Knowledge E
Engaging minds

Conference Paper

Modeling Number of Third-party Legal Liability Insurance Claims

Fia Fridayanti Adam

Insurance Administration and Actuarial Department, Vocational Program, Universitas Indonesia, Depok, Indonesia

Abstract

General insurance is insurance that bears financial losses due to the destruction of some or all of the property of the insured. In general insurance, there are 13 types of insurance products that are commonly marketed, one of which is motor vehicle insurance. Risks guaranteed in motor vehicle insurance include, among others, loss and damage to vehicles covered by accident, fire, theft, and third-party liability. In third-party liability insurance, the owner of the vehicle bears the claim of another driver, passenger or pedestrian, resulting from an accident involving the owner of the insured vehicle. In insurance, there is a claim term. The occurrence of this claim may occur at any time so that the claim is a random variable. Since claim data is a count data so that ordinary regression that states there must be a linear relationship between the free and dependent variables, the error must be normally distributed, and so forth cannot be used. So, GLM might be used. Modeling the number of third-party liability insurance claims with independent variables is the number of accidents using negative binomial distribution gives good results. This is indicated by the value of deviance approaching 1. Interpretation of the model indicates if the number of accidents increased 10%, then the number of claims will increase by 18.5%.

Keywords: insurance, third-party liability, claim

Corresponding Author:
Fia Fridayanti Adam
fia@vokasi.ui.ac.id

1. Introduction

Insurance is a contractual agreement between two parties, the one party is obliged to pay the dues and the other party is obliged to provide full guarantee to the payer of contributions in case something happens to the first party or his property in accordance with the agreement made. n general there are 2 types of insurance, namely life insurance and general insurance. Life insurance bears financial losses due to the death of the insured while the general insurance covers financial losses due to the destruction of some or all of the insured's property.

🔓 **OPEN ACCESS**

In general insurance there are 13 types of insurance products are commonly marketed, one of which is motor vehicle insurance. Motor vehicle insurance is the owner of the motor vehicle to enter into an agreement to the insurer to make the payment of the insurance premium (premium) usually per year, but can also be done in several years at once and the vehicle insurance party is also obliged to provide assurance to the insured vehicle in the case of minor damage or heavy damage and replacement of part or all of the insured motor vehicle if experiencing various damages caused by various matters in accordance with the agreement in the insurance.

The amount of motor vehicle insurance premium is calculated based on the percentage of the vehicle's prices, either the new one or the old one. The more expensive of the vehicle's price the smaller the percentage. This percentage result is calculated for insurance premiums that must be paid by the user of insurance services to the insurer. In this insurance agreement, the owner of the vehicle/premium payer will have an insurance policy as a valid certificate of insurance participants by having a policy number and containing about the type of insurance coverage selected. This policy is useful when going to claim insurance.

Risks guaranteed in motor vehicle insurance include, among others, loss and damage to vehicles covered by accident, fire, theft, and third-party liability. In insurance with legal liability a third-party insurer provides compensation for the insured's liability to the loss suffered by a third party, which is directly caused by a motor vehicle as a result of risk guaranteed through the process of deliberation, mediation, arbitration or trial, provided that it has been awarded Prior written approval from the insurer, in the form of property damage, medical expenses, bodily injury and or death; which maximum of the coverage price for guarantee of legal liability to third parties as stated in the policy

In addition to having the obligation to pay a premium, the insured is also entitled to claim in case of events that cause financial loss. The occurrence of this claim may occur at any time so that the claim is a random variable. In addition, the claim event is also a data count, so the usual regression that states there must be a linear relationship between the independent variables and the dependent variable, the error must be normally distributed, and so can't be used.

The Generalized Linear Model (GLM) developed by McCullach and Nelder (1989) can be used to overcome the relationship between independent variables and response variables which is non-normal distribution as long as they are exponential family distributions members.

