Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models

Noriszura Ismail and Abdul Aziz Jemain

Abstract
In actuarial literature, researchers suggested various statistical procedures to estimate the parameters in claim count or frequency model. In particular, the Poisson regression model, which is also known as the Generalized Linear Model (GLM) with Poisson error structure, has been widely used in the recent years. However, it is also recognized that the count or frequency data in insurance practice often display overdispersion, i.e., a situation where the variance of the response variable 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. This paper suggests the Negative Binomial and Generalized Poisson regression models as alternatives for handling overdispersion. If the Negative Binomial and Generalized Poisson regression models are fitted by the maximum likelihood method, the models are considered to be convenient and practical; they handle overdispersion, they allow the likelihood ratio and other standard maximum likelihood tests to be implemented, they have good properties, and they permit the fitting procedure to be carried out by using the Iterative Weighted Least Squares (IWLS) regression similar to those of the Poisson. In this paper, two types of regression model will be discussed and applied; multiplicative and additive. The multiplicative and additive regression models for Poisson, Negative Binomial and Generalized Poisson will be fitted, tested and compared on three different sets of claim frequency data; Malaysian private motor third party property damage data, ship damage incident data from McCullagh and Nelder, and data from Bailey and Simon on Canadian private automobile liability.

Keywords: Overdispersion; Negative Binomial; Generalized Poisson; Multiplicative; Additive; Maximum likelihood.

1. INTRODUCTION

In property and liability insurance, the determination of premium rates must fulfill four basic principles generally agreed among the actuaries; to calculate "fair" premium rates whereby high risk insureds should pay higher premium and vice versa, to provide sufficient funds for paying expected losses and expenses, to maintain adequate margin for adverse deviation, and to produce a reasonable return to the insurer. The process of establishing "fair" premium rates for insuring uncertain events requires estimates which were made of two important elements; the probabilities associated with the occurrence of such event, i.e., the frequency, and the magnitude of such event, i.e., the severity. The frequency and severity estimates were usually calculated through the use of past experience for groups of similar risk characteristics. The process of grouping risks with similar risk characteristics to establish "fair" premium rates in an insurance system is also known as risk classification. In this paper, risk classification will be applied to estimate claim frequency rate which is equivalent to the claim count per exposure unit.

In the last forty years, researchers suggested various statistical procedures to estimate the parameters in risk classification model. For example, Bailey and Simon [1] suggested the minimum chi-squares, Bailey [2] devised the zero bias, Jung [3] produced a heuristic method for minimum modified chi-squares, Ajne [4] proposed the method of moments also for minimum modified chi-squares, Chamberlain [5] used the weighted least squares, Coutts [6] produced the method of orthogonal weighted least squares with logit transformation, Harrington [7] suggested the maximum likelihood procedure for models with functional form, and Brown [8] proposed the bias and likelihood functions for minimum bias and maximum likelihood models.

In the recent actuarial literature, research on the estimation methods for risk classification model is still continuing and developing. For example, Mildenhall [9] merged the models which were introduced by Bailey and Simon, 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, Feldblum and Brosius [10] summarized the minimum bias procedure and provided intuition for several bias functions, which include zero bias, least squares, minimum chi-squares and maximum likelihood, for practicing actuary. Anderson *et al.* [11] provided foundation for GLMs statistical theory also for practicing actuary. Their study provided practical insights and realistic output for the analysis of GLMs. Fu and Wu [12] developed the models 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 Jemain [13] 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 more generalized form. They solved the parameters by using weighted equation, regression approach and Taylor series approximation.

Besides statistical procedures, research on multiplicative and additive models has also been carried out. Among the pioneer studies, Bailey and Simon [1] compared the systematic bias of multiplicative and additive models and found that the multiplicative model overestimates the high risk classes. Their result was later agreed by Jung [3] and Ajne [4] who also found that the estimates for multiplicative model are positively biased. Bailey [2] compared the multiplicative and additive models by producing two statistical criteria, namely, the minimum chi-squares and average absolute difference. In addition, he also suggested the multiplicative model for percents classes and additive model for cents classes. Freifelder [14] predicted the pattern of over and under estimation for multiplicative and additive models if true models were misspecified, Jee [15] compared the predictive accuracy of multiplicative and additive models, Brown [8] discussed and summarized the additive and multiplicative models which were derived from the maximum likelihood and minimum bias approaches, Holler *et al.* [16] compared the initial values sensitivity of multiplicative and additive models, Mildenhall [9] identified the Generalized Linear Models for identity and log link functions with the additive and multiplicative models which were discussed in Brown [8], and Ismail and Jemain [13] discussed and compared the parameter estimates and goodness-of-fit of the additive and multiplicative regression models.

In insurance practice, the Poisson regression model, which is also known as the Generalized Linear Model with Poisson error structure, has been widely used for modeling claim count or frequency data in the recent years. For example, Aitkin *et al.* [17] and Renshaw [18] each respectively fit the Poisson model to two different sets of U.K. motor claim count data. For insurance practitioners, the Poisson regression model has been considered as practical and convenient; besides allowing the statistical inference and hypothesis tests to be determined by statistical theories, the model also permits the fitting procedure to be carried out easily by using any statistical package containing a routine for the Iterative Weighted Least Squares (IWLS) regression.

However, at the same time it is also recognized that the count or frequency data in insurance practice often display overdispersion or extra-Poisson variation, a situation where the variance of the response variable 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.

Based on the actuarial literature, the Poisson quasi likelihood model has been suggested to accommodate overdispersion in claim count or frequency data. For example, McCullagh and Nelder [19], using the data provided by Lloyd's Register of Shipping, applied the quasi likelihood model for damage incidents caused to the forward section of cargo-carrying vessels, to allow for possible inter-ship variability in accident proneness. The same quasi likelihood model was also fitted to the count data of U.K. own damage motor claims by Brockman and Wright [20], to take into account the possibility of within-cell heterogeneity.

For insurance practitioners, the most likely reason for using Poisson quasi likelihood is that the model can still be fitted without knowing the exact probability function of the response variable, as long as the mean is specified to be equivalent to the mean of Poisson, and the variance can be written as a multiplicative constant of the mean. To account for overdispersion, the Poisson quasi likelihood produces parameter estimates equivalent to the Poisson, and standard errors larger than those of the Poisson.

On the contrary, the maximum likelihood approach suggested in this paper differs from the quasi likelihood approach such that it requires the complete probability of the response variable, thus, allowing the likelihood ratio and other standard maximum likelihood tests to be implemented. With this objective in mind, this paper suggests the Negative Binomial and Generalized Poisson regression models for handling overdispersion. If the Negative Binomial and Generalized Poisson were fitted by the maximum likelihood method, the models may also be considered as convenient and practical; they allow the likelihood ratio and other standard maximum likelihood tests to be implemented, they have good properties, they permit the fitting procedure to be carried out by using Iterative Weighted Least Squares (IWLS) regression similar to those of the Poisson, and last but not least, they handle overdispersion. In this paper, two types of regression models will be discussed and applied; multiplicative and additive models. Specifically, the multiplicative and additive regression models for Poisson, Negative Binomial and Generalized Poisson will be fitted, tested and compared on three different sets of claim frequency data; Malaysian private motor third party property damage data, ship damage incident data from McCullagh and Nelder [19], and data from Bailey and Simon [1] on Canadian private automobile liability.

2. MULTIPLICATIVE REGRESSION MODELS

2.1 Poisson

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

$$\Pr(Y_i = y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}, \quad y_i = 0,1,\ldots \tag{2.1}$$

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

To incorporate covariates and to ensure non-negativity, the mean or the fitted value is assumed to be multiplicative, i.e., $E(Y_i \mid \mathbf{x}_i) = \lambda_i = e_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, where e_i denotes a measure of exposure, $\mathbf{x}_i$ a $p \times 1$ vector of explanatory variables, and $\boldsymbol{\beta}$ a $p \times 1$ vector of regression parameters.

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

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} = \sum_i (y_i - \lambda_i) x_{ij} = 0, \quad j = 1,2,\ldots,p. \tag{2.2}$$

Since Eq.(2.2) is also equal to the weighted least squares, the maximum likelihood estimates, $\hat{\boldsymbol{\beta}}$, may be solved by using the Iterative Weighted Least Squares (IWLS) regression.

2.2Negative Binomial I

Under the Poisson, the mean, λ_i, is assumed to be constant or homogeneous within the classes. However, by defining a specific distribution for λ_i, heterogeneity within the classes is now allowed. For example, by assuming λ_i to be a Gamma with mean $E(\lambda_i) = \mu_i$ and variance $Var(\lambda_i) = \mu_i^2 v_i^{-1}$, and $Y_i \mid \lambda_i$ to be a Poisson with conditional mean $E(Y_i \mid \lambda_i) = \lambda_i$, it can be shown that the marginal distribution of Y_i follows a Negative Binomial distribution with probability density function,

$$\Pr(Y_i = y_i) = \int \Pr(Y_i = y_i \mid \lambda_i) f(\lambda_i) d\lambda_i = \frac{\Gamma(y_i + v_i)}{\Gamma(y_i + 1)\Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i}\right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i}\right)^{y_i}, \tag{2.3}$$

where the mean is $E(Y_i) = \mu_i$ and the variance is $Var(Y_i) = \mu_i + \mu_i^2 v_i^{-1}$.

Different parameterization can generate different types of Negative Binomial distributions. For example, by letting $v_i = a^{-1}$, Y_i follows a Negative Binomial distribution with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = \mu_i(1 + a\mu_i)$, where a denotes the dispersion parameter (see Lawless [21]; Cameron and Trivedi [22]).

If a equals zero, the mean and variance will be equal, $E(Y_i) = Var(Y_i)$, resulting the distribution to be a Poisson. If $a > 0$, the variance will exceed the mean, $Var(Y_i) > E(Y_i)$, and the distribution allows for overdispersion as well. In this paper, the distribution will be called as Negative Binomial I.

If it is assumed that the mean or the fitted value is multiplicative, i.e., $E(Y_i \mid \mathbf{x}_i) = \mu_i = e_i \exp(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta})$, the likelihood for Negative Binomial I regression model may be written as,

$$\ell(\boldsymbol{\beta}, a) = \sum_i \left\{ \sum_{r=1}^{y_i - 1} \log(1 + ar) \right\} - y_i \log(a) - \log(y_i!) + \\ y_i \log(a\mu_i) - (y_i + a^{-1}) \log(1 + a\mu_i). \tag{2.4}$$

Therefore, the maximum likelihood estimates, $(\hat{\boldsymbol{\beta}}, \hat{a})$, may be obtained by maximizing $\ell(\boldsymbol{\beta}, a)$ with respect to $\boldsymbol{\beta}$ and a. The related equations are,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i) x_{ij}}{1 + a\mu_i} = 0, \quad j = 1, 2, \ldots, p, \tag{2.5}$$

and,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial a} = \sum_i \left\{ \sum_{r=1}^{y_i - 1} \left(\frac{r}{1 + ar} \right) \right\} + a^{-2} \log(1 + a\mu_i) - \frac{(y_i + a^{-1})\mu_i}{(1 + a\mu_i)} = 0. \tag{2.6}$$

The maximum likelihood estimates, $(\hat{\boldsymbol{\beta}}, \hat{a})$, may be solved simultaneously, 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 equation is Eq.(2.5) which is also equivalent to the weighted least squares. Therefore, with a slight modification, this task can be performed by using the IWLS regression similar to those of the Poisson. 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 equation is Eq.(2.6), and the task can be carried out by using the Newton-Raphson iteration. By iterating and cycling between holding a fixed and holding $\boldsymbol{\beta}$ fixed, the maximum likelihood estimates, $(\hat{\boldsymbol{\beta}}, \hat{a})$, will be obtained. Further explanation on the fitting procedure will be discussed in Section 4.

An easier approach to estimate a is by using the moment estimation suggested by Breslow [23], i.e., by equating the Pearson chi-squares statistic with the degrees of freedom,

$$\sum_i \frac{(y_i - \mu_i)^2}{\mu_i (1 + a\mu_i)} = n - p, \tag{2.7}$$

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, $(\hat{\boldsymbol{\beta}}, \tilde{a})$.

In this paper, when a is estimated by the maximum likelihood, the model will be called as Negative Binomial I (MLE). Likewise, when a is estimated by the method of moment, the model will be called as Negative Binomial I (moment).

2.3 Negative Binomial II

By letting $v_i = \mu_i a^{-1}$, another type of Negative Binomial distribution is produced, this time with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = \mu_i(1+a)$ (see Nelder and Lee [24]; Cameron and Trivedi [22]). If a equals zero, the mean and variance will be equal, resulting the distribution to be a Poisson. If $a > 0$, the variance will exceed the mean and the distribution allows for overdispersion as well. In this paper, the distribution will be called as Negative Binomial II.

If it is assumed that the mean or the fitted value is multiplicative, i.e., $E(Y_i \mid \mathbf{x}_i) = \mu_i = e_i \exp(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta})$, the likelihood for Negative Binomial II regression model may be written as,

$$\begin{aligned} \ell(\boldsymbol{\beta}, a) = \sum_i \log(\Gamma(y_i + \mu_i a^{-1})) - \log(\Gamma(\mu_i a^{-1})) - \log(y_i!) - \\ \mu_i a^{-1} \log(a) - (y_i + \mu_i a^{-1}) \log(1 + a^{-1}). \end{aligned} \tag{2.8}$$

Therefore, the maximum likelihood estimates, $(\hat{\boldsymbol{\beta}}, \hat{a})$, may be obtained by maximizing $\ell(\boldsymbol{\beta}, a)$ with respect to $\boldsymbol{\beta}$ and a. The related equations are,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial \beta_j} = \sum_i \mu_i x_{ij} a^{-1} \left\{ \sum_{r=0}^{y_i - 1} (\mu_i a^{-1} + r)^{-1} - \log(1+a) \right\} = 0, \quad j = 1,2,\ldots,p, \tag{2.9}$$

and,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial a} = -\sum_i \mu_i a^{-2} \left\{ \sum_{r=0}^{y_i - 1} (\mu_i a^{-1} + r)^{-1} - \log(1+a) \right\} + \sum_i \frac{y_i - \mu_i}{(1+a)a} = 0. \tag{2.10}$$

However, the maximum likelihood estimates, $\hat{\boldsymbol{\beta}}$, are numerically difficult to be solved because the related equation, Eq.(2.9), is not equal to the weighted least squares. As an alternative, since the Negative Binomial II has a constant variance-mean ratio, the method of weighted least squares is suggested, i.e., by equating,

$$\sum_i \frac{y_i - \mu_i}{Var(Y_i)} \frac{\partial \mu_i}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{1+a} = 0, \quad j = 1,2,\ldots, p, \tag{2.11}$$

to produce the least squares estimates, $\tilde{\boldsymbol{\beta}}$.

