Sigma J Eng & Nat Sci 38 (3), 2020, 1543-1553

Sigma Journal of Engineering and Natural Sciences
Sigma Mühendislik ve Fen Bilimleri Dergisi

Research Article

ON COMPARISON OF MODELS FOR COUNT DATA WITH EXCESSIVE ZEROS IN NON-LIFE INSURANCE

Övgücan KARADAĞ ERDEMİR*[1], Özge KARADAĞ[2]

[1]*Department of Actuarial Sciences, Hacettepe University, ANKARA;* ORCID: 0000-0002-4725-3588
[2]*Department of Statistics, Hacettepe University, ANKARA;* ORCID: 0000-0002-2650-1458

Received: 23.07.2019 Revised: 03.06.2020 Accepted: 13.06.2020

ABSTRACT

Modeling of the claim frequency is crucial from many respects in the issues of non-life insurance such ratemaking, credibility theory, claim reserving, risk theory, risk classification and bonus-malus system. For analysing claims in non-life insurance the most used models are generalized linear models, depending on the distribution of claims. The distribution of the claim frequency is generally assumed Poisson, however insurance claim data contains zero counts which effects the statistical estimations. In the presence of excess zero, there are more appropriate distributions for the claim frequency such as zero-inflated and hurdle models instead of a standard Poisson distribution. In this study, using a real annual comprehensive insurance data, the zero-inflated claim frequency is modeled via several models with and without consideration of zero-inflation. The underlying models are compared using information criteria and Vuong test. Parameter estimations are carried out using the maximum likelihood.
Keywords: Claim frequency, count data, generalized linear model, hurdle model, zero-inflation.
2000 Mathematics Subject Classification: 62J12, 62J99, 62P05.

1. INTRODUCTION

In non-life insurance mathematics, aggregate loss is total amount paid on all claims by an insurance system and there are two components of aggregate loss. The claim frequency is the counts of claims in an insurance pool through the insurance duration, while the claim severity shows the monetary losses of insurance claims. Modeling claim frequency is essential from many respects in different analyses of non-life insurance such as ratemaking, credibility, claim reserving, classification, risk and bonus-malus system.

For instance, in the ratemaking of the property and casualty insurance, it is aimed to determine the net premium. The net premium can be calculated under two assumptions. When the claim frequency and claim severity assumed to be independent, the components are modeled separately and the selection of risk factors to be used in the ratemaking is based on significant variables according to the distribution of claim frequency [1], [2]. Net premium also can be calculated in the case of violation of independence. Then, claim frequency can be taken as an explanatory variable for the claim severity and total claim is estimated [3], [4]. In both dependency and independency cases, it is crucial to model the claim frequency correctly.

* Corresponding Author: e-mail: ovgucan@hacettepe.edu.tr, tel: (312) 297 61 60 / 119

In the credibility theory which combines the recent and the past experience for the pricing of insurance, the distribution of the claim frequency is important for the determination of the credibility model [5].

Moreover, in the risk theory, while using the collective risk model, it is assumed that the claim frequency is a random variable and has a distribution function. According to the collective risk model, the aggregate loss has a joint distribution as the primary distribution being the claim frequency and the secondary one being the claim severity. To determine the appropriate collective risk model, the distribution of the claim frequency should be determined accurately. On the other hand, for the calculations such as convolution and Panjer recursion, the distribution of claim frequency is also important [6].

In addition to these calculations, actuarial modeling of claim counts is substantial in non-life insurance with respect to risk classification and bonus-malus systems [7]. While risk classification divides a portfolio into homogeneous risk groups, bonus-malus systems fairly determine the insurance premium considering the claim history of insureds. Thus, both of these regulations aim the fair premium.

Afore-mentioned reasons, in almost all fields of non-life insurance, modeling of the claim frequency is crucial and the statistical distribution should be determined appropriately. The claim frequency is commonly assumed as to be distributed Poisson. Although, Poisson distribution is widely used, it has a strong assumption such as the equality of the mean and variance. However, this assumption, can be violated due to the heterogeneous nature of the real data and overdispersion is occurred.

Moreover, claim frequency data may contain high density of zero counts. For instance, in the automobile insurance which is one of the branches of non-life, the implementations such as bonus-malus system, the owners of the policy do not inform the company of any small consistent damages. The hunger of bonus which emerge when the insureds do not report all of the claims to save bonus on premiums may appear to be undamaged and may cause an increase in the number of zero claims [8]. Hence in such situations, insurance claim data may show aberration such as overdispersion and zero-inflation.

