Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models

Noriszura Ismail, Ph.D. and Hossein Zamani, Ph.D.

Abstract: This study relates negative binomial and generalized Poisson regression models through the mean-variance relationship, and suggests the application of these models for overdispersed or underdispersed count data. In addition, this study relates zero-inflated negative binomial and zero-inflated generalized Poisson regression models through the mean-variance relationship, and suggests the application of these zero-inflated models for zero-inflated and overdispersed count data. The negative binomial and generalized Poisson regression models were fitted to the Malaysian OD claim count data, whereas the zero-inflated negative binomial and zero-inflated generalized Poisson regression models were fitted to the German healthcare count data.

Keywords: negative binomial regression, generalized Poisson regression, zero-inflated negative binomial regression, zero-inflated generalized Poisson regression.

1. INTRODUCTION

Poisson regression has been widely used for fitting count data. As examples, in insurance area, Aitkin et al. (1990) and Renshaw (1994) fitted Poisson regression to two different sets of U.K. motor claim data, whereas in healthcare area, Riphahn et al. (2003) fitted the model to German Socioeconomic Panel (GSOEP) data. However, count data often display overdispersion and inappropriate imposition of Poisson regression may underestimate the standard errors and overstate the significance of regression parameters. Quasi Poisson regression has been suggested to accommodate overdispersion in count data and the advantage of using this model is that the model can be fitted without knowing the exact probability function of the response variable, as long as the mean is specified to be equivalent to Poisson mean and the variance can be written as a linear proportion of Poisson mean. To account for overdispersion, quasi Poisson regression produces regression estimates equivalent to Poisson regression, but standard errors larger than Poisson regression (Ismail and Jemain 2007). In insurance applications, McCullagh and Nelder (1989) fitted quasi Poisson regression to damage incidents of cargo-carrying vessels and Brockman and Wright (1992) fitted the model to U.K. own damage motor claim data.

Besides quasi Poisson regression, several mixed Poisson regressions have been considered as alternatives for handling overdispersed count data. The details of mixed Poisson regressions such as Poisson-inverse Gaussian (PIG), Poisson-lognormal (PLN) and Poisson-gamma (or negative binomial) can be found in Denuit et al. (2007), who applied these models for accommodating overdispersed claim data. For the case of both overdispersed and underdispersed count data, generalized Poisson and COM-Poisson regression models have been suggested and fitted by several researchers (Consul 1989, Consul and Famoye 1992, Famoye et al. 2004, Wang and Famoye 1997, Conway and Maxwell 1962, Shmueli et al. 2005, Lord et al. 2008, Lord et al. 2010).

Several parameterizations have been performed for negative binomial (NB) regression, and the two well known models, NB-1 and NB-2, have been applied (Cameron and Trivedi 1986, Lawless 1987, McCullagh and Nelder 1989, Cameron and Trivedi 1998, Winkelmann 2008, Hilbe 2007). However, both NB-1 and NB-2 regressions are not nested, and appropriate statistical tests to choose a better model cannot be carried out. Recently, the functional form of NB regression has been extended and introduced as NB-P regression, where both NB-1 and NB-2 regressions are special cases of NB-P when $P = 1$ and $P = 2$ respectively (Greene 2008). The advantage of using NB-P regression is that it parametrically nests both NB-1 and NB-2 regressions, and hence, allowing statistical tests of the two functional forms against a more general alternative. In particular, likelihood ratio test can be implemented for choosing between NB-1 against NB-P regressions, or NB-2 against NB-P regressions. Applications of several parameterizations for NB, PIG and PLN regressions can be found in Boucher et al. (2007), who fitted NB (NB-1, NB-2 and NB-K+1), PIG (PIG-1, PIG-2 and PIG-K+1) and PLN (PLN-1, PLN-2 and PLN-K+1) regressions to Spanish motor claim data.

Several parameterizations have also been performed for generalized Poisson (GP) regression where the two well known models, GP-1 and GP-2, have been applied for dealing with overdispersed as well as underdispersed count data (Consul and Jain 1973, Consul 1989, Consul and Famoye 1992, Famoye et al. 2004, Wang and Famoye 1997, Ismail and Jemain 2007). Similar to NB-1 and NB-2 regressions, both GP-1 and GP-2 regressions are not nested and appropriate statistical tests to choose a better model cannot be performed. Recently, the functional form of GP regression has been extended and introduced as GP-P regression, where both GP-1 and GP-2 regressions are special cases of GP-P when $P = 1$ and $P = 2$ respectively (Zamani and Ismail 2012). The advantage of using GP-P is that it parametrically nests both GP-1 and GP-2, and therefore, allowing statistical tests of the two functional forms against a more general alternative. In particular, likelihood ratio test can be implemented for choosing between GP-1 against GP-P regressions, or GP-2 against GP-P regressions.

In terms of properties, there is no big difference between NB and GP distributions when the means and variances are fixed, but, GP distribution has heavier tail whereas NB distribution has more mass at zero (Joe and Zhu 2005). In addition, the score tests for overdispersion in Poisson vs. NB (NB-1 and NB-2) regressions and Poisson vs. GP (GP-1 and GP-2) regressions are equal (Yang et al. 2007, Yang et al. 2009a).

In several cases, count data often have excessive number of zero outcomes than are expected in Poisson regression. As an example, the proportion of zero claims in motor insurance data may increase due to the conditions of deductible and no claim discount that discourage insured drivers to report small claims (Yip and Yau 2005). In healthcare area as another example, services of psychiatric outpatient may report a large proportion of zero utilization of such services for many patients (Neelon et al. 2010). Zero-inflated datasets can also be found in other areas such as environmental sciences (Agarwal et al. 2002), medicine (Bohning et al. 1996) and manufacturing (Lambert 1992). Zero-inflation phenomenon is a very specific type of overdispersion, and zero-inflated Poisson (ZIP) regression has been suggested to handle purely zero-inflated data. ZIP regression mixes a distribution degenerate at zero with a Poisson distribution, by allowing the incorporation of explanatory variables in both the zero process and the Poisson distribution.

As an alternative to ZIP regression, one may consider zero-inflated negative binomial (ZINB) regression if the count data continue to suggest additional overdispersion. ZINB regression is obtained by mixing a distribution degenerate at zero with a NB distribution, by allowing the incorporation of explanatory variables in both the zero process and the NB distribution. Applications of ZINB-1 and ZINB-2 regressions can be found in Ridout et al. (2001).

Besides ZINB, zero-inflated generalized Poisson (ZIGP) regression has been proposed as an alternative to handle zero-inflation and additional overdispersion in count data. ZIGP regression, which mixes a distribution degenerate at zero with a GP distribution and allows the incorporation of explanatory variables in both the zero process and the GP distribution, have been applied by Famoye and Singh (2006) for domestic violence data and by Yang et al. (2009b) for apple shoot propagation data. However, both ZIGP-1 and ZIGP-2 regressions are not nested and appropriate statistical tests to choose a better model cannot be carried out. Recently, the functional form of ZIGP regression has been extended and introduced as ZIGP-P regression (Zamani and Ismail 2013a), where ZIGP-1 and ZIGP-2 regressions are special cases of ZIGP-P regression when $P=1$ and $P=2$ respectively. The advantage of using ZIGP-P regression is that it parametrically nests both ZIGP-1 and ZIGP-2 regressions, and hence, allowing statistical tests of the two functional forms against a more general alternative. In particular, likelihood ratio test can be implemented for choosing between ZIGP-1 against ZIGP-P regressions, or ZIGP-2 against ZIGP-P regressions.

Even though ZIGP regression is a good competitor of ZINB regression, in several cases, ZINB regression may not provide converged values in the iterative technique of the fitting procedure, and thus, ZIGP regression may be considered as an alternative (Famoye and Singh 2006). In addition, when both zero-inflation and overdispersion exist in count data, ZIGP regression behaves similarly to ZINB regression. As examples, Yang et al. (2009b) proved that the score statistics for testing overdispersion in ZIP against ZIGP (ZIGP-1 and ZIGP-2) regressions and ZIP against ZINB (ZINB-1 and ZINB-2) regressions are equal, whereas Joe and Zhu (2005) proved that ZIGP distribution provides better fit than ZINB distribution when there is a large zero fraction and heavy tail, implying that the ZIGP regression can be used as an alternative for modeling zero-inflated and overdispersed count data.

The objectives of this study are:

- to relate NB and GP regressions through the mean-variance relationship,
- to suggest applications of these models for overdispersed or underdispersed claim count data,
- to relate ZINB and ZIGP regression models through the mean-variance relationship, and finally,
- to suggest applications of these zero-inflated models for zero-inflated and overdispersed claim count data.

