Development of rental property insurance models with Generalized Linear Models (GLM)

S Sudarwanto*, L Ambarwati and I Hadi

Department of Mathematics, Faculty of Mathematics and Science, Universitas Negeri Jakarta, Jalan Rawamangun Muka, Jakarta 13220, Indonesia

*sudarwanto@unj.ac.id

Abstract. In this paper we developed an insurance model on property rental business using Generalized Linear Models (GLM). GLM was chosen because the data on property rental match with the characteristics of GLM. Analysis of the model is done by looking at the relationship between the distribution functions that are most widely used in the analysis of property rental data. The models for the relations of distribution functions are: Poisson-Gamma, Poisson-Inverse Gauss, Negative Binomial-Gamma and Negative Binomial-Inverse Gauss. The relationship between the distributions is seen from the root-mean-squared error (RMSE) and absolute mean error (MAE). The simulation results show that the model formed in the case of random effects increases in standard deviation values also increases the variation value on the average estimator. This is in accordance with the condition that the addition of standard deviation means that there is an increase in the variance of the data.

1. Introduction

In recent years, general insurance businesses have experienced a very significant growth rate along with the acceleration of development in all aspects. Chairman of AAUI (Indonesia General Insurance Association), said that the growth of the general insurance premium in 2018 was 72.8 trillion rupiahs. The general insurance business growth is in line with the pace of business growth in other fields such as infrastructure, property and other fields where the 2017 national property capitalization reached 318 trillion rupiahs, a 15% increase compared to 2016 which reached 277 trillion rupiahs. The problem that appears frequently in the apartment rental business is that the tenant deposit is lower than the repairment cost for the damage caused by the tenant. This paper proposes a way to overcome this problem by minimizing the loss risk incurred for both the owner and the tenant focusing on developing a general insurance model that is able to cover the risks that arise in terms of property leases, especially apartments.

The above condition opens opportunities for the emergence of new types of insurance products in Indonesia. The development of insurance models especially property insurance can be developed based on the existing insurance model. One method for developing insurance models is Generalized Linear Model (GLM) as stated by Haberman’s in his paper in 1996[1]. Some researchers applied GLM to calculate vehicle insurance premium [2,3] claim in general insurance [4,5,6] In this article, a property insurance model setting will be developed using GLM.
2. Basic theory and methods

To the claim model, we use the following assumptions: 1) The \((i)^{th}\) event claim at \(T_i\) satisfy \(0 \leq T_i \leq T_2 \leq \cdots\) and Event claim process called arrival time; 2) The occurrence claim at \(T_i\) related to size of claim \(X_i\). Assume that the sequence \((X_i)\) random variable is mutually independent and have independent and identical distributions (i.i.d). 3) \(T_i\) and \(X_i\) are mutually independent.

Based on those assumptions, if \(N(t)\) represents the number of occurrences of some events \(t\) then we have \(N(t) = \#\{i \geq 1; T_i \leq t\}; t > 0\). Total event claims at \(t\), denoted by \(S(t)\), can be formulated by \(S(t) = \sum_{i=1}^{\infty} X_i I[0,t](T_i) = S_N(t)\). Since \(N(0) = 0\) a.s. and \(\mu(0) = 0\), then \(N(t) \sim \text{Pois}(\mu(t))\), \(\mathbb{P}(N(t) = k + h \mid N(s) = k) = p_{k,k+h}(s,t)\); for \(0 \leq s < t\) and \(h, k\) in \(\mathbb{N}\cup\{0\} = \mathbb{N}_0\), where \(p_{k,k+h}\) is a transition probability in the Markov process \(N\) with state space \(\mathbb{N}_0\). Next \(\lim_{s \downarrow 0} \frac{p_{k,k+h}(t,t+s)}{s} = \lambda_{k,k+h}(t)\) is called intensity of Markov process \(N\).

