ESTIMATION OF CLAIM COST DATA USING ZERO ADJUSTED GAMMA AND INVERSE GAUSSIAN REGRESSION MODELS

Yulia Resti, Noriszura Ismail and Saiful Hafizah Jamaan

1Department of Mathematics, Faculty of Mathematics and Science, University of Sriwijaya, Jalan Raya Palembang Prabumulih Km 32 Inderalaya 30662 Sumatera Selatan, Indonesia
2School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Received 2013-04-01, Revised 2013-06-18; Accepted 2013-07-11

ABSTRACT

In actuarial and insurance literatures, several researchers suggested generalized linear regression models (GLM) for modeling claim costs as a function of risk factors. The modeling of claim costs involving both zero and positive claims experience has been carried out by fitting the claim costs collectively using Tweedie model. However, the probability of zero claims in Tweedie model is not allowed to be fitted explicitly as a function of explanatory variables. The purpose of this article is to propose the application of Zero Adjusted Gamma (ZAGA) and Zero Adjusted Inverse Gaussian (ZAIG) regression models for modeling both zero and positive claim costs data. The models are fitted to the Malaysian motor insurance claims experiences which are divided into three types namely Third Party Bodily Injury (TPBI), Own Damage (OD) and Third Party Property Damage (TPPD). The fitted models show that both claim probability and claim cost are affected by either the same or different explanatory variables. The fitted models also allow the relative risk of each rating factor to be compared and the low or high risk vehicles to be identified, not only for the claim cost but also for the claim probability. The AIC and BIC indicate that ZAIG regression is the best model for modeling both positive and zero claim costs for all claim types.

Keywords: Claim Probability, Claim Cost, Zero Adjusted, Regression Model

1. INTRODUCTION

In actuarial and insurance literatures, several researchers suggested generalized linear regression models (GLM) for modeling claim costs as a function of risk factors and such studies can be found in Brockman and Wright (1992); Renshaw (1994); MacCullagh and Nelder (1989) and Ismail and Jemain (2009). Due to the common properties of claim costs distributions which have positive support and right skewness (Hogg and Klugman, 2009), Gamma and Inverse Gaussian regression models have been used by researchers for fitting insurance claim costs. Nonetheless, Gamma and Inverse Gaussian regression models can only be fitted to claim cost data with non-zero claims.

A distribution that includes both positive and zero claim costs is a distribution with discrete and continuous mixture, where the discrete probability distribution represents cases of zero claims (or cases of making no claim) and the continuous distribution represents cases of positive claims whose distribution is skewed to the right. The modeling of claim costs involving both zero and positive claims experience can be carried out by fitting the claim costs collectively using Tweedie model. As examples, Czado (2005); Jorgensen and Souza (1994) and Smyth and Jorgensen (2002) applied Tweedie model for modeling claim costs, while Peters et al. (2008) and Wuthrich...
fitted Tweedie model for payments of outstandings in claims reserves. However, the probability of zero claims in Tweedie model is not allowed to be fitted explicitly as a function of explanatory variables (or a function of regression covariates). As an alternative, a zero adjusted regression model, which is a regression model with a mixed discrete and continuous distributions, can be used to model both zero and positive claim costs and at the same time, allows the zero claim probability to be modeled explicitly as a function of explanatory variables. The discrete distribution of zero adjusted regression model is represented by Bernoulli distribution, whereas the continuous distribution can be represented by any continuous distribution with a positive range and right skewness. If the continuous distribution is represented by Gamma distribution, the model is called Zero Adjusted Gamma (ZAGA) regression model and if the continuous distribution is represented by Inverse Gaussian distribution, the model is called Zero Adjusted Inverse Gaussian (ZAIG) regression model.