It is shown that in the presence of a modest amount of overdispersion, the least squares estimates were highly efficient for the estimation of a moment parameter of an exponential family distribution (Cox [25]). Since Eq.(2.11) is also equivalent to the likelihood equation of the Poisson, i.e., Eq.(2.2), the same IWLS regression which is used for the Poisson can be applied to estimate the least squares estimates, $\tilde{\boldsymbol{\beta}}$. As a result, the least squares estimates are also equal to the maximum likelihood estimates of Poisson, but the standard errors are equal or larger than the Poisson because they are multiplied by $\sqrt{1+a}$ where $a \geq 0$.

For simplicity, a is suggested to be estimated by the method of moment, i.e., by equating the Pearson chi-squares statistic with the degrees of freedom,

$$\sum_i \frac{(y_i - \mu_i)^2}{(1+a)\mu_i} = n - p, \tag{2.12}$$

which involves a straightforward calculation and produces a moment estimate, $\tilde{a}$.

In this paper, the estimates which were produced by the multiplicative regression models of Negative Binomial I (MLE), Negative Binomial I (moment) and Negative Binomial II will be denoted respectively by $(\hat{\boldsymbol{\beta}}, \hat{a})$, $(\hat{\boldsymbol{\beta}}, \tilde{a})$ and $(\tilde{\boldsymbol{\beta}}, \tilde{a})$.

2.4 Generalized Poisson I

The advantage of using the Generalized Poisson distribution is that it can be fitted for both overdispersion, $Var(Y_i) > E(Y_i)$, as well as underdispersion, $Var(Y_i) < E(Y_i)$. In this paper, two different types of Generalized Poisson will be discussed; each will be referred to as Generalized Poisson I and Generalized Poisson II. For Generalized Poisson I distribution, the probability density function is (Wang and Famoye [26]),

$$\Pr(Y_i = y_i) = \left(\frac{\mu_i}{1+a\mu_i}\right)^{y_i} \frac{(1+ay_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\mu_i(1+ay_i)}{1+a\mu_i}\right), \quad y_i = 0,1,\ldots, \tag{2.13}$$

with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = \mu_i(1 + a\mu_i)^2$:

The Generalized Poisson I is a natural extension of the Poisson. If a equals zero, the Generalized Poisson I reduces to the Poisson, resulting $E(Y_i) = Var(Y_i)$. If $a > 0$, the variance is larger than the mean, $Var(Y_i) > E(Y_i)$, and the distribution represents count data with overdispersion. If $a < 0$, the variance is smaller than the mean, $Var(Y_i) < E(Y_i)$, so that now the distribution represents count data with underdispersion.

If it is assumed that the mean or the fitted value is multiplicative, i.e., $E(Y_i \mid \mathbf{x}_i) = \mu_i = e_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, the likelihood for Generalized Poisson I regression model may be written as,

$$\ell(\boldsymbol{\beta}, a) = \sum_i y_i \log\left(\frac{\mu_i}{1 + a\mu_i}\right) + (y_i - 1)\log(1 + ay_i) - \frac{\mu_i(1 + ay_i)}{1 + a\mu_i} - \log(y_i!) . \quad (2.14)$$

Therefore, the maximum likelihood estimates, $(\hat{\boldsymbol{\beta}}, \hat{a})$, may be obtained by maximizing $\ell(\boldsymbol{\beta}, a)$ with respect to $\boldsymbol{\beta}$ and a. The related equations are,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{(1 + a\mu_i)^2} = 0, \quad j = 1,2,\dots, p, \quad (2.15)$$

and,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial a} = \sum_i -\frac{y_i \mu_i}{1 + a\mu_i} + \frac{y_i(y_i - 1)}{1 + ay_i} - \frac{\mu_i(y_i - \mu_i)}{(1 + a\mu_i)^2} = 0 . \quad (2.16)$$

The sequential iteration procedure similar to the Negative Binomial I regression model may also be implemented to obtain the maximum likelihood estimates, $(\hat{\boldsymbol{\beta}}, \hat{a})$. For the sequential iteration, the IWLS regression can be applied because Eq.(2.15) is also equal to the weighted least squares.

An easier approach to estimate a is by using the moment estimation, i.e., by equating the Pearson chi-squares statistic with the degrees of freedom,

$$\sum_i \frac{(y_i - \mu_i)^2}{\mu_i(1 + a\mu_i)^2} = n - p, \quad (2.17)$$

producing $(\hat{\boldsymbol{\beta}}, \tilde{a})$.

In this paper, when a is estimated by the maximum likelihood, the model will be called as Generalized Poisson I (MLE). Likewise, when a is estimated by the method of moment, the model will be called as Generalized Poisson I (moment).

2.5 Generalized Poisson II

For Generalized Poisson II, the probability density function may be written in the form of (Consul and Famoye [27]),

$$\Pr(Y_i = y_i) = \begin{cases} \mu_i(\mu_i + (a-1)y_i)^{y_i-1} a^{-y_i} \dfrac{\exp(-a^{-1}(\mu_i + (a-1)y_i))}{y_i!}, & y_i = 0,1,\ldots \\ 0, & y_i > m, a < 1 \end{cases}, \quad (2.18)$$

where $a \geq \max(\frac{1}{2}, 1 - \frac{\mu_i}{4})$, and m the largest positive integer for which $\mu_i + m(a-1) > 0$ when $a < 1$. For this distribution, the mean is equal to $E(Y_i) = \mu_i$, whereas the variance is equivalent to $Var(Y_i) = a^2\mu_i$.

The Generalized Poisson II is also a natural extension of the Poisson. If a equals one, the Generalized Poisson II reduces to the Poisson. If $a > 1$, the variance is larger than the mean and the distribution represents count data with overdispersion. If $\frac{1}{2} \leq a < 1$ and $\mu_i > 2$, the variance is smaller than the mean so that now the distribution represents count data with underdispersion.

If it is assumed that the mean or the fitted value is multiplicative, i.e., $E(Y_i \mid \mathbf{x}_i) = \mu_i = e_i \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$, the likelihood for Generalized Poisson II regression model may be written as,

$$\ell(\boldsymbol{\beta}, a) = \sum_i \log(\mu_i) + (y_i - 1)\log(\mu_i + (a-1)y_i) - \\ y_i \log(a) - a^{-1}(\mu_i + (a-1)y_i) - \log(y_i!). \quad (2.19)$$

Therefore, the maximum likelihood estimates, $(\hat{\boldsymbol{\beta}}, \hat{a})$, may be obtained by maximizing $\ell(\boldsymbol{\beta}, a)$ with respect to $\boldsymbol{\beta}$ and a. The related equations are,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial \beta_j} = \sum_i \left\{ \mu_i^{-1} - a^{-1} + \frac{y_i - 1}{\mu_i + (a-1)y_i} \right\} \mu_i x_{ij}, \quad j = 1,2,\ldots,p, \quad (2.20)$$

and,

$$\frac{\partial \ell(\boldsymbol{\beta},a)}{\partial a}=\sum_i \frac{y_i(y_i-1)}{\mu_i+(a-1)y_i}-y_i a^{-1}+(\mu_i-y_i)a^{-2}=0. \tag{2.21}$$

However, the maximum likelihood estimates, $\hat{\boldsymbol{\beta}}$, are numerically difficult to be solved because the related equation, Eq.(2.20), is not equal to the weighted least squares. Since the Generalized Poisson II has a constant variance-mean ratio, the method of weighted least squares is suggested as an alternative, i.e., by equating,

$$\sum_i \frac{(y_i-\mu_i)x_{ij}}{a^2}=0, \quad j=1,2,\ldots,p, \tag{2.22}$$

to produce the least squares estimates, $\tilde{\boldsymbol{\beta}}$. The same Poisson IWLS regression may be used to estimate $\tilde{\boldsymbol{\beta}}$ because Eq.(2.22) is also equivalent to the Poisson likelihood equation, i.e., Eq.(2.2). As a result, the least squares estimates are also equal to the maximum likelihood estimates of Poisson. However, the standard errors could be equal, 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, a is suggested to be estimated by the method of moment, i.e., by equating the Pearson chi-squares statistic with the degrees of freedom,

$$\sum_i \frac{(y_i-\mu_i)^2}{a^2\mu_i}=n-p, \tag{2.23}$$

involving a straightforward calculation and producing a moment estimate, $\tilde{a}$.

In this paper, the estimates which were produced by the regression models of Generalized Poisson I (MLE), Generalized Poisson I (moment) and Generalized Poisson II will be denoted respectively by $(\hat{\boldsymbol{\beta}},\hat{a})$, $(\hat{\boldsymbol{\beta}},\tilde{a})$ and $(\tilde{\boldsymbol{\beta}},\tilde{a})$.

To summarize the multiplicative regression models which were discussed in this section, Table 1 shows the methods and equations for solving the estimates of $\boldsymbol{\beta}$ and a.

Table 1. Methods and equations for solving $\boldsymbol{\beta}$ and a in multiplicative regression models

Models	Estimation of β		Estimation of a	
	Method	Equation	Method	Equation
Poisson	Maximum Likelihood	$\sum_i (y_i - \mu_i)x_{ij} = 0$	-	-
NBI(MLE)	Maximum Likelihood	$\sum_i \frac{(y_i - \mu_i)x_{ij}}{1 + a\mu_i} = 0$	Maximum Likelihood	$\sum_i \left\{ \sum_{r=1}^{y_i - 1} \left(\frac{r}{1 + ar} \right) + a^{-2}\log(1 + a\mu_i) - \frac{(y_i + a^{-1})\mu_i}{(1 + a\mu_i)} \right\} = 0$
NBI(moment)	Maximum Likelihood	$\sum_i \frac{(y_i - \mu_i)x_{ij}}{1 + a\mu_i} = 0$	Moment	$\sum_i \left\{ \frac{(y_i - \mu_i)^2}{\mu_i(1 + a\mu_i)} \right\} - (n - p) = 0$
NBII	Weighted Least Squares	$\sum_i \frac{(y_i - \mu_i)x_{ij}}{1 + a} = 0$	Moment	$\sum_i \left\{ \frac{(y_i - \mu_i)^2}{\mu_i(1 + a)} \right\} - (n - p) = 0$
GPI(MLE)	Maximum Likelihood	$\sum_i \frac{(y_i - \mu_i)x_{ij}}{(1 + a\mu_i)^2} = 0$	Maximum Likelihood	$\sum_i \left\{ -\frac{y_i\mu_i}{1 + a\mu_i} + \frac{y_i(y_i - 1)}{1 + ay_i} - \frac{\mu_i(y_i - \mu_i)}{(1 + a\mu_i)^2} \right\} = 0$
GPI(moment)	Maximum Likelihood	$\sum_i \frac{(y_i - \mu_i)x_{ij}}{(1 + a\mu_i)^2} = 0$	Moment	$\sum_i \left\{ \frac{(y_i - \mu_i)^2}{\mu_i(1 + a\mu_i)^2} \right\} - (n - p) = 0$
GPII	Weighted Least Squares	$\sum_i \frac{(y_i - \mu_i)x_{ij}}{a^2} = 0$	Moment	$\sum_i \left\{ \frac{(y_i - \mu_i)^2}{\mu_i a^2} \right\} - (n - p) = 0$

3. GOODNESS-OF-FIT TESTS

In this section, several goodness-of-fit measures will be briefly discussed, including the Pearson chi-squares, deviance, likelihood ratio test, Akaike Information Criteria (AIC) and Bayesian Schwartz Criteria (BSC). Since these measures are already familiar to those who used the Generalized Linear Model with Poisson error structure for claim count or frequency analysis, the same measures may also be implemented to the regression models of Negative Binomial and Generalized Poisson as well.

3.1 Pearson chi-squares

Two of the most frequently used measures for goodness-of-fit in the Generalized Linear Models are the Pearson chi-squares and the deviance. The Pearson chi-squares statistic is equivalent to,

$$\sum_i \frac{(y_i - \mu_i)^2}{Var(Y_i)}. \tag{3.1}$$

For an adequate model, the statistic has an asymptotic chi-squares distribution with $n-p$ degrees of freedom, where n denotes the number of rating classes and p the number of parameters.

3.2 Deviance

The deviance is equal to,

$$D = 2(\ell(\mathbf{y};\mathbf{y}) - \ell(\boldsymbol{\mu};\mathbf{y})), \tag{3.2}$$

where $\ell(\boldsymbol{\mu};\mathbf{y})$ and $\ell(\mathbf{y};\mathbf{y})$ are the model's log likelihood evaluated respectively under $\boldsymbol{\mu}$ and $\mathbf{y}$. For an adequate model, D also has an asymptotic chi-squares distribution with $n-p$ degrees of freedom. Therefore, if the values for both Pearson chi-squares and D are close to the degrees of freedom, 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 same values by fitting a simplified version of the model. The chi-squares statistic is equal to $(D_2 - D_1)/(df_2 - df_1)$ and it should be compared to a chi-squares distribution with $df_2 - df_1$ degrees of freedom.

3.3 Likelihood ratio

The advantage of using the maximum likelihood method is that the likelihood ratio test may be employed to assess the adequacy of the Negative Binomial I (MLE) or the Generalized Poisson I (MLE) over the Poisson because both Negative Binomial I (MLE) and Generalized Poisson I (MLE) will reduce to the Poisson when the dispersion parameter, a, equals zero.

For testing Poisson against Negative Binomial I (MLE), the hypothesis may be stated as $H_0 : a = 0$ against $H_1 : a > 0$. The likelihood ratio statistic is,

$$T = 2(\ell_1 - \ell_0), \tag{3.3}$$

where ℓ_1 and ℓ_0 are the model's log likelihood under the respective hypothesis. T has an asymptotic distribution of probability mass of one-half at zero and one-half-chi-squares distribution with one degrees of freedom (see Lawless [21]; Cameron and Trivedi [22]). Therefore, to test the null hypothesis at the significance level of α, the critical value of chi-squares distribution with significance level 2α is used, i.e., reject H_0 if $T > \chi^2_{(1-2\alpha,1)}$.

