KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF SCIENCE

DEPARTMENT OF MATHEMATICS

APPLICATION OF BÜHLMANNS CREDIBILITY THEORY TO AN AUTOMOBILE INSURANCE CLAIMS

By

GAMADEKU MAWULI

(BSc. Mathematics)

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN ACTUARIAL SCIENCE

October 18, 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 acknowledgment had been made in the text.

Gamadeku Mawuli (20360192)	....	....
Student	Signature	Date

Certified by:

Nana Kena Frempong	....	....
Supervisor	Signature	Date

Certified by:

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

DEDICATION

This thesis is dedicated to my lovely and supportive wife Mrs Angela Gamadeku and our children Klenam and Selorm; and to my extended family, most especially my brother; Robert Selorm Gamadeku and father Nelson Gamadeku and all my family members.

ABSTRACT

In this thesis, we study auto-mobile insurance claim counts past data for commercial vehicles having third party liability cover.There are various literature on how this can be done. But in this study experience rating was used. Observations of claim events of the past were used to forecast the future claim counts of the policyholders without considering risk factors that are associated with them.The main objective of this thesis is to estimate the number of auto insurance claims occurring under motor accidents using Bühlmans credibility theory. Specific objectives were to adopt Bühlmans credibility model and use the model to calculate credibility weighted estimates for the expected number of claims per year for each policyholder. Finally, the weighted estimates were used to estimate the number of claim counts in 2015.The R software was used to compute the Bühlmans credibility factor and the credibility weighted estimates for each policyholder for 2015. Conclusively, the findings of this study revealed that out of the 600 policy holders about 197 of them are likely to make claims in 2015. It was recommend that the insurance company should make claims reservation for 197 policyholders in 2015.

ACKNOWLEDGMENT

I am grateful to the Almighty God who sustained me through my university education. A special acknowledgement is reserved for Mr Nana Kena Frempong, a lecturer in the Mathematics Department of Kwame Nkrumah University of Science and Technology, Kumasi(KNUST), who generously took time off his busy schedule to give comprehensive advice and constructive remarks which made this thesis a reality.

I wish to express a special thank you to all the lecturers in the Mathematics Department of KNUST, who traveled from Kumasi to Accra every week to teach me all the courses I read. Finally, I would like to express my sincere appreciation to everyone especially Mr Gilbert Agudze, Mr Doe-Dekpey Wonder and all their able men who consciously or unconsciously contributed to the successful completion of my MSc. Actuarial Science study at Kwame Nkrumah University of Science and Technology, Kumasi.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Background to the study

Insurance is a significant part of risk management strategy for institutions and individuals. It is considered as a risk financing transfer under which an insurer agrees to accept financial burdens arising from loss. In general , insurance is defined as a contractual agreement between two parties: the insurer and the insured. The use of Vehicle has an implanted mortality and morbidity risk in addition to loss through fire and theft in a lot of countries worldwide Maddala (2005).

In view of this Motor vehicle insurance was developed to take care of the possible loss that may arise from use of motor vehicles (Amoo, 2002 Onafalujo, et al, 2011). This takes the form of equal transfer of the risk of a loss, from vehicle owner (insured) to the insurer in exchange of an amount called premium, and can be considered as a compensation for any devastating loss. Motor insurance, also known as Automobile insurance, is interestingly becoming the most patronized form of insurance.

This insurance was introduced to protect vehicle users from potentially great financial loss from operating their vehicles. Therefore, policy makers require motorist to purchase auto insurance cover to protect innocent third parties as well as the at fault motorist from liability (Outreville,1990 Aeron-Thomas, 2002).

In Ghana, there are primarily two types of motor insurance; namely, third party insurance and comprehensive motor insurance. Third party insurance covers only the cost of repairing the vehicle following an accident while comprehensive motor insurance policy covers the risk of the cost of repairing the vehicle when an accident occurs, the cost of a new vehicle when stolen or damaged beyond economic repairs as well as legal liability claim against the driver or owner of the vehicle following the vehicle causing damage to a third party (Esho,et al., 2004).

Comprehensive car insurance is not compulsory in most countries worldwide. However, if a motor owner wants protection against financial loss, comprehensive auto insurance is the best because it covers compensation for car accidents and other kinds of misfortunes. If damage is made to another party, the motor insurance policy will protect the insured in that kind of situation. It also covers the insured's own car and medical expenses. This policy has the highest premiums compared to the third party auto insurance policy.

Moreover in addition to the coverage for the damages to insured's car and the other party's vehicle, the comprehensive insurance also covers damages that consequences to other non-car crash incidents. Insurance is well accepted in Ghana, as everyone needs financial protection against risks of loss. The insurance needs in Ghana for example transcends the following competitive areas : marine insurance, fire insurance, life insurance, motor insurance, travel insurance etc.

In Ghana, Insurance Act 58 states that all vehicle owners must insure their vehicles against accidental risks and damages. Regardless of all the posing questions asked concerning insurance, Ghanaians are steadily accepting the fact of paying premiums to protect themselves in case of loss and damage.

Insurance benefits are not yet being fully realized in developing countries as stated in USAID 2006 report on insurance. Evaluating the potential to

develop insurance schemes, annuity markets and insurance products for low-income populations; it is believe some countries like Ghana may have great potential to benefit from the development of specific lines of insurance, given adequate infrastructure and conditions.
Outreville (1990) posits that the share of total insurance premium generated in developing countries remain at low figure even though these countries contain more than 75 percent of the world population.

Statistical methods and principles are needed in allocating insurance funds since the process involves risk Equities ,bonds, shares ,loans and many others which are the assets in which insurance funds are invested with aim getting real returns claim liability and other financial demands. Statistical methods of estimation are used by insurance companies to acquire tangible information about these uncertain liabilities where decisions regarding expected claim counts, expected monthly claims, payment targets, assets allocations as well as future insurance pricing are determined. This is in insurance is referred to as the expected claim liability estimation problem.

One major concern of every insurer when determining a base premium is to ensure that it is adequate enough to meet its financial obligations. Experience rating systems in general and credibility theory for that matter constitute an efficient step to determine a fair premium. Experience rating system takes into account the past individual experience of an insured when establishing the insured's premium.

1.2 1.2 Statement of the problem

Insureds expect a cushion in the event of economic loss as stipulated in an insurance contract. In view of this, the challenge of meeting the claim payment terms becomes an issue of much concern of the insurer. Claims and risks have long been estimated using some stochastic techniques by Actuaries and these techniques result in poor estimations on the data (Wüthrich Merz (2008). The Bühlmanns credibility model which easily computes a certainty estimates for future claim expectation using past experiences of the claim count on the policyholder results in good estimation, hence its adoption in this study.

1.3 Objectives of the study

The main objective of the study will be to estimate the future claim counts of the insureds using claim experiences of the insureds by using Bühlmans credibility theory.
The specific objectives of the study included the following:

1. to determine the credibility factor.
2. to determine the credibility weighted estimates per year for each policy holder.
3. to estimate the expected number of claims in 2015

1.4 Justification of the Study

Insurance is the most important parts of welfare system. Insurance provides financial protections against unplanned losses that are unforeseen and unexpected and facilitates a redistribution of social benefits in society. The basic idea of insurance is that individuals transfer their risk to an insurance company, and for this they pay premiums.

In the light of the contributions of the insurance companies in the country's developmental agenda, this study would serve as a source of information on insurance claim liabilities for most of the insurance companies in Ghana. Furthermore, the findings of the study would also enlighten policy holders to understanding some issues on insurance claims.