Several applications of ZAGA and ZAIG regression models can be found in Tong et al. (2011); Heller et al. (2006); Ferreira (2008) and Bortoluzzo et al. (2009) also compared ZAIG regression model with Tweedie model and found that ZAIG regression model is better than Tweedie model. The purpose of this article is to propose the application of ZAGA and ZAIG regression models for modeling both zero and positive claim costs data. The models are fitted to the Malaysian motor insurance claims experience which are divided into three types; Third Party Bodily Injury (TPBI), Own Damage (OD) and Third Party Property Damage (TPPD).

2. MATERIALS AND METHODS

2.1. Zero Adjusted Regression Models

Let \( W_i \) be the binary variable that indicates the occurrence of at least one claim and \( \pi_i \) the probability of at least one claim in the \( i \)th rating class, \( i = 1, 2, \ldots, n \). The probability function for \( W_i \) can be defined as:

\[
f(w_i) = (\pi_i)^w_i (1-\pi_i)^{1-w_i}, \quad w_i = 0, 1
\]

Let \( e_i, 0 < e_i < 1 \), be the exposure in the \( i \)th rating class which is defined as the proportion of observation period for which the policy has been in force. Assuming \( e_i \) is known, let \( y_i \) be the number of claims in the observation period and assume \( y_i \) follows a Poisson process with mean (or average) number of claims \( \pi_i \). Then:

\[
y_i | e_i \sim \text{Poisson}(e_i \pi_i)
\]

And:

\[
K_b(y_i = 0 | e_i) = \exp(-e_i \pi_i^*) = 1 - \pi_i^*
\]

so that Equation (1):

\[
f(w_i) = \pi_i^{w_i} (1-\pi_i)^{1-w_i}, \quad w_i = 0, 1
\]

Is a Bernoulli event with \( \pi_i = e_i \pi_i^*; 0 < \pi_i < 1 \).

If \( C_i \) is the random variable for average claim costs in the \( i \)th rating class which is represented as:

\[
\begin{align*}
C_i & = 0 \quad \text{with probability} \quad (1 - \pi_i) \\
> \pi_i \quad \text{with probability} \quad \pi_i
\end{align*}
\]

Then \( C_i \) has a mixed discrete-continuous probability function Equation (2):

\[
f(c_i) = 1 - \pi_i, \quad c_i = 0 \\
\pi_i g(c_i), \quad c_i > 0
\]

where, \( g(c_i) \) is the density function of a continuous and right skewed distribution and \( \pi_i \) is the probability of claim from a Bernoulli event defined in (1). The regression model of a mixed discrete-continuous probability function defined in (2) is called the zero adjusted regression model.

2.2. ZAGA and ZAIG Regression Models

Let \( g(c_i) \) be the density function of Gamma distribution defined as Equation (3):

\[
g(c_i) = \frac{\frac{1}{\sigma^2} \exp\left(-\frac{c_i}{\sigma \mu_i}\right)}{(\sigma \mu_i)^{\frac{1}{2}} \Gamma\left(\frac{1}{\sigma^2}\right)}
\]

where, \( \sigma \) is the scalar parameter. Therefore, the mean and variance for ZAGA regression model are \( E(C_i) = \pi_i \mu_i \) and \( \text{Var}(C_i) = \pi_i \mu_i^2 (\pi_i + \sigma^2) \) and the covariates can be incorporated via a logit link Equation (4):

\[
\pi_i = \frac{\exp\left(x_i^T \beta_i\right)}{1 + \exp\left(x_i^T \beta_i\right)}
\]

and a log link Equation (5):
\[ \mu_i = \exp\left(x_i^T \beta_{\mu}\right) \]  

(5)

where, \( \beta_{\pi} \) and \( \beta_{\mu} \) are the vectors of regression parameters for \( \pi_i \) and \( \mu_i \), respectively and \( x_i \) is the vector of explanatory variables.

