Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

An Application Comparison of Two Negative Binomial Models on Rainfall Count Data

To cite this article: L. H. Hashim *et al* 2021 *J. Phys.: Conf. Ser.* **1818** 012100

View the article online for updates and enhancements.

An Application Comparison of Two Negative Binomial Models on Rainfall Count Data

L. H. Hashim[1], N. K. Dreeb[2], K. H. Hashim[3], Mushtak A. K. Shiker [4 *]

[1]Department of Planning and Studies, University of Babylon, Babil-Iraq.

[2,3,4]Department of Mathematics, College of Education for Pure Sciences, University of Babylon, Babil-Iraq.

E-mail: [1]luayinfo@gmail.com , [2]nabiha250222@gmail.com ,
[3]teacher89karrar@gmail.com , [4]mmttmmhh@yahoo.com

[*] Corresponding Author.

Abstract. Counts data models cope with the response variable counts, where the number of times that a certain event occurs in a fixed point is called count data, its observations consists of non-negative integers values {0,1,2,}. Due to the nature of the count data, it is generally considered that response variables do not follow normal distribution. Therefore, because of the skewed distribution, linear regression is not an effective method for analyzing counting results. And hence, the use of the linear regression model to analyse count data is likely to bias the outcomes, "Negative binomial regression" is likely to be the optimal model for analyzing count data under these limitations. Researchers may sometimes count more zeros than expected. Going to count data with several Zeros gives rise to the "Zero-inflation" concept. In health, marketing, finance, econometrics, ecology, statistical quality control, geographical and environmental fields, data with abundant zeros is common when counting the incidence of certain behavioural and natural events, such as the frequency of alcohol consumption, drug consumption, the amount of cigarettes smoked, the incidence of earthquakes, rainfall, etc. The Negative Binomial, "Zero-Inflated Negative Binomial" (ZINB), and "Zero-Altered Negative Binomial" (ZANB) models were used in this paper to analyse rainfall data.

Keywords: Negative Binomial. Zero-Inflated. Zero-Altered. Counts data. Excess zero.

1. Introduction

In a diverse range of applications, count data, including zero counts arise, so count models have become widely common in many fields. In the field of statistics, the count data can be defined as that type of observation that only takes the value of non-negative integers, Researchers may sometimes count more zeros than expected [1]. One can describe surplus zero as Zero-Inflation. Excess zero sometimes may be the reason of occurs Over-dispersion (variance a lot larger than mean) [2]. In the study of discrete data, the Over-dispersion principle is widely used. Therefore, linear regression is not applicable procedure to estimate the parameters of predictors due to the asymmetric distribution of the response variable. Under these limitations, "Poisson" regression and "Negative binomial" regression are used to model the Count data [3, 4].

Lambert (1992) [3] addressed this problem and proposed the "zero-inflated Poisson" model with an application in production quality also proposed by Greene (1994) and the "zero-altered Poisson" model (Another popular approach for modelling excess zeros in count data is the use of hurdle models (also referred to as a zero-altered model) developed by Cragg (1971)) [6] that have been proposed to cope with an overabundance of zeros (also called a zero- altered model). In last years, zero-Inflation models have become so important [6- 17]. The authors have many papers dealing with optimization, transportation problems and nonlinear systems, see [18- 30], but in this work we focus on the excess zero.

In some commonly used discrete distributions the mean of the distribution related to the variance, the reason of exhibit Over-dispersion [7]. That is, in the data in which there is evidence that the variance of the dependent variable is greater than the mean, over-dispersion occurs.

In health, marketing, finance, econometrics, ecology, statistical quality control, geographical and environmental fields, data with abundant zeros is common when counting the incidence of certain behavioural and natural events, such as the frequency of alcohol consumption, drug consumption, the amount of cigarettes smoked, the incidence of earthquakes, rainfall, etc.

Famoye and Consul (1992) [31] proposed "generalized Poisson" distribution which can take consideration of "over-dispersion" of Poisson distribution. The extension of generalized Poisson distribution is "zero-inflated generalized Poisson" (ZIGP) suggested by Famoye and Singh (2006) [4].