Finally, the study is expected to add knowledge to the existing insurance claim liabilities procedures since they need information on claim counts to estimate insurance premiums and would also serves as a basis for future research.

1.5 Methodology

The data for the study were obtained from the motor department of the Quality Insurance Company. The data were on insured vehicles specifying claim counts of the policy holder. Other literature materials on estimation of auto insurance claims were collected from National Insurance Commission. In this study Bühlmans credibility model would be employed as the main methodology to analyze the data.

1.6 Scope of the Study

The study will cover only the Quality Insurance Company and its insurance claim liabilities due to the limited time period available within the academic calendar to conduct this study. But the findings from the study can be generalized to insurance companies in the country.

1.7 Organisation of the study

This study would be organized into five chapters. Chapter one will provide introduction which include, background of the study, statement of the problem, objective of the study, justification of this study, methodology of the study and the scope of the study on auto insurance claim liabilities.

In chapter two, we presented a review of some related literatures on insurance companies of Ghana and auto insurance claim liabilities. Methodology which would contain a development Bühlmans credibility model was done in chapter three of the study. The model implementation and discussion of results obtained were detailed in Chapter four. Lastly, the summary, conclusions and recommendations of the study were presented in chapter five.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews the various literatures related to the auto insurance claims and the findings which have already been identified by previous researchers and numerous studies in and around the world.

It will also assess the existing mathematical models of auto accidents insurance claims. It is specifically sub-headed under the following: Insurance Claims, Auto insurance claims, Concept of auto insurance in Ghana, Buhlmanns credibility model, Logistic Regression Models, Count Regression Models and Accident Year Statistics.

2.2 Insurance Claims

In their study Viaene et al (2007) found out that the probability and claim size forecast is significant, since an insurance company can use these estimates to offer or not offer premium discounts depending on an individual client's characteristics or create strategies for detecting fraudulent claims.

In similar vein, Meulbroek (2001) posits that insurance companies need to treat risk management as a series of related factors and events. Furtherance, Boland (2007), said in order to handle claims arising from incidents that have already occurred insurers must employ predictive methods to deal with the extent of the liability. Therefore, an insurer has to find ways to predict claims and appropriately charge a premium to cover this risk. Insurers have to look for better ways to capture the characteristics of individuals that affect claim size and probability, and consequently identify insured persons that have a higher propensity for generating losses.

According to Weisberg and Tomberlin (1982) insurance companies attempt to estimate reasonable prices for insurance policies based on the losses reported for certain kinds of insureds . The estimate has to based on consider past data in order to grasp the trends that have occurred.

2.3 Auto insurance claims

In the late 1980's and early 1990's there was a distinct effort for actuaries to combine financial theory with stochastic methods into their established models D'arcy (1989). The profession, today both in practice and in the education syllabi of many actuarial establishments combine stochastic methods of claim models, loss models, life tables, and financial theory. Feldblum (2001). However the articles written about actuarial modeling for insurance claim severity from different angles are very scanty.

Wright (2005) used the steps involved in actuarial modeling to fit models to 490 claim amounts drawn from seven successive periods, fitted analytic loss distributions using maximum likelihood estimation for each of the seven years. He then used P-P plots and the kolmogorov-smirnov tests (K-S test) to assess the absolute quality of fit.

Meyers (2005) used actuarial modeling to fit a statistical distribution to 250 claims. He tested the claim data using the following statistical distributions gamma, weibull and log-normal distributions. He also used the maximum likelihood estimation to fit the distributions, an idea that was useful to this research, only that he based his methodology on the Bayesian solution. That is, after he calculated the likelihoods, he used them to get the posterior probabilities of each model.

Renshaw (2004) also use actuarial modeling in claim amounts for nonlife insurance using the concept of quasi-likelihood and extended quasi likelihood. Maximum likelihood estimate or was used to fit the model structure; because the quasi-likelihood parameter estimates according to him have similar asymptotic properties to the maximum likelihood parameter estimates.
This study however did not touch on the quasi-likelihood approach but used the maximum likelihood estimates and the steps followed by renshaw in his modeling procedure goes in line with the procedures involved in all actuarial modeling process.

According to Amoo, (2002) Onafalujo, et al, (2011) motor vehicle insurance was developed to take care of the possible loss that may arise from use of motor vehicles. Motor insurance is an important form of contract arising out of or in connection with the use of motor vehicle.
This takes the form of equitable transfer of the hazard of a los, from an insured to the insurer in exchange for an amount called premium, and can be thought of as a guarantee for extreme loss. Motor insurance, which is also known as motor insurance, is becoming the most popular form of insurance. Similarly, Outreville, (1990) Aeron-Thomas, (2002) posit that motor insurance was introduced to protect motorist from potentially enormous financial loss from using a vehicle. Therefore, policies are made requesting motorist to purchase automobile insurance cover to protect innocent third parties as well as the art fault motorist from liability.

Insurance is the purchase of financial security or it can be considered as the insured, anxious to protect oneself against a risk of a loss. It is a form of risk management strategy used to hedge against risk of contingent loss. Vehicle usage has embedded morbidity and mortality risk plus loss through theft and fire in several countries Maddala, (2005).

In their study, Cummins and Weiss (2000) found out that the first stochastic nature of non-life loses is likely to create an "errors-in-variables problem" that would distort the measurement of scale economies, particularly where the insurance includes catastrophe risk. Secondly, when claims incurred are used, the output quality of loss control and risk management is ignored, which is undesirable.

Therefore, the insurer that is very successful in loss-prevention would be measured as having less output. So given both premiums and claims incurred are imperfect measures of insurance output, we follow the suggestion that says alternate between both measures of insurance output. Estimating a loss severity distribution that is distribution of individual claim sizes from historical data is an important actuarial activity in insurance. Thus it is essential to find a good statistical model for the largest observed historical losses.
It is less important that the model explains smaller loses. A good model is the extreme-value theory or generalized Pareto distribution to model exceedances over a high threshold.

2.4 Concept of auto insurance in Ghana

Assurance began in Ghana in the colonial era since 1924, when the Royal Guardian Enterprise now known as the Enterprise Insurance Company Limited was established. The Gold Coast Insurance Company was the first indigenous private Insurance to be established in 1955 and in 1962 the State Insurance Company was also set up. By 1971 eleven more companies were also established. Quality Isurance Company.

The 1989 insurance law which is the PNDC 229 led to the founding of the National Insurance Commission (NIC). National Insurance Commission is the regulatory body that has been authourised to regulate and supervise activities of insurance companies in the country. Insurance companies in Ghana divided into :(i) Life Insurance, (ii) Non-Life Insurance, and (iii) Composite Insurance (a combination of Life and Non-Life insurance).

There are two types of motor insurance in Ghana; namely, third party insurance and comprehensive motor insurance .Comprehensive motor insurance is not compulsory in most countries of the world. However, if one wants to guard against financial loss, the best policy is the comprehensive motor insurance because it covers compensation for car accidents and other kinds of future misfortunes.

It covers the following: damage made to another party, harm to the person, this policy will safeguard insured in that kind of situation. It also takes care of the insureds own vehicle and medical expenses. The only challenge with policy is that the premiums are high compared to the third party auto insurance policy. The needs of insurance in Ghana for example transcends travel insurance, fire insurance, motor insurance, marine insurance etc.