For testing Poisson against Generalized Poisson I (MLE), the hypothesis may be stated as $H_0 : a = 0$ against $H_1 : a \neq 0$. The likelihood ratio is also equal to Eq.(3.3) and under null hypothesis, T has an asymptotic chi-squares distribution with one degrees of freedom (see Wang and Famoye [26]).

3.4 AIC and BIC

When several maximum likelihood models are available, one can compare the performance of alternative models based on several likelihood measures which have been proposed in the statistical literature. Two of the most regularly used measures are the Akaike Information Criteria (AIC) and the Bayesian Schwartz Information Criteria (BIC). The AIC is defined as (Akaike [28]),

$$\text{AIC} = -2\ell + 2p, \tag{3.4}$$

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.

The BIC is defined as (Schwartz [29]),

$$\text{BIC} = -2\ell + p\log(n), \tag{3.5}$$

where ℓ denotes the log likelihood evaluated under $\boldsymbol{\mu}$, p the number of parameters and n the number of rating classes. For this measure, the smaller the BIC, the better the model is.

4. FITTING PROCEDURE

As mentioned previously, the estimates of $\boldsymbol{\beta}$ and a for Negative Binomial I (MLE), Negative Binomial I (moment), Generalized Poisson I (MLE) and Generalized Poisson I (moment), may be solved simultaneously and the fitting procedure involves sequential iterations. The sequential iterations involve two steps of maximization in each sequence; maximizing $\ell(\boldsymbol{\beta}, a)$ with respect to $\boldsymbol{\beta}$ by holding a fixed, and maximizing $\ell(\boldsymbol{\beta}, a)$ with respect to a by holding $\boldsymbol{\beta}$ fixed.

4.1 Maximizing $\ell(\boldsymbol{\beta}, a)$ with respect to $\boldsymbol{\beta}$

By using the Newton-Rahpson iteration and the method of Scoring, the iterative equation in the standard form of IWLS regression may be written as,

$$\boldsymbol{\beta}_{(r)} = \boldsymbol{\beta}_{(r-1)} + \mathbf{I}_{(r-1)}^{-1}\mathbf{z}_{(r-1)}, \tag{4.1}$$

where $\boldsymbol{\beta}_{(r)}$ and $\boldsymbol{\beta}_{(r-1)}$ denote the vectors for $\boldsymbol{\beta}$ in the *r*th and *r*-1th iteration, $\mathbf{I}_{(r-1)}$ the information matrix containing negative expectation of the second derivatives of log likelihood evaluated at $\boldsymbol{\beta}_{(r-1)}$, and $\mathbf{z}_{(r-1)}$ the vector containing first derivatives of log likelihood evaluated at $\boldsymbol{\beta}_{(r-1)}$.

For an easier demonstration, an example for Poisson's IWLS regression will be shown and the notation for Poisson mean, λ_i, will be replaced by μ_i. The first derivatives of Poisson log likelihood, which is shown by Eq.(2.2), can also be written as,

$$\mathbf{z} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{k}, \tag{4.2}$$

where $\mathbf{X}$ denotes the matrix of explanatory variables, $\mathbf{W}$ the diagonal weight matrix whose *i*th diagonal element is,

$$w_i^P = \mu_i, \tag{4.3}$$

and $\mathbf{k}$ the vector whose *i*th row is equal to,

$$k_i = \frac{y_i - \mu_i}{\mu_i}. \tag{4.4}$$

The negative expectation of the second derivatives of Poisson log likelihood may be derived and it is equivalent to,

$$-E\left(\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_s}\right) = \sum_i \mu_i x_{ij} x_{is}, \quad j,s = 1,2,\ldots,p. \tag{4.5}$$

Therefore, the information matrix, $\mathbf{I}$, which contains negative expectation of the second derivatives of log likelihood, may be written as,

$$\mathbf{I} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}, \tag{4.6}$$

where the *i*th diagonal element of the weight matrix is also equal to Eq.(4.3).

Finally, the iterative equation shown by Eq.(4.1) may be rewritten as,

$$\boldsymbol{\beta}_{(r)} = \boldsymbol{\beta}_{(r-1)} + (\mathbf{X}^{\mathrm{T}}\mathbf{W}_{(r-1)}\mathbf{X})^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{W}_{(r-1)}\mathbf{k}_{(r-1)}). \tag{4.7}$$

It can be shown that with a slight modification in the weight matrix, the same iterative equation, i.e., Eq.(4.7), can also be used to obtain the maximum likelihood estimates, $\hat{\boldsymbol{\beta}}$, for Negative Binomial I and Generalized Poisson I as well.

The related equations for the first derivatives of log likelihood for Negative Binomial I and Generalized Poisson I are shown by Eq.(2.5) and Eq.(2.15). Both equations may also be written as Eq.(4.2), where the *i*th row of vector $\mathbf{k}$ is also the same as Eq.(4.4). However, the *i*th diagonal element of the weight matrix is,

$$w_i^{NBI} = \frac{\mu_i}{1 + a\mu_i}, \tag{4.8}$$

for Negative Binomial I and,

$$w_i^{GPI} = \frac{\mu_i}{(1+a\mu_i)^2}, \tag{4.9}$$

for Generalized Poisson I.

The negative expectation of the second derivatives of log likelihood for Negative Binomial I and Generalized Poisson I may be derived, and the respective equations are,

$$-E\left(\frac{\partial^2 \ell(\boldsymbol{\beta}, a)}{\partial \beta_j \partial \beta_s}\right) = \sum_i \frac{\mu_i x_{ij} x_{is}}{1+a\mu_i}, \quad j,s = 1,2,\ldots,p, \tag{4.10}$$

and,

$$-E\left(\frac{\partial^2 \ell(\boldsymbol{\beta}, a)}{\partial \beta_j \partial \beta_s}\right) = \sum_i \frac{\mu_i x_{ij} x_{is}}{(1+a\mu_i)^2}, \quad j,s = 1,2,\ldots,p. \tag{4.11}$$

Therefore, the information matrix may also be written as Eq.(4.6), where the *i*th diagonal element of the weight matrix for Negative Binomial I and Generalized Poisson I are respectively equal to Eq.(4.8) and Eq.(4.9).

The same iterative equation for the Poisson may also be used for Negative Binomial II and Generalized Poisson II because the weighted least squares equations, i.e., Eq.(2.11) and Eq.(2.22), are equivalent to the likelihood equations of the Poisson, i.e., Eq.(2.2).

The matrices and vectors for solving $\boldsymbol{\beta}$ in multiplicative regression models are summarized in Table 2.

Table 2. Matrices and vectors for solving $\boldsymbol{\beta}$ in multiplicative regression models

Models	Matrices and vectors for $\boldsymbol{\beta}_{(r)} = \boldsymbol{\beta}_{(r-1)} + \mathbf{I}_{(r-1)}^{-1}\mathbf{z}_{(r-1)}$, where $\mathbf{I}_{(r-1)} = \mathbf{X}^{\mathbf{T}}\mathbf{W}_{(r-1)}\mathbf{X}$, $\mathbf{z}_{(r-1)} = \mathbf{X}^{\mathbf{T}}\mathbf{W}_{(r-1)}\mathbf{k}_{(r-1)}$, *js*-th element of matrix $\mathbf{I} = i_{js} = -E\left(\frac{\partial^2 \ell}{\partial\beta_j\partial\beta_s}\right)$, *j*-th row of vector $\mathbf{z} = z_j = \frac{\partial \ell}{\partial \beta_j}$.	
Poisson/ NBII/ GPII	matrix $\mathbf{I}$	$i_{js} = \sum_i \mu_i x_{ij} x_{is} \rightarrow \mathbf{I} = \mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{X}$
	weight matrix $\mathbf{W}$	$w_i^P = \mu_i$
	vector $\mathbf{z}$	$z_j = \sum_i \mu_i \frac{(y_i - \mu_i)}{\mu_i} x_{ij} \rightarrow \mathbf{z} = \mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{k}$
	vector $\mathbf{k}$	$k_i = \frac{y_i - \mu_i}{\mu_i}$
NBI(MLE)/ NBI(moment)	matrix $\mathbf{I}$	$i_{js} = \sum_i \frac{\mu_i}{1 + a\mu_i} x_{ij} x_{is} \rightarrow \mathbf{I} = \mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{X}$
	weight matrix $\mathbf{W}$	$w_i^{NBI} = \frac{\mu_i}{1 + a\mu_i}$
	vector $\mathbf{z}$	$z_j = \sum_i \frac{\mu_i}{1 + a\mu_i} \frac{(y_i - \mu_i)}{\mu_i} x_{ij} \rightarrow \mathbf{z} = \mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{k}$
	vector $\mathbf{k}$	$k_i = \frac{y_i - \mu_i}{\mu_i}$
GPI(MLE)/ GPI(moment)	matrix $\mathbf{I}$	$i_{js} = \sum_i \frac{\mu_i}{(1 + a\mu_i)^2} x_{ij} x_{is} \rightarrow \mathbf{I} = \mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{X}$
	weight matrix $\mathbf{W}$	$w_i^{GPI} = \frac{\mu_i}{(1 + a\mu_i)^2}$
	vector $\mathbf{z}$	$z_j = \sum_i \frac{\mu_i}{(1 + a\mu_i)^2} \frac{(y_i - \mu_i)}{\mu_i} x_{ij} \rightarrow \mathbf{z} = \mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{k}$
	vector $\mathbf{k}$	$k_i = \frac{y_i - \mu_i}{\mu_i}$

4.2 Maximizing $\ell(\boldsymbol{\beta}, a)$ with respect to a

The maximization of $\ell(\boldsymbol{\beta}, a)$ with respect to a can be carried out by applying one-dimensional Newton-Raphson iteration,

$$a_{(r)} = a_{(r-1)} - \frac{f'(a_{(r-1)})}{f''(a_{(r-1)})}, \tag{4.12}$$

where f' denotes the first derivatives of function f and f'' the second derivatives of function f. The respective f' for Negative Binomial I (MLE), Negative Binomial I (moment), Generalized Poisson 1 (MLE) and Generalized Poisson I (moment) are Eq.(2.6), Eq.(2.7), Eq.(2.16) and Eq.(2.17).

The f'' equations for Negative Binomial I (MLE), Negative Binomial I (moment), Generalized Poisson I (MLE) and Generalized Poisson (moment) may be derived, and the respective equations are,

$$\sum_i \left\{ -\sum_{r=0}^{y_i-1} \left(\frac{r}{1+ar} \right)^2 - 2a^{-3}\log(1+a\mu_i) + \frac{2a^{-2}\mu_i}{1+a\mu_i} + \frac{(y_i + a^{-1})\mu_i^2}{(1+a\mu_i)} \right\}, \tag{4.13}$$

$$-\sum_i \frac{(y_i - \mu_i)^2}{(1+a\mu_i)^2}, \tag{4.14}$$

$$\sum_i \frac{y_i\mu_i^2}{(1+a\mu_i)^2} - \frac{y_i^2(y_i-1)}{(1+ay_i)^2} + \frac{2\mu_i^2(y_i-\mu_i)}{(1+a\mu_i)^3}, \tag{4.15}$$

and,

$$-2\sum_i \frac{(y_i - \mu_i)^2}{(1+a\mu_i)^3}. \tag{4.16}$$

The process of finding the moment estimate, $\tilde{a}$, for Negative Binomial II and Generalized Poisson II does not involve any iteration. The moment estimate can be obtained directly from Eq.(2.12) and Eq.(2.23).

The equations for solving a in multiplicative regression models are summarized in Table 3.

Table 3. Equations for solving a.

Models	Equations for $a_{(r)} = a_{(r-1)} - \frac{f'(a_{(r-1)})}{f''(a_{(r-1)})}$	
NBI(MLE)	$f'(a)$	$\sum_i \left\{ \sum_{r=0}^{y_i-1} \left(\frac{r}{1+ar} \right) + a^{-2} \log(1+a\mu_i) - \frac{(y_i + a^{-1})\mu_i}{(1+a\mu_i)} \right\}$
	$f''(a)$	$\sum_i \left\{ -\sum_{r=0}^{y_i-1} \left(\frac{r}{1+ar} \right)^2 - 2a^{-3} \log(1+a\mu_i) + \frac{2a^{-2}\mu_i}{1+a\mu_i} + \frac{(y_i + a^{-1})\mu_i^2}{(1+a\mu_i)} \right\}$
NBI(moment)	$f'(a)$	$\sum_i \frac{(y_i - \mu_i)^2}{\mu_i(1+a\mu_i)} - (n-p)$
	$f''(a)$	$-\sum_i \frac{(y_i - \mu_i)^2}{(1+a\mu_i)^2}$
NBII	Straightforward calculation	$a = \sum_i \frac{(y_i - \mu_i)^2}{\mu_i(n-p)} - 1$
GPI(MLE)	$f'(a)$	$\sum_i \left\{ -\frac{y_i\mu_i}{1+a\mu_i} + \frac{y_i(y_i-1)}{1+ay_i} - \frac{\mu_i(y_i - \mu_i)}{(1+a\mu_i)^2} \right\}$
	$f''(a)$	$\sum_i \frac{y_i\mu_i^2}{(1+a\mu_i)^2} - \frac{y_i^2(y_i-1)}{(1+ay_i)^2} + \frac{2\mu_i^2(y_i-\mu_i)}{(1+a\mu_i)^3}$
GPI(moment)	$f'(a)$	$\sum_i \frac{(y_i - \mu_i)^2}{\mu_i(1+a\mu_i)^2} - (n-p)$
	$f''(a)$	$-2\sum_i \frac{(y_i - \mu_i)^2}{(1+a\mu_i)^3}$
GPII	Straightforward calculation	$a = \sqrt{\sum_i \frac{(y_i - \mu_i)^2}{\mu_i(n-p)}}$

4.3 Restrictions on Generalized Poisson I

The iterative programming for Generalized Poisson I distribution should also allows for restrictions on a because the probability density function, Eq.(2.13), indicates that the value of a must satisfy both $1+a\mu_i > 0$ and $1+ay_i > 0$. Therefore, after obtaining estimate of a in each iteration, the program should check that when $a < 0$ (underdispersion), a must also fulfilled the condition for both $1+a\mu_i > 0$ and $1+ay_i > 0$.

