$\eta = \frac{1}{\mu} = \eta = c + a + b \tag{3.37}$$

3.5.2 THE OVER DISPERSED POISSON MODEL (ODP)

Poisson distribution is often used to model the counting data. The for a Poisson distribution with parameter μ the pdf, expected value and the variance are as below;

$$f(y,\mu) = \frac{\mu^y e^{-\mu}}{y!} \quad \text{for some} \quad y = 0,1,2\ldots, \quad and \quad \mu > 0 \tag{3.19}$$

$$E(Y) = Var(Y) = \mu \tag{3.20}$$

However when the data is overly disbursed, as is often the case with the claim data, the variance and the expected value are often not equal. In such a case the data follow the Over Disbursed Poisson distribution. The variance is bigger than the expected value. Usually the Consider a random variable Y which follows the Over Disbursed Poisson distribution then;

$$Y \sim ODP(\mu, \phi) \tag{3.21}$$

And the expected value and the variance are given by

$$E(Y) = \mu \tag{3.22}$$

$$Var(Y) = \phi * \mu \tag{3.23}$$

If the scale parameter, $\phi = 1$, then we get the Poisson distribution. If $\phi < 1$, then we get an under-disbursed Poisson distribution.

If $\phi > 1$, then we have an Over-disbursed Poisson distribution.

3.6 BOOTSTRAPPING

3.6.1 INTRODUCTION

Bootstrapping is a sampling method (with replacement), from an observed data set, that is used to create a pseudo data set which can be used to determine the distribution of the the parameters. The sampling is done with replacement. Bootstrap can either be;

1. Paired bootstrap-this is where the observed data are used directly in re-sampling
2. Residual bootstrap- this is where the residuals of the model are determined and used in re-sampling.

The residual bootstrapping is preferred since the bootstrapping method requires data to have the following properties;

1. The data should be identically and independently distributed.
2. To be indifferent to resembling the data(residue) or the data (residue) multiplied by a constant.

Scaled Pearson residuals will be used in this project. As stated in McCullagh and Nelder (1989) the Pearson residue can be defined as;

$$r_p = \frac{(y_i - \mu_i)}{\sqrt{(V(\mu_i))}} \tag{3.38}$$

The scaled Pearson residual for a run off triangle is given by;

$$r_{ij}^{sp} = \frac{(C_{ij} - \hat{\mu_{ij}})}{\sqrt{(a(\phi) * V(C_{ij}))}} = \frac{(C_{ij} - \hat{\mu_{ij}})}{\sqrt{(a(\phi) * V(\hat{\mu_{ij}}))}} \tag{3.39}$$

The dispersal parameter ϕ can either be a constant or dependant on the development period j in which case it is denoted as ϕ_j. When bootstrapping the GLM the following procedure should be followed;

(a) Define the GLM and fit it, obtaining parameters c, α_i, β_j $(i, j = 1, 2, \ldots n)$ and ϕ and fitted values for the observed data, $\hat{\mu_{ij}}$

(b) Calculate the scaled Pearson residuals, r_{ij}^{sp} of your fitted model.

$$r_{ij}^{sp} = \frac{(C_{ij} - \hat{\mu_{ij}})}{\sqrt{(a(\Phi) * V(C_{ij}))}} \tag{3.40}$$

(c) Sample from the residuals,with replacement,(bootstrapping bit) i.e.construct a triangle of bootstrapped residuals,

(d) Invert these residues to obtain a set of pseudo-data

$$C_{ij} = r_{ij}^{sp} * \sqrt{(a(\Phi) * V(\mu_{ij}))} + \hat{\mu_{ij}} \tag{3.41}$$

(e) Re-fit the GLM using this pseudo-data set, to obtain another set of forecast output.

(f) Repeat steps 3,4and 5 many times (e.g. 10,000) to derive a forecast output for each pseudo dataset incorporating the randomness present in the residuals.

(g) Apply some method e.g.the Bornhuetter-Ferguson Method, the chain ladder etc(in this project chain ladder method is applied) in carrying out projections for each of the sets of forecast output(alternative past data).

(h) Calculate the bootstrap reserves for each of the alternative projections.

(i) Determine the distribution of the possible reserve estimates incorporating the randomness in the residuals.

3.6.2 THE PREDICTION ERROR

The above procedure of bootstrapping can gives us the prediction error estimates. We can break prediction error into;

1. The process error; this arises from the the assumption made concerning the underlying distribution that the claim data follows. This is determined by the scale parameter in the ODP.

2. The parameter error; this arises from step 5 where we estimate the parameters for each pseudo data set obtained, giving a distribution of the parameter.

The prediction variance can then be given as;

$$Predictionvariance = Estimationvariance + Processvariance \tag{3.42}$$

3.6.3 BOOTSTRAPPING THE ODP MODEL

This model is a members of the exponential families. We model the claim amount on the assumption that they follow the ODP distribution. As stated in the earlier sections the ODP distribution has variance that is proportional to the mean. The main assumptions on the claim data are as follows;

1. incremental claim data are independent.

