Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

Comparison of Poisson Regression and Conway Maxwell Poisson Models Using Simulation

Sarah Adnan Radam
Department of Statistics /College of Administration and Economics / University of Baghdad
Baghdad, Iraq
Sara.adnan1201a@coadec.uobaghdad.edu.iq

Lamyaa Mohammed Ali Hameed
Department of Statistics /College of Administration and Economics / University of Baghdad
Baghdad, Iraq
lamiaa.mohammed@coadec.uobaghdad.edu.iq

Received: 28/2/2023 **Accepted: 30/3/2023** **Published: June / 2023**

Abstract

Regression models are one of the most important models used in modern studies, especially research and health studies because of the important results they achieve. Two regression models were used: Poisson Regression Model and Conway-Max Well- Poisson), where this study aimed to make a comparison between the two models and choose the best one between them using the simulation method and at different sample sizes (n = 25,50,100) and with repetitions (r = 1000). The Matlab program was adopted.) to conduct a simulation experiment, where the results showed the superiority of the Poisson model through the mean square error criterion (MSE) and also through the Akaiki criterion (AIC) for the same distribution.

Paper type: Research paper.

Keywords: Poisson regression ,Conway-MaxWell Poisson, Maximum Likelinood Method , Mean square error, Akakie, Inverse Simulation.

1. Introduction

The primary goal of building probabilistic statistical models is to select the appropriate model that appropriately describes the set of data obtained from experiments, observational studies, field surveys, etc. Most of these model building techniques depend on finding the most appropriate probability distribution that explains the basic structure of the set. However, there is no single probability distribution suitable for the different data set. Therefore, a comparison was made between the Poisson regression model and the Conway-Maxwell Poisson model. The simulation experiment was carried out using the inverse method of the cumulative distribution function (c.d.f), and the Maximum Likelinood Method (MLE) was adopted to estimate Parameters of the two distributions with the use of the mean squared error (MSE) criterion and the Akaki criterion (AIC) for comparison between the two models.

1.1 Literature Review:

i. Long (1997) was the first researcher who addresses the study of the Poisson regression model by building the basics of the model as well as the process of estimating the parameters, where the specifications of the dependent variable were in the form of a descriptive variable.

ii. While the researcher Winkelmann (2008) presented a research on the analysis of counted data, where he dealt with the estimation of the parameters of the regular Poisson regression model according to the logarithmic formula and without it, and it was shown that the larger the sample size, the Poisson distribution turned into a normal distribution, and the researcher also touched on The most important assumptions of the Poisson regression model.

iii. Mansson (2011) introduced a Poisson estimator using the Ridge regression method as a means to address instability while taking the method of the Maximum lik lihood estimation (MLE) to compare the two methods to estimate the parameters of the Poisson regression model, where the average square error (MSE) was calculated As a criterion for comparison, and by using the Monte-Calo simulation, it has been shown from the simulation that the Poisson letter regression method is better than the greatest possibility method.

iv. Hameed and Jewad (2017) used two regression models Poisson and Hierarchical Poisson to study the indicator of maternal mortality through comparison between The Maximum Likelihood Method and Full Maximum Likelihood method by simulation and (MSE) they conclude that full maximum likelihood method is the best method.

v. Hameed and Jewad (2017) compared between Bayesian and full maximum likelihood methods to estimate hierarchical Poisson regression model through real data on maternal mortality they conclude that hierarchical Poisson regression model is the best for representing maternal mortalities data.

vi. Avci (2018) suggested the use of counting models to determine the factors that affect the number of people with schizophrenia, which is called (Schizophrenia). Conway Maxwell - Poisson and using two standards for comparison, first (Log-likelihood) and secondly (Akakie) and a sample of (205) observations taken from 2011 to (2014).

