International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 7, No. 2, pp. 65 - 75
https://doi.org/10.52968/28309288

On Zero inflated models with applications to maternal healthcare utilization

Uduak E. Bassey[1], Mary I. Akinyemi [2*], Kelechi F. Njoku [3]

1, 2*, 3. Department of Mathematics, University of Lagos, Nigeria.
*Corresponding author: makinyemi@unilag.edu.ng, uduakbasseya@gmail.com, njokukelechifrancisca@gmail.com

Article Info
Received: 8 August 2021 Revised: 18 February 2022
Accepted: 28 February 2022 Available online: 03 March 2022

Abstract

We consider the problem of modelling count data with excess zeros and over-dispersion which are commonly encountered in various disciplines that limit the use of traditional models for count outcomes. Our research work applies the Zero-inflated Poisson and Negative Binomial models in modelling Maternal Health Care (MHC) utilization in Nigeria, employing the Andersen's behavioural model to examine the effect of predisposing, enabling, and need factors on MHC utilization. The performance of these models are compared to the traditional Poisson and negative binomial models. The Vuong test and AIC suggests that the Zero-inflated Negative Binomial model provided the most significant improvement over traditional models for count outcomes.

Keywords: Zero-inflated, Count data, Vuong test, Anderson's behavioural model.
MSC2010: 62E15.

1 Introduction

The deaths of new-born babies in Nigeria represents a quarter of the total number of deaths of children under five. Similarly, a woman's chance of dying from pregnancy and childbirth in Nigeria is 1 in 13. The majority of these occur within the first week of life, mainly due to complications during pregnancy and delivery, reflecting the intimate link between new-born/mother survival and the utilization and quality of maternal care [5]. Many of these deaths can be prevented by coverage, quality health care services and the decision to utilize the health care services.

When healthcare utilization is measured by the number of hospital visits, such count data are typically very skewed and exhibit over-dispersion and a lot of zero count observations [8, 12]. Over-dispersion in count data occurs frequently due to excess zeroes, unexplained heterogeneity, and/or temporal dependency. Count data are routinely modelled using Poisson and Negative Binomial (NB) regression but zero-inflated models may be advantageous in this setting [9–11]. These models are mixture models in which the complete distribution of the outcome is represented by two separate components, a first part modelling the probability of excess zeros and a second part accounting for the non-excess zeros and non-zero counts [11].

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 7, No. 2, pp. 65 - 75
https://doi.org/10.52968/28309288

The concept of zero inflation is studied extensively in relation to count data with larger number of zero outcomes than expected. Zero inflated models have been applied in various ways to handle zero inflated data. For example, imajo, applied zero-inflated Poisson models to a micro-level data and made comparisons to existing panel data models for count data. They showed that separately controlling for whether outcomes are zero or positive in one of the two years does make a difference for counts with a layer number of zeros. [19] studied recreation demands and visitor characteristics of urban green spaces in the Sapporo city area in Northern Japan. Recreation demands for 21 large urban green spaces were estimated using a zero-inflated negative binomial model. [2] applied the zero inflated poisson model to fishing data. [1] modelled accident risk at road level of multiple road networks in multiple cities of the Valencian Community (Spain) using the ZINB model. [20], applied the zero inflated poisson factor model to microbiome read counts using data on oral infections, glucose intolerance and insulin resistance. Their model assumed that the microbiome counts follow a ZIP process with library size as offset and Poisson rates negatively related to the inflated zero occurrences. [3] provided and extensive study on the robustness of ZIP regression with varying zero inflated submodels, they proposed a more flexible alternative link function for the ZIP model. [4] applied a multivariate zero inflated endemic epidemic model to measles count data from 16 Germany states. They found that by extending the HHH, endemic-epidemic model using the Zero inflated model, they were able to capture seasonality and serial correlation which in turn improves probability forecasts.In the application of a Poisson regression model to examine spatial patterns in antenatal care (ANC) utilization in Nigeria by gaya, results revealed a significant difference in ANC between never-married and married-women respondents. Results also showed substantial spatial variation with a distinct north-south divide in ANC utilization. Mohebbi, explored disease mapping and regression with count data in the presence of over dispersion and spatial autocorrelation. The outcome suggested that modelling strategies based on the use of generalized Poisson and negative binomial with spatial autocorrelation worked well and provided a robust basis for inference. In 2016, gayaetal investigated the spatial distribution of antenatal care utilization in West Africa using a geo-additive zero-inflated count model and the results revealed a tie, transcending boundaries especially among regions of Mali, Niger and northern Nigeria where utilization remains persistently lower.

