ABUAD INTERNATIONAL JOURNAL OF NATURAL AND APPLIED SCIENCES
ISSN: 2955-1021
AIJNAS Volume 3, Issue 2, 2023 pp 1-15
https://doi.org/10.53982/aijnas.2023.0302.01-j

https://journals.abuad.edu.ng/index.php/aijnas

AIJNAS

A new reduced contagious zero-inflated model: An application to count data

[1,2]M. I. Adarabioyo* and [2]G. M. Oyeyemi
[1]Department of Mathematical and Physical Sciences, Afe Babalola University, Ado-Ekiti, Nigeria
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria
*Author for correspondence: adarabioyomi@abuad.edu.ng

Abstract
In this paper, a reduced one-parameter contagious distribution was developed from the joint distribution of three-parameter gamma and Poisson distributions, on which Lakshmi's three-parameter gamma distribution is based to model a count data. The distribution properties and some common descriptive measures relating to this contagious distribution are derived. The behavior of the probability mass function with changes in parameters was also studied. The parameter estimation by the maximum likelihood and moment-generating function methods is discussed. A simulation study was carried out with the proposed model to check for consistency and bias. The new model show consistency as the sample size increases. The model was applied to a real-life dataset and was seen to be more flexible in capturing excess zero, under, and over-dispersion in count data and proved to be a useful alternative to some existing zero-inflated models.
Key Words: *Gamma, Poisson, Model, contagious, moment*

INTRODUCTION

Poisson distribution has been used extensively in modelling count data (June *et al*.,1999). However, application of Poisson distribution may not give valid results due to the fact that Poisson assumption may be invalidated by other features dominant in the data thereby paving the way for other valid models to be examined. Among these features is over-dispersion due to presence of many zeros in the data. The models for consideration include hurdle models and zero-inflated models (Lambert, 1992, Famoye *et al*.,2004 and 2006, Agrest *et al*.2004). They are class of models that can handle data with under-, equi- and over-dispersion in the data, both the hurdle model and the zero-inflated fit into this scenario. The zero-inflated is similar to the hurdle model in the sense that a zero-inflated model separates the zero part from the positive part whereas the hurdle model allows some zeros to be analysed with the nonzero. Zero-inflation regression model has been extensively used in the literature to account for excess zero that may arise in count data (Kazembe, 2013). The models such as the zero-inflated Poisson and zero-inflated negative binomial were proposed when there is excess zero and over-dispersion in the count data. However, non-zero observations may be over-dispersed, in this circumstance, parameter estimated would become biased and the standard errors underestimated. However, according to Mullahy (1986), zero-inflated negative binomial regression model better accounts for these characteristics compared to zero-inflated Poisson (ZIP). Moreover, when the number of zero in the sample exceeds what can be predicted by either Poisson or Negative Binomial, it is mostly preferred to use the ZIP or Zero-Inflated Negative Binomial (ZINB). Other important models such as Zero-truncated Poison and Zero-inflated Negative Binomial can also be used to model excess zero count data where the zero counts are truncated and the data is strictly positive (Adarabioyo and Ipinyomi, 2019). Furthermore, in the literature, the Poisson and Negative binomial have been modelled after zero-inflated and zero-truncated and have been widely used to model some real life data (Adarabioyo and Ipinyomi, 2020). However, this study provides alternative model known as a contagious zero-inflated model derived from the joint distribution of three-parameter Gamma distribution proposed by Lakshmi and Vaidyanathan 2016 and a Poisson distribution.

Contagious zero-inflated models for count data have gained considerable attention in recent years due to their ability to handle data with excessive zeros and over-dispersion (Famoye and Singh, 2006)). In this section, we will discuss some key points related to these models and their implications for count data analysis.One of the primary advantages of contagious zero-inflated models is their ability to capture over dispersion in count data (Showkat *et al*., 2022).

In many real-world scenarios, count data exhibits contagious behavior, where the occurrence of an event influences the likelihood of subsequent events. Contagious zero-inflated models explicitly account for this behavior by incorporating contagion parameters. This enables researchers to understand the spread or transmission of events, which is particularly relevant in fields such as epidemiology, social sciences, and finance (Peer and Mohammad, 2023).

Contagious zero-inflated models have found extensive applications in public health research. For example, in disease outbreak analysis, these models can capture the excess zeros due to individuals who are not susceptible to the disease (Frank *et al.*, 2022). Additionally, they can account for the contagion effect, where the occurrence of a disease in one individual increases the likelihood of infection in nearby individuals.

In economics and finance (Francisco *et al.*, 2017), count data often arises in the context of rare events, such as financial crises or market crashes. Contagious zero-inflated models provide a valuable framework for understanding the occurrence and propagation of such events. By considering the contagion effect, these models allow researchers to assess the impact of one event on the likelihood of subsequent events, providing insights into systemic risk and contagion dynamics.

Count data analysis is a fundamental aspect of many research domains, including public health, ecology, economics, and social sciences. However, traditional count models, such as Poisson or negative binomial, often encounter limitations when analyzing data with excessive zeros and contagious events. Contagious zero-inflated models offer an innovative solution to overcome these limitations by providing a comprehensive framework to handle both excess zeros and over-dispersion contagion. The objective of this study is to explore the advantages and applications of contagious zero-inflated models for count data.

Among those who have explored the zero-inflated and or hurdle models include Lambert (1992) who proposed Zero-Inflated Poison with 71.8% zeros and it performed better than the Zero-Inflated Negative Binomial model. Greene (1994) also proposed ZIP with 0.894 zeros and heavy skews of 4.02. Slymen *et al* (2006) modeled ZIP and negative binomial ZIP with identical results and with uniform event stage distribution. Warton (2005) findings showed that negative binomial fits better than the ZIP only when zero-inflation and over dispersion are present. His results were inconsistence with Lamber (1992). Mendonca and Kirchner (1999) findings showed that the ZIP actually fit better than the Poisson given 0.216 and 0.289 zero-deflation. These findings contradicted the suggestion that the hurdle model and not the ZIP appropriate for zero-deflated counts. However, Zorn (1996) found that given 0.958 zero-inflation and with skews of 7.97 and 1.86, the ZIP fitted better than the Poisson.

Further, it has been suggested that Hurdle and ZIP models should be chosen given a priori research about the source and nature of the zeros. It was also suggested that the negative binomial formulations are meant to handle additional over-dispersion in the event stage (Jeffrey M. M, 2007). Famoye, (1999) fitted the restricted generalized Poisson regression model. It is a three parameter model that account for excess zeros and over-dispersion. In 2024, he also fitted the On the Generalized Poisson Regression Model with an Application to Accident Data with the aim of identifying the relationship between the number of accidents and some covariates. Their results were compared with negative binomial regression and found to perform better than the Poisson Regression in identifying demographic factors and some other variables.

Further suggestions such as the proportion of zeros, the nature of the event stage distribution, varying the size of the skew (from heavily positive skew, moderately positive skew and uniformly distributed) and the number of predictors may affect the choice of model(s) for the count data (Jeffero-rey, 2007).

By reviewing existing literature, analyzing statistical properties, and examining real-world applications, this study aims to provide a comprehensive understanding of the potential of this model.

MATERIALS AND METHODS

The Density Function of a one-Parameter Poisson distribution

Let Y denote a random variable having one-parameter Poisson distribution with probability density function given as

$$f(Y=y,\theta)=\begin{cases}(y,\theta)=e^{-\theta}\frac{\theta^{y}}{y!}\ for\ y=0,1,2\ldots.\\ f(y,\theta)=0,\ elsewhere\end{cases} \quad 1$$

for y = 1, 2, 3, … We recall that for a Poisson variable $E(Y)=\theta$ and $Var(Y)=\theta$.

The Density Function of a three-Parameter Gamma distribution

$$g(\theta;\alpha,\beta,\mu)=\frac{(\theta-\mu)^{\alpha-1}e^{(-\frac{(\theta-\mu)}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)},\quad -\infty<\mu<\infty \quad 2$$

$$-\infty<\theta<\infty,\ \theta>0,\beta>0,\alpha>0,$$

Equation 2 is a three-parameter Gamma distribution as proposed by Lakshmi and Vaidyanathan 2016).

3.2.3 Proposed reduced The Contagious model

Given a sequence of density functions which are either discrete density or probability density function $f_0(.), f_1(.), f_2(.), \ldots, f_n(.), \ldots$ which may or may not depend on parameters and sequence of parameters $p_0, p_1, p_2 \ldots, p_n$ where $p_i \geq 0$ and $\sum_{i=1}^{\infty} p_i = 1$, and therefore, $\sum_{i=1}^{\infty} p_i f(x)$ is a density function which is also known as contagious distribution or mixture distribution proposed by Mood and Alexander (1913) which is a two-parameter gamma and Poisson distribution. By applying this method, the contagious distribution function of the three-parameter gamma and Poisson distributions was obtained in equation 5.

$$f(y_i,\theta)g(\theta;\ \alpha,\beta,\mu) = \int_0^{\infty} \frac{e^{-\theta}\,(\mu-\theta)^y}{y!} \frac{\theta^{\alpha-1} e^{-\frac{(\theta-\mu)}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}\, d\theta \qquad 3$$

Setting $\mu = 0$

$$f(y_i,\theta)g(\theta;\ \alpha,\beta,\mu) = \int_0^{\infty} \frac{e^{-\theta}\,\theta^y}{y!} \frac{\theta^{\alpha-1} e^{-\frac{\theta}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}\, d\theta \qquad 4$$

$$= \frac{1}{y!\beta^{\alpha}\Gamma(\alpha)} \int_0^{\infty} \theta^{\alpha+y-1}\, e^{-\left(\frac{\beta+1}{\beta}\right)\theta} d\theta$$

