Vol.11.Issue.3.2023 (July-Sept.)
©KY PUBLICATIONS

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com

Email:editorbomsr@gmail.com

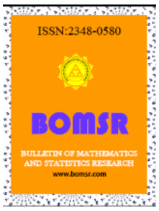
INTERNATIONAL STANDARD SERIAL NUMBER
2348-0580

RESEARCH ARTICLE

Experience Rating in Motor Insurance Industry using Generalized Linear Models

Amos Otunga[1], George Muhua[2*]
·[1]Department of Mathematics, University of Nairobi
P.O Box 30197-00100, Nairobi, Kenya
E-mail: amosotunga@mail.com
[2]Department of Mathematics, University of Nairobi
P.O Box 30197-00100, Nairobi, Kenya
*E-mail: gmuhua@uonbi.ac.ke

DOI:10.33329/bomsr.11.3.58



ISSN:2348-0580

BOMSR

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH
www.bomsr.com

ABSTRACT

Motor insurance is a necessity in most States and thus records an overwhelming number of claims in any given period. Actuaries therefore need to determine a reward structure in a manner that is fair to the policyholder and with certainty of maximum profits to the insurer. This paper aims at reviewing the methodology behind the generalized linear models used in the pricing of premiums paid in by policyholders in the motor insurance industry based on the general risk factors, policyholder and vehicle characteristics, policy type among others that an insurer may wish to include in their rating plan. Negative binomial regression is presented with comparison to the Poisson regression as rating techniques among others such as credibility, BM, multi-state and BS approaches. It is clear that motor insurance industry still experiences heterogeneity within the given risk categories such as the drinking habits, temperament or knowledge of the traffic rules of every policyholder. An accurate rating system is crucial to the actuary as it would precisely reflect the observed losses. Classifying the observed losses according to the appropriate risk factors is very substantial in determining how accurate the rating system is, in the sense that, the risk factors tell us exactly which level of which risk factor causes more claims which would result to the biggest loss and therefore to be charged the highest premium, and which causes little claims which would result to the smallest loss, to be charged the lowest premium.

Keywords: Motor insurance, GLM, Claim frequency, Over-dispersion, Count data, Risk factors, heterogeneity, homoscedasticity

1. INTRODUCTION

Kenya's Insurance industry has flourished since the conquest of Kenya as a British colony. The settlers initiated various economic activities like farming and extraction of various agricultural products. Such substantial investments needed some form of protection against various risk exposures. The Kenya Insurance Industry is governed by the Insurance Act (KIA) which was enacted in 1985. The Insurance Act of 2006 then established the Insurance Regulatory Authority (IRA), a body that ensures the effective administration, supervision, regulation and control of insurance and reinsurance business in Kenya. Section 4 of the Kenya insurance Act expressly provides that no person shall use, cause to use or permit any person to use a motor vehicle unless there is in force a policy of insurance or such a security in respect of third party risks.

General insurance under which auto insurance is classified is perhaps the fastest growing investment areas for actuaries (Boland, 2006). It is also known as the non-life insurance. The policies under this insurance would include personal insurance (such as home and automobile or car insurance), mortgage protection insurance, business insurance, travel insurance among others. Motor policy is a non-life insurance policy. Motor insurance is a necessity in most States thus records an overwhelming number of claims in any given period. It remains a very competitive industrial sector with high possibility of insured exit. Such exit may be as a result of overcharge or simply for better incentives from other insurers.