Some other models, such as the "negative binomial" model, were used to analyse count data. The "zero-altered negative binomial" (ZANB) model discussed by Heilbron (1994) is a natural stretch of the "negative binomial" model to accommodate increased zeros in the data. In this paper, I focus on the models, Negative Binomial, ZINB, and ZANB to analyze rainfall data.

2. Negative Binomial Regression Model (NBRM)

Negative binomial regression is one of types of generalized linear models in which the "dependent variable" is a count of the number of times an event occurs [2]. Negative binomial regression is similar to the multiple regression excepting that the response variable (y) is an observed count that follows the " negative binomial distribution". Therefore, the possible values of (y) are "nonnegative integers".

Suppose that $y_1,\ldots,y_n$ are a random sample from the Negative binomial distribution, then the p.m.f of y_1 is expressed as

$$p\left(y_i;\frac{1}{\alpha},\mu_i\right)=\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}\qquad ;y=0,1,2,\ldots \tag{1}$$

By assumptions of GLM [1,5,10,11], We have

$Y_i\sim NB\left(\mu_i,\frac{1}{\alpha}\right)$; $E(Y_i)=\mu_i$, $Var(Y_i)=\mu_i+\alpha\mu_i^2$ and $\mu_i=e^{\eta(X_{i1},\ldots,X_{iq})}=e^{X'\beta}$

Where $X'\beta=\alpha+\beta_1X_{i1}+\cdots+\beta_qX_{iq}$ and $X_{i1},\ldots,X_{iq}$ are the independent variables.

Given the p.m.f in (1) and using the method of maximum likelihood and assuming independence of the observations, We can estimate regression parameters as follow

$$L=\prod_i p(y_i\,;\mu_i)$$

$$L=\prod_i^n\left[\frac{\Gamma(y_i+\frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}\right]$$

$$log(L)=\sum_{i=1}^{n}\begin{bmatrix} y_i log\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)-\frac{1}{\alpha}log(1+\alpha\mu_i) \\ +log\Gamma\left(y_i+\frac{1}{\alpha}\right) \\ -log\Gamma(y_i+1)-log\Gamma\left(\frac{1}{\alpha}\right)\end{bmatrix}$$

$$log(L)=\sum_{i=1}^{n}\begin{bmatrix} y_i log\left(\frac{\alpha e^{X'\beta}}{1+\alpha e^{X'\beta}}\right)-\frac{1}{\alpha}log\left(1+\alpha e^{X'\beta}\right) \\ +log\Gamma\left(y_i+\frac{1}{\alpha}\right) \\ -log\Gamma(y_i+1)-log\Gamma\left(\frac{1}{\alpha}\right)\end{bmatrix}$$

By taking partial derivatives of the parameters and equalizing the likelihood equation to zero

$$\frac{\partial log\ (L)}{\partial\beta}=\frac{\partial}{\partial\beta}\left[\sum_{i=1}^{n}\begin{bmatrix} y_i log\left(\frac{\alpha e^{X'\beta}}{1+\alpha e^{X'\beta}}\right)-\frac{1}{\alpha}log\left(1+\alpha e^{X'\beta}\right) \\ +log\Gamma\left(y_i+\frac{1}{\alpha}\right) \\ -log\Gamma(y_i+1)-log\Gamma\left(\frac{1}{\alpha}\right)\end{bmatrix}\right]=0 \quad (2)$$

$$\frac{\partial log\ (L)}{\partial\alpha}=\frac{\partial}{\partial\alpha}\left[\sum_{i=1}^{n}\begin{bmatrix} y_i log\left(\frac{\alpha e^{X'\beta}}{1+\alpha e^{X'\beta}}\right)-\frac{1}{\alpha}log\left(1+\alpha e^{X'\beta}\right) \\ +log\Gamma\left(y_i+\frac{1}{\alpha}\right) \\ -log\Gamma(y_i+1)-log\Gamma\left(\frac{1}{\alpha}\right)\end{bmatrix}\right]=0 \quad (3)$$

Applying numerical methods such as “Newton Raphson” to solve equations (2) and (3).

3. Zero-Inflated Models (ZI)

In certain populations, excess zeros lead to zero-inflation which is made up two types of data subgroups (data generation), the first subgroup is a set of only zeros count (true zeros and false zeros), and the second subgroup is a set of count variables (with true zeros) that distributed according to Poisson distribution (Lambert 1992, Van den Broek 1995) [1, 3, 7].

