International Journal of Health and Medical Sciences
volume 7 pp. 25-39 doi: https://dx.doi.org/10.20469/ijhms.7.30005

Fertility Determinants among Reproductive Age Women in Nigeria: Evidence from Some Modelling Techniques

Lawal Olumuyiwa Mashood *
Department of Statistics, Faculty of Science, Air Force Institute of Technology, Nigerian Air Force Base, Mando, Kaduna State, Nigeria

Abstract: This paper presented four parametric count distributions: Poisson (P), Negative Binomial (NB), Poisson Hurdle (PH) and Negative Binomial Hurdle (NBH) regression models. Data used was extracted from the 2018 National Demographic and Health Survey. The LRT, Vuong test, rootograms, Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) were used as goodness-of-fit and model selection measures. The objectives of this study were to examine the models for analyzing ideal number of children data exhibiting overdispersion, evaluate their performance and interpret the result of the best model selected that significantly assess some factors contributing to fertility preferences in Nigeria. It was revealed that Poisson-type (P and PH) models were more appropriate in handling of the overdispersion in the ideal number of children data than the NB-type (NB and NBH) models. The result further showed that there was no difference between the PH and NBH models ($Z = 0.2435$, $p = 0.4038$). According to both AIC and BIC of the four competing models, it shows that PH model provided a good fit to the ideal number of children data best than the other models (P, NB and NBH). The finding from this study was that mother's current age, age at first birth, age at first intercourse, place of residence, region of residence except South-West; middle wealth quintile category and Muslim women were found to be significant factors for mothers choosing no child and at least a child as the ideal number of children to have for their whole life in Nigeria.

Keywords: *Count data, overdispersion, National Demographic and Health Survey (NDHS), fertility, Vuong, hurdle model*

Received: 24 November 2020; ***Accepted:*** 02 February 2021; ***Published:*** 24 March 2021

I. INTRODUCTION

Most of data gathered by scientists, medical statisticians and economists is in the form of whole numbers which are all potentially response variables for study. Count data is the kind of data that is non-negative, discrete and skewed to the right. The number zero (0) in count data frequently occurs as a value in the response variable. This obliges meticulous chosen of parametric model to set of experimental observations and evaluation of the fitted selected model.

It is typically achievable to opt for model parameters in such a manner that the theoretical population mean of the model is roughly the same as the sample mean. When one or more significant factors have not been measured, then a problem of overdispersion arises. Overdispersion is the presence of more prominent variability (statistical dispersion) in a dataset than would be likely free-based on a specified simple statistical model. Whereas, underdispersion means that there was a lesser amount of discrepancy in the data than predicted. One may determine to employ Generalized Poisson (GP), a hurdle (e.g., Poisson-logit, or even NB-logit), or a Generalized Negative Binomial (GNB) model such as Waring NB (WNB) regression or Conway-Maxwell Generalized Pois-

*Correspondence concerning this article should be addressed to Lawal Olumuyiwa Mashood, Department of Statistics, Faculty of Science, Air Force Institute of Technology, Nigerian Air Force Base, Mando, Kaduna State, Nigeria. E-mail: maslaw008@gmail.com

son (COMPoisson). A number of researchers have proposed the use of Generalized Poisson and COM-Poisson regression models when there exist both over dispersed and under dispersed in the count data [1, 2, 3, 4]. Because overdispersion is frequent, numerous models have been formulated and applied, amongst which are: Negative Binomial, Quasi-Poisson and Zero-Inflated models; which they can equally deal with under dispersed count data. Relationships among some of the distributions can be bringing into being in [5] and [6].

If at hand there are extremely zeros in the response term, given the value of its mean, then, Zero-Inflated (ZI) mixture model is appropriate. [7] presented the general approach at anytime analyzing count data as to firstly carry out Poisson regression analysis in order to get foremost intuition way to look for the best fitted model. When modeling count data [8] suggested that robust (or sandwich) variance adjustment be used as a default for standard errors. [9] demonstrate the use of Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) models to data with extra zeros and portraits overdispersion. The researchers similarly provided different option which estimates the distributions with excess zeros, hurdle models. [10] describes the functions implemented in the statistical software R useful to implement ZIP, ZINB, and hurdle models. Suppose the data containing far-off zero counts that could be accommodated by the distributional assumptions of the Poisson or Negative Binomial models, and then a Zero-Inflated (ZI) set of models may possibly have to be considered. Hence, Poisson family, zero inflated and hurdle models have been applied into analyzing count data with excess zeroes and overdispersion that occurs when the sample variance exceeds the sample mean [11].

A. *Objectives*

The objectives of this study were to: examine the models for analyzing ideal number of children data exhibiting overdispersion; evaluate the performance of the competing models; select the best of the models significant for assessing some factors contributing to fertility preferences in Nigeria; and interpret the result of the best model.