In other words, for condition $1+a\mu_i > 0$, the program should checks if $a > -\frac{1}{\max(\mu_i)}$ is true. If this condition is not true, a new estimate for a is set as $-\frac{1}{\max(\mu_i)+1}$. A similar check is then implemented for $1+ay_i > 0$. Finally, if both conditions of $a > -\frac{1}{\max(\mu_i)}$ and $a > -\frac{1}{\max(y_i)}$ are not true, a new estimate for a is set as $\min(-\frac{1}{\max(\mu_i)+1}, -\frac{1}{\max(y_i)+1})$.

4.4 Variance-covariance matrix for $\hat{\boldsymbol{\beta}}$

The variance-covariance matrix, $Var(\hat{\boldsymbol{\beta}})$, for Negative Binomial I and Generalized Poisson I regression models is also equal to the variance-covariance matrix of Poisson regression model, i.e.,

$$Var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}. \tag{4.17}$$

However, the *i*th diagonal element of the weight matrix differs for each model, i.e., it is equal to Eq.(4.3) for Poisson, Eq.(4.8) for Negative Binomial I and Eq.(4.9) for Generalized Poisson I.

The variance-covariance matrix for Negative Binomial II and Generalized Poisson II is multiplied by a constant and they are equal to,

$$Var(\hat{\boldsymbol{\beta}}) = (1+a)(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}, \tag{4.18}$$

for Negative Binomial II and,

$$Var(\hat{\boldsymbol{\beta}}) = a^2(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}, \tag{4.19}$$

for Generalized Poisson II, where the *i*th diagonal element of the weight matrix is equal to the Poisson weight matrix, i.e., Eq.(4.3).

Examples of S-PLUS programming for Negative Binomial I (moment) and Generalized Poisson I (moment) are given in Appendix A and Appendix B. Similar programming can also be used for all of the multiplicative regression models which were discussed in this paper. Each programming is differentiated only by four distinguishable elements:

- Types of iteration.

  The sequential iterations are required for Negative Binomial I and Generalized Poisson I. For Poisson, Negative Binomial II and Generalized Poisson II, the standard iterations are adequate.
- Weight matrix.

  The weight matrix for Negative Binomial II and Generalized Poisson II is equal to the Poisson. Each of Negative Binomial I and Generalized Poisson I has its own weight matrix.
- Equation for estimating a.

  Each of Negative Binomial I (MLE), Negative Binomial I (moment), Generalized Poisson I (MLE) and Generalized Poisson I (moment) has its own equation for estimating a.
- Restriction on a.

  The restriction on a is required only in Generalized Poisson I.

5. EXAMPLES

5.1 Malaysian data

In this paper, the data for private car Third Party Property Damage (TPPD) claim frequencies from an insurance company in Malaysia will be considered. Specifically, the TPPD claim covers the legal liability for third party property loss or damage caused by or arising out of the use of an insured motor vehicle. The data, which was based on 170,000 private car policies for a three-year period of 1998-2000, was supplied by the General Insurance Association of Malaysia (PIAM). The exposure was expressed in terms of a car-year unit and the incurred claims consist of claims which were already paid as well as outstanding. Table 4 shows the rating factors and rating classes for the exposures and incurred claims, and altogether, there were $2 \times 2 \times 3 \times 4 \times 5 = 240$ cross-classified rating classes of claim frequencies to be estimated. The complete data, which contains the exposures, claim counts, rating factors and rating classes, is shown in Appendix C.

Table 4. Rating factors and rating classes for Malaysian data

Rating factors	Rating classes
Coverage type	Comprehensive Non-comprehensive
Vehicle make	Local Foreign
Vehicle use and driver's gender	Private-male Private-female Business
Vehicle year	0-1 year 2-3 year 4-5 year 6+ year
Location	Central North East South East Malaysia

The claim counts were first fitted to the Poisson multiplicative regression model. The fitting involves only 233 data points because seven of the rating classes have zero exposures. Several models were fitted by including different rating factors; first the main effects only, then the main effects plus each of the paired interaction factors. By using the deviance and degrees of freedom, the chi-squares statistics were calculated and compared to choose the best model. Table 5 gives the results of fitting several Poisson regression models to the count data.

Table 5. Analysis of deviance for Poisson

Model	deviance	df	Δdeviance	Δdf	χ^2	*p*-value
Null	2202	232	-	-	-	-
+ Coverage type	1924	231	278	1	278	0.00
+ Use-gender	997	229	927	2	464	0.00
+ Vehicle year	522	226	475	3	158	0.00
+ Vehicle location	369	222	153	4	38	0.00
+ Vehicle make	358	221	11	1	11	0.00
+ Vehicle make*vehicle year	255	218	103	3	34	0.00

Based on the deviance analysis, the best model indicates that all of the main effects are significant and one of the paired interaction factors, i.e., vehicle make and vehicle year, is also significant. Therefore, it is suggested that the rating factors for both vehicle make and vehicle year are combined to take into account the interaction between these two rating factors. The number of rating factors is now reduced from five to four. Table 6 shows the parameter estimates for the four-factor model.

Table 6. Parameter estimates for Poisson four-factor model

Parameter		estimate	std.error	p-value
β_1	Intercept	-2.37	0.04	0.00
β_2	Non-comprehensive	-0.68	0.07	0.00
β_3	Female	-0.51	0.03	0.00
β_4	Business	-6.04	1.00	0.00
β_5	Local, 2-3 year	-0.48	0.04	0.00
β_6	Local, 4-5 year	-0.82	0.05	0.00
β_7	Local, 6+ year	-1.06	0.05	0.00
β_8	Foreign, 0-1 year	-0.59	0.07	0.00
β_9	Foreign, 2-3 year	-0.68	0.05	0.00
β_{10}	Foreign, 4-5 year	-0.77	0.06	0.00
β_{11}	Foreign, 6+ year	-0.84	0.05	0.00
β_{12}	North	-0.22	0.03	0.00
β_{13}	East	-0.43	0.06	0.00
β_{14}	South	-0.01	0.04	0.78
β_{15}	East Malaysia	-0.50	0.06	0.00
Df		218.00		
Pearson χ^2		404.67		
Deviance		254.60		
Log L		-387.98		

The p-value for β_{14} (South) is equivalent to 0.78, and this value indicates that the parameter estimate is not significant. Therefore, the location for South is suggested to be combined with Central (Intercept) because both locations have almost similar risks. Table 7 shows the parameter estimates for the four-factor-combined-location model.

Table 7. Parameter estimates for Poisson four-factor-combined-location model

Parameter		estimate	std.error	p-value
β_1	Intercept	-2.37	0.03	0.00
β_2	Non-comprehensive	-0.68	0.07	0.00
β_3	Female	-0.51	0.03	0.00
β_4	Business	-6.04	1.00	0.00
β_5	Local, 2-3 year	-0.48	0.04	0.00
β_6	Local, 4-5 year	-0.82	0.05	0.00
β_7	Local, 6+ year	-1.06	0.05	0.00
β_8	Foreign, 0-1 year	-0.59	0.07	0.00
β_9	Foreign, 2-3 year	-0.68	0.05	0.00
β_{10}	Foreign, 4-5 year	-0.77	0.06	0.00
β_{11}	Foreign, 6+ year	-0.84	0.05	0.00
β_{12}	North	-0.22	0.03	0.00
β_{13}	East	-0.42	0.06	0.00
β_{14}	East Malaysia	-0.50	0.06	0.00
Df		219.00		
Pearson χ^2		404.47		
Deviance		254.67		
Log L		-388.02		

The result shows that all of the parameter estimates are significant. As a conclusion, based on the deviance analysis and parameter estimates, the best model is provided by the four-factor-combined-location model if the claim counts were fitted to the Poisson.

If the same four-factor-combined-location model was fitted to the multiplicative regression models of Negative Binomial and Generalized Poisson, the parameter estimates and standard errors may be compared. The comparisons are shown in Table 8 and Table 10.

Table 8. Poisson vs. Negative Binomial

Parameters		Poisson			Negative Binomial I (MLE)			Negative Binomial I (moment)			Negative Binomial II		
		est.	std. error	p-value	est.	std. error	p-value	est.	std. error	p-value	est.	std. error	p-value
a		-	-	-	0.02	-	-	0.15	-	-	0.85	-	-
β_1	Intercept	-2.37	0.03	0.00	-2.36	0.07	0.00	-2.37	0.15	0.00	-2.37	0.05	0.00
β_2	Non-comp	-0.68	0.07	0.00	-0.73	0.09	0.00	-0.79	0.13	0.00	-0.68	0.09	0.00
β_3	Female	-0.51	0.03	0.00	-0.54	0.05	0.00	-0.57	0.09	0.00	-0.51	0.04	0.00
β_4	Business	-6.04	1.00	0.00	-6.05	1.00	0.00	-6.06	1.00	0.00	-6.04	1.36	0.00
β_5	Local,2-3	-0.48	0.04	0.00	-0.51	0.09	0.00	-0.49	0.18	0.01	-0.48	0.06	0.00
β_6	Local,4-5	-0.82	0.05	0.00	-0.87	0.09	0.00	-0.87	0.19	0.00	-0.82	0.07	0.00
β_7	Local,6+	-1.06	0.05	0.00	-1.04	0.09	0.00	-0.98	0.18	0.00	-1.06	0.07	0.00
β_8	Foreign,0-1	-0.59	0.07	0.00	-0.62	0.10	0.00	-0.63	0.20	0.00	-0.59	0.09	0.00
β_9	Foreign,2-3	-0.68	0.05	0.00	-0.69	0.09	0.00	-0.65	0.19	0.00	-0.68	0.07	0.00
β_{10}	Foreign,4-5	-0.77	0.06	0.00	-0.76	0.10	0.00	-0.76	0.19	0.00	-0.77	0.08	0.00
β_{11}	Foreign,6+	-0.84	0.05	0.00	-0.81	0.09	0.00	-0.76	0.18	0.00	-0.84	0.07	0.00
β_{12}	North	-0.22	0.03	0.00	-0.16	0.06	0.00	-0.12	0.11	0.28	-0.22	0.04	0 00
β_{13}	East	-0.42	0.06	0.00	-0.43	0.08	0.00	-0.46	0.13	0.00	-0.42	0.08	0.00
β_{14}	EastM'sia	-0.50	0.06	0.00	-0.51	0.08	0.00	-0.49	0.13	0.00	-0.50	0.08	0.00
Df			219.00			218.00			218.00			218.00	
Pearson χ^2			404.47			293.71			219.00			-	
Deviance			254.67			155.99			90.72			-	
Log L			-388.02			-368.72			-391.64			-	

Table 8 shows the comparison between Poisson and Negative Binomial multiplicative regression models. The regression parameters for all models give similar estimates. The Negative Binomial I (MLE) and Negative Binomial II give similar inferences about the regression parameters, i.e., their standard errors are slightly larger than the Poisson's. However, the Negative Binomial I (moment) gives a relatively large values for the standard errors and hence, resulted in an insignificant regression parameter for β_{12}.

The deviance for Poisson regression model is relatively larger than the degrees of freedom, i.e., 1.16 times larger, and thus, indicating possible existence of overdispersion. To test for overdispersion, the likelihood ratio test of Poisson against Negative Binomial I (MLE) is implemented. The likelihood ratio statistic of $T = 38.6$ is significant, implying that the Negative Binomial I (MLE) is a better model. Further comparison can be made by using the results of likelihood ratio, AIC and BIC as shown in Table 9. Based on the likelihood ratio, AIC and BIC, the Negative Binomial I (MLE) is better than the Poisson.

Table 9. Likelihood ratio, AIC and BIC

Test/Criteria	Poisson	Negative Binomial I (MLE)
Likelihood ratio	-	38.6
AIC	804.0	767.4
BIC	809.2	773.0

Table 10 shows the comparison between Poisson and Generalized Poisson multiplicative regression models. Both Negative Binomial II and Generalized Poisson II give equal values for parameter estimates and standard errors. However, this result is to be expected because both regression models were fitted by using the same procedure.

The comparison between Poisson and Generalized Poisson also shows that the regression parameters for all models give similar estimates. The Generalized Poisson I (MLE) and Generalized Poisson II give similar inferences about the regression parameters. The Generalized Poisson I (moment) gives a relatively large values for the standard errors and this resulted in an insignificant regression parameter for β_{12}.

Based on the likelihood ratio test of Poisson against Generalized Poisson I (MLE), the likelihood ratio statistic of $T = 37.7$ is significant. Therefore, the Generalized Poisson (MLE) is also a better model compared to the Poisson.

Table 11 gives further comparison between Poisson and Generalized Poisson I (MLE). The comparison, which was based on the likelihood ratio, AIC and BIC, indicates that the Generalized Poisson I (MLE) is also a better model compared to the Poisson.