In Nigeria, maternal healthcare utilization is an important marker of access to and coverage of maternal health services. The level of utilization varies within a population, differing amongst various social groups or people with different behaviors [13]. The Andersen's Behavioral Model of Health Care Utilization, initially developed in the late 1960's, suggests that people's use of health services is a function of their predisposition to use services, factors which enable or impede use, and their need for care. In this model, use of services is defined as a function of 3 main elements: need, enabling, and predisposing factors. Need factors include the individual's perceived health care need and other indicators of their health status. Enabling factors include items such as the individual's income, health insurance status, and access to a source of regular care. Finally, predisposing factors include demographic variables, socioeconomic status, attitudes, and beliefs [14,15]. The rest of this article is structured as follows, Section 2 describes the methods employed in the research, Section 3 gives an extensive description of the data and data properties, Section 4 discusses the results of the analysis and Section 5 concludes.

2 Methods

Count response variables are non-normal responses hence the need for the Generalized linear models(GLM) which extend standard linear regression models to incorporate non-normal response distributions. GLM has three components viz: the random component, the linear predictor and the link function given as:

$$f(\lambda) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n \tag{2.1}$$

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 7, No. 2, pp. 65 - 75
https://doi.org/10.52968/28309288

Where $X_1, X_2, \cdots, X_n$ are explanatory variables, $\beta_i \quad, i = 0, 1, \cdots, n$ are the intercept and regression coefficients. λ is the link function. The random component of a GLM consists of a response variable y with independent observations $(y_1, \ldots, y_n)$. The conditional distribution of each y_i on a vector of regressors is a linear exponential family with probability density function
given by:

$$f(y; \lambda, \phi) = \exp\left\{\frac{y \;.\; \lambda - c(\lambda)}{\alpha(\phi)} + n(y, \phi)\right\}, \tag{2.2}$$

where λ is the canonical parameter or link function; $c(\lambda)$, the cummulant and $\alpha(\phi)$ is the scale parameter, set to one in discrete and count models and $n(y, \phi)$ is the normalization term. The exponential family of distributions include the Normal, poisson, Gamma, Binomial Negative Binomial etc. Common choices for the link function include, identity $(f(\lambda) = \lambda)$, log $(f(\lambda) = ln\lambda)$ and logit $(f(\lambda) = ln\frac{\lambda}{1-\lambda})$.

This paper considered standard and zero inflated versions of the Poisson and negative binomial models with link functions given as:

- Poisson $=$ log(λ): The link function here results in a log-linear relationship between mean and linear predictor. Recall that the variance in the Poisson model is identical to the mean, hence, the dispersion is fixed at $\phi = 1$.
- Negative binomial $=$ log(λ): Similar to the Poisson model, the dispersion is fixed at $\phi = 1$.

We considered four models for modeling the number of ANC visits made by a woman during her last pregnancy. These models are Poisson and Negative Binomial (which we refer to as the standard models), zero-inflated Poisson (ZIP) and zero-inflated Negative binomial (ZINB)(which we refer to as zero inflated models). We use the logit model for the zero-inflated portions of all zero-inflated models.

2.1 Zero Inflated Poisson (ZIP) Model

According to cameron, zero-inflated models can be viewed as finite mixture models with a degenerate distribution whose mass is concentrated at zero. Excess zeroes arise when the event of interest is not experienced by many of the study individuals. This model allows for over-dispersion assuming that there are two different types of individuals in the data: those who have a zero count with a probability of 1 (Always-0 group), and those who have counts predicted by the standard Poisson (Not always-0 group). An observed zero could be from either group, and if the zero is from the Always-0 group, it indicates that the observation is free from the probability of having a positive outcome ([24]).

For a response vector of counts $y_i, i = 1, \ldots, n$, y_i are independent; the probability p_i that an observation is in the Always-0 group is predicted by the characteristics of observation i, written as:

$$p_i = \mathcal{G}(z_i'\gamma)$$

where z_i is the vector of covariates and γ is the vector of coefficients of the logit model. Then the probability that observation i is in the Not always-0 group becomes $1 - p_i$. For observations in the Not always-0 group, their positive count outcome is predicted by the standard Poisson model, given as:

$$Pr(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, \cdots \tag{2.3}$$

where $\lambda_i = E(y_i) = e^{x_i'\beta}$ and x_i is also a vector of covariates and β the vector of regression parameters. Hence, the zero inflated Poisson distribution is given by:

$$Pr(Y_i = y_i \mid x_i, z_i) = \begin{cases} p_i + (1-p_i)e^{-\lambda_i}\;, & y_i = 0 \\ \\ (1-p_i)\dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, & y_i > 0 \end{cases} \tag{2.4}$$