Though quite competitive, the governing laws of insurance have also increased with time. In Ghana, Insurance Act 58 states that all vehicle owners insure their vehicles against accidental risks and damages. Regardless of all the questions asked on insurance, Ghanaians are steadily accepting the fact of paying premiums to protect themselves in case of loss or damage.

Statistics taken on industries in Ghana shows that insurance companies have increased considerably over the past ten years and the desire of Ghanaians to purchase policies for their vehicles, businesses and households is irresistible. That is why most Insurance Companies are repositioning themselves to meet the future challenges of demand for Investments, Risk Management, Insurance Policies, and as well as Claim Payment Demands.

Although every insured expects a cushion in the event of his/her economic loss as stipulated in an Insurance Contract, the challenge of meeting the payment terms becomes an issue of much concern to the Insurer. The estimation of Expected Claim Liabilities of the Insurance Policies enables decisions on asset allocation and claim payment to be taken without much error thereby rising above the said challenges.

In actuarial automobile insurance applications, the outcome of interest often is the number of claims or the amount of money that the claims cost. These two variables are then modeled to depend on different attributes of the car and its owner of. These types of attributes can be the car make, the age of the driver, the age of the car, the geographical area where the driver lives etc. The result of the model is then used to set a fair price on the insurance.

The statistical framework of generalized linear models (GLMs) allows explicit assumptions to be made about the nature of insurance data and the relationships with predictive variables. Multivariate methods, such as GLMs can adjust for correlations and allow for investigations into interaction effects. Haberman Renshaw (1996) show the GLM to be applied to a wide area of application in insurance such as; survival modeling, multiple-state models in health insurance, fitting loss distributions for claim severity, premium rating and claims reserving.

The frequency and severity estimates were usually calculated through the use of past experience for groups of similar risk characteristics. Insurance data sets are typically extremely large and thus usually represented in tabular or grouped format according to variables and cases to present it easily. The areas of interest from a statistical point of view are inferences such as Extreme values.

This is interesting because by looking for outliers and extreme values, insurers can see the highest claim amounts and protect themselves in the future. Number of policies that do not submit a claim.

This is important because as a result all the premiums will become a profit and the insurer will be happy. Predicting outstanding claims will involve looking at data including responses such as the time the claim is submitted and the time the claim has been settled and can give the insurer an estimate of which claims are due to be soon paid. In this project I will look at motor insurance analysis focusing on modeling the number of claims i.e. the claim frequency and will touch on loss distributions i.e. claim sizes towards the end.

Credibility models are common in one regard: losses are independent among entities. Risk parameters are assumed to be independent for different entities. Losses are also assumed to be independent conditional on risk parameters. This means losses are unconditionally independent among entities.

Extensions of credibility models to allow for dependence among entities through risk parameters are common in the literature. A prominent example is the Jewell's hierarchical credibility model Jewell(1975) which introduces dependence among entities with its hierarchical Bayesian model formulation. Is the common effects model, considered by Yeo Valdez (2006) and Wen et al. (2009), which introduces dependence through an additional cross-entity latent variable. More generally, there are the crossed classification models by Dannenburg (1995) Goulet (2001).

Finally, Wen Wu (2011) propose a general framework through dependence among risk parameters. Another method to introduce dependence among losses is to alter the assumption of conditional independence of losses. Temporal conditional dependence has been studied extensively. Some examples are Frees et al. (1999) studied credibility predictors under the longitudinal data structure, Frees Wang (2006) used the elliptical copula to model temporal dependence, and Lo et al. (2006) and others also used estimation of regression credibility models with temporal dependence. It is very uncommon to have conditional cross-sectional dependence in credibility theory.

2.5 The Credibility Factor

According to Waters (1994) Credibility theory is a technique, or set of techniques, for calculating premiums for short term insurance contracts. The technique calculates a premium for a risk using two ingredients: past data from the risk itself and collateral data, i.e. data from other sources considered to be relevant. The important features of a credibility premium are that it is a linear function of the past data from the risk itself and that it allows for the premium to be regularly updated as more data are collected in the future.

The main purpose of credibility models is to predict the loss (or any other quantities of interest) in a future time period for a certain entity (for example, policy or region) using previously observed data.

Credibility factor Z is a measure of how much reliance the company is prepared to place on the data from the policy itself. The Z value is a number between 0 and 1 and is called the credibility weight or credibility factor. It expresses how credible or acceptable an item or estimate is. In credibility of data, 0 credibility is given to data that is too small to be used for premium rate making.

If some data has credibility of 1, this means the data is fully credible. Z value reflects how much "trust" is placed in the data from the risk itself (Y) compared with the data from the larger group μ. As an estimate of next year's expected aggregate claims or number of claims, the higher the value of Z, the more trust is placed in Y compared withμ and vice versa. It can be seen that, in general terms, the credibility factor would be expected to behave as follows: The more data there are from the risk itself, the higher should be the value of the credibility factor. When the collateral data is relevant, the lower should be the value of the credibility factor. While the value of the credibility factor should reflect the amount of data available from the risk itself, its value should not depend on the actual data from the risk itself.

One final point to be made about the credibility factor is that, while its value should reflect the amount of data available from the risk itself, its value should not depend on the actual data from the risk itself, i.e. on the value of Y.

2.6 Greatest Accuracy

The greatest accuracy credibility theory method uses both the variance of observations within each company and the variance across one company to another. Policy detail of each insurance company is needed whenever calculations involving greatest accuracy credibility theory approach is to used by actuaries . Since other company's data is classified, a company needs a statistical agent to access the other company data needed for the greatest accuracy credibility theory method. It is also called the Least Squares or Bühlman's Credibility. Greatest accuracy credibility theory originated from two seminal papers writen by Bailey (1945 1950). In his 1945 paper, Bailey obtains a credibility formula that seems to anticipate the nonparametric universe to be explored two decades later by Bühlman's.

Unfortunately, the paper suffered due to a somewhat awkward notation that made it difficult to read. The 1950 paper, on the other hand, was better understood and is considered as the pioneering paper in greatest accuracy credibility.

The credibility is given by the formula:$Z = N/(n+k)$. As the number of observations N increases, the credibility Z approaches 1. In order to apply Bühlman's Credibility to various real-world situations, one is typically required to calculate or estimate the so-called Bühlman's Credibility Parameter K. This involves being able to apply analysis of variance: the calculation of the expected value of the process variance and the variance of the hypothetical means. It is theoretically complete, and meets the criteria for a credibility method with one shortcoming.

That shortcoming is that additional information about industry experience (beyond what is customarily collected and published) is required. Without these practical difficulties, Greatest Accuracy Credibility Theory (GACT) would likely be the preferred credibility method to use in determining the expected valuation mortality assumption. There are several versions of GACT. One of the simplest is the Bühlman model.

2.7 Fundamentals of Bühlman's Credibility Theory

Bayesian ideas and techniques were introduced into actuarial science in a big way in the late 1960s when the papers of Bühlman (1967, 1969) and Bühlman Straub (1970) laid down the foundation to the empirical Bayes credibility approach, which is still being widely used in the insurance industry. Bayesian methodology is used in various areas within actuarial science.

The Bühlman credibility model uses a Bayesian framework and assumes the risk parameters of each unit to be independent and to follow a common distribution. In addition, conditional on the risk parameter of an unit, the losses are considered to be independent and identically distributed. The expected quadratic loss of a linear predictor is then minimized to produce the Bühlman credibity premium.