II. MATERIALS AND METHOD

A. *Source of Data*

The data employed for this study were extracted from the 2018 NDHS, which is a nationally representative cross-sectional household survey that provides up-to-date information on demographic and health indicators and the sixth of its kind. The survey was conducted by the National Population Commission (NPC) with funding support from the US Agency for International Development (USAID), the United Nations Population Fund (UNFPA), and the United Kingdom Department for International Development (DFID). Technical support was provided by ICF International through measure DHS (https://www.dhsprogram.com/). The survey consists of all women age 15–49 in the sample households, who were either permanent residents of the selected households or visitors who stayed in the households the night before the survey were eligible to be interviewed. Two stage stratified design was used for sample selection in the survey consisting of 1,389 clusters. In the 2018 NDHS dataset, a total of 41,668 households were selected, of which 40,666 were engaged. Out of the engaged households, 40,427 were successfully interviewed, yielding a response rate of 99%. In the households interviewed, 42,121 women age 15–49 were identified for individual interviews; and interviews were completed with 41,821 women, yielding a response rate of 99% [12]. Therefore, the sample used for the data analysis was from 40,670 women of childbearing age (15–49 years) after deletion of the non-numeric responses and outliers from the ideal family size.

B. *Data Description*

1) Response variable: The dependent variable was the ideal family size (ideal number of children) counts ranged from 0 to 30. The mean and variance of the dependent variable are 6.1187 and 8.4311 respectively. Women without children were asked “If you could choose exactly the number of children to have in your whole life, how many would that be?” Women who had children were asked “If you could go back to the time when you did not have any children and could choose exactly the number of children to have in your whole life, how many would that be?”

C. *Explanatory variables:*

The explanatory variables of interest selected from the survey consist of fifteen variables that may have influence on the ideal family size, both continuous and categorical. Mother’s current age, mother’s Body Mass Index (BMI), mother’s age at first cohabitation, mother’s age at first intercourse and mother’s age at first birth are the continuous covariates. The categorical variables considered were: region of residence, defined as North East, North West, North Central, South East, South West and South-South; place of residence, is the designation of the EA as an urban area or a rural area; educational level of the mother, reported as the highest level of education attended (not necessarily completed) by the woman in categories of no education, primary, secondary, and higher than secondary; religion, is the religious group to which the woman associates herself classify as Catholic, Islam, other Christian, and traditionalist/others; mother's rank among partner's wives, categorize as first wife, second wife and third wife and above; wealth quintile, is a composite measure of a household's cumulative living standard define as lowest, second, middle, fourth, and highest; current marital status, is categorize as never in union, married, living together, widowed, divorced, not living together/separated; contraceptive use, divided into using and not using contraceptives; mother's working status, is whether the woman worked in the past 12 months (but not currently) or not working in the past 12 months as at the period of the interview; and mother's number of other wives, categorized as has co-wife and no co-wife.

These variables are akin with those considered in earlier researches [13, 14, 15, 16, 17].

D. Method

The models that are good candidate where the ratio of variance and mean is more than 1 are the negative binomial distribution, Poisson Quasi-likelihood, COM-Poisson and Zero-Inflated models. In this study, four different competing count models were fitted to the data: Poisson (P), Negative Binomial (NB), Poisson Hurdle (PH) and Negative Binomial Hurdle (NBH) regression models were employed in order to make inference on the count data (ideal family size/ideal number of children) using fifteen factors. The NB regression models solve the problem of overdispersion by including a dispersion parameter that relaxes the presumption of equal mean and variance in the distribution while the hurdle regression model was employed to address the distribution of count outcome with extra zeroes, see [18, 19, 20, 21].

The results were evaluated in threefold; the first of its kind is to compare the performance of these four statistical model distributions using NDHS data of 2018. The Goodness-of-fit of the four competing models was compared and tested using Likelihood Ratio Test (LRT) for the nested models (NB against P and NBH against PH), the Vuong test of non-nested models (NBH against P, PH against P, NBH against NB and PH against NB), rootograms and Akaike Information Criterion (AIC), Bayesian Schwartz Information Criteria (BIC). Secondly, the model validation techniques such as Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used for evaluating the out-of-sample performance of the four models. Lastly, the result of the best model selected based on the selection techniques at the modeling stage was interpreted in order to examine the relationship between ideal number of children and the selected covariates in Nigeria. Statistical software R [20] were used for data analyses [22] and modeling [10, 23, 24, 25, 26].

E. Model Selection, Validation and Evaluation

One important issue in statistical analysis is how to select the best model or model comparison among multiple competing models. When several models are available, one can compare the models' performance on the platform of several likelihood methods which have been anticipated in statistical literatures [27].