Historically, actuarial science has been limited to using the standard Gaussian linear regression in order to quantify the exogenous variables impact over the phenomenon of interest such as those of the motor insurance industry. The linear model has taken the lead in econometrics, but the applicability of this model in insurance has been found to be difficult as it implies a series of assumptions that are not compatible with the reality imposed by the claim frequency and costs of the damages generated by the risk occurrence. Considering this, the most important assumptions of such linear regression models are the Gaussian probability density, the linearity of the predictor and homoscedasticity. The count regression analysis is seen to allow for the identification of the risk factors and the prediction of the expected frequency of claims given the characteristics of the policyholders. The other technique evident in the auto insurance industry is the Bonus-malus system, known to be one of the simplest rating techniques used in most countries. The method explores certain principles and methods for adjusting insurance premium as claim experience is obtained. The future premiums are adjusted according to the insured claim history. Black and Scholes model also known as the BS model has now gained application in the pricing of auto insurance policies. The model, however, has some restrictive limitations that limit its use in the insurance industry. The multi-state models also present its application in the motor insurance industry. The drivers are rated into various states such as Preferred, standard and sub-standard with the possibilities of moving back and forth among the states by a constant force of transition at any given time due to certain factors that may contribute to the same. The credibility approach to auto rating studies the automobile insurance claim counts past data. Observation of the claim events of the past are used to forecast the future claim counts of the policyholders by considering the risk factors associated with them. In this study, the generalized linear model (GLM) will be used to explore the idea that the claim frequency and pure premium in the motor insurance industry can be modeled based on the policyholders' claim experience for the given risk exposure.

2. LITERATURE REVIEW

2.1 Generalized Linear Model

Nelder and Wedderburn (1972) discovered that the regression models where the response variable is distributed as a member of the exponential family share the same characteristics. The normal, binomial, Poisson, gamma and inverse gamma belong to the exponential dispersion family. It extends the framework of linear regression models with normal distribution to the class of distributions from the exponential family (Silvie and Lenka, 2014). The GLM are not limited by inflexible preconditions. The GLM specifies the distribution of the dependent variable. It has the ability to specify a non-normal distribution and non-identity link function unlike the general linear model.

GLMs extend the framework of linear regression models with normal distribution to the class of distributions from the exponential family. It therefore allows for the modeling of large number of variable types such as counts, frequencies, and binary and even to treat skewed probability distributions of the data (Kafkova and Krivankova, 2014). The paper assumes that the number of claims is a dependent variable which follows Poisson distribution and depends on known and observable predictors that characterize the insured individual or vehicle, i.e. vehicle body type, vehicle age, area of residence, gender of policyholder and age band of the policyholder.

The auto insurance is seen to hold an increased interest because it manages a large number of situations i.e. both the number of insured vehicles and of accidents with a variety of risks (David and Jemna, 2015). Boucher and Guillen (2009) express the need to model the claim frequency as it is the basis of premium calculation by the insurers. Antonio et al (2012) proposed the Poisson distribution for modeling the claim frequency. In connection to that, Cameroon and Trivedi (1999) argued out that the Poisson distribution has an equi-dispersion property i.e. the equality of the mean and variance. Unobserved heterogeneity is usually a common feature in an automobile insurance and therefore we would expect over-dispersion.

Jong and Heller (2008) analyzed an automobile portfolio using GLM. Drivers in the motor insurance portfolio are observed and the numbers of claims produced over the past years are recorded. Both the binomial and Poisson distributions are used. The Poisson regression model has gained application in the analysis of ship damage rates. The GLM fits a Poisson regression for the analysis of count data (Collet, 2003). The incident counts are then modeled as occurring at a Poisson rate given the values of the predictors. The data under consideration contains information on certain types of damage caused by the waves. The risk of damage is associated with three variables; the ship type, the year of ship construction and the block of years the ship saw service. Collect's idea can be extended to model insurance claim count data.

2.2 Credibility theory

This is one of the commonly used experience rating method that involves premium rating for a risk class that lies within a risk group (Whitney, 1918). Typically there are some mix data for the risk class, some data for the other risk classes within the risk group and some data for the risk group as a whole. Behan (2009) describes the need for ensuring a balance between class-experience on one hand and risk-experience on the other. A recent study by Mawuli (2016) on the application of Buhrmann's credibility theory to an automobile insurance claims considers insurance claims counts past data for commercial vehicles having third party liability cover. The number of automobile insurance claims is then estimated using Buhrmann credibility theory without actually considering the risk factors that are associated with them.

