UNIVERSITY OF CAPE COAST

A MARKOV-MODULATED TREE-BASED GRADIENT BOOSTING MODEL FOR AUTO-INSURANCE RISK PREMIUM PRICING IN GHANA

BY

DENNIS ARKU

Thesis submitted to the Department of Statistics of the School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfilment of the requirements for the award of Doctor of Philosophy degree in Statistics

JULY 2018

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:.. Date: 25/02/19

Name: Dennis Arku

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature:.................................. Date: 25/02/19

Name: Dr. Kwabena Doku-Amponsah

Co-Supervisor's Signature: .... Date: 27/02/2019

Name: Dr. Nathaniel Kwamina Howard

ABSTRACT

In most sub-Saharan African countries, the mechanism for pricing auto insurance policies is tariff based. This means that the key factor that influences price changes is usually based on margins, regulation and legislative dynamics. Additionally, where pricing is risk based, analysis has in most cases focused on internal historical data. These policy regimes have led to unfair price distortions among policyholders and have increased risk of portfolios for most insurance companies. In this study we consider historical risk and location risk that is influential to loss cost. The study develops a Markov-modulated Tree-based Gradient Boosting (MMGB) model for pricing auto-insurance premiums. The Markov-modulated Tree-based Gradient Boosting is a Tweedie generalized model-based pricing algorithm with a compound-Poisson distribution whose rate varies according to accident risk in a Markov process. Thus, the study extends the existing premium pricing framework by integrating both historical and location risk into the main pricing framework. The study applies the model to a motor insurance data set from Ghana. The results show that the proposed method is superior to other competing models since it generates relatively fair premium predictions for the non-life auto-insurance companies, helping to mitigate more the insured risk for the firm and the industry.

KEY WORDS

Accident risk

Auto-insurance pricing

Gradient boosting

Markov-modulated

Risk modelling

Risk premium

ACKNOWLEDGEMENTS

First, I am thankful to the God of Abraham, Isaac and Jacob, the creator of the universe, for granting me the mental fortitude and good health throughout this journey and the grace to finish this work successfully.

I am also thankful to my supervisors, Dr Kwabena Doku-Amponsah and Dr Nathaniel Kwamina Howard for their unflinching support and guidance at various stages of the work. Special mention to Dr Isaac Baidoo, but for him this work would have been abandoned. He has been of an immense help to me in getting access to data.

My thanks also go to my brother, Sampson Arku who has been exceptionally supportive emotionally and economically. He ensured that all the necessary tools I needed for the success of my education were provided. My profound gratitude also goes to my loving wife for taking care of my home and allowing me absolute concentration on my studies.

Finally, to all lecturers at the Department of Statistics and Actuarial Science, University of Ghana, Legon who encouraged and offered pieces of ideas and advice, I say God richly bless you.

DEDICATION

This research work is dedicated to my mother Felicia Arku and late father S. K. Arku.

TABLE OF CONTENTS

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Brief Introduction

This Chapter presents the introductory section of the study. It starts with the background to the study, the problem statement, the objectives and the theoretical framework. It also presents an overview of the evolution of the insurance industry in Ghana and finally outlines the concept relating to the study.

Background to the Study

The operational process of non-life Insurance assumes different risks profiles for the insured influenced by instabilities within the business environment. Predicting the financial obligation for claims in non-life insurance is quite complicated and usually depends on the structure of insurers' liabilities. Pricing premiums is one of the most critical issues facing the non-life insurance industry across the world. The issue of charging fair premiums or price differentiation among policyholders has been the bane for most insurance companies (Mihaela, 2015). The question has always been, how much premium to allocate to a policyholder to ensure fairness on the part of the insured and how much to allocate to the insured to avoid bankruptcy. For researchers the most critical task in insurance pricing is how to accurately predict the risk of claim and the expected coverage for the insured. The task of modeling claims has been a major challenge because of data structure which is usually highly skewed with many zeroes but high claim severity.

By insurance contract economic risk is transferred from the policyholder to the insurer. Due to the law of large numbers, the loss of the insurance company, being the sum of many comparatively small independent losses is much more predictable than that of an individual. This means that the loss should not be too far from its expected value. This leads us to the generally applied principle that the premium should be based on the expected (average) loss that is transferred from the policyholder to the insurer (Ohlsson & Johansson, 2010).

Several statistical approaches have been used to approximate the risk of accident. In the 1990's, British actuaries introduced generalized linear models (GLMs) as a tool for tariff analysis which was first proposed by Nelder and Wedderburn (1972) and later developed by McCullagh and Nelder, (1989). This has now become the standard approach in most insurance companies of the developed world.

Generalized linear modeling (MacCullagh & Nelder, 1989) employs exponential family distributions where spread and shape are related to the mean. This means that any equation for the mean in terms of risk factors is indirectly also a model for shape and or spread. But the connection is implicit and hence does not permit the explicit flexibility of modeling the spread and or shape in terms of the risk factors. Joint modeling of the mean and dispersion has recently become popular in the statistical literature (Nelder & Lee, 1992; Rigby & Stasinopoulos 1996). It is important to note that, with normal distribution, the variance controls the spread and there is a similar shape for all values of the variance. However, for exponential family distributions, the variance is of the form $\phi Var(\mu)$, where μ is the mean, ϕ is the dispersion parameter and $Var(\mu)$ is a function of the mean that is dictated by the particular distribution. The standard generalized linear model specifies a model for μ and employs a constant ϕ. Joint modeling of mean and dispersion in terms of risk factors is not useful, this is because for many distributions occurring in insurance, variance is not a useful quantity for modeling.

Several other authors have come up with distributions for claims sizes (Hogg & Klugman ,1984; Habberman & Renshaw, 1996). According to Hogg and Klugman (1984), the distribution of total claim amount can be modelled as a mixed discrete-continuous process with a probability mass at zero corresponding to the probability of number of claims, and a continuous right-skewed component to the right of zero for the density of positive claims, corresponding to total of one or more claims. In other instances, models for claim occurrence and claim severity are typically treated separately. This according to the authors is to help identify different factor scenarios affecting claims severity and frequency. The model for claim occurrence produces an estimate for the probability of a claim while the model for claim severity produces an estimate of the expected claim size. The product of these is the total expected policy claims cost known as the risk premium. It has also been argued that the propensity of claim depends on several explanatory variables.

Most empirical models which have been developed are based on GLMs. For instance, Haberman and Renshaw (1996), used GLMs to model total nonzero claims as a function of risk factors. Mihaela (2015) analyzed claims severity and frequency using the GLM technique. The literature on joint modeling of both zero and nonzero is rare. Jorgensen and de Souza (1994) and Smyth and Jorgensen (2002) considered models for claim sizes, that includes zero claims. This modeling framework assumes that claims arrival has a Poisson distribution and claim severity follows a gamma distribution such that the total claim structure could be modelled with Tweedie compound Poisson. Also, Zuanetti, Diniz, and Leite (2006) considered a model without explanatory variables. The authors proposed Poisson distribution for modeling number of claims and lognormal for modeling the claim size. This method is woefully inadequate since it fails to recognize risk factors inherent in claim processes.

Even though Tweedie GLM is often used, it has a major drawback. The major drawback is that link between the variate and covariates that is usually constrained to a linear form is rare in practice. For instance, in auto-insurance, risk of claim is not necessarily inversely related with age (McCartt, Shabanova, & Leaf, 2003; Anstey, Wood, Lord & Smith, 2005). To correct this drawback, various procedures have been proposed. Wood (2006) proposed Generalized Additive Models (GAM) to overcome some of the deficiencies of the GLM such as the linear link to a more general form. However, with GAM the structure of the model must be specified. The main and interaction effects must be specified by the researcher. This often results in specification bias which likely affects the predictive power.

To overcome the deficiencies of GAM, and GLM, Friedman (2001) and Friedman (2002) extended the work of Freud and Schapire (1996), Freund and Shapire (1997), Brieman (1998), Brieman (1999), Friedman, Hastie and Tibshirani (2000) laid the ground work for a new generation of boosted modeling framework to allow for a variety of loss functions. It is an iterative approach where several "weak" models are created and aggregated to form a final prediction model. The framework requires a specification of loss function. Using the connection between boosting and optimization, Ridgeway (2012) proposed an R integrated framework of Gradient Boosting Machine (gbm). Guelman (2012) analyzed and developed an auto insurance loss cost modeling using the gradient boosting machine.

However, this model is insufficient in capturing the data structure since it only implements square error loss function for claims size modeling and binomial deviance for probability of claim. Yang, Qian and Zou (2016) improved on the idea of Friedman by proposing a TDboost model; an important tweak to the procedure specified by Friedman in 2001 with an implementation option of Tweedie-based error loss function. Despite its strong predictive power, its outcome depends on the data generating process and the associated variables. The mechanism also fails to account for external risk factor effects such as environmental risk. More so, for most of these frameworks described the main source of data on which decisions are based use historical claims data generated from the insurance company. An actuary uses this data generated by the company or the industry to determine the model which describes how the claim cost of an insurance policy behaves. The need for statistical methods comes from the fact that the expected loss may vary among policies. This means that rate of accident rate is not the same for all policy holders in the environment in which it operates. More so, once a claim has occurred, the expected damages may also vary among policyholders.

This study seeks to improve on the work of Yang, Qian and Zou, (2016) by obtaining an auxiliary variate x_i closely related to the study variable, claims (y_i). The auxiliary variable is used as proxy to incorporate location risk which is an aspect of latent risk within the TDboost model framework. According to Loisel (2004), the business of an insurance company may have influenced and modulated within a Markovian environment process. Thus, the study proposes a hybrid approach for deriving risk premium for an auto insurance policy by considering both latent and historical risk within the insurance industry. The fast-paced changes in business environment and technological advancement require an all-inclusive dynamic risk treatment, especially in Non-Life insurance. Cramer (1955) (as cited in Djuric (2013)) said that "the goal of risk theory is to provide a mathematical analysis of the fluctuations in the insurance business and to suggest various means of protection against their adverse effects". This motive is what this study seeks to achieve.

Statement of the Problem

The motivation of the study is described in two ways. Firstly, differential pricing or risk-based auto insurance premium pricing has been in existence for a while, at least for most developed countries and some other countries such as India. In risk-based pricing the idea is to charge higher premiums for vehicles that are more likely to get into an accident and therefore higher propensity to claim and a lower for vehicles less likely to be involved in accidents. It is very common to identify premium-modeling framework for most developed countries that are based on generalized linear model in most actuarial literatures.

Such models mainly depend on historical claims data to determine which model describes the claim frequency and size. However, it is not common to encounter studies that delve into situations where internal historical data is augmented with external data to improve predictive power.

Secondly, in most sub Saharan countries, premiums are tariff based. This means that price charges are mainly due to regulatory and economic dynamics of the period, such as inflation, exchange rate as well as other regulatory changes. For instance, in Ghana, the current premium computation regime is based on 2015 underwriting and implementation guideline issued by Ghana Insurers Association in consultation with General Insurance Council. The premium computation is done in five steps:

1. Quote the basic premium indicated in the tariff class
2. Add cubic capacity and vehicle age loadings
3. Apply the approve discounts
4. Add other underwriting loading factors (where applicable) to arrive at the office premium
5. Add NIC contributions, NHIS, NRSC and ECOWAS levies to arrive at the total premium chargeable

The risk loading considerations for age of vehicles have been classified as follows: 1 to 5 years free, 5 to 10 years 5% of base premium indicated in the tariff, above 10 years is 7.5%. For cubic capacity loading: up to 1600 is free, 1600 to 2000 is 5%, above 2000 is 10%. Extra seat loading: vehicles with seating capacities of up to 5 shall or do not attract seat loadings. For vehicles with seats above 5, the extra loadings are as follow: private cars charged Ghc 5 per seat above 5 seats. For hiring vehicles and other commercial vehicles, Ghc

8 per seat above 5 seats. Others such as underwriting discounts, fleet rebates, ECOWAS peril as well as other statutory charges are considered. The guideline is firmly controlled, and the industry is compelled to follow the guideline strictly and uniformly. Failure of which sanction is meted out to members who undercut to gain unfair advantage.

This pricing theory is fraught with a lot of challenges because the risk captured in this pricing framework is inadequate. The price charge may not cover the expected risk leading to unpalatable outcomes. For instance, PricewaterhouseCoopers report (2010) indicated that the insurance claim ratio relating to third party insurance is very high for commercial vehicles. For mini buses and taxis, the claims always outrun the tariffs which make the insurance company digging into other funds (compensation fund) to meet the liabilities. The statistics indicate that for every cedi of premium paid, the claim ratio for mini buses on third party was 92.26%, maxi buses were 62%, whilst 44.9% was for taxis. For comprehensive taxis, it was 118.52%. This means that for every 100 premiums paid the insurance company would have to use up 92.26 for mini buses and incurred a loss of 18.52 for taxi on comprehensive at the end of the year. Claims ratio is a measure of underwriting efficiency. It indicates the gross premium available to contribute towards profit. The lower the ratio the better the underwriting efficiency.

In addition, according to the National Insurance Commission report (2016) the claims ratio which is a measure of how well an insurance company pays claim suggest that about 84% of the insurance companies are underperforming in terms of payment of claims compared to acceptable standards (60-80%). This may be due to high claims, underpricing or excessive expenditure. More so, the combined ratio which is a single measure of profitability, discounting other investments, suggest that the industry average has worsened considerably from 7% in 2011 to 19% in 2016. Out of the 25 non-life insurance companies, about 80% of these companies were worse off in 2016. This means that the combined ratio was beyond 100%. The challenges facing the insurance industry include both underpricing and overpricing of premiums due to lack of actuarial pricing framework. Thus the continuous use of the tariff system is probably because there are no efficient models that could adequately capture the encumbrances inherent in the insurance industry data structure in Ghana despite the existence of several models.

It is known that the most efficient way for a reliable predictive modeling is based on well-designed choice of framework and appropriate statistical model, which reflects not only the design of the study but also the characteristics of the data. If the existing distributions do not fit the data well, the implication is that different sets of characteristics are associated with the data and largely differ from that used to build these models. Hence, inference targeted at such models will be out of place and inefficient. The modeling framework envisaged in this study consider two data structures to capture the policy characteristics and the policy's usual operating environment. The study proposes a differential premium pricing strategy based on these two key ingredients for deriving risk premium for an auto insurance policy in Ghana and alternative framework for loss cost modeling in non-life insurance in general. Sound risk evaluation is essential in terms of company profitability. Beyond the company level, risk management is key to sustainable growth of the industry for the development of the broader economy and socio-economic wellbeing within the country (Akotey & Abor, 2013).

Research Objectives

The study seeks to assess the performance of insurance companies in Ghana and develop a model that is robust, risk-based and semi-parametric for pricing auto-insurance premiums.

Specific Objectives

Considering the gaps and challenges, the general objectives has been broken down into three specific objectives:

1. Assess the financial performance of the insurance companies via grey systems theory.
2. To develop a robust, risk-based and semi-parametric pricing model for premium determination
3. To demonstrate the superiority of the model using an underwriting claims data from a local insurance company.

Relevance of the Study

The study examines how to price auto-insurance with practical data from Ghana. Almost all the literature reviewed had used data generated for an insurance company. One hardly finds a modeling framework that utilizes both claim experience and relevant information outside the insurance industry. This study is thus significant in two-fold. The study seeks to contribute to existing literature on insurance pricing. Its usefulness is not only to statisticians but also to researchers and practitioners in a wide variety of fields.

Secondly, as earlier mentioned the continuous use of the tariff system is probably because there are no efficient models that could adequately capture factors inherent in the insurance industry database. This study is therefore useful for the auto-insurance industry in Ghana and could be the bases to consider risk-based premium pricing regime. This will guarantee the avoidance of eventual ruin; the glimpses seen in annual financial reports.

Data Acquisition and Source

The study utilized data from three various sources. Data set was also obtained from National Insurance Commission report from 2012 to 2016 for purpose of analyzing financial performances of the industry.

The data set for practical application of the theory was obtained from two sources. Data was obtained from a major Auto-insurance company in Ghana. The data spans from 2013 to 2016. Auxiliary data was also obtained from National Road Safety Commission. The National Road Safety Commission is a state agency responsible for safety education and accident statistics in Ghana. The data includes records of both vehicle and motor accidents in all the ten regions of Ghana from 1991 to 2015.

Overview of Insurance Industry in Ghana

The history of insurance in Ghana dates to 1920s by courtesy of the British merchants with the establishment of the British of the Guardian Royal Exchange Assurance Ghana Limited in 1924 now known as Enterprise Insurance Ghana (Akotey & Abor, 2013). Towards independence, local insurance companies begun to emerge. The first was Gold Coast Insurance Company which was formed in 1955. Cooperative Insurance society also formed in 1958. Government of Ghana bought Gold Coast Insurance and merged with Cooperative Insurance Society to form State Insurance Company (SIC) and incorporated in 1962 (Abor, 2005). Through structural and legislative changes, non-life insurance companies consisted of 26 as at June 2016.

The Ghanaian insurance industry is governed by National Insurance Commission (NIC) under the Insurance Act 2006 (Act, 724). The object of the NIC as detailed in Act 724 is to ensure effective administration, supervision, regulation and control of the business of insurance in Ghana among others (Act.724). There are three broad categorizations of insurance business namely: Life and Non-life insurance, Reinsurance and Reinsurance broker business and Loss Adjustment business which are briefly described.

Life and Non-Life Insurance

Life insurance is a contract in which the insurer, in consideration of certain premium, either in a lump sum or by other periodical payments (yearly, semi-annually etc.), agrees to pay to the assured. (Gupta, 2011). Non-life or general insurance on the other hand is basically the policy that protects the insured against losses and damages other than those covered by life insurance.

Reinsurance and Reinsurance brokers

This is insurance for insurers. Reinsurance companies share the risk borne by the insurance companies in return for part of the premium paid by the insured. Reinsurance enables a client to get coverage that would be too great for anyone company to assume. That is giving off part or whole of a risk underwritten by one company to one or more companies.

Reinsurance brokers play an integral intermediary role between the insurance company and the reinsurance company. It brings the two entities together for business of mutual benefit.

Loss Adjustment Business

Loss adjustment business is not too prevalent in Ghana. Loss adjusters are independent claim specialists who assist in just and fair settlement of claims, including complex and contentious claims on behalf of the company. They also help policyholders to restore their proper and full working order. Loss adjustment firms investigate at the scene of an incident to establish the causes of the loss and ascertain whether it is covered by the insurance policy. They engage legal experts and forensic scientists as appropriate to execute their businesses. They write reports for the insurer, assess the validity of the claim and recommend appropriate payment.

According to the National Insurance Commission database, as at June 2016, the insurance industry comprised 26 non-life companies, 23 life companies, 3 Reinsurance companies, 70 Broking companies, 1 Reinsurance Broker, 1 Oil and Gas company, 7000 insurance agents.

Limitation and Scope of the Study

There are twenty-six (26) non-life insurance companies in Ghana. Efforts were made to obtain policy and claims data from at least 50% of these companies. However, due to competition and the classified nature of data, only one major company oblige to release data for this purpose. This limitation notwithstanding, given the nature of the industry and the company involve, the analysis and findings are still valid.

Theoretical Framework

Actuarial models are composed of equations that represent the functioning of insurance companies, accounting for the probabilities of the events covered by policies and the costs each event presents to the company (Klugman, Panjer & Willmot, 2004). Events covered by policies are random in nature, hence evaluation of their financial impact requires probabilistic methods. Insurance companies attempt to estimate reasonably priced insurance policies based on the losses reported by policyholders. The estimation considers past data in order to grasp the trends that occurred (Weisberg & Tomberlin, 1982).

Information available to predict the price for a period in the future usually consists of the claim experience for a population or a large sample from the population over a period in the past. Precise estimation may consider many exposures in a data set and a stable claim generation process over time. According to Boland (2006), modeling claims is crucial since a good understanding and interpretation of loss distribution is the back-bone of all the decisions made in the insurance industry regarding premiums and its loadings, expected profits, the reserves necessary to insure profitability and the impact of re-insurance.

The basic probability models for actuarial situations tend to be either continuous or discrete. In both cases the situation calls for counting something which falls under discrete case or paying something which falls under continuous case. In both cases the model does not assume negative values.

Models are characterized by how much information is available; the number of parameters needed to describe a complex model. When simple models are used, it is more likely that each item can be determined more accurately but the model itself may be too superficial. It is required that a model is more likely to be stable across time and settings. A simple model may provide necessary smoothing to offset the irregularities in the data. Complex models, however, can more closely mimic reality. Regarding complex models with many parameters, its specification can more closely match irregularities in the data. This study derived its support from aspect of risk theory; the concept of gradient boosting in the field of statistical learning, Markov theory and credibility theory.

The concept of gradient boosting emerged from the field of machine learning. The basic idea is to boost the predictive accuracy of weak models by aggregating various instances of such weak models into a more accurate predictive framework.

The Markov theory helps to incorporate price differential strategy based on location for each policyholder given the level of policy operational risk for a given period.

Credibility theory is a set of statistical tools which allows an insurer an opportunity to adjust future premiums based on experience on a risk or group of risks. If the experience of a policyholder is consistently better than that assumed to be in the underlying rate, the policyholder may demand rate reduction. The reasoning is that the manual rate is designed to reflect the expected experience of the entire rating class and implicitly assumes that the risks are homogenous. However, no rating system is perfect. There always remain some heterogeneity in the risk levels after all the underwriting criteria are accounted for (Klugman et al., 2004).

The question has always been how much of the difference in experience of a given policyholder is due to random and how one can identify whether the policyholder is really a better or worse than average for the given rating class? In other words, how credible is the policyholder's own experience? Two facts are considered by Klugman et. al. (2004).

The more past information the insurer has on a given policyholder, the more credible the policyholders own experience ceteris paribus. In like manner, in group insurance the experience of larger groups is more credible than that of smaller groups. Competitive considerations may force the insurer to give full (using the past experience of the policyholder only and not the manual rate) or near full credibility to a given policyholder to retain the business. Buhlmann (1967) provided a statistical framework within which credibility theory has developed and flourished.

Several aspects of credibility theory have been proposed by several authors, such as limited fluctuation credibility theory, greatest accuracy theory, full credibility and partial credibility briefly explained below.

Limited fluctuation credibility theory (as cited in Klugman et al., 2004) was initially suggested by Mowbray (1914) and improved by Buhlmann (1967) and the greatest accuracy credibility theory proposed by Whitney (1918). The problem of limited fluctuation theory is formulated as follows. Suppose that a policyholder has experienced R_j claims in the past, where $j \in \{1,2,3,\ldots,n\}$.

Suppose that $E(R_j) = \omega$, this quantity would be the premium to charge net of expenses and other loadings. Suppose $Var(R_j) = \sigma^2$ for all j. The past experience could be summarized by an average $\overline{R} = (R_1 + R_2 + ... + R_n)/n$. It is known that $E(\overline{R}) = \omega$ and if the R_j are independent, then $Var(\overline{R}) = \sigma^2 / n$. The insurer's goal is to decide on the value of ω. One possibility is to ignore past data (no credibility) and simply charge the manual rate M.

Another possibility is to ignore M and charge $\overline{R}$ (full credibility). A third possibility is to choose some combination of $\overline{R}$ and M (partial credibility). From the point of view of the insurer, $\overline{R}$ will be more appropriate if the experience is more stable (σ^2 is small). This means that $\overline{R}$ is a more useful predictor of next year's outcome. On the contrary, if the experience is more volatile, then choice M makes more sense. It is important to note that, the factors contributing to each R_j could arise from a single policyholder, a class of policyholders possessing similar underwriting characteristics, or a group of insureds assembled for some other reason. The approach to deciding to assign full credibility or partial credibility depends on the dynamics of past experience.

Full credibility is assigned when $\overline{R}$ is stable. One method of quantifying the stability of $\overline{R}$ is to infer that $\overline{R}$ is stable if $(\overline{R} - \omega)$ is small relative to ω with high probability.

This means that, two numbers $q > 0$, $0 < r < 1$ with q close to 0 and r close to 1. Common choices are $q = 0.05$, $p = 0.9$. Thus, full credibility is assigned if

$$P(-q\omega \le \overline{R} - \omega \le q\omega) \ge r \tag{1.1}$$

In other words when

$$P\left(\left|\frac{\overline{R}-\omega}{\sigma/\sqrt{n}}\right| \le \frac{q\omega\sqrt{n}}{\sigma}\right) \ge r, \tag{1.2}$$

If full credibility is inappropriate, then it may be desirable to reflect the experience $\overline{R}$ in the risk premium computation as well as externally obtained mean *M*. An intuitive approach is by using a weighted average,

$$P_c = Z\overline{R} + (1-Z)M, \tag{1.3}$$

where the credibility factor $Z \in [0,1]$, need to be specified.

While the limited fluctuation method provides simple solutions to premium determination, there are theoretical challenges. First there is no underlying theoretical model for the distribution of R_js and hence no justification of a premium of the form (1.2). Secondly, even where it is appropriate there is no guidance for the selection of q and r. Finally, the limited fluctuation method does not examine the difference between ω and M. When Equation (1.2) is employed, it is essentially stating that the value of M is accurate as fair representation of the expected value given no information about a given policyholder.

An approach known as greatest accuracy credibility theory, which is an improvement over that of Buhlmann (1967) was also promulgated. According to this theory, it is possible that the given policyholder may be different from what has been assumed. If this is the case, how should one choose an appropriate rate for the policyholder. With greatest accuracy credibility theory, it is assumed that the risk level of each policyholder in the rating class may be characterized by a risk parameter ϑ, vector valued, but the value of ϑ varies by policyholder. This allows us to quantify the differences between policyholders with respect to the risk characteristics. Because all observable underwriting characteristics have already been used, ϑ may be viewed as the representative of the residual, the unobserved factors that affect the risk level.

Because risk levels vary within a population, clearly, the experience of the policyholder varies in a systematic way with ϑ. Suppose that the experience of a policyholder picked (at random) from the population arises from a two-stage process. First the risk parameter ϑ is selected from the distribution $\Pi(\vartheta)$. Then the claims or losses R arise from the conditional distribution of R given ϑ, $f_{R|\Theta}(r \mid \vartheta)$. Thus, the experience varies with ϑ via the distribution given the risk parameter ϑ. The distribution of claims thus differs from policyholder to policyholder to reflect the differences in the risk parameters.

Definition of Relevant Terms

Stochastic process

Suppose $\{X(t),\ t \geq 0\}$ is a family of random variables indexed by the time parameter t. The process $X(t)$ is called a stochastic process. The values assumed by the process are called the states, and the set of possible values is called the state space. The state space can be continuous or discrete. The set of values of indexing parameter is the parameter space. The parameter space can also be either continuous or discrete.

Markov Property

Suppose $\{X(t)\}$ is a strictly stationary d-dimensional time series process, where d is a positive integer. It follows a Markov process if the conditional probability distribution of X_{t+1} given the information set $\tau_t = \{X_t, X_{t-1,} ...\}$ is the same as the conditional probability distribution of X_{t+1} given X_t only. Formally expressed as $\mathrm{P}(X_{t+1} \le x \mid \tau_t) = \mathrm{P}(X_{t+1} \le x \mid X_t)$ almost surely for all $x \in \Re^d$ and all $t \ge 1$.

Actuarial Risk

Actuarial risk results from the selling of insurance policies and other liabilities to raise funds. Actuarial risk according to Santomero and Babbel (1997) is the risk that the firm is paying too much for the funds it receives or, alternatively, the risk that the firm has received is too little for the risk it has agreed to absorb. Actuarial risk may have adverse effects on the long-term profitability of an insurance company due to underwriting losses and overpricing liabilities.

Boosting

Boosting is a general procedure that combines simple models iteratively to improve the predictive performance of a model instead of a single model as in the case of traditional predictive modeling techniques that uses QR-decomposition or factorization.

QR-Decomposition

Given a matrix A, its QR-decomposition is a matrix of decomposition of the form $A = QR$, where R is an upper triangular matrix and Q is an orthogonal matrix, satisfying $QQ' = I$, where I is the identity matrix. This matrix decomposition can be used to solve systems of equations. QR-decomposition is implemented in several packages to calculate the least squares fit.

Gradient Boosting

Gradient boosting is an approach where new models are created and predicted on the residuals or errors of the prior model and then added together to make final prediction. It is called gradient boosting because it uses gradient descent algorithm to minimize the loss when adding new models.

Organization of the Study

This study is structured into five chapters as follows; Chapter One describes the background, the statement of the problem, objectives, the significance, limitations of the study and the theoretical framework on which the study hinges.

Chapter Two presents a structured review of literature relevant to the study. It also provides a brief history of the insurance evolution in general in Ghana, and discusses studies related to it. It also reviews literature on the various pricing techniques with a focus on tree-based methods, bagging, random forest and gradient boosting. Relevant codes for the benefit of the reader is presented in APPENDIX D.

Chapter Three presents the methodology used to achieve the objectives; the methods adopted and how it was modified to achieve the objective is discussed in this chapter. The chapter also discusses numerical methods for estimation as well as the procedure for evaluating the model.

Chapter Four presents a detailed analysis of the results of the study. In order to get a fair view of the health of the non-life insurance sector, the study adopted grey relational theory to assess the financial performance of the non-life insurance companies as a prelude to studying the main objective of the study. It then provides a step-wise approach to the model buiding processes and asseessment.

Chapter Five presents the discussion of the key findings and appropriate conclusions and recommendations. The limitations of the study to guide further study in this area is captured in this chapter.

Chapter Summary

This chapter examined critically the task in insurance pricing. It specifies the problem and identifies the research objectives. It identifies various ways that total claim amount and size can be modelled and identified the main drawback of the various strategies. The chapter provides a section on the theoretical framework in which the study hinges.

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter presents a structured review of the literature related to the study. It starts with an evolution of insurance and overview of the non-life insurance industry in Ghana. It also presents relevant literature on the on various assessment methods of financial performance in the insurance industry. It then examines the core aspect of insurance pricing; the theoretical aspect of non-life insurance pricing and the existing methods. The chapter also reviews papers that involves both traditional and algorithmic modeling framework. The chapter ends with a summary.

Evolution of Insurance

Protecting against risk dates back to earliest civilization and it all ties in with major events, newly introduced legislation and the industry of the time. It is believed that King Hammurabi, the 6th Babylonian King introduced the first basic insurance policy around 2100 B.C. The policy was paid by traders in the form of loan to guarantee the safe arrival of their goods (Gadahn, 2010). As history progressed, the desire for insurance increased. The Phoenicians and the Greeks wanted the same type of insurance their neighbors had. The Romans were the first to have burial insurance where people joined burial clubs which paid funeral expenses to surviving family members and this has progressed till today.

Formally, marine insurance appears to be the earliest form of insurance (Afriyie, 2006; Gadahn, 2010). Around 13th or 14th Century in Genoa and

Palermo in Italy, insurance policies were secured on landed estates. By around 1500, marine insurance was in use in England, Spain, France, Italy, etc. It was London that held the reins in the insurance industry with 30 sworn brokers in the capital by the later part of 16th Century. Despite the competition throughout the 17th Century, the city of London's commitments to marine risk overseas had an annual total of several millions of pounds. Policies were signed by individuals, either alone or in a group. They wrote their name and the amount of risk they were willing to assume under the insurance proposal. The term "underwriter" begun during this era.

In 1693, the astronomer Edmond Halley created a basis for underwriting life insurance by developing the first mortality table. He combined the statistical laws of mortality and the principle of compound interest to come up with this table. The problem is that this table used the same rate for all ages and for both sexes. In 1756, Joseph Dodson corrected the error and made it possible to rescale the premium rate to age. During this time, the practice of insuring cargo while being shipped was widespread throughout Europe. In 1688, the first insurance company was formed and started in Lloyd Coffee House in London; a place where merchants, ship owners, and underwriters met to transact business (Gadahn, 2010).

Moreover, in 1666 the Great Fire of London that rampaged through the city, destroying 13,000 houses also catalyzed the development of fire insurance. At that time, the people of London did not have fire insurance so that if one's house was destroyed one would have to personally fund the rebuilding of it. Fourteen years after the Great fire, a well-known physician and construction entrepreneur Nicholas Barbon was compelled to do something to help people to protect their property against such disasters. He founded the "Fire office" in London which was to be the very first fire insurance provider in the UK. In exchange for a yearly premium, he would pay to rebuild a home should it burn down. Within 10 years, 1 in 10 houses were insured. During 18th and 19th Centuries, various other types of insurance sprang up, though out of Europe: From hailstorm insurance designed to protect farmers and gardeners, to livestock and steam broiler insurance. It was clear that the type of cover matched with the predominant industry of the time. An early form of employers' insurance was also introduced around the mid-19th Century called 'Fidelity Insurance'. It was designed to protect employers from staff fraud or embezzlement and originated in the UK, as did the earliest form of employer's liability cover which was devised in response to the Employers Liability Act.

The first American insurance company was founded in the British colony of Charleston, SC. in 1787 and the first fire insurance companies were formed in New York city and Philadelphia in 1794. Other needs for insurance were discovered and the practice of risk classification begun. The Workmen's Compensation Act of 1897 in Britain required employers to insure their employees against industrial mishaps and risks. This has brought about what is today known as public liability insurance.

Towards the close of the 19th Century, motor insurance was introduced. It was initially designed for horse-drawn vehicles. This sector of insurance cover became the fastest growing insurance in the 20th Century with various levels of cover (Gadahn, 2010).

It is clear from the foregoing that, industry transforms, legislation changes, and major events and natural disasters are the defining metrics of insurance around the world. Insurance cover evolves to stay relevant and to protect the newly emerging risks.

Financial Performance Analysis

There are diverse ways of measuring financial performance. The most basic ones are based on statistical procedures which usually depend on normality assumption of the data. In most cases such assumptions are not met. Sometimes the financial ratio data is limited, and this could make the outcome bias. The grey system theory proposed by Deng (1982) can be used in cases where a limited amount of data is available, and the normality assumption is not satisfied (Kung & Cheng, 2004; Wen, 2004). It is especially suitable for the determining and assessing the financial performance of companies (Kung & Cheng, 2004). Kung and Cheng (2004); Wu, Hsiao and Tsai (2008) and Huang and Peng (2011), used GRA methodology to measure the performance of companies in various industries.

For instance, Kung et al. (2006) examined the health of 16 non-life insurance companies in Taiwan during the 2000 to 2004 period. The study selected 24 financial ratios as a basis of performance evaluation. These ratios were segmented into five performance indicators namely profitability, capital operational capability, capital structure, solvency and management efficiency. The findings revealed that return on assets (ROA) ratio, funds utilization efficiency ratio, current debt to capital ratio, equity ratio and retention premium ratio exhibited high impact on the performance the selected companies. Tsai, Huang and Wang (2008) proposed a performance evaluation model for the property-liability insurance companies using a combination of Analytic Hierarchy Process (AHP) and GRA. The authors used three main evaluation criteria that comprise of business, profitability and whole company operating indices. The authors also used eleven (11) sub-evaluation criteria in the analysis and ranked fourteen Taiwanese property-liability insurance companies based on the results of the analysis.

Yan and Kung (2011) also applied the GRA method to rank the business performance of 15 larger scale Tawainese insurance companies based on the grey relational grade using data from 2004 to 2008. Twenty-four financial ratios were selected for this study and these ratios were categorized into five. These categories are capital structure, profitability, debt-paying ability, business performance and capital employment. The results and many other studies have indicated that the GRA method is a more flexible and convenient method for assessing performance of companies.

Elitas, Eleren, Yildiz and Dogan (2012) as well as Kula, Kandemir and Baykut (2015) have analyzed the performance of insurance companies using the GRA method. These authors used 10 financial ratios and segmented these ratios into three. Based on the analysis Perker and Baki (2011) ranked the financial performance of three leading insurance companies operating in Turkey using a one-year data (2008). The GRA method equally suggested that an insurance company that has high liquidity ratios may have high performance. In addition, Elitas et al. (2012) research study investigated the financial performance of seven insurance companies traded in ISE for the 2010 to 2011 period and found that the most important ratios in the financial success of insurance companies are the liquidity ratios.

More so, Kula et al. (2015) in their application of GRA method used ten (10) financial indicators namely; current ratio, net profit margin, earnings per share, equity to total assets ratio, return on assets, return on equity, market value, total assets, short term debt to total debt ratio and debt to total assets ratio to study and asses the financial performance of eight insurance companies traded using the 2013 BIST data. The findings of the study indicate that the profitability ratios have tremendous effect on financial performance an insurance company.

In this study the financial performance of 25 non-life insurance companies in Ghana as at 2016 was examined using the Grey Relational Analysis approach.

The Concept of Non-Life Insurance Pricing

Insurance is based on risk. When you get an insurance policy, the insurance company is taking on some of your risk. Having an auto-insurance policy means that if your car gets damaged in an accident, the insurance company will pay for the repairs. The insurance company makes up for the risk it takes on by charging premiums and setting deductibles. If a company charges too little, it could go bankrupt and when it charges too much it could lose business to its competitors. In Ghana such argument is curtailed because prices are set by NIC and punishes companies that do undercutting; charging prices below what the NIC have set forth.