Take $Z = \left(\frac{\beta+1}{\beta}\right)$

$$= \frac{1}{y!\beta^{\alpha}\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+y)}{Z^{\alpha+y}} \int_0^{\infty} [Z\theta]^{\alpha+y-1}\, e^{-Z\theta} d[Z\theta]$$

$$= \frac{1}{y!\beta^{\alpha}\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+y)}{Z^{\alpha+y}}$$

$$= \frac{1}{y!\beta^{\alpha}\Gamma(\alpha)} \cdot \frac{(\alpha+y-1)!}{Z^{\alpha}Z^{y}}$$

$$= \left(\frac{1}{\beta+1}\right)^{\alpha} \frac{\Gamma(\alpha+y)}{y!\Gamma(\alpha)} \left(\frac{\beta+1}{\beta}\right)^{-y}$$

$$h(y_i,\alpha,\beta) = \left\langle {\alpha+y-1 \atop y} \right\rangle \left(\frac{1}{\beta+1}\right)^{\alpha} \left(\frac{\beta+1}{\beta}\right)^{-y}, \alpha \geq 1, \beta > 0 \qquad 5$$

Equation 5 is a special case of negative binomial distribution

Setting $\alpha = 1$,, we have the reduced Model as follows

$$h(y_i,\beta) = \left(\frac{1}{\beta+1}\right) \left(\frac{\beta+1}{\beta}\right)^{-y}, \beta > 0 \qquad 6$$

Equation 6 is a special case of geometric distribution

3.2.4 Moment Generating Function of the Reduced (One-Parameter) Model

The moment generating function of a random variable y of the distribution in (6) is defined as

$$M_t = E(e^{ty}) \qquad 7$$

$$= \sum_{x=0}^{\infty} e^{ty}\, P(Y=y)$$

$$= \sum_{x=0}^{\infty} e^{ty} \left(\frac{1}{\beta+1}\right) \left(\frac{\beta}{\beta+1}\right)^{y}, \beta > 0$$

$$= \left(\frac{1}{\beta+1}\right) \Sigma_0^{\infty} \left(\frac{\beta}{\beta+1} e^t\right)^{y}$$

By sum of infinite Geometric sequence, for the sum to be convergent, we must have

$$|(1 - \left(\frac{1}{\beta+1}\right) e^t\ | < 1$$

$$M_t = \left(\frac{1}{\beta+1}\right) \sum_{x=0}^{\infty} e^{ty} \left(\frac{\beta}{\beta+1} e^t\right)^{y} = \left(\frac{1}{\beta+1}\right) \left(1 - \frac{\beta}{\beta+1} e^t\right)^{-1} \qquad 8$$

Where $|\ \frac{1}{\beta+1} e^t| < 1$

The first and second moment are obtained as

$$M_t' = \frac{d}{dt}(M_t) = \left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right) e^t \left(1 - \frac{\beta}{\beta+1} e^t\right)^{-2}$$

Setting t=0, we obtained the mean of the distribution as

$$M_0' = \beta \qquad 9$$

The second moment is obtained by

$$M_t'' = \frac{d}{dt}(M_t') = \left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right) e^t \left[\frac{1+\left(\frac{\beta}{\beta+1}\right)e^t}{\left(1-\left(\frac{\beta}{\beta+1}\right)e^t\right)^3}\right] \qquad 10$$

Setting t=0, we obtained the second moment of the distribution as

$$M_0'' = \left[\frac{\beta}{(\beta+1)^2}\right] \left[\frac{1+\left(\frac{\beta}{\beta+1}\right)}{\left(1-\left(\frac{\beta}{\beta+1}\right)\right)^3}\right] \qquad 11$$

$$= 2\beta^2 + \beta \qquad 12$$

Hence the variance is obtained as

$$Variance = M_0'' - (M_0')^2 = 2\beta^2 + \beta - \beta^2 = \beta(\beta+1)$$

The third moment was obtained as follows,

$$M_t''' = \frac{d^3}{dt^3}(M_t'') = \left[\frac{\beta}{(\beta+1)^2}\right] \left[\frac{1+\frac{4\beta e^t}{\beta+1}+\frac{\beta^2 e^{2t}}{(\beta+1)^2}}{\left(1-\frac{\beta e^t}{\beta+1}\right)^4}\right] \qquad 13$$

The fourth moment was equally derived by the same procedure as follows

$$\frac{d}{dt}(M_t''') = \frac{\beta e^t}{(\beta+1)^2} \left[\frac{1+\frac{11\beta e^t}{\beta+1}+\frac{11\beta^2 e^{2t}}{(\beta+1)^2}+\frac{\beta^3 e^{3t}}{(\beta+1)^3}}{\left(1-\frac{\beta e^t}{\beta+1}\right)^5}\right] \qquad 14$$

The coefficients of variation (γ), the skewness $(\sqrt{\beta_1})$ and the kurtosis (β_2) of the contagious distribution were obtained as follows:

$$C.V = \gamma = \frac{\sigma}{\mu_I'} = \frac{\sqrt{\beta(\beta+1)}}{\beta},\ Skewess = \frac{{\mu_3}^2}{{\mu_2}^3} = \frac{(\beta^3+8\beta^2+\beta)^2}{(\beta(\beta+1))^3},\ Kurtosis =_2 \frac{\mu_4}{{\mu_2}^2} = \frac{37\beta^4+64\beta^3+29\beta^2+\beta}{(\beta(\beta+1))^2}$$

3.2.5 The Cumulative Distribution Function of the Proposed reduced Model

For a random variable $Y=y_1, y_2, \ldots, y_n$, the cumulative distribution function of the distribution in equation 6 was derived as follows:

$$\sum_{i=k}^{\infty}\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)^{y}=\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)^{k}+\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)^{k+1}+\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)^{k+2}+\cdots$$

$$=\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)^{k}\left[1+\left(\frac{\beta}{\beta+1}\right)+\left(\frac{\beta}{\beta+1}\right)^{2}+\left(\frac{\beta}{\beta+1}\right)^{3}+\cdots\right]$$

$$=\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)^{k}\left[\frac{1}{1-\left(\frac{\beta}{\beta+1}\right)}\right]$$

$$\sum_{i=k}^{\infty}\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)^{y}=\left(\frac{\beta}{\beta+1}\right)^{k} \qquad 15$$

3.2.6 The Maximum Likelihood Estimation Method the Proposed Reduced Model

For an independent and identically distributed (i.i.d) sample of $Y = y_1, y_2, y_3, \ldots, y_n$, the likelihood of parameter is defined as follow:

$$L(\beta|X) \propto f(X|\beta) = \prod_{i=1}^{n}\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)^{y}$$

$$=\left(\frac{1}{\beta+1}\right)^{n}\prod_{i=1}^{n}\left(\frac{\beta}{\beta+1}\right)^{y_i} \qquad 16$$

Taking the log of the likelihood function, we have

$$\log(\beta|Y) = nlog\left(\frac{1}{\beta+1}\right)+\sum_{i=1}^{n} y_i log\left(\frac{\beta}{\beta+1}\right)$$

$$= nlog\left(\frac{1}{\beta+1}\right)+\left(\frac{\beta}{\beta+1}\right)\sum_{i=1}^{n} y_i \qquad 17$$

The score which is the first derivative of the likelihood function is derived as

$$S(\beta) = n(\beta+1)-\left(\frac{\beta+1}{\beta}\right)\sum_{i=1}^{n} y_i \qquad 18$$

The score is set to zero to solve for β

$$S(\beta) = n(\beta+1)-\left(\frac{\beta+1}{\beta}\right)\sum_{i=1}^{n} y_i = 0$$

$$n(\beta+1) = \left(\frac{\beta+1}{\beta}\right)\sum_{i=1}^{n} y_i$$

$$n\beta(\beta+1) = (\beta+1)\sum_{i=1}^{n} y_i$$

$$n\beta = \sum_{i=1}^{n} y_i$$

$$\beta = \frac{\sum_{i=1}^{n} y_i}{n} \qquad 19$$

$$\beta = \mu_y \qquad 20$$

The second derivative is obtained as

$$log''(\beta|Y) = -n(\beta+1)^2-\left(\frac{\beta+1}{\beta}\right)^{2}\sum_{i=1}^{n} {y_i}^2 \qquad 21$$

This captures the steepness of the distribution around probability (The Fisher information is the negation of the second derivative derived as:

$$Jn(\beta) = -log'' = n(\beta+1)^2+\left(\frac{\beta+1}{\beta}\right)^{2}\sum_{i=1}^{n} {y_i}^2 \qquad 22$$

The Zero-Inflation of the Proposed Reduced Model

Let be a nonnegative values of a random variable and if y = 0 is observed with frequency significantly higher such that it cannot be modeled by a Poisson or negative binomial models, thus the distribution is defined by

$$P(Y=y_i/x_i,z_i) = \begin{cases}\omega_i+(1-\omega_i)f(y_i=0), & y_i=0\\ (1-\omega_i)f(y_i=0), & y_i\geq 0\end{cases} \qquad 23$$

Where $f(y_i = 0)$, $y_i = 0, 1, 2, \ldots$ is the pdf of Y_i and $0 < \varphi_i < 1$. The function $\omega_i = \omega_i(Z_i)$

Satisfying the logit $(\varphi_i) = log\frac{\omega_i}{1-\omega_i} = \sum_{j=1}^{m} Z_{ij}\delta_j$ where $Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{im})$ the ith row of the covariate matrix Z and $\delta = \delta_1, \delta_2, \ldots, \delta_m)$ are the unknown m-dimensional column vector of parameters. The nonnegative function is modeled through link function that allows being negative may be used and $f(y_i = 0)$ is the distribution defined as

$$f(y_i) = \Pr(Y_i = y_i/\alpha,\beta) = \left(\frac{1}{1+\beta}\right)\left(\frac{\beta}{1+\beta}\right)^{y_i} \qquad 24$$

We consider the zero-inflation of the 3-Parameter Gamma-Poisson model in which the response variable Y_i (i= 1, 2, …, n) has the distribution

$$P(Y=y_i/x_i,z_i) = \begin{cases}\omega_i+(1-\omega_i)\left(\frac{1}{1+\beta}\right), y_i=0\\ (1-\omega_i)\left(\frac{1}{1+\beta}\right)\left(\frac{\beta}{1+\beta}\right)^{y_i}, y_i\geq 0\end{cases} \qquad 25$$

The mean and variance are obtained as

$$E(Y_i) = (1-\omega_i)\beta_t \text{ and } Var(Y_i) = (1-\omega_i)\beta_t[(\beta_t+1)+\omega\beta] \qquad 26$$

Where ω_i is the logit link function and β_t the exposure rate is modeled through the log link function defined below

$$\log(\beta_t) = X\vartheta \text{ , and } \operatorname{logit}\left(\frac{\omega_i}{1-\omega_i}\right) = Z\gamma. \qquad 27$$

where X are Z matrices of covariates and ϑ *and* γ are vectors of parameters. The two sets of covariates may or may not be the same. When they do, more parsimonious models may be developed by assuming that the two linear predictors are related in some way.