Buhlmann and Gisler (2005) proposed the concept of credibility approach to designing an experience rating system of the bonus type used in the motor insurance. A group of motor vehicles risks which is homogeneous with respect to some directly observable risk factors (vehicle type, residence, vehicle use…). Within a group, there will still be accident proneness differentials due to unobservable risk factors (skill, temperament of the driver...). Both the individual driver and group claim experiences of risk should therefore be considered for the purpose of fair premium calculations.

2.3 Bonus-malus system

Future premiums are adjusted by certain motor insurers according to the insured claim history (Mahmoudvand and Aziznasiri, 2014). According to Norberg (1979), at the outset, all automobile drivers in a particular classification group are charged the same premium. The premiums are thereafter adjusted annually according to the insurance company bonus rules, which are usually to the effect that drivers with a favorable claims record are allowed a premium deduction (bonus), while those with an unfavorable ones experience premium increase (malus).

The NCD schemes represent an attempt to categorize the policyholders into relatively homogeneous risk groups who pay premiums relative to their claim experience. Those who have made few claims in recent years are rewarded with discounts on their initial premium. Denuit et al (2007) outlines a wide range of variables an actuary would consider when calculating a motorist's insurance premium such as age, gender and type of vehicle. The rating system penalizes insured responsible for one or more accidents by premium surcharges, and rewarding the claim-free policyholders by awarding them discount. The premium amounts are adjusted each year on the basis of the individual claims experience. The discussion revolves around a closed portfolio. The same approach is seen in the works of (Soren Asmussen, 2013), Arato and Martinek (2014) and Pinquet (2012).

2.4 Multi-state model

Boland (2006), Machado et al (2009) and Noor et al (2014) explains how an actuary would frequently use Markov chain methods to investigate how premiums and insured movements would take place over time. The model is used to predict the dynamics of the insurance purchase made by a policyholder using the transition probability matrix.

Motor car insurance can be a function of many factors such as the type of the car, mileage, age of the driver, region, and sex as presented by Amico et al (2010). Transition matrix with constant forces of transition i.e. a time-homogenous Markov process is obtained from the available data with the payment of a claim by the insurer to the insured seen as a lump sum (impulse or transition reward. In many actuarial applications though, this is impractical as we would require forces which vary with age.

Daniel (2004) outlines the areas of application of the multi-state model as basic survival model, multi-decrement survival models, multiple-life models, disability, Continuous Care Retirement Communities (CCRC's) and the driver ratings. In modeling insured motor driver's ratings, one would consider the states such as Preferred, Standard and Sub-standard. The model describes the probabilities of moving back and forth among the states. An additional state 'Gone' can be included to represent a state for those no longer insured.

2.5 Black-Scholes model

The BS formula is a well-known pricing formula for the put and call options developed in the early 1970s by Black and Scholes (1973). Some actuarial researchers have opted that the payoff functions of a European call option and a stop-loss reinsurance contract are similar, and have proposed an option-pricing approach to pricing insurance risks (Wang, 2002). The model applies the lognormal distributions of market returns.

Holtman (2004) demonstrates the pricing of non-life insurance contracts within a financial option pricing context. The claims risk of an insurance customer is interpreted as an option object. Holder of an insurance contract gives the right to get covered all the incurred insurance claims within a predetermined date (maturity date) and at a predetermined price. In addition, a complete market set up with perfectly efficient buying and selling of insurance contracts is presented with the pure risk and cost based premiums as the sufficient pricing tasks to handle for an insurance company. Robustein (1999) relaxes the requirement of risk-neutrality of BS and that it holds for the risk-averse investors like the investors in the insurance sector with the other conditions unchanged. The same concept is explained by Goodwin et al (2016).