2. Materials and Methods

2.1 Poisson Regression Model

The Poisson regression model is one of the most important linear logarithmic regression models, and it is considered one of the important means for modeling the dependent variable when the values of the variable are countable by Long (1997). This model assumes that the dependent variable $(\boldsymbol{y_i})$ is a response variable that follows the Poisson distribution with a parameter of $(\boldsymbol{\vartheta})$ and also trace the random faults in the Poisson distribution model as a parameter of magnitude $(\boldsymbol{\vartheta})$ (Winkelmann, 2008) and (Cameron, 1998).

The naming of this model came according to the distribution of its random breakdowns and also as a logarithmic - linear from taking the natural logarithm of the model formula and thus transforming it into a linear formula. Of course, the Poisson regression is considered the basic model for numerical models.

Where it is highly relied upon, and in this context, the probability function will be defined according to the following formula: (2011:Mansson) (2008 : Winkelmann)

$$Y=e^{x\underline{\beta}+\vartheta} \quad \ldots (1)$$

where:
Y: the vector of the dependent variable of degree n×1.
X: Matrix of independent variables with degree n×(p+1).
$\boldsymbol{\beta}$: vector of the model with degree (n×1).
$\boldsymbol{\vartheta}$ **:** random debris of degree (n×1).

2.1.1 Assumptions of the Poisson regression model:

The Poisson regression model is based on three basic assumptions that are used to build this model:

<u>First assumption:</u>

The conditional probability function of the dependent variable (Y) follows the Poisson distribution when the distribution parameter ($\boldsymbol{\vartheta}$) follows Winkelmann(2008) the Poisson distribution with a parameter of value ($\boldsymbol{\vartheta}$) as in the previously mentioned formula in formula (Cameron, 1998).

<u>Second assumption:</u>

The distribution parameter of the dependent variable (y) is equal to: (Winkelmann, 2008), (Mansson, 2011), and (Mansson, 2012).

$$\vartheta_{i} = e^{x_i'\underline{\beta}} \quad \ldots (2)$$

where:
x_i: represents row (i) of the matrix of independent variables X.

<u>Third assumption:</u>

The pairs of the two variables (yi, xi) have independence between them, that with the adoption of these three assumptions in addition to the characteristics of the Poisson distribution, the arithmetic mean and variance will be as follows: (1997:Long)

$$E\,(y_i/x) = var(y_i/x) = \vartheta_i = e^{\acute{x}_i\underline{\beta}} \quad \ldots (3)$$

For the process of estimating the parameters of the Poisson regression model, which were mentioned in equation (1), and depending on the assumptions that were previously clarified, the random error limit follows the Poisson distribution as well. Therefore, the Maximum Likelinood Method was used to estimate the parameters of the model.

2.2 Conway-Maxwell-Poisson Distribution

1.It is a discrete probability distribution (discrete) that allows the modeling of counting data that is characterized by either excessive or insufficient dispersion. This distribution was defined and presented in 1962 by Conway-Maxwell-Poisson, hence the name (Com-Poisson) or Poisson-Maxwell (cmp) and was developed and the application of this distribution in multiple fields by the world sellers in 2012 (Shmueli, 2005). This distribution has relationships with other distributions under certain conditions and when the parameter s = 1 approaches the Poisson distribution, and when it approaches the Bernoulli distribution and for $\boldsymbol{s} \to \infty$a parameter of its magnitude (Lord, 2008).

$$\frac{\lambda}{1+\lambda} \quad \ldots (4)$$

And when V = 0, the geometric distribution approaches success $\boldsymbol{1-\lambda}$ and that the probability function for it will be as follows:

$$P(Y=y) = f(y;\lambda,s) = \frac{\lambda^y}{(y!)^s}\,\frac{1}{Z(\lambda,s)} \quad ;\ y=0,1,\ldots \quad ; \quad s>0\ ;\ \lambda > 0 \quad \ldots (5)$$

otherwise:

$$\text{Where } Z(d,v) = \sum_{r=0}^{\infty}\frac{\lambda^r}{(s!)^s} \quad \ldots (6)$$

In order to estimate the parameters of the Conway-Maxwell-Poisson regression model, the greatest possibility method of estimation was adopted, as it is the most reliable and optimal method.