Two of the most regularly used measures which shall be employ in the course of this work are Akaike Information Criteria (AIC) and Bayesian Schwartz Information Criteria (BIC). AIC and BIC are equally penalized-likelihood criteria. However, the Poisson regression model underestimates the Standard Errors (SEs) when there is presence of overdispersion and this leads to inaccurate inference. A way out of this is to select between these models by comparing them based on a few criterion, such as AIC and BIC according to Lui [28]. The number of parameters in the model is represented by p and n is the sample size. Generally, the AIC or BIC for a model is written as: [$-2\times$logL+kp], where L is the likelihood function and k is 2 for AIC and Log(n) for BIC.

The model that is preferred is the one with the least Information Criteria (IC) value [29]. AIC and BIC are both roughly accurate in line with a different target and a different set of asymptotic assumptions. The assumptions have been criticized as impractical. Accepting the difference in the practical performance of both is easiest if one chew over the uncomplicated case of comparing two nested models. In that scenario, Information Criterions turn out to be the same as Likelihood Ratio Tests (LRT) with different alpha levels. Hence, the utilizing $-2\times$logL and the LRT for nested models [27], that is, NB vs. P and NBH vs. PH. In LRT, the null hypothesis is for the null model and the alternative hypothesis is for the full model as demonstrated by Hilbe [30].

The Vuong [31] test is an LR-based test for model selection which can be used to test between pairs of non-nested models (i.e., NBH vs. P, PH vs. P, NBH vs. NB, and PH vs. NB) and it helps in testing if the presence of overdispersion in count data is as a result of large numbers of zero counts. The null hypothesis tests that two models are indistinguishible against the alternative hypothesis that the two models are distinguishible. The Vuong non-nested test statistic is asymptotically distributed N(0, 1); the null hypothesis is rejected if P-value is lesser than 0.05 and conclude that the two models are distinct from each other. Similarly using critical value to test these non-nested models NB against PH at 95% confidence interval, when the Vuong value > +1.96 indicates that NB is the preferred model but if Vuong < –1.96 means that PH is the preferred model. Hilbe [30] shown that neither model is preferred over the other if –1.96 < Vuong < 1.96.

The dataset was divided into training and testing sets in the 80% and 20% respectively in order to investigate the out-of-sample performance of the four models using the following model validation techniques: The MSE, RMSE and MAE to examine the closeness of the predicted values to the observed values [32]. The four different competing models were visualized using rootograms, which has been established that it is very useful for diagnosing and treating count data with overdispersion and/or extra zeros. The hanging style of rootograms was used because it can emphasize the fitted values and the respective residuals [26].

III. DATA SUMMARY

The descriptive statistic of the selected variables was presented in Table 1, frequencies and percentages for categorical variables and means with Standard Deviations (SDs) for continuous variables of women characteristics. The Fig. 1 presents the bar plot of the ideal number of children of women age 15–49 years to actually reveal the skewness of the counts and its frequency table is presented in Table 2.