The Maximum Likelihood Estimation of equation 3.65 is obtained as follows

$$L(\vartheta,\gamma,y,X,Z) = \sum_{i=1}^{n}\left(1+e^{Z_i'\gamma}+\sum_{i:y=0}^{n} log\left(e^{Z_i'\gamma}+\left(e^{X_i'\beta}+1\right)^{-1}\right)+\sum_{i:y>0}\left(\log\left(e^{X_i'\beta}+1\right)+ y_i log\left(1+e^{X_i'\beta}\right)\right)\right. \qquad 28$$

where $X = (x_1, x_2, \ldots, x_n)$ and $Z = (z_1, z_2, \ldots, z_n)$.

The first order derivative of L_z with respect to the $\theta = (\vartheta, \gamma)'$ parameters was obtained as

$$\frac{\delta L_z}{\delta\vartheta_j} = \begin{cases}\sum_{i=1}^{n}\left(\frac{\beta_t}{\beta_t(1+q_{ir_i})}X_{ij}\right), y=0\\ \sum_{i=1}^{n}\left(1-\frac{1+y_i}{\beta_t+1}\right)X_{ij}\ , y>0\end{cases} \qquad 29$$

$$\frac{\delta L_z}{\delta\gamma_j} = \begin{cases}\sum_{i=1}^{n}(-)\left(\frac{1}{1+e^{Z_i'\gamma}}+\frac{1}{1+q_{ir_i}}Z_{ij}\right), y=0\\ \sum_{i=1}^{n}\left(\frac{q_i}{1+q_i}Z_{ij}\right)\ , y>0\end{cases} \qquad 29$$

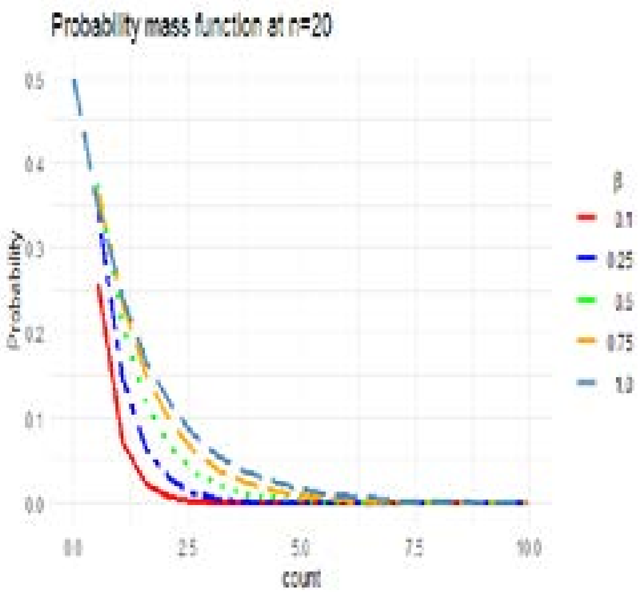
Where

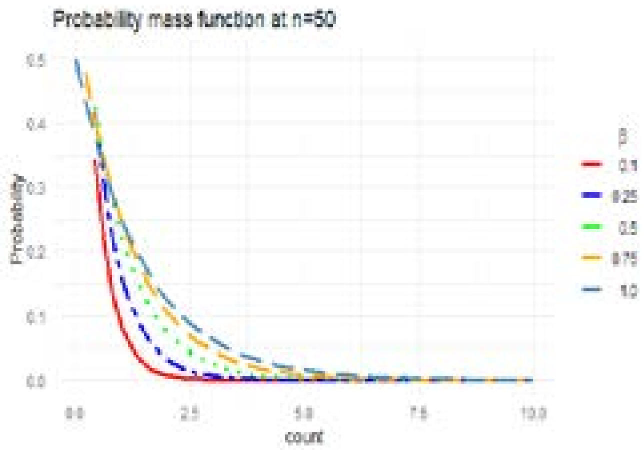
$$\beta_t = e^{X_i'\vartheta}, r_i = \beta_t+1,\ q = e^{Z_i'\gamma} \qquad 30$$

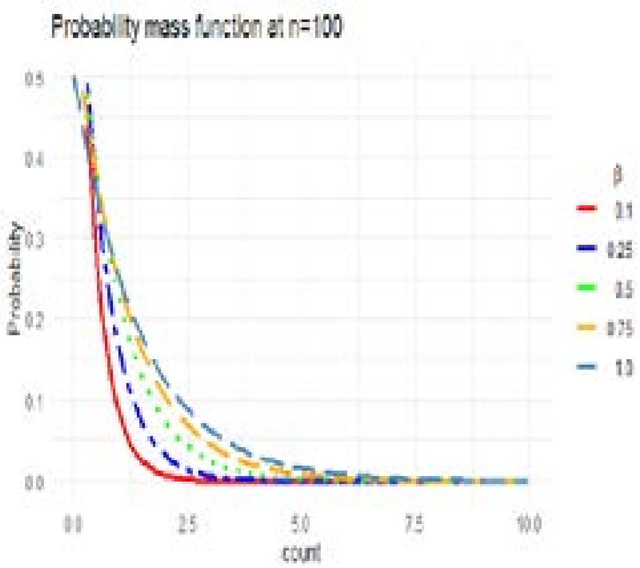
The parameter estimation is done by BFGS algorithm adopted by Nocedal and Wright (pp190-202). The BFGS is a quasi-Newton optimization technique which can be implemented in the Optim-R software package.

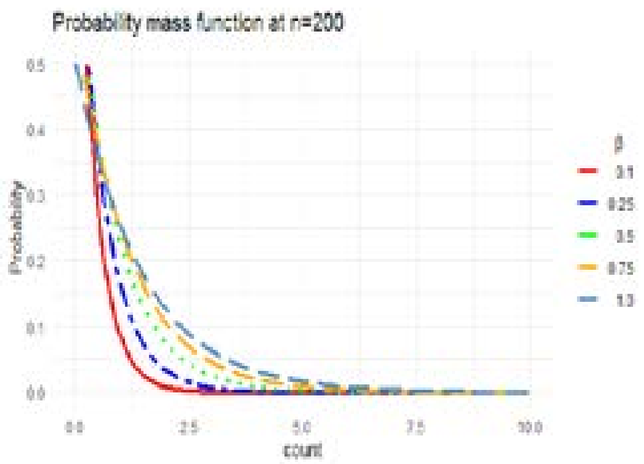
Density Plots

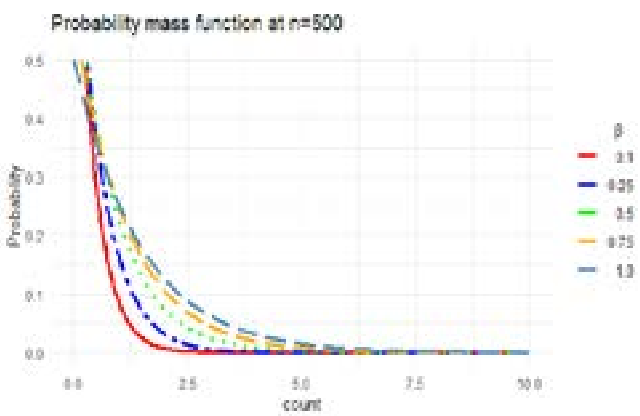
The density plots were obtained for the three-Parameter-Gamma-Poisson Distribution by assigning different values to the parameter (β) of the model. The density plots were obtained as follows:

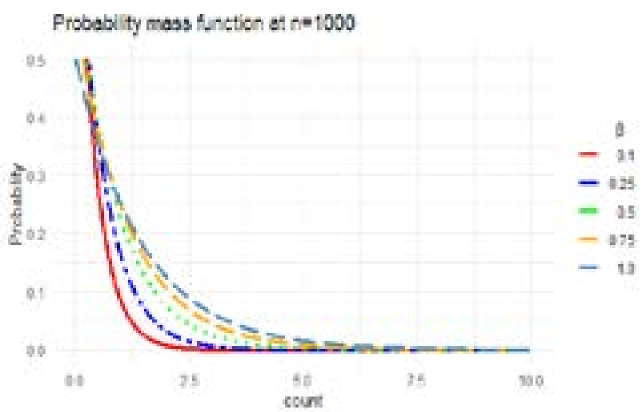

Probability mass function at n=20
Probability
count
β
0.1
0.25
0.5
0.75
1.0

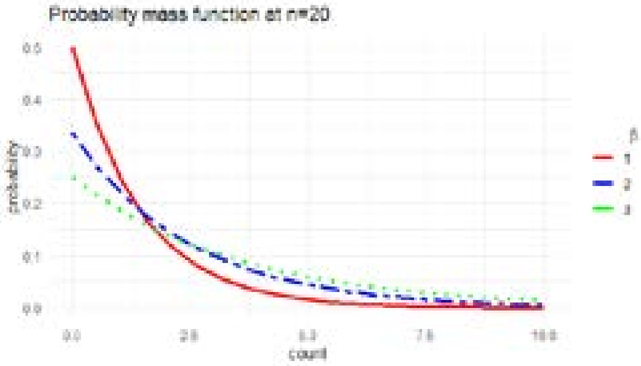

Probability mass function at n=50
Probability
count
β
0.1
0.25
0.5
0.75
1.0

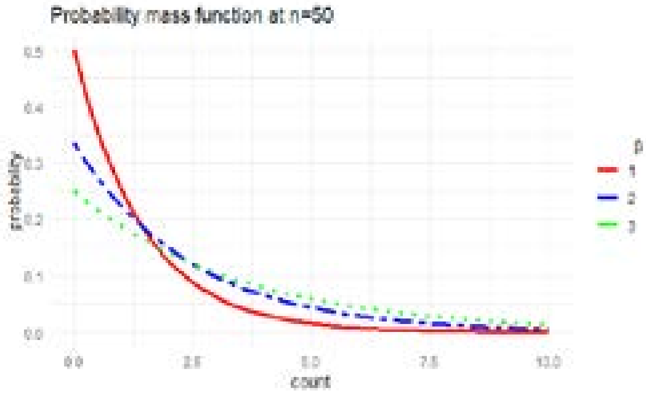

Probability mass function at n=100
Probability
count
β
0.1
0.25
0.5
0.75
1.0

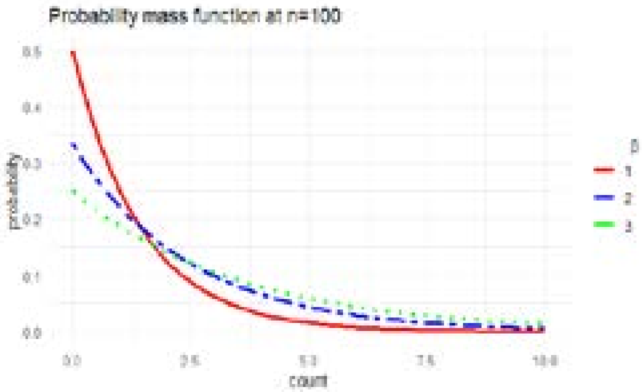

Probability mass function at n=200
Probability
count
β
0.1
0.25
0.5
0.75
1.0

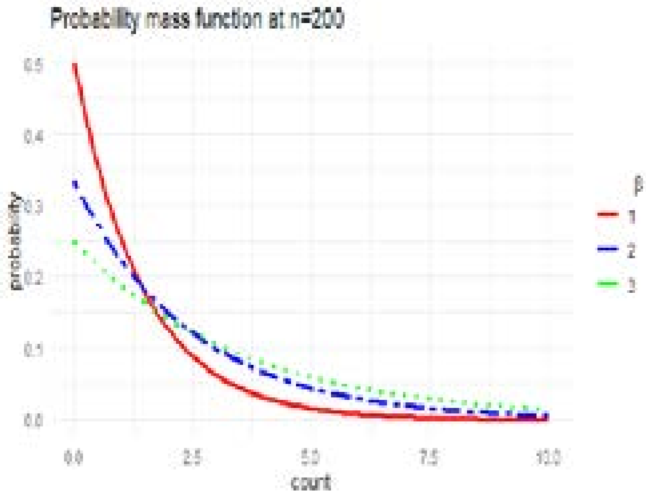

Probability mass function at n=500
Probability
count
β
0.1
0.25
0.5
0.75
1.0

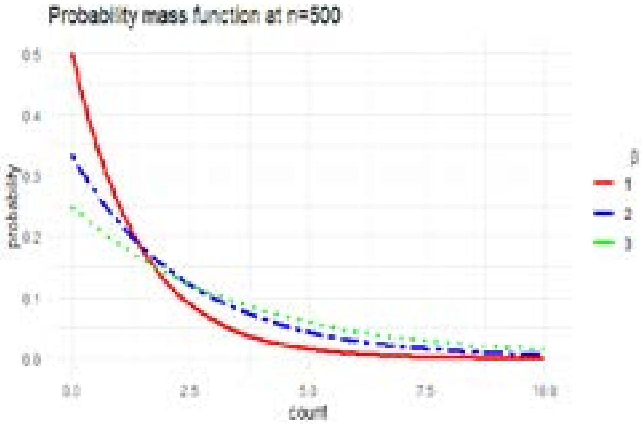

Probability mass function at n=1000
Probability
count
β
0.1
0.25
0.5
0.75
1.0

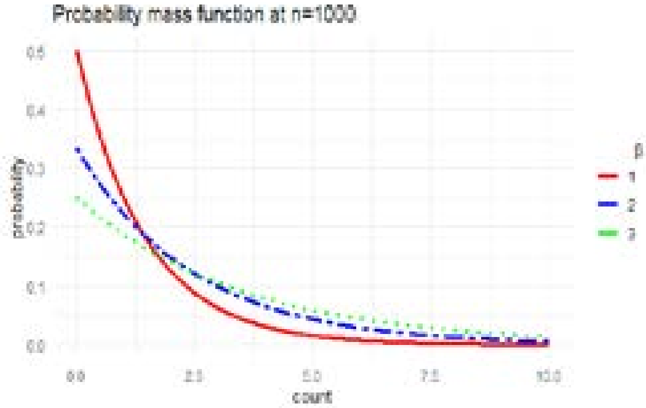

Probability mass function at n=20
probability
count
β
1
2
3


Probability mass function at n=50
probability
count
β
1
2
3


Probability mass function at n=100
probability
count
β
1
2
3


Probability mass function at n=200
probability
count
β
1
2
3


Probability mass function at n=500
probability
count
β
1
2
3


Probability mass function at n=1000
probability
count
β
1
2
3

RMSE of the proposed reduced model and ZIG

The Root Mean Square Error of the proposed reduced model and the (ZIG) were obtained by setting and varying the size of and *n* in the model. The model fit was performed by R software and the RMSE were obtained and tabulated in tables 4.1to 4.3 below.

Table 1: RMSE the Proposed Reduced Model and the ZIG at β=1

β=1	Model	n=20	n=50	n=100	n=200	n=500	n=1000
W = 0.1	ZIG	0.04281	0.07565	0.05432	0.03804	0.02310	0.01681
	Proposed	0.04281	0.07563	0.05414	0.03765	0.02291	0.01672
W=0.25	ZIG	0.03891	0.07047	0.04981	0.03509	0.02231	0.01566
	Proposed	0.03890	0.07044	0.0972	0.03488	0.02217	0.01548
W=0.5	ZIG	0.02634	0.04471	0.04033	0.02964	0.01890	0.01259
	Proposed	0.02634	0.04456	0.04010	0.02951	0.01973	0.01237
W=0.75	ZIG	0.03931	0.03485	0.02794	0.02092	0.01335	0.00988
	Proposed	0.03931	0.03482	0.02787	0.02078	0.01321	0.00972
W=0.9	ZIG	0.04253	0.03858	0.02923	0.01822	0.0088	0.0059
	Proposed	0.04253	0.03858	0.02911	0.01811	0.0074	0.0038

Table 1 consists of root mean square error of zero-inflated Geometric (ZIG) and the proposed reduced model. At β=1 at different levels of zero fractions (ω). However, the RMSEs of the proposed model were smaller to that of ZIG as the sample size increases from 100 to 1000. The result equally shows that as sample size increases, RMSE decreases (RMSE) →0 and as increases the RMSE also decreases ((RMSE).

Table 2: RMSE the Proposed Reduced Model and the ZIG at β=2

β=2	Model	20	50	100	200	50	1000
ω = 0.1	ZIG	0.8241	0.7565	0.0527	0.0373	0.0229	0.0169
	Proposed	0.8241	0.7562	0.0521	0.0369	0.0221	0.0168
ω =0.25	ZIG	0.0760	0.0667	0.0499	0.0371	0.0219	0.0159
	Proposed	0.0760	0.0652	0.0491	0.0370	0.0218	0.0148
ω =0.5	ZIG	0.0575	0.0468	0.0404	0.0296	0.0179	0.0133
	Proposed	0.0575	0.0466	0.0403	0.0277	0.0179	0.0132
ω =0.75	ZIG	0.0492	0.0431	0.0398	0.0211	0.0131	0.0094
	Proposed	0.0492	0.0427	0.0397	0.0209	0.0130	0.0091
ω =0.9	ZIG	0.0583	0.0402	0.0321	0.0109	0.0084	0.0061
	Proposed	0.0583	0.0402	0.0319	0.0107	0.0067	0.0052

Table 2 consists of root mean square error of the ZIG and the proposed reduced model. At β=2t the RMSEs of the proposed reduced model were smaller to that of ZIG as the sample sizes increases from 100 to 1000.The result equally shows that as sample sizes increases, RMSE decreases (RMSE →0) and as W increases the RMSE also decreases ((RMSE →0).

Table 3: RMSE the Proposed Reduced Model and the ZIG at 3

β=3	Model	20	50	100	200	50	1000
ω = 0.1	ZIG	0.0845	0.0751	0.0549	0.03731	0.0244	0.0272
	Proposed	0.0845	0.075	0.0541	0.03728	0.0243	0.0268
ω =0.25	ZIG	0.0810	0.0809	0.0598	0.0366	0.0263	0.0099
	Proposed	0.0807	0.0806	0.0575	0.0356	0.026	0.0099
ω =0.5	ZIG	0.0834	0.0572	0.0432	0.02801	0.0185	0.0133
	Proposed	0.0833	0.057	0.0428	0.0029	0.0185	0.01329
ω =0.75	ZIG	0.0743	0.0719	0.0711	0.02131	0.0139	0.00985
	Proposed	0.0743	0.0718	0.0710	0.02129	0.0129	0.00981
ω =0.9	ZIG	0.0537	0.0441	0.0382	0.0147	0.0086	0.00595
	Proposed	0.0537	0.0441	0.0381	0.0139	0.0078	0.00539

Table 3 consists of root mean square error (RMSE) of the proposed reduced model and ZIG when β=3. The result equally shows that as sample size increases, RMSE decreases (RMSE →0) and as increases the RMSE also decreases ((RMSE →0).

