Jurnal Teknologi, 43(C) Dis. 2005: 39–54

GENERALIZED POISSON REGRESSION: AN ALTERNATIVE FOR RISK CLASSIFICATION

NORISZURA ISMAIL[1] & ABDUL AZIZ JEMAIN[2]

Abstract. The Poisson regression model has been widely used for risk classification in the recent years. However, the Poisson regression model assumes that the mean and variance of the dependent variable is equal, whereas in practice, the data may display overdispersion or extra-Poisson variation, i.e., a situation where the variance exceeds the mean. Inappropriate imposition of the Poisson may underestimate the standard errors and overstate the significance of the regression parameters, and consequently, giving misleading inference about the regression parameters. Therefore, the objective of this paper is to suggest the Generalized Poisson regression model as an alternative for risk classification. In this paper, the Poisson and Generalized Poisson regression models are fitted, tested and compared on two types of Malaysian motor insurance claims count data; Own Damage (OD) and Third Party Bodily Injury (TPBI). The Poisson regression model for OD claims gives large values for Pearson chi-squares and deviance, indicating possible existence of overdispersion. Based on the results of goodness-of-fit tests, the Generalized Poisson is superior to the Poisson. On the contrary, the small deviance for Poisson regression model in TPBI claims implies that the model is adequate. Based on the likelihood ratio test, the likelihood ratio is insignificant, implying that the Poisson is adequate.

Keywords: Risk classification, Generalized Poisson, claim frequency

Abstrak. Semenjak beberapa tahun yang lepas, model regresi Poisson telah diguna secara meluas untuk pengkelasan risiko. Namun, model regresi Poisson mengandaikan bahawa min dan varians pemboleh ubah bersandar adalah sama, sedangkan secara praktis, masalah lebih-serakan mungkin wujud di dalam data, yakni, situasi di mana varians adalah lebih besar daripada min. Penggunaan model Poisson yang kurang bersesuaian mungkin boleh menyebabkan nilai sisihan piawai terkurang-anggar dan kesignifikan parameter regresi terlebih-anggar, yang akhirnya, boleh memberikan penta'abiran parameter regresi yang agak mengelirukan. Oleh itu, tujuan kertas ini adalah untuk mencadangkan model regresi Poisson Teritlak sebagai alternatif terhadap pengkelasan risiko. Dalam kertas ini, model regresi Poisson dan Poisson Teritlak akan disuai, diuji dan dibandingkan terhadap dua jenis data bilangan tuntutan insurans motor di Malaysia; Kerosakan Sendiri (OD) dan Kecederaan Badan Pihak Ketiga (TPBI). Model regresi Poisson bagi tuntutan OD memberikan nilai khi-kuasa dua Pearson dan devians yang besar, dan ini mengimplikasikan kemungkinan wujudnya masalah lebih-serakan. Berdasarkan hasil ujian kebagusan, model Poisson Teritlak adalah lebih baik daripada model Poisson. Sebaliknya, nilai devians yang kecil bagi model regresi Poisson tuntutan TPBI mengimplikasikan bahawa model yang disuai adalah memadai. Berdasarkan ujian nisbah kebolehjadian, nisbah kebolehjadian adalah tidak signifikan, dan ini menunjukkan bahawa model Poisson adalah memadai.

Kata kunci: Pengkelasan risiko, Poisson Teritlak, kekerapan tuntutan

[1&2]Pusat Pengajian Sains Matematik, Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan. Tel: 03-89215754; Fax: 89254519: E-mail: [1]ni@pkrisc.cc.ukm.my, [2]azizj@pkrisc.cc.ukm.my

1.0 INTRODUCTION

Risk classification is the process of statistical modeling that classifies risks with similar risk characteristics into cross-classified classes. The risks may be categorized according to risk or rating factors. In motor insurance for instance, the driver's gender, claim experience and age, or the vehicle's model, capacity and year may be considered as rating factors. The goal of risk classification in an insurance system is to estimate "fair" classification rates, i.e., high risk insureds should be classified into higher risk classes and vice versa. Failure to achieve this goal may lead to adverse selection to insureds and economic losses to insurers.

In an insurance system, the premium charged to the policyholders comprises two components; risk premium and related expenses. The risk premium, which exclude the expenses, is equivalent to the product of expected claim frequency and expected claim severity. In this paper, risk classification will be used to estimate claim frequency rates and to classify the frequency rates into supposedly homogeneous rating classes. For motor insurance, the claim frequency rate is equal to the claim count per exposure unit, and the exposure is usually expressed in terms of car-year unit [1].

In the last forty years, researchers suggested various statistical procedures to estimate the parameters in risk classification model. For example, Bailey and Simon [2] suggested the minimum chi-squares, Bailey [3] devised the zero bias, Jung [4] produced a heuristic method for minimum modified chi-squares, Ajne [5] proposed the method of moments for minimum modified chi-squares, Chamberlain [6] applied the weighted least squares, Coutts [7] suggested the method of orthogonal weighted least squares with logit transformation, Harrington [8] applied the maximum likelihood procedure for models with functional form, and Brown [9] suggested the bias and likelihood functions for minimum bias and maximum likelihood models.

