A RISK-BASED ADJUSTMENT MODEL FOR EXPERIENCE RATING OF MOTOR INSURANCE IN NIGERIA

BY

MESIKE, GODSON CHUKWUNWIKE
(960202026)

Department of Actuarial Science & Insurance
University of Lagos

August, 2017

A RISK-BASED ADJUSTMENT MODEL FOR EXPERIENCE RATING OF MOTOR INSURANCE IN NIGERIA

BY

MESIKE, GODSON CHUKWUNWIKE
(960202026)

B.Sc. (Hons.) Actuarial Science, University of Lagos, Nigeria 2001
M.Sc. Actuarial Science, University of Lagos, Nigeria 2006
M.Sc. Applied Actuarial Science, University of Kent, United Kingdom 2014

A THESIS SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES, UNIVERSITY OF LAGOS IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D) IN ACTUARIAL SCIENCE

Department of Actuarial Science & Insurance
University of Lagos

August, 2017

SCHOOL OF POSTGRADUATE STUDIES
UNIVERSITY OF LAGOS

CERTIFICATION

This is to certify that the Thesis:

A RISK-BASED ADJUSTMENT MODEL FOR EXPERIENCE RATING OF MOTOR INSURANCE IN NIGERIA

Submitted to the
School of Postgraduate Studies
University of Lagos

For the award of the degree of

DOCTOR OF PHILOSOPHY (Ph.D.)

Is a record of original research carried out
By:

MESIKE, GODSON CHUKWUNWIKE

In the Department of Actuarial Science & Insurance

……………………………..….	…………………	…………………
AUTHOR'S NAME	**SIGNATURE**	**DATE**
……………………………..…..	…………………	…………………
1ST SUPERVISOR'S NAME	**SIGNATURE**	**DATE**
………………………….. …..	…………………	…………………
2ND SUPERVISOR'S NAME	**SIGNATURE**	**DATE**
……………………………..…..	…………………	…………………
1ST INTERNAL EXAMINER	**SIGNATURE**	**DATE**
………………………………....	…………………	…………………
2ND INTERNAL EXAMINER	**SIGNATURE**	**DATE**
…………………………..……	…………………	…………………
EXTERNAL EXAMINER	**SIGNATURE**	**DATE**
…………………………..……	…………………	…………………
SPGS REPRESENTATIVE	**SIGNATURE**	**DATE**

DEDICATION

I dedicate this work to the glory of Almighty God, and my beloved parents.

ACKNOWLEDGEMENT

It is the grace, mercy, charity, forgiveness, help and kindness of the almighty God that has made me to be alive, achieve this success and strength to go through all the difficult time. So, all praise is to the Almighty God, without whom I am nobody today! Thank you Lord.

This work was written with help of many individuals who have given invaluable assistance. I would like to express my gratitude to all of them and notably to my supervisors at the Department of Actuarial science and Insurance. I would like to express my deep and sincere gratitude to my supervisors Professor RasheedkolawoleOjikutu and Dr.IsmailaAdedejiAdeleke for their unreserved support, intellectual inspirations, personal guidance and encouragement through my Ph.D study. Their wide knowledge and extensive discussions, as well as their constructive comments about my work have been of great value for me and provided a sound basis for my thesis.

In addition to my supervisors I have been very fortunate to have had understandable, and encouraging colleagues at the department of Actuarial Science and Insurance, University of Lagos. I am grateful to you all. I also appreciate the efforts and contributions of Dr.OlufemiAkintunde and Dr AkeemAkinwale.

Also, I am grateful to my teachers, mentors, role models and well wishers from whose various spectra of expertise and depths of intellectual knowhow I have drawn scholarly inspiration and academic craftsmanship. Mentioning a few of such personalities but without prejudice to other invaluable individuals I may not hint at here, I acknowledge the following: Professor Richard Ayorinde, Professor Ade Ibiwoye, Professor Joseph Mojekwu, Professor Malcom Brown, Professor Paul Sweeting, Professor Martin Ridout, Professor Jian Zhang, Professor Philip Brown. John Millet, Dr.PradipTapadar, Dr. Diana Cole, Dr. Alfred Kume, Dr. Owen Lyne, CaireBellis, Nick Wood, Peter McQuire, JaideepOberoi, Andrew Jackson and Peter Duffet. As actuarial professors, statistician and actuaries, they have provided a solid knowledge of actuarial science and embedded a deep appreciation for scholarly rigor.

I thank immensely all my friends and colleagues for their supports and encouragement, Samuel Ojogbo, RasakiBakare, TundeElegbede, SeyiSode, Thaddeus Olufayo, and DoyinSonuga.

This research manuscript has also been trivial without a commitment, material and moral supports I have received from my family, friends nearby and class mates. So, my earnest gratitude and gratefulness goes to my family members, especially to my parents Mr Leonard Mesike and Mrs. Victoria Mesike for their unreserved encouragement and belief in me since my childhood days, and my siblings: Dr. Samuel Mesike, Engr. Mike Mesike (late), Engr. Kingsley Mesike, Engr. Emmanuel Mesike, Mrs. Stella Mesike-Osazuwa. It is because of their subtle fostering encouragement and inspiring example that have seen me through this work. I cannot overlook the helping hands that Prof. OwolabiKuye,Dr.LanreAdebiyi and Sunday Olaniyan has extended to me – they are just more than friends and colleagues!

My gratitude is incomplete without acknowledging my adorable wife, Sonia Mesike and loving son, Gerald Mesike (Virus jnr) for letting profound love flow ceaselessly unto me from both sides. Finally; my warmest thanks go to all my friends and well-wishers, specifically whose best wishes encourage me all the time and people who in their own ways contributed to the successful completion of this research but could not be listed here due to the limitation of space. Nevertheless, I acknowledge their efforts and contribution and extend my utmost thanks. Speechless- you did a good stuff Guys.

Abstract

In establishing the burden of costs, the principle underlying the calculation of differentiated premium in the insurance portfolio is characterized by a pricing process that involves the classification of all risks regarding factors of influence. This classification is based on known observable characteristics of the insured. However, there are many other important factors that are unobservable by the insurer which cannot be taken into account a-priori when pricing motor liability insurance products but may represent significant risk factors. Also, it has become extremely difficult recently, for insurance companies to maintain cross subsidies between different risks categories in a competitive market. As competition between insurance companies intensifies, higher efficiency and greater focus on profitability are required. While the potential for cost reductions is limited, improvements in profitability and growth can be achieved through appropriate pricing mechanisms. Experience rating which is popularly referred to as No Claim Discount or Bonus-Malus Systems involves modifying premiums using claims records. Risk-based adjustment pricing is an experience rating technique commonly used in motor insurance to categorize policyholders into relatively homogenous group who pay premium relative to their claims experience. In this study, a risk-based adjustment model that incorporates costs in a fair and equitable manner given the individual characteristics of the insured for experience rating is adopted using generalized linear model. Claim cost and frequency data from motor insurance liability portfolio in Nigeria as well as the insured characteristics were collected and analysed using the generalized negative binomial and gamma regressions. Individual risk weights from the fitted models were used to compute the risk scores. This was subsequently used to determine the relative costs of an insured based on their individual characteristics and claims history. Results show that the claims data from automobile insurance scheme is highly peaked and leptokurtic. The claims data also vary significantly across age groups, gender, occupation, and nature of loss, as well as the place of residence, type of product and customer type. The study established that motor insurance risks are influenced by individual risk characteristics and a risk-based adjustment pricing be introduced to establish fair and equitable costs among the insured. It is recommended that a risk-based adjustment pricing be employed to estimate accurately the average expected loss in order to charge adequate price for motor insurance.

Keywords: Experience rating, Risk-based adjustment model, Generalized linear model, Motor insurance, Fair pricing

Table of Contents

Title Page	i
Approval page	ii
Certification page	iii
Dedication	iv
Acknowledgement	v
Abstract	vii
Table of Contents	viii
List of Tables	xi
List of Figures	xii
Abbreviations	xiii
Chapter One	1
Introduction	1
1.1. Background to the Study	1
1.2. Statement of the Problem	4
1.3. Aim and Objectives of the Study	7
1.4. Research Questions	7
1.5. Significance of the Study	7
1.6. Scope and Delimitation of the Study	8
1.7. Operational Definition of Terms	9
Chapter Two	11
Literature Review	11
2.1. Preamble	11
2.2. Theoretical Framework	11
2.2.1. Generalized Linear Model Theory	11
2.3. Conceptual framework	12
2.4. Empirical Review of Non-life Insurance Pricing	14
2.5. Theoretical Review	17
2.5.1. Credibility Theory	17
2.5.1.1. Limited fluctuation credibility	18
2.5.1..2. Greatest accuracy credibility	19

2.5.2. Markov chain theory 21
2.6. Claim Counts Model 23
2.6.1. Poisson Regression 24
2.6.2. Mixed Poisson Model 24
2.7. Claim Severity Model 26
2.8. Brief Historical Background of Insurance in Nigeria 27
2.8.1. Nigeria Insurance Market Industry 29
2.9. Motor Insurance Policy in Nigeria 30
2.10 Experience Rating System 32
2.10.1. The No Claim Discount System in Nigeria 34

Chapter Three 36
Methodology 36
3.1. Preamble 36
3.2. Research Design 36
3.3. Population of the Study 37
3.4. Type and Sources of Data Collection 37
3.5. Method of Data Analysis 38
3.5. Generalized Linear Model 38
3.5.1. Maximum Likelihood Estimation 40
3.5.2. Exponential Family of Distributions 40
3.5.3. The Variance Function 41
3.5.4. Standard distributions in the exponential family form 42
3.5.4.1. Binomial 42
3.5.4.2. Normal 43
3.5.4.3. Poisson 43
3.5.4.4. Negative binomial 43
3.5.4.5. Gamma 44
3.5.4.6 Inverse Gaussian 44
3.6 . The Proposed Risk-based Adjustment Model 45
3.6.1. Estimation Model of Claim Frequency 45
3.6.2. Estimation Model of Claim Cost 47

3.6.3. Criteria for Assessing the Models' Goodness of Fit 48
3.6.4. Risk Premium Modelling 48
3.7. The Risk-based Adjustment Modelling 49

Chapter Four 51
Data Presentation and Analysis 51
4.1. Descriptive Statistics for the Insured Portfolio 47
4.2. Automobile Claims Modelling 56
4.3. Implementation of the risk adjustment model 67

Chapter Five 70
Summary Conclusion and Recommendations 70
5.1. Summary of Findings 70
5.2. Conclusion 72
5.3. Recommendations 72
5.4. Contribution to Knowledge 74

Biblography 75
Appendix 92

List of Tables

Table 2.1: Level of NCD System of NIA 34

Table 4.1: Frequency distribution of policyholder in the portfolio 52

Table 4.2: Descriptive analysis of claim cost, claim frequency and premiums by age 53

Table 4.3: Descriptive analysis of claim cost, claim frequency and premiums by gender 53

Table 4.4: Descriptive analysis of claim cost, claim frequency and premiums by product type 54

Table 4.5: Descriptive analysis of claim cost, claim frequency and premiums by district 54

Table 4.6: Descriptive analysis of claim cost, claim frequency and premiums by occupation 55

Table 4.7: Descriptive analysis of claim cost, claim frequency and premiums by loss type 55

Table 4.8: Descriptive analysis of claim cost, claim frequency and premiums by customer type 56

Table 4.9: Likelihood Ratio Statistics for Type 3 Analysis 57

Table 4.10:Goodness of fit test 58

Table 4.11:Analysis of Parameter Estimates 59

Table 4.12:Goodness of fit test 60

Table 4.13:Analysis of Parameter Estimates 62

Table 4.14:Wald Statistics for Type 3 Analysis 62

Table 4.15:Negative binomial regression analysis of automobile claim frequency 63

Table 4.16:Wald Statistics for Type 3 Analysis 64

Table 4.17:Analysis of Parameter Estimates 64

Table 4.18: Goodness of fit test 65

Table 4.19: Gamma regression analysis of automobile claims costs 66

Table 4.20: Adjusted coefficient for age group based on the frequency and cost models 67

Table 4.21: Adjusted coefficient for all rating factors based on claim frequency and cost models 69

List of Figures

Figure 2.1: Conceptual Model for Risk-Based Adjustment Pricing 13
Figure 2.2: Transition diagram of discount levels of NIA 35

Abbreviations

AIC	AKAIKE'S INFORMATION CRITERIA
BIC	BAYESIAN INFORMATION CRITERIA
BMS	BONUS MALUS SYSTEMS
EFInA	ENHANCINGFINANCIAL INNOVATION & ACCESS
FSAP	FINANCIAL SECTOR ASSESSMENT PROGRAM
GDP	GROSS DOMESTIC PRODUCT
GLM	GENERALIZED LINEAR MODEL
IMF	INTERNATIONAL MONETARY FUND
LB	LINEAR BAYES
LR	LIKELIHOOD RATIO
MSE	MEAN SQUARE ERROR
NAICOM	NATIONAL INSURANCE COMMISSION
NCD	NO CLAIM DISCOUNT
NIA	NIGERIAN INSURER'S ASSOCIATION

CHAPTER ONE

INTRODUCTION

1.1. Background to the Study

Generally, in today's competitive market environment, insurance companies are faced with situation of risk where decision must be made in the face of uncertainty (Boland, 2007). For example, in considering a new insurance product, one readily challenge will be the adjustment to be made to the price structure to enhance its profitability, yet at the same time maintain a reasonable degree of security and competitiveness. A characteristic aspect of insurance is that it is a product whose cost to the provider is unknown at inception of the policy; hence this makes it imperative to estimate future claim costs with much credibility and precision as the accuracy of these estimates are germane in determining the underwriting profits of non-life insurance companies (Mesike&Adeleke, 2015). In an insurance portfolio, the potential risks exposed by policyholders vary, particularly for automobile insurance.

One of the major tasks of the actuary is the design of a tariff structure that fairly distribute the burden of claims among policyholders as the insurers aim to sell coverage at prices that are sufficient enough to cover anticipated claims, administrative expenses, and an expected profit to compensate for the cost of capital necessary to support the sale of the coverage. If risks are not equal in an insurance scheme, it seems fair and perhaps essential to require insured parties to contribute premiums approximately in proportion to their relative risk, for example as the risk of a motor accident which often gives rise to an insurance claim varies from driver to driver (Boland, 2007).

The calculation of a differentiated premium within the motor insurance portfolio based on risk classification is essentially an important tool of insurance pricing in many countries where the insurance market is mature and highly competitive. As competition between insurance companies intensifies, higher efficiency and greater focus on profitability are required. While the potential for cost reductions is limited, improvements in profitability and growth can be achieved through sophisticated pricing management mechanisms (see, for example, Schmidt-Gallas&Lauszus, 2005, and Pratt, 2010). This strong competition therefore induces insurers to classify risks they underwrite to receive fair premium for the risk undertaken (Antonio, Frees & Valdez, 2010). This classification is based on known observable characteristics of the insured such as age, sex, engine capacity, etc. However, there are many other important factors that are unobservable (unobserved heterogeneity) by the insurer which cannot be taken into account a-priori when pricing motor liability insurance products. For example, aggressiveness behind the wheel, the swiftness of reflexes, and knowledge of the Highway Code or accident-proneness of a person are difficult to integrate into risk classification (Pitrebois, Denuit&Walhin, 2003). It is logical to believe that these hidden characteristics become apparent only by the number of claims reported after an accident or a series of an accident has taken place, hence the adjustment of individual premiums according to the accident history of the insured to restore fairness among policyholders.

In automobile insurance, among general insurance policies, it is a widespread practice to reduce the amount of premium by a factor in case the insured does not make any claim in a given period. Adjusting premiums with claims history is known as experience or merit rating, popularly referred to as No Claim Discount (NCD) or Bonus-Malus Systems (BMS). (For more extensive surveys on application of bonus-malus systems see, Loimaranta, 1972;Norberg, 1976; Vepsalainen, 1972; Lemaire, 1995, 1998; Dionne &Vanasse, 1992;

Lemaire&Zi, 1994;Pitrebois, Denuit&Walhin, 2005; Boland, 2007; Ibiwoye&Adeleke, 2011; and Mesike&Adeleke, 2016). Experience ratings are posteriori rating system commonly used in motor insurance as an attempt to categorize policyholders into relative homogeneous group who pay premium relative to their claims experience. It allows the matching of individual premium to risk and increases incentives for road safety by taking past record into consideration. They are justified by asymmetrical information between the policyholders and the insurance company as it encourage policyholders to drive carefully by reducing the inefficiencies associated with moral hazard and also respond to adverse selection in automobile insurance (Antonio *et al*, 2010).

Experience ratings were introduced in Europe in the early 1960s, following the seminal works of Delaporte (1965), Bichsel (1964), and Buhlmann (1964). The use of Markovian analysis on experience rating systems has been widely considered in several actuarial applications (see Hastings, 1976; Kolderman&Volgenant, 1985; Heras, Villar& Gil, 2002; Pitrebois*et al,* 2003; Aggoun&Benkherouf, 2006; Denuit, Xavier, Pitrebois&Walhin, 2007; Boland, 2007, Ibiwoye&Adeleke, 2011; Nath&Sinha, 2014; Chen & Li, 2014 and Mesike&Adeleke, 2016). Extensive studies have discussed the problem of how to design an optimal experience rating system. For example, optimal scales have been infered by Norberg (1976), Borgan, Hoem and Norberg (1981), and Gilde and Sundt (1999) while Centeno and Andrade (2002) deduced the optimal scales for bonus system that were not first order Markovian processes. Lemaire and Zi (1994) compared the validity of 30 bonus-malus systems using four different tools, such as the relative stationary average premium level, the coefficient of variation of the insured's premiums, the efficiency of the bonus-malus system, and the average optimal retention. Ibiwoye and Adeleke (2011) examined the no claim discount operation in Nigeria in a finite state Markovian framework, Mesike and Adeleke (2016) studied the desirability of a multi-layer premium system, where the state space consists of the different level of premium and the state of a particular insured shift randomly from one year to the next.

Most of these researchers considered the claim frequency as the most important factor and used the Bayesian estimation. The Bayesian estimator not only presents a rather irregular pattern but also may result unfairly without taking the severity of each claim into account. Frangos and Vrontos (2001) designed an optimal rating system that integrates both the frequency and the severity of the claim, Mahmoudvand and Hassani (2009) developed the system to a generalized form with a frequency and a severity component based both on the a priori and on the posteriori classification criteria.

Antonio and Valdez (2012) considered the difficulty of the phenomenon to be modelled and some methodological aspects related to the insurance data, in David (2015), and showed that Generalized Linear Models (GLMs) constitute an efficient tool for risk classification. GLMs allow modelling a non-linear behaviour and a non-Gaussian distribution of residuals which is very useful for the analysis of non-life insurance, where the claim frequency and costs follow an asymmetric density that is clearly non-Gaussian. In this regard, this study proposed a risk-based adjusted rate setting process that imposes costs in a fair and equitable manner using Generalized Linear Model in order to determine the premiums applied to each insured.

1.2. Statement of the Problem

Sustainable insurance pricing requires that the total amount of premium collected in aggregate covers the losses generated by all policyholders, the expenses of the insurer and provide an adequate rate of return to the insurance company. In establishing the burden of costs, the principle underlying the calculation of differentiated premium in the insurance portfolio is characterized by a pricing process that involves the classification of all risks regarding factors of influence where the actuary determines the impact of the observable factors on the insured risk and the correlations between data. However, there are other important information on the insured that cannot be seen by the insurer, which cannot be integrated into the premium calculation but may represent significant risk factors.

The potential risks exposed by policyholders vary, particularly for automobile insurance. Risk of accident which often give rise to claims is heterogeneous and not observable to the insurer, causing adverse selection (Boland, 2007; Pitrebois*et al*., 2005). The behaviour of the occupants varies widely with each member bringing into the class a different level of risk from that of the other members. This introduces a lot of heterogeneity and the premium assessment based on class membership in such cases cannot be fair or equitable (Lemaire, 1995). Also, within motor insurance classification, it has become extremely difficult recently, for insurance companies to maintain cross subsidies between different risks categories in a competitive market.