4. Zero-Inflated Negative Binomial Regression Model (ZINB)

The “zero-inflated Negative binomial” regression is used for modelling count data that show over-dispersion and zero counts (excess zeros). This model takes into account that there are two types of data sources, the first source is zero type and the second is comes from data follows Negative binomial distribution [1].

According to Lambert (1992), response variable Y_i is independent with

$Y_i \sim 0$ with probability (θ_i) and $Y_i \sim$Negative binomial $(\mu_i, \frac{1}{\alpha})$ with probability $(1+\theta_i)$

Therefore,

$$\Pr(Y_i=0)=\theta_i+(1-\theta_i)\times \Pr(\text{Count process at (i)gives a zero}) \quad (4)$$

by assuming the Yi follows a Negative binomial distribution with mean μ_i

Journal of Physics: Conference Series **1818** (2021) 012100 doi:10.1088/1742-6596/1818/1/012100

$$p\left(y_i;\frac{1}{\alpha},\mu_i|y_i\geq 0\right)=\frac{\Gamma(y_i+\frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}$$

Subsequently

The term $\Pr(\text{Count process at (i)gives a zero})$ is given by

$$p\left(y_i=0;\frac{1}{\alpha},\mu_i\middle|y_i\geq 0\right)=\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}$$

Hence, Equation (4) can now be written as

$$\Pr(Y_i=0)=\theta_i+(1-\theta_i)\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} \quad (5)$$

For the probability that Y_i is a non-zero count;

$$\Pr(Y_i=y_i)=(1-\theta_i)\times\Pr(\text{Count process}) \quad (6)$$

Hence, Equation (6) can be rewritten as follows

$$p(Y_i=y_i|y_i>0)=(1-\theta_i)\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \quad (7)$$

Therefore, the probability density function for a ZINB model is given by

$$P(Y_i=y_i)=\begin{cases}\theta_i+(1-\theta_i)\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} & \text{if} \quad y_i=0\\ (1-\theta_i)\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} & \text{if} \quad y_i>0\end{cases} \quad (8)$$

By GLM[1,5,10,11], $\mu_i=e^{X'_i\beta_i}$, where X'$_i$ are knows independent variables, Lambert (1992) suggested the functional form for modelling the parameter θ_i as logistic function, which is given by

$$\text{Log}\left(\frac{\theta_i}{1-\theta_i}\right)=z'_i\gamma_i$$

and therefore,

$$\theta_i=\frac{e^{z'_i\gamma_i}}{1+e^{z'_i\gamma_i}}>0$$

Where; Z : the covariates and γ : are regression coefficients.

The corresponding Log-Likelihood function of (8) is given as follow

$$\log(L)=$$

$$\Sigma_i^n\begin{bmatrix} I(y_i=0)log\left(\theta_i+(1-\theta_i)\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\right)+\\ I(y_i>0)(\log((1-\theta_i)+\log\left(\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_i+1)}\right)-\\ \left(y_i+\frac{1}{\alpha}\right)\log(1+\alpha\mu_i)+y_i\log(\alpha\mu_i))\end{bmatrix} \quad (9)$$

Subsequently

$$E(Y_i)=\mu_i(1-\theta_i)$$
$$Var(Y_i)=(1-\theta_i)\left(\mu_i+\alpha\mu_i^2\right)+\mu_i^2(\theta_i^2+\theta_i)$$

5. Zero-Altered Models (ZA)

Zero-altered models known as a two-part models, Where the first part is a binary outcome model governs with binomial probability, and the second part is a truncated count model [7]. In zero-inflated models assumed that count data consist of two types of data subgroups, the first subgroup is a set of only zeros count (true zeros and false zeros), and the second subgroup is a set of count variables (with true zeros) [32]. While, zero-altered models do not discriminate between the types of zeros; they are