RMSE of the Proposed Reduced Model and ZINB

The Root Mean Square Error of the proposed reduced model and ZINB were obtained by setting β=1,2 and 3 and varying the size of *W* and *n* in the model. The model fit was performed by R software and the RMSE were obtained and tabulated in tables 4.3.3.1 to 4.3.3.6 below.

Table 4: RMSE of the Proposed Reduced Model and ZINB at β=1

Weight (W) β=1	Model	20	50	100	200	500	1000
ω = 0.1	ZINB	0.2437	0.1903	0.1407	0.1022	0.0630	0.0451
	Proposed	0.2437	0.1882	0.1405	0.1018	0.0628	0.0443
ω =0.25	ZINB	0.2386	0.1785	0.1244	0.0900	0.0573	0.0404
	Proposed	0.2386	0.1783	0.1223	0.0882	0.0572	0.0400
ω =0.5	ZINB	0.2262	0.1547	0.1059	0.0782	0.0501	0.0351
	Proposed	0.2262	0.1541	0.1043	0.0781	0.0489	0.0348
ω =0.75	ZINB	0.1826	0.1409	0.0784	0.0559	0.0367	0.0258
	Proposed	0.1826	0.1409	0.0782	0.0558	0.0366	0.0242
ω =0.9	ZINB	0.1545	0.1201	0.0638	0.0286	0.0228	0.0157
	Proposed	0.1545	0.1199	0.0621	0.0277	0.0219	0.0149

Table 4 consists of root mean square error (RMSE) of the zero-inflated Negative Binomial (ZINB) and the proposed reduced model at β=1. However, the RMSEs of the proposed reduced model were smaller to that of ZINB as the sample size increases from 50 to 1000.

The result equally shows that as sample size increases, RMSE decreases (RMSE →0) and as ω increases the RMSE also decreases ((RMSE →0).

Table 5: RMSE of the Proposed Reduced Model and ZINB at β=2

Weight (W) β=2	Model	20	50	100	200	500	1000
ω = 0.1	ZINB	0.3062	0.1992	0.1378	0.0990	0.0639	0.0448
	Proposed	0.3062	0.1989	0.1364	0.0981	0.0633	0.0439
ω =0.25	ZINB	0.2273	0.1771	0.1345	0.1089	0.0584	0.0402
	Proposed	0.2273	0.1769	0.1342	0.1078	0.0582	0.0402
ω =0.5	ZINB	0.1736	0.1603	0.1055	0.0738	0.0478	0.0343
	Proposed	0.1736	0.1584	0.1049	0.0737	0.0478	0.0342
ω =0.75	ZINB	0.0389	0.0531	0.0779	0.0566	0.0351	0.0257
	Proposed	0.0389	0.0530	0.0776	0.0564	0.0350	0.0256
ω =0.9	ZINB	0.0342	0.0325	0.0300	0.0280	0.0233	0.0165
	Proposed	0.0321	0.0312	0.0310	0.0269	0.0232	0.0159

Table 5 consists of root mean square error (RMSE) of the zero-inflated Negative Binomial (ZINB) and the proposed reduced model at β=2. Likewise, the RMSEs of the proposed reduced model were smaller to that of ZINB as the sample size increases 1000. The result, in summary, equally shows that as sample size increases, RMSE decreases (RMSE →0) and as W (proportion of zeros) increases the RMSE also decreases ((RMSE →0).

Table 6: RMSE of the Proposed Reduced model and ZINB at β=3

Weight (W) β=3	Model	20	50	100	200	50	1000
ω = 0.1	ZINB	0.2242	0.2131	0.1315	0.1003	0.06571	0.04246
	Proposed	0.2237	0.2032	0.1315	0.1003	0.0644	0.04023
ω =0.25	ZINB	0.2022	0.2021	0.1297	0.09073	0.0583	0.04012
	Proposed	0.201	0.2	0.1286	0.09701	0.0583	0.04012
ω =0.5	ZINB	0.1679	0.1578	0.1089	0.07748	0.0497	0.01324
	Proposed	0.1643	0.1578	0.1089	0.07746	0.0493	0.01329
ω =0.75	ZINB	0.1203	0.09955	0.08351	0.05738	0.0351	0.02439
	Proposed	0.1203	0.09955	0.0835	0.05729	0.0351	0.02429
ω =0.9	ZINB	0.0482	0.0734	0.0511	0.0363	0.0235	0.0424
	Proposed	0.0482	0.0733	0.0508	0.036	0.0234	0.04118

Table 6 consists of root mean square error (RMSE) of the zero-inflated Negative Binomial (ZINB) and the proposed reduced model at β=3. However, the result equally shows that as sample sizes increases, RMSE decreases (RMSE→0)) and as W (proportion of zeros) increases the RMSE also decreases ((RMSE→0)).

RMSE of Proposed Reduced model and ZIP

The Root Mean Square Error of the proposed reduced Model and ZIP were obtained by setting β=1,2 and 3 and varying the size of *W* and *n* in the model. The model fit was performed by R software and the RMSE were obtained and tabulated in tables 7-9 below.

Table 7: RMSE of the proposed Reduced Model and ZIP at β=1

Weight (W) β=1	Model	20	50	100	200	500	1000
ω = 0.1	ZIP	0.1855	0.1239	0.0877	0.0643	0.0394	0.0274
	Proposed	0.1855	0.1237	0.0875	0.0632	0.0392	0.0272
ω =0.25	ZIP	0.1616	0.1188	0.0824	0.6136	0.0370	0.0263
	Proposed	0.1616	0.1183	0.0823	0.0611	0.0368	0.0261
ω =0.5	ZIP	0.1569	0.1059	0.0718	0.0533	0.0342	0.0244
	Proposed	0.1569	0.1058	0.0717	0.0531	0.0341	0.0242
ω =0.75	ZIP	0.0895	0.0825	0.0552	0.0407	0.0245	0.0179
	Proposed	0.0895	0.0821	0.0551	0.0405	0.0226	0.0177
ω =0.9	ZIP	0.0900	0.0589	0.0432	0.0264	0.0166	0.0117
	Proposed	0.0900	0.0589	0.0432	0.0249	0.0145	0.0115

Table 7 consists of root mean square error (RMSE) of the zero-inflated Poisson (ZIP) and the proposed reduced model at β=1. However, the result equally shows that as sample sizes increases, RMSE decreases (RMSE→0) and as W (proportion of zeros) increases the RMSE also decreases ((RMSE→0).

Table 8: RMSE of the proposed Reduced Model and ZIP at β=2

Weight (W) β=3	Model	20	50	100	200	500	1000
ω = 0.1	ZIP	0.1939	0.1239	0.0878	0.0642	0.0396	0.0290
	Proposed	0.1939	0.1238	0.0874	0.0664	0.0395	0.0276
ω =0.25	ZIP	0.1798	0.1470	0.0861	0.0718	0.0367	0.0264
	Proposed	0.1798	0.1470	0.0855	0.0715	0.0367	0.0264
ω =0.5	ZIP	0.1412	0.1041	0.0827	0.0574	0.0289	0.0256
	Proposed	0.1412	0.1031	0.0826	0.0262	0.0276	0.0251
ω =0.75	ZIP	0.0813	0.0721	0.0566	0.0401	0.0256	0.0174
	Proposed	0.0813	0.0719	0.0565	0.0382	0.0243	0.0164
ω =0.9	ZIP	0.0722	0.0538	0.0312	0.0210	0.0165	0.0114
	Proposed	0.0722	0.0538	0.0311	0.0202	0.0162	0.0101

Table 8 consists of root mean square error (RMSE) of the zero-inflated Poisson (ZIP) and the proposed reduced model at β=2. The result equally shows that as sample sizes increases, RMSE decreases (RMSE→0) and as ω (proportion of zeros) increases the RMSE also decreases ((RMSE→0).

Table 9: RMSE of the proposed Reduced Model and ZIP at β=3

Weight (W) β=3	Model	20	50	100	200	500	1000
ω = 0.1	ZIP	0.1338	0.1214	0.0876	0.06048	0.0387	.02722
	Proposed	0.1332	0.1209	0.0870	0.06047	.0387	0.0272
ω =0.25	ZIP	0.1406	0.1499	0.0814	0.06421	0.0375	0.0263
	Proposed	0.1398	0.1496	0.0814	0.06418	0.0373	0.0244
ω =0.5	ZIP	0.0766	0.1061	0.0756	0.05198	0.0338	0.0246
	Proposed	0.0765	0.1046	0.0749	0.05188	0.0335	0.0239
ω =0.75	ZIP	0.0754	0.0725	0.0566	0.03975	0.0265	0.0178
	Proposed	0.07538	0.0725	0.05651	0.03973	0.0265	0.01759
ω =0.9	ZIP	0.0557	0.0432	0.0363	0.02709	0.0164	0.01184
	Proposed	0.0553	0.0428	0.0339	0.02636	0.0164	0.01092

Table 9 consists of the root mean square error (RMSE) of the zero-inflated Poisson (ZIP) and the proposed reduced model at β=3. However, the result equally shows that as sample sizes increases, RMSE decreases (RMSE→0) and as ω (proportion of zeros) increases the RMSE also decreases ((RMSE→0).

Model with covariates

The proposed reduced model was fitted to four simulated covariates x_1,x_2,x_3 and x_4 generated from NB (n, 0.25, 0.5), B(n, 0.5), P(0.5) and N(4, 2) respectively.Model validity was carried out by deriving the AIC, BIC, the Mean Absolute Bias and the Root Mean Square Error (RMSE).

