University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Actuarial Science

Optimal Pricing of Medical Insurance for Formal Employees using Empirical Bayes Credibility Model

Research Report in Actuarial Science, Number 37, 2019

Oloo Collins Singei
I56/88515/2016

July 2019

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Actuarial Science

Optimal Pricing of Medical Insurance for Formal Employees using Empirical Bayes Credibility Model

Research Report in Actuarial Science, Number 37, 2019

Oloo Collins Singei
I56/88515/2016

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis
Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Actuarial Science

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

Through the premium, insurance cover is made possible . The study purposed to estimate optimal price of premiums paid to an insurance company to cover for medical expenditure based on historical expenditure of an organization/company. Specifically, this was to be achieved by first, coming up with a credible risk premium values of institutional claim experience to be used as an average premium regulator, second, by projecting future financial cost of providing medical cover due to rising costs brought about by inflation and technological advancement, third, modelling group expenditure claim cost using distribution based techniques to ascertain estimates of cost of expenditure. To derive credibility premium, Buhlmann credibility and Buhlmann Straub credibility models were adopted. Five contracts which represent the hospitals were simulated for a five year period each with it claim size, weight and ratios. Projections for the credibility premiums is done by computation of between and within portfolio variances with the view of finding credibility premiums by linear estimation from the credibility formula. Unbiased estimators for the mean and variance functions for both the Buhlmann and Buhlmann-Straub procedures are obtained.Results reveals that Buhlmann Straub procedure yields higher premium amounts for all the contracts. Credibility factors with the Buhlmann procedure were constant while Straub credibility premium varied with associated weights. It was discovered from the two models suffered that outliers who had hamper correctness of mean and variances which should be addressed. A longer data period is strongly recommended clean of outliers in further studies.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

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

Signature Date

OLOO COLLON SINGEI
Reg No. I56/88515/2016

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

Signature Date

Prof. Philip Ngare
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: pngare@uonbi.ac.ke

Dedication

This research project report is dedicated to my family for their material; financial and moral support throughout the course of my studies at the University of Nairobi. I sincerely offer my thanks to all of them and May the good Lord bless them for their unwavering support and generosity.

Contents

Acknowledgments

My sincere gratitude goes to my supervisor Prof. Philip Ngare for his support in moderating this research project report from its seminal phase up to completion. I am indebted to the University of Nairobi, for providing a conducive, ambiance environment for the duration which I have been at the university undertaking my studies. I am also thankful to my lecturers and other university staff who have contributed towards my academic and intellectual formation.

I want to thank my classmates of 2016 for engaging me in a fruitful academic discourse through class discussions and group assignments. Your team spirit is worthy, emulating and will sustain me as a prospective project manager.

I am grateful to my family: my wife Marian Obote, my dad Samuel Singei, my mum Linet Akinyi Singei, my brothers Bonnie, Jaap and Villa, dear sister Shirley friend Geoffrey who helped me reach this far.

Oloo Collon Singei

Nairobi, 2019.

1 Introduction

1.1 Background Information

Insurance is a financial service that involves the transfer of risk from one party, referred to as the insured, to another party, the insurer. The price paid by the insured for the security provided is referred to as the premium which may be a single payment or a series of payments. In return for the premiums, the insurer compensates the insured following the occurrence of the insured event. However, no payment is made if the insured event does not occur. Through selling insurance policies to a large number of policyholders, the insurer spreads the risk to a large number of exposed individuals which by law of large numbers in statistics ensures that actual results converge to the expected value. The law allows the use of the expected value principle in determining the premiums to charge for a certain level of coverage. The insurer will determine how much the policyholder will have to pay as premiums. These premiums are paid mostly in advance (that is paying before using) and in different periods like monthly, yearly and even on daily basis depending on what the insured prefer or comfortable with.

Studies have shown that individual policies are 30% more costly than group insurance. Kenya's vision 2030 and its constitution advocates for universal health care for its population, irrespective of status of employment of its citizens. The main objective of this is to reduce the cost of medical that is placing lots of financial challenges to its people that eventually ends up consuming a large portion of the employees' income. An insurance premium is an amount of money a person pays to an insurance company for an insurance policy. This payment could be regarded as transferring some or all of the risk of loss or damage. The cost of an insurance premium needs to take into account the expected number of claims and the expected average claim size.

A number of factors must be considered in determining premium calculation, they include:

- Occupation- some occupations may be considered more risky and dangerous. This calls for rated premiums above the normal premiums
- Medical history of an individual
- Individual's pre-existing condition, an existing or previous illness that the insurer believes is likely to worsen or recur – there may be some insurers that will not cover this risk.

We have emerging trends in the market with respect to medical insurance coverage. Short term rules indicate new requirements for benefits, such as providing coverage for adult children through age 26. There are also new minimum standards for loss ratios, and rebates that must be paid if those standards are not met. By 2030 the changes will become more significant due to rising cost of medical insurance and numerous compensations. Many insurance providers will be willing to provide cover with the most effective and efficient services.