simply zeros. For the zero-altered models, the basic concept is that the outcomes are treated as absence and presence zeros data'. This means that the outcomes are divided into two groups, the first includes all zeros, the second includes non-zero counts [1].
Where, The binomial distribution is used to model the absence and presence, and a Negative Binomial distribution for the counts [2, 7]. To measure a non-zero count should be modified the distribution and exclude the possibility of a zero observation, and this is called a zero-truncated distribution.
Assume that the zeros are follow the probability mass function (p.m.f) $f_1(.)$ with $P(y=0)=f_1(0)$ and $P(y>0)=1-f_1(0)$, while the positive outcomes are formed by the probability mass function truncated at zero given by

$$f_2(y|y>0)=f_2(y)/[1-f_2(0)]$$

Hence, the Hurdle (Altered) probability mass function as follow

$$P(y)=\begin{cases} f_1(0) & ; y=0 \\ \frac{1-f_{1(0)}}{1-f_2(0)} f_2(y) & ; y>0 \end{cases} \tag{10}$$

6. Zero-Altered Negative binomial regression Model (ZANBM)
Suppose that the probability of measuring zero observation in the first part of Hurdle structure is modelled with a binomial distribution', Where θ_i is the probability that $y_i=0$.
Suppose that be the response variable for the positive counts' (truncated at zero) with Negative binomial probability mass function (1).
Furthermore, let the probability of observing $y_i=0$ in the first part of Hurdle model (zero count) as follow

$$P(y_i=0)=f_1(0)=\theta_i \tag{11}$$

Where, the probability of observing ($y_i>0$) in the second part of Hurdle model (positive counts) as follow

$$p(y_i;\mu_i|y_i>0)=f_2(y)=\frac{\Gamma(y_i+\frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \tag{12}$$

Therefore, substituting (1), (11), and (12) in Zero-Altered (6), we have

$$P(Y_i=y_i)=\begin{cases} \theta_i & ; y_i=0 \\ \dfrac{(1-\theta_i)\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}}{\left(1-\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\right)} & ; y_i>0 \end{cases} \tag{13}$$

By GLM[1,5,10,11], $\mu_i=e^{X'_i\beta_i}$ where X_i are knows independent variables, Lambert (1992) suggested the functional form for modelling the parameter θ_i as logistic function, which is given by

$$Log\left(\frac{\theta_i}{1-\theta_i}\right)=z'_i\gamma_i$$

and therefore,

$$\theta_i=\frac{e^{z'_i\gamma_i}}{1+e^{z'_i\gamma_i}}>0$$

Where; Z : the covariates and γ : are regression coefficients.
The corresponding Log-Likelihood function is given as follow

$$log(L)=\sum_i^n\left[I(y_i=0)\,log(\theta_i)+I(y_i>0)log\left(\frac{(1-\theta_i)\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}}{\left(1-\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\right)}\right)\right] \tag{14}$$

The mean and variance for ZANB are

$E(Y_i) = \frac{1-\theta_i}{1-P_0}\mu_i$ where $P_0 = \left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}$

$$Var(Y_i) = \frac{1-\theta_i}{1-P_0}(\mu_i^2 + \mu_i + \alpha\mu_i^2) - \left(\frac{1-\theta_i}{1-P_0}\mu_i\right)^2$$

7. Model Selection

It is important that we have one or more a criterion to consider the best results and choose the appropriate model for data representation. There are several methods that provide a measure for selecting the appropriate model, The following four methods will be used: AIC is an evaluating model fit for a given data among different types of non-nested models , and its formula is given as $AIC = -2logL + 2k$, BIC is another estimator for evaluating model fit for a given data among different types of non-nested models, and its formula is given as $BIC = -2logL + klogn$, Likelihood ratio test (LR) is a statistical test used to compare two nested models, its formula is given as $LR = -2log\,(L_1/L_2)$ and Vuong test (V) is a statistical test used to compare non-nested models, It is defined as

$$V = (\sqrt{n}(\frac{1}{n}\sum_{i}^{n} m_i)/(\sqrt{\frac{1}{n}\sum_{i}^{n}(m_i - \bar{m})^2})$$

Where $m_i = log\big(P_1(Y_i|X_i)\big) - log\big(P_2(Y_i|X_i)\big)$.

If V>1.96 , then the first model is preferred. If V<-1.96, then the second one is preferred. If | V | <1.96, none of the models are preferred.