In the presence of overdispersion, negative binomial distribution can be an alternative to Poisson distribution. If the zero counts cause the overdispersion, it is better to use zero-inflated models [9]. Zero-inflated models can be considered as a mixture of a zero-point mass and any count data regression under the original generalized linear modeling framework [10].

Zero-inflated models are used in many studies to model data which has high zero density. Yip and Yau [11] studied on zero-inflated distributions for claim frequency and they used the generalized Pearson χ^2 statistic and information criteria. Tüzel and Sucu [12] investigated zero-inflated regression models using Turkish insurance data. Mouatassim and Ezzahid [13] compared Poisson and zero-inflated Poisson model for health insurance and they used Vuong test for model comparison. Ismail and Zamani [14] investigated negative binomial and generalized Poisson regression model to fit German health insurance count data. Covrig and Badea [15] compared different models for insurance claim counts and they investigated the effect of overdispersion. A new zero-inflated regression model for zero-inflated count data and a new regression model so called Poisson quasi-Lindley regression model for over-dispersed count data are proposed by Altun [16], [17].

There are also hurdle models as an alternative to zero-inflated models. Boucher et al. [8] used compound frequency models and they examined different risk classification models for count data by using Score and Haussmann tests. Yang et al. [18] proposed new link functions for hurdle Poisson and hurdle negative binomial to deal with zero-inflation, overdispersion and missing observations in clinical trials. Saffari et al. [19] proposed negative binomial hurdle model. Sarul and Şahin [20] compared Poisson models, zero-inflated models and hurdle models for claim frequency data. Gilenko and Mironova [21] used hurdle model to model claim frequency for a pricing study and they model claim frequency and severity separately to calculate the price of a policy using a real Russian motor own damage insurance data. Baetschmann and Winkelmann [22] introduced a new dynamic hurdle model for zero-inflated count data related to stochastic process. Sakthivel and Rajitha [23] compared methods with back propagation neural network for modeling the count data which has excessive number of zeros by using mean square error for model selection.

In this study, a more comprehensive comparative statistical analysis is performed by considering both overdispersion and zero-inflation for actuarial count data models. Models are evaluated by information criteria and Vuong test. This study is organized as follows: Firstly, models such as generalized linear models, zero-inflated models and hurdle models for count data are examined in Section 2. The properties of the count data models for actuarial sciences are briefly given in Section 3. An application study is carried out using a real annual comprehensive insurance data of a non-life insurance company with a high density number of zero claims in Section 4. In the application part, models are established, compared and the results of the models have been interpreted. Finally, concluding remarks are briefly given in Section 5.

2. METHOD

Linear models assume that the residual errors follow a normal distribution. However, especially in the field of applied sciences, the data are frequently distributed discrete and limited to non-negative values. These type of data are called as count data and usually analysed by generalized linear models (GLMs).

2.1. Generalized Linear Models

GLMs are an extended family of linear models for non-normal error distributed data from the exponential family such as Bernoulli, Poisson, Gamma, etc. [24]. The methodology of the generalization is to transforming the mean response to a linear predictor and relating the model parameters to the predictors by a link function [25].

GLMs have a long history and over time the models of that family have been specialized and extended depending on the properties of the error distribution. Poisson and negative binomial models are the most commonly used cases when the structure of the studied data is count.

Poisson Model

Let $y_i = 0, 1, 2, \dots,\ (i = 1, 2, \dots, n)$ be a scalar variable distributed Poisson with probability density function

$$P(Y_i = y_i|\boldsymbol{x_i}) = \frac{\mu_i^{y_i} exp(-\mu_i)}{y_i!} \tag{2.1}$$

where μ_i is the conditional mean of y_i on k covariates $\boldsymbol{x_i} = (x_{i1}, x_{i1}, \dots, x_{ik})'$, $E(y_i|\boldsymbol{x_i}) = \mu_i$. Then the Poisson model can be written as,

$$\mu_i = exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}) \tag{2.2}$$

via the canonical logarithmic link function $g(\mu_i) = log(\mu_i)$.