Many government and private sectors in Kenya provide group medical insurance cover for its employees. This is an incentive for its worker population that enjoys eighty percent of free cover and only twenty percent out-of-pocket contribution toward offsetting medical cost on drugs whenever a member falls sick. This cost-sharing attitude relieves the employer of extraneous budgetary allocation towards medical cover.

Pricing of random claims has ever been one of the core subjects in both actuarial and financial mathematics and there exist various approaches for calculating lower premiums. The actuarial way of pricing usually considers the classical premium calculation principles that consist of net premium and safety loading: Thus apart from the cost of paying benefits, there must be stipulated some loading factor usually a percentage of the discretized monthly/annual premium.

1.2 Statement of the Problem

With group health insurance, the focus is on the aggregate cost of the group. Except for the very smallest firms, health plans focus on the historical claim levels for the group, rather than on the health of specific employees. Institutions allocate a large portion of their income in paying medical expenditure for its employees. Many insurers normally make payments on behalf of institutions that in turn are responsible for payment of premiums. The question that we need to answer here is "how much of a premium is to be paid projected from the past expenditure history on medical cover?" This is an ingredient to selection of insurance providers by concept of optimal pricing. An organization/company seeks to find an insurance medical provider with a minimal premium cost but effective and satisfactory service provision to its members. This would largely lead to saving/investing of large amounts of money that would otherwise been used to cover medical costs. For all customers, the need for being charged best premium is important. This study makes use of credibility theory in pooling of risk with the assumption that premium pricing is independent of age, sex and health status-homogeneity of population.

1.3 Objectives

1.3.1 General Objective

The general objective of the study is to find the optimal price of premiums paid to an insurance company to cover for medical expenditure based on historical expenditure of an organization/company

1.3.2 Specific Objectives

1. To come up with a credible risk premium values of institutional claim experience to be used as an average premium regulator.
2. To project future financial cost of providing medical cover due to rising costs brought about by inflation and technological advancement.
3. To model group expenditure claim cost using distribution based techniques to ascertain estimates of cost of expenditure.

1.4 Justification

The cost of healthcare in Kenya is becoming increasingly high year in year out due to several factors affecting the economy. These include the cost of inflation and technological advancement. Organizations/companies hire the services of insurance companies in providing medical cover to its employees. A large number of insurance providers would want to take advantage of the lack of enough knowledge that determines the correct premiums. This has led to dishonesty insurance companies taking advantage of clients by charging high cost of insurance. By doing that, it allows them make super normal profits. The aim of this study is to provide an optimal premium paying function obtained by experience rating- premium which is based on the group's own experience for past years claims are projected forward and used as the basis for this year's premiums. The obtained limits of premium payments would aid the employer in determining the 'fair' price that is charged by the health insurer.

The study will provide substantial knowledge to actuarists in determining gross premiums affected by fluctuation in incidental economic factors, an extension of credibility theory.

1.5 Limitation

There were limited information/details of the data. The data used was of a short period of time which is not really good since for more accurate credibility premium, a large data is required.

2 Literature Review

2.1 Introduction

The insurance history is dated back to the nineteen century during the famous era of London's fire. In the aftermath, London's residents made contributions to a group account in form of savings. Some insurance companies directly paid for some of Londons' city fire brigades. People who paid insurance companies to insure their homes were given a "fire plate" showing the insurance company's logo. This fire plate was fixed near the front door of their house. If a house caught fire, the fire brigade would check if the house had a fire plate. If the house did not have a fire plate, or had the fire plate of another insurance company, the fire brigade would let the house burn down. People who wanted the fire brigade to help them if their home caught fire, would each put in a little money to help pay for the fire brigade to protect their house. The problem that was of concern was; residents were of different classes of wealth and the kind of lifestyle one lived was dependent upon the class (the population was heterogeneous in nature). The size of wealth one owned was evidenced by the kind of housing. The main issue now was "How much one is expected to pay monthly that is proportional to the value of the property". Since this era, actuaries have developed much theory to determine the premium amounts per period of insurance. Much of the theories developed include credibility theory-measure of predictive value attached to a particular class of data based on experience rating-and premium rating-process of determining premium estimates of expected values of future costs per unit time of exposure for group of risks.

2.2 Credibility Theory

Credibility theory is a set of techniques of calculating insurance premiums for short-term nonlife insurance contracts. This technique makes use of:

1. Historical data related to the actual risk.
2. Data from other related but relevant sources commonly referred to as collateral data.

The credibility premium as derived by Waters (1987), in the special note, is of the form shown in Equation 2.2.1

$$\hat{M} = z\bar{x} + (1-x)\mu \tag{2.2.1}$$

Where , $\hat{M}$ - the premium
z - the weight or credibility factor usually between zero and 1. The credibility factor here

is an increasing function for large values of n
$\bar{x}$ -the observed mean claim amounts per unit risk exposed for individual contract/risk itself. μ - the parametric estimate of the proposed data in the case than an assumption of the underlying distribution is made. For a series of risks
x -the corresponding portfolio (set of risks) mean.