Research on risk classification in the recent actuarial literature is still continuing and developing. For example, Mildenhall [10] merged the models which were introduced by Bailey and Simon [2], i.e., the minimum bias models, with the Generalized Linear Models (GLMs), i.e., the maximum likelihood models. Besides providing strong statistical justifications for the minimum bias models which were originally based on a non-parametric approach, his effort also allowed a variety of parametric models to be chosen from. Later, Fu and Wu [11] also developed the model of Bailey and Simon by following the same approach which was created by Bailey and Simon, i.e., the non-parametric approach. As a result, their research offers a wide range of non-parametric models to be created and applied. Ismail and Abdul Aziz [12] found a match point that merged the available parametric and non-parametric models, i.e., minimum bias and maximum likelihood models, by rewriting the models in a generalized form. The parameters were solved by applying weighted equation, regression approach and Taylor series approximation.

The Poisson regression model has been widely used for risk classification in the recent years. For instance, McCullagh and Nelder [13] proposed the model for number of damage incidents of cargo-carrying vessels in marine insurance. In motor insurance, Brockman and Wright [14] applied the model to the U.K. motor own damage claims, and Renshaw [15] suggested the model for motor claims which were provided by a leading insurance company in the U.K. For local claims experience, the Poisson regression model was fitted by Ismail and Jemain [16] to a set of private car own damage claims which was provided by an insurance company in Malaysia.

However, the Poisson regression model assumes that the mean and variance of the dependent variable is equal. In practice, the data may display overdispersion or extra-Poisson variation, i.e., a situation where the variance exceeds the mean. Inappropriate imposition of the Poisson may underestimate the standard errors and overstate the significance of the regression parameters, and consequently, giving misleading inference about the regression parameters. Therefore, the objective of this paper is to suggest the Generalized Poisson regression model as an alternative for risk classification. In this paper, the Poisson and Generalized Poisson regression models are fitted, tested and compared on two types of Malaysian motor insurance claims count data; Own Damage (OD) and Third Party Bodily Injury (TPBI).

2.0 POISSON REGRESSION MODEL

Let Y_i be the random variable for claim counts in the ith class, $i = 1,2,...,n$, where n is the number of rating classes. If Y_i follows a Poisson distribution, the probability density function is,

$$\Pr(Y_i = y_i) = \frac{\exp(-\mu_i)\,\mu_i^{yi}}{y_i\,!}, \qquad y_i = 0, 1, \ldots, \tag{1}$$

with mean and variance, $E(Y_i) = Var(Y_i) = \mu_i$.

To incorporate covariates and to ensure non-negativity, the mean or the fitted value is assumed to be $E(Y_i \mid \mathbf{x_i}) = \mu_i = e_i \exp(\mathbf{x_i^T}\boldsymbol{\beta})$, where e_i denotes a measure of exposure, $\mathbf{x_i}$, a vector of covariates and $\boldsymbol{\beta}$, a vector of regression parameters.

The estimates of $\boldsymbol{\beta}$ may be obtained by using the maximum likelihood method. For Poisson regression model, the likelihood is,

$$\frac{\partial \ell}{\partial \beta_j} = \sum_i \frac{y_i - \mu_i}{\mu_i} \frac{\partial \mu_i}{\partial \beta_j} = 0 \qquad j = 1, 2, \ldots, p, \tag{2}$$

where p is the number of regression parameters. Since Equation (2) is also equivalent to the weighted least squares, the estimates of $\boldsymbol{\beta}$ may be solved by using the Iteratively Weighted Least Squares (IWLS) procedure.

3.0 GENERALIZED POISSON REGRESSION MODEL

In this paper, two different types of Generalized Poisson regression models will be discussed; each will be referred as the Generalized Poisson I (GPI) and Generalized Poisson II (GPII).

3.1 GPI

Let Y_i be the random variable for GPI distribution. The probability density function is [17],

$$\Pr(Y_i = y_i) = \left(\frac{\mu_i}{1+a\mu_i}\right)^{y_i} \frac{(1+ay_i)}{y_i!} \exp\left(-\frac{\mu_i(1+ay_i)}{1+a\mu_i}\right), \qquad y_i = 0,1,\ldots. \tag{3}$$

The mean is also assumed to be equal to $E(Y_i \mid \mathbf{x_i}) = \mu_i = e_i \exp(\mathbf{x_i^T}\boldsymbol{\beta})$. However, the conditional variance is equivalent to $Var(Y_i \mid \mathbf{x_i}) = \mu_i(1+a\mu_i)^2$.