8. Data Analysis

Data were collected from database of the meteorology and seismology organization in Iraq for Hilla weather station. The weather station are located in central Iraq, specifically in the city of Hilla (about 116 kilometers south of Baghdad).

The count response variable of interest to be modeled "Rainfall hours" measured at Hilla weather station. The predictor variables consists of six climate variables derived from Iraqi Meteorological Organization and Seismology database, which include measurements of rainfall, sea pressure, station pressure, wind speed, temperature, and humidity, as shown in Table (1). Data contain observations of (731) for two years.

Table 1. Summary statistics of explanatory variables and response used in our count data regression models in Hilla weather station.

variables	Minimum value	First quarter	Median	Mean	Third quarter	Maximum value
Rainfall (hours)	0	0	0	0.6553	0	20
Wind speed (m/s)	0	0.6	1.4	1.619	2.3	9.3
Temperature (°C)	3	15.8	25	23.97	32.85	40.5
Humidity (%)	17	31.8	40.6	44.54	56	94
Station pressure (1bar/1000)	0.9908	1.0007	1.0068	1.0074	1.0131	1.3804
Sea pressure (1bar/1000)	0.9947	1.0046	1.0108	1.0109	1.0171	1.0287

The distribution of the number of non-rainfall hours in Hilla weather stations for the two years is shown in figure 1

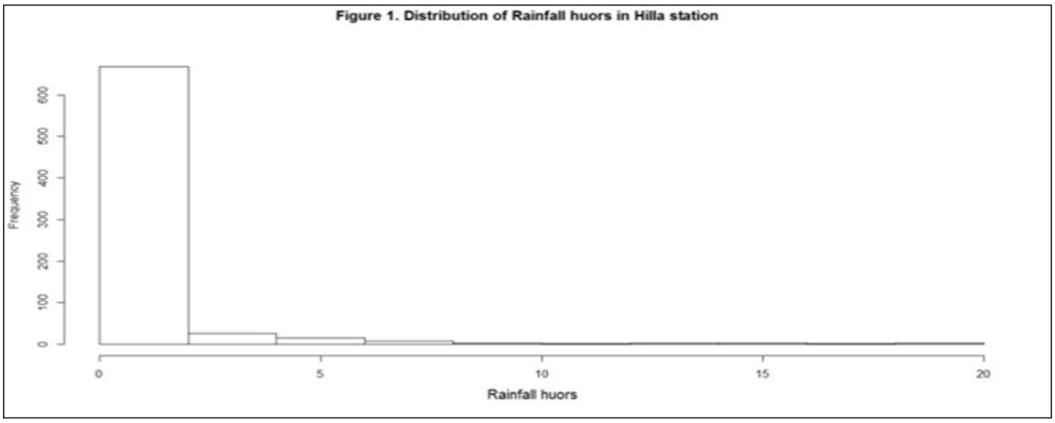

Figure 1. distribution of rainfall hours in Hilla station.

We conducted a test of over-dispersion and the results of this test are shown below likelihood ratio test of ($H_\circ$: Mean=Variance), (H_1: Mean < Variance), as restricted NB model, Critical value of test statistic at the alpha= 0.00 level: 2.7055, For Hilla weather station, Chi-Square test statistic= 579.6014, p-value = <2.2e-16. The significance of X^2-statistics implies the existence of over-dispersion. Therefore, in the next section, we develop Negative Binomial model to handle the issue of over-dispersion.

9. Negative Binomial Regression

In order to address the issue of over-dispersion, we used The model fit statistics and estimated coefficients of Negative Binomial regression model are given in Table 2 and Table 3.

Table 2. Fit statistics of Negative Binomial regression model, Rainfall count data

criterions	Hilla weather station
-2Log Likelihood	892.7366
AIC	906.7386
BIC	938.8995

Table 3. Estimated coefficients of Negative Binomial regression model, Rainfall count data in Hilla weather station

Parameter	Estimate	Standard Error	z Value	Pr > |z|
Intercept	72.08011	46.31368	1.556	0.12
Wind speed	0.43955	0.09276	4.738	2.15e-06
Temperature	-0.05902	0.04534	-1.302	0.193
Humidity	0.0921	0.01372	6.715	1.88e-11
Station pressure	-5.48177	47.6056	-0.115	0.908
Sea pressure	-70.94321	65.8155	-1.078	0.281
Alpha	0.15	0..0248		

Lambert (1992) and Mullahy (1986) indicated that Negative Binomial regression might not be an appropriate model for count data with excess zeros because it increases the probabilities of both zero and non-zero counts [3]. Since the initial data analysis of our data implied excess zeros (more than 87.8% of the responses in Hilla weather station , have non- Rainfall days (rainfall hours are zeros)), we develop Zero-inflated regression to handle excessive number of zeros.