3. METHODOLOGY

3.1 Data

Secondary data is used from an Insurance Brokerage firm in Kenya regarding the Third party compulsory, comprehensive, third party property and fire and theft policies for 2014-2016. Three assumptions were made on the data before used including;

1. All the claims came from the same distribution (they were independent and identically distributed)
2. There were no zero claims for any motor vehicle registered under the given policies
3. All future claims were to be generated from the same distribution.
4. The data will be collected from the auto insurance claims department with a random selection using the inclusion criteria with the policyholders selected known to have been in contract for at least two consecutive years with or without a claim during the period. If no claim will be made in a given policy year, then it will be recorded as zero (0), and if a claim will be made in policy year then it will be recorded as one (1).

3.2 Premium calculation

The annual frequency of claims is calculated from the number of auto claims on the contract. The numbers of claims depend on several factors believed to have direct impact on the expected cost of future claims. The number of claims is a random variable with a Poisson-gamma distribution. Policyholders are usually classified into groups depending on the expected values of claims incurred.

3.2.1 Generalized linear model.

The model is seen to extend the framework of linear regression models with normal distribution to the class of distributions from the exponential family. The GLM are not limited by inflexible preconditions. The GLM specifies the distribution of the dependent variable. It has the ability to specify a non-normal distribution and non-identity link function unlike the general linear model.

GLMs are a means of modeling the relationship between a variable whose outcome we wish to predict and one or more explanatory variables under consideration. The predicted variable is the target variable denoted by in this care the pure premium. The explanatory variables, also called the predictors, are denoted as $X_1, X_2, X_3 ... X_p$, where p is the number of predictors in the GLM. The predictors in this case will be the automobile policyholder characteristics that an insurer may wish to include in their rating plan, e.g. the type of vehicle, age, marital status etc.

Not all the statistical analyses involve data with normal error (Crawley, 2007). Many kinds of data have non- normal errors, for example, the strongly skewed errors, kurtotic errors, strongly bounded and those that cannot lead to negative fitted values as in the case of count data. A GLM allows for the specification of a variety of different error distributions including the Poisson error which is useful with data count and therefore will be our main point of focus. Other commonly used error distributions include the binomial errors that are useful with data on proportions, gamma errors that are useful with data showing a constant coefficient of variation and the exponential errors that are useful with data on time to death (survival analysis).

3.2.2 Assumptions of GLM

I. The error term follows any distribution from the exponential family not necessarily the normal distribution

II. The variance does not have to be assumed as constant

3.2.3 Negative binomial regression model

A way of dealing with the over-dispersion in the Poisson distribution resulting from the random variation is to treat the Poisson mean from any given risk as a random variable itself. This will involve another probability distribution to model the Poisson mean; for this study the gamma distribution resulting to a Poisson-gamma distribution also called the negative binomial distribution. The negative binomial GLM allows for the variance to be non-proportional to the mean.

Over-dispersion is often seen in practice in the motor insurance. Portfolio heterogeneity is a common feature in automobile insurance (Wangui, 2015). Every policyholder is expected to have a constant but unequal underlying risk of having an accident. The mixed Poisson distributions usually have thicker tail than the commonly used Poisson distribution and therefore will provide a good fit to claim frequency data when the portfolio is heterogeneous. Such Poisson mixture distributions that are relevant in the actuarial modeling of claim frequency include the Poisson-gamma distribution, Poisson-exponential distribution, Poisson-erlang distribution and the Poisson-lindley distribution.

The negative binomial distribution has the following properties,

Probability Mass Function;

$$P(Y = y \mid X_1, X_2, X_3, k) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)}\left(\frac{k}{k+\mu}\right)^k \left(\frac{\mu}{k+\mu}\right)^y \qquad y = 0,1,2,...$$

$E(Y) = \mu$

$V(Y) = \mu+(\mu^2/k)$

Link Function: $g(\mu) = \log(\mu)$

Negative binomial distribution includes a third parameter, **k,** called the **over-dispersion parameter** which is related to the variance of the gamma distribution.