However a great deal of work has been done to extend Bühlman's model, including Bühlman Straub (1970) who generalize the model in cases where volume is involved, and the Hachemeister regression credibility model Hachemeister (1975) which introduces covariates to the conditional mean of losses.

Klugman et al.(2008) Denuit et al. (2007) were interested in studying the frequency distribution of insurance claims, and also the parameter estimation methods. Bühlman (1967) presented the credibility approach in the form of a linear function to estimate and predict the expected claim counts in future periods, using past data on claims as a risk class or connected risk classes. Bühlman's credibility theory is very interesting and can be extended to other approaches, such as the Bühlman-Straub model, Jewell's model or the exact credibility approach.

Calculating the expected claim counts using the credibility approach only depends on the information from prior experience of claim counts, and independent of the occurrence behaviour of claim counts over time frame. Other authors have discovered an alternative approach to claim counts relating to a specified time or their behaviour over time, for example, Mikosch (1972) viewed the claim counting process as a homogeneous Poisson process (HPP) in the Cramér-Lundberg model, a very popular and useful risk models in non-life insurance, and Matsui Mikosch (2009) also considered a Poisson cluster model for the modeling of a total claims amount by a point of claim counts as an HPP with a constant rate of occurrence called the constant intensity.

In non-life insurance portfolios, the claim counts during a certain time period are caused by periodic phenomena . These claim counts are modeled in terms of a non-homogeneous Poisson process (NHPP) with a period time-dependent intensity rate.

2.8 Credibility Theory

Credibility theory is a technique that can be used to determine premiums or claim frequencies (number of claims) in general insurance. This technique uses; Historical data related to the actual risk Data from other related but relevant sources commonly referred to as collateral data. The credibility premium formula as derived by Waters (1987) is of the form; $m = ZX + (1 - Z)\mu$
Where; m is the premium, z is the weight or credibility factor and is usually between zero and one. The credibility factor here is an increasing function for large value of n. The mean parameter x is the observed mean claim amounts per unit risk exposed for individual contract/risk itself. is the parametric estimate of the proposed data in the case than an assumption of the underlying distribution is made. For a series of risks, is the corresponding portfolio (set of risks) mean.

Here are some features of credibility formula;

- It is a linear combination of estimates to a pure premium policy based on observed data from the risk itself and the other based on projected risks.
- The credibility factor z, shows the degree of reliability of the observed risk data in the sense that high values of z implies high reliability.
- The credibility z is a dependent function of the number of claims. This

implies, the higher the claim number the larger the credibility factor.

- The value of credibility factor lies between zero and one.

2.9 Credibility theory development

Credibility theory was originally developed for a long time by actuarists from North America in the early 20th century. Mowbray (1914) put it into practical solution to premium calculation and it came to be called the American credibility theory. It is at times called the "limited credibility theory" or "the Fixed effect credibility". This study assumed that the yearly claims $X_1, X_2 \ldots X_n$are independently and identically distributed random variables from a probabilistic model with means $m(\alpha)$ and variance $s^2(\alpha)$. The assumption is that the data follows a normal distribution.

Whitney (1918) and other researchers criticized a lot this theory. Whitney proposed that claims are random in nature and hence assumption of fixed effects model was invalid. In addition, the theory also faced the problem of partial credibility since it was difficult to determine the value of the credibility factor. After the World War II revolution, Whitney's random effect model came into place.

Later on, Nelder and Verall derived credibility functions by the generalized linear model approach and consequently included the random effects model. This has provided a wide range of actuarial application among them is premium rating and reserving. Though a lot of research was done that yield several findings, it was found that the fixed effect credibility was not able to solve the problem of credibility. It is said that part of it was due to undeveloped or poor statistical background.

In 1967 and 1970, the real thing came when Bühlman derived the credibility premium formula in a distribution free-way such that there was no assumption of prior distribution of claims. Bühlman later clarified in this work the several assumptions of using the credibility premium formula (see Bühlman 1971).

This major breakthrough has seen much of the research tilting to the development of Bayesian estimation techniques by Jewell (1974, 1975), Hachmeister (1975), Devylder (1976, 1986) Gooverts Hoogstad (1987). Jewell (1974) showed that for exponential family distribution, the best linear approximation to Bayesian estimate is obtained using quadratic loss functions.Hachmeister (1975) extended the Bühlman Straub model by use of matrix methods.

2.10 Logistic Regression Models

Logistic regression was proposed as an alternative in estimating claims in the late 1960s and early 1970s Cabrera, (1994), and it became routinely available in statistical packages in the early 1980s. Since that time, the use of logistic regression has increased in the social sciences (Chuang, 1997; Janik Kravitz, 1994; Tolman Weisz, 1995) and in educational research, especially in higher education (Austin, Yaffee, Hinkle, 1992; Cabrera, 1994; Peng So, 2002; Peng, So, Stage, St. John, 2002).

However with worldwide accessibility of complicated statistical software for speedy computers, logistic regression is increasingly becoming useful. This expanded use demands that researchers, editors, and readers be attuned to what to expect in an article that uses logistic regression techniques.

The central mathematical concept that underlies logistic regression is the logit, the natural logarithm of odds ratio. The simplest example of a logit is derives from a 2×2 contingency table. Consider a situation in which the distribution of a dichotomous outcome variable (a child from an inner city school who is recommended for remedial reading classes) is paired with a dichotomous predictor variable (gender).

Generally, logistic regression is well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables.

Logistic regression also referred the logistic model or logit model, analyzes the association between multiple independent variables and a categorical dependent variable, and also estimates the probability of occurrence of an event by fitting data to a logistic curve. There are basically two types logistic regression model that is: binary logistic regression and multinomial logistic regression. Binary logistic regression is usually applied when the dependent variable is dichotomous and the independent variables are either continuous or categorical. When the dependent variable is not dichotomous and is made of more than two categories, a multinomial logistic regression can be used.

Jovanis Chang (1986) found a number of problems with the use of linear regression in their study applying Poisson regression as a means to predict accidents.

For example, they discovered that as the distance travelled by the vehicle increases, so does the variance of the accident frequency. Thus, this analysis violates the homoscedasticity assumption of linear regression. In a well-summarized review of models predicting accident frequency, Milton Mannering (1997) state that "the use of linear regression models is inappropriate for making probabilistic statements about the occurrences of vehicle accidents on the road". They proved that the negative binomial regression is a powerful predictive tool and one that should be increasingly applied in future accident frequency studies.

Nassar et al. developed an integrated Accident Risk Model (ARM) for policy decisions using risk factors affecting both accident occurrences on road sections, and injury severity of occupants involved in the accidents. The use negative binomial regression and a sequential binary logit model, the models they developed are realistic and easy to use.

Mercier et al. (1997) used logistic regression to determine whether either age or gender (or both) was a factor influencing severity of injuries suffered in head-on automobile collisions on countryside highways. Logistic regression was also used by Veilahti et al. (1989) in predicting automobile driving accidents of young drivers.

They examined the predictive values of the Cattel 16-factor personality on the occurrence of automobile accidents among conscripts during 11-month military service in a transportation section of the Finnish Defense Forces. James Kim (1991) developed a logistics regression model for describing the use of child safety seats for children involved in crashes in Hawaii from 1986 through 1991. The model shows that children riding in automobiles are less likely to be restrained and that drivers that use seat belts are far more likely to restrain their children; and one-and-two-year olds are less likely to be restrained.

