Journal of Economics and Administrative Sciences
2024; 30(142), pp. 492-508

P-ISSN 2518-5764
E-ISSN 2227-703X

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq
DOI: https://doi.org/10.33095/7bt1f714

Comparison of Estimation Methods for Zero Truncated Poisson Regression Model

Mohammed Sabri Al-zubaidi*
Department of Statistics
College of Administration and Economics
University of Baghdad,
Baghdad, Iraq
mohammed.sabri2101m@coadec.uobaghdad.edu.iq
***Corresponding author**

Ebtisam K. Abdulah
Department of Statistics
College of Administration and Economics
University of Baghdad,
Baghdad, Iraq
ekabdullah@coadec.uobaghdad.edu.iq

Received: 14/11/2023 **Accepted: 17/1/2024** **Published Online First: 30 /8/ 2024**

Abstract:

Count data represents a number of defined events that occur within a specific time frame for explanatory variables in the form of integers, and discrete distributions are among the probability distributions that use count data. The most famous of these distributions is the Poisson distribution. However, sometimes a change in the pattern of the random variable period may occur, such as the absence of the zero value, which requires finding a distribution that fits such a change, and that is the Zero Truncated Poisson Distribution (ZTPD).

This research aims to find a suitable model for the effect of non-zero value of data on any phenomenon and to use it to build a Zero Truncated Poisson Regression Model. This is done by selecting the best method out of three methods: Gauss-Newton (GN), Iteratively re-weighted least squares (IRWLS), and the Newton-Raphson algorithm method embedded in Maximum Likelihood (N-RAMEML), using the Mean Square Error (MSE) criterion, by simulating the Monte Carlo method using the R language program. This is done by changing different factors such as sample size (30, 70, 100, and 150) and the number of explanatory variables, repeating each experiment 1000 times. The study showed that the IRWLS method outperforms the N-RAMEML and GN methods.

Paper type: Research paper
Keywords: Zero-Truncated Poisson Distribution , The Iterative Reweighted Least Squares method , the Gauss -Newton method and the Newton –Raphson Algorithm Method Embedded in Maximum Likelihood , Mean square error .

1. Introduction :

The Poisson distribution is considered one of the important discrete probability distributions for addressing rare phenomena (AL-Shareefi and Al Baldawi, 2023) and uses the count data (Shanker and Shukla, 2019). That can be described descriptively as $x_i \sim P\ (\lambda_i)\ \ ,\forall\ \ i = 0,1,2,\dots$ where x represents i the observation of the phenomenon and λ represents and the distribution parameter the probability mass function (Manjula and Uma, 2020):

$$P_r(x|\lambda) = \frac{e^{-\lambda}\ \lambda^x}{x!} \quad , \quad x = 0,1,2,\dots\dots \quad ,\lambda > 0 \qquad (1)$$
$$= 0 \quad , \quad \text{otherwise}$$

And $E(x) = V(x) = \lambda$

David and Johnson (1952) considered to be the first to introduce the zero - truncated Poisson distribution (ZTPD), which is distributed as $x_i \sim ZTP\ (\lambda_i)\ \ ,\forall\ \ i = 1,2,\dots$ and is considered as one of the models of logarithmic linear regression (Al-doori, 2018).

1.1 Literature Review:

There are many studies on the characteristics of the zero - truncated Poisson distribution, including:

Abodyand Nuimai (2016) presented a combination of two distributions: Fréchet distribution with Poisson Lindely and Rayleigh distribution with Poisson Lindely. They proved that the Fréchet distribution with Poisson Lindely is the best by using comparative criteria. They also compared the Maximum Likelihood Estimate (MLE) method and the Least Square Method (OLS) in estimating the parameters of the Fréchet distribution, and found that the MLE method was the best. Real data representing earthquakes that occurred in the Badra region in Wasit province from 1994 to 2014 were used in the comparison.

Mohammed and Hamodi (2017) explained the most widely used regression model is Com-Poisson regression. When data display over-dispersion, the common solution is to Geometric regression. In order to model the effect of variables the number of Pneumonia patients, Geometric, Hurdle - Geometric and Zero inflated -Geometric regression models are fitted respectively. The results of Log likelihood and AIC indicated the Zero inflated –Geometric distribution is the best fit for this model.

Mohammed and Hussain (2017) presented a comparison between Bayesian method and full maximum likelihood to estimate hierarchical Poisson regression model. By simulation using different sample sizes (n = 30, 60, 120) and different frequencies (r = 1000, 5000) for the experiments as was the adoption of the Mean Square Error to compare the preference estimation methods and concluded that hierarchical Poisson regression model that has been appreciated full maximum likelihood with sample size (n = 30) is the best to represent the maternal mortality data that obtained from the Ministry of Health.

Mohammed and Hussain (2017) presented a study of the health institutions reality in Baghdad to identify the main reasons that affect the increase in maternal mortality by using two regression models, Poisson regression model and Hierarchical Poisson regression model. By a comparison between the estimation methods of the used models. The maximum likelihood method was used to estimate the Poisson regression model whereas the full maximum likelihood method were used for the Hierarchical Poisson regression model. The comparison was made through the use of simulation technique, A conclusion was reached, that the Hierarchical Poisson regression model - which was estimated by full maximum likelihood" method is the most excellent model for representing maternal mortalities data.

Mohammed and Aduri (2017) illustrated a formulation for the distribution of Poisson with Gamma distribution resulting in a Negative Binomial distribution, which is a discrete distribution. Its parameters were estimated using four methods: The maximum likelihood method, The moment method, The Downhill Simplex algorithm, and The EM algorithm. The second formulation is for the distribution of Poisson with Weibull distribution resulting in a compound Poisson-Weibull distribution, which is a continuous distribution. Its parameters were estimated using the maximum likelihood method based on the failure rate function and the percentage method.