Negative Binomial Model

The probability of a negative binomial model distributed event count y_i is,

$$P(Y_i = y_i|\boldsymbol{x_i}) = \frac{\Gamma(y_i+\theta)}{\Gamma(\theta) y_i!} \frac{\mu_i^{y_i} \theta^{\theta}}{(\mu_i+\theta)^{y_i+\theta}} \tag{2.3}$$

where μ_i is again the conditional mean of y_i and θ is the shape parameter. Then the negative binomial model is written as,

$$\mu_i = exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}) \quad (2.4)$$

via again the canonical logarithmic link function $g(\mu_i) = log(\mu_i)$. Poisson model and negative binomial models are nested models since when $\frac{1}{\theta} = \alpha \to 0$ the negative binomial model reduces to Poisson.

In count data analysis, in addition to non-normally distributed errors, data are also composite with issues such as excessive zeros and overdispersion. Overdispersion describes the observation that variation is higher than would be expected. Negative binomial model is alternative to Poisson model when the sample variance is larger than the sample mean because of being longer and fatter tailed distribution compared to Poisson [26], [27]. For capturing the excess number of zeros, zero-inflated and hurdle models have been developed and generated for both distributions. The difference between two different approaches is the treatment to sources of zeros.

In applied sciences, the zeros split into two kinds, structural zeros and sampling zeros. As is evident from its name, sampling zeros are the random zeros from the sampling distribution. On the other hand, structural zeros are the only possible values. Depending on the distinction of zeros, zero-inflated models treat zeros as a combination of both types whereas hurdle models treat all zeros only as structural.

2.2. Zero Inflated Models

Zero inflated models allow for modeling the outcome as a mixture of Bernoulli distribution and a count data distribution. In zero-inflated models, Bernoulli distribution is defined for structural zeros and the other count data distribution governs for random sampling zeros. Thus, zero inflated models drive two different zero generating processes.

Following Lambert [10] zero-inflated Poisson (ZIP) model incorporating covariates can be written as a mixture of two components:

$$P(Y_i = 0|\boldsymbol{x_i}) = \pi_i + (1 - \pi_i)\exp(-\mu_i) \quad (2.5)$$

$$P(y_i = z_i|\boldsymbol{x_i}) = (1 - \pi_i)\frac{{\mu_i}^{z_i}\exp(-\mu_i)}{z_i!}, \quad z_i \geq 1 \quad (2.6)$$

where $y_i = 0$ with probability π_i. Hence the number events has a Poisson (μ_i) distribution with probability $(1 - \pi_i)$.

In the presence of overdispersion, the zero-inflated negative binomial model (ZINB) usually provides more accurate results. ZINB model is given by,

$$P(Y_i = 0|\boldsymbol{x_i}) = \pi_i + (1 - \pi_i)(1 + \alpha\mu_i)^{-\frac{1}{\alpha}} \quad (2.7)$$

$$P(y_i = z_i|\boldsymbol{x_i}) = (1 - \pi_i)\frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)y_i!}\frac{(\alpha\mu_i)^{y_i}}{(1 + \alpha\mu_i)^{y_i + \alpha}}, \quad z_i \geq 1 \quad (2.8)$$

2.3. Hurdle Models

Hurdle models are two part models. In the first part model, a binary model is used for zeros or non-zero values and in the following part of the model, zero-truncated distributions, is defined for the positive values.

Hurdle Poisson (HP) model uses truncated Poisson distribution. Let y_i be truncated Poisson (μ_i) distributed random variable then HP model is set as,

$$P(Y_i = 0|\boldsymbol{x_i}) = \pi_i \quad (2.9)$$

$$P(y_i = z_i|\boldsymbol{x_i}) = \frac{(1 - \pi_i)}{1 - exp(-\mu_i)}\frac{{\mu_i}^{z_i}\exp(-\mu_i)}{z_i!}, \quad z_i \geq 1 \quad (2.10)$$

In a similar way hurdle negative binomial (HNB) model uses truncated negative binomial distribution.