The following are features of the credibility formula:

- It is a linear combination of estimates to a pure premium policy based on observed data from the risk itself and the other based on projected risks
- The credibility factor Z shows the degree of reliability of the observed risk data in the sense that high values of Z implies high reliability
- The credibility factor is a dependent function of the number of claims. This implies that the higher the claim numbers, the larger the credibility factor
- The value of Z is between zero and one, i.e; $0 \leq Z \leq 1$

2.3 Credibility Theory Development

Credibility theory was originally developed for a long time by actuarists from North America in the early 20th century. Mowbray (1914) put it into practical solution to premium calculation and it came to be called the American credibility theory. It is sometimes referred to as "limited credibility theory" or "the Fixed effect credibility". In this work it was assumed that the annual claims $X_1, X_2, X_3, \ldots, X_n$ are independently and identically distributed random variables from a probabilistic

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

Later on, Nelder and Verall derived credibility functions by the generalized linear model approach and consequently included the random effects model. This has provided a wide range of actuarial application among them is premium rating and reserving. Though a lot of research was done that yield several findings, it was found that the fixed effect credibility was not able to solve the problem of credibility. It is said that part of it was due to undeveloped or poor statistical background. In 1967 and 1970, the real thing came when Bulhmann derived the credibility premium formula in a distribution free-way such that there was no assumption of prior distribution of claims.

Bulhmann later clarified in this work the several assumptions of using the credibility premium formula (see Bulhmann 1971). This major breakthrough has seen much of the research tilting to the development of Bayesian estimation techniques by Jewell (1974, 1975), Hachmeister (1975),Devylder (1976, 1986) and Gooverts and Hoogstad (1987). Jewell (1974) showed that for exponential family distribution, the best linear approximation to Bayesian estimate is obtained using quadratic loss functions. Hachmeister (1975) extended the Bulhmann Straub model by use of matrix method.

2.4 Demand for medical insurance

Gius (2010) argued that health insurance coverage among the young people in America is mostly influenced by socio-demographic characteristics (for example age, gender, health status, religion, and locality, level of education, race, income and price of related commodities). Moreover, though the premium cost had a significant influence on health policy coverage, individuals who believed that they were healthy were relactant to taking the health since they assumed they had low chance of benefiting from the policy.

Past studies such as (Chankov et al, 2008; Ito and Kono, 2010) showed that there is a positive significant relationship between health insurance demand and age. Schneider and Diop (2001) showed a positive relation between gender, number of dependants in a household and the uptake of health insurance policy. According to Chankov et al, (2008) there is a positive significant relationship between both occupation and wealth status of an individual and purchase of a health insurance cover. Schneider and Diop (2001) argued that a health policy seeker understanding influences purchase of medical cover positively.

According to Gin et al,(2007) there is a positive significant relationship between trust as measured by other household known to have purchased the policy, previous group membership in which some members have taken the insurance policy, credibility of the claim payment as well as individual insurance seekers perception) and purchase of agricultural micro insurance.

Ito and Kono (2010) showed a positive significant between purchase of health insurance and previous experience as measured by death experience, illness experience in the family as well as the health status of an individual and family member.

Huber (2012) applied probit analysis to determine micro insurance determinants in Indonesia; the study findings showed an inverse relationship with life cycle, positive significant relationship with occupation status, an inverse relationship in a household with multiple earner status, asset endowment and purchase of micro insurance. Moreover, the findings showed positive insignificant relationship between product level of literacy as well as product knowledge with demand for micro insurance. The level of trust as measured by trust degree, participation in non-formal group as well as membership of microfinance had positive but insignificant relationship with demand for micro insurance while both brand recognition and client experience had a positive significant relationship with demand for micro insurance.

3 Methodology

The credibility premium is a linear function of the form shown in

$$\hat{M} = z\bar{x} + (1-x)\mu$$

For this case, z is the amount of credibility that is assigned to a certain data set originating from past experience data. The main problem of actuarists is how much information/observations are required for one to attain 100% credibility. This leads us to determining conditions necessary to attain full credibility and partial credibility. In most practical situations, full credibility is a rare phenomenon.