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 7, No. 2, pp. 65 - 75
https://doi.org/10.52968/28309288

Where, $0 \leq p_i < 1$, $\lambda_i > 1$, with mean $E(Y_i) = \lambda_i(1-p_i)$, $V(Y_i) = \lambda_i(1-p_i)(1+p_i\lambda_i)$. The vectors of covariates x_i and z_i are such that the log link for λ_i and the logit link for p_i are:

$$\lambda_i = e^{x_i'\beta}, \qquad \left(\frac{p_i}{1-p_i}\right) = e^{z_i'\gamma} \qquad \text{respectively.} \tag{2.5}$$

The full set of parameters of β and γ can be estimated by Maximum Likelihood. However, with larger sample sizes and over dispersed data (i.e. variance much larger than the mean), the ZIP model could be found to fall short of expectations. The zero inflated negative binomial model has been recommended in this instance.

2.2 Zero Inflated Negative Binomial (ZINB) Model

The ZINB model is also a finite mixture model just like the ZIP model, but instead of using the poisson distribution, the negative binomial distribution is used. The Negative binomial distribution introduces a dispersion parameter(α) to a poisson pdf by including a gamma noise variable which has a mean of 1. The Poisson-gamma mixture (negative binomial) distribution that results is the negative binomial distribution(NB2 form) given as:

$$Pr(Y_i = y_i) = \binom{y_i + \frac{1}{\alpha} - 1}{\frac{1}{\alpha} - 1}\left(\frac{1}{1+\lambda_i\alpha}\right)^{\frac{1}{\alpha}}\left(\frac{\lambda_i\alpha}{1+\lambda_i\alpha}\right)^{y_i}, \qquad y_i = 0, 1, 2, \cdots \tag{2.6}$$

The Negative binomial model is generally appropriate for handling over-dispersion due to unobserved heterogeneity and temporal dependency, but may be inappropriate for over-dispersion resulting from excess zeroes. ZINB models are sometimes preferred because they allow for additional flexibility in the variance. Using Equation 2.6 we can express the ZINB as:

$$Pr(Y_i = y_i \mid x_i, z_i) = \begin{cases} p_i + (1-p_i)\left(\dfrac{1}{1+\lambda_i\alpha}\right)^{\frac{1}{\alpha}}, & y_i = 0 \\ \\ (1-p_i)\dbinom{y_i + \frac{1}{\alpha} - 1}{\frac{1}{\alpha} - 1}\left(\dfrac{1}{1+\lambda_i\alpha}\right)^{\frac{1}{\alpha}}\left(\dfrac{\lambda_i\alpha}{1+\lambda_i\alpha}\right)^{y_i}, & y_i > 0 \end{cases} \tag{2.7}$$

with mean $E(Y_i) = \lambda_i(1-p_i)$, $V(Y_i) = \lambda_i(1-p_i)(1+\alpha\lambda_i + p_i\lambda_i)$.

NB regressions reduce to Poisson regression in the limit as the dispersion parameter $\alpha \to 0$, and display overdispersion when $\alpha > 0$. Furthermore, ZINB regressions reduce to NB regressions when the probability that an observation is in the always 0 group $p_i = 0$ and reduce to ZIP regression in the limit as the overdispersion parameter $\alpha \to 0$. The variance of ZINB regressions exhibits overdispersion when $\alpha > 0$ and $p_i > 0$ allowing the models to be used for handling both zero-inflated and overdispersed count data. The link functions for ZINB regressions can also be written as in Equation 2.5. Finally, Maximum likelihood estimates can be also be obtained by maximizing the log likelihood.

