

Zero-Truncated Models applied to the Nigerian National Health Insurance Data

Bayo H. Lawal

*LOYAB Statistical Consulting & Data Analytics, Adeleke Road,
Lamodi, Offa, Kwara State.*

(Received: 03 November 2021; Accepted: 21 April 2022)

Abstract. In this paper we have applied zero-truncated univariate and bivariate distributions to the NHIS data with two outcomes variables that are strongly over dispersed. Because of strong correlation between the outcome variables, we have fitted bivariate Poisson model that encompass covariance between the two variables. In the univariate case, we found that the Type II zero truncated generalized Poisson will be most suitable to each of the outcome variables. Other models also equally perform well, but the type II zero-truncated generalized Poisson (ZTGP2) is most preferable because of the ease of its implementation. For the bivariate models, we recognize that the zero-truncated Marshall & Olkin (1985) bivariate NB model does not perform well. Our preferred model would be the version of zero-truncated bivariate Poisson model proposed in Holgate(1964) and recently represented in AlMuhayfith *et al.* (2016). Our results indicate that this model is most suitable. Further it captures the covariance between the two outcome variables. All the models are implemented in SAS PROC NLMIXED. For each distribution considered, MLE estimation based on the log-likelihood functions are obtained using the Adaptive Gaussian Quadrature (usually with 32 q-points) and then optimized by using the Newton-Raphson algorithm. Starting values are obtained from those obtained from employing the Poisson or Negative binomial models.

Keywords: Bivariate Poisson, Overdispersion, Quasi negative binomial, Zero-truncated Bivariate Poisson, SAS PROC NLMIXED

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

Adesina *et al.* (2021) re-analyzed the The National Health Insurance Scheme (NHIS) data that is fully described in Mendeley Data web site, <https://data.mendeley/z7wznk53cf/8>. The data, obtained from health facilities in Ota, Ogun State, Nigeria has 1647 patients. The response variables of interest are Y_1 -th number of encounter visits to doctors and Y_2 -the number of diagno-

sis a patient had for the period of encounter. The predictors in the data set are: the covariates: Eclass-class of admission (in patient=1, outpatient=0), follow-up (follow-up=1, no follow-up=0), Sex(male=1, female=0), number of diagnosis (Y_2) which represents the number of diagnosis a patient had for the period of encounter and age of patient.

The predictors here are:

- eclass- nature of admission (1=for in patients, 0=for outpatients)
- sex- gender of patients (male=1, female=0)
- sge- age of patients
- Followup- (followup=1, no-follow-up=0)-designated here as fup.

The variable Y_1 has $\bar{y}_1 = 3.3892$; $s_1^2 = 11.5987$, giving a dispersion index of $3.4223 > 1$ thus indicating strong over-dispersion.

Further, Y_1 has the range $[1, 27]$, thus, it is truncated at $Y = 0$. The variable Y_2 -the number of diagnosis a patient had during the period also has, $\bar{y}_2 = 2.60777$; $s_2^2 = 3.7586$, giving a dispersion index of $1.4413 > 1$. This therefore also indicates over-dispersion though not as strong as that of variable Y_1 . Again the range of Y_2 is $[1, 15]$, also indicating truncation at $Y = 0$.

The question arises-why bother about truncation for this data set. For outcome variable Y_1 , the expected number of zeros under the Poisson model is $1647 \times e^{-3.3892} = 56$ observations out of the 1647 observations in the data which is substantial. On the other hand, for the outcome variable Y_2 , the expected number of zeros is $1647 \times e^{-2.60777} = 122$ observations which is quite substantial. Thus, we need to account for the non-zero occurrences in the two outcome variables by implementing zero-truncated models to the data. The two outcome variables have a sample correlation coefficient $r_{y_1 y_2} = 0.8849$ which is very high. Thus this can not be ignored in our model implementation.

The zero-truncation, when ignored, might lead to the over-dispersion that we observed. Various models have been suggested for modeling zero-truncated data. In the first section of this paper,

- we shall consider zero-truncated distributions (namely, Poisson, Negative binomial, Type 2 Generalized Poisson, Quasi-negative binomial, IT) to the frequency count data, Y_1 and Y_2 separately, ignoring the covariates-to see what distribution better explains the variations in the response variables.
- Apply the zero-truncated models to the full data with the covariates
- The expected means and variances under these models are computationally obtained as described in Lawal (2017, 2018 & 2021).
- Truncated bivariate Poisson model which assumes independence between the outcome variables (Chowdhury & Islam, 2016) and a truncated version of the Famoye (2010) bivariate Poisson model that assumes presence of covariance among the two variables will be employed and a corresponding bivariate negative binomial model based on the Marshall & Olkin (1985) parameterization.

2. Materials and Method

We present below the model formulations of the zero-truncated (ZT) models employed in this paper. For a random variable Y with a discrete distribution, where the value of $Y = 0$ can not be observed, then the zero-truncated random variable Y_t has the probability mass function

$$\Pr(Y_t = y) = \frac{\Pr(Y = y)}{\Pr(Y > 0)}, \quad y = 1, 2, 3, \dots \quad (1)$$

For the zero-truncated Poisson, with parameter μ , $\Pr(Y > 0) = 1 - \Pr(Y = 0) = 1 - \exp(-\mu)$. Hence, the pmf of zero-truncated Poisson random variable Y_t becomes

$$\Pr(Y_t = y) = \frac{\exp(-\mu)\mu^y}{y![1 - \exp(-\mu)]}, \quad y = 1, 2, 3, \dots, \quad (2)$$

Note that the mean of zero-truncated Poisson model is not equal to μ . Suppose the parameter μ in (2) is replaced by $\mu_i = \exp(\beta_i)$, the log-likelihood for the zero-truncated Poisson model in a sample of size n is

$$\log(L) = \sum_{i=1}^n (-\mu_i + y_i \log \mu_i - \log y_i! - \log[1 - \exp(-\mu_i)]).$$

The probability mass function for the zero-truncated negative binomial model is given by

$$\Pr(Y_i = y_i \mid y_i > 0) = \frac{f(y_i; \mu_i, \alpha)}{1 - (1 + \alpha\mu_i)^{-\alpha-1}}, \quad y = 1, 2, 3, \dots, \quad (3)$$

In this section, we will examine the corresponding zero truncated regression models for each of the count models discussed in the preceding section. The zero-truncated regression models arise in those situations where there is no zero by nature of the data. An example for instance, is the length of stay at an hospital. Once you are admitted, it is deemed that you have spent at least one day in the hospital (if number of days is the metric or outcome variable). Thus the zeros can not be observed in this for all patients admitted into the hospital. For a discrete distribution, and for a random variable Y , where the value of $Y = 0$ can not be observed, then, the zero truncated random variable Y_T has the probability density function:

$$\Pr(Y_T = y) = \frac{\Pr(Y = y)}{\Pr(Y > 0)}, \quad y = 1, 2, \dots, \quad (4)$$

Zhao *et al.* (2010) has employed the zero-truncated generalized Poisson using score tests, while, Lawal (2011) has implemented the ZTP, ZTNB, and ZTGP to the *medpar* data. Applications of zero-truncated models abound in the literature.

