


# Modelling Auto Insurance Claims in Singapore

Yogita S. Wagh and Kirtee K. Kamalja

Department of Statistics, School of Mathematical Sciences,  
North Maharashtra University, Jalgaon-425001, Maharashtra, India

\*Corresponding Author: email address: [kirteekamalja@gmail.com](mailto:kirteekamalja@gmail.com)

 <http://orcid.org/0000-0001-8200-7596>

Received: 10<sup>th</sup> February, 2017/ Revised: 10<sup>th</sup> October, 2017/ Accepted: 28<sup>th</sup> October, 2017  
©IAppStat-SL2017

## ABSTRACT

*Claim frequency data in general insurance may not follow the traditional Poisson distribution when there are many zeros. When the number of observed zeros exceeds the number of expected zeros under the Poisson distribution, extra dispersion appears. This paper summarizes several dispersed and zero-inflated count data models, which are used to handle dispersion and excess zeros. We model the insurance claim count data with excess zeros with these models. We use chi-square goodness-of-fit, to test the validity of the assumption of the count data distribution and fit count data regression model with predictors. We compare the fits through AIC and BIC. The generalized Poisson model and Negative binomial model provide a good fit to the data.*

**Keywords:** Claim frequency, zero-inflated Poisson, zero-inflated Negative Binomial, zero-inflated generalized Poisson model.

## 1. Introduction

The modeling of count data is of a primary interest in many fields such as insurance, public health, epidemiology, psychology, and many other research areas. The Poisson model is most commonly used for modelling count data. It assumes that the mean and variance are equal. However, this assumption is violated in many applications. The situation in which the variance is larger (smaller) than the mean is known as overdispersion (under-dispersion). The dispersion occurs when the single parameter  $\theta$  of the Poisson distribution is unable to fully describe event counts. For modelling the dispersed data, a



choice of analysis is a Negative binomial (NB) model and a Generalized Poisson (GP) model (introduced by Consul and Jain, 1973).

Generally, two sources of over-dispersion are determined: heterogeneity of the population and excess of zeroes. The heterogeneity is observed when the population can be divided into many homogeneous subpopulations. The excess of zeroes is detected when the number of observed zeroes exceeds largely the number of zeroes produced by the fitted Poisson model, i.e. when the frequency of 'zero' is significantly higher than the one predicted by the Poisson model.

Generally, in insurance a precise ratemaking system allows insurers to cover expected losses, expenses and make adequate provision for contingencies. The first step in ratemaking is to model the claim frequency distribution. Traditionally, the claim count distribution in general insurance is assumed to follow the Poisson and NB distributions (Thomas & Samson, 1987; Renshaw, 1994). Under the usual deductible agreement in general insurance policies, a claim will not be created unless the loss exceeds the prescribed deductible limit. Furthermore, the no claim discount (NCD) system, which is widely adopted by automobile insurers, leads to excess zero claims because policyholders seldom make a claim if the amount to be claimed is small. Such practice results in excess zeros in the observed claim frequency distribution, even though the original accident count distribution follows a Poisson or NB distribution. Yip and Yau (2005) discussed the claims in motor insurance data may contain excess zeros due to the conditions of deductible and no claim discount that discourage insured drivers to report small claims. Neyman (1939) and Feller (1943) introduced the concept of zero-inflation for the problem of more zeros. The zero-inflation phenomenon is a very particular type of over-dispersion, which is specifically handled by the zero-inflated count data models.

There are many situations in insurance, econometric, medical, engineering, manufacturing, public health, road safety, epidemiology, etc. where zero-inflation can be observed. Lambert (1992) studies some of the real life situations (manufacturing). Some other work can be found in Yip and Yau, (2005) (motor insurance), Mouatassim and Ezzahid, (2012) (health insurance), Neelon *et al.*, (2010) (healthcare data), Pandey, (1964) (number of flowers of the plants with *Primula Veris*), Yip, (1988) (number of insects per leaf), Bohning *et al.*, (1999) (dental epidemiology) etc.