2. As stated above, the variance and the mean of the incremental claim data are proportional.

3. For all the development years the claim data are positive (if negative, this will lead to negative variance since variance is a proportion of the mean with the constant of proportionality of ϕ).

4. The same runoff pattern for each origin year.

Suppose that we take the incremental claim amount to be represented by the random variable C_{ij} , where i and j are as defined before, then;

$$C_{ij} = ODP(\theta_{ij}, \phi_{ij}) \tag{3.43}$$

Where;

$$E(C_{ij}) = \mu_{ij} \tag{3.44}$$

$$Var(C_{ij}) = \phi_j * \mu_{ij} \tag{3.45}$$

ϕ, scale parameter, can be a constant for all the development years or can be defined as ϕ_j for each development year. The linear predictor can be defined as η=$c + a_i + b_j$. The Pearson residual is defined as ;

$$r_{ij}^{sp} = \frac{(C_{ij} - \hat{\mu_{ij}})}{\sqrt{(\phi * V(\mu_{ij}))}} \tag{3.46}$$

This is then inverted to compute the pseudo claim data set C_{ij}, which are bootstrap cumulative claims given by the form;

$$C_{ij} = r_{ij}^{sp} * \sqrt{(\phi * V(\mu_{ij}))} + \hat{\mu_{ij}} \tag{3.47}$$

where r_{ij}^{sp} is the sampled scaled Pearson residue for the origin year i and development year j.

3.6.4 BOOTSTRAPPING THE GAMMA MODEL

As stated above, the link function of gamma can be considers as the reciprocal link, which is a canonical link. This leads to;

$$\mu_{ij} = \frac{1}{c + a_i + b_j} \tag{3.48}$$

For Gamma model,the scaled Pearson residue if given by;

$$r_{ij}^{sp} = \frac{(C_{ij} - \hat{\mu_{ij}})}{\sqrt{(V(\hat{\mu_{ij}}))}} = \frac{(C_{ij} - \hat{\mu_{ij}})}{\hat{\mu_{ij}}\sqrt{\hat{\alpha})}} \tag{3.49}$$

for $0 \leq i + j \leq I$. This is then inverted to compute the pseudo claim data set C_{ij}, the same way as in the Over-Dispersed Poison model.

Chapter 4

APPLICATION

In this section we look at how we can apply the over dispersed Poisson model, using the bootstrapping technique in data analysis. We will use 10,000 loops to construct the predictive distribution of the ultimate reserve. This distribution will help in calculating some measures of risk such as the *VaR* and *CVaR* of the reserve. The analysis is achieve in the environment of *R* statistical software.

4.1 THE DATA SET

We use the incremental claim data as used in Taylor and Ashe(1983). This is shown in table 4.1 below. The development triangle is in the USD

	Development Years									
Origin year	1	2	3	4	5	6	7	8	9	10
1	357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
2	352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
3	290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
4	310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
5	443,160	693,190	991,983	769,488	504,851	470,639				
6	396,132	937,085	847,498	805,037	705,960					
7	440,832	847,631	1,131,398	1,063,269						
8	359,480	1,061,648	1,443,370							
9	376,686	986,608								
10	344,014									

Table 4.1: Development triangle of incremental claims

As per the previous chapters the number of origin years and development years are I=J=10 The cumulative claim triangle for the same data is as shown in table 4.2 below

	Development Years									
Origin year	1	2	3	4	5	6	7	8	9	10
1	357,848	1,124,788	1,735,330	2,218,270	2,745,596	3,319,994	3,466,336	3,606,286	3,833,515	3,901,463
2	352,118	1,236,139	2,170,033	3,353,322	3,799,067	4,120,063	4,647,867	4,914,039	5,339,085	
3	290,507	1,292,306	2,218,525	3,235,179	3,985,995	4,132,918	4,628,910	4,909,315		
4	310,608	1,418,858	2,195,047	3,757,447	4,029,929	4,381,982	4,588,268			
5	443,160	1,136,350	2,128,333	2,897,821	3,402,672	3,873,311				
6	396,132	1,333,217	2,180,715	2,985,752	3,691,712					
7	440,832	1,288,463	2,419,861	3,483,130						
8	359,480	1,421,128	2,864,498							
9	376,686	1,363,294								
10	344,014									

Table 4.2: Development triangle of cumulative claims

From the above it is clear that the value of the cumulative claim amount rises with the accident year, however the rise is less with letter development year than the rise at the beginning.