Denuit (2003) posits that the pricing process is a procedure for determining a fair premium corresponding to the insured's risk profile. In other words, the insurance pricing process can be understood as an ensemble of methods that establish the price paid by the insured to the insurance company in exchange for risk transfer. Within the context of insurance, the necessity of different charging tariffs is emphasized by the insurance portfolio heterogeneity that leads directly to the so-called concept of asymmetrical information. The information problems between the insurance company and the policyholder arise when the insurer has difficulties in evaluating the risk level of the insured.

In Economics literature, two aspects of asymmetrical information exist, namely moral hazards and adverse selection. According to Denuit (2007), adverse selection occurs when the policyholders have a better knowledge of their claim behavior than the insurer and they take advantage of the information unknown to the insurer, while moral hazard; according to Chiappori and Salanie (2000) arises when the probability of risk occurrence depends on the insured behavior and his decisions. The difference between the two is explained by Dionne et al. (2001), that adverse selection is the effect of unobserved differences among individuals that affect the optimality of insurance contract while moral hazard is the effect of contracts on individuals' unobserved behavior. Put differently, problems arise in insurance as a result of effect of applying the same premium for the entire portfolio or class of portfolios.

Chiappori and Salanie (2000) and Dionne, Gourieroux and Vanasse (2001) believed insurance pricing is efficient to combat the asymmetrical information by dividing the insurance portfolio into sub-portfolios, where the risks can be considered independent. This leads to defining risk classes that will have assigned different premium depending on the gravity of risk that define them. In this respect, an important notion is emphasized by risk classification criteria.

Hence if risks are grouped based on a priori information regarding the insured or the insured assets, the obtained groups are called a priori class. Conversely if claim history is considered for every insured, the groups are called a posteriori risk class. The two concepts a priori and a posteriori pricing are briefly explained.

Charpentier and Denuit (2005) suggest that the fundamental idea in a priori pricing is to segment the insured risk in several categories so that within each category the risks are considered equivalent and governed by same law. According to Delaporte (1972), a priori pricing allows grouping the risks assembly in tariff classes, each group including policyholders with identical risk profile that will pay the same reasonable premium. An important remark on independence of risk is given by Buhlmann (1967) who states that the independence assumption is so natural, that many authors forget to mention.

The first a priori pricing in non-life insurance is the minimum bias risk classification procedure proposed by Bailey and Simon (1960). This method utilizes an iterative algorithm in calculating the optimal values for each risk level by minimizing the bias function. Although it was created outside a recognized statistical framework, the actuarial literature argues that this "heuristic" iteration approach is a case of Generalized Linear Models (GLMs).

GLMs have become a common statistical industry practice for non-life insurance pricing. McCullagh and Nedler (1989), highlighted the two main advantages of GLM techniques. Firstly, the generalization of the linear modeling allows the deviation from the assumption of normality, regression being extended to distributions from the exponential family (Normal, Poisson, Binomial and Gamma distributions). Secondly, the GLMs allow the linear regression to be related to the dependent variable through a link function, modeling the additive effect of independent variables on a transformation of the mean, instead of the mean itself. In other words, this function connects the linear predictor or the score with the mean of the dependent variable. Comparing to the minimum bias procedure techniques, the GLM models have the advantage of a theoretical framework that allows the usage of statistical tests in order to evaluate the fit of the models.

Lemaire (1985) in his literature demonstrated the effectiveness of the methods used to estimate the insured risks. A remarkable contribution goes also to Charpentier and Denuit (2005) who have succeeded to cover in modern perspective of the insurance business. Ohlson and Johansson (2010) also treated in an exhaustive manner the methods considered to be the base in insurance risk classification, giving attention to statistical techniques for calculating the auto-insurance premium. Other studies have pointed out the contribution and the merits of Jong and Heller (2008), Kaas et al., (2009) and Frees (2010) who have emphasized the theoretical and practical aspects of the pricing methods to assess the insurance premium. Antonio and Valdez (2012) also presented a conceptual view and an empirical approach for insurance pricing process.

In terms of posterior pricing, most actuarial literature has demonstrated that the usage of a priori pricing involves the lack of causality between some tariff variables and risk occurrence. Certain important risk factors cannot be observed, leading to violation of homogeneity assumption of an effective risk classification system. The limits of this type of pricing require the approach of posterior actuarial models that take into consideration additional information about the individual claims history of policyholders.

Posterior pricing is based on credibility theory. Savage (1954) emphasized that the notion of credibility theory is closely related to perception of risk; individuals are given different degrees of credibility to the occurrence of certain events depending on perception of risk. Whitney (1918) introduced the concept of partial credibility, arguing that the problem of assessing the experience comes from the need to find a balance between collective experience, on one hand, and individual risk experience on the other hand. Hence Whitney declares that the basic principle of credibility is to establish a weighting factor, emphasizing the definition of pure premium as a balance between experience of an individual risk and that of a risk class.

Buhlmann (1967) solves the problem of finding an optimal estimation for the premium corresponding to the n^{th} period, by considering the observations regarding the risks registered in previous periods. He succeeded in revolutionizing the credibility theory by introducing a credibility factor. Buhlmann (1970) together with Erwin, developed the Buhlmann-Straub famous model. The main improvement of the initial model being the definition of the structural parameter estimators. Most of the principles of credibility theory aligns with the basic model proposed by Buhlmann.

The fundamental idea of system is detailed in Lemaire (1995). Lemaire indicated that within the bonus-malus system, described as a scale that consists of a finite number of levels, policyholders are given a certain place according to transition rules and to the number of claims at fault. Each level corresponds with a certain coefficient that will be applied to the premium calculated in the a priori stage of analysis. McClenahan (2001) observed that, in the 18th century, the determination of fire insurance premiums was based upon the roof type and the structure of buildings, and the premium for marine insurance considered to be the oldest form of insurance, was based on the design characteristics of the ship. The author argues that, considering the presence of uncertain events that may occur depending on certain risk factors, actuaries have always aimed to find a mathematical formulation to determine the probability of risk occurrence and to establish the insurance premium. Under the notable influence of Lundberg's (1932) and Cramer's (1930) studies, who are considered as the founders of mathematical theory of risk, actuaries were interested in approaching the risks from the insurance company's perspective.

The famous monograph published by Buhlmann (1970) requires the recognition of actuarial mathematics as a fundamental subject in probability theory and applied statistics of non-life insurance. Gerber (1979) in his paper indicated that the determination of the probability law of risk occurrence cost has always been the central topic in risk theory literature.

Retrospectively, actuarial science was limited to the use of Gaussian linear models. Thus, using linear regression models to quantify the impact of explanatory variables on a phenomenon of interest. The linear model, proposed by Legendre and Gauss in the 19th century, has taken lead in areas such as econometrics and finance, but the applicability of this model in insurance has been suspect. Linear modeling implies a series of hypotheses that are not compatible with the reality imposed by the frequency and cost of the damages generated by the risk's occurrence. While the complexity of the statistical criteria has become more pronounce, the actuaries had to solve the problem of finding some models that explain as realistic as possible the risk of occurrence. Admittedly, no mathematical model will describe completely the reality, the model analyses and the confrontation of theoretical properties of the studied phenomenon with those observed is a pragmatic way to acquire a better understanding of reality and to predict the future responses of analyzed events.

Denuit (2007), however, explained that although the credibility theory can be seen as the art to combine different collections of data to obtain an accurate overall estimate, its specific methods are difficult to implement in practice due to their mathematical complexity. Therefore, insurance companies have approached some methods which simplified versions of are imposed by credibility theory. In this sense, one of the commercial versions of the credibility theory is the bonus-malus system introduced by Pesonen (1962). He tried to establish the rules for obtaining optimal premiums for each risk classes depending on the bonus-malus levels. According to Denuit (2007), bonus-malus systems allow premiums to be adapted for hidden individual risk factors by taking into consideration the past claim record. Hence, in the context of insurance markets, the main purpose of the bonus-malus system is to assess in an equitable manner the individual degree of risk so that the insurance company will demand a premium corresponding to the insured risk profile and claim history.

Modeling Techniques in Insurance

Traditional modeling techniques such as generalized linear modeling (GLM) technique as proposed by Nelder and Wedderburn (1972) has been the major tool used for loss cost modeling and solving other insurance related problems (McCullagh & Nelder, 1989). Haberman and Renshaw (1996) demonstrated how GLMs can be used for solving a variety of actuarial statistical problems such as survival modeling, loss distributions, risk classifications, premium rating and claims reserving in non-life insurance. Konstantinides et al., (2002), compared the adequacies of the Poisson model to each of the mixed Poisson models with Belgian motor third party insurance portfolio concluded that the mixed Poisson model fits better than the Poisson model. However, a comparison amongst the mixed Poisson models, such as Poisson-Normal, Poisson-Lognormal and Poisson-Inverse Gaussian revealed no significant differences. The Gamma distribution has also been found to be extremely useful for risk analysis, especially, for claim size modeling (Hogg and Craig, 1995).

Bourcher et al. (2008) studied models of insurance claim counts with time dependence. The study concluded that Negative Binomial distribution models exhibit better fits statistically than Poisson distributions. Boucher et al. (2008) also implemented a zero-inflated generalized Poisson regression to estimate the disability severity score of victims involved in motor accidents on Spanish roads. The model showed that the estimated mean of the severity score for the disability of motor cycles was 2.8 times higher than that computed for car victims. It was hence concluded that the settlement of bodily injury claims represents the largest aggregate claim cost faced by motor insurers. Moncher and Fu (2004), presented in Dzakwasi (2014), also presented a work on severity distributions for generalized linear models considering Gamma and lognormal distributions. In additon, a Monte Carlo simulation technique was applied to examine the unbiasedness and stability of the generalized linear model's classification relativities when gamma, lognormal and normal distributions are assumed. The gamma distribution provided better predictive accuracy and efficiency. Willmot (1987), compared the Poisson-Inverse Gaussian (PIG) distribution to the negative binomial distribution and concluded that the negative binomial fit is superior compared with the Poisson-Inverse Gaussian. This agrees with the paper by Dadey et al., (2011). Ruohonen (1987) also considered a model for the claim number process. He considered a mixed Poisson process with three-parameter Gamma distribution as the structure function. The three-parameter Gamma was compared with the two-parameter Gamma model giving the negative Binomial distribution.

Panjer and Willmot (1992), proposed the Generalized Poisson-Pascal distribution with three parameters for modeling automobile claims. The fits obtained were satisfactory. However, Denuit et al., (2007), cautions that the Poisson-Inverse and the negative binomial are special cases for Generalized Poisson-Pascal distribution.

Chernobai et al. (2005) also worked on a procedure for constituent estimation of the severity and frequency distributions based on insurance data. The findings of the study revealed that thresholds lead to a serious underestimation of the ruin probabilities. The theoretical study proposed a practical solution to the problem and suggested that using truncated distributions instead of regular (un-conditional) distributions provides for accurate distributional parameters.

Dadey et al. (2011) compared the Poisson distributions and the negative binomial distribution to determine which distribution best fit the auto-insurance claims in Ghana. The results revealed that the negative binomial distribution appear to be more effective than Poisson distribution for fitting insurance claims and therefore, provides somewhat reliable estimates for planning and decision - making as far as claim reserving is concerned. The paper also compared bootstrap estimates with the normal estimates and indicated that the bootstrap estimates did not vary from the estimates obtained by the probability models in terms of claim modeling.

Mihaela (2015) also analyzed claims severity and frequency using the GLM technique. This paper used Poisson regression distribution to model the claim counts, whilst the gamma regression was used in modeling claim severity. The data was categorized into four main types; third party property damage, third party bodily injury, own damage and theft. This study focus on non-zero counts.

Adeleke and Ibiwoye (2011) used data from prominent lines of non-life insurance business in Nigeria to determine appropriate models for claim amounts by fitting theoretical distributions to the various data. The risk premiums for each class of business were also estimated. The authors fitted various distributions such as exponential, Lognormal and Gamma distributions. In coming out with which distribution fits the data well, the researchers employed goodness-of-fit test. Thus, the study made use of chi-square goodness-of-fit, which is found to be suitable for both discrete and continuous distributions. The study, however, did not consider the many zeros inherent in the data. The chi-square goodness-of-fit was chosen over the Kolmogorov-Smirnov test because the Kolmogorov-Smirnov test is often not good at detecting tail discrepancies (Boland, 2006). The study established that a Gamma distribution would be best for property, fire and commercial insurance products, whilst lognormal is best to model theft and motor insurance. Weibull was also found to be best fit for armed robbery plan. The researchers' choice of candidate models was purely subjective, because there are several other candidate models such as Pareto, inverse-gaussian etc., that were left untested which may have turned out to be better. Hence the study was not comprehensive enough. Guiahi (2001), presented a paper on issues and methodologies for fitting alternative statistical distributions to samples of insurance data. His illustration was based on sample of data with log-normal as the underlying distribution. He used the method of maximum likelihood to estimate model parameters and used Akaike's Information Criterion (AIC) to compare which probability distributions fits the data set best.

Consul (1990) as cited in Denuit et al. (2007) fitted six data sets by the Generalized Poisson distribution. Although the Generalized Poisson law is not rejected by a Chi-square test, the fits obtained by Kestemont and Paris (1985), for instance was better. Elvers (1991) as cited in Boucher et al. (2008), reported that the Generalized Poisson distribution did not fit the data observed in a motor third party liability insurance portfolio very well. Consul and Famoye (1992) suggested that the Consul distribution as a probabilistic model for the distribution of the number of claims in automobile insurance. However, Sharif and Panjer (1993) found serious flaws embedded in the fitting of the Consul model, the restricted parameter space and some theoretical problems in the derivation of the maximum likelihood estimators. Their findings also revealed that Generalized Poisson-Pascal or the Poisson-Inverse Gaussian fits quite well.

Denuit (2007) demonstrated that the Poisson-Goncharov distribution introduced by Lefevre & Picard (1996) provide an appropriate probability model to describe the annual number of claims incurred by an insured motorist. The findings revealed that the Poisson-Goncharov distribution successfully fit the six observed claims as well as other insurance data sets. Wright (2005) fitted models to 490 claim amounts which were drawn from seven (7) consecutive years. He fitted loss distributions using maximum likelihood estimation for each of the 7 years, after which he used P-P plots and Kolmogorov-Smirnov test to assess the quality of fit. Wright employed several statistical distributions which included the inverse Pareto, Pareto, bur, Pearson VI, inverse bur, and lognormal.

It is worth noting that all the modeling approaches discussed above fit data to theoretical distributions without recourse to other relevant factors that influence the outcome and judge for the best using goodness of fit tests. Other approaches such as the GLM overcome the deficiency by considering factors relating to claims. Even though GLMs are more intuitive relative to univariate modeling techniques, it has its drawback. In fact, GLM may be thought of as a linear model for a transformation of the expected response or as nonlinear regression model for the response. Several authors have proposed other procedures. For instance, Wood (2006) proposed Generalized Additive Models (GAM) to overcome some of the deficiencies of the GLM such as the linear link to a more general form. However, with GAM the structure of the model must be specified. The main and interaction effects must be specified by the researcher. This often results in specification bias which likely affects the predictive power.

It has now become customary practice in auto-insurance to let the risk premium per unit exposure vary with geographic area when all other risk factors are held constant. In most developed countries, auto-insurance companies have adopted risk classification according to the geographical zone where policyholder lives (urban/non-urban or according to zip codes). The spatial variation may relate to geographic factors (eg. traffic density or proximity to arterial roads) or socio-demographic factors. In such cases it is desirable to estimate the spatial variation in risk premium and to price accordingly. Spatial postcode methods for insurance rating attempt to extract information in addition to that contained in standard factors (like age or gender). Often, claim characteristics tend to be similar in neighboring post code areas (after other factors have been accounted for). The idea of postcode rating models is to exploit this spatial smoothness by allowing for information transfer to and from neighboring regions (Brouhns et al., 2002).

Patrik (1980) indicated that the risk associated with each district or region can be assessed with the help of statistical models for spatial data. Dimakos and Rattalma (2002) pointed out that this procedure is nevertheless not totally satisfactory since it mixes a frequentist approach to estimate the effect of all the risk factors (except location) with a Bayesian approach to evaluate the riskiness of each geographical zone. Dimakos and Rattalma (2002) proposed a fully Bayesian approach to non-life insurance rate making. This approach still relies on GLMs and thus suffers from the challenges mentioned earlier: continuous covariates as policyholders age enter linearly in the model where as it is now well-established that the effect of some continuous variables is far from linear (typically when it comes to the age of policyholder or age of vehicle).

To correct for the drawback of the fully Bayesian approach to non-life insurance rating, Denuit and Lang (2004) proposed a new modeling framework known as Bayesian GAMs. The Bayesian GAMs allows to estimate simultaneously possible nonlinear effects of an arbitrary number of continuous risk factors, the risk variation, unit or cluster -specific heterogeneity and complex interactions between risk factors. The Bayesian GAM was developed from a Bayesian point of view, primarily because it allows for a unified treatment of linear effects of categorical covariates, nonlinear effects of continuous risk factors, spatial and or cluster-specific heterogeneity. This means that the risk represented by each policyholder is assessed in a single model, which avoids possible distortions in the process. The procedure yields a detailed rating that can be used in the back office to monitor the portfolio (the amount of premium is regarded as risk measure in that context). It also provides the actuary with the appropriated tools to decide about the commercial tariff that will be effectively applied to customers (Markov, 2002). Denuit and Lang (2004) performed numerical analysis of Belgian dataset using Bayesian GAMs. The nonlinear effects are modelled by Penalized-splines (Eilers & Marx, 1996). The authors performed the modeling of Bayesian GAM by incorporating spatial heterogeneity. The method revealed key features of the claim process which are not easily detected by traditional methods.

It could be observed from the literature that auto-insurance claim modeling has been approached in two main ways: the first one disregards observable covariates altogether and lumps all the individual characteristics into random latent variable. The second one disregards random individual risk characteristics and trries instead to catch all relevant individual variations by covariates. The paper by Denuit and Lang (2004) takes the second view, employing contemporary, advance data analysis. It is worth mentioning again that with all the covariates included, there remains substantial risk differentials between individual drivers (due to temper and skill, aggressiveness behind the, knowledge of the highway code, vehicle density etc). Random effects could be added on the score scale to take this residual heterogeneity into account, in the spirit of the model pioneered by Dionne and Vanasse (1992).

Cossette, Landriault and Marceau (2004) also proposed a compound binomial model defined in a Markovian environment which is an extension of the compound binomial model proposed by Gerber (1988). One interesting feature of Gerber's model is that it can be used as a proxy to the classical compound Poisson risk model (Dickson et al., 1995). Moreover, since in

Gerber's model, there exist simple recursive formulas to compute the aggregate claim amount distribution and the ruin probabilities, risk measures under the classical risk model can be approximated by the corresponding ones obtained under the compound binomial model. Other methods have been proposed in the actuarial literature to approximate the risk measures. Among the proposed alternatives are simulation procedures and approximation methods (e.g. Diffusion method), which rely both on complex mathematical tools, to approximate ruin probabilities (Asmussen, 1989; Rolski et al., 1999).

The compound binomial model defined in a markovian environment differs from discrete-time models defined in a markovian environment by Lehtonen and Nyrhinen (1992), Nyrhinen (1998), Lillo and Semeraro (2004). In their models, the outcome variable corresponds to aggregate claims over a single period. However, in the compound binomial model in a markovian environment, the periods are of smaller lengths and at most one claim can occur in each period in computing ruin probabilities.

Other alternative models have employed Tobit models by treating zero outcomes as censored below some cutoff points (Showers & Shotick,1994), these approaches also rely on a normality assumption of the latent response. Jorgensen and de Souza (1994) and Smyth and Jorgensen (2002) used GLMs with a Tweedie distributed outcome to simultaneously model frequency and severity of insurance claims. Due to its ability to simultaneously model the zeros and the continuous positive outcomes, the Tweedie GLM has been widely used method in actuarial studies (Mildenhall, 1999; Murphy et al., 2000; Peters et al., 2008). Despite the popularity of the Tweedie GLM, a major limitation is that the structure of the link function is restricted to a linear form, which can be too rigid for real applications. For instance, it is known that risk does not monotonically decrease as age increases (Owsley et al., 1991; McCartt et al., 2003; Anstey et al., 2005). Although non-linearity may be modelled by adding splines (Zhang, 2011), low-degree splines are often inadequate to capture the non-linearity in the data, while high-degree splines often result in the over-fitting issue that produces unstable estimates. Generalized Additive Models was introduced by Hastie and Tibshirani, (1990), Wood (2006) to overcome the restrictive nature of GLMs.

Statistical Learning Techniques in Estimation

The last three decades have given rise to many new statistical learning methods, including Classification and Regression Trees (Breiman et al., 1984), Random Forest (Breiman, 2001), Neural Networks (Bishop, 1995), Support Vector Machines (Boser, 1992) and high dimensional regression (Fan & Li, 2001; 2002; Gui & Li, 2005; Hastie & Tibshirani, 1990). Boosting has emerged as a powerful framework for statistical modeling. It was originally introduced into the field of machine learning for classifying binary outcomes (Freund and Schapire, 1996). Later, its connection with statistical estimation was established by Friedman et al. (2000). Friedman (2001) proposed a gradient boosting framework for regression settings. Buhlmann and Yu (2003) proposed component-wise boosting framework based on cubic smoothing splines for squared error loss functions. Buhlmann and Horthon (2007) demonstrated that the boosting procedure works well in high-dimensional settings. For censored outcome data, Ridgeway (1999) applied boosting to fit proportional hazard models and Li and Luan (2005) developed a boosting procedure for modeling potentially non-linear functional forms in proportional hazard models.

Thus, the rapid development in computation and information technology has created an immense amount of data which has revolutionized the field of statistics by the creation of new tools that helped analyze the increasing size and complexity in the data structures. Most of these tools originated from an algorithmic modeling culture as opposed to a data modeling culture (Brieman, 2001). In contrast to data modeling, algorithmic modeling does not assume any specific model for the data but treats the data mechanism as unknown. As a result, algorithmic models significantly increase the class of functions that can be approximated relative to data models. They are more efficient in handling large and complex data sets and in fitting non-linearities to the data. Model validation is measured by the degree of predictive accuracy and this objective is usually emphasized over producing interpretable models. It is probably due to this lack of interpretability in most algorithmic models, that their application to insurance problems has been very limited. In terms of practical applications Chapados et al. (2001) used several data-mining methods to estimate car insurance premiums. Francis (2001) illustrates the application of neural networks to insurance pricing problems such as the prediction of frequencies and severities. Kolyshkina, Wong and Lim (2004) demonstrated the use of Multivariate Adaptive Regression Splines (MARS) to enhance GLM building.

Decision trees are intuitive concepts for decision making. They work by splitting the observations into many regions and predictions are made based on the mean or mode of the training observations in that region (James et al., 2015).

Most regression problems have often relied on linear regression. A strictly linear model is a poor fit for the data if the relationship between the response and the predictors appear to be quadratic. Decision trees work through a process of stratification as follows:

1) Divide the predictor space $X = (X_1, X_2, ..., X_p)$ into J distinct and non-overlapping regions $R_1, R_2, ..., R_J$.
2) For every observation in region R_j, we make the same prediction which is the mean of the response variable Y for all observations in R_j.

Estimation of Decision Trees

Decision trees use stratification principle to divide the observations into R_J regions. As in linear regression, the goal is to minimize the residual sum of squares (RSS) which is defined for a decision tree as

$$\sum_{j=1}^{J}\sum_{i\in R_j}(y_i - \hat{y}_{R_j})^2, \qquad (2.1)$$

where $\hat{y}_{R_j}$ is the mean response for the observations in the j^{th} region. To do this, decision trees implement a recursive binary strategy. The process begins at the top of the tree (top-down) and successively splits the data into new regions. This split generates two new branches in the tree. Rather than selecting the optimal split among all future possibilities, this approach is greedy in that it selects the best split at that particular step. Given all the potential splits that could be performed on one of the predictors $(X_1, X_2, ..., X_p)$, the algorithm assigns a cut-off point that splits the data in the manner that reduces the RSS by the largest amount. As the number of predictors and observations increase, the more potential cut-off points the algorithm must consider. However, even with relatively large number of predictors and observations, the computational process is quite efficient. This process continues until some designated stopping criteria is reached, otherwise it could continue until each training observation is sorted into its own node resulting in overfitting. Once this iterative process stops, we can generate predicted values for the response of a given test observation by calculating the mean of the training observations for the region in which the test observation belongs.

Pruning the Tree

Decisions trees are highly susceptible to overfitting due to its natural complexity. This means that if we set the stopping criteria at a higher level we may miss crucial branches later in the process. Instead, we want a method that allows to grow a large tree but preserve the most important branches or elements. This is what is meant by pruning.

Cost complexity pruning is one predominant method for achieving this goal. Thus, cost complexity pruning uses a tuning parameter to selectively prune or snip branches that do not contribute significant predictive accuracy, resulting in a subtree generated from the full tree. Different tuning parameter values will lead to different trade-offs between model complexity and model accuracy. Friedman (2001) recommends pruning in conjunction with K-fold cross-validation to select a cost complexity parameter that optimally balances the trade-off for the specific dataset.

Building Classification and Regreesion Trees

A classification tree is like a regression tree, except that the response variable is qualitative. In making predictions, we predict for a test set observation that are the most commonly occurring class value in the given region. However, we could also consider the class proportions or the proportion of the training observations in the region R_j that fall into a given class. Rather than using RSS to grow the tree, we have three options for minimizing error. An obvious choice might be the classification error rate, or the proportion of training observations in a given region that do not belong to the most common class: $E = 1 - \max_k(\hat{p}_{mk})$, where $\hat{p}_{mk}$ is the proportion of training observations in region m that do not belong to the most common class k. In practice, two other methods grow better and more accurate trees. The Gini index is defined as

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1-\hat{p}_{mk}).$$

It is a measure of node purity. The higher the proportion of observations belonging to a single class, the closer this value will be to zero.

The alternative is cross-entropy:

$$D = \sum_{k=1}^{K} \hat{p}_{mk} \log(\hat{p}_{mk})$$

As more observations are closer to or near 0 or 1, cross-entropy will shrink towards zero. This means that for classification trees, each split can be evaluated using one of these criteria but typically it is either Gini index or cross-entropy.

Decision Trees versus Linear Regression

Linear regression and decision trees utilize entirely different functional forms. Linear regression assumes linear and additive relationships between predictors and the response is given by

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \ , \qquad (2.2)$$

whereas decision trees assume that the observations can be partitioned into the feature space as

$$f(X) = \sum_{m=1}^{M} c_m I_{X \in R_m} \ , \qquad (2.3)$$

$I_{X \in R_m}$ is an indicator function which is 1 when $X \in R_m$ and 0, otherwise.

If the relationship between the predictor(s) and the response are truly linear and additive, then linear regression will likely perform better than a decision tree. If the relationship is highly complex and non-linear, then decision trees may be the better option (Haistie, Tibshirani & Friedman, 2009).

Bagging

According to Hastie et al., (2009), decision trees suffer from high variance. This means that slight change in training set or test set may lead to substantial changes in the estimated model as well as the resulting fit. Bootstrap aggregation or **bagging** is a general method for reducing variance in estimates. Bootstrap involves repeatedly sampling with replacement from a sample, estimating a parameter or set of parameters for each bootstrap sample, then averaging across the bootstrap samples to form a bootstrap estimate of the parameter. By averaging across all the bootstrap samples, reduce the variance σ^2 in the final estimate. Thus, we estimate $\hat{f}^1(x), \hat{f}^2(x), \ldots, \hat{f}^B(x)$ using B separate training sets, and average across the models to generate a low-variance model:

$$\hat{f}_{avg}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^b(x)$$

In bagging, we estimate a decision tree model on each bootstrap sample and average the results of the models to generate the bagged estimate:

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^b(x)$$

Each tree is grown without pruning, so they have high variance but low bias. However, by averaging across the results give an estimate that has low bias and low variance.

For regression trees this is straightforward but for classification trees, we estimate B trees and for a given test observation assign it the majority-class result: the overall prediction is the most commonly occurring predicted outcome across all the B predictions. Compared to the error rate for the corresponding classification tree, bagged estimates generally have slightly lower error rates.

Random Forests

Random forests improve upon bagging by decorrelating the individual trees. The problem with bagging is that if there is a single dominant predictor in the dataset, most trees will use the same predictor for the first split and will bring about correlation and similarity among the trees. The goal of bagging is to reduce the variance of the estimates of the response variable, however, averaging across a set of correlated trees will not substantially reduce variance, at least not as much as if the trees were uncorrelated. To resolve this problem, when splitting a tree, random forests consider a random sample m of the total possible predictors p. Thus, it intentionally ignores a random set of variables. Every time a new split is considered, a new random sample is drawn. The main question then becomes how to select the size of m. The random forests use $m=\sqrt{p}$ for classification trees and $m=p/3$ for regression trees. In comparison with bagging, it is observed that the out-of-bag (OOB) error rate is smaller on random forests model. Also, the Gini index associated with each variable is generally smaller using the random forest method compared to bagging. This is because of the variable restriction imposed when considering splits.

Boosting

The concept of boosting emerged from the field of supervised learning, which is automated learning of data with observed outcome to make predictions for unobserved data. The success story of boosting began with a question, not with an algorithm. The theoretical question is, could any weak learning tool for classification be transformed to become a strong learner (Kearns and Valiant, 1989). In binary classification a weak learner is defined to yield a correct classification rate at least slightly better than random guessing (>50%). A strong learner, on the other hand, is to a considerable extent nearly perfect classification (eg. 99% accuracy). This theoretical question is of high practical relevance as it is typically easy to construct a weak learner, but difficult to get a strong one (Zhou, 2012). The answer that laid the ground for the concept of boosting is that any weak base-learner can be potentially iteratively improved (boosted) to become a strong learner.

To develop this concept Schapire (1990) and Freund (1990) developed the first boosting algorithm. Schapire and Freund later compared the general concept of boosting with garnering *"wisdom from a council of fools"* (Shapire & Freund, 2012). The "fools" here signify the solutions of the simple base learners. These solutions only classify slightly better than a coin flip. The task of boosting is thus to learn from the iterative application of a 'weak' learner and to use this information to combine it to an accurate classification. The idea with boosting is not to manipulate the base-learner itself to improve this performance but to manipulate the underlying data by iteratively re-weighting the observations (Shapire & Freund, 2012). This means that the base learner in every iteration will identify a new solution $h^{[m]}(\bullet)$ from the data. That is, through repeated application of the weak-base learner on the observations that are weighted based on the base-learner's success in the previous rounds, the algorithm is forced to concentrate on objects that are hard to classify since the observations that were misclassified before gets higher weights. Boosting the accuracy is achieved by way of increasing the importance of "difficult" observations. For instance, for each iteration $m = 1,...,m_{stop}$, the weight vector $w^{[m]} = (w_1^{[m]},...,w_n^{[m]})$ indicates the individual weights of all observations depending on the success of their classification in previous iterations. In the process of iteration, the focus is shifted towards observations that were misclassified up to the current iteration m. In the final analysis, all previous results of the base-learner are combined to form a more accurate prediction model: the weights of better performing solutions of the base-learner are increased via an iteration-specific coefficient, which depends on the corresponding miscalculation rate. The resulting weighted majority vote (Littlesone and Warmuth, 1989) chooses the class most often selected by the base-learner while taking the error rate in each iteration into account. From the foregoing it could be deduced that all weak-learners can be potentially boosted to become strong learners.

Boosting is seen as another approach to improve upon the result of a single decision tree. More so, instead of creating multiple independent decision trees through a bootstrapping process, boosting grows trees sequentially, using information form previously grown trees. Rather than fitting a model to the response variable *Y*, boosting fits many decision trees $\hat{f}^1, \ldots, \hat{f}^B$ to current residuals. Each time a new decision tree is estimated, the residuals are updated combining the results of all previous decision trees in preparation for fitting the next tree. Thus, rather than learning hard and fast like in bagging and random forests, boosting learns slowly over time as new trees are added. Because boosting is additive and slow, we can estimate relatively small trees and still gain considerable predictive power. The three main tuning parameters when boosting are:

1) The number of trees B. If B is too large, boosting can overfit. Typically, cross-validation estimate of the error rate or MSE is used to select the optimal B.
2) The shrinkage parameter (ς), which is a small positive number (i.e. 0.01, 0.001 or 0.005). which controls the rate at which boosting learns. As ς gets smaller, B must increase.
3) The number of splits in each tree represented by d. essentially when $d=1$ it is essentially an additive model (each tree is a stump with a single predictor)

Statistical Perspective on Boosting

In statistical learning the major goal is the estimation of functional relationship $y_i \approx h(x_i)+a$ between an outcome variable y_i belong to some set Υ and a vector of explanatory variables $x_i=(x_{i,1},\dots,x_{i,k})\in\Re^p$. The function h and the intercept parameter a are unknown. The estimate of (h,a) is used to get predictions of an unobserved outcome y_{new} based on an observed value of x_{new}. The classical assumption in statistical learning is that, the training data (x_i,y_i) are independent and identically distributed from underlying unknown distribution for a pair of random variables $(X_i,Y_i), 1\le i\le n.$. The quality of the predictor $h(x_i)+a$ is measured by some loss function defined by $L(y_i,h(x_i)+a)$. The goal is to find a predictor $h_P(x_i)+a_P$ that minimizes the expected loss, i.e

$$E_P L(Y,h_P(X)+a_P=\min_h E_P L(Y,h(X)+a) \tag{2.4}$$

where $E_P L(Y, h(X)+a) = \int L(y, h(x))dP(x,y)$ denotes the expectation of L with respect to P. In the case of binary classification, we have $y_i \in \Upsilon := \{-1, +1\}$ and in regression situations $y_i \in \Upsilon \subseteq \mathbb{R}$. If P is unknown, it is in general not possible to solve (2.4).

Adaboost Classification Framework

The early boosting algorithms by Schapire (1990) and Freund (1990) were theoretical constructs for proving the idea of boosting than being suitable for practical usage.

Table 1: Schematic Overview of Adaboost Algorithm

1	Initialize Set the iteration counter $m=0$ and the individual weights w_i for observations $i=1,\ldots,n$ to $w_i^{[0]} = \frac{1}{n}$
2	Choose base-learner Set $m := m+1$ and compute the base-learner for the weighted data set. Re-weight observations with $w_1^{[m-1]},\ldots,w_n^{[m-1]} \xrightarrow{base-learner} \hat{h}^{[m]}(\cdot)$
3	Update weights Compute error rate and update the iteration-specific coefficient $\alpha_m \longrightarrow$ high values for small error rates. Update individual weights $w_i^{[m]} \longrightarrow$ higher values if observation was misclassified.
4	Iterate Iterate steps 2 and 3 until $m = m_{stop}$.
5	Final aggregation Compute the final classifier for a new observation x_{new}: $f_{Adaboost}(x_{new}) = sign\left(\sum_{m=1}^{m_{stop}} \alpha_m \hat{h}^{[m]}(x_{new})\right)$

This notwithstanding it paved the way and hence formed the bases for the first concrete and most important boosting algorithm. Adaboost was the first adaptive boosting algorithm as it automatically adjusts its parameters to the data based on the actual performance in the current iterations: both the weights (w_i) for re-weighting that data as well as the weights for the final aggregation are re-computed iteratively. The setting for adaboost is devided into five. The schematic overview of the Adaboost algorithm is in Table 1.

The introduction of Adaboost (Freund and Schapire, 1996) revolutionized the success of boosting in the field of classification and machine learning. Even though Adaboost results was very accurate the problem with it is the fact that, the predictions are difficult to interpret. This is because the focus of classical supervised learning approaches is often restricted to getting reliable predictions for a new observation. How the prediction for the new observation is derived in most cases is not considered important.

In practice Adaboost is often used with simple classification trees or stumps as base-leaners and typically results in a dramatically improved performance compared to the classification by one tree or any other single base-learner (Ridgeway, 1999). Bauer and Kohavi (1999) reported an average of 27% relative improvement in the misclassification error for Adaboost compared with a single decision tree. They also compared the result with bagging and conclude that boosting algorithms are able to reduce not only the variation in the base-learners' prediction error resulting from the used of different training data set but also the average difference between the predicted and the true classes (bias). Breiman who is a pioneer and leading expert in machine learning supported the view of Bauer and Kohavi. To Breiman, "Boosting is the best off-the-shelf classifier in the world" (Hastie, et. al., 2009).