2.3 Model Evaluation

We used several criteria to compare and evaluate the considered models. Nested models were tested using a likelihood ratio test ($\mathcal{LR}$). Since the NB model reduces to the Poisson when $\alpha = 0$, and as a result we can test if there is over-dispersion due to heterogeneity by testing if the dispersion parameter is necessary. We tested if the dispersion parameter was necessary by comparing the Poisson versus NB and ZIP versus ZINB. The $\mathcal{LR}$ test statistic is asymptotically distributed as a chi-squared random variable, with degrees of freedom equal to the difference in the number of parameters between the two models. The $\mathcal{LR}$ test statistic, which is frequently used is:

$$\mathcal{LR} = -2(\mathcal{L}_{reduced} - \mathcal{L}_{full}) \tag{2.8}$$

where, $\mathcal{L}_{reduced}$ and $\mathcal{L}_{full}$ are the log-likelihoods of the reduced and full models respectively.

green, points out that Poisson and ZIP are not nested. For the ZIP model to reduce to the Poisson, it is necessary for p_i to equal zero. This does not occur when $\gamma = 0$ since $p_i = \mathcal{G}(z_i'.0) = 0.5$. Similarly, the NB and the ZINB are not nested. Consequently, Greene proposes using a test by vuong; for non-nested models. This test considers two models, where $\widehat{Pr}_1(y_i|x_i)$ is the predicted probability of observing y in the first model and $\widehat{Pr}_2(y_i|x_i)$ is the predicted probability for the second model. Defining:

$$m_i = ln\left[\frac{\widehat{Pr}_1(y_i|x_i)}{\widehat{Pr}_2(y_i|x_i)}\right] \tag{2.9}$$

let $\overline{m}$ be the mean and let S_m be the standard deviation of m_i, then the Vuong statistic is given by:

$$\mathcal{V} = \frac{\overline{m}\sqrt{n}}{S_m} \tag{2.10}$$

The statistic $\mathcal{V}$ asymptotically follows a standard normal distribution. If $\mathcal{V} > 1.96$, the first model is favoured; if $\mathcal{V} < -1.96$ the second model is favoured. This means that a large, positive test statistic provides evidence of the superiority of model 1 over model 2, while a large, negative test statistic is evidence of the superiority of model 2 over model 1. Lastly, we used the Akaike (AIC) and Bayesian (BIC) information criteria to compare all models. Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) have been extensively used for model selection [25]. The AIC statistic is generally found in two forms:

$$AIC = \frac{-2\mathcal{L} + 2k}{n} \qquad \text{or} \qquad AIC = -2\mathcal{L} + 2k \tag{2.11}$$

where $\mathcal{L}$ is the model log-likelihood, k is the number of predictors including the intercept, and n represents the number of model observations. In both parameterizations $2k$ is referred to as a penalty term which adjusts for the size and complexity of the model [25]. The BIC statistic has also undergone a variety of parameterizations. The most commonly used is:

$$BIC = -2\mathcal{L} + kln(n) \tag{2.12}$$

with k indicating the number of predictors, including intercept, in the model. Given a set of candidate models for the data, the one with the lowest AIC and BIC is the preferred model.

3 Data

The data used for this paper was from the Nigeria Demographic and Health Survey (NDHS), implemented from 2008-2013 [6]. The survey was conducted by the National Population Commission (NPC) Federal Republic of Nigeria Abuja, Nigeria (Internet: www.population.gov.ng). ICF International provided financial and technical assistance for the survey through the USAID-funded MEASURE DHS program (Internet: www.dhsprogram.com), which is designed to assist developing countries to collect data on fertility, family planning, and maternal and child health. Financial support for the survey was provided by USAID, the United Kingdom Department for International Development (DFID) through PATHS2, and the United Nations Population Fund (UNFPA).

Authorization to extract the data was obtained from MEASURE DHS/ICF International by providing a brief description of the study through their website (Internet: www.dhsprogram.com).

According to Andersen's behavioral model ([15], [7]), healthcare utilization is a function of three major elements: predisposing factors (socio-demographic factors), enabling factors and need factors. The Predisposing factors selected for consideration from the data source include: age, level of education, marital status, religion, resident location, region of residence and number of children

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 7, No. 2, pp. 65 - 75
https://doi.org/10.52968/28309288

ever born. Wealth index, distance to health facility and health insurance was considered as Enabling factors, while visiting the health facility for any other reason in the year prior to the survey was selected as a proxy for the Need factor. The number of Antenatal care (ANC) visits made by a woman pertaining to the last birth was used as the outcome variable.