**Theorem 1**

If a Poisson process \(N = (N(t))_{t \geq 0}\) has continue intensity \(\lambda\) in \([0, \infty)\), the for \(k \geq 0\)

\[
\lambda_{k,k+l}(t) = \lambda(t) \text{ if } l = 1 \text{ and } 0 \text{ if } l > 1
\]

Based on Theorem 1, then \(P(N(t, t + s) \geq 2) = 1 - e^{-\lambda(t,t+s)} - \mu(t, t + s)e^{-\mu(t,t+s)}\). Since \(\lambda\) is continue then \(\mu(t, t + s) = \int_t^{t+s} \lambda(y) dy = s\lambda(t)(1 + o(1)) \rightarrow 0\) (when \(s \downarrow 0\)), so that we obtained \(P(N(t, t + s) \geq 2) = o(\mu(t, t + s)) = o(s)\). Therefore \(P(N(t, t + s) = 1) = \lambda(t)s(1 + o(1))\). Since the total claims \(S(t) = \sum_{i=1}^{N(t)} X_i ; t \geq 0\), then \(\mathbb{E}(S(t)) = \mathbb{E}(N(t)\mathbb{E}(X_1))\) and \(\text{Var}(S(t)) = \mathbb{E}(\text{Var}(N(t)\mathbb{E}(X_1)^2)\).

With this model, the claim process and the numbers of claims on property insurance can be modeled as a stochastic process. Therefore, the model will be obtained in the form of a mathematical equation.

2.1. General insurance premium

Any type of general insurance policy must mention the risk that will be guaranteed. Some general insurance guarantee all types of losses or all risks with some protected, exceptions to policies in any type of policy. There is also a policy that states in detail what risks the insurance company wishes to guarantee along with the loss guaranteed or replaced will be in accordance with the insurance price stated in the policy.

In general, there are deductibles or risks that must be borne alone by the insured party in each loss insurance claim. Therefore, this deductible value will reduce claim payments. This is one of the formulae for calculating risks [7].

\[
\text{Risk} = [\text{vulnerability likelihood} \times \text{value}]x[1 - \% \text{risk already controlled uncertainty}] \tag{2}
\]

In general insurance, pure premiums represent the expected costs of claims reported by policy holders during the insurance period. Pure premium can be formulated as:

\[
E[N|x] \cdot E[Z|x] \tag{3}
\]

Where \(E[N|x]\) is the expectation of the number of claims during the insurance period and \(E[Z|x]\) is the expectation value of the cost of each claim during the insurance period and \(x\) is the risk factor.

2.2. Calculating premium general insurance

Insurance data is basically represented in the forms of: 1) The number of claims in the form of integers which can be seen as a continuous random variable. These data histograms are usually clustered on the left and the more right they shrink; 2) Legal representation which usually takes the form of "yes" or "no" answers. This data can be seen as binomial distribution data; 3) The damage level code data that is
usually categorized in numbers 1 to 9 to describe the lowest to most severe level of damage.; 4) Data about claim settlement time which can be seen as a continuous random variable.

Generalized Linear Models (GLM) method is an effective model to calculate insurance premiums [8]. GLM modeling is based on a linear model that has the following models:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i = x_i \beta + \epsilon_i \]

(4)

where \( x_i = [1, x_{i1}, x_{i2}, \ldots, x_{in}] \) and \( \beta = [\beta_0, \beta_1, \beta_2, \ldots, \beta_k]^T \), while \( \epsilon_i, i = 1, 2, \ldots n \) is an error that is assumed to be mutually independent and identically distributed (iid) especially normally distributed with averages 0 and variance \( \sigma^2 \). Based on the assumption of \( \epsilon_i, i = 1, 2, \ldots n \) we then obtained:

\[ E(Y_i|x_i) = x_i \beta + E[\epsilon_i] = x_i \beta = \mu \quad \text{and} \quad \text{var}(Y_i|x_i) = \text{var}[\epsilon_i] = \sigma^2. \]

So that \( Y_i \sim N(x_i \beta, \sigma^2) \).