2.1 Zero-Truncated Poisson Models-ZTP

For the zero-truncated Poisson model therefore if we apply the expression in (4), the truncated Poisson with parameter μ , becomes:

$$\begin{aligned}\Pr(Y = 0) &= \exp(-\mu) \\ \Pr(Y > 0) &= 1 - \Pr(Y = 0) = 1 - \exp(-\mu)\end{aligned}$$

Thus the conditional probability of observing Y events given that $y > 0$, that is, the pdf of a zero-truncated Poisson random variable Y_T is given by:

$$\Pr(Y_T = y \mid y > 0) = \frac{\Pr(Y = y)}{\Pr(Y > 0)} = \frac{\exp(-\mu) \mu^y}{y! [1 - \exp(-\mu)]}, \quad y = 1, 2, 3, \dots, \quad (5)$$

Here,

$$\begin{aligned}E(Y_T \mid y > 0) &= \frac{\mu}{1 - \exp(-\mu)} \\ \text{Var}(Y_T \mid y > 0) &= \frac{\mu}{1 - \exp(-\mu)} \left[1 - \frac{\mu \exp(-\mu)}{1 - \exp(-\mu)} \right]\end{aligned} \quad (6)$$

For the zero-truncated Poisson distribution described in (6), the log-likelihood for a single observation i is:

$$LL = -\mu + y \log \mu - \log y! - \log[1 - \exp(-\mu)] \quad (7)$$

2.2 The Zero-truncated Negative Binomial Model-ZTNB:

The negative binomial distribution has the pdf,

$$\Pr(\mu_i, k, y_i) = \frac{\Gamma(y_i + \frac{1}{k})}{\Gamma(y_i + 1) \Gamma(\frac{1}{k})} \left(\frac{1}{1 + k\mu_i} \right)^{1/k} \left(\frac{k\mu_i}{1 + k\mu_i} \right)^{y_i}, \quad y_i = 0, 1, \dots \quad (8)$$

Here, the dispersion parameter $k > 0$. Thus, for the ZT, we have from (8),

$$\begin{aligned}\Pr(Y_T = 0) &= \left(\frac{1}{1 + k\mu_i} \right)^{1/k} \\ \Pr(Y_T > 0) &= 1 - \left(\frac{1}{1 + k\mu_i} \right)^{1/k} = 1 - (1 + k\mu_i)^{-1/k}\end{aligned}$$

and hence, the zero-truncated negative binomial pdf becomes:

$$\Pr(Y_T | y_i > 0) = \frac{\frac{\Gamma(y_i + \frac{1}{k})}{\Gamma(y_i + 1) \Gamma(\frac{1}{k})} \left(\frac{1}{1 + k\mu_i}\right)^{1/k} \left(\frac{k\mu_i}{1 + k\mu_i}\right)^{y_i}}{1 - (1 + k\mu_i)^{-1/k}} \quad (9)$$

The contribution of the i_{th} observation to the log-likelihood function for the truncated negative binomial using (9) is:

$$\begin{aligned} LL = & \log[\Gamma(y_i + \frac{1}{k})] + \frac{1}{k} \log\left(\frac{1}{1 + k\mu_i}\right) + y_i \log\left(\frac{k\mu_i}{1 + k\mu_i}\right) \\ & - \log[\Gamma(y_i + 1)] - \log[\Gamma(\frac{1}{k})] - \log[1 - (1 + k\mu_i)^{-1/k}] \end{aligned} \quad (10)$$

Its mean and variance are given by:

$$E(Y_T) = \frac{\mu_i}{1 - (k\mu_i + 1)^{-1/k}} \quad (11a)$$

$$\text{Var}(Y_T) = (1 + k\mu_i + \mu_i)E(Y_T) - [E(Y_T)]^2 \quad (11b)$$

2.3 Zero-truncated Generalized Poisson Distribution-ZTGP:

The type I generalized Poisson regression (GP1) model has the following pdf:

$$\Pr(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \frac{(1 + \alpha y_i)^{y_i-1}}{y_i!} \exp\left\{-\frac{\mu_i(1 + \alpha y_i)}{(1 + \alpha\mu_i)}\right\}, \quad y_i = 0, 1, \dots \quad (12)$$

with mean

$$E(Y_i) = \mu_i; \quad \text{and} \quad \text{Var}(Y_i) = \mu_i(1 + \alpha\mu_i)^2. \quad (13)$$

Consul and Famoye (1989) have also considered the GPI model for over-dispersed data because like the NB model, the GP also has a dispersion parameter α . The GP reduces to the Poisson when $\alpha = 0$.

For zero-truncated version, employing the expression for the GP distribution in (12), we have for when $Y = 0$,

$$\Pr(Y = 0) = \exp \left(-\frac{\mu_i}{(1 + \alpha\mu_i)} \right)$$

$$\Pr(Y > 0) = 1 - \exp \left(-\frac{\mu_i}{(1 + \alpha\mu_i)} \right)$$

and hence, the zero-truncated generalized Poisson pdf becomes:
Consequently,

$$\Pr(Y | y_i > 0) = \frac{\left(\frac{\mu_i}{1 + \alpha\mu_i} \right)^{y_i} \frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!} \exp \left(-\frac{\mu_i(1 + \alpha y_i)}{(1 + \alpha\mu_i)} \right)}{1 - \exp \left[-\frac{\mu_i}{(1 + \alpha\mu_i)} \right]} \quad (14)$$

Its corresponding log-likelihood function for a single observation having the generalized Poisson (GP) model is also given by:

$$\begin{aligned} \text{LL} = & y_i \log \left(\frac{\mu_i}{1 + \alpha\mu_i} \right) + (y_i - 1) \log(1 + \alpha y_i) - \frac{\mu_i(1 + \alpha y_i)}{1 + \alpha\mu_i} - \log(y_i!) \\ & - \log \left\{ 1 - \exp \left[\frac{-\mu_i}{(1 + \alpha\mu_i)} \right] \right\} \end{aligned} \quad (15)$$

2.4 Zero-Truncated Quasi-Negative Binomial Distribution-ZTQNB

The quasi-negative binomial distribution recently employed in Li *et al.* (2011) has the pmf given by:

$$P(Y = y) = \begin{cases} \frac{\Gamma(y + \alpha)}{y! \Gamma(\alpha)} \left(\frac{1}{1 + cy} \right) \left(\frac{1 + cy}{1 + b + cy} \right)^y \left(\frac{b}{1 + b + cy} \right)^\alpha, & y = 0, 1, \dots \\ & \text{for } y > m \text{ if } c < 0 \end{cases} \quad (16)$$

Lawal (2017) has employed an alternative formulation of the QNB model. From (16) therefore, the corresponding zero-truncated model therefore has the form:

$$\Pr(Y_t = y) = \frac{\Gamma(y + \alpha) \left(\frac{1}{1 + cy} \right) \left(\frac{1 + cy}{1 + b + cy} \right)^y \left(\frac{b}{1 + b + cy} \right)^\alpha}{y! \Gamma(\alpha) \left[1 - \left(\frac{b}{1 + b} \right)^\alpha \right]} \quad (17)$$

and the corresponding log-likelihood for a single observation becomes:

$$\begin{aligned} \text{LL4} = & \log \Gamma(y + \alpha) - \log(y!) - \log[\Gamma(\alpha)] + \log \left(\frac{1}{1 + cy} \right) + y \log \left(\frac{1 + cy}{1 + b + cy} \right) \\ & + \alpha \log \left(\frac{b}{1 + b + cy} \right) - \log \left[1 - \left(\frac{b}{1 + b} \right)^\alpha \right] \end{aligned} \quad (18)$$