2. NB REGRESSION MODELS

Let $(Y_1, Y_2, ..., Y_n)^T$ be the vector of count random variables and n be the sample size. The probability mass function (p.m.f.) for Poisson regression is,

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

with mean and variance $E(Y_i) = Var(Y_i) = \mu_i$. The mean or the fitted value can be assumed to follow a log link, $E(Y_i) = \mu_i = \exp(\mathbf{x_i}'\boldsymbol{\beta})$, where $\mathbf{x_i}$ denotes the vector of explanatory variables and $\boldsymbol{\beta}$ the vector of regression parameters. The maximum likelihood estimates can be obtained by maximizing the log likelihood.

The latent heterogeneity can be incorporated by rewriting the conditional mean of Poisson regression as (Cameron and Trivedi 1986, Cameron and Trivedi 1998, Winkelmann 2003, Greene 2008),

$$E(Y_i \mid \varepsilon_i) = \exp(\mathbf{x_i}'\boldsymbol{\beta} + \varepsilon_i) = k_i \mu_i ,$$

where $k_i = \exp(\varepsilon_i)$ is assumed to follow gamma distribution with mean 1 and variance $v^{-1} = a$, with probability density function (p.d.f.),

$$f(k_i) = \frac{v^v k_i^{v-1} \exp(-v k_i)}{\Gamma(v)}, \qquad k_i \ge 0, v > 0,$$

so that the conditional Poisson regression is,

$$\Pr(Y_i = y_i \mid k_i) = \frac{\exp(-k_i \mu_i)(k_i \mu_i)^{y_i}}{y_i !}, \quad y_i = 0,1,\ldots$$

The marginal distribution is NB regression with p.m.f.,

$$\begin{aligned} \Pr(Y_i = y_i) &= \int_0^\infty \frac{e^{-k_i \mu_i} (k_i \mu_i)^{y_i}}{y_i !} \frac{v^v e^{-v k_i} k_i^{\,v-1}}{\Gamma(v)} \, dk_i \\ &= \frac{\Gamma(y_i + v)}{y_i ! \Gamma(v)} \left(\frac{v}{v + \mu_i} \right)^v \left(\frac{\mu_i}{v + \mu_i} \right)^{y_i}, \qquad y_i = 0,1,2,\ldots, \end{aligned} \qquad (2)$$

where the mean is $E(Y_i) = \mu_i$, the variance is $Var(Y_i) = \mu_i(1 + v^{-1}\mu_i) = \mu_i(1 + a\mu_i)$, and $v^{-1} = a$ denotes the dispersion parameter. The NB regression in (2) is also referred as NB-2 regression. Another parameterization for NB regression is by letting $v = a^{-1}\mu_i$ in (2) to produce NB-1 regression, with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = \mu_i(1 + a)$. Another parameterization is the NB-P regression, which is produced by letting $v = a^{-1}\mu_i^{2-P}$ in (2), so that the mean is $E(Y_i) = \mu_i$ and the variance is $Var(Y_i) = \mu_i(1 + a\mu_i^{P-1})$, where a denotes the dispersion parameter and P the functional parameter (Greene 2008).

NB regressions reduce to Poisson regression in the limit as $a \to 0$, and display overdispersion when $a > 0$. In addition, NB-P regression reduces to NB-1 when $P = 1$ and reduces to NB-2 when $P = 2$. Therefore, NB-P regression parametrically nests both NB-1 and NB-2, and allows statistical tests of the two functional forms against a more general alternative. The mean of NB regressions can also be assumed to follow the log link, $E(Y_i) = \mu_i = \exp(\mathbf{x_i}'\boldsymbol{\beta})$, and the maximum likelihood estimates can be obtained by maximizing the log likelihood.

NB-P regression can be fitted to count data using *R program*. For faster convergence, estimated parameters from fitting Poisson regression can be used as initial values. Both NB-1 and NB-2 regressions can also be fitted using the same fitting procedure, by letting the functional parameter, *P*, to be fixed at *P*=1 and *P*=2 respectively for NB-1 and NB-2.

We can also use *MASS package* in *R* to fit Poisson and NB-2 regression models. If we use *glm* function and set *family=poisson*, we can get Poisson regression. If we use *glm.nb* function, we can get NB-2 regression.

3. GP REGRESSION MODELS

The p.m.f of GP distribution is (Consul and Famoye 1992),

$$\Pr(Y_i = y_i) = \frac{\theta(\theta + v y_i)^{y_i - 1} \exp(-\theta - v y_i)}{y_i!}, \quad y_i = 0, 1, 2, \ldots, \tag{3}$$

where $\theta > 0$ and $\max(-1, -\frac{\theta}{4}) < v < 1$. The mean and variance are $E(Y_i) = \mu = (1-v)^{-1}\theta$ and $Var(Y_i) = (1-v)^{-3}\theta = (1-v)^{-2}\mu$, where $(1-v)^{-2}$ denotes the dispersion factor and v the dispersion parameter.

There are two well known parameterizations for GP regression, referred as GP-1 and GP-2. By letting $\theta_i = (1-v)\mu_i$ in (3), GP-1 regression is produced with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = (1-v)^{-2}\mu_i$. GP-1 regression reduces to Poisson regression when $v = 0$, allows overdispersion when $v > 0$, and allows underdispersion when $v < 0$. A new form of GP-1 regression, which has the same properties but different form of p.m.f., was recently proposed in Zamani and Ismail (2012) by rewriting $v = a(1+a)^{-1}$ and $\theta_i = (1+a)^{-1}\mu_i$ in (3). The mean and variance of new GP-1 regression are $E(Y_i) = \mu_i$ and $Var(Y_i) = (1+a)^2\mu_i$, where a denotes the dispersion parameter. The other parameterization is GP-2 regression, involving the parameterization of $v = (1+\mu_i a)^{-1} a\mu_i$ and $\theta_i = (1+\mu_i a)^{-1}\mu_i$ in (3), with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = (1+a\mu_i)^2\mu_i$. Another parameterization is GP-P regression, which was recently proposed in Zamani and Ismail (2012), obtained using $\theta_i = (1+a\mu_i^{P-1})^{-1}\mu_i$ and $v = (1+a\mu_i^{P-1})^{-1}a\mu_i^{P-1}$ in (3).

The mean and variance of GP-P regression are $E(Y_i) = \mu_i$ and $Var(Y_i) = (1 + a\mu_i^{P-1})^2 \mu_i$, where a denotes the dispersion parameter and P the functional parameter.

GP regressions reduce to Poisson regression when $a = 0$, display overdispersion when $a > 0$, and display underdispersion when $a < 0$. In addition, GP-P reduces to new GP-1 when $P = 1$, and reduces to GP-2 when $P = 2$. Therefore, GP-P regression parametrically nests both new GP-1 and GP-2 regressions, and allows statistical tests of the two functional forms against a more general alternative. The mean of GP regressions can be assumed to follow the log link, $E(Y_i) = \mu_i = \exp(\mathbf{x_i}'\boldsymbol{\beta})$, and the maximum likelihood estimates can be obtained by maximizing the log likelihood.

GP-P regression can be fitted using *R program*. For faster convergence, estimated parameters from fitting Poisson regression can be used as initial values. The Poisson, new GP-1 and GP-2 regressions can also be fitted using the same fitting procedure, by letting the dispersion parameter, *a*, to be fixed at *a*=0 for Poisson and by letting the functional parameter, *P*, to be fixed at *P*=1 and *P*=2 respectively for new GP-1 and GP-2. An example of *SAS code* for fitting GP-1 and GP-2 regressions can be found in Yang et al. (2007) and Yang et al. (2009a) respectively.

We can also use *ZIGP package* in *R* to fit GP-2 regression models. If we use *est.zigp* function and set *fm.Z=NULL* (which is the weight for zero-inflation), we can get GP-2 regression. Table 1 summarizes the p.m.f., mean and variance of Poisson, NB and GP regression models.