In actuarial science the use of GLM is widely used, including David (2015) using GLM in calculating motor vehicle insurance, Xacur and Garrido (2015) using GLM in modeling

aggregate claims. Previously Haberman and Renshaw (1993) has explained the use of GLM In various actuarial areas. In addition Cahyandari (2014) describes the over dispersion problem in GLM with Poisson distribution.

This article aims to model the amount of claims on third-party liability insurance by using GLM. The data used are secondary data obtained from http://www.acst.mq.edu.au/GLMforInsuranceData/.

2. Literature Study

Generalized Linear Model (GLM) was introduced by Nelder dan Wedderbun (1972) and then by McCullagh dan Nelder (1989). GLM differs from ordinary regression models in the following two important areas:

1. The distribution of the response variable comes from the exponential distribution family. Thus the response variable distribution does not have to be from normal distribution or close to normal distribution and may be explicitly stated not from the normal distribution

2. The transformation of the mean response variable is linearly correlated with the explanatory variables

McCullagh and Nelder (1989) state that generalized linear modeling has three main component:

1. Random component Y as $E(Y) = \mu$.

   The distribution of Y is a member of an exponential family, such as the Gaussian, Binomial, Poisson, Gamma, or Inverse Gaussian distributions.

2. Systematic component $\eta = \sum_1^p x_j \beta_j$

3. Link functiong $g(\mu) = \eta$.

   This link function is invertible and smooth. So μ can be stated as $\mu = g^{-1}(\eta)$. The inverse of this link function is called the mean function.

The parameter estimate β_jj is performed using the maximum likelihood method, which in turn results in a scoring or newton-raphson method in which the approximation solution is obtained numerically. It is also be obtained through the Iterated Weighted Least Squares procedure which is derived from the Newton–Raphson approximation.

Meanwhile, Poisson distribution, $Y \sim$ Poisson (μ), is used to mode count response data. Here are some examples count data modeling:

- In mortality study, number of death associated with age, sex, and lifestyle variables.
- The number of defective parts in a computer or a batch of manufactured goods.
- The number of diseases a person has.
- Number of insurance claims submitted by the insured.

Supposed $Y_1, Y_2, \ldots Y_N$ are an independent random variable with Y_i representing the number of occurrences at one observation time and supposed μ_i is defined as the average number of occurrences i, the Poisson regression model is $Y_i \sim P(\mu_i)$, $g(\mu_i) = x_i^T \beta$ with the link function $g(\mu_i)$ usuually can be identity function $g(\mu_i) = \mu_i$ or logarithmic function $g(\mu_i) = \log\mu_i$.

If the identity link function used then the value of μ_i is not necessarily positive, while if the logarithmic link function is used then the value of μ_i i must be positive

For the identity link function, interpretation is an additive, while the logarithmic link function of the interpretation is multiplicative. If the logarithmic link function $g(\mu_i) = \log\mu_i = x_i^T \beta$ is used then

$\mu_i = \exp x_i^T \beta$ so the likelihood function becomes $l_i = y_i x_i^T \beta - \exp x_i^T \beta - \log y_i!$

Thus, by using IWLS algorithm we got b as parameter estimator β.

If the rate you want to model then the Poisson regression can be used to model events related to other explanatory variables that may be categorical or continuous data.

The second situation when the number of observations does not change, and if the data is presented in a cross-classified table then the log-linear regression used to model the event.

Supposed $Y_1, Y_2, \ldots Y_N$ are an independent random variable with Y_i representing the number of occurrences in a number of observations n_i for covariate i. The expectation value of Y_i is $E(Y_i) = \mu_i = n_i\theta_i$. For instance, Y_i is the number of insurance claims for a car with a certain type and model. Then Y_i will depend on the number of cars of that type that will be insured, n_i, and other influencing variables

θ_i, like car's age and car's location. Subscript i states a different combination of model, age, location, and so on.