2.5 The Inverse Trinomial Distribution-IT

The IT distribution, (Shimizu and Yanagimoto, 1991), which is derived from the Lagrangian expression has the pmf of the form

$$P(Y = y) = \frac{\lambda p^\lambda q^y}{y + \lambda} \sum_{t=0}^{\lfloor y/2 \rfloor} \frac{(y + \lambda)!}{t!(t + \lambda)!(y - 2t)!} \times \left(\frac{pr}{q^2} \right)^t, \quad (19)$$

$y = 0, 1, \dots; \lambda > 0, p \geq r$ and $p + q + r = 1$. It is so named because its cumulant generating function is the inverse of that for the trinomial distribution, Yanagimoto (1989). The IT model was employed for overdispersed medical count data in Phang and Loh (2014). A zero-truncated application of the model was also proposed in Phang and Ong (2006), while Phang *et al.* (2013) observed that the IT distribution presents “a stochastic formulation as a classical one dimensional random walk distribution and is another example of a distribution in the Takac family (Letac and Mora, 1990) with a cubic variance function of the mean”. The IT is modeled as a three-parameter distribution (λ, p, r) . Its corresponding zero-truncated pmf is given by:

$$P(Y_T = y) = \frac{\lambda p^\lambda q^y}{(y + \lambda)(1 - p^\lambda)} \sum_{t=0}^{\lfloor y/2 \rfloor} \frac{(y + \lambda)!}{t!(t + \lambda)!(y - 2t)!} \cdot \left(\frac{pr}{q^2} \right)^t, \quad y = 1, 2, \dots \quad (20)$$

since $P(Y = 0) = p^\lambda$. Hence, its log-likelihood is:

$$LL = \log(\lambda) + \lambda \log(p) + y \log(q) - \log(y + \lambda) + \log Q(y, \lambda) - \log(1 - p^\lambda). \quad (21)$$

where

$$Q(y, \lambda) = \sum_{t=0}^{\lfloor y/2 \rfloor} \frac{(y + \lambda)!}{t!(t + \lambda)!(y - 2t)!} \cdot \left(\frac{pr}{q^2} \right)^t$$

2.6 The Negative Binomial-Generalized Exponential Distribution NB-GE

The NB-GE distribution with parameters r, α, β is a mixture of the NB and the generalized exponential distributions, viz:

$$Y|\pi \sim NB(r, \pi = \exp(-\lambda)), \quad \text{and} \quad \lambda \sim GE(\alpha, \beta)$$

with the resulting unconditional pmf being given by:

$$f(y; r, \alpha, \beta) = \binom{r+y-1}{y} \sum_{j=0}^y (-1)^j \binom{y}{j} \left[\frac{\Gamma(\alpha+1)\Gamma(1+\frac{r+j}{\beta})}{\Gamma(\alpha+\frac{r+j}{\beta}+1)} \right] \quad (22)$$

for $y = 0, 1, \dots$, and $r, \alpha, \beta > 0$.

The means and variances of the NB-GE distribution in (22) are provided in Aryuyuen & Bodhisuwan (2013).

The NBGE has been applied to both over-dispersed and under-dispersed count data successfully (see Aryuyuen & Bodhisuwan, 2013; Lawal, 2018) and the zero-truncated pmf for the NBGE is therefore given in (23) as:

$$P(Y_t = y|r, \alpha, \beta) = \binom{r+y-1}{y} \cdot \left(\frac{1}{G}\right) \sum_{j=0}^y (-1)^j \binom{y}{j} \left[\frac{\Gamma(\alpha+1)\Gamma(1+\frac{r+j}{\beta})}{\Gamma(\alpha+\frac{r+j}{\beta}+1)} \right] \quad (23)$$

where, $y = 1, 2, \dots$ and G in (23) is defined as:

$$G = 1 - \left[\frac{\Gamma(\alpha+1)\Gamma(1+\frac{r}{\beta})}{\Gamma(\alpha+\frac{r}{\beta}+1)} \right]$$

The corresponding log-likelihood is:

$$\begin{aligned} LL = \log[(y_i + r - 1)] + \log \left[\sum_{j=0}^y (-1)^j \binom{y}{j} \left(\frac{\Gamma(\alpha+1)\Gamma(1+\frac{r+j}{\beta})}{\Gamma(\alpha+\frac{r+j}{\beta}+1)} \right) \right] \\ - \log y_i! - \log[(r-1)!] - \log \left[1 - \left(\frac{\Gamma(\alpha+1)\Gamma(1+\frac{r}{\beta})}{\Gamma(\alpha+\frac{r}{\beta}+1)} \right) \right] \end{aligned} \quad (24)$$

2.7 ZTCOM-NB

We also present here the zero-truncated model for the class of extended Com-Poisson models (Chakraborty et al.). We particularly focus on the COM-Negative binomial distribution which has the following pmf in (25) with parameters (ν, p, α) :

$$f(y; \nu, p, \alpha) = \frac{(\nu)_y p^y}{(y!)^\alpha {}_1H_{\alpha-1}(\nu, 1, p)} = \frac{\Gamma(\nu + y)}{\Gamma(\nu) {}_1H_{\alpha-1}(\nu, 1, p)} \cdot \frac{p^y}{(y!)^\alpha}; \quad y = 0, 1, 2, \dots \quad (25)$$

where

$${}_1H(\nu, 1, p) = \sum_{k=0}^{\infty} \frac{(\nu)_k p^k}{(k!)^\alpha} = \sum_{k=0}^{\infty} \frac{\Gamma(k + \nu) p^k}{\Gamma(\nu) (k!)^\alpha}$$

and the distribution is defined in the parameter space

$$\Theta_{COM-NB} = \{\nu > 0, p > 0, \alpha > 1\} \cup \{\nu > 0, 0 < p < 1, \alpha = 1\}$$

Hence, its zero-truncated pmf has:

$$f_{ZT}(y; \nu, p, \alpha) = \frac{\Gamma(\nu + y)}{\Gamma(\nu) (H - 1)} \cdot \frac{p^y}{(y!)^\alpha}; \quad y = 1, 2, \dots \quad (26)$$

Its corresponding log-likelihood is therefore given by:

$$LL = \log[\Gamma(\nu + y_i)] + y_i \log(p) - \log[\Gamma(\nu)] - \log(H - 1) - \alpha \log(y_i!) \quad (27)$$

where H is as defined above.

2.8 Estimation

With log likelihoods of a single observation i from **ZIP**, **ZTNB**, **ZTGP**, **ZTQNBD**, **ZTIT**, **ZTNBGE** and **ZTCOM-NB**, given in expressions (7), (10), (15), (18), (21), (24) and (17) respectively.