The graphical representation of the cumulative claim amounts per each origin year is as shown in figure 4.1 below; As stated in section 3.2 the development

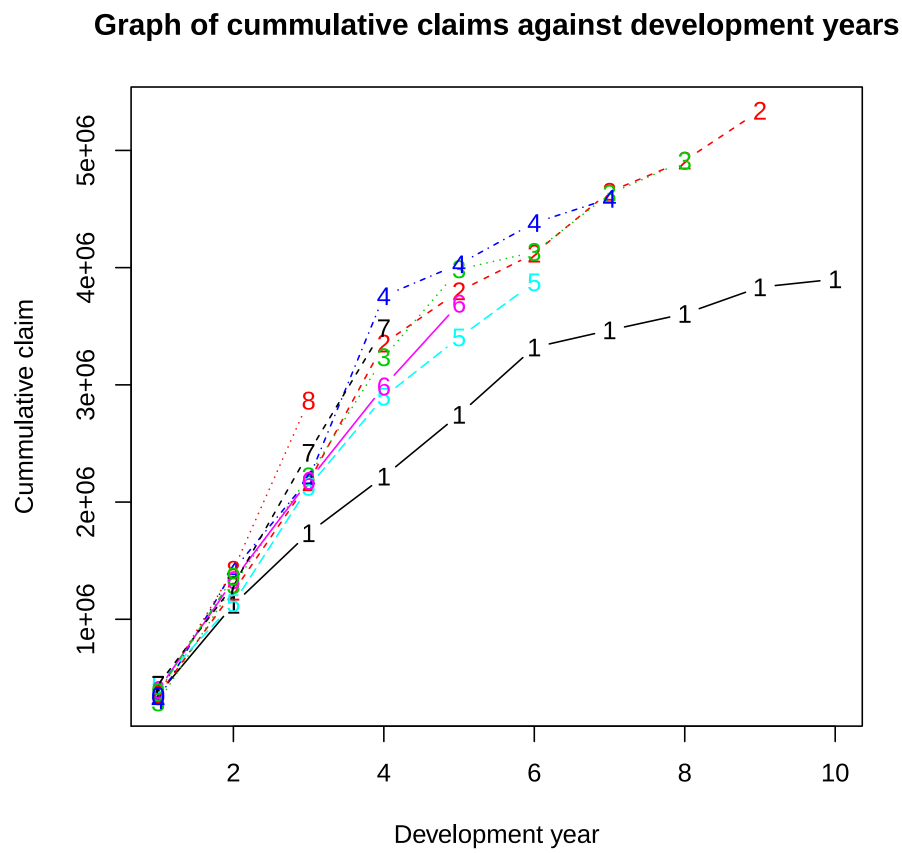

Figure 4.1: Graph of cumulative claims against development years

factors f_j can be determined as:

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j+1} C_{ij}}{\sum_{i=1}^{I-j+1} C_{ij-1}}$$

When applied on the above data we get the development factors as shown in table 4.3 below;

The resulting estimated future claim amounts, using the above development factors is as shown in table 4.4 below;

Development Year	1	2	3	4	5	6	7	8	9
Development factors	3.491	1.747	1.457	1.174	1.104	1.086	1.054	1.077	1.018

Table 4.3: Table of development factors

	Development Years									
Origin year	1	2	3	4	5	6	7	8	9	10
1	357,848	1,124,788	1,735,330	2,218,270	2,745,596	3,319,994	3,466,336	3,606,286	3,833,515	3,901,463
2	352,118	1,236,139	2,170,033	3,353,322	3,799,067	4,120,063	4,647,867	4,914,039	5,339,085	5,433,587
3	290,507	1,292,306	2,218,525	3,235,179	3,985,995	4,132,918	4,628,910	4,909,315	5,285,369	5,378,920
4	310,608	1,418,858	2,195,047	3,757,447	4,029,929	4,381,982	4,588,268	4,835,576	5,205,981	5,298,127
5	443,160	1,136,350	2,128,333	2,897,821	3,402,672	3,873,311	4,207,578	4,434,366	4,774,039	4,858,539
6	396,132	1,333,217	2,180,715	2,985,752	3,691,712	4,074,912	4,426,577	4,665,169	5,022,521	5,111,420
7	440,832	1,288,463	2,419,861	3,483,130	4,088,846	4,513,269	4,902,764	5,167,023	5,562,817	5,661,278
8	359,480	1,421,128	2,864,498	4,174,719	4,900,703	5,409,396	5,876,227	6,192,956	6,667,336	6,785,348
9	376,686	1,363,294	2,382,084	3,471,649	4,075,368	4,498,392	4,886,603	5,149,991	5,544,480	5,642,617
10	344,014	1,200,815	2,098,185	3,057,894	3,589,662	3,962,269	4,304,213	4,536,210	4,883,683	4,970,125

Table 4.4: The Full Development Triangle of cumulative claims for both the original claims and projected claims

The claim amount can be determined as per the section formulae; The outstanding reserve per origin year can be calculated using the formula;

$$R_i = C_{ij} - C_{i,n-i+1} \quad \text{for} \quad 1 \le i \le J$$

This gives reserves as shown in table 4.5;

Origin Year	1	2	3	4	5	6	7	8	9	10
Reserve amount	0	94,634	469,511	709,638	984,889	1,419,459	2,177,641	3,920,301	4,278,972	4,625,811

Table 4.5: The outstanding reserve per origin year

The total claim reserve outstanding, R, define by R=$\sum_{i=1}^{I} R_i$ as per section 3.2 gives 18,680,856/=.

In the following sections we will look at the stochastic analysis of the models. We will compare the results we get to those of chain ladder reserving.