Let y_i be truncated Negative Binomial (μ_i) distributed random variable then HNB model can be given as,

$$P(Y_i = 0|\boldsymbol{x_i}) = \pi_i \tag{2.11}$$

$$P(y_i = z_i|\boldsymbol{x_i}) = (1-\pi_i)\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)y_i!}\frac{(1+\alpha\mu_i)^{-\left(\frac{1}{\alpha}+y_i\right)y_i+\alpha}(\alpha\mu_i)^{y_i}}{1-(1+\alpha\mu_i)^{-\frac{1}{\alpha}}}, \quad z_i \geq 1 \tag{2.12}$$

In ZIP, ZINB, HP and HNB models, the canonical link functions $log(\mu_i)$ and $logit(\pi_i) = log\frac{\pi_i}{(1-\pi_i)}$ are used for mean and Bernoulli probability of success, respectively.

3. COUNT DATA MODELS FOR ACTUARIAL SCIENCES

In most of the non-life insurance studies, the first step is to analyse the relationship between the number and risk factors. Thus, the dependent variable y_i is usually defined as the claim frequency and the vector of $\boldsymbol{x_i} = (x_{i1}, x_{i1}, \dots, x_{ik})'$ is consisted of k explanatory variables such as age, gender, properties of car, engine capacity, previous claim experiences and etc ($i = 1,2,\dots,n$).

As it is mentioned in previous sections, y_i is usually observed with an excessive number of zeros more than what standard distributions can yield and zero-inflated or hurdle versions of the standard models are the appropriate statistical techniques in modeling of claim frequency.

As in the case of standard GLMs, in zero-inflated and hurdle version of models, parameters are estimated by maximum likelihood (MLE). In hurdle models the specification and the maximization of the likelihoods of two components can be carried out separately [28].

The best fit the claim frequency is usually found out by a model selection analysis. The commonly used model selection methods are information criteria such as AIC and BIC. For a model with intercept, they are defined as follows:

$$AIC = -2Log - Likelihood + 2p \tag{3.1}$$

$$BIC = -2Log - Likelihood + pln(n) \tag{3.2}$$

where $p = k + 1$ and n represent the number of parameters and observations, respectively. Especially in actuarial sciences, depending on the structure of the claim frequency data, the mentioned models are usually nested for instance Poisson GLM and negative binomial GLM. Thus, a Voung test can be used as an alternative to AIC and BIC for the comparison of non-nested models. The Vuong test basically compares the predicted probabilities of two non-nested models [29]. Generally, zero-inflated models are compared with their non-zero-inflated versions such as ordinary Poisson GLM versus ZIP or ordinary negative binomial GLM versus ZINB.

4. CASE STUDY

A comprehensive insurance (motor own damage insurance) data set taken from an insurance company for year 2014 is used for application. After regulations, a portfolio comprised 15767 policies is obtained. The data contains the claim counts of policyholders and risk factors, some of them are related to policyholders and some are related to vehicles.

The age of policyholders is ranged between 21 and 90. While 11127 policyholders are male, the rest 4640 are female. Vehicles are categorized into four groups: car, pickup, rental vehicle and taxi, and age of vehicles are ranged 1 to 18. Type of usage is classified as private, commercial and leasing. The levels of No Claim Discount (NCD) are assigned as 0, 30, 40, 50 and 60. Only vehicles with two types of fuel (benzine and diesel), are taken into consideration by excepting

LPG powered vehicles. The status of new or renewal and the price of vehicle are considered as additional risk factors. The frequency distribution of the claim counts is given in Table 1.

Table 1. Number of Claim Frequencies

Frequency	Observed Claim Count
0	13955
1	1333
2	399
3	70
4	9
5	1

In this study, claim frequency ranged 0 to 5 is taken as the response variable. The distribution of claim counts is visualized by the histogram as given in Figure 1.

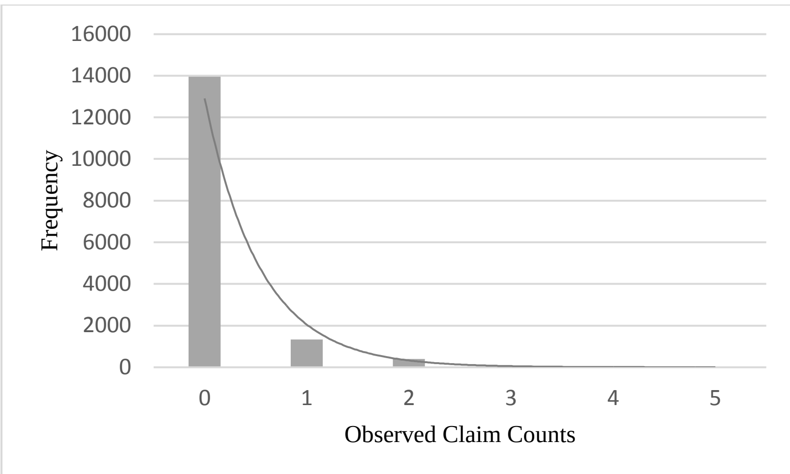

Figure 1. Histogram of Claim Frequency

According to Table 1 and Figure 1, 80.1 % of policyholders made no claims or do not report the claims during the year 2014. The high ratio of zeros is resulted as zero-inflation. Remainder 19.9 % of policyholders made a least one claim.