TABLE 1
FREQUENCY (PERCENTAGES IN PARENTHESIS) DISTRIBUTION OF ALL EXPLANATORY VARIABLES ACROSS IDEAL NUMBER OF CHILDREN

		Ideal Number of Children								
Covariates	Levels	0	1	2	3	4	5	6	7+	Total
Mother's current age	15 – 19	162 (0.39)	9 (0.02)	232 (0.55)	849 (2.03)	2237 (5.35)	1433 (3.43)	3293 (7.87)	208 (0.50)	8423 (20.14)
	20 – 24	112 (0.27)	5 (0.01)	128 (0.31)	616 (1.47)	1673 (4.00)	1115 (2.67)	3051 (7.30)	144 (0.34)	6844 (16.36)
	25 – 29	156 (0.37)	2 (0.00)	103 (0.25)	518 (1.24)	1639 (3.92)	1088 (2.60)	3536 (8.46)	161 (0.38)	7203 (17.22)
	30 – 34	145 (0.35)	4 (0.01)	105 (0.25)	391 (0.93)	1269 (3.03)	918 (2.20)	2997 (7.17)	168 (0.40)	5997 (14.34)
	35 – 39	123 (0.29)	5 (0.01)	91 (0.22)	323 (0.77)	1088 (2.60)	794 (1.90)	2806 (6.71)	176 (0.42)	5406 (12.93)
	40 – 44	142 (0.34)	3 (0.01)	56 (0.13)	179 (0.43)	680 (1.63)	550 (1.32)	2313 (5.53)	134 (0.32)	4057 (9.70)
	45 – 49	170 (0.41)	3 (0.01)	48 (0.11)	139 (0.33)	569 (1.36)	506 (1.21)	2296 (5.49)	160 (0.38)	3891 (9.30)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
	Mean (SD)	31.80 (10.41)	29.06 (10.38)	27.07 (9.70)	26.49 (8.91)	27.66 (9.27)	28.41 (9.54)	30.31 (9.78)	30.98 (10.24)	
Mother's age at first cohabitation	14 and below	265 (0.63)	7 (0.02)	25 (0.06)	86 (0.21)	476 (1.14)	595 (1.42)	4867 (11.64)	193 (0.46)	6514 (15.58)
	15 – 22	635 (1.52)	21 (0.05)	559 (1.34)	2228 (5.33)	6603 (15.79)	4656 (11.13)	13658 (32.66)	791 (1.89)	29151 (69.70)
	23 – 30	95 (0.23)	2 (0.00)	143 (0.34)	598 (1.43)	1768 (4.23)	1001 (2.39)	1544 (3.69)	148 (0.35)	5299 (12.67)
	31 – 38	12 (0.03)	1 (0.00)	33 (0.08)	97 (0.23)	276 (0.66)	141 (0.34)	193 (0.46)	17 (0.04)	770 (1.84)
	39 – 48	3 (0.01)	0 (0.00)	3 (0.01)	6 (0.01)	32 (0.08)	11 (0.03)	30 (0.07)	2 (0.00)	87 (0.21)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
	Mean (SD)	17.41 (4.29)	18.63 (5.37)	20.42 (4.52)	20.48 (4.27)	20.19 (4.43)	19.32 (4.33)	17.24 (3.97)	18.44 (4.33)	
Mother's age at first intercourse	Never had sex	165 (0.39)	7 (0.02)	231 (0.55)	792 (1.90)	1953 (4.67)	1166 (2.79)	2271 (5.44)	164 (0.39)	6749 (16.15)
	17 and below	658 (1.57)	18 (0.04)	225 (0.54)	829 (1.98)	3415 (8.17)	2877 (6.89)	13522 (32.36)	583 (1.40)	22127 (52.96)
	18 – 28	183 (0.44)	5 (0.01)	292 (0.70)	1354 (3.24)	3714 (8.89)	2321 (5.56)	4422 (10.58)	397 (0.95)	12688 (30.37)
	29 and above	2 (0.00)	1 (0.00)	15 (0.04)	38 (0.09)	68 (0.16)	36 (0.09)	50 (0.12)	7 (0.02)	217 (0.52)
	TOTAL	1008 (2.41)	31 (0.07)	763 (1.83)	3013 (7.21)	9150 (21.90)	6400 (15.32)	20265 (48.50)	1151 (2.75)	41781
	Median (IQR)	15 (3.0)	14 (4.5)	16 (20.0)	17 (20.0)	16 (5.0)	16 (4.0)	15 (3.0)	16 (4.0)	
Mother's age at first birth	15 and below	205 (0.49)	5 (0.01)	23 (0.05)	87 (0.21)	408 (0.98)	484 (1.16)	3188 (7.62)	148 (0.35)	4548 (10.87)
	16 – 20	623 (1.49)	19 (0.05)	519 (1.24)	1987 (4.75)	5648 (13.51)	4027 (9.63)	13275 (31.74)	707 (1.69)	26805 (64.09)
	21 – 25	136 (0.33)	4 (0.01)	112 (0.27)	517 (1.24)	1909 (4.56)	1294 (3.09)	2957 (7.07)	219 (0.52)	7148 (17.09)
	26 – 30	37 (0.09)	2 (0.00)	76 (0.18)	308 (0.74)	896 (2.14)	458 (1.10)	693 (1.66)	57 (0.14)	2527 (6.04)
	31 and above	9 (0.02)	1 (0.00)	33 (0.08)	116 (0.28)	294 (0.70)	141 (0.34)	179 (0.43)	20 (0.05)	793 (1.90)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
	Mean (SD)	18.61 (3.52)	19.66 (4.65)	21.27 (4.10)	21.22 (3.86)	20.98 (3.93)	20.23 (3.79)	18.72 (3.44)	19.43 (3.80)	
Mother's rank among partner's wives	1st wife	136 (1.57)	0 (0.00)	16 (0.18)	48 (0.55)	229 (2.64)	282 (3.25)	2728 (31.45)	134 (1.55)	3573 (41.20)
	2nd wife	139 (1.60)	1 (0.01)	36 (0.42)	100 (1.15)	407 (4.69)	466 (5.37)	3018 (34.80)	124 (1.43)	4291 (49.48)
	3rd wife and above	24 (0.28)	1 (0.01)	6 (0.07)	23 (0.27)	74 (0.85)	95 (1.10)	550 (6.34)	36 (0.42)	809 (9.32)
	TOTAL	299 (3.45)	2 (0.02)	58 (0.67)	171 (1.97)	710 (8.19)	843 (9.72)	6296 (72.59)	294 (3.39)	8673
	Mean (SD)	2.81 (5.24)	5.98 (17.08)	2.95 (3.46)	3.10 (4.90)	3.22 (6.13)	3.10 (5.72)	2.65 (3.61)	3.26 (7.96)	