Although, actuarial literature presents an impressive list of works and trends in order to improve the pricing methods applied in insurance, there is a lack of such studies, if any, in Nigeria as research in this area does not have yet a well-defined structure. The insurance market is bedevilled by series of challenges, one of which is unprincipled underwriting, which has led to its present abysmal state and in some cases leading to insolvency of the insurance companies (Adeleke&Mesike, 2015). The bid to attract more customers and higher market share among rival companies has given rise to rate cutting and premium purchases, resulting in inability to meet obligation to insurance consumers when the need arises (Ibiwoye&Adeleke, 2011). Many companies with poor underwriting results rely on investment income to pay claims. This can lead to total collapse of the industry if the trend continues, particularly during adverse investment market conditions.

In some countries, the insurance industry shares responsibility for preventing road injuries, and organizations funded by the insurance industry (such as, the Insurance Institute for Highway Safety in the United States) make a valuable contribution to road safety (Gonulal, 2009). Motor insurance has the potential to be a powerful tool in the promotion of personal responsibility. If communicated effectively, the link between the consequences of causing an accident and the economics of paying for those consequences will of itself gradually lead to improved driving. Many more developed economies work extensively with risk-based pricing model, which has a dramatic effect on making the driver feel responsible for his or her driving (Gonulal, 2009).

There is a definite classification of hazardous, non-hazardous and extra hazardous risk in fire, marine, cargo and even life insurances, but for motor insurance, a mere reference to mechanical specifications of vehicle which can hardly be satisfactorily to elicit not-easily verifiable answers to couple of vague and general questions in the application for insurance concerning past accident and convictions by traffic regulator is not the optimal way out. Road accidents do not happen; they are caused, either directly or indirectly and wholly or partly by human error. It may be an innocent or deliberate violation of a traffic rule or plain incompetent driving.

In view of the influence of motor insurance in developing insurance markets, and especially the complexity it is of utmost importance to gain the trust of the motoring public by developing a rating system that is seen to be transparent, efficient, and equitably run. Such a system would be free of unfair market practices and promote the timely settlement of claims. The current study seeks to empirically provide criteria for analysis and development of a risk-based adjustment model that strikes a reasonable balance between fair premium and collective liability which imposes costs on insured in a fair and equitable manner.

1.3. Aim and Objectives of the Study

The general aim of this study is to develop a risk-based adjustment model for experience rating of the motor insurance sector in Nigeria in order to determine the premiums applied to each insured in an equitable and reasonable manner. The specific objectives of the study are to:

i. explore the relevant risk factors for motor insurance claims occurrence in Nigeria
ii. determine the risk factors that influence motor insurance costs in Nigeria
iii. evaluate the use of risk-based adjustment model in determining the future costs of motor insurance policies in Nigeria
iv. To establish the risk profile of policyholders for experience rating of motor insurance policy in Nigeria

1.4. Research Questions

i. What are the relevant risk factors for motor insurance claims occurrence in Nigeria?
ii. What are the risk factors that influence motor insurance cost in Nigeria?
iii. To what extent does risk-based adjustment model determine the future costs for motor insurance?
iv. What determines the risk profile of policyholders for experience rating of motor insurance policy in Nigeria?

1.5. Significance of the Study

This study will provide empirical and theoretical methods that would help insurers to formulate decisions that will ensure the effectiveness of tariffication process. The model will also help insurers to reduce and manage risks (cost) associated with moral hazard and adverse selection, and its introduction is expected to create more incentives for safe driving, as it links individual premiums to past reported accidents. This study would make useful contributions to policy formulation on the issue of insurance pricing and penetration. Such policies would enable the insurance companies to design appropriate pricing strategy and system that will be transparent, efficient, fair and competitive. It will also help in reducing the burden of road traffic accidents. To the regulatory authorities in the industries, the findings will provide guidance regarding the various approaches that may be adopted to help developing countries to increase insurance awareness, market deepening and insurance penetration and operational effectiveness of motor insurance and improve the overall social welfare in the economy.

1.6. Scope and Delimitation of the Study

The study covers the criteria for analysis and development of a characteristic-based risk adjustment model for effective computation of experience rating for the general insurance sector in Nigeria. The scope of this research study will, however, be limited to use of risk-based adjustment in determining the automobile insurance pure premium; hence does not cover all aspects of general insurance tariffication process in estimating the office premium. Also, the issue under investigation is limited by the problem of absence or shortage of fully organised data and poor data integration which makes it practically impossible to compute the bonus malus coefficient for all the insured in the portfolio.

1.7. Operational Definition of Terms

Accident: It is used to classify claim event (i.e, nature of loss) such as fire accident, hit a pole or hit a wall

Couple: This used to describe married policyholders whose gender classification are unknown

Collision: It is usedin this study to describe claim event (nature of loss) that involve a head on collision with another car.

District: It is used in this studyto describe the geo-political zone that the policyholder resides

Entity: This is used in this study to represent corporate organisation who has purchased a motor insurance cover

Experience rating: It is used to describe a posteriori pricing system where each risk is judged based on the claim experience of the motorist and the individual premium modified accordingly.

Insured: This is used in this study to describe an individual who has purchased a motor insurance policy or cover

Other account: This is used to describe customers (policyholders) other than individual, corporate organisation and government parastals or agencies such as NGOs, and co-operative groups

Policy: This is used in this study to represent motor insurance cover

Premium: This is used to describe the price paid by an individual for motor insurance cover

Privately employed: This is used in this study to represent policyholders who are employed in the private sector

Publicly employed: It is used in this study to represent policyholders who are employed in the public sector

Risk adjustment: is the process by which insured-level information is used in assigning relative risk factors to individuals or groups based on expected auto claim liability and by which those factors are taken into consideration and applied.

Tariff structure: This is used to describe the set of procedures used to determine how to charge different categories of motor insurance consumers

CHAPTER TWO

LITERATURE REVIEW

2.1 Preamble

This section presents the theoretical framework and the conceptual framework of the study as well as empirical review of the literature.

2.2 Theoretical Framework

There are many different relevant theories related to experience rating in non-life insurance pricing such as Credibility theory, Markov theory, Generalized linear model theory. This study however is based on generalized linear model theory.

2.2.1 Generalized Linear Model Theory

The theory and implementation merits of Generalized Linear Models (GLM) both in actuarial science and statistics was developed by Nelder and Wedderburn (1972) when they demonstrated the generalization of the existing theory of the classical normal linear model, by allowing deviation from its restrictive assumption of normality, and extending the Gaussian model to a particular family of distribution, namely the exponential family. A feature of this model is that it expresses the mean response as a function of linear combinations of explanatory variables. Given the distribution of the exponential family as:

$$f(y_i|\theta_i, \quad) = exp\left\{\frac{y_i\theta_i \quad a(\theta_i)}{} + c(y_i, \quad)\right\}, \qquad y_i \epsilon S \tag{2.1}$$

where$S$ represents a subassembly that belongs to or set, θ_i is the natural parameter and is the scale parameter. Traditionally, $\mu_i = E(y_i)$ is used for the mean response and $\eta_i =$

$x_i^t\beta$the systematic component of the model. The error structure allow writing a function (g) for the mean (μ_i) of the variable Y_i as a linear combination of the exogenous variables X_i;

$$g(\mu) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = x_i^t\beta = \eta_i (2.2)$$

The monotonous and differentiable function g is known as a link function because it connects the linear predictor η_i with the mean μ_i. Its inverse,$\mu_i = g^{-1}(x_i^t\beta)$ is known as the mean function. Risk premium modelling fits very naturally within the generalized linear model framework, especially when split into its constituent parts (i.e. frequency or average cost by claim type).Generalized linear models have become standard industry practice for non-life insurance pricing (David, 2015).

2.3. Conceptual framework

The conceptual framework for this study covers the specific empirical properties of the research on relationships between the risk factors for risk-based adjustment model of motor insurance claims. The relationship in insurance markets is appropriately described by the concept proposed by Akerlof (1970) and Rothschild and Stilglitz (1976) which refers to a situation in which the insured's private information relating to their overall risk level, although important to the insurer, cannot be introduced in the insurance premium calculation because they are not accessible to them when considering their underwriting decision. Thus, important related risks are not factored into the decision- making process. This implies that the drivers who purchase insurance cover are likely to be at greater risk of being involved in accident, thereby indicating a positive correlation between coverage and risk. These correlations must be regarded as constitutive elements in establishing a fair pricing structure as the main purpose of pricing is the accurate individual risk assessment, where insured drivers pay premium corresponding to the frequency and severity of the reported risks. By placing individual into risk categories and pooling risks within these categories, insurers adjust premium such that they reflect the average of the expected claim cost within a risk category. These concepts are applied to risk adjustment process, which is the process of transforming insured-level information into a factor indicating relative risk level. Statistical models were developed to evaluate the explanatory power of the risk-based adjustment system. The risk adjustment system was used to compute and develop risk score based on reported claim data, which were then related to the insured claim costs. This is illustrated in figure 2.1

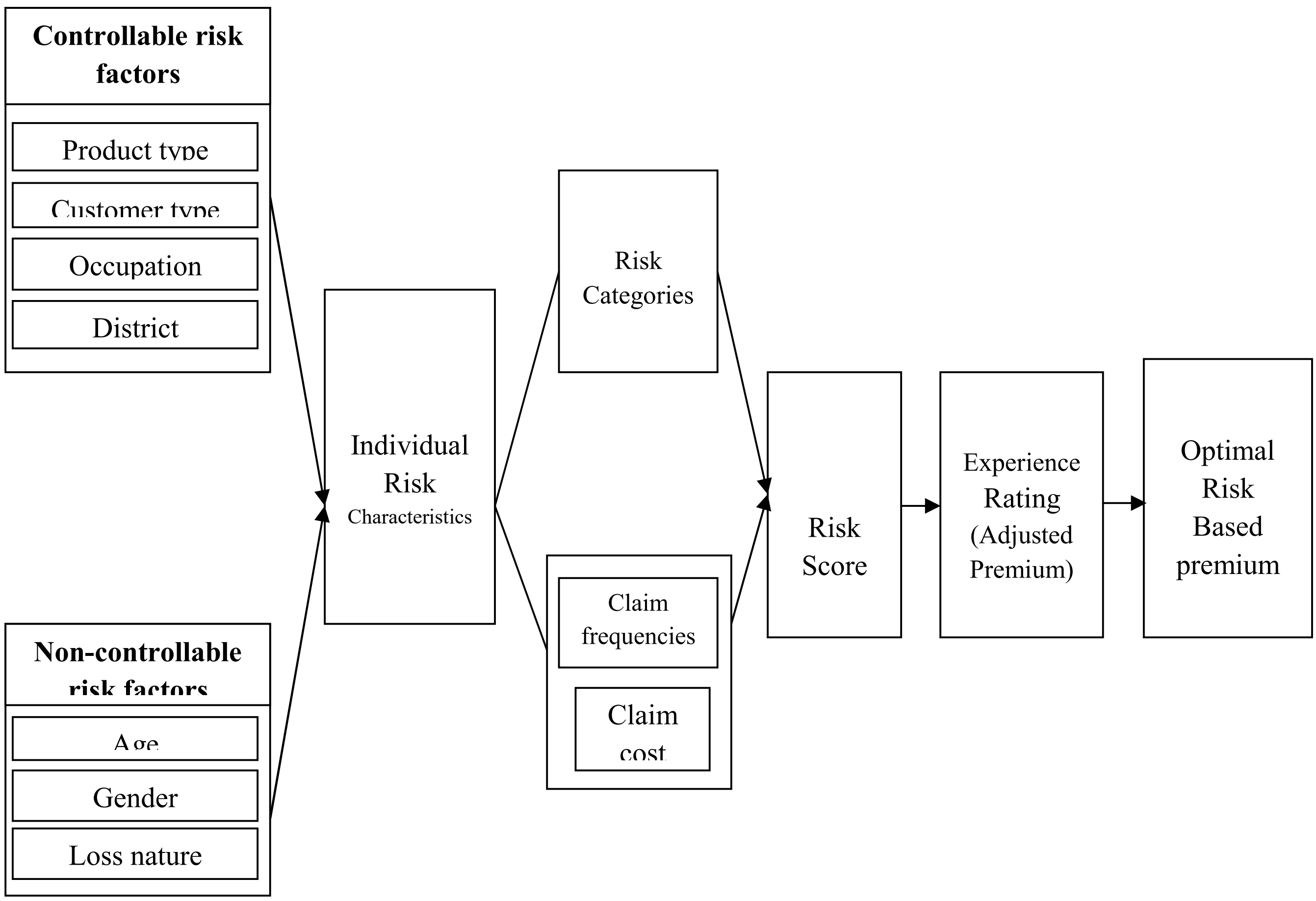

FIGURE 2.1: Conceptual Model for Risk-Based Adjustment Pricing

Source: Researcher's design

2.4 Empirical Review of Non-life Insurance Pricing

According to Denuit (2003), the pricing process within insurance business consists of the procedure for determininga fair premium corresponding to the insured's individual risk profile. The importance of pricing for non-life insurance arises in an attempt to challenge the anti-selection problem. The insurance portfolio is sub-divided into classes based on certain influencing risk factors where each class contain policyholders with identical risk profile who pays the same premium. A considerable body of literature exists about the theory of risk classification, especially its effects on adverse selection, its profitability, costs, fairness, and efficiency (see, e.g., Doherty, 1981, Hoy, 1982, Abraham, 1985, Crocker & Snow, 1986). David (2015) argued that the need for this differentiated tariff is highlighted by the insurance portfolio heterogeneity which leads to the concept of asymmetrical information. There exist two aspects of asymmetrical information presented in many relevant literatures, namely moral hazard and adverse selection (see for example Dionne, Michaud &Pinquet, 2012; David, 2015). The adverse selection according to Denuit*et al.* (2007), occurs when the policyholders take advantage of better knowledge of their claim behaviour information unknown to the insurer, while Chiappori, Jullien, Salanie and Salanie, (2006) emphasize the fact that when the probability of risk occurrence depends on the insured behaviour and his decisions, it gives rise to moral hazard. The difference between the two concepts was highlighted by Dionne, *et al*. (2012) who argued that adverse selection is the effect of unobserved differences among individual that affect the optimality of insurance transaction, while moral hazard is the effect of contracts on individual's unobserved behaviour.

In view of this, the actuarial literature presents two concepts of pricing (a priori and a posteriori pricing) with focus on finding adequate methods or tools for each of the types of pricing applied in non-life insurance. The main idea of a priori pricing as suggested by

Charpentier and Denuit (2004) is the partitioning of the insured risks into several categories so that each group contains equivalent risks. The a priori pricing divides policies into homogeneous classes, allowing all policyholders with identical risk profile paying the same premium. Extant literature has demonstrated that risk classes are still quite heterogeneous despite the use of a priori pricing due to some important unobservable factors that cannot be taken into account at this pricing stage (see for example, Pitrebois*et. al.*2005*,*Denuit*et. al.*, 2007 and Boland, 2007). This drawback necessitates the actuarial approach of a posteriori pricing where additional information about the individual claims history of the policyholders is considered. The a posteriori pricing is based on credibility theory originated by Mowbray (1914). The concept of credibility was linked to risk perception by Savage (1954), where individuals give different degrees of credibility to the occurrence of certain events. Whitney (1918) argue that the problem of assessing the experience arises from the need to strike a balance between collective experience (risk class) and individual experience (risk). According to Denuit (2006), the experience rating allow the adjustment of premiums for hidden individual risk factors by considering the past claim record, with the aim to assess the individual degree of risk in order to charge premium corresponding to the insured risk profile and claim history.

Traditionally, actuarial science has been limited to the use of Gaussian linear model in quantifying the impact of explanatory variable on the variable of interest, but the applicability of this model has been proven difficult as the linear modelling infers some set of assumptions that are not compatible with the reality imposed by the frequency and severity of damages generated by risks occurrence (see David, 2015). Although no mathematical model will describe completely the reality, David (2015) indicated that model analysis and the confrontation of theoretical properties of the studied occurrence with those observed is a pragmatic way in acquiring better understanding of reality and to predict the future responses of analysed events. One of the predominant methods developed to analyse approaches for the selection of classification criteria and calculation of the actuarial price in non-life insurance is the minimum bias procedure employed by Bailey and Simon (1960) for multiplicative tariff models. This consists of defining randomly the link between the explanatory variables, the risks levels and the distance between the predicted values and the observed ones. This approach was further developed by Bailey (1963), Jung (1968), and Ajne (1975), Ismail and Jemain, (2006) among others. Although the iterative algorithm method used was created outside a recognised statistical framework, but this approach has been found to be a particular case of GLMs (see, Bailey & Simon, 1960; and Bailey, 1963; Buhlmann (1967); Nelder&Verrall 1997; Mildenhall 1999; as well as Ohlsson, 2008) which has become a standard statistical industry practice for non-life insurance pricing. Another approach that attracted a great deal of attention is based on experience rating and credibility theory (see, Lass, Schmeiser& Wagner, 2016). The most famous credibility model was introduced by Buhlmann (1967), and Buhlmann and Straub (1970). This model, the parameters estimation, and its possible enhancements have been examined in a large number of subsequent research works, some of which includes Bichsel and Straub (1970), Sundt (1988), De Vylder and Goovaerts (1992), Dannenburg (1994), and Young and De Vylder (2000).

Comparing to the minimum bias techniques, the GLM models have the advantage of a theoretical framework that allows the usage of statistical tests in evaluating the fitting of models (Jong & Heller, 2008; David, 2015).As mentioned by Cameron and Trivedi (1998) in David and Jemna (2015), an important milestone in the development of models for count data is reached by the emergence of Generalized Linear Models (GLMs).The implementation merits of these Models was later developed by Nelder and Wedderburn (1972) who demonstrated that the generalization of the linear modelling allows the deviation from the assumption of normality, extending the Gaussian model to a particular family of distribution, called the exponential family. McCullagh (1976) offered detailed information on the iterative algorithm and the asymptotic properties of the parameter estimation of the model. Many studies in actuarial literature have emphasized the theoretical and practical aspects of the pricing methods in assessing the insurance premium (see for example, Jong& Heller, 2008; Kass*et al*, 2009; Frees, 2010, Antonio& Valdez, 2012).

2.5. Theoretical Review

There are many different experience rating systems, including bonusmalus systems, merit-demerit systems, participating policies and commissions in reinsurance, no claim discount (see, for example Buhlmann, 1967, 1969). The most widely used methods however are based on credibility theory and Markov theory.

2.5.1. Credibilty Theory

Credibility theory in general insurance is essentially a technique of experience rating that allows the use of data in hand, together with the experience of others in determining rates and premium (Boland, 2007). The advent of credibility theory as a technique for predicting future expected claims of a risk class; given past claims and related risk classes has a long history in actuarial literature, with elemental contributions dating back to Mowbray (1914). Whitney (1918) developed the first formal logical concept of using a weighted average for average claims from the risk class and overall risk classes to predict future expected claims to address the problem of assessing the risk premium *m*, defined as the expected claims expenses per unit of risk exposed, for an individual risk selected from a portfolio of similar risks (see, Norberg 2004 and Mesike&Adeleke 2015). The weight associated with the risk class under consideration is known as the credibility factor. The basic formula for calculating credibility weighted estimates is:

$$\bar{m} = z\hat{m} + (1 \quad z)\mu \qquad 0 \leq z \leq 1 \qquad (2.3)$$

where $\hat{m}$ is the observed mean claim amount per unit of risk exposed for the individual contract, μ is the corresponding overall mean in the insurance portfolio. The weight z is called the credibility factor since it measures the amount of credence attached to the individual experience, and $\bar{m}$ was called the credibility premium (see, Longley-Cook, 1962; Miller & Hickman, 1975; Boland, 2007; Klugman, Panjer & Willmot, 2008; Adeleke & Mesike, 2012).

Credibility theory uses two main approaches, each representing a different method of incorporating individual experience in the ratemaking process (Goulet 1998; Norberg, 2004). The first approach is called limited fluctuation credibility where an insured's premium is based solely on its own experience provided the experience is significant and stable enough to be considered credible. The second approach, called the greatest accuracy credibility does not concentrate on the stability of the experience but rather focuses on the homogeneity of the experience within the portfolio.