Zero-inflation is also analysed by testing the fit between observed and expected number of zeros. The zero test [30] gives a chi-square test statistic on one degree of freedom. The chi-square test statistic is calculated as 1136.258 and the presence of zero-inflation is reasonable (p-value<2.2e-16).

Moreover, dispersion test introduced by Cameron and Trivedi [31] is used to test overdispersion. The null hypotheses of no dispersion is tested against the overdispersion. According to the dispersion test, the overdispersed claim frequency is considered to be significant (z=15.435, p-value<2.2e-16).

For modeling the mentioned claim frequency, with and without consideration of zero-inflation, 6 different GLMs are analysed. The underlying models are Poisson GLM, negative binomial GLM, ZIP, ZINB, HP and HNB. Following a pre-modeling process, three explanatory variables are selected for the comparison of models. Age of insured, age of vehicle and NCD are detected as significant and summarized by Table 2 as follows.

Table 2. Summary of Explanatory Variables

Variables	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Age of insured	21.0	40.0	48.0	48.9	57.0	90.0
Age of vehicle	1.0	4.0	7.0	6.9	9.0	18.0
NCD	**0**	**30**	**40**	**50**	**60**	
	536	875	1034	2060	11262	

The parameter estimations for considered models are given by Table 3 where β_0 is the intercept, β_1, $\beta_2$and β_3 are the model parameters representing age of vehicle, NCD and age of insured, respectively.

Table 3. MLEs of the Parameters for Poisson GLM, Neg. Bin. GLM, ZIP, ZINB, HP, HNB

	Models					
Parameters	**Poisson GLM**	**Neg.Bin.GLM**	**ZIP**	**ZINB**	**HP**	**HNB**
β_0	-0.747262	-0.717794	-0.904352	-0,90447	-0.663402	-0.663402
s.e.	(0.103627)	(0.131239)	(0.307765)	(0.30777)	(0.129550)	(0.129550)
p-value	5.55e-13*	4.52e-08*	0.003299*	0.003295*	3.04e-07*	3.04e-07*
β_1	-0.013131	-0.014202	-0.025468	-0.02546	-0.00601	-0.006601
s.e.	(0.006436)	(0.007790)	(0.018121)	(0.01812)	(0.007858)	(0.007858)
p-value	0.04134*	0.06830	0.159877	0.160121	0.400907	0.400907
β_2	-0.186382	-0.188004	0.294989	0.29487	-0.218193	-0.218193
s.e.	(0.016605)	(0.021458)	(0.050891)	(0.05089)	(0.020946)	(0.020946)
p-value	<2e-16*	<2e-16*	6.77e-09*	6.85e-09*	<2e-16*	<2e-16*
β_3	-0.005203	-0.005518	0.016699	0.021671	-0.008053	-0.008053
s.e.	(0.001821)	(0.002213)	(0.004429	(0.00443)	(0.002239)	(0.002239)
p-value	0.006756*	0.003466*	0.000163*	0.000162*	0.000321*	0.000321*

* significant at $\alpha = 0.05$.

For model evaluation; the Log-Likelihood, AIC and BIC statistics are used. Furthermore, for non-nested comparisons Vuong test is carried out. The model selection criteria and the results of Voung test are given in Table 4 and Table 5.

Table 4. Comparison of Models

Model	Log-Likelihood(df)	AIC	BIC
Poisson GLM	-7250.409 (df=4)	14492.36	13712.41
Neg. Bin. GLM	-6899.806 (df=5)	13799.98	14523.03
ZIP	-6852.822 (df=8)	13708.82	13770.14
ZINB	-6852.823 (df=9)	13710.82	13779.81
HP	-6855.543 (df=8)	13710.41	13771.74
HNB	-6855.553 (df=9)	13712.41	13781.40

df represents the degree of freedom.