AIC, BIC and Mean Absolute Bias

Table 10: AIC at β=1,ω=0.9

Model	20	50	100	200	500	1000
ZIG	42.314	55.515	90.938	90.939	351.543	562.399
ZIP	40.209	54.937	91.472	91.473	351.790	558.869
ZINB	50.209	64.937	100.751	100.752	361.220	568.867
Reduced Model	42.414	55.525	90.886	90.671	348.421	560.453

Table 10 presents the AIC of the proposed reduced model at β=1,W=0.1 and that of the ZIG, ZIP and ZINB. The proposed model reduced has least AIC at sample 100 and above when compared to the other three models. This shows that the proposed model performed better at higher sample size than at lower sample based on the AIC model validation procedure.

Table 11: BIC at β=1,ω=0.9

Model	20	50	100	200	500	1000
ZIG	26.9194	40.12028	75.5438	113.244	336.148	547.005
ZIP	24.8144	39.54242	76.0781	112.554	336.396	543.475
ZINB	27.1172	41.8451	77.6596	114.856	338.128	545.778
Proposed Model	27.0194	40.22328	75.3317	113.143	336.063	543.003

Table 11 presents the BIC of the proposed reduced model at β=1,W=0.1 and that of the ZIG, ZIP and ZINB. The proposed reduced model has least BIC at sample 100 and above when compared to the other three models. For instance, at sample size 20, the BIC of 3GP was 27.01941 indicating the least when compared with other three models. Also at sample size 100, 500 and 1000, the BIC of the proposed reduced model were 75.3317, 336.0628 and 543.0032 respectively, these were least when compared to the other three models. This shows that the proposed model performed better at higher sample size than at lower sample based on the BIC model validation procedure.

The Mean Absolute Bias of the Models

Table 12: Mean Absolute Bias (MAB) β=1,ω=0.1

Sample	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Proposed (Reduced Model)
20	0.7204	0.2667	0.2837	0.1987
50	0.7413	0.7773	0.7762	0.5688
100	0.7702	0.7467	0.7241	0.5798
200	0.7252	0.7575	0.7221	0.5696
500	0.7097	0.7694	0.76	0.5526
1000	0.7508	0.7975	0.7974	0.5973

Table 12: presents the mean absolute bias (MAB) of the proposed reduced model and that of the other three models at varied sample sizes and at β=1,W=0.1 . The mean absolute bias (MAB) of the proposed reduced model was compared with the ZIG, ZIP and ZINB and was found to be relatively better across sample sizes as the MABs were the least when compared to others. For instance at n = 20 the MAB of the proposed model was 0.1987 and was least compared to other models. Also at n = 100 and 1000, the MAB of the proposed model was 0.5798 and 0.5973 respectively and were the least compared to other models.

Table 13: Mean Absolute Bias (MAB) β=1 ,ω=0.5

Sample	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Proposed (Reduced Model)
20	0.7202	0.2668	0.2833	0.1983
50	0.7411	0.7768	0.7761	0.5688
100	0.77	0.7464	0.7238	0.5799
200	0.7249	0.7571	0.7218	0.5701
500	0.7093	0.7691	0.7602	0.5532
1000	0.7508	0.7972	0.7968	0.5971

Table 13 presents the mean absolute bias (MAB) at β=1 ,W=0.5 of the proposed reduced model and that of the other three models at varied sample sizes and the results indicate that the MAB of the proposed model was least compared with other three models. For instance at sample size 20, the MAB was 0.1983 and at sample size 500, the MAB was 0.5532. Therefore based on the MAB criterion, the proposed model is adjudged better in modelling both zero-inflation and zero deflation.

Table 14: Mean Absolute Bias (MAB) β=1,ω=0.9

Sample	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Proposed (Reduced Model)
20	0.6463	0.6625	0.5575	0.4457
50	0.7225	0.7785	0.7785	0.5868
100	0.7798	0.7284	0.7785	0.6235
200	0.7521	0.7988	0.7584	0.6569
500	0.745	0.7859	0.7392	0.7731
1000	0.7537	0.785	0.785	0.7474

Table 14 presents the mean absolute bias (MAB) at β=1,ω=0.9 of the proposed reduced model and that of the other three models at varied sample sizes and the results indicate that the MAB of the proposed model was least compared with other three models in all sample sizes. For instance at sample size 20, the MAB was 0.4457 and at sample size 500, the MAB was 0.7731. Therefore based on the MAB criterion, the proposed model is adjudged better in modelling both zero-inflation and zero deflation.

Table 15: Mean Absolute Bias (MAB) β=2,ω=0.1

MODEL	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Proposed Reduced Model
20	1.3646	2.0693	2.0693	0.2428
50	1.8832	1.0265	1.4195	0.1286
100	1.0047	2.5128	7.4779	0.0609
200	1.1909	0.1572	8.2415	0.0679
500	1.8676	0.0513	0.0861	0.1121
1000	0.2426	0.0602	0.1048	0.0556

Table 15 presents the mean absolute bias (MAB) at β=2,W=0.1 of the proposed reduced model and that of the other three models at varied sample sizes and the results indicate that the MAB of the proposed model was least compared with other three models in all sample sizes. For instance at sample size 20, the MAB was 0.2428 and at sample size 1000, the MAB was 0.0556. However, at sample size 500, the MAB of ZIP and ZINB were the least. Therefore based on the MAB criterion, the proposed model is adjudged better in modelling both zero-inflation and zero deflation.

Table 16: Mean Absolute Bias (MAB) β=2,ω=0.5

MODEL	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Proposed Reduced Model
20	4.6565	7.1699	0.9367	0.2781
50	9.2431	1.0265	1.4195	0.2783
100	2.4553	89.1412	87.4779	0.0437
200	1.092	0.1012	0.2743	0.1431
500	7.8676	0.0513	0.0963	0.1121
1000	0.0651	0.0989	0.0879	0.0556

Table 17 presents the mean absolute bias (MAB) at β=2,ω=0.5 of the proposed reduced model and that of the other three models at varied sample sizes. The results indicate that the MAB of the proposed model was least compared with other three models in all sample sizes. For instance at sample size 20, the MAB was 0.0.2781 and at sample size 500, the MAB was 0.1121. Therefore based on the MAB criterion, the proposed `model is adjudged better in modelling both zero-inflation and zero deflation.

Table 18: Mean Absolute Bias (MAB) β=2,ω=0.5

Model	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Propose Reduced Model
15	0.2002	0.0226	0.0682	0.1668
25	1.3646	2.0693	1.2543	0.2428
50	1.8832	1.0265	1.4195	0.1286
100	1.0047	2.5128	1.4779	0.0609
150	0.4193	0.2745	0.5872	0.0355
300	1.1909	0.1572	1.2415	0.0679
500	1.8676	0.0513	0.0861	0.0121
1000	0.2426	0.0602	0.1048	0.0556

Table 17 presents the mean absolute bias (MAB) at β=2,ω=0.5 of the proposed reduced model and that of the other three models at varied sample sizes. The results indicate that the MAB of the proposed model was least compared with other three models in all sample sizes. For instance at sample size 20, the MAB was 0.0.2781 and at sample size 500, the MAB was 0.1121. Therefore based on the MAB criterion, the proposed `model is adjudged better in modelling both zero-inflation and zero deflation.

Table 19: Mean Absolute Bias (MAB) β=3,ω=0.1

MODEL	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Proposed Reduced Model
20	0.5096	3.955	2.2284	0.0132
50	1.6561	1.2033	11.4939	0.1027
100	1.4553	1.9623	12.9456	0.043
200	1.2171	1.1903	1.476	0.1399
500	1.6103	0.255	0.4893	0.0835
1000	1.6674	0.7709	1.341	0.04

Table 19 presents the mean absolute bias (MAB) at β=3,ω=0.1 of the proposed model (3GP) and that of the other three models at varied sample sizes. The results indicate that the MAB of the proposed model was least compared with other three models in all sample sizes. For instance at sample size 20, the MAB was 0.0132 and at sample size 500, the MAB was 0.0835. Therefore based on the MAB criterion, the proposed model is adjudged better in modelling both zero-inflation and zero deflation.

Table 20: Mean Absolute Bias (MAB) β=3,ω=0.5

MODEL	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Proposed Reduced Model
20	0.1045	0.0908	0.1284	0.0078
50	0.0881	0.1898	0.028	0.0791
100	0.0524	0.1251	0.227	0.0904
200	0.1292	0.1499	0.121	0.0529
500	0.0817	0.0396	0.0886	0.0979
1000	0.13	0.0794	0.0613	0.0057

Table 20 presents the mean absolute bias (MAB) at β=3,ω=0.5 of the proposed reduced model and that of the other three models at varied sample sizes. The results indicate that of the proposed reduced Model model was least compared with other three models in all sample sizes. For instance at sample size 20, the MAB was 0.0078 and at sample size 500, the MAB was 0.0979. Therefore based on the MAB criterion, the proposed model is adjudged better in modelling both zero-inflation and zero deflation.

Table 21: Mean Absolute Bias (MAB) β=3,ω=0.9

MODEL	Zero-Inflated Geometric (ZIG)	Zero-Inflated Poisson (ZIP)	Zero-Inflated NB (ZINB)	Proposed Reduced Model
20	0.1045	0.0908	0.1321	0.0107
50	0.0881	0.0891	0.03	0.0481
100	0.0434	0.0521	0.146	0.0304
200	0.1302	0.1499	0.121	0.0621
500	0.0817	0.0396	0.086	0.0734
1000	0.132	0.0794	0.061	0.0572

Table 21 presents the mean absolute bias (MAB) at β=3,ω=0.9 of the proposed reduced model and that of the other three models at varied sample sizes. The results indicate that the MAB of the proposed model was least compared with other three models in all sample sizes except at sample size 50 where the MAB of ZINB was the least. For instance at sample size 20, the MAB was 0.0107 and at sample size 500, the MAB was 0.0572. Therefore based on the MAB criterion, the proposed reduced model is adjudged better in modelling both zero-inflation and zero deflation.