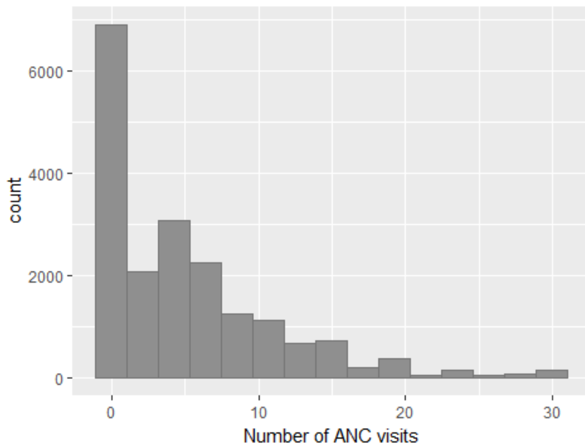

Figure 1: Histogram of Number of ANC Visits

Outcome # Variable	Factors	Counts	%
Number # of ANC # Visits	0	6529	33.9
	1 - 3	2456	12.7
	4 - 5	3070	15.9
	6 - 8	3119	16.2
	9 - 15	2668	3.8
	> 16	1438	7.2

Figure 2: Distribution of respondents by ANC Visits

Figure 1 shows a positively skewed distribution of the number of ANC visits, with a preponderance of zeros, this can also be observed from Figure 2 with 33% of the women making no ANC visits.

	Variable	Factors	Frequency	Percent(%)
Predisposing# factors	Age Group	15 - 24	4961	25.7
		25 - 34	9024	46.8
		35 - 49	5295	27.5
	Level of education	No education	8904	46.2
		Primary	3907	20.3
		Secondary	5217	27.1
		Higher	1252	6.5
	Marital Status	Never Married	508	2.6
		Married/Co-habiting	18161	94.2
		Previously# Married/Co-habiting	358	1.9
		Widowed	253	1.3
	Religion	Islam	11171	57.9
		Catholic	1540	8.0
		Protestant	6273	32.5
		Traditonalists/Others	296	1.5
	Resident Location	Rural	12887	66.8
		Urban	6393	33.2
	Region of Residence	North central	2935	15.2
		North east	3867	20.1
		North west	6063	31.4
		South South	2260	11.7
		South east	1618	8.4
		South west	2537	13.2
	Children ever Born	1 - 4	11909	61.8
		5 - 8	5910	30.7
		9 - 12	1395	7.2
		> 13	66	0.3
Enabling# factors	Wealth Index	Poor	8720	45.2
		Middle	3867	20.1
		Rich	6693	34.7
	Distance to# Health Facility	Big problem	6193	32.1
		Not a big problem	13087	67.9
	Health Insurance	Yes	334	1.7
		No	18946	98.3
Need# factors	Visited # health facility # in past year	Yes	5938	30.8
		No	13342	69.2

Table 1: Distribution of respondents by selected Variables

A total of 19280 women aged 15-49 whose information was complete on both explanatory variables and the outcome variable was selected out of 119,386 respondents. The distribution of the study population is presented in Table 1. About 47% of the women were aged 25-34 and other age groups were somewhat evenly distributed. 46% of the women had no education and less than 7%

had higher education. We observed that a greater percentage (>94%) of the women were married or co-habiting and about 61% had 1-4 children. More than 65% of the women were from northern regions and 57% were Muslim. Futhermore, we observed that a majority of the women live in rural areas (67%) and belonged to the poor wealth index (45%). More women (68%) indicated that distance was not a problem in their utilization of health services and a huge percentage (98%) had no health insurance. Finally, about 69% of the women visited a health facility in the 12 months period prior to the interview.

4 Results

We computed the mean and variance for the number of ANC visits. The observed mean and variance using all 19280 observations are 5.3 and 37.3, respectively. Our observed variance to mean ratio is 7.04, which indicates over-dispersion.

We fitted the Poisson and NB to all explanatory variables regardless of significance; ZIP and ZINB models to all explanatory variables regardless of significance in both parts of the models.

We compared the Poisson versus NB, ZIP versus ZINB to determine if the over-dispersion parameter was significant by testing if the dispersion parameter equals zero, using the likelihood ratio test of self. For all $\mathcal{LR}$ tests the dispersion parameter is highly significant (p-value $< 2.2e\text{-}16$), which provides strong evidence for preferring the NB and ZINB over the Poisson and ZIP models respectively. The AIC and BIC for all models are presented in Table 2. The ZINB model has the smallest AIC and BIC, it outperforms the other models. Additionally, the Vuong test also suggests that the ZINB model is a significant improvement over the other models. Our test results indicate that there is significant over-dispersion due to both heterogeneity and excess zeroes since NB models are preferred over Poisson models and zero-inflated models are preferred over the standard Poisson and NB models.