2.5.1.1 Limited fluctuation credibility

The limited fluctuation credibility also known as the frequentist approach was originated by Mowbray (1914) when he suggested determining the amount of individual risk exposure needed for $\hat{m}$ to be a fully reliable estimate of m. Using an annual claim amounts $X_i, \dots., X_n$, assumed to be independently and identically distributed with a probability density function $f((x|\theta))$, mean $m(\theta)$ and variance $s^2(\theta)$ and taking $\hat{m}=\frac{1}{n}\sum_{t=1}^{n} Xj$, he sought to determine how many years n of observation are needed to make $P_{\theta}[|\hat{m} \quad m(\theta)| \leq km(\theta)] \geq 1 \quad \epsilon$ for some given k and ϵ. The parameter θ was viewed as non-random. Using the normal approximation $\hat{m} \sim N(m(\theta), \frac{s(\theta)}{\sqrt{n}})$, the criterion $km(\theta) \geq z_{i-\epsilon}/2 \frac{s(\theta)}{\sqrt{n}}$, was inferred where, z_{i}-$\epsilon/2$is the fractile in the standard normal distribution (Norberg, 2004). Ceiling in the empirical estimates $\hat{m}$ and $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i \quad \hat{m})^2$ for the unknown parameters, he arrived at

$$n \geq \frac{Z_{1-\epsilon}^{2}/2^{\hat{s}2}}{k^2 \hat{m}^2} \tag{2.4}$$

This solution paved way for the issue of partial credibility on how to choose z when n does not satisfy the above equation. Whitney (1918) develop the first partial credibility formula based on the homogeneity of the portfolio with the assumption that the individual averages are distributed according to the normal distribution and obtain an expression for the credibility factor of the form

$$z = \frac{n}{n+k} \tag{2.5}$$

where$k$ is a constant which is an explicit expression that depends on the various parameters of the model. The determination of k was suggested to be determined by the actuary's judgement rather than by its correct mathematical formula and thus has no open unifying principle for significant generalizations. Therefore, the limited fluctuation approach according to Norberg (2004), despite its grand scale, does not really constitute a theory in the usual sense.

2.5.1..2 Greatest accuracy credibility

The greatest accuracy credibility theory was developed following the works of Bailey (1945, &1950). The experience rating problem is seen as a matter of estimating the random variable m(Θ) with some function $m(X)$ of the individual data X, with the objective to minimize the mean square error (MSE)

$$\rho(\breve{m}) = E[m(\Theta) \quad \breve{m}(X)]^2 \tag{2.6}$$

The calculation of the above equation shows that the optimal estimator is the conditional mean

$$\widetilde{m}(X) = E[m|\mathrm{X}], \tag{2.7}$$

and its MSE is

$\tilde{\rho} = E\,Var[m(\Theta)|X] = Var\,m \quad Var\widetilde{m}$ (see, Norberg, 2004).

Buhlmann (1967, 1969) set out clearly the programme of the theory when he emphasized that the optimal estimator and its mean square error depend only on the first and second moments that are usually easy to estimate from statistical data. Considering a non-parametric model conditional onΘ, the annual claim amounts $X_1, \ldots., X_n$which are independent and identically distributed with mean m(Θ) and variance s^2(Θ).

$$\widehat{m} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_J \tag{2.8}$$

which is the best linear unbiased estimator (BLUE) of $m(\theta)$ in the conditional model, givenΘ $= \theta$, he arrived at the credibility formula:

$$\overline{m} = z\widehat{m} + (1 \quad z)\mu$$

with

$$\mu = E(m(\quad)) = E(X_j), \tag{2.9}$$

$$z = \frac{\lambda n}{\lambda n +} \tag{2.10}$$

where $\lambda = var[m(\ \ )]\ and \quad = E(s^2(\ \ ))$

The credibility factor z increases and tends to 1 as sample size n increases. It increases with λ and decreases with , which means that the larger the process variance of the observations between the different risk parameters the lesser the weight put on the sample mean. The Linear Bayes (LB) risk is

$$\bar{\rho} = var\ m \quad \frac{cov^2[m,\hat{m}]}{var\ \hat{m}} \tag{2.11}$$

This measures the accuracy of a LB estimator which is

$$\bar{m}=\mathrm{E}(m(\ \ )) + \frac{cov[m,\hat{m}]}{var\hat{m}}(\hat{m}\text{-}E(\hat{m})) \tag{2.12}$$

and becomes

$$\bar{\rho} = \frac{\lambda}{\lambda n+} = (1 \quad z)\lambda \tag{2.13}$$

This approach which is sometimes called the least squares approach to credibility is an empirical Bayes approach (see Boland, 2007). It reflects the similarity to Bayesian estimation using squared error loss, but here the prior distribution is unobservable hence, full credibility is never achieved. Therefore, this approach to credibility has limited effectiveness, because the assumptions about the distributions are rarely met in practice (Behan, 2009).

2.5.2 Markov chain theory

Markovian theory came into existence following the work of Markov (1913) when he extended the theory of probability in a new direction to chains of linked events (where what happens next depends on the current state of the system). A Markov chain is a discrete-time stochastic process $X_1, X_2 \ldots\ldots$taking values in an arbitrary state space that has the Markov property and stationary transition probabilities; where the conditional distribution of $X_n$given $X_1, \ldots, X_{n-1}$ is the same as the conditional distribution of $X_n$given X_{n-1}only, and the conditional distribution of X_n given X_{n-1}does not depend on n. The conditional distribution of X_n given X_{n-1} specifies the transition probabilities of the chain.

A stochastic process $\{X_n\}$ is a Markov chain if for all times $n \geq 0$ andall states $i_0, \ldots., i, j \in S$

$$P(X_{n+1} = J \mid \mathrm{X_n} = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)$$

$$= P(X_{n+1} = J \mid \mathrm{X_n} = i)$$

$$= P_{ij} \qquad (2.14)$$

P_{ij}denotes the probability that the chain, whenever in state i, moves next (one unit of time later) into state j, and is referred to as a one-step transition probability. The square matrix $\boldsymbol{P} = (P_{ij}), i, j \in S$, is called the one-step transition matrix, and since when leaving state i the chain must move to one of the states $j \in S$, each row sums to one. For each $i \in S$ $\sum_{j \in S} P_{ij} = 1$. The n step which is the probability that in n time the chain will be in state j given that it is now in state i is denoted by:

$$P^n = \left(P_{ij}^n\right), \qquad n \geq 1 \qquad (2.15)$$

where$P_{ij}^n = P(X_{m+n} = J \mid \mathrm{X_m} = i)$

These n step probabilities can be computed by the Chapman-Kolmogorov equation:

$$P_{i,j}^{n+m} = \textstyle\sum_{k \in S} P_{i,k}^n P_{k,j}^m \text{for any } n, m \geq 0, i \in S, j \in S, \qquad (2.16)$$

Markovian analysis has been the basis of the works on experience rating, which assumes that the NCD forms a Markov chain which is a stochastic process in which the future development depends only on the present state and not the history of the process or the manner in which the present state was reached. For a given Markov chain NCD model, irrespective of the initial distribution $\boldsymbol{P}^0$ there is a stationary distribution$\pi = (\pi_0, \pi_1, \ldots., \pi_k)$ to which $\boldsymbol{P}^0$ converges as n becomes large, that is, there exist the limiting probabilities $\pi_j = lim_{n\to\infty} P_j^n$ for all j (see Boland, 2007). The stationary probability vector is unique and satisfies

$$\pi = \lim_{n\to\infty} P^n = \lim_{n\to\infty} P^{n+1}$$

$$= \lim_{n\to\infty} P^n . P$$

$$= \pi . P$$

2.6 Claim Counts Model

This section considers count data models where the number of loss events occurs in unit time, that is, an event where the response variable is a count. In general insurance, for example, the count variable of interest could be the number of a claim made on a motor vehicle insurance policies or the number of losses to the insurer/insured in a year. These count variables of losses represent individual risks, and need to be predicted, paticularly when the pure premium is to be computed for new policyholders, or when future premiums are adjusted based on past experience. A well-known method in determining the basic elements of the pure premium is multiplying the conditional expectation of the claim frequency with that of the expected cost of claims. Thus, modelling count data represents an essential step of non-life insurance pricing, as noted in Boucher and Guillen (2009), that count regression analysis permits the identification of the risk factors and the prediction of the expected frequency of claims given the risk characteristics. In actuarial literature over the years, there has been considerable interest in count data models (see for example, Nelder&Wedderburn, 1972; Gourieroux, Monfort&Trognon, 1984a, 1984b; Hausman, Hall &Griliches, 1984; McCullagh&Nelder,

1989; Dionne &Vanasse, 1989, 1992; Gourieroux&Jasiak, 2004; Jong & Heller, 2008; Antonio& Valdez,2012; David, 2015; David &Jemna, 2015).

2.6.1. Poisson Regression

Cameron and Trivedi (1998) demonstrated the particularities of Poisson regression approach in modelling claim frequency as a particular case of GLMs. With Poisson regression, the mean μ is explained in terms of explanatory variables x via an appropriate link, If $y \sim P(\mu)$

$$f(y) = \mu^{y} \frac{e^{-\mu}}{y!} \quad , \qquad y = 0,1,2, \dots, \tag{2.9}$$

Within the framework of GLMs, the mean of the response variable is related to the linear predictor through the log link function:

$$\mathrm{g}(\mu) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = x_i' \beta \tag{2.10}$$

The estimation of the parameters is done by maximum likelihood and the resulting equation forming the system solved numerically by using iterative algorithm such as Newton-Raphson or Fisher information (see, Charpentier&Denuit, 2005). Though Poisson distribution is often considered as a benchmark model in modelling claim count but in practice there are some idiosyncratic risks related to individual insurance contract that make the underlying assumption of the model seem quite unrealistic (see, Gourrieroux&Jasiak, 2007; Jong &Heller, 2008; David &Jemna, 2015).

2.6.2. Mixed Poisson Model

The Poisson distribution is often suggested for count data but found to be inadequate because the data displays far greater variance than that predicted by the Poisson. This phenomenon is known as overdispersionor extra-Poisson variation which may be modelled using compound Poisson distributions. The weakness of the Poisson distribution in accommodating heavy tails was recognized in the early twentieth century, when Greenwood and Yule (1920) postulated a heterogeneity model for the overdispersion, in the context of disease and accident frequencies. This is thefirst appearance of the negative binomial as a compound Poisson distribution,as opposed to its derivation as the distribution of the number of failures till the *r*th success. Newbold (1927) and Arbous and Kerrich (1951) illustrated compound Poisson distributions in the context of modelling industrial accidents while Lundberg (1940) further considered the negative binomial as a compound Poisson distribution, as a result of heterogeneity of risk over either time or individuals, as a model for claim frequencies. With this model the count $y$is Poisson distributed with meanλ, and the mean λ itself a positive continuous random variable with probability function $g(\lambda)$. Given λ, the count is distributed as P (λ). Then the probability function of y is:

$$f(y) = \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!}\, g(\lambda)d\lambda \tag{2.11}$$

Within the actuarial literature, a suitable choice for the mixing distribution $g(\lambda)$ is the gamma probability function $G(\mu, \nu)$, implying (2.11) is $NB(\mu, \kappa)$ where $\kappa = 1/\nu$. There are alternative choices to the gamma for the mixing distribution $g(\lambda)$.Two which have appeared in the actuarial literature are the generalized inverseGaussian and inverse Gaussian distributions (see, Jong &Heller, 2008). The generalized inverse Gaussian is a three-parameter distribution which is highly flexible, but has the drawback that its computation is complex. Its two-parameter version, the inverse Gaussian, is computationally somewhat simpler. Willmot (1987) compared their performance in fitting claim frequency distributions, and found that the Poisson-inverse Gaussian was more successful in accommodating heavy tails than the negative binomial. However, this difference appears only to be a marginal improvement and the benefit of the Poisson-inverse Gaussian over the negative binomial was disputed by Lemaire (1991). In recent years the negative binomial has been widely used as the distribution of choice when modelling overdispersed count data in many fields, possibly because of its appealing properties and availability in standard softwares.

2.7 Claim Severity Model

We consider here continuous responses of interest to insurers which include claim size and time between the reporting of a claim and settlement. Continuous insurance variables are usually non-negative and skewed to the right. Generalized linear modelling can be used to model these variables using a response distribution that is concentrated on the non-negative axis such as the gamma and inverse Gaussian distributions (see, Jong& Heller, 2008). Traditionally, most experience rating modelling takes only the numbers of claims into account under the assumption that the number of accidents per year is independent of its severity, but it is closer to reality to incorporate the claim severity into the risk measure. Picard (1976) first proposed a model to distinguish claims that cause only property damage from those that caused both bodily injury and property damage by generalizing the negative binomial model to account for the subdivision of claims into small and large losses. Pinquet (1997) incorporated the severity of claims by including the rating factors and two heterogeneity components in the scale parameter under the assumption that the costs of claims follow gamma or lognormal model. Frangos and Vrontos (2001) model the cost of claims according to the Pareto distribution while Jong and Heller (2008) illustrated the modelling of claim cost using the gamma and inverse Gaussian model.

2.8 Brief Historical Background of Insurance in Nigeria

Prior to the introduction of modern insurance, there were some forms of traditional social insurance and mutual schemes that existed in Nigeria, which evolved through the African communal channels like the extended family system, age grades, and clan unions African cultures (Obasi, 2010). According to Adeyemi (2005), the origin of modern insurance can be traced to the advent of British trading companies in the West African region which culminated into increased inter-regional trade that compel some of the foreign firms to handle some of their risks locally. This increased trade commerce led to the trading companies being granted insurance agency licenses by foreign insurance companies, with the first of such agency in Nigeria created in 1918 when the Africa and East trade companies established the Royal Exchange Assurance Agency (Jegede, 2005). It was not until 1958 that the first indigenous insurance company, the African Insurance Company Limited, was established.

In response to the dominance of non-indigenous insurance companies witnessed in the country's post-independence era, where out of the twenty five firms in existence at independence, only four were indigenous, the Obadan Commission of 1961 which gave rise to the establishment of Insurance Companies Act of 1961 was set up and there was upsurge in the number of the indigenous insurance companies by 1976 (see, Ujunwa&Modebe, 2011; Oke, 2012). Of the 70 insurance industries in 1976, fourteen were foreign owned, ten were wholly owned while forty six were indigenously owned. The introduction of Structural Adjustment Programme, led to a remarkable increase in the number of insurance companies in Nigeria, with the number increasing to 110 in 1990. The financial system reforms of 2004, led to a dramatic change in the insurance industry and as at September 2005, there were one hundred and four insurance companies and four reinsurance companies in existence before recapitalization (see, Oke, 2012).

In the past two decades, regulation of Nigeria insurance industry has become considerably intensified according to Ezekiel (2005) due to the presence of risks of potential abuse, poor market penetration, low level awareness, low operating capital, as well as low capacity for retention and writing of foreign risks, all of which led to massive regulation of the insurance sector of Nigeria financial system. The first major attempt at regulating insurance in the country was the promulgation of the Nigerian Insurance Decree, 1976, with the biggest development in the industry being the establishment of the National Insurance Commission in 1997. Nigeria undertook an initial Financial Sector Assessment Program (FSAP) in December 2001, which included a review of the structure of Nigeria's insurance market and the supervisory framework and approach (IMF, 2013). Nigeria has also undertaken reviews of its observance of international accounting and auditing standards (2004 and 2011), and corporate governance (2008).

The first major recapitalization process was introduced by the insurance Act 2003, followed by the 2005 recapitalization which changed the landscape of insurance companies operating in Nigeria by compelling many insurers to merge in compliance with the new capital base directive of National Insurance Commission (NAICOM) (see Oke, 2012). This is to ensure the matching their capital according to the risks they underwrite to allow insurers concentrate on businesses they have core competence. Following the recapitalization of insurers and reinsurers in 2007, NAICOM introduced initiatives that will considerably improve the regulatory environments, including a voluntary code on corporate governance, operational guidelines, risk management framework, and the adoption of international financial reporting standards (IMF, 2013)

2.8.1 Nigeria Insurance Market Industry

The Nigerian insurance market, like any other market, is intricately linked to the socio-economic, demographic and macroeconomic context within which it operates (Vos, Hougaard, & Smith, 2011). The Oxford Business Group (2010) report noted that Nigeria is the most populous nation in Africa, and the eighth-most populous country in the world with an estimated growth rate of about 3.2% per year, yet the current insurance usage according to the Enhancing Financial Innovation & Access (EFInA) survey in 2010, is extremely low as the insurance sector serves less than 1% of the adult population. Currently, the insurance sector contributes a mere 0.72% to GDP, much lower than the African average of 3.3% and the global average of 7% (Swiss Re, 2010). The insurance sector is an underdeveloped part of the Nigerian financial sector with less than 2% of GDP in assets and assets of the life business are about half of the assets of the non-life sector reflecting a low level of savings and investment insurance products (IMF, 2013).

In terms of gross written premium according to the international monetary fund report (2013), the total sector grew at an average rate of 23% from 2001 to 2010 but remains very small with a total premium income of 192 billion, representing 0.7% of GDP in 2010 and the gross written premium is estimated to be 232 billion in 2011. The non-life insurance sector, which is about three times the size of the life insurance sector, dominates with only seven specialised life insurers, compared to 22 non-life and 20 composite underwriters. Non-life insurance accounts for 84% of premiums (of which motor insurance has been the dominant source of premiums for more than five years). Nevertheless, according to the Nigeria Insurers' Association, the insurance industry is quite profitable, with a sustained average profit of around 25% which is driven by low claims (underwriting losses) relative to premiums (see, Vos, *et al*., 2011). One reason often cited for the low claims ratio is the onerous administrative process. Management/administrative and marketing expenses are disproportionately high and exceed claims ratios – contrary to international best practice. This leads to low consumer value, which in turn undermines trust in the industry.

2.9. Motor Insurance Policy in Nigeria

Motor vehicles first appeared on the roads during the 1880s and the first motor insurance policies were issued during the 1890s (Ellis, 1983). Generally, it is noteworthy to mention that the early years of the twentieth century saw the formation of insurance companies in which the main emphasis was upon motor insurance and thus, the motor tariff came into operation within the framework of the Accident Offices' Association (Ajemunigbohun&Oreshile, 2014). Nigeria has witnessed a magnificent growth in the past two decades with appreciable levels of urbanization and the law evolving with it. In Nigeria, motor insurance is normally offered in two categories namely comprehensive and third party as it is mandatory to own car insurance before driving your vehicle. However, some companies offer a sort of extension called the third party fire & theft. The first kind of insurance, called Comprehensive is sort of a master service because it includes the third party cover and protect against damage. Most of the insurance companies offer cover against accidental collision, fire, explosion, theft and malicious acts, with an option to buy additional protection against riots, damage against floods, liability to passengers and expenses incurred if you happen to damage your vehicle.

The second kind, which is called the third party policy protects from death or injury from the use of the car and the obvious damage to a third party´s property. As for the third one, the title practically describes it but, the services offered may vary from one company to another. However, the Nigeria's motor tariff prescribes the standard format for underwriting motor insurance and general regulations applicable to all types of motor vehicle including those belonging to or held in trust by motor trade. According to Akintayo (2004), some of the general regulations are: value of vehicles; period of insurance; short period rates; cancellation of policies; No claim discount; joint insureds/policies; vehicles paidup; and vehicles hire under contract for not less than twelve months and not being a hire purchase contract. Ngwuta (2007) thus posit that motor insurance is usually grouped according to the usage of vehicles, i.e. private cars; commercial vehicles; passenger carrying vehicles; goods carrying vehicles; public authorities vehicles; agricultural and forestry vehicles; and mechanical plants of special design.

As rightly noted by Ozioko (2007), a market where pricing is tariff-driven without sufficient proof or statistics to back up the adequacy of charges is bound to suffer the fate of our motor insurance pricing. An important attributes of insurance operation is evidenced by the needs to make some basic assumptions concerning the expected cost of assuming a risk by the insurer when pricing such risk or group of risks.This infers that some degree of uncertainty is involved in the cost of insurance operation. According to Asokere and Nwankwo (2010), the workability of insurance pricing is hinged upon certain factors such as adequacy, reasonableness, equity, technical profitability and induced loss prevention. Trieschmann, Hoyt and Sommer (2005) described insurance premium as the total cost of insurance, determined by multiplying the rate by the number of units covered.