Application of the Proposed Model to Number of Under-Five Death in Nigeria

A The Dataset

The study uses data from a set of Demographic and Health Surveys (DHS) that were carried out in Nigeria in 2008. Over the years, the DHS project has offered technical support to surveys in a number of developing nations, enhancing understanding of demographic and health patterns worldwide. To help survey processes, ensuring that the data accurately represent the settings they are intended to depict, and to make sure that they are comparable between nations and across time, DHS has created standard protocols, methodology, and manuals. The population and housing censuses carried out by the various agencies and commissions given such authority by the respective country constitutions served as the basis for the sampling frames employed for the surveys. The enumeration areas (EAs) from the Census frames served as the foundation for defining the primary sampling units. The typical method for choosing DHS samples is a two-stage stratified design. At the second stage, the selection of the households was done once the numbers of clusters had been chosen from the list of EAs.

Model Fitting to Child-Mortality dataset

The under-five deaths dataset was fitted to the proposed model and other three zero-inflated models; 2GP, ZINB and ZIP and the RMSE and the absolute Bias of the models' parameters were compared in order to suggest a model of good-fit.

Table 22: The Mean and RMSE of the proposed models and the existing models to the child-mortality dataset

Weight	Parameter	Proposed Reduced Model	ZIG	ZINB	ZIP
	Mean	0.2213	0.2213	0.75	0.72
W = 0.1	RMSE	0.0265	0.028	0.0915	0.0419
	Mean	0.182	0.182	0.595	0.6167
W=0.25	RMSE	0.0369	0.0388	0.0825	0.0648
	Mean	0.121	0.121	0.426	0.381
W=0.5	RMSE	0.0246	0.0259	0.074	0.0535
	Mean	0.0625	0.0663	0.2033	0.1887
W=0.75	RMSE	0.0163	0.0177	0.0428	0.0406
	Mean	0.025	0.0253	0.076	0.0521
W=0.9	RMSE	0.01028	0.0103	0.0402	0.0251

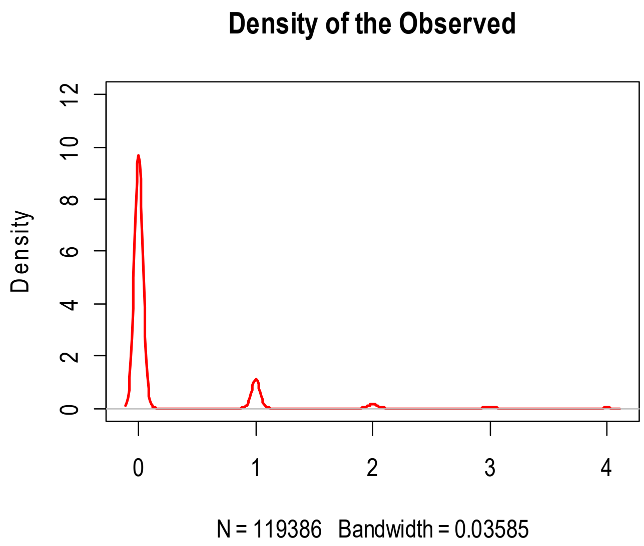
Table 22, presents the mean and the RMSE of the proposed reduced model , the ZIG, ZINB and ZIP The result indicated that at W=0.1 the mean of the proposed reduced model equal the mean of ZIG but different from the mean of the ZINB and ZIP. However, the RMSE of the proposed reduced model was lesser than the RMSEs of the ZIG, ZINB and ZIP. The result further showed that at W=0.25, the mean of the proposed reduced model equal the mean of the ZIG but different from the mean of ZINB and ZIP. The RMSE of the proposed reduced model was equally lesser than the RMSEs of the ZIG, ZINB and ZIP. At W=0.5, the mean of the proposed reduced model equal that of ZIG but also different from the other models, the RMSE of the proposed reduced was lesser that the rest models. Highlights of the result revealed that at W=0.75, the mean of the proposed reduced model approximate the mean of the ZIG but different from the mean of the rest models. However, the RMSE of the proposed reduced model was lesser than the RMSE of the existing models. Finally, at W=0.9, the mean of the proposed model was equal to the mean of ZIG while the RMSE of the proposed reduced model was lesser than the RMSEs of ZP, ZINB and ZIG. The deduction from these results is that the proposed reduced model outperformed the ZIG, ZINB and ZIP.

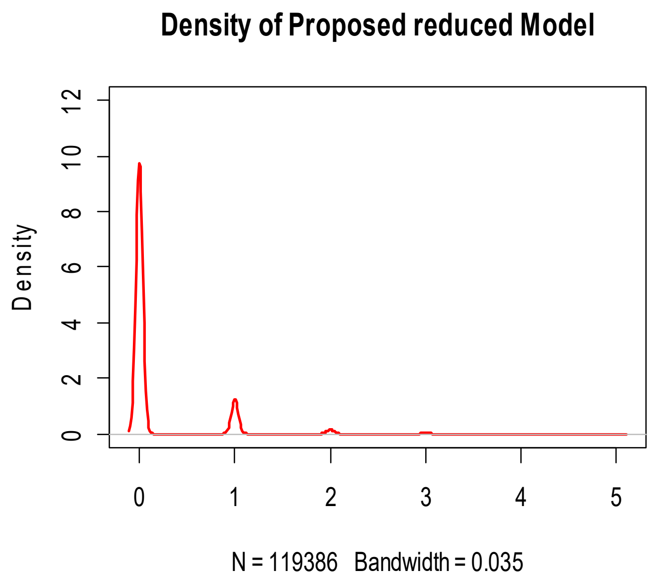
The Expected Frequencies of the models' Parameter

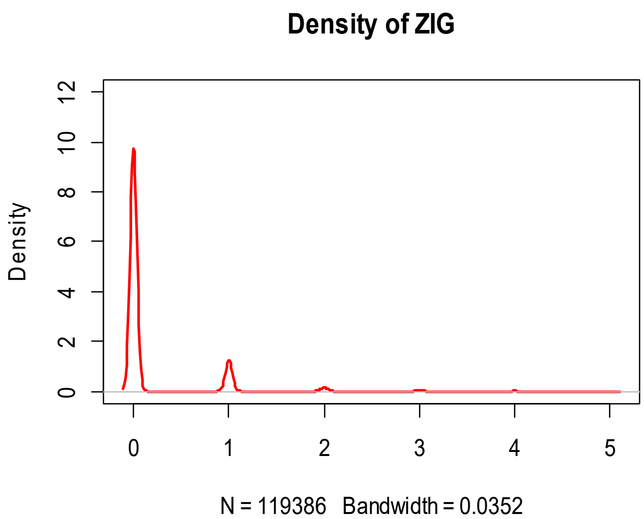
In this section we generate the expected frequencies from the fitted models; the proposed reduced model, ZIG, ZINB and ZIP. The expected frequencies are tabulated bellow:

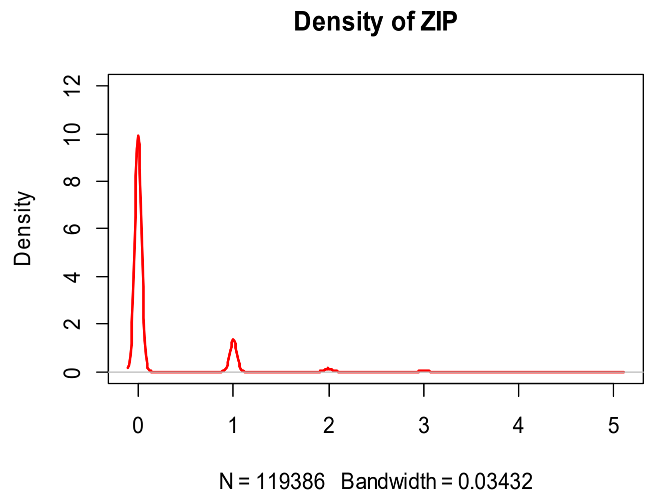
Table 23: Expected Frequencies of the Models

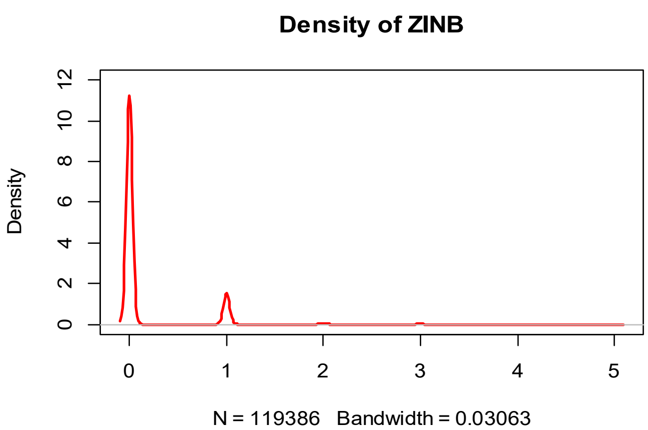
Model	0	1	2	3	4	5	Total	χ^2	P<0.05
Observed	105093	12141	1843	249	60	0	119386		
Proposed	104728	12841	1594	198	23	2	119386	152.985	0.00001
ZIG	104488	13036	1645	188	25	4	119386	161.575	0.00001
ZINB	104623	14251	421	83	7	1	119386	5851.853	0.00001
ZIP	103851	14051	1309	150	15	10	119386	702.669	


Density of the Observed
Density
0 2 4 6 8 10 12
0 1 2 3 4
N = 119386 Bandwidth = 0.03585


Density of Proposed reduced Model
Density
0 2 4 6 8 10 12
0 1 2 3 4 5
N = 119386 Bandwidth = 0.035