10. Zero-Inflated Regression Models

To fixable the excess zeros problem in non-Rainfall days (rainfall hours are zeros), We used Zero-inflated regression models.

11. Zero-Inflated Negative Binomial Regression (ZINBR) Model

We used the same explanatory variables in both parts of the ZINBR model. The model fit statistics and estimated coefficients of ZINBR model are given in Table 4 and Table 5.

Table 4. Fit statistics of Zero-Inflated Negative Binomial Regression (ZINBR) model, Rainfall count data

criterions	Hilla weather station
-2Log Likelihood	774.8
AIC	800.7555
BIC	814.3665

Table 5. Estimated coefficients of Zero-Inflated Negative Binomial Regression (ZINBR) model, Rainfall count data in Hilla weather station

Parameter	Estimate	Standard Error	z Value	Pr > |z|
NB _ Intercept	4.450441	28.89126	0.154	0.877577
NB _ Wind speed	-0.01452	0.047221	-0.307	0.758472
NB _ Temperature	0.025615	0.023359	1.097	0.272823
NB _ Humidity	0.032441	0.007135	4.547	5.44e-06
NB _ Station pressure	18.075792	27.86469	-0.627	0.520412
NB _ Sea pressure	-23.29534	28.50922	0.775	0.418057
Logit _ Intercept	-144.89025	51.27398	-2826	0.00472
Logit _ Wind speed	-0.66723	0.10557	-6.320	2.61e-10
Logit _ Temperature	0.07038	0.04973	1.415	0.15701
Logit _ Humidity	-0.10567	0.01591	-6.643	3.07e-11
Logit _ Station pressure	-0.12663	31.50265	-0.004	0.99679
Logit _ Sea pressure	150.17569	59.40599	2.528	0.01147
Log (Alpha)	0.937272	0.280696	3.339	0.000841

12. Zero-Altered Regression Models (ZARM)

To fixable the excess zeros problem in non-Rainfall days (rainfall hours are zeros), We used Zero-Altered regression models.

13. Zero-Altered Negative Binomial Regression (ZANBR)

We used the same explanatory variables in both parts of the ZANBR 'model. The model fitting statistics and parameters estimation of ZANBR model are given in Table 6 and Table 7.

Table 6. Fit statistics of Zero- Altered Negative Binomial Regression (ZANBR) model, Rainfall count data

criterions	Hilla weather station
-2Log Likelihood	772
AIC	797.948
BIC	811.6556

Table 7. Estimated coefficients of Zero- Altered Negative Binomial Regression (ZANBR) model, Rainfall count data in Hilla weather station

Parameter	Estimate	Standard Error	z Value	Pr > |z|
NB _ Intercept	6.161e+00	3.296e+01	0.187	0.8517
NB _ Wind speed	-2.728e-02	5.060e-02	-0.539	0.5898
NB _ Temperature	2.584e-02	3.519e-02	0.734	0.4627
NB _ Humidity	3.275e-02	7.582e-03	4.319	1.57e-05
NB _ Station pressure	1.801e+01	2.010e+03	0.009	0.9929
NB _ Sea pressure	-2.493e+01	2.013e+03	-0.012	0.9901
Logit _ Intercept	146.38914	49.35864	2.966	0.00302
Logit _ Wind speed	0.65227	0.0991	6.582	4.64e-11
Logit _ Temperature	-0.0656	0.04771	-1.375	0.1691
Logit _ Humidity	0.1103	0.01536	7.182	6.89e-13
Logit _ Station pressure	-1.09452	42.64166	-0.026	0.97952
Logit _ Sea pressure	-150.90185	64.65376	-2.334	0.0196
Log (Alpha)	8.158e-01	3.201e-01	2.548	0.0108

14. Model Comparison
We used Vuong test to compare non-nested models and Likelihood ratio test to compare nested models, The results of all the Vuong tests are summarized in Table 8 and the results of all Likelihood ratio tests are summarized in Table 9. Furthermore, the results of all information criterions (fit statistics) for all models were summarized in Table 10.