Maximum-likelihood estimations from these log-likelihoods are carried out with PROC NLMIXED in SAS, which minimizes the function $-LL(y, \Theta)$ over the parameter space Θ numerically. Our choice optimization algorithm here is the Newton-Raphson techniques. Convergence is often a major problem here and the choice of starting values is very crucial. By carefully setting initial values in grid format, the initial values can be obtained very easily by continually adjusting the initial parameters.

One common feature of the distributions described above, and indeed for most distributions employed for count regression models is that they all have infinite range. Consequently for a real life data that takes values $Y = 0, \dots, m$, it is most common to observe that the expected probabilities under any of the above models are not necessarily summing to 1.00 within the range $0 \leq Y \leq m$ as expected for a probability mass function, and consequently, the expected values will also not sum to n , the sample size. Lawal (2017, 2019) has provided a remedy for this anomaly, and we are not focusing on this in this paper.

3. Application:

In Tables 1 and 2 are the results of implementing the one parameter (ZTP), some two parameter zero truncated distributions (ZTNB, ZTGP) and some three parameter zero-truncated distributions (ZTNBD, ZTIT, ZTNBGE and ZTCOMNB) to the frequency counts of Y_1 and Y_2 respectively of the NHIS Data. Except for the ZTP, all the other six probability models do not have their estimated cumulative probabilities summing to 1.00 within the observed range [1,27] of the outcome variable Y_1 and [1,15] for Y_2 . This is characteristic of all discrete distributions, see Lawal (2017, 2018, 2019). While expressions exist for the means and variances of the ZTP, ZTNB and ZTGP, the other distributions do not have such and we have therefore obtained the means and variances of all distributions employed here computationally using the method of moments as discussed in Lawal (2017,2019). Suffice to say that results obtained from these computational procedures agree with those of the ZTP, ZTNB and ZTGP.

We observed immediately, why the ZTP does not fit our data. The variance of the outcome variable Y_1 is grossly under estimated, viz, $2.9478 \ll 11.5987$. The other distributions on the other hand provide estimated variances and means that are very close to the observed values in the response variable. Of the other six models, two models stand out clearly better than the others. These are the ZTGP and the ZTIT. Both provide lower -2LL as well as lower Wald's test statistic

$$X^2 = \sum_{i=1}^{1647} \frac{(y_{1i} - \hat{\mu})^2}{\hat{\sigma}^2}, \text{ The difference between the } X^2 \text{ for the two distributions}$$

is the fact that ZTIT estimates the variance much higher than the observed variance in the data and does lower the Wald's Test statistic. The ZTGP on the other hand produces a matched mean with the observed mean and a much closer variance to the observed variance.

Table 1: ZT models Applied to frequency counts of variable Y_1

Parm.	ZTP	ZTNB	ZTGP	ZTQNB	ZTIT	ZTNBGE	ZTComNB
	$\hat{b}_0=1.1814$	$\hat{b}_0=0.4561$	$\hat{b}_0=0.7316$	$\hat{\alpha}=0.5271$	$\hat{\lambda}=1.0019$	$\hat{\alpha}=3541.41$	$\hat{\alpha}=0.9336$
		$\hat{k}=2.6021$	$\hat{\tau}=0.5715$	$\hat{b}=0.3126$	$\hat{p}=0.3867$	$\hat{\beta}=5.8707$	$\hat{\nu}=0.6669$
				$\hat{c}=0.0123$	$\hat{r}=0.0617$	$\hat{r}=0.4933$	$\hat{p}=0.6671$
-2LL	8577.5	6716.5	6712.8	6715.1	6712.6	6715.6	6714.2
AIC	8579.5	6720.5	6716.8	6721.1	6718.6	6721.6	6720.2
X^2	6476.5867	1709.7236	1620.5949	1610.1446	1609.6312	1615.4079	1615.2553
G^2	4091.6430	4091.6446	4091.6440	4088.2381	4091.6444	4099.1970	4091.7401
d.f.	1646	1645	1645	1644	1644	1644	1644
$\hat{\mu}$	3.3892	3.3892	3.3892	3.3902	3.3892	3.3869	3.3892
$\hat{\sigma}^2$	2.9478	11.1664	11.7806	11.8570	11.8608	11.8184	11.8195

Similarly, from Table 2 however, the most parsimonious models for outcome variable Y_2 are the ZTQNB and ZTCOMNB. Both produce means and variances that are much closer to the observed moments of Y_2 . Both models do not have close form estimates of means and variances, thus, these are computationally obtained using the method of moments. We observe that for the ZTQNB, $\hat{c} = -0.01574$, thus, the moments exist only in the range $0 \leq Y_2 \leq \text{int}(-1/\hat{c})$ which is 63 in this case. The ZTNBGE model overestimates the mean and variance here.

Table 2: ZT models Applied to frequency counts of variable Y_2

Parm.	ZTP	ZTNB	ZTGP	ZTQNBD	ZTIT	ZTNBGE	ZTComNB
	$\hat{b}_0=0.8595$	$\hat{b}_0=0.5654$ $\hat{k}=0.7465$	$\hat{b}_0=0.6153$ $\hat{\tau}=0.2685$	$\hat{\alpha}=1.0332$ $\hat{b}=0.5850$ $\hat{c}=-0.0157$	$\hat{\lambda}=3.2801$ $\hat{p}=0.7023$ $\hat{r}=0.0553$	$\hat{\alpha}=4404.33$ $\hat{\beta}=11.7770$ $\hat{r}=1.5626$	$\hat{\alpha}=1.1942$ $\hat{\nu}=0.5843$ $\hat{p}=0.9220$
-2LL	6090.7	5714.3	5716.2	5713.4	5715.7	5715.8	5713.2
AIC	6092.7	5718.3	5720.2	5719.4	5721.7	5721.8	5719.2
X^2	3145.2316	1617.7790	1600.0258	1644.5274	1607.4468	1593.7658	1648.8278
G^2			1991.7980	-	1991.7981	1988.56	1991.7989
d.f.	1646	1645	1645	1644	1644	1644	1644
$\hat{\mu}$	2.6078	2.6078	2.6078	2.6077	2.6078	2.6088	2.6078
$\hat{\sigma}^2$	1.9670	3.8241	3.8666	3.7619	3.8487	3.8818	3.7521

3.1 ZT Models with Explanatory Variables-GLM

In this section, we have the applied the ZT models discussed in the previous sections to the full data with four explanatory variables: sex, age, fup and ecs. Here, both ZTP and ZTNB are modeled with $\mu_i = \exp(a_0 + a_1 + a_2 + a_3 + a_4)$. The zero-truncated generalized Poisson employed here is the GP type II, proposed in Consul and Famoye (1992) and Consul (1989) for implementing GLM generalized Poisson model and has the pmf given by:

$$f_{ZT}(y_i; \theta_i, \delta) = \frac{\theta_i(\theta_i + \delta y_i)^{y_i-1} e^{-\theta_i - \delta y_i}}{y_i!(1 - e^{-\theta_i})} \quad (28)$$

with $y_i = 1, \dots, \theta_i > 0, 0 \leq \delta < 1$. Expressions for the mean and variance of the un-truncated version are provided in Joe & Zhu (2005). Thus the distribution can be modeled in the form:

$$\log\left(\frac{\theta_i}{1 - \delta}\right) = \mathbf{x}'\beta \quad (29)$$

The log-likelihood for a single observation would therefore be given by:

$$LL = \log(\theta_i) + (y_i - 1) \log(\theta_i + \delta y_i) - (\theta_i + \delta y_i) - \log(y_i!) - \log(1 - e^{-\theta_i}) \quad (30)$$