2.11 Accident Year Statistics

An accident year embraces the entire population of claims incurred with accident dates in a particular calendar year, whether reported to the company in that year or subsequently (i.e., incurred/not reported). Under the accident year method of establishing claim reserves, there is an automatic grouping or segregation of all like accident year claims through successive calendar years as the total population of reported claims for each accident year moves to ultimate settlement status, that is, as reported claims increase in number to attain settlement or closure status and the balance or number of unpaid or unsettled claims approaches to (in that future year when all claims occurring in a specified accident year have been settled and closed.)

The accident year method of establishing aggregate reserves involves procedures for making successive annual (and quarterly for internal management and balance sheet reports) estimations of probable settlement costs of remaining unpaid claims in each accident year grouping of claims received. A study of past lags in claim reporting gives the needed statistical count from which one can proceed by the use of mean values as for known claims. The general method is to determine, for each accident year, the estimated total losses incurred, Ghana cedis L. From this total Ghana cedis L is subtracted the actual total losses paid for each respective accident year giving a balance, which is the claim reserve for claims that are not paid for.

The researcher has found through long experience that more satisfactory and adequate results are obtained by such estimation of losses incurred for the entire accident year population of claims and deducting there from the actual claims payments, than with methods by which unpaid losses on unsettled claims are directly estimated by multiplying unpaid loss averages by the number of claims not paid for. When the mean values for all claims incurred in an accident year are used it gives more comparability and stability than does the use of the unpaid claims only. However, in later stages of development the aggregate of individual claim estimates by claims experts do attain increasing reliability.

An important basic statistic is the count of claim notices received. The whole population of claims including notices of claims which ultimately may be settled without payment because of triviality, being less than the policy contract deductible, or non-liability. The method of counting the entire number of these random events (claim notices) has distinct advantages, the most important of which is being able to use the earliest possible indication of a claim liability incurred by the insurance company. The definition of the claim notice must be broad enough to permit immediate and exact identification of the claims. It is more appropriate to use a broad definition in order to be able to use claim notice statistics immediately, rather than to so refine the definition such that first notices of claims cannot be immediately classified.

For example, separate counts of Auto Liability (B I) and (PD) can be made for most U.S. claim notices but such may not be possible for single line third party liability. The use of individual company patterns to measure the rate of development is recommended because of the wide variations in individual insurance company practices of recording and counting claims reported, speed of settlement, distribution of business, seasonal traffic patterns, and differences in geographic areas as well as differences in general claims management practices.

2.12 Count Regression Models

The breaking up of data with more than 2 unique values into a dichotomous variable (e.g. "presence" or "absence") allows any dataset to be incorporated in a logistic regression model, such an approach considerably minimizes variance across spatial units. This will influence the statistical power of the model, which may increase the chance of accepting the null hypothesis when a significant relationship exists between the dependent variable and independent variable(s) a situation commonly referred to as a "Type II" error (Britt Weisburd(2010): 313).

In other words, when influenced, this statistical model may find the dependent and independent variable(s) to be unrelated even when a significant relationship really exists. Using the technique you will able to incorporate non-categorical data preserves the statistical power of the analysis, and may be preferable to use logistic regression in certain instances. Many analysts first consider linear regression models when working with non-categorical data.

Linear regression, particularly Ordinary Least Squares (OLS) regression, is one of the most traditional statistical techniques use in applied research. However, OLS regression models are based on particular assumptions which often-times are not satisfied with criminology data (Maxfield Babbie, (2001): 404). Ordinary Least Square regression assumes that the dependent variable is a continuous value, normally distributed, and linearly related to the independent variables McClendon (1994). Data on crime, in particular, rarely adheres to these assumptions. Most crime incidents are distributed as "rare event counts." Said differently, smaller values are much more common across spatial units than larger values with zero often being the most commonly observed value. Such a distribution violates the aforementioned assumptions of OLS regression.

Poisson and negative binomial regression models are designed to be used in analyzing count data. The "rare events" of nature crime counts are controlled in the formulas of both Poisson and negative binomial regression. However, Poisson and negative binomial regression models differ in regards to their assumptions of the conditional mean and variance of the predictive variable. Poisson models assume that the conditional mean and variance of the distribution are equal. Negative binomial regression models do not assume an equal mean and variance and particularly correct for over dispersion in the data, which is when the variance is greater than the conditional mean (Osgood, 2000); Paternoster Brame, (1997). Many have noted that criminological data rarely exhibits equal means and variances, leading to the increased popularity of negative binomial regression in contemporary studies of crime (MacDonald Lattimore, (2010).

Count regression analysis allows identification of risk factors and prediction of the expected frequency given the characteristics of the policyholders. A usual way to calculate the premium is to obtain the conditional expectation of the number of claims given the risk characteristics and to combine it with the expected claim amount. Some insurers may also consider experience rating when setting the premiums, so that the number of claims reported in the past can be used to improve the estimation of the conditional expectation of the number of claims for the following year

2.13 Generalized Linear Models (GLM)

Nelder Wedderburn (1972) were the first to introduce the generalized linear model (GLM). The GLM provides a unified framework to study various regression models, rather than a separate study for each individual regression. Generalized linear model (GLM) is an extension of the classical linear models. It includes linear regression models, Zero-inflated Poisson regression models, analysis of variance models, Poisson regression models, , Negative Binomial regression models, log-linear models, , Logistic regression models as well as many other models. These models have a number of unique properties in common, such as linearity and a method for parameter estimation. A generalized linear model has these three components:

1. A random component, which specify the conditional distribution of response variable, Y_i, given the explanatory variables, x_ij
2. A linear function of the regression variables, called the linear predictor,$Y_i = \alpha+\beta_i X_{i1}+......+\beta_n X_{in} = X_i^{'}\beta$ On which the expected value μ_i of Y_i depends.
3. An invertible link function, $g(\mu_i) = \gamma_i$ Which transforms the expectation of the response the linear predictor? The inverse of the link function sometimes called the mean function: $g^{-1}(\gamma_i) = \mu_i$

For traditional linear model for which the random component consists of the assumption that the response variable follows the Normal distribution, the conical link function is the identity link. The identity link specifies that the expected mean of the response variable is identical to the linear predictor, rather than to a non-linear function of the predictor. That is, for the normal linear model, the link function $g(u_i) = u_i$ The Generalized Linear Model is an extension of the Linear Model to include response variables that follow any probability distribution in the exponential family of distributions.

Many commonly used distributions in the exponential family are the inverse Gaussian distributions, normal, exponential, binomial, Poisson, gamma and. In addition, several other distributions are in the exponential family and they include the beta, multinomial, Pareto and Dirichlet. There are other several distributions which are not in the exponential family but are used for statistical modeling and they include the student's t and uniform distributions.

2.14 Probability Models for Claims Estimation

When an insured is observed over period of time, longitudinal data model could be used to estimate the frequency of claims. This consists of repeated observations of individual units that are studied over time. Each individual is assumed to be independent of others but correlation between observations relating to the same individual is permitted. Here it is assumed that the number of claims per year follows Poisson distribution with a parameter specific to each policy holder.

Specifically, $N_{i,j}$ is assumed to be Poisson distributed with mean $\theta_i \lambda_{i,j} i = 1, ..., T$. The expected annual claim frequency is a product of a static factor θ_i and a dynamic factor $\lambda_{i,j}(\theta_i \lambda_{i,j})$. The formal account for the dependence between observations relating the same policyholder. The later introduces the observable characteristic (that allows varying with time.