Supposed $\theta_i = e^{x_i^T \beta}$ then $E(Y_i) == n_i e^{x_i^T \beta}$, $Y_i \sim$ Poisson (μ_i)

The natural link function is logarithmic function $\log\mu_i = \log n_i + x_i^T \beta$ where

$\log n_i$ is called as offse$t$. The parameter estimator β can be done by common scoring methods by putting $\log \mu_i = \log n_i + x_i^T \beta$ into the likelihood function.

The fitting model with data is a question that usually appears in statistical modeling.

The principles of significance test, model selection, and diagnostic tests in generalized linear models are similar to normal regressions, although there are technically differences. One way to check the suitability of the model is to check the value of deviance D divided by the degrees of freedom.

Deviance, also known as log-likelihood statistics defined as

$$D = 2\left[l\left(b_{max}; y\right) - l\left(b; y\right)\right]$$

Deviance sample distribution is:

$$D = 2\left[l\left(b_{max}; y\right) - l\left(b; y\right)\right]$$

$$= 2\left[l\left(b_{max}; y\right) - l\left(b; y\right)\right] - 2\left[l\left(b; y\right) - l\left(\beta; y\right)\right] + 2\left[l\left(\beta_{max}; y\right) - l\left(\beta; y\right)\right]$$

where $2\left[l\left(b_{max}; y\right) - l\left(b; y\right)\right] \sim \chi^2\left(m\right)$, $2\left[l\left(b; y\right) - l\left(\beta; y\right)\right] \sim \chi^2\left(p\right)$,

and $\upsilon = 2\left[l\left(\beta_{max}; y\right) - l\left(\beta; y\right)\right]$ is a positive constant number whose value will be close to o if the model fit is of concern to the data as well as the saturated model.

When the deviance value is divided the model degrees of freedom is greater than one then it says the model indicates unfit Under these circumstances the model indicates over dispersion.

In Poisson distribution, in theory known $E\left(Y\right) = \mu$ and Var $E\left(Y\right) = \mu$. However, often the data show variance is a greater than expected value, Var $E\left(Y\right) > E\left(Y\right)$. Over dispersion in the Poisson distribution can be overcome by the Binomial Negative distribution

3. Data and Methods

Third-party liability insurance is additional insurance purchased by the owner of the vehicle. In this insurance the owner of the vehicle bears the claim of another driver, passenger or pedestrian, resulting from an accident involving the owner of the insured vehicle.

In this research data used are secondary data obtained from secondary data http: //www.acst.mq.edu.au/GLMforInsuranceData/. This third-party insurance data is data from 176 geographical areas in New South Wales, Australia, taken over 12 months between 1984 and 1986.

Suppose that we will model the number of claims in an area with the number of reported accidents. So the response variable is the number of claims and the independent variable is the number of accidents.

The steps taken are:

1. calculating mean and variance of response variable.

2. Modeling data based on Poisson distribution.

3. Modeling data based on Negative Binomial distribution.

4. Comparing the result on point (2) and (3).

Data processing was done by PROC GLM on SAS.

4. Result and Discussion

It can be seen from Table 1 that the mean and variety of claims amounts are 586.7 and 1.03×10^6, respectively. So obviously Poisson model cannot be used to model this problem. If we continue to force the Poisson model, it will be seen that the deviance is not fit enough to model this third-party claim problem.

From Table 3 the estimated parameter value is significant, but in Table 2 the deviance value that divided by degree of freedom shows 52.2531 which is well above the number 1. This indicates the over dispersion of the model. Thus, modeling using Poisson distribution is not suitable. The solution is to model the data with negative binomial distribution.

TABLE 1: Variable: Claims.

Moments			
N	176	Sum Weights	176
Mean	586.6875	Sum Observations	103257
Std. Deviation	1013.53928	Variance	1027261.87
Skewness	2.97201203	Kurtosis	10.3092367
Uncorrected SS	240350419	Corrected SS	179770828
Coeff. Variation	172.756243	Std. Error Mean	76.3983978

The binomial negative regression model is used to model third-party claim issues is

$$Y \sim NB\left(\mu, \kappa\right),$$

TABLE 2: Goodness of fit Criteria: Poisson.