Table 10. Poisson vs. Generalized Poisson

Parameters		Poisson			Generalized Poisson I (MLE)			Generalized Poisson I (moment)			Generalized Poisson II		
		est.	std. error	*p*-value	est	std. error	*p*-value	est.	std. error	*p*-value	est.	std. error	*p*-value
a		-	-	-	0.007	-	-	0.035	-	-	1.359	-	-
β_1	Intercept	-2.37	0.03	0.00	-2.35	0.07	0.00	-2.37	0.16	0.00	-2.37	0.05	0.00
β_2	Non-comp	-0.68	0.07	0.00	-0.74	0.09	0.00	-0.80	0.13	0.00	-0.68	0.09	0 00
β_3	Female	-0.51	0.03	0.00	-0 55	0.05	0.00	-0.59	0.09	0.00	-0.51	0.04	0.00
β_4	Business	-6.04	1.00	0.00	-6.06	1.00	0.00	-6.08	1.00	0.00	-6.04	1.36	0.00
β_5	Local,2-3	-0.48	0.04	0.00	-0.52	0.09	0.00	-0.49	0.20	0.01	-0.48	0.06	0.00
β_6	Local,4-5	-0.82	0.05	0.00	-0.89	0.09	0.00	-0.88	0.19	0.00	-0.82	0.07	0.00
β_7	Local,6+	-1.06	0.05	0.00	-1.05	0.09	0.00	-0.94	0.19	0.00	-1.06	0.07	0.00
β_8	Foreign,0-1	-0.59	0.07	0.00	-0.63	0.10	0.00	-0.63	0.19	0.00	-0.59	0.09	0 00
β_9	Foreign,2-3	-0.68	0.05	0.00	-0.71	0.10	0.00	-0.64	0.19	0.00	-0.68	0.07	0.00
β_{10}	Foreign,4-5	-0.77	0.06	0.00	-0.77	0.10	0.00	-0.75	0.19	0.00	-0.77	0.08	0.00
β_{11}	Foreign,6+	-0.84	0.05	0.00	-0.81	0.09	0.00	-0.74	0.18	0.00	-0.84	0.07	0.00
β_{12}	North	-0.22	0.03	0.00	-0.14	0.06	0.00	-0.09	0.12	0.46	-0.22	0.04	0.00
β_{13}	East	-0.42	0.06	0.00	-0.43	0.08	0.00	-0.45	0.13	0.00	-0.42	0.08	0.00
β_{14}	EastM'sia	-0.50	0.06	0.00	-0.51	0.08	0.00	-0.51	0.12	0.00	-0.50	0.08	0.00
Df			219.00			218.00			218.00			218.00	
Pearson χ^2			404.47			294.72			219.00			-	
Deviance			254.67			159.21			98.52			-	
Log L			-388.02			-369.19			-392.92			-	

Table 11. Likelihood ratio, AIC and BIC

Test/Criteria	Poisson	Generalized Poisson I (MLE)
Likelihood ratio	-	37.7
AIC	804.0	766.4
BIC	809.2	773.9

The deviance analysis should also be implemented to both Negative Binomial I (MLE) and Generalized Poisson I (MLE) multiplicative regression models because the aim of our analysis is to obtain the simplest model that reasonably explains the variation in the data.

Following the same procedure as the Poisson, several Negative Binomial I (MLE) and Generalized Poisson I (MLE) regression models were fitted by including different rating factors; first the main effects only, then the main effects plus each of the paired interaction factors. By using the deviance and degrees of freedom, the chi-squares statistics were calculated and compared to choose the best model. Table 12 and Table 13 give the results of fitting several Negative Binomial I (MLE) and Generalized Poisson I (MLE) multiplicative regression models to the count data.

Table 12. Analysis of deviance for Negative Binomial I (MLE)

Model	deviance	df	Δdeviance	Δdf	χ^2	p-value
Null	207	231	-	-	-	-
+ Use-gender	166	229	41.63	2	20.82	0.00
+ Covarage type	149	228	16.54	1	16.54	0.00

Table 13. Analysis of deviance for Generalized Poisson I (MLE)

Model	deviance	df	Δdeviance	Δdf	χ^2	p-value
Null	262	231	-	-	-	-
+ Use-gender	180	229	81.90	2	40.95	0.00
+ Covarage type	159	228	20.73	1	20.73	0.00

Based on the deviance analysis, the best model indicates that only two of the rating factors, i.e., coverage type and use-gender, are significant and none of the paired interaction factor is significant. The parameter estimates for the two-factor models are shown in Table 14.

The two-factor models give significant parameter estimates. As a conclusion, based on the deviance analysis and parameter estimates, the best model for Negative Binomial I (MLE) and Generalized Poisson I (MLE) regression models is provided by the two-factor model.

Table 14. Parameter estimates for Negative Binomial I (MLE) and Generalized Poisson I (MLE)

Parameter		Negative Binomial I (MLE)			Generalized Poisson I (MLE)		
		estimate	std.error	p-value	estimate	std.error	p-value
a		0.16	-	-	0.04	-	-
β_1	Intercept	-3.15	0.06	0.00	-3.17	0.07	0.00
β_2	Non-comprehensive	-0.94	0.12	0.00	-0.92	0.12	0.00
β_3	Female	-0.55	0.09	0.00	-0.55	0.09	0.00
β_4	Business	-6.02	1.00	0.00	-6.01	1.00	0.00
Df			228.00			228.00	
Pearson χ^2			259.53			275.51	
Deviance			149.12			158.91	
Log L			-423.69			-425.97	

Based on the comparison between Poisson, Negative Binomial and Generalized Poisson multiplicative regression models, several remarks can be made:

- The Poisson, Negative Binomial and Generalized Poisson regression models give similar parameter estimates.
- The Negative Binomial and Generalized Poisson regression models give larger values for standard errors. Therefore, it is shown that in the presence of overdispersion, the Poisson overstates the significance of the regression parameters.
- The best regression model for Poisson indicates that all rating factors and one paired interaction factor are significant. However, the best regression model for Negative Binomial I (MLE) and Generalized Poisson I (MLE) indicates that only two rating factors are significant. Therefore, it is shown that in the presence of overdispersion, the Poisson overstates the significance of the rating factors.

5.2 Ship damage data

The ship damage incidents data of McCullagh and Nelder [19] was based on the damage incidents caused to the forward section of cargo-carrying vessels. The data provides information on the number and exposure for ship damage incidents, where the exposure was expressed in terms of aggregate number of month service. The risk of ship damage incidents was associated with three rating factors; ship type, year of construction and period of operation. The fitting procedure only involves thirty-four data points because six of the rating classes have zero exposures. The data, which was provided by Lloyd's Register of Shipping, can also be accessed from the Internet by using the following website address, http://sunsite.univie.ac.at/statlib/datasets/ships.

Since the same data was analyzed in some detail by both McCullagh and Nelder [19] and Lawless [21], the related remarks and discussions from their studies will be reported here. McCullagh and Nelder detected that there was some inter-ship variability in accident-proneness which could lead to overdispersion. For these reasons, McCullagh and Nelder assumed that,

$$Var(Y_i) = a\mu_i,$$

where,

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

i.e., a is equal to the Pearson chi-squares divided by the degrees of freedom.

By using the fitting procedure which is similar to the Poisson IWLS regression, the McCullagh and Nelder's model was fitted to the main effects data. The parameter estimates for the model are equal to the Poisson, but the standard errors are equal or larger than the Poisson because they are multiplied by $\sqrt{a}$ where $a \geq 0$.

The same main effects data was also fitted to the multiplicative regression models of Negative Binomial I (MLE) and Negative Binomial I (moment) by Lawless [21]. However, the Negative Binomial I (MLE) produced $a = 0$, and this result is equivalent to fitting the data to the Poisson multiplicative regression model.

To confirm Lawless's result, we also run the S-PLUS programming for Negative Binomial I (MLE) to the ship data. We found that the parameter estimates for the ship data did not converge and therefore concluded that the data is better to be fitted by the Poisson. Table 15 shows the comparison between Poisson, Negative Binomial and McCullagh and Nelder multiplicative regression models.

Table 15. Poisson, Negative Binomial and McCullagh and Nelder regression models

Parameters		Poisson/Negative Binomial I (MLE)			Negative Binomial I (moment)			Negative Binomial II/ McCullagh and Nelder		
		est.	std. error	p-value	est.	std. error	p-value	est.	std. error	p-value
a			-	-	0.15	-	-	0.69/ 1.69	-	-
β_1	Intercept	-6.41	0.22	0.00	-6.45	0.41	0.00	-6.41	0.28	0.00
β_2	Ship type B	-0.54	0.18	0.00	-0.50	0.30	0.10	-0.54	0.23	0.02
β_3	Ship type C	-0.69	0.33	0.04	-0.56	0.41	0.18	-0.69	0.43	0.11
β_4	Ship type D	-0.08	0.29	0.79	-0.11	0.41	0.79	-0.08	0.38	0.84
β_5	Ship type E	0.33	0.24	0.17	0.46	0.35	0.19	0.33	0.31	0.29
β_6	Const'n 65-69	0.70	0.15	0.00	0.72	0.35	0.04	0.70	0.19	0.00
β_7	Const'n 70-74	0.82	0.17	0.00	0.91	0.34	0.01	0.82	0.22	0.00
β_8	Const'n 75-79	0.45	0.23	0.05	0.46	0.42	0.27	0.45	0.30	0.13
β_9	Oper'n 75-79	0.38	0.12	0.00	0.34	0.23	0.14	0.38	0.15	0.01
Df			25.00			24.00			24.00	
Pearson χ^2			42.28			25.00			-	
Deviance			38.70			25.01			-	
Log L			-68.28			-72.83			-	

The parameter estimates and standard errors for both Negative Binomial II and McCullagh and Nelder are equal because the models were fitted by using the same procedure.

McCullagh and Nelder [19] found that the main effects model fits the data well, i.e., all of the main effects are significant and none of the paired interaction factor is significant. According to McCullagh and Nelder, if the Poisson regression model was fitted, there was an inconclusive evidence of an interaction between ship type and year of construction. However, this evidence vanished completely if the data is fitted by the overdispersion model.

Lawless [21] reported that the regression models for both McCullagh and Nelder and Negative Binomial I (moment) fit the data well. According to Lawless, both models gave the same estimates for the regression parameters and similar inferences about the regression effects. Lawless also remarked that even though there was no strong evidence of overdispersion under the Negative Binomal I (moment) or McCullagh and Nelder regression models, the method for fitting the models has a strong influence on the standard errors. In particular, the Poisson and Negative Binomial I (moment) respectively produced the smallest and largest standard errors, whereas the McCullagh and Nelder's were somewhere in between. In addition, the effects of ship type are not significant under the Negative Binomial I (moment), whereas they are under the Poisson and to a lesser extent under the McCullagh and Nelder.

If the same main effects data was fitted to the multiplicative regression models of Generalized Poisson, the parameter estimates and standard errors may also be compared. The comparisons are shown in Table 16.

Table 16. Poisson vs. Generalized Poisson

Parameters		Poisson/Generalized Poisson I (MLE)			Generalized Poisson I (moment)			Generalized Poisson II		
		est.	std. error	p-value	est.	std. error	p-value	est.	std. error	p-value
a		0.00	-	-	0.06	-	-	1.30	-	-
β_1	Intercept	-6.41	0.22	0.00	-6.46	0.45	0.00	-6.41	0.28	0.00
β_2	Ship type B	-0.54	0.18	0.00	-0.49	0.33	0.14	-0.54	0.23	0.02
β_3	Ship type C	-0.69	0.33	0.04	-0.56	0.41	0.17	-0.69	0.43	0.11
β_4	Ship type D	-0.08	0.29	0.79	-0.11	0.41	0.80	-0.08	0.38	0.84
β_5	Ship type E	0.33	0.24	0.17	0.49	0.36	0.17	0.33	0.31	0.29
β_6	Const'n 65-69	0.70	0.15	0.00	0.73	0.41	0.07	0.70	0.19	0.00
β_7	Const'n 70-74	0.82	0.17	0.00	0.94	0.39	0.02	0.82	0.22	0.00
β_8	Const'n 75-79	0.45	0.23	0.05	0.46	0.46	0.31	0.45	0.30	0.13
β_9	Oper'n 75-79	0.38	0.12	0.00	0.34	0.26	0.19	0.38	0.15	0.01
Df			25.00			24.00			24.00	
Pearson χ^2			42.28			25.00			-	
Deviance			38.70			25.29			-	
Log L			-68.28			-74.22			-	

The parameter estimates and standard errors for Generalized Poisson II, Negative Binomial II and McCullagh and Nelder are equal because the regression models were fitted by using the same procedure.

Similar to the Negative Binomial I (MLE), the Generalized Poisson I (MLE) also does not give converged values for its parameter estimates. Therefore, it will be assumed that the

Generalized Poisson I (MLE) produces $a = 0$ and this is also equivalent to fitting the data to the Poisson.

All models give similar estimates for the regression parameters. The Poisson and Generalized Poisson I (moment) respectively produced the smallest and largest standard errors, whereas the Generalized Poisson II's were somewhere in between. The effects of ship type are also not significant under the Generalized Poisson I (moment), whereas they are under the Poisson and to a lesser extent under the Generalized Poisson II.

5.3 Canadian data

The Canadian private automobile liability insurance data from Bailey and Simon [1] provides information on the number of claims incurred and exposures, where the exposure was expressed in terms of number of earned car years. The data was classified into two rating factors, merit rating and class rating. Altogether, there were twenty cross-classified rating classes of claim frequencies to be estimated. The data can also be accessed from the Internet by using the following website address, http://www.casact.org/library/astin/vol1no4/ 192.pdf.

Table 17 and Table 18 show the comparison between Poisson, Negative Binomial and Generalized Poisson multiplicative regression models for the main effects data.

The Negative Binomial II and Generalized Poisson II give equal values for parameter estimates and standard errors. The regression parameters for all models give similar estimates. The smallest standard errors are given by the Poisson, the largest are by the Negative Binomial II and Generalized Poisson II, whereas the standard errors for Negative Binomial I (MLE), Negative Binomial I (moment), Generalized Poisson I (MLE) and Generalized Poisson I (moment) lie somewhere in between.

The likelihood ratio test for Poisson against Negative Binomial I (MLE) produces likelihood ratio statistic of $T = 514.94$. The likelihood ratio is very significant, indicating that the Negative Binomial I (MLE) is a better model compared to the Poisson.

The likelihood ratio test for Poisson against Generalized Poisson I (MLE) also produces a very significant likelihood ratio statistic, $T = 525.44$. Therefore, the Generalized Poisson I (MLE) is also a better model compared to the Poisson.