Mowbray (1914) deduced a criterion for determining the sample size required for partial credibility. This approach came under so much criticism due to its fixed effects. This led to the adoption of Whitney's random effects model that mainly focused on the estimation of the credibility function. This opened a wide area of research where experience rating problems were seen to be a matter of estimating the random variables $\hat{M}$ from observed mean of information, μ , of the individual data sets. The main aim was to minimizing the Mean square error.

$$\rho(m) = E[m(\theta) - mx] \tag{3.0.1}$$

The optimal estimator,μ , is obtained by conditional approach $\hat{M} = E(m/x)$. The most important computational functions include:
E(X) = E(E(X|Y))
Var (X) = E[Var(X|Y)] + Var [E(X|Y)]
Thus the MSE is thus obtained as follows:

$$\begin{aligned} \rho(\hat{M} &= E[Var(m(\varphi)|Y] \\ &= Var(m) - Var(\hat{M} \\ &= E[m(\varphi - E(m|X)]^2 + E[E(m|X) - m(X)]^2 \\ &= E[m(\varphi) - \hat{M}(X)]^2 \end{aligned} \tag{3.0.2}$$

The above derivation (3.0.2) of the MSE mostly gave restrictions on distribution functions. This form of MSE was then modified to avoid much restriction on distribution function which eventually gave rise to a linear credibility function of the form:

$$m(X) = a + b\hat{m}(X)$$

The linear Bayes estimator of the form shown in equation 3.0.3:

$$\bar{m} = E(m(\varphi)) + \frac{Cov[m,\hat{m}]}{var(\hat{m})}(\hat{m} - E(\hat{m})) \tag{3.0.3}$$

The Linear Bayes risk is thus given by the function

$$\bar{\rho} = var(\hat{m}) - \frac{Cov^2[m,\hat{m}]}{var(\hat{m})} \tag{3.0.4}$$

The linear Bayes risk approaches zero with increasing amounts of data. The sufficient conditions that must hold include:

$$E(\hat{m} - E(\hat{m}))^2 \to 0 \tag{3.0.5}$$

$$E[Var(\frac{\hat{m}}{\varphi})] \to 0 \tag{3.0.6}$$

$$E[(\frac{\hat{m}}{\varphi})] = m(\varphi) \tag{3.0.7}$$

For these conditions in place
$E(\hat{m}) = E(m)$;
$Cov[m,\hat{m}] = Var(m)$;
$Var(\hat{m}) = Var(m) + E[Var(\frac{m(\varphi)}{Y})]$

The credibility function Z is thus given as

$$Z = \frac{Var(m(\varphi))}{Var(m) + [m(\varphi)|Y)]} \tag{3.0.8}$$

Various models have been suggested for calculation for the credibility premiums in the vast literature of empirical Bayes credibility. The model assumptions are that the aggregate claims are independent and identically distributed in nature. In most life situations, this is not normally the case since to analyze for risk, we need different variables that are not necessarily dependent on one another. We relax this assumption of independence and we assert that the aggregate claims are not necessarily identically distributed.

3.1 Empirical Bayes Credibility

Denote a given random data set of aggregate claims for successive years for a particular class of risk, say $Y_1, Y_2, \ldots Y_n$ for successive years. Let $P_1, P_2, \ldots, P_n$ be a corresponding sequence of known constants, in this case the number of policies issued in a year. Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables such that:

$$X_j = \frac{Y_j}{P_j}$$

The assumption on the distribution of the random variable X_j is dependent on a fixed parameter ϕ and is denoted by $U(\phi)$

3.2 Derivation of the Credibility Premium

3.2.1 Buhlmann's Credibility

By Bulhamann's approach, the credibility premium is a linear function of observed values X_j which gives the best approximation to $E[m(\phi)|X]$
The observed values are linear and are of the form

$$a + \sum_{j=1}^{n} a_j X_j \qquad\qquad j = 1, 2, \ldots, n$$

We seek to find the constant a_j that minimize the mean square error

$$E\{E[m(\phi)|X] - [a + \sum_{j=1}^{n} a_j X_j\}^2$$

We solve this by differentiating the function above with respect to a_j S.

This leads to a new set of equations that can be solved iteratively:

$$E[m(\phi) - a_0 - \sum_{j=1}^{n} a_j X_j = 0 \tag{3.2.1}$$

$$E[X_k m(\phi) - a_0 X_k - \sum_{j=1}^{n} a_j X_j X_K = 0 \qquad\qquad K = 1, 2, \ldots n \tag{3.2.2}$$

Equation 3.2.1 and 3.2.2 are reducible to the forms:

$$a_0 = \{\sum_{j=1}^{n} a_j\}E(m(\phi)) \tag{3.2.3}$$

$$E[m^2(\phi) - a_0E(m(\phi)) - a_kE[\frac{s^2(\phi)}{P_j} - \sum_{j=1}^{n} a_jE(m^2(\phi))] = 0 \tag{3.2.4}$$

Substituting Equation 3.2.3 into 3.2.4, we obtain Equation 3.2.5

$$P_K Var[m(\phi)]\{1 - \sum_{j=1}^{n} a_j\} = a_k[S^2(\phi)] \tag{3.2.5}$$

Using these equation a_0 and a_k can be obtained as Equation 3.2.6 and 3.2.8

$$\sum_{j=1}^{n} a_j = \frac{\sum_{j=1}^{n} P_j}{\{\sum_{j=1}^{n} P_j + \frac{E[S^2(\phi)]}{Var[m(\phi)]}\}} \tag{3.2.6}$$

$$a_j = \frac{P_k}{\{\sum_{j=1}^{n} P_j + \frac{E[S^2(\phi)]}{Var[m(\phi)]}\}} \tag{3.2.7}$$