The GPI is a natural extension of the Poisson. When the dispersion parameter, a, is equal to zero, the probability density function, which is shown by Equation (3), reduces to the Poisson so that the mean is equal to the variance, i.e., $E(Y_i \mid \mathbf{x_i}) = Var(Y_i \mid \mathbf{x_i})$. For $a > 0$, the variance is larger than the mean, i.e., $Var(Y_i \mid \mathbf{x_i}) > E(Y_i \mid \mathbf{x_i})$, and for this situation, the regression model represents count data with overdispersion. For $a < 0$, the variance is smaller than the mean, i.e., $Var(Y_i \mid \mathbf{x_i}) < E(Y_i \mid \mathbf{x_i})$, so that now the regression model represents count data with underdispersion.

If $\boldsymbol{\beta}$ is estimated by the maximum likelihood method, the related equations are,

$$\ell(\boldsymbol{\beta}, a) = \sum_i y_i \log\left(\frac{\mu_i}{1+a\mu_i}\right) + (y_i - 1)\log(1+ay_i) - \frac{\mu_i(1+ay_i)}{1+a\mu_i} - \log(y_i!), \tag{4}$$

$$\frac{\partial \ell}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)}{\mu_i(1+a\mu_i)^2} \frac{\partial \mu_i}{\partial \beta_j} = 0 \qquad j = 1,2,\ldots,p. \tag{5}$$

Since Equation (5) is also equivalent to the weighted least squares, with a slight modification, the estimates of $\boldsymbol{\beta}$ may also be solved using the IWLS procedure.

In this paper, two methods are suggested for solving the dispersion parameter, a; maximum likelihood method and method of moments. Under the maximum likelihood method, the related equations are,

$$\frac{\partial \ell}{\partial a} = \sum_i -\frac{y_i \mu_i}{1+a\mu_i} + \frac{y_i(y_i - 1)}{1+ay_i} - \frac{\mu_i(y_i - \mu_i)}{(1+a\mu_i)^2} = 0, \tag{6}$$

$$\frac{\partial^2 \ell}{\partial a^2} = \sum_i \frac{y_i \mu_i^2}{(1+a\mu_i)^2} - \frac{y_i^2 (y_i - 1)}{(1+ay_i)^2} + \frac{2\mu_i^2 (y_i - \mu_i)}{(1+a\mu_i)^3}. \tag{7}$$

The maximum likelihood estimate of a maybe solved simultaneously with $\boldsymbol{\beta}$ and the procedure involves sequential iterations. In the first sequence, by using an initial value of a, $a_{(0)}$, $\ell(\boldsymbol{\beta}, a)$ is maximized with respect to $\boldsymbol{\beta}$, producing $\boldsymbol{\beta}_{(1)}$. The related equations are Equations (4) and (5). In the second sequence, by holding $\boldsymbol{\beta}$ fixed at $\boldsymbol{\beta}_{(1)}$, $\ell(\boldsymbol{\beta}, a)$ is maximized with respect to a, producing $a_{(1)}$. The related equations are Equations (6) and (7). By sequentially iterating between holding a fixed and holding $\boldsymbol{\beta}$ fixed, the maximum likelihood estimates of $\boldsymbol{\beta}$ and a will be obtained.

Under the method of moments, a may be estimated by equating the Pearson chi-squares with the degrees of freedom, as suggested by Breslow [18],

$$\sum_i \frac{(y_i - \mu_i)^2}{\mu_i (1+a\mu_i)^2} = n - p, \tag{8}$$

where n denotes the number of rating classes and p the number of regression parameters. The sequential iteration procedure similar to the one mentioned above can also be used, this time producing maximum likelihood estimates of $\boldsymbol{\beta}$ and moment estimate of a.

In this paper, when a is estimated by the maximum likelihood, the model will be denoted by GPI(MLE). Likewise, when a is estimated by the method of moments, the model will be denoted by GPI(moment).

3.2 GPII

Let Y_i be the random variable for GPII distribution. The probability density function is [19],

$$\Pr(Y_i = y_i) = \begin{cases} \mu_i (\mu_i + (a-1) y_i)^{y_i - 1} a^{-y_i} \dfrac{\exp\left(-\dfrac{\mu_i + (a-1) y_i}{a}\right)}{y_i!}, & y_i = 0,1,\ldots, \\ 0, & y_i > m, a < 1 \end{cases} \tag{9}$$

where $a \geq max\left(\frac{1}{2}, 1 - \frac{\mu_i}{4}\right)$, and m is the largest positive integer for which $\mu_i + m(a-1) > 0$ when $a < 1$. The mean is also assumed to be equal to $E(Y_i \mid \mathbf{x_i}) = \mu_i = e_i \exp(\mathbf{x_i^T} \boldsymbol{\beta})$. However, the conditional variance is equivalent to $Var(Y_i \mid \mathbf{x_i}) = a^2 \mu_i$.

When the dispersion parameter, a, is equal to one, the probability density function of GPII, which is shown by Equation (9), reduces to the Poisson so that the mean is equal to the variance. For $a > 1$, the variance is larger than the mean, and for this situation, the regression model represents count data with overdispersion. For $\frac{1}{2} \leq a < 1$ and $\mu_i > 2$, the variance is smaller than the mean, so that now the regression model represents count data with underdispersion.