According to the Oxford Business Group (2010), motor insurance has accounted for the majority of premiums yet, motor insurance usage is still only a fraction of total motor ownership. This presents a significant untapped opportunity, not only for better enforcement of compulsory third party vehicle insurance, but also for comprehensive auto insurance. Of concern is the high incidence of fake compulsory insurance, such as third party motor vehicle insurance (Vos, Hougaard, & Smith, 2011), as these products are sold by companies that have not officially registered as insurance companies and therefore will not make any insurance pay-out. Earlier research such as World Bank (2008) estimated that 60-80% of all motor vehicle insurance policies were provided in this way and recent industry conversations suggest that the practice is still rife (see, Vos, *et al*., 2011). The motor insurance business in Nigeria is forecasted to grow 7% in the following three years with an estimate of over 40 million vehicles to be roaming the streets by 2020, and the challenges will grow accordingly, which includes a bigger possibility of accidents occurring (NIA, 2013). Therefore, motor insurance in Nigeria needs to arise to the challenge and do not underestimate these changes if they want to remain successful and gain the trust of those that are not yet convinced.

2.10 Experience Rating System

Experience ratings were introduced in Europe in the early 1960s, following the seminal works of Delaporte (1965), Bichsel (1964), and Buhlmann (1964). Many studies have discussed the problem of how to design an optimal experience rating system. For example, formulas for some asymptotic properties of bonus systems were developed by Loimaranta (1972), where bonus systems are understood as Markov chains. Vepsäläinen (1972) used this method to study the bonus systems used in Denmark, Norway, Sweden, Finland, Switzerland and Germany. Lemaire (1976) derived an algorithm for obtaining the optimal strategy for a policy holder. This algorithm was applied to compare the bonus systems used in Denmark, Norway, Sweden, Finland, Switzerland and West Germany. Under the assumption that the frequency of claims is Poisson and the severity of damage is negative exponentially distributed, Hastings (1976) formulated a simple model as a Markov decision problem which was solved by dynamic programming. Lemaire (1979) computed a merit-rating system for motor third party liability insurance. The results are applied to the portfolio of a Belgian company and compared to the premium system provided by the expected value principle.

The use of Markovian analysis on BMS has been widely considered in several actuarial applications (see Kolderman&Volgenant, 1985; Heras, Villar& Gil, 2002; Pitrebois*et al*, 2003; Aggoun&Benkherouf, 2006; Denuit, Xavier, Pitrebois&Walhin, 2007; Boland, 2007, Ibiwoye&Adeleke, 2011; Chen & Li, 2014 and Mesike&Adeleke, 2016). Optimal scale scales have been infered by Norberg (1976), Borgan, Hoem and Norberg (1981), Gilde and Sundt (1989) while Centeno and Andrade (2002) deduced the optimal scales for bonus system that were not first order Markovian processes. The analysis on experience rating mechanism for motor insurance was carried out by Lemaire (1988) when he compare the bonus-malus system of 13 European countries using three metrics: the relative stationary average premium level, the efficiency of the bonus-malus systems, and the average optimal retention. Based on this analysis, five guidelines were noted for the construction of a good bonus-malus system.

Lemaire and Zi (1994) analysed 30 BMS from all over the world and concluded that the design of a BMS is influenced by economic development and culture. The toughness of 16 Asian BMS towards consumers and correlation with cultural and economic variables were evaluated by Park, Lemaire and Chua (2009) using principal component analysis and regression analysis. The study found that Common Law legal system and cultural variables such as uncertainty avoidance influences BMS. In a Markovian study of the no claim discount in India using the India regulatory and development authority, Nath and Sinha (2014) found that the probability of claims and difference NCD rates are not parallel.

2.10.1 The No Claim Discount System in Nigeria

There are many different experience rating systems, including bonus malus systems, merit-demerit systems, participating policies and commissions in reinsurance, no claim discount (see, Buhlmann, 1967, 1969).There are wide variants of it in place in different countries of the world, from total freedom to government-imposed systems, with many intermediate situations (Lemaire, 1998; Boland, 2007). Some are known to be soft while others are referred to as severe depending on the transition rules applied. In Nigeria, insurance companies appear to follow an experience rating system of tariffication imposed by the Nigeria Insurance Association (NIA) known as the No Claim Discount system (NIA, 2006). An insured enters the system, in the initial class, when he or she obtains a driving license. Then, throughout the entire driving lifetime, the transition rules are applied upon each renewal to determine the new class as a function of claims history. This definition assumes that the NCD forms a Markov chain which is a stochastic process in which the future development depends only on the present state and not the history of the process or the manner in which the present state was reached.

NIA Transition Rule of NCD System

Table 2.1: Level of NCD System of NIA

Starting Level	No Claim Discount Saving
0	0%

1	20%
2	25%
3	33.5%
4	40%
5	50%

Source: NIA

The six level of discount of NIA (for private cars) are $0\%, 20\%, 25\%, 33\ 1/2\ \%, 40\%, 50\%$. At the end of each policy year, policyholders change levels according to the following rules:

i. A policyholder who has made no claim(s) during a policy year moves to the next level higher discount level or remain at 50% if already at the highest level.

ii. A policyholder who has made at least one claim during a policy year the period of classification for discount commences de novo as from the next renewal date. That is whatever the class a policyholder is when making a claim, he loses all the bonus and starts at the 0% discount level.

The rule of the NCD system described above can be summarized in a transition matrix showing the probabilities of movements among each level, see figure 2.1, for the general notation, where $P_0$is the probability of no claim and $(1 \quad P_0)$ is the probability of at least one claim. Here, $P = (P_i)_{5\times5}, P_{i0} = (1 \quad P_0),\ i = 0,1,\dots,5; = P_0,\ i = 0,\dots.,4\ \ andP_{55} = P_0$.

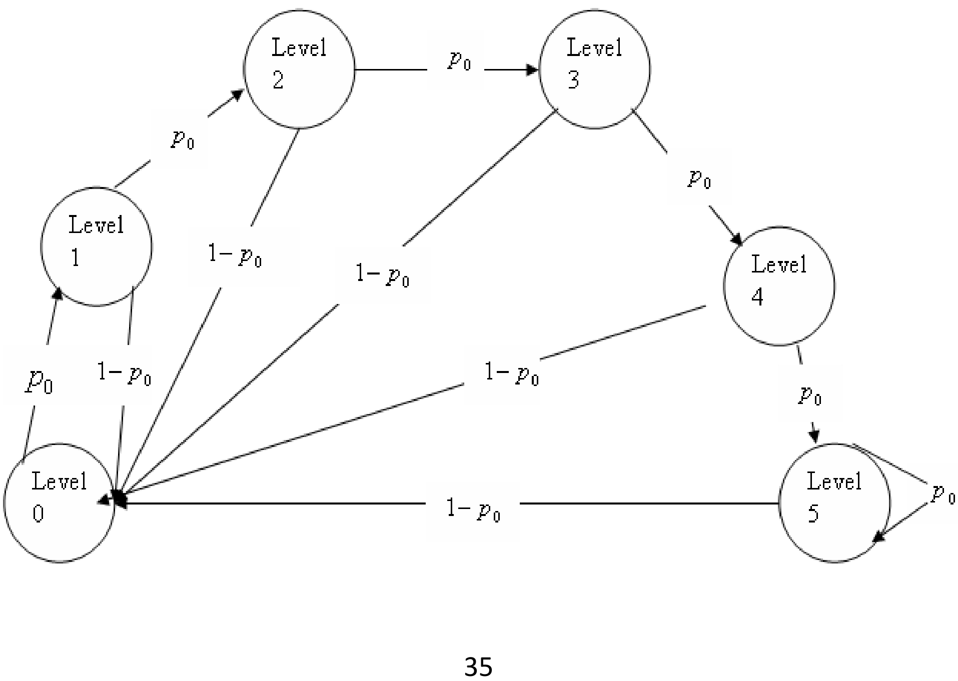

Figure 2.2: Transition diagram of discount levels of NIA

CHAPTER THREE

METHODOLOGY

3.1 Preamble

This section includes the presentation of the data used, the procedures for gathering and processing the data, based on which a numerical illustration of the statistical techniques is performed in the conduct of the research. It specifies the research design, population of the study, process of data collection, and sampling design.

3. 2 Research Design

This study adopted the exploratory and cross-sectional descriptive research design. The design was selected based upon existing differences in the sample population information (premium and claim amount) and the capability of the research design of using data from a large number of subjects (policyholders). The main purpose of such design according to Kothari (2004) is formulating a problem for more precise investigation and developing the working hypotheses from an operational point of view. The major emphasis is on the discovery of ideas and insights, which in this case, trying to understand how various characteristics of the insured can help develop appropriate rate that is proportionate to the risk they bring into the pool. As such the research design appropriate for this kind of study must be flexible enough to provide opportunity for considering different aspects of a problem under study. Inbuilt flexibility in research design is needed because the research problem, broadly defined initially, is transformed into one with more precise meaning in exploratory studies, which fact may necessitate changes in the research procedure for gathering relevant data (Burns & Grove, 1993).

3.3 Population of the Study

The population of the study comprises all the insured in motor liability portfolios of motor insurance service providers in Nigeria business environment, for which the insurance is covering the losses within the limits of the insured amount. There are 41 insurance companies licensed to issue motor insurance cover by the national insurance commission (NAICOM, 2015).

The data were sourced from the registered policies through the Nigerian insurance industry database (NIID). All registered insurance companies operating in Nigeria subscribe to the NIID and regularly upload the details of vehicle covers issued.The database consists of 2.7 million registered policies (www.niid.org). However, for the avoidance of practical difficulties owing to the absence of unified collection of data by motor insurance service providers used in the underwriting process for pricing, some variables required for this study were not available. Hence, the data were profiled and screened. Then, only the usable data from the policies that have adequate and sufficient information which are presented in the format suitable for the analysis in addressing the study objective was used for the purpose of this study. Finally, a total number of 15,979 registered policies of motor insurance liability portfolios were found useful for the purpose of this study.

3.4 Type and Sources of Data Collection

Secondary data were collected for the purpose of this study. The data used were extracted from the registered policies of motor insurance portfolio observed during the year 2015. The elements included in the policies are the factors considered in this study. The covariates used are considered risk factors, known a priori by the insurer which reflects the insured characteristics: policyholder's age (four classes: <24 years, 24-30 years, 31-60 years and > 60 years), gender (male, female, entity, joint gender), occupation (self-employed, publicly-employed, privately-employed, unemployed), the geo-political zone where the policyholder lives (federal capital territory, south-west, south-east, south-south, north-west, north-east, north-central), product type (commercial vehicle, comprehensive, third party, motorcycle), customer type (individual, companies, government, others account), nature of loss (theft, collision, accident, vandalisation, others)

3.5 Method of Data Analysis

The variables entered are taken into consideration as risk factors and the models fitted using the Statistical Package for Social Sciences (SPSS 20) software by means of GENLIN procedure which enables the use of type 3 analysis that allows the impact assessment of each risk factor, considering all other explanatory variables. The type 3 analysis provides the values of Chi-square statistics for each variable by calculating two times the difference between the log-likelihood of the model which includes all the independent variables and the log-likelihood of the model obtained by deleting one of the specified variables. This test statistic value the impact of each risk factor on the studied interest and follow the asymptotic χ^2 distribution with p degrees of freedom, representing the number of parameters related to the analysed variable.

3.6 Generalized Linear Model

This study used the generalized linear model (GLM) in developing the risk-adjusted model. A feature of this model is that the GLM provides methods for the modelling of non-linear behaviour and non-Gaussian distribution of residuals which are very important and useful for the analysis of non-life insurance data, where claim frequencies, claim costs and the occurrence of a claim on a single policy are all outcomes that follows an asymmetric density that is clearly non-Gaussian (see, Jong & Heller, 2008; David, 2015). It expresses the mean response as a function of linear combinations of explanatory variables. Generalized linear modelling is used to assess and quantify the relationship between a response variable and explanatory variables. The purpose is to estimate an interest variable (Y) depending on a certain number of explanatory variables (X_i) that have the probability density generated by the expression (see, Jong & Heller, 2008):

$$f(y_i|\theta_i, \ ) = exp\left\{\frac{y_i\theta_i \quad a(\theta_i)}{\quad} + c(y_i, \ )\right\}, \qquad y_i \epsilon S \tag{3.1}$$

where$S$ represents a subassembly that belongs to or set, θ_i is the natural (canonical) parameterand is the scale parameter. The searched parameters$\beta_1, \beta_2 \dots , \beta_p$, allow writing a function (g) for the mean (μ_i) of the variable Y_i as a linear combination of the exogenous variables X_i;

$$g(\mu) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = x_i^t \beta = \eta_i (3.2)$$

the monotonous and differentiable function g is known as a link function because it connects the linear predictor η_i with the mean μ_i. The choice of $a(\theta)$ determines the response distribution and the choice of $g(\mu)$, which is called the link determines how the mean is related to the explanatory variable x. Constructing interpretable models for connecting (or linking) such responses to variables can often give added insight into the complexity of the relationship which may often be hidden in a huge amount of data as multivariate methods such as GLM adjust for correlations and allow investigation into interaction effects.

3.5.1 Maximum Likelihood Estimation

The maximum likelihood estimation (MLE) of β *and*   are derived by maximizing the log-likelihood function, $\ell(\beta, \ )$ which is the logarithm of the likelihood as

$$\ell(\beta, \ ) = \sum_{i=1}^{n} \ln f(y_i; \beta, \ ) = \sum_{i=1}^{n} \left\{ \ln c(y_i, \ ) + \frac{y_i\theta_i \quad a(\theta_i)}{} \right\}, \tag{3.3}$$

which assumes independent exponential family responses y_i. Consider the MLE of β_j, to find the maximum $\ell(\beta, \ )$ is differentiated with respect to β_j:

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta_j},$$

where

$$\frac{\partial \ell}{\partial \theta_i} = \frac{y_i - a(\theta_i)}{} = \frac{y_i - \mu_i}{}, \qquad \frac{\partial \theta_i}{\partial \beta_j} = \frac{\partial \theta_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \theta_i}{\partial \eta_i} x_{ij}.$$

Here $\eta_i = x_i' \beta$ *and* x_{ij} is component i of x_j. Setting $\frac{\partial \ell}{\partial \beta_j} = 0$ yields the first order conditions for likelihood maximization:

$$\sum_{i=1}^{n} \frac{\partial \theta_i}{\partial \eta_i} x_{ij}(y_i \quad \mu_i) = 0 \qquad\qquad X'D(Y \quad \mu) = 0$$

where D is the diagonal matrix with diagonal entries $\partial\theta_i / \partial\eta_i$ (Jong & Heller, 2008),

$$\left(\frac{\partial \theta_i}{\partial \eta_i}\right)^{-1} = \frac{\partial \eta_i}{\partial \theta_i} = \frac{\partial \eta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \theta_i} = g(\mu_i) a(\theta_i) = g(\mu_i) V(\mu_i)$$

3.5.2 Exponential Family of Distributions

The concept of the exponential family of distributions is one of the key constructs that's fundamental to the theory of generalized linear models. The response variable Y has a density function $f(y)$ that can be expressed in the form.

$$f(y) = c(y,\ \ )exp\left\{\frac{y\theta \quad a(\theta)}{}\right\} \qquad ,$$

$$g(\mu) = x'\beta \tag{3.4}$$

whereθ and are the parameters. The parameter θ is called the canonical parameter and the dispersion parameter. The choice of the functions $a(\theta)$ and $c(y,\ \ )$ determine the actual probability function such as the negative binomial or gamma. In terms of $a(\theta)$,

$$E(y) = a(\theta), \qquad Var(y) = \ \ a(\theta) \tag{3.5}$$

Where $a(\theta)$ and $a(\theta)$ are the first and second derivatives of $a(\theta)$with respect to θ, respectively.

3.5.3 The Variance Function

For exponential family response distributions

$$a(\theta) = \frac{\partial a(\theta)}{\partial \theta} = \frac{\partial \mu}{\partial \theta} \equiv V(\mu),$$

and so one can always write $Var(y) = \ \ V(\mu)$ where $V(\mu)$is called the variance function, indicating the relationship between the mean and variance. In generalized linear modelling, the mean μ is related to explanatory variables, and thus the mean varies with the explanatory variables. As the mean varies, so does the variance, through$V(\mu)$. A model connecting the mean to explanatory variables is thus, at the same time, a model for the relationship between the variance and the explanatory variables. Although, there are many mean–variance relationships that cannot be captured with an exponential family density. However, this issue is addressed with quasi-likelihood methods (see Jong & Heller, 2008).

To show the relationship of the mean-variance expression, we define $f(y)$ and $f(y)$ as the first and second derivatives of $f(y)$ with respect to θ. Then

$$f(y) = f(y)\left\{\frac{y-a(\theta)}{}\right\}, \qquad f(y) = f(y)\left\{\frac{y-a(\theta)}{}\right\}^2 \quad f(y)\frac{a(\theta)}{}$$

Integrating both sides of each of these expressions with respect to y yields

$$0 = \frac{E(y)-a(\theta)}{}, \; 0 = \frac{E[\{y-a(\theta) \quad \}^2]}{2} - \frac{a(\theta)}{}. \qquad (3.6)$$

The left hand sides are zero since

$$\int f(y)dy = \frac{\partial}{\partial\theta}\int f(y)dy\,, \quad \int f(y)dy = \frac{\partial^2}{\partial\theta^2}\int f(y)dy\,,$$

Where $\int f(y)dy = 1$ and assuming integration and differentiation can be interchanged, the stated relations follows (Jong & Heller, 2008).

3.5.4 Standard distributions in the exponential family form

This section shows how the probability functions fit into the exponential family framework. For this family

$$ln\{f(y)\} = ln\{c(y,\phi)\} + \frac{y\theta - a(\theta)}{\phi} \qquad (3.7)$$

3.5.4.1 Binomial.

Suppose $y \sim B(n,\pi)$. Then $f(y) = \binom{n}{y}\pi^y(1 \quad \pi)^{n-y} \quad , \; y = 0,1,\ldots.,n$

It follows that $ln\{f(y)\}$ is

$$ln\{\pi^y(1 \quad \pi)^{n-y}\} = yln\left(\frac{\pi}{1-\pi}\right) + nln(1 \quad \pi) = \frac{y\theta - a(\theta)}{\phi}, \qquad (3.8)$$

Where $\theta = ln\{\pi/(1 \quad \pi\}, a(\theta) = nln(1+e^{\theta}) and \quad = 1$. It follows that the binomial is in the exponential family.

$$E(y) = a(\theta) = \frac{ne^{\theta}}{1+e^{\theta}} = n\pi \;, \qquad Var(y) = \quad a(\theta) = n\pi(1 \quad \pi)\,.$$

The binomial proportion y/n has exponential family probability function with the same θbut $a(\theta) = \ln\left(1 + e^{\theta}\right)$ and $\quad = 1/n$.

3.5.4.2 Normal

Suppose $y \sim N(\mu, \sigma^2)$, $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(y-\mu)^2}{2\sigma^2}\right)}$ $\quad \infty < y < \infty$

Apart from a numerical constant, $ln\{f(y)\}$ is

$$ln\sigma \quad \frac{(y-\mu)^2}{2\sigma^2} = \quad ln\sigma \quad \frac{y^2/2}{\sigma^2} + \frac{y\mu - \mu^2/2}{\sigma^2} \qquad (3.9)$$

The first two terms on the right which involve only y $and\sigma$ serve to define $ln\{y, \quad\}$ with $\quad = \sigma^2$ while the final term on the right is equivalent to the second term in equation (3.9) if $\theta = \mu$ and $a(\theta) = \theta^2/2$.We see that y belongs to an exponential family, moreover $a(\theta) = \theta = \mu$, $Var(y) = \quad a(\theta) = \sigma^2$.

3.5.4.3 Poisson

If $y \sim P(\mu)$, that is $f(y) = \frac{\mu^y e^{-\mu}}{y!}$ $\quad , \; y = 0,1,2, \dots,$

$$ln\{f(y)\} = \quad \mu + yln\mu \quad lny! = \quad lny! + \frac{y\theta \quad a(\theta)}{}, \qquad (3.10)$$

Provided $\quad = 1, \theta = \ln(\mu)$ and $a(\theta) = e^{\theta}$. This shows that the Poisson is in the exponential family and $a(\theta) = e^{\theta} = \mu = E(y) = a(\theta) = Var(y)$.