Model on the equation (4) can be written in the matrix form \( Y = X \beta + \epsilon \), Where \( Y = [Y_1, Y_2, \ldots, Y_n]^T; \) \( X = [x_1^T, x_2^T, \ldots, x_n^T]^T; \) \( \epsilon = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n]^T \), so that \( Y \sim N(X \beta, \sigma^2 I_n) \) when \( I_n \) is identity matrix with an order of \( n \times n \). GLM formula can be written in the matrix form:

\[ \eta = g(\mu) = X \beta \]

(5)

Where \( \eta = [\eta_1, \eta_2, \ldots \eta_n]^T : = [\mu_1, \mu_2, \ldots, \mu_n] \). Insurance data to be modeled in this study follows the form of this log-links \( \log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} \) or \( \mu_i = e^{k \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}} \) and \( Y_1, Y_2, \ldots, Y_n \) mutually independent and have identical distribution.

**Theorem 2**

If \( Y_i, i = 1, 2, \ldots n \) have exponential dispersion, then \( E(Y_i) = b'(\theta_i) \) and var \( (Y_i) = a(\phi)b''(\theta_i) \)

The relationship between the mean value \( \mu_i \) and its linear estimator \( \eta_i \) is symbolized as a smooth and monotone function \( g(\cdot) \). This function is called the link function. If \( g(\mu) = \theta = x_i \beta \) then \( g \) is called the canonical link function. Therefore, the average value of \( \mu_i \) can be expressed as an inverse function of its linear estimator meaning or in other words \( \mu = g^{-1}(\eta) \) or \( \mu_i = g^{-1}(x_i \beta) \). Furthermore, according to De Jong, P. et al. [9] and Ohlsson, E. at al. [8] the number of general insurance claims with Poisson or Binomial distribution is negative. They also stated that the number of claim size has a Gamma and inverse Gaussian distribution.

**3. Results and discussion**

Suppose that \( Y = y \) is insurance property data where \( y = (y_1, y_2, \ldots, y_n) \) and \( y_i; i = 1, 2, \ldots n \) which is assumed to be identical and independent distribution (iid) and have cdf \( f(y_i); i = 1, 2, \ldots n \). Then the likelihood function of \( Y \) can be formulated as:

\[ L(\beta) = \prod_{i=1}^{n} f(y_i) \]

Since \( f(y_i); i = 1, 2, \ldots n \) exponential dispersion so we have,

\[ f(y_i) = e^{k \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)} \]

Then \( l(\beta) = \log L(\beta) = \sum_{i=1}^{n} \log f(y_i) = \sum_{i=1}^{n} \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi) \). The maximum of \( L(\beta) \) is obtained by solving some equations \( \frac{\partial l(\beta)}{\partial \beta_j} = 0; j = 0, 1, \ldots k \). Since \( E(Y_i) = \mu \) and \( \eta_i = x_i \beta \),

then we obtained \( \frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} \), \( \frac{\partial l_i}{\partial \beta_j} \). Then because \( \frac{\partial l_i}{\partial \beta_j} = \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} = \frac{y_i \theta_i - b(\theta_j)}{a(\phi)} \),

\[ \frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} \frac{\partial \mu_i}{\partial \theta_i} + \frac{\partial \mu_i}{\partial \eta_i} = x_{ij} \frac{\partial \mu_i}{\partial \eta_i}, \quad \text{var}(\eta_i)^{-1}(\eta_i - \mu_i) = 0 \]

(6)

The equation (6) can be written by \( X'DV^{-1}(y - \mu) = 0 \), where:
\[
X = \begin{bmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{nn} \end{bmatrix}; \quad D = \begin{bmatrix} \frac{\partial \mu_1}{\partial \eta_1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \frac{\partial \mu_n}{\partial \eta_n} \end{bmatrix} \quad \text{dan} \quad V = \begin{bmatrix} \text{Var}(Y_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \text{Var}(Y_n) \end{bmatrix}
\]