Let \( g(c_i) \) be the density function of inverse Gaussian distribution defined as Equation (6):

\[
g(c_i) = \frac{1}{\sqrt{2\pi c_i^3}} \exp\left[ -\frac{\left(\frac{c_i - \mu_i}{\sigma}\right)^2}{2} \right] 
\]

(6)

where, \( \sigma \) is the scalar parameter. Therefore, the mean and variance of ZAIG regression model are \( \mathrm{E}(C_i) = \pi_i \mu_i \) and \( \mathrm{Var}(C_i) = \pi_i \mu_i (1 - \pi_i + \mu_i \sigma^2) \) and the covariates are incorporated in the regression model also via logit and log links in (4)-(5).

2.3. Maximum Likelihood Estimation

The regression parameters, \( \beta_{\pi} \) and \( \beta_{\mu} \) and the scalar parameter, \( \sigma \), for both ZAGA and ZAIG regression models can be estimated using maximum likelihood procedure. The maximum likelihood estimates of, \( \beta_{\pi} \beta_{\mu} \) and \( \sigma \) for ZAGA regression model can be obtained by maximizing likelihood of \( f(c_i) \) shown in (2):

\[
L(\beta_{\pi}, \beta_{\mu}, \sigma) = \prod_{i=1}^{n} f(c_i) 
\]

\[
= \prod_{i=0}^{\infty} \prod_{j=0}^{\infty} \frac{c_i^{(\frac{1}{\sigma^2})} \exp\left[ -\frac{c_i - \mu_i}{\sigma} \right]}{(\sigma^2 \mu_i)^{\frac{3}{2}}} \prod_{j=0}^{\infty} \Gamma\left(\frac{1}{\sigma}\right) 
\]

Or log likelihood:

\[
\log L(\beta_{\pi}, \beta_{\mu}, \sigma) 
= \sum_{i=0}^{\infty} \log(1 - \pi_i) + \sum_{i=0}^{\infty} -\log(\mu_i) + \left(\frac{1}{\sigma^2} - 1\right) \log(c_i) 
- \frac{c_i - \mu_i}{\sigma^2 \mu_i} - \log(\sigma^2 \mu_i) - \log \Gamma\left(\frac{1}{\sigma}\right) 
\]

The maximum likelihood estimates for ZAIG regression model can also be obtained in a similar manner.

2.4. Goodness of Fit

Several measures can be used for comparing ZAGA and ZAIG regression models such as Akaike Information Criteria (AIC) and Schwartz Bayesian Information Criterion (BIC). Let \( n \) be the number of observations, \( m \) the number of estimated parameters and \( \ell \) the log likelihood. The AIC and BIC can be calculated respectively as:

\[
\text{AIC} = -2\ell + 2m \\
\text{BIC} = -2\ell + m \ln(n) 
\]

3. RESULTS

The database for the Malaysian motor insurance claims costs experience is supplied by Insurance Services Malaysia Berhad (ISM), providing information on private car insurance portfolios of ten general insurance companies in 2001-2003 and containing 1,099,175 policies with 117,586 (or 9.7%) claims. The claim costs, which are in Ringgit Malaysia (RM) currency, are divided into three types namely OD, TPPD and TPBI. In this study, we consider five rating factors, each with two, five, five, five and five rating classes, producing a total of \( 2 \times 5 \times 5 \times 5 \times 5 = 1250 \) rating classes. Therefore, each rating class corresponds to eighteen explanatory variables (covariates), including the intercept. The rating factors and classes are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Rating factors and classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating factors</td>
</tr>
<tr>
<td>Coverage</td>
</tr>
<tr>
<td>Vehicle age</td>
</tr>
<tr>
<td>Vehicle cubic capacity (cc)</td>
</tr>
<tr>
<td>Vehicle make</td>
</tr>
<tr>
<td>Location</td>
</tr>
</tbody>
</table>
Table 2. ZAGA and ZAIG regression models (TPBI)

