Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 1787–1793
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.7064

Compare some estimation methods for zero-inflated Poisson regression models with simulation

Zahraa Abdulameer Ali AL-Mosawy, Abdul Hussian Habeab AL-Tai*

Department of Statistics, Collage of Administration and Economics, University of Karbala, Iraq

(Communicated by Javad Vahidi)

Abstract

Car accidents are an important phenomenon because of their direct relationship to the living conditions of different population centres in cities; it is known that a single accident causes increased human and material losses. For the purpose of researching the topic of predicting the number of accidents, the zero-inflated Poisson distribution was studied. In this thesis, several methods were searched, namely (maximum likelihood estimation, and moments) methods, in order to estimate the zero-inflation parameter of the Poisson regression. In this research, a number of simulation experiments were carried out according to the assumed distribution (Poisson's zero inflation) and the methods for estimating the assumed zero inflation parameter. And for a number of sample sizes (60, 80,100) according to different values of the zero inflation parameters (0.1, 0.2) and the second parameter of the zero-inflated Poisson distribution (1, 2). The comparison between the results of the different simulation experiments was done through the mean square error due to the estimations of each of the two parameters of zero inflation and the second parameter of the zero-inflated Poisson distribution according to each of (estimation method, distribution parameter and sample size).

Keywords: Poisson's zero inflation, maximum likelihood estimation method, moment's method, Poisson regression, mean square error, simulation experiments
2020 MSC: Primary 90C33; Secondary 26B25

1 General Introduction

Car accidents certainly need in-depth studies, as they represent one of the most important growing problems with the increase in road users and the multiplicity of approved means of transport, and the impact of these accidents extends and increases with the increase in these accidents.

Therefore, the process of estimating these accidents is one of the main goals, which contributes a lot to solving the problems associated with the accident, as well as reducing the rates of accidents in the future.

There are many types of research and studies that included research on the subject of the research, the most important of which are:- The research presented by (Lambert, Diane) in the year (1992), included the presentation of a Poisson regression model, assuming that the occurrence of the event represents (1) with a probability (P), while

*Corresponding author
Email address: zahraa.abdulameer@s.uokerbala.edu.iq (Zahraa Abdulameer Ali AL-Mosawy)

Received: April 2022 *Accepted:* August 2022

the non-occurrence with a value of (0) is with a probability of (q=1-p). The research included the logarithmic transformation of the assumed model with the adoption of the greatest possible method, which was adopted for a series of simulation experiments for samples of (25, 50,150) and the research also included a comparison of the Poisson regression model with a binomial regression model, as the (ZIP - Regression) model was presented, which represents A linear combination of the previous two models [3].

The research presented by (Cook, Richard J) and others in the year (1996) and within the regression models in which the evidence possesses Poisson processes. The research included the comparison between the parametric and semi-parametric perspectives of the assumed model. The research included presenting the theoretical aspects of the (A Specific Parametric Model) method, which was presented by The researcher accepted (Williams) in the year (1981), which includes models with random effects. The theoretical aspects of the Poisson regression model were also presented according to the hypotheses presented by (Cox) in the year (1972) [1].

In the year (2012), the researcher (Månsson, Kristofer) presented a paper in which the improved Liu Estimators were presented to estimate the parameters of the Poisson regression model. Linearity (Multicollinearity) included the presentation of several simulation experiments according to the assumed estimation methods and by adopting the mean squares of error (MSE) and the mean absolute error (MAE) as measures to compare the presented results [5].

In the year (2021), the researcher (Omer, Talha) and others presented research that included providing improved estimations for the zero-inflated Poisson according to the presence of the problem of multicollinearity and simulation experiments, as well as the practical application of maternal mortality data. The results were compared with Liu's estimations and the extent of convergence and divergence between them was noted according to each simulation experiment. The results of simulation estimators were relied on to treat the real data of maternal mortality [7].

As for the presented research, the zero-inflated parameter and the second parameter of the zero-inflated Poisson regression model were estimated through (estimation methods (maximum likelihood estimation, moments), sample sizes (60,80,100) and zero inflation parameter (0.1, 0.2) The second parameter (1,2). In this research, the theoretical aspects of the inflated Poisson regression models were presented. And different estimation methods, as well as comparing the estimations of the two parameters of the model through mean squares of error.