The smallest values of AIC and BIC belong to ZIP model. According to Log Likelihood and information criteria given in Table 4, zero-inflated and hurdle models fit better data than standard GLMs in the presence of zero-inflation.

In this paper, the ordinary GLMs and zero-inflated versions are categorized as non-nested models, hence Vuong test is carried out for the comparison of Poisson GLM versus ZIP and negative binomial GLM versus ZINB model.

Table 5. The Results of Vuong Test

	Vuong z-Statistic	**Model Comparison**	**p-value**
Raw	-12.24049	ZIP>Poisson GLM	<2.22e-16
AIC-corrected	-12.11678	ZIP>Poisson GLM	<2.22e-16
BIC-corrected	-11.64261	ZIP>Poisson GLM	<2.22e-16
Raw	-5.473100	ZINB>Neg.Bin. GLM	2.2112e-08
AIC-corrected	-5.022451	ZINB>Neg.Bin. GLM	2.5508e-07
BIC-corrected	-3.295186	ZINB>Neg.Bin. GLM	0.00049178

According to the negative test statistic zero-inflated versions of Poisson and negative binomial model better fit the data than standard ones according to the results of Vuong test.

The QQ plots of residuals shown by Figure 2 is used for assessing the goodness of fit and in models which consider zero-inflation, the data are more concentrated. In Figure 2, quantile residuals are plotted versus theoretical quantiles. In the presence of overdispersion, the deviance and the residuals are not usually normal. In this case, it might better to use randomized quantile residuals for GLMs. As it is seen in the left panel of Figure 2, the ordinary models have deviations especially in the upper half of the plots.

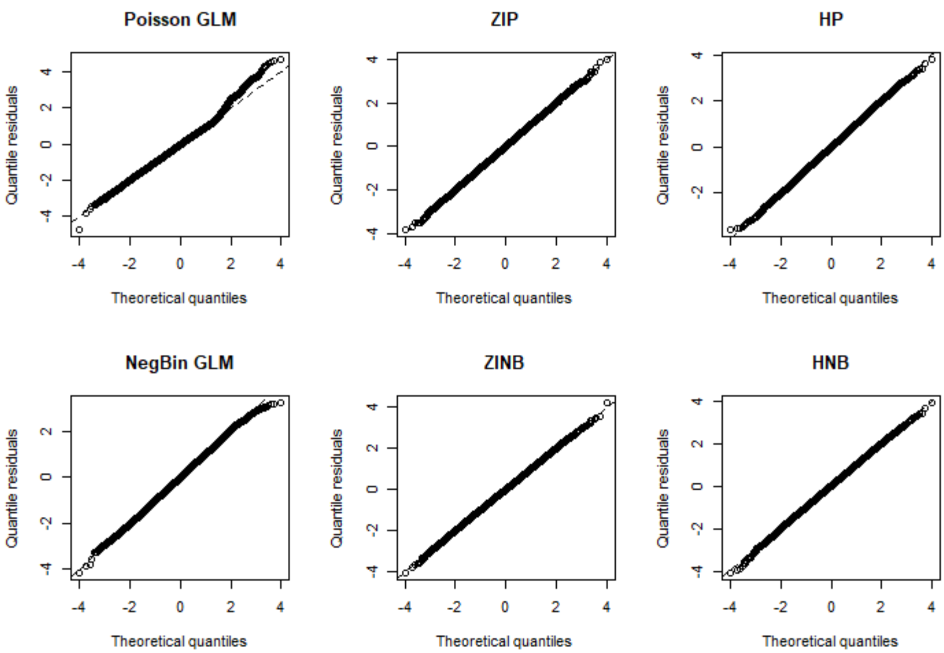

Figure 2. Residual Plots of Claim Frequency for Underlying Models

Figure 3 displays the rootograms for comparing the square roots of empirical frequencies with the fitted ones. Based on the rootograms; ZIP, ZINB, HP and HNB capture the underlying distribution better than standard GLMs. In the rootograms of Poisson GLM and negative binomial GLM, zero-inflation leads to deviations.