Table 17. Poisson vs. Negative Binomial

Parameters		Poisson			Negative Binomial I (MLE)			Negative Binomial I (moment)			Negative Binomial II		
		est.	std. error	p-value	est.	std. error	p-value	est.	std. error	p-value	est.	std. error	p-value
a		-	-	-	0.001	-	-	0.002	-	-	47.15	-	-
β_1	Intercept	-2.53	0.00	0.00	-2.45	0.02	0.00	-2.45	0.03	0.00	-2.53	0.01	0.00
β_2	Class 2	0.30	0.01	0.00	0.24	0.03	0.00	0.24	0.03	0.00	0.30	0.05	0.00
β_3	Class 3	0.47	0.01	0.00	0.43	0.03	0.00	0.43	0.03	0.00	0.47	0.03	0.00
β_4	Class 4	0.53	0.01	0.00	0.46	0.03	0.00	0.46	0.03	0.00	0.53	0.04	0.00
β_5	Class 5	0.22	0.01	0.00	0.14	0.03	0.00	0.14	0.04	0.00	0.22	0.07	0.00
β_6	Merit X	0.27	0.01	0.00	0.22	0.03	0.00	0.22	0.03	0.00	0.27	0.05	0.00
β_7	Merit Y	0.36	0.01	0.00	0.27	0.03	0.00	0.27	0.03	0.00	0.36	0.04	0.00
β_8	Merit B	0.49	0.00	0.00	0.41	0.02	0 00	0.41	0.03	0.00	0.49	0.03	0.00
Df		12 00			11.00			11.00			11.00		
Pearson χ^2		577.83			17.56			12.00			-		
Deviance		579.52			17.67			12.08			-		
Log L		-394.96			-137.49			-138.11			-		

Table 18. Poisson vs. Generalized Poisson

Parameters		Poisson			Generalized Poisson I (MLE)			Generalized Poisson I (moment)			Generalized Poisson II		
		est.	std. error	p-value	est.	std. error	p-value	est.	std. error	p-value	est.	std. error	p-value
a		-	-	-	0.0002	-	-	0.0002	-	-	6.94	-	-
β_1	Intercept	-2.53	0.00	0.00	-2.41	0.03	0.00	-2.41	0.03	0.00	-2.53	0.01	0.00
β_2	Class 2	0.30	0.01	0.00	0.22	0.03	0.00	0.22	0.03	0.00	0.30	0.05	0.00
β_3	Class 3	0.47	0.01	0.00	0.42	0.02	0.00	0.42	0.03	0.00	0.47	0.03	0.00
β_4	Class 4	0.53	0.01	0.00	0.43	0.02	0.00	0.43	0.03	0.00	0.53	0.04	0.00
β_5	Class 5	0.22	0.01	0.00	0.12	0.03	0.00	0.12	0.03	0.00	0.22	0 07	0.00
β_6	Merit X	0.27	0 01	0.00	0.20	0.02	0.00	0.20	0.02	0.00	0.27	0.05	0.00
β_7	Merit Y	0.36	0.01	0.00	0.24	0.02	0.00	0.24	0.02	0.00	0.36	0.04	0.00
β_8	Merit B	0.49	0.00	0.00	0.38	0.02	0.00	0.38	0.38	0.00	0.49	0.03	0 00
Df		12.00			11.00			11.00			11.00		
Pearson χ^2		577.83			15.04			12.00			-		
Deviance		579.52			15.31			12.20			-		
Log L		-394.96			-132.24			-132.46			-		

Table 19 gives further comparison between Poisson, Negative Binomial I (MLE) and Generalized Poisson I (MLE). The comparison, which was based on the likelihood ratio, AIC and BIC, indicates that the Negative Binomial I (MLE) and Generalized Poisson I (MLE) are better models compared to the Poisson.

Table 19. Likelihood ratio, AIC and BIC

Test/Criteria	Poisson	Negative Binomial I (MLE)	Generalized Poisson I (MLE)
Likelihood ratio	-	514.94	525.44
AIC	805.92	292.98	282.48
BIC	800.33	286.69	276.19

6. ADDITIVE REGRESSION MODELS

In this section, the estimation procedure for the additive regression models will be briefly discussed. However, a slightly different approach is taken to compute the regression parameters.

6.1 Poisson

Let r_i, y_i and e_i denote the claim frequency rate, claim count and exposure for the *i*th class so that the observed frequency rate is equal to,

$$r_i = \frac{y_i}{e_i}. \tag{6.1}$$

If the random variable for claim count, Y_i, follows a Poisson distribution, the probability density function can be written as,

$$f(y_i) = g(r_i) = \frac{\exp(-e_i f_i)(e_i f_i)^{e_i r_i}}{(e_i r_i)!}, \tag{6.2}$$

where the mean and variance for the claim count is equal to $E(Y_i) = Var(Y_i) = e_i E(R_i) = e_i f_i$.

For Poisson regression model, the likelihood equations are equal to,

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} = \sum_i \frac{e_i(r_i - f_i)}{f_i} \frac{\partial f_i}{\partial \beta_j} = 0, \qquad j = 1,2,\ldots, p. \tag{6.3}$$

If the Poisson follows an additive model, the mean or the fitted value for frequency rate can be written as,

$$E(R_i) = f_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}, \tag{6.4}$$

so that,

$$\frac{\partial f_i}{\partial \beta_j} = x_{ij}. \tag{6.5}$$

Therefore, the first derivatives of log likelihood for Poisson are,

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} = \sum_i \frac{(r_i - f_i) e_i x_{ij}}{f_i} = 0, \qquad j = 1,2,\ldots p, \tag{6.6}$$

and the negative expectation of the second derivatives of log likelihood are,

$$-E\left(\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_s} \right) = \frac{e_i}{f_i} x_{ij} x_{is}, \qquad , j,s = 1,2,\ldots, p. \tag{6.7}$$

The information matrix, $\mathbf{I}$, which contains negative expectation of the second derivatives of log likelihood, may be written as,

$$\mathbf{I} = \mathbf{X}^{\mathrm{T}} \mathbf{W} \mathbf{X}, \tag{6.8}$$

where $\mathbf{X}$ denotes the matrix of explanatory variables, and $\mathbf{W}$ the diagonal weight matrix whose *i*th diagonal element is equal to,

$$w_i^P = \frac{e_i}{f_i}. \tag{6.9}$$

The first derivatives of log likelihood, i.e., Eq.(6.6), can be written as,

$$\mathbf{z} = \mathbf{X}^{\mathrm{T}} \mathbf{W} \mathbf{k}, \tag{6.10}$$

where $\mathbf{W}$ is the diagonal weight matrix whose *i*th diagonal element is also equivalent to Eq.(6.9), and $\mathbf{k}$ the vector whose *i*th row is equal to,

$$k_i = r_i - f_i . \tag{6.11}$$

6.2 Negative Binomial I

If the mean or the fitted value for frequency rate is assumed to follow an additive regression model, the first derivatives of log likelihood for Negative Binomial I are,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial \beta_j} = \sum_i \frac{(r_i - f_i) e_i x_{ij}}{f_i (1 + a e_i f_i)} = 0, \qquad j = 1,2,\ldots p , \tag{6.12}$$

and the negative expectation of the second derivatives of log likelihood are,

$$-E\left(\frac{\partial^2 \ell(\boldsymbol{\beta}, a)}{\partial \beta_j \partial \beta_s} \right) = \frac{e_i}{f_i (1 + a e_i f_i)} x_{ij} x_{is} , \qquad j, s = 1,2,\ldots, p . \tag{6.13}$$

Therefore, the information matrix, $\mathbf{I}$, may also be written as Eq.(6.8). However, the *i*th diagonal element of the weight matrix, $\mathbf{W}$, is equal to,

$$w_i^{NBI} = \frac{e_i}{f_i (1 + a e_i f_i)} . \tag{6.14}$$

The first derivatives of log likelihood, i.e., Eq.(6.12), can also be written as Eq.(6.10) where $\mathbf{k}$ is the vector whose *i*th row is equal to Eq.(6.11). However, the *i*th diagonal element of the weight matrix, $\mathbf{W}$, is equal to Eq.(6.14).

6.3 Negative Binomial II

The maximum likelihood estimates, $\hat{\boldsymbol{\beta}}$, for Negative Binomial II additive regression model are numerically difficult to be solved from the likelihood equations. However, the regression parameters are easier to be approximated by using the least squares equations,

$$\sum_i \frac{e_i (r_i - f_i)}{f_i (1 + a)} \frac{\partial f_i}{\partial \beta_j} = \sum_i \frac{(r_i - f_i) e_i x_{ij}}{f_i (1 + a)} = 0, \qquad j = 1,2,\ldots, p , \tag{6.15}$$

because the distribution of Negative Binomial II has a constant variance-mean ratio. Since Eq.(6.15) is also equal to the likelihood equations of the Poisson, i.e., Eq.(6.6), the least squares estimates, $\tilde{\boldsymbol{\beta}}$, are also equivalent to the Poisson maximum likelihood estimates. However, the standard errors are equal or larger than the Poisson because they are multiplied by $\sqrt{1+a}$ where $a \geq 0$.

6.4 Generalized Poisson I

If the mean or the fitted value for frequency rate is assumed to follow an additive regression model, the first derivatives of log likelihood for Generalized Poisson I are,

$$\frac{\partial \ell(\boldsymbol{\beta}, a)}{\partial \beta_j} = \sum_i \frac{(r_i - f_i)e_i x_{ij}}{f_i(1+ae_i f_i)^2}, \qquad j = 1,2,\ldots p, \tag{6.16}$$

and the negative expectation of the second derivatives of log likelihood are,

$$-E\left(\frac{\partial^2 \ell(\boldsymbol{\beta}, a)}{\partial \beta_j \partial \beta_s}\right) = \frac{e_i}{f_i(1+ae_i f_i)^2} x_{ij} x_{is}, \qquad , j,s = 1,2,\ldots, p. \tag{6.17}$$

Therefore, the information matrix, $\mathbf{I}$, may also be written as Eq.(6.8). However, the *i*th diagonal element of the weight matrix, $\mathbf{W}$, is equal to,

$$w_i^{GPI} = \frac{e_i}{f_i(1+ae_i f_i)^2}. \tag{6.18}$$

The first derivatives of log likelihood, i.e., Eq.(6.16), can be written as Eq.(6.10), where $\mathbf{k}$ is the vector whose *i*th row is equal to Eq.(6.11). However, the *i*th diagonal element of the weight matrix, $\mathbf{W}$, is equivalent to Eq.(6.18).

6.5 Generalized Poisson II

The maximum likelihood estimates, $\hat{\boldsymbol{\beta}}$, for Generalized Poisson II additive regression model are also numerically difficult to be solved from the likelihood equations. However, by using the least squares equations,

$$\sum_i \frac{(r_i - f_i)e_i x_{ij}}{a^2 f_i} = 0, \qquad j = 1,2,\ldots, p. \tag{6.19}$$

the regression parameters are easier to be calculated because the distribution of Generalized Poisson II also has a constant variance-mean ratio. Since Eq.(6.19) is equal to the likelihood equations of the Poisson, i.e., Eq.(6.6), the the least squares estimates, $\tilde{\boldsymbol{\beta}}$, are also equivalent to the Poisson maximum likelihood estimates. However, the standard errors are equal, larger or smaller than the Poisson because they are multiplied by a where $a \geq 1$ or $\frac{1}{2} \leq a < 1$.

The methods and equations for solving $\boldsymbol{\beta}$ in additive regression models are summarized in Table 20. The matrices and vectors for solving $\boldsymbol{\beta}$ in additive regression models are summarized in Table 21. An example of S-PLUS programming for the additive regression model of Negative Binomial I (moment) is given in Appendix D.

Table 20. Methods and equations for solving $\boldsymbol{\beta}$ in additive regression models

Models	Estimation of β	
	Method	Equation
Poisson	Maximum Likelihood	$\sum_i \frac{(r_i - f_i)e_i x_{ij}}{f_i} = 0$
Negative Binomial I	Maximum Likelihood	$\sum_i \frac{(r_i - f_i)e_i x_{ij}}{f_i(1+ae_i f_i)} = 0$
Negative Binomial II	Weighted Least Squares	$\sum_i \frac{(r_i - f_i)e_i x_{ij}}{(1+a)f_i} = 0$
Generalized Poisson I	Maximum Likelihood	$\sum_i \frac{(r_i - f_i)e_i x_{ij}}{f_i(1+ae_i f_i)^2} = 0$
Generalized Poisson II	Weighted Least Squares	$\sum_i \frac{(r_i - f_i)e_i x_{ij}}{a^2 f_i} = 0$

Table 21. Matrices and vectors for solving $\boldsymbol{\beta}$ in additive regression models

Models	Matrices and vectors for $\boldsymbol{\beta}_{(r)} = \boldsymbol{\beta}_{(r-1)} + \mathbf{I}_{(r-1)}^{-1}\mathbf{z}_{(r-1)}$, where $\mathbf{I}_{(r-1)} = \mathbf{X}^{\mathrm{T}}\mathbf{W}_{(r-1)}\mathbf{X}$, $\mathbf{z}_{(r-1)} = \mathbf{X}^{\mathrm{T}}\mathbf{W}_{(r-1)}\mathbf{k}_{(r-1)}$, *js*-th element of matrix $\mathbf{I} = i_{js} = -E\left(\frac{\partial^2 \ell}{\partial\beta_j \partial\beta_s}\right)$, *j*-th row of vector $\mathbf{z} = z_j = \frac{\partial \ell}{\partial \beta_j}$.	
Poisson/ NBII/ GPII	matrix $\mathbf{I}$	$i_{js} = \sum_i \frac{e_i}{f_i} x_{ij} x_{is} \;\; \rightarrow \mathbf{I} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}$
	weight matrix $\mathbf{W}$	$w_i^P = \frac{e_i}{f_i}$
	vector $\mathbf{z}$	$z_j = \sum_i \frac{e_i}{f_i}(r_i - f_i) x_{ij} \;\; \rightarrow \mathbf{z} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{k}$
	vector $\mathbf{k}$	$k_i = r_i - f_i$
NBI	matrix $\mathbf{I}$	$i_{js} = \sum_i \frac{e_i}{f_i(1 + a e_i f_i)} x_{ij} x_{is} \;\; \rightarrow \mathbf{I} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}$
	weight matrix $\mathbf{W}$	$w_i^{NBI} = \frac{e_i}{f_i(1 + a e_i f_i)}$
	vector $\mathbf{z}$	$z_j = \sum_i \frac{e_i}{f_i(1 + a e_i f_i)}(r_i - f_i) x_{ij} \;\; \rightarrow \mathbf{z} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{k}$
	vector $\mathbf{k}$	$k_i = r_i - f_i$
GPI	matrix $\mathbf{I}$	$i_{js} = \sum_i \frac{e_i}{f_i(1 + a e_i f_i)^2} x_{ij} x_{is} \;\; \rightarrow \mathbf{I} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}$
	weight matrix $\mathbf{W}$	$w_i^{GPI} = \frac{e_i}{f_i(1 + a e_i f_i)^2}$
	vector $\mathbf{z}$	$z_j = \sum_i \frac{e_i}{f_i(1 + a e_i f_i)^2}(r_i - f_i) x_{ij} \;\; \rightarrow \mathbf{z} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{k}$
	vector $\mathbf{k}$	$k_i = r_i - f_i$

6.6 Examples

Following the same examples as the multiplicative regression models, the additive regression models were also fitted on three different sets of claim frequency data; Malaysian data, ship damage incident data, and Canadian data. Unfortunately, the Malaysian data did not give converged parameter solutions for any of the Poisson, Negative Binomial and Generalized Poisson regression models. However, the parameter solutions are obtainable for both ship damage incident data and Canadian data. Table 22 shows the comparison between Poisson, Negative Binomial and Generalized Poisson additive regression models for the ship damage incident data.