If $\boldsymbol{\beta}$ is estimated by the maximum likelihood method, the related equations are,

$$\ell(\boldsymbol{\beta}, a) = \sum_i \log(\mu_i) + (y_i - 1)\log(u_i + (a-1)y_i) - y_i \log(a) - \frac{\mu_i + (a-1)y_i}{a} \tag{10}$$
$$-\log(y_i!),$$

$$\frac{\partial \ell}{\partial \beta_j} = \sum_i \left(\frac{1}{\mu_i} - \frac{1}{a} + \frac{y_i - 1}{\mu_i + (a-1)y_i} \right) \frac{\partial \mu_i}{\partial \beta_j} = 0, \qquad j = 1, 2, \ldots, p. \tag{11}$$

The maximum likelihood estimates of $\boldsymbol{\beta}$ are numerically difficult to be solved because Equation (11) is not equal to the weighted least squares. Since the GPII distribution has a constant variance-mean ratio, the method of weighted least squares is suggested, i.e., by equating,

$$\sum_i \frac{(y_i - \mu_i)}{a^2 \mu_i} \frac{\partial \mu_i}{\partial \beta_j} = 0, \qquad j = 1, 2, \ldots, p. \tag{12}$$

The same IWLS procedure of the Poisson can also be used to solve for $\boldsymbol{\beta}$ because Equation (12) is equivalent to the likelihood equation of the Poisson which is shown by Equation (2). As a result, the least squares estimates of $\boldsymbol{\beta}$ for GPII are also equal to the maximum likelihood estimates of Poisson, but the standard errors could be larger or smaller than the Poisson because they are multiplied by a where $a \geq 1$ or $\frac{1}{2} \leq a < 1$.

For simplicity, the estimate of a may be obtained by using the method of moments, i.e., by equating the Pearson chi-squares with the degrees of freedom,

$$\sum_i \frac{(y_i - \mu_i)^2}{a^2 \mu_i} = n - p. \tag{13}$$

An example of S-PLUS programming for solving estimates of $\boldsymbol{\beta}$ and a for GPI(moment) is available in Ismail and Jemain [20]. Similar programming can also be used for the Poisson, GPI(MLE) and GPII.

4.0 MODEL EVALUATION

The goodness-of-fit of the models may be measured by several statistical criteria; some of them are discussed briefly below.

4.1 Pearson Chi-squares

The Pearson chi-squares is equal to $\sum_i \frac{(y_i - \mu_i)^2}{Var(Y_i)}$. For an adequate model, the Pearson chi-squares has an asymptotic chi-squares distribution with $n - p$ degrees of freedom.

4.2 Deviance

The deviance is equivalent to $2\left(\ell(\mathbf{y};\mathbf{y}) - \ell(\boldsymbol{\mu};\mathbf{y})\right)$, where $\ell(\boldsymbol{\mu};\mathbf{y})$ and $\ell(\mathbf{y};\mathbf{y})$ are the log likelihoods evaluated under $\boldsymbol{\mu}$ and $\mathbf{y}$, respectively. For an adequate model, the deviance also has an asymptotic chi-squares distribution with $n - p$ degrees of freedom. Thus, if the values for both Pearson chi-squares and deviance are close to the degrees of freedom, $n - p$, the model may be considered as adequate.

The deviance could also be used to compare between two nested models, one of which is a simplified version of the other. Let D_1 and df_1 be the deviance and degrees of freedom for such model, and D_2 and df_2 be the values by fitting a simplified version of the model. The chi-squares is $\frac{D_2 - D_1}{df_2 - df_1}$ and it should be compared to a chi-squares distribution with $df_2 - df_1$degrees of freedom.

4.3 Likelihood Ratio Test

The advantage of using the maximum likelihood method is that the likelihood ratio test can be implemented to assess the adequacy of the GPI(MLE) over the Poisson because the GPI(MLE) will reduce to the Poisson when the dispersion parameter, a, is equal to zero.

For testing Poisson against GPI(MLE), the hypothesis can be stated as H_0: $a = 0$ against H_1: $a \neq 0$. The likelihood ratio is $T = 2(\ell_1 - \ell_0)$, where ℓ_1 and ℓ_0 are the model's log likelihood under the respective hypothesis. Under null hypothesis, T has an asymptotic chi-squares distribution with one degree of freedom (see [17]).

4.4 Other Tests

When several maximum likelihood models are available, one can also compare the performance of alternative models based on several likelihood measures. One commonly used measure is the Akaike information criteria [21] which is defined as $AIC = -\ell + p$, where ℓ denotes the log likelihood evaluated under $\boldsymbol{\mu}$ and p the number of parameters. For this measure, the smaller the AIC, the better the model is.

Another goodness-of-fit measure is by using the Bayesian-Schwartz criteria [22] which is defined as $BSC = \ell - p \log\left(\frac{n}{2\pi}\right)$, where ℓ denotes the log likelihood evaluated under $\boldsymbol{\mu}$, p is the number of parameters and n is the number of rating classes. For this measure, the larger the BCS, the better the model is.