Abbas and Ahmed (2020) introduced to use the method of moment to estimate the reliability function for truncated skew normal distribution, This distribution represents a parameterized distribution that is characterized by flexibility in dealing with data that is distributed normally and other a new distribution is derived from the original skew distribution that achieves the characteristics of the skew normal distribution function.

Kadhum and Abdullah (2021) presented the Poisson distribution with two parameters (the distribution parameter and the dispersion parameter). The features and characteristics of this distribution were displayed, along with estimating the distribution parameters using five methods: the method of moments, the maximum likelihood method, the method of weighted differences, the method of minimum repetitions, and the method of reduction. A simulation process and analysis of real data, represented by the number of suicide cases in Baghdad province, were conducted.

Handique and Chakraborty et al (2021) introduced a new truncated Poisson distribution family (with order G) by adding two additional parameters. The probability mass function and some statistical properties were derived. Additionally, random descriptions of the proposed family were studied based on moments, hazard function, and inverse hazard function. The family parameters were estimated using the method of maximum likelihood, and simulation was conducted to assess the bias and mean squared error of the maximum likelihood estimators.

Irshad and Chesneau et al (2022) introduced a new Lagrangian discrete distribution, named the Lagrangian zero truncated Poisson distribution (LZTPD). It can be presented as a generalization of the zero truncated Poisson distribution (ZTPD) and an alternative to the intervened Poisson distribution (IPD).

Kim and Kim Dae et al (2022) presented a generalization of the results of the zero-truncated negative binomial distribution to the case of a truncated negative binomial distribution with parameter r, considering its potential applications to the COVID-19 virus pandemic, probabilistic methods for studying certain special numbers and boundary multiples, as well as finding two different expressions for the probability generating function of a finite sum of independent truncated Poisson variables with equal parameters.

Li and Sun et al (2023) introduced a new mean regression model for the ZTP distribution with a clear interpretation about the regression coefficients an embedded Newton–Raphson algorithm is developed to calculate the MLEs of regression coefficients. The construction of bootstrap confidence intervals is presented and three hypothesis tests (i.e., the likelihood ratio test, the Wald test and the score test) are considered. Furthermore, the ZTP mean regression model is generalized to the mean regression model for the k-truncated Poisson distribution. Simulation studies are conducted and two real data are analyzed to illustrate the proposed model and methods.

Niyomdecha and Srisuradetchai (2023) presented a study on the Complementary Gamma Zero-Truncated Poisson Distribution (CGZTPD) as a new continuous lifetime distribution with three parameters. Its properties were discussed along with evidence such as probability density function, cumulative distribution function, survival function, hazard function, moment function, and maximum likelihood estimation of its regression parameters. Wald confidence intervals for the parameters were also calculated.

When is no value or more than one observation, either at the beginning, middle, or end of the period, that is called the truncation or the cut-off (Shamsur and Mohd, 2005) that has a significant effect on changing the probability mass function of the Poisson distribution and its properties, from which another distribution branches out with different characteristics depending on the cut and location. This research aims to find a model that fits the effect of observed data values devoid of zero value for any phenomenon and employs it to build a zero-truncated Poisson regression model by choosing the best method from three specific methods using simulation.

2. Material and Methods:

2.1 Zero-Truncated Poisson Regression Model with Assumptions ZTPRM:

The zero-truncated Poisson regression model ZTPRM is one of the logarithmic linear models of the response variable (y) and is defined according to the following formula (Al-doori, 2018):

$$y = e^{X\beta+U} \qquad , \qquad U \sim P(\lambda)$$
$$Log\, y = X\beta + U \tag{2}$$

Whereas: y is the dependent variable vector of degree (n×1), x is the independent (explanatory) variable matrix of degree ((p+1) × n), β is the parameter vector of degree (1 × (p+1)), U is the random error vector of degree (n ×1), P: is the number of explanatory variables, n is the sample size, e is the natural logarithm base.

And this model is built on three assumptions, as follows (Irshad and Chesneau, 2022):

The conditional probability function of the dependent variable (y_i) with parameter (λ) for the truncated zero Poisson distribution is:

$$P_r(y|\lambda) = \frac{e^{-\lambda}\ \lambda^{y}}{y!\ (1-e^{-\lambda_i})} \qquad y = 1,2,\dots\dots \quad ,\lambda > 0 \tag{3}$$

We can express the distribution parameter of the dependent variable (y_i).

$$\lambda_i = e^{x_i^T \beta} \tag{4}$$

Where x_i^T represents row i of the matrix transformer x_i

The distribution of ordered pairs of observations (y_i, x_i) is an independent distribution. By substituting the distribution parameter in equation (4) into the zero-truncated Poisson distribution function in equation (2), we obtain the following conditional probability function.

$$f(y_i|x_i) = \frac{e^{-e^{(x_i^T\beta)}}\, e^{\,y_i x_i^T \beta}}{\left(1-e^{-e^{(x_i^T\beta)}}\right) y_i\,!} \qquad i = 1,2,\dots,\dots \tag{5}$$

2.2 Newton–Raphson algorithm method embedded in ML:

If the zero value is missing from observations of (y_i) then the probability mass function of the zero - truncated Poisson distribution $\{y_i\}_1^n \sim ZTP\,(\lambda_i)$ it takes the following form (Shamsur and Mohd, 2005):

$$P_r(y|\lambda) = \frac{e^{-\lambda}\ \lambda^{y}}{y!\ (1-e^{-\lambda_i})} \qquad ,y = 1,2,\dots\dots \ ,\lambda > 0 \tag{6}$$

$$E(x) = \frac{\lambda}{1 - e^{-\lambda}} \; or \; \frac{\lambda \, e^{\lambda}}{e^{\lambda} - 1} \simeq \mu_i \tag{7}$$

$$Var(x) = \frac{\lambda}{(1 - e^{-\lambda})} \left[1 - \frac{\lambda\left(e^{-\lambda}\right)}{(1 - e^{-\lambda})}\right] = \mu_i - \mu_i^2\left(e^{-\lambda}\right) \tag{8}$$