	AIC	**BIC**		
Poisson	119709.09	119897.89		
Negative binomial	95653.76	95850.43		
ZIP	92517.98	92893.61		
ZINB	85488.35	85883.47		
			Vuong # Z-Statistic	**Vuong # P-value**
Poisson vs NB			-42.6819	< 2.22e-16
Poisson vs ZIP			-59.6425	< 2.22e-16
NB vs ZIP			-8.4830	< 2.22e-16
ZIP vs ZINB			-26.1889	< 2.22e-16

Table 2: Poisson Model

Since the ZINB model outperforms the other models, we will only consider this model for subsequent results. To assess the goodness of fit, we compare the ZINB model to a null model (intercept only model) without predictors using a chi-squared test on the difference of log likelihoods. Results showed that the ZINB model fits the data statistically better than the null model ($p < 1e\text{-}16$).

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 7, No. 2, pp. 65 - 75
https://doi.org/10.52968/28309288

	Variable	IRR	95% C.I	P-value
	Intercept	5.727	5.212 - 6.294	< **2e-16**
Children ever Born		0.988	0.983 - 0.994	**0.00012**
Age group	25 - 34	1.047	1.017 - 1.077	**0.00178**
	35 - 49	1.123	1.08 - 1.168	**5.99e-09**
Level of education	No education	0.834	0.796 - 0.873	**4.92e-15**
	Primary	0.881	0.797 - 0.917	**6.12e-10**
	Secondary	0.954	0.921 - 0.988	**0.00896**
Marital Status	Married/Co-habiting	0.998	0.928 - 1.072	0.9474
	Never Married	0.918	0.838 - 1.005	0.06263
	Widowed	1.042	0.937 - 1.16	0.4471
Religion	Islam	1.069	1.025 - 1.160	**0.00173**
	Protestant	1.029	0.992 - 1.066	0.12312
	Traditionalists/Others	1.066	0.966 - 1.176	0.2050
Resident Location	Urban	1.003	0.979 - 1.027	0.8151
Region of Residence	North east	0.742	0.715 - 0.770	< **2e-16**
	North west	0.733	0.706 - 0.761	< **2e-16**
	South South	1.158	1.114 - 1.204	**1.16e-13**
	South east	1.335	1.282 - 1.390	< **2e-16**
	South west	1.884	1.823 - 1.947	< **2e-16**
Wealth Index	Middle	0.489	0.438 - 0.546	< **2e-16**
	Rich	0.288	0.249 - 0.334	< **2e-16**
Distance to# Health Facility	Not a big problem	1.143	1.114 - 1.173	< **2e-16**
Health Insurance	Yes	1.050	0.987 - 1.116	0.1221
Visited health facility # in past year	Yes	1.010	0.989 - 1.031	0.3407

Table 3: Results from the ZINB Model: **Count Model coefficients**

From Table 3, it can be seen that the predictors : the number of children ever born, age groups, levels of education, Islam, wealth index, region of residence, and distance to health facility were statistically significant for the negative binomial model part predicting the number of ANC visits.

The reference group in this case is previously married Catholic women, aged 15 - 24 with higher education, residing in rural parts of the North central region; with a poor living standard, having a big distance to health facility problem, no health insurance and never visited the health facility in the past year prior to the interview. These women made on average 6 (see Table 3) ANC visits during their last pregnancy with the remaining predictor values held constant. Holding all other variables constant, the effects of level of education show that women with no education make 17% fewer ANC visits, women with primary education 12% fewer ANC visits and women with secondary education make 5% fewer ANC visits than women with higher education. Also, women aged 25 - 34 make 5% more ANC visits and women aged 35 - 49 make 12% more ANC visits than women aged 15 - 24. The women belonging to the middle wealth index make 10% more ANC visits and the rich women make 20% more ANC visits than the poor women. Among those who have the opportunity to make ANC visits, not having a big distance to health facility problem increases the expected number of visits by 14% holding all other factors constant. Women from the South west region make 88% more ANC visits than the North central women with the remaining predictor values held constant.

Table 4, shows results from the zero inflated component of the ZINB model. It can be seen that the predictors : the number of children ever born, age groups, levels of education, wealth index, place of residence, distance to health facility etc. were statistically significant.