3.5.4.4 Negative binomial

If $y \sim NB(\mu, \kappa)$ with density function $f(y) = \frac{\Gamma(y+\frac{1}{\kappa})}{y!\Gamma(1/\kappa)} \left(\frac{1/\kappa}{\mu+1/\kappa}\right)^{1/\kappa} \left(\frac{\mu}{\mu+1/\kappa}\right)^{y}$ $\qquad y = 0,1,2$

The log of $f(y)$, apart from a constant involving y *and* κ is

$$y\ln\left(\frac{\mu}{1+\kappa\mu}\right) \quad \frac{1}{\kappa}\ln(1+\kappa\mu) = \frac{y\theta-a(\theta)}{\phi}, \tag{3.11}$$

$\phi = 1, \theta = ln\{\mu/(1+\kappa\mu)\}$and $a(\theta) = \quad (1/\kappa)\ln(1 \quad \kappa e^{\theta})$. For known κ, the negative binomial is thus in the exponential family with

$$E(y) = a(\theta) = \frac{e^{\theta}}{1 \quad \kappa e^{\theta}} = \mu\,, \qquad Var(y) = \phi a(\theta) = \frac{e^{\theta}}{(1 \quad \kappa e^{\theta})^2} = \mu(1+\kappa\mu).$$

3.5.4.5 Gamma

If $y \sim G(\mu, \nu)$, then $f(y) = \frac{y^{-1}}{\Gamma(\nu)}\left(\frac{y\nu}{\mu}\right)^{\nu} e^{-\frac{y\nu}{\mu}} \qquad y > 0$

$\ln\{f(y)\}$is

$$(\nu \quad 1)lny \quad ln\Gamma(\nu) + \frac{y(\quad \mu^{-1})}{\nu^{-1}} \quad \frac{ln\mu}{\nu^{-1}} + \nu ln\nu \tag{3.12}$$

$$= \left\{\frac{y\theta \quad a(\theta)}{\phi}\right\} + (\nu \quad 1)lny \quad ln\Gamma(\nu) + \nu ln\nu,$$

with$\theta = \quad 1/\mu, a(\theta) = \quad \ln(\quad \theta)$ *and* $\phi = 1/\nu$. It follows that gamma densities are in the exponential family with

$$E(y) = a(\theta) = \quad \frac{1}{\theta} = \mu\;, \qquad Var(y) = \phi a(\theta) = \nu^{-1}\frac{1}{\theta^2} = \frac{\mu^2}{\nu}.$$

3.5.4.6 Inverse Gaussian

Suppose $y \sim IG(\mu, \sigma^2)$ with density function $f(y) = \frac{1}{\sqrt{2\pi y^3}\sigma} e^{\left\{-\frac{1}{2y}\left(\frac{y-\mu}{\mu\sigma}\right)^2\right\}} y > 0.$

Then the log of the density function is

$$\frac{1}{2}\ln(2\pi y^3) \quad ln\sigma \quad \frac{1}{2y}\left(\frac{y-\mu}{\mu\sigma}\right)^2 \tag{3.13}$$

$$= \quad \frac{y}{2\mu^2\sigma^2} + \frac{1}{\mu\sigma^2} \quad \frac{1}{2y\sigma^2} \quad \frac{1}{2}\ln(2\pi y^3) \quad ln\sigma$$

$$= \frac{y\theta \quad a(\theta)}{} + terms\ involving\ only\ y\ and\ \sigma^2$$

Where $\theta = \quad 1/(2\mu^2)$, $a(\theta) = \quad \sqrt{\ 2\theta}$ and $\quad = \sigma^2$. Thus the Inverse Gaussian is therefore in the exponential family with

$$E(y) = a(\theta) = \frac{1}{\sqrt{-2\theta}} = \mu, \qquad Var(y) = \quad a(\theta) = \frac{\sigma^2}{(-2\theta)^{3/2}} = \sigma^2\mu^3 \ .$$

3.6 The Proposed Risk-based Adjustment Model

3.6.1 Estimation Model of Claim Frequency

The Poisson regression model is often suggested for count data but found to be inadequate because the data displays far greater variance than that predicted by the Poisson. Thus a Poisson model for the number of claims is inappropriate since the observed varianceis much larger than the mean. One alternative to Poisson regression is negative binomial regression. Within the actuarial literature, the negative binomial distribution is employed as a functional form that relaxes the equidispersion restriction of the Poisson model. It has been shown that the negative binomial distribution may be viewed as a statistical model for counts, in the situation where overdispersion is explained by heterogeneity of the mean over the population (see, Jong & Heller, 2008, David &Jemna, 2015). The negative binomial regression model, using the log link, is y ~NB(μ, κ) , lnμ = ln$n$ + x β.Another alternative choice is the quasi-likelihood. The negative binomial is intuitivelymore appealing than quasi-likelihood, because it explains the mechanismunderlying the overdispersion. However, quasi-likelihood provides estimateswhich are comparable and the results of the two analyses are usually equivalent. The only difference between the Poisson and quasi-likelihood (Poissonvariance) models is an inflation factor on the standard errors of the Poissonparameter estimates.In recent years the negative binomial has gained popularity as the distribution of choice when modelling overdispersed count data in many fields,possibly because of its simpler computational requirements and its availabilityin standard software.

Extant literature present various ways of constructing the negative binomial distribution,nevertheless Boucher, Denuit and Guillen (2008) argued that an intuitive way is the introduction of a random heterogeneity term θ with mean 1 and variance α in the mean parameter of the Poisson distribution. For an intensive discussion of this approach see Gourieroux *et al*. (1984a), Cameron and Trivedi (1990, 1998). The negative binomial is derived from a Poisson-gamma mixture distribution. Given λ, if the count y is Poisson distributed

$$y|\lambda \sim P(\lambda) \Rightarrow f(y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} \tag{3.14}$$

Suppose λ is a continuous random variable with probability density function (pdf) $g(\lambda)$ where $g(\lambda) = 0$ for $\lambda < 0$, then the unconditional pdf of y is

$$f(y) = \int_0^\infty f(y|\lambda)\, g(\lambda) d\lambda \text{(3.15)}$$

If $\lambda \sim G(\mu, v)$,

$$f(y) = \int_0^\infty \frac{e^{-\lambda}\lambda^{y}}{y!} \frac{\lambda^{-1}}{\Gamma(v)} \left(\frac{\lambda v}{\mu}\right)^{v} e^{-\lambda v/\mu} d\lambda \text{(3.16)}$$

$$= \frac{1}{y!\,\Gamma(\nu)} \left({}^{\nu}/_{\mu}\right)^{\nu} \int_0^\infty \lambda^{y+\nu-1}\, e^{-\lambda(1+\frac{\nu}{\mu})} d\lambda$$

$$= \frac{\Gamma(\nu + y)}{y!\,\Gamma(\nu)} \left(\frac{\nu}{\nu + \mu}\right)^{\nu} \left(\frac{\mu}{\nu + \mu}\right)^{y} \quad y = 0,1,2, \ldots.$$

Substituting $\kappa = 1/\nu$ results in the $NB(\mu, \kappa)$ (see, Jong & Heller, 2008). The first two moments of the negative binomial are $E(y) = \mu,\ Var(y) = \mu(1 + \kappa\mu)$.The standard estimator for this model is the maximum likelihood estimator.

3.6.2 Estimation Model of Claim Cost

The classical method for econometric modelling of claim cost is the gamma model due to parameters $\mu \ and \ \upsilon$ which offers more flexibility while estimating the cost of claims. Pinquet (1997) described a simple, realistic parametric model based on gamma distribution in modelling auto insurance claim cost. Letting $c_1, c_2, \dots, c_i$ be the cost of claims initiated by insured i, and assuming that they are independently gamma distributed, the probability density function (pdf) is given by :

$$f(c_i) = \frac{1}{\Gamma(\upsilon)}\left(\frac{\upsilon c_i}{\mu_i}\right)^{\nu} exp\left(\frac{\nu c_i}{\mu_i}\right), \qquad c_i > 0(3.17)$$

the mean $E(c_i) = \mu_i$ and the $Var(c_i) = \mu_i^2/\nu$ and the log-likelihood function for the Gamma model is given as:

$$(\beta) = \prod_{i|y_{i>0}} \prod_{k=1}^{y_i} \left(\frac{1}{\Gamma(\nu)}\left(\frac{\nu c_{ik}}{\mu_i}\right)^{\nu} \exp\left(\quad \frac{\nu c_{ik}}{\mu_i}\right)\frac{1}{c_{ik}}\right)(3.18)$$

The equations of the log-likelihood function for obtaining the estimators $\hat{\beta}_j$ are given by:

$$\mu_i \frac{\partial LL(\beta|c)}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \sum_{i|y_{i>0}} \sum_{k=1}^{y_i} \left(\quad \nu ln\mu_i \quad \frac{\nu c_{ik}}{\mu_i}\right) = 0(3.19)$$

which can be simplified as

$$\sum_{i|y_{i>0}} \sum_{k=1}^{y_i} x_{ij}\left(1 \quad \frac{\nu c_{ik}}{\mu_i}\right)$$

Defining $c_i = \mu_i = x_i'\beta$ the estimated cost of the claims for the insured i, the solution of the equation:

$$\sum_{i|y_{i>0}} \left(y_i \quad \frac{c_{i\cdot}}{c_i}\right) x_i = 0 \qquad (3.20)$$

is the maximum likelihood estimates $\hat{\beta}$.

3.6.3 Criteria for Assessing the Models' Goodness of Fit

There exists many statistics in the literature that can be used to select and assess the performance of the regression models, however Denuit and Lang (2004) described the likelihood ratio (LR) as the standard measure of goodness of fit for assessing the adequacy of various models. The test statistics follows a $\chi^2_{\alpha,p}$ distribution for a significance level α of 0.05 and p degrees of freedom, where p represents the number of explicative variables included in the regression model. This statistics test is obtained from the difference between the deviance of the regression model without covariates (D_0) and that of the deviance of the model including the independent variables (D_p):

$$LR = D_0 \quad D_p \qquad (3.21)$$

The deviance was defined by Charpentierand Denuit (2005) as twice the difference between the maximum log-likelihood achievable ($y_i \quad \lambda_i$) and the log-likelihood of the fitted model:

$$D = 2(LL(y_i|y_i) \quad LL(y_i|y_i) \qquad (3.22)$$

A value of the likelihood ratio higher than the statistics theoretical value ($LR > \chi^2_{\alpha,p}$) indicates that the regression model explains well the analysed data.

3.6.4 Risk Premium Modelling

Within non-life insurance, the risk premium represents the expected cost of all claims initiated by insured during the cover period. The calculation of the premium is based on statistical methods that incorporate all available information about the accepted risk with emphasis on better accurate assessment of tariffs ascribed to each insured.

The basis for calculating the risk premium is the econometric modelling of the frequency and cost of claims based on the characteristics that define the insurance contract. The risk premium is the mathematical expectation of the annual cost of claims declared by the insured and this is obtained by multiplying the estimated claim frequency and cost for the claims amount $(C_1, C_2, \ldots)$ independent of their number (Y):

$$E[\sum_{i=1}^{N} C_i] = E(Y) \times E(C_i)(3.23)$$

The separate approach of frequency and cost of claims is particularly relevant as shown by Charpentier*et al*. (2005), because the risk factors which influence the two components of the risk premium are usually different. Basically, the separate analysis of the two components gives a clearer perspective on how the risk factors are affecting the premium as it provide a better understanding of the way in which factors affect the cost of claims, and can more easily allow the identification and removal of certain random effects from one element of the experience.

3.7 The Risk-based Adjustment Modelling

Here, we describe the construction of the risk adjustment model. Claim-based risk modelling in automobile insurance is the process of determining the relative costs of an insured based on individual characteristics and claims history. Typically, the process involves grouping the claims history of an insured into categories. These classifications are intended to be as homogeneous as possible with respect to rating factors characteristics and cost. The categories serve as indicators for whether a person has that characteristic. A general approach for this model for n defined characteristics is represented as

$$Y_i = I + \beta_1 X_{1i} + \beta_2 X_{2i} + \quad + \beta_n X_{ni} + \varepsilon$$

Where

Y_i – risk-adjusted expected claims cost for policyholder i

I – intercept which is the minimum claim cost

β_i – coefficient for the *i*th classification

i – policyholder's value of 0 or 1 representing whether or not policyholder i possesses certain characteristics.

$X_{1i}, ..X_{ni}$ arethe predictor variables (risk factors)

ε the error term

CHAPTER FOUR

DATA PRESENTATION AND ANALYSIS

The results obtained through the application of the aforementioned models, based on which the risk premium is determined are presented and interpreted.

4.1 Descriptive Statistics for the Insured Portfolio

The preliminary descriptive analysis of the data is presented in tables 4.1 - 4.8. Table 4.1 presents the frequency distribution of policyholder in the portfolio. The observed mean claim frequency and mean claim cost for the portfolio are 14.09% and 284117.71 naira respectively. The age structure of the portfolio as described in Table 1 shows that most policyholders were middle-aged as 7730 insured drivers (representing 48.4% of the portfolio) were in the age bracket of 31 and 60 years.

Only 1458 insured drivers (representing 9.1% of the portfolio) were over 60 years. The young drivers represent 28% of the portfolio (4472), and the remaining 2318 insured drivers (14.5% of the portfolio) were in the age range of 24 to 30 years. There were 9672 male policyholders (representing 60.5 % of the portfolio) and 4958 female policyholders (representing 31.0 % of the portfolio) while it is 1248 for an entity and 100 for couples (representing 7.8% and 0.6 % of the portfolio respectively). The descriptive analysis of the data by claim costs, claim frequency and premiums for each of the rating factors are presented in Tables 4.2 to 4.8 respectively. There is evidence that the claims data is heavily tailed and highly peaked which suggest that the data is significantly non-normal.

Table 4.1: Frequency distribution of policyholder in the portfolio

Variables		*Frequency*	*Percentage*
Age group			
	Less than 24 years	4472	28.0
	24 - 30 years	2318	14.5
	31 - 60 years	7730	48.4
	61 years and Above	1458	9.1
Classification of Policyholder			
	Male	9672	60.5
	Female	4958	31.0
	Entity	1248	7.8
	Couple	100	.6
Geo-political zone			
	FCT	976	6.1
	South-west	13144	82.3
	South-east	327	2.0
	South-south	981	6.1
	North-east	57	.4
	North-west	296	1.9
	North-central	197	1.2
Occupation			
	Self-employed	1340	8.4
	Publicly employed	6078	38.0
	Privately employed	8210	51.4
	Unemployed	350	2.2
Product type			
	Commercial Vehicle	2783	17.4
	Comprehensive	12520	78.4
	Third party	641	4.0
	Motorcycle	34	.2
Nature of loss			
	Theft	306	1.9
	Collision	14261	89.3
	Accident	391	2.4
	Vandalisation	767	4.8
	Others	253	1.6
Customer type			
	Individual	13283	83.1
	Companies	2611	16.3
	Government	77	.5
	All account	7	.0

Source: Researcher's computation 2016

From Table 4.2, one can see that on the average claim costs decreases initially with age and then increases along the age group. This may be attributed to the fact that younger drivers on average have larger claims because they have less driving experience and take more risks, older individuals on the other hand are riskier drivers due to a deterioration of their cognitive and sensory skills (McKnight & McKnight, 1999, 2003; Kelly & Nielson, 2006). It can be noticed that the policyholders aged less than 24 years with observed average claim frequency of 19.46% tends to report more claims on the average than the policyholders aged between 24 and 30 (observed mean claim frequency of 9.08%).

Table 4.2: Descriptive analysis of claim cost, claim frequency and premiums by age group

AGE GROUP		Mean	N	Std. Deviation	Kurtosis	Skewness
< 24 years	CLAIMS COST	401330.9142	4472	951355.82177	53.170	5.608
	CLAIM FREQUENCY	19.46	4472	39.524	11.676	3.320
	PREMIUM	7229804.8883	4472	13834803.33631	9.494	2.963
24 - 30 years	CLAIMS COST	172641.9702	2318	410523.22618	38.346	5.297
	CLAIM FREQUENCY	9.08	2318	20.209	32.306	5.003
	PREMIUM	76074.0459	2318	114142.64100	114.794	9.288
31 - 60 years	CLAIMS COST	209692.6571	7730	585718.04689	92.733	7.547
	CLAIM FREQUENCY	10.65	7730	25.706	30.778	5.119
	PREMIUM	115915.6951	7730	228360.25275	333.667	14.811
≥61 years	CLAIMS COST	496414.6381	1458	1120956.83538	16.751	3.727
	CLAIM FREQUENCY	23.78	1458	48.159	7.671	2.870
	PREMIUM	15000794.3528	1458	25093219.11026	.539	1.544

Source: Researcher's computation 2016

Table 4.3: Descriptive analysis of claim cost, claim frequency and premiums by gender

GENDER		Mean	N	Std. Deviation	Kurtosis	Skewness
Male	CLAIMS COST	258423.4138	9672	666670.26578	39.477	5.461
	CLAIM FREQUENCY	13.04	9672	30.293	22.010	4.404
	PREMIUM	2839125.8332	9672	10913413.8530	27.765	5.256
Female	CLAIMS COST	317002.2362	4958	871005.60168	82.400	7.090
	CLAIM FREQUENCY	15.36	4958	34.884	16.641	3.923
	PREMIUM	5258711.5213	4958	13912482.8924	11.478	3.373
Entity	CLAIMS COST	366038.6440	1248	909480.35777	28.403	4.816
	CLAIM FREQUENCY	17.76	1248	38.753	14.859	3.732
	PREMIUM	1388533.5632	1248	2333449.62718	6.463	2.584
Joint Gender	CLAIMS COST	116481.4381	100	222596.77872	17.724	3.903
	CLAIM FREQUENCY	6.26	100	11.105	17.809	3.912
	PREMIUM	96069.1945	100	82334.89117	8.006	2.182

Source: Researcher's computation 2016

From the exploratory data analysis result displayed in Tables 4.2 to 4.8, very positive skewness and heavy tailed kurtosis were observed for all the rating factors. Surprisingly, the mean claim cost for female was higher than for male and the female policyholders tends to report more claim than their male counterpart as presented in Table 4.3.