TABLE 1 CONTINUE

		Ideal Number of Children								
Covariates	Levels	0	1	2	3	4	5	6	7+	Total
Mother's Body Mass Index	Under weight	59 (0.14)	0 (0.00)	40 (0.10)	113 (0.27)	333 (0.80)	271 (0.80)	1135 (0.65)	66 (0.16)	2017 (4.83)
	Healthy weight	227 (0.54)	10 (0.02)	166 (0.40)	649 (1.55)	1933 (4.62)	1361 (3.26)	4499 (10.76)	271 (0.65)	9116 (21.81)
	Overweight	691 (1.65)	21 (0.05)	529 (1.27)	2126 (5.09)	6464 (15.47)	4502 (10.77)	14154 (33.86)	778 (1.86)	29265 (70.02)
	Moderate obesity	20 (0.05)	0 (0.00)	16 (0.04)	76 (0.18)	280 (0.67)	171 (0.41)	335 (0.80)	22 (0.05)	920 (2.20)
	Severe obesity	7 (0.02)	0 (0.00)	6 (0.01)	32 (0.08)	103 (0.25)	64 (0.15)	115 (0.28)	11 (0.03)	338 (0.81)
	Very severe obesity	6 (0.01)	0 (0.00)	5 (0.01)	18 (0.04)	36 (0.09)	29 (0.07)	44 (0.11)	2 (0.00)	140 (0.33)
	TOTAL	1010 (2.42)	31 (0.07)	762 (1.82)	3014 (7.21)	9149 (21.89)	6398 (15.31)	20282 (48.53)	1150 (2.75)	41796
	Median (IQR)	25.41 (1.65)	25.41 (2.32)	25.41 (1.20)	25.41 (0.54)	25.41 (0.31)	25.41 (0.60)	25.41 (1.61)	25.41 (1.87)	
Place of residence	Urban	388 (0.93)	12 (0.03)	478 (1.14)	1852 (4.43)	4918 (11.76)	2953 (7.06)	5945 (14.22)	438 (1.05)	16984 (40.61)
	Rural	622 (1.49)	19 (0.05)	285 (0.68)	1163 (2.78)	4237 (10.13)	3451 (8.25)	14347 (34.31)	713 (1.70)	24837 (59.39)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
Region of residence	North Central	80 (0.19)	6 (0.01)	141 (0.34)	575 (1.37)	2012 (4.81)	1399 (3.35)	3307 (7.91)	252 (0.60)	7772 (18.58)
	North East	149 (0.36)	5 (0.01)	81 (0.19)	165 (0.39)	585 (1.40)	748 (1.79)	5502 (13.16)	404 (0.97)	7639 (18.27)
	North West	491 (1.17)	5 (0.01)	82 (0.20)	158 (0.38)	609 (1.46)	954 (2.28)	7727 (18.48)	103 (0.25)	10129 (24.22)
	South East	134 (0.32)	0 (0.00)	41 (0.10)	340 (0.81)	1775 (4.24)	1339 (3.20)	1895 (4.53)	47 (0.11)	5571 (13.32)
	South West	51 (0.12)	6 (0.01)	295 (0.71)	1222 (2.92)	2353 (5.63)	746 (1.78)	664 (1.59)	293 (0.70)	5630 (13.46)
	South-South	105 (0.25)	9 (0.02)	123 (0.29)	555 (1.33)	1821 (4.35)	1218 (2.91)	1197 (2.86)	52 (0.12)	5080 (12.15)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
Educational level	No education	503 (1.20)	11 (0.03)	76 (0.18)	167 (0.40)	826 (1.98)	1191 (2.85)	11182 (26.74)	442 (1.06)	14398 (34.43)
	Primary	189 (0.45)	6 (0.01)	75 (0.18)	236 (0.56)	1175 (2.81)	1141 (2.73)	3320 (7.94)	241 (0.58)	6383 (15.26)
	Secondary	266 (0.64)	13 (0.03)	407 (0.97)	1774 (4.24)	5496 (13.14)	3394 (8.12)	4944 (11.82)	404 (0.97)	16698 (39.93)
	Higher	52 (0.12)	1 (0.00)	205 (0.49)	838 (2.00)	1658 (3.98)	678 (1.62)	846 (2.02)	64 (0.15)	4342 (10.38)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
Religion	Islam	697 (1.67)	11 (0.03)	212 (0.51)	649 (1.55)	2182 (5.22)	2327 (5.56)	14244 (34.06)	637 (1.52)	20959 (50.12)