By expressing the PMF of negative binomial distribution in the form of the frequency function of an exponential family distribution, we have;

$$f(y,\theta,\phi)=\exp\left[y\log\left(\frac{k\mu}{1+k\mu}\right)-\frac{1}{k}\log(1+k\mu)\right]$$

But since the frequency function of any distribution belonging to the exponential family takes the form;

$$f(y,\theta,\phi)=e^{\left[\frac{y\theta-b(\theta)}{a(\phi)}\right]+c(y,\phi)}$$

$$\Rightarrow\theta=\log\left(\frac{k\mu}{1+k\mu}\right)$$

$$a(\phi)=1$$

$$c(y,\phi)=-\frac{1}{k}\log(1+k\mu)$$

4. RESULTS

The paper estimates the annual claim amounts from which the premium is derived. It is considered that the annual claim amount depend on many risk factors. Twelve factors are taken into account; the customer life value, education level, income, employment status, gender, location, marital status, type of policy, channel of the contract, class and size of the insured vehicle.

The main part of the research involves the application of GLM in the motor insurance based on the policyholders' claims experiences. A data set based on 1,500 policyholders' claims experiences is observed for the periods 2014-2015. The drivers are divided into groups on the basis of the risk factors. For each group, the expected claim amounts per policyholder are modeled. The aim is to find a well-fitting GLM for the claim amounts in terms of the risk factors. Both the Poisson and the negative binomial are assumed for the claim amounts and the log-link function is used. GLM extends the framework of the linear regression models with normal distribution to the class of distributions from the exponential family. Annual claim amount of the policyholders is the response variable and whose outcome is predicted from the given explanatory variables in the dataset.

The two-parameter negative binomial regression used in this study is not a standard member of the exponential family. Dispersion parameter ϕ is treated as a known and a fixed constant so as to make the distribution a member of the family. Log link is used as the canonical link for both Poisson and negative generalized linear models instead of the canonical link for the negative binomial distribution so as to facilitate comparison with the Poisson regression model.

The R software is used to obtain the coefficients of the explanatory variables with the output displayed in a table. The coefficients show how statistically the various variables would influence the change in the annual claim amounts. Some of these variables lead to an increase while for others it is to a decrease in the annual claim amount. It was also used to obtain the p-values for testing the significance of the parameters and the deviance for testing the goodness of fit of our model. The main predictor variables among them the customer life value, location and the marital status of the insured are also outlined. Analysis of deviance, based on a comparison of goodness of fit is used to select the best model between the Poisson and the negative binomial generalized linear models.

5. CONCLUSION

Every person, when applying for a vehicle insurance policy, is usually assigned to a class that is known to be homogeneous in terms of the risk. One of the criteria used for assigning an individual to a certain class is the number of claims recorded for the policyholder. Modeling of the annual claim amounts therefore remains an important task for the motor insurance companies that are known to record more claims compared to other general insurance policies.

Based on the study findings, the research concludes that the motor insurance actuary needs to completely know and understand the essentials of decision and game theory so as to thrive well in the competitive insurance market. An understanding of probability and statistical distribution is necessary to absorb and evaluate risk when balancing claims, reserves and premiums. Considering that the real data from vehicle insurance is not normally distributed, we cannot use the standard linear regression model. This research represents a work devoted to better understand, using data of motor insurance, and how GLM can be used to explain the relation of annual claim amounts on given risk factors.

Overall, the study concludes that the Credibility theory and Bayesian statistics play a big role in evaluating the sample and collateral information in introducing and developing new insurance products. Markov chains on the other hand are essential in predicting the success of the rating methods, including the No Claim Discount, also called the bonus-malus. The time-series methods are used in various ways to predict trends. The Generalized Linear models are considered the essential tools in finding the risk factors for fair premium calculations.

The study also concludes that the negative binomial regression is the best for modeling the motor claims data due to the expected over-dispersion in the distribution of such data. It also accommodates several factors of interest by the insurer with a key on the main parameters for premium estimation as well as considering interaction between them.