Table 4.4: Descriptive analysis of claim cost, claim frequency and premiums by product type

PRODUCT TYPE		Mean	N	Std. Deviation	Kurtosis	Skewness
Commercial Vehicle	*CLAIMS COST*	514330.3445	2783	1073323.65449	13.949	3.486
	CLAIM FREQUENCY	24.91	2783	47.319	7.647	2.828
	PREMIUM	9453698.6903	2783	19496649.0307	4.494	2.451
Comprehensive	*CLAIMS COST*	211841.8735	12520	604817.01584	143.937	8.899
	CLAIM FREQUENCY	10.72	12520	25.639	29.731	5.007
	PREMIUM	1248764.7010	12520	6138406.65700	111.524	9.683
Third party	*CLAIMS COST*	705222.3754	641	1250866.11870	17.563	3.580
	CLAIM FREQUENCY	33.32	641	50.857	4.327	2.183
	PREMIUM	20762294.4163	641	20224342.8616	-1.630	.441
Motor Cycle	*CLAIMS COST*	115987.3162	34	383932.14250	25.018	4.866
	CLAIM FREQUENCY	6.26	34	19.111	25.127	4.877
	PREMIUM	658981.0382	34	539777.63776	-1.905	.134

Source: Researcher's computation 2016

Table 4.5: Descriptive analysis of claim cost, claim frequency and premiums by district

DISTRICT		Mean	N	Std. Deviation	Kurtosis	Skewness
FCT	*CLAIMS COST*	276329.8562	976	744046.96642	58.800	6.693
	CLAIM FREQUENCY	13.51	976	30.363	26.264	4.755
	PREMIUM	1556257.8726	976	3700582.81843	9.457	3.005
Southwest	*CLAIMS COST*	290219.0663	13144	764725.01972	59.981	6.080
	CLAIM FREQUENCY	14.40	13144	33.144	18.655	4.116
	PREMIUM	4022015.0710	13144	12641762.31483	17.358	4.164
Southeast	*CLAIMS COST*	349979.8987	327	1153605.28730	77.690	7.693
	CLAIM FREQUENCY	15.43	327	37.538	18.910	4.185
	PREMIUM	407363.8006	327	1365382.47806	40.131	6.108
South south	*CLAIMS COST*	233442.7999	981	577557.23185	39.368	5.353
	CLAIM FREQUENCY	11.98	981	27.295	22.822	4.466
	PREMIUM	671996.1053	981	2378447.93545	15.266	4.136
Northeast	*CLAIMS COST*	224186.4833	57	480904.28262	23.447	4.321
	CLAIM FREQUENCY	11.65	57	23.982	23.460	4.323
	PREMIUM	147686.8893	57	271856.22359	21.882	4.631
Northwest	*CLAIMS COST*	215772.8639	296	524438.96518	31.334	4.823
	CLAIM FREQUENCY	11.16	296	25.245	23.686	4.331
	PREMIUM	181037.0312	296	516820.44945	41.046	6.186
North central	*CLAIMS COST*	178665.7845	197	386351.19490	23.446	4.347
	CLAIM FREQUENCY	9.42	197	19.279	23.537	4.354
	PREMIUM	185271.8541	197	479569.07579	21.861	4.549

Source: Researcher's computation 2016

Table 4.6: Descriptive analysis of claim cost, claim frequency and premiums by occupation

OCCUPATION		Mean	N	Std. Deviation	Kurtosis	Skewness
Self	*CLAIMS COST*	265276.7378	1340	716639.84457	47.324	5.908
	CLAIM FREQUENCY	13.20	1340	31.195	20.612	4.314
	PREMIUM	128740.2949	1340	237204.50189	152.307	10.702
Public	*CLAIMS COST*	415578.9545	6078	980665.78877	41.087	5.059
	CLAIM FREQUENCY	20.12	6078	41.326	10.994	3.267
	PREMIUM	8742300.5670	6078	17493124.76674	5.778	2.585
Private	*CLAIMS COST*	191669.9504	8210	528868.40013	115.710	8.160
	CLAIM FREQUENCY	9.83	8210	23.402	34.864	5.351
	PREMIUM	228580.1342	8210	1143383.44592	133.666	10.858
Unemployed	*CLAIMS COST*	241893.4559	350	571187.65993	15.010	3.784
	CLAIM FREQUENCY	12.55	350	28.546	15.024	3.786
	PREMIUM	258159.5573	350	642200.49103	15.929	4.075

Source: Researcher's computation 2016

Table 4.7: Descriptive analysis of claim cost, claim frequency and premiums by loss type

LOSS TYPE		Mean	N	Std. Deviation	Kurtosis	Skewness
Theft	*CLAIMS COST*	1163962.8554	306	1154331.75127	1.959	1.308
	CLAIM FREQUENCY	57.87	306	54.929	.289	.992
	PREMIUM	649733.8739	306	2768907.90697	56.950	7.245
Collision	*CLAIMS COST*	274724.3224	14261	748521.25543	73.103	6.744
	CLAIM FREQUENCY	13.59	14261	31.694	21.474	4.381
	PREMIUM	3831754.6856	14261	12172273.89369	18.993	4.326
Accident	*CLAIMS COST*	413686.4916	391	987879.82150	15.970	3.738
	CLAIM FREQUENCY	20.18	391	44.211	9.449	3.113
	PREMIUM	170344.4995	391	647893.56104	125.657	10.813
Vandalisation	*CLAIMS COST*	91856.8173	767	249443.68721	207.602	12.628
	CLAIM FREQUENCY	5.06	767	11.971	182.443	11.828
	PREMIUM	128675.4838	767	539668.84504	295.180	16.468
Others	*CLAIMS COST*	132058.9865	253	243728.15203	32.525	4.740
	CLAIM FREQUENCY	7.08	253	12.193	32.719	4.754
	PREMIUM	1053133.0269	253	6364258.48667	97.573	9.545

Source: Researcher's computation 2016

The mean number of claims per product type was 24.91 for commercial vehicle, 10.72 in an auto comprehensive, 33.32 in auto third party liability and 6.26 for a motorcycle. On average, policyholders paid annual premiums of 9453698 naira in commercial vehicle, 1248764 naira in auto comprehensive, 20762294 naira in auto third party liability and 658981 in motorcycle.

Table 4.8: Descriptive analysis of claim cost, claim frequency and premiums by customer type

CUSTOMER TYPE		Mean	N	Std. Deviation	Kurtosis	Skewness
Individual	CLAIMS COST	249739.6168	13283	683966.88816	95.638	7.438
	CLAIM FREQUENCY	12.49	13283	29.208	23.667	4.542
	PREMIUM	1990692.1712	13283	8453476.41793	39.822	5.987
Companies	CLAIMS COST	465314.7008	2611	1035500.17669	18.107	3.856
	CLAIM FREQUENCY	22.50	2611	44.948	8.910	3.007
	PREMIUM	10820895.7322	2611	19788390.11086	3.853	2.309
Government	CLAIMS COST	85362.3155	77	132598.08025	11.684	3.122
	CLAIM FREQUENCY	4.79	77	6.638	11.793	3.123
	PREMIUM	7520266.5124	77	4701386.11663	-1.611	-.589
All account	CLAIMS COST	118829.2857	7	155438.42483	2.526	1.612
	CLAIM FREQUENCY	6.57	7	7.721	2.459	1.597
	PREMIUM	61457.1429	7	3928.89176	-2.739	-.392

Source: Researcher's computation 2016

The preliminary exploratory data analysis findings are that the automobile liability claims are heavily tailed and highly peaked suggesting the suitability of generalized linear modelling (Jong & Heller, 2008; Frees, 2010).

4.2 Automobile Claims Modelling

The regression models fitted considered the two components of insurance risk premium (frequency and severity). For these two components, seven different models are fitted depending on the predictor variables captured. For claims frequency, model 1 includes all the rating factors as the predictors of the number of claims, model 2 consist of age and gender characteristics as the predictors, while model 3 covers age characteristics as the only predictors; and model 4 considers gender as the predictors of claims frequency. Model 5, 6 and 7 uses the district of the insured, occupational types and customer types respectively as the predictors of frequency of claims. For automobile claims cost, model 1 uses all the risk factors (characteristics) in building the models, while model 2 incorporates the age and gender characteristics as the predictors of claims cost and model 3, 4, 5,6 and 7 comprises the age, gender, district, type of loss and product types only in constructing the models respectively.

Poisson

The results of the type 3 analysis are presented in Table 4.9. This enables the contribution evaluation of each variable taking into consideration all the other exogenous variables. The p-value column indicates the probability associated to the likelihood ratio test which appreciates the impact of each risk factor on the studied phenomenon. It can be observed that all the rating variables are statistically significant with a p-value (<.0.05), which clearly underlines their influence on the claims frequency.

Table 4.9: Likelihood Ratio Statistics for Type 3 Analysis

Source	Likelihood Ratio Chi-Square	df	P-value
(Intercept)	787.572	1	.000
Age	2520.802	3	.000
Gender	419.494	3	.000
District	385.777	6	.000
Occupation	1008.901	3	.000
Product type	18012.051	3	.000
Loss type	37553.284	4	.000
Customer type	648.441	3	.000

Source: Researcher's computation 2016

The goodness-of-fit statistics displayed in Table 4.10 provides measures that are useful for comparing competing models. Additionally, the Values for the Deviance and Pearson Chi-Square statistics divided by its degree of freedom gives corresponding estimates for the scale parameter. To verify if the data are overdispersed, the most common way is the interpretation of the deviance and Pearson statistics values.These values should be near 1.0 for a Poisson regression; the fact that they are greater than 1.0 (28.877 and 57.799 respectively) indicates an inequality between the mean and variance of the claim frequency, and thus the overdispersion hypothesis is confirmed.The analysis of parameter estimates of the Poisson regression coefficients for each of the predictors variables along with their standard errors, Wald chi-square values and p-values for the coefficients are presented in

Table 4.11.

Table 4.10: Goodness of fit test

Criterion	Value	df	Value/df
Deviance	460638.564	15952	28.877
Pearson Chi-Square	922017.507	15952	57.799
Log Likelihood	-256858.784		
Akaike's Information Criterion (AIC)	513769.568		
Finite Sample Corrected AIC (AICC)	513769.656		
Bayesian Information Criterion (BIC)	513969.222		
Consistent AIC (CAIC)	513995.222		

Source: Researcher's computation 2016

Table 4.11: Analysis of Parameter Estimates

Parameter	Estimate	Std. Error	Wald Chi-Square	P-value
(Intercept)	0.659	.1720	14.661	0.000
<24 years	-.114	.0070	262.616	.000
24 - 30 years	-.498	.0111	2018.276	.000
31 - 60 years	-.389	.0090	1872.954	.000
≥61 years	0a			
Male	.374	.0402	86.774	.000
Female	.444	.0403	121.714	.000
Entity	.322	.0409	62.136	.000
Couple	0a			
FCT	.325	.0249	170.439	.000
South-west	.259	.0234	121.855	.000
South-east	.408	.0272	224.298	.000
South-south	.329	.0251	172.113	.000
North-east	.191	.0456	17.525	.000
North-west	.150	.0291	26.731	.000
North-central	0a			
Self- employed	.086	.0170	25.543	.000
Publicly- employed	.087	.0160	29.827	.000
Privatly-employed	-.094	.0156	36.444	.000
Uemployed	0a			
Commercial vehicle	1.523	.0687	491.621	.000
Comprehensive	0.873	.0687	161.364	.000
Third party	1.801	.0690	680.118	.000
Motor cycle	0a			
Theft	2.205	.0249	7864.756	0.000
Collision	.408	.0238	293.155	.000
Accident	1.147	.0262	1914.115	0.000
Vandalisation	-.183	.0286	40.726	.000
Others	0a			
Individual	.011	.1476	.006	.941
Companies	.035	.1477	.055	.814
Government	-1.081	.1567	47.565	.000
All account	0a			
(Scale)	1b			
Dependent Variable: CLAIMS FREQUENCY				

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Source: Researcher's computation 2016

Negative binomial

Tables 4.12, 4.13, 4.14 and 4.15 present the results of the claim frequency modelling based on the negative binomial regression analysis. These show that the different age categories, gender, occupation, district, product type, loss type and customer type are significant in determining the number of claims reported. The results presented suggest that the fitted model is significant based on the goodness of fit tests, at the value/df column for the Pearson chi-square test. The results of the type 3 analysis presented in Table 4.11 shows that each of the rating variables is statistically significant. The table includes the six degree of freedom test which indicates that as a whole, the rating variable district is a significant predictor of the number of claims occurrence. The likelihood ratio chi-square statistic test of the overall model against a null model shows that our model is a significant improvement over the model without any predictors by looking at the p-value (< 0.000) of this test.

Table 4.12: Goodness of fit test

Criterion	Value	df	Value/df
Deviance	29164.489	15952	1.828
Pearson Chi-Square	68622.029	15952	4.302
Log Likelihood	-56092.990		
Akaike's Information Criterion (AIC)	112237.979		
Finite Sample Corrected AIC (AICC)	112238.067		
Bayesian Information Criterion (BIC)	112437.633		
Consistent AIC (CAIC)	112463.633		

Source: Researcher's computation 2016

Analysing the results presented in Table 4.9, it is noted that the value of deviance and Pearson divided by the number of degrees of freedom are now closer to 1.0 (1.828 and 4.302 respectively). This is a significant improvement over the Poisson model.

Testing for Poisson overdispersion.

One problem with the overdispersed Poisson regression is that there is no formal way to test it versus the standard Poisson regression. However, one suggested formal test to determine whether there is overdispersion is to perform a likelihood ratio test between a standard Poisson regression and a negative binomial regression with all other settings equal.With a negative binomial fit, an estimated κ close to zero suggests a Poisson response. A formal test of $\kappa = 0$is based on the likelihood ratio test. Since $\kappa = 0$ is at the boundary of thepossible range $\kappa \geq 0$, the distribution of the test statistic is non-standard and requires care. The likelihood ratio test statistic is -2(P NB) where P and NB are the values of the log-likelihood under the negative binomial and Poisson models, respectively. The distribution of the statistic has a mass of 0.5 atzero, and a half Chi-square one degree of freedom distribution above zero. A test at the 100α% significancelevel, requires a rejection region corresponding to the upper 2α point of the Chi-square one degree of freedomdistribution (Cameron and Trivedi 1998).The Poisson and negative binomial regressions yieldP =-256858.784, NB = -56092.990. Hence the likelihood ratio statistic is 401531.588.The hypothesis $\kappa = 0$ is rejected, at all significance levels. The conclusion isthat overdispersion is indeed present. For a significance level $\alpha = 0.05$, thehypothesis $\kappa = 0$ is rejected if the likelihood ratio statistic is greater than theupper 10% point of the Chi-square one degree of freedom distribution, which is 2.71.

Table 4.13 : Analysis of Parameter Estimates

Parameter	Estimate	Std. Error	Wald Chi-Square	P-value
(Intercept)	.672	.4782	1.977	.160
<24 years	-.010	.0341	.092	.762
24 - 30 years	-.437	.0448	95.416	0.000
31 - 60 years	-.341	.0399	72.874	.000
≥61 years	0[a]			
Male	.408	.1093	13.944	.000
Female	.452	.1100	16.868	.000
Entity	.357	.1142	9.749	.002
Couple	0[a]			
FCT	.196	.0831	5.553	.018
South-west	.045	.0766	.352	.553
South-east	.144	.0958	2.273	.132
South-south	.181	.0832	4.743	.029
North-east	.134	.1593	.703	.402
North-west	-.039	.0979	.158	.691
North-central	0[a]			
Self- employed	.192	.0637	9.042	.003
Publicly- employed	.183	.0607	9.069	.003
Privately-employed	-.041	.0579	.491	.483
Unemployed	0[a]			
Commercial vehicle	1.561	.1879	69.045	.000
Comprehensive	.922	.1875	24.196	.000
Third party	1.800	.1920	87.950	0.000
Motor cycle	0[a]			
Theft	2.231	.0891	626.905	0.000
Collision	.344	.0683	25.267	.000
Accident	1.121	.0856	171.493	0.000
Vandalisation	-.187	.0785	5.704	.017
Others	0[a]			
Individual	.040	.4067	.010	.921
Companies	.103	.4078	.063	.801
Government	-1.115	.4270	6.815	.009
All account	0[a]			
(Scale)	1[b]			
(Negative binomial)	1.710	.0175		
Dependent Variable: CLAIMS FREQUENCY				

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Source: Researcher's computation 2016

Table 4.14: Wald Statistics for Type 3 Analysis

Source	Wald Chi-Square	df	P-value
(Intercept)	387.920	1	.000
Age	183.715	3	.000
Gender	23.978	3	.000
District	36.240	6	.000
Occupation	116.094	3	.000
Product type	836.374	3	.000
Loss type	1469.071	4	.000
Customer type	86.584	3	.000

LR Chi-Square: (5406.714, p-value<0.000)

Source: Researcher's computation 2016

The analysis of parameter estimates table contains the negative binomial regression coefficients for each of the predictor variables along with their standard errors, Wald chi-square values and p-values for the coefficients. Analyzing the result from Table 4.10, a decrease of the claims frequency can be observed along with an increase in the age of the insured. On the contrary, when the gender coefficient increases, the frequency of claims increases as well. Additionally, there is an estimate of the dispersion coefficient, (Negative binomial). The parameter 95% confidence interval does not include zero, suggesting that the model fitted is more appropriate than the Poisson.

Table 4.15: Negative binomial regression analysis of automobile claim frequency

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
Parameter	B	P value	B	P value	B	P value	B	P value	B	P value	B	P value	B	P value
(Intercept)	.672	.160	2.512	.000	3.169	.000	1.834	.000	2.243	.000	2.530	.000	1.883	.000
<24 years	-.010	.762	-.193	.000	-.200	.000								
24 - 30 years	-.437	.000	-.959	.000	-.963	.000								
31 - 60 years	-.341	.000	-.798	.000	-.804	.000								
≥61 years	0^{a}		0^{a}		0^{a}									
Male	.408	.00	.652	.000			.734	.000						
Female	.452	.00	.668	.000			.897	.000						
Entity	.357	.002	.621	.000			1.043	.000						
Couple	0^{a}		0^{a}				0^{a}							
FCT	.196	.018							.360	.000				
South-west	.045	.553							.424	.000				
South-east	.144	.132							.493	.000				
South-south	.181	.029							.240	.003				
North-east	.134	.402							.212	.177				
North-west	-.039	.691							.169	.079				
North-central	0^{a}								0^{a}					
Self- employed	.192	.003									.050	.420		
Publicly-employed	.183	.003									.472	.000		
Privately-employed	-.041	.483									-.245	.000		
Unemployed	0^{a}										0^{a}			
Commercial vehicle	1.561	.000												
Comprehensive	.922	.000												
Third party	1.800	.000												
Motor cycle	0^{a}													
Theft	2.231	.000												
Collision	.344	.000												
Accident	1.121	.000												
Vandalisation	-.187	.017												
Others	0^{a}													
Individual	.040	.921											.642	.114
Companies	.103	.801											1.231	.002
Government	-1.115	.009											-.316	.457
All account	0^{a}												0^{a}	
(Scale)	1^{b}		1^{b}		1^{b}		1^{b}		1^{b}		1^{b}		1^{b}	
(Negative binomial)	1.71	0.018	1.89	0.019	1.892	0.019	1.972	0.02	1.978	0.02	1.897	0.019	1.939	.019
Dependent Variable: CLAIMS FREQUENCY														
Goodness of fit test														
Deviance	29164.49		32722.80		32756.21		34368.65		34496.82		32865.37		33703.72	
Pearson Chi-Square	68622.03		74551.78		74801.40		78374.21		78561.75		75287.60		77355.76	
Log Likelihood	-56092.99		-57872.14		-57888.85		-58695.1		-58759.2		-57943.43		-58362.61	
(AIC)	112237.98		115758.3		115785.7		117398.1		117532.3		115894.86		116733.21	
(AICC)	112238.07		115758.3		115785.7\		117398.1		117532.3		115894.86		116733.21	
(BIC)	112437.63		115812		115816.4		117428.9		117586.2		115925.57		116763.93	
(CAIC)	112463.63		115819.		115820.4		117432.9		117593.1		115929.57		116767.93	
Df	15952.00		15971		15974		15974		15971		15974		15974	

Source: Researcher's computation 2016

Gamma

The next step in establishing the risk premium is estimating the cost of claims based on the risk factors considered. For the Gamma model, Tables 4.16, 4.17, 4.18 and 4.19 present the analysis. Table 4.13 shows that the cost of claims is influenced by the age of the insured, the gender and the district where the insured resides, the profession of the insured, the type of product, the nature of the loss type and the customer type. The influence factors of the claims cost are similar to the factors corresponding to the frequency of claims, fact that refutes the assumption suggested by the actuary literature regarding the separate analysis of these two components of risk premium. This could be attributed to the type of risk factors considered in the rating process.

Table 4.16: Wald Statistics for Type 3 Analysis

Source	Wald Chi-Square	df	P-value
(Intercept)	5130.532	1	0.000
Age	117.670	3	0.000
Gender	20.676	3	.000
District	36.448	6	.000
Occupation	73.091	3	.000
Product type	446.059	3	0.000
Loss type	788.187	4	0.000
Customer type	58.898	3	.000

LR Chi-Square: (2865.765, p-value<0.000)

Source: Researcher's computation 2016

Table 4.17 : Analysis of Parameter Estimates

Parameter	Estimate	Std. Error	Wald Chi-Square	P-value
(Intercept)	10.273	.6512	248.887	0.000
<24 years	-.002	.0489	.002	.966
24 - 30 years	-.495	.0635	60.554	.000
31 - 60 years	-.363	.0572	40.215	.000
≥61 years	0^a			
Male	.475	.1476	10.335	.001
Female	.543	.1485	13.379	.000
Entity	.422	.1549	7.416	.006
Couple	0^a			
FCT	.239	.1149	4.341	.037
South-west	.050	.1058	.226	.635
South-east	.348	.1330	6.864	.009
South-south	.199	.1151	3.005	.083
North-east	.163	.2205	.548	.459
North-west	-.051	.1352	.140	.708
North-central	0^a			
Self- employed	.230	.0887	6.736	.009
Publicly- employed	.221	.0846	6.829	.009
Privately-employed	-.028	.0804	.119	.730
Unemployed	0^a			
Commercial vehicle	1.689	.2547	43.993	.000
Comprehensive	1.025	.2538	16.299	.000
Third party	1.934	.2608	55.023	.000
Motor cycle	0^a			
Theft	2.312	.1248	343.557	0.000
Collision	.388	.0932	17.382	.000
Accident	1.226	.1188	106.540	0.000
Vandalisation	-.204	.1064	3.665	.056
Others	0^a			
Individual	.086	.5539	.024	.876
Companies	.142	.5554	.066	.798
Government	-1.205	.5806	4.309	.038
All account	0^a			
(Scale)	2.140^b	.0196		
Dependent Variable: CLAIMS COST				

a. Set to zero because this parameter is redundant.

b. Maximum likelihood estimate

Source: Researcher's computation 2016

The results obtained in measuring the quality of Gamma regression model using Fisher statistic, which is the last step of claim cost analysis are shown in Table 4.15.