4.2 THE STOCHASTIC ANALYSIS OF THE MODELS

Here we will look at the analysis of the fitted Over Dispersed Poisson and Gamma models as stated sections .3.5. We will look at residual analysis,bootstrapping of both models to determine the reserve distribution and some risk measures from the distributions of the reserves.

4.2.1 DETERMINING THE SCALED DEVIANCE STATISTIC AND THE PEARSON χ^2 STATISTIC

We can also compute the scaled deviance and the scaled Pearson χ^2 statistic. This can be achieved as describe d in section These are as summarized in the table;

	ODP MODEL	GAMMA MODEL
PEARSON RESIDUAL	1893649	3.606954
DEVIANCE RESIDUAL	1903014	3.986787

The ODP model has very high values of both the Deviance residual and the Pearson chi square residual compared to the Gamma model. This indicates that Gamma model fits the data set better than ODP model. For the purpose of this model we will continue with the data analysis for both models.

4.2.2 CLAIM RESERVE CALCULATION USING THE BOOTSTRAPPING TECHNIQUE

We will calculate the bootstrap reserve and the reserve distribution for both the ODP and the Gamma models. The analysis is done in the environment of R statistical package. The R codes used are as presented in the appendix. As described in chapter 4, the bootstrap technique is used with 10,000 repetitions of loops. For each origin year, we will calculate the claim reserves as well as the total reserve, by applying bootstrapping technique to both the ODP and the Gamma models, and compare them with the claim reserves we got using basic chain ladder technique. This will be done for both the ODP and the Gamma models. We will also calculate the various values of Value at Risk (VaR) as one of the quantitative measures of risks.

For the ODP model the table below shows the ODP Bootstrap mean ultimate claim, the ODP bootstrap mean IBNR, the SD IBNR ,the IBNR $VaR(99\%)$ and the IBNR $VaR(99.5\%)$

Accident year	Chain ladder Ultimate claim	Chain ladder IBNR	ODP Bootstrap Ultimate Claim	ODP Bootstrap IBNR	ODP Bootstrap SD IBNR	VaR(99) IBNR 99	VaR(99.5) IBNR 99.5
1	3,901,463	0	3,901,463	0	0	0	0
2	5,433,719	94,634	5,434,272	95,187	114,147	460,238.10	539,786.40
3	5,378,826	469,511	5,386,116	476,801	219,570	1,084,274.70	1,168,452.80
4	5,297,906	709,638	5,302,154	713,886	260,733	1,442,143.20	1,551,285.30
5	4,858,200	984,889	4,866,294	992,983	307,996	1,859,272.90	1,967,047.90
6	5,111,171	1,419,459	5,124,104	1,432,392	380,347	2,464,676.20	2,594,727.70
7	5,660,771	2,177,641	5,675,610	2,192,480	497,252	3,527,018.40	3,685,942
8	6,784,799	3,920,301	6,800,769	3,936,271	798,480	6,027,880.50	6,324,337.50
9	5,642,266	4,278,972	5,671,828	4,308,534	1,056,522	7,137,294.60	7,614,843.60
10	4,969,825	4,625,811	5,075,308	4,731,294	2,038,800	10,907,262	11,624,802.40
Total	53,038,946	18,680,856	53,237,918	18,879,828	5,673,847	34,910,061	37,071,226

The empirical cumulative distribution function for the total bootstrap reserve for the ODP bootstrap model is as shown below in figure 4.2.