By taking the natural logarithm of the maximum Likelihood function and substituting the linear predictor into it, we will notice that there are estimators (β) as exponent in the exponential function of the first and second derivatives as follows (Irshad and Chesneau, 2022):

$$Log\ L(\beta) = -\sum_{i=1}^{n} Log\ \Gamma(y_i + 1) + \sum_{i=1}^{n}\left[y_i Log\left(e^{x_i^T\beta}\right) - log\left[\frac{\left(1 - e^{-e^{x_i^T\beta}}\right)}{e^{-e^{x_i^T\beta}}}\right]\right] \tag{9}$$

And taking the first derivative (Gradient) for equation (9) for (β) and equating it to zero, we get to:

$$\frac{\partial Log\ L(\beta)}{\partial(\beta)} = \sum_{i=1}^{n}\left[y_i - \frac{e^{x_i^T\beta}}{1 - e^{-e^{x_i^T\beta}}}\right]x_i = 0 \tag{10}$$

Taking the second derivative (Hessian) , we get the following:

$$H = \frac{\partial^2 Log\ L(\beta)}{\partial\beta\partial\beta^T} = -\left[\sum_{i=1}^{n}\left\{\frac{\left(e^{x_i^T\beta}\right)}{\left(1 - e^{-e^{x_i^T\beta}}\right)} + \frac{\left(e^{-e^{x_i^T\beta}}\right)\left(e^{x_i^T\beta}\right)\left(e^{x_i^T\beta}\right)}{\left(1 - e^{-e^{x_i^T\beta}}\right)^2}\right\}\right]x_i x_i^T$$

Because of the existence of these estimators as exponents, we resort to Newton-Raphson algorithm and its embedding in the maximum likelihood method to calculate the regression parameters of a distribution ZTP as shown in (11) below (Yong 2012):

$$\hat{\beta}_{t+1} = \hat{\beta}_t - \left[H(\hat{\beta}_t)\right]^{-1} S(\hat{\beta}_t) \tag{11}$$

2.3 Gauss–Newton algorithm Method:

It is one of the numerical methods as it represents an extension of Newton's method used to reduce the sum of the quadratic function without calculating the second derivatives (Yong, 2012):

$F = (f_1, f_2, \ldots\ldots, f_m)$

Let F be a vertical vector representing (m) of functions and every function depends on a vector of variables (x):

$\beta = (\beta_1, \beta_2, \ldots\ldots, \beta_n)$

And β is a vertical vector representing n parameters. The steps of this method can be summarized as follows (Lai and Kek, 2017):

2.3.1 Newton method :

$$x_{k+1} \cong x_k - H(x_k)^{-1}\,\partial f(x_k) \tag{12}$$

2.3.2 Gauss Approximation :

$$S(\beta) = \frac{1}{2}\sum_{i=1}^{m} f_i(\beta)^2$$
$$\partial S = J_F^T F(\beta)$$
$$H \cong J_F^T J_F$$

Whereas:

$$J_F = \begin{bmatrix} \frac{\partial f_1}{\partial \beta_1} & \frac{\partial f_1}{\partial \beta_2} & \cdots\cdots & \frac{\partial f_1}{\partial \beta_n} \\ \frac{\partial f_2}{\partial \beta_1} & \frac{\partial f_2}{\partial \beta_2} & & \frac{\partial f_2}{\partial \beta_n} \\ & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial \beta_1} & \frac{\partial f_m}{\partial \beta_2} & \cdots\cdots & \frac{\partial f_m}{\partial \beta_n} \end{bmatrix} = \begin{bmatrix} \partial^T f_1 \\ \partial^T f_2 \\ \vdots \\ \vdots \\ \partial^T f_m \end{bmatrix}$$

2.3.3 Gauss – Newton Algorithm :

Where we find that J_F has been replaced by a matrix $H(x_k)^{-1}$ (Hessian) equation (7) is to facilitate the solution by avoiding finding the second derivative.

$$\beta^{k+1} \cong \beta^k - (J_F^T J_F)^{-1} J_F^T F(\beta^k)$$

2.4 Iteratively re-weighted least squares (IRWLS) (Green 1984):

This method relies on reformulating the probability mass function of the zero-truncated Poisson distribution equation (6) by taking the log and exp to transform it into the general form of the exponential family equation (13), where $\alpha(\phi) = 1$, as follows.

$$P(y, \theta, \phi) = \exp\left\{y \, log \, \lambda - log \left[\frac{(1 - e^{-\lambda})}{e^{-\lambda}}\right] + [-log(y!)]\right\}$$
$$\therefore P(y, \theta, \phi) = \exp\{y\,\theta - b(\theta) + C(y, \phi)\} \qquad (13)$$

As y represents the response variable, the function in terms of observations and dispersion parameter can be written as $C(y, \phi) = -log(y!)$ and θ represents the location parameter or the law parameter $\lambda = e^{\theta} \iff \theta = log\,\lambda$. Similarly, ϕ represents the dispersion parameter with $\alpha(\phi) = 1$. The function in terms of observations and dispersion parameter can be written as $C(y, \phi) = -log(y!)$. Finally, $b(\theta) = log\,\left[(1 - e^{-\lambda})/e^{-\lambda}\right]$ represents the location parameter function.