	Catholic	93 (0.22)	5 (0.01)	82 (0.20)	323 (0.77)	1456 (3.48)	1025 (2.45)	1393 (3.33)	59 (0.14)	4436 (10.61)
	Other Christian	214 (0.51)	15 (0.04)	461 (1.10)	2026 (4.84)	5397 (12.90)	2989 (7.15)	4518 (10.80)	450 (1.08)	16070 (38.43)
	Traditionalist/ Others	6 (0.01)	0 (0.00)	8 (0.02)	17 (0.04)	120 (0.29)	63 (0.15)	137 (0.33)	5 (0.01)	356 (0.85)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
Wealth quintile	Lowest	222 (0.53)	4 (0.01)	44 (0.11)	131 (0.31)	553 (1.32)	681 (1.63)	5871 (14.04)	241 (0.58)	7747 (18.52)
	Second	251 (0.60)	9 (0.02)	74 (0.18)	241 (0.58)	1102 (2.64)	1087 (2.60)	5290 (12.65)	292 (0.70)	8346 (19.96)
	Middle	259 (0.62)	6 (0.01)	122 (0.29)	483 (1.15)	1907 (4.56)	1547 (3.70)	4302 (10.29)	233 (0.56)	8859 (21.18)
	Fourth	171 (0.41)	7 (0.02)	193 (0.46)	825 (1.97)	2584 (6.18)	1722 (4.12)	3113 (7.44)	225 (0.54)	8840 (21.14)
	Highest	107 (0.26)	5 (0.01)	330 (0.79)	1335 (3.19)	3009 (7.19)	1367 (3.27)	1716 (4.10)	160 (0.38)	8029 (19.20)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
Current marital status	Never in union	200 (0.48)	12 (0.03)	388 (0.93)	1521 (3.64)	3519 (8.41)	1849 (4.42)	2943 (7.04)	237 (0.57)	10669 (25.51)
	Married	726 (1.74)	14 (0.03)	296 (0.71)	1193 (2.85)	4807 (11.49)	3951 (9.45)	16036 (38.34)	818 (1.96)	27841 (66.57)
	Living together	14 (0.03)	1 (0.00)	27 (0.06)	133 (0.32)	306 (0.73)	216 (0.52)	325 (0.78)	25 (0.06)	1047 (2.50)
	Widowed	45 (0.11)	1 (0.00)	18 (0.04)	51 (0.12)	221 (0.53)	184 (0.44)	572 (1.37)	25 (0.06)	1117 (2.67)
	Divorced	18 (0.04)	2 (0.00)	9 (0.02)	33 (0.08)	109 (0.26)	81 (0.19)	269 (0.64)	22 (0.05)	543 (1.30)
	Not living together/ Separated	7 (0.02)	1 (0.00)	25 (0.06)	84 (0.20)	193 (0.46)	123 (0.29)	147 (0.35)	24 (0.06)	604 (1.44)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
Contraceptive use	Not using	947 (2.26)	28 (0.07)	616 (1.47)	2411 (5.77)	7377 (17.64)	5311 (12.70)	18574 (44.41)	915 (2.19)	36179 (86.51)
	Using	63 (0.15)	3 (0.01)	147 (0.35)	604 (1.44)	1778 (4.25)	1093 (2.61)	1718 (4.11)	236 (0.56)	5642 (13.49)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
Mother's working status	Not working	573 (1.37)	9 (0.02)	298 (0.71)	1064 (2.54)	2942 (7.03)	2121 (5.07)	7407 (17.71)	352 (0.84)	14766 (35.31)
	Working	437 (1.04)	22 (0.05)	465 (1.11)	1951 (4.67)	6213 (14.86)	4283 (10.24)	12885 (30.81)	799 (1.91)	27055 (64.69)
	TOTAL	1010 (2.42)	31 (0.07)	763 (1.82)	3015 (7.21)	9155 (21.89)	6404 (15.31)	20292 (48.52)	1151 (2.75)	41821
Number of other wives	No co-wife	438 (1.05)	12 (0.03)	264 (0.63)	1147 (2.75)	4365 (10.46)	3301 (7.91)	10037 (24.06)	541 (1.30)	20105 (48.20)
	Has co-wife	569 (1.36)	18 (0.04)	498 (1.19)	1860 (4.46)	4752 (11.39)	3080 (7.38)	10227 (24.52)	602 (1.44)	21606 (51.80)
	TOTAL	1007 (2.41)	30 (0.07)	762 (1.83)	3007 (7.21)	9117 (21.86)	6381 (15.30)	20264 (48.58)	1143 (2.74)	41711