In general, $ln\lambda_{i,j}$ is expressed as a linear combination of observable characteristics, that is$\lambda_{i,j} = exp(\beta_0 + \beta X_{i,j}), where\beta_0$ in the intercept, $(\beta' = \beta_i\beta_2, ..., \beta_p)$is a vector of regression parameters for explanatory variables $x_{i,j} = (X_{i,j1}, X_{i,j2}, ..., X_{i,jp})^T$ Boucher Denuit(2005).

Chapter 3

METHODOLOGY

3.1 Introduction

This chapter deals with the model and the model development. Bühlmann'scredibility model would be used to analyze the auto insurance claim data.

3.2 Data Collection

Data was collected from the auto insurance claims department of the Quality Insurance Company in Ho Municipality. The data was randomly selected using the inclusion criteria. The policyholders selected are those who renew their third party policies for at least three consecutive years and also made at least one or no claims in the three successive periods .If no claim was made in a policy year it was recorded as zero(0) and if a claim was made in a policy year it was recorded as one(1). The data was collected over the 2012 to 2014 policy years.

3.3 Introduction to Credibility

Credibility theory provides important tools to help the actuary deal with the randomness inherent in the data that he or she analyzes. Actuaries use past data to predict what can be expected in the future, but the data usually arises from a random process. In insurance, the loss process that generates claims is random. Both the number of claims and the size of individual claims can be expected to vary from one time period to another. The actuary wants to know how much to believe the data that's being analyzed. To use the data to predict the future, this "belief in the data" must be quantified so that calculations can be made. This leads us to actuarial credibility which is the weight to be given to data relative to the weight to the other data.

Bühlmann's approach of updating the predicted premium is based on a linear predictor using past observations. It is also called the greatest accuracy approach or the least squares approach. The Bühlmann's credibility factor Z depends on the sample size n and the expected value of the process variance (EPV) to Variance of hypothetical means (VHM) ratio k. In particular, Z varies with n and k as follows:

- Z increases with the sample size n of the data.
- Z increases with the distinctiveness of the risk groups. The risk groups are more distinguishable when k is small. Thus, Z increases as k decreases.

The fundamental formula for calculating credibility weighted estimates is: $Estimate = Z \times [Observeddata] + (1 - Z) \times [collateraldata], and 0 \leq z \leq 1$. If our body of data is so large that we can give full weight to it in making our estimate, then we would set $Z = 1$.

3.4 Bühlmann's Credibility Model Assumptions

We now state formally the assumptions of the Bühlmann's model and derive the updating formula as the least mean-squared-error (MSE) linear predictor.

- The Bühlmann's model assumes that there are n observations of losses, denoted by $Y_1, ..., Y_n$.
- The observations may be losses recorded in n periods and they are assumed to be independent and identically distributed (iid) as Y, which depends on the parameter θ.

- The task is to update the prediction of Y for the next period, that is ., $Y_{(n+1)}, based on Y_1, ..., Y_n$.

- In the Bühlmann approach the solution depends on the variation between the conditional means as well as the average of the conditional variances of the risk groups.

3.5 Bühlmann's Credibility Model Development

The Buhlmann model assumes that for any selected risk, the random variables $Y_1, Y_2, ..., Y_n, Y_{n+1}, ...$ are independently and identically distributed. For the selected risk, each Y_t has the same probability distribution for any time period t, both for the $X_1, X_2, ..., X_n, X_{n+1}, ...$random variables in the experience period, and future outcomes $Y_{n+1}, Y_{n+2}, ...$ As Hans Bühlman described it, "homogeneity in time" is assumed.

Considering a contract, observed during T years. Let Y_t denote the number of claims and assume $Y_t \epsilon (0, 1)$ so that $Y_t \sim Bernoulli(P)$ where $i = 1, ..., m$
and
$t = 1, ..., T$ If we assume a priori beta distribution $B(\alpha, \beta)$ for p then where α and β are the scale and shape parameters in the Beta distribution T is the period under investigation.

$$E(P|Y) = \frac{\alpha + T\bar{Y}_T}{\alpha + \beta + T} \tag{3.1}$$

$$C = \left(\frac{T}{\alpha + \beta + T}\right)\bar{Y}_T + \left(1 - \frac{T}{\alpha + \beta + T}\right)\frac{\alpha}{(\alpha + \beta)} \tag{3.2}$$

Where C is the credibility estimates

$$Z = \frac{T}{\alpha + \beta + T} \tag{3.3}$$

Where $z\epsilon[0,1]$,

$$E(P|Y) = PvarE(Y|P) = P(1-P) \tag{3.4}$$

$$\frac{E(Var(Y_t|P))}{Var(EY_t|P)} = \frac{E(P(1-P))}{(Var(P))} \tag{3.5}$$

$$\frac{E(P(1-P))}{(Var(P))} = \frac{E(P) - [Var(P) + E(P)^2]}{Var(P)} = \alpha + \beta \tag{3.6}$$

$$Z = \frac{T}{(\alpha + \beta + T)} = \frac{T}{(K+T)} \tag{3.7}$$

$$K = \frac{E(Var(Y_t|P))}{Var(EY_t|P)} \tag{3.8}$$

If insured I was observed T years with past experience $Y_t = Y_1, ..., Y_T$ then The Bühlmann credibility model is

$$E(Y_{T+1}|Y_T) = Z\bar{Y}_{i,T} + (1-Z)\mu \tag{3.9}$$

$$\bar{Y}_T = \frac{1}{T}\sum_{t=1}^{T} Y_{i,t} \tag{3.10}$$

$\bar{Y}_T$ is mean of all the claims recorded in the data. $\bar{Y}_{i,T}$is the mean of each of the risks within the three years

$$S^2 = \frac{1}{m(T-1)}\sum_{t=1}^{m}\sum_{t=1}^{T}(X_{i,t} - \bar{X}_t)^2 \tag{3.11}$$

$$a = \frac{1}{m-1}\sum_{t=1}^{m}(\bar{X}_t - \bar{X})^2 - \frac{1}{n}S^2 \tag{3.12}$$

$$\tilde{Z} = \frac{T}{T + \frac{S^2}{a}} \tag{3.13}$$

Bühlmann $\tilde{Z}$is the credibility factor estimator

From equation (3.7), we estimate the credibility factor in equation (3.13). To obtain the Bühlmann credibility weighted estimates from equation (3.9), we use :

$$\hat{\theta}_i = \hat{Z}\bar{Y}_{i,T} + (1 - \hat{Z})\mu \tag{3.14}$$

In equation (3.9) μ is the average on the whole portfolio. The optimal linear forecast Y_{T+1} given in equation (3.9) is also called the premium. Equation (3.7) is the share of the premium base on past information for a given insured and is related to the amount of credibility we give to past information. It also called credibility factor. The factor calculated with equation (3.8) is credibility parameter. Note that k depends only on the parameters of the model, while Z is a function of k and the size n of the data. For predicting claim frequency N, the sample size n is the number of periods over which the number of claims is aggregated.

For predicting claim severity X, the sample size n is the number of claims. As aggregate loss S refers to the total loss payout per period, the sample size is the number of periods of claim experience.

Equation (3.11) is the structure parameter and is a global measure of the stability of the portfolio's claim experience. It is sometimes called the homogeneity within the insureds. Equation (3.12) is a measure of the variation of the various individual risk premiums and is sometimes called the homogeneity between the insureds.