The product of the first and second derivatives of the location parameter function $b'(\theta)$, $b''(\theta)$ should be equal to the expected value and the variance of the zero truncated Poisson distribution. To find the first derivative of the logarithmic likelihood function, it can be obtained using the chain rule as follows:

$$\frac{\partial L}{\partial \beta_j} = \left[\frac{\partial L}{\partial \lambda}\right]\left[\frac{\partial \lambda}{\partial \mu}\right]\left[\frac{\partial \mu}{\partial \eta}\right]\left[\frac{\partial \eta}{\partial \beta_j}\right] \qquad (14)$$

$$\frac{\partial L}{\partial \beta_j} = \frac{[y - \mu]}{[V(y)]}[b''(\theta)]\, x_{ij} = [y - \mu]\, x_{ij} \qquad (15)$$

Equation (15) represents the first derivative (gradient) for a single observation of β_j. However, in the case of n observations, it can be found according to equation (16), which is represented by the symbol U below:

$$U = \frac{\partial L}{\partial \beta} = X^T A(y - \mu) \qquad (16)$$

Whereas

$$A = W\left(\frac{\partial \eta}{\partial \mu}\right) \quad , \quad W = \frac{1}{V(y)}\left(\frac{\partial \mu}{\partial \eta}\right)^2$$

For the second derivative (Hessian), represented by the symbol H, it can be found from either the Fisher matrix or the Information matrix according to the following equation.

$$H = -E\left[\frac{\partial^2 L}{\partial \beta_j \partial \beta_k}\right] = \frac{1}{V(y)}\left(\frac{\partial \mu}{\partial \eta}\right)^2 x_{ij}\, x_{ik} = X^T W X \qquad (17)$$

We can find the values of the parameter β_r using the Newton-Raphson algorithm to obtain:

$$\beta_r = \beta_{r-1} + H^{-1}U \qquad (18)$$

By substituting equations (16) and (17) into equation (18), we obtain:

$$\beta_r = (X^T W X)^{-1} X^T W z. \qquad , where\ z$$

$$= \eta + \frac{\partial \eta}{\partial \mu}(y - \mu) \qquad (19)$$

Where z is called Working Variate or Adjusted Dependent Variate

3. Discussion of Results:

3.1 Simulation:

Simulation is an image that represents the real-life situation of any system, where it is used in modeling realistic problems and solving them using computer programs, such as ready-made software or programming languages like Matlab or R. This facilitates obtaining the optimal solution to identify the changes that occur in this solution, saving a lot of time and effort (Fishman and Gross, 1976), through the use of Monte Carlo method in simulation, for example.

3.2 Results of Simulations:

For the purpose of implementing estimation methods for the truncated Poisson regression model and determining the best method, the results representing the estimated values will be presented according to the Poisson distribution and the MSE values for each method, based on the assumed sample sizes and the normal Poisson distribution as follows:

Table 1 : Represents the Assumed Values for the Parameters

Case	Number of Repressors	β_0	β_1	β_2	β_3	β_4	β_5	β_6
I	p=1	2.349	- 0.005	-	-	-	-	-
II	p=2	2.194	- 0.004	0.0001	-	-	-	-
III	p=3	2.061	- 0.006	0.0001	0.0014	-	-	-
IV	p=4	2.026	- 0.006	0.0001	0.0015	0.0021	-	-
V	p=5	2.132	- 0.006	0.0001	0.0014	0.0022	- 0.026	-
VII	p=6	1.892	- 0.005	0.0001	0.0014	0.0018	0.0279	- 0.0015

The tables below illustrate the results of a simulation experiment obtained by executing a program in the language (R 4.3.0) for three methods (N-RAMEML, GN and IRWLS) and comparing them with the default parameter values for the zero-truncated Poisson distribution from table (1) with different sample sizes (30, 70, 100, 150). Each experiment was repeated 1000 times, and the results representing the estimated values and MSE values for each method according to the assumed sample sizes and for each case are presented.

Through table (2), It is evident that the IRWLS method has the lowest MSE compared to the GN and N-RAMEML methods, With varying sample sizes for a single parameter (β_1), indicating that it is the best method. as shown in figure (1).

Table 2 : Shows the estimations of the parameters and MSE(Y) for the three methods of one explanatory variable (Case I)

Sample Size	Methods	β_0	β_1	MSE(Y)
Assume Values (Real)		1	0.3	-
30	N-RAMEML	0.99238	0.27326	10.146
	GN	0.99983	0.27294	10.116
	IRWLS	1.06329	0.27476	**9.892**
70	N-RAMEML	1.00510	0.27676	6.714
	GN	1.01162	0.27745	7.41
	IRWLS	1.07223	0.27814	**5.896**
100	N-RAMEML	1.00181	0.28290	5.376
	GN	1.00279	0.28300	6.63
	IRWLS	1.06863	0.28248	**5.164**
150	N-RAMEML	1.01280	0.28479	4.308
	GN	1.01630	0.28508	4.284
	IRWLS	1.07703	0.28520	**3.948**

However, in the model case of two parameters (β_1، β_2) as in table (3), we notice that the IRWLS method has the least MSE compared to the GN method and the N-RAMEML method for all sample sizes, indicating that it is the best method. as shown in Figure (2).

P-ISSN 2518-5764
E-ISSN 2227-703X

Table 3 : Shows the estimations of the parameters and MSE(Y) for the three methods of two explanatory variables (Case II)

Sample Size	Methods	β_0	β_1	β_2	MSE(Y)
Assume Values (Real)		1	0.8	-0.5	-
30	N-RAMEML	0.94455	0.84110	-0.52020	**11.532**
	GN	0.94442	0.84354	-0.51991	11.898
	IRWLS	1.03857	0.81100	-0.53868	12.51
70	N-RAMEML	1.01105	0.77216	-0.51416	7.182
	GN	1.01124	0.77157	-0.51283	6.534
	IRWLS	1.08976	0.86087	-0.54324	**6.444**
100	N-RAMEML	0.98356	0.84217	-0.51623	5.556
	GN	0.98470	0.84294	-0.51718	**4.902**
	IRWLS	1.06540	0.82485	-0.50667	5.418
150	N-RAMEML	0.98818	0.80197	-0.49337	5.052
	GN	0.99018	0.80273	-0.49382	4.818
	IRWLS	1.06786	0.80181	-0.49461	**4.242**

However, in the model case of three parameters (β_1، β_2، β_3) as shown in table (4), we notice that the mean squared error MSE is very close for all three methods, indicating the consistency property is achieved. It should be noted that the MSE for the IRWLS method is the lowest compared to the N-RAMEML and GN methods for all sample sizes, as shown in Figure (3).