Table 8. Model comparison by Vuong test for non-nested models for Hilla weather station

Model	Vuong Statistic	Best model
ZINB vs NB	5.943327	ZINB
ZINB vs ZANB	-2.14748	ZANB

Note: “If $V > 1.96$, the first model is preferred. If $V < -1.96$, then the second one is preferred. If $|V| < 1.96$, none of the models are preferred ”.

Table 9. Model comparison by likelihood ratio test for nested models for Hilla weather station

Model	Likelihood Ratio Test (p-value)	Best model
NB vs ZANB	0.29	ZANB

Note:
H_o : the simpler model is preferred.
H_1 : the more complex model is preferred.
If p-value < 0.05, we reject H_o , H_1 is preferred.

Table 10. Fit statistics of all models, Rainfall count data Hilla weather station

models	criterions		
	-2Log Likelihood	AIC	BIC
NB regression	892.7366	906.7386	938.8995
ZINBR	774.8	800.7555	814.3665
ZANBR	772*	797.984*	811.5665*

*The best model.

15. Application results

After estimating the regression parameters for all models using real counting data. The test criteria values for all models were obtained for the purpose of comparing these models and selecting the best ones to represent our data. The results in Table 10 indicated that Zero-Altered Negative Binomial (ZANBR) regression model was the best count data model for our data, Although it is hard to distinguish Zero-Inflated Negative Binomial, (ZINBR) regression models, it is better than Negative Binomial regression model.

References

[1] Zuur A F et.al. 2009 *Mixed effects models and extensions in ecology with R*, Springer, NY: New York.
[2] Greene W H 1994 *Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models*, Leonard N. Stern School of Business, New York University.
[3] Lambert D 1992 Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing, *Technometrics*, **34**: 1, P. 1-14.
[4] Famoye F and singh KP 2006 Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data, *Journal of Data Science*, **4**, P. 117-130.
[5] Cragg J G 1971 Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods, *Econometrica: Journal of the Econometric Society*, **39**: 5, P. 829-844.
[6] Heilbron D C 1994 Zero-Altered and Other Regression Models for Count Data with Added Zeros, *Biometrical Journal*, **36**, P. 531-547.

[7] Hassan Z A H H and Shiker M A K 2018 Using of generalized baye's theorem to evaluate the reliability of aircraft systems, Journal of Engineering and Applied Sciences, (Special Issue 13), pp. 10797-10801.

[8] Abed S A, Sulaiman H K. and Hassan Z A H 2019 Reliability Allocation and Optimization for (ROSS) of a Spacecraft by using Genetic Algorithm, *J. Phys.: Conf. Ser*. **1294** 032034.

[9] Hassan Z A H and Balan V 2015 Reliability extrema of a complex circuit on bi-variate slice classes, *Karbala International Journal of Modern Science*, **1**: 1, p. 1- 8.

[10] Hassan Z A H and Balan V 2017 Fuzzy T-map estimates of complex circuit reliability, International Conference on Current Research in Computer Science and Information Technology (ICCIT-2017), *IEEE, Special issue*, p .136- 139.

[11] Abdullah G and Hassan Z A H 2020 Using of particle swarm optimization (PSO) to addressed reliability allocation of complex network, *J. Phys.: Conf. Ser*. **1664** 012125.

[12] Abdullah G and Hassan Z A H 2020 Using of Genetic Algorithm to Evaluate Reliability Allocation and Optimization of Complex Network, *IOP Conf. Ser.: Mater. Sci. Eng*. **928** 042033.

[13] Hassan Z A H and Mutar E K 2017 Geometry of reliability models of electrical system used inside spacecraft, 2017 Second Al- Sadiq International Conference on Multidisciplinary in IT and Communication Science and Applications (*AIC-MITCSA*), p. 301-306

[14] Shiker M A K and Sahib Z 2018 A modified trust-region method for solving unconstrained optimization. Journal of Engineering and Applied Sciences, 13:22, p. 9667– 9671.