2 Problem of Research

Car accidents represent an increasing problem that is difficult to solve with the increase in road users and the presence of many reasons for their occurrence, and the increasing costs and increasing human losses as an inevitable result accompanying these accidents, which represented a burden on the proper management of the road.

3 Aim of research

The thesis aims to provide estimations for the parameters of inflation and the second parameter of the zero-inflated Poisson regression model (which is one of the Poisson regressions models suffering from inflation in recording the number of zero incidents). It also aims to test the best of these estimations using mean square error (MES).

4 poisson distribution

In the late 1830s, the famous French mathematician Simon Denis Poisson (1837) presented this distribution. It is an exponential distribution of the discrete data where the number of successes per unit time describes the occurrence of a particular event during a unit time. For a long time the Poisson distribution was used only to represent rare events. The probability mass function is [4]

$$p(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!} \tag{4.1}$$

$x=0,1,2,\ldots,$ with
(λ) represent distribution parameter
(x) represent the discrete random variabel
$P(X=x)$ represent the probability mass function whereas

$$\Sigma p(x)=1 \quad \text{and } p(x)\geq 0.$$

5 Zero-inflated Poisson model (ZIPM)

The zero-inflated model is a statistical model based on an inflated zero-probability distribution, i.e. the. Distribution High-frequency views are allowed with a zero value. One of the famous inflated zero-value models is the Diane Lambertis Poisson model with oversized which relates to the occurrence of a random variable containing inflated zero data per unit of time approved. The model is based on the fact that the number of accidents within a certain category will not include the inflated zeros of a category Among the observations that make the zero-inflated (ZIPM) model includes two operations to generate the zero (the process The first generates zeros and it gives a result of zero for any corresponding probability value, the second operation is governed by a distribution A Poisson that generates some of the operations generates some of the operations recorded as an observation equal to zero and a mixture function (which is the combination of zeros and non-zeroes) (ZIPM)zero, The probability of zeros is (π) And the probability of other than zeros is $(1-\pi)$, so the probabilistic mass of the probability of zeros is [2, 4].

$$\Pr(X=0)=\pi+(1-\pi)e^{-\lambda} \tag{5.1}$$

The probability function for non-zeroes is

$$\Pr\left(X=x_i\right)=(1-\pi)\frac{\lambda^{x_i}e^{-\lambda}}{x_i!} \tag{5.2}$$

$x_i=1,2,3,\ldots$. Since the random variable (x_i) has any positive integer greater than zero,
(λ) represents the expected parameter of the Poisson process dependent on observation (i).
(π) represents the inflated zero probability (which is the probability of an operation that gives the value of a variable opposite equal to zero)
The prediction of the number of accidents can be expressed as (μ), and when previous equation is applied then (μ) will be

$$\mu=\Sigma x(1-\pi)\frac{\lambda^{x}e^{-\lambda}}{x!} \tag{5.3}$$

$$\mu=(1-\pi)\Sigma\lambda\frac{\lambda^{x-1}e^{-\lambda}}{(x-1)!} \tag{5.4}$$

$$\mu=(1-\pi)\lambda \tag{5.5}$$

To find the variance, the following relationship can be applied

$$v(x)=\left(E\left(x^2\right)-(E(x))^2\right) \tag{5.6}$$

$$v(x)=(1-\pi)\lambda(1+\pi\lambda) \tag{5.7}$$

6 Zero-inflated Regression (ZIPR)

Inflated zero regression models are one of the most popular models for countable data, however the behavior of zero numbers in the observed data can create difficulties for these models. Among these difficulties is what is known (zero-truncation) is defined as(It is the smallest value that gives the transformation of the distribution from the normal regression to the zero inflation regression)[4].