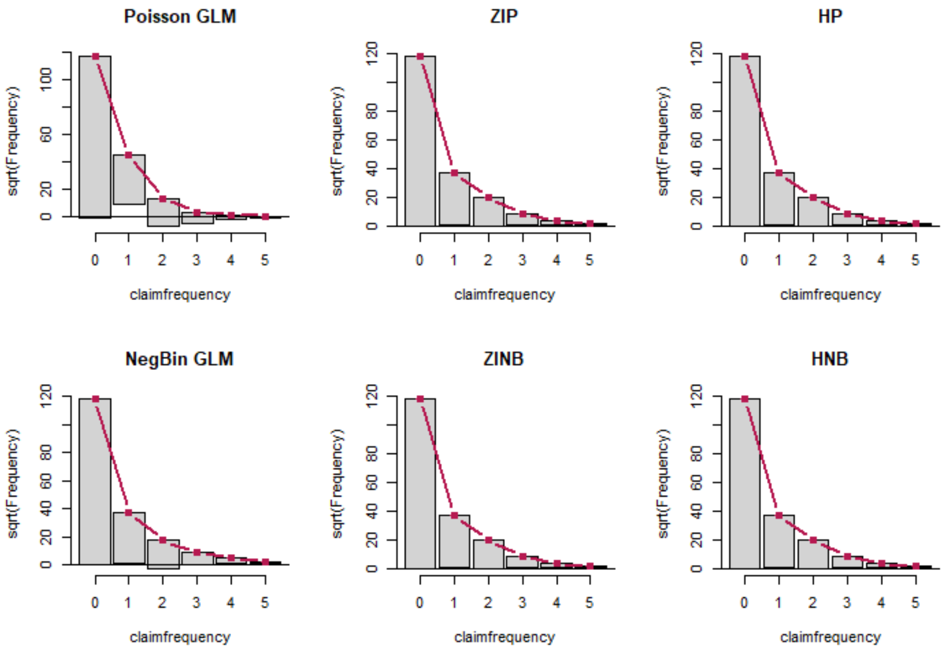

Figure 3. Rootograms of Claim Frequency for Underlying Models

5. CONCLUSION

In non-life insurance, it is essential to model claim frequency accurately in many respects. GLMs are mostly preferred in terms of convenience and applicability in actuarial science and generally Poisson GLM is used for modeling claim frequency under the assumption that the number of claim has a Poisson distribution. However, actuarial count data may contain too many zeroes. Zero inflated and hurdle models are flexible methods dealt with the zero-inflation. In order to find out the most suitable model for claim frequency, the models are established using a real data set.

Basically, both residual plots and rootograms point out the zero-inflated and hurdle models without an explicit distinction. To obtain the superior fitted model, the information criteria are consulted and it is noticed that ZIP with age of vehicle, NCD and age of insured is the best fit model to mentioned insurance count data. Although, ZIP has the smallest AIC and BIC, the values of hurdle model are quite similar. The reason is that, zeros in non-life insurance are generally sampling zeros. Therefore, HP may be an alternative to ZIP in the presence of zero-inflation. The results of model comparisons are also supported by Vuong test. Based on the test results, zero-inflated versions of models better accommodates excess zeros in the claim frequency compared to ordinary ones.

While the present models are set up by considering the overdispersion, models based on Poisson fit better than their corresponding models with negative binomial distribution. Based on comparison of ZIP and ZINB and comparison of HP and HNB; models with Poisson distribution are better fit than negative binomial-distribution. According to the observed mean and variance of claim frequency, there is no extreme overdispersion.

REFERENCES

[1] Charpentier, A. (Ed.). (2014) Computational actuarial science with R, *CRC press, Chapman and Hall.*

[2] David, M. (2015) Auto insurance premium calculation using generalized linear models, *Procedia Economics and Finance*, 20, 147-156.

[3] Gschlößl, S. and Czado, C. (2007) Spatial modelling of claim frequency and claim size in non-life insurance, *Scandinavian Actuarial Journal*, 2007(3), 202-225.

[4] Garrido, J., Genest, C. and Schulz, J. (2016) Generalized linear models for dependent frequency and severity of insurance claims, *Insurance: Mathematics and Economics*, 70, 205-215.

[5] Tse, Y. K. (2009) Nonlife actuarial models: theory, methods and evaluation, *Cambridge University Press.*

[6] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008) Modern actuarial risk theory: using R, (Vol. 128), *Springer Science & Business Media*.