Table 3.1shows the credibility weighted estimate format for the expected number of claims per year. The following is the definition of the parameters used in the table. θ_i : the credibility weighted estimate for the expected number of claim per year for each Policyholder. $Y_{i,j}$: number of claims for policyholder i at year j. $\hat{\theta}$ is the mean of all the credibility weighted estimates The expected estimate of claim counts is calculated in equation (3.15) below:

$$\hat{\theta} = \frac{\sum \theta_i}{N} \tag{3.15}$$

Table 3.1: credibility weighted estimate format for the expected number of claims per year for each risk

	Time period				
Policy holder	1	2	...	K	Risk Parameter
1	Y_{11}	Y_{12}	...	Y_{1N}	θ_1
2	Y_{21}	Y_{22}	.....	Y_{2N}	θ_2
:	:	:	....	:	:
N	Y_{N1}	Y_{N2}	.....	Y_{Nk}	θ_N

Chapter 4

DATA PRESENTATION, ANALYSIS AND DISCUSSION OF FINDINGS

4.1 Introduction

In this chapter, we shall deal with the model implementation and discussion of results obtained from the Bühlman credibility model. The data for this study has been obtained from the Quality Insurance Company in Ho Municipality.

4.2 The model data

The Table 4.1 below shows the data for claim counts data from third party auto insurance for commercial vehicles in Ho municipality.

Table 4.1: Claim Counts

Year	Claim (%)	No Claim (%)
2012	236 (39.3)	364 (60.7)
2013	116 (19.3)	484 (80.7)
2014	236 (39.3)	364 (60.7)

From Table 4.1, the least number of claims of (116) representing 19.3% was recorded in 2013.There was no difference in claim counts in 2012 and 2014 respectively. Meanwhile in 2012 and 2014 a total of 236 claims representing 39.3% was recorded .It can also be seen that there is a difference between the number of claims and no claims from 2012 to 2013.

Table 4.2: Claim Frequency over the three year period

Claim	Frequency (%)
0	138 (23.00%)
1	338 (56.33%)
2	124 (20.67)
3	0
Total	600

From table 4.2 above 23%(138 policyholders) received no claims in the three policy years. 56.33%(338 policyholders) received exactly one claim during the three years.20.67%(124 policyholders) made exactly two claims in the three policy years.

There is no policy holder who had exactly three claims in the three years .The highest proportion of claims paid are those who had exactly one claim followed by those who had no claims over the period and the least being those who had cumulative two claims over the period.

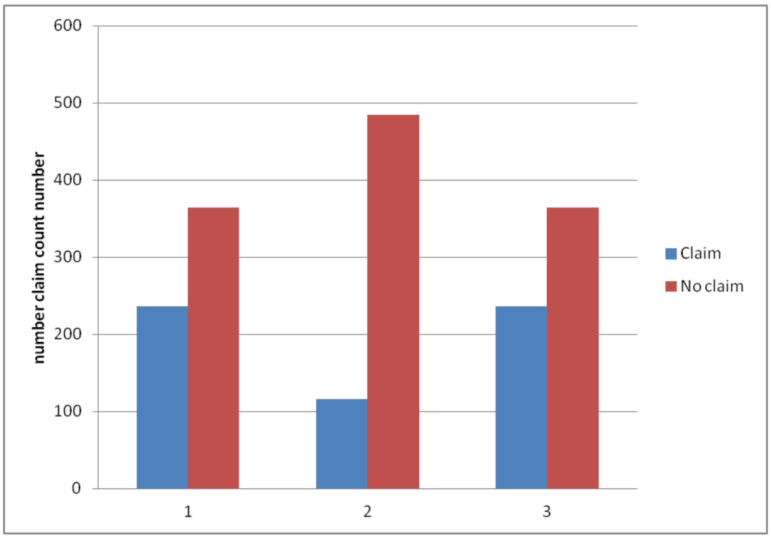

Figure 4.1: Number of Claim counts

Figure 4.1 Visualize the information in table 4.1 It shows that in each year there are less claims received than no claim. The distribution 0f 2012 claims data is similar to that of 2014 this may be coincidal reported claims. A policyholder who makes a claim in 2012 is likely to make a claim 2013.

A policyholder who makes no claim 2012 is likely to make a claim in 2013.There are certain policyholders who did not make claims in any of the policy years. A policyholder who did not make a claim in 2013 will probably make a claim in 2014. A policyholder who makes no claim in 2012 may probably make a claim in 2013.

Table 4.3: Credibility Estimates for the first 20 policyholders

Credibility Factor Z= 0.33	
Policyholder	2015 Weighted Estimates
1	0.331
2	0.331
3	0.331
4	0.441
5	0.221
6	0.331
7	0.331
8	0.331
9	0.441
10	0.221
11	0.331
12	0.331
13	0.331
14	0.331
15	0.331
16	0.331
17	0.331
18	0.331
19	0.442
20	0.221

Table 4.3 above shows the of the credibility weighted estimates for 2015 for the first twenty policyholders with the credibility factor of 0.33 or 33%. The outcomes of the estimates are independent of each other. Policyholders (5, 10, 20) had the lowest weighted credibility estimates of 0.221 which means that they will pay the least premium because they less probable to make a claim.

Majority of the of the policyholders have a credibility estimate of 0.331 which means they will have more weight to their premium because they are more likely to make more claims compare with 0.221 credibility estimates. However policyholders with credibility estimate of 0.441 will pay the highest premium because they are more probable to make a claim more any other insured within the policy years.

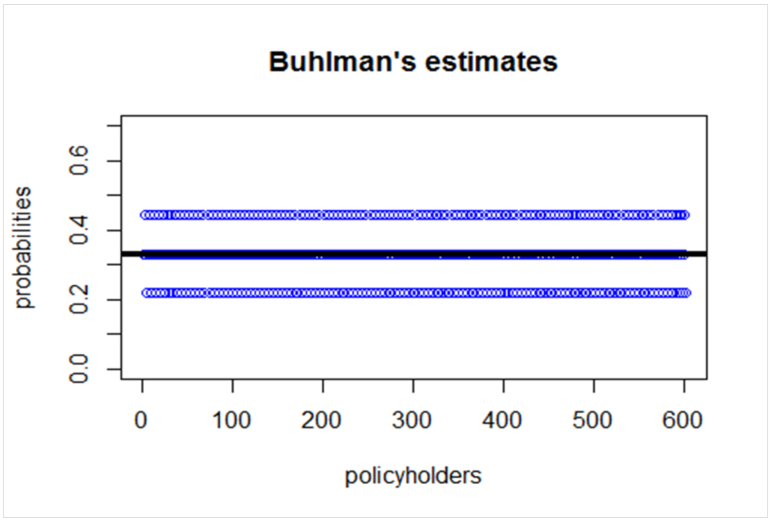

Figure 4.2: Bühlman estimates.

From the figure 4.2 above, the actual credibility factor is the middle thicker line (the threshold). About 23% of the 600 policyholders fall below the threshold implying that they have less weight than the credibility factor , whiles 44% of them fall above the threshold indicating that they are more are likely to make claims. The estimated weighted claims on the average $\hat{\theta}$ is 0.329.

Table 4.4: Estimated Number of Claim Counts for 2015

Number of Policyholders	Estimated Claim Counts for 2015
100	33
200	66
300	99
400	131
500	164
600	197

From the Bühlman credibility model employed, the estimated number of claim counts in 2015 has been computed as shown in table 4.4 above. The model predicts that in 2015, out of 100 policyholders 33 of them may probably make claims. The same trend follows for 200, 300, 400 and 500 policyholders. It also observed that claim counts also increases with number of Policyholders. Finally for the 600 policyholders used in the study 197 representing 33% may probably make claims in 2015.