5.0 RESULTS AND DISCUSSION

In Malaysian practice, motor insurance claims can give rise to multiple types such as Own Damage (OD), Third Party Property Damage (TPPD) and Third Party Bodily Injury (TPBI). For practical and statistical reasons, the claims are strongly suggested to be treated separately. In this paper, risk classification is carried out on two types of motor insurance claims count data; OD and TPBI. Specifically, the OD claims are the claims for loss or damage to the insured vehicle, and the TPBI claims are the claims for death or bodily injury to any person.

The claims count data, which was based on 170,000 private car policies for a three-year-period of 1998 to 2000, was supplied by the Persatuan Insurans Am Malaysia (PIAM). The data, which enclosed the exposures and number of incurred claims for each claim type, also contained information for the rating factors and rating classes. The rating factors and rating classes for each claim type are summarized in Table 1.

The OD claims have only four rating factors because the claims may only occur in comprehensive coverage. Therefore, the total number of cross-classified rating classes is $2 \times 3 \times 4 \times 5 = 120$. For TPBI claims, the total number of cross-classified rating classes is $2 \times 2 \times 3 \times 4 \times 5 = 240$.

The rating factor for use-gender represents vehicles which are used for private and business purposes. The vehicles used for private purposes are further classified by the driver's gender. However, for vehicles used for business purposes, the driver's gender is not provided.

The rating factor for location corresponds to the postcode written in the driver's policy. Specifically, the location for Central are represented by Kuala Lumpur and Selangor, North by Perlis, Kedah, Pulau Pinang and Perak, East by Terengganu, Kelantan and Pahang, South by Negeri Sembilan, Melaka and Johor, and East Malaysia by Sabah and Sarawak.

5.1 Own Damage Claims (OD)

The claim counts were first fitted to the Poisson regression model. Several models were fitted by including different rating factors; first the main effects only, then the main effects plus each of the paired interactions. By using the deviance and degrees of freedom, the chi-squares statistics were calculated and compared to choose the

Table 1 Rating factors and rating classes

OD claims		TPBI claims	
Rating factors	**Rating classes**	**Rating factors**	**Rating classes**
Vehicle make	Local Foreign	Coverage	Comprehensive Non-comprehensive
Use-gender	Private-male Private-female Business	Vehicle make	Local Foreign
Vehicle year	0-1 year 2-3 years 4-5 years 6+ years	Use-gender	Private-male Private-female Business
Location	Central North East South East Malaysia	Vehicle year	0-1 year 2-3 years 4-5 years 6+ years
		Location	Central North East South East Malaysia

best model. The best model suggests that two of the rating factors, i.e., use-gender and vehicle year, are significant and none of the paired interaction is significant. The two-factor model was then fitted to the regression models of GPI(MLE), GPI(moment) and GPII. The results are summarized in Table 2.

The regression parameters for all models give similar estimates. The Poisson, GPI(MLE) and GPII give almost similar inferences about their regression parameters, i.e., their standard errors are almost similar. On the contrary, the GPI(moment) gives a relatively large values for standard errors, and hence, resulted in an insignificant regression parameter for β_6.

The Pearson chi-squares and deviance were reduced significantly if the GPI(MLE) or GPI(moment) were fitted. In particular, the GPI(moment) gives the smallest values for Pearson chi-squares and deviance, and the GPI(MLE) gives the largest value for log likelihood. However, this result is to be expected because the estimation of a in GPI(moment) was carried out by equating the Pearson chi-squares with the degrees of freedom, whereas in GPI(MLE), a was estimated by maximizing the log likelihood.

The deviance for the Poisson is relatively larger than the degrees of freedom, i.e., 2.17 times larger, indicating possible existence of overdispersion. To test for

Table 2 Results for two-factor model

Parameter	Poisson			GPI(MLE)			GPI(moment)			GPII		
	Est.	**Std. error**	***p*-value**	**Est.**	**Std. error**	***p*-value**	**Est.**	**Std. error**	***p*-value**	**Est.**	**Std. error**	***p*-value**
a	–	–	–	0.004	–	–	0.023	–	–	1.577	–	–
β_1 (Intercept)	–2.86	0.03	0.00	–2.83	0.04	0.00	–2.80	0.08	0.00	–2.86	0.05	0.00
β_2 (Female)	–0.30	0.02	0.00	–0.32	0.04	0.00	–0.34	0.08	0.00	–0.30	0.04	0.00
β_3 (Business)	–1.46	0.08	0.00	–1.46	0.09	0.00	1.40	0.11	0.00	–1.46	0.13	0.00
β_4 (2-3 years)	0.29	0.04	0.00	0.27	0.05	0.00	0.19	0.10	0.06	0.29	0.06	0.00
β_5 (4-5 years)	0.28	0.04	0.00	0.30	0.05	0.00	0.31	0.10	0.00	0.28	0.06	0.00
β_6 (6+ years)	0.13	0.04	0.00	0.13	0.05	0.00	0.13	0.10	0.20	0.13	0.06	0.02
Degrees of freedom		114			113			113			113	
Pearson chi-squares		283.57			198.16			114.00			–	
Deviance		247.18			166.69			93.42			–	
Log L		–420.23			–407.67			–432.58			–	
Likelihood ratio		–			25.12			–			–	
AIC		426.23			414.67			–			–	
BSC		–437.93			–428.31			–				

overdispersion, the likelihood ratio test of Poisson against GPI(MLE) is implemented. The likelihood ratio of 25.12 is significant, implying that the GPI(MLE) is a better model. For further comparison, the results of *AIC* and *BSC* also indicate that the GPI(MLE) is a better model compared to the Poisson.