| Parameter          | ZAGA          |                |                | ZAIG          |                |  \\
|--------------------|---------------|---------------|---------------|---------------|---------------|
|                    | estimated std error | p-value       | Estimated std error | p-value       |                |  \\
| **Claim cost:**    |               |               |               |               |               |  \\
| intercept          | 3.76          | 0.22          | 0.00          | 0.04          | 0.09          | 0.64  \\
| non-comprehensive  | 1.48          | 0.15          | 0.00          | 2.67          | 0.26          | 0.00  \\
| 2-3 years          | 0.96          | 0.16          | 0.00          | -             | -             | -     \\
| 8+ years           | -0.81         | 0.17          | 0.00          | -             | -             | -     \\
| 0-1000 cc          | 0.39          | 0.23          | 0.08          | 4.84          | 1.32          | 0.00  \\
| 1301-1500 cc       | -0.89         | 0.22          | 0.00          | -             | -             | -     \\
| 1501-1800 cc       | -1.01         | 0.22          | 0.00          | 1.17          | 0.18          | 0.00  \\
| 1801+ cc           | -1.06         | 0.20          | 0.00          | 2.23          | 0.27          | 0.00  \\
| Local type 2       | 2.18          | 0.24          | 0.00          | 6.42          | 1.88          | 0.00  \\
| Foreign type 1     | -0.5          | -             | 0.00          | 3.66          | 0.40          | 0.00  \\
| Foreign type 2     | 0.91          | 0.16          | 0.00          | 4.41          | 0.59          | 0.00  \\
| Foreign type 3     | 1.10          | 0.28          | 0.00          | 2.10          | 0.58          | 0.00  \\
| north              | 0.36          | 0.19          | 0.05          | 0.71          | 0.15          | 0.00  \\
| east               | 1.22          | 0.19          | 0.00          | 2.07          | 0.25          | 0.00  \\
| south              | 0.72          | 0.19          | 0.00          | 0.82          | 0.16          | 0.00  \\
| east Malaysia      | 0.55          | 0.23          | 0.01          | 1.39          | 0.20          | 0.00  \\
| **Claim probability:** |               |               |               |               |               |  \\
| intercept          | -1.13         | 0.28          | 0.00          | -1.13         | 0.28          | 0.00  \\
| non-comprehensive  | 1.46          | 0.15          | 0.00          | 1.46          | 0.15          | 0.00  \\
| 2-3 years          | -1.48         | 0.23          | 0.00          | -1.48         | 0.23          | 0.00  \\
| 4-5 years          | -1.39         | 0.23          | 0.00          | -1.39         | 0.23          | 0.00  \\
| 6-7 years          | -1.48         | 0.23          | 0.00          | -1.48         | 0.23          | 0.00  \\
| 8+ years           | -2.06         | 0.24          | 0.00          | -2.06         | 0.24          | 0.00  \\
| 0-1000 cc          | 0.51          | 0.23          | 0.03          | 0.51          | 0.23          | 0.03  \\
| 1301-1500 cc       | -0.37         | 0.22          | 0.09          | -0.37         | 0.22          | 0.09  \\
| 1501-1800 cc       | -0.70         | 0.23          | 0.00          | -0.70         | 0.23          | 0.00  \\
| 1801+ cc           | -0.93         | 0.23          | 0.00          | -0.93         | 0.23          | 0.00  \\
| Local type 2       | 2.06          | 0.20          | 0.00          | 2.06          | 0.20          | 0.00  \\
| Foreign type 2     | 0.53          | 0.19          | 0.01          | 0.53          | 0.19          | 0.01  \\
| Foreign type 3     | 3.17          | 0.24          | 0.00          | 3.17          | 0.24          | 0.00  \\
| north              | 0.59          | 0.23          | 0.01          | 0.59          | 0.23          | 0.01  \\
| east               | 1.21          | 0.23          | 0.00          | 1.21          | 0.23          | 0.00  \\
| south              | 0.87          | 0.23          | 0.00          | 0.87          | 0.23          | 0.00  \\
| east Malaysia      | 1.89          | 0.24          | 0.00          | 1.89          | 0.24          | 0.00  \\