Chapter 5

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

The main objective of the study was to estimate the future claim counts of the insureds using claim experiences of the insureds by using Bühlman credibility theory and to use the model to calculate credibility weighted estimates for the expected number of claims in 2015. Data for the study was collected from Quality insurance company in Ho Municipality.

Zero (0) was assigned for no claim whiles one (1) represents a claim made. A bar chart was used to visualize the yearly claim and no claim for the policyholders. The distribution of 2012 claims data was coincidally similar to that of 2014.Thus a policyholder who makes a claim in 2012 may probably make no claim in 2014 .Similarly a policy holder who makes no claim in 2012 can probably make a claim in 2014.

The data was categorize into the policy holders who made no claim, one claim, two claims and three within the three periods .The table shows that policy holders with exactly one claim have highest proportion of claims followed by those with zero claim then those with exactly two claims within the policy years.

The R software was used to compute the credibility factor and the credibility weighted estimates for 2015. It was also used to compute the overall mean of the credibility weighted estimates and the graphical presentations of these estimates.

The graph shows that the proportion of policyholders who are weighted with high premiums was the same as those who are rated with low premiums .The middle thicker line is the credibility factor. Finally the model predicted estimates for 100, 200, 300, 400, and 500 policyholders

5.2 Conclusion

The Buhlmanns credibility was developed and used to estimate the weighted credibility estimates of each of the policyholders. The model estimated that for the 600 policyholders used in the study, 197 of them may probably make claims in 2015.

5.3 Recommendations

Based on the findings of the study, we therefore recommend that the insurance company should make claims reservation for 197 policyholders in 2015.
We also recommend that Bühlman's credibility theory be used by all insurance companies in estimating future claim counts. This is because it does not take into account all the risk factors in calculating insurance claims as the case in other statistical methods.

5.4 REFERENCES

- Aeron-Thomas, A. (2002)."The Role of the Motor Insurance Industry in Preventing and Compensating Road Casualties", Scoping Study Report PR/INT/243/02,

- Amoo G. K (2002). Going the extra mile: The challenge of providing insurance cover for loss of Use of motor vehicle in a developing economy. A dissertation summated to Chartered Insurance Institute.

- Boland A. Lawrence, (2007).The identification problems and the validity of economic models. South African Journal of Economics 36 (Issue 3):236-240.

- Bühlman H., Gisler A. (2005). A course in Credibility Theory and its Applications. Springer-Verlag Berlin Heidelberg: Netherlands

- Bühlman H, (1967) "Introduction Report Experience Rating and Credibility," ASTIN Bulletin, Vol. 4, No. 3, , pp. 199-207.

- Bühlman H, "Credibility Procedures," Proceedings of the Sixth Berkeley Symposium on thematical Statistics and Probability, University of California Press, Berkeley, Vol. 1,(1972) pp. 515-525.

- Esho, N., Kirievsky, A., Ward, D., and Zurbrueg, R. (2004). "Law and determinants of Property Causality insurance", Journal of Risk and Insurance 71(2) 713-734.

- Maddala G.S. (2005) "Introduction to Econometrics" 3rd Edition. John Wiley and Sons Ltd. The Atrium, Southern Gate, England PP 318-323

- Onafalujo, A.K., Abass, O.A., and Dansu, S.F. (2011). "Effect of Risk Perception on the Demand for Insurance: Implication on Nigerian Road Users", Journal of Emerging Trends in Economics and Management Sciences 2(4) 285-290

- Outreville, J.F (1990) "The Economic Significance of Insurance Markets in Developing Countries", Journal of Risk and Insurance, 63(2): 263-278.,
- USAID, (2006). Assesment on How Strengthening. The Insurance in Developing Countries.
- James, J. L. and Kim, K. E. Restraint Use by Children Involved in Crashes in Hawaii(1986- 1991). Transportation Research Record 1560, 8-11, 19.
- Jørgensen, B., and Souza, M. C. P. de (1994). Fitting Tweedie´s compound Poisson model to insurance claims data. Scandinavian Actuarial Journal, 1(1),(69-93).
- Jovanis, P. and Chang, H. (1986) Modeling the relationship of accidents to miles traveled. Transportation Research Record 1086, 42-51.
- M. Matsui and T. Mikosch, "Prediction in a Poisson Cluster Model," Journal of Applied Probability, Vol. 47, No. 2,(2010), pp. 350-366. doi:10.1239/jap/1276784896
- Myth, G. K., and Jørgensen, B. (2002). Fitting tweedie's compound Poisson model to insurance claims data: dispersion modeling. Actuarial Studies in Non-life insurance (ASTIN) Bulletin
- Milton, J. and Mannering, F. The Relationship (1997). Among Highway Geometric, Traffic-32(1), 143-157. doi: 10.2143/AST.32.1.1020
- Mercier, C. R., Shelley, M. C., Rimkus, J., and Mercier, J. M. (1997). Age and Gender as Predictors of Injury Severity in Head-On Highway Vehicular Collisions. The 76th Annual Meeting, Transportation Research Board, Washington, D.C.. Related Elements and Motor-Vehicle Accident Frequencies. The 76th Annual Meeting, Transportation Research Board, Washington, D.C.

- Meulbroek L(2001), "The Efficiency of Equity-Linked Compensation: Understanding the Full Cost of Awarding Executive Stock Options," Financial Management (Summer 2001), 5-30.

- M. Morales, “On a Surplus Process under a Periodic Environment: A Simulation Approach,” Nassar, S. A., Saccomanno, F.F., and Shortreed, J.H (1997). Integrated Risk Model (ARM) of Ontario. The 76th Annual Meeting, Transportation Research Board, Washington, D.C.,.

- North American Actuarial Journal, Vol. 8, No. 2, (2004), pp. 76-87.

- R Development Core Team (2007). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Retrieved January 12, 2008, from http://cran.r-project.org/doc/manuals/refman.pdf

- S. A. Klugman, H. H. Panjer and G. E. Willmot, “Loss Models from Data to Decisions,” 3rd Edition, John Wiley and Sons, Hoboken,(2008).

- T. Mikosch, “Non-Life Insurance Mathematics,” 2nd Edition, Springer-Verlag, Berlin, 2009.doi:10.1007/978-3-540-88233-6

- Viaene et al, (2007). Strategies for Detecting Fraudulent Claims in the Automobile Insurance Industry. European Journal of Operational Research, Elsevier, vol.176 (1), 565-583.

- Wüthrich, M. V., and Merz, M. (2008). Stochastic claims reserving methods in insurance. West Sussex: John Wiley and Sons.

- Weisberg H, Tomberlin T (1982). A Statistical Perspectwe on Actuarial Methods for Estimating Pure Premtums from Cross-Classified Data Journal of Rtsk and Insurance 49, 539 563

5.5 Appendix A

R code for computation of data

```
>T <- ncol(mydata)
(m <- mean(mydata))
(s2 <- mean(apply(mydata,1,var)))
(a <- var(apply(mydata, 1,mean))-s2/T)
#Bühlman's credibility factor
(Z <- T/(T+s2/a))
#Bühlman's estimates
> Z*apply(mydata,1,mean)+(1-z)*m
```