Table 1: Poisson, NB and GP regression models

Regression model	P.m.f.	Mean and variance
Poisson	$\frac{\mu_i^{y_i}}{y_i!}\exp(-\mu_i)$	$E(Y_i)=\mu_i$ $Var(Y_i)=\mu_i$
NB-1	$\frac{\Gamma(y_i+\mu_i a^{-1})}{y_i!\Gamma(\mu_i a^{-1})}\left(\frac{\mu_i a^{-1}}{\mu_i a^{-1}+\mu_i}\right)^{\mu_i a^{-1}}\left(\frac{\mu_i}{\mu_i a^{-1}+\mu_i}\right)^{y_i}$	$E(Y_i)=\mu_i$ $Var(Y_i)=\mu_i(1+a)$
NB-2	$\frac{\Gamma(y_i+a^{-1})}{y_i!\Gamma(a^{-1})}\left(\frac{a^{-1}}{a^{-1}+\mu_i}\right)^{a^{-1}}\left(\frac{\mu_i}{a^{-1}+\mu_i}\right)^{y_i}$	$E(Y_i)=\mu_i$ $Var(Y_i)=\mu_i(1+a\mu_i)$
NB-P	$\frac{\Gamma(y_i+a^{-1}\mu_i^{2-P})}{y_i!\Gamma(a^{-1}\mu_i^{2-P})}\left(\frac{a^{-1}\mu_i^{2-P}}{a^{-1}\mu_i^{2-P}+\mu_i}\right)^{a^{-1}\mu_i^{2-P}}\left(\frac{\mu_i}{a^{-1}\mu_i^{2-P}+\mu_i}\right)^{y_i}$	$E(Y_i)=\mu_i$ $Var(Y_i)=\mu_i(1+a\mu_i^{P-1})$
GP-1	$\frac{((1-v)\mu_i)((1-v)\mu_i+vy_i)^{y_i-1}}{y_i!}\exp(-((1-v)\mu_i+vy_i))$	$E(Y_i)=\mu_i$ $Var(Y_i)=\mu_i(1-v)^{-2}$
New GP-1	$\frac{\mu_i(\mu_i+ay_i)^{y_i-1}}{(1+a)^{y_i}y_i!}\exp\left(-\frac{\mu_i+ay_i}{1+a}\right)$	$E(Y_i)=\mu_i$ $Var(Y_i)=\mu_i(1+a)^2$
GP-2	$\frac{\mu_i(\mu_i+a\mu_i y_i)^{y_i-1}}{(1+a\mu_i)^{y_i}y_i!}\exp\left(-\frac{\mu_i+a\mu_i y_i}{1+a\mu_i}\right)$	$E(Y_i)=\mu_i$ $Var(Y_i)=\mu_i(1+a\mu_i)^2$
New GP-P	$\frac{\mu_i(\mu_i+a\mu_i^{P-1}y_i)^{y_i-1}}{(1+a\mu_i^{P-1})^{y_i}y_i!}\exp\left(-\frac{\mu_i+a\mu_i^{P-1}y_i}{1+a\mu_i^{P-1}}\right)$	$E(Y_i)=\mu_i$ $Var(Y_i)=\mu_i(1+a\mu_i^{P-1})^2$

Several general comparisons can be made regarding NB and GP regression models:

- NB regressions reduce to Poisson regression in the limit as $a \to 0$, whereas GP regressions reduce to Poisson regression when $a = 0$.
- The mean-variance relationships of NB-1 and GP-1 are linear, NB-2 is quadratic, GP-2 is cubic, NB-P is to the *p*th power, and GP-P is to the (2*p*-1)th power.
- Likelihood ratio test (LRT) can be performed for testing overdispersion in Poisson vs. NB (NB-1 and NB-2) regressions, $H_0 : a = 0$ vs. $H_1 : a > 0$. Since the null hypothesis is on the boundary of parameter space, the asymptotic distribution for LRT statistic is a mixture of half of probability mass at zero and half of chi-square with one degree of freedom (Lawless 1987).
- Likelihood ratio test (LRT) can also be performed for testing dispersion (over or underdispersion) in Poisson vs. GP (GP-1 and GP-2) regressions, $H_0 : a = 0$ vs. $H_1 : a \neq 0$, where the LRT statistic is asymptotically distributed as a chi-square with one degree of freedom (Consul and Famoye 1992, Wang and Famoye 1997).
- In terms of properties, there is no big difference between NB and GP distributions when the means and variances are fixed. However, GP distribution has heavier tail, whereas NB distribution has more mass at zero (Joe and Zhu 2005).

4. ZIP REGRESSION MODEL

Zero-inflated Poisson (ZIP) regression has been used by researchers for handling purely zero-inflated count data. ZIP regression can be obtained by mixing a distribution degenerate at zero with the Poisson distribution, by allowing the incorporation of explanatory variables in both the zero process and the Poisson distribution. The p.m.f. of ZIP regression is,

$$\Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)\exp(-\mu_i), & y_i = 0 \\ (1-\omega_i)\dfrac{\mu_i^{y_i}}{y_i!}\exp(-\mu_i), & y_i > 0 \end{cases} \tag{4}$$

where $0 \leq \omega_i < 1$ and $\mu_i > 0$, with mean $E(Y_i) = (1-\omega_i)\mu_i$ and variance $Var(Y_i) = (1-\omega_i)\mu_i(1+\omega_i\mu_i)$. ZIP regression reduces to Poisson regression when $\omega_i = 0$, and exhibits overdispersion when $\omega_i > 0$. The covariates can be incorporated by using a log link for μ_i and a logit link for ω_i,

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \text{ and } \log\left(\frac{\omega_i}{1-\omega_i}\right) = \mathbf{z}_i^T \boldsymbol{\gamma}, \qquad (5)$$

where $\mathbf{x}_i$ and $\mathbf{z}_i$ are the vectors of explanatory variables, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the vectors of regression parameters. Maximum likelihood estimates can be obtained by maximizing the log likelihood.

5. ZINB REGRESSION MODELS

Zero-inflated negative binomial (ZINB) regressions have been used by researchers for handling both zero-inflation and overdispersion in count data. ZINB-1 regression can be obtained by mixing a distribution degenerate at zero with the NB-1 distribution, by allowing the incorporation of explanatory variables in both the zero process and the NB-1 distribution. The mean and variance of ZINB-1 regression are $E(Y_i) = (1-\omega_i)\mu_i$ and $Var(Y_i) = (1-\omega_i)\mu_i(1+a+\omega_i\mu_i)$.

ZINB-2 regression can be obtained by mixing a distribution degenerate at zero with the NB-2 distribution, by allowing the incorporation of explanatory variables in both the zero process and the NB-2 distribution. The mean and variance of ZINB-2 regression are $E(Y_i) = (1-\omega_i)\mu_i$ and $Var(Y_i) = (1-\omega_i)\mu_i(1+a\mu_i+\omega_i\mu_i)$.

The ZINB-P regression can be obtained by mixing a distribution degenerate at zero with the NB-P distribution, by allowing the incorporation of explanatory variables in both the zero process and the NB-P distribution. The mean and variance of ZINB-P regression are $E(Y_i) = (1-\omega_i)\mu_i$ and $Var(Y_i) = (1-\omega_i)\mu_i(1+a\mu_i^{P-1}+\omega_i\mu_i)$.

ZINB regressions reduce to NB regressions when $\omega_i = 0$, and reduce to ZIP regression in the limit as $a \to 0$. The variance of ZINB regressions exhibits overdispersion when $a > 0$ and $\omega_i > 0$, allowing the models to be used for handling both zero-inflated and overdispersed count data. The link functions for ZINB regressions can also be written as (5). Maximum likelihood estimates can be obtained by maximizing the log likelihood.

ZINB-P regression can be fitted using *R program*. For faster convergence, estimated parameters from fitting ZIP regression can be used as initial values. Both ZINB-1 and the ZINB-2 regressions can also be fitted using the same fitting procedure, by letting the functional parameter, *P*, to be fixed respectively at *P*=1 and *P*=2 for ZINB-1 and ZINB-2.

We can also use *PSCL package* in *R* to fit ZIP and ZINB-2 regression models. If we use *zeroinfl* function and set *dist="poisson"*, we can get ZIP regression. If we set *dist="negbin"*, we can get ZINB-2 regression.

6. ZIGP REGRESSION MODELS

As an alternative for handling zero-inflation and overdispersion in count data, zero-inflated generalized Poisson (ZIGP) regressions can be fitted. ZIGP-1 regression is obtained by mixing a distribution degenerate at zero with the GP-1 distribution, by allowing the incorporation of explanatory variables in both the zero process and the GP-1 distribution. The mean and variance of ZIGP-1 regression are $E(Y_i) = (1-\omega_i)\mu_i$ and $Var(Y_i) = (1-\omega_i)\mu_i((1-v)^{-2} + \omega_i\mu_i)$. A new form of ZIGP-1 regression has been proposed by Zamani and Ismail (2013a), by mixing a distribution degenerate at zero with the new GP-1 distribution, and allowing the incorporation of explanatory variables in both the zero process and the new GP-1 distribution. The mean and variance of new ZIGP-1 are $E(Y_i) = (1-\omega_i)\mu_i$ and $Var(Y_i) = (1-\omega_i)\mu_i((1+a)^2 + \omega_i\mu_i)$.