Table 4 : Shows the estimations of the parameters and MSE(Y) for the three methods of three explanatory variables (Case III)

Sample Size	Methods	β_0	β_1	β_2	β_3	MSE(Y)
Assume Values (Real)		2	-0.5	-0.5	1.5	-
30	N-RAMEML	1.99869	-0.50974	-0.48586	1.48519	16.426
	GN	1.99700	-0.50756	-0.48591	1.48692	15.764
	IRWLS	2.00184	-0.50706	-0.48324	1.48790	**14.831**
70	N-RAMEML	1.99388	-0.49729	-0.51688	1.51890	10.41
	GN	1.99661	-0.49772	-0.51654	1.51812	10.788
	IRWLS	1.99706	-0.49450	-0.51427	1.51163	**9.9488**
100	N-RAMEML	1.99190	-0.50193	-0.48885	1.49618	9.3224
	GN	1.99301	-0.50111	-0.48835	1.49556	8.8544
	IRWLS	1.99460	-0.49924	-0.48644	1.49967	**8.0804**
150	N-RAMEML	1.99721	-0.50828	-0.49160	1.49970	5.796
	GN	1.99551	-0.50665	-0.49146	1.50178	5.7548
	IRWLS	1.99992	-0.50579	-0.48914	1.49332	**5.464**

But in the case the model of four parameters (β_1، β_2، β_3، β_4) as in table (5), we notice that the IRWLS method has the least MSE compared to the GN and N-RAMEML methods for all sample sizes, indicating that it is the best method. as shown in Figure (4)

Table 5 : Shows the estimations of the parameters and MSE(Y) for the three methods of four explanatory variables (Case IV)

Sample Size	Methods	β_0	β_1	β_2	β_3	β_4	MSE(Y)
Assume Values (Real)		1.5	0.5	0.3	-1.5	0.8	-
30	N-RAMEML	1.48381	0.48385	0.33676	-1.52406	0.79071	19.768
	GN	1.48062	0.48496	0.33786	-1.52234	0.79420	18.592
	IRWLS	1.52092	0.45053	0.31897	-1.52445	0.79874	**18.053**
70	N-RAMEML	1.46482	0.50182	0.31679	-1.51424	0.81879	17.7149
	GN	1.46556	0.50196	0.31614	-1.51620	0.82189	16.97679
	IRWLS	1.49767	0.47499	0.29833	-1.52291	0.79345	**16.46596**
100	N-RAMEML	1.45918	0.52385	0.29287	-1.48681	0.83438	11.8237
	GN	1.46064	0.52491	0.29194	-1.48720	0.83413	10.8528
	IRWLS	1.49211	0.49690	0.27720	-1.50286	0.80998	**10.52338**
150	N-RAMEML	1.48380	0.52205	0.29056	-1.50259	0.81371	7.3073
	GN	1.48498	0.52285	0.28944	-1.50285	0.81453	7.2751
	IRWLS	1.51471	0.49504	0.27603	-1.50133	0.80205	**7.2317**

However, in the case the model of five parameters (β_1، β_2، β_3، β_4، β_5) as shown in table (6), We notice that the IRWLS method has the lowest MSE for a sample size of (30, 70) compared to the GN and N-RAMEML methods, While we find that the N-RAMEML method has the lowest MSE compared to the other two methods for a sample size of (100). At a sample size of (150), We notice that the GN method has the lowest MSE, as shown in Figure (5).

Table 6 : Values of estimations estimates and MSE(Y)
for the three methods for five explanatory variables (Case V)

Sample Size	**Methods**	**β_0**	**β_1**	**β_2**	**β_3**	**β_4**	**β_5**	**MSE(Y)**
Assume Values (Real)		2	0.7	-0.8	0.5	-1.8	0.2	-
30	N-RAMEML	1.98775	0.70459	-0.81873	0.51679	-1.85669	0.20644	28.3488
	GN	1.99077	0.70370	-0.81806	0.51592	-1.85690	0.20623	27.4578
	IRWLS	1.98926	0.64467	-0.74599	0.47110	-1.68252	0.19215	**25.7529**
70	N-RAMEML	1.99295	0.69084	-0.82937	0.53415	-1.81724	0.20192	15.5568
	GN	1.99427	0.69134	-0.82908	0.53396	-1.81657	0.20080	15.3105
	IRWLS	1.99998	0.69756	-0.79668	0.49377	-1.66559	0.20461	**14.7936**
100	N-RAMEML	1.96125	0.70349	-0.78403	0.54188	-1.78137	0.20993	12.6759
	GN	1.96177	0.70283	-0.78370	0.54165	-1.78067	0.20937	12.8214
	IRWLS	1.96667	0.69056	-0.79346	0.50196	-1.63669	0.20689	**12.7473**
150	N-RAMEML	2.00427	0.70922	-0.79887	0.50946	-1.80466	0.20969	11.3544
	GN	2.00342	0.70881	-0.79851	0.51108	-1.80326	0.20073	11.4078
	IRWLS	2.00802	0.69825	-0.79997	0.50310	-1.80962	0.20614	**11.3424**

However, In the case the model of six parameters(β_1، β_2، β_3، β_4، β_5، β_6) with data close to the real data as in table (7), We notice that the IRWLS method has the lowest MSE compared to the GN and N-RAMEML methods, and all methods have achieved consistency for all sample sizes. as shown in Figure (6).