The zero-inflation phenomenon is common in general insurance practice and appropriate modelling of such a data structure is necessary to precisely fit the claim frequency distribution. However, the use of zero-inflated Poisson (ZIP)

model receives little attention in the insurance and actuarial literature (Lambert, 1992). Among mostly discussed zero-inflated models, most common models are zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB) and zero-inflated generalized Poisson (ZIGP). In this paper, an application of the ZIP, ZIGP and ZINB distribution on the modeling of the claim frequency is discussed. The paper is organized as follows.

Description of the automobile insurance dataset is given in Section 2. In Section 3, we describe various count data models and zero-inflated models, which are most common for modeling over-dispersed and zero-inflated count data. In Section 4, we discuss goodness of fit measures, fit various count data models to automobile insurance dataset and compare the fitted models. In Section 5, we demonstrate the performance of different zero-inflated models and compare the results with the Poisson, GP and NB models via the log-likelihood and related statistics. Finally, in Section 6, we provide concluding remarks.

## **2. Insurance claim data from Singapore**

These data are from a portfolio of year 1993, which includes 7,483 automobile insurance policies from a major insurance company in Singapore. The data are described in Table 2 and provide the distribution of the claim counts. The variable of interest is the number of insurance claims per policyholder. For this dataset, it turns out that the maximum number of accidents in a year was three. There were on an average 0.06989 accidents per person. Frees and Valdez (2008) investigated hierarchical models of Singapore driving experience. Here, we consider a subset of this data, focusing on counts of automobile accidents and claims in 1993. The purpose of the analysis is to understand the impact of vehicle and driver characteristics on accident experience. We compare the performance of different zero-inflated models with the Poisson, GP and NB models to find the most suitable fit for modeling automobile insurance data.

### **2.1 Description of data**

The number of observations in the present database is 7,483. Around 90.64% of the policyholders are males. From this database, several characteristics were available to explain the number of claims. These characteristics include different vehicle variables, such as type and age of vehicle, as well as the variables related to driver, such as age, sex, and prior driving experience.

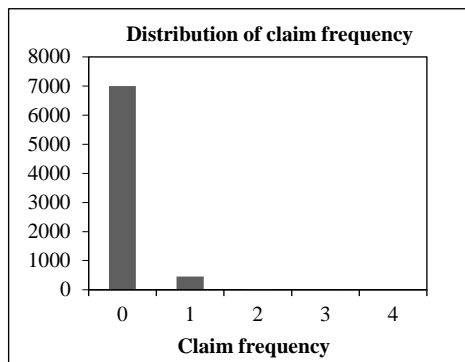
**Table 1:** Variable description of the automobile insurance data set

<b>Variable</b>	<b>Type</b>	<b>Description</b>
Clm_Count	Discrete	Number of claims during the year
Female	Binary	Takes the value 1 if the insured is female, 0 if the insured is male
AgeCat	Categorical	The age of the policyholder, in years grouped into seven categories. 0-6 indicate age groups 21 and younger, 22-25, 26-35, 36-45, 46-55, 56-65, 66 and over, respectively
NCD	Categorical	No claims discount. This is based on the previous accident record of the policyholder. It indicates the categories from 0,10,20,30, 40 and 50. The higher the discount, the better is the prior accident record
VAgeCat	Categorical	The age of the vehicle, in years, grouped into seven categories. 0-6 indicates groups 0, 1, 2, 3-5, 6-10, 11-15, 16 and older, respectively.
PC	Binary	Takes the value 1 if private vehicle, 0 otherwise
Auto Age	Categorical	It is a combination of Private vehicle takes the value 1 and VAgeCat as 0, 1 or 2, otherwise it is 0
Vehicle Type	Categorical	The type of vehicle being insured with levels A, G, M, P, Q, S, T, W, Z
Exp_weights	Numeric	Exposure weight or the fraction of the year that the policy is in effect
LNWEIGHT	Numeric	Logarithm of exposure weight, (log(Exp_weights))