3. Discussion of Results

3.1 Definition of Simulation

Simulation is defined as imitating the operation of a system or a process from the real reality, over time. Simulation is the most appropriate way to solve complex physical and mathematical problems, which are difficult to solve in reality as a result of not obtaining real data that represents the approved study or to increase the material cost, as this method is used when All other methods and methods fail by using another systemic behavior that simulates the first, any attempt to repeat a specific process in artificial conditions somewhat similar to natural conditions, and computers are used to implement simulation, and the simulation process also requires the use of models and their development over time, and there are many methods to generate Evaluate the explanatory variables through the use of simulations, including (the inverse method, the Monte-Carlo method...etc).

3.2 Generating Random Variables:

i. Random variables were generated that follow the orderly distribution ($u \sim (0,1)$).

ii. The simulation experiment was implemented using the inverse method of the cumulative distribution function (c.d.f) by taking the inverse (c.d.f) of the two distributions as follows:

$$F(y) = \sum_{n=0}^{y} \frac{\vartheta^{y} e^{-\vartheta}}{y_!} \quad \ldots (7)$$

$$\vartheta = F(y) \quad \ldots (8)$$

$$u = \sum_{n=0}^{y} \frac{\vartheta^{y} e^{-\vartheta}}{y_!} \quad \ldots (9)$$

iii. Three different sample sizes were used (n = 25, 50, 100).

iv. Repeatability of this experiment (r = 1000).

v. The number of explanatory variables for this experiment are variables (p = 6) [(2003: Minka)] .

$$F(x) = p(x+1) = p(x)\frac{\lambda}{(x+1)^{s}} \quad \ldots (10)$$

vi. Akaike Information Criterion (AIC)

$$\mathbf{AIC = 2P\text{-}2\ Ln(L)} \quad \ldots (11)$$

3.3 Results of Simulation

The tables below show the results of the simulation experiment, which were obtained through the implementation of the program and assuming values for the parameters and for the three distributions according to Table (1-3) and with different sizes n = 25, 50, and 100 and at repetitions r = 1000, as the results reflect the parameter values and the mean squared error (MSE) and the values of the Akaki criterion (AIC) for each distribution and under the two proposed models under study.

Table (1) default values for the parameters of the two distributions

Distribution	Poisson	COM-Poisson	
	θ	λ	s
I	3	2.5	3
II	3.5	3.2	4

Table (2) represents the parameter estimators for the first model.At a sample size of n = 25

Distribution	Poisson -Model		COM-Poisson- Model	
	poisson	COM-Poisson	poisson	COM-Poisson
β_0	-1.21074	-2.93716	-1.37221	-2.59567
β_1	-0.177	-0.177	-0.11187	-0.18659
β_2	0.304289	0.304289	0.369423	0.645777
β_3	0.230114	0.230114	0.144324	0.456621
β_4	0.348476	0.348476	0.187006	0.391705
β_5	0.138043	0.138043	0.203177	0.128453
β_6	-0.18379	-0.18379	-0.11866	0.116419
MSE	0.361158	0.00935	0.089266	0.481738
AIC	21.70934	19.25474	18.2565	21.31997

The Poisson model was the best and under the com-poisson distribution through the mean square error criterion. The Poisson model achieved the lowest AIC under the Poisson distribution.