Dealing with overfitting and underfitting in Adaboost

One critical issue of Adaboost has been its overfitting and underfitting behavior. Overfitting describes the common phenomenon when a prediction rule concentrates too much on peculiarities of the specific sample of the data it will often perform poorly out-of-sample. To avoid overfitting, the task of the algorithms should not be focused on the best possible classifier for the underlying sample but to find the best prediction rule for a new set of observations. The main control mechanism for overfitting is the stopping iteration m_{stop}. Very late stopping of Adaboost algorithm will favour overfitting and too early stoppage also lead to underfitting. Too early stopping often lead to higher error on the training data resulting in as well as poorer prediction on a new data set (underfitting).

One way to explain Adaboost's overfitting behavior is based on the margin interpretation (Mier & Ratsch, 2003; Ratsch et al., 2001). The margin of the final boosting solution, in brief, can be interpreted as the confidence in the prediction. With higher values of this margin may still increase and lead to better predictions on the test data even if the training errors is already zero. This theory was questioned by Breiman who developed the "arc-gv" algorithm which should yield a higher margin than Adaboost, but clearly failed to outperform it in practice with respect to prediction accuracy. Reyzin and Schapire (2006) explained the findings with other factors such as complexity of the base-learner.

For more on margin interpretation see Zhou (2012), Schapire and Freund (2012).

Another explanation on the overfitting behavior of boosting is the use of the wrong performance criteria for evaluation (Mease and Wyner, 2008). In most instances the performance of Adaboost has been measured by evaluating the classification rate, and the resistance to overfitting has been demonstrated usually by focusing on this specific criterion. However, the criterion that is optimized by Adaboost is not correct classification rate by the exponential loss function, and the two criteria are not necessarily optimized by the same predictions. It is for this reason that some authors have argued that the overfitting behavior of Adaboost should be analyzed by focusing on the exponential loss function. It has been suggested by Buhlmann and Yu (2008) that too many iterations can lead to overfitting regarding the exponential loss without affecting the misclassification rate.

In recent times, boosting algorithms have been used in to estimate unknown statistical quantities in general statistical models (statistical boosting). The remainder of this section will broaden the scope for any outcome random variable; count or continuous. The conceptual view and interpretation of boosting from statistical view point has been promulgated by Friedman et al. (2000). The authors provided the basis for understanding the boosting concept in general by showing that Adaboost fits an additive model.

According to Mayr and Schimid (2012) most solutions of machine learning algorithms, such as Adaboost, SVM, Bagging, Random forest etc. are usually regarded as black-box prediction schemes. This is because even though they might yield very accurate predictions, the way those results were produced and which role each single predictor played is difficult to explain. In contrast, a statistical model aims at quantifying the relation between one or more observed predictors and the expectation of the response through and interpretable function $E(Y \mid X = x) = f(x)$. With more than one predictor variables, the different effects of each predictor are typically added, forming an additive model;

$$E(Y \mid X = x) = f(x) = \beta_0 + h_1(x_1) + \ldots + h_p(x_p), \qquad (2.8)$$

where β_0 is an intercept and $h_1(\cdot), \ldots, h_p(\cdot)$ are the effects of the predictors $x_1, \ldots, x_p$ which are components of X. The corresponding model class is called generalized additive model (GAM); (see Hastie & Tibshirani, 1990). The aim is to model the expected value of the response variable, given the observed predictors via a link function $g(\cdot)$.

$$g(E(Y \mid X = x)) = \beta_0 + \sum_{j=1}^{p} h_j(x_j) \qquad (2.9)$$

In this wise, GAMs are not black- boxes but contain interpretable additive predictors. For instance, the partial effect of x_1 is given by $h_1(\cdot)$. The direction, size and shape of the effect can be shown pictorially and interpreted. This is the main difference towards many tree-based machine learning approaches. The core message in this review with statistical boosting is that the original Adaboost algorithm with regression type base-learners (eg. Linear models, smoothing splines) fits a GAM for dichotomous outcomes via exponential loss in a stage-wise manner. The work by Friedman et al. (2000) provided the link between a successful machine-learning and the world of statistical modeling. It is worth to note that there are two approaches to boosting namely: Gradient boosting and likelihood-based boosting which we briefly discuss.

Gradient Boosting

Gradient boosting is one of the most successful machine learning algorithm for non-parametric regression and classification (Freund and Schapire, 1996, 1997). What Boosting does is that it adaptively combines many relatively simple predictive models called base learners and aggregate them into an ensemble learner to achieve a high predictive performance. The seminal work on the boosting algorithm called Adaboost was originally proposed for classification problems. Breiman (1998, 1999) pointed out an important connection between the Adaboost algorithm and a functional gradient descent algorithm. Friedman et al. (2000), Friedman (2001) and Hastie et al. (2009) developed a statistical view of boosting and developed a gradient boosting method for both classification and regression.

Gradient boosting is a more flexible statistical boosting technique that does not depend on a likelihood but depends on the gradient of the loss function. The most popular model class for survival data, the semi-parametric Cox proportional hazard model, can be fitted both by gradient boosting (Ridgeway, 1999) as well as by likelihood-based boosting (Hofner et al., 2013). Statistical boosting has also been made available for fitting Fine and Gray models in the presence of competing risks (Binder et al., 2008). Schmid and Hothorn (2008) extended the tool box for boosting survival data to fully parametric accelerated failure time (AFT) models. A popular discriminatory measure for the evaluation of prediction model is the concordance index (C- index) by Harrell et al. (1982).

Friedman et al. (2000) developed a more general, statistical framework which yields a direct interpretation of boosting as a method for statistical estimation. Friedman presented a boosting framework by optimizing the empirical risk through steepest gradient descent in a function space. According to the authors it is a "stage-wise" additive modeling approach. In general, the optimization problem for estimating the regression function $f(\cdot)$ of a statistical model relating predictor variable X with the outcome variable Y, is expressed as

$$\hat{f}(\cdot) = \underset{f(\cdot)}{\arg\min}\left\{E_{Y,X}[\rho(Y, f(X))]\right\} \tag{2.10}$$

where, $\rho(\cdot)$ denotes a loss function. The most common loss function is the L_2 of the form

$$\rho(y, f(\cdot)) = (y - f(\cdot))^2. \tag{2.11}$$

This leads to classical least squares regression of the mean

$$f(x) = E(Y \mid X = x). \tag{2.12}$$

Practically, with a sample of observations, the empirical risk is minimized as

$$\hat{f}(\cdot) = \underset{f(\cdot)}{\arg\min}\left\{\frac{1}{n}\sum_{i=1}^{n}\rho(y_i, f(x_i))\right\}. \tag{2.13}$$

The fundamental idea of gradient boosting is to fit the base-learner not to re-weighted observations, as in Adaboost, but to the negative gradient vector $u^{[m]}$

of the loss function $\rho(y, \hat{f}(x))$ evaluated at previous iteration $m-1$, of the form

$$u^{[m]} = \left(-\frac{\partial}{\partial f} \rho(y, f) |_{f=\hat{f}^{[m-1]}} \right)_{i=1,\ldots,n} \tag{2.14}$$

In the case of L_2, the loss function $\rho(y, \hat{f}(x)) = \frac{1}{2}(y-\hat{f})^2$. This leads to re-fitting of the residuals $(y-\hat{f})$. This means that for every boosting iteration m, the base-learner is fitted directly to the errors made in the previous iteration $(y-\hat{f}(\cdot)^{[m-1]})$. This suggests that both Adaboost and gradient boosting follow the same fundamental idea. Both algorithms boost the performance of a simple base-learner by iteratively shifting the focus towards problematic observations that are 'difficult' to predict. With regard to Adaboost, this shift is done by up-weighting observation that was misclassified earlier. Gradient boosting however identifies the difficult observations by large residuals computed in the previous iterations. See Table 2 for a schematic view of gradient boosting framework.

Table 2: Schematic Overview of Gradient Boosting Algorithm

- Initialization

1) Let the iteration counter $m = 0$ and initialize the additive predictor $\hat{f}^{[0]}$ with a starting value, e.g. $\hat{f}^{[0]} := (0)_{i=1,\dots,n}$. As well as specify a set of base-learners $h_1(x_1),\dots,h_p(x_p)$

- Fit the negative gradient

2) Let $m := m+1$ and compute the negative gradient vector u of the loss function evaluated at the previous iteration:

$$u^{[m]} = \left(u_i^{[m]}\right)_{i=1,\dots,n} = \left(-\frac{\partial}{\partial f}\rho(y_i, f)\,|_{f=\hat{f}^{[m-1]}(\cdot)}\right)_{i=1,\dots,n}$$

3) fit the negative gradient $u^{[m]}$ separately to every base-learner:

$$u^{[m]} \rightarrow h_j^{[m]}(x_j);\ \text{for } j = 1,\dots,p$$

- Update one component

4) select the component j^* that best fits the negative gradient vector:

$$j^* = \underset{1\le j\le p}{\arg\min} \sum_{i=1}^{n} (u_i^{[m]} - h_j^{[m]}(x_j))^2$$

5) update the additive predictor $\hat{f}$ with this component

$\hat{f}^{[m]}(\cdot) = \hat{f}^{[m-1]}(\cdot) + sl \cdot h_j^{[m]}(x_{j^*})$; where sl is a small step length $(0 < sl \le 1)$.

- Interation

Likelihood-based Boosting

In considering statistical models, particularly in estimations that involves low-dimensional settings, estimation is performed by maximizing a likelihood. While such likelihood can also be used to define a loss function in gradient boosting, a boosting approach could also be built on base-learners that directly maximize an overall likelihood in each boosting step. This is the underlying idea of likelihood -based boosting (Tutz and Binder, 2006). When the effects of the predictors $x_1, x_2, \ldots, x_p$ can be specified by a joint parameter vector β, the task is to maximize the overall log-likelihood $l(\beta)$.

Given a starting value or estimate from a previous boosting step $\hat{\beta}$, likelihood-based boosting approaches use base-learners for estimating parameters in a log-likelihood that contains the effect of a fixed effect offset for obtaining small updates, like gradient boosting, a penalty term is attached to $l(\gamma)$ by a base learner becomes standard least-squares estimation with respect to these residuals. In this special case, likelihood-based boosting coincides with gradient boosting for L_2 loss function. Component-wise likelihood-based performs variable selection in each step. This means there is a separate base-learner for fitting a candidate model for each predictor x_j by maximizing a log-likelihood $l(\gamma)$. The overall parameter estimate $\hat{\gamma}$ is then only updated for that predictor x_j which result in that candidate model with the largest log-likelihood. In linear models, γ_j is a scalar and the penalized log-likelihood takes the form $l(\gamma_j) - \lambda\gamma_j^2$, where λ is a penalty parameter that determines the size of the updates. Component-wise likelihood boosting then generalizes state-wise regression. See Table 3 for a shematic view.

Table 3: Algorithm for Component-wise Likelihood Boosting

- Initialization

1) Let the iteration counter $m = 0$ and initialize the additive predictor $\hat{f}^{[0]}$ with a starting value, e.g. $\hat{f}^{[0]} := (0)_{i=1,\dots,n}$. Or the maximum likelihood estimate $\hat{\beta}_0$ from an intercept model (if the overall regression model includes an intercept term).

- Candidate models

2) Let $m := m+1$

3) for each predictor x_j, $j = 1,..,p$ estimate the corresponding functional term $\hat{h}_j(\cdot)$, as determined by parameter γ_j, by attaching a \penalty term to the log-likelihood $l(\gamma_j)$, which includes as an offset.

- Update one component

4) select the component j^* that results in the candidate model With the argest log-likelihood $l(\gamma_{j^*}) : j^* = \arg\max_{1 \le j \le p} l(\hat{\gamma}_j)$

5) update $\hat{f}^{[m]}$ to $\hat{f}^{[m]}(\cdot) = \hat{f}^{[m-1]}(\cdot) + h_j^{[m]}(x_{j^*})$; potentially adding an intercept term from maximum likelihood estimation.

- Iteration

6) iterate steps (2) to (5) until $m = m_{stop}$

Chapter Summary

As discussed there have been instances where tree-based methods and gradient boosting have been used in analyzing insurance data. For instance, Guelman (2012) used gradient boosting methodology to model auto insurance loss cost "at fault" accident using data from a major Canadian insurer. The framework relies on squared error loss function and the findings indicated that the level of accuracy in prediction was shown to be higher for Gradient Boosting than the conventional GLM approach. Yang et al. (2016) used gradient boosting approach called 'TDBoost' that relied on Tweedie loss function. The findings indicated the TDboost method performed better that the one with squared error loss function by Guelman. The TDboost model also peformed better than the standard GLM approach. This is not surprising because GLMs are relatively simple linear models and thus constrained by the class of functions they can approximate. Besides, Gradient Boosting provides interpretable results through the plot of relative influence of the input variables and their partial dependence plots. This is critical aspect to consider in a business environment, where models usually must be approved by non-statistically trained decision makers who need to understand how the output was produced. The methodology also requires very little data preprocessing compared with the traditional methods, which is one of the most time-consuming activities in data mining project. Last, but not least, the model selection is done as an integral part of the Gradient Boosting procedure, and so it requires little "detective" work on the part of the analyst as opposed to the conventional data modeling or the conventional traditional methods where analyst depends on p-values or statistical tables to make decisions. In summary, Gradient Boosting is a suitable alternative approach to conventional methods for building insurance loss cost models. Appendex D is a detailed and illustrative overview of the tree-based, bagging, random forest and boosting methods.

CHAPTER THREE

RESEARCH METHODS

Introduction

The previous section discussed the relevant literature of the study. This chapter describes the methodology derived to achieve the main objectives of the study. It starts with a description of the grey systems theory used in assessing the financial performances of insurance industry as well as a brief description of design of the research and how the data was obtained. It provides the theoretical basis of the methodology and its application to actuarial modeling, the model building processes and finally discusses the model validation procedures.

Description of Methods used for Financial PerformanceAssessment

Financial statements are the most reliable sources that give current and periodical information about the financial situation of a business. These statements help business partners and stakeholders do financial analysis of the related period or current period. Financial analysis is fulfilled with the basic purpose of making business decisions in a healthier way. Where as financial ratios can be used in financial planning, they can also be used in measurement of realizing activities. The most commonly used method to make financial analysis is the ratio analysis. Ratio analysis is the expression and mathematical interpretation of the relationship between two items aimed to be examined in the financial statements. It helps companies and business to ascertain their obligations, profitability, liquidity status, financial structure and effective use of assets (Kaya, 2016). Financial ratios seen in Table 1 and used in the study show similarities with ratios that were used in a limited number of studies in literature intended to measure performance.

Table 4: Financial Ratios Used in Performance Evaluation

Indicators	Ratio	Code	Aim
Capital Adequacy	Gross Written Premiums/Equity Capital (Gross insurance risk)	R1	Smaller the better
	Technical Reserve cover (Technical provisions to liquid investments)	R2	Smaller the better
	Net Written Premiums/Equity Capital (Net insurance risk)	R3	Smaller the better
	Premium Retention Ratio (NWP/GWP)	R4	Smaller the better
Operating Efficiency	Combined Ratio	R5	Nominal target
	Loss Ratio	R6	Nominal target
Profitability	Return on Assets	R7	Larger the better
	Return on Equity	R8	Larger the better

*NWP : Net written premium, GWP: Gross written premium

The financial ratios were calculated using the annual data of the companies. These data were obtained from the annual reports of the companies. The financial ratios used in the analysis were selected taking into account the availability of data and based on the National Insurance Commission Annual Reports on key financial indicators that summarizes the activities of insurance companies. (National Insurance Commission report, 2016, 2015). The variables are briefly explained:

1. Gross insurance risk (R1) is the ratio of Gross Written Premium to Equity capital.
2. Technical Reserve Cover (R2) is the ratio of Technical provisions to liquid investments. It is an indicator of whether sufficient liquid assets are being held to cover technical provisions, because claims should be paid as when they fall due. Ratios above 100% mean that the company do not have enough liquid investment backing their technical provisions.
3. Net insurance risk ratio (R3) measures the ability or capacity of the insurer's capital and surplus to absorb unforeseen shocks. It is calculated as a ratio of Net Written Premium to Equity. The higher the ratio, the less conservative the insurer, and hence the greater the potential risk that the insurer cannot absorb shock or losses. Retention ratio (R4) is computed as the Net Written Premium over the Gross Written Premium. It represents the portion of the risk that insurers have not passed onto reinsurers. High retentions are usually considered riskier. High retention will require sufficient capital to support the insurer.
4. Combined Ratio (R5) is the summation of claims ratio and the total expense ratio. It is the single best measure of an insurers underwriting and operational efficiency. Generally, a ratio of less than 100% indicates underwriting profitability, while a ratio of more than 100% indicate a loss. This may not necessary be the case for companies with huge investment and other related earnings to boost profitability.
5. Claim or Loss ratio (R6) is calculated as the net claims incurred divided by the Net Earned Premiums. It is a key ratio which indicates how well an insurance company pays claims and to some extent, of fair customer

treatment. The optimal ratio ranges between 40% and 60%. Claims ratio beyond 100% is regarded as inefficient.

6. Return of Assets (R7), This ratio is an indicator of general profitability of the insurer. It is calculated as after-tax profits divided by total assets. It seeks to measure the efficiency with which management utilize the assets of the company to generate returns of the various stakeholders. In practice, however, high ratios may not always be an indication of good performance, as factors such as capital inadequacy can boost the ratio.
7. Return on Equity (R8), this ratio measures the return on the shareholders' funds over a period. It also indicates how effective management is growing and funding the operations of an insurance company using equity financing.

Grey Relational Analysis (GRA)

GRA is a method that can be used in decision making in situations where there are many criteria by ordering them as to relational grade. It is especially useful in ordering the alternatives in situations in which the sample is small and sample distributions is not known. Grey Relational Analysis (GRA) is an analytical method in Grey System theory which was founded by Professor Deng Julong (Deng, 1982; Wu and Chen's, 1999). The term "Grey" means lack of information or not being known at all. This method enables us to determine the level of relation between each factor that come across in a grey system and the compared factor (reference series). It is a distinct similarity measurement that uses data series to obtain grey relational order to describe the relationship between the related series (Kaya, 2016). It can be used to measure the correlations between the reference series and other compared series.

Accordingly, effect degree between factors is called grey relation grade. One of the purposes of usage of GRA is to separate important variables in groups between themselves by recognizing unimportant ones among various variables. Besides, when data set is large and has a normal distribution, methods such as factor analysis, cluster analysis and discriminant analysis can be used in statistics.

Nevertheless, when sample is little and whether distribution is normal or not known the reliability of these analysis decreases and hence GRA become a useful alternative. The steps of GRA as summarized in (Kaya, 2016; Wu, 2002; Zhai, Khoo & Zhong, 2009; Wu and Chen's 1999), is briefly outlined.

Step 1. Construction of the decision matrix

Given that there are n (number of companies) data sequences characterized by m criteria (comprising 8 financial ratios). The sequences can be represented in a matrix form as

$$X = \begin{pmatrix} x_1(1) & x_1(2) & \cdot & \cdot & \cdot & x_1(m) \\ x_2(1) & x_2(2) & \cdot & \cdot & \cdot & x_2(m) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_n(1) & x_n(2) & \cdot & \cdot & \cdot & x_n(m) \end{pmatrix} \tag{3.1}$$

where, $x_i(j)$ is the value of the i^{th} insurance company corresponding to the j^{th} financial ratio. $i = (1,\ldots,n); j = (1,\ldots,m)$.

Step 2. Normalization of the data set

To obtain comparable scales, the data set is normalized (Feng & Wang, 2000). The data can be normalized by one of the three regimes: larger the better, smaller the better and nominal the best.

For larger is the better normalization, we transform $x_i(j)$ to $x_i^*(j)$ as follows.

$$x_i^*(j) = \frac{x_i(j) - \min_{i=1}^{n}[x_i(j)]}{\max_{i=1}^{n}[x_i(j)] - \min_{i=1}^{n}[x_i(j)]}; \quad (3.2)$$

Where, $\min_{i=1}^{n}[x_i(j)]$ is the minimum value of the j^{th} financial ratio and $\max_{i=1}^{n}[x_i(j)]$ is the maximum value of the j^{th} financial ratio.

For smaller is the better normalization, the formula is defined as

$$x_i^*(j) = \frac{\max_{i=1}^{n}[x_i(j)] - x_i(j)}{\max_{i=1}^{n}[x_i(j)] - \min_{i=1}^{n}[x_i(j)]} \quad (3.3)$$

For nominal is the best normalization, the formula is defined as

$$x_i^*(j) = 1 - \frac{|x_i(j) - x_{obj}(j)|}{\max\{\max_{i=1}^{n}[x_i(j)] - x_{obj}(j), x_{obj}(j) - \min_{i=1}^{n}[x_i(j)]\}}, \quad (3.4)$$

where $x_{obj}(j)$ is the target (ideal) value of the j^{th} financial ratio and $\min_{i=1}^{n}[x_i(j)] \le x_{obj}(j) \le \max_{i=1}^{n}[x_i(j)]$

Step 3. Construction of the normalized matrix and generation of the reference sequence

After the normalization process, the normalized matrix which is the revised version of the original matrix or the initial decision matrix is presented in Equation (4.5). Equation (4.6) also shows the reference sequence.

$$X^* = \begin{pmatrix} x_1^*(1) & x_1^*(2) & \cdot & \cdot & \cdot & x_1^*(m) \\ x_2^*(1) & x_2^*(2) & \cdot & \cdot & \cdot & x_2^*(m) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_n^*(1) & x_n^*(2) & \cdot & \cdot & \cdot & x_n^*(m) \end{pmatrix}$$

(3.5)

$$x_0^* = x_0^*(1), x_0^*(2), \ldots, x_0^*(j), \ldots, x_0^*(n) \quad (3.6)$$

$$x_0^*(j) = \max_{i=1}^{n}[x_i^*(j)] \quad (3.7)$$

Step 4. Construction of the difference matrix

Let $\Delta_{0i}(j)$ represent the absolute value of difference between the normalized value and the reference value of the j^{th} financial ratio and is calculated using Equation (4.8)

$$\Delta_{0i}(j) = \left| x_0^*(j) - x_i^*(j) \right| \quad (3.8)$$

After computing $\Delta_{0i}(j)$ values, the constructed difference matrix is shown in Equation (4.9)

$$\Delta = \begin{pmatrix} \Delta_{01}(1) & \Delta_{01}(2) & \cdot & \cdot & \cdot & \Delta_{01}(m) \\ \Delta_{02}(1) & \Delta_{02}(2) & \cdot & \cdot & \cdot & \Delta_{02}(m) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \Delta_{0n}(1) & \Delta_{0n}(2) & \cdot & \cdot & \cdot & \Delta_{0n}(m) \end{pmatrix} \quad (3.9)$$

Step 5. Calculation of the grey relational coefficient

The grey relational coefficient of the j^{th} financial ratio is computed using Equation (3.10)

$$\tau_{0i}(j) = \frac{\min_{i=1}^{n} \min_{j=1}^{m} \Delta_{0i}(j) + \xi \max_{i=1}^{n} \max_{j=1}^{m} \Delta_{0i}(j)}{\Delta_{0i}(j) + \xi \max_{i=1}^{n} \max_{j=1}^{m} \Delta_{0i}(j)} \quad (3.10)$$

where $\tau_{0i}(j)$ is the grey relational coefficient of the j^{th} financial ratio and ξ is the distinguishing coefficient. This coefficient is value between 0 and 1. However, ξ usually ranges between 0 and 0.5 and it reduces the effect of extremely large $\max_{i=1}^{n} \max_{j=1}^{m} \Delta_{0i}(j)$ in cases where the data variation is large (Chang & Lin, 1999).

Step 6. Calculation of the grey relational grades

Given weights of financial ratios, the grey relational grade is calculated as follows;

$$\eta_{0i} = \sum_{j=1}^{m} [(w(j))(\tau_{0i}(j))]; \sum_{j=1}^{m} w(j) = 1 \quad (3.11)$$

With equal weights the degree of grey coefficient is computed by

$$\eta_{0i} = \frac{1}{n} \sum_{j=1}^{m} [(\tau_{0i}(j))] \quad (3.12)$$

where η_{0i} is the grey relational grade and $w(j)$ is the weight of the j^{th} financial ratio in this study. For decision-making processes, if any alternative has the highest η_{0i} value, then it is the most important alternative (Wen, 2004;

Lin & Chang, 2010). Consequently, the performance of the insurance companies can be ranked according to the grey relational grades.

Statistical Perspectives of Premium Pricing Variables

An insurance company collects a lot of information for each single policy holder for each year and period. It is often accompanied with hundreds of variables available for each customer. Most of this information belong to one of the following categories: personal information such as name, type of policy, policy number, other insurance policies. Demographic information such as gender, age, place of residence, population density of the region where the customer is living, occupation type, etc. Driver information such as main user, driving distance within a year, car kept in a garage, etc. Family information such as age and gender of other people using the same car, income, etc. History include count and size of previous claims, property damage, physical injury, occurrence of a loss. Vehicle information such as type, age, engine capacity or strength etc. Response information such asclaim (yes/no), number of claims and claim size. In practice, the claim size is not always known exactly apriori. For instance, if a big accident occurs in January, the exact claim size will often not be known at the end of the year and perhaps not even at the end of the following year. The possible reasons are law-suits or the case of physical injuries. In such a case the statistician or actuary will have to use a more appropriate estimation of the exact claim size to construct a new insurance tariff for the following year. Hence, the empirical distribution of the claim sizes is in general a mixture of what is really observed and estimated.

An insurance company may be interested in determining the actual premium charged to the customer. In principle, the actual premium is the sum of the pure or risk premium plus safety loading such as administrative cost and desired profit.

The focus of this study is on pure premium. Given a set of explanatory variables, the primary response variable for the study is the conditional expectation of the pure premium given the explanatory variables. The secondary response variable is the conditional probability that the policy holder will have at least one claim within one year given the information contained in the explanatory variables. An estimate for the expected pure premium should have the following four attributes;

It is fair: the expectation of the estimated pure premium $E(\hat{Z} \mid \boldsymbol{X} = \boldsymbol{x})$ should be approximately unbiased for the entire population and in sub populations.

It has high precision: some of the precision criteria include mean squared error (MSE) defined by $E[(Z - \hat{Z})^2 \mid \boldsymbol{X} = \boldsymbol{x})]$, which should be small relative to those of other competing models. It is robust against moderate violations of the statistical model assumptions and the impact of outliers on the estimation is bounded. It has the simplicity property, because too complex tariff structure with many interaction terms may only have a reduced practical significance.

If the tariff is not fair, this will lead to bias which is of course bad from the view point of the policyholder, because the premium is too high, and the policyholder will have to pay too much. At first sight this case looks great for the insurance company, but there is a danger that the customer will turn to another insurance company.

Research Approach

According to Saunders et al. (2007), the two overall approaches to conducting research are through induction and deduction. When data is first collected and after analyzing the data a theory is developed, the approach is inductive. On the other hand, if a theory or hypothesis is first developed and later a research strategy is designed to test the hypothesis, the approach is of deductive nature. In this thesis, *an inductive research approach* was adopted to achieve the desired objectives of the study.

Model Description and Assumptions

The Poisson distribution is commonly used to model claim number distributions in non-life insurance. Such a choice assumes, among others, mutual dependence among the number of claims occurring in a period. Nonetheless, if one considers background factors such as weather conditions or location risk and other claim-causing events, the risk propensity may vary significantly from one policy to another as well as location. In as much as these variations are deterministic, the Poisson distribution still applies. On the contrary, if the intensity variations are considered as random, the independence assumption no longer holds. More so, when a count data has variance greater than mean (over-dispersion), assumption of independence does not apply.

In the classical risk model, it is assumed that the claim arrival process is a Poisson process. This assumption implies a constant claim arrival intensity and in most practical cases, such assumption is inadequate. In such situations, Cox process can be used as an alternative. Cox process is a doubly stochastic Poisson process which has stochastic claim arrival intensity rate. A general treatment of Cox process is presented in Rolski et al., (1999). The study considers a process (Markov-modulated), where the claim intensity is assumed to be homogenous within each state but heterogenous among states. A brief pictorial description is shown in Figure 4.

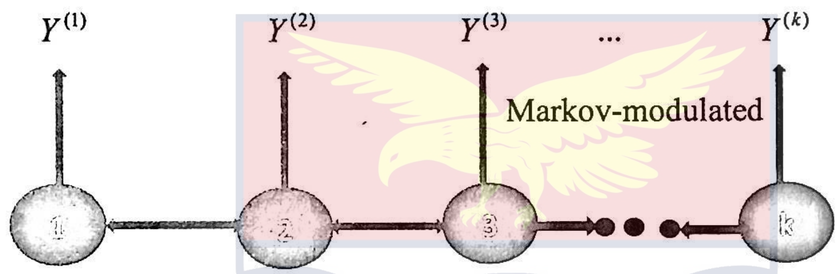


Figure 1: The conceptual framework of the study

where $Y^{(i)}=\sum_{j=1}^{n}Y_j^{(i)}$, is the total claim size in state i, $Y_j^{(i)}$ is a non-negative random variable that represents the amount of j^{th} claim in state i $i=(1,2,...,k)$, $j=(1,2,...n)$ and $N^{(i)}(t)$ represents the number of claims in state i at time t. The total number of claims at time t is given as $N(t)=\sum_{i=1}^{k}N^{(i)}(t)$ and the total claim size at time t, is given as $S(t)=\sum_{i=1}^{k}Y^{(i)}$.

Suppose we assume that the number of claims is equally defined as a function of the Markovian process M at time t such that the number of claims that occurs depends on the behavior of the events in the states. Hence, the random variable $Y^{(i)}$ will have a probability density function $f_{Y^{(i)}}$ and cumulative distribution function $F_{Y(i)}$. In order to define the total claim amount process $S = \{S(t), t > 0\}$, denote $Y_j^{(i)}$ as the amount of j^{th} claim occurring when the Markovian process is in state i. Suppose that $\{Y_j^{(i)},\ j \in \Re^+\}$ is a sequence of independent and identically distributed random variables with probability density function $f_{Y(i)}$ and cumulative distribution function $F_{Y(i)}$. Hence S can be defined as

$$S(t) = \sum_{i=1}^{k} \sum_{j=1}^{n} Y_j^{(i)}(t).$$

Model Specification

The study considers a portfolio of policies of the form $\{(y_j, \boldsymbol{X}_j, w_j, \gamma_i)\}_{j=1}^{n}, j = 1 \ldots, n; i = 1, \ldots, k$ from n independent insurance contracts, where for the j^{th} contract, y_j is the policy's claim amount, $\boldsymbol{X}_j$ (historical and location risk factors) or the set of explanatory variables that characterize the policyholder and the risk being insured, w_j is the duration of policy and γ_i is the risk factor that characterizes risk specific to geographical location i. Thus, the study assumes that the expected risk premium μ_j is determined by a predictor function such that

$$E(Z_j \mid H, L) = \mu_j). \tag{3.13}$$

We thus define the expected risk premium for a policy in location L as

$$\eta(\mu) = \vartheta L + \alpha H,$$

where, ϑ is the associated risk parameter for a policy in location L. The α is the historial risk parameter associated with the claim history and η is the link function that establishes the relationship between the expected premium and the covariates. Historical risk is the risk specific to the policy characteritics. We define location risk as the risk specific to the geographical location or environment in which the policy usually operates. The study derives the location risk for each policy from the distribution of crash data obtained via Markov chain of which we briefly explain.

Given that occurrence of accident at time t is a stochastic sequence, we initially characterize and discretize our risk model into ten states $i = 1,\ldots,10$ based on geographic segregation. Given the initial probabilities x_0 of event E in state i during period t, we compute the transition matrix for period t via Bayes theorem.

Denote $P(E)$ as the long run proportion of times the event E occurs upon repeated sampling during period t, or how likely it is that the event E will occur in state i. Denote by $P(E_{t1})$ the risk of occurrent of an event in state 1 at time t and $P(E_{t2})$ the risk of occurrence of an event at time t in state 2, etc. In the context of our study and for mathematical tractability, define $E_{t1}E_{t2}$ as the event that a vehicle operates between two states. This means that $P(E_{t1}E_{t2})$ represents the risk involved when moving from state one to state two. Given that E_{t1} and E_{t2} are independent, $P(E_{t1}E_{t2}) = P(E_{t1})P(E_{t2})$ representing the risk of accidents between states.

Next, by noting that the number of times accident occurs represent a sequence of random variables. The probability distribution of transitions between states can be expressed as a transition matrix Q. If $i=1,...,k$ the transition matrix Q is shown as

$$Q=\begin{bmatrix} q_{11} & q_{12} & q_{13} & \cdots & q_{1k} \\ q_{21} & q_{22} & q_{23} & \cdots & q_{2k} \\ q_{31} & q_{32} & q_{33} & \cdots & q_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{k1} & q_{k2} & q_{k3} & \cdots & q_{kk} \end{bmatrix}$$

where q_{ij} represents accident risk from state i to state j, q_{kk} represents the risk within state k.

If a discrete time Markov chain $\{Q(t)\}$ is irreducible and aperiodic, then it has a limiting distribution and this distribution is stationary. As a consequence, Q is a $k \text{x} k$ transition matrix of the chain and $x=(x_1,...,x_k)$ is the eigenvector of Q such that $\sum_i q_i = 1$, then we get $\lim_{n\to\infty} Q^n = \gamma_i$, representing accident risk within each of the states.

Estimation of Model-Based Risk Premium

In non-life insurance, the risk premium represents the expected cost of all claims declared by policyholders during the insured period. The calculation of the premium is based on statistical models that seek to incorporate all available information about the accepted risk, thereby aiming at a more accurate assessment of tariffs attributed to each insured.

The basis for calculating the risk premium is the statistical modeling of frequency and cost of claims that depends on the characteristics defined in the insurance contracts. The risk premium is the mathematical expectation of the annual cost of claims declared by policyholders and it is obtained by multiplying the two components; the estimated frequency and cost of claims: the risk premium for the i^{th} policyholder is

$$\sum_{j=1}^{N_i} Y_j = E(Y_i)E(N_i),$$

The claim amount $(Y_1, Y_2, ..., Y_N)$ are independent of the claim number N_i. Several authors have considered a separate evaluation of frequency and cost of claims in computing risk premium with the view to provide a perspective on how the risk factors are influencing the two components. Others recommend a convolution of the two components with the aim of ensuring data consistency and accurate models. Suffice to say that the modeling choice depends on structure of the data and the overall objective. The task of modern actuary is to analyze and find the right formulation of his problem. The study adopted the principle of convolution of the two components defined in a Markovian environment. The formulation of the problem in this study is briefly described below:

The Compound Poisson distribution and the Tweedie Model

This section briefly introduces the compound Poisson distribution and Tweedie model which is essential for the methodology development. Let the claim number N be a Poisson random variable denoted by $Pois(\lambda)$.

Let the claim amount Y be independent and identically distributed gamma random variables denoted by $Gamma(\alpha, \varpi)$ with mean $\alpha\varpi$ and variance $\alpha\varpi^2$. Define a random variable Z such that

$$Z = \begin{cases} 0, & N = 0 \\ Y_1 + Y_2 + Y_3 + \ldots + Y_N, & N = 1,..n \end{cases}$$

Thus Z is a Poisson sum of independent gamma random variables. The resulting distribution Z is referred to as compound Poisson distribution (Feller,1968, Smyth and Jorgensen, 2002). This distribution, F_Z, is connected to exponential dispersion models. The function could be expressed as

$$f_Z(z) = \mathrm{P}(N=0)d_0(z) + \sum_{j=1}^{\infty} P(N=\delta) f_Z(z \mid N=\delta)$$

Given that $E(N(t)) = \xi$,

$$f_Z(z) = \exp(-\xi)d_0(z) + \sum_{\delta=1}^{\infty} \frac{\xi^{\delta} e^{-\xi} z^{\delta\alpha-1} e^{-z/\varpi}}{\delta!\varpi^{\delta\alpha}\Gamma(\delta\alpha)}$$

where $\mathrm{P}(Z=0) = \exp(-\lambda)$, $P(Z \mid N=\delta) = Gamma(\delta\alpha, \varpi)$, $f_Z(z \mid N=\delta)$ is the conditional density of Z given $N=\delta$, and d_0 represents the Dirac delta function at zero.

The exponential dispersion model plays a significant role in actuarial modeling as it is the underlying response distribution (Jorgensen, 1987). A two-parameter representation of the exponential dispersion model is

$$P(z \mid \theta, \phi) = a(z, \phi)\exp\left(\frac{z\theta - k(\theta)}{\phi}\right) \tag{3.14}$$

where a and k are known functions, θ is the natural or canonical parameter belonging to the open interval $k(\theta) < \infty$ and $\phi > 0$ is the dispersion parameter. The function $k(\cdot)$ is called the cumulant function of the exponential dispersion model since if $\phi = 1$, the derivatives of $k(\cdot)$ give the successive cumulants of the distribution. For instance, $E(z) = \mu = k'(\theta)$ and $Var(z) = \phi k''(\theta)$ (McCullagh and Nelder, 1989). The function $k''(\theta)$ can also be expressed in terms of μ denoted by $V(\mu)$. This is what is referred to as the variance function that uniquely defines an exponential dispersion model.