**Table 2:** Observed claim counts in the motor insurance data

Observed Claim Counts	Frequency	% claim count
0	6996	93.49
1	455	6.08
2	28	0.37
3	4	0.05
4	0	0.00

**Fig.1.** Claim count distribution of the motor insurance data set



### 3. Dispersed and zero-inflated count data models

If the count data contains excess zeros along with over-dispersion, one may consider ZIGP or ZINB model to fit the data. Several forms of GP model have been introduced by the researchers. The primary interest to introduce the different functional forms of GP model is to use the proper mean-variance interrelation in a regression context. The other objective is to achieve flexibility in the development of various test procedures and inferential advantages.

Zamani and Ismail (2012) introduced a functional form of the GP regression model, which is referred as the GP-P model, that parametrically nests the Poisson and the two well-known GP regression models (GP-1 and GP-2). Zamani and Ismail (2014) proposed a functional form of ZIGP regression model, which mixes a distribution degenerate at zero with the GP-P distribution. The ZIGP-1 and ZIGP-2 regression models are particular cases of ZIGP-P model with  $P = 1$  and  $P = 2$ , respectively. We summarize the functional forms of several dispersed and zero-inflated count data models in Table 3, along with its mean and variance.

**Table 3:** Various forms of dispersed and zero-inflated count data regression models with their mean and variance

Sr. No.	Count data regression Models with introducer	Parameter of the regression model	pmf of $Y: P(Y_i = y_i   \text{parameters}), y_i = 0, 1, 2, \dots, \theta_i > 0$	(Mean, variance)
1.	Poisson (Frome <i>et al.</i> (1973))	$\theta_i$	$\frac{\theta_i^{y_i}}{y_i!} \exp(-\theta_i)$	$E(Y_i) = \theta_i$ $V(Y_i) = \theta_i$
2.	GP-1 (Consul and Jain (1973))	$(\theta_i, \varphi)$	$\frac{\theta_i(\theta_i + \varphi y_i)^{y_i-1}}{(1 + \varphi)^{y_i} y_i!} \exp\left(-\frac{\theta_i + \varphi y_i}{1 + \varphi}\right)$	$E(Y_i) = \theta_i$ $V(Y_i) = \theta_i(1 + \varphi)^2$ Here $\varphi$ is dispersion parameter
3.	GP-1 (Consul (1989))	$(\theta_i, \varphi)$	$((1 - \varphi)\theta_i + \varphi y_i)^{y_i-1} \frac{\theta_i}{y_i!} \exp(-(1 - \varphi)\theta_i - \varphi y_i)$	$E(Y_i) = \theta_i$ $V(Y_i) = \varphi E(Y_i)$ , Here $\varphi = \frac{1}{(1 - \varphi)^2}$ is dispersion factor
4.	GP (Consul and Famoye (1992))	$(\theta_i, \varphi)$	$\theta_i(\theta_i + y_i(\varphi - 1))^{y_i-1} \frac{\varphi^{-y_i}}{y_i!} \exp(-(\theta_i + y_i(\varphi - 1))/\varphi)$	$E(Y_i) = \theta_i$ $V(Y_i) = \theta_i \varphi^2$
5.	GP-2 (Famoye (1993))	$(\theta_i, \varphi)$	$\left(\frac{\theta_i}{1 + \varphi \theta_i}\right)^{y_i} \frac{(1 + \varphi y_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\theta_i(1 + \varphi y_i)}{1 + \varphi \theta_i}\right)$	$E(Y_i) = \theta_i$ $V(Y_i) = \theta_i(1 + \varphi \theta_i)^2$