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 7, No. 2, pp. 65 - 75
https://doi.org/10.52968/28309288

	Variable	OR	95% C.I	P-value
	Intercept	0.134	0.061 - 0.297	**7.23e-07**
Children ever Born		1.052	1.029 - 1.075	**6.41e-06**
Age group	25 - 34	0.855	0.765 - 0.955	**0.00559**
	35 - 49	0.714	0.607 - 0.839	**4.53e-05**
Level of education	No education	17.039	8.323 - 34.8817	**8.69e-15**
	Primary	6.491	3.182 - 13.244	**2.73e-07**
	Secondary	4.305	2.114 - 8.768	**5.76e-05**
Marital Status	Married/Co-habiting	1.014	0.753 - 1.364	0.928
	Never Married	1.171	0.785 - 1.747	0.440
	Widowed	1.170	0.733 - 1.867	0.511
Religion	Islam	0.823	0.666 - 1.017	0.0711
	Protestant	0.798	0.653 - 0.976	**0.0283**
	Traditonalists/Others	1.260	0.891 - 1.782	0.19126
Resident Location	Urban	0.646	0.571 - 0.731	5.01e-12
Region of Residence	North east	0.904	0.781 - 1.048	0.18092
	North west	2.286	1.982 - 2.636	$<$ **2e-16**
	South South	3.948	3.301 - 4.721	$<$ **2e-16**
	South east	0.335	0.249 - 0.451	**6.59e-13**
	South west	0.790	0.638 - 0.979	**0.03092**
Wealth Index	Middle	0.489	0.438 - 0.546	$<$ **2e-16**
	Rich	0.288	0.249 - 0.334	$<$ **2e-16**
Distance to# Health Facility	Not a big problem	0.476	0.439 - 0.517	$<$ **2e-16**
Health Insurance	Yes	0.345	0.161 - 0.739	**0.00622**
Visited health facility # in past year	Yes	0.290	0.259 - 0.325	$<$ **2e-16**

Table 4: Results from the ZINB Model: **zero inflation Model coefficients**

Results from Table 4 show that the odds of membership in the "always zero" group (i.e., the group of women who would never make ANC visits) is estimated to be 17.04 higher in uneducated women than in highly educated women. The odds of membership in the group of women who would never make ANC visits is estimated to be 0.49 lower in the women belonging to the middle wealth index and 0.288 lower in the women belonging to the rich wealth index than in the women belonging to the poor wealth index. For each additional child born the odds of belonging to the no ANC group increases by 5.2%. The odds of membership in the "always zero" group is 66% lower for women who have health insurance than for women who do not. Also, the women who do not have a distance to health facility problem have some 0.48 times lower odds of not making ANC visits than does the women who do have a problem. Furthermore, the odds of belonging to the no ANC group is 35% lower for women residing in urban areas than for women who live in rural areas. The 25 - 34 year old women group has some 0.86 times lower odds of not making ANC visits than does the 15 - 24 year age group. Similarly, for the 35 - 49 age group their odds of not making ANC visits is 0.7 times lower than the 15 -24 age group.

5 Conclusion

This paper applied the Zero-inflated Poisson and Negative Binomial models in modelling Maternal Health Care (MHC) utilization in Nigeria. The data consisted of 19280 observations with observed variance to mean ratio computed as 7.04, indicative of over-dispersion. The model parameters were obtained by maximum likelihood and model selection was done using the AIC and BIC information criteria as well as the Vuong statistic. It was observed that the ZINB model outperformed the other models.

Furthermore, the analysis revealed the odds that a woman would never make ANC visits is estimated to be 17.04 higher in uneducated women than in highly educated women, while the odds that a woman would never make ANC visits is estimated to be 0.49 lower in women belonging to the middle wealth index and 0.288 lower in the women belonging to the rich wealth index than in women belonging to the poor wealth index. In addition, for each additional child born to a woman, the odds that she will not attend ANC increases by 5.2%. Further analysis revealed that the odds of not attending ANC is 66% lower for women who have health insurance than for women who do not. Also, the women who do not have to travel a distance to a health facility have 0.48 times lower odds of not making ANC visits than the women who do have a distance to travel. Finally, the odds of belonging to the no ANC group is 35% lower for women residing in urban areas than for women who live in rural areas. The odds of women 25 - 34 and 35-49 year attending ANC visits is 0.86 and 0.7 times lower respectively than 15 - 24 year old women.

Competing Financial Interests

The authors declare no competing financial interests.

References

[1] Álvaro Briz-Redón, Jorge Mateu & Francisco Montes Modeling accident risk at the road level through zero-inflated negative binomial models: A case study of multiple road networks. *Spatial Statistics.* **43**, 100503 (2021).

[2] Buu-Chau Truong, Kim-Hung Pho, Cong-Chanh Dinh & Michael Mcaleer Zero-Inflated Poisson Regression Models: Applications In The Sciences And Social Sciences. *Annals of Financial Economics.* **16**, 2150006 (2021).

[3] Essoham Ali A simulation-based study of ZIP regression with various zero-inflated submodels. *Communications in Statistics - Simulation and Computation.* **0**, 1–16 (2022).