To solve equation (6) we will use Newton-Raphson method, and therefore we need Taylor expansion series. We use first order Taylor expansion \( (\beta) \approx l(\beta^0) + l'(\beta^0)(\beta - \beta^0) + \frac{1}{2} l''(\beta^0)(\beta - \beta^0)^2 \). Suppose that \( l(\beta^0) = 0 \), then we obtained \( \hat{\beta} = \beta^0 - \frac{l'(\beta^0)}{l''(\beta^0)} \). Let \( \hat{\beta}^s \) be estimator of \( \beta \) after \( s \) iterations, then we got \( \hat{\beta}^{s+1} = \hat{\beta}^s - \frac{l'(\hat{\beta}^s)}{l''(\hat{\beta}^s)} \). That means the iteration from \( \beta^0 \) to \( |\hat{\beta}^{s+1} - \hat{\beta}^s| < \varepsilon \) so \( |l'(\hat{\beta}^s)| < \varepsilon; \forall \varepsilon > 0 \). Then we obtained:
\[
l(\beta) \approx l(\beta^0) + S(\beta^0)(\beta - \beta^0) + \frac{1}{2} (\beta - \beta^0)^t H(\beta^0)(\beta - \beta^0)
\]
where \( S(\beta^0) \) is a score vector in \( \beta^0 \) and \( H(\beta^0) \) is a Hessian matrix in \( \beta^0 \), while estimator \( \beta \) after \((s+1)^{\text{th}}\) iteration is:
\[
\hat{\beta}^{s+1} = \hat{\beta}^s - [H(\hat{\beta}^s)]^{-1} S(\hat{\beta}^s)
\]
Scoring Fisher algorithm on equation (8) obtained by replacing the Hessian matrix with an information matrix, so we obtain:
\[
\hat{\beta}^{s+1} = \hat{\beta}^s - [I(\hat{\beta}^s)]^{-1} S(\hat{\beta}^s)
\]

We will make a simulation model based on Table 1. To do a simulation, a sheet of data is created that contains a simulation of the number of claims, the claim size and the matrix that contains eleven variables (covariates) in the regression model. The eleven variables are: location based on NJOP (Nilai Jual Objek Pajak/Tax Object Selling Values), construction class, building use, type of building, type of room, building price, room completeness, safety factor, age of building, size or area, and loss experience. The eleven variables are stated in the information matrix. Next simulation is based on two models, namely: 1) \( \mu = eks(x\beta) \) : model with constant effect; 2) \( \mu = eks(x\beta + \gamma) \) : model with irregular effect.

### Table 1. Simulation model.

| No | Number of claim model | Size of claim model | Random Effect |
|----|-----------------------|---------------------|---------------|
| 1  | Poisson               | Gamma               | No            |
| 2  | Poisson               | Gaussa inverse      | No            |
| 3  | negative binomial     | Gamma               | No            |
| 4  | negative binomial     | Gaussa inverse      | No            |
| 5  | Poisson               | Gamma               | Yes           |
| 6  | Poisson               | Gaussa inverse      | Yes           |
| 7  | negative binomial     | Gamma               | Yes           |
| 8  | negative binomial     | Gaussa inverse      | Yes           |