Where,

$$\theta_i = \exp(\mathbf{x}'\beta + \text{offset}) \text{ and } \text{offset} = \log(1 - \delta).$$

This version of the generalized Poisson or its zero-truncated counterpart are the ones implemented SAS PROC HPFMM, STATA and R package *glmmTMB*. For the ZQNBD, its parameter b is modeled for the four explanatory variables as:

$$b = \exp(a_0 + a_1 + a_2 + a_3 + a_4) \quad (31)$$

Similarly, the λ , β and p parameters in the ZIIT, ZTNBGE and ZTCOMNB are modeled respectively in the form of b in expression (31). Results in Tables 3 and 4 are those from implementing the GLM (with explanatory variables) versions of all the zero-truncated models with outcome variables Y_1 and Y_2 respectively. From Table 3 it appears that models ZTGP2, ZTQNB and ZTCOMNB are suitable candidates for parsimony, with the ZTGP2 having one less parameter and much easier to model. It does not have the convergence problems the other two have in terms of not having suitable initial parameter estimates. Thus, the zero truncated generalized Poisson Type II model will be preferred here. The model has reduced the dispersion index of 3.9000 under the ZTP to 0.9844 under the ZTGP2. The ZTGP2 performs better than the ZTNB and both are computational easier than the other distributions.

Table 3: ZT models on Y_1 with covariates

Parm.	* Sig at 5%						
	ZTP	ZTNB	ZTGP2	ZTQNB	ZTIT	ZTNBGE	ZTCOMNB
Int	1.0766*	0.3267	0.5377*	1.0451*	-0.2076	1.4695	-0.4143
sex	-0.0126	-0.0144	0.0204	-0.0112	0.0213	0.0043	-0.0034
age	0.0032*	0.0043*	0.0058	0.0040*	0.0058	-0.0023	0.0008
fup	0.0726	0.1046	0.0470	0.0901	0.0460	-0.0424	0.0171
ecs	0.2146	0.3323	0.2312	0.2919	0.2300	-0.1423	0.0474
		$\hat{k}=2.5302$	$\hat{\delta}=0.5416$	$\hat{\alpha}=0.5211$	$\hat{p}=0.3812$	$\hat{\alpha}=100.03$	$\hat{\alpha}=0.9406$
				$\hat{c}=0.0107$	$\hat{r}=0.0591$	$\hat{r}=0.7021$	$\hat{\nu}=0.6471$
-2LL	8550.7	6709.3	6707.7	6708.1	6707.5	6710.4	6707.6
AIC	8560.7	6721.3	6719.7	6722.1	6721.5	6724.4	6721.6
X^2	6403.81	1699.84	1615.35	1615.90	1603.73	1635.27	1622.38
G^2	4068.56	4067.29	4072.11	4064.68	4072.17	4146.59	4070.46
d.f.	1642	1641	1641	1640	1640	1640	1640

The corresponding results in Table 4 also indicate that most of the models apart from the ZTP are suitable candidates. However, because of the ease of computation and interpretation, the ZTNB and the ZTGP2 would be highly recommended. In particular, the ZTGP2 is better. We note here that for ZTQNB, $\hat{c} = -0.01778$ which indicates that the moments exist for $1 \leq Y_2 \leq 56$, with 56 being the $\text{int}(-1/\hat{c})$.

Table 4: ZT models on Y_2 with covariates

Parm.	Sig at 5%						
	ZTP	ZTNB	ZTGP2	ZTQNB	ZTIT	ZTNBGE	ZTCOMNB
Int	0.7735	0.4680*	0.4877*	0.4528	1.0700*	2.3177*	-0.1114
sex	0.0021	0.0023	0.0132	0.0008	0.0137	-0.0029	0.0010
age	0.0025*	0.0029	0.0040*	0.0030	0.0043*	-0.0021	0.0013
fup	0.0575	0.0714	0.0167	0.0906	0.0205	-0.0397	0.0297
ecs	0.0827	0.1139	0.1053	0.1225	0.1110	-0.0767	0.0430
		$\hat{k}=0.7354*$	$\hat{\delta}=0.3303*$	$\hat{\alpha}=1.0119*$	$\hat{p}=0.7080*$	$\hat{\alpha}=400.01$	$\hat{\alpha}=1.2023*$
				$\hat{c} = -0.0178$	$\hat{r}=0.0559*$	$\hat{r}=1.7973*$	$\hat{\nu}=0.5720$
-2LL	6080.7	5709.2	5711.7	5708.0	5711.1	5712.3	5707.9
AIC	6090.7	5712.2	5723.7	5722.0	5725.1	5726.3	5721.9
X^2	3122.69	1611.92	1600.64	1640.09	1608.23	1586.5370	1644.05
G^2	2403.09	1984.22	1984.88	1984.00	1984.81	1984.99	1984.25
d.f.	1642	1641	1641	1640	1640	1640	1640

In Table 5 are presented the first and last five estimated means and variances under the ZTGP2 model for the outcome variables Y_1 and Y_2 . The average means and variances over the entire 1647 observations are very close to the observed

values in the outcome variables. We have also presented the ranges of these estimated moments.

Table 5: Estimated Means and Variances under th GLM-ZTGP2 Models

Obs	Outcome Y_1		Outcome Y_2	
	\bar{y}_1	$\hat{\sigma}_{y_1}^2$	\bar{y}_2	$\hat{\sigma}_{y_2}^2$
1	3.4253	11.9086	2.6170	3.8747
2	3.4884	12.2404	2.6576	3.9809
3	3.6447	13.0655	2.7532	4.2316
4	3.5146	12.3785	2.6739	4.0236
5	3.4543	12.0611	2.6363	3.9251
\vdots	\vdots	\vdots	\vdots	\vdots
1643	3.2862	11.1797	2.5460	3.6895
1644	3.5633	12.6352	2.7250	4.1574
1645	3.2132	10.7993	2.4974	3.5634
1646	3.2132	10.7993	2.4974	3.5634
1647	3.5361	12.4914	2.7079	4.1127
Average	3.1485	11.7215	2.6078	3.8512
Range	[3.1485,4.0558]	[10.4632,15.2527]	[2.4526,2.9250]	[3.4473,4.6841]

4. Bivariate Zero-truncated Models

The data in our study here is an example of count data exhibiting two response variables Y_1 and Y_2 , with several covariates (sex, age, followup, ecs). It has been suggested that the Bivariate Poisson (BVP) is the underlying model for these outcomes. Such variables may be correlated or independent. There are several examples of data exhibiting bivariate outcomes (e.g. traffic data often give rise to two outcomes, viz: number of traffic accidents and number of injuries or fatalities occurring during a specified period). Other examples have been provided in (Famoye, 2010a, 2010b).