To find the best model for GPI(MLE), several models are also fitted and compared, including the main effects only, then the main effects plus each of the paired interactions. Based on the deviance and degrees of freedom, the best model suggests that none of the rating factors are significant. Therefore, if the claim counts were based on GPI(MLE), all insureds are proposed to be classified under only one rating class. The results for the null-factor model are summarized in Table 3.

Comparison between observed and fitted frequency rates for the best regression model of Poisson and GPI(MLE) is given in Appendix 1. As a conclusion, the estimates of $\boldsymbol{\beta}$ do not change drastically if the Poisson is used. However, their standard errors do, and consequently, the Poisson overstates the significance of the rating factors. For instance, the Poisson regression model for OD claim counts data indicates that two of the rating factors are significant. This result contradicts with the result of GPI(MLE), which implies that none of the rating factor is significant.

Table 3 Results for GPI(MLE) null-factor model

Parameter	**Estimate**	**Std. error**	***p*-value**
a	0.049	–	–
β_1 (Intercept)	–3.04	0.05	0.00
Degrees of freedom		118	
Pearson chi-squares		119.10	
Deviance		160.83	
Log L		–509.55	

5.2 Third Party Bodily Injury Claims (TPBI)

The claim counts were first fitted to the Poisson regression model. The fitting involves only 221 data points because nineteen of the rating classes have zero exposures. Again, several models were fitted by including different rating factors, first the main effects only, then the main effects plus each of the paired interactions. By using the deviance and degrees of freedom, the chi-squares statistics were calculated and compared to choose the best model. The best model suggests that only three rating factors, i.e., coverage, use-gender and location, are significant and none of the paired interaction is significant. The three-factor model was then fitted to the models of GPI(MLE), GPI(moment) and GPII. The results are summarized in Table 4.

All models give similar values for parameter estimates and similar inferences about the regression parameters.

Table 4 Results for three-factor model

Parameter	Poisson			GPI(MLE)			GPI(moment)			GPII		
	Est.	**Std. error**	***p*-value**	**Est.**	**Std. error**	***p*-value**	**Est.**	**Std. error**	***p*-value**	**Est.**	**Std. error**	***p*-value**
a	–	–	–	0.006	–	–	0.024	–	–	1.094	–	–
β_1 (Intercept)	–3.55	0.05	0.00	–3.54	0.06	0.00	–3.53	0.09	0.00	–3.55	0.06	0.00
β_2 (NonComp)	–0.75	0.11	0.00	–0.78	0.12	0.00	–0.83	0.14	0.00	–0.75	0.12	0.00
β_3 (Female)	–0.30	0.07	0.00	–0.30	0.08	0.00	–0.31	0.09	0.00	–0.30	0.08	0.00
β_4 (Business)	–3.43	0.58	0.00	–3.44	0.58	0.00	–3.45	0.58	0.00	–3.43	0.63	0.00
β_5 (North)	0.41	0.07	0.00	0.41	0.08	0.00	0.42	0.11	0.00	0.41	0.07	0.00
β_6 (East)	0.36	0.11	0.00	0.36	0.12	0.00	0.34	0.15	0.02	0.36	0.12	0.00
β_7 (South)	0.23	0.08	0.00	0.24	0.09	0.01	0.27	0.12	0.03	0.23	0.09	0.01
β_8 (EastMsia)	–0.68	0.16	0.00	–0.68	0.17	0.00	–0.67	0.19	0.00	–0.68	0.18	0.00
Degrees of freedom		213			212			212			212	
Pearson chi-squares		255.09			240.16			213.00			–	
Deviance		168.47			153.89			128.89			–	
Log *L*		–266.60			–265.66			–269.31			–	
Likelihood ratio		–			1.88			–			–	
AIC		274.60			274.66			–			–	
BSC		–295.08			–297.70			–				

The Pearson chi-squares and deviance were slightly reduced if the GPI(MLE) or GPI(moment) were fitted. As expected, the GPI(moment) gives the smallest values for Pearson chi-squares and deviance, and the GPI(MLE) gives the largest value for log likelihood.

The Poisson gives a small deviance, indicating that the model is adequate. To test for overdispersion, the likelihood ratio test of Poisson against GPI(MLE) is implemented. The likelihood ratio of 1.88 is insignificant, implying that the Poisson is adequate. On the contrary, the *AIC* and *BSC* values for Poisson and GPI(MLE) are almost similar, indicating that there is not much difference between choosing either the Poisson or the GPI(MLE).