We thus substitute Equations 3.2.6 and 3.2.8 in their linear form of credibility premium to obtain the pure premium per unit volume of risk. This equation is of the form;

$$E(m(\phi))[\frac{E[S^2(\phi)]}{Var[m(\phi)]} + \frac{\sum_{j=1}^{n} Y_j}{\sum_{j=1}^{n} P_j + \frac{E[S^2(\phi)]}{Var[m(\phi)]}} \tag{3.2.8}$$

Taking

$$\bar{X} = \frac{\sum_{j=1}^{n} P_j X_j}{\sum_{j=1}^{n} P_j}$$

$$Z = \frac{\sum_{j=1}^{n} P_j}{\sum_{j=1}^{n} P_j + \frac{E[S^2(\phi)]}{Var[m(\phi)]}}$$

The linear credibility can be written as in equation 2.2.1 as:

$$\hat{m} = z\bar{x} + (1 - z)\mu$$

3.2.2 Parameter Estimation

In this section, we estimate parameters contained in the credibility premium formula from a suitable data set. The parameter estimates are proved to be unbiased estimators.

Suppose we define a single risk from a class of N risks. Let $Y_{i1}, Y_{i2}, \ldots, Y_{in}$ denote the aggregate claims in successive years from the risk with $P_{i1}, P_{i2}, \ldots, P_{in}$ being the corresponding risk volumes of known constant value. Define $X_{ij} = \frac{Y_{ij}}{P_{ij}}$

The assumptions under this model are that:

- $X_{i1}/\phi_i, X_{i2}/\phi_i, \ldots, X_{in}/\phi_i$ are independent but not necessarily identically distributed
- $\phi_1, \phi_2, \ldots, \phi_n$ are independent and identically distributed.

We possess the same number of observed risks. Assuming
$M(\phi_i) = E[\frac{X_{ij}]}{phi_j}$ and $S^2(\phi_j)|P_{ij} = Var[X_{ij}|\phi_j]$ Because of the assumption of identical distribution of ϕ_j, the distribution of $m(\phi_j)$ and $S^2(\phi)$ are the same for all i's. We adopt the following notations:
$\bar{P}_i = \sum_{j=1}^{n} P_{ij}$

$\bar{P} = \sum_{j=1}^{N} \bar{P}_i$

$P^* = [\sum_{j=1}^{N} \frac{\bar{P}_i(1-\bar{P}_i|\bar{P})}{Nn-1}]$

$\bar{X}_j = \sum_{j=1}^{n} \frac{P_{ij}X_{ij}}{\bar{P}}$

$\bar{X} = \sum_{j=1}^{n} \frac{P_{ij}X_{ij}}{\bar{P}} = \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{P_{ij}X_{ij}}{\bar{P}}$
Note that the unbiased estimators were of the form:

$$E[S^2(\phi)] = \sum_{j=1}^{n} \frac{P_i j(X_{ij} - \bar{X})^2}{N \sum_{i=1}^{n}(n-1)} \tag{3.2.9}$$

$$Var[m(\phi)] = P^{*-1}\{[\sum_{i=1}^{N} \sum_{j=1}^{n} \frac{P_i j(X_{ij} - \bar{X})^2}{Nn-1)}] - [\sum_{j=1}^{n} \frac{P_{ij}(X_{ij} - \bar{X})^2}{N \sum_{i=1}^{n}(n-1)}]\} \tag{3.2.10}$$

3.2.3 Bulhmann-Straub Model

This is a generalization of the classical credibility premium of Bulhmann(1969) and has been used to rate reinsurance treaties where much of this has been applied in auto and reinsurance sectors.

This consists of a portfolio of N insured each characterized by an unobservable random risk parameter ϕ_i and let X_{it} be a list of available observations such as the average claim amount or claim loss ratio for $t = 1, 2, \ldots, Ti$and $i = 1, 2, \ldots, N$. The number of periods of experience depends on the insured. To each X_{it} , a weight W_{it} is assigned. The weights can be valid measures such as no of claims in one year or the premium volumes.

3.2.4 Model Assumptions

- The insured's $X_{i1}, X_{i2}, \ldots X_{iTi}, \phi_i$ vectors are mutually independent
- The risk parameters ϕ_i are independent and identically distributed
- The variables X_{it} have finite variance, for $i = 1, 2, \ldots, N$

$$E(X_{it}|(\phi) = \mu(\phi) \tag{3.2.11}$$

$$Cov(X_{it}, X_{it}/(\phi_i) = \frac{\sigma^2(\phi_i)}{W_{it}} \tag{3.2.12}$$

Equation 3.2.12 reflects the non-correlation between the insured's claim experience across the years and the homogeneity in time.