For simulation purposes, the number of claims \( \beta^N \) and size of the claims \( \beta^Z \) need to be determined. The random effect of \( \beta^N \) is \( y^N = [y^N_0,y^N_1]^t \) and \( \beta^Z \) is \( y^Z = [y^Z_0,y^Z_1]^t \). The link function for GLM is:
\[
\mu = eks(x\beta) = eks[\beta_0, \beta_1 x_1 + \ldots + \beta_n x_n] = eks[\beta_0], eks[\beta_1 x_1], \ldots, eks[\beta_n x_n].
\]
The value of \( \beta^N \) and \( \beta^Z \) are:
\[
\beta^N = [-1.98, 0.31, 0.35, 0.24, 0.71, -0.53, 0.23, -0.13, 0.12, -0.13, 0.49, 0.61, -0.42, -0.62, 1.20, 0.25]^t
\]
\[
\beta^Z = [-2.44, 0.32, -0.19, 0.28, 0.32, -0.22, 0.42, 0.23, 0.29, 0.27, 0.23, 0.31, -0.21, 0.12, 0.64, 0.24]^t
\]
While the random effect is obtained from normal distribution:
\[
y^N_0 \sim N(0, v_1^2); \quad y^N_1 \sim N(0, v_2^2); \quad y^Z_0 \sim N(0, v_3^2) \quad \text{dan} \quad y^Z_1 \sim N(0, v_4^2)
\]
Where \( v_1; v_2; v_3 \) and \( v_3 \) are standard deviation of random effect, in this case we used 0;0.1;0.2;0.4. Premium calculation uses equation (3) with \( E[N|x] = eks[x\beta^N] \) and \( E[Z|x] = E[\mu|x] = eks[x\beta^Z] \). The algorithms to calculate premium and premium estimator as follows:
Algorithm to calculate premium

Input: $\beta^N$ and $\beta^Z$; $v_1; v_2; v_3; v_4$

1. Draw X
2. $\lambda_{asi} = eks(X\beta^N)$
3. $\mu_{asi} = eks(X\beta^Z)$
4. $p_{asi} = \lambda_{asi} \times \mu_{asi}$
5. Return $p_{asi}$

Algorithm to calculate premium estimator

Input: n, m, $\beta^N$ and $\beta^Z$; $v_1; v_2; v_3; v_4$

1. Generate N and Z
2. Set N and Z
3. $\lambda = eks(X\beta^N)$
4. $\mu = eks(X\beta^Z)$
5. $p_{asi} = \lambda \times \mu$
6. Return $p_{asi}$

Evaluations for premiums and estimators are seen through bias tests, root-mean-squared error (RMSE) and absolute mean errors (mean absolute error / MAE) with the following formulas:

$$bias = \frac{1}{N}\sum_{i=1}^{N}(\widehat{p}_{asi} - p_{asi})$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{p}_{asi} - p_{asi})^2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\widehat{p}_{asi} - p_{asi}|.$$ 

Next we will make a simulation by comparing two data models with different distributions.

3.1. The Simulation of poisson-gamma

The results of this simulation show that Negative Binomial-Inverse Gaus gives an average estimator that is closer to the original data, while the Binomial Negative appears to produce results that are not much different from Poisson. Thus, Negative Binomial is a good alternative for Poisson distribution data. In general, it can also be seen that the model that contains Gamma distribution data gives better results than Invers Gauss distribution data, both of simulating data with or without random effects. This happens because the data is the size of the gamma distribution claim. When the standard deviation is 0.4, the results of the data with random effects show that the estimator of bias for the Inverse Gaus average is better than Gamma.

3.2. The simulation of poisson-inverse gauss

The model for the number of fixed claims is a Poisson distribution, but the measure of the claim is an inverse Gaussian distribution. At a standard deviation of 0.1 with a random effect there is no significant difference in the model. Likewise, when the standard deviation is increased to 0.2 for models with random effects it can be seen that the results are similar to the previous results. The results obtained by Negative Binomial-gamma with Poisson-Gamma are almost the same. This is because the data that has a negative binomial distribution contains dierse parameters that have an effect on random effects. When the standard deviation is 0.4 for models with random effects, the results are almost similar to those in the Poisson-gamma model.