Table 4.18: Goodness of fit test

Criterion	Value	df	Value/df
Deviance	43908.281	15952	2.753
Pearson Chi-Square	104295.893	15952	6.538
Log Likelihood	-209046.713		
Akaike's Information Criterion (AIC)	418147.425		
Finite Sample Corrected AIC (AICC)	418147.520		
Bayesian Information Criterion (BIC)	418354.757		
Consistent AIC (CAIC)	418381.757		

Source: Researcher's computation 2016

The obtained value of Fisher statistic test within the studied portfolio is much higher than the theoretical value, meaning that the proposed Gamma model fits well the data and its employment is significant in explaining the variation of claim cost.

Table 4.19: Gamma regression analysis of automobile claims costs

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
Parameter	B	P value	B	P value	B	P value	B	P value	B	P value	B	P value	B	P value
(Intercept)	10.273	.000	12.399	.000	13.115	.000	11.665	.000	12.093	.000	11.661	.000	11.791	.000
<24 years	-.002	.966	-.209	.000	-.213	.000								
24 - 30 years	-.495	.000	-1.048	.000	-1.056	.000								
31 - 60 years	-.363	.000	-.853	.000	-.862	.000								
≥61 years	0[a]		0[a]		0[a]									
Male	.475	.001	.701	.000			.797	.000						
Female	.543	.000	.738	.000			1.001	.000						
Entity	.422	.006	.692	.000			1.145	.000						
Couple	0[a]		0[a]				0[a]							
FCT	.239	.037							.436	.002				
South-west	.050	.635							.485	.000				
South-east	.348	.009							.672	.000				
South-south	.199	.083							.267	.054				
North-east	.163	.459							.227	.395				
North-west	-.051	.708							.189	.248				
North-central	0[a]								0[a]					
Self- employed	.230	.009												
Publicly-employed	.221	.009												
Privately-employed	-.028	.730												
Unemployed	0[a]													
Commercial vehicle	1.689	.000									1.489	.000		
Comprehensive	1.025	.000									.602	.042		
Third party	1.934	.000									1.805	.000		
Motor cycle	0[a]										0[a]			
Theft	2.312	.000											2.176	.000
Collision	.388	.000											.733	.000
Accident	1.226	.000											1.142	.000
Vandalisation	-.204	.056											-.363	.004
Others	0[a]												0[a]	
Individual	.086	.876												
Companies	.142	.798												
Government	-1.205	.038												
All account	0[a]													
(Scale)	2.140[b]	.0196	3.011[b]		3.013[b]		3.138[b]		3.150[b]		2.972[b]		3.037[b]	
Dependent Variable: CLAIMS COSTS														
							Goodness of fit test							
Deviance	43908.3		48090.6		48135.8		50131.2		50307.3		48287.8		48502.9	
Pearson Chi-Square	104295.9		105251.5		105956.6		109844.8		110780.7		108057.1		114891.5	
Log Likelihood	-209046.7		-215449.0		-215471.6		-216469.3		-216557.3		-215547.6		-215655.1	
(AIC)	418147.4		430911.9		430951.1		432946.5		433128.6		431103.2		431320.2	
(AICC)	418147.5		430911.9		430951.1		432946.5		433128.6		431103.2		431320.2	
(BIC)	418354.8		430965.7		430981.8		432977.2		433182.4		431133.9		431358.6	
(CAIC)	418381.8		430972.7		430985.8		432981.2		433189.4		431137.9		431363.6	
Df	15952		15971		15974		15974		15971		15974		15973	

Source: Researcher's computation 2016

The regression models fitted are then used to determine coefficients for each characteristic which are ultimately used to compute the risk score for any insured (usually, risk scores are stated relative to 1.0, with 1.0 being equal to the average expected risk score across the entire population). These scores are the sum of coefficient values for present situations. They represent the expected relative cost of an insured.

4.3 Implementation of the risk adjustment model

An illustration of the implementation of the risk adjustment model is discussed here. Consider the risk score for the four different age groups. The average of these risk scores is computed as:

$$\frac{2.969 + 2.206 + 2.365 + 3.169}{4} = 2.6771$$

The adjustment coefficient would be computed for all age group by dividing the age group's risk score by the average risk score as follow:

Table 4.20: Adjusted coefficient for age group based on the frequency and cost models

Age group	*<24 years*	*24 - 30 years*	*31 - 60 years*	*≥61 years*
Risk adjusted coefficient (Negative binomial)	1.1090	0.8240	0.8834	1.1837
Risk adjusted coefficient (Gamma)	1.0255	0.9584	0.9738	1.0423

Source: Researcher's computation 2016

From the computed adjusted coefficient displayed in table 4.20, it could be noticed that the risk scores vary across age groups for all insured. This justifies the introduction of the risk adjustment as the risks brought into the pool by different age group are different significantly. To minimize risk selection by the insurer a risk adjustment model can be employed for adjusting the premium. The risk adjusted coefficients for all the predictors (risk factors) for the claims frequency model (negative binomial) and claims cost model (gamma) are computed and presented in table 4.21. The adjusted coefficients results displayed using the negative binomial for the frequency model shows that an insured aged less than 24 years tends to report about 11% more claims than the portfolio average reported claims. This implies that insured in this age group are more risky and therefore should pay a little more than the average premium relative to their risk level. Policyholders who reside in southern region of the country are riskier than those in the northern region

Looking at the frequency and severity components of automobile insurance risk premium from table 4.18, it is interesting to note that the relatively high importance of the place of residence is in line with the high regional differences noted for motor insurance premiums (see, for example, K¨ohne, 2011). Thus, this implies that motor insurers still need to heavily rely on regional classes for tariff schemes. The reason being that the frequency and severity of accidents differ substantially between places due to different traffic volume and road conditions (higher frequency in urban areas, worse consequences in rural regions, see, e.g., Etgar, 1975, and Sipulskyte, 2012). Considering the relationship within the analysed insurance portfolio, the profile of policyholders with the higher risk for the insurance company can be established. For example, an insured male aged less than 24 years, that's publicly employed, who resides in the southwest region and purchased a third party motor insurance policy, but loss his car through theft would be expected to report a claim 6.8 times an average member.

Table 4.21: Adjusted coefficients for all rating factors based on claim frequency and cost models

	AGE		GENDER		DISTRICT		OCCUPATION		PRODUCT TYPE		LOSS TYPE		CUSTOMER TYPE	
	Freq	Cost	Freq	Cost	Freq	Cost	Freq	Cost	Freq	Cost	Freq	Cost	Freq	Cost
<24 years	1.11	1.03												
24 - 30 years	0.82	0.96												
31 - 60 years	0.88	0.97												
≥61 years	1.18	1.04												
Male			1.03	1.00										
Female			1.09	1.02										
Entity			1.15	1.03										
Couple			0.73	0.94										
FCT					1.04	1.01								
South-west					1.06	1.01								
South-east					1.09	1.03								
South-south					0.99	1.00								
North-east					0.98	0.99								
North-west					0.96	0.99								
North-central					0.89	0.97								
Self- employed							0.99	1.00						
Publicly-employed							1.15	1.04						
Privately-employed							0.88	0.97						
Unemployed							0.97	0.99						
Commercial vehicle									1.20	1.04				
Comprehensive									0.91	0.97				
Third party									1.35	1.07				
Motor cycle									0.53	0.92				
Theft											1.53	1.11		
Collision											0.98	1.00		
Accident											1.13	1.03		
Vandalisation											0.61	0.91		
Others											0.74	0.94		
Individual													1.11	1.02
Companies													1.37	1.08
Government													0.69	0.94
All account													0.83	0.96

Source: Researcher's computation 2016

CHAPTER FIVE

SUMMARY CONCLUSION AND RECOMMENDATIONS

5.1 Summary of Findings

This study sought to develop a risk-adjustment model for the motor insurance risks. The model is expected to account for the varying levels of risk using individual socio-demographic characteristics as well as motor risk factors. The main purpose of the model is to provide the motor insurance scheme with incentives to produce efficient services by minimizing risk selection so that motor insurance products in a competitive market can be priced on the basis of a risk-based tariff regime, and also insurance service providers can compete on the basis of quality service and sound tariff administrative efficiency which will ideally improve the profitability of companies, while encouraging improved behaviour on the part of drivers rather than on the ability to select risk. This will be achieved by rewarding motor insurers and policyholders equitably and fairly for the risks they assume and protect the financial sustainability of the insurance market. Also, the model facilitated the consolidation of the present and historical data of the Nigeria Insurance Industry database management system by providing pathways for common analysis approach to enable the sharing of experiences and useful data which will improve pricing capabilities, effective administrative information, actuarial valuation and in-depth statistical analysis. The research findings arising from this study indicated the following:

1. There is evidence that automobile liability claims are heavily tailed and highly peaked showing that the data is significantly non-normal, suggesting the suitability of generalized linear modelling.

2. The descriptive statistics on motor claims shows that claim frequency and cost decreases on the average with age initially but then increases along the age group, which support the fact noted in studies such as McKnight and McKnight (1999, 2003), Kelly and Nielson (2006) that younger drivers on the average have larger claims because of less driving experience and taking more risks, while older individual on the other hand are riskier drivers due to a deterioration of their cognitive and sensory skills.
3. Motor insurance policyholders who resides in the southern region of the country are riskier than those in the northern region
4. The influence factors of the claim cost are similar to the factors corresponding to the frequency of claims, fact that refutes the assumption suggested by actuarial literature regarding the separate analysis of these two components of motor risk premium
5. Female policyholders has higher claim costs and tends to report more claim than their male counterpart
6. From the computed risk scores displayed, a variation of risk scores across age group was discovered and high regional differences noted for motor insurance premium which suggest the need for calculation of differentiated premium within the insurance portfolio so that each insured pays the same reasonable insurance premium to insured with similar risk profile.
7. The regression results revealed that the age of the insured, the gender type, the geographical region where the insured resides, the profession of the insured, the type of product, the nature of the loss type and the customer type significantly predict the frequency and cost of automobile claims occurrence.

5.2 Conclusion

The basic idea of the entire process of non-life insurance pricing comprises in establishing an equitable premium or a tariff method to be paid by the insured to the insurer in exchange for the risk of contingent transfer. Problems caused by risk selection can lead to a number of problems for regulatory authorities, insurers as well as the society at large; more so for developing insurance markets like the Nigeria insurance market, which is bedevilled by unprincipled underwriting where pricing is tariff-driven without sufficient proof or statistics to back up the adequacy of charges. Chief amongst these problems is that aggregate motor claim costs can increase for portfolios which can undermine the solvability of the insurance company business offering multiple coverage when risk selection occurs.

Using historical claims data, it was established that the claims data from automobile insurance scheme is highly peaked and heavily-tailed and vary significantly among age groups, gender, occupation, nature of loss, geographical region, product type and customer type. This demonstrated that the usual normal regression based model for risk adjustment might not be adequate for the data coverage and risk adjustment. The use of generalized negative binomial and gamma regression models to fit automobile claims data and risk-based adjustment model to establish fair and equitable risk premium rates is suggested as it will assist in appropriate premium determination, mitigate the impact of potential adverse selection and stabilize premiums in the individual and aggregate portfolio.

5.3 Recommendations

i. For the insurance companies to be able to reliably estimate their future profits or losses they have to first accurately estimate the burden of costs that precisely reflects the risk profile, hence a risk-based adjustment pricing should be employed to calculate accurately the average expected loss and charge adequate price for motor insurance accordingly.

ii. There is need for insurance companies to carry out more research about the expected future claim amounts in their regions of location. This is because the study has acknowledged that future claim frequency and amounts are dependent on a number of factors experienced by the policyholder and not just by computational analysis.

iii. Viable and sustainable motor liability insurance needs to be founded on intelligent and risk-related pricing foundations; hence pricing it requires careful research and analysis of the complex function, and a large number of more detailed variables that need to be properly established and actuarially monitored.

iv. In order to implement sound tariff, it is also crucial for insurance companies to improve the standards of data collection as this is essential to a well-managed scheme. With better data, tariff can be set more precisely, and more risk sensitive rating factors can be introduced. This should ideally improve the profitability of companies, while encouraging improved behaviour on the part of drivers.

v. Supervisory authorities should ensure the creation of a central database that stores and provides access to the insurance information of policyholders, including claims. Supporting standard practices using data in the same format for everyone in the market and ensuring a high level of transparency are beneficial for the sake of both the process and supervision as it will help to keep companies and individuals from engaging in deleterious market practices. Such a system is useful for identifying uninsured drivers, unifying motor liability insurance practices, preventing fraud, and to establish correct pricing, and creating confidence in the sector.

With all of these in place, there is a sound foundation for a sophisticated actuarial analysis that will enable the pricing of motor insurance risks to be conducted on a sustainable basis. This, in turn, will enable the motor liability insurance schemes to fulfil its proper role in helping developing countries to manage their motor risks and gradually improve their response to the challenge presented by motoring.

5.4 Contribution to Knowledge

i. The study developed a risk-based adjustment model for establishing fair and equitable risk premium rates
ii. A conceptual model has been developed for determining optimal risk-based adjustment premium
iii. This study demonstrates that similar risk factors influences the risk component (claim frequency and costs), fact that refutes the assumption suggested by actuarial literature regarding the separate analysis of these two components of motor risk premium
iv. This study estimated the risk scores, which provides useable profile of policyholders for appropriate risk assessment.

Biblography

Abraham, K. S. (1985). Efficiency and Fairness in Insurance Risk Classification.*Virginia Law Review,* 71(3):403–451.

Adeleke, I. A. &Mesike, G. C. (2012). From data to decisions: developing an innovative industry-wide statistical information system for credible pricing. *Elixir Statistics,* 50, 10206-10209

Adeleke, I. A &Mesike, G. C. (2015). An innovative industry-wide credible pricing model using statistical information system.*Nigeria Journal of Management Studies,* 15(1), 169 – 176.

Adeyemi M (2005). "An overview of the insurance act 2003". issues in merger and acquisition for the insurance industry. Ezekiel. O.C. (ED). Being Proceeding of the 2003 NIA Workshop on Insurance ACT 2003, Nigeria Insurance Association.61 -78.

Agbonkhese, O, Yisa, G. L, Agbonkhese, E. G, Akanbi, D. O, Aka, E. O &Mondigha, E, B (2013) Road Traffic Accidents in Nigeria: Causes and Preventive Measures, *Civil and Environmental Research,* 3(13),

Aggoun, L &Benkherouf, L. (2006).'An Insurance Model with Bonus-Malus System'. *Paper presented at International Mathematical Forum,* 1(26) 1255-1272.

Ajemunigbohun, S. S. &Oreshile, A. S. (2014) Risk attitude and demand for motor insurance: An examination of selected motorist in Lagos state, Nigeria, *Developing Country Studies*, 4(21), Ajne, B. (1975). A note on the multiplicative ratemaking model.*ASTIN Bulletin,* 8(2), 144–153.

Akintayo, L.A. (2004).*Introduction to general insurance underwriting.* 2nd ed. Lagos: CSS Bookshops Limited.

Akerlof, G. A. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics, 84* (3), 488–500.

Antonio, K., Frees, E. W & Valdez, E. A. (2010). A multilevel analysis of intercompany claim counts. *ASTIN Bulletin*, 40(1), 151-177.

Antonio, K. & Valdez, E. A.(2012). Statistical concepts of a priori and a posteriori risk classification in insurance.*AstA Advances in Statistical Analysis,* 96, 187 -224.

Arbous, A. G, Kerrich, J. E (1951).Accident statistics and the concept of accident proneness. *Biometrics,* 7(2), 340 - 342.

Arosanyin, G. T, Olowosulu, A. T &Oyeyemi, G. M. (2012). An examination of some safety issues among commercial motorcyclists in Nigeria: a case study, *International Journal of Injury Control and Safety Promotion*, DOI:10.1080/17457300.2012.686040

Asokere, A. S. &Nwankwo, S. I. (2010).*Essentials of insurance: A modern approach*. Lagos: Fevas Publishing.

Bailey, A. L. (1945). A generalized theory of credibility.*Proceedings of the Casualty Actuarial Society,* 32, 13-20.

Bailey, A. L. (1950): Credibility procedures, La Place's generalization of Bayes' rule, and the combination of collateral knowledge with observed data. *Proceedings of the Casualty Actuarial Society,* 37, 7 - 23. Discussion in 37, 94 - 115.

Bailey, R. A. (1963). Insurance rates with minimum bias.*Proceedings of the Casualty Actuarial Society*, (93):4–11.

Bailey, R. A. & Simon, L. J. (1960). Two studies in automobile insurance ratemaking. *ASTIN Bulletin*, 1(4), 192 – 217.

Behan, D. F. (2009). "Statistical credibility theory" *Proceeding of Southeastern Actuarial Conference,*

Bichsel, F (1964).Erfahrungs-Tarifierung in der Motorfahrzeug-halfplichtversicherung,*Milleilungen der VereinigungSchweizerischerVersicherungsmathematiker*, 119-129.

Bichsel, F. & Straub, E. (1970).Erfahrungstarifierung in der Kollektivkrankenversicherung, unpublished manuscript.

Boland, P. J (2007).*Statistical and probabilistic methods in actuarial science.* New York: Chapman & Hall/CRC.

Borgan, O., Hoem, J. M.&Norberg, R. A. (1981).A nonasymptotic criterion for the evaluation of automobile bonus systems.*Scandinavian Actuarial Journal,* 165-178

Boucher, J. P, Denuit, M. &Guillen, M. (2008) Risk classification for claims counts- A comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal,* 11(4), 110-131.

Boucher, J. P. &Guillen, M. (2009).A survey on models for panel count data with applications to insurance. RACSAM, 103 (2), 277–294.

Buhlmann, H. (1964).‘‘OptimalePramienstufensysteme,’’ *Milleilungen der Vereinigung SchweizerischerVersicherungsmathematiker*, 193–213.

Buhlmann, H. (1967). Experience rating and credibility.*Astin Bulletin,* 4(2), 199 - 207.

Buhlmann, H. (1969). Experience rating and credibility.*Astin Bulletin,* 5(3), 157-165.

Bühlmann, H. and Gisler, A.(2005). *A course in credibility theory and its applications*, Berlin: Springer Universitext.

Buhlmann, H. & Straub, E. (1970*). Glauburdigkeit fur Schadensatze. Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 70, 111-133.

Burns, N. & Grove, S. K. (1993).*The practice of nursing research: Conduct, critique and utilization.* 2nd ed. Philadelphia: Saunders.

Cameron, A. C. &Trivedi, P. K. (1998).*Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Centeno, M. L. & Andrade e Silva, J. M. (2002). Optimal bonus scales under path-dependent bonus rules, *Scandinavian Actuarial Journal,* 2, 129-136.

Charpentier, A. &Denuit, D. (2004).Mathématiques de l'Assurance Non-Vie, Tome I: Principe fondamentaux de théorie du risqué. Economica, Paris.

Charpentier, A. &Denuit, M. (2005).Mathématiques de l'assurance Non-Vie, Tome II: Tarificationetprovisionnement. Economica, Paris.

Chiappori, P.A., Jullien, B., Salanié, B. &Salanié, F. (2006). Asymmetric information in insurance: general testable implications. *RAND Journal of Economics,* 37(1), 783-798.

Crocker, K. J. & A. Snow. (1986). The efficiency effects of categorical discrimination in the insurance industry. *Journal of Political Economy*, 94(2), 321–344.

Dannenburg, D. (1994). Some results on the estimation of the credibility factor in the classical B¨uhlmann model. *Insurance: Mathematics and Economics,* 14(1), 39–50.