6.1 Zero-inflated Poisson Regression models (ZIPRM)

Let the discrete random variable (X ∈ N) represent the number of accidents in a given experiment. Let (C) be an indicator by taking a value of $(0,1)$ for a latent category within the conditional distribution as follows [6]:-

$$Y/C=c\sim\begin{cases} p(y;\mu.\vartheta) & c=0\\ 0 & c=1\end{cases}\quad\ldots(9) \tag{6.1}$$

where ($p(y;\mu.\vartheta)$) represents a probability mass function with parameters ($\mu.\vartheta$)
And that there is an additional parameter(heterogeneity parameter) that may appear in the negative binomial regression model. will be (Y)) for the variable The marginal distribution), so the marginal distribution It is a dependent variable and depending on equation (17-2) then (Y) where [4]

$$f_Y(y;\mu.\vartheta) = \shortmid(C=1)\mathbb{P}\left(Y=\frac{y}{C}=1\right) \tag{6.2}$$

$$+\shortmid(C=0)\mathbb{P}\left(Y=\frac{y}{C}=0\right) \tag{6.3}$$

Let

$$\pi = \shortmid(C=0) \tag{6.4}$$

and

$$(1-\pi) = \shortmid(C=1) \tag{6.5}$$

Then

$$f_Y(y;\mu,\vartheta) = \pi I\{y=0\} + (1-\pi)p_Y(y;\mu,\vartheta) \tag{6.6}$$

(π) represent the (zero-inflation)

7 Estimation of(ZIPRM) parameters

There are a number of reliable methods for estimating model parameters such that:-

7.1 Maximum Likelihood Estimation Method(MLEM)

The Maximum Likelihood Estimation for parameters of the zero-inflated Poisson regression model for the observations that have Independent identically distributions (IID) such that [6]

$$\tilde{X} = (X_1 \cdot X_2 \quad X_n)$$

and each of them has a ZIP model with parameters ($\pi.\lambda$), therefore, the maximum likelihood function will be

$$L(\pi,\lambda/\tilde{X}) = \prod_{i=1}^{n} p\left(X = X_i\right) \tag{7.1}$$

Assuming that (Y) represents the number of observations (variables) that have the value (0) of the values of (X_i), then

$$L(\pi\cdot\lambda/\tilde{X}) = \left[\pi + (1-\pi)e^{-\lambda}\right]^Y \prod_{i=1}^{n}\prod_{X_{i\neq 0}}^{n}(1-\pi)e^{-\lambda}\frac{\lambda^{X_i}}{X_i!} \tag{7.2}$$

The logarithm of the maximum likelihood will be

$$Ln = Y\,\mathrm{Ln}\left(\pi + (1-\pi)e^{-\lambda}\right) + (n-Y)\,\mathrm{Ln}(1-\pi) - (n-Y)\lambda + n\bar{X}\,\mathrm{Ln}(\lambda) - Ln\left(\prod_{i=1}^{n} X_i!\right) \tag{7.3}$$

partial derivative with respect to (λ) will be

$$\frac{\partial L_n}{\partial\lambda} = \frac{-Y(1-\pi)e^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}} - (n-Y) + \frac{n\bar{X}}{\lambda} \quad \ldots(18) \tag{7.4}$$

Making the previous function equal to zero we get

$$\frac{n\bar{X}}{\lambda} = \frac{Y(1-\pi)e^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}} + n - Y \quad \ldots(19) \tag{7.5}$$

The partial derivative with respect to (π) will be

$$\frac{\partial L_n}{\partial\pi} = \frac{Y}{\pi+(1-\pi)e^{-\lambda}}\left(1-e^{-\lambda}\right) - \frac{n-Y}{1-\pi} \tag{7.6}$$

We note that the above equations are functions of each (π,λ) and we note that they are non-linear, so the (numerical iterative) methods are used to obtain the values of the estimators (λ_{mle},π_{mle}) and by substitution, as (Newton-Rafson) method.