Table 22. Poisson, Negative Binomial and Generalized Poisson for ship data

Parameters		Poisson/NBI(MLE)/ GPI(MLE)			NBI(moment)			GPI(moment)			NBII/GPII		
		est. ($\times 10^3$)	std. error ($\times 10^3$)	p-value	est. ($\times 10^3$)	std. error ($\times 10^3$)	p-value	est. ($\times 10^3$)	std. error ($\times 10^3$)	p-value	est. ($\times 10^3$)	std. error ($\times 10^3$)	p-value
a		0.00	-	-	133.73	-	-	52.94	-	-	599.25/ 1264.61	-	-
β_1	Intercept	2.60	0.72	0.00	2.19	1.04	0.03	2.16	1.05	0.04	2.60	0.91	0.00
β_2	Ship type B	-1.73	0.71	0.01	-1.33	1.01	0.19	-1.30	1.02	0.20	-1.73	0.90	0.05
β_3	Ship type C	-1.89	0.86	0.03	-1.52	1.12	0.17	-1.52	1.11	0.17	-1.89	1.09	0.08
β_4	Ship type D	-0.79	1.10	0.47	-1.05	1.32	0.43	-1.13	1.28	0.38	-0.79	1.39	0.57
β_5	Ship type E	1.87	1.30	0.15	2.72	1.91	0.15	2.87	1.92	0.13	1.87	1.64	0.25
β_6	Cons. 65-69	1.05	0.24	0.00	0 87	0.56	0.12	0.78	0.63	0.22	1.05	0.31	0.00
β_7	Cons. 70-74	1.58	0.38	0.00	2.15	0.76	0.00	2.33	0.84	0.01	1.58	0.47	0.00
β_8	Cons. 75-79	0.69	0.55	0.22	0.77	0.94	0.42	0.76	0.98	0.44	0.69	0 70	0.33
β_9	Oper. 75-79	0.79	0.24	0.00	0.79	0.52	0.13	0.81	0.58	0.16	0.79	0.31	0.01
Df		25.00			24.00			24.00			24.00		
Pearson χ^2		39.98			25.00			25.00			-		
Deviance		38.44			25.65			26.12			-		
Log L		-68.15			-72.44			-73.48			-		

After running the S-PLUS programming for Negative Binomial I (MLE) and Generalized Poisson I (MLE) to the ship data, we found that the models did not give converged parameter solutions and concluded that the data is better to be fitted by the Poisson. Since the Poisson is a special case of the Negative Binomial I (MLE) and Generalized Poisson I (MLE), the result of fitting the Poisson is also equivalent to the result of fitting the Negative Binomial I (MLE) or Generalized Poisson I (MLE) which produces $a = 0$.

The parameter estimates and standard errors for Negative Binomial II and Generalized Poisson II are equal because both models were fitted by using the same procedure.

The smallest standard errors are given by the Poisson, the largest are by the Negative Binomial I (moment) and Generalized Poisson I (moment), whereas the standard errors for Negative Binomial II and Generalized Poisson II are somewhere in between.

Table 23 shows the comparison between Poisson, Negative Binomial and Generalized Poisson additive regression models for the Canadian data.

The parameter estimates and standard errors for Negative Binomial II and Generalized Poisson II are equal because both models were fitted by using the same procedure.

The smallest standard errors are given by the Poisson, the largest are by the Negative Binomial I (moment) and Generalized Poisson I (moment), whereas the standard errors for Negative Binomial I (MLE), Generalized Poisson I (MLE), Negative Binomial II and Generalized Poisson II are somewhere in between.

Table 23. Poisson, Negative Binomial and Generalized Poisson for Canadian data

Parameters		Poisson			NBI(MLE)			NBI(moment)		
		est. (×10²)	std. error (×10²)	p-value	est. (×10²)	std. error (×10²)	p-value	est. (×10²)	std. error (×10²)	p-value
a		-	-	-	0.06	-	-	0.12	-	-
β_1	Intercept	7.88	0.02	0.00	7.98	0.17	0.00	8.00	0.22	0.00
β_2	Class 2	3.13	0.09	0.00	2.99	0.25	0.00	2.99	0.32	0.00
β_3	Class 3	5.24	0.07	0.00	5.66	0.27	0.00	5.70	0.35	0.00
β_4	Class 4	6.53	0.08	0.00	6.36	0.28	0.00	6.34	0.36	0.00
β_5	Class 5	2.17	0.12	0.00	1.88	0.26	0.00	1.81	0.32	0.00
β_6	Merit X	2.76	0.08	0.00	2.74	0.24	0.00	2.72	0.30	0.00
β_7	Merit Y	3.86	0.08	0.00	3.55	0.24	0.00	3.50	0.31	0.00
β_8	Merit B	5.88	0.06	0.00	5.63	0.25	0.00	5.59	0.32	0.00
Df			12.00			11.00			11.00	
Pearson χ^2			95.93			19.19			12.00	
Deviance			96.07			19.36			12.10	
Log L			-153.24			-132.31			-133.24	

Parameters		GPI(MLE)			GPI(moment)			NBII/GPII		
		est. (×10²)	std. error (×10²)	p-value	est. (×10²)	std. error (×10²)	p-value	est. (×10²)	std. error (×10²)	p-value
a		0.02	-	-	0.02	-	-	699.38/ 282.73	-	-
β_1	Intercept	8.24	0.28	0.00	8.29	0.35	0.00	7.88	0.05	0.00
β_2	Class 2	2.84	0.30	0.00	2.84	0.35	0.00	3.13	0 24	0.00
β_3	Class 3	5.84	0.31	0.00	5.90	0.38	0.00	5.24	0.19	0.00
β_4	Class 4	6.21	0.31	0.00	6.19	0.38	0.00	6.53	0.23	0.00
β_5	Class 5	1.72	0.31	0.00	1.64	0.36	0.00	2.17	0.33	0.00
β_6	Merit X	2.55	0.27	0.00	2 51	0.32	0.00	2.76	0.23	0.00
β_7	Merit Y	3.28	0.27	0.00	3.19	0.32	0.00	3.86	0.22	0.00
β_8	Merit B	5.37	0.28	0.00	5.28	0.33	0.00	5.88	0.18	0.00
Df			11.00			11.00			11.00	
Pearson χ^2			17.08			12.00			-	
Deviance			17.51			12 34			-	
Log L			-132.12			-132.61			-	

7. CONCLUSIONS

This paper proposed the Negative Binomial and Generalized Poisson regression models as alternatives for handling overdispersion. Specifically, four types of distributions, i.e., Negative Binomial I, Negative Binomial II, Generalized Poisson I and Generalized Poisson II, and two types of regression models, i.e., multiplicative and additive, were discussed. Since the likelihood equations for the multiplicative and additive regression models of the Negative Binomial I and Generalized Poisson I are equal to the weighted least squares, the fitting procedure can be carried out easily by using the Iterative Weighted Least Squares (IWLS) regression.

The estimation of the dispersion parameter, a, can be implemented by using either the maximum likelihood method or the method of moment. In this paper, the models where a is estimated by the maximum likelihood method are denoted by Negative Binomial I (MLE) and Generalized Poisson I (MLE). Similarly, the Negative Binomial I (moment) and Generalized Poisson I (moment) represent the models where a is estimated by the method of moment.

The maximum likelihood estimates for Negative Binomial II and Generalized Poisson II are numerically difficult to be solved because their likelihood equations are not equal to the weighted least squares. As an alternative, the method of least squares is suggested because both Negative Binomial II and Generalized Poisson II have constant variance-mean ratios.

Table 1 and Table 20 summarize the methods and equations for solving $\beta_j, j = 1,2,..., p$, in multiplicative and additive regression models. The matrices and vectors for solving $\boldsymbol{\beta}$ in multiplicative and additive regression models are summarized in Table 2 and Table 21. Finally, Table 3 summarizes the equations for solving a.

This paper also briefly discussed several goodness-of-fit measures which were already familiar to those who used Generalized Linear Model with Poisson error structure for claim count or frequency analysis. The measures, which are also applicable to the Negative Binomial as well as the Generalized Poisson regression models, are the Pearson chi-squares, deviance, likelihood ratio test, Akaike Information Criteria (AIC) and Bayesian Schwartz Information Criteria (BIC).

In this paper, the multiplicative and additive regression models of the Poisson, Negative Binomial and Generalized Poisson were fitted, tested and compared on three different sets of claim frequency data; Malaysian private motor third party property damage data, ship damage incident data from McCullagh and Nelder [19], and data from Bailey and Simon [1] on Canadian private automobile liability. Unfortunately, none of the additive regression models give converged parameter solutions for the Malaysian data.

This paper shows that even though the Poisson, the Negative Binomial and the Generalized Poisson produce similar estimate for the regression parameters, the standard errors for the Negative Binomial and the Generalized Poisson are larger than the Poisson. Therefore, the Poisson overstates the significance of the regression parameters in the presence of overdispersion. An example can be seen from the results of fitting the Poisson, the Negative Binomial and the Generalized Poisson to the ship damage data. The effects of ship type are not significant under the NBI-moment or the GPI-moment, whereas they are under the Poisson, and to a lesser extent under the McCullagh and Nelder or the NBII or the GPII.

This paper also shows that in the presence of overdispersion, the Poisson overstates the significance of the rating factors. An example can be seen from the results of implementing the deviance analysis to the Malaysian data. The best regression model for the Poisson indicates that all rating factors and one paired interaction factor are significant. However, the best regression model for NBI-MLE and GPI-MLE indicates that only two rating factors are significant. Another example can be seen from the ship damage data. According to McCullagh and Nelder [19], there was an evidence of interaction between ship type and year of construction if the Poisson regression was fitted. However, the evidence vanished completely if the data is fitted by the overdispersion model.

In addition, this paper shows that the maximum likelihood approach has several advantages compared to the quasi likelihood approach, which was suggested in the actuarial literature, to accommodate overdispersion in claim count or frequency data. Besides having good properties, the maximum likelihood approach allows the likelihood ratio and other standard maximum likelihood tests to be implemented.

The Negative Binomial and the Generalized Poisson models are not that difficult to be understood. Even though the probability density function for both Negative Binomial and Generalized Poisson involve mathematically complex formulas, the mean and variance for both models are conceptually simpler to be interpreted. The mean for both Negative Binomial and Generalized Poisson models are equal to the Poisson. The variance of the Negative Binomial is equal or larger than the Poisson, and this allows the Negative Binomial model to handle overdispersion. The variance of the Generalized Poisson is equal, larger or smaller than the Poisson, and this allows the Generalized Poisson to handle either overdispersion or underdispersion.

The Negative Binomial and Generalized Poisson are also not that difficult to be fitted. The fitting procedure can be carried out by using the Iterative Weighted Least Squares regression which was used in the Poisson fitting procedure. The only difference is that the Negative Binomial and the Generalized Poisson has their own weight matrix, and the iteration procedure for calculating the dispersion parameter, a, has to be added in the fitting procedure.

Acknowledgment

The authors gratefully acknowledge the financial support received in the form of a research grant (IRPA RMK8: 09-02-02-0112-EA274) from the Ministry of Science, Technology and Innovation (MOSTI), Malaysia. The authors are also pleasured to thank the General Insurance Association of Malaysia (PIAM), in particular Mr. Carl Rajendram and Mrs. Addiwiyah, for supplying the data.

Appendix A: S-PLUS programming for Negative Binomial I (moment) multiplicative regression model

```
NB.moment <- function(data)
{
# To identify matrix X, vector count and vector exposure from the data
    X          <- as.matrix(data[, -(1:2)])
    count      <- as.vector(data[, 1])
    exposure   <- as.vector(data[, 2])
# To set initial values for a and beta
    new.a      <- c(0.001)
    new.beta   <- rep(c(0.001), dim(X)[2])
# To start iterations
    for (i in 1:50)
    {
# To start the first sequence
            a           <- new.a
            beta        <- new.beta
            miul        <- exposure*exp(as.vector(X%*%beta))
            W           <- diag(miul/(1+a*miul))
            I.inverse   <- solve(t(X)%*%W%*%X)
            k           <- (count-miul)/miul
```

```
            z           <- t(X)%*%W%*%k
            new.beta    <- as.vector(beta+I.inverse%*%z)
            new.miul    <- exposure*exp(as.vector(X%*%new.beta))
# To start the second sequence
            G           <- sum((count-new.miul)^2/(new.miul*(1+a*new.miul)))-
                           (dim(X)[1]-dim(X)[2])
            G.prime     <- -(sum((count-new.miul)^2/(1+a*new.miul)^2))
            new.a       <- a-G/G.prime
    }
# To calculate the variance and standard error
    varians     <- as.vector(diag(I.inverse))
    std.error   <- sqrt(varians)
# To list the programming output
    list (a=new.a, beta=new.beta, std.error=std.error, df=dim(X)[1]-dim(X)[2]-1)
}
```