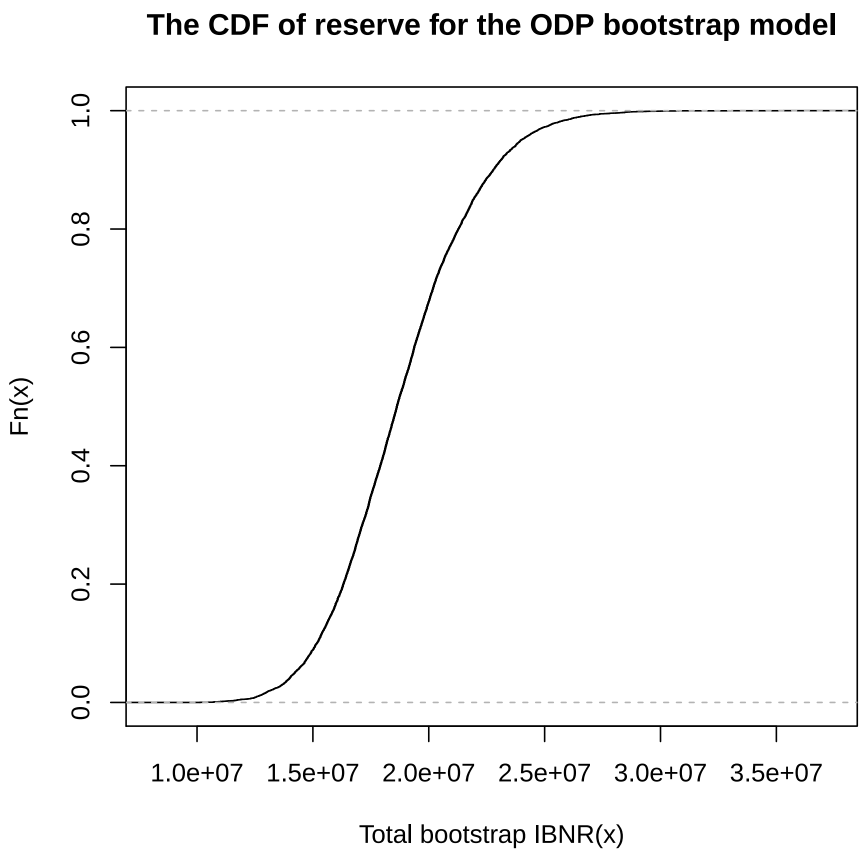

Figure 4.2: The CDF of reserve for the ODP bootstrap model

The histogram of the total bootstrap reserve for the bootstrap ODP model is as shown in figure 4.3 below.

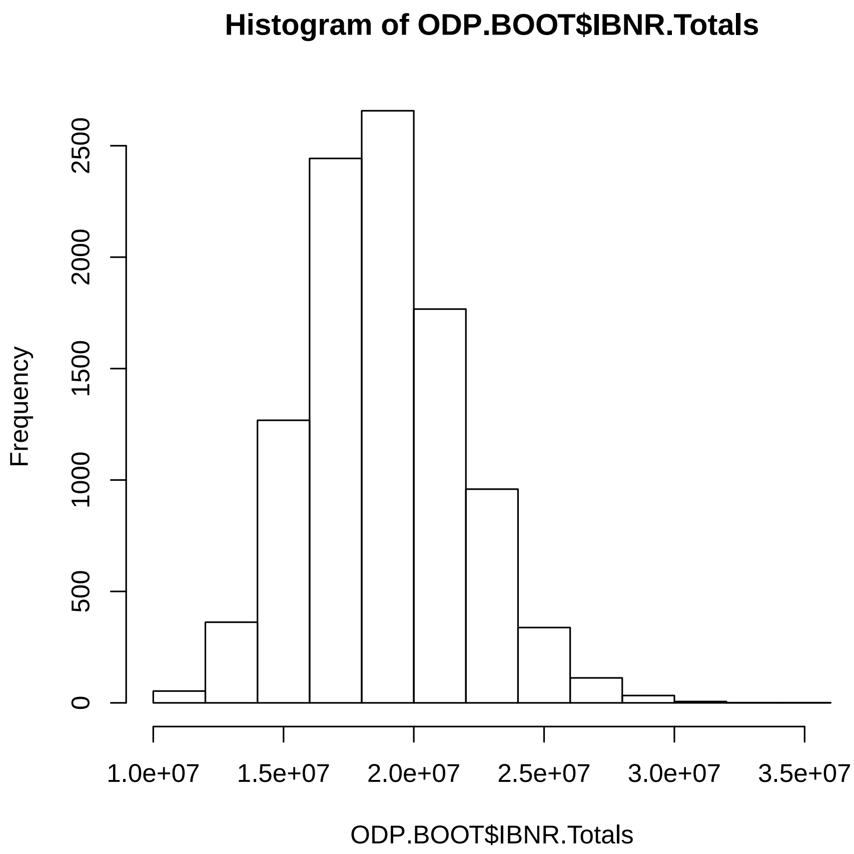

Figure 4.3: The Histogram of reserve for the ODP bootstrap model

For the Gamma model the table below shows the Gamma Bootstrap mean ultimate claim, the Gamma bootstrap mean IBNR, the SD IBNR ,the IBNR $VaR(99\%)$ and the IBNR $VaR(99.5\%)$

Accident year	Chain ladder Ultimate claim	Chain ladder IBNR	GAMMA Bootstrap Ultimate Claim	GAMMA Bootstrap IBNR	GAMMA Bootstrap SD IBNR	VaR(99) IBNR 99	VaR(99) IBNR 99.5
1	3,901,463	0	3,901,463	0	0	0	0
2	5,433,719	94,634	5,437,585	98,500	115,865	481,297.20	538,662.6
3	5,378,826	469,511	5,382,686	473,371	217,196	1,100,210.10	1,203,454.3
4	5,297,906	709,638	5,307,142	718,874	266,834	1,454,054.0	1,527,939.0
5	4,858,200	984,889	4,864,456	991,145	306,831	1,816,066.7	1,910,242.7
6	5,111,171	1,419,459	5,119,415	1,427,703	378,857	2,443,722.2	2,552,053.5
7	5,660,771	2,177,641	5,671,678	2,188,548	500,764	3,533,452.6	3,695,441
8	6,784,799	3,920,301	6,799,116	3,934,618	797,320	6,047,578.4	6,324,309.5
9	5,642,266	4,278,972	5,671,116	4,307,822	1,062,967	7,155,475.1	7,551,060.1
10	4,969,825	4,625,811	5,039,623	4,695,609	2,004,311	10,692,659	11,352,051.8
Total	53,038,946	18,680,856	53,194,280	18,836,190	5,650,945	34,724,515	36,655,215

The values of the ultimate claim and the IBNR are also close to the values of the chain ladder. However for the gamma model these values of closer to the ones of chain ladder model compared to the values of obtained by the ODP model. Just like the ODP model, Bootstrapping technique helps us to calculate the IBNR standard deviation and the $VaR(75\%)$ and $VaR((99.5\%)$, i.e. the 75% and 99.5% quantiles.