Notes

Note 1.

In order to obtain the estimated value of the claim amounts for these groups, we have to take into consideration that the link function for the negative binomial distribution is the logarithm function as presented in the methodology section of this paper, that is;

$$g(\mu)=\log(\mu)=\alpha+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\ldots\ldots\ldots+\beta_nX_n$$

$$\Rightarrow \quad \mu=e^{\alpha+\beta_1X_1+\beta_2X_2+\beta_3X_{3+\ldots+}\beta_nX_n}=e^{x'\beta} \quad \left(x'=\begin{bmatrix}1 & X_1 & X_2 & X_3X_{4\ldots\ldots}X_n\end{bmatrix}\right)$$

Note 2.

Given the explanatory variables X_1, X_2, X_3, X_4…X_n (CLV, EdL, Income, MS, loc_code...). Considering the regression coefficients for the negative binomial regression, the predicted mean claim amount for insurance policy i from the Binomial regression model is given by;

$$\mu_i=e^{4.428}.e^{0.016CLV_i}.e^{-0.321\mathrm{EdL(college)_i}}.e^{0.956EdL(Doctor)_i}\ldots..e^{0.911\mathrm{Size(small)_i}}$$

Table 1. Analysis of variables

Variable	Mean	Median	Std.Dev	Min	Max
Claim_amt	424.7571	379.2	284.7106	0.3821	2893.24
CLV	7892	5759	6506.949	2004	58167
Income	37986	34486	30016.51	0	99960

Mean of 424.7571 and variance of 81060.13 shows that the variance of claim amount exceeds its mean, an indication of overdispersion. In addition, the distribution of the independent variable, Income for example, shows that a good number of the policyholders records no income but with high claim amounts recorded. This indicates lack of homogeneity in the data.

Table 2. Parameter Estimation

Estimate Std. Error t value Pr (>|t|)				
Intercept)	4.428	0.09754	45.393	<2E-16***
CLV	0.016	1.98E-06	11.659	<2E-16***
EdLCollege	-0.321	0.0333	-0.712	0.47647
EdLDoctor	0.956	0.07247	0.975	0.32961
EdL(High School or Below)	0.151	0.03356	0.333	0.73884
EdL(Master)	0.726	0.05226	1.027	0.30476
Emp(StatusEmployed)	0.138	0.07344	0.038	0.96981
Emp(Status Medical Leave)	-0.113	0.08705	-0.026	0.97925
EmpStatus(Retired)	0.982	0.09571	0.758	0.44842
EmpStatus(Unemployed)	0.396	0.07393	1.457	0.14531
GenderM	0.511	0.02592	1.458	0.14516
Income	-0.00398	7.41E-07	-0.589	0.55588
loc_code(Suburban)	1.594	0.03619	44.035	<2E-16***
loc_code(Urban)	1.156	0.04341	26.63	<2E-16***
marital_status(Married)	0.505	0.03797	0.982	0.3264
marital_status(Single)	0.523	0.04368	3.257	0.00115
Policy(Personal Auto)	-0.462	0.03176	-1.074	0.28283
Policy(Special Auto)	-0.138	6.238	-0.163	0.87059
Channel(Branch)	-2.257	0.03219	-0.07	0.94412
Channel(Call Center)	-1.056	0.03587	-2.175	0.02981
Channel(Web)	0.431	0.03988	0.217	0.82805
Size(Medsize)	0.205	0.04411	0.34	0.73157
Size(Small)	0.911	0.05088	1.34	0.18581

The table shows the R output for the negative binomial regression model that relates the annual claim amount to the risk factors by estimating a parameter for each of the risk factors. Reviewing the coefficient signs, an increase in the claim amount can be observed along with an increase in the

Customer life value, the Education level, gender, location, marital status and the size of the vehicle. When the level of education, income, type of policy and channel of business increase, there is an expected decrease in claim amount recorded from the policyholder.