The BVP was originally proposed by Holgate(1964) and further discussed in Johnson *et al.* (1997).. To formulate the BVP, let, the random variables X_1, X_2 and X_3 follow three independently distributed Poisson with parameters λ_1, λ_2 and λ_3 respectively, with $\lambda_k > 0, k = 1, 2, 3$. The random variables $Y_1 = X_1 + X_3$, and $Y_2 = X_2 + X_3$ is (employing the trivariate reduction method) jointly distributed as bivariate Poisson JBP($\lambda_1, \lambda_2, \lambda_3$) The bivariate joint probability distribution (AlMuhayfith *et al.*, 2016) is given by:

$$f_{JBP}(Y_1 = y_1, Y_2 = y_2) = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{i=0}^{\min(y_1, y_2)} \frac{\lambda_3^i}{i!} \frac{\lambda_1^{y_1-i}}{(y_1 - i)!} \frac{\lambda_2^{y_2-i}}{(y_2 - i)!} \quad (32)$$

We may note here that there have been several definitions of the bivariate Poisson (Famoye, 1999) and several approaches have been discussed in Kocherlakota and Kocherlakota (1992). However, we are adopting the definition defined in (32). It can be shown that Y_1 and Y_2 in (4.1) are marginally distributed as Poisson with means $\lambda_1 + \lambda_3$ and $\lambda_2 + \lambda_3$ respectively. The covariance between Y_1 and Y_2 is

$$\text{COV}(Y_1, Y_2) = \text{COV}(X_1 + X_3, X_2 + X_3) = \text{Var}(X_3) = \lambda_3, \quad (33)$$

and thus the correlation between Y_1 and Y_2 is given by:

$$\text{corr}(Y_1, Y_2) = \frac{\lambda_3}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}}, \quad \text{with } \lambda_3 > 0 \quad (34)$$

An alternative formulation for the BVP is provided by Famoye (2010), which, following Lakshminarayana *et al.* (1999) defines the probability function as a product of Poisson marginals with a multiplicative factor and has the form:

$$f_F(y_1, y_2) = \frac{\theta_1^{y_1} \theta_2^{y_2} e^{-(\theta_1 + \theta_2)} [1 + \lambda(e^{-y_1} - e^{-d\theta_1})(e^{-y_2} - e^{-d\theta_2})]}{y_1! y_2!}, \quad y_1, y_2 = 0, 1, 2, \dots \quad (35)$$

where $d = 1 - e^{-1}$. The covariance between Y_1 and Y_2 is $\lambda\theta_1\theta_2d^2e^{-d(\theta_1+\theta_2)}$ and the corresponding correlation coefficient $\rho = \lambda\sqrt{\theta_1\theta_2d^2e^{-d(\theta_1+\theta_2)}}$. Which implies that the correlation can be positive, zero or negative depending on the value of λ . However, since λ is modeled as e' , it is usually unlikely to be negative or zero.

5. Zero-Truncated BVP

Following Chou & Steenhard (2011), the zero truncated bivariate distribution takes the form:

$$ZT_{bvp} = \frac{f(y_1, y_2)}{\phi} \quad (36)$$

for $\phi = 1 - f(y_1 = 0) - f(y_2 = 0) + f(y_1 = 0, y_2 = 0)$. Thus, the log-likelihoods of the zero-truncated versions of the BVP models in (32) and (35) become:

$$\log LL = \log[f(y_1, y_2)] - \log \phi$$

For the model in (32) therefore, we have,

$$\begin{aligned} f(y_1 = 0) &= e^{-\frac{\lambda_2^{y_2}}{y_2!}} \\ f(y_2 = 0) &= e^{-\frac{\lambda_1^{y_1}}{y_1!}} \\ f(y_1 = 0, y_2 = 0) &= e^{-\lambda} \end{aligned} \quad (37)$$

where $\lambda = (\lambda_1 + \lambda_2 + \lambda_3)$.

Thus, the log-likelihood LL for the model becomes:

$$LL1 = - \sum_{k=1}^3 \lambda_k + \sum_{k=1}^2 [y_k \log(\lambda_k) - \log(y_k!)] + \log(\mathbf{Q}) - \log(\phi) \quad (38)$$

where

$$\mathbf{Q} = \sum_{j=0}^{\min(y_1, y_2)} \frac{\lambda_3^j}{j!} \frac{\lambda_1^{y_1-j}}{(y_1-j)!} \frac{\lambda_2^{y_2-j}}{(y_2-j)!}$$

5.1 ZTFamoye

For the Famoye BPV formulation in (35) we also have the following:

$$\begin{aligned} f(y_1 = 0) &= \frac{\theta_2^{y_2} e^{-(\theta_1 + \theta_2)} [1 + \lambda(1 - c_1)(e^{-y_2} - c_2)]}{y_2!}, \\ f(y_2 = 0) &= \frac{\theta_1^{y_1} e^{-(\theta_1 + \theta_2)} [1 + \lambda(e^{-y_1} - c_1)(1 - c_2)]}{y_1!}, \end{aligned} \quad (39)$$

$$f[(y_1, y_2) = 0] = e^{-(\theta_1 + \theta_2)} [1 + \lambda(1 - c_1)(1 - c_2)]$$

where $c_1 = e^{-d\theta_1}$, $c_2 = e^{-d\theta_2}$, and $d = 1 - e^{-1}$. The log-likelihood therefore becomes:

$$LL2 = \sum_{i=1}^n \left\{ \sum_{t=1}^2 [y_t \log(\theta_t) - \log(y_{it})] + \log[1 + \lambda(e^{-y_1} - c_1)(e^{-y_2} - c_2)] - \log(\phi) \right\} \quad (40)$$

While the ZT_{bvp} models describe by the log-likelihoods in (38) and (40), have a covariance structure between Y_1 and Y_2 , the ZT_{bvp} proposed by Chowdhury & Islam (2016) assumes independence between the two outcome variables. Their proposed model is presented in (41).

5.2 ZTBVP Model-Chowdhury & Islam

Chowdhury & Islam (2016) discussed the zero-truncated marginal bivariate Poisson distribution of the form:

$$f_{ztbp}(y_1, y_2) = \frac{e^{-(\lambda_1 + \lambda_2)} \lambda_1^{y_1} \lambda_2^{y_2}}{y_1! y_2! (1 - e^{-\lambda_1})(1 - e^{-\lambda_2})}, \quad (y_1, y_2) = 1, 2, \dots \quad (41)$$

while its conditional distribution is given by:

$$g_{zt}(y_1, y_2) = f_{zt}(y_2|y_1)f_{zt}(y_1) = \frac{(\lambda_2 y_1)^{y_1} (\lambda_1)^{y_1}}{y_1! y_2! (e^{\lambda_2 y_1} - 1)(e^{\lambda_1} - 1)} \quad (42)$$

where:

$$f_{zt}(y_1, y_2) = \frac{(\lambda_2 y_1)^{y_1}}{y_1! (e^{\lambda_2 y_1} - 1)}$$

with mean and variance of (42) being:

$$\begin{aligned} \mu_{Y_2|Y_1} &= E[Y_2|Y_1, Y_2 > 0] = \frac{\lambda_2 y_1 e^{\lambda_2 y_1}}{e^{\lambda_2 y_1} - 1}, \text{ and} \\ \sigma_{Y_2|Y_1}^2 &= \text{Var}[Y_2|Y_1, Y_2 > 0] = \frac{\lambda_2 y_1 e^{\lambda_2 y_1}}{e^{\lambda_2 y_1} - 1} \left[1 - \frac{\lambda_2 y_1}{e^{\lambda_2 y_1} - 1} \right] \end{aligned}$$

and $f_{zt}(y_1)$ in (42) being similarly defined as:

$$f_{zt}(y_1) = \frac{\lambda_1^{y_1}}{y_1! (e^{\lambda_1} - 1)}$$

with mean and variance

$$\begin{aligned} \mu_{Y_1} &= E[Y_1|Y_1 > 0] = \frac{\lambda_1 e^{\lambda_1}}{e^{\lambda_1} - 1}, \text{ and} \\ \sigma_{Y_1}^2 &= \text{Var}[Y_1|Y_1 > 0] = \frac{\lambda_1 e^{\lambda_1}}{e^{\lambda_1} - 1} \left[1 - \frac{\lambda_1}{e^{\lambda_1} - 1} \right] \end{aligned}$$

Chowdhury & Ismail (2016) proposed the model in (42) and have followed it up in Chowdhury & Ismail (2019) with and R package *bpglm*. However the marginal formulation of the ZTBP in (41) and its corresponding conditional distribution in (42) and as implemented in the package *bpglm* assumes that Y_1 and Y_2 are independently distributed and hence has $\text{Cov}(Y_1, Y_2) = 0$ which may not necessarily be the case.

We shall implement both the zero-truncated models of f_{JBP} and the Famoye $f_{BP}(y_1, y_2)$ as well as the marginal and conditional versions proposed by Chowdhury & Ismail using SAS PROC NLMIXED. The former two provide us with the covariances of the two variables while the Chowdhury models assume independence. We shall compare the results from SAS PROC NLMIXED with those from the application of the package *bpglm*.

5.3 Zero-Truncated Bivariate Negative Binomial

We shall fit the zero-truncated Marshall & Olkin (1985) parameterization of the bivariate negative binomial. The Marshall-Olkin parameterization of the negative binomial has the probability mass function given by

$$f(y_1, y_2 | \lambda_1, \lambda_2, \alpha) = \frac{\Gamma(y_1 + y_2 + \alpha)}{y_1! y_2! \Gamma(\alpha)} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2 + 1} \right)^{y_1} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2 + 1} \right)^{y_2} \left(\frac{1}{\lambda_1 + \lambda_2 + 1} \right)^\alpha \quad (43)$$

where λ_1, λ_2 are the two marginal means and α is the common over dispersion parameter. Again, for the zero truncated version of the model,

$$\begin{aligned} g(y_1 = 0) &= \frac{\Gamma(y_2 + \alpha)}{y_2! \Gamma(\alpha)} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2 + 1} \right)^{y_2} \left(\frac{1}{\lambda_1 + \lambda_2 + 1} \right)^\alpha \\ g(y_2 = 0) &= \frac{\Gamma(y_1 + \alpha)}{y_1! \Gamma(\alpha)} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2 + 1} \right)^{y_1} \left(\frac{1}{\lambda_1 + \lambda_2 + 1} \right)^\alpha \\ g(y_1 = 0, y_2 = 0) &= \left(\frac{1}{\lambda_1 + \lambda_2 + 1} \right)^\alpha \end{aligned} \quad (44)$$

Thus leading to the ZT model:

$$f_{ZT}(y_1, y_2 | \lambda_1, \lambda_2, \alpha) = \frac{\Gamma(y_1 + y_2 + \alpha)}{y_1! y_2! \Gamma(\alpha)} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2 + 1} \right)^{y_1} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2 + 1} \right)^{y_2} \left(\frac{1}{\lambda_1 + \lambda_2 + 1} \right)^\alpha - \phi \quad (45)$$

for $(y_1, y_2) = 1, 2, \dots$ and $\phi = 1 - g(y_1 = 0) - g(y_2 = 0) + g(y_1 = 0, y_2 = 0)$.

6. Application to the NHIS Data

We have earlier established the two outcomes are zero truncated. Further, the two outcome variables have a sample correlation coefficient $r_{y_1 y_2} = 0.8849$ which is very high. Thus this can not be ignored in our model implementations. To implement the f_{ztJBP} model, the λ s are modeled as:

$$\begin{aligned} \lambda_1 &= \exp(a_0 + a_1 + a_2 + a_3 + a_4) \\ \lambda_2 &= \exp(b_0 + b_1 + b_2 + b_3 + b_4) \\ \lambda_3 &= \exp(c_0) \end{aligned} \quad (46)$$

We observe here that λ_3 is modeled as constant. If so desired, we could model it as a function of one or any number of the covariates. Our preliminary analysis indicates that the constant covariate is most suitable.

6.1 Results

In Table 6 are the results of implementing the zero-truncated bivariate Poisson and negative binomial to the NHIS data. based on these results, we would prefer either the ZTBJP or the ZTFamoye model. Of the two however, we prefer the ZTBJP because its correlation coefficient is much closer to the observed correlation coefficient between the outcome variables. The Famoye model underestimates this correlation coefficient and its convergence is a major issue. The -2LL in the two models can not be compared as they are based on different formulations. There is not much difference between the ZTBJP that utilizes a constant correlation parameter λ_3 or one that incorporates the predictors **fup** and **age** in its formulation-thus varying across the 1647 observations in the data. The Chowdhury & Ismail (2016) model ignores the dependence of the outcome variables and is therefore not suitable in this case. based on results from ZTBJP, we observe that:

- Predictors, sex, age, followup and eclass are all significant on outcome variable Y_1 -the number of doctor visits.
- Only predictors age and followup are significant on the Y_2 outcome variable-the number of diagnosis encounters
- the estimated average correlation coefficient is $\bar{\rho} = 0.7520$ and has a range of [0.5944, 0.7844] across the 1647 observations in the data set.