Table (3) represents the parameter estimators for the first model.At a sample size of n = 50

Distribution	Poisson - Model		COM-Poisson- Model	
	poisson	COM-Poisson	poisson	COM-Poisson
β_0	-0.80173	-6.53088	-0.6456	-6.45105
β_1	0.114094	0.114094	-0.2326	-0.03521
β_2	0.185056	0.185056	0.341185	0.035753
β_3	-0.25108	-0.25108	-0.60957	-0.40038
β_4	0.51193	0.51193	0.398217	0.362626
β_5	0.008784	0.008784	0.164912	0.088616
β_6	0.026403	0.026403	0.182531	-0.1229
MSE	0.204174	0.016163	0.154358	1.293239
AIC	23.26122	19.34629	11.80292	21.90505

The Poisson model was the best under the com-poisson distribution through the lower MSE and also the lower AIC for the Poisson distribution and under the com-poisson mode

Table (4) represents the parameter estimators for the first model.. At a sample size of n = 10

Distribution	Poisson- Model		COM-Poisson- Model	
	poisson	COM-Poisson	poisson	COM-Poisson
β_0	-0.99977	-4.50718	-0.98942	-4.10103
β_1	0.017786	0.017786	0.048545	0.423936
β_2	0.040016	0.040016	0.042147	-0.39907
β_3	0.026306	0.026306	0.465241	-0.0341
β_4	0.196534	0.196534	0.635469	-0.24255
β_5	0.21081	0.21081	-0.49014	-0.22827
β_6	-0.01285	-0.01285	-0.21477	0.03468
MSE	0.008694	0.009966	0.037197	0.582029
AIC	23.79953	21.73292	16.78471	25.02851

The Poisson model was the best by the MSE criterion and under the Poisson distribution and it achieved the lowest AIC for the Poisson distribution. And under the com-poisson model.

Table (5) represents the parameter estimators for the second modelAt a sample size of n = 25

Distribution	Poisson- Model		COM-Poisson- Model	
	poisson	COM-Poisson	poisson	COM-Poisson
β_0	-0.13289	-3.64902	-0.19799	-2.74106
β_1	0.003272	0.003272	-0.06183	0.911233
β_2	0.250036	0.250036	0.184933	-0.26886
β_3	0.002947	0.002947	-0.06216	0.12525
β_4	-0.04907	-0.04907	-0.11417	0.858894
β_5	0.125975	0.125975	0.060872	0.894103
β_6	0.076474	0.076474	0.011371	-0.05046
MSE	0.036138	0.035493	0.637259	3.797011
AIC	21.36842	15.54949	14.33616	23.03846

The Poisson model achieved less MSE under the com-poisson distribution and also the com-poisson model achieved less AIC under the Poisson distribution.

Table (6) represents the parameter estimators for the second model. At a sample size of n = 50

Distribution	Poisson- Model		COM-Poisson- Model	
	poisson	COM-Poisson	poisson	COM-Poisson
β_0	-0.05347	-5.58242	-0.23814	-4.80217
β_1	0.160133	0.160133	0.5308	0.784042
β_2	0.008156	0.008156	0.173688	0.788402
β_3	0.34188	0.34188	0.15721	0.870736
β_4	-0.07792	-0.07792	0.29275	0.511277
β_5	-0.04246	-0.04246	-0.22713	-0.73297
β_6	0.040295	0.040295	-0.14438	0.087343
MSE	0.047139	0.037461	0.01143	0.614065
AIC	23.07424	22.60882	12.01635	23.34373

The com-poisson model was the best by the MSE criterion and also by the AIC criterion and under the Poisson distribution.

Table (7) represents the parameter estimators for the second model. At a sample size of n = 100

Distribution	Poisson- Model		COM-Poisson-Model	
	poisson	COM-Poisson	poisson	COM-Poisson
β_0	-0.01128	-3.93895	0.84246	-3.00936
β_1	0.093836	0.093836	0.947579	0.294426
β_2	0.118541	0.118541	-0.14708	-0.74863
β_3	-0.0375	-0.0375	-0.46618	-0.90468
β_4	0.18057	0.18057	-0.64825	-0.6866
β_5	0.05552	0.05552	-0.20174	-0.46659
β_6	-0.05986	-0.05986	-0.88868	-0.92703
MSE	0.005961	0.01383	1.343002	7.398824
AIC	24.24246	25.11142	16.38712	27.70217

Poisson model achieves lower MSE under Poisson distribution while compoisson model achieves lower AIC under Poisson distribution.