3.2.5 Parameter Estimation of the B-S Credibility Premium

The structural parameters are as follows.

$$m = E(X_{it}|\phi_i) = \mu(\phi_i$$

$$S^2 = E(\sigma^2(\phi_i))$$

$$a = Var(\mu(\phi_i))$$

$$W_{i\bullet} = \sum_{t=1}^{T_i} W_{it}$$

$$W_{\bullet\bullet} = \sum_{i=1}^{I} \sum_{t=1}^{T_i} W_{it}$$

$$X_{i\bullet} = \sum_{t=1}^{T_i} (\frac{W_{it}}{W_{i\bullet}}) X_{it}$$

$$X_{\bullet\bullet} = \sum_{i=1}^{I} \sum_{t=1}^{T_i} (\frac{W_{it}}{W_{\bullet\bullet}}) X_{it}$$

$$Z_{\bullet} = \sum_{i=1}^{I} Z_i$$

$$Z^z_{\bullet\bullet} = \sum_{i=1}^{I} \frac{Z_i}{Z_0} \sum_{t=1}^{T_i} (\frac{W_{it}}{W_{\bullet\bullet}}) X_{it}$$

The credibility premium is found by minimizing the mean square error. This is estimated as:

$P_i = Z_i X_{i\bullet} + (1 - Z_i)m$ *where* $Z_i = \frac{W_{i\bullet}}{W_{i\bullet} + K}$ *and* $K = \frac{S^2}{a}$

S^2 is a measure of the stability of portfolio claim. The lower the value, the larger the credibility factor. An increase in this leads to an increase in the credibility factor, Gowell, (1998). The estimates of the structural parameters are:

$m = Z^z_{\bullet\bullet}$ which is the pseudo estimator, a function of unknown parameters S^2 and a

$S^2 = \frac{1}{N-1} \sum_{i=1}^{1} \sum_{t=1}^{T_i} W_{it}(X_{it} - X_i$; the unbiased estimator of S^2

$\hat{a} = \frac{W_{\bullet\bullet}}{W^2_{\bullet\bullet} - \sum_{i=1}^{I} W^2_{i\bullet}} [\sum_{i=1}^{I} W_{i\bullet}(X_{i\bullet} - X_{\bullet\bullet})^2 - (I-1)S^2]$, the estimator obtained by ANOVA which is sometimes negative.

$\hat{a} = \frac{1}{I-1} \sum_{i=1}^{I} Z_i (X_{i\bullet} - X^Z_{\bullet\bullet})^2$, the Bichsel-Straub Estimator which is always positive

3.3 Model Application and Methodology

A family of distributions for the number of claims N can be generated by assuming that the Poisson parameter is random variable with pdf $f(\lambda)$ with >0 . The conditional distribution of N is also a Poisson with parameter λ. When the variance of the number of claims exceeds its mean, the Poisson distribution is not appropriate. Rather, a negative binomial distribution is used (Bowers et al., 1997).

For a collective risk model, assuming a random process that generates claims for a portfolio of policies. Each of the claim amounts Xi, then $S = X_1 + X_2 + \ldots + X_N$ represents the aggregate claims for the portfolio for the period under study. The random variables $X_1, X_2, \ldots, X_N$ also measure the severity of the claims. For this reason of stability, a simulation of claim numbers from a Poisson distribution where the mean and variance components are equal need to be done. i.e. $E(\mu(\phi_1)) = Var(\phi_1) = \phi_i$

For many insurance claims, the claim amount random variable is only positive and its distribution is usually skewed to the right. These properties resemble the properties to the gamma distribution. In this study we adopt the above two essential properties to perform simulation of claim amounts. The distribution of claim amounts may not be of a simple form, but the convolution of claim amounts may yield a compound Poisson distribution. We may opt to choose a discrete claim distribution and calculate the required convolutions numerically. This may be a new line of investigation for credibility premium calculations.

3.4 Data Simulation

In this study, five contracts are simulated depicting five different insureds/contracts, whereby in this case, the contracts are the Hospitals. These contracts (hospitals) each involves observed data for a period of five years with each claim size and the respective number of claims in that year being obtained.

The simulation procedure first begins with the generation of weights, W_{it} from a Uniform distribution such that on (a,b), $0 \leq a < b$. In this case, simulation is done for 10 variables from uniform distribution in the range (500,1000) by the function, unif (10,(500, 1000)).These weights may be the total number of claims in the respective year of interest or any chosen function of the claim amount, say the square root of the total claim amount in that year.

This is then followed by generation of the risk levels from a gamma distribution function using the R function, rgamma (3,2). The risk levels are also functions of weights generated as above.

The aggregate claims N_{it} are for different contracts and/or insureds is obtained from a *Poisson distribution* using the function rpois (weights * contracts), where each claim is generated from a gamma distribution of parameters α and β such that each claim amount is also gamma distributed.