The study focused on the exponential dispersion model with a power variance function $V(\mu) = \mu^{\xi}$ such that the index parameter ξ lies in the interval $(1,2)$. This is referred to as Tweedie distribution which is generated by a compound Poisson-Gamma distribution with high probability mass at zero with a skewed continuous distribution on the positive real line. Exponential dispersion models could also be expressed in terms of moment generating functions, a fact which is exploited by this study.

From Equation (3.4) the moment generating function is given by

$$M(t) = \int \exp(zt) f(z;\mu,\phi)dz \tag{3.15}$$

Hence the conditional expectation and variance for the compound Poisson random variable $Z \mid N$ when $w_i \neq 1$ is

$$\begin{aligned} \mu_i &:= E(Z_i) = E(E(Z_i \mid N_i)) = w_i \lambda_i \alpha \varpi_i \\ Var(Z_i) &= E(Var(Z_i \mid N_i)) + Var(E(Z_i \mid N_i)) = \frac{1}{w_i}\left(\lambda_i \alpha \varpi_i^2 + \lambda_i \alpha^2 \varpi_i^2\right) \end{aligned} \tag{3.16}$$

The conditional expectation and variance for the compound Poisson random variable $Z \mid N$ when $w_i = 1$ is

$$\mu_i := E(Z_i) = E(E(Z_i \mid N_i)) = \lambda_i \alpha \varpi_i$$
$$Var(Z_i) = E(Var(Z_i \mid N_i)) + Var(E(Z_i \mid N_i)) = \lambda_i \alpha \varpi_i^2 + \lambda_i \alpha^2 \varpi_i^2 \quad (3.17)$$

This means the Tweedie distribution is scale invariant. w_i is the duration for policyholder i

The cumulant function of the compound Poisson-Gamma model is

$$M_Z(t) = \exp\left[\lambda\left\{(1-\varpi t)^{-\alpha} - 1\right\}\right] \ ; \ 1 < \xi < 2, \mu > 0. \quad (3.18)$$

Given an exponential dispersion model of the form $V(\mu) = \mu^{\xi}; \xi \geq 1$, the cumulant function $k(\cdot)$ for the Tweedie model could be found by

$k''(\theta) = \dfrac{d\mu}{d\theta} = \mu^{\xi}$ and solve for $k(\cdot)$. Without loss of generality, we choose $k(\theta) = 0, \mu = 1$ at $\theta = 0$

This gives the following results:

$$\theta = \begin{cases} \dfrac{\mu^{1-\xi} - 1}{1-\xi}, & \xi \neq 1 \\ \log(\mu), & \xi = 1 \end{cases} \qquad k(\theta) = \begin{cases} \dfrac{\mu^{2-\xi} - 1}{2-\xi}, & \xi \neq 2 \\ \log(\mu), & \xi = 2 \end{cases}$$

$$\mu = \{\theta(1-\xi) + 1\}^{1/(1-\xi)}.$$

In the context of the study variable in Equations (3.17) and (3.18), the parameter representations for the Tweedie model is given by

$$\mu = \lambda \alpha \varpi \qquad \lambda = \frac{\mu^{2-\xi} - 1}{\phi(2-\xi)}$$

$$\alpha = \frac{2-\xi}{\xi - 1} \qquad \xi = \frac{\alpha + 2}{\alpha + 1}$$

$$\phi = \frac{\lambda^{1-\xi}(\alpha\varpi)^{2-\xi}}{2-\xi} \qquad \varpi = \phi(\xi-1)\mu^{\xi-1}$$

Numerical approximation to the density function

From (3.3), the joint distribution of the compound Poisson can be derived as

$$P(z,t \mid \lambda,\alpha,\varpi) = P(z \mid t,\ \alpha,\varpi)P(t \mid \lambda) = \begin{cases} \exp(-\lambda), & 0 \\ \dfrac{z^{t\alpha-1}\exp(-z/\varpi)\lambda^{t}\exp(-\lambda)}{\Gamma(t\alpha)\varpi^{t\alpha}t!}, & \Re^{+} \end{cases}$$

Thus, Z is a Poisson sum of independent gamma random variables and for each policyholder $Z_i \sim Tw(\mu_i, \phi/w_i, \xi)$, where ϕ is the dispersion parameter, ξ index parameter $(1<\xi<2)$ and w_i is the duration for the policy. The loglikelihood of the Tweedie function is given as,

$$\begin{aligned} l(F(\cdot),\phi,\xi \mid \{z_i,x_i,w_i\}_{i=1}^{n}) &= \sum_{i=1}^{n} \log f_Z(z_i \mid \mu_i,\phi/w_i,\xi) \\ &= \sum_{i=1}^{n} \frac{w_i}{\phi}\left(z_i \frac{\mu_i^{1-\xi}}{1-\xi} - \frac{\mu_i^{2-\xi}}{2-\xi} \right) + \log a(z_i,\phi/w_i,\xi) \end{aligned} \tag{3.19}$$

To find the marginal distribution of Z, integrate out t in (3.19). That is,

$p(z \mid \lambda,\alpha,\varpi) = \sum_{t=0}^{\infty} p(z,t \mid \lambda,\alpha,\varpi)$. The normalizing quantity in Equation (3.14) is expressed as

$$a(z,\phi,\xi) = \frac{1}{z}\sum_{t=1}^{\infty} \frac{z^{t}}{(\xi-1)^{t\alpha}\phi^{t(1+\alpha)}(2-\xi)^{t}\,t!\,\Gamma(t\alpha)} = \frac{1}{z}\sum_{t=1}^{\infty} W_t \tag{3.20}$$

According to Dunn and Smyth (2005), the normalizing quantity in (3.20) does not have a closed form expression. However, methods have been proposed to approximate the quantity reasonably well. Smyth (1996) suggests that for the compound Poisson model, the mean μ is orthogonal to ξ and ϕ. This means that the parameter estimates vary slowly as ϕ and ξ change. Thus, the index parameter (ξ) significantly impacts the estimation of the dispersion parameter (ϕ), which, in turn, has substantial influence on the estimation of asymptotic standard errors of the model coefficients, and predictive measures. For this reason, it is reasonable to specify the optimal index parameter based on the data at hand or based on expert judgement.

Estimating the Premium Function via Gradient Boosting

To estimate the predictor function requires that the index parameter (ξ) of the Tweedie distribution is estimated. If the index parameter ξ is known, the compound Poisson GLM can be estimated using the Fisher's scoring algorithm (McCullagh & Nelder, 1989) and or the TDboost algorithm (Yang et al., 2016). For unknown ξ, parameter estimation can be proceeded using the profile likelihood approach (Cox & Reid, 1987). Thus, to estimate the predictor function in Equation (3.1), optimal index parameter ξ need to be estimated first. The index and dispersion parameters (ξ, ϕ) jointly determine the mean-variance relation $Var(Z_i) = \phi \dfrac{\mu^{\xi}}{w_i}$ of the risk premium. However, in Tweedie models the estimation of μ depends only on ξ. This means that given a fixed ξ, the estimate $\hat{\mu}(\xi)$ can be solved without knowledge of ϕ.

Denote $\sigma = (\xi, \phi)'$. For a given value of σ, the maximum likelihood estimation $\hat{B}(\sigma)$ can be determined using the scoring algorithm, B profiles out of the likelihood and maximize the profile likelihood to estimate σ as

$$\hat{\sigma} = \arg\max_{\sigma} \ell(\sigma \mid z, \hat{B}(\sigma)).$$

Since in general there are no closed forms for Tweedie densities, in likelihood evaluation one must deal with an infinite summation in the normalizing function

$$a(z,\sigma) = \frac{1}{z}\sum_{i=1}^{\infty} W_i .$$

Dunn and Smyth (2005) proposed a series expansion approach, which sums an infinite series arising from Taylor expansion of the characteristic function. Another alternative approach also promulgated by Dunn and Smyth (2008) was based on Fourier inversion approach, which consists of an inversion of the characteristic function based on numerical integration methods for oscillating functions. It is noted that, the two numerical approaches coincide: considering the case where $(1 < \xi < 2)$, the series approach performs very well for small z but gradually loses computational efficiency as z increases, whereas the inversion approach performs very well for large z but gradually fails to provide accurate results as z decreases. For this reason, the inversion approach is preferred for large z and the series approach for small z. This means that the two perform best in different regions of the parameter space. In this study, the algorithm provided in R package "Tweedie" was adopted for our profile likelihood computation. This is because it provides an effective interpolation scheme which blends the inversion and series methods to provide comprehensive evaluation of Tweedie densities across the parameter space.

Conditional on the index parameter, the premium function $F(x_i)$ is estimated as follows;

$$F(x) = \arg\min\{-\ell(F(\cdot), \phi, \xi \mid \{z_i, x_i, w_i\}_{i=1}^{n}) = \arg\min \sum_{i=1}^{n} \psi(z_i, F(x_i) \mid \xi), \quad (3.21)$$

where

$$\psi(z_i, F(x_i) \mid \xi) = \left\{ \frac{z_i \exp(1-\xi) F(x_i)}{1-\xi} + \frac{\exp(2-\xi) F(x_i)}{2-\xi} \right\}$$

Due to non-closed nature of Equation (3.19), the function is estimated iteratively using the forward stage-wise algorithm.

The first initial estimate of $F(x_i)$ is chosen as a constant function that minimizes the negative log-likelihood.

$$F^{[0]} = \arg\min \sum_{i=1}^{n} \psi(z_i, \eta \mid \xi)$$
$$= \left(\log \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i} \right)$$

This corresponds to the best estimate of $F(\cdot)$ without any covariates. Let $\hat{F}^{[m-1]}$ be the current estimate before iteration m. At the m^{th} stage, the base learner $h(x_i; \tau^{[m]})$ is fitted such that

$$\hat{\tau}^{[m]} = \arg\min \sum_{i=1}^{n} [u_i^{[m]} - h(x_i; \tau^{[m]})]^2$$

where, $(u_1^{[m]}, \ldots, u_n^{[m]})$ is the current negative of $\psi(\cdot \mid p)$, thus

$$u_i^{[m]} = \frac{\partial \psi(z_i, F(x_i) \mid \xi)}{\partial F(x_i)} \Big|_{F(x_i)=\hat{F}^{[m-1]}(x_i)} \tag{3.22}$$

$$= w_i \{-z_i \exp[(1-\xi)\hat{F}^{[m-1]}(x_i)] + \exp[(2-\xi)\hat{F}^{[m-1]}(x_i)]\}$$

use L -terminal node regression tree, such that;

$$h(x;\tau^{[m]}) = \sum_{l=1}^{L} u_l^{[m]} I(x \in R_l^{[m]}) \tag{3.23}$$

With the parameters $\xi^{[m]} = \{R_l^{[m]}, u_l^{[m]}\}_{l=1}^{L}$ as the base learner. To find $R_l^{[m]}$ and $u_l^{[m]}$, we use a procedure with a least squares splitting criterion (Friedman et al., 2000), to identify the splitting variables and the corresponding split locations that determine the fitted terminal regions $\{R_l^{[m]}\}_{l=1}^{L}$. $\{R_l^{[m]}\}_{l=1}^{L}$ is estimated jointly with $u_l^{[m]}$ so that

$$\bar{u}_l^{[m]} = mean_{i:x_i \in \hat{R}_l^{[m]}}(u_i^{[m]}), \qquad l = 1,\dots,L$$

Once the base learner $h(x_i;\tau^{[m]})$ has been estimated, the optimal value of the expansion coefficient $\beta^{[m]}$ is determined by a line search.

$$\beta^{[m]} = \arg\min \sum_{i=1}^{n} \psi(z_i, \hat{F}^{[m-1]}(x_i) + \beta h(x_i;\tau^{[m]} \mid \xi)$$

$$= \arg\min_{\beta} \sum_{i=1}^{n} \psi(z_i, \hat{F}^{[m-1]}(x_i) + \beta \sum_{l=1}^{L} u_l^{[m]} I(x_i \in \hat{R}_l^{[m]}) \mid \xi) \tag{3.24}$$

The regression tree in Equation (3.24) predicts a constant value $\bar{u}_l^{[m]}$ within each region $\hat{R}_l^{[m]}$. This means that Equation (3.24) can be solved by a separate line search performed within each respective region $\hat{R}_l^{[m]}$. Thisreduces to finding a best constant $\eta_l^{[m]}$ to improve the current estimate in each region $\hat{R}_l^{[m]}$ based on the following criterion

$$\hat{\eta}_l^{[m]} = \arg\min_{\eta} \sum_{i:x_i \in \hat{R}_l^{[m]}} \psi(z_i, \hat{F}^{[m-1]}(x_i) + \eta \mid \xi), \tag{3.25}$$

The parameters are found by the function

$$\hat{\eta}_l^{[m]} = \log \left\{ \frac{\sum_{i:x_i \in \hat{R}_l^{[m]}} (w_i z_i \exp[(1-\xi)\hat{F}^{[m-1]}(x_i)]}{\sum_{i:x_i \in \hat{R}_l^{[m]}} (w_i \exp[(2-\xi)\hat{F}^{[m-1]}(x_i)]} \right\}, \tag{3.26}$$

Having estimated the parameters $\{\eta_l^{[m]}\}_{l=1}^{L}$, the current estimate of the function $\hat{F}^{[m-1]}(x_i)$ is then updated in each corresponding region

$$\hat{F}^{[m]}(x_i) = \hat{F}^{[m-1]}(x_i) + \zeta \hat{\eta}_l^{[m]} I(x \in \hat{R}_l^{[m]}),, \qquad l = 1, \ldots, L \tag{3.27}$$

where ζ is the shrinkage parameter $0 < \zeta \leq 1$.

Following Freidman (2001), ζ is set to 0.005 at the implementation stage. The steps are repeated M times and the final estimate $\hat{F}^{[M]}(x)$ is thus reported. Summary of the TDboost model is in Table 5.

Table 5: Schematic Overview of TDboost Algorithm.

Initialize with $\hat{F}^{[0]} = \log \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$

for $m = 1, \ldots, M$ repeat steps 2a to 2d

a. compute the negative gradient $\{\mu_i^{[m]}\}_{i=1}^{n}$

$$= w_i \left\{ -z_i \exp[(1-\xi)\hat{F}^{[m-1]}(x_i)] + \exp[(2-\xi)\hat{F}^{[m-1]}(x_i)] \right\}$$

$;\ i = 1, \ldots, n$

b. fit the negative gradient vector $\{\mu_i^{[m]}\}_{i=1}^{n}$ to $x_1, \ldots, x_n$ by L terminal node regression tree. This gives us the partitions $\{\hat{R}_l^{[m]}\}_{l=1}^{L}$

compute the terminal node predictions $\eta_l^{[m]}$ for each region $\{\hat{R}_l^{[m]}\}_{l=1}^{L}$, where

$$\eta_l^{[m]} = \log \left\{ \frac{\sum_{i:x_i \in \hat{R}_l^{[m]}} w_i z_i \exp(1-\xi)\hat{F}^{[m-1]}(x_i)}{\sum_{i:x_i \in \hat{R}_l^{[m]}} w_i \exp(2-\xi)\hat{F}^{[m-1]}(x_i)} \right\}; l = 1, \ldots, L$$

c. update $\hat{F}^{[m]}(x)$ for each region $\{\hat{R}_l^{[m]}\}_{l=1}^{L}$

$$\hat{F}^{[m]}(x) = \hat{F}^{[m-1]}(x) + \zeta \hat{\eta}_l^{[m]} I(x \in \hat{R}_l^{[m]}), l = 1, \ldots, L$$

d. Report $\hat{F}^{[M]}(x)$ as the final estimate

Model Interpretaion

This section discusses the interpretation of the risk premium model described in the previous section. The relevance of a statistical model is not only its ability to predict accurately but also how well model parameters could be interpreted in terms of the contribution of each predictor quantitatively as well as pictorial views through partial dependency plots and variable imporatance.

This section briefly explains these concepts in the context of the gradient boosting framework.

Marginal Effects of Predictors

The main and interaction effects of the variables in the boosted Tweedie model can be extracted. A tree with L terminal nodes produces a function approximation of p predictors, with interaction order of at most $\min(L-1, p)$. For instance, when $L=2$, we have an additive TDboost model with only the main effects of the predictors, since it is a function based on a single splitting variable in each tree. When $L=3$, it allows both main effects and second order effects.

Given the training data set $\{y_i, x_i\}_{i=1}^{n}$ with p-dimensionpal input vector $x=[x_1,\dots,x_p]'$, let q_S be a subset of size s, such that $q_S=\{q_1,\dots,q_S\}\subset\{x_1,\dots,x_p\}$.

To study the main effect of the variable j, let $q_s=\{q_j\}$ and to study the second order interaction of variables i and j, let $q_s=\{q_i,q_j\}$. Let $q_{\setminus s}$ be the complement set of q_S such that $q_{\setminus s}\cup q_s=\{x_1,\dots,x_p\}$. Let the prediction function $\hat{F}(q_s\mid q_{\setminus s})$ be a function of the subset q_S conditioned on specific values of $q_{\setminus s}$. The partial dependence of $\hat{F}(x)$ on $q_{\setminus s}$ can be formulated as $\hat{F}(q_S\mid q_{\setminus S})$ averaged over the marginal density of the complement subset $q_{\setminus s}$

$$\hat{F}_S(q_S) = \int \hat{F}(q_S \mid q_{\setminus s}) p_{\setminus s}(q_{\setminus s}) dq_{\setminus s} \tag{3.28}$$

where $p_{\setminus s} = \int p(x) dq_{\setminus s}$ is the marginal density of $q_{\setminus s}$.

Equation (3.18) is estimated by $\bar{F}_s(q_s) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}(q_s \mid q_{\setminus s,i})$, where $\{q_{\setminus s,i}\}_{i=1}^{n}$ are evaluated at the training dataset. Pictorial view can be obtained by plotting $(F_s(q_s), q_s)$.

Variable Importance

One of the key ingredients in model building is ability to extract relevant predictors. The 'TDboost' model (Yang, et al., 2016, Friedman, 2001) defines a variable importance measure for each candidate predictor X_j in the set $X = \{X_1, ..., X_p\}$ in terms of proportion of influence on response variable Y. The main advantage of this variable selection procedure as compared to univariate screening methods, is that the approach considers the impact of each individual predictor as well as multivariate interaction among predictors simultaneously.

Breiman et al., (1984) defined the variable importance measure as $I_{X_j}(T_m)$ of the variable X_j. In a single tree T_m is defined as the total heterogeneity reduction of the response variable Y produced by X_j, which can be estimated by adding up all the decreases in the squared error reductions $\hat{\delta}_l$ obtained in all $L-1$ internal nodes when X_j is chosen as the splitting variable.

Let $I_{jl} = I(\nu(X_j = l))$. Then

$$I_{X_j}T_m = \sum_{l=1}^{L-1} \hat{\delta}_l I_{jl} \tag{3.29}$$

where $\hat{\delta}_l$ is defined as the squared error difference between the constant fit and the two sub-region fits achieved by splitting the region associated with internal node l into the left and right regions.

Friedman (2001) extended the variable importance measure I_{X_j} for the boosting model with the combination of M regression trees, by average Equation over $\{T_1,...,T_M\}$ such that

$$I_{X_j} = \frac{1}{M}\sum_{m=1}^{M} I_{X_j}(\mathrm{T}_m) \tag{3.30}$$

Despite the wide use of the variable importance measure as introduced by Breiman et al., (1984), Kestemont and Paris (1985), White and Liu (1994) among others suggested that the variable importance described in Equations (3.29) and (3.30) are biased in the sense that even if X_j is not informative to the response variable, X_j may still be used as a splitting variable. Hence the variable importance measure in Equation (3.30) is not zero.

Following Sandri and Zuccolotto (2010) and to avoid variable selection bias, the study computed an adjusted variable importance measure for each explanatory variable by permutating each X_j described in six steps below;

1) For $s = 1,...,S$ repeat steps (2) to (4)

2) Generate a matrix Q^s by randomly permutating (without replacement) the n rows of the design matrix X, while keeping the order of columns unchanged

3) Create an $n \times 2p$ matrix $\overline{x}^s = [x, Q^s]$ by binding Q^s with matrix X by column

4) Use the data $\{y, \overline{x}^s\}$ to fit the model and compute the variable importance measures $I^s_{X_j}$ for X_j and $I^s_{Q_j}$ for Q^s_j, where Q^s_j (j^{th} column of Q^s) is the pseudo-predictor corresponding to X_j.

5) Compute the variable importance measure $\overline{I}_{X_j}$ as the average of $I^s_{X_j}$ and the baseline $\overline{I}_{Q_j}$ as the average of $\overline{I}^s_{Q_j}$, ,where

$$\overline{I}_{X_j} = \frac{1}{S}\sum_{s=1}^{S} I^s_{X_j} ; \overline{I}_{Q_j} = \frac{1}{S}\sum_{s=1}^{S} I^s_{Q_j}.$$

6) Report the variable importance measure as $\overline{I}^{adj}_{X_j} = \overline{I}_{X_j} - \overline{I}_{Q_j}$ for the variable X_j

The basic idea of the above procedure is that, the permutation breaks the association between the response variable and each of the pseudo-predictors Q^s_j while preserving the association between Q^s_j and $Q^s_k, (k \neq j)$. Since Q^s_j is re-shuffled from X_j, Q^s_j has the same number of possible splits as the corresponding predictor X_j and has approximately the same probability of being selected in split nodes. Hence $\overline{I}_{Q_j}$ could be viewed as bias approximation of the importance of the predictor variable X_j.

Implementation Technique

This section discusses the choice of meta parameters required to enhance the model framework in Figure 3: L (the size of the trees), M (the number of boosting steps or iteration) and ν (the shrinkage factor). The study set up the optimal number of boosting iterations to avoid over-fitting as well as improve out-of-sample predictions. Zhang and Yu (2005) recommend regulating the boosting procedure by limiting the number of boosting iterations M and the shrinkage factor.

Friedman (2001), Ridgeway (2007), Elith et al. (2008), have shown that the predictive accuracy is almost always better with smaller shrinkage factor ν at the cost of more computing time. Thus, smaller shrinkage factor usually requires a larger number of boosting iterations and hence more computing time. The study chose $\nu = 0.005$ and determine M by cross validation using the data. Lastly, the value of L reflects the true interaction order in the underlying model. Given no prior knowledge on such information on optimal L. The study thus chose the optimal L and M using K - fold cross validation. Starting with a fixed value of L the data are split into K equal parts roughly.

Let an index function $\pi(i):\{1,...,n\} \rightarrow \{1,...,K\}$ indicate the fold to which observation i is allocated. Each time the k^{th} fold of the data is removed $(k = 1,...,K)$, the model is trained using the remaining $K-1$ folds.

Formally, denote resulting model by $\hat{F}_{-k}^{[M]}(\mathrm{x})$ and compute the validation loss by predicting on each k^{th} fold of the data removed:

$$CV(M,L) = \hat{F}^{[M]}_{-\pi(i)}(x_i; L \mid \xi) \tag{3.31}$$

The optimal M is thus selected at the point which the minimum validation loss was reached.

$$\hat{M}_L = \underset{M}{\arg\min}\, CV(M,L) \tag{3.32}$$

To select L, the process is repeated for several number of L, for instance $(L = 2,\ldots,n)$ and choose the one with the smallest minimum generalization error which is a measure of how accurately an algorithm is able to predict outcome values for test data or previously unseen data.

$$\hat{L} = \underset{L}{\arg\min}\, CV(L, \hat{M}_L) \tag{3.33}$$

It is important to note that for a given shrinkage parameter ν, fitting trees with higher L leads to smaller M required to reach the optimal threshold (minimum error).

Model Evaluation Technique

One major ingredient in statistical models is the model ability to predict well out-of-sample. After fitting on the training data set, the study predicts the risk premium $P(x) = \hat{\mu}(x)$ by applying the MMGB with location model with location risk to an independent held-out sample called testing set. The result is compared with other competing models. It is worth noting that when measuring the differences between predicted premiums $P(x)$ and real losses y, the mean squared error loss or the mean absolute loss is not appropriate because the losses have high percentage of zeros and high skewed to the right. Hence, an alternative measure is required. The study thus considered the ordered Lorenz curve and the associated Gini index proposed by Frees et al., (2011). The efficiency of different predictive models can be assessed by comparing their Gini indices.

To explain briefly, the idea of the ordered Lorenz curve. Let $S(x)$ be the "based premium", which is calculated using the existing premium prediction model, and let $P(x)$ be the "competing premium" which is computed based on an alternative premium prediction model. The ordered Lorenz curve is such that, both the distribution of losses and the distribution of premiums are sorted based on the relative premium. $R_i = R(x_i) = \frac{S(x_i)}{P(x_i)}$ (Frees et al., 2011; Werner, Modlin and Claudine, 2010) . The distribution for the ordered premium and loss are respectively given by

$$\hat{F}_P(s) = \frac{\sum_{i=1}^{n} P(x_i) I(R_i \le z)}{\sum_{i=1}^{n} P(x_i)} \quad \text{and} \quad \hat{F}_L(s) = \frac{\sum_{i=1}^{n} P(s_i) I(R_i \le z)}{\sum_{i=1}^{n} P(s_i)} \; ; \qquad (3.34)$$

$I(\cdot)$ is an indicator function, that returns a 1, if the event is true and return a 0, if the event is false. We then compute the Ordered Lorenz curve and Gini coefficient for the portfolio. The two empirical distributions are based on the same sort order, which makes it possible to compare the premium and loss distributions for the same policyholder. We then compute the ordered Lorenz curve and Gini coefficient for the portfolio. Thus, the graph of $(\hat{F}_P(s), \hat{F}_L(s))$ is called ordered Lorenz curve, where

When the proportion of losses equals the proportion of premiums for the insurer, the curve results in 45-degree line. This is known as "the line of equality". Twice the area between the ordered Lorenz curve and the line of equality measures the discrepancy between the premium and loss distributions, which is defined as the Gini index. Curves below the line of equality indicate that, given prior information of the relative premium, an insurer could identify the profitable contracts, whose premiums are greater than losses. Hence a larger Gini index or larger area between the line of equality and the curve below suggest a more favorable model.

The Gini index could be computed using Trapezium or Simpson's rule as follows; suppose that the empirical ordered Lorenz curve is given by $\{(a_0 = 0, b_0 = 0), (a_1, b_1), \ldots, (a_n = 1, b_n = 1)\}$ for a sample of size n, we use $a_j = \hat{F}_P(R_j)$ and $b_j = \hat{F}_L(R_j)$.

The Gini coefficient could also be used to assess the profitability of portfolios.

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{F}_P(R_i) - \hat{F}_L(R_i)\right) \approx \frac{\hat{Gini}}{2} \tag{3.35}$$

This means that insurers that adopt a pricing structure with a large Gini index are more likely to enjoy profitable portfolios. Denoting the tariff -based score by $P1 = P(x)$, other competing models represented by $P2 = S(x)$, and and another represented by $P3 = T(x)$. The relativity premium is given by $R_1(x) = \frac{S(x)}{P(x)}$ and $R_2(x) = \frac{S(x)}{T(x)}$.

The desired expected loss is denoted by $S(x)$. If $R(x)$ is small, then a small loss is expected relative to the premium suggested. If $R(x)$ is large, then a large loss is expected relative to the premium suggested. Gini index becomes larger as one uses a "more refined" insurance score.

Following Frees et al. (2013), the study specifies and uses the predictions from each model as base premium $S(x)$ and use the predictions from the remaining models as the competing premium $P(x)$ to compute the Gini indices. The entire procedure of the data splitting and Gini index computation are repeated 20 times and a matrix of the averaged Gini indices and standard errors are reported. Using "minimax" strategy the "best" model is selected. Thus, the minimax strategy selects the base premium model that are least vulnerable to competing premium models. This means that the model that provides the smallest of the maximal Gini indices is selected.

Chapter Summary

The chapter described the methods used in the study. The section describes the metods used in assessing the industry performance. It also explains the compound-Poisson distributions and in particular the Tweedie model as the appropriate technique to be used in predicting the risk premium considering the nature of dataset for the non-life insurance industry. Given the difficiencies identified in the literature the study utilized three theories; the Markov theory which helped to integrate the external risk factor (the location risk), the exponential dispersion theory was used to establish the model framework and the theory of gradient boosting for estimation of the model parameters. In the classical models reviewed, it is assumed that claim arrivals process is constant for all geographical regions. However, the MMGB model assumes that the number claims vary from one region to another and from time to time. The contribution of each factor considered in the model is measured by 'variable importance'. The model validation utilizes Ordered Lorenze curve and Gini coefficient by Frees, Myers and Cummings (2011, 2013).

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

This chapter presents the results and test efficiency of the methodology developed using datasets obtained from two sources. It first starts with a background and preliminary analysis of the performance of the insurance industry during the period 2012 to 2016 using Grey Systems Theory. The chapter then presents the findings in line with the objectives of the study. The insurance data was obtained from the database of a major insurance company in Ghana and the accident data obtained from Ghana Road Safety Commission (NRSC). The chapter ends with a discussion of the model, its efficiency and usefulness.

Assets and Liabilities

Aggregate insurance industry assets stood at GHS 3.76bn as at the end of 2016, representing a growth of 23% from GHS 3.06bn in 2015. Life insurance contribution to total industry assets as at 2016 was GHS 2.25bn representing 60% as against non-life insurance contribution of GHS 1.51bn representing 40%. Total assets mainly consist of Investments, Cash, Property, Plant and Equipment (PPE). Figure 2 shows pictorial analysis of yearly aggregate industry contribution of total assets between life and non-life from 2014 to 2016, while Figure 3 shows the composition of non-life insurance total asset from 2014 to 2016.

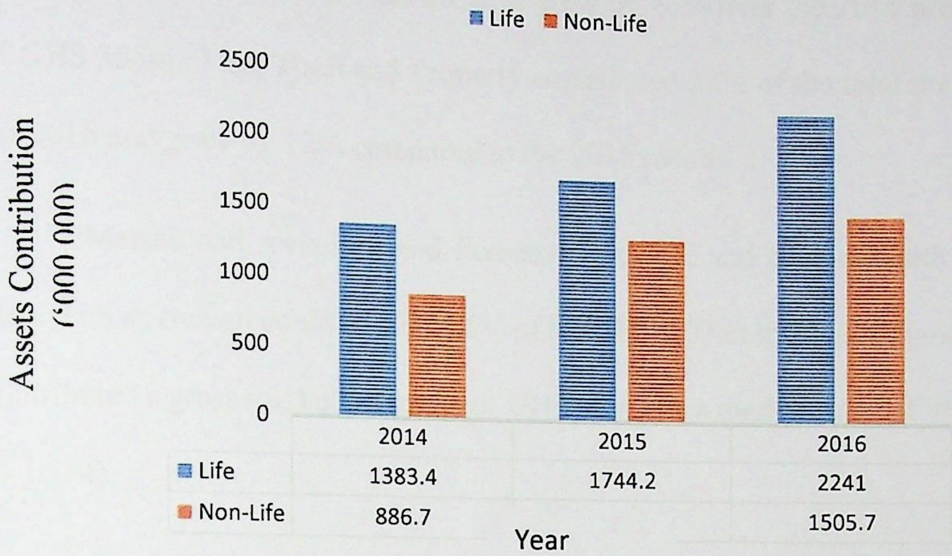


Figure 2: Yearly Aggregate Industry Contribution of Total Assets (Ghc).

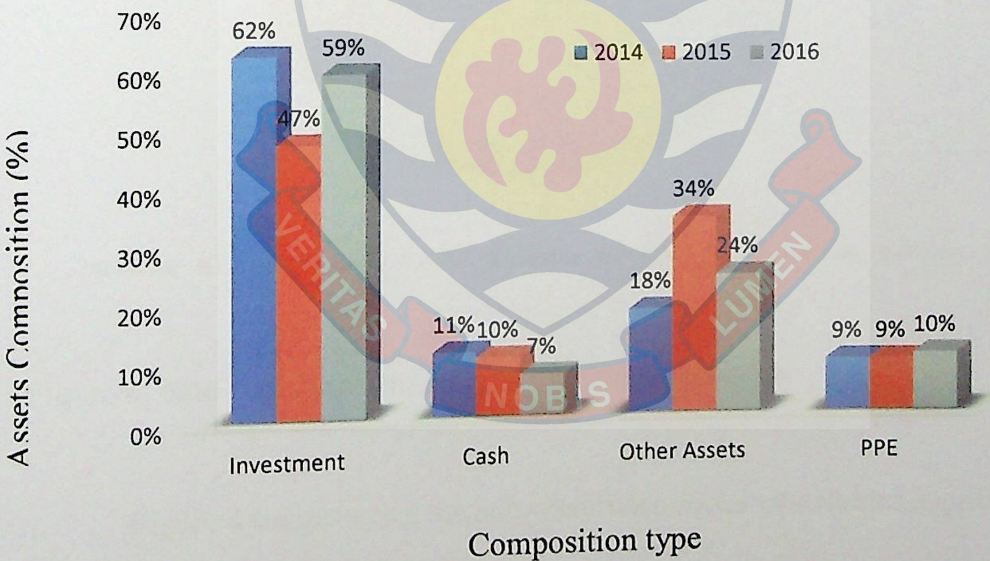


Figure 3: Composition of Total Non-Life Insurance Industry Asset.

Premium Contribution by Class of Business

The motor insurance class of business has dominated the non-life insurance sector with a premium income of GHS 518m, representing 50% of the overall premium income for 2016. It grew by 46% over the 2015 premium of GHS 355m. Fire, Theft and Property contributed 22% of the total premium for 2016 and grew by 12% compared to the 2015 results.

Marine and Aviation, and Personal Accident and Medical both came third with an overall contribution of 6% of the GHS1.07bn in 2016. Engineering contributed a gross market premium of GHS 90m and a market share of 9% (see Figure 4).

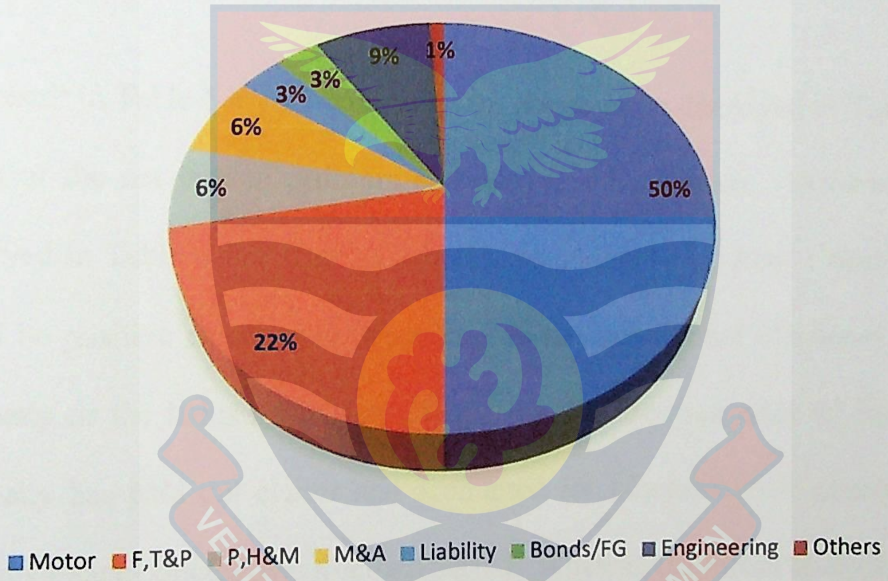


Figure 4: Non-Life Premium income by class of business.

Figure 4 suggests that the auto-insurance sector contributes significantly (50%) in terms of aggregate premium income for the industry and thus have the largest market share within the non-life industry.

Table 6: Growth in Gross Premium for Non-Life Insurance (2012-2016)

Year	Premium Income (GHS)	Growth Rate (%)
2012	494,891,864	-
2013	582,456,306	17.7
2014	659,262,969	13.2
2015	854,825,825	29.7
2016	1,070,057,051	25.2

The result in Table 6 is consistent with the information displayed in Table 7 in terms of the net written primium and net earned premium. However, it is observed in Table 6 that there is aconsistent underwriting loss. Underwriting could be positive or negative depending on the operational efficiency of the company or the industry. Underwriting losses ix incurred after an insurance company has paid out claims and accounted for administrative expenses for their insurance policies over a certain period. When an insurance company must pay out more claims than envisaged, and the premiums brought in do not cover the overall expenses, it results in underwriting loss. The amount reflects the inefficiency of the insurance company's underwriting activities. Underwriting losses mainly arise due to the result of huge claims and disproportionate expenses. Table 6 records underwriting losses of about 24million in 2013, 89million in 2014, 65million in 2015 and 25million in 2016. This suggests an operational inefficiency or disproportionate pricing regime.