The empirical cumulative distribution function for the total bootstrap reserve for the Gamma bootstrap model is as shown in figure 4.4 below.

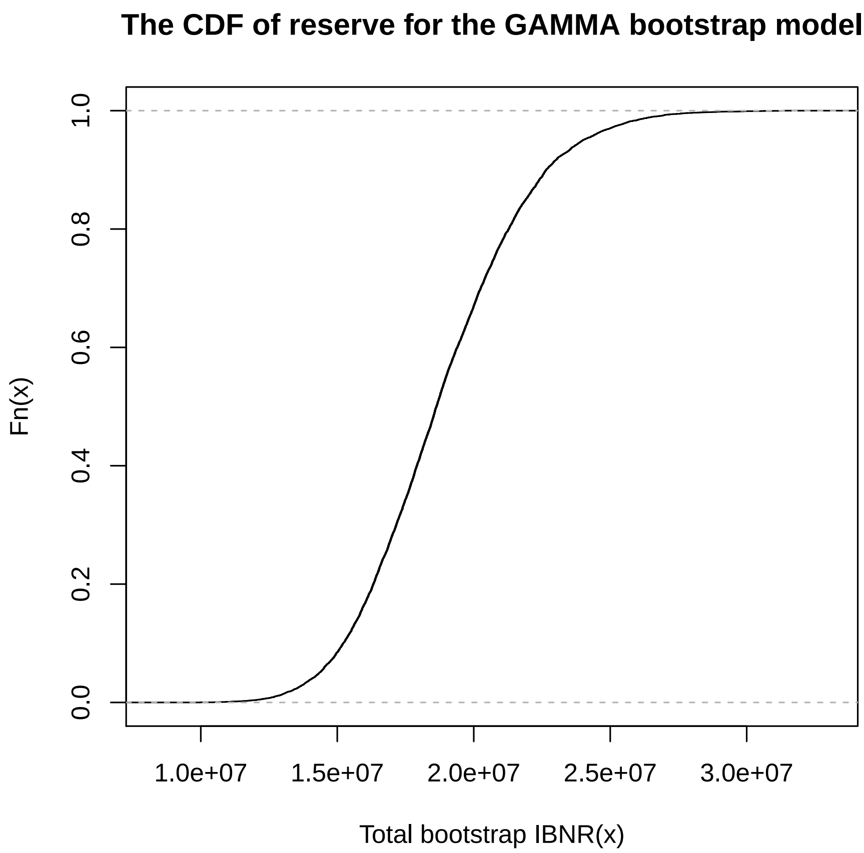

Figure 4.4: The CDF of reserve for the GAMMA bootstrap model

The histogram of the total bootstrap reserve for the bootstrap ODP model is as shown in figure 4.5 below.