ZIGP-2 regression is obtained by mixing a distribution degenerate at zero with the GP-2 distribution, by allowing the incorporation of explanatory variables in both the zero process and the GP-2 distribution. The mean and variance of ZIGP-2 regression are $E(Y_i) = (1-\omega_i)\mu_i$ and $Var(Y_i) = (1-\omega_i)\mu_i((1+a\mu_i)^2 + \omega_i\mu_i)$.

A functional form of ZIGP, which is referred as ZIGP-P, was recently proposed by Zamani and Ismail (2013a). ZIGP-P regression is obtained by mixing a distribution degenerate at zero with the GP-P distribution, by allowing the incorporation of explanatory variables in both the zero process and the GP-P distribution. The mean and variance of ZIGP-P regression are $E(Y_i) = (1-\omega_i)\mu_i$ and $Var(Y_i) = (1-\omega_i)\mu_i((1+a\mu_i^{P-1})^2 + \omega_i\mu_i)$. When $\omega_i = 0$, ZIGP regressions reduce to GP regressions, and when $a = 0$, ZIGP regressions reduce to ZIP regression. In addition, when $P = 1$ and $P = 2$, ZIGP-P regression reduces to ZIGP-1 and ZIGP-2 regressions respectively. The variance of ZIGP regressions exhibit overdispersion when $a > 0$ and $\omega_i > 0$, indicating that they can be used for handling both zero-inflated and overdispersed count data. The link functions for ZIGP regressions can also be written as (5), and the maximum likelihood estimates can be obtained by maximizing the log likelihood.

ZIGP-P regression can be fitted using *R program*. For faster convergence, estimated parameters from fitting ZIP regression can be used as initial values. The ZIP, ZIGP-1 and ZIGP-2 regressions can also be fitted using the same fitting procedure, by letting the dispersion parameter, a, to be fixed at a=0 for ZIP and by letting the functional parameter, P, to be fixed respectively at P=1 and P=2 for ZIGP-1 and ZIGP-2. An example of *SAS code* for fitting ZIGP-2 regression can be found in Yang et al. (2009b).

We can also use *ZIGP package* in *R* to fit ZIP and ZIGP-2 regression models. If we use *est.zigp* function and set *fm.W=NULL* (which is the weight for dispersion), we can get ZIP regression. We can get ZINB-2 regression if both *fm.Z* (weight for zero-inflation) and *fm.W* (weight for dispersion) are not set to *NULL*. Table 2 summarizes the p.m.f., mean and variance of ZIP, ZINB and ZIGP regression models.

Several general comparisons can be made regarding ZINB and ZIGP regression models:

- ZINB regressions reduce to ZIP regression in the limit as $a \to 0$, whereas ZIGP regressions reduce to ZIP regression when $a = 0$.
- LRT can be performed for testing overdispersion in ZIP vs. ZINB regressions, where $H_0 : a = 0$ vs. $H_1 : a > 0$. Since the null hypothesis is on the boundary of parameter space, the asymptotic distribution for LRT statistic is a mixture of half of probability mass at zero and half of chi-square with one degree of freedom (Stram and Lee 1994, 1995).
- The signed square root of likelihood ratio test (SSN-LRT) can be performed for testing overdispersion in ZIP vs. ZIGP regressions, $H_0 : a = 0$ vs. $H_1 : a > 0$, where the statistic is asymptotically distributed as a standard normal (Yang et al. 2009b, Famoye and Singh 2006).
- Since GP distribution has heavier tail and NB distribution has more mass at zero, both ZINB and ZIGP regressions can be used for fitting zero-inflated and overdispersed count data (Joe and Zhu 2005).

Table 2: ZIP, ZINB and ZIGP regression models

Model	P.m.f.	Mean and variance
ZIP	$\begin{cases} \omega_i + (1-\omega_i)\exp(-\mu_i), & y_i = 0 \\ (1-\omega_i)\dfrac{\mu_i^{y_i}}{y_i!}\exp(-\mu_i), & y_i > 0 \end{cases}$	$E(Y_i) = (1-\omega_i)\mu_i$ $Var(Y_i) = E(Y_i)(1+\omega_i\mu_i)$
ZINB-1	$\begin{cases} \omega_i + (1-\omega_i)\left(\dfrac{\mu_i a^{-1}}{\mu_i a^{-1} + \mu_i}\right)^{\mu_i a^{-1}}, & y_i = 0 \\ (1-\omega_i)\dfrac{\Gamma(y_i + \mu_i a^{-1})}{y_i!\Gamma(\mu_i a^{-1})}\left(\dfrac{\mu_i a^{-1}}{\mu_i a^{-1} + \mu_i}\right)^{\mu_i a^{-1}}\left(\dfrac{\mu_i}{\mu_i a^{-1} + \mu_i}\right)^{y_i}, & y_i > 0 \end{cases}$	$E(Y_i) = (1-\omega_i)\mu_i$ $Var(Y_i) = E(Y_i)(1+a+\omega_i\mu_i)$
ZINB-2	$\begin{cases} \omega_i + (1-\omega_i)\left(\dfrac{a^{-1}}{a^{-1} + \mu_i}\right)^{a^{-1}}, & y_i = 0 \\ (1-\omega_i)\dfrac{\Gamma(y_i + a^{-1})}{y_i!\Gamma(a^{-1})}\left(\dfrac{a^{-1}}{a^{-1} + \mu_i}\right)^{a^{-1}}\left(\dfrac{\mu_i}{a^{-1} + \mu_i}\right)^{y_i}, & y_i > 0 \end{cases}$	$E(Y_i) = (1-\omega_i)\mu_i$ $Var(Y_i) = E(Y_i)(1+a\mu_i+\omega_i\mu_i)$
ZINB-P	$\begin{cases} \omega_i + (1-\omega_i)\left(\dfrac{a^{-1}\mu_i^{2-P}}{a^{-1}\mu_i^{2-P} + \mu_i}\right)^{a^{-1}\mu_i^{2-P}}, & y_i = 0 \\ (1-\omega_i)\dfrac{\Gamma(y_i + a^{-1}\mu_i^{2-P})}{y_i!\Gamma(a^{-1}\mu_i^{2-P})}\left(\dfrac{a^{-1}\mu_i^{2-P}}{a^{-1}\mu_i^{2-P} + \mu_i}\right)^{a^{-1}\mu_i^{2-P}}\left(\dfrac{\mu_i}{a^{-1}\mu_i^{2-P} + \mu_i}\right)^{y_i}, & y_i > 0 \end{cases}$	$E(Y_i) = (1-\omega_i)\mu_i$ $Var(Y_i) = E(Y_i)(1+a\mu_i^{P-1}+\omega_i\mu_i)$

Model	P.m.f.	Mean and variance
ZIGP-1	$\begin{cases} \omega_i + (1-\omega_i)\exp(-(1-v)\mu_i), & y_i = 0 \\ (1-\omega_i)\dfrac{((1-v)\mu_i)((1-v)\mu_i + vy_i)^{y_i-1}}{y_i!}\exp(-((1-v)\mu_i + vy_i)), & y_i > 0 \end{cases}$	$E(Y_i) = (1-\omega_i)\mu_i$ $Var(Y_i) = E(Y_i)((1-v)^{-2} + \omega_i\mu_i)$
New ZIGP-1	$\begin{cases} \omega_i + (1-\omega_i)\exp\left(-\dfrac{\mu_i}{1+a}\right), & y_i = 0 \\ (1-\omega_i)\dfrac{\mu_i(\mu_i + ay_i)^{y_i-1}}{(1+a)^{y_i}\, y_i!}\exp\left(-\dfrac{\mu_i + ay_i}{1+a}\right), & y_i > 0 \end{cases}$	$E(Y_i) = (1-\omega_i)\mu_i$ $Var(Y_i) = E(Y_i)((1+a)^2 + \omega_i\mu_i)$
ZIGP-2	$\begin{cases} \omega_i + (1-\omega_i)\exp\left(-\dfrac{\mu_i}{1+a\mu_i}\right), & y_i = 0 \\ (1-\omega_i)\dfrac{\mu_i(\mu_i + a\mu_i y_i)^{y_i-1}}{(1+a\mu_i)^{y_i}\, y_i!}\exp\left(-\dfrac{\mu_i + a\mu_i y_i}{1+a\mu_i}\right), & y_i > 0 \end{cases}$	$E(Y_i) = (1-\omega_i)\mu_i$ $Var(Y_i) = E(Y_i)((1+a\mu_i)^2 + \omega_i\mu_i)$
New ZIGP-P	$\begin{cases} \omega_i + (1-\omega_i)\exp\left(-\dfrac{\mu_i}{1+a\mu_i^{P-1}}\right), & y_i = 0 \\ (1-\omega_i)\dfrac{\mu_i(\mu_i + a\mu_i^{P-1} y_i)^{y_i-1}}{(1+a\mu_i^{P-1})^{y_i}\, y_i!}\exp\left(-\dfrac{\mu_i + a\mu_i^{P-1} y_i}{1+a\mu_i^{P-1}}\right), & y_i > 0 \end{cases}$	$E(Y_i) = (1-\omega_i)\mu_i$ $Var(Y_i) = E(Y_i)((1+a\mu_i^{P-1})^2 + \omega_i\mu_i)$