4. Conclusions

i. From the results of the simulation experiment at three different sample sizes (n = 25, 50, 100) and with repetitions (r = 1000) and for six parameters p = 6, it was found that the (Poisson) model was the best at the three sizes, through the standard (MSE) either from During the (AIC) standard, the (COM-POISSON) model was the best.

ii. As for the second model (Com-poisson), the (Poisson) model achieved less (MSE) at the sample size (n = 25,100), while the (com-poisson) model achieved less (AIC) for all the selected sizes.

References

1.Avci , Esin , (2018),"Using Count Regression Models to Determine the Factors Which Effects the Hospitalization Number of people with Schizophrenia" ,Giresun University .

2.Cameron , A .C and Trivedi, P .K (1998) "Regression Analysis of Count Data", Cambridge University Press, New York USA.

3.Conway, R.W., Maxwell, W.L. (1962). “A queuing model with state dependent service rates”, Journal of Industrial Engineering, 12:132-136.

4.Hameed, L. M. A. and Jawad, E. H.(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.

5.Hameed, L. M. A. and Jawad, E. H.(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.

6.Long, J . S (1997) "Regression Models for Categorical and Limited Independent Variables", SAGE Publicayion Inc, USA.

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

8.Mansson, K & Shukur , G (2011). “A Poisson Ridge Regression Estimator" , Economic Modeling, Vol. 28, Issue 4 ,pp. 1475-1491.

9.Mansson, K , Kebria, B . M , Sjolander , P and Shukur, G (2012). “New Liu Estimates for the Poisson Regression Model: Methods and Application, International Journal of Statistics and Probability, Vol. 1, No. 1.

10. Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright, P. (2005). “A Useful Distribution for Fitting Discrete Data: Revival of the Conway–Maxwell–Poisson Distribution”, Journal of The Royal Statistical Society. Series C (Applied Statistics), 54(1): 127-142.

11. Winkelmann, R (2008) "Economic Analysis of Count Data ", 5 th ed., Springer, Verlag Berlin Heidelberg Germany.

مقارنة نماذج انحدار بواسون و كونواي ماكسويل بواسون باستعمال المحاكاة

سارة عدنان ردام
جامعة بغداد/ كلية الادارة والاقتصاد/ قسم الاحصاء.
بغداد، العراق
Sara.adnan1201a@coadec.uobaghdad.edu.iq

لمياء محمد علي حميد
جامعة بغداد/ كلية الادارة والاقتصاد/ قسم الاحصاء.
بغداد، العراق
lamiaa.mohammed@coadec.uobaghdad.edu.iq

Received: 28/2/2023 Accepted: 30/3/2023 Published: June / 2023

مستخلص البحث

تعد نماذج الانحدار إحدى أهم النماذج المستخدمة في الدراسات الحديثة, لاسيما الابحاث والدراسات الصحية لما يحقق من نتائج مهمة, فقد تم استخدام انموذجين من انماذج الانحدار وهما :انموذج انحدار بواسون وانموذج كونواي ماكسويل-بواسون حيث هدفت هذه الدراسة إلى إجراء مقارنة بين النموذجين واختيار الأفضل بينهما بأستخدام اسلوب المحاكاة وعند احجام عينات مختلفة (n=25,50,100) وبتكرار(r=1000) وقد تم اعتماد برنامج الماتلاب لأجراء تجربة المحاكاة حيث اظهرت النتائج تفوق انموذج بواسون من خلال معيار متوسط مربعات الخطأ وايضا من خلال معيار الأكايكي لنفس التوزيع .

نوع البحث: ورقة بحثية.
المصطلحات الرئيسة للبحث: انحدار بواسون ، كونواي ماكسويل بواسون ، طريقة الامكان الاعظم , متوسط مربعات الخطأ , الاكايكي ,محاكاة المعكوس.

*البحث مستل من رسالة ماجستير