Density of ZIG
Density
0 2 4 6 8 10 12
0 1 2 3 4 5
N = 119386 Bandwidth = 0.0352


Density of ZIP
Density
0 2 4 6 8 10 12
0 1 2 3 4 5
N = 119386 Bandwidth = 0.03432


Density of ZINB
Density
0 2 4 6 8 10 12
0 1 2 3 4 5
N = 119386 Bandwidth = 0.03063

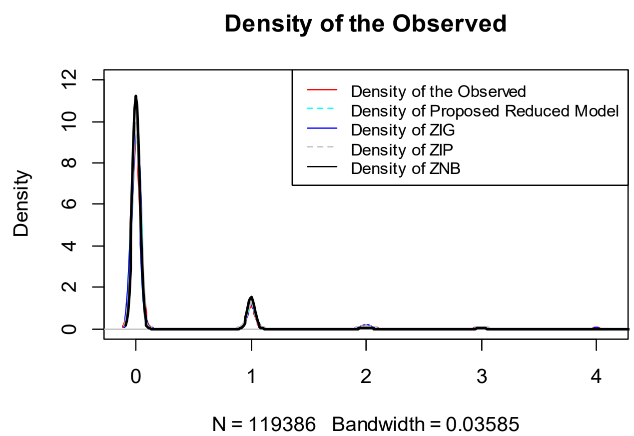

Density of the Observed
Density of the Observed
Density of Proposed Reduced Model
Density of ZIG
Density of ZIP
Density of ZNB
Density
0 2 4 6 8 10 12
0 1 2 3 4
N = 119386 Bandwidth = 0.03585

Table 24: Percentage Expected Predictions by the Models

Model	0	1	2	3	4	5	Total
Observed (%)	88.03	10.17	1.54	0.21	0.05	0	100
Proposed Reduced Model	87.72	10.76	1.34	0.17	0.02	0	100
ZIG	87.52	10.92	1.38	0.16	0.02	0	100
ZINB	87.63	11.94	1.41	0.25	0.02	0.01	100
ZIP	86.99	11.77	1.22	1.22	0.01	0.01	100

Table 23 and 24; present the expected frequencies and percentage predicted values by the proposed model reduced, 2GP, ZINB, and ZIP respectively. The total observed zeros was 105093 (88.03%) of the 119386 (100%) total number of child death, the proposed model predicted 104728 (87.72%) total zeros, 2GP predicted 104488 (87.52%) total zeros, ZINB predicted a total zero frequencies of 104623 (87.63%) and ZIP predicted total zero frequencies of 103851 (86.99%). The proposed 3GP model has the least Chi-square value of 152.985 with p=0.00001(p<0.05) is said to provide a better fit even though other models also fitted well into the data with each having probability value less than 0.05.

Absolute Bias of the Models Parameters

Table 25: Absolute Bias of the Models' Parameters

Parameters	Proposed Reduced Model	ZIP	ZINB	ZIG
Intercept	0.58	2.21	7.84	2.46
Child is alive	0.17	0.22	3.67	2.03
Mother's age	0.03	0.1	0.27	5.44
Region	0.05	0.07	0.06	2.03
Religion	0.36	0.23	0.18	4.26
Birth place	0.3	0.78	3.5	2.46
Highest education	0.06	0.45	0.82	2.03
Water Source	0.22	0.35	1.02	5.44
Wealth index	0.07	0.33	0.38	4.26
Total number of children	1.42	2.3	8.28	2.46
Visits	0.24	0.12	1.47	2.03
Smoking	0.12	0.08	1.28	5.44

Table 25, presents the absolute bias of the models' parameters. The results indicate that the absolute bias of the proposed reduced model was the least under the following parameters; intercept, child is alive, mothers' age, region, place of birth, mother's highest education, water source, wealth index total number of children in the household, number of visit during antenatal care and smoking behaviour of the mother. The results are showed that the proposed reduced model provides a better fit to the number of deaths per household within the study population.

DISCUSSION ON FINDINGS

The results generated from the simulation studies and real data from the proposed reduced model and the existing models provided great insight on the strength of the proposed reduced model at different level of parameters against the existing models; the ZIG, ZINB and ZIP. The results on the Root Mean Square Error show that the proposed reduced model is asymptotically consistent as sample size increases from 20 to 1000. The results also show that the proposed reduced model outperformed the ZIG, ZINB and ZIP at both zero-inflation and zero deflation and at some sample points. The Alkaike information criterion (AIC) and indicated that the proposed reduced model outperformed the ZIG, ZINB and ZIP at some sample points. At sample point 20 the proposed reduced model outperformed ZINB, however, at sample points 100 and above, the proposed reduced model outperformed ZIG, ZINB and ZIP. Similarly, the results from the Bayesian Information criterion (BIC) show that the proposed reduced model outperformed the ZINB at sample size 20 and 50. However, at sample size 100 and above the proposed reduced model outperformed ZIG, ZINB and ZIP.

The results from the mean absolute bias (MAB), at different levels of parameters indicated that the proposed reduced model outperformed the ZIG, ZINB and ZIP with least MAB when compared with ZIG, ZINB and ZIP at different sample points. The results from the analysis of the empirical data indicated that the proposed reduced model and ZIG have similar RMSE at different levels of zero-inflation parameters. The ZINB and ZIP also showed similar RMSE at different zero-inflation parameters. The results from the expected frequencies from the fitted distributions showed that the proposed reduced model gave the highest prediction of zeros closed to the observed frequency compared to ZIG, ZINB and ZIP. The chi-square value of the proposed model (χ^2=152.9815) was the minimum compared to chi-square values of ZIG, ZINB and ZIP. Similarly out of 88.03% zeros present in the original data, the proposed reduced model predicted 87.72%, ZIG predicted 87.52%, ZINB and ZIP predicted 87.63% and 86.99% respectively.

CONCLUSION

The new reduced model is a zero-inflated model and has proven to adequately fit into count data with excess zero which result in over-dispersion than some standard zero-inflated counts models. It also shown that at zero deflation, the model performed comparatively well alongside the standard count models. This model is recommended for both under-, moderately and over-dispersed count data and count data with over 90% excess zeros.

ACKNOWLEDGMENT

We acknowledge the contribution of Dr Laide Saka Balogun for her contribution. in making this research work possible. We also acknowledge ABUAD for providing favourable environment for research work.

CONFLICT OF INTEREST

No conflict of interest.

REFERENCE

Famoye, F., & Singh, K. (2006). Zero-inflated generalized Poisson regression model with an Application to Domestic Violence Data. Journal of Data Science. 4(2006), 117-130

Felix Famoye, John T. Wulu, Jr and Karan P. Singh (2004); On the Generalized Poisson Regression Model with an Application to Accident Data. Journal of Data Science 2

Francisco louzada, Fernaudo Moreira and Mauro Ribeiro (2017). A zero-inflated non-default rate regression model for credit scoring data. Communication in Statistics-Theory and methods DOI:10.1080/03610962.2017.1346803.

Frank badu Osei, Alfred Stein and Veronica Andreo (2022). A zero-inflated mixture spatially varying coefficient modeling of cholera incidences. Spatial Statistics. Volume 48, 100635. https://doi.org/10.1016/.spata.2022.100635

Idowu Adarabioyo and Reuben Ipinyomi (2020). Bayesian Modified Zero Hurdle and Zero-Inflated Models with Application to Under-Five Mortality in Nigeria. IEEE explore.wwwonacademic.com/detail/journal_1000049128719599_cfbd.html

Ma, Jiaqi and Goulias. K. G (1999). Application of Poisson Regression Models to Activity frequency Analysis and prediction. Journal of the Transportaion Research Board. Volume 1676, issue 1.

https://doi.org/10.3141/1676-11

Kazembe LN (2013) A Bayesian Two Part Model Applied to Analyze Risk Factors of Adult Mortality with Application to Data from Namibia. PLoS ONE 8(9): e73500. doi:10.1371/journal.pone.0073500

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics, 34,* 1-14.

Min, Y., Agresti, A. (2004). *Random effects models for repeated measures of zero-*

Modelling abundance of rare species: Statistical models for counts with extra zeros. *Ecological Modelling, 88,* 297-308.

Mullahy, J. (1986). Specification and testing of some modified count data models. of statistics, 30, 1-16

Partha Deb, Murat. K. Munkin and Pravin. K. Trivedi (2006). Bayesian Analysis of the Two-Part model with Endogeneity: Application to Health Care Expenditure; Journal of Applied Econometrics. 21:1081-1099

Peer Bllal Ahmead and Mohammad Kafeel Wani (2023). A new compound distribution and its application in over-dispersed count data. Annals of Data Science. DOI:10.1007/s40745-023-00478-0

P´erez S´anchez. J. M and G´omez–D´eniz. E (2015). Bayesian analysis of zero-inflated regression models:Journal of Statistical Planning and Science. (136) 1360-1375)

R Vani Lakshmi & V S Vaidyanathan (2016) Three-parameter gamma distribution: Estimation using likelihood, spacings and least squares approach, Journal of Statistics and Management Systems, 19:1, 37-53

Showkat Ahmad Dar, Anwar Hassan, Peer Bilal Ahmad and Bilal Ahmad Para (2022). A new compound probability model. Application to count data. The Philippine Statistician Vol 70 No. 2, pp. 11-21

Welsh, A. H., Cunningham, R. B., Donnelly, C. F., & Lindenmayer, D. B. (1996).

Modelling abundance of rare species; Statistical models for counts with extra zeros. *Ecological Modelling, 88,* 297-308.

Swarup *et. al.*(2014). Fitting truncated geometric distributions in large scale real world networks: Theoretical Computer Science (551):22-38

Zorn, C. J. W. (1996). Evaluating zero-inflated and hurdle Poisson specification. Midwest Political Science Association, 1-16.