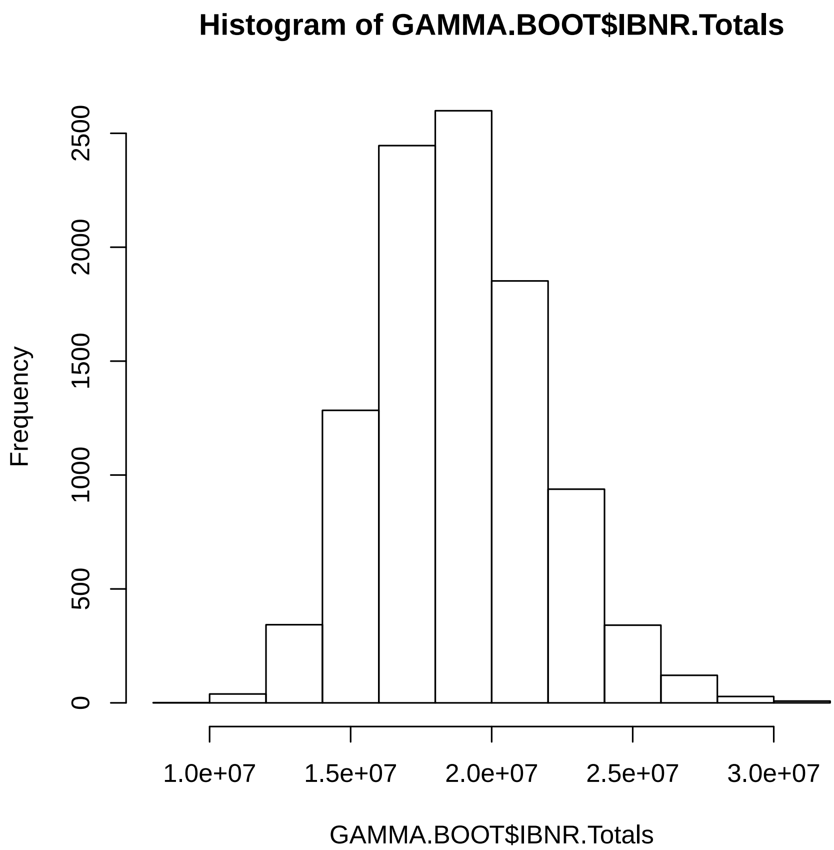

Figure 4.5: The Histogram of reserve for the GAMMA bootstrap model

From the values given for $VaR(99.5\%)$,the loss of the insurance company would excess the amount of 18,191,398 with low probability of 0.5From this analysis it is clear that stochastic reserving techniques are a great tool in determining loss reserves, the predictive distributions of the reserves and to the quantitative measures of risks live Value at Risk when analyzing reserves.

4.3 FITTING A DISTRIBUTION TO THE IBNR

The distribution for both the ODP and the Gamma models can be fitted to a known distribution. In this case the reserve distribution seems to be fitting well to the log normal distribution. The Maximum likelihood technique is applied to calculate parameters of the chosen distribution.

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

It has been shown that the algorithm based loss reserving models overlooks some key factors that need to be incorporates in loss reserving. It has also been demonstrated that that the stochastic loss reserving can provide an improved alternative to the algorithm based methods..

The concept of loss reserving was introduced. The algorithm based claim reserving method was discussed. However it was clear that Chain ladder loss reserving method does not allow us to calculate the various risk measure quantities. This makes it necessary to introduce other loss reserving techniques that would take care of this.

Stochastic loss reserving is introduced as an alternative to chain ladder. Claim reserving is done within the frame work the Generalized Linear Model. The various components of GLM are introduced as well as the exponential family of distribution.

Over dispersed Poisson model and gamma are used as the underlying distributions that the claim amounts follow. Any other distributions both discrete and continuous could be used. The models are validate by comparing their deviance residuals and the Pearson chi square statistics. For purpose of this thesis ,both the Over dispersed Poisson model and gamma were used in data analysis.

The bootstrapping technique is applied on the Pearson residuals of the runoff triangles for both the Over dispersed Poisson model and Gamma distributions. This helped in getting the predictive distributions of the loss reserve. From here it is possible to get risk measures such as the VaR, as may be required by the company management, or for further analysis.

Data analysis was done in the environment of R statistical package. This was done to help compare the results of chain ladder loss reserving and stochastic loss reserving using Over dispersed Poisson distribution and Gamma distribution as the underlying GLM distributions. Residual analysis was done on the data set and both the Pearson chi square and the Deviance residuals were calculated for both the Over dispersed Poisson distribution and Gamma distribution were done. The results of the analysis was summarized using graphically and using tabular form.

Solvency II requires distribution of the expected value of the liability after year 1 for the 1 year ahead balance sheet. This can be achieved using stochastic loss reserving methods. The traditional techniques like chain ladder don't give the distribution properties of the reserves. Bootstrapped distribution gives a technique of getting the unknown distribution of the reserves. This can be fitted to get the exact distribution that it represents.

5.2 RECOMMENDATIONS

From the thisis it is clear that stochastic loss reserving gives quite a satisfactory results and details that can help improve risk management in general insurance reserving. However, there are several ways in which stochastic loss reserving can be improved. One area that need further research is on the other risk analysis measure and various ways that stochastic loss reserving can help improve enterprise risk management.Another area of further research is how stochastic loss reserving can be incorporated in the the International Accounting Standards for better reporting of the financial positions of insurance companies.

References

[1] S Haberman and A E Renshaw (1988) Generalized linear models andactuarial work. The Staple Inn Actuarial Society(1988).

[2] Duncan Anderson,SholomFeldblum, Claudine Modlin,DorisSchirmacher, Ernesto Schirmacher and NeezaThandi (2007): "Practitioner's Guide to Generalized Linear Models" , third edition. Towers Watson.

[3] Tom Hoedemakers , Jan Beirlant , Marc J. Goovaerts and Jan Dhaene(2005): "On the distribution of discounted loss reserves using generalized linear models". Scandinavian Actuarial Journal, 2005:1,25-45

[4] Schmidt (2006) in his work,Methods and Models of Loss Reserving Based on Run-Off Triangles: A Unifying Survey, Casualty Actuarial Society, Fall 2006.

[5] Greg Taylor and Grainne McGuire(2004): Loss reserving with GLMs: a case study, Casualty Actuarial Society, Spring 2004.