The total amounts made for claims, S_{it} is the sum of all N_{it} for all insureds or contracts. Finally, we obtain the claim ratios from dividing the total claim amount S_{it} by the weights obtained above.

This is represented as

$$X_{it} = \frac{S_{it}}{W_{it}}$$

In order to fit the credibility model to the above data as obtained from simulation, the procedure requires that a package *actuar* be loaded such that the linear modeling parameters be installed. We thus extract the data and process it in a simplified form as shown in the appendix. In the analysis, we attempt to make projections of credibility predictions based on both the Buhlman's credibility and the Buhlman Straub credibility approaches. We tabulate results for the weights that are obtained, the various variance Where possible, we attempt to provide graphical comparison for the same procedures above.

3.5 Analysis

In this study, I analyze tabulated information that gives projections for the credibility premiums. This involves computation of between and within portfolio variances with the view of finding credibility premiums by linear estimation from the credibility formula. In the process, individual means are thus tabulated with respect to each of the above procedures. The results are obtained by simulation procedure using the R.13.0 package.

I then find the unbiased estimators for the mean and variance functions for both the Buhlmann and Buhlmann-Straub procedures.

4 Data Analysis and Results

4.1 Introduction

This study makes use of simulated data and the summary statistics is as shown in Tables 4.1 and ??. Tables generated by R script have have been modified. The result in Table 4.1 and ?? and discussion thereafter are similar for all types of scenarios that can be generated.

4.2 Aggregate Claim Amount

Table 4.1. Aggregate Claim Amount for Five Years

Contract	Year 1	Year 2	Year 3	Year 4	Year 5	Totals
1	146725.5	265668.3	777109.1	964314.3	407412.7	2771198.4
2	1138941.4	797176.9	454067.8	603411.5	511138.0	4054773.6
3	1186919.9	516176.6	545918.6	435903.4	5896443.5	5331438.09
4	549769.0	892476.2	549556.8	2987047.2	868446.8	5417457.6
5	944445.9	7662378.6	87387643.6	9854284.9	1069117.0	6555250.5

Table 4.2. Ratios

ratio.1	ratio.2	ratio.3	ratio.4	ratio.5	sum(ratios)
498	774	1224	1212	873	4581
1093	912	732	721	680	4138
1204	756	918	749	1522	5249
983	1336	1176	2100	1217	6812
1302	2009	1998	1461	1491	8261

Table 4.3. Weights

weight.1	weight.2	weight.3	weight.4	weight.5	sum(weights)
277.096	414.694	684.730	742.200	448.351	2567.070
1316.470	1126.29	866.212	864.640	781.148	4954.759
925.607	614.648	638.256	555.314	1169.209	3903.034
528.786	697.950	645.899	1136.701	672.597	3681.933
642.13	921.732	924.778	639.747	721.806	3850.193

Now, Table ?? shows the fitting the Buhlman's credibility yields after predicting for the sixth year claim experience: The structural parameters being the component. Between contract Variance/covariance: 108981.8
Within contract variance: 118167.5
Collective premium: 1219.12

Table 4.4. Buhlmann's Results

Contract	$\bar{X}_{i\cdot}$	Cred Fac ($Z_{i\cdot}$)	Cred premium
1	1041.4	0.82179	1073.072
2	827.6	0.82179	897.373
3	1089.8	0.82179	1112.846
4	1362.4	0.82179	1336.866
5	1774.4	0.82179	1675.443

Fitting the Buhlman-Straub credibility for the sixth year claim experience yields the results shown on Table ?? linear prediction for the five contracts: Summary Collective premium: 1297.027
Between contract variance/covariance: 109431.8
Within contract variance: 91987995

Table 4.5. Buhlman-Straub Results

Contract	individual.mean	Weight	Cred.Factor	Cred.premium
1	1157.425	2567.070	0.753322	1191.862
2	858.716	4954.759	0.8549534	922.292
3	1186.625	3903.034	0.8227947	1206.189
4	1471.362	3681.933	0.8141313	1438.959
5	1819.456	3850.193	0.8207985	1725.837

Results show that the Buhlmann Straub procedure yields higher premium amounts. For all contracts, the individual premiums are higher than in the case of Buhlmann procedure. This may be due to weighting of claim amounts thus reducing variance components. The individual contract means above were obtained by,

$$\bar{X} = \sum_{i=1}^{5} \frac{(sumoftheratio)}{5}$$

This gives rise to the group mean

$$\bar{X} = \sum_{i=1}^{5} \bar{X}_{i\bullet}$$

The unbiased estimator for the contract variances is the sample variance, 109431.8

It can readily be observe that the credibility factor for the Buhlmann procedure is a constant for all the hospitals/contracts. In the case of the Buhlmann Straub procedure, the credibility premium varies with the associated weights.

Table **??** shows the numerical summary of the structural parameters for the Buhlmann straub procedure for credibility theory.