[7] Denuit, M., Mar´ echal, X., Pitrebois, S. and Walhin, J. F. (2007) Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems, *John Wiley & Sons.*

[8] Boucher, J. P., Denuit, M. and Guill ´en, M. (2007) Risk classification for claim counts: a comparative analysis of various zeroinflated mixed Poisson and hurdle models, *North American Actuarial Journal*, 11(4), 110-131.

[9] Karazsia, B. T. and Van Dulmen, M. H. (2008) Regression models for count data: Illustrations using longitudinal predictors of childhood injury, *Journal of pediatric psychology*, 33(10), 1076-1084.

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

[11] Yip, K. C. and Yau, K. K. (2005) On modeling claim frequency data in general insurance with extra zeros, *Insurance: Mathematics and Economics*, 36(2), 153-163.

[12] Tüzel, S. and Sucu, M. (2012) Zero-inflated discrete models for claim frequencies, *Journal of Statisticians: Statistics and Actuarial Sciences*, 5(1), 23-31.

[13] Mouatassim, Y. and Ezzahid, E. H. (2012) Poisson regression and zero-inflated Poisson regression: application to private health insurance data, *European actuarial journal*, 2(2), 187-204.

[14] Ismail, N. and Zamani, H. (2013) Estimation of claim count data using negative binomial, generalized Poisson, zero-inflated negative binomial and zero-inflated generalized Poisson regression models, *In Casualty Actuarial Society E-Forum* (Vol. 41, No. 20, pp. 1-28).

[15] Covrig, M. and Badea, D. (2017) Some Generalized Linear Models for the Estimation of the Mean Frequency of Claims in Motor Insurance, *Economic Computation and Economic Cybernetics Studies and Research*, 51(4), 91-107.

[16] Altun, E. (2018) A new zero-inflated regression model with application, *Journal of Statisticians: Statistics and Actuarial Sciences*, 11 (2), 73-80.

[17] Altun, E. (2019) A new model for over-dispersed count data: Poisson quasi-Lindley regression model, Mathematical Sciences, 13 (3), 241-247.

[18] Yang, J., Li, X. and Liu, G. F. (2012) Analysis of zero-inflated count data from clinical trials with potential dropouts, *Statistics in Biopharmaceutical Research*, 4(3), 273-283.

[19] Saffari, E.S., Adnan, R., and Greene, W., (2012) Hurdle negative binomial regression model with right cencored count data, *Statistics and Operations Research Transactions (SORT)*, 36(2), pp.181-194.
[20] Sarul, L. S. and Sahin, S. (2015) An application of claim frequency data using zero inflated and hurdle models in general insurance, *Journal of Business Economics and Finance*, 4(4).
[21] Gilenko, E. V. and Mironova, E. A. (2017) Modern claim frequency and claim severity models: An application to the Russian motor own damage insurance market, *Cogent Economics and Finance*, 5(1), 1311097.
[22] Baetschmann, G. and Winkelmann, R. (2017) A dynamic hurdle model for zero-inflated count data, *Communications in Statistics-Theory and Methods*, 46(14), 7174-7187.
[23] Sakthivel, K. M. and Rajitha, C. S. (2017) A Comparative Study of Zero-inflated, Hurdle Models with Artificial Neural Network in Claim Count Modeling, *International Journal of Statistics and Systems*, 12(2), 265-276.
[24] McCullagh, P. and Nelder, J. A. (1989) Generalised linear modelling, *Chapman and Hall: New York*.
[25] De Jong, P. and Heller, G.Z., 2008. Generalized Linear Models for Insurance Data, *Cambrige: Cambridge University Press*.
[26] Mullahy, J. (1986). Specification and testing of some modified count data models, *Journal of econometrics*, 33(3), 341-365.
[27] Cameron, C. and Trivedi, P. (1998) Models for count data, Regression Analysis of Count Data, *Cambridge University Press*, New York.
[28] Zeileis, A., Kleiber, C. and Jackman, S. (2008) Regression models for count data in R, *Journal of statistical software*, 27(8), 1-25.
[29] Vuong, Q. H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica: Journal of the Econometric Society*, 307-333.
[30] Van den Broek, J. (1995) A Score Test for Zero Inflation in a Poisson Distribution. Biometrics, 51(2), 738-743.
[31] Cameron, A.C. and Trivedi, P.K. (1990) Regression-based Tests for Overdispersion in the Poisson Model. Journal of Econometrics, 46, 347–364.