Table 7: Non-Life insurance industry performance indicators (2013-2016)

Indicator	2016 (GHS'm)	2015 (GHS'm)	2014 (GHS'm)	2013 (GHS'm)
Gross Premium	1,070	854	659	582
Reinsurance	402	355	272	204
Net Written Premium	668	509	396	366
Net Earned Premium	597	475	394	325
Gross Claims incurred	329	555	237	173
Management Expenses	445	273	274	210
Commissions	124	103	80	64
Underwriting Results	-25	-65	-89	-24
Investment Income	160	118	89	53
Other Income	25	30	22	20
Profit After Tax	88	61	22	73

Source: NIC annual report (2016)

The study observed combined ratio which is also a significant indicator and the single best measure of an insurer's underwriting and operational efficiency. It is a summation of total expense ratio and claims ratio. Though the ratio does not entirely measure total profitability since it does account for other investment incomes, it is an important measure should there be any systemic failure such as any fiscal crisis or budgetary crisis. In general, a ratio of less than 100% indicates underwriting profitability, while a ratio of more than 100% usually indicate a loss.

Figure 5 displays the industry level combined ratio from 2010 to 2016. A ratio beyond 100% represent operational loss. The result suggests that industry players need to improve their underwriting efficiencies for a robust profitability. On average all the 25 insurance companies have combined ratio of above 100% for the year under review as shown in Table 8.

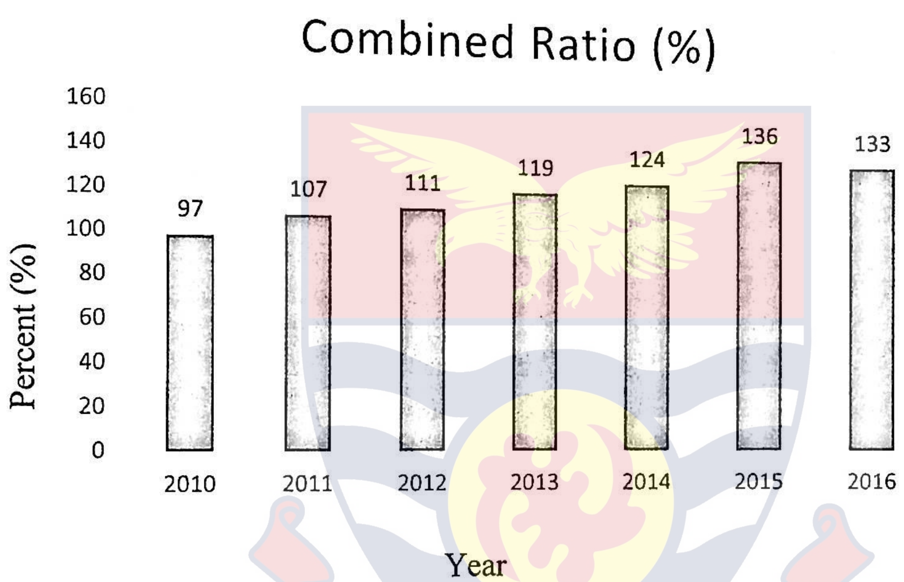


Figure 5: Industry underwriting performance measure with combined ratio

Table 8: Combined ratio for 25 Non-life Insurance Companies (2012-2016)

No	Company	2012	2013	2014	2015	2016	Average
1	Activa International Insurance.	106	107	143	151	142	1.298
2	Allianz Insurance Company Limited	149	154	170	196	158	1.654
3	Best Assurance Company.	-	-	-	-	151	1.51
4	Donewell Insurance Comp.	132	107	97	80	99	1.03
5	Entreprise Insurance Company Limited	108	103	111	114	96	1.064
6	Equity Assurance Company	92	91	97	103	99	0.964
7	Ghana Union Assurance Company Limited	130	134	129	125	135	1.306
8	Glico General Insurance Limited	115	114	153	130	116	1.256
9	Heritage Insurance Company Limited	98	246	-	-	-	1.72
10	Hollard Insurance Company Limited	135	133	150	136	122	1.352
11	Imperial General Insurance Company Limited	-	160	252	166	153	1.462
12	Milennium Insurance Company Limited	98	94	139	185	127	1.286
13	NSIA Ghana Insurance Company Limited	96	178	185	140	176	1.55
14	Phoenix Insurance Company Limited	96	150	122	138	141	1.294
15	Prime Insurance Company Limited	-	364	593	132	129	2.436
16	Priority Insurance Company Limited	-	175	131	109	100	1.03
17	Provident Insurance Company Limited	111	142	140	142	108	1.286
18	Quality Insurance Company Limited	94	108	104	109	105	1.04
19	RegencyNEM Insurance Ghana Limited	95	95	90	117	106	1.006
20	Saham insurance Company Limited	142	128	115	141	221	1.494
21	SIC Insurance Company	134	131	158	137	158	1.436
22	Star Assurance Company Limited	100	108	135	130	128	1.202
23	Unique Insurance Company	168	151	119	128	119	1.37
24	Vanguard Assurance Company Limited	100	101	117	131	113	1.124
25	Wapic Insurance (Gh), Limited	104	130	106	197	202	1.478

Table 8 summarizes the combined ratio recorded for 25 licensed companies in Ghana from 2012 to 2016. It could be observed that the combined ratio for most of the companies recorded above 100%. This is an indication that most of the companies underwriting operations are inefficient. This is quite risky and the way to minimize it is to look at reducing expenditure, improving income or reducing risk. One of the key ways in reducing actuarial risk is to charge actuarially fair premiums.

Empirical Analysis of Financial Data

Table 9 shows the description of codes used to represent the insurance companies in the analysis.

Table 9: Description of Codes Used in the Analysis

Code	Company Name
C1	Activa International Insurance Company
C2	Allianz Insurance Comp. Ltd
C3	Best Assurance Company Limited
C4	Donewell Insurance Company Limited
C5	Entreprise Insurance
C6	Equity Assurance Co.
C7	Ghana Union Assurance Company
C8	Glico General Insurance Limited
C9	Heritage Insurance Company Limited
C10	Hollard Insurance Company Limited
C11	Imperial General Insurance Co.
C12	Milennium Insurance Company Limited
C13	NSIA Ghana Insurance Company Limited
C14	Phoenix Insurance Company Limited
C15	Prime Insurance Company Limited
C16	Priority Insurance Company Limited
C17	Provident Insurance Company Limited
C18	Quality Insurance Company Limited
C19	RegencyNEM Insurance Ghana
C20	Saham insurance Company Limited
C21	SIC Insurance Co.
C22	Star Assurance Co.
C23	Unique Insurance Company Limited
C24	Vanguard Assurance C
C25	Wapic Insurance (Gh), Limited

As indicated chapter four, financial ratios are usually preferred in firm's performance evaluations. In this study 25 non-life insurance companies registered by NIC in Ghana as at 2016 is evaluated for the 2012 to 2016 period with eight financial ratios. The financial ratios were limited to eight, due to the inability to obtain healthy data for other computations. More so, it is limited to 2012- 2016 since 2017 financial statement is not yet published. The five-year averages of the financial ratios for each company were computed. The summary of financial ratios used in the analysis is displayed in Table 10.

Table 10: Financial ratios of Non-Life Insurance Companies (2012-2016)

	Capital Asset Ratio				Operating Eficiency		Profitability	
	R1	R2	R3	R4	R5	R6	R7	R8
C1	2.602	1.352	0.822	0.314	1.298	0.290	0.048	0.178
C2	4.616	1.122	1.368	0.318	1.654	0.526	-0.046	-0.150
C3	0.32	0.070	0.240	0.770	1.510	0.026	0.020	0.030
C4	1.67	1.232	1.280	0.776	1.030	0.336	0.078	0.176
C5	2.042	0.624	1.316	0.648	1.064	0.556	0.120	0.256
C6	2.03	0.914	1.718	0.842	0.964	0.236	0.092	0.234
C7	0.634	4.852	0.242	0.374	1.306	0.564	0.034	0.058
C8	2.77	1.126	1.340	0.486	1.256	0.538	-0.002	-0.012
C9	3.235	2.270	2.600	0.795	1.720	0.144	-0.305	0.875
C10	3.5	1.054	1.494	0.442	1.352	0.532	0.054	0.178
C11	0.426	0.304	0.336	0.606	1.462	0.164	-0.008	-0.018
C12	0.85	0.314	0.690	0.812	1.286	0.386	0.062	0.100
C13	0.966	1.096	0.792	0.796	1.550	0.428	-0.038	-0.144
C14	1.892	0.654	1.352	0.706	1.294	0.402	0.060	0.134
C15	0.352	1.396	0.344	0.610	2.436	0.314	-0.468	0.388
C16	0.252	0.234	0.216	0.838	1.030	0.864	0.080	0.108
C17	0.802	0.960	0.620	0.758	1.286	0.344	0.120	0.284
C18	2.098	1.468	1.712	0.814	1.040	0.278	0.058	0.162
C19	2.416	0.862	2.014	0.836	1.006	0.242	0.094	0.224
C20	1.902	0.938	1.106	0.572	1.494	0.382	-0.026	-0.048
C21	1.682	1.566	1.068	0.642	1.436	0.440	-0.008	-0.012
C22	1.644	0.884	1.004	0.612	1.202	0.326	0.080	0.176
C23	0.928	2.102	0.750	0.816	1.370	0.378	-0.088	0.092
C24	2.772	0.662	1.940	0.698	1.124	0.416	0.058	0.178
C25	1.666	0.608	1.590	0.868	1.478	0.510	-0.014	-0.056

Table 10 summarizes the general outlook of the non-life insurance companies. The maximum and minimum gross written premium was 4.62 and 0.252, respectively, corresponding to Allianz Assurance limit ed (C2) and Priority insurance (C16). The maximum and minimum net insurance risk 2.6 and 0.216 corresponding to Heritage Insurance Company and Priority insurance.

Table 11: Comparison Matrix

	R1	R2	R3	R4	R5	R6	R7	R8
Reference	0.252	0.07	0.216	0.314	0.964	0.026	0.12	0.875
C1	2.602	1.352	0.822	0.314	1.298	0.290	0.048	0.178
C2	4.616	1.122	1.368	0.318	1.654	0.526	-0.046	-0.150
C3	0.32	0.070	0.240	0.770	1.510	0.026	0.020	0.030
C4	1.67	1.232	1.280	0.776	1.030	0.336	0.078	0.176
C5	2.042	0.624	1.316	0.648	1.064	0.556	0.120	0.256
C6	2.03	0.914	1.718	0.842	0.964	0.236	0.092	0.234
C7	0.634	4.852	0.242	0.374	1.306	0.564	0.034	0.058
C8	2.77	1.126	1.340	0.486	1.256	0.538	-0.002	-0.012
C9	3.235	2.270	2.600	0.795	1.720	0.144	-0.305	0.875
C10	3.5	1.054	1.494	0.442	1.352	0.532	0.054	0.178
C11	0.426	0.304	0.336	0.606	1.462	0.164	-0.008	-0.018
C12	0.85	0.314	0.690	0.812	1.286	0.386	0.062	0.100
C13	0.966	1.096	0.792	0.796	1.550	0.428	-0.038	-0.144
C14	1.892	0.654	1.352	0.706	1.294	0.402	0.060	0.134
C15	0.352	1.396	0.344	0.610	2.436	0.314	-0.468	0.388
C16	0.252	0.234	0.216	0.838	1.030	0.864	0.080	0.108
C17	0.802	0.960	0.620	0.758	1.286	0.344	0.120	0.284
C18	2.098	1.468	1.712	0.814	1.040	0.278	0.058	0.162
C19	2.416	0.862	2.014	0.836	1.006	0.242	0.094	0.224
C20	1.902	0.938	1.106	0.572	1.494	0.382	-0.026	-0.048
C21	1.682	1.566	1.068	0.642	1.436	0.440	-0.008	-0.012
C22	1.644	0.884	1.004	0.612	1.202	0.326	0.080	0.176
C23	0.928	2.102	0.750	0.816	1.370	0.378	-0.088	0.092
C24	2.772	0.662	1.940	0.698	1.124	0.416	0.058	0.178
C25	1.666	0.608	1.590	0.868	1.478	0.510	-0.014	-0.056

For a detailed company analysis, the research forms the Comparison matrix shown in Table 11 and computes the Normalized matrix in Table 12. Absolute value Tables shown in Table 13, and finaly obtain the grey relational coefficient and grey rational grades in Table 14 and Table 15, respectively. Gross insurance risk (R1) is the ratio of Gross written Premium to Equity capital. Technical Reserve cover (R2) is the ratio of Technical provisions to liquid investment. The company that recorded highest for R1 is C2 (Allianze insurance). R2: records the technical reserve cover. The maximum recorded was 4.852 (485.5%) and the minimum was 0.07 (7%) corresponding to Ghana Union Insurance and Best Insurance Company. Ratios above 100% mean that the company do not have enough liquid investment backing their technical provisions. Thus about 48% of the companies have the technical research beyond 100% (See R2 in Table 11). Net insurance risk (R3) measures the ability of the company to absorb unforeseen shocks. This figure suggests that Priority insurance can withstand shocks and losses than the rest of the companies.

R4 is premium retention ratio, the maximum rate was 0.868 (86.8%) and the minimum rate is 0.314 (31.4%) corresponding to Wapic Insurance and Activa Insurance, respectively. High retentions are usually considered riskier. High retention will require sufficient capital to support the insurer. This means that Wapic Insurance is relatively riskier than Activa Insurance. R5 is combined ratio, the maximum ratio recorded among the companies is 2.436 (243.6%) and the minimum is 0.964 (96.4%). This correspond to the companies' Prime Insurance and Equity Assurance. A ratio of more than 100% represent a loss. From Table 10, apart from Equity Assurance, the rest of the companies at operating at loss ceteris paribus; discounting other investments.

R6: Loss ratio is a key ratio which indicates how well an insurance company pays claims and to some extent, of fair customer treatment. From the Table 14, Priority insurance loss ratio on average is 86.4%, higher than the rest of the companies.

According to Asset profitability R7 and R8: Return on Asset, This, ratio is an indicator of general profitability of the insurer. It is calculated as after-tax profits divided by total assets. It seeks to measure the efficiency with which management utilize the assets of the company to generate returns of the various stakeholders. The highest profit was recorded by Enterprise Insurance of 12%. Allianz Insurance company however experience the biggest loss (46.8%). Normalized values and the reference sequence are presented in Table 12. The reference sequence was determined by selecting the largest normalized value for each financial ratio. Thus, after the comparison matrix is formed, normalized matrix is obtained. By considering that business owners and managers usually prefer the capital adequacy to be low, operating efficiency indicators nominal (within some target range) while the profitability indicators are expected to be high. Accordingly, Equation (3.3) for capital adequacy indicators, Equation (3.2) for profitability indications and Equation (3.4) for operating efficiency indicators are used in the formation of the normalized matrix, shown in Table 12.

Table 12: Normalized Matrix

	R1	R2	R3	R4	R5	R6	R7	R8
Reference	1	1	1	1	1	1	1	1
C1	0.4615	0.7319	0.7458	1.0000	0.1983	2.1742	0.8776	0.3200
C2	0.0000	0.7800	0.5168	0.9928	0.3802	1.1742	0.7177	0.0000
C3	0.9844	1.0000	0.9899	0.1769	0.5741	0.2803	0.8299	0.1756
C4	0.6751	0.7570	0.5537	0.1661	0.4956	2.1742	0.9286	0.3180
C5	0.5898	0.8841	0.5386	0.3971	0.2342	1.0000	1.0000	0.3961
C6	0.5926	0.8235	0.3700	0.0469	0.2527	0.1667	0.9524	0.3746
C7	0.9125	0.0000	0.9891	0.8917	0.1983	1.3788	0.8537	0.2029
C8	0.4230	0.7792	0.5285	0.6895	0.3845	0.1364	0.7925	0.1346
C9	0.3165	0.5399	0.0000	0.1318	0.3573	0.2348	0.2772	1.0000
C10	0.2557	0.7942	0.4639	0.7690	0.6100	1.7273	0.8878	0.3200
C11	110.9601	0.9511	0.9497	0.4729	0.4096	0.2576	0.7823	0.1288
C12	0.8630	0.9490	0.8012	0.1011	0.4695	1.6515	0.9014	0.2439
C13	0.8364	0.7854	0.7584	0.1300	0.3736	0.8106	0.7313	0.0059
C14	0.6242	0.8779	0.5235	0.2924	0.5174	0.6515	0.8980	0.2771
C15	0.9771	0.7227	0.9463	0.4657	0.3780	0.7500	0.0000	0.5249
C16	1.0000	0.9657	1.0000	0.0542	1.0000	1.0833	0.9320	0.2517
C17	0.8740	0.8139	0.8305	0.1986	0.2342	1.0000	1.0000	0.4234
C18	0.5770	0.7077	0.3725	0.0975	0.3736	0.9697	0.8946	0.3044
C19	0.5041	0.8344	0.2458	0.0578	0.2397	1.2197	0.9558	0.3649
C20	0.6219	0.8185	0.6267	0.5343	0.2211	1.3561	0.7517	0.0995
C21	0.6723	0.6872	0.6426	0.4079	0.4869	0.8258	0.7823	0.1346
C22	0.6810	0.8298	0.6695	0.4621	0.4553	0.6061	0.9320	0.3180
C23	0.8451	0.5751	0.7760	0.0939	0.3279	1.0379	0.6463	0.2361
C24	0.4225	0.8762	0.2768	0.3069	0.4194	0.8409	0.8946	0.3200
C25	0.6760	0.8875	0.4237	0.0000	0.2854	0.6970	0.7721	0.0917

After forming the normalized matrix, Absolute Value Table is constructed by using Equation (4.8). Thus, the distance between normalized values and reference values are calculated. Table 13 is constructed by subtracting normalized values from reference values.

Table 13: Absolute Values Table

	R1	R2	R3	R4	R5	R6	R7	R8
C1	0.5385	0.2681	0.2542	0.0000	0.8017	1.1742	0.1224	0.6800
C2	1.0000	0.2200	0.4832	0.0072	0.6198	0.1742	0.2823	1.0000
C3	0.0156	0.0000	0.0101	0.8231	0.4259	0.7197	0.1701	0.8244
C4	0.3249	0.2430	0.4463	0.8339	0.5044	1.1742	0.0714	0.6820
C5	0.4102	0.1159	0.4614	0.6029	0.7658	0.0000	0.0000	0.6039
C6	0.4074	0.1765	0.6300	0.9531	0.7473	0.8333	0.0476	0.6254
C7	0.0875	1.0000	0.0109	0.1083	0.8017	0.3788	0.1463	0.7971
C8	0.5770	0.2208	0.4715	0.3105	0.6155	0.8636	0.2075	0.8654
C9	0.6835	0.4601	1.0000	0.8682	0.6427	0.7652	0.7228	0.0000
C10	0.7443	0.2058	0.5361	0.2310	0.3900	0.7273	0.1122	0.6800
C11	0.0399	0.0489	0.0503	0.5271	0.5904	0.7424	0.2177	0.8712
C12	0.1370	0.0510	0.1988	0.8989	0.5305	0.6515	0.0986	0.7561
C13	0.1636	0.2146	0.2416	0.8700	0.6264	0.1894	0.2687	0.9941
C14	0.3758	0.1221	0.4765	0.7076	0.4826	0.3485	0.1020	0.7229
C15	0.0229	0.2773	0.0537	0.5343	0.6220	0.2500	1.0000	0.4751
C16	0.0000	0.0343	0.0000	0.9458	0.0000	0.0833	0.0680	0.7483
C17	0.1260	0.1861	0.1695	0.8014	0.7658	0.0000	0.0000	0.5766
C18	0.4230	0.2923	0.6275	0.9025	0.6264	0.0303	0.1054	0.6956
C19	0.4959	0.1656	0.7542	0.9422	0.7603	0.2197	0.0442	0.6351
C20	0.3781	0.1815	0.3733	0.4657	0.7789	0.3561	0.2483	0.9005
C21	0.3277	0.3128	0.3574	0.5921	0.5131	0.1742	0.2177	0.8654
C22	0.3190	0.1702	0.3305	0.5379	0.5447	0.3939	0.0680	0.6820
C23	0.1549	0.4249	0.2240	0.9061	0.6721	0.0379	0.3537	0.7639
C24	0.5775	0.1238	0.7232	0.6931	0.5806	0.1591	0.1054	0.6800
C25	0.3240	0.1125	0.5763	1.0000	0.7146	0.3030	0.2279	0.9083

Grey Relational Coefficient Matrix given in Table 14 is obtained by taking Grey relation coefficient $\xi = 0.5$ and using Equation (3.10). It also shows the relational rank.

Table 14: Grey Relation Coefficients Matrix

	R1	R2	R3	R4	R5	R6	R7	R8	GRA	Rank
C1	0.522	0.687	0.698	1.000	0.423	0.333	0.827	0.463	0.619	10
C2	0.370	0.727	0.549	0.988	0.486	0.771	0.675	0.370	0.617	12
C3	0.974	1.000	0.983	0.416	0.580	0.449	0.775	0.416	0.699	3
C4	0.644	0.707	0.568	0.413	0.538	0.333	0.892	0.463	0.570	22
C5	0.589	0.835	0.560	0.493	0.434	1.000	1.000	0.493	0.676	4
C6	0.590	0.769	0.482	0.381	0.440	0.413	0.925	0.484	0.561	23
C7	0.870	0.370	0.982	0.844	0.423	0.608	0.801	0.424	0.665	6
C8	0.504	0.727	0.555	0.654	0.488	0.405	0.739	0.404	0.559	24
C9	0.462	0.561	0.370	0.403	0.477	0.434	0.448	1.000	0.519	25
C10	0.441	0.740	0.523	0.718	0.601	0.447	0.840	0.463	0.597	18
C11	0.936	0.923	0.921	0.527	0.499	0.442	0.730	0.403	0.672	5
C12	0.811	0.920	0.747	0.395	0.525	0.474	0.856	0.437	0.646	8
C13	0.782	0.732	0.708	0.403	0.484	0.756	0.686	0.371	0.615	13
C14	0.610	0.828	0.552	0.453	0.549	0.628	0.852	0.448	0.615	14
C15	0.962	0.679	0.916	0.524	0.486	0.701	0.370	0.553	0.649	7
C16	1.000	0.945	1.000	0.383	1.000	0.876	0.896	0.440	0.817	1
C17	0.823	0.759	0.776	0.423	0.434	1.000	1.000	0.505	0.715	2
C18	0.581	0.668	0.483	0.394	0.484	0.951	0.848	0.458	0.608	15
C19	0.542	0.780	0.438	0.384	0.436	0.728	0.930	0.480	0.590	19
C20	0.608	0.764	0.611	0.558	0.430	0.622	0.703	0.395	0.586	20
C21	0.642	0.652	0.622	0.498	0.534	0.771	0.730	0.404	0.607	16
C22	0.648	0.775	0.640	0.522	0.519	0.598	0.896	0.463	0.633	9
C23	0.791	0.580	0.724	0.393	0.466	0.939	0.624	0.435	0.619	11
C24	0.504	0.826	0.448	0.459	0.503	0.787	0.848	0.463	0.605	17
C25	0.644	0.839	0.505	0.370	0.451	0.660	0.720	0.393	0.573	21

Finally, we show the Grey relational grades for the three financial indicators; capital adequacy ratio, operating efficiency and profitability in Table 15, using Equation (3.12). In order to make a better evaluation intended for the financial performances of the non-life insurance companies, we compute the overall grey relational grade shown in Table 15. It was noticed that Priority

Insurance Company Limited (C16) is the non-life insurance company that has the best financial performance in Ghana within the period under study. Provident Insurance Company Limited (C17) takes the second position, followed by Best Assurance Company Limited (C3) and Enterprise Insurance Company Limited (C5) fourth. However, Heritage Insurance Company Limited scored relatively high in terms of profitability (3rd). This is an indication that even though, operationally very poor, other investment activities are helping to cushion and sustain it. Table 17 shows the results of grey relational grade for each of the three performance ratios used.

Table 15: Results of the Grey Relational Grades

Code	Name	CAR	Rank	Oper. Effic.	Rank	Profitability	Rank
C1	Activa International Insurance Company	0.73	6	0.38	25	0.64	14
C2	Allianz Insurance Comp. Ltd	0.66	9	0.63	8	0.52	24
C3	Best Assurance Company Limited	0.84	1	0.51	18	0.60	16
C4	Donewell Insurance Company Limited	0.58	20	0.44	23	0.68	7
C5	Entreprise Insurance	0.62	14	0.72	4	0.75	2
C6	Equity Assurance Co.	0.55	22	0.43	24	0.70	5
C7	Ghana Union Assurance Company	0.77	5	0.52	17	0.61	15
C8	Glico General Insurance Limited	0.61	16	0.45	22	0.57	17
C9	Heritage Insurance Company Limited	0.45	25	0.46	21	0.72	3
C10	Hollard Insurance Company Limited	0.60	17	0.52	16	0.65	11
C11	Imperial General Insurance Co.	0.83	3	0.47	20	0.57	19
C12	Milennium Insurance Company Limited	0.72	7	0.50	19	0.65	13
C13	NSIA Ghana Insurance Company Limited	0.66	10	0.62	9	0.53	23
C14	Phoenix Insurance Company Limited	0.61	15	0.58	11	0.65	12
C15	Prime Insurance Company Limited	0.77	4	0.59	10	0.46	25
C16	Priority Insurance Company Limited	0.83	2	0.94	1	0.67	8
C17	Provident Insurance Company Limited	0.69	8	0.72	3	0.75	1
C18	Quality Insurance Company Limited	0.53	24	0.72	2	0.65	10
C19	RegencyNEM Insurance Ghana	0.53	23	0.58	12	0.70	4
C20	Saham insurance Company Limited	0.63	12	0.53	15	0.55	21
C21	SIC Insurance Co.	0.60	18	0.65	6	0.57	18
C22	Star Assurance Co.	0.65	11	0.56	13	0.68	6
C23	Unique Insurance Company Limited	0.62	13	0.70	5	0.53	22
C24	Vanguard Assurance C	0.56	21	0.65	7	0.66	9
C25	Wapic Insurance (Gh), Limited	0.59	19	0.56	14	0.57	20

Based on the findings in Table 15, on average the most important financial indicator impacting on financial performances of the non-life insurance companies in Ghana is the capital adequacy indicator (65.0%). In addition, profitability and operating efficiency follows with 62.5% and 57.6% respectively. The Table shows that Best Assurance Company Limited (C3) is the most efficient (84.3%) in terms of capital adequacy. It is the company with the ability to withstand shocks or loss with Heritage Insurance Company Limited (C9) being least capable to withstand losses (44.9%). Priority Insurance Company Limited (C16) holds the second position with adequacy (83.2%).

When ordered operating efficiency indicators is examined, it is found that Priority Insurance Company Limited (C16) was the best in terms of underwriting and operational efficiency (93.8%). The second position is taken by Quality Insurance Company Limited (C18), and third is Provident Insurance Company Limited (C17). The worst in terms of underwriting and operational efficiency is Activa Insurance Company Limited (C1) with 37.8%.

According to the profitability indicators, Provident Insurance Company Limited (C17) is more efficient (75.2%). The second position is taken by Enterprise Insurance Company Limited (C5) with 74.6%. This is interpreted as good despite its poor capital adequacy position. The worst is Prime Insurance Company Limited (C15) (46.1%).

Assessment of financial performances of insurance companies is quite important for regulators and industry players. Assessment of insurance companies is rarely seen within the Ghanaian insurance industry and hence this study is the first attempt to measure the financial performance regarding the non-life insurance companies in Ghana using financial ratios. The findings of the study showed that the most important financial indicator is the capital adequacy indicators, followed by profitability and then operating efficiency. The results also indicated Priority Insurance Company Limited has the most successful performance among the 25 non-life insurance companies in Ghana.

The performance of Priority Insurance Company Limited is driven by very high underwriting and efficiency indicators as well as on capital adequacy with relatively low profitability. In the light of these findings, it can be suggested that Priority Insurance Company Limited is the best managed financially.

Provident Insurance Company Limited is the second best also driven by high profitability and operating efficiency, second and third, respectively, in the ordering table. This company is 8th on CAR ordering. This means that this company is high net insurance risk. Heritage Insurance Company Limited exhibited worse performance overall. Heritage Insurance Company Limited was also ranked last in terms of capital adequacy requirement and 21st in terms of operating efficiency. This means that the company was not managed well during the period under review.

Developinga robust, risk-based and semi-parametric pricing model for premium determination

The previous section quantitatively analyzed the financial performance of the insurance companies. This is important because it gives a fair view of the non-life insurance industry operational performances and lays the foundation for this section. Using accident and claims data sets as described in Chapter Three, this section presents the analysis and results of the modeling framework in line with the objectives of the study. The summary of regional distribution of the accident datasets is presented in the Tables in APPENDIX A-1, APPENDIX A-2, APPENDIX A-3 and APPENDIX A-4.

Summary of Accident Data

In APPENDIX A, Table A1 summarizes the road traffic crashes by region from 1991 to 2015. Using the most recent three-year datasets (2013-2015) consistent with accident rate computation, it was established that about 15.5% of crashes occurred in Ashanti region, 5.9% crashes occurred in Brong Ahafo, 8.9% occurred in Central region, 10.5% occurred in Eastern, 42.4% in Greater Accra region, 2.9% Northern, 1.4% in Upper East, 1.5% in Upper West, 4.9% in Volta region and 6.1% occurred in Western region. Majority of the accident crashes occurred in Greater Accra followed by Ashanti region. This is not surprising because the vehicle density in Greater Accra and Ashanti regions are higher than the rest of the regions (see APPENDIX A-1 for details).

The Table in APPENDIX A-2 presents the annual record of the number of deaths as a result of accidents in the various regions. If we focus on data from 2013-2015, out of a total of 4676 deaths recorded, about 19.8% of deaths were recorded in Ashanti region, 9.8% in the Brong Ahafo region, 9.3% in Central, about 11.1% in Eastern, 24.6% in Greater Accra region, 4.8% in the Northern region, 3.3% in Upper East, 3.0% in Upper West, 6.6% in Volta region and 7.8% in the Western region. The highest number of deaths were recorded in the Greater Accra region followed by Ashanti, Eastern, Brong Ahafo, Central etc. This statistic correlates with what was observed in APPENDIX A-1.

Data on actual annual traffic fatalities was also obtained, which isthe number of traffic crashes that resulted in death. The summary is provided in APPENDIX A-3. The annual distribution of casualties; crashes that involve injuries is also summarized in APPENDIX A-4.

The regional rank order when considering fatal crashes during the three, year period 2013-2015 remains nearly the same (see Appendix A-2). Greater-Accra region remains the highest in terms of fatal crashes with 24.6% followed by Ashanti region (19.8%), Eastern (11.1%), Brong-Ahafo (9.8%) and Central (9.3%). Expectedly, in 2015, the highest number of fatal crashes occurred in Greater Accra (439; 27.6%), followed by Ashanti Region (310; 19.5%), Eastern (181; 11.4%) and then Central region (149; 9.4%). Together, these four regions accounted for slightly over two-thirds (67.9%) of all the fatal crashes in Ghana.

From year 2014 to 2015, five (5) regions recorded increases in fatal crashes. Greater-Accra region recorded the highest percentage increase of 14.6% followed by Ashanti region (13.6%), Western (13.1%), Upper West (8.1%) and then Eastern region (7.1%). Reductions in fatal crashes were however recorded in the Upper East region (-33.8%), Northern region (-13.7%), Volta region (-13.0%), and Brong Ahafo region (-8.4%). Central region recorded no change in fatal crashes from 2014 to 2015.

In terms of the distribution of road traffic fatalities among the regions, the pattern virtually follows that of fatal crashes. During the year 2015, the highest regional number of road traffic fatalities of 458 deaths, representing 25.4%, were recorded in Greater Accra region followed by Ashanti region (352 deaths; 19.5%), Eastern region (197 deaths; 10.9%), Central region (185 deaths;10.3%) and Brong Ahafo region (170 deaths;9.4%). These five regions alone contributed slightly over 75% of all the road traffic fatalities in Ghana. It should be noted that Greater Accra region has been the worst region, in terms of regional fatalities in Ghana.

Compared with the 2014 records, in year 2015, fatality increases were recorded in three (3) regions, namely Greater Accra Region (9.3%), Western region (19.7%) and marginally in Ashanti region (0.3%). Figure 10 gives a pictorial view of road crash and fatalities trend from 1991 through 2015.

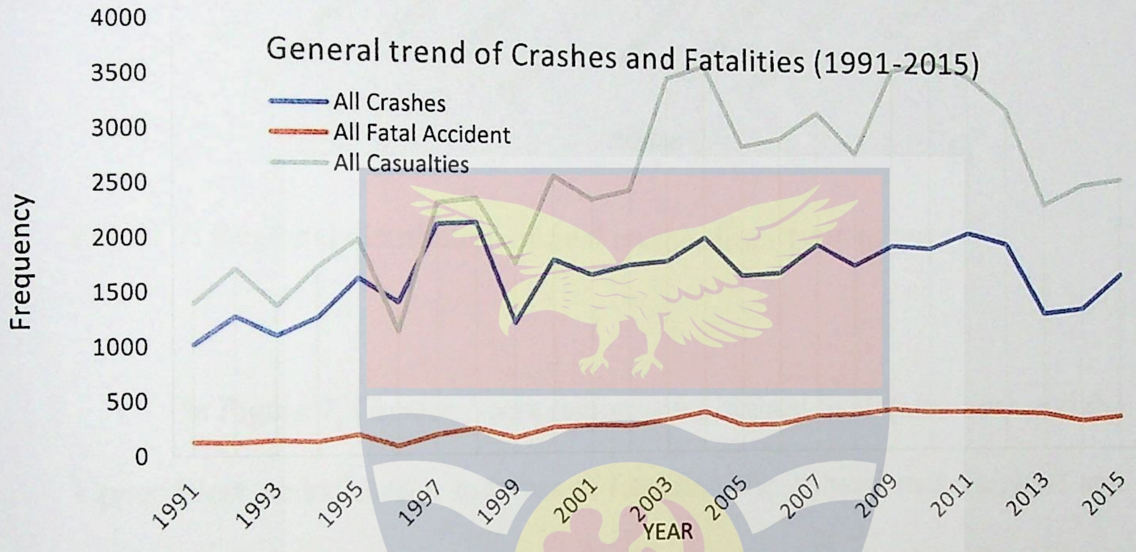


Figure 61: General trend of crashes and fatalities from 1991 to 2015.

Figure 6 shows the plot of crashes, fatalities and casualties from 1991 to 2015. Generally, the graph shows an upward trend during the year under consideration. This means that accident have increased from 1991 to 20015. This may have corresponding impact in the claim's records for the insurance companies.

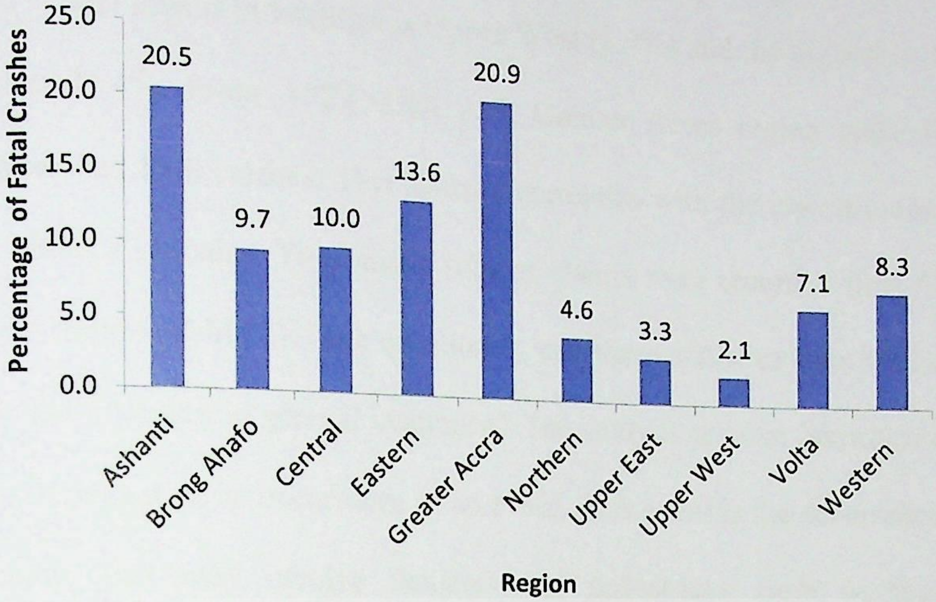


Figure 7: Regional distribution of fatal crashes from 1991 to 2015.

In Figure 7, Greater Accra region ranks highest in all fatal crash and the Upper West ranks lowest. In terms of fatalities, the Ashanti ranks highest and the Upper West region ranks lowest whilst in terms of casualties the Greater Accra ranks highest. The trend is an indication that accident risk is high when there is high traffic intensity.