Criteria for Assessing Goodness of Fit			
Criterion	DF	Value	Value/DF
Deviance	174	9092.0384	52.2531
Scaled Deviance	174	9092.0384	52.2531
Pearson Chi-Square	174	9970.0721	57.2993
Scaled Pearson X2	174	9970.0721	57.2993
Log Likelihood		647738.0293	

TABLE 3: Parameter estimates: Poisson distribution.

Analysis of Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	–4.0797	0.0263	–4.1312	–4.0282	24099.0	< 0.0001
I_accident	1	1.7458	0.0076	1.7309	1.7606	53239.4	< 0.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

TABLE 4: Goodness of fit criteria: NegBin.

Criteria for Assessing Goodness of Fit			
Criterion	DF	Value	Value/DF
Deviance	174	192.0868	1.1039
Scaled Deviance	174	192.0868	1.1039
Pearson Chi-Square	174	213.8970	1.2293
Scaled Pearson X2	174	213.8970	1.2293
Log Likelihood		651904.2688	

TABLE 5: Parameter estimates: NegBin.

Analysis of Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	–4.1653	0.1418	–4.4431	–3.8874	863.40	< 0.0001
I_accident	1	1.7811	0.0509	1.6814	1.8809	1224.44	< 0.0001
Dispersion	1	0.1267	0.0149	0.0974	0.1559		

$$\log \mu = \log n + \beta_1 + \beta_2 z$$

where Y is third-party claim amount and Z is the number of accidents, and n the size of the population in an area.

As can be seen from Table 5, the average number of claim estimates related to the number of accidents can be modeled as

$$\hat{\mu} = ne^{-4.1653+1.7811\log z}.$$

If z rise by factor a toi az then $rate$ μ/n is estimated to increase as factor $e^{1.7811\log a} = a^{1.7811}$. Suppose number of accident increase 10%, $a = 1.1$, then the claim number mean will increase to $1.1^{1.781} = 1.185017$ which increase 18.5%. This model has shown the value of deviance divided by degree of freedom is 1.1039. This can be seen in Table 4.

This value indicates that the model is good enough to describe the data. Much better compared to using Poisson distribution.

5. Conclusion

This study estimates the number of claims from third-party liability insurance. The response variable is number of claims and the independent variable is number of accidents. The model using Poisson distribution shows over dispersion, which is indicated by a deviance value greater than 1.

Models using Negative Binomial Distributions show better results. Model $\hat{\mu} = ne^{-4.1653+1.7811 log z}$ shows an increase in the number of accidents by 10% will raise the claim amount by 18.5%.

References

[1] David M. 2015. Auto insurance premium calculation using generalized linear models. Procedia Economics and Finance 20 (2015) 147 – 156

[2] De Jong P., Heller G. 2008. *Generalized Linear Model for Insurance Data*. Cambridge University Press. UK

[3] Dobson A. 2002. *An Introduction to Generalized Linear Model*. Second Edition. Chapman and Hall, NY. USA

[4] Haberman S, Renshaw A. 1996. Generalized Linear Models and Actuarial Science. Journal of the Royal Statistical Society. Series D (The Statistician),Vol. 45, No. 4.

[5] Mc Cullagh P, Nelder, JA. 1989. *Generalized Linear Models*. Second Edition. Chapman and Hall, NY. USA

[6] Xacur O, Garrido J. 2015. Generalized linear models for aggregate claims: to Tweedie or not?. European Actuarial Journal.5th edition page 181–202.

[7] Cahyandari R. 2014. Pengujian Overdispersi pada Model Regresi Poisson (Studi Kasus: Laka Lantas Mobil Penumpang di Provinsi Jawa Barat). Statistika, Vol. 14 No. 2, halaman 69 – 76.