Even though the *AIC* and *BSC* values indicate that there is no difference between choosing the Poisson or the GPI(MLE), the actuaries or practitioners may choose their models based on the results of profitability analysis. The profitability analysis, which is also similar to the sensitivity analysis, allows the actuarial management to check whether the model assumptions agree or disagree with the actual experience by translating the difference between actual experience and model assumptions into effects on profits. However, further discussion on profitability analysis will not be included here because it is outside the scope of this study.

Comparison between observed and fitted frequency rates for the best regression model of Poisson is given in Appendix (Table 1).

6.0 CONCLUSION

This paper proposed the Generalized Poisson regression model as an alternative for risk classification. Even though the Poisson regression model has been widely used for risk classification in the recent years, this paper has shown that for an overdispersed claim data, the Generalized Poisson is superior to the Poisson. It is suggested that the discussions and results from this paper would also encourage similar analysis by the practitioners, especially those involved in the rating of premium for their insurance companies, towards contributing to a more accurate measure of claim frequency rates, and ultimately a "fair" premium rates for all.

ACKNOWLEDGEMENTS

This research was supported by IRPA RMK8 (Code: 09-02-02-0112-EA274). We are also grateful to Persatuan Insurans Am Malaysia (PIAM), in particular Mr. Carl Rajendram and Mrs. Addiwiyah, for supplying the data.

REFERENCES

[1] Hossack, I. B., J. H. Pollard, and B. Zehnwirth. 1983. *Introductory Statistics with Applications in General Insurance.* Great Britain: Cambridge University Press.

[2] Bailey, R. A., and L. J. Simon. 1960. Two Studies in Automobile Insurance Ratemaking. *ASTIN Bulletin*. 4(1): 192-217.

[3] Bailey, R. A. 1963. Insurance Rates with Minimum Bias. Proceedings of the Casualty Actuarial Society. 50(93): 4-14.

[4] Jung, J. 1968. On Automobile Insurance Ratemaking. *ASTIN Bulletin*. 5(1): 41-48.

[5] Ajne, B. 1975. A Note on the Multiplicative Ratemaking Model. *ASTIN Bulletin*. 8(2): 144-153.

[6] Chamberlain, C. 1980. Relativity Pricing Through Analysis of Variance. *Casualty Actuarial Society Discussion Paper Program*. 4-24.

[7] Coutts, S. M. 1984. Motor Insurance Rating, an Actuarial Approach. *Journal of the Institute of Actuaries*. 111: 87-148.

[8] Harrington, S. E. 1986. Estimation and Testing for Functional Form in Pure Premium Regression Models. *ASTIN Bulletin*. 16: 31-43.

[9] Brown, R. L. 1988. Minimum Bias with Generalized Linear Models. Proceedings of the Casualty Actuarial Society. 75(143): 187-217.

[10] Mildenhall, S. J. 1999. A Systematic Relationship between Minimum Bias and Generalized Linear Models. Proceedings of the Casualty Actuarial Society. 86(164): 93-487.

[11] Fu, L., and C. P. Wu. 2005. Generalized Minimum Bias Models. Casualty Actuarial Society Forum. Winter: 72-121.

[12] Ismail, N., and A. A. Jemain. 2005. Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation. Casualty Actuarial Society Forum. Spring: 367-394.

[13] McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2nd Edition. London: Chapman and Hall.

[14] Brockman, M. J., and T. S. Wright. 1992. Statistical Motor Rating: Making Effective Use of Your Data. *Journal of the Institute of Actuaries*. 119(3): 457-543.

[15] Renshaw, A. E. 1994. Modeling the Claims Process in the Presence of Covariates. *ASTIN Bulletin*. 24(2): 265-285.

[16] Ismail, N., and A. A. Jemain. 2003. Modeling Frequency of Auto Insurance Claims with Overdispersion Correction. Prosiding Seminar Sains Aktuari Matematik Kewangan. Bangi: Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia. 23-35.

[17] Wang, W., and F. Famoye. 1997. Modeling Household Fertility Decisions with Generalized Poisson Regression. *Journal of Population Economics*. 10: 273-283.

[18] Breslow, N. E. 1984. Extra-Poisson Variation in Log-linear Models. *Journal of the Royal Statistical Society (Applied Statistics)*. 33(1): 38-44.

[19] Consul, P. C., and F. Famoye. 1992. Generalized Poisson Regression Model. *Communication Statistics (Theory & Methodology)*. 2(1): 89-109.

[20] Ismail, N., and A. A. Jemain. 2004. Modeling Insurance Claim Frequency with Generalized Poisson Regression. Prosiding Simposium Kebangsaan Sains Matematik ke-12. Kuala Lumpur: Department of Computational and Theoretical Sciences, International Islamic University Malaysia. 1-17.

[21] Akaike, H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. Second International Symposium on Inference Theory. Budapest: Akademiai Kiado. 267-281.

[22] Schwartz, G. 1978. Estimating the Dimension of a Model. *Annals of Statistics*. 6: 461-464.