[4] Junyi Lu & Sebastian Meyer A zero-inflated endemic-epidemic model with an application to measles time series in Germany. (2022).

[5] Government of Nigeria, Ministry of Health, UNICEF Nigeria, DFID, UNFPA, USAID & WHO Mother, Newborn and Child Health Mortality in Nigeria. *Government of Nigeria and Ministry of Health* (2008).

[6] National Population Commission (NPC) & ICF International Nigeria Demographic and Health Survey 2013. *NPC and ICF International, Abuja, Nigeria, and Rockville, Maryland, USA* (2014).

[7] Hirschman, Karen B. & Naylor, Mary D. Adapting Andersen's expanded behavioral model of health services use to include older adults receiving long-term services and supports. *BMC Geriatrics* **20**, 1471-–2318 (2020).

[8] Freund, D.A., Kniesner, T.J. & LoSasso, A.T. Dealing with the common econometric problems of count data with excess zeros, endogenous treatment effects, and attrition bias. *Economics Letters* **62**, 7-–12 (1999).

[9] Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–17 (1992).

[10] Cameron, A. C. & Trivedi, P. K. Regression analysis of count data. *Cambridge Press* (1998).

[11] Loeys, Tom, Moerkerke, Beatrijs, Olivia De Smet & Buysse, Ann The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression. *British Journal of Mathematical and Statistical Psychology* **65**, 163–180 (2012).

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 7, No. 2, pp. 65 - 75
https://doi.org/10.52968/28309288

[12] Wang, Peiming A bivariate zero-inflated negative binomial regression model for count data with excess zeros. *Economics Letters* **78**, 373–378 (2003).

[13] Babalola, Stella & Fatusi, Adesegun Determinants of use of maternal health services in Nigeria - looking beyond individual and household factors. *Economics Letters* (2009).

[14] Andersen, R. M. & Newman, J. F. Societal and individual determinants of medical care utilization in the United States. *The Milbank Memorial Fund Quarterly, Health and Society* **51**, 95–124 (1973).

[15] Andersen, R. M. Revisiting the behavioral model and access to medical care: does it matter. *Journal of Health and Social Behavior* **36**, 1–10 (1995).

[16] Andersen, R. M., Pamela, L. & Baumeister, S. E. Improving Access to Care in: Changing the U.S. Health Care System. *John Wiley and Sons* (2013).

[17] Gayawan, E. & Omolofe, O. T. Analyzing Spatial Distribution of Antenatal Care Utilization in West Africa Using a Geo-additive Zero-Inflated Count Model. *Spatial Demography* **4**, 245–262 (2016).

[18] Kimani, D. N., Mug, M. G. & Kioko, U. M. An Econometric Analysis of Health Care Utilization In Kenya. *European Scientific Journal* **12**, (2016).

[19] Andersen, R. M. Understanding recreation demands and visitor characteristics of urban green spaces: A use of the zero-inflated negative binomial model. *Urban Forestry & Urban Greening* **65**, 127–332 (2021).

[20] Xu, T., Demmer, R. T. & Li, G. Zero-inflated Poisson factor model with application to microbiome read counts. *Biometrics* **77**, 91–101 (2021).

[21] Majo, M. C. & Soest, A. H. A Fixed-Effects Zero-Inflated Poisson Model with an Application to Health Care Utilization. *CentER Discussion Paper* **83**, (2011).

[22] Mohebbi, M., Wolfe, R. & Forbes, A. Disease Mapping and Regression with Count Data in the Presence of Overdispersion and Spatial Autocorrelation: A Bayesian Model Averaging Approach. *International Journal of Environmental Research and Public Health* (2014).

[23] Gayawan, E. A Poisson Regression Model to Examine Spatial Patterns in Antenatal Care Utilization in Nigeria. *Population, Space and Place* **20**, 485–497 (2014).

[24] Long, J. S. Regression models for categorical and limited dependent variables. *Sage* **7**, (1997).

[25] Hilbe, J. M. Negative Binomial Regression. *Cambridge University Press* (2011).

[26] Greene, W. H. Some Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. *SSRN 1293115* (1994).

[27] Vuong, Q. H. Likelihood Ratio Tests for Model Selection and non-nested Hypotheses. *Econometrica* **57**, 307—333 (1989).

[28] Simple random talk Tilte of dataset. *http: simplerandomtalk. com/ data* (2017).

[29] Self, S. G. & Liang, K. Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *JASA* **82**, (1987).