Sr. No.	Count data regression Models with introducer	Parameter of the regression model	pmf of Y: $P(Y_i = y_i   \text{parameters}), \theta_i > 0$	(Mean, variance)
6.	ZIP (Lambert (1992))	$(\theta_i, \omega_i)$	$\begin{cases} \omega_i + (1 - \omega_i)e^{-\theta_i}, & y_i = 0 \\ (1 - \omega_i) \frac{\theta_i^{y_i}}{y_i!} e^{-\theta_i}, & y_i > 0 \end{cases}$	$\begin{aligned} E(Y_i) &= (1 - \omega_i)\theta_i \\ V(Y_i) &= (1 - \omega_i)\theta_i(1 + \omega_i\theta_i) \end{aligned}$
7.	ZIGP-1 (Yang, Hardin and Addy (2009))	$(\theta_i, \varphi, \omega_i)$	$\begin{cases} \omega_i + (1 - \omega_i)e^{-(1-\varphi)\theta_i} & y_i = 0 \\ (1 - \omega_i)((1 - \varphi)\theta_i + \varphi y_i)^{y_i-1} \frac{(1 - \varphi)\theta_i}{y_i!} \exp(-(1 - \varphi)\theta_i - \varphi y_i) & y_i > 0 \end{cases}$	$\begin{aligned} E(Y_i) &= (1 - \omega_i)\theta_i \\ V(Y_i) &= (1 - \omega_i)\theta_i \left( \omega_i\theta_i + \frac{1}{(1 - \varphi)^2} \right) \end{aligned}$
8.	ZIGP-2 (Yang, Hardin and Addy (2009))	$(\theta_i, \varphi, \omega_i)$	$\begin{cases} \omega_i + (1 - \omega_i) \exp\left(-\left(\frac{\theta_i}{1 + \varphi\theta_i}\right)\right) & y_i = 0 \\ (1 - \omega_i) \left(\frac{\theta_i}{1 + \varphi\theta_i}\right)^{y_i} \frac{y_i!}{(1 + \varphi y_i)^{y_i-1}} \exp\left(-\frac{\theta_i(1 + \varphi y_i)}{1 + \varphi\theta_i}\right) & y_i > 0 \end{cases}$	$\begin{aligned} E(Y_i) &= (1 - \omega_i)\theta_i \\ V(Y_i) &= (1 - \omega_i)\theta_i(\omega_i\theta_i + (1 + \varphi\theta_i)^2) \end{aligned}$
9.	ZINB (Yau, Wang and Lee (2003))	$(\theta_i, \varphi, \omega_i)$	$\begin{cases} \omega + (1 - \omega) \left(\frac{1}{1 + \varphi\theta}\right)^{\varphi-1} & y = 0 \\ (1 - \omega) \left(\frac{1 + y - 1}{y}\right) \left(\frac{1}{1 + \varphi\theta}\right)^{\varphi-1} \left(\frac{\varphi\theta}{1 + \varphi\theta}\right)^y & y > 0 \end{cases}$	$\begin{aligned} E(Y_i) &= (1 - \omega_i)\theta_i \\ V(Y_i) &= (1 - \omega_i)(\theta_i + \theta_i^2\varphi) \end{aligned}$

#### 4. Goodness-of-fit statistic and model selection criteria

Data is said to be over-dispersed if the conditional variance exceeds the conditional mean. An indication of the magnitude of over-dispersion or under-dispersion can be obtained simply by comparing the sample mean and variance of the dependent count variable using the dispersion index, which is the ratio of variance and mean. Here the dispersion index is calculated as, 1.0832, it shows that the data is over-dispersed. The zero-inflation index (ZI index) is a measure of detecting zero-inflation from the Poisson distribution. If  $Y$  is a nonnegative integer random variable (count variable) with mean  $\theta$  and  $p_0$  is the proportion of zeros in a random sample of size  $n$  then ZI index,  $z_i$  for the sample is defined as,

$$z_i = 1 + \log(p_0)/\theta$$

For the large sample from Poisson distribution  $z_i$  is close to zero with high probability.