7.2 Moment Estimation Method (MOM)

Within this method, the estimators are found by equalizing the moment of the sample to the corresponding parameters of the distribution to be estimated, This is done by taking expectations and to different degrees of the assumed distribution, which will be a function of its parameters Equating these moments with the moments of the sample in order to find their own estimations [6]

$$E(X)=(1-\pi)\lambda \tag{7.7}$$

$$V(X)=(1-\pi)\lambda(1+\pi\lambda) \tag{7.8}$$

There is a sample of size (n) which is $(x_1, x_2, \ldots\ldots, x_n)$ which each has Independent and second moment $\left(E(X), E(X)^2\right)$, the arithmetic mean and variance $\left(\bar{x}, s^2\right)$ can be obtained

$$\bar{X}=\frac{\sum_{i=1}^{n} X_i}{n} \tag{7.9}$$

$$s^2=\frac{\sum_{i=1}^{n}\left[X_i-\bar{X}\right]^2}{n-1} \tag{7.10}$$

By using $(E(X)=\bar{X})$ and $\left(V(X)=s^2\right)$ equations we get

$$\bar{X}=(1-\pi)\lambda \tag{7.11}$$

$$s^2=(1-\pi)\lambda(1+\pi\lambda) \tag{7.12}$$

By solving the previous equations we get

$$\lambda_{\text{mom}}=\frac{s^2-\bar{X}}{\bar{X}} * \frac{1}{\frac{s^2}{\bar{X}}+\bar{X}-1} \tag{7.13}$$

$$\pi_{mom}=\frac{s^2-\bar{X}}{\bar{X}} * \frac{1}{\frac{s^2+\bar{X}-1}{\bar{X}}} \tag{7.14}$$

8 Simulation Experiments

The simulation experiments that were carried out based on generating a Poisson distribution according to a specific sample size $(n=60, 80, 100)$n, a specific zero inflation parameter $(\lambda=0.1, 0.2)$ and a specific second parameter $(\beta=1, 2)$ and (MLE, MOM) estimation methods, each simulation experiment was repeated (1000) times and the estimation methods was compared by using (MSE) such that [2]

$$MSE=\frac{\sum_{i=1}^{1000}\left[\hat{\theta}_i-\theta\right]^2}{1000}. \tag{8.1}$$

So that (θ) equals (λ) once and (π) equals again From the previous table, it appears that the best method for each of the sixth sub-experiments within the simulation experiment. At (n = 60), $(\beta=1)$ the best method is the (MOM) method and it gave the mean squares error is the least and it reached (8.91E-11) and so on for other sub-experiments related to estimating the inflation parameter equal to (0.1) From the previous table, it appears that the best method for each of the sixth sub-experiments within the simulation experiment

At (n = 60), $(\beta=1)$ the best method is the (MOM) method and it gave the mean squares error is the least and it reached (2.45E-09) and so on for other sub-experiments related to estimating the inflation parameter equal to (0.2) From the tables (1) and (2) the figures (1) and (2) , we note that for the twelve experiments, the best method is(MOM) with because it was the best in (8) experiments.

Table 1: The (estimators and The mean square error) for Inflation parameter estimator ($\lambda = 0.1$)

β	n	The Estimators		The MSE		The Best Estimation	Index
		MLE	MOM	MLE	MOM		
1	60	0.10332	0.100009	1.10E-05	8.91E-11	8.91E-11	2
1	80	0.100153	0.100002	2.33E-08	3.52E-09	3.52E-09	2
1	100	0.100007	0.10102	4.43E-11	2.18E-08	4.43E-11	1
2	60	0.116497	0.10001	2.72E-04	1.58E-09	1.58E-09	2
2	80	9.98E-02	0.100002	4.93E-08	3.70E-07	4.93E-08	1
2	100	9.90E-02	0.100013	9.79E-07	1.01E-13	1.01E-13	2

Table 2: The (estimators and The mean square error) for Inflation parameter estimator ($\lambda = 0.2$)

β	n	The Estimators		The MSE		The Best Estimation	Index
		MLE	MOM	MLE	MOM		
1	60	2.67E-01	2.00E-01	4.45E-03	2.45E-09	2.45E-09	2
1	80	2.00E-01	2.00E-01	6.84E-09	1.66E-11	1.66E-11	2
1	100	2.91E-01	2.00E-01	8.35E-03	3.80E-07	3.80E-07	2
2	60	1.98E-01	2.00E-01	3.67E-06	9.03E-03	3.67E-06	1
2	80	1.99E-01	2.00E-01	4.46E-07	1.97E-05	4.46E-07	1
2	100	1.97E-01	2.00E-01	1.04E-05	2.67E-08	2.67E-08	2