Appendix B: S-PLUS programming for Generalized Poisson I (moment) multiplicative regression model

```
GP.moment <- function(data)
{
# To identify matrix X, vector count and vector exposure from the data
    X           <- as.matrix(data[, -(1:2)])
    count       <- as.vector(data[, 1])
    exposure    <- as.vector(data[, 2])
# To set initial values for a and beta
    new.a       <- c(0.001)
    new.beta    <- rep(c(0.001), dim(X)[2])
# To start iterations
    for (i in 1:50)
    {
# To start the first sequence
            a           <- new.a
            beta        <- new.beta
            miul        <- exposure*exp(as.vector(X%*%beta))
```

```
        W           <- diag(miul/(1+a*miul)^2)
        I.inverse   <- solve(t(X)%*%W%*%X)
        k           <- (count-miul)/miul
        z           <- t(X)%*%W%*%k
        new.beta    <- as.vector(beta+I.inverse%*%z)
        new.miul    <- exposure*exp(as.vector(X%*%new.beta))
# To start the second sequence
        G           <- sum((count-new.miul)^2/(new.miul*(1+a*new.miul)^2))-
                                                         (dim(X)[1]-dim(X)[2])
        G.prime     <- -(sum(2*(count-new.miul)^2/(1+a*new.miul)^3))
        new.a       <- a-G/G.prime
# To set restrictions for a
          if ((new.a<0)*(new.a<=-1/max(count)))
            new.a <- -1/(max(count)+1)
            else
              if ((new.a<0)*(new.a<=-1/max(new.miul)))
                new.a <- -1/(max(new.miul)+1)
                else
                  if ((new.a<0)*(new.a<=-1/max(count))*
                                                   (new.a<=-1/max(new.miul)))
                    new.a <- min(-1/(max(count)+1),-1/(max(new.miul)+1))
                    else
                      new.a <- new.a
    }
# To calculate the variance and standard error
    varians <- as.vector(diag(I.inverse))
    std.error <- sqrt(varians)
# To list the programming output
    list(a=new.a, beta=new.beta, std.error=std.error, df=dim(X)[1]-dim(X)[2]-1)
}
```

Appendix C: Malaysian data

Rating factors					Exposures	Claim counts
Coverage type	Vehicle make	Use-gender	Vehicle year	Location		
Comprehensive	Local	Private-male	0-1 year	Central	4243	381
				North	2567	146
				East	598	44
				South	1281	161
				East Malaysia	219	8
			2-3 year	Central	6926	422
				North	4896	203
				East	1123	41
				South	2865	164
				East Malaysia	679	19
			4-5 year	Central	6286	276
				North	4125	145
				East	1152	29
				South	2675	115
				East Malaysia	700	17
			6+ year	Central	6905	223
				North	5784	150
				East	2156	39
				South	3310	89
				East Malaysia	1406	33
		Private-female	0-1 year	Central	2025	165
				North	1635	55
				East	301	12
				South	608	23
				East Malaysia	126	6
			2-3 year	Central	3661	147
				North	2619	72
				East	527	12
				South	1192	39
				East Malaysia	359	8
			4-5 year	Central	2939	56
				North	1927	36
				East	439	7
				South	959	23
				East Malaysia	376	2
			6+ year	Central	2215	51
				North	1989	38
				East	581	5
				South	937	23
				East Malaysia	589	9
		Business	0-1 year	Central	290	0
				North	66	0
				East	24	0
				South	52	0
				East Malaysia	6	0
			2-3 year	Central	572	0
				North	148	0
				East	40	0
				South	91	0
				East Malaysia	17	0
			4-5 year	Central	487	0
				North	100	0
				East	40	0
				South	59	0

			East Malaysia	22	0
		6+ year	Central	468	0
			North	93	1
			East	33	0
			South	77	0
			East Malaysia	25	0
Foreign	Private-male	0-1 year	Central	1674	94
			North	847	47
			East	377	21
			South	740	38
			East Malaysia	518	6
		2-3 year	Central	3913	202
			North	1930	85
			East	618	21
			South	1768	65
			East Malaysia	833	23
		4-5 year	Central	4002	157
			North	1777	85
			East	534	15
			South	1653	73
			East Malaysia	840	24
		6+ year	Central	6891	245
			North	4409	151
			East	1345	44
			South	2735	113
			East Malaysia	2108	64
	Private-female	0-1 year	Central	1222	29
			North	632	11
			East	209	2
			South	452	17
			East Malaysia	345	6
		2-3 year	Central	2111	46
			North	1068	41
			East	283	5
			South	857	13
			East Malaysia	493	10
		4-5 year	Central	1699	39
			North	793	15
			East	188	0
			South	637	16
			East Malaysia	367	11
		6+ year	Central	1922	47
			North	1376	35
			East	336	6
			South	710	9
			East Malaysia	792	10
	Business	0-1 year	Central	457	0
			North	135	0
			East	70	0
			South	86	0
			East Malaysia	101	0
		2-3 year	Central	1134	0
			North	315	0
			East	113	0
			South	284	0
			East Malaysia	205	0
		4-5 year	Central	1030	0
			North	252	0
			East	70	0

				South	208	0
				East Malaysia	221	0
			6+ year	Central	1075	0
				North	297	0
				East	78	0
				South	231	0
				East Malaysia	282	0
Non-comprehensive	Local	Private-male	0-1 year	Central	8	0
				North	14	0
				East	5	0
				South	8	0
				East Malaysia	3	0
			2-3 year	Central	34	3
				North	65	0
				East	26	0
				South	51	1
				East Malaysia	21	0
			4-5 year	Central	71	1
				North	180	5
				East	47	0
				South	48	1
				East Malaysia	39	0
			6+ year	Central	349	9
				North	496	5
				East	143	2
				South	233	4
				East Malaysia	141	2
		Private-female	0-1 year	Central	2	0
				North	6	0
				East	6	0
				South	3	0
				East Malaysia	3	0
			2-3 year	Central	12	0
				North	23	1
				East	22	0
				South	14	0
				East Malaysia	21	0
			4-5 year	Central	36	0
				North	66	1
				East	19	0
				South	13	0
				East Malaysia	29	0
			6+ year	Central	133	1
				North	213	0
				East	50	0
				South	55	0
				East Malaysia	85	1
		Business	0-1 year	Central	1	0
				North	2	0
				East	0	0
				South	0	0
				East Malaysia	0	0
			2-3 year	Central	1	0
				North	5	0
				East	1	0
				South	1	0
				East Malaysia	1	0
			4-5 year	Central	18	0
				North	8	0

			East	1	0
			South	1	0
			East Malaysia	0	0
		6+ year	Central	57	0
			North	27	0
			East	1	0
			South	133	0
			East Malaysia	3	0
Foreign	Private-male	0-1 year	Central	4	0
			North	11	0
			East	2	0
			South	5	0
			East Malaysia	8	0
		2-3 year	Central	41	0
			North	54	3
			East	7	0
			South	30	2
			East Malaysia	25	0
		4-5 year	Central	68	0
			North	132	3
			East	20	0
			South	55	0
			East Malaysia	48	3
		6+ year	Central	3164	49
			North	3674	71
			East	920	6
			South	2067	56
			East Malaysia	1985	22
	Private-female	0-1 year	Central	2	0
			North	8	0
			East	1	0
			South	3	0
			East Malaysia	6	0
		2-3 year	Central	10	0
			North	47	0
			East	0	0
			South	12	0
			East Malaysia	26	0
		4-5 year	Central	29	0
			North	66	0
			East	2	0
			South	14	0
			East Malaysia	25	0
		6+ year	Central	875	14
			North	1177	15
			East	190	2
			South	411	6
			East Malaysia	555	3
	Business	0-1 year	Central	1	0
			North	1	0
			East	0	0
			South	2	0
			East Malaysia	2	0
		2-3 year	Central	4	0
			North	6	0
			East	0	0
			South	5	0
			East Malaysia	14	0
		4-5 year	Central	17	0

		North	14	0
		East	4	0
		South	7	0
		East Malaysia	20	0
	6+ year	Central	157	0
		North	141	0
		East	22	0
		South	89	0
		East Malaysia	152	0
Total			170,749	5,728

Appendix D: S-PLUS programming for Negative Binomial I (moment) additive regression model

```
NBmoment.add <- function(data)
{
# To identify matrix X, vector count, vector exposure and vector frequency from the data
    X <- as.matrix(data[,-(1:2)])
    count <- as.vector(data[,1])
    exposure <- as.vector(data[,2])
    rate <- count/exposure
# To set initial values for a and beta
    new.beta <- rep(c(0.001), dim(X)[2])
    new.a <- c(0.001)
# To start iterations
    for (i in 1:50)
    {
# To start the first sequence
        beta <- new.beta
        a <- new.a
        fitted <- as.vector(X%*%beta)
        W <- diag(exposure/(fitted*(1+a*exposure*fitted)))
        I.inverse <- solve(t(X)%*%W%*%X)
        k <- rate-fitted
        z <- t(X)%*%W%*%k
        new.beta <- as.vector(beta+I.inverse%*%z)
        new.fitted <- as.vector(X%*%new.beta)
```

```
# To start the second sequence
        G <-   sum((exposure*(rate-new.fitted)^2)/(new.fitted*(1+a*exposure*new.fitted)))-
                      (dim(X)[1]-dim(X)[2])
        G.prime <- -sum((exposure^2*(rate-new.fitted)^2)/(1+a*exposure*new.fitted)^2)
        new.a <- a-G/G.prime
    }
# To calculate the variance and standard error
    varians <- as.vector(diag(I.inverse))
    std.error <- sqrt(varians)
    }
# To list the programming output
    list(a=new.a, beta=new.beta, std.error=std.error, df=dim(X)[1]-dim(X)[2]-1)
}
```

8. REFERENCES

[1] R.A. Bailey, L.J. Simon, "Two Studies in Automobile Insurance Ratemaking", *ASTIN Bulletin*, **1960**, Vol. 4, No. 1, 192-217.

[2] R.A. Bailey, "Insurance Rates with Minimum Bias", *Proceedings of the Casualty Actuarial Society*, **1963**, Vol. 50, No. 93, 4-14.

[3] J. Jung, "On Automobile Insurance Ratemaking", *ASTIN Bulletin*, **1968**, Vol. 5, No. 1, 41-48.

[4] B. Ajne, "A Note on the Multiplicative Ratemaking Model", *ASTIN Bulletin*, **1975**, Vol. 8, No. 2, 144-153.

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

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

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

[8] R.L., Brown, "Minimum Bias with Generalized Linear Models", *Proceedings of the Casualty Actuarial Society*, **1988**, Vol. 75, No. 143, 187-217.

[9] S.J. Mildenhall, "A Systematic Relationship Between Minimum Bias and Generalized Linear Models", *Proceedings of the Casualty Actuarial Society*, **1999**, Vol.86, No. 164, 93-487.

[10] S. Feldblum, J.E. Brosius, "The Minimum Bias Procedure: A Practitioner's Guide", *Proceedings of the Casualty Actuarial Society*, **2003**, Vol. 90, No. 172, 196-273.

[11] D. Anderson, S. Feldblum, C. Modlin, D. Schirmacher, E. Schirmacher, N. Thandi, "A Practitioner's Guide to Generalized Linear Models", *Casualty Actuarial Society Discussion Paper Program*, **2004**, 1-115.

[12] L. Fu, C.P. Wu, "Generalized Minimum Bias Models", *Casualty Actuarial Society Forum*, **2005**, Winter, 72-121.

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

[14] L. Freifelder, "Estimation of Classification Factor Relativities: A Modeling Approach", *Journal of Risk and Insurance*, **1986**, Vol. 53, 135-143.

[15] B. Jee, "A Comparative Analysis of Alternative Pure Premium Models in the Automobile Risk Classification System", *Journal of Risk and Insurance*, **1989**, Vol. 56, 434-459.

[16] K.D. Holler, D. Sommer, G. Trahair, "Something Old, Something New in Classification Ratemaking with a Novel Use of GLMs for Credit Insurance", *Casualty Actuarial Society Forum*, **1999**, Winter, 31-84.
[17] M. Aitkin, D. Anderson, B. Francis, J. Hinde, *Statistical Modelling in GLIM*, **1990**, Oxford University Press, New York.
[18] A.E. Renshaw, "Modelling the Claims Process in the Presence of Covariates", *ASTIN Bulletin*, **1994**, Vol. 24, No. 2, 265-285.
[19] P. McCullagh, J.A. Nelder, *Generalized Linear Models (2nd Edition)*, **1989**, Chapman and Hall, London.
[20] M.J. Brockmann, T.S. Wright, "Statistical Motor Rating: Making Effective Use of Your Data", *Journal of the Institute of Actuaries*, **1992**, Vol. 119, No. 3, 457-543.
[21] J.F. Lawless, "Negative Binomial and Mixed Poisson Regression", *The Canadian Journal of Statistics*, **1987**, Vol. 15, No. 3, 209-225.
[22] A.C. Cameron, P.K. Trivedi, "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests", *Journal of Applied Econometrics*, **1986**; Vol. 1, 29-53.
[23] N.E. Breslow, "Extra-Poisson Variation in Log-linear Models", *Journal of the Royal Statistical Society (Applied Statistics)*, **1984**, Vol. 33, No. 1, 38-44.
[24] J.A. Nelder, Y. Lee, "Likelihood, Quasi-likelihood and Pseudolikelihood: Some Comparisons", *Journal of the Royal Statistical Society (B)*, **1992**, Vol. 54, No. 1, 273-284.
[25] D.R. Cox, "Some Remarks on Overdispersion", *Biometrika*, **1983**, Vol. 70, No. 1, 269-274.
[26] W. Wang, F. Famoye, "Modeling Household Fertility Decisions with Generalized Poisson Regression", *Journal of Population Economics*, **1997**, Vol. 10, 273-283.
[27] P.C. Consul, F. Famoye, "Generalized Poisson Regression Model", *Communication Statistics (Theory & Methodology)*, **1992**, Vol. 2, No. 1, 89-109.
[28] H. Akaike, "Information Theory and An Extension of the Maximum Likelihood Principle", *Second International Symposium on Inference Theory*, **1973**, Akademiai Kiado, Budapest, 267-281.
[29] G. Schwartz, "Estimating the Dimension of a Model", *Annals of Statistics*, **1978**, Vol. 6, 461-464.

Biographies of the Authors

Noriszura Ismail is a lecturer in National University of Malaysia since July 1993, teaching Actuarial Science courses. She received her MSc. (1993) and BSc. (1991) in Actuarial Science from University of Iowa, and is now currently pursuing her PhD in Statistics. She has presented several papers in Seminars, including those locally and in South East Asia, and published several papers in local and Asian Journals.

Abdul Aziz Jemain is an Associate Professor in National University of Malaysia, teaching in Statistics Department since 1982. He received his MSc. (1982) in Medical Statistics from London School of Hygiene and Tropical Medicine and PhD (1989) in Statistics from University of Reading. He has written several articles, including local and international Proceedings and Journals, and co-authors several local books.