7. TESTS FOR NB AND GP REGRESSION MODELS

7.1 Likelihood Ratio Test (LRT)

Since NB regressions reduce to Poisson regression in the limit as $a \to 0$, the test of overdispersion in Poisson vs. NB-1 regressions and Poisson vs. NB-2 regressions, $H_0 : a = 0$ vs. $H_1 : a > 0$, can be performed using LRT, $T = 2(\ln L_1 - \ln L_0)$, where $\ln L_1$ and $\ln L_0$ are the models' log likelihood under their respective hypothesis. Since the null hypothesis is on the boundary of parameter space, the LRT is asymptotically distributed as half of probability mass at zero and half of chi-square with one degree of freedom (Lawless 1987). In other words, to test the null hypothesis at significance level α, the critical value of chi-square distribution with significance level 2α is used, or reject H_0 if $T > \chi^2_{1-2\alpha,1}$. As an example, for 0.05 significance level, the critical value is $\chi^2_{0.90,1} = 2.7055$ instead of $\chi^2_{0.95,1} = 3.8415$.

A two-sided test can be performed for dispersion (over or underdispersion) in Poisson vs. new GP-1 regressions and Poisson vs. GP-2 regressions, $H_0 : a = 0$ vs. $H_1 : a \neq 0$, using LRT which is asymptotically distributed as a chi-square with one degree of freedom. If we are interested in a one-sided test for overdispersion in Poisson vs. new GP-1 regressions and Poisson vs. GP-2 regressions, $H_0 : a = 0$ vs. $H_1 : a > 0$, we can use signed square root of likelihood ratio statistic (SSR-LRT), $\operatorname{sgn}(a)\sqrt{T} = \operatorname{sgn}(a)\sqrt{2(\ln L_1 - \ln L_0)}$, where $\operatorname{sgn}(.)$ is a sign function which indicates value of 1 when $a > 0$ and value of -1 when $a < 0$. Under H_0, SSR-LRT is asymptotically distributed as a standard Normal distribution.

A two-sided test for testing NB-1 vs. NB-P, NB-2 vs. NB-P, new GP-1 vs. GP-P and GP-2 vs. GP-P regressions can be performed using LRT. The hypothesis are $H_0 : P = 1$ vs. $H_1 : P \neq 1$ or $H_0 : P = 2$ vs. $H_1 : P \neq 2$, and the LRT is asymptotically distributed as a chi-square with one degree of freedom.

7.2 Wald Test

The test of overdispersion in Poisson vs. NB-1 regressions and Poisson vs. NB-2 regressions can also be performed using Wald statistic which is, $\frac{\hat{a}^2}{Var(\hat{a})}$, where $\hat{a}$ is the estimate of dispersion parameter and $Var(\hat{a})$ is its variance. Since the null hypothesis is on the boundary of parameter space, the Wald statistic is asymptotically distributed as half of probability mass at zero and half of chi-square with one degree of freedom. For testing dispersion (over or underdispersion) in Poisson vs. GP-1 regressions and Poisson vs. GP-2 regressions, the Wald statistic can also be applied, and the statistic is asymptotically distributed as a chi-square with one degree of freedom.

Using similar approach, the adequacy of NB-1 vs. NB-P, NB-2 vs. NB-P, new GP-1 vs. GP-P and GP-2 vs. GP-P regression can be performed using Wald statistic which is, $\frac{\hat{P}^2}{Var(\hat{P})}$, where $\hat{P}$ is the estimate of functional parameter and $Var(\hat{P})$ is its variance. The Wald statistic is asymptotically distributed as a chi-square with one degree of freedom.

7.3 Vuong Test

For non-nested models, a comparison between models with p.m.f. $p_1(.)$ and $p_2(.)$ can be performed using Vuong test (Vuong 1989), $V = \frac{\bar{m}\sqrt{n}}{sd(m)}$, where $\bar{m}$ is the mean of m_i, $sd(m)$ is the standard deviation of m_i, n the sample size and $m_i = \ln\left(\frac{p_{1i}(y_i)}{p_{2i}(y_i)}\right)$. The Vuong test statistic follows a standard normal. As an example, for 0.05 significance level, the first model is "closer" to the actual model if V is larger than 1.96. In the other hand, the second model is "closer" to the actual model if V is smaller than -1.96. Otherwise, neither model is "closer" to the actual model and there is no difference between using the first or the second model.

For models with unequal number of parameters, the equation for m_i in Vuong test is slightly modified to account for the difference in the number of parameters, $m_i = \ln\left(\frac{p_{1i}(y_i)}{p_{2i}(y_i)}\right) - \frac{k_1 - k_2}{2}\ln(n)$, where k_1 and k_2 are the number of parameters in model 1 and model 2 respectively.

7.4 AIC and BIC

When several models are available, one can compare the models' performance based on several likelihood measures which have been proposed in statistical literatures. Two of the most regularly used measures are Akaike Information Criteria (AIC) and Bayesian Schwartz Information Criteria (BIC). The AIC penalizes a model with larger number of parameters, and is defined as $\text{AIC} = -2\ln L + 2p$, where $\ln L$ denotes the fitted log likelihood and p the number of parameters. The BIC penalizes a model with larger number of parameters and larger sample size, and is defined as $\text{BIC} = -2\ln L + p\ln(n)$, where $\ln L$ denotes the fitted log likelihood, p the number of parameters and n the sample size.

8 TESTS FOR ZINB AND ZIGP REGRESSION MODELS

8.1 Likelihood Ratio Test

Since ZINB regressions reduce to ZIP regression in the limit as $a \to 0$, the test of overdispersion in ZIP vs. ZINB-1 regressions and Poisson vs. ZINB-2 regressions, $H_0 : a = 0$ vs. $H_1 : a > 0$, can be performed using LRT. Since the null hypothesis is on the boundary of parameter space, the LRT statistic is asymptotically distributed as half of probability mass at zero and half of chi-square with one degree of freedom (Stram and Lee 1994, 1995).

A two-sided test can be performed to test dispersion (over or underdispersion) in ZIP vs. new ZIGP-1 regressions and ZIP vs. ZIGP-2 regressions using LRT, $H_0 : a = 0$ vs. $H_1 : a \neq 0$. Since ZIP regression model is nested within new ZIGP-1 and ZIGP-2 regressions, the boundary problem does not exist and the statistic is asymptotically distributed as a chi-square with one degree of freedom. If we are interested in a one-sided test for overdispersion in ZIP vs. new ZIGP-1 regressions and ZIP vs. ZIGP-2 regressions, $H_0 : a = 0$ vs. $H_1 : a > 0$, we can use SSR-LRT which is asymptotically distributed as a standard Normal distribution.

A two-sided test for adequacy of ZINB-1 vs. ZINB-P, ZINB-2 vs. ZINB-P, new ZIGP-1 vs. ZIGP-P and ZIGP-2 vs. ZIGP-P regressions can be performed using LRT, where the hypothesis are $H_0 : P = 1$ vs. $H_1 : P \neq 1$ or $H_0 : P = 2$ vs. $H_1 : P \neq 2$. Since both ZINB-1 and ZINB-2 regressions are nested within ZINB-P regression, and both ZIGP-1 and ZIGP-2 regressions are nested within ZIGP-P regression, the LRT statistic is asymptotically distributed as a chi-square with one degree of freedom.

8.2 Wald Test

The test of overdispersion in ZIP vs. ZINB-1 regressions and ZIP vs. ZINB-2 regressions can also be performed using Wald statistic. Since the null hypothesis is on the boundary of parameter space, the Wald statistic is asymptotically distributed as half of probability mass at zero and half of chi-square with one degree of freedom. For dispersion (over or underdispersion) in ZIP vs. ZIGP-1 regressions and ZIP vs. ZIGP-2 regressions, the Wald statistic can also be applied, where the statistic is asymptotically distributed as a chi-square with one degree of freedom.

Using similar approach, the adequacy of ZINB-1 vs. ZINB-P, ZINB-2 vs. ZINB-P, new ZIGP-1 vs. ZIGP-P and ZIGP-2 vs. ZIGP-P regressions can be performed using Wald statistic, where the statistic is asymptotically distributed as a chi-square with one degree of freedom.