David, M. &Jemna, D. (2015).Modelling the frequency of auto insurance claims by means of Poisson and negative binomial models.*Scientific Annals of the "AlexandruIoan Cuza" University of Iasi Economic Sciences,* 62(2), 151-168

David, M. (2015a). Automobile insurance pricing using generalized linear models. *Procedia Economics and Finance*, 20, 147-156

David, M. (2015b). A review of theoretical concepts and empirical literature of non-life insurance pricing, *Procedia Economics and Finance*, 20,157-162

Delaporte, P. (1965). ''Tarification du risqueindividueld'accidents par la prime modele´esur le risque,'' *ASTIN Bulletin* 3, 251–271.

Denuit, M. (2003).Tarification Automobile surdonnées de panel, *ASTIN Bulletin,* 36(1), 51-81.

Denuit, M. (2006). An actuarial analysis of the french bonus-malus system. *Scandinavian Actuarial Journal,* 20(5), 247-264.

Denuit, M. & Lang, S. (2004). Nonlife ratemaking with bayesian GAM's. *Insurance: Mathematics and Economics*, 35(3), 627-647.

Denuit, M. Xavier, M. Pitrebois, S. &Walhin, J.F. (2007).*Actuarial modelling of claim counts: risk classification, credibility and bonusmalus scales*. Chichester: John Wiley & Sons.

De Vylder, F. &Goovaerts, M. J. (1992).Estimation of the heterogeneity parameter in the B¨uhlmann-Straub credibility theory model.*Insurance: Mathematics and Economics,* 10(4),233–238.

Dionne, G. &Vanasse, C. (1989). A generalization of automobile insurance rating models: The negative binomial distribution with a regression component. *ASTIN Bulletin,* 19, 199- 212.

Dionne, G. &Vanasse, C. (1992).Automobile insurance ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics,* 7, 149 – 165

Dionne, G. Michaud, P-C.&Pinquet.J. (2012).A review of recent theoretical and empirical analyses of asymmetric information in road safety and automobile insurance. Centre Interuniversitariesur le Risque, les PolitiquesÉconomiques et l’Emploi. Working Paper 12- 04.

Doherty, N. (1981). Is rate classification profitable.*The Journal of Risk and Insurance*, 48(2), 286–295.

Ellis, R. (1983).*Insurances of transportation.*Cambridge: The Burlington Press (Cambridge) Limited.

Etgar, M. (1975).Unfair price discrimination in public liability insurance and the reliance on loss ratios.*The Journal of Risk and Insurance*, 42(4), 615–624.

Ezekiel, O. C. (2005). “The nigerian insurance market in the context of the insurance act 2003”. Ezekiel, O.C. (ED). Issues in merger and acquisition for the insurance industry.Being Proceeding of the 2003 NIA Workshop on Insurance ACT 2003. Nigeria Insurance Association.61 -78.

Frangos, N. E &Vrontos, S. D. (2001).Design of optimal bonus-malus systems with a frequency and a severity component on an individual bases in automobile insurance.*Astin bulletin,* 33, 1 – 22

Frees, E. W. (2010).*Regression modeling with actuarial and financial applications*. Cambridge: Cambridge University Press.

Gilde, V. &Sundt, B. (1999) “On bonus systems wiyh credibility scales”.*Scandinavian Actuarial Journal*, 13-22

Gonulal, S. O. (2009). *Motor third-party liability insurance in developing countries: Raising awareness and improving safety.* Washington: The International Bank for Reconstruction and Development / The World Bank.

Goulet, V. (1998). Principles and application of credibility theory. *Journal of Actuarial Practice*, 6(1-2), 5–62.

Gourieroux, C., Montfort, A. &Trognon, A. (1984a). Pseudo maximum likelihood methods: theory, *Econometrica*, 52, 681-700.

Gourieroux, C., Montfort, A. &Trognon, A. (1984b), Pseudo maximum likelihood methods: application to Poisson models, *Econometrica*, **52**, 701-720.

Gourieroux, C. &Jasiak, J. (2004).Heterogeneous model with application to car insurance.*Insurance: Mathematics and Economics*, 34(2), 177- 192.

Gourieroux, C. &Jasiak, J. (2007).*The econometrics of individual risk:credit, insurance and marketing.* New Jersey: Princeton University Press.

Greenwood, M. & Yule, G. (1920).An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents.*Journal of the Royal Statistical Society* Series, 83(2), 255–279.

Haberman S.& Renshaw A.E. (1998). *Actuarial applications of generalized linear models. instatistics in Finance*, Hand D.J. & Jacka S.D. (eds), London: Arnold.

Hastings, N. A. J. (1976). Optimal claiming on vehicle insurance.*Operational Research Quarterly journal,* 27, 805-813.

Hausman, N. A. J., Hall, B. H. &Griliches, Z. (1984). Econometric models for count data with an application to the patents – r&d relationship. *Econometrica,* **52**, 909-938.

Heras, A., Villar, J. L. & Gil, J. A. (2002).“Asymptotic fairness of bonus-malus systems and optimal scale of premiums.”*Geneva Papers on Risk and Insurance*, 27, 61-82.

Hogg, R. V.&Klugman, S. A. (1984).*Loss Distributions*. New York:Wiley.

Hoy, M. (1982). Categorizing risks in the insurance industry.*The Quarterly Journal of Economics,* 97(2):321–336.

Ibiwoye, A &Adeleke, I. A. (2011).Markovian framework for effective bonus-malus rating system in nigeria. *Journal of Scientific Research & Development,* 13, 62-74.

Ibiwoye, A., Adeleke, I. A. &Aduloju, S. A. (2011). Quest for optimal bonus-malus in automobile insurance in developing economies: an actuarial perspective. *International Business Research,* 4(4), 74-83.

International Monetary Fund. (2013). Financial sector assessment program update- detailed assessment of observance of insurance core principles, International Monetary Fund (IMF) report no. 13/145

Ismail, N. &Jemain, A. A. (2006).A comparison of risk classification methods for claim severity data.*Journal of Modern Applied Statistical Methods,* 5(2), 513–528.

Jegede, M. I. (2005). “A comprehensive analysis of the insurance act 2003 and its implications on the insurance business environment” Ezekiel, O. C. (ED), Issues in merger and acquisition for the insurance industry.A Proceeding of the NIA Workshop on Insurance ACT 2003. Nigerian Insurance Association, 61 -78

Jong, P. & Heller, G. (2008).*Generalized linear models for insurance data*. Cambridge: Cambridge University Press.

Jørgensen, B. (1997). *The theory of dispersion models*, London: Chapman and Hall.

Jørgensen, B. & Paes de Souza, M. C. (1994). Fitting Tweedie's compound Poisson model to insurance claim data. *Scandinavian Actuarial Journal*, 1: 69–93.

Jung, J. (1968). On automobile insurance ratemaking.*ASTIN Bulletin,* 5(1), 41–48.

Kaas, R., Goovaerts, M., Dhaene, J. &Denuit, M. (2009).*Modern Actuarial Risk Theory: Using R.* 2nd ed. Berlin: Springer.

Karm, P., Bodia, B. S. &Garg, M. C. (2007). “*Insurance Management Principle and Practices*”.New Delhi: Deep and Deep Publication private Limited.

Kelly, M. & Nielson, N. (2006). Age as a variable in insurance pricing and risk classification. *The Geneva Papers: Issues and Practice*, 31(2), 212–232.

Klugman, S. A., Panjer, H.H. &Willmot, G.E. (2008). *Loss models: from data to decisions*, 3rd ed., New Jersey: John Wiley.

K¨ohne, T. (2011).ZurPreissituationimdeutschenKfz-Versicherungsmarkt, StudieimAuftrag der Direct Line Versicherung AG. (Available at: http://blog.directline.de/wpcontent/uploads/Studie-ZurPreissituation-im-deutschen-Kfz-Versicherungsmarkt.pdf, accessed August 26th 2016).

Kolderman, J. &Volgenant, A. (1985).Optimal claiming in an automobile insurance system with bonus-malus structure.*Journal of the Operational Research Society,* 36: 239-247.

Kothari, C. R. (2004). *Research Methodology: Methods and Techniques, 2nd ed*. New Delhi: New age International Publishers.

Lass, D., Schmeiser, H. & Wagner, J. (2016).Empirical findings on motor insurance pricing in Germany, Austria, and Switzerland.*The Geneva Papers on Risk and Insurance-issues and practice,* 41(3), 398-431.

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistics Society B*, 58(4), 619–678.

Lemaire, J. &Zi, H. (1994). “High deductibles instead of bonus-malus, can it work?” *ASTIN BULLETIN*, 24(1), 75-88.

Lemaire, J. (1976). Driver versus company, optimal behaviour of the policyholder, *Scandinavian Actuarial Journal,* 59, 209-219.

Lemaire, J. (1977). La soif du bonus, *The ASTIN Bulletin,* 9, 181-190.

Lemaire, J. (1979). How to define a bonus-malus system with an exponential utility function. *The ASTIN Bulletin,* 10, 274-282.

Lemaire, J. (1985). *Automobile Insurance: Actuarial Models.* Huebner International Series on Risk, Insurance and Economic Security. Boston: Kluwer Academic Publishers.

Lemaire, J. (1988).A comparative analysis of most European and Japanese bonus-malus systems.*Journal of Risk and Insurance,* LV, 660-681

Lemaire, J. (1991). Negative binomial or Poisson-inverse Gaussian?*ASTIN Bulletin 21*, 167–168.

Lemaire, J. (1995). *Bonus-malus systems in automobile insurance*. Massachusetts: Kluwer Academic Publishers.

Lemaire, J. (1998). Bonus-Malus systems: The European and Asian approach to merit-rating. *North American Actuarial Journal,* 2(1),

Loimaranta, K. (1972). Some asymptotic properties of bonus systems.*The ASTIN Bulletin,* 6, 233-245.

Longley-Cook, L. H. (1962). An introduction to credibility theory. *Proceedings of the Casualty Actuarial Soc*iety, 49, 194-221

Lundberg, O. (1940). On random processes and their application to sickness and accident statistics.Thesis, Stockholm. Reprinted by Almqvist and Wiksell, Uppsala, 1964

Mahmoudvand, R. &Hassani, H.(2009). Generalized bonus-malus system with a frequency and a severity component on an individual basis in automobile insurance.*ASTIN Bulletin*, *39*, 307–315.

Markov, A. A. (1913). *The calculation of probabilities*.Tip. ImperatorskoiAkademiiNauk, St. Petersburg.

McCullagh, P. (1976). Analysis of ordered categorical data.(Unpublisheddoctoral thesis) Imperial college, London.

McCullagh, P. &Nelder, J. (1989).*Generalized linear models*, Boca Raton: Chapman and Hall.

McKnight, J. A. & McKnight, S. A. (1999).Multivariate analysis of age-related driver ability and performance deficits. Accident Analysis and Prevention, 31(5), 445–454.

McKnight, J. A. & McKnight, S. A. (2003). Young novice drivers: careless or clueless? Accident Analysis and Prevention, 35(6), 921–925.

Mesike, G. C., Adeleke, I. A &Ibiwoye, A. (2012). Predictive actuarial modeling of health insurance claims costs. *International Journal of Mathematics and Computation*, 14(1), 34-45.

Mesike, G. C. &Adeleke, I. A. (2012). From data to decisions: developing an innovative industry-wide statistical information system for credible pricing. *Elixir International Journal,* 50, 10206 -10209

Mesike, G. C. &Adeleke, I. A. (2015).An empirical bayesian credibility estimation of future aggregate claims for non-life insurance.*Nigeria Journal of Management Studies,* 13(1), 32-34.

Millenhall, S.J. (1999). A systematic relationship between minimum bias and generalized linear models. *Proceedings of the Casualty Actuarial Society,* 86, 393–487.

Miller, R. B. & James C. H. (1975). Insu*rance Credibility Theory and Bayesian Estimation, in*: Paul, M. K., ed. *Credibility Theory and Applications,* New York: Academic Press, 249-270.

Mowbray, A. H. (1914): *How extensive a payroll exposure is necessary to give a dependable pure premium?* Proceedings of the Casualty Actuarial Society, 1, 24-30.

Nath, D, C. &Sinha, P. (2014).A markovian study of no claim discount system of insurance regulatory and development authority and its application.*Thailand Statistician,*12(2), 223-236.

National Insurance Commission. (2015). Annual report and audited accounts. Abuja, Nigeria

National Bureau of Statistics.(2014). *Annual Abstracts of Statistics.*National Bureau of Statistics.

Nelder, J. A. & Verrall, R. J. (1997). Credibility theory and generalized linear models. *ASTIN Bulletin*, 27(1), 71–82.

Nelder, J. A. &Wedderburn, R.W. M. (1972).Generalized linear interactive models.*Journal of the Royal Statistical Society*, A 135(3), 370-384.

Newbold, E. M. (1927). Practical application of statistics of repeated events, particularly to industrial accident.*Journal of the Royal Statistical Society*, 90(3), 487 – 547

Ngwuta, O. (2007). *Modern insurance business management.* Lagos: Nigerian Insurers Association.

Nigerian Insurers Association. (2006). *Nigeria motor tariff*, Lagos: Nigerian Insurers Association.

Nigerian Insurers Association.(2013).Nigerian insurance digest.*A Statistical Journal of Nigerian Insurers Association,*15-18

Nigerian Insurance Industry Database (2015).Retreived March 15, 2016, from http:/www.niid.org

Norberg, R. (1976**).**A credibility theory for automobile bonus system.*Scandinavian Actuarial. Journal*, 92–107.

Norberg, R. (2004). *Credibility theory,* Department of Statistics, London School of Economics, London WC2A 2AE,

Norris, J. R. (1998). "*Markov chains*". London: Cambridge University Press.

Obasi, N. (2010). Policies, challenges, reforms and nigerian disposition to insurance contracts. *The Fronteira Post,* 1-6.

Ohlsson, E. & Johansson, B. (2006a). Exact credibility and Tweedie models. *ASTIN Bulletin*, 36(1), 121–133.

Ohlsson, E. (2008). Combining generalized linear models and credibility models in practice. *Scandinavian Actuarial Journal,* 2008(4), 301-314

Oke, M. O. (2012). Insurance sector development and economic growth in Nigeria.*African Journal of Business Management,* 6(23), 7016-7023.

Oxford Business Group. (2010). *The Report: Nigeria 2010*

Ozioko, E. M. (2007). *General insurance companion.* Lagos: Chimoke Printing & Publishing Company Limited.

Park, S., Lemaire, J., & Chua, C. T. (2009). Is the design of bonus-malus systems influenced by insurance maturity or national culture? Evidence from Asia. *The Geneva Papers*, *35*(S1), 7-27.

Pesonen, E. (1962). A numerical method of finding a suitable bonus scale.*ASTIN Bulletin*, 2(1), 102-108.

Picard, P. (1976). Generalisation de l'etudesur la survenance des sinistres en assurance automobile.*Bulletin Trimestriel de l'Institute des ActuairesFrancais*, 204-267.

Pinquet, J. (1997). Allowance for cost of claims in bonus-malus systems.*ASTIN Bulletin,* 27(1), 33-57.

Pitrebois, S., Denuit, M. &Walhin, J.F. (2003). Fitting the Belgian bonus-malus system.*Belgian Actuarial Bulletin*, 3(1), 58-62.

Pitrebois, S.,Denuit, M. &Walhin, J. F. (2005).Bonus-malus system with varying deductibles.*ASTIN Bulletin*, 35, 261-274

Pratt, K. (2010). Pricing.*The Journal of the Chartered Insurance Institute,* 19–21.

Rothschild, M. &Stiglitz, J. (1976). Equilibrium in competitive insurance markets: an essay on the economics of imperfect information, *Quarterly Journal of Economics* 90, 629-650.

Savage, L. J. (1954). *The Foundation of Statistics*. New York: Dover Publications.

Schmidt-Gallas, D. &Lauszus, D. (2005).*MehrMarkt: Die Pricing-Prozesse der Versichererm¨ussenbesserwerden, Ergebnisse der Branchenbefragung und Empfehlungenf¨ur das Management*. Bonn: Simon Kucher& Partners.

Seal, H. (1982). Mixed Poisson – an ideal distribution of claim numbers? *Insurance: Mathematics and Economics 8*, 35–46.

Sipulskyte, R. (2012). Development of a motor vehicle classification scheme for a New Zealand based insurance company. *New Zealand Society of Actuaries Conference*.

Sundt, B. (1988). Credibility estimators with geometric weights.*Insurance: Mathematics and Economics,* 7(2), 113–122.

Sundt, B. (1987). Two credibility regression approaches for the classification of passenger cars in a multiplicative tariff. *ASTIN Bulletin*, 17(1), 42–70.

Swiss Re. (2010). *Sigma report* #2: World insurance in 2009. Available at: http://media.swissre.com/documents/sigma2_2010_en.pdf.

Trieschmann, J. S., Hoyt, R. E. &Sommer, D. W. (2005). *Risk management and insurance.* 12th ed. USA: South-Western Thomson Corporation.

Ujunwa, A. &Modebe, N. J. (2011). Repositioning insurance industry for operational efficiency:the Nigeria case. *Journal of Applied Finance and Banking,* 1(3), 15-32.

Ukoji, V. N. (2014). Trends and patterns of fatal road accidents in Nigeria.*Institute for Research in Africa, IFRA-Nigeria working papers,*35.

Vepsäläinen, S. (1972).Applications to a theory of bonus systems.*The ASTIN Bulletin,* 6, 212-221.

Vos, M. Hougaard, C. & Smith, A. (2011). Opportunities for insurance inclusion in Nigeria: Exploring the potential in the Nigeria insurance market using data from the EFInA Access to Finance in Nigeria 2010 survey. Centre for Financial Regulation and Inclusion

Whitney, A. W. (1918). *The theory of experience rating*. Proceedings of the Casualty Actuarial Society 4, 274-292.

Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial.*Scandinavian Actuarial Journal 87*, 113–127.

World Bank (2008).Nigeria access to finance diagnostic.Presentation to the FSS2020 Insurance Implementation Committee.

Young, V. R. & De Vylder, E. F. (2000).Credibility in favor of unlucky insureds.*North American Actuarial Journal,* 4(1), 107–113.

Appendix 1
SOCIO DEMOGRAPHIC

AGE GROUP

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Less than 24 years	4472	28.0	28.0	28.0
	24 - 30 years	2318	14.5	14.5	42.5
	31 - 60 years	7730	48.4	48.4	90.9
	61 years and Above	1458	9.1	9.1	100.0
	Total	15978	100.0	100.0	

GENDER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	9672	60.5	60.5	60.5
	Female	4958	31.0	31.0	91.6
	Entity	1248	7.8	7.8	99.4
	Joint Gender	100	.6	.6	100.0
	Total	15978	100.0	100.0	

LGA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	FCT	976	6.1	6.1	6.1
	South West	13144	82.3	82.3	88.4
	South East	327	2.0	2.0	90.4
	South South	981	6.1	6.1	96.6
	North East	57	.4	.4	96.9
	North West	296	1.9	1.9	98.8
	North Central	197	1.2	1.2	100.0
	Total	15978	100.0	100.0	

OCCUPATION

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Self	1340	8.4	8.4	8.4
	Public	6078	38.0	38.0	46.4
	Private	8210	51.4	51.4	97.8
	Unemployed	350	2.2	2.2	100.0
	Total	15978	100.0	100.0	

PRODUCT NAME

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Commercial Vehicle	2783	17.4	17.4	17.4
	Comprehensive	12520	78.4	78.4	95.8
	Third party	641	4.0	4.0	99.8
	Motor Cyc le	34	.2	.2	100.0
	Total	15978	100.0	100.0	

LOSS TYPE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Theft	306	1.9	1.9	1.9
	Collision	14261	89.3	89.3	91.2
	Accident	391	2.4	2.4	93.6
	Vandalisation	767	4.8	4.8	98.4
	Others	253	1.6	1.6	100.0
	Total	15978	100.0	100.0	

CUSTOMER TYPE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Individual	13283	83.1	83.1	83.1
	Companies	2611	16.3	16.3	99.5
	Government	77	.5	.5	100.0
	All account	7	.0	.0	100.0
	Total	15978	100.0	100.0	

Appendix 2

RISK SCORES