For example, the variable CLV has a coefficient of 0.016 implying that for each unit increase in the Customer life Value, the expected count of claim amount increases by 0.016 units when holding other explanatory variables constant, which is statistically significant. A unit increase in EdL (college) will result to a decrease in the expected count of claims by 0.321. The same applies for the other explanatory variables.

Table 3. Test of Goodness of Fit

	Deviance	Degree of freedom
Null deviance	920.51	1499
Residual deviance	395.92	1477

Additionally to the null and residual deviances, we get that the AIC is 20679.

The residual deviance is not significantly large and the model is therefore good as far as the residuals are concerned.

The negative binomial regression model also leads to an estimated dispersion of ϕ=0.2442958 which is clearly larger than one confirming that overdispersion is present in our claims data.

To check the residuals, we do the hypothesis test:

H_0: the residual deviance is not significantly large and the model is good as far as the residuals are concerned

H_1: otherwise

Since D_*= 395:92 on 1477 degree of freedom, then $\chi_{2395,92,1477}$ = 1.000 > 0.05.

Hence we have no evidence to reject H_0 at 5% significance level and the model is good as far as the residuals are concerned.

Table 4. Significance of parameters

Parameter	**Pr(>|z|)**
CLV	< 2e-16 ***
EdL(College)	0.47647
EdL(Doctor)	0.32961
EdL(High School or Below)	0.73884
EdL(Master)	0.30476
EmpStatus(Employed)	0.96981
EmpStatus(Medical Leave)	0.97925
EmpStatus(Retired)	0.44842
EmpStatus(Unemployed)	0.14531

GenderM	0.14516
Income	0.55588
loc_code(Suburban)	< 2e-16 ***
loc_code(Urban)	< 2e-16 ***
marital_status(Married)	0.3264
marital_status(Single)	0.00115 **
Policy(Personal Auto)	0.28283
Policy(Special Auto)	0.87059
Channel(Branch)	0.94412
Channel(Call Center)	0.02981 *
Channel(Web)	0.82805
Size(Medsize)	0.73157
Size(Small)	0.18581

From Table 4, we can test the significance of the parameters by looking at their **Pr(>|z|)** values. The parameters EdL (Doctor, High School or Below, Master), Emp Status (Employed, Medical Leave, Retired, Unemployed), Gender (M), Income,marital_status (Married), Policy (Personal Auto, Special Auto), Channel (Branch,Web) and Size (Medsize, Small) are statistically insignificant at **5%** significance level, because their Pr(>|z|) values are greater than 0.05. For example, the variable denoting the Education level (College) is not statistically significant as it yields a p-value of 0.47647 which is greater than the level of significance α of 0.05. In consequence, such variables can be excluded from the model to obtain an optimal combination of factors with p-values<0.05 which can explain the variation of the claim amount.

The variables CLV, loc_code (Suburban), loc_code (Urban), marital status (Single) and Channel (Call Centre) are seen to have p-values less than 0.05 and thus considered significant predictors that will contribute to the process of understanding and predicting the amount of claims made on the motor insurance policies.

Table 5. Link Functions

Error	Link Function
Normal	Identity
Poisson	log
Binomial	logit
Gamma	reciprocal
Negative Binomial	1/x or log

In order to obtain the estimated value of the claim amounts for the groups, the link function is taken into consideration. In this paper, the logarithm link function is used for the negative binomial distribution.

References

[1]. Nelder, J. A and Wedderburn, R.W.M., (1972), Generalized linear models, Journal of the Royal Statistical Society, (Series A), 135, 370-384.

[2]. De Jong, P and Heller, G. Z. (2008), Generalized Linear Models for Insurance Data, Cambridge University Press, and Cambridge.

[3]. Boland, P. J. (2006), Statistical methods in general insurance.