Table 6: MLE Estimates parameters for various ZT Bivariate Models
Sig at 5%

	ZTBJP		Chow	Famoye	Molkin
Parameter	Constant	Variable	Constant		
Int (Y_1)	1.0555	1.0555	1.0766	1.0740	0.0948
sex	-0.0436*	-0.0436*	-0.0126*	-0.0700*	-0.0096
age	0.0029*	0.0029*	0.0032*	0.0042	0.0027*
followup	0.1683*	0.1684*	0.0726	0.0900*	0.0648
eclass	0.3467*	0.3467*	0.2146	0.5700*	0.2038
Int (Y_2)	0.5404	0.5403	0.7735	0.9600	-0.1426
sex	-0.0136	-0.0136	0.0021	0.1500*	0.0025
age	0.0013*	0.0013*	0.0025*	-0.0008	0.0018
followup	0.1511*	0.1512*	0.0575	0.0370	0.0442
eclass	0.0499	0.0500	0.0827	0.1200	0.0702
	$\hat{\lambda}_3=1.8145^*$			$\hat{\lambda}=5.3122^*$	$\hat{\alpha}=2.8369^*$
-2LL	82,157	82,099	14,631	14,749	12,708
AIC	82,179	82,125	14,651	14,771	12,730
X^2	9870.7742	9871.6394	7926.2108	8224.0417	na
$\bar{\rho}$	0.7576	0.7520	0	0.5933	na

7. Conclusions:

We have demonstrated in this paper that the zero-truncated distributions when applied to each outcome separately produce similar results, with the type II generalized Poisson regression model being the most suitable in the univariate case. However, because the data has two outcomes, bivariate models applied to the data seem to be more appropriate here. The zero-truncated BJP model behaves well and clearly performs better than the parameterization presented in Famoye (2010b). All the models are implemented with SAS PROC NLMIXED. The zero truncated model proposed in Chowdhury and Islam has a corresponding R package which makes it easier to implement but is based on the assumption that both outcome variables are independent which is a rare occurrence in bivariate data. One major difference between the results here and those in Adesina *et al.* (2021) is that the response variable Y_1 is over-dispersed rather than being under-dispersed and some of the -2LL or AIC reported are not realizable.

References

- Adesina, O. S., Agunbiade, D. A. and Oguntunde, P. E. (2021). Flexible Bayesian Dirichlet mixtures of generalized linear mixed models for count data. *Scientific African*, 13, 1-11.
- AlMuhayfith, F. E., Alzaid, A. A. and Omair, M. A. (2016). On bivariate Poisson regression Models. *Journal of King Saudi University-Science*, 28, 178-189.
- Aryuyuen, S. and Bodhisuwan, W. (2013). Negative binomial-generalized exponential (NB-GE) distribution. *Applied Mathematical Sciences*, 7(22), 1093-1105.
- Chakraborty, S. and Imoto, T. (2016). Extended Conway-Maxwell-Poisson distribution and its properties and applications. *Journal of Statistical Distributions and Applications*, 3(5), 1-199. DOI:10.1186/s40488-016-0044-1
- Chakraborty, S. and Ong, S. H. (2016). A COM-type generalization of the negative binomial distribution. *Communications in Statistics-Theory and Methods*, 45, 4117-4135.
- Chou, N. and Steenhard, Z. (2011). SAS Global Forum 2011. Paper 355.
- Chowdhury, R.I. and Islam, M. A. (2016). Zero-truncated bivariate model: marginal-Conditional modelling Approach with an application to Traffic Accident Data. *Applied Mathematics*, 7, 1589-1598.
- Chowdhury, R. I. and Islam, M. A. (2019). *bpglm*: R package for bivariate Poisson GLM with Covariates.
- Consul, P.C. (1989). *Generalized Poisson Distributions; Properties and Applications*. Marcel Dekker, New York.
- Consul, P. C. and Famoye, F. (1992). Generalized Poisson Regression Model. *Communication in Statistics. Theory Methods*, 21(1), 89-109.
- Famoye, F. (2010). A new bivariate generalized Poisson distribution. *Stat. Neerl.*, 64(1), 112-124.
- Famoye, F. (2010). On the bivariate negative binomial regression model. *Journal of Applied Statistics*, 37(6), 969-981.
- Famoye, F. and Consul, P. C. (1995). Bivariate generalized Poisson distribution with some applications. *Metrika*, 42, 127-138.
- Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica*, 52, 701-720.
- Holgate, P. (1964). Estimation for the bivariate Poisson Distribution. *Biometrika*, 51, 241-245.
- Imoto, T. (2014). A generalized Conway-Maxwell-Poisson distribution which includes the negative binomial distribution. *Appl. Math. Comput.*, 247, 825-834.
- Janardan, K. G. (1975). Markov-Polya urn-model with predetermined strategies. *Gujarat Statist. Rev.*, 2(1), 17-32.

- Jang, R. C., and Winkelmann, R. (1993). Two aspects of labor mobility: a bivariate Poisson regression approach. *Empir. Econom.*, 18, 543-556.
- Joe, H., and Zhu, R. (2005). Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial distribution. *Biomet. J.*, 47, 219-229.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*, John Wiley and Sons, Inc., New York.
- Karlis, D. and Ntzoufras, I. (2005). Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software*, 14, 1-36.
- Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. Marcel Dekker, New York.
- Lakshminarayana J, Pandit, N. N. and Rao, K. S. (1999). On a bivariate Poisson distribution. *Commun. Stat. Theory*, 28(2), 267-276.
- Lawal, H. B. (2011). On zero-truncated generalized Poisson count regression model. *Journal of Theoretical and Applied Statistics*, 13-24.
- Lawal, H. B. (2017). On the negative binomial-generalized exponential distribution and its applications. *Applied Mathematical Sciences*, 11, 345-360. doi.org/10.12988/ams.2017.612288
- Lawal, H. B. (2018). Correcting for Non-Sum to 1 Estimated Probabilities in Applications of Discrete Probability Models to Count Data. *International Journal of Statistics and Probability*, 6, 119-131.
- Lawal, H. B. (2019). Quasi-Negative Binomial, Inverse Trinomial and Negative Binomial Generalized Exponential Distributions and their Applications. *Jour. Prob. Statist. Science*, 17(2), 205-221.
- Lawal, H. B. (2021). On Com-Negative Binomial Distributions. *Benin Journal of Statistics*, 4, 90-113.
- Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions. *Annals of Statistics*, 18, 1-37.
- Li, S., Yang, F., Famoye, F., Lee, C., and Black, D. (2011). Quasi-negative binomial distribution: Properties and applications. *Computational Statistics and Data Analysis*, 55(20), 2363-2371.
- Marshall, A. W. and Olkin, I. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution. *Jour. Amer. Statist. Assoc.*, 80, 332-338.
- Phang, Y. N. and Loh, E. F. (2014). Statistical Analysis for over-dispersed medical Count Data. *International Scholarly and Scientific Research*, 8(2), 292-294.
- Phang, Y. N., Sim, S. Z. and Ong, S. H. (2013). Statistical Analysis for the inverse Trinomial Distribution, *Communications in Statistics-Simulation and Computation*, 42(9), 2073-2085.
- Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Comp. and Graomegacal Statistics*, 4, 12-35.
- Sen, K. and Jain, R. (1996). Generalized Markov-Polya urn-model with pre-determined strategies. *Jour. Statist. Plann. Infer.*, 54, 119-133.
- Shimizu, K. and Yanagimoto, T. (1991). The inverse trinomial Distribution. *Japan Journal of Applied Statistics*, 20, 89-96.
- Vangala, P. (2015). Negative binomial-generalized exponential distribution: generalized linear model and its applications. M.S. thesis, Texas A & M University, College Station, TX, USA.
- Yanagimoto, T. (1989). The inverse binomial distribution as a statistical model. *Communications in Statistics theory and Methods*, 18(10), 3629-3633.
- Yip, K. C. H. and Yau, K. K. W. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, 36(2), 153-163.
- Zhao, W., Feng, Y. and Li, Z. (2010). Zero-truncated generalized Poisson regression model and its score tests. *Journal of East China Normal University*, 1, 17-23.