[6] England and Verrall (2002):Stochastic claim reserving in general insurance,Presented to the Institute of Actuaries, 28 January 2002

[7] Mark R. Shapland (2016):Using the ODP bootstrap model: A practitioner's guide,Casualty Actuarial Society Monograph Series, 2016

Chapter 6

APPENDIX

6.1 R Code

6.1.1 Importing The Data

```
## Importing the date setwd("C:/Users/edmond ochieng/Desktop/UON/loss analysis using GLM/data analysis")
data<-read.csv("Book1.csv",header=F)
data.incremental<-as.triangle(data)
data.incremental

## Transforming triangles from incremental to cumulative
data.cumulative <- incr2cum(data.incremental)
data.cumulative

# # Plotting cummulative claims against development years
plot(data.cumulative,ylab="Cummulative claim",xlab="Development year",main="Graph of cummulative claims against development years")

## Creating a data frame
claims <- as.vector(data.incremental)
number.origin.years <- nrow(data.incremental)
number.development.years <- ncol(data.incremental)
origin <- factor(row <- rep(1:number.origin.years, number.development.years))
dev <- factor(col <- rep(1:number.development.years, number.origin.years))
data.frame <-data.frame(claims.amount=claims, origin.years=origin, development.years=dev)
data.frame
```

6.1.2 Chain Ladder Reserving

```
# Chain Ladder Reserving
# Development factors
n <- 10
development.factors <- sapply(1:(n-1),
function(i)
sum(data.cumulative[c(1:(n-i)),i+1])/sum(data.cumulative[c(1:(n-i)),i])

)
development.factors

    ## Getting the full cumulative triangle
fulldata.cumulative <- cbind(data.cumulative)
for(k in 1:n)
fulldata.cumulative[(n+1-k):n, 1+k] <- fulldata.cumulative[(n+1-k):n,k]*development.factors[k]

round(fulldata.cumulative)

    ## Ultimate claims per origin year
fulldata.cumulative[ ,10]
cbind(fulldata.cumulative[ ,10])

    ## calculating the yearly reserves(per origin year)and the total reserves
fulldata.cumulative[ ,10] - getLatestCumulative(data.cumulative,na.values = 0)

    sum(fulldata.cumulative[ ,10] - getLatestCumulative(data))
```

6.1.3 Stochastic Reserving

```
# Stochastic Reserving
# Residual analysis and Checking the goodness of fit of the data

    # Fiting the gamma model
glm
gamma.model<- glm(claims   origin + dev, family = Gamma,subset=is.na(claims),
data=data.frame)
gamma.model
summary(gamma.model)
```

# pearson and deviance residuals for Gamma model
gamma.model.pearson.residuals<-residuals(gamma.model,'pearson')
gamma.model.deviance.residuals<-residuals(gamma.model,'deviance')

# Fit over-dispersed poison model (ODP)
ODP.model <- glm(claims origin + dev, family = quasipoisson(),subset=!is.na(claims), data=data.frame)

summary(ODP.model)

# pearson and deviance residuals for over-dispersed poison model
ODP.model.pearson.residuals<-residuals(ODP.model,'pearson')
ODP.model.deviance.residuals<-residuals(ODP.model,'deviance')

# Goodness of fit tests Gamma Model
# Model deviance
gamma.model.deviance <- $sum(residuals(gamma.model, "deviance")^2)$
$gamma.model_deviance$
Scaled Pearson Chisq statistic
$gamma.model_Pearson < -sum(residuals(gamma.model, "pearson")^2)$ $gamma.model_Pearson$
# Goodness of fit tests Over dispersed poisson Model
# Model deviance statistic
ODP.model.deviance<-$sum(residuals(ODP.model, "deviance")^2)$
ODP.model.deviance
# Scaled Pearson Chi square statistic
ODP.model.Pearson<- $sum(ODP.model.pearson.residuals^2)$
ODP.model.Pearson

Getting The IBNR Distribution Using Bootstrap Technique

# Assuming gamma distribution
GAMMA.BOOT <- BootChainLadder(data.cumulative, R=10000, process.distr="gamma")
GAMMA.BOOT

# Gamma Model graphs
# The CDF of reserve for the GAMMA bootstrap model
plot(ecdf(GAMMA.BOOT$IBNR.Totals),xlab="Total bootstrap IBNR(x)", main="The CDF of reserve for the GAMMA bootstrap model")

Histogram of GAMMA bootstrap model
plot(hist(GAMMA.BOOT$IBNR.Totals),xlab="Total bootstrap IBNR(x)",ylab="Frequency",

```
main="The Histogram of reserve for the GAMMA bootstrap model")

    # Assuming ODP MODEL distribution
ODP.BOOT <- BootChainLadder(data.cumulative, R=10000, process.distr="od.pois")
ODP.BOOT

    # ODP model graphs
# The CDF of reserve for the ODP bootstrap model
plot(ecdf(ODP.BOOT$IBNR.Totals),xlab="Total bootstrap IBNR(x)", main="The
CDF of reserve for the ODP bootstrap model")
# Histogram of ODP bootstrap model
plot(hist(ODP.BOOT$IBNR.Totals),xlab="Total bootstrap IBNR(x)",ylab="Frequency",
main="The Histogram of reserve for the ODP bootstrap model")
summary(ODP.BOOT)

    # Analysis Of Reserves
# Calculating Value at Risk of the bootstrap IBNR for both Gamma and ODP
models
quantile(GAMMA.BOOT, c(0.75,0.95,0.99, 0.995))
quantile(ODP.BOOT, c(0.75,0.95,0.99, 0.995))
```