APPENDIX

Table 1 Observed and fitted rates for comprehensive coverage and local vehicle make

Use-gender	Vehicle year	Location	OD				TPBI		
			Exposure	Observed	Fitted Poisson	Fitted GPI (MLE)	Exposure	Observed	Fitted Poisson
Male	0-1 yr	Central	4243	0.061	0.058	0.048	1447	0.040	0.029
		North	2567	0.045	0.058	0.048	1027	0.040	0.043
		East	598	0.080	0.058	0.048	178	0.034	0.041
		South	1281	0.062	0.058	0.048	557	0.038	0.036
		EastMsia	219	0.041	0.058	0.048	117	0.000	0.015
	2-3 yr	Central	6926	0.080	0.077	0.048	2560	0.024	0.029
		North	4896	0.077	0.077	0.048	441	0.038	0.043
		East	1123	0.098	0.077	0.048	319	0.038	0.041
		South	2865	0.083	0.077	0.048	898	0.032	0.036
		EastMsia	679	0.081	0.077	0.048	177	0.011	0.015
	4-5 yr	Central	6286	0.075	0.076	0.048	1310	0.033	0.029
		North	4125	0.071	0.076	0.048	880	0.038	0.043
		East	1152	0.098	0.076	0.048	247	0.036	0.041
		South	2675	0.076	0.076	0.048	559	0.041	0.036
		EastMsia	700	0.081	0.076	0.048	140	0.007	0.015
	6+ yr	Central	6905	0.063	0.066	0.048	1442	0.028	0.029
		North	5784	0.067	0.066	0.048	1252	0.051	0.043
		East	2156	0.064	0.066	0.048	439	0.050	0.041
		South	3310	0.063	0.066	0.048	615	0.033	0.036
		EastMsia	1406	0.080	0.066	0.048	261	0.023	0.015
Female	0-1 yr	Central	2025	0.048	0.043	0.048	756	0.024	0.021
		North	1635	0.035	0.043	0.048	544	0.033	0.032
		East	301	0.053	0.043	0.048	115	0.035	0.031
		South	608	0.046	0.043	0.048	222	0.023	0.027
		EastMsia	126	0.032	0.043	0.048	51	0.000	0.011
	2-3 yr	Central	3661	0.061	0.057	0.048	1373	0.017	0.021
		North	2619	0.066	0.057	0.048	672	0.033	0.032
		East	527	0.046	0.057	0.048	135	0.044	0.031
		South	1192	0.059	0.057	0.048	351	0.020	0.027
		EastMsia	359	0.064	0.057	0.048	85	0.024	0.011
	4-5 yr	Central	2939	0.048	0.057	0.048	425	0.033	0.021
		North	1927	0.050	0.057	0.048	320	0.034	0.032
		East	439	0.043	0.057	0.048	81	0.012	0.031
		South	959	0.052	0.057	0.048	151	0.026	0.027
		EastMsia	376	0.040	0.057	0.048	74	0.000	0.011

cont.

Table 1 continued

Use-gender	Vehicle year	Location	OD				TPBI		
			Exposure	Observed	Fitted Poisson	Fitted GPI (MLE)	Exposure	Observed	Fitted Poisson
	6+ yr	Central	2215	0.038	0.049	0.048	427	0.012	0.021
		North	1989	0.051	0.049	0.048	403	0.040	0.032
		East	581	0.046	0.049	0.048	121	0.050	0.031
		South	937	0.046	0.049	0.048	155	0.026	0.027
		EastMsia	589	0.044	0.049	0.048	94	0.000	0.011
Business	0-1 yr	Central	290	0.021	0.013	0.048	127	0.000	0.001
		North	66	0.030	0.013	0.048	38	0.000	0.001
		East	24	0.000	0.013	0.048	7	0.000	0.001
		South	52	0.038	0.013	0.048	19	0.000	0.001
		EastMsia	6	0.000	0.013	0.048	4	0.000	0.000
	2-3 yr	Central	572	0.012	0.018	0.048	203	0.000	0.001
		North	148	0.034	0.018	0.048	42	0.000	0.001
		East	40	0.000	0.018	0.048	15	0.000	0.001
		South	91	0.033	0.018	0.048	18	0.000	0.001
		EastMsia	17	0.000	0.018	0.048	7	0.000	0.000
	4-5 yr	Central	487	0.023	0.018	0.048	145	0.000	0.001
		North	100	0.020	0.018	0.048	28	0.000	0.001
		East	40	0.100	0.018	0.048	8	0.000	0.001
		South	59	0.068	0.018	0.048	13	0.000	0.001
		EastMsia	22	0.091	0.018	0.048	7	0.000	0.000
	6+ yr	Central	468	0.019	0.015	0.048	121	0.000	0.001
		North	93	0.075	0.015	0.048	24	0.000	0.001
		East	33	0.061	0.015	0.048	5	0.042	0.001
		South	77	0.013	0.015	0.048	20	0.000	0.001
		EastMsia	25	0.040	0.015	0.048	2	0.000	0.000