The proportion of zeros in the insurance claim count data is 0.95 and ZI index is 0.037150. We fit various count data models to insurance claim count data to identify the best fit. The Chi-squared statistic is used to assess the goodness of fit of the distribution. For large sample sizes, the distribution of the Chi-squared statistic is approximately a Chi-squared with  $n - k - 1$  degrees of freedom, where  $n$  is the number of observations and  $k$  is the number of parameters. A significant  $p$  value indicates that the model does not fit the data well and another model with an additional parameter or parameters may be considered a significant improvement over the nested model.

In assessing the performance of the models and for model selection, the zero-inflated models are compared with the Poisson, GP, and NB models by means of the log-likelihood, Akaike's information criteria (AIC) and Bayesian information criterion (BIC). In general, the smaller is the AIC and BIC, the better is the model fit. If  $k$  is the number of parameters estimated and  $n$  is the number of observations in the data, the AIC and BIC are defined as follows.

$$\begin{aligned} AIC &= -2 \log \text{likelihood} + 2k \\ BIC &= -2 \log \text{likelihood} + k \ln(n) \end{aligned}$$

Results of fitting the claim frequency distribution by using various models are given in Table 4. Based on the chi-square statistic, the Poisson distribution does



not provide an adequate fit to the automobile insurance data. The Poisson distribution is inappropriate in modeling the automobile claim frequency data. Based on the chi-square statistic ( $\chi^2 = 1.7599$ ), the ZIGP-1 distribution provides a good fit to the data. However, it should be noted that the Poisson variance is inflated by the dispersion parameter of the GPD. Therefore, the GPD, NB distribution indicate an adequate fit according to the  $\chi^2$  statistic when the underlying count, random variable is in fact zero-inflated.

**Table 4:** Comparison of Fitted Models

$x$	Freq.	Poisson	GP-1	GP-2	NB	ZIP	ZIGP-1	ZIGP-2	ZINB
0	6996	6977.86	6996.43	6996.44	6996.41	6996.00	6996.44	6996.38	6996.00
1	455	487.69	452.83	452.82	452.78	452.69	452.83	452.90	452.69
2	28	17.04	31.24	31.24	31.41	32.68	31.24	31.22	32.68
3	4	0.40	2.30	2.30	2.23	1.57	2.30	2.30	1.57
4	0	0.01	0.18	0.18	0.16	0.06	0.16	0.16	0.06
Total	7483	7483	7483	7483	7483	7483	7483	7483	7483
Proportion of zeros		0.9325	0.9350	0.9350	0.9350	0.9349	0.9350	0.9350	0.9349
$\hat{\theta}$		0.0699	0.0699	0.0699	0.0699	0.1444	0.0699	0.0699	0.1443
$\hat{\phi}$		NA	0.0381	0.6609	1.1441	NA	0.0381	0.6609	0.0002
$\hat{\omega}$		NA	NA	NA	NA	0.5159	0.000008	0.000005	0.5158
$\chi^2$		41.9843	1.7786	1.7831	1.9502	4.4851	1.7599	1.7631	4.4840
$df$		3	2	2	2	2	1	1	1
$pvalue$		0.00000	0.410951	0.410903	0.377149	0.106187	0.184640	0.184232	0.034215
Log – likelihood		-1941.18	1932.35	1932.35	-1932.38	-1933.17	1932.35	1932.35	-1933.00
AIC		3884.35	3868.70	3868.70	3868.77	3870.34	3870.70	3870.70	3872.34
BIC		38.23	3882.50	3868.70	3872.51	3874.08	3868.70	3868.70	3877.62

## 5. Regression model fitting results

The zero-inflated models can be extended to accommodate covariates in a regression setting. The estimation of the proportion of zeros and the parameters can be divided into two parts. A logit part can be used to model the odds of structural zeros proportion and a Poisson part to model the counts that follow the Poisson distribution. Covariates could enter both the logit part and the Poisson part (Lambert, 1992). Estimation of parameters for zero-inflated models are done through the R and SAS software via the optimization of relevant log-likelihood functions using the NLMIXED procedure.