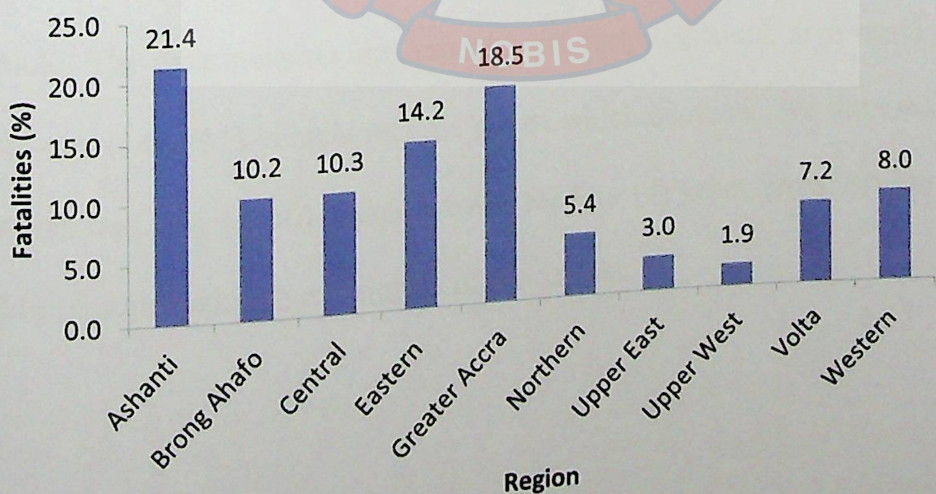


Figure 8: Regional distribution of fatalities from 1991 to 2015.

The lowest in fatalities is Upper West (1.9%) and the highest is Ashant region 21.4%. From APPENDIX A-5, Greater Accra region ranks first in crashes and fatal crashes. This statistic correlates with the claims recorded in the claim's statistics. Thus about 70% of claims were recorded from Greater Accra region (Table 8). The question is, can these statistics form the basis of risk classification in general insurance? The study notes that insurance claims are contingent on an occurrence of an event. This event is the occurrence of an accident that may involve vehicles and pedestrians from an insurance perspective. In view of the statistics, the study wishes to test the effectiveness of a new classification model based on location in which the policy usually operates. We argue that the risk of claim is not only influenced by policy or vehicles characteristics but also by the environment in which the policy operates. The study thus examines the distribution of risk of accident across all the ten regions and used that as a basis of model refinement.

Summary of Insurance Data Used

The insurance data consists of policy and claim information for each vehicle. The data contains one hundred and forty thousand, nine hundred and sixty-one (140,961) vehicle records out of which contains five thousand, four hundred and fifty (5450) claims records for four (4) years, from 2013 to 2016. Table 16 summarizes the variables of the dataset.

Table 16: Insurance Policy Variables

Policy characteristics	Vehicle characteristics	Claims history
Use type 1. Commercial 2. Private Usage category 1. Taxis, 2. Ambulance, 3. Tanker, 4. General cartage, 5. Maxi-bus, 6. Mini-bus, 7. Private individual, 8. Corporate individual, 9. Motor, 10. Own goods carrying, 11. Hiring, 12. Special types	Vehicle make 1. Opel 2. Nissan 3. Mitsubishi 4. Kia 5. Tata 6. Toyota 7. BMW 8. Hyundai 9. Daewoo 10. Honda 11. Audi 12. Peugeot 13. Ford 14. Daf 15. Mercedes 16. Mazda 17. Make.Other	No. of claims
Policy coverage 1. Comprehensive 2. Third Party 3. Third Party Fire and Theft	Vehicle age Region	Claim amount

Table 16 presents the insurance policy variables considered as historical risk in this study. These are the available variables associated with the data collected. The variables include policy characteristics, vehicle characteristics, claims history, policy coverage, claim amount where claim has occurred, vehicle age, region and claim amount.

Distribution of Claims and Premiums by Region

Table 17 below gives the regional distribution of claims for 2013 to 2016.

Table 17: Regional Distribution of Claims (2013-2016)

Region of incidence	No. of Claim (N)	No. of Vehs. Insured (R)	N/R	Claim Proportion (N/T)	Claim Size (Ghc)	Premium Income
Ashanti	512	21143	0.0242	0.0796	3,453,901	547,326
Brong Ahafo	55	1962	0.0280	0.0059	254,983	79,426
Central	60	1731	0.0347	0.0074	322,927	75,762
Eastern	154	3921	0.0393	0.0298	1,293,609	214,743
Greater Accra	4407	106232	0.0415	0.8532	36,997,062	6,337,100
Northern	27	759	0.0356	0.0029	127,307	37,581
Upper East	24	696	0.0345	0.0027	116,862	36,828
Upper West	4	416	0.0096	0.0009	37,883	7,558
Western	149	3148	0.0473	0.0104	451,503	201,352
Volta	58	953	0.0609	0.0071	308,335	92,581
Total (T)	5450	140961	0.03866	1.0000	43,364,372	7,630,261

Table 17, records 140,961 insured customers of which about 5450 (3.9%) filed for claims. About 85% of the claims in the study period occurred in Greater Accra region and the lowest claims frequency occurred in Upper West region. The total claim size recorded for the study period was Ghc43,364,372 while the premium income accrued from claimants is Ghc7,630,261 representing about 17.6% of aggregate claims. This means that less than 20% of claims was accounted for by the premium income for the period. In terms of regional distribution of total claims, Greater Accra region recorded the highest of the aggregate claims (85%) and contributed about 83% of the premium income. Upper West recorded the lowest (0.09%) proportion of claims. These figures are not surprising since nearly 75.4% of the total number of vehicles insured are in the Greater Accra region, and less than 0.3% of same are in the Upper West region. In terms of third party and comprehensive policy types, the data showed that a claim of approximately GHc 5,284,407 was recorded for Third Party representing 12.06% of total claims and the remaining majority for Comprehensive. No claims were recorded for Third Party Fire and Theft.

In line with the study objectives, it is required that the predictor function considered in Equations (3.12) and (3.19) is estimated. The study first considers the distribution of accident risk across the ten regions of Ghana. To the best of our knowledge no such information has been considered in pricing insurance policies.

The study reckons that several factors influence insurance outcomes. Some of the factors include regulation and legislative changes, claim trends, vehicle density, interest rate and investment. Legislative and regulation, inflation, interest rates and investment could be regarded as fixed. However, claim trends, vehicle congestion and for that matter accidents risk in different geographical zones could vary from one region to the other as well as from one period to the other. More so, factors such as roadway design, roadway maintenance have been shown to contribute significantly to road accidents which vary from one region to another. For these reasons, these phenomena are characterized and incorporated into the pricing framework. The study used Markov theory to categorize the claims data set based on accident risk derived for each region. By noting that the number of times accident occurs represent a sequence of random variables, the probability of transitions between states or movement of vehicles between and within regions could be expressed as a transition probability matrix with finite state space $\{i = 1,...,10\}$. The initial state distribution was obtained using (2013-2015) distribution of accident claims, with initial probabilities given in Table 18.

Table 18: Accident Distribution (2013-2015)

Region	All Crashes (%)
Greater Accra	42.4
Ashanti	15.5
Eastern	10.5
Central	8.9
Western	6.1
Brong Ahafo	5.9
Volta	4.9
Northern	2.9
Upper West	1.5
Upper East	1.4

Given that occurrence of an accident at any given time is stochastic and heteroskedastic between regions, the study employes Bayes theorem and conditional probabilities to compute the accident risk between and within states to generate the transition probabilities for the Markov chain. Using the crash data, risk of accidents within each state were computed to generate the transition probability matrix. Table 18 displays accident risk distributions for the ten states, which is the maximum likelihood estimate of risk in the various regions. The Bayes formula has two interpretations: given the probability of an event A, the first interpretation is the long run proportion of times that the event A occurs upon repeated sampling.

The second interpretation is the subjective belief in how likely it is that the event A will occur. If A and B are two events, and $P(B) > 0$, then the conditional probability of A given B is $P(A \mid B) = P(AB) / P(B)$, where AB denotes the event that both A and B occurs which in the context of the study denotes the probability of accident occurrence between states or within two states. The frequency interpretation of $P(A \mid B)$ is the long run proportion of times of accident occurrence when we restrict our attention to movement between states. The subjective probability interpretation is that $P(A \mid B)$ represents the updated belief of how likely it is that accident occurs if it is known that the vehicle is moving from state B to A. $P(A \mid B)P(B) = P(A \cap B)$. Table 15 represents the values of the crash matrix using the initial states probabilities.

Table 19: Accident Matrix for the Ten Geographical Zones

	GR	AS	ER	CR	WR	BA	VR	NR	UW	UE	Total
GR	0.4240	0.0657	0.0445	0.0377	0.0259	0.0250	0.0208	0.0123	0.0064	0.0059	0.6682
AS	0.0657	0.1550	0.0163	0.0138	0.0095	0.0091	0.0076	0.0045	0.0023	0.0022	0.2860
ER	0.0445	0.0163	0.1050	0.0093	0.0064	0.0062	0.0051	0.0030	0.0016	0.0015	0.1990
CR	0.0377	0.0138	0.0093	0.0890	0.0054	0.0053	0.0044	0.0026	0.0013	0.0012	0.1701
WR	0.0259	0.0095	0.0064	0.0054	0.0610	0.0036	0.0030	0.0018	0.0009	0.0009	0.1183
BA	0.0250	0.0091	0.0062	0.0053	0.0036	0.0590	0.0029	0.0017	0.0009	0.0008	0.1145
VR	0.0208	0.0076	0.0051	0.0044	0.0030	0.0029	0.0490	0.0014	0.0007	0.0007	0.0956
NR	0.0123	0.0045	0.0030	0.0026	0.0018	0.0017	0.0014	0.0290	0.0004	0.0004	0.0572
UW	0.0064	0.0023	0.0016	0.0013	0.0009	0.0009	0.0007	0.0004	0.0150	0.0002	0.0298
UE	0.0059	0.0022	0.0015	0.0012	0.0009	0.0008	0.0007	0.0004	0.0002	0.0140	0.0278

All things being equall, the risk of accident within Greater Accra region is 0.4240 and the risk of accident if a vehicale is moving from Greater Accra region to Ashanti region is 0.0657. From Table 19, we derive the transition probability shown in Table 20.

The transition probability matrix is aperiodic and irreducible and therefore the limiting distribution exist. Thus, from the Table 20 it is possible to events between states or a vehicle can freely move from one state to another without any hindrance (irreducible) and observing the event within each state is also possible (aperiodic). On the basis of this it possible to find the limiting and stationary distribution of the transistin matrix. This means that there exist, $\lim_{n\to\infty} M^n = \gamma_j$. The summary of long run distribution ($\lim_{n\to\infty} M^n = \gamma_j$) is shown in Table 21.

Table 20: A 10-Dimensional Discrete Time Markov Chain Defined by the Ten Geographical Regions

	GR	AS	BA	CR	ER	NR	UE	UW	VR	WR
GR	0.5566	0.1119	0.0532	0.0585	0.0822	0.0240	0.0170	0.0106	0.0389	0.0472
AS	0.1131	0.5559	0.0531	0.0584	0.0820	0.0240	0.0169	0.0106	0.0389	0.0471
BA	0.1068	0.1056	0.5251	0.0552	0.0775	0.0227	0.0160	0.0100	0.0367	0.0445
CR	0.1074	0.1061	0.0504	0.5277	0.0779	0.0228	0.0161	0.0100	0.0369	0.0448
ER	0.1098	0.1085	0.0516	0.0567	0.5398	0.0233	0.0165	0.0102	0.0377	0.0458
NR	0.1040	0.1027	0.0488	0.0537	0.0754	0.5110	0.0156	0.0097	0.0357	0.0433
UE	0.1033	0.1021	0.0485	0.0533	0.0749	0.0219	0.5077	0.0096	0.0355	0.0431
UW	0.1027	0.1015	0.0482	0.0530	0.0745	0.0218	0.0154	0.5048	0.0353	0.0428
VR	0.1054	0.1042	0.0495	0.0544	0.0765	0.0224	0.0158	0.0098	0.5181	0.0439
WR	0.1062	0.1050	0.0499	0.0548	0.0771	0.0159	0.0099	0.0365	0.3651	0.5221

Table 21: Long run distribution of accident risk.

GR	AS	BA	CR	ER	NR	UE	UW	VR	WR
0.196	0.1943	0.0978	0.1069	0.1469	0.0453	0.0323	0.0202	0.072	0.0873

Table 21 shows the long run distribution of accident risks across the ten regions of Ghana. Thus, in the long run accidents risk within Greater Accra region, for example is 19.64% and 19.43% inAshanti region. Based on the outcome in Table 21, the risk for the ten regions are classified into three risk zones based on risk similarities described in Table 21 and consistent with Occam's razor.

Table 22: Classification of Risk Zones

Range	Classification	Region
0.0 – 0.05	Low risk	NR, UW, UE
0.06 – 0.15	Medium risk	WR, VR, CR, ER, BA
0.15 – 0.25	Considerable risk	GA, AS

As shown in Table 22 the result of the risk classification is considerable risk (Greater Accra and Ashanti regions), medium risk (Brong Ahafo, Central region, Eastern region, Western region and Volta regions), and low risk zone (Northern, Upper West and Upper East regions).

Two indicator variables q_1 and q_2, were adopted to integrate the location risk, thus

$$\eta(z) = \vartheta_1 l_1 + \vartheta_2 l_2 + \sum_{i=1}^{p} \kappa_i R_i \quad (4.13)$$

where l_1 and l_2, are the levels of risk into a single model, and R_i are the historical risk. The levels of the indicator variables are in Table 19.

Table 23: Location Risk Variable Classification

q_1	q_2	Variable Description
0	0	if the observed claim is in State 3
1	0	if the observed claim is in State 2
0	1	if the observed claim is in State 1

This means State 3 is used as the reference. This implies that when policyholder is in low-risk zone (state 3), the expected premium is related by the function

$$\eta(z) = \sum_{i=1}^{p} \kappa_i R_i \quad (4.14)$$

For any other states, the risk premium for a policyholder depends on the dynamics of the risk associated within the environment in which the policy usually operates. It is therefore required that the relevant factors that influence premium function and the magnitude of their effects on the premium function is determined.

Estimation of the Premium Function

The study estimates the premium function using the method developed in Chapter Three. Consistent with statistical model framework, 70% of the data that was randomly selected was used in building the model, while 30% is used for out-of-sample validation of the model. The first choice in building a modelwith gradient boosting algorithm involves selection of the appropriate loss function which we specify as Tweedie. The use of Tweedie loss function requires specification of the index parameter (ξ), the shrinkage, the optimal number of trees and the interaction depth. The optimal index parameter was obtained using profile likelihood estimation procedure. Figure 9 shows that, the optimal ξ obtained was 1.61 at 5% level of significance.

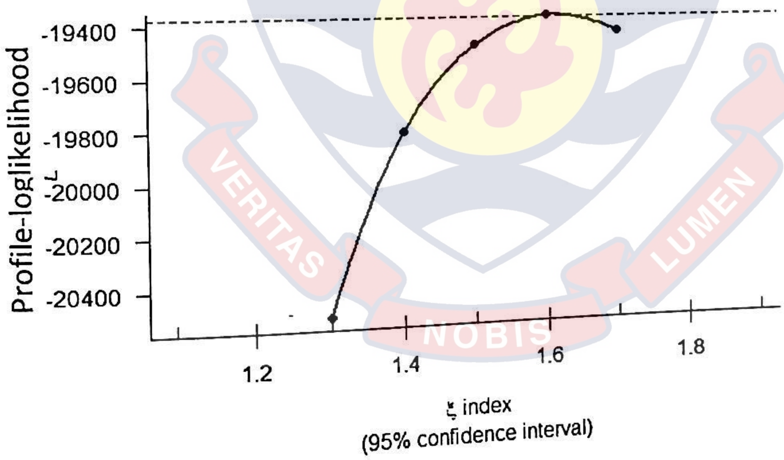


Figure 9: Estimating index parameter (ξ)

Secondly, it is necessary to select the shrinkage or the learning rate ζ. The learning rate (ζ) was set at 0.05 as suggested by Friedman (2001) to give optimal results. The next tuning parameter for the model is the size of trees or the number of boosting iterations M. This was specified via 5-fold cross validation. To illustrate the selection procedure, we first grew many trees with $M = 5000$. The result is displayed in Figure 10.

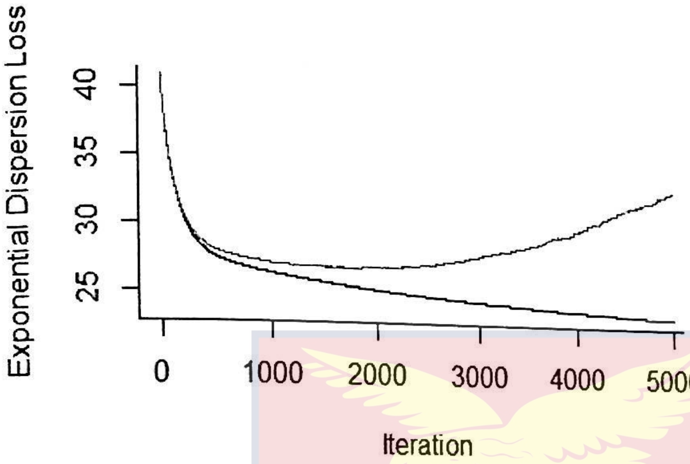


Figure 10: A Plot of CV error showing the optimal iteration number.

Figure 10 plots the error rate obtained at each level of iteration. The study used a 5-fold cross validation. This means that the data was randomly divided into five (5) samples not necessarily equal size. Each of the five (5) samples is fitted separately to the model, out of which the optimal number of trees is obtained. Figure 10 shows two lines one black and the other green. The green line above black line plots the error rate at various levels of iterations. As the tree grows from point zero (0) onwards the error rate reduces suggesting an improvement of error reduction rate. As the model moves beyond 2000 iterations, the curve turn upwards suggesting diminishing returns in model accuracy. The point where the error rate reaches minimum and begins to rise is the optimal tree number which from the Figure 14 is about $M = 1788$.

One other important parameter consideration is the shrinkage or the learning rate that ranges between 0 and 1. It must be noted that according to the literature the shrinkage parameter slows down the learning rate. It controls the rate at which boosting learns. It was observed from the analysis that the number of trees is inversely related to the shrinkage parameter. As the shrinkage parameter gets smaller the number of trees increase at the expense of computational time. Thus, the parameter was tested at various levels; 0.1, 0.01, and 0.05. From the analysis the shrinkage parameter at 0.05 ($\varsigma = 0.05$) was most approapriate.

Thirdly we evaluated the optimal number of splits for each tree or size of interaction effects. A model with a one-way interaction effects is simply an additive model, without any interaction effect. The study set out to investigate interaction effects specification that gives the best results, given $\varsigma = 0.05$ and $M = 1788$. We evaluated 20, 10, 5,4,3 and 2-way interaction effects using the training data set. The result showed that the ten (10) way interaction effects $(L = 10)$ gives the best results. The model framework makes it possible for higher order interaction effects to be investigated. Given $M = 1788$, $\varsigma = 0.05$, and $L = 10$ the predictor function in Equation (4.13) is estimated using the function described in Equation (3.11).

APPENDIX B-1 to APPENDIX B-6 illustrate the procedures involved in getting the outcome in Table 24 which presents the relevant predictors arranged in order of magnitude.

Table 24: Summary of Model Variables and their Relative Importance

Variable No.	Variable Name	Relative influence
1	Vehicle Age	29.4643
2	Comprehensive	23.8642
3	Usage.Own.Goods.Carrying	7.8849
4	Truck	3.8131
5	Car	3.3395
6	UsageMaxi.Buses	2.9536
7	Private_cars(individual)	2.4073
8	Bus	2.3541
9	Usage.Tankers	2.2603
10	Toyota	1.9629
11	Make.Other	1.4143
12	Ford	1.3968
13	Daf	1.3809
14	Usage.Ambulance/Hearse	1.2331
15	Mercedes	1.1269
16	Motor	1.0991
17	HGV	1.0517
18	UsagePrivate_cars(corporate)	1.0124
19	Tata	0.9235
20	Mitsubishi	0.8395
21	Type.Private	0.7919
22	Type. Commercial	0.7497
23	Pickup	0.7357
24	State2	0.6034
25	Make.Nissan	0.5927
26	Third.Party	0.5436
27	Usage.Taxis	0.5348
28	Kia	0.5124
29	General cartage	0.5040
30	UsageMini.Buses	0.4187
31	LGV	0.3326
32	Make. Opel	0.3177
33	Usage.Motor.Cycle (Corporate)	0.2933
34	Hyundai	0.2671
35	Audi	0.2576
36	SUV	0.2068
37	Usage. Hired	0.1173
38	Honda	0.1102
39	Van	0.1033

Table 24: Cont'd.

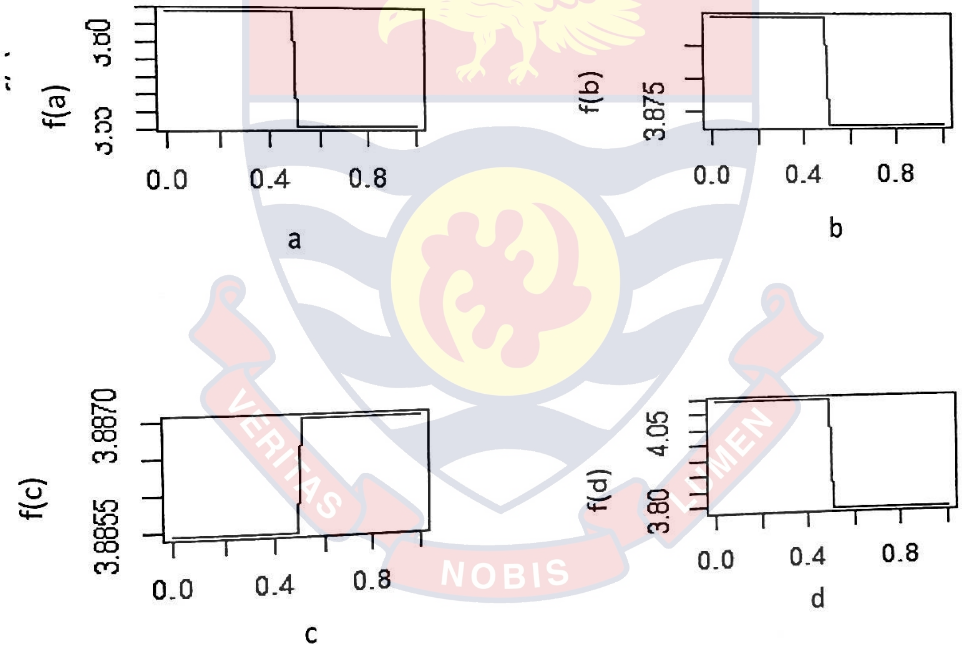
Variable No.	Variable Name	Relative influence
40	State 1	0.0767
41	Usage.Special Types (Road)	0.0649
42	Daewoo	0.0604
43	Usage.Motor. Cycle (Individual)	0.0114
44	Make.Motor	0.0074
45	Mazda	0.0017
46	Third.Party(ft)	0.0000
47	Stype	0.0000
48	Bmw	0.0000
49	Peugeot	0.0000

Thus, Table 24 presents the summary of the relative importance of model parameters. This is an attempt to assess how important each variable is to the model. In this procedure, in regression trees we calculate the total amount of reduction in the residual sum of squares (RSS) attributable to splits caused by the predictor, averaged over the number of trees. In classification trees, we do the same using average reduction in the Gini index. The variables that are relevant to the model have been arranged in order of magnitude. Out of the 49 variables tested out-of-sample, 4 of variables were found not relevant to the model (variable importance recorded zero as in Table 24). The location effect as specified in the model was significant. Thus, while State 1 contributes about 0.08%, State 2 contributes about 0.6% (Table 24). This is significant considering the nature of business of the non-life insurance business and the severity of claims when it occurs. This also means that latent claims that would have occured as a result of risk associated with risk specific geographical location has been accounted for by the model summarized in Table 24. The results in Table 24 also shows that the most principal factors influencing the model is vehicle age (29.46%). The rest include comprehensive policyholders (23.86%), own goods carrying (7.88%), etc.

Partial Dependence Plot

Partial dependence plots offer additional insights into the way the variables influence the dependent variable. Recall that one of the criticisms of the previous study utilizing procedures such as generalized linear models is that the fuctional form or relationship of claims with some other variables in the predictor space may not be necessarily monotonic and linear. The partial dependence plot therefore seeks to display or validate among others this assertion or otherwise. In this plot the vertical scale is designated as the log odds and the hash marks at the base of each plot show the decile of the distribution of the corresponding variable. The partial dependence accounts for the average joint effect of the other predictors in the model. The variable "vehicle age" have a roughly monotonically decreasing partial dependence (Figure 11). The nature of dependence of vehicle age and price of the vehicle is natural. This is because newer and more expensive cars would cost more to repair in the event of accident (collision). The nature of dependence of vehicle age is fairly linear over the vast majority of the data (see Figures 11 and 12). As expected, there is no partial dependence of third-party fire and theft on the premium model since thereis a horizontal relationship.

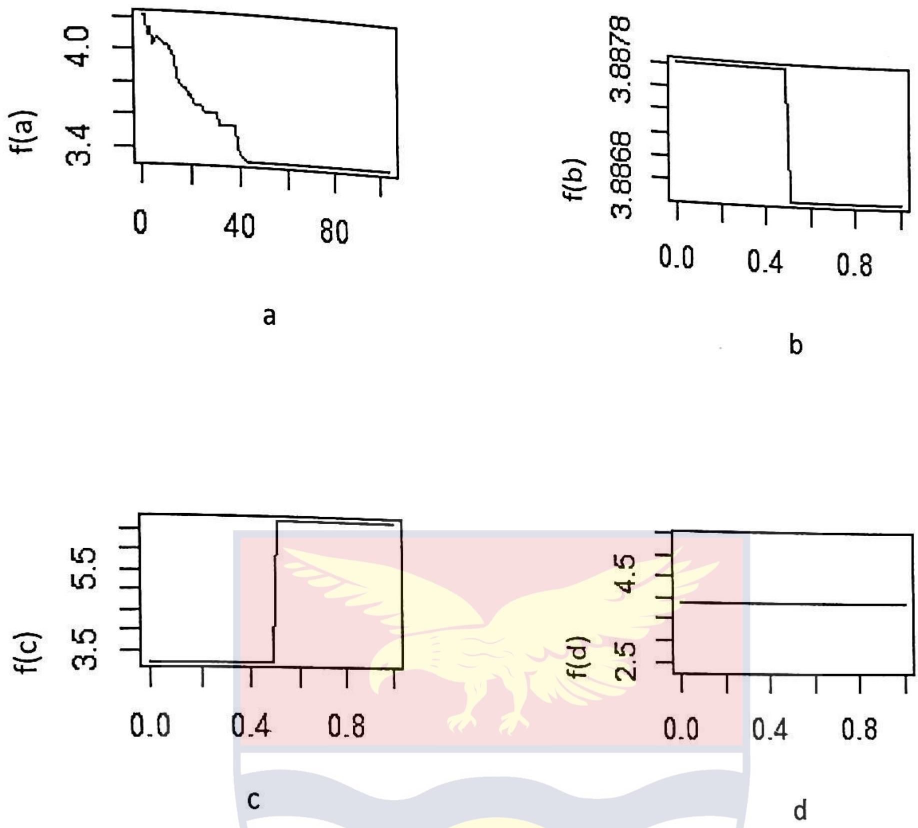
The age of the vehicle is widely recognized as an important predictor of claims (Brockman & Wright, 1992), since it is believed to be negatively associated with annual mileage. It is not a widespread practice to use annual mileage directly as an input in the model due to the difficulty in obtaining a reliable estimate for this variable. Annual mileage was however not included in this study. Figures 11 and 12 show the partial dependence plots for the premium model. Positive relationship means that when the risk in State 1 goes up, the average premium also goes up.



Key

a) SUV b) State 2 c) State 1 d) Private type vehicles

Figure 11: Partial Dependence Plots (a)



Key

a) Vehicle age b) Third party c) Comprehensive d) Third party (fire and theft)

Figure 12: Partial Dependence Plots (b)

These plots are not necessarily smooth, since there is no smoothness constraint imposed on the fitting procedure. This is the consequence of using a tree-based model. If a smooth trend is observed, then it is the result of the estimated nature of the dependence of the predictors on the response and it is purely dictated by the data. The rest of the variables have binary classifications. For instance, for State 1, responses close to zero signify that the observation is either in State 2 or State 3 and responses close to 1 means the observed response is in State 1. The partial dependency plot suggests a logit function. We interpret the partial dependence plot for the dummy variables in terms of its direction of relationship with the premium function. For instance, the charts for sports utility vehicles (SUV), private vehicles have, State 2, etc. suggests negative partial dependence with claims. The nature of the curve for State 1, suggests a positive relationship with the claims function.

Model Evaluation

One of the objectives of the study is to demonstrate the superiority of the model summarized in Table 24 using underwriting data from a local insurance company. According to Klugman et al. (2004), model-based approach to solving real life problems should be considered in the context of the objectives of any given problem. Many problems in actuarial science involve the building of a mathematical model that can be used to forecast or predict insurance costs in the future. A model is a simplified mathematical expression or description which is constructed based on the knowledge and experience of the actuary combined with data from past. The data guide the actuary in selecting the form of the model as well as in calibrating unknown quantities usually called parameters. This means the model provides the balance between simplicity and conformity with the available data. The simplicity is measured in terms of such things as the number of unknown parameters. The conformity of data is measured in terms of the discrepancy between the data and the model. Model selection is usually based on the balance between the two criteria: fit and simplicity (Klugman et al., 2004).

In line with the objectives the study compares the model developed against other competing models such as the conventional Generalized Linear

Model (GLM), TDboost model proposed by Yang et al. (2016) and the gradient boosting approach proposed by Guelman (2012) based on a squared error loss function. Thus, using the same training and test data, the study fitted and predicted premiums using the three competing models: the TDboost model, GLM, GBM model and MMGB were compared using the same data.

To examine the performance of these competing models, after fitting each on the training data, we predict the risk premium $Z(X) = \hat{\mu}(X)$ by applying each model on the independent out-of-sample data. It would not be appropriate to measure differences between predicted premiums $Z(X)$ and real losses y by depending on mean square error loss or the mean absolute loss. This is because the losses or claims have high proportion of zeros and very much positively skewed. An alternative statistical measure was considered. The ordered Lorenz curve and the associated Gini index proposed by Frees et al., (2011) was considered. This measure captures the discrepancy between premium and loss distributions without the influence of either the zeros nor skewness. As discussed in Chapter Four, we compute the Gini index and calculate the ratio of the rate we would charge based on MMGB model and the rate we would charge based on GLM, TDboost and GBM.

Based on the discussions, the prediction from each model is successively specified as the base premium and use the predictions from the remaining models as the competing premium to compute the Gini indices. Using "minimax" strategy to select the best performing model, the study selected the model that provides the smallest of the maximal indices over the competing models.

Table 25 presents the gini scores which is used as a measure of model performance and Table 26 presents the associated standard errors of the gini coefficients.

Table 25: Matrix of Gini Indices

	MMGB(P1)	TDBOOST(P2)	GLM(P3)	GBM(P4)
MMGB	0.000	-8.013	1.380	-2.115
TDBOOST	12.027	0.000	2.950	-2.470
GLM	28.406	27.225	0.000	19.837
GBM	36.296	35.729	30.025	0.000

Table 26: Matrix of Standard Errors

	MMGB(P1)	TDBOOST(P2)	GLM(P3)	GBM(P4)
MMGB	0.000	3.062	3.555	3.107
TDBOOST	3.045	0.000	3.570	3.140
GLM	3.246	3.266	0.000	3.412
GBM	2.273	2.304	2.869	0.000

From Tables 25 and 26, we have the Gini indices and the standard errors respectively. We find that the maximal Gini index is 1.38 when using MMGB as the base premium, 12.07, when using TDboost as base premium, 28.406 is when using GLM as base premium and 36.296 when using GBM as the base. MMGB is the smallest. Therefore, MMGB has the smallest maximum Gini index of 1.380, hence it is the least vulnerable to alternative scores. Figure 14 also shows that when GLM is selected as the base premium, the area between the line of equality and the ordered Lorenz curve is large when choosing MMGB as the competing premium, indicating that the MMGB model represents the most favorable choice. See Figure 13 and 14 for a pictorial overview.

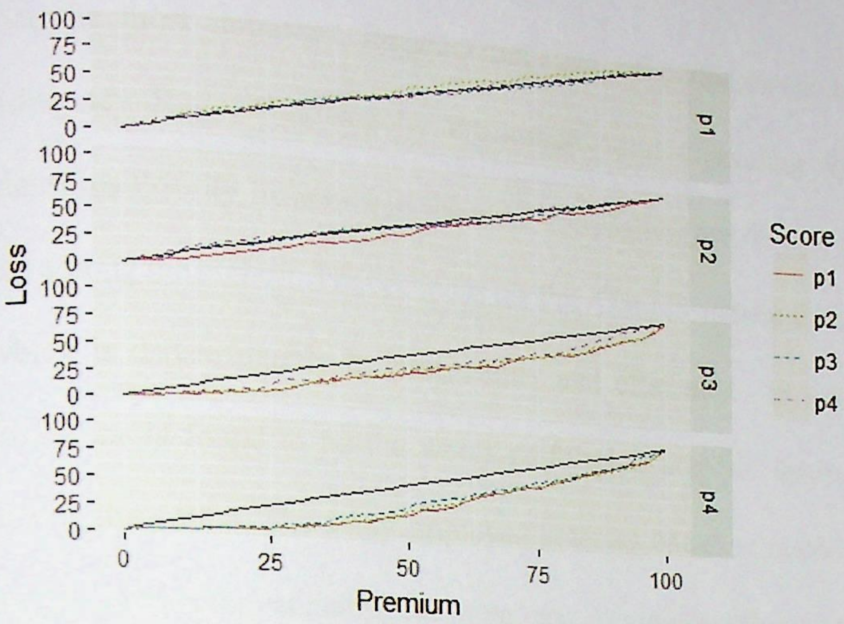


Figure 13: Plot of aggregate ordered Lorenz curve.

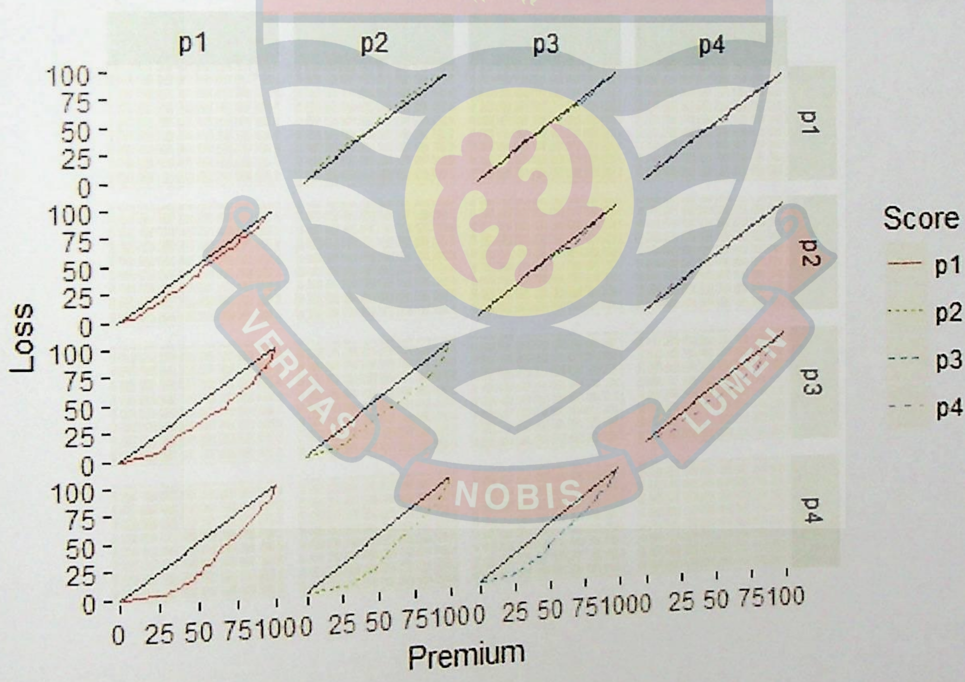


Figure 14: Plot of ordered Lorenz curve by model score.