Table 1 depicted the frequency counts of each of the explanatory variables considered for this study. The ideal number of children was segregated into 0, 1, 2, 3, 4, 5, 6, 7 and more which were the thought of the reproductive age women in Nigeria if they could go back to the time when they did not have any children and could choose exactly the number of children to have in their whole life. The descriptive table was obtained before deletion of non-numeric response. From the six geo-political regions, almost half of the women (48.52%) thought 6 is the ideal number of children; whereas, 3809 (9.11%) stated that if they could go back to the time when they did not have any children and they could choose between a child and three children to have in their whole life. Contrarily, 2.42% choose not to have children in their whole life. It could be noticed that 61.07% of the reproductive age women were from the Northern region of the country, with most of the women (24.22%) from the North West. This could be because Northern states are predominantly practicing Islam with noticeable number of Christians in some northern regions and Muslims being minority in the southern states, see [33]. Islamic religion has the most percentage (50.12%), Catholic and other Christianity practitioners are with 49.03% and less than 400 of the women (0.85%) were followers of indigenous and other religions.

The age of individual woman plays a vital role; since there are reproductive period when individual is at her prime. Two thousand, three hundred and twelve (2,312) of the women in the survey were 25 years old; the most age, making up with the proportion of 5.56%. It was obvious from the Table 1 presented that, more than twenty percent of the women were in the age bracket (15 – 19 years); which is expected to comprise of women in their last year of secondary school or higher institution. Most especially, those in their late secondary or higher institution could have started thinking about the specific number of children to have in their whole life; this was evident from their percentages respectively obtained as 39.93% and 10.38%. This is more reason why more than half (64.09%) of them had their first birth between the age of 16 and 20 years. Women between the age of 45 and 49 years being the least amongst the age brackets with 9.30% were supposed to have finished bearing all the children they planned to have in their life, except otherwise, if they had one or two delays and getting closer to the menopause (shown that close to just 800 (1.90%) of the women had their first child at the age of at least 31 year). The women in the following age groups 20–24, 25–29, 30-34 and 35-39 years had 16.36%, 17.22%, 14.34% and 9.70% respectively. Women between 15 and 22 years had the most entry age for marriage; thereby couples are expected to dialogue and reach compromise on the number of children ideal for them to have. This further cements reason why the current age of the women were mostly between 15 and 19 years and slightly above half of the women (52.96%) had their first intercourse at the age of at most 17 year. Though, 16.15% of them never had sexual intercourse.

It is no doubt that women that reside in the rural area will have the highest percentage of the ideal number of children of (59.39%) due to the fact that they are majorly farmers because of that they need more hands, whilst their counterpart in the urban had (40.61%) having a considerable size family because of their occupation, educational level or somewhat. Married women had the most percentage (66.57%) because the choice of number of children to have is dominantly and rest upon the marriage and not when ones still single, whereas, 25.51% of them were never in union. About 50% of the married women were second fiddle and 41.20% were the first wife amongst their husband's wives. Not exceeding fifty-two percent have one or more rivalry that share husband, whereas 48.20% of the women have no co-wife. More than thirty-six thousand (86.51%) of the women were not using any contraceptive and nearly sixty-five percent of the women were working in the past 12 months (but not currently) or not working in the past 12 months as at the period of the interview. Though, the highest proportion of the women (21.18%) was in the middle category of a household's cumulative living standard. Most of the women were overweight with the proportion of 70.02%, whereas 21.81% had healthy weights but 3.34% were obese across the three classes: moderate, severe and very severe obesity.