The fitted ZAGA and ZAIG regression models for TPBI, OD and TPPD claims are presented in Table 2-4. The results indicate that both ZAGA and ZAIG models produce either same or different significant factors. As an example, the claim cost for TPBI from ZAGA model imply that the rating factor for foreign type 1 vehicle is not significant, while ZAIG regression model show that the rating factors for 2-3 years, 8+ years and 1301-1500 cc vehicles are not significant. On the other hand, the claim probabilities for TPBI from both ZAGA and ZAIG regression models have the same significant rating factors.

4. DISCUSSION

From Table 2-4, the fitted claim cost and claim probability for each claim type can be calculated respectively as:
Table 3. ZAGA and ZAIG regression models (OD)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ZAGA</th>
<th>ZAIG</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimated</td>
<td>std error</td>
<td>p-value</td>
<td>estimated</td>
</tr>
<tr>
<td>Claim cost:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.41</td>
<td>0.16</td>
<td>0.00</td>
<td>2.79</td>
</tr>
<tr>
<td>6-7 years</td>
<td>-0.38</td>
<td>0.17</td>
<td>0.02</td>
<td>-</td>
</tr>
<tr>
<td>1301-1500 cc</td>
<td>-1.35</td>
<td>0.22</td>
<td>0.00</td>
<td>-</td>
</tr>
<tr>
<td>1501-1800 cc</td>
<td>-1.88</td>
<td>0.20</td>
<td>0.00</td>
<td>-2.73</td>
</tr>
<tr>
<td>1801+ cc</td>
<td>-1.55</td>
<td>0.20</td>
<td>0.00</td>
<td>-1.85</td>
</tr>
<tr>
<td>Local type 2</td>
<td>3.84</td>
<td>0.22</td>
<td>0.00</td>
<td>4.97</td>
</tr>
<tr>
<td>Foreign type 2</td>
<td>1.52</td>
<td>0.18</td>
<td>0.00</td>
<td>1.94</td>
</tr>
<tr>
<td>Foreign type 3</td>
<td>2.31</td>
<td>0.27</td>
<td>0.00</td>
<td>9.87</td>
</tr>
<tr>
<td>North</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.03</td>
</tr>
<tr>
<td>East</td>
<td>1.55</td>
<td>0.19</td>
<td>0.00</td>
<td>2.55</td>
</tr>
<tr>
<td>South</td>
<td>0.53</td>
<td>0.19</td>
<td>0.01</td>
<td>1.08</td>
</tr>
<tr>
<td>east Malaysia</td>
<td>1.06</td>
<td>0.19</td>
<td>0.00</td>
<td>2.01</td>
</tr>
<tr>
<td>σ</td>
<td>0.38</td>
<td>0.03</td>
<td>0.00</td>
<td>-0.91</td>
</tr>
<tr>
<td>Claim probability:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.50</td>
<td>0.08</td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>1501-1800 cc</td>
<td>-0.56</td>
<td>0.15</td>
<td>0.00</td>
<td>-0.56</td>
</tr>
<tr>
<td>1801+ cc</td>
<td>-0.50</td>
<td>0.15</td>
<td>0.00</td>
<td>-0.50</td>
</tr>
<tr>
<td>Foreign type 3</td>
<td>1.22</td>
<td>0.18</td>
<td>0.00</td>
<td>1.22</td>
</tr>
<tr>
<td>log likelihood</td>
<td>-3149.03</td>
<td>-3073.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>6330.06</td>
<td>6177.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>6412.16</td>
<td>6254.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$$\hat{\mu}_i = \exp \left( \sum \beta_k x_{ik} \right)$$