 P-ISSN 2518-5764
E-ISSN 2227-703X

Table 7 : Shows the estimations of the parameters and MSE(Y) for the three methods of six explanatory variables (Case VII)

Sample Size	Methods	β_0	β_1	β_2	β_3	β_4	β_5	β_6	MSE(Y)
Assume Values(Real)		2	-0.005	0.0001	0.0015	0.0018	0.003	-0.0015	-
30	N-RAMEML	2.0158	-0.0062	0.0074	-0.0393	-0.0449	0.0054	0.0143	28.635
	GN	2.0201	-0.0068	0.0075	-0.0402	-0.0459	0.0046	0.0135	28.335
	IRWLS	2.0168	-0.0062	0.0074	-0.0389	-0.0446	0.0054	0.0142	**26.013**
70	N-RAMEML	1.9853	0.0159	-0.0233	0.0063	0.0129	0.0277	-0.0016	15.714
	GN	1.9860	0.0168	-0.0223	0.0064	0.0134	0.0274	-0.0019	15.765
	IRWLS	1.9862	0.0158	-0.0232	0.0063	0.0129	0.0276	-0.0017	**14.943**
100	N-RAMEML	2.0078	-0.0134	-0.0258	0.0090	0.0243	0.0072	-0.0053	**12.804**
	GN	2.0074	-0.0131	-0.0261	0.0094	0.0232	0.0079	-0.0036	12.951
	IRWLS	2.0085	-0.0133	-0.0257	0.0089	0.0242	0.0072	-0.0053	12.876
150	N-RAMEML	1.9721	-0.0084	0.0184	0.0145	0.0065	0.0451	-0.0062	**11.469**
	GN	1.9716	-0.0076	0.0184	0.0138	0.0066	0.0459	-0.0059	11.523
	IRWLS	1.9730	-0.0084	0.0183	0.0144	0.0065	0.0448	-0.0062	11.475

We notice in all six cases above that the MSE decreases as the sample size increases.