Chapter Summary

The chapter presents the analysis and results of the research findings. Financial analysis of the insurance companies is performed. The study found that the most important financial indicator of performance is the Capital Adequacy Ratio, followed by Profitability and Operating Efficiency. It identifies Priority Insurance as the most successful even though profitability is not among the highest. The second is the Provident Insurance Company Limited which is driven mainly by profitability and efficienty. In general Heritage Insurance is found to be the worse performing non-life insurance company during the period. The study employed gradient boosting methods to estimate the parameters for estimating the premium. Among the relevant variables is the location, particularly, Greater Accra region and Ashanti regions. The study used the Gini Index to compare the MMGB model with three other models: GLM, TDBoost and GBM. The study found that the MMGB model is a preferred alternative method for building insurance cost models.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview

In this study, we have discussed the theory of gradient boosting, the Markov-modulated approach which the study developed and its application to risk premium pricing. The practical steps required to build a model using this methodology has been described. This chapter discusses the findings of the study and make appropriate conclusions based on the study outcome. It outlines some challenges of the study and outline recommendations and suggestions for further research.

Summary

Several statistical techniques have been proposed to price premiums such as modeling the frequency and severity of claims and computing the product of its expectations. Models such as Poisson, negative binomial and generalized Poison have been suggested for claim frequency data whereas Gamma, lognormal, Pareto etc. have been recommended for claim severity (Renshaw, 1994; Mihaela, 2015). There is constantly the need to improve on ways in which policy premiums are priced. In many actuarial risk models, consideration is mostly placed on internal historical data that are obtained in lieu of the data generating processes within the insurance firm or industry. Typically, in insurance pricing, analyst usually perform extrapolations based on internal historical data to project future claim cost or profitability as discussed in section one. Thus, geographical risk and for that matter external data is rarely used. In cases where external data is used, the focus has been quantifying the effects of seasonal variations inherent in insurance portfolios, given that the external data is observable. With unobservable phenomenon, researchers usually employ hidden Markov chains to make extrapolations (Guillou, et. al. 2013).

However, besides financials, legislation and regulatory mechanisms, and internal historical data and other seasonal considerations, it is important to note that additionally in Auto insurance external observable phenomenon such as policy operational risk as a factor contribute significantly to loss cost. Practical tools for studying such phenomenon in a more flexible way has been a challenge. The substantial number of categorical and sometimes numerical predictors, the presence of nonlinearities in the data and complex interactions among the predictors has been norm. Additionally, missing values for some predictors could pose a challenge. As oppose to other non-linear statistical learning methods such as neural networks and support vector machines, gradient boosting provides interpretable results via the relative influence of the input variables and their partial dependence plots (Guelman, 2012).

By considering location risk of a policy, the flexible Markov Modulated Tree- Based Gradient Boosting method is designed to integrate location risk factor into insurance pricing framework. Thus, location effect as specified in the model in Equation (4.13) was significant. Thus, while State 1 contributes about 0.08% reduction in errors, State 2, contributes about 0.6% reduction in errors in the premium function. This is significant considering the nature of business of the non-life insurance business. The results in Table 4.18 also shows that the most principal factors influencing the model is vehicle age (29.46%). The rest include comprehensive policyholders (23.86%), own goods carrying (7.88%)

etc. According to the literature about the claims ratio for comprehensive insured policy was about 118%. A fair pricing regime that consider such risk will reduce the phenomena of probability of ruin.

Based on the sample data used in this analysis, the level of accuracy in predictions was shown to be higher for MMGB relative to other models including the conventional generalized linear model approach. This is not surprising since GLMs are relatively simpler linear models and are thus constrained by the class of functions they can approximate. The MMGB method can enhance capacity of insurers to refined insurance premium predictions to reflect policy operational risk. For the industry that still rely on tariffs, the model presents a framework for evaluating their portfolios at both individual and aggregate level. It is worth noting that the Markov modulated (MMGB) framework can be an important complement to the Gradient Boosting model and the traditional Tweedie Generalized Linear Model (GLM) in insurance pricing. Even under strict circumstance where there is no regulation on risk-based pricing, our approach will still be helpful to policy makers and insurance companies to refine the tariff-based premiums based on the behavior of location risk. We find that our model performs relatively better and complements other competing models.

Conclusions

The most important financial indicator exhibiting high influence on financial performances of non-life insurance companies in Ghana is the capital adequacy ratio indicators such as net insurance risk, premium retention ratio and technical reserve cover. This is closely followed by the profitability ratio indicators and efficiency ratios. Out of the 49 variables considered, 45 were found to be significant by the model including location risk.

A risk-based pricing model (MMGB) that considers geographical location risk is developed. The MMGB model was compared with other competing models and found to be relatively more efficient than others. This could assist non-life insurance companies in Ghana in underwriting and claims management. It is believed that the present study will contribute in evaluating the insurance companies in Ghana from economic perspective.

Recommendations

It is recommended that the personal characteristics of drivers be examined and incorporated into the model in future research.

The gradient boosting framework within which MMGB is situated is an efficient framework for statistical predictive modeling due to its predictive power and flexibility and as such actuaries could use it as part of their tool box in research and analysis of insurance loss cost as well as in other related field.

In view of conclusions to the study, to ensure sustainability and fairness in pricing, a model-based MMGB approach is recommended for risk premium pricing in Ghana and other sub-saharran countries where accident risk is diverse across geographical locations. This is because the model is flexible and fairly captures the distribution of the data structure and accounts for location risk for any given policy.

It is also recommended based on the findings that the non-life insurance companies use the risk-based model as an additional tool to ascertain the level of risk of its clients. After calibrating the model to their own experience, which will involve reweighting of the factors as opposed to full model development, the applicant's score should be used as a precautionary measure. The results of the trial could then be assessed against current business practice in a champion-challenger comparison strategy. In this wise, it is expected that insurers will more likely to enjoy more stable profitable portfolio.

If the insurance industry is to spread risk across both good and bad sub-populations of applicants, there is the need for uniformity and actuarial fairness in decision-making. The National Insurance Authority should consider enacting a law that will ensure that all insurance companies in Ghana move from tariff-based to risk-based premium pricing for better loss control measure such as renewal underwriting restrictions.

It is also recommended that National Insurance Commission passes a policy that collates databases of all insurance companies that could track the behavior of drivers during bonus-malus transactions. As per the existing regime, insurers in Ghana are challenged in the implementation of bonus-malus. This is because of the ease with which policyholders can change from one company to another to avoid premium increment as a result of claim occurrence.

While we feel that the resulting model is structurally representative of the national non-life insurance experience, the model needs to be validated on a geographically diverse sample of companies in the non-life insurance sector before the model is deployed in an industrial application. This will allow for idiosyncratic differences associated with specific through-the-door clients to be captured in the model application. Further research could be done on computing the probability of ruin under MMGB model and compared with other competing models.

REFERENCES

Abor, J. (2005). The effect of capital structure on profitability. An empirical analysis of listed firms in Ghana. *Journal of Finance*. 6 (5), 438-447.

Adeleke, I. A., & Ibiwoye, A. (2011). Modeling claim size in personal line non-life insurance. *International business and Economists Research Journal*, 10 (2), 21-39.

Afriyie, D. K. (2006). *Ghana Insurance Market: Past, Present and the Future*. Accra: University of Ghana.

Akotey, J. O., & Abor, J.Y. (2013). Risk management in the Ghanaian insurance industry. *Qualitative research in financial markets*. Available at https//www.researchgae.net/publication/262852484 (accessed, January, 12, 2018).

Anstey, K. J., Wood, J., Lord, S., & Walker, J. G. (2005). Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical psychology review*, 25, 45-65.

Antonio, K. & Valdez, E.A. (2012). Statistical concepts of a priori and a posteriori risk classification in insurance. *Advances in Statistical Analysis* 96(2), 187-224.

Asmussen, S. (1989). Risk process in a Markovian environment. *Scandanavian. Actuarial Journal.* 66–100.

Bailey, R. A. & Simon, L.R.J. (1960). Two Studies in Automobile Insurance

Ratemaking. *ASTIN Bulletin*, 1(4), 192-217.

Bauer, E. & Kohavi, R. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Journal of Machine Learning*, 36: 105-139.

Binder, H. & Schumacher, M. (2008). Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. *BMC Bioinformatics*, 9 (1), 9-14.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.

Boland, P. J. (2006). Statistical and probabilistic methods in actuarial science. Boca Raton: Chapman & Hall/CRC.

Boser, B.E. (1992). A training algorithm for optimal margin classifiers. *In: Proceedings of the Fifth Annual*. ACM Workshop on Computational Learning Theory, 144–152.

Boucher, J., Denuit, M. & Guillen, M. (2008). Models of insurance claim counts with time dependence based on generalization of Poisson and Negative Binomial distributions. *Variance*. 2(1), 135-162.

Breiman, L. (1998). Arcing classifier (with discussion and a rejoinder by the author. *The Annals of Statistics*, 26, 801-849.

Brieman, L. (1999). Prediction games and arcing algorithms. Neural

Computation, 11, 1493-1517.

Brieman, L. (2001a). *Statistical modeling*: The two cultures. Statistical Science, 16, 199–231.

Brieman, L. (2001b). Random forests. *Machine Learning*, 45, 5-32.

Brieman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. CRC Press.

Brockman, M., & Wright, T. (1992). Statistical motor rating: Making effective use of your data. *Journal of the Institute of Actuaries*, 119, 457–543.

Brouhns, N., Denuit, M., Masuy, B. & Verrall, R. (2002). Ratemaking by geographical area in the Boskov and Verrall model: a case study using Belgian car insurance data. *Actuarial journal-L* 2, 3–28.

Buhlmann, H. (1967). *Experience Rating and Credibility*. ASTIN Bulleting, 4, 199-207.

Buhlmann, H. (1970). *Mathematical Methods in Risk Theory*. New York: Springer Verlag.

Buhlmann, P. & Yu, B. (2008). Sparse boosting. *J. Ma chine Learning Research*. 7, 1001-1024. MR2274395.

Buhlmann, P. and Yu, B. (2003). Boosting with the L2 loss: Regression and classification. J. *Amer. Statist. Ass*. 98, 324-339. *Casualty Actuarial Society*. Vol 67,57-109.

Buhlmann, P. & Horthorn, T. (2007). Boosting algorithms: regularization, prediction and model fitting. *Statistical Science,* 22, 477-505.

Chang, T. C., & Lin, S. J. (1999). Grey Relation Analysis of Carbon Dioxide Emissions from Industrial Production and Energy Uses in Taiwan. *Journal of Environmental Management*, 56 (4), 247-257.

Chapados, N., Bengio, Y., Vincent, P., Ghosn, J., Dugas, C., Takeuchi, I. (2001). Estimating car insurance premia: A case study in high-dimensional data inference. *University of Montreal, DIRO Technical Report,* 1199.

Charpentier, A. & Denuit, M. (2005). *Mathématiques de l'Assurance Non-Vie, Tome II: Tarification et provisionnement*. Economica, Paris.

Chernobai, A., Menn, C., Truck, S. & Rachev, S. (2005). A note on the estimation of the frequency and severity distribution of operational losses. *Mathematical Scientist,* 30 (2). Retrieved on April 20, 2017 from https://ssrn.com/abstract=675936

Chiappori, P. A. & Salanie, B. (2000). Testing for Asymmetric Information in Insurance Markets. *The Journal of Political Economy*, 108, 56-78.

Consul, P.C. & Famoye, F. (1992). Generalized Poisson regression model. *Communication Statistics (Theory & Methodology)* **2**(1), 89-109.

Cossette, H., Landriault, D., Marceau, E. (2004). Compound binomial risk model in a Markovian environment. *Insurance: Mathematics and Economics*. 35, 425-443.

Cox. D.R. and Reid, N. (1987). Parameter Orthogonality and Approximate Conditional Inference. *Journal of the Royal Statistical Society*, Series B, 49, 1-39.

Cramer, H. (1930). *On the mathematical theory of risk*. Stockholm, Skandia Jubilee Volume.

Dadey, E., Ablebu, G. & Agboda, K. (2011). Probability modeling and simulation of insurance claim. *Unpublished manuscript.*

Daykin, C. D, Pentikäinen, T. & Pesonen, M. (1994). *Practical Risk Theory for Actuaries.* London, UK: Chapman & Hall.

Delaporte, J.P. (1972). Les Mathématiques de l'Assurance Automobile. *ASTIN Bulletin* 6(3),185-190.

Deng, J. L. (1982). Control Problems of grey systems. *Systems & Control Letters*, I (5), 288-294.

Denuit, M. (2003). Tarification Automobile sur Données de Panel. *ASTIN Bulletin.* 36(1), 51-81.

Denuit, M. (2007). An Actuarial Analysis of the French Bonus-Malus System. *Scandinavian Actuarial Journal* 2006(5), 247-264.

Denuit, M. and Lang, S. (2004). Nonlife Ratemaking with Bayesian GAM'S.

Mathematics and Economics, **35** (3),627-647.

Denuit, M., Marechal, X., Pitrebois, S. & Walhin, J. F. (2007). Actuarial modelling of claim counts: Risk classification, Credibility and Bonus-malus scales. Wiley, New York.

Dickson, D.C.M., Egidio Dos Reis, A.D. & Waters, H.R. (1995). *Some stable algorithms in ruin theory and their applications*. ASTIN Bull. 25, 153-175.

Dimakos, X. K, & Rattalma, A. (2002). Bayesian premium rating with latent structure *Scandinavian Actuarial Journal.* 3, 162-184.

Dionne, G. & Vanasse, C. (1992). Automobile insurance ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics*, 7(2), 149-165.

Dionne, G., Gourieroux, C., & Vanasse, C. (2001). Testing for evidence of adverse selection in the automobile insurance market: A comment. *Journal of Political Economy*, 109, 444-453.

Djuric, Z. (2013). Collective risk models in non-life insurance. Economic Horizon. 15(2), 165-175.

Dunn, P. K. & Smyth, G. K. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15, 267-280.

Dunn, P. K. & Smyth, G. K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, 18, 73-86.

Dzakwasi, W.S. (2014). Actuarial Modeling of Motor Insurance Claim Severity and Frequency. *Masters Thesis*.

Elitas, C., Eliren, A., Yildiz, F., & Dogan, M. (2012). Grey relational analysis and determination of performance of insurance companies. *Proceedings of 16 National Finance Symposium,* (521-530). Erzurum, Turkey.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression Trees. *Journal of Animal Ecology*, 77, 802-813.

Eilers, P. H. C. & Mark, B. D. (1996). Flexible Smoothing using B-splines and Penalized likelihood (with comments and rejoinder). *Statistical Science.* 11 (2), 89-121.

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association*, 96, 1348–1360.

Fan, J. & Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Annals of Statistics*., 30, 74–99.

Feller, W. (1968). *An Introduction to probability theory and its applications.* Wiley & Sons, New York.

Feng, T. & Wang, Z. (2000). Performance evaluation for airlines including the consideration of financial ratios. *Journal of Air Transport Management*, 6 (3), 133-142.

Francis, L. (2001). Neural networks demystified. *Casualty Actuarial Society* Forum, Winter, 252–319.

Frees, E. W. (2010). Regression Modeling with actuarial and financial applications. New York: Cambridge University Press.

Frees, E.W., Meyers, G. & Cummings, A.D. (2012). Predictive modeling of multi peril homeowners' insurance. To appear in *Variance*.

Frees, E.W., Meyers, G., & Cummings, A.D. (2011). Summarizing Insurance Scores Using a Gini Index. *Journal of the American Statistical Association*, 495,1085-1098.

Freud, Y. (1990). Boosting a weak learning algorithm by majority. *A third annual workshop on computational learning.* University of Rochester, Rochester. 202-216.

Freund, Y. & Schapire, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55, 119-139.

Freund, Y. & Schapire, R. (1996). Experiments with a new boosting algorithm. Prentice Hall.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting

Machine. *Annals of Statistics* **29**(5):1189-1232.

Friedman, J. H. (2002). Stochastic Gradient Boosting. *Computational Statistics and Data Analysis* **38**(4):367-378.

Friedman, J.H., Hastie, T. & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals of Statistics* 28(2): 337-374.

Gadahn, A. (2010). *The History of Insurance*, Retrieved on April 20, 2017, Retrieved from http://righttruth.typepad.com/right_truth/2010/03/history-of-insurance.html.

Gerber, H. U. (1977). Mathematical fun with ruin theory. *Ins. Math. Econ*. 7, 15-23.

Gerber, H. U. (1988). Mathematical fun with the compound binomial process. *ASTIN Bull*. 18, 161–168.

Gerber, H. U. (1979). An introduction to Mathematical risk theory. *Philadelphia,* Pa: Huebner Foundation.

Guelman, L. (2012). Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert systems with applications*.39, 3659-3667.

Gui, J. & Li, H. (2005). Penalized cox regression analysis in the high dimensional and low-sample size settings with application to microarray gene expression data. *Bioinformatics*, 21, 3001-3008.

Guiahi, F. (2001). Fitting loss distributions in the presence of rating variables. *Jouranl of Actuarial Practice*, 9, 97-130, Absalom Press.

Guillou, A., Loisel, S. & Stupfler, G. (2013). Estimation of the parameters of a Markov-modulated process in insurance. *Insurance: Mathematics and Economics*.

Gupta, P.K. (2011). Risk management in Indian companies: EWRM concerns and issues, *The Journal of Risk Finance*, Vol. 12 No. 2, 21-39

Haberman, S. & Renshaw, A. E. (1996). Generalized linear models and actuarial science. *Journal of the Royal Statistical Society*, Series D, 45,407-436

Harrell, F.E., Califf, R.M. Pryor, D. B., Lee, K. L. & Rosati, R. A. (1982). Evaluating the Yield of Medical Tests. *Journal of the American Medical Association*. 1982;247(18):2543-2546.

Hastie, T. and Tibshirani, R. and Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction*. New York, NY: Springer.

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized additive models*. Taylor and Francis

Hofner, B., Mayr, A., Robinzonov, N. & Schmid, M. (2013). *Model-based boosting in R: A Hands-on Tutorial Using the R Package. Mboost.* Technical Report, Department of Statistics, University of Munich.doi:10.1007/s00180- 012-0382-5

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

Hogg, R.V. & Craig, A. T. (1995). *Introduction to Mathematical Statistics*. 5th edition, New Jersey, Prentice Hall.

Huang J. H. & Peng, K. H. (2011). Using Rasch Model and GRA to assess international tourist hotel industry performance. *Journal of Grey System*, 23(3), 299-310.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2015). *An introduction to statistical learning with applications in R*. Springer Texts in Statistics.

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

Jorgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society Series* B (Methodological), 127-162.

Jorgensen, B. and de Souza, M.C. (1994). Fitting Tweedie's compound Poisson model to insurance claim data. *Scandinavian Actuarial Journal*, 69-93.

Kaas, R., Goovaerts, M. J., Dhaene, J. & Denuit, M. (2009). *Modern Actuarial*

Kaya, E.O. (2016). Financial Performance assessment of Non-life Insurance Companies Traded in Borsa Instanbul via Grey Relational Analysis. *International Journal of Economics and Finance*, 8 (4).

Kestemont, R. M., & Paris, J. (1985). *Introduction to Mathematical Statistics*, 5th edition Printice Hall.

Klugman, A. S., Panjer, H. H. & Willmot, E.G. (2004). Loss models: from data to decisions. John Wiley & Sons, Inc.

Konstantinides, D., Tang, Q., & Tsitsiashvili, G., (2002). Estimates for the ruin probability in the classical risk model with constant interest force in the presence of heavy tails. *Insurance Math. Economics*. 31, no. 3, 447-460.

Kolyshkina, I., Wong, S. & Lim, S. (2004). Enhancing generalized linear models with data mining. *Casualty Actuarial Society*, Discussion paper program.

Kula, V., Kandemir, T., & Baykut, E. (2015). Borsa Instanbul Bunyesinde Islem Goren Sigorta. Proceedings of *19th National Finance Symposium*. 217-233.

Kung, C.Y., & Cheng, C. R. (2004). Grey assessing the performance of enterprise outsourcing management. *Journal of Grey System*, 16 (1), 63-72.

Kung, C. Y., Yan, T. M., & Chuang, S. C. (2006). GRA to assess the operating performance of non-life insurance companies in Taiwan. Journal of Grey System, 18(2), 155-160.

Lefevre, C. & Picard, P. (1996). Collective epidemic models. *Mathematical Biosciences*. 134, 51-70.

Lehtonen, T. & Nyrhinen, H. (1992). On asymptotically efficient simulation of ruin probabilities in a Markovian environment. *Scand. Act. J.* 60–75.

Lemaire, L. (1985). Automobile Insurance: Actuarial *Models. Huebner International Series on Risk, Insurance and Economic Security.*

Lemaire, L. (1995). *Bonus-malus systems in automobile insurance*. Kluwer Academic Publisher, Boston.

Li, H. & Luan, Y. (2005). Boosting Proportional Hazards Models Using Smoothing Splines with Applications to High-Dimensional Microarray Data. *Bioinformatics* 2005; 21 (10): 2403 -2409.

Lillo, R. E. & Semeraro, P. (2004). Stochastic bounds for discrete-time claim processes with correlated risks. *Scand. Act. J.* 1–13.

Littlestone, N. & Warmuth, M. K. (1989). The Weighted Majority Algorithm. *In: Foundations of Computer Science*, 30th Annual Symposium on. IEEE, 256-261.

Loisel, S. (2004). Finite-time ruin probabilities in the Markov-modulated multivariate Compound Poisson model with common shocks and impact of dependence. *Cahiers de recherche de l'ISFA, WP 2027.*

Lin, C. T., & Chang, H. F. (2010). Performance evaluation for medical industry in Taiwan: Apply Grey Situation Decision making. *Journal of Grey System.* 22 (3), 219-226.

Lundberg, F. (1932). Some supplementary research on the collective risk theory. *Skandinavisk Aktuarietidskrift, 15,* 137- 158.

Markov, U.E. (2002). Principal applications of Bayesian methods in actuarial science: a perspective. *North Am. Actuarial J.* 5, 53–57.

Mayr A, Hofner B. & Schmid M. (2012). The importance of knowing when to stop a sequential stopping rule for component-wise Ggadient boosting methods. *Inf Med*, 51 (2): 178 -186.

McCartt, A. T., Shabanova, V. I., & Leaf, W. A. (2003). Driving experience, crashes and traffic citations of teenage beginning drivers. *Accident Analysis & Prevention*, 35, 311-320.

McClenahan C. L. (2001). Ratemaking. *Foundations of Casualty Actuarial Science*: 25-90. New York: R&S Financial Printing.

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*. Chapman Hall, Boca Raton, 2nd edition.

McCullagh, P., & Nelder, J. (1989). Generalized linear models (2nd ed.). Boca Raton, Chapman and Hall.

Mease, D. & Wyner, A. (2008). Evidence contrary to the statistical view of boosting. *The Journal of Machine Learning Research*, 9: 131-142.

Meir, R. & Ratsch, G. (2003). An introduction to boosting and leveraging. In *Advanced Lectures on Machine Learning* (S. Mendelson and A. Smolaeds.). Springer, Berlin.

Mihaela, D. (2015). Insurance pricing using generalized linear models. *Procedia Economics and Finance*, 20, 147-156.

Mildenhall, S. J. (1999). A systematic relationship between minimum bias and generalized linear models. *in Proceedings of the Casualty Actuarial Society*, vol. 86, pp. 393-487.

Moncher, B. R. & Fu, L. (2004). *Severity distribution for GLMs*: Gamma or Lognormal? Evidence from Monte Carlo Simulations.

Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a differentiable pure premium? *Proceedings of the Casualty Actuarial Society*. 181-221.

Murphy, K. P., Brockman, M. J., & Lee, P. K. (2000). Using generalized linear models to build dynamic pricing systems. *in Casualty Actuarial Society Forum*, Winter, pp. 107-139.

Nelder, J. A. & Lee, Y. (1992). Likelihood, quasi-likelihood and pseudo-likelihood: some comparisons. *J. R. Statist. Soc*. B., 54: 273-284.

Nelder, J. & Wedderburn, R. (1972). Generalized Linear Models. *Journal of Statistical Society*. Series A, 135,370-384.

NIC (2015), *Annual Report*, 2015, National Insurance Commission at: www.nic.gh.org (accessed on January, 15,2018).

NIC (2016). *Annual Report*, 2016, National Insurance Commission office.

Nyrhinen, H. (1998). Rough descriptions of ruin for a general class of surplus processes. *Adv. Appl. Prob.* 1008–1026.

Ohlsson, E. & Johansson, B. (2010), *Non-life insurance pricing with generalized linear models*, Springer Verlag.

Owsley, C., Ball, K., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1991). Visual cognitive correlates of vehicle accidents in older *drivers. Psychology and aging*, 6, 403.

Panjer, H. & Willmot, G. (1992). *Tnsurallce risk models.* Society of Actuaries. Schaumburg.

Patrik, G. (1980). Estimating casualty insurance loss amount distributions. *Calsualty Actuarial Society.* 67, 57-109.

Peker, I., & Baki, B. (2011). Performance Evaluation in Turkish Insurance Sector with Grey Relational Analysis. *International Journal of Economic and Administrative Studies,* 4 (7), 1-17.

Pesonen, E. (1962). A Numerical Method of Finding a Suitable Bonus Scale. *ASTIN Bulletin*, 2(1), 102-108.

Peters, G. W., Shevchenko, P. V., & Wuthrich, M. V. (2008). Model risk in claims reserving within Tweedie's compound Poisson models. *ASTIN Bulletin*, to appear.

PricewaterhouseCoopers Ghana report (2010).

Ratsch, G., Onoda, T. & Muller, K. R. (2001). Soft margins for Adaboost. *Neurocolt technical report* nc-tr-98-021.

Renshaw, A. E. (1994). Modeling the claims process in the presence of covariates. *ASTIN Bulletin*, 24, 265-285.

Reyzin, L. & Schapire R.E. (2006). How boosting the margin can also boost classifier complexity. *In proceeding of the 23rd International Conference on Machine Learning*, 753-760.

Ridgeway, G. (1999). The state of boosting. *Computing Science and Statistics*. 31:172-181

Ridgeway, G. (2007). *Generalized boosted models:* a guide to the gbm package. Available from http://cran.r-project.org/web/packages/gbm/index.html

Ridgeway, G. (2012). *Genearlized Boosted Models*: A guide to the gbm package. Accessed at Comprehensive R Application Network (CRAN).

Rigby, R. A. & Stasinopoulos, D. M. (1996). Smooth Centile Curves for Skew and Kurtotic Data Modelled Using the Box–Cox Power Exponential Distribution. *Statistics in Medicine*, 23, 3053-3076.

Rolski, T., Schmidli, H., Schmidt, V. & Teugels, J. (1999). *Stochastic processes for insurance and finance*. Wiley, New York.

Ruohonen, M. (1987). On a model for the claim number process. *ASTIN Bulletin* 18,57-68.

Sandri, M. & Zuccolotto, P. (2010). Analysis and correction of bias in total decrease in node impurity measures for tree-based algorithms. *Statistics and Computing*, 20, 393-407.

Santomero, A. M. & Babbel, D. F. (1997). Financial risk management by insurers: an analysis of the process. *The Journal of Risk and Insurance*, Vol. 64, No. 2, 231-70.

Saunders, M., Lewis, P. & Thornhill, A. (2007). *Research methods for business students* (4th edn). Harlow: FT. Prentice Hall.

Schmid, M. & Hothorn, T. (2008). Boosting additive models using component-wise P-splines. *Computational Statistics & Data Analysis*, 53: 298 –311.

Schapire, R. (1990). The strength of weak learnability. *Machine Learning*, 5, 197-227.

Schapire, R. E. & Freund, Y. (2012). *Boosting: Foundations and Algorithms*. MIT Press.

Sharif, A. H. & Panjer, H. H. (1993). A probabilistic model for automobile claims: a comment on the article by Islam and Consul. *Bulleting of the Swiss Association of Actuaries*, 279-282.

Showers, V. E. & Shotick, J. A. (1994). The effects of household characteristics on demand for insurance: A tobit analysis. *Journal of Risk and Insurance*, 61(3),492-502.

Smyth, G. & Jorgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: Dispersion modeling. *ASTIN Bulletin*, 32, 143-157.

Smyth, G. K. (1996). Regression analysis of quantity data with exact zeros. in Proceedings of the second Australia-Japan workshop on stochastic models in engineering, technology and management, *Citeseer*, pp. 572-580.

Tsai, H. Y., Huang, B. H. & Wang, A. S. (2008). Combining AHP and GRA Model for Property-Liability Insurance Companies to Rank. *Journal of Grey System*, 20 (1), 65-78.

Tutz, G. & Binder, H. (2006). generalized additive modeling with implicit variable selection by likelihood-based boosting. *Biometrics*. 62:961–971.

Weisberg, H. I., & Tomberlin, T.J. (1982). A Statistical perspective on actuarial methods for estimating pure premiums from cross-classified data. *The Journal of Risk and Insurance*, 49(4), 539-563.

Wen, K. L. (2004). The Grey System Analysis and its application in gas breakdown and var compensator finding (invited Paper). *International Journal of Computational Cognition*, 2(1), 21-44.

Werner, G.& Modlin, C. (2010). *Basic ratemaking*. Fifth Edition. Casualty Actuarial Society.

White, A. P., & Liu, W. Z. (1994). Technical note: Bias in information-based measures in decision tree induction. *Machine Learning*, 15, 321-329.

Whitney, A.W. (1918). The Theory of Experience Rating. *Proceedings of the Casualty Actuarial Society* 4, 274-292.

Willmot, G. (1987). The Poisson-inverse Gaussian as an alternative to the negative binomial. *Scand. Actuarial J. 113-127.*

Wood, S. N. (2006). *Generalized additive models: an introduction to R.* Chapman & Hall/CRC.

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

Wu, H. H. (2002). A comparative study of using grey relational analysis in multiple attribute decision-making problems. *Quality Engineering*, 15 (2), 209-217.

Wu, J. H., & Chen, C. B. (1999). An alternative form for grey relational grades. *Journal of Grey System*, 11(1), 7-12.

Wu, W.Y., Hsiao, S. W., & Tsai, C.H. (2008). Forecasting and evaluating the Tourist Hotel industry.

Yan, T. M., & Kung, C. Y. (2011). Business performance assessment of insurance company via grey relation analysis. *Journal of Grey System*, 23 (1), 83-90.

Yang, Y., Qian, W., & Zou, H. (2016): Insurance premium prediction via gradient tree-based Tweedie coumpound Poisson models, *Journal of Business and Economics*. 10.1080/07350015.2016.1200981.

Zhai, L. Y., Khoo, L. P., & Zhong, Z. W. (2009). Decision concept evaluation in product development using rough relation analysis. *Expert Systems with Applications*, 36 (3), 7072-7079.

Zhang, T. & Yu, B. (2005). Boosting with early stopping: Convergence and consistency. *The Annals of Statistics*, 1538-1579.

Zhang, W. (2011), cplm: *Monte Carlo EM algorithms and Bayesian methods for fitting Tweedie compound Poisson linear models.* R package, http://cran.r- project.org/web/packages/cplm/index.html.

Zhou, Z. H. (2012). *Ensemble Methods: Foundations and algorithms*. CRC Machine Learning & Pattern Recognition. Chapman & Hall.

Zuanetti, D.A., Diniz, R. A. C., & Leite, J. G. (2006). A lognormal model for insurance claims data. *REVSTAT - Statistical Journal 4* (2), 131-1

APPENDICES

APPENDIX A

SUMMARY OF ACCIDENT DATA IN GHANA

APPENDIX A-1:

TREND IN CAR CRASHES BY REGION

Year	Region										
	Ashanti	Brong Ahafo	Central	Eastern	Greater Accra	Northern	Upper East	Upper West	Volta	Western	Total
1991	1021	493	542	1162	3713	148	129	65	366	731	8370
1992	1277	232	631	1314	2097	84	138	68	344	737	6922
1993	1102	247	490	1072	2187	34	129	52	382	772	6467
1994	1272	261	599	1078	2302	72	88	32	192	688	6584
1995	1636	343	498	730	3645	103	130	85	422	721	8313
1996	1415	346	643	1052	3654	119	146	75	423	615	8488
1997	2144	496	686	1052	4231	130	96	50	266	767	9918
1998	2161	534	654	1261	4963	134	106	54	350	780	10997
1999	1222	591	760	1186	3414	171	154	78	363	823	8762
2000	1818	630	918	1421	5234	188	169	103	509	724	11714
2001	1680	494	955	1397	5003	225	173	86	594	684	11291
2002	1774	588	831	1469	4230	193	209	66	546	809	10715
2003	1811	562	907	1383	4110	203	225	64	517	756	10538
2004	2037	691	1026	1703	4624	323	209	72	682	800	12167
2005	1680	655	916	1445	4983	224	181	82	567	595	11328
2006	1706	621	883	1351	5454	266	125	62	522	678	11668
2007	1975	541	709	1349	5936	255	136	73	495	569	12038
2008	1779	691	756	1295	5044	257	155	79	503	655	11214
2009	1971	693	917	1340	5588	220	119	61	554	836	12299
2010	1944	543	982	1182	5122	257	88	90	599	699	11506
2011	2094	726	1046	1100	4311	260	110	118	481	641	10887
2012	1990	669	1041	1220	5247	223	114	155	488	936	12083
2013	1312	628	804	971	3925	274	110	172	477	527	9200
2014	1358	560	866	1076	3745	292	115	132	438	570	9152
2015	1690	474	838	913	4259	259	169	105	478	611	9796
Total (2013-2015)	4360	1662	2508	2960	11929	825	394	409	1393	1708	28148
%	0.155	0.059	0.089	0.105	0.424	0.029	0.014	0.015	0.049	0.061	1.00

APPENDIX A-2:

TREND IN FATALITIES BY REGION

Year	Region										
	Ashanti	Brong Ahafo	Central	Eastern	Greater Accra	Northern	Upper East	Upper West	Volta	Western	Total
1991	128	76	86	150	90	31	20	11	54	61	707
1992	124	50	86	138	140	23	28	8	45	75	717
1993	143	47	69	155	111	12	14	12	53	88	704
1994	131	54	91	139	113	13	17	3	23	48	632
1995	190	70	81	100	176	22	18	10	62	84	813
1996	85	82	120	138	165	41	25	15	82	73	826
1997	187	86	100	150	160	32	13	5	37	65	835
1998	243	113	104	210	243	34	10	9	72	90	1128
1999	152	102	144	185	158	49	26	19	50	94	979
2000	247	107	134	201	214	44	47	16	79	110	1199
2001	263	87	156	205	220	54	32	14	112	114	1257
2002	251	157	141	226	150	47	39	18	108	108	1245
2003	306	109	148	196	207	76	47	22	114	120	1345
2004	377	151	176	240	253	96	54	21	111	121	1600
2005	249	130	156	236	259	71	63	21	90	116	1391
2006	257	172	138	174	305	76	45	21	129	102	1419
2007	332	163	146	218	363	80	63	25	118	114	1622
2008	343	136	150	238	351	77	54	33	131	134	1647
2009	388	168	181	261	385	63	52	33	117	142	1790
2010	362	142	167	199	385	84	44	42	122	139	1686
2011	367	219	163	207	360	65	51	42	103	161	1738
2012	355	168	172	219	524	60	58	67	87	203	1913
2013	342	161	139	169	327	89	54	61	123	103	1568
2014	273	155	149	169	383	73	59	37	100	122	1520
2015	310	142	149	181	439	63	39	40	87	138	1588
Total (2013-2015)	925	458	437	519	1149	225	152	138	310	363	4676
%	0.198	0.098	0.093	0.111	0.246	0.048	0.033	0.030	0.066	0.078	1.00

APPENDIX -3:

ANNUAL DISTRIBUTION OF TRAFFIC FATALITIES BY REGION

Year	Region										
	Ashanti	Brong Ahafo	Central	Eastern	Greater Accra	Northern	Upper East	Upper West	Volta	Western	Total
1991	183	96	98	183	126	41	23	13	92	65	920
1992	153	61	122	204	164	30	32	8	50	90	914
1993	168	100	97	207	115	17	14	16	59	108	901
1994	161	69	123	186	155	31	20	3	27	49	824
1995	234	86	118	152	190	28	21	13	80	104	1026
1996	95	108	151	196	186	57	31	15	126	84	1049
1997	220	100	131	181	174	35	14	6	43	111	1015
1998	283	144	145	285	258	71	13	9	101	111	1420
1999	178	124	165	294	172	76	30	22	72	104	1237
2000	332	141	199	272	237	60	85	18	89	145	1578
2001	379	152	206	279	240	66	34	17	152	135	1660
2002	359	190	215	346	169	71	44	20	130	121	1665
2003	377	140	188	263	232	138	53	35	152	138	1716
2004	577	202	234	325	299	131	68	24	167	158	2185
2005	315	192	183	299	313	97	79	30	122	154	1784
2006	388	244	184	216	335	112	44	21	169	143	1856
2007	463	207	190	280	407	105	69	27	145	150	2043
2008	416	155	150	294	385	95	59	36	179	169	1938
2009	469	259	246	343	429	113	54	40	140	144	2237
2010	454	169	167	259	424	114	45	54	143	157	1986
2011	474	297	203	248	408	123	54	50	139	203	2199
2012	432	221	207	316	535	99	61	71	96	202	2240
2013	406	201	200	197	363	140	57	72	142	120	1898
2014	351	198	198	209	419	97	60	42	125	137	1836
2015	352	170	185	197	458	85	46	40	105	164	1802
Total (2013-2015)	1109	569	583	603	1240	322	163	154	372	421	5536
%	0.200	0.103	0.105	0.109	0.224	0.058	0.029	0.028	0.067	0.076	1.00

APPENDIX - 4

ANNUAL DISTRIBUTION OF CASUALTIES BY REGION

Year	Region										
	Ashanti	Brong Ahafo	Central	Eastern	Greater Accra	Northern	Upper East	Upper West	Volta	Western	Total
1991	1389	904	958	2133	2002	306	217	97	745	933	9684
1992	1713	481	1165	2461	1905	205	259	140	619	1082	10030
1993	1376	514	850	2088	1703	86	199	140	688	934	8578
1994	1738	530	1130	1862	1819	143	137	76	346	707	8488
1995	2000	588	981	1588	2453	199	223	147	828	1124	10131
1996	1141	916	1323	2487	2612	269	294	125	842	916	10925
1997	2341	814	1276	2181	2764	277	132	96	495	1072	11448
1998	2387	1021	1350	2884	3182	293	133	104	730	1125	13209
1999	1783	931	1409	2608	2278	321	264	142	619	1083	11438
2000	2608	920	2101	2899	3295	335	312	198	905	1091	14664
2001	2386	952	1681	3013	3420	439	339	185	1325	1093	14833
2002	2482	1168	1991	3185	2798	473	304	121	1187	1365	15074
2003	3548	1039	2193	2882	3136	623	323	133	1226	1211	16314
2004	3676	1451	1943	3148	3782	806	322	134	1828	1346	18436
2005	2913	1346	1602	2995	3566	448	291	187	1445	1045	15838
2006	2604	1261	1170	2501	3880	594	144	86	1189	1063	14492
2007	3243	1121	1324	2662	4857	615	245	117	1056	1176	16416
2008	2856	1512	1438	2749	4267	745	241	174	1271	1216	16469
2009	3663	1538	1862	2897	4971	524	199	144	1155	1543	18496
2010	3752	1303	1595	2483	4293	812	145	183	1123	1215	16904
2011	3614	1514	1550	2331	3794	689	197	210	1168	1152	16219
2012	3298	1308	1397	2345	4090	464	191	257	671	1220	15241
2013	2372	1139	1056	1813	3169	786	177	296	936	765	12509
2014	2561	1118	1195	2037	3419	578	179	205	767	804	12863
2015	2616	857	1344	1664	3417	475	268	158	625	943	12367
Total (2013-2015)	7549	3114	3595	5514	10005	1839	624	659	2328	2512	37739
%	0.200	0.083	0.095	0.146	0.265	0.049	0.017	0.017	0.062	0.067	1.00

APPENDIX - 5

REGIONAL CRASH FATALITIES AND CASUALTIES BY RANK

Region	All Crashes	Rank	Region	All Fatal Clashes	Rank	Region	Fatalities	Rank	Region	Casualties	Rank
Greater Accra	42.4	1.0	Greater Accra	20.9	1.0	Ashanti	21.4	1.0	Greater Accra	26.5	1.0
Ashanti	15.5	2.0	Ashanti	20.5	2.0	Greater Accra	18.5	2.0	Ashanti	20.0	2.0
Eastern	10.5	3.0	Eastern	13.6	3.0	Eastern	14.2	3.0	Eastern	14.6	3.0
Central	8.9	4.0	Central	10.0	4.0	Central	10.3	4.0	Central	9.5	4.0
Western	6.1	5.0	Brong Ahafo	9.7	5.0	Brong Ahafo	10.2	5.0	Brong Ahafo	8.3	5.0
Brong Ahafo	5.9	6.0	Western	8.3	6.0	Western	8.0	6.0	Volta	6.2	6.0
Volta	4.9	7.0	Volta	7.1	7.0	Volta	7.2	7.0	Northern	4.9	7.0
Northern	2.9	8.0	Northern	4.6	8.0	Northern	5.4	8.0	Upper West	1.7	8.0
Upper West	1.5	9.0	Upper East	3.3	9.0	Upper East	3.0	9.0	Upper East	1.7	9.0
Upper East	1.4	10.0	Upper West	2.1	10.0	Upper West	1.9	10.0	Western	6.7	10.0

Source: National Road Safety Commission (2017)

APPENDIX B

PROCEDURE FOR MODEL PARAMETER ESTIMATION

Appendix B-1 to Appendix B-5 displays the five-fold cross validation results for 1000 trees at step size 0.05. The last column shows the marginal contribution to reduction in error sum of squares for each iteration per sample.