The following Table 5 shows the various count data models fitted to claim frequency data with covariates. To further analyze the automobile insurance data, the use of zero-inflated regression models is examined. The choice of variable is achieved through an exhaustive search of the database in which variables providing a significant improvement in the Poisson's log-likelihood function at convergence are chosen. Among these nine explanatory variables, the NCD, VAgeCat, gender of policyholders is shown to be significant in the Poisson regression model. For comparison, the same set of variables is used in other models. Results from fitting the Poisson, GP, NB, ZIP, ZIGP and ZINB regression models are given in Table 5. We consider the Poisson, GP, NB, ZIP, ZIGP and ZINB models and the performance of these models is evaluated via the log-likelihood, AIC, BIC.

**Table 5:** Results of fitting Poisson, GP, NB and zero-inflated regression models

Parameter/ coefficient of	Parameter estimate and its standard error when fitted model is						
	Poisson	GP-1	GP-2	NB	ZIP	ZIGP-1	ZIGP-2
(Intercept)	-2.2991 (0.1615)*	-2.3602 (0.1698)*	-2.3013 (0.1678)*	-2.3012 (0.1681)*	-1.7655 (0.2316)*	-2.3585 (0.1920)*	-2.3013 (0.1678)*
Female	-0.1493 (0.1550)	-0.1616 (0.1591)	-0.1524 (0.1598)	-0.1522 (0.1599)	-0.1480 (0.1593)	-0.1616 (0.1591)	-0.1525 (0.1598)
AutoAge	0.0654 (0.1573)	0.1242 (0.1652)	0.0689 (0.1638)	0.0687 (0.1640)	0.0573 (0.1630)	0.1240 (0.1653)	0.0689 (0.1638)
(NCD)10	-0.3726 (0.1261)*	-0.3648 (0.1295)*	-0.3774 (0.1305)*	-0.3772 (0.1306)*	-0.3768 (0.1300)*	-0.3650 (0.1295)*	-0.3772 (0.1305)*
(NCD)20	-0.5190 (0.1305)*	-0.5129 (0.1336)*	-0.5195 (0.1345)*	-0.5193 (0.1343)*	-0.5192 (0.1342)*	-0.5137 (0.1336)*	-0.5192 (0.1345)*