And:

$$\hat{\pi}_i = \frac{\exp \left( \sum \beta_k x_{ik} \right)}{1 + \exp \left( \sum \beta_k x_{ik} \right)}$$

where, $\beta_k$ is the regression parameter and $x_{ik}$ the explanatory variable with a value of zero or one. As an example, the fitted claim cost and claim probability for TPBI based on ZAIG regression model for vehicles with comprehensive coverage, age 0-1 year, cubic capacity 0-1000, local (type 1) make and North location respectively are:

$$\hat{\mu}_i = \exp(0.04 + 2.67 + 4.84 + 0.71) = \text{RM}267.74$$

And:

$$\hat{\pi}_i = \frac{\exp(-1.13 + 0.51 + 0.59)}{1 + \exp(-1.13 + 0.51 + 0.59)} = 0.4925$$

so that the expected TPBI claim cost that take into account both zero and positive claims is $E(C_1) = \hat{\pi}_i \hat{\mu}_i = \text{RM}131.86$.

The results in Table 2-4 can also be used to compare the relative risk of each rating factor and therefore, identifying low or high risk vehicles. As an example, the fitted claim cost and claim probability for TPBI based on ZAIG regression model for vehicles with non-comprehensive coverage, age 0-1 year, cubic capacity 0-1000, local (type 1) make and North location respectively are:

$$\hat{\mu}_i = \exp(0.04 + 2.67 + 4.84 + 0.71) = \text{RM}3866.09$$

And:

$$\hat{\pi}_i = \frac{\exp(-1.13 + 0.51 + 0.59)}{1 + \exp(-1.13 + 0.51 + 0.59)} = 0.8069$$

Indicating that non-comprehensive coverage has higher risk in both claim cost and claim probability than the comprehensive coverage. Therefore, the claim cost that take into account both zero and positive claims increases and the expected value is $E(C_1) = \hat{\pi}_i \hat{\mu}_i = \text{RM}3119.56$. 