Table 4.6. Summary Data-Buhllmann Parameters

$W_{i\bullet}$	$X_{i\bullet}$	$S_{i\bullet}$	$\bar{X}_{i\bullet}$	$Z_{\bullet}$	Cred.Premium
2567.070	5207	2971189.9	1157.425	0.753	1191.862
4954.759	4138	4254730.9	858.716	0.855	922.292
3903.034	5449	4631436	1186.625	0.823	1206.189
3681.933	6812	5417458	1471.362	0.814	1438.959
3850.193	8872	7005259	1819.456	0.821	1725.836

5 Conclusion and Recommendations

5.1 Conclusion

This study has its focus on computation of credibility premiums by Buhlmann and Buhlmann Straub credibility theory. These methods are rather linear approximation techniques as opposed to techniques that are normally parametric in nature.

In the case of health insurance claims, there are the risks levels being the outpatient and inpatient entities. The procedure looks at both sides of scenario using the data

Real data with full information or details is of high importance in determining the physical financial scenario of a company. In the medical sector of life insurance, detailed data is difficult to obtain. In general, data from many insurance companies in Kenya was difficult to obtain. The reason behind this is the oath of secrecy to hold on to information that is deemed ethically private in the medical sector. If real data with full information or details is observed, then the findings may be varied due to the different claims experience. In addition, real data claims amount is inclusive of expenses like administration and commission costs that may have been incurred. Also the amount of claim may contain so errors because some claims that may be made in one month may not be paid till the next month. The claim amount is recorded in the exiting month while the claim number is in the correct month of entry of claim.

Health and age are very important factors in determining the cost of health insurance. An individual health may occur seasonally since it is a variable of time and so claim experience may vary highly at different seasons. These fluctuations may lead to inflated premium amount. Age consideration in the premium computation is very vital for it helps in obtaining accurate credibility premium.

The Buhlmann and Buhlmann-Straub procedures is faced the problem of outliers which distort the mean and variance functions. This will in turn affect the accuracy of the credibility premium.

5.2 Recommendations

This study recommend that the data given by any insurance firm for the purpose carrying out research should contain a reasonable information in order to get good if not best results.

It further recommends that data should be smoothened of any outliers in order to increase accuracy of the credibility premiums.

5.3 Further Research

In the models under study above, the assumption of homogeneity within the cohorts has been made. In most cases, if we assume that the years' claim total is heterogeneous in nature, we need to account for heterogeneity in the model in calculation of premiums. This is because the claim experience for individuals is not the same in the different cohorts. This is normally referred to as over-dispersion. This is a more interesting field that much research can be done. This involves the estimation of the over-dispersion parameters and factoring this into the required credibility mode.

Bibliography

[A1] Bailey, A.L. (1945). A generalized theory of credibility. *Proceedings of the Casualty Actuarial Society 32, 13-20.*

[A2] Buhlmann H. (1967). Experience rating and credibility. Astin bulletin, 4, 199-207

[A3] Buhlmann H, (1969). Experience rating and credibility. *Astin bulletin*, 5, 157-165

[A4] Buhlmann, H., & Gisler, A. (2005). A Course in Credibility Theory and its Application. *Springer-Verlag, Berlin.*

[A5] Buhlmann, H. E., & Straub E. (1970). Credibility for loss ratios. schweizerischer mathematica, 70, 111-133.

[A6] Gooverts M. N., & Hoogstad P. W. (1987). *Surveys of actuarial studies, 4, Rotterdam, Holland.*

[A7] Hachemeister, C. (1975). Credibility for regression models with application to trend. In Credibility: Theory and applications (ed. P.M. Kahn), Academic Press, 129-163

[A8] Hogg, R.V & Klugmann, S .A (1984). *Loss distributions, John Willey and sons, New York.*

[A9] Jewell, W. (1974). Credible means are exact Bayesian for exponential families. Astin Bulletin, 8(1),77–90.

[A10] Jewell, W.S. (1974c). Regularity conditions for exact credibility. *Astin Bulletin 8, 336-341.*

[A11] Longley-Cook, L.H. (1962). An introduction to credibility theory. Proceedings of the Casualty Actuarial Society, 49 , 194-221.

[A12] Mowbray, A., (1914). How extensive a payroll is necessary to give dependable pure premium? *Proceedings of the Casualty Actuarial Society, 1, 24–30.*

[A13] Norberg, R. (1979). The credibility approach to experience rating. Scandinavian Actuarial journal, 181-221.

[A14] Sundt, B. (1984). an introduction to non-life insurance mathematics, *verlag versicherunswirtschaft.*

[A14] Vincent G. (1998). Principles and Application of credibility theory. Journal of actuarial practice, 6(2).

[A15] Waters, H.R (1987). An introduction to credibility theory, *Special note, vol .2*

[A16] Whitney, A. (1918). The theory of experience rating. Proceedings of the Casualty Actuarial Society, 4, 274–292.