Parameter/ coefficient of	Parameter estimate and its standard error when fitted model is						
	Poisson	GP-1	GP-2	NB	ZIP	ZIGP-1	ZIGP-2
(NCD)30	-0.3742	-0.3855	-0.3724	-0.3720	-0.3688	-0.3879	-0.3720
	(0.1938)*	(0.2001)*	(0.2002)	(0.1999)*	(0.1996)*	(0.2003)*	(0.2002)*
(NCD)40	-0.7391	-0.7101	-0.7351	-0.7355	-0.7361	-0.7087	-0.7356
	(0.2427)*	(0.2436)*	(0.2484)*	(0.2480)*	(0.2480)*	(0.2434)*	(0.2485)*
(NCD)50	-0.6639	-0.6531	-0.6632	-0.6631	-0.6616	-0.6534	-0.6631
	(0.1370)*	(0.1394)*	(0.1411)*	(0.1409)*	(0.1406)*	(0.1394)*	(0.1411)*
(VAgeCat)1	0.2764	0.2758	0.2730	0.2732	0.2725	0.2754	0.2732
	(0.1475)*	(0.1521)*	(0.1531)	(0.1532)*	(0.1522)*	(0.1521)*	(0.1530)*
(VAgeCat)2	0.5037	0.5334	0.5041	0.5041	0.4963	0.5334	0.5040
	(0.1400)*	(0.1424)*	(0.1462)*	(0.1459)*	(0.1455)*	(0.1425)*	(0.1462)*
(VAgeCat)3	0.2143	0.2779	0.2186	0.2182	0.2079	0.2779	0.2181
	(0.2007)	(0.2087)	(0.2085)	(0.2083)	(0.2076)	(0.2088)	(0.2085)
(VAgeCat)4	-0.1781	-0.1183	-0.1726	-0.1727	-0.1826	-0.1192	-0.1727
	(0.2080)	(0.2160)	(0.2151)	(0.2148)	(0.2143)	(0.2162)	(0.2151)
(VAgeCat)5	-0.9925	-0.9241	-0.9896	-0.9897	-0.9996	-0.9258	-0.9894
	(0.2363)*	(0.2438)*	(0.2420)*	(0.2420)*	(0.2414)*	(0.2440)*	(0.2420)*
(VAgeCat)6	-1.3693	-1.2815	-1.3627	-1.3628	-1.3724	-1.2819	-1.3628
	(0.5241)*	(0.5267)*	(0.5296)*	(0.5288)*	(0.5289)*	(0.5268)*	(0.5296)*
$\alpha_0$	NA	NA	NA	NA	-0.3679	-6.8487	-11.5501
$\hat{\omega}$	NA	NA	NA	NA	0.4091	0.0011	9.64E-06
	-	-	-	-	(0.0936)	(0.0822)	(0.0040)
$\hat{\phi}$	NA	0.0301	0.3783	0.7639	NA	0.0301	0.3782
	-	(0.0115)*	(0.1444)*	(0.5060)*	-	(0.0119)*	(0.1444)*
Log-likelihood	-1883.39	-1878.40	-1878.15	-1878.20	-1878.85	-1878.40	-1878.15
AIC	3794.78	3786.80	3786.30	3786.39	3787.70	3788.80	3788.30
BIC	3891.67	3890.60	3890.10	3890.20	3891.50	3899.50	3899.10

Bracketed figures denote the standard error of the parameter estimates while \* denotes the significant estimates at the 5% level of significance. Based on the AIC values, the GP2, NB and the zero-inflated regression models fit to the automobile insurance data reasonably well.

## 6. Discussion and conclusions

Accurate modeling of the claim count distribution is one of the essential steps in calculating policy rates. Motivated by the dispersion and zero-inflation problem in the claim counts of the automobile insurance dataset, this study proposes the use of several count data models. The method accommodates the extra zeros possibly caused by the unreported minor losses. The Poisson, GP, NB, ZIP, ZIGP and ZINB models are considered and the performance of these models is evaluated via the log-likelihood, AIC, BIC. Based on the findings shown in the previous section, the GP-2, NB and the zero-inflated regression models fit the automobile insurance data reasonably well. The ZIGP distribution provides the best fit to the data. Other than the zero-inflated models, parametric methods such as the mixture of distributions can be used to model the claim frequency distribution with extra zeros. Hurlimann (1990) discussed the use of several pseudo compound Poisson distributions in modelling the claim count data. Dobbie and Welsh (2001) considered the use of the Neyman type-A distribution to model zero-inflated counts. Accordingly, the Neyman type-A distribution models the count data via two Poisson parameters and it becomes more flexible in modelling multimodal data.

Due to the possible over-dispersion in the Poisson part, the baseline ZIP model may not be adequate. As such, the Poisson part in ZIP has been modified by using NB distributions. Referring to this over-dispersion problem in the Poisson part when fitting the claim count data, the quasi-likelihood (QL) model defined by Wedderburn (1974) can be an alternative in modelling the extra-dispersion.

## Acknowledgements

The authors would like to thank the Editor of the journal and reviewers for their constructive comments which helped to improve the paper.