190
Table 4. ZAGA and ZAIG regression models (TPPD)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ZAGA</th>
<th>ZAIG</th>
<th>ZAGA</th>
<th>ZAIG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimated</td>
<td>std error</td>
<td>p-value</td>
<td>estimated</td>
</tr>
<tr>
<td>Claim cost:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intercept</td>
<td>3.21</td>
<td>0.17</td>
<td>0.00</td>
<td>4.00</td>
</tr>
<tr>
<td>non-comprehensive</td>
<td>1.37</td>
<td>0.14</td>
<td>0.00</td>
<td>2.54</td>
</tr>
<tr>
<td>4-5 years</td>
<td>-0.29</td>
<td>0.17</td>
<td>0.09</td>
<td>-</td>
</tr>
<tr>
<td>6-7 years</td>
<td>-1.11</td>
<td>0.18</td>
<td>0.00</td>
<td>-0.96</td>
</tr>
<tr>
<td>8+ years</td>
<td>-1.65</td>
<td>0.18</td>
<td>0.00</td>
<td>-1.30</td>
</tr>
<tr>
<td>0-1000 cc</td>
<td>0.68</td>
<td>0.20</td>
<td>0.00</td>
<td>-</td>
</tr>
<tr>
<td>1301-1500 cc</td>
<td>-0.82</td>
<td>0.17</td>
<td>0.00</td>
<td>-2.43</td>
</tr>
<tr>
<td>1801-1800 cc</td>
<td>-0.48</td>
<td>0.17</td>
<td>0.00</td>
<td>-2.49</td>
</tr>
<tr>
<td>Local type 2</td>
<td>2.43</td>
<td>0.20</td>
<td>0.00</td>
<td>4.89</td>
</tr>
<tr>
<td>Foreign type 1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.51</td>
</tr>
<tr>
<td>Foreign type 2</td>
<td>1.21</td>
<td>0.17</td>
<td>0.00</td>
<td>-</td>
</tr>
<tr>
<td>Foreign type 3</td>
<td>1.58</td>
<td>0.24</td>
<td>0.00</td>
<td>0.67</td>
</tr>
<tr>
<td>north</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.97</td>
</tr>
<tr>
<td>east</td>
<td>0.64</td>
<td>0.17</td>
<td>0.00</td>
<td>2.31</td>
</tr>
<tr>
<td>south</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.24</td>
</tr>
<tr>
<td>east Malaysia</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.32</td>
</tr>
<tr>
<td>σ</td>
<td>4.27</td>
<td>0.02</td>
<td>0.00</td>
<td>-0.75</td>
</tr>
<tr>
<td>Claim probability:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intercept</td>
<td>0.01</td>
<td>0.2</td>
<td>0.96</td>
<td>0.01</td>
</tr>
<tr>
<td>non-comprehensive</td>
<td>0.94</td>
<td>0.13</td>
<td>0.00</td>
<td>0.94</td>
</tr>
<tr>
<td>2-3 years</td>
<td>-1.19</td>
<td>0.21</td>
<td>0.00</td>
<td>-1.19</td>
</tr>
<tr>
<td>4-5 years</td>
<td>-1.17</td>
<td>0.21</td>
<td>0.00</td>
<td>-1.17</td>
</tr>
<tr>
<td>6-7 years</td>
<td>-1.34</td>
<td>0.21</td>
<td>0.00</td>
<td>-1.34</td>
</tr>
<tr>
<td>8+ years</td>
<td>-1.38</td>
<td>0.21</td>
<td>0.00</td>
<td>-1.38</td>
</tr>
<tr>
<td>1501-1800 cc</td>
<td>-1.09</td>
<td>0.18</td>
<td>0.00</td>
<td>-1.09</td>
</tr>
<tr>
<td>1801+ cc</td>
<td>-1.16</td>
<td>0.18</td>
<td>0.00</td>
<td>-1.16</td>
</tr>
<tr>
<td>Local type 2</td>
<td>1.08</td>
<td>0.17</td>
<td>0.00</td>
<td>1.08</td>
</tr>
<tr>
<td>Foreign type 1</td>
<td>-0.40</td>
<td>0.18</td>
<td>0.03</td>
<td>-0.40</td>
</tr>
<tr>
<td>Foreign type 3</td>
<td>2.08</td>
<td>0.20</td>
<td>0.00</td>
<td>2.08</td>
</tr>
<tr>
<td>east</td>
<td>0.80</td>
<td>0.18</td>
<td>0.00</td>
<td>0.80</td>
</tr>
<tr>
<td>south</td>
<td>0.31</td>
<td>0.18</td>
<td>0.08</td>
<td>0.31</td>
</tr>
<tr>
<td>east Malaysia</td>
<td>0.46</td>
<td>0.18</td>
<td>0.01</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Based on both AIC and BIC, ZAIG regression model is better than ZAGA regression model for all TPBI, OD and TPPD claims.

5. CONCLUSION

This study proposes the application of ZAGA and ZAIG regression models for modeling both positive and zero claim costs for three types of motor insurance claims; TPBI, OD and TPPD. The main advantage of using ZAGA and ZAIG regression models compared to Tweedie model is that the probability of claim can be expressed in a function of explanatory variables. The fitted models show that both claim probability and claim cost are affected by either the same or different explanatory variables. The fitted models also allow the relative risk of each rating factor to be compared and the low or high risk vehicles to be identified, not only for the claim cost but also for the claim probability. The application of ZAGA and ZAIG regression models proposed in this study can also be used for other lines of insurance (besides motor insurance) or any other data...
(besides insurance data), as long as the covariates for both positive and zero costs are available. Further applications can also be performed to other distributions with positive range and right skewness.

6. ACKNOWLEDGEMENT

The researchers gratefully acknowledge the financial support received in the form of research grants (GUP-2012-024 and LRGS/TD/2011/UKM/ICT/03/02) from the Ministry of Higher Education (MOHE), Malaysia. The authors are also pleased to thank Insurance Services Malaysia Berhad (ISM) for supplying the data.

7. REFERENCES