[4]. Kafkova Silvie and Krivankova Lenka. (2014), Generalized Linear Models in Vehicle Insurance .Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 62(2): 383-388.

[5]. Commonwealth of Pennsylvania, (2013), Automobile insurance guide. Retrieved 14 March, 2013, from www.portal.state.pa.us/.../auto_insurance/.../ _insurance_guide/53652

[6]. David, M. and Jemna, D. (2015), Modeling the Frequency of Auto Insurance Claims by Means of Poisson and Negative Binomial Models. Annals of the Alexandru Ioan Cuza University - Economics, 62(2), pp. 151-168. Retrieved 26 Jun. 2017, from doi:10.1515/aicue-2015-0011

[7]. Boucher, J. P., Denuit, M., and Guillen, M., (2008), Models of insurance claim counts with time dependence based on generalization of Poisson and Negative Binomial Distributions. Advancing the Science of Risk Variance, 2(1), 135-162

[8]. Cameron, A. C., and Trivedi, P. K., (1999), Essentials of Count Data Regression (Chapter 15). In B. B.H. (Ed.), A Companion to Theoretical Econometrics. Malden, MA: Blackwell Publishing Ltd.

[9]. Collet, D. (2003), Modelling binary data. Chaman and Hall.

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

[11]. Behan, Donald F. (2009), Statistical Credibility Theory", Southeastern Actuarial Conference, June 18, 2009.

[12]. Buhlmann, H. and Gisler, A. (2005), A Course in Credibility Theory and its Applications, Springer, Berlin, Heidelberg.

[13]. Mahmoudvand, R. and Aziznasiri, S. (2014), Bonus-Malus Systems in Open and Closed Portfolios .Mellat Insurance Company, Tehran and Bu-Ali Sina University,Hamedan.

[14]. Norberg, R. (1979), Empirical Bayes credibility. (Submitted for publication.) Preliminary version as Staristical Research Report l977-2. Institute of Mathematics,University of Oslo.

[15]. Gamadeku, Mawuli,(2016), Application of Buhlmann’s Credibility Theory to an Automobile Insurance Claims’,Post graduate theses / dissertations submitted to the College of Science.

[16]. Denuit, M., Marchal, X., Pitrebois, S., Walhin, J.F(2007): Actuarial Modelling of Claim Counts’.Wiley, Chichester

[17]. Pinquet, J. (2000), Experience Rating through Heterogeneous Models, Handbook of Insurance 459-500, Kluwer Academic Publishers. Huebner International Series on Risk, Insurance and Economic Security (Editor: Georges Dionne).

[18]. Meira-Machado L, de Una-Alvarez J, Cadarso-Suarez C, Andersen PK.,(2009), Multi-state models for the analysis of time-to-event data. Stat Methods Med Res.195-222

[19]. G D'Amico, J Janssen, R Manca(2010), Homogeneous semi-Markov reliability models for credit risk management ,Decisions in Economics and Finance 28 (2), 79-93

[20]. James W. Daniel (2014), Multi-state transition models with actuarial applications,Casualty Actuarial Society and the Society of Actuaries.

[21]. Mohd Rahimie Bin Md Noor and Zaidi mat Isa (2014) Predicting Number of Purchasing Life Insurance Using Markov Chain Method, Applied Mathematical Sciences, Vol. 8, 2014, no. 82, 4087 – 4095.

[22]. Black, F. and M. Scholes. (1973), The Pricing of Options and Corporate Liabilities. Journal of Political Economy. 81, 637-659.

[23]. Holtan, J. (2004), Optimal insurance coverage under bonus-malus contracts. ASTIN Bulletin 31, 179-190.

[24]. Rubinstein, Mark, (1976), The Valuation of Uncertain Income Streams and the Pricing of Options, Bell Journal of Economics and Management Science 7, pp. 407-425.

[25]. Wang, S.S. (2002a), A Universal Framework for Pricing Financial and Insurance Risk (2002),ASTIN Bulletin, Vol. 32, No. 2