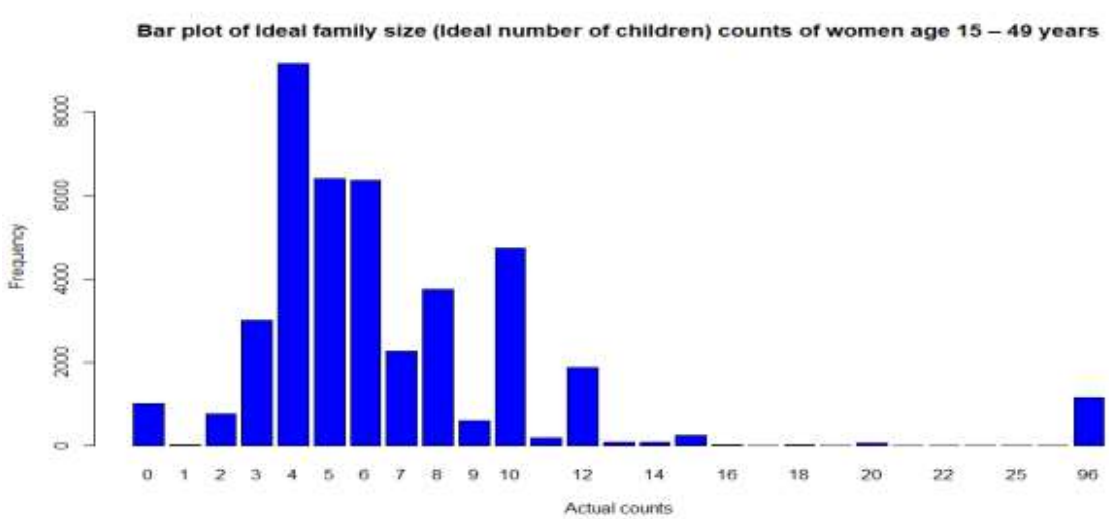

Fig. 1. The bar plot of Ideal family size counts of women age 15–49 years

TABLE 2
THE FREQUENCIES COUNT OF THE IDEAL NUMBER OF CHILDREN

Actual Count	Frequency	Proportion	Cumulative Proportion
0	1010	0.02415	0.02415
1	31	0.00074	0.02489
2	763	0.01824	0.04314
3	3015	0.07209	0.11523
4	9155	0.21891	0.33414
5	6404	0.15313	0.48727
6	6356	0.15198	0.63925
7	2261	0.05406	0.69331
8	3738	0.08938	0.78269
9	590	0.01411	0.79680
10	4744	0.11344	0.91024
11	188	0.00450	0.91473
12	1876	0.04486	0.95959
13	82	0.00196	0.96155
14	75	0.00179	0.96334
15	258	0.00617	0.96951
16	17	0.00041	0.96992
17	5	0.00012	0.97004
18	15	0.00036	0.97040
19	5	0.00012	0.97052
20	72	0.00172	0.97224
21	1	0.00002	0.97226
22	1	0.00002	0.97229
24	2	0.00005	0.97233
25	2	0.00005	0.97238
30	4	0.00010	0.97248
Non-numeric	1151	0.02752	1.00000
TOTAL	41821		

Fig. 1 presents the bar plot of ideal number of children counts of women age 15-49 years. The frequency table is presented in Table 2. The distribution of the data is highly positively skewed [mean (6.1187) > median (5) > mode (4)], with a high spike on the left and a long tail on the right. Most subjects (9,155 of them) thought that the ideal number of children to have is 4; 31 respondents were of the view that a child is the ideal number in a family size; and among the respondents, 1010 (2.42%) believed that remaining childless is their own stance which might be as a result of their infecundity or other factors, which shows that there are minimal zeros count in the data; but 4 of them had thirty as the ideal number of children is the way and 1,151 (2.75%) of the women gave a non-numeric responses. Whereas, 11.5% of the respondent were on the side of having at most three (3) children is the ideal situation and the rest percent (88.5%) were on the side of having more than three children is the ideal number.

Now, the plot of observed frequency and the predicted Poisson frequency are respectively plotted on the left-hand panel and the right-hand panel as depicted in Fig. 2 below:

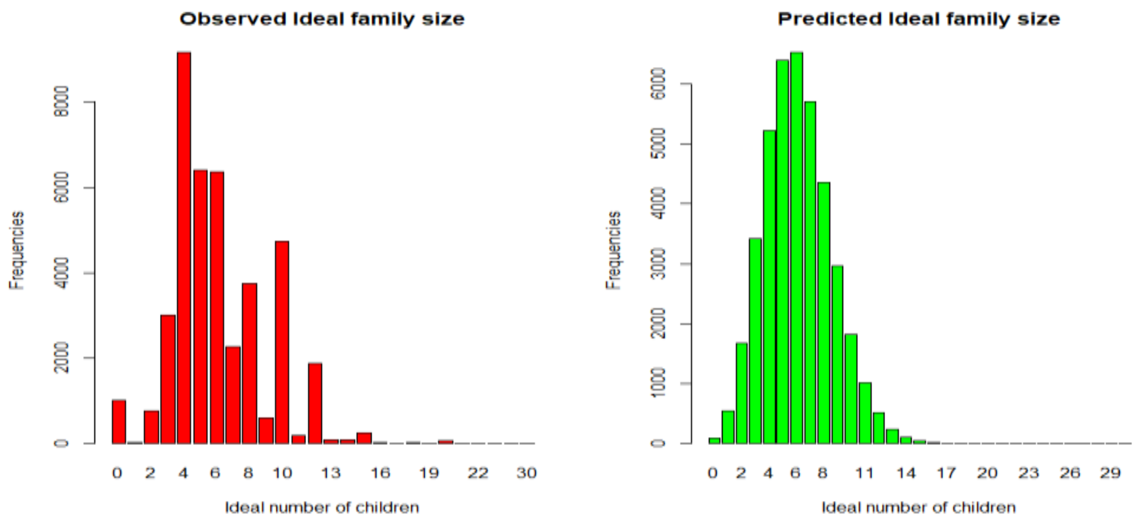

Fig. 2. The plots showing the observed frequency and the poisson predicted frequency of Ideal family size

The distributions are very different: the mode of the observed data is 4, but the mode of the predicted poisson with the same mean is 6. The observed ideal family size is highly aggregated; have a variance-mean ratio (1.3779) greater than 1; this suggests the presence of an overdispersion in the data since the conditional variance is much greater than its conditional mean, with mean (6.1187), variance (8.4311), median (5) and interquartile range (4). The observed data contained significantly 18, 19, and 20 ideal numbers of children, but these would be highly unlikely under a Poisson process. Negative binomial distribution has been recommended as substitutes to Poisson regression when there exist over-dispersed in the data [34] and according to Osgood [35].

IV. RESULTS

All the four competing models: Poisson (P), Negative