9. EXAMPLES

9.1 Malaysian Own Damage Claim Counts

The dataset for private car Own Damage (OD) claim counts analyzed in Zamani and Ismail (2012) is reconsidered here for fitting NB and GP regression models. The data was based on 1.01 million private car policies for a three-year period of 2001-2003, the exposures were expressed in car-year units, and the incurred claims consisted of claims already paid as well as outstanding. Table 3 shows the rating factors and rating classes for the exposures and incurred claims. The estimates of Poisson regression were used as initial values for fitting NB and GP regressions.

Table 3: Rating factors and rating classes (Malaysian OD data)

Rating factors	Rating classes
Vehicle age	0-1 year
	2-3 years
	4-5 years
	6-7 years
	8+ years
Vehicle c.c.	0-1000
	1001-1300
	1301-1500
	1501-1800
	1801+
Vehicle make	Local type 1
	Local type 2
	Foreign type 1
	Foreign type 2
	Foreign type 3
Location	North
	East
	Central
	South
	East Malaysia

Table 4 shows the parameter estimates and their *t*-ratios for the fitted models. The best Poisson model was chosen using backward stepwise based on both AIC (chose model 2 if $AIC_{\text{model 2}} < AIC_{\text{model 1}}$) and *p*-values (drop a covariate if it is not significant). The same Poisson covariates were then utilized for fitting NB and GP regressions to ensure that the LRT can be performed. The results in Table 4 indicate that the regression parameters for all models have similar estimates. As expected, NB and GP models provide similar inferences for the regression parameters, i.e. their t-ratios, in absolute value, are smaller than Poisson model. In particular, the estimates and *t*-ratios of NB-1 are closer to GP-1, and similar results are also observed between NB-2 and GP-2, and between NB-P and GP-P. These results are expected since the mean-variance relationships for NB-1 and GP-1 are linear, NB-2 is quadratic, GP-2 is cubic, and both NB-P and GP-P are in the *p*th power and the (2*p*-1)th power respectively.

For testing overdispersion in Poisson versus NB-1 regressions, $H_0 : a = 0$ vs. $H_1 : a > 0$, the likelihood ratio and the Wald *t* respectively are 2[-2182.75-(-3613.71)]=2861.92 and 14.44, indicating that the null hypothesis is rejected and the NB-1 is more adequate. The likelihood ratio and Wald *t* for testing overdispersion in Poisson versus NB-2, Poisson versus GP-1 and Poisson versus GP-2 regressions are (2840.50, 2896.96, 2691.08) and (9.93, 19.22, 12.27) respectively, also indicating that the data is overdispersed and NB-2, GP-1 and GP-2 regressions are better than Poisson regression.

Table 4: NB and GP regression models (Malaysian OD data)

Parameters	Poisson		NB-1		NB-2		NB-P		GP-1		GP-2		GP-P	
	Est.	*t*-ratio	Est.	*t*-ratio	Est.	*t*-ratio	Est.	*t*-ratio	Est.	*t*-ratio	Est.	*t*-ratio	Est.	*t*-ratio
Intercept	-3.04	-195.43	-3.05	-70.55	-3.20	-45.78	-3.09	-54.19	-3.06	-67.78	-3.17	-48.22	-3.09	-54.03
2-3 year	0.51	41.09	0.53	15.07	0.57	8.75	0.54	11.36	0.53	14.62	0.58	9.11	0.55	11.47
4-5 year	0.52	39.96	0.51	14.03	0.52	8.02	0.52	10.70	0.51	13.29	0.51	8.10	0.52	10.56
6-7 year	0.43	33.63	0.45	12.55	0.40	6.14	0.44	9.19	0.46	12.20	0.37	5.92	0.45	9.34
8+ year	0.24	19.11	0.24	6.84	0.27	4.19	0.25	5.40	0.24	6.55	0.29	4.48	0.26	5.45
1001-1300 cc	-0.31	-24.64	-0.28	-8.30	-0.12	-2.03	-0.24	-5.20	-0.27	-7.69	-0.08	-1.36	-0.23	-5.05
1301-1500 cc	-0.16	-14.83	-0.16	-5.17	0.10	1.69	-0.12	-2.86	-0.15	-4.81	0.19	3.03	-0.12	-3.04
1501-1800 cc	0.14	12.93	0.13	4.36	0.25	4.61	0.15	3.90	0.13	4.08	0.27	5.18	0.14	3.69
1801+ cc	0.11	10.83	0.11	3.77	0.33	5.92	0.16	4.05	0.11	3.64	0.34	6.05	0.15	3.95
Local type 2	-0.46	-32.41	-0.45	-11.55	-0.27	-4.04	-0.40	-7.94	-0.45	-11.07	-0.35	-5.47	-0.41	-8.23
Foreign type 1	-0.20	-19.17	-0.18	-6.12	-0.27	-5.38	-0.20	-5.32	-0.17	-5.54	-0.34	-6.79	-0.20	-5.16
Foreign type 2	0.18	11.50	0.21	4.80	0.34	6.34	0.26	5.19	0.22	4.96	0.31	5.93	0.25	5.10
Foreign type 3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
East	0.35	19.91	0.39	8.29	0.31	4.71	0.36	6.28	0.41	8.37	0.28	4.83	0.37	6.56
Central	0.32	29.43	0.31	10.21	0.31	5.28	0.31	7.50	0.31	9.63	0.28	4.76	0.31	7.39
South	0.26	20.40	0.26	7.20	0.36	5.87	0.30	6.36	0.26	6.92	0.39	6.77	0.30	6.42
East Malaysia	0.13	8.87	0.12	3.04	0.11	1.80	0.12	2.26	0.12	2.80	0.10	1.77	0.11	2.15
a	-	-	7.07	14.44	0.14	9.93	1.57	6.60	2.01	19.22	0.02	12.27	0.64	7.95
P	-	-	1.00	-	2.00	-	1.39	39.00	1.00	-	2.00	-	1.27	46.23
Log likelihood		-3613.71		-2182.75		-2193.46		-2114.38		-2165.23		-2268.17		-2111.91
AIC		7259.42		4399.50		4420.92		4264.77		4364.45		4570.33		4259.83
BIC		7328.30		4472.67		4494.10		4342.25		4437.63		4643.51		4337.31

For testing the adequacy of NB-1 against NB-P and GP-1 against GP-P regressions, $H_0 : P = 1$ vs. $H_1 : P \neq 1$, the likelihood ratio and Wald *t* respectively are (136.74, 106.64) and (39.00, 46.23), implying that the null hypothesis is rejected and both NB-P and GP-P are more adequate. The likelihood ratios for testing adequacy of NB-2 versus NB-P and GP-2 versus GP-P regressions, $H_0 : P = 2$ vs. $H_1 : P \neq 2$, are (158.16, 312.52), also indicating that both NB-P and GP-P are better models.

Based on AIC and BIC, GP-P model has the lowest value for both criteria, indicating that the GP-P is the best model. However, based on Vuong test between GP-P as the first model and NB-P as the second model, the statistic is 1.0647 (less than 1.96), indicating that neither model is preferred over the other.

9.2 German Healthcare Data

The German Socioeconomic Panel (GSOEP) data (Riphahn et al. 2003) which was analyzed in Zamani and Ismail (2013a) is reconsidered here for fitting ZINB and ZIGP regression models. For an illustration purpose, only the first 438 individual data were fitted, where the response variable is the number of doctor visit in the last three month and the covariates were gender, age, health satisfaction, marital status, working status and education years. Table 5 shows the mean and standard deviation of the selected variables.

Table 5: Descriptive summary (German healthcare data)

Variable	Measurement	Mean	Standard deviation
DOCVIS	Number of doctor visit in last three months	2.93	33.09
GENDER	Female=1; Male=0	0.51	0.25
AGE	Age in years	43.30	116.23
HSAT	Health satisfaction	6.84	4.89
MARRIED	Married=1; else=0	0.54	0.25
WORKING	Employed=1; else=0	0.79	0.17
EDUC	Years of schooling	12.45	9.77

As mentioned previously, the covariates of μ_i and ω_i for ZIP, ZINB and ZIGP regressions can be included via the log and logit link functions, as shown in (5), where the vectors $\mathbf{x}_i$ and $\mathbf{z}_i$ may or may not share the same components. For the case where the covariates are not utilized in ω_i, the log and logit functions can be rewritten as (Yip and Yau, 2005; Ozmen and Famoye, 2007),

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \text{ and } \log\left(\frac{\omega}{1-\omega}\right) = \tau . \tag{6}$$

The data were fitted to ZIP, ZINB and ZIGP regressions using link functions (6), and the estimates of ZIP regression were used as initial values for fitting ZINB and ZIGP regressions.