[15] Mahdi M M and Shiker M A K 2020 Three-Term of New Conjugate Gradient Projection Approach under Wolfe Condition to Solve Unconstrained Optimization Problems, Journal of Advanced Research in Dynamical and Control Systems, 12:7, p. 788- 795.

[16] Mahdi M M and Shiker M A K 2020 A new projection technique for developing a Liu-Storey method to solve nonlinear systems of monotone equations, J. Phys.: Conf. Ser. 1591012030.

[17] Mahdi M M and Shiker M A K 2020 Three terms of derivative free projection technique for solving nonlinear monotone equations, *J. Phys.: Conf. Ser*. 1591012031.

[18] Dwail H H and Shiker M A K 2020 Reducing the time that TRM requires to solve systems of nonlinear equations, *IOP Conf. Ser.: Mater. Sci. Eng*. **928** 042043.

[19] Mahdi M M and Shiker M A K 2020 A New Class of Three-Term Double Projection Approach for Solving Nonlinear Monotone Equations, *J. Phys.: Conf. Ser*. 1664 012147.

[20] Wasi H A and Shiker M A K 2020 A new hybrid CGM for unconstrained optimization problems, *J. Phys.: Conf. Ser*. **1664** 012077.

[21] Dwail H H and Shiker M A K 2020 Using a trust region method with nonmonotone technique to solve unrestricted optimization problem, *J. Phys.: Conf. Ser*. 1664 012128.

[22] Hussein H A, Shiker M A K and Zabiba M S M 2020 A new revised efficient of VAM to find the initial solution for the transportation problem, *J. Phys.: Conf. Ser*. **1591** 012032.

[23] Hussein H A and Shiker M A K 2020 Two New Effective Methods to Find the Optimal Solution for the Assignment Problems, *Journal of Advanced Research in Dynamical and Control Systems*, **12**: 7, p 49- 54.

[24] Hussein H A and Shiker M A K 2020 A modification to Vogel's approximation method to solve transportation problems, *J. Phys.: Conf. Ser*. **1591** 012029.

[25] Wasi H A and Shiker M A K 2020 Proposed CG method to solve unconstrained optimization problems, "in press", accepted paper for publication in Journal of Physics: Conference Series, International Conference of Modern Applications on Information and Communication Technology (ICMAICT) - Iraq.

[26] Shiker M A K and Amini K 2018 A new projection-based algorithm for solving a large scale nonlinear system of monotone equations, *Croatian operational research review*, **9**: 1, p. 63- 73.

[27] Mahdi M M and Shiker M A K 2020 Solving systems of nonlinear monotone equations by using a new projection approach, “in press”, accepted paper for publication in Journal of Physics: Conference Series, International Conference of Modern Applications on Information and Communication Technology (ICMAICT) - Iraq.

[28] Dwail H H, Mahdi M M, Wasi H A, Hashim K H, Dreeb N K, Hussein A H and Shiker M A K 2020 A new modified TR algorithm with adaptive radius to solve a nonlinear systems of equations, “in press”, accepted paper for publication in Journal of Physics: Conference Series, International Conference of Modern Applications on Information and Communication Technology (ICMAICT) - Iraq.

[29] Hashim K H and Shiker M A K 2020 Using a new line search method with gradient direction to solve nonlinear systems of equations, “in press”, accepted paper for publication in Journal of Physics: Conference Series, International Conference of Modern Applications on Information and Communication Technology (ICMAICT) - Iraq.

[30] Wasi H A and Shiker M A K 2020 A modified of FR method to solve unconstrained optimization, “in press”, accepted paper for publication in Journal of Physics: Conference Series, International Conference of Modern Applications on Information and Communication Technology (ICMAICT) - Iraq.

[31] Consul P C and Famoye F 1992 Generalized Poisson regression model, *Communications in Statistics – Theory and Methods*, **21**, P. 89-109.

[32] Oliveira, M and Others 2016 Zero-inflated regression models for radiation-induced chromosome aberration data: *A comparative study, Biometrical Journal*, **58**: 2, P. 259–279.