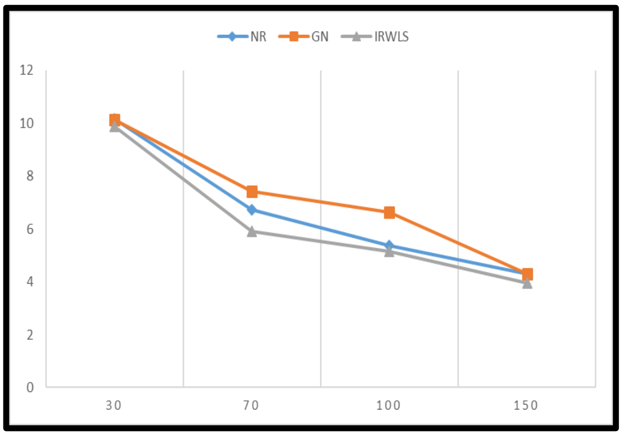

Figure 1: Elucidation of the MSE for case I at different sample sizes

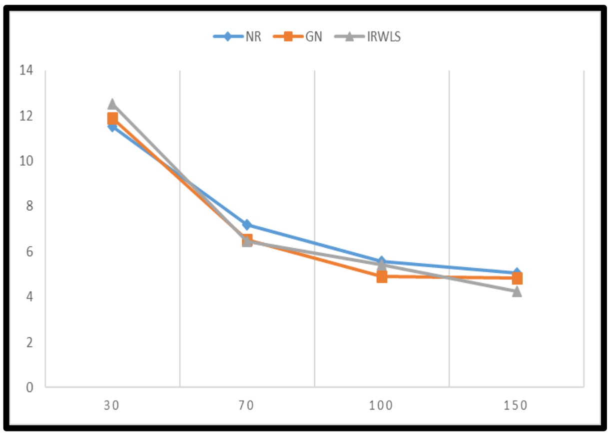

Figure 2: Elucidation of the MSE for case II at different sample sizes

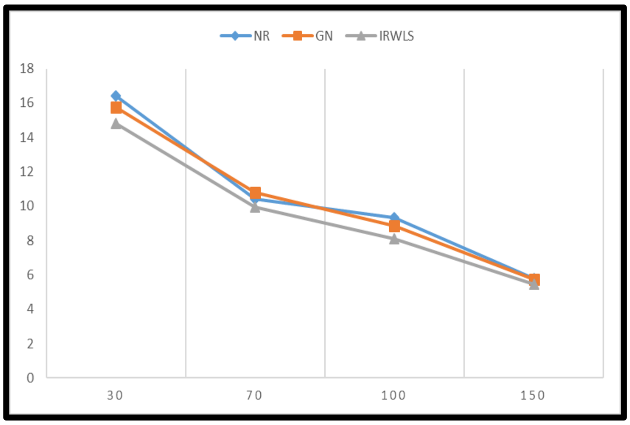

Figure 3: Elucidation of the MSE for case III at different sample sizes

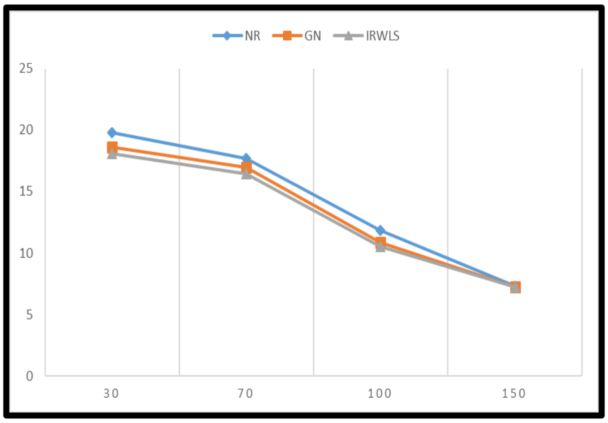

Figure 4: Elucidation of the MSE for case IV at different sample sizes

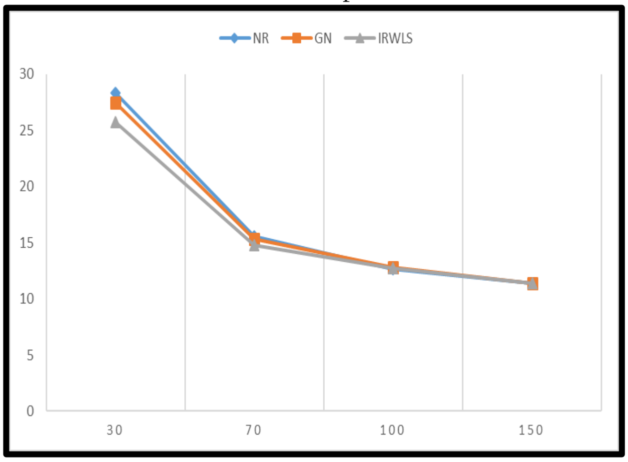

Figure 5: Elucidation of the MSE for case V at different sample sizes

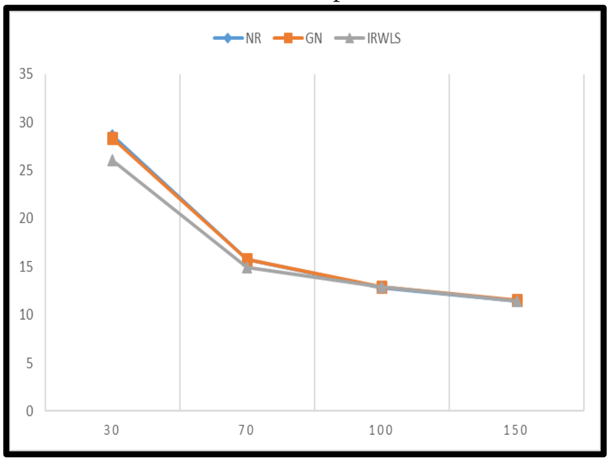

Figure 6: Elucidation of the MSE for case VII at different sample sizes

4. Conclusion:

One of the most important conclusions reached in this research is that simulation results indicate that the best method for estimating of the zero-truncated Poisson regression model is the Iteratively Reweighted Least Squares (IRWLS) method with repeated weightings, compared to the Gauss-Newton (GN) method and the Newton-Raphson embedded in the Maximum Likelihood (N-RAMEML) method because the IRWLS method has the lowest MSE value in most cases. The behavior of the N-RAMEML method approaches that of the GN method, and the behavior of all three methods approaches with increasing sample size. So the parameter estimate values become closer to each other as the explanatory variables increase .The MSE values decrease with increasing sample size, indicating that the three methods have the consistency property.

P-ISSN 2518-5764
E-ISSN 2227-703X

Authors Declaration:

Conflicts of Interest: None

-We Hereby Confirm That All The Figures and Tables In The Manuscript Are Mine and Ours. Besides, The Figures and Images, Which are Not Mine, Have Been Permitted Republication and Attached to The Manuscript.

- Ethical Clearance: The Research Was Approved By The Local Ethical Committee in The University.

References:

1- Abbas, H. K. and Ahmed, D. A. (2020). “Use The moment method to Estimate the Reliability Function of the Data of Truncated Skew Normal Distribution”, Journal of Economics and Administrative Sciences, Vol.26, No. 124, pp. 481-492.

2- Abody, E. H. and Nuimai, A. B. (2016). "Comparison Between Two Approaches (MLE & DLS) to Estimate Frechet Lindley Poisson Distribution Compound by Using Simulation", Ibn Al-Haitham Journal for Pure & Application Sciences, Vol. 29, No. 3, pp. 401-414 .

3- Al-doori, E. A. and Muhammed, N.A. (2017). “Comparison of some methods for estimating Poisson-Weibull distribution parameters”, Journal of Economics and Administrative Sciences, Vol. 23, No. 101, pp. 452-475.

4- Al-doori, E. A. (2018). “New Robust Estimation in Compound Exponential Weibull-Poisson Distribution for both contaminated and non-contaminated Data”, journal of Economics and Administrative Sciences, Vol. 24, No. 126, pp. 30-48 .

5- Al-Shareefi, S. N. and Al Baldawi, T. H.(2023).“A Computational Bayesian Approach to the Poisson Regression Models and the Proportional Hazard Models”,AIP Publishing Conf. Proc.2834, pp.080146,1-14.

6- Best, D. J. and Rayner, J.C. and Thas O. (2007). “Goodness of Fit for the Zero Truncated Poisson Distribution”. Journal of Statistical Computation and Simulation Vol. 77, No. 7, pp. 585-591.

7- David, F. N. and Johnson, N. L. (1952). "The Truncated Poisson" International Biometric Society, Vol. 8, No. 4, pp. 275-285.

8- Fishman, G.D. and Gross, D. (1976). “Concepts and Methods in Discrete Event Digital Simulation”. IEEE Transactions on Systems, Man, and Cybernetics Vol.: SMC-6 Issue: 5.

9- Green, P. J. (1984). "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives". Journal of the Royal Statistical Society: Series B (Methodological), Vol. 46, No. 2, pp. 149 -170.

10- Handique, L. and Chakraborty, S. and Eliwa, M.S. and Hamedani, G.G. (2021). “Poisson Transmuted-G Family of Distributions: Its Properties and Applications”, Pakistan Journal of statistics and Operation research Vol. 17, No. 1, pp. 309-332 .

11- Irshad, M. R., Chesneau, C., Shibu, D. S., Monisha, M. and Maya, R. (2022).“Lagrangin Zero–Truncated Poisson Distribution: Properties Regression Model and Application :”, Symmetry ,Vol.14,No.1775,pp.1-25.

12- Kadhum, M. K. and Abdullah, E. K. (2021). "Estimating the Parameters of the Poisson Distribution with Application", Tikrit Journal of Administration and Economics Sciences, Vol. 17, No. 56 Part (4), pp. 547-557.

13- Li, X. and Sun, Y. and Tian G. and Liang J. and Shi J. (2023). “Mean Regression Model for the Zero – Truncated Poisson Distribution and its Generalization”, Computational Statistics and Data Analysis, ELSEVIER, Vol. 179, pp. 107650.