Table 6 shows the parameter estimates and their t-ratios for the fitted models. The best ZIP model was chosen based on *p*-values (drop a covariate if it is not significant). The same ZIP covariates were then utilized for fitting ZINB and ZIGP models to ensure that the LRT can be performed. Since ZINB-P model did not provide converged solutions, the results are not displayed in the table. The results in Table 6 indicate that the regression parameters for all models have similar estimates. As expected, both ZINB and ZIGP regressions provide similar inferences for the regression parameters, i.e. their t-ratios, in absolute value, are smaller than ZIP.

For testing overdispersion in ZIP versus ZINB-1 and ZIP versus ZINB-2 regressions, $H_0 : a = 0$ vs. $H_1 : a > 0$, the likelihood ratio and Wald *t* respectively are (393.40, 525.30) and (6.16, 3.99), indicating that the null hypothesis is rejected and both ZINB-1 and ZINB-2 regressions are more adequate. The Vuong test for choosing between non-nested models of ZINB-2 as the first model and ZINB-1 as the second model is 5.5287 (more than 1.96), indicating that the ZINB-2 is a better model.

The likelihood ratio and Wald *t* for testing overdispersion in ZIP versus ZIGP-1 and ZIP versus ZIGP-2 regressions are (524.00, 521.44) and (7.68, 6.31) respectively, also indicating that the data is overdispersed, and both ZIGP-1 and ZIGP-2 are better models than ZIP. For testing the adequacy of ZIGP-1 against ZIGP-P regressions, $H_0 : P = 1$ vs. $H_1 : P \neq 1$, and ZIGP-2 versus ZIGP-P regressions, $H_0 : P = 2$ vs. $H_1 : P \neq 2$, the likelihood ratio and Wald *t* respectively are (7.34, 9.90) and 9.22, implying that the null hypothesis is rejected and ZIGP-P is more adequate.

Table 6: ZIP, ZINB and ZIGP regression models (German healthcare data)

Parameter	ZIP		ZINB-1		ZINB-2		ZIGP-1		ZIGP-2		ZIGP-P	
	est.	*t*-ratio	est.	*t*-ratio	est.	*t*-ratio	est.	*t*-ratio	est.	*t*-ratio	est.	*t*-ratio
Intercept	2.49	26.26	2.33	13.46	2.29	8.00	2.33	10.10	2.34	8.36	2.42	9.13
GENDER	0.30	4.85	0.21	1.90	0.58	3.83	0.52	3.73	0.52	3.52	0.60	3.90
HSAT	-0.22	-18.67	-0.16	-7.23	-0.25	-7.95	-0.22	-8.19	-0.24	-7.44	-0.25	-8.10
MARRIED	0.26	4.25	0.14	1.42	0.20	1.37	0.13	1.06	0.18	1.23	0.17	1.17
WORKING	0.15	2.35	0.13	1.13	0.05	0.27	-0.06	-0.44	0.14	0.81	0.00	-0.02
τ	-0.33	-3.16	-0.22	-2.28	-1.69	-2.65	-1.38	-4.92	-0.99	-4.25	-1.29	-4.49
a	-	-	2.53	6.16	1.39	3.99	1.52	7.68	0.33	6.31	0.77	3.80
P	-	-	1.00	-	2.00	-	1.00	-	2.00	-	1.46	9.22
Log likelihood	-1136.48		-939.78		-873.83		-874.48		-875.76		-870.81	
AIC	2284.96		1893.56		1761.66		1762.97		1765.52		1757.62	
BIC	2309.45		1922.13		1790.24		1791.54		1794.09		1790.28	

Based on AIC and BIC, ZIGP-P model has the lowest AIC but ZINB-2 model has the lowest BIC. For chosing between ZINB-2 as the first model and ZIGP-P as the second model which involves non-nested models with different number of parameters, the Vuong test statistic is 1.7303 (less than 1.96), indicating that neither model is preferred over the other.

Table 7 shows the parameters, log likelihood, AIC and BIC for the best ZIGP-P and ZINB-2 models, chosen based on *p*-values (drop a covariate if it is not significant). Based on AIC and BIC, ZIGP-P model has the lowest value for both criteria. However, the Vuong test statistic for comparing between ZINB-2 as the first model and ZIGP-P as the second model, which are non-nested models, is 1.6803 (less than 1.96), indicating that neither model is preferred over the other.

Table 7: ZINB-2 and ZIGP-P models with significant covariates (German healthcare data)

Parameter	ZINB-2			ZIGP-P		
	est.	*t*-ratio	*p*-value	est.	*t*-ratio	*p*-value
Intercept	2.46	9.87	0.00	2.53	11.01	0.00
GENDER	0.55	3.68	0.00	0.59	3.91	0.00
HSAT	-0.26	-8.16	0.00	-0.26	-8.27	0.00
τ	-1.75	-2.67	0.00	-1.28	-4.72	0.00
a	1.43	4.11	0.00	0.78	3.94	0.00
P	2.00	-	-	1.46	9.31	0.00
Log likelihood	-874.77			-871.50		
AIC	1759.54			1755.01		
BIC	1779.95			1779.50		

10. CONCLUSIONS

This study has related NB and GP regressions through the mean-variance relationship and has shown applications of these models for overdispersed count data. In addition, this study has related ZINB and ZIGP regressions through the mean-variance relationship and has shown applications of these zero-inflated models for zero-inflated and overdispersed count data.

The Malaysian data for private car Own Damage (OD) claim counts analyzed in Zamani and Ismail (2012) has been reconsidered for fitting Poisson, NB and GP regression models. The results indicate that the regression parameters of all models have similar estimates and the t-ratios, in absolute value, for NB and GP models are smaller than Poisson model. The likelihood ratio and Wald t for testing overdispersion indicate that the data is overdispersed, and NB-1, NB-2, GP-1 and GP-2 models are better than Poisson model. The likelihood ratio and Wald t also imply that both NB-P and GP-P are the best two models. Based on AIC and BIC, GP-P model has the lowest value for both criteria. However, based on Vuong test between GP-P and NB-P models, neither model is preferred over the other.

The German Socioeconomic Panel (GSOEP) data (Riphahn et al. 2003) which was analyzed in Zamani and Ismail (2013a) has been reconsidered for fitting ZIP, ZINB and ZIGP regression models. Unfortunately, ZINB-P model did not provide converge solutions and the results were not displayed. For other models, the results indicate that the regression parameters of all models have similar estimates and the t-ratios, in absolute value, for ZINB and ZIGP models are smaller than ZIP model. The likelihood ratio and Wald t for testing overdispersion indicate that the data is overdispersed, and ZINB-1, ZINB-2, ZIGP-1 and ZIGP-2 models are better than ZIP model. For ZINB models, the Vuong test indicates that ZINB-2 regression is better than ZINB-1 regression, whereas for ZIGP models, the likelihood ratio and Wald t indicate that ZIGP-P regression is better than both ZIGP-1 and ZIGP-2 regressions. Based on AIC and BIC, ZIGP-P model has the lowest AIC but ZINB-2 model has the lowest BIC. However, based on Vuong test between ZINB-2 and ZIGP-P regressions, neither model is preferred over the other.

Several general comparisons can be made regarding NB and GP regression models. Firstly, NB regression reduces to Poisson regression in the limit as $a \to 0$, whereas GP regression reduces to Poisson regression when $a = 0$. Secondly, the mean-variance relationships of NB-1 and GP-1 are linear, NB-2 is quadratic, GP-2 is cubic, NB-P is to the pth power, and GP-P is to the $(2p\text{-}1)$th power. Thirdly, LRT can be performed for testing overdispersion in Poisson vs. NB regressions, where the statistic is asymptotically distributed as a mixture of half of probability mass at zero and half of chi-square with one degree of freedom (Lawless 1987). LRT can also be performed for testing dispersion (over or underdispersion) in Poisson vs. GP regressions, where the statistic is asymptotically distributed as a chi-square with one degree of freedom (Consul and Famoye 1992, Wang and Famoye 1997). And finally, in terms of properties, there is no big difference between NB and GP distributions when the means and variances are fixed. However, GP distribution has heavier tail, whereas NB distribution has more mass at zero (Joe and Zhu 2005).

Several general comparisons can also be made regarding ZINB and the ZIGP models. Firstly, ZINB regression reduces to ZIP regression in the limit as $a \to 0$, whereas ZIGP regression reduces to ZIP regression when $a = 0$. Secondly, LRT can be performed for testing overdispersion in ZIP vs. ZINB regressions, where the statistic is asymptotically distributed as a mixture of half of probability mass at zero and half of chi-square with one degree of freedom (Stram and Lee 1994, 1995). SSN-LRT can be performed for testing overdispersion in ZIP vs. ZIGP regressions, where the statistic is asymptotically distributed as a standard normal (Yang et al. 2009b, Famoye and Singh 2006). And finally, in terms of properties, GP distribution has heavier tail whereas NB distribution has more mass at zero, indicating that both ZINB and ZIGP regressions can be used for fitting zero-inflated and overdispersed count data (Joe and Zhu 2005).