APPENDIX B - 1:

OBTAINING OPTIMAL TREE NUMBER VIA CROSS VALIDATION (SAMPLE 1)

Iter	Train Deviance	Valid Deviance	Step Size	Improve
1	41.1767	39.7637	0.0050	0.1246
2	41.0557	39.6497	0.0050	0.1243
3	40.9295	39.5344	0.0050	0.1156
4	40.8023	39.4215	0.0050	0.1102
5	40.6842	39.3148	0.0050	0.1103
6	40.5734	39.2139	0.0050	0.1111
7	40.4545	39.1057	0.0050	0.1102
8	40.3363	38.9988	0.0050	0.1083
9	40.2187	38.8898	0.0050	0.1093
10	40.1065	38.7867	0.0050	0.1057
100	33.3692	32.3567	0.0050	0.0510
200	30.0408	29.0340	0.0050	0.0276
300	28.3740	27.3240	0.0050	0.0079
400	27.5524	26.4893	0.0050	0.0056
500	27.0140	25.9981	0.0050	0.0017
600	26.6393	25.6805	0.0050	-0.0010
700	26.3215	25.4456	0.0050	0.0002
800	26.0496	25.2621	0.0050	0.0009
900	25.7816	25.0925	0.0050	0.0017
1000	25.5398	24.9102	0.0050	-0.0003

APPENDIX B-2

OBTAINING OPTIMAL TREE NUMBER VIA CROSS VALIDATION (SAMPLE 2)

Iter	Train Deviance	Valid Deviance	Step Size	Improve
1	41.1767	39.7637	0.0050	0.1246
2	41.0557	39.6497	0.0050	0.1243
3	40.9295	39.5344	0.0050	0.1156
4	40.8023	39.4215	0.0050	0.1102
5	40.6842	39.3148	0.0050	0.1103
6	40.5734	39.2139	0.0050	0.1111
7	40.4545	39.1057	0.0050	0.1102
8	40.3363	38.9988	0.0050	0.1083
9	40.2187	38.8898	0.0050	0.1093
10	40.1065	38.7867	0.0050	0.1057
100	33.3692	32.3567	0.0050	0.0510
200	30.0408	29.0340	0.0050	0.0276
300	28.3740	27.3240	0.0050	0.0079
400	27.5524	26.4893	0.0050	0.0056
500	27.0140	25.9981	0.0050	0.0017
600	26.6393	25.6805	0.0050	-0.0010
700	26.3215	25.4456	0.0050	0.0002
800	26.0496	25.2621	0.0050	0.0009
900	25.7816	25.0925	0.0050	0.0017
1000	25.5398	24.9102	0.0050	-0.0003

APPENDIX B-3

OBTAINING OPTIMAL TREE NUMBER VIA CROSS VALIDATION (SAMPLE 3)

Iter	Train Deviance	Valid Deviance	StepSize	Improve
1	40.9766	40.5660	0.0050	0.1265
2	40.8541	40.4534	0.0050	0.1266
3	40.7313	40.3420	0.0050	0.1215
4	40.6166	40.2405	0.0050	0.1198
5	40.5029	40.1375	0.0050	0.1190
6	40.3823	40.0288	0.0050	0.1129
7	40.2621	39.9197	0.0050	0.1176
8	40.1405	39.8108	0.0050	0.1135
9	40.0296	39.7097	0.0050	0.1146
10	39.9100	39.5989	0.0050	0.1197
100	33.2072	33.5401	0.0050	0.0427
200	29.7732	30.2904	0.0050	0.0199
300	28.0618	28.6481	0.0050	0.0146
400	27.1983	27.9430	0.0050	0.0013
500	26.6538	27.5120	0.0050	0.0024
600	26.2815	27.2230	0.0050	0.0013
700	25.9833	27.0320	0.0050	0.0030
800	25.7369	26.8394	0.0050	0.0003
900	25.4927	26.6826	0.0050	-0.0005
1000	25.2594	26.6468	0.0050	0.0007

APPENDIX B-4

OBTAINING OPTIMAL TREE NUMBER VIA CROSS VALIDATION (SAMPLE 4)

Iter	Train Deviance	Valid Deviance	StepSize	Improve
1	40.9878	40.5380	0.0050	0.1176
2	40.8593	40.4191	0.0050	0.1168
3	40.7401	40.3030	0.0050	0.1095
4	40.6198	40.1914	0.0050	0.1148
5	40.4948	40.0711	0.0050	0.1307
6	40.3680	39.9502	0.0050	0.1236
7	40.2492	39.8383	0.0050	0.1112
8	40.1330	39.7298	0.0050	0.1084
9	40.0210	39.6228	0.0050	0.1073
10	39.9044	39.5139	0.0050	0.1115
100	33.2641	33.1160	0.0050	0.0381
200	29.8707	29.8726	0.0050	0.0137
300	28.1727	28.2860	0.0050	0.0098
400	27.2752	27.5100	0.0050	0.0023
500	26.7320	27.1195	0.0050	0.0032
600	26.3152	26.8910	0.0050	0.0038
700	25.9854	26.7276	0.0050	0.0014
800	25.7037	26.5996	0.0050	0.0005
900	25.4310	26.5252	0.0050	0.0010
1000	25.2123	26.4659	0.0050	-0.0000

APPENDIX B-5

OBTAINING OPTIMAL TREE NUMBER VIA CROSS VALIDATION (SAMPLE 5)

Iter	Train Deviance	Valid Deviance	StepSize	Improve
1	41.4014	38.9014	0.0050	0.1214
2	41.2701	38.7863	0.0050	0.1262
3	41.1421	38.6701	0.0050	0.1249
4	41.0151	38.5606	0.0050	0.1227
5	40.8829	38.4417	0.0050	0.1258
6	40.7609	38.3316	0.0050	0.1172
7	40.6409	38.2279	0.0050	0.1107
8	40.5218	38.1192	0.0050	0.1167
9	40.3826	37.9962	0.0050	0.1325
10	40.2587	37.8882	0.0050	0.1100
100	33.2764	31.8120	0.0050	0.0515
200	29.8321	29.0609	0.0050	0.0213
300	28.1416	27.8214	0.0050	0.0119
400	27.2470	27.2035	0.0050	0.0045
500	26.7150	26.9244	0.0050	0.0013
600	26.3224	26.7349	0.0050	0.0019
700	26.0061	26.5953	0.0050	0.0014
800	25.7409	26.4818	0.0050	0.0002
900	25.4865	26.3860	0.0050	-0.0001
1000	25.2415	26.2792	0.0050	-0.0001

APPENDIX B-6

TESTING THE OPTIMAL TREE NUMBER

Iter	Train Deviance	ValidDeviance	StepSize	Improve
1	40.8970	nan	0.0050	0.1204
2	40.7745	nan	0.0050	0.1257
3	40.6556	nan	0.0050	0.1200
4	40.5340	nan	0.0050	0.1184
5	40.4082	nan	0.0050	0.1193
6	40.2899	nan	0.0050	0.1157
7	40.1793	nan	0.0050	0.1175
8	40.0624	nan	0.0050	0.1137
9	39.9476	nan	0.0050	0.1151
10	39.8340	nan	0.0050	0.1113
100	33.1457	nan	0.0050	0.0420
200	29.7316	nan	0.0050	0.0205
300	28.0736	nan	0.0050	0.0062
400	27.2193	nan	0.0050	0.0057
500	26.7207	nan	0.0050	-0.0008
600	26.3607	nan	0.0050	0.0028
700	26.0482	nan	0.0050	0.0007
800	25.7912	nan	0.0050	0.0007
900	25.5704	nan	0.0050	0.0000
1000	25.3474	nan	0.0050	0.0006

APPENDIX C

STEPS FOR OPTIMAL NUMBER OF ITERATIONS

A three-fold cross validation with number of iterations (iter) or trees specified as 1788 and 0.05 step size. The last column shows the marginal contribution to reduction in error sum of squares for each iteration per sample.

APPENDIX C-1

SEARCHING FOR OPTIMAL ITERATION NUMBER (SAMPLE 1)

Iter	Train Deviance	Valid Deviance	Step Size	Improve
1	40.8583	40.9910	0.0050	0.1171
2	40.7320	40.8618	0.0050	0.1214
3	40.6112	40.7467	0.0050	0.1140
4	40.4983	40.6320	0.0050	0.1130
5	40.3837	40.5184	0.0050	0.1046
6	40.2696	40.4031	0.0050	0.1114
7	40.1552	40.2868	0.0050	0.1057
8	40.0379	40.1682	0.0050	0.1066
9	39.9272	40.0599	0.0050	0.1026
10	39.8147	39.9445	0.0050	0.1044
100	33.0013	33.2720	0.0050	0.0480
200	29.5573	30.0916	0.0050	0.0265
300	27.8710	28.6270	0.0050	0.0064
400	27.0280	27.8933	0.0050	0.0062
500	26.4897	27.5728	0.0050	0.0008
600	26.0971	27.3547	0.0050	0.0021
700	25.7683	27.1425	0.0050	-0.0000
800	25.4854	26.9842	0.0050	0.0001
900	25.2202	26.8252	0.0050	0.0006
1000	24.9770	26.7249	0.0050	0.0001
1100	24.7644	26.6370	0.0050	0.0002
1200	24.5739	26.5964	0.0050	0.0010
1300	24.3709	26.6677	0.0050	-0.0000
1400	24.1888	26.6768	0.0050	-0.0006
1500	24.0270	27.0130	0.0050	0.0008
1600	23.8869	27.2785	0.0050	-0.0002
1700	23.7471	27.5573	0.0050	-0.0006
1788	23.6200	27.9230	0.0050	-0.0008

APPENDIX C-2

SEARCHING FOR OPTIMAL ITERATION NUMBER (SAMPLE 2)

Iter	Train Deviance	Valid Deviance	Step Size	Improve
1	40.3172	42.0866	0.0050	0.1158
2	40.2041	41.9677	0.0050	0.1110
3	40.0868	41.8418	0.0050	0.1075
4	39.9793	41.7260	0.0050	0.1078
5	39.8665	41.6058	0.0050	0.1091
6	39.7556	41.4836	0.0050	0.1089
7	39.6441	41.3634	0.0050	0.1071
8	39.5415	41.2473	0.0050	0.0980
9	39.4324	41.1219	0.0050	0.1016
10	39.3217	41.0019	0.0050	0.1051
100	32.9259	33.9730	0.0050	0.0463
200	29.6225	30.4037	0.0050	0.0163
300	27.9614	28.7059	0.0050	0.0069
400	27.0411	27.8572	0.0050	0.0056
500	26.4699	27.4581	0.0050	0.0025
600	26.0481	27.1920	0.0050	0.0014
700	25.7039	27.0146	0.0050	0.0005
800	25.4061	26.8555	0.0050	0.0000
900	25.1395	26.7316	0.0050	0.0000
1000	24.8948	26.6595	0.0050	0.0003
1100	24.6771	26.6176	0.0050	0.0004
1200	24.4673	26.6020	0.0050	-0.0001
1300	24.2792	26.7675	0.0050	-0.0003
1400	24.1088	27.0265	0.0050	-0.0003
1500	23.9520	27.3893	0.0050	0.0000
1600	23.7992	28.4527	0.0050	0.0002
1700	23.6487	29.4877	0.0050	0.0028
1788	23.5231	30.4047	0.0050	-0.0028

APPENDIX C-3

SEARCHING FOR OPTIMAL ITERATION NUMBER (SAMPLE 3)

Iter	Train Deviance	Valid Deviance	Step Size	Improve
1	41.5043	39.6874	0.0050	0.1292
2	41.3767	39.5785	0.0050	0.1290
3	41.2419	39.4668	0.0050	0.1190
4	41.1095	39.3546	0.0050	0.1347
5	40.9766	39.2447	0.0050	0.1246
6	40.8416	39.1294	0.0050	0.1259
7	40.7171	39.0210	0.0050	0.1233
8	40.5887	38.9081	0.0050	0.1237
9	40.4647	38.8042	0.0050	0.1181
10	40.3458	38.7035	0.0050	0.1116
100	33.3631	32.8472	0.0050	0.0407
200	29.8048	29.8511	0.0050	0.0225
300	28.1134	28.4743	0.0050	0.0091
400	27.2272	27.7879	0.0050	0.0035
500	26.6744	27.4545	0.0050	-0.0017
600	26.2638	27.1856	0.0050	0.0008
700	25.9125	26.9361	0.0050	0.0020
800	25.5994	26.7449	0.0050	0.0011
900	25.3259	26.5982	0.0050	-0.0003
1000	25.0987	26.4803	0.0050	0.0015
1100	24.8668	26.3620	0.0050	-0.0003
1200	24.6649	26.2815	0.0050	0.0002
1300	24.4767	26.2022	0.0050	0.0001
1400	24.2975	26.1297	0.0050	-0.0004
1500	24.1280	26.0798	0.0050	0.0001
1600	23.9956	26.0384	0.0050	-0.0001
1700	23.8480	26.0527	0.0050	0.0002
1788	23.7338	26.0436	0.0050	0.0002

APPENDIX D

PRACTICAL ILLUSTRATION OF STATISTICAL LEARNING.

In this section, we demonstrate the application of classification and regression trees within the R framework.

The 'tree' and "ISLR" were used to illustrate the construct of classification and regression trees. In this demonstration, we use classification trees to analyze the Carseats data set. The data set contains eleven variables.

```
>library(tree)
> library(ISLR)
> data("Carseats")
> names(Carseats)
 [1] "Sales"     "CompPrice" "Income"    "Advertising" "Population"
 [6] "Price"     "ShelveLoc" "Age"       "Education"   "Urban"
[11] "US"
```

The "Sales" variableis a continous variable and hence we recode it as a binary variable for the purpose of illustration. The function *ifelse()* is used to create a variable called High, which takes on a value "Yes" if sales is more that 10 and takes on a value of "No" otherwise.

```
> attach(Carseats)
> High<-ifelse(Sales<=10,"No","Yes")
```

We then use the *data.frame ()* function to merge High with the rest of the data.

```
>Carseats<-data.frame(Carseats,High)
```

```
> names(Carseats)

[1] "Sales"       "CompPrice"
[3] "Income"      "Advertising"
[5] "Population"  "Price"
[7] "ShelveLoc"   "Age"
[9] "Education"   "Urban"
[11] "US"         "High"
```

The *tree()* function is then used to fit a classification in order to predict High using all variables but Sales.

```
> tree.carseats<-tree(High~.,-Sales, data=Carseats)
```

The *summary()* is used to view the variables used as internal nodes in the tree, the number of terminal nodes and the error rate.

```
> summary(tree.carseats)
Classification tree:
tree (formula = High ~ . - Sales, data = Carseats)
Variables used in tree construction:
[1] "High."       "ShelveLoc"  "Price"   "Income"   "CompPrice"
[6] "Advertising" "Education"  "Age"     "Population"
Number of terminal nodes:  17
Residual mean deviance:  0.2782 = 106.5 / 383
Misclassification error rate: 0.0675 = 27 / 400
```

From the information above, the training error rate is 6%. one attractive prope rty of trees is that, the results can be graphically displayed using the *plot()* fun ction and the *text()* function can be used to display the nodes labels.

To properly evaluate the performance of a classification tree, we split the obse rvations into a training set and a test set, build a tree using the training set and evaluate its performance on the test set. The *predict()* function can be used for this purpose. In the case of classification tree, the argument type= "class" instr uct R to return actual class prediction.

```
> set.seed(234)
> train<-sample(1:nrow(Carseats),200)
> Carseats.test=Carseats[-train,]
> High.test<-High[-train]
> tree.carseats=tree(High~.-Sales,Carseats,subset =train)
> test.pred<-predict(tree.carseats,Carseats.test,type="class")
> table(test.pred,High.test)
```

The result of the test is shown in Table D1.

Table D1: Testing predictive accuracy of decision trees

Response	High.test (Actual)	
	Yes	No
Test prediction		
Yes	119	47
No	0	34

```
> (119+34)/200
[1] 0.765
```

This means that the classification described gives correct prediction rate of about 76.5%.

The next is to illustrate whether pruning the tree might lead to improved results. The function *cv.tree()* performs -validation in order to determine the optimal level of tree complexity; cost complexity pruning is used in order to select a sequence of trees for consideration. The appropriate argument is *FUN=prune.misclass* to indicate that we want the classification error rate to guide the cross-validation and pruning process rather than the default for the *cv.tree()* , which is deviance. The *cv.tree()* function will report the number of terminal nodes of each tree considered (size) as well as the corresponding error rate and the value of the cost-complexity parameter used. The procedure the tree that gives optimal results is described below (see $dev and $size); the tree size that gives the minimum deviance.

```
> set.seed(124)
> cv.carseats<-cv.tree(tree.carseats,FUN = prune.misclass)
> names(cv.carseats)
[1] "size"   "dev"   "k"
[4] "method"

cv.carseats
$size
[1] 12  8  6  4  3  1

$dev
```

[1] 33 33 29 27 40 41

$k

[1] -Inf 0.0 1.5 3.0 5.0 6.5

$method

[1] "misclass"

attr(,"class")

[1] "prune" "tree.sequence"

In the figure above "dev" corresponds to the cross-validation error rate in this i nstance. The tree with 4 terminal nodes results in the lowest cross-validation e rror rate with 27 cross-validation errors. We plot the error rate as a function of both *size* and *k*.

We now apply the prun.misclass() function to prun the tee to obtain the four-n ode tree.

> *prune.carseats<-prune.misclass(tree.carseats,best = 4)*

We now examine how well this pruned tree perform on the test data set.

> *tree.pred<-predict(prune.carseats,Carseats.test,type="class")*

> *table(tree.pred,High.test)*

Table D2: Testing predictive accuracy of pruned decision tree

Response	High.test (Actual)	
	Yes	No
Test prediction	119	35
Yes	0	46
No		

From the figure above now, 82.5% of the test observations were correctly clas sified. This means that pruning has improved the classification accuracy. If w e increase the value of the "best", we obtain a larger pruned tree with lower cla ssification rate or accuracy.

3.6.2 Fitting Regression Trees

In this demonstration we fit a regression tree to the Boston data set.

As required we create a training set and fit the tree to the training data with the procedure below

```
>library(MASS)
> data("Boston")
> attach(Boston)
names(Boston)
 [1] "crim"    "zn"     "indus"
 [4] "chas"    "nox"    "rm"
 [7] "age"     "dis"    "rad"
[10] "tax"     "ptratio" "black"
[13] "lstat"   "medv"
> set.seed(2)
> train<-sample(1:nrow(Boston),nrow(Boston)/2)
> tree.boston<-tree(medv~.,data=Boston,subset=train)
```

The summary of gives the following outcome

```
> summary(tree.boston)
Regression tree:
tree(formula = medv ~ ., data = Boston, subset = train)
```

```
Variables actually used in tree construction:
[1] "lstat" "rm"   "dis"  "nox"
[5] "crim"
Number of terminal nodes:  9
Residual mean deviance:  13.95 = 3404 / 244
Distribution of residuals:
     Min.   1st Qu.   Median    Mean    3rd Qu.    Max.
-16.66000 -2.16900  0.03077  0.00000  1.87700  14.64000
```

Notice that the output of summary () indicates that five out of the fourteen variables have been used in constructing the tree. In the context of a regression tree, the deviance is simply the sum of squared errors for the tree plot. We now use the *cv.tree()* function to see whether pruning the tree will improve performance.

```
> cv.boston<-cv.tree(tree.boston)
> cv.boston
$size
[1] 9 8 7 6 5 4 3 2 1
$dev
[1]  7174.098  7755.065  8050.935
[4]  8108.430  8235.968 10810.645
[7] 11476.320 11484.222 20367.622
$k
[1]      -Inf  383.6171  398.6721
[4]  408.3703  480.0938 1618.8987
```

```
[7] 1836.4258 2492.5477 9026.7308

$method

[1] "deviance"

attr(,"class")

[1] "prune"         "tree.sequence"
```

In keeping with the cross-validation results, we use the unpruned tree to make predictions on the test set.

```
> yhat<-predict(tree.boston,newdata=Boston[-train,])

> boston.test<-Boston[-train,"medv"]

> plot(yhat,boston.test)

> mean((yhat-boston.test)^2)

[1] 23.93114

> sqrt(mean((yhat-boston.test)^2))

[1] 4.891947
```

This means that the test set MSE associated with the regression tree is 23.93. the square root is about 4.8, indicating that this model leads to test predictions that are within around $4892 of the true median home value for the suburb.

3.6.3 Bagging and Random Forests

In this section we demonstrate usefulness of bagging and random forest to the Boston data. We recall that bagging is simply a special case of a random forest with $m = p$. Hence *randomForest()* function can be used to perform both random forests and bagging. We perform bagging as follows:

```
> library(randomForest)
> bag.boston<-randomForest(medv~.,data=Boston,subset=train,mtry=13,impo
rtance=TRUE)
> bag.boston

Call:
randomForest(formula = medv ~ ., data = Boston, mtry = 13, importance = TR
UE, subset = train)
Type of random forest: regression
Number of trees: 500
No. of variables tried at each split: 13

Mean of squared residuals: 17.93316
% Var explained: 77.37
```

The argument mtry=13 indicates that all 13 predictors should be considered for each split of the tree. In otherwords bagging should be done. How well does this bagged model perform on the test set

```
> yhat.bag<-predict(bag.boston,newdata=Boston[-train,])
> mean((yhat.bag-boston.test)^2)
[1] 10.78826
```

The test set MSE associated with the bagged regression tree is 10.788, less than half that obtained using an optimally pruned single tree. We could change the number of trees grown by *randomForest()* using *ntree* argument:

```
> bag.boston<-randomForest(medv~.,data=Boston,subset=train,mtry=13,ntree
=2)
> yhat.bag<-predict(bag.boston,newdata=Boston[-train,])
> mean((yhat.bag-boston.test)^2)
[1] 12.29288
```

Growing a random forest proceeds in the same way, except that we use a smaller value of the *mtry* argument. By default, randomForest() uses $p/3$ variables when building a random forest of regression trees, and $\sqrt{p}$ variables when building a random forest of classification trees. In this illustration we use *mtry=6.*

```
> rf.boston<-randomForest(medv~.,data=Boston,subset=train,mtry=6, importa
nce=TRUE)
> rf.boston


Call:
randomForest(formula = medv ~ ., data = Boston, mtry = 6, importance = TR
UE,      subset = train)
Type of random forest: regression
Number of trees: 500
No. of variables tried at each split: 6
```

Mean of squared residuals: 15.70296

% Var explained: 80.18

```
> yhat.rf<-predict(rf.boston,newdata=Boston[-train,])
> plot(yhat.bag,boston.test)
> abline(0,1)
```

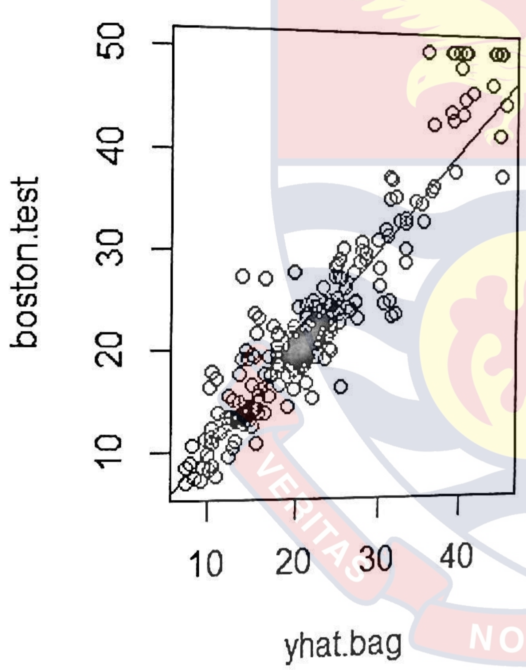


Figure D1: Predictive response versus actual random forest

```
> mean((yhat.rf-boston.test)^2)
[1] 11.28141
```

The test set MSE in this case is 11.28 which indicates that random forests yielded an improvement over bagging. Using the *importance()* function, we can view the importance of each variable.

Table D3: Variable importance representation with random forest

	%IncMSE	IncNodePurity
crim	11.821569	1250.74334
zn	1.579428	39.09829
indus	11.782228	1277.75846
chas	3.210381	129.41726
nox	10.063418	1040.68952
rm	26.666688	4588.96899
age	10.515628	683.33947
dis	13.559475	1524.73820
rad	4.093151	160.41895
tax	9.077987	501.12685
ptratio	10.048951	528.26802
black	6.839026	309.38997

In Table B3, two measures of variable important are shown. The first is based upon the mean decrease of accuracy in predictions on the out of bag samples when a given variable is excluded from the model. The second is a measure of the total decrease in the node impurity that results from splits over that variable, averaged over all trees. In the case of regression trees, the node impurity is measured by the training RSS, and for classification trees by the deviance. Plots of these importance measures can be produced using *varImpPlot()* function. From the results in the table among the trees considered in the random forest, the wealth level of the community (lstat) and the house size(rm) are by the two most important variables.

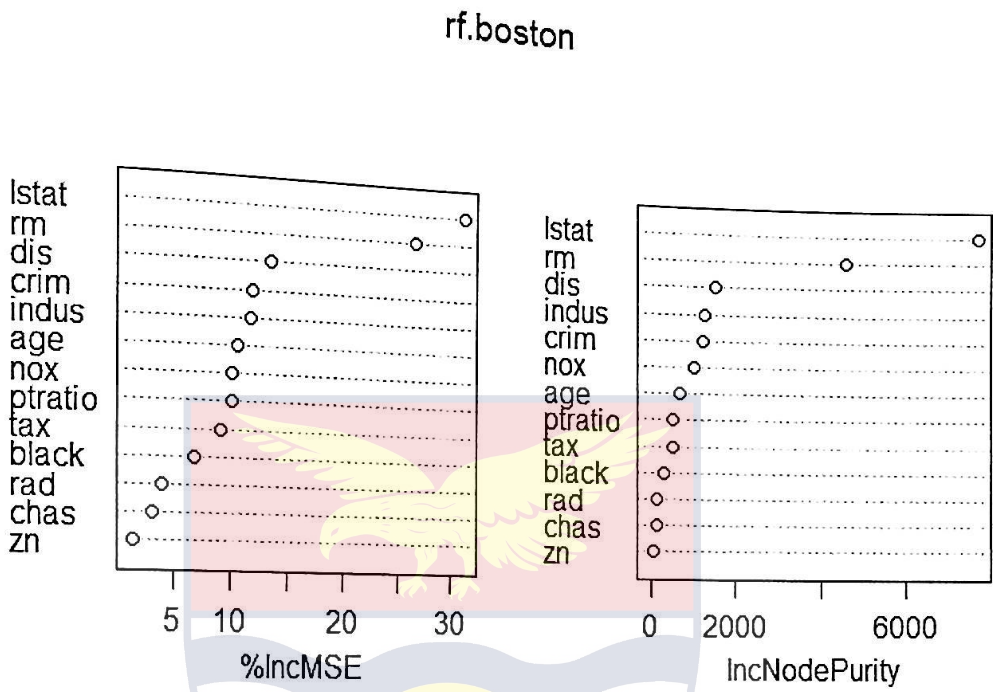


Figure D2: Graph depicting variable importance for the two methods

Boosting

In this illustration we use the *gbm* library, and the function *gbm()* to fit boosted regression trees to the Boston data set. We run the gbm function with the option of *distribution="gaussian"* if the response distribution is assumed to be normally distributed. For classification problems we would use *distribution="Bernoulli"*. The argument *n.trees=5000* indicates that we want to grow 5000 trees. The *interaction. depth=4* option limites the depth of each tree.

```
> set.seed(1)
> boost.boston<-gbm(medv~.,data=Boston[train,],distribution = "gaussian",n.t
rees = 5000,interaction.depth = 4)
```

The summary() function produces a relative influence plot and relative influence statistics.

```
> summary(boost.boston)
```

Table D4 : Relative variable importance with boosting method

variable	Relative influence
lstat	54.27
rm	20.09
dis	7.76
crim	4.48
nox	3.15
ptratio	2.78
age	1.89
black	1.77
indus	1.63
tax	1.47
rad	0.46
chas	0.14
zn	0.03

We notice that the variables *lstat* and *rm* are by far the most important variables. Partial dependence plots can also be produced for the two variables. The partial dependence plots illustrate the marginal effect of the selected variables on the response after integrating out other variables.

```
> par(mfrow=c(1,2))
>plot(boost.boston,i="rm")
```

```
> plot(boost.boston,i="lstat")
```

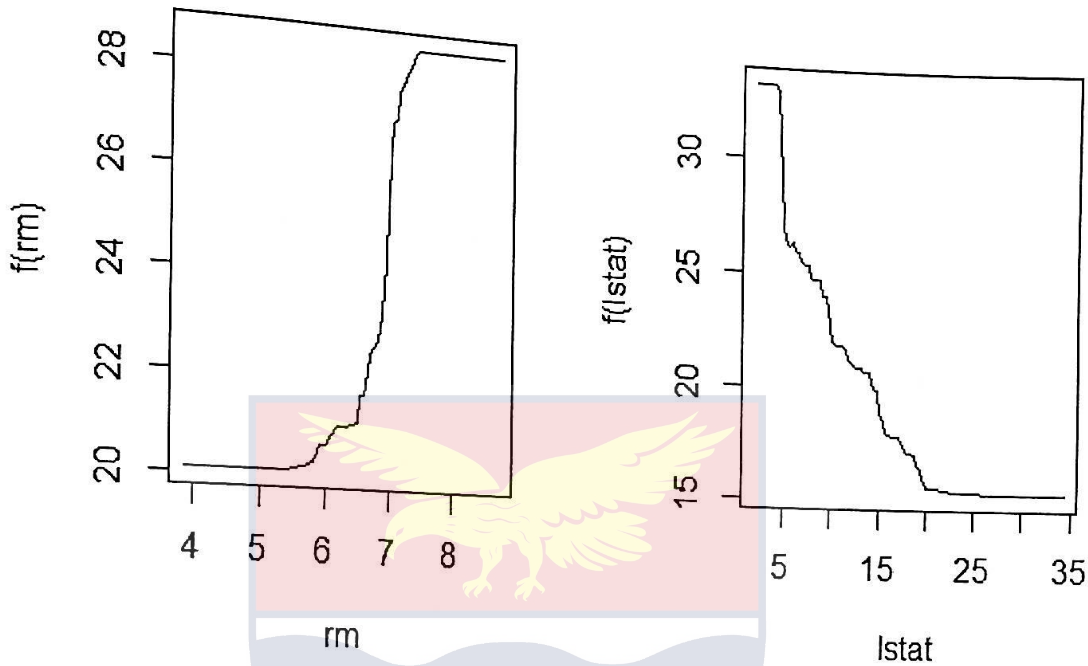


Figure D3: Variable dependence plots for "rm" and "lstat".

As expected median house prices are increasing with *rm* and decreasing with *lstat*.

we now use the boosted model to predict *medv* on the test set.

```
> yhat.boost<-predict(boost.boston,newdata=Boston[-train,],n.trees=5000)
> mean((yhat.boost-boston.test)^2)
[1] 13.24306
```

The test MSE obtained is 13.24. This is higher than that obtained from random forest. It is inferior to both bagging and random forest. This is probably due to the violation of assumption of non-normality. We can also investigate at various level of shrinkage parameter other than the default = 0.001, to see if there would be improvement.

```
> boost.boston1<-gbm(medv~.,data=Boston[train,],distribution = "gaussian",n
.trees = 5000,interaction.depth = 4,shrinkage = 0.1, verbose = F)
> yhat.boost1<-predict(boost.boston1,newdata=Boston[-train,],n.trees=5000)
> mean((yhat.boost1-boston.test)^2)
[1] 13.37469
```

It could be noticed that the test MSE has gone with increase in the shrinkage parameter=0.1 an indication of no improvement. This means the best shrinkage parameter $\lambda = 0.01$, yielded worse result than the result produced by bagging and random forest. The response distribution may have been wrongfully specified. This is an indication that, even though gradient boosting, versatile in terms of model building and interpretability, wrongful specification of the response distribution could bias the outcome. In most of the literature on lost cost modeling, using gbm, the loss distribuition has been assumed to be normaly distributed which is rare in practice (details found in chapter five).