14- Manjula, D. and Uma, G. and Nandhinidevi, R. (2020). “Derivation of Zero –One Truncate Poisson Distribution”. International Journal of Applied Research Vol. 5,No. 6, pp. 253-255.

15- Mohammed, A. H. and Hussain, J. E. (2017). “Analytical Study Compared Between Poisson and Poisson Hierarchical Model and Applied in Healthy Field”, Journal of Economics and Administrative Sciences Vol. 23, No.100 .

16- Mohammed, J. M. and Hamodi, H. A. (2017). “Comparison count regression models for the number of infected of pneumonia” Global Journal of Pure and Applied Mathematics Vol. 13, No. 9 , pp.5359-5366.
17- Mohammed, A. H. and Hussain, J. E. (2017). “A comparison between Bayesian method and full maximum likelihood to estimate Poisson regression model hierarchy and its application to the maternal deaths in Baghdad”, Journal of Economics and Administrative Sciences Vol. 23, No.101.
18- Niyomdecha, A. and Srisuradetchai, P. (2023). “Complementary Gamma Zero-Truncated Poisson Distribution and Its Application”, Mathematics, Vol.11, No. 11, pp. (1-13) 2584.
19- Shanker, R. and Shukla, K. K. (2019). “A Generalization of Zero-Truncated Poisson-Sujatha Distribution”, Hungarian Statistical Review Vol. 2, No. 2, pp. 32–50.
20- Shamsur, R. and Mohd, Y. (2005). “Truncated Distributions and their Applications”. Dissertation submitted for the award of the degree of master philosophy in statistics in Department of statistics and Operations Research Aligarh Muslim University (India) (https://core.ac.uk/download/pdf/144520251.pdf).
21- Kim, T. , Kim, D. S., Park, J., Lee, S., Park, S., ALqawba, M. S. and Jang, L. (2022). “Some Results on r-Truncated Degenerate Poisson Random Variables”, World Scientific Publishing Company Vol. 30, No. 10 , pp. 1-7 .
22- Lai, W. H., Kek, S. L. and Tay, K. G. (2017). “Solving Nonlinear Least Squares Problem Using Gauss-Newton Method”, International Journal of Innovative Science, Engineering & Technology, Vol. 4, No. 1, pp. 107650.
23- Yong, W. (2012). “Gauss–Newton method”. Department of Statistics, The University of Auckland, Auckland, New Zealand, Wiley Periodicals, Inc. Vol. 4 No. 4 , pp. 415-420.

Journal of Economics and Administrative Sciences
2024; 30(142), pp. 492-508
P-ISSN 2518-5764
E-ISSN 2227-703X

مقارنة طرائق تقدير أنموذج إنحدار بواسون المبتور صفرياً

محمد صبري محمود الزبيدي
كلية الادارة والاقتصاد/ جامعة بغداد/ قسم الاحصاء
بغداد، العراق
mohammed.sabri2101m@coadec.uobaghdad.edu.iq

ابتسام عبد الله كريم
كلية الادارة والاقتصاد/ جامعة بغداد/ قسم الاحصاء
بغداد، العراق
ekabdullah@coadec.uobaghdad.edu.iq

Received: 14/11/2023 Accepted: 17/1/2024 Published Online First: 30 /8/ 2024

مستخلص البحث:

تمثل البيانات القابلة للعد (Count Data) عدد من الاحداث المحددة التي تحدث في مدة زمنية معينة للمتغيرات التوضيحية على شكل اعداد صحيحة و تعتبر التوزيعات المتقطعة من بين التوزيعات الاحتمالية التي تستخدم البيانات القابلة للعد و من أشهر هذه التوزيعات هو توزيع بواسون. لكن في بعض الأحيان قد تطرأ على هذا التوزيع تغير في نمط فترة المتغير العشوائي (الذي يأخذ القيم من الصفر الى ما لا نهاية بالنسبة لتوزيع بواسون) كخلوه مثلا من القيمة الصفرية ذلك يستدعي إيجاد توزيع يتلائم مع هكذا تغير وهو توزيع بواسون المبتور صفرياً Zero Truncated Poisson (ZTPD) Distribution .

تهدف هذه الورقة الى إيجاد نمذجة تلائم تاثير قيم البيانات الخالية من القيمة الصفرية لاي ظاهرة و توظيفها لبناء أنموذج إنحدار بواسون المبتور صفرياً و ذلك من خلال إختيار أفضل طريقة من ثلاث طرائق هي طريقة كاوس– نيوتن Gauss – Newton (GN) و طريقة المربعات الصغرى معادة الوزن التكرارية Iteratively re-weighted least squares (IRWLS) و خوارزمية نيوتن رافسون المضمنة في طريقة الامكان الاعظم Newton –Raphson algorithm method embedded in Maximum Likelihood (N-RAMEML) بواسطة معيار متوسط مربعات الخطأ Mean Square Ererr (MSE) ، باستخدام إسلوب محاكاة مونتي كارلو عن طريق برنامج لغة (R) و ذلك بتغيير عوامل مختلفة مثل حجم العينة (30 ، 70 ، 100 و 150) و عدد المتغيرات التوضيحية بتكرار كل تجربة 1000 مرة ، حيث أظهرت الدراسة تفوق طريقة المربعات الصغرى معادة الوزن التكرارية IRWLS على طريقة خوارزمية نيوتن رافسون المضمنة في طريقة الامكان الاعظم N-RAMEML و طريقة كاوس– نيوتن GN .

نوع البحث: ورقة بحثية.

المصطلحات الرئيسة للبحث: توزيع بواسون المبتور صفرياً، طريقة خوارزمية نيوتن رافسون المضمنة في الامكان الاعظم، طريقة كاوس – نيوتن، طريقة المربعات الصغرى معادة الوزن التكرارية، متوسط مربع الاخطاء .