In this study, we have fitted a variety of models to two different datasets, involving several forms of NB, GP, ZINB and ZIGP regressions. The selection of the best model depends on many considerations. First, we can check for overdispersion, and if the data is slightly overdispersed, we can fit the data to quasi-Poisson regression. Secondly, if our data is largely overdispersed and it is not caused by excessive zeros but due to variation in the data, we can fit NB and GP regressions. Thirdly, if our data is both overdispersed and zero-inflated, we can fit zero-inflated (ZINB, ZIGP) and hurdle (HNB, HGP) regressions. Nevertheless, the hurdle regression models were not discussed in this study. Finally, the choice between zero-inflated (ZINB, ZIGP) and hurdle (HNB, HGP) models should be based upon a priori knowledge of the cause of excessive zeros in the data. Zero-inflated models are interpreted as a mix of structural and sampling zeros from two processes; the process that generates excess zeros from a binary distribution which are the structural zeros, and the process that generates both non-negative and zero counts from Poisson or NB distributions which are the sampling zeros. In the other hand, hurdle models assume that all zeros are sampling zeros. Therefore, as a crude guideline, if occurrences of count event do not depend on any condition and may occur at any time, the hurdle models should be fitted. However, if occurrences of count events depend on specific conditions and/or time, such as the case of deductible or no claim discount in insurance data, the zero-inflated models are more appropriate.

Acknowledgement

The authors gratefully acknowledge the financial support received in the form of research grants (UKM-ST-06-FRGS0108-2009 and GUP-2012-024) from the Ministry of Higher Education (MOHE), Malaysia. The authors are also pleasured to thank Insurance Services Malaysia Berhad (ISM) for supplying the data.

REFERENCES

[1.] Agarwal, D.K., Gelfand, A., Citron-Pousty, S. 2002. Zero-inflated model with application to spatial count data. *Environmental and Ecological Statistics*. 9: 341-355.

[2.] Aitkin, M., Anderson, D., Francis, B., Hinde, J. 1990. *Statistical modelling in GLIM*. New York: Oxford University Press.

[3.] Bohning, D., Dietz, E., Schlattman, P., Mendonca, L., Kirchner, U. 1996. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A*. 162: 195-209.

[4.] Boucher, J.P., Denuit, M., Guillen, M. 2007. Risk classification for claim count: a comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal*. 11(4): 110-131.

[5.] Brockmann, M.J., Wright, T.S. 1992. Statistical motor rating: making effective use of your data*, Journal of the Institute of Actuaries*. 119(3): 457-543.

[6.] Cameron, A.C., Trivedi, P.K. 1986. Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*. 1: 29-53.

[7.] Cameron, A.C., Trivedi, P.K. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.

[8.] Consul, P.C. 1989. *Generalized Poisson Distribution: Properties and Application*. New York: Marcel Dekker.

[9.] Consul, P.C., Famoye, F. 1992. Generalized Poisson regression model. *Communications in Statistics (Theory & Method)*. 2(1): 89-109.

[10.] Consul, P.C., Jain, G.C. 1973. A generalization of the Poisson distribution. *Technometrics*. 15: 791-799.

[11.] Conway, R. W., Maxwell, W. L. 1962. A queuing model with state dependent service rates. *Journal of Industrial Engineering*. 12: 132–136

[12.] Denuit, M., Marechal, X., Pitrebois, S., Walhin, J.F. 2007. *Actuarial Modeling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. John Wiley and Sons: England.

[13.] Famoye, F., Wulu, J.T., Singh, K.P. 2004. On the generalized Poisson regression model with an application to accident data. *Journal of Data Science*. 2: 287-295.

[14.] Famoye, F., Singh, K.P. 2006. Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*. 4: 117-130.

[15.] Greene, W. 2008. Functional forms for the negative binomial model for count data. *Economics Letters*. 99: 585-590.

[16.] Hilbe, J. 2007. *Negative Binomial Regression*. Cambridge, UK: Cambridge University Press.

[17.] Ismail, N., Jemain, A.A. 2007. Handling overdispersion with negative binomial and generalized Poisson regression models. *Casualty Actuarial Society Forum*. Winter: 103-158.

[18.] Joe, H., Zhu, R. 2005. Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal*. 47: 219-229.

[19.] Lambert, D. 1992. Zero-inflated Poisson regression, with an application to random defects in manufacturing. *Technometrics*. 34: 1-14.

[20.] Lawless, J.F. 1987. Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*. 15(3): 209-225.

[21.] Lord, D., Guikema, S.D., Geedipally, S.R. 2008. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis & Prevention*. 40(3): 1123–1134.

[22.] Lord, D., Geedipally, S.R., Guikema, S.D. 2010. Extension of the application of Conway-Maxwell-Poisson models: analyzing traffic crash data exhibiting under-dispersion. *Risk Analysis*. 30(8): 1268-1276.

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

[24.] Neelon, B.H., O'Malley, A.J., Normand, S.T. 2010. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modelling*. 10(4): 421–439.

[25.] Ozmen, I., Famoye, F. 2007. Count regression models with an application to zoological data containing structural zeros. *Journal of Data Science*. 5: 491-502.

[26.] Renshaw, A.E., 1994. Modelling the claims process in the presence of covariates. *ASTIN Bulletin*. 24(2): 265-285.

[27.] Ridout, M.S., Hinde, J.P., Demetrio, C.G.B. 2001. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*. 57: 219-223.

[28.] Riphahn, R., Wambach, A., Million, A. 2003. Incentive effects in the demand for health care: a bivariate panel count data estimation. *Journal of Applied Econometrics*. 18(4): 387-405.

[29.] Shmueli, G., Minka T., Kadane, J.B., Borle, S., Boatwright, P.B. 2005. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54: 127–142.

[30.] Stram, D.O., Lee, J.W. 1994. Variance components testing in the longitudinal mixed effects model. *Biometrics*. 50: 1171-1177.

[31.] Stram, D.O., Lee, J.W. 1995. Correction to "Variance components testing in the longitudinal mixed effects model". *Biometrics*. 51: 1196.
[32.] Vuong, Quang H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*. 57(2): 307–333.
[33.] Wang, W., Famoye, F. 1997. Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics*. 10: 273-283.
[34.] Winkelmann, R. 2008. *Econometric Analysis of Count Data*. Heidelberg: Springer Verlag.
[35.] Yang, Z., Hardin, J.W., Addy, C.L. Vuong, Q.H. 2007. Testing approaches for overdispersion in Poisson regression versus the generalized Poisson model. *Biometrical Journal*. 49: 565-584.
[36.] Yang, Z., Hardin, J.W., Addy, C.L. 2009a. A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. *Journal of Statistical Planning and Inference*. 139: 1514-1521.
[37.] Yang, Z., Hardin, J.W., Addy, C.L. 2009b. Testing overdispersion in the zero-inflated Poisson model. *Journal of Statistical Planning and Inference*. 139: 3340-3353.
[38.] Yip, K.C.H., Yau, K.K.W. 2005. On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*. 36: 153-163.
[39.] Zamani, H., Ismail, N. 2012. Functional form for the generalized Poisson regression model. *Communications in Statistics (Theory and Methods)*. 41(20): 3666–3675.
[40.] Zamani, H., Ismail, N. 2013a. Functional Form for the zero-inflated generalized Poisson regression model. *Communications in Statistics (Theory and Methods)*. (in press)
[41.] DOI: 10.1080/03610926.2012.665553.
[42.] Zamani, H., Ismail, N. 2013b. Score tests for overdispersion in Poisson regression model against generalized Poisson alternatives and zero-inflated Poisson regression model against zero-inflated generalized Poisson alternatives. (submitted).

Biographies of Authors

Noriszura Ismail is an Associate Professor in Universiti Kebangsaan Malaysia, teaching Actuarial Science courses since July 1993, and is currently pursuing research activities in Actuarial Science and Statistics areas with her PhD students. She received her PhD (Statistics) in 2006 from Universiti Kebangsaan Malaysia, and MSc (Actuarial Science) and BSc (Actuarial Science) in 1993 and 1991 from University of Iowa.

Hossein Zamani is a lecturer in Hormozgan University, Iran, teaching Statistics courses and is currently pursuing research activities in Statistics. He received his PhD (Statistics) in 2011 from Universiti Kebangsaan Malaysia, MSc (Statistics) in 2004 from Shiraz University, Iran and BSc (Statistics) in 2002 from Razi University, Iran.