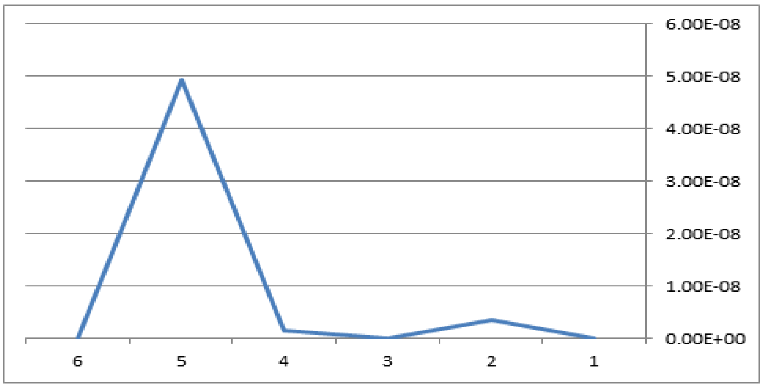

Figure 1: the minimum mean square error for each simulation experiment with inflation parameter equal to (0.1)

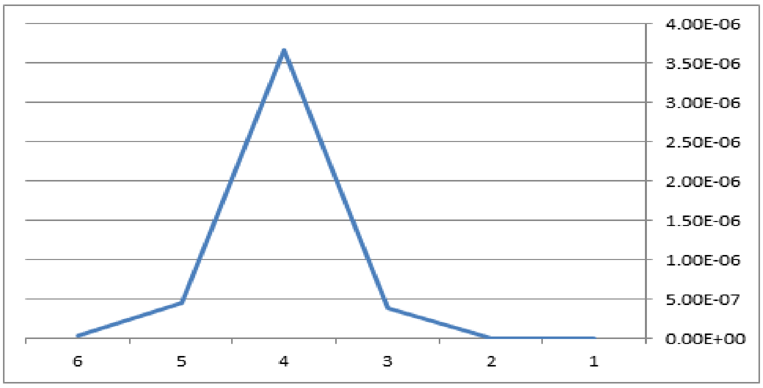

Figure 2: the minimum mean square error for each simulation experiment with inflation parameter equal to (0.2)

9 Conclusions and recommendations

After the results came out, a number of conclusions and recommendations emerged

1- The estimation method is affected by (sample size, the inflation parameter value and the distribution parameter value)

2- The (MOM) estimation method provided minimum mean squares of error for a greater number of simulation experiments

3- In general, an increase in the inflation parameter leads to an increase in the mean squares of error

4- It is possible to apply the estimation methods (MLE,MOM) in estimating the inflation parameter for each of (the binomial distribution and the negative binomial distribution)

5- It is possible to apply the estimation methods (shrinkage ,percentage) in estimating the inflation parameter for (Poisson regression)

References

[1] R.J. Cook, J.F. Lawless and C. Nadeau. *Robust tests for treatment comparisons based on recurrent event responses*, Biometrics **52** (1996), no. 2, 557–571.

[2] B.G. Kibria, K. Månsson and G. Shukur, *Some ridge regression estimators for the zero-inflated Poisson model*, J. Appl. Statist. **40** (2013), no. 4, 721–735.

[3] D. Lambert, *Zero-inflated Poisson regression, with an application to defects in manufacturing*, Technometrics **34** (1992), no. 1, 1–14.

[4] T. Loeys, B. Moerkerke, O. De Smet and A. Buysse, *The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression*, Br. J. Math. Statist. Psych. **65** (2012), no. 1, 163–180.

[5] K. Månsson, *On ridge estimators for the negative binomial regression model*, Econ. Modell. **29** (2012), no. 2, 178–184.

[6] G. Nanjundan and T. Raveendra Naika, *Asymptotic comparison of method of moments estimators and maximum likelihood estimators of parameters in zero-inflated Poisson model*, Appl. Math. **3** (2012), no. 6, Article ID: 20356, 7 pages.

[7] T. Omer, P. Sjölander, K. Månsson and B.G. Kibria, *Improved estimators for the zero-inflated Poisson regression model in the presence of multicollinearity: simulation and application of maternal death data*, Commun. Statist.: Case Studies, Data Anal. Appl. **7** (2021), no. 3, 394–412.