## References

1. Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L., and Kirchner, U. (1999). The Zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of Royal Statistical Society Series A*, 162(2):195-209.DOI:10.1111/1467-985X.00130
2. Consul, P.C. (1989). *Generalized Poisson distribution: Properties and Applications*. New York, Marcel Dekker.
3. Consul, P.C. and Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 21(1):89-109.DOI:10.1080/03610929208830766
4. Consul, P.C. and Jain, G.C. (1973). A Generalization of the Poisson distributions. *Technometrics*, 15:791-799.DOI: 10.1080/00401706.1973.10489112
5. Dobbie, M.J. and Welsh A.H. (2001). Modeling correlated zero-inflated count data. *Australian & New Zealand Journal of Statistics*, 43(4):431-444.DOI: 10.1111/1467-842X.00191
6. Famoye, F. (1993). Restricted generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 22:1335-1354. DOI: 10.1080/03610929308831089
7. Feller, W. (1943). On a general class of contagious distributions. *Annals of Mathematical Statistics*, 14(4):389-400.DOI:10.1214/aoms/1177731359
8. Frees and Valdez (2008). Hierarchical Insurance Claims Modeling. *Journal of the American Statistical Association*, 103(484):1457-1469.DOI: 10.1198/016214508000000823
9. Hurlimann, W. (1990). On maximum likelihood estimation for count data models. *Insurance: Mathematics and Economics*, 9:39-49. DOI: 10.1016/0167-6687(90)90014-5
10. Lambert, D. (1992). Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing. *Technometrics*,34(1):1-14. DOI: 10.2307/1269547
11. Mouatassim, Y. and Ezzahid, E.H. (2012). Poisson regression and Zero-inflated Poisson regression: application to private health insurance data. *European Actuarial Journal*, 2:187-204.DOI 10.1007/s13385-012-0056-2
12. Neelon, B.H., Malley, A.J. and Normand, S.L. (2010). A Bayesian model for repeated measures zero-inflated count data with application to

- outpatient psychiatric service use. *Statistical Modelling*, 10(4):421–439. DOI:10.1177/1471082X0901000404
13. Neyman, J. (1939). On a new class of contagious distributions applicable in entomology and bacteriology. *Annals of Mathematical Statistics*, 10(1):35–57. DOI:10.1214/aoms/1177732245
  14. Pandey, K.N. (1964). Generalized inflated Poisson distribution. *Banaras Hindu University Journal of Scientific Research*, 15(2):157–162.
  15. Renshaw, A.E. (1994). Modeling the claims process in the presence of covariates. *Astin Bulletin*, 24:265–285. DOI: 10.2143/AST.24.2.2005070
  16. Thomas, H. and Samson, D. (1987). Linear models as aids in insurance decision making: the estimation of automobile insurance claims. *Journal of Business Research*, 15:247–256. DOI: 10.1016/0148-2963(87)90027-0
  17. Wedderburn, R.W.M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, 61(3):439–447. DOI: 10.1093/biomet/61.3.439
  18. Yang, Z., Hardin, J.W. and Addy, C.L. (2009). Testing overdispersion in the zero-inflated Poisson model. *Journal of Statistical Planning and Inference*, 139:3340–3353. DOI:10.1016/j.jspi.2009.03.016
  19. Yip, P. (1988). Inference about the mean of a Poisson distribution in the presence of a nuisance parameter. *Australian Journal of Statistics*, 30(3):299–306. DOI:10.1111/j.1467-842X.1988.tb00624.x
  20. Yip, K.C.H. and Yau, K.K.W. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance Mathematics and Economics*, 36:153–163. DOI:10.1016/j.insmatheco.2004.11.002
  21. Zamani, H. and Ismail, N. (2012). Functional form for the generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 41(20):3666–3675. DOI:10.1080/03610926.2011.564742
  22. Zamani, H. and Ismail, N. (2014). Functional form for the zero-inflated generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 43(3): 515–529. DOI:10.1080/03610926.2012.665553
  23. Yau, K., Wang, K., Lee A. (2003). Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. *Biometrical Journal*, 45 (4): 437–452. DOI: 10.1002/bimj.200390024