3.3. The simulation of negative binomial-gamma

When the standard deviation of 0.1 shows that bias, RMSE and MAE are all different on all four models. While the standard deviation of 0.2 the results obtained are not very different from the results obtained when the standard deviation is 0.1. Here also can be seen that Poisson-Gamma yields the greatest bias on the model. When the standard deviation becomes 0.4, it can be seen that the results obtained by Poisson-gamma and Poisson-inverse Gaussian are almost similar. Broadly speaking, it can be seen that the addition of standard deviation sizes will add to the impact of variance on the data. The results shown by RMSE and MAE can be interpreted that measure of claim has more influence on binomial negative distribution data.

3.4. The simulation of negative binomial-inverse gauss

When the standard deviation of 0.1 shows that the Negative Binomial-Gamma model shows the greatest bias while for RMSE and MAE it is similar from the other models. Furthermore, when the standard deviation of 0.2 shows that the difference in the results obtained by each model is not much different. In the standard deviation of 0.4, it can be seen that the bias is obtained by the Poisson-Gamma model as well as the value of RMSE and MAE. This indicates that in this model the average estimator for
Inverese-Gauss appears to be better than the average estimator for the gamma distribution, which is not unusual because the measure of the claim is an Inverse Gauss distribution.

4. Conclusion

Property Rental insurance model is done based on a generalized linear model GLM because the insurance data match the characteristics of GLM. The compiled model was analyzed by looking at how the influence of the random model effect. In the case of random effects, the simulation results show that the former model increases the standard deviation values and the variation value on the average estimator. This is in accordance with the condition that the addition of standard deviation means that there is an increase in the variance of the data. The simulation also shows that the premium calculation is not much different between the data that has Binomial and Gamma and Inverse Gauss distribution.

Acknowledgement

This work was funded by Universitas Negeri Jakarta trough PSNI research fund with contract number of 10/SPK Penelitian/5.FMIPA/2018.

References

[1] S Haberman and A E Renshaw 1996 Generalized linear models and actuarial science, The Statistician 45(4) 407-436
[2] S Kařková and L Krivánková 2014 Generalized Linear Models in Vehicle Insurance, Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis, 62(2) 383-388
[3] D Mihaela 2015 Auto insurance premium calculation using generalized linear model, Procedia Economics and Finance 20 147 – 156
[4] J Garrido, C Genest and J Schulz 2016 Generalized linear models for dependent frequency and severity of insurance claims Insurance: Mathematics and Economics 70 205-215
[5] G W Peters, A X D Dongc and R Kohnd 2014 A copula based Bayesian approach for paid–incurred claims models for non-life insurance reserving, Insurance: Mathematics and Economics 59 258–278
[6] O A Q Xacur and J Garrido 2015 Generalised linear models for aggregate claims: to Tweedie or not?, European Actuarial Journal. 5 181–202
[7] M E Whitman and H J Mattoro 2007 Principles of Incident Response and Disaster Recovery, Thomson Course Technology, New York
[8] E Ohlsson and B Johansson 2010 Non-Life Insurance Pricing with Generalized Linear Models, Springer-Verlag Berlin
[9] P D J De Jong, and H Z Gillian 2008 Generalized Linear Models For Insurance Data, Cambridge University Press, Cambridge UK
[10] N L Bowers, H U Gerber et al. 1997 Actuarial Mathematics, Second Edition, The Society of Actuaries, USA
[11] T Mikosch 2009 Non-LifeInsurance Mathematics, Second Edition, Springer-Verlag Berlin
[12] S D Promislow 2011 Fundamentals of Actuarial Mathematic, Second Edition, John Wiley and Sons, Chichester
[13] J Yao 2013 Generalized Linear Models for Non-life Pricing-Overlooked Facts and Implications, A Report from GIRO Advanced Pricing Techniques (APT) Working Party, Institute and Faculty of Actuaries.
[14] https://keuangan. Kontan.co.id/ news / 2018-general-targeted-up-insurance premiums) -10
[15] http://www.beritasatu.com / occupancy /396849-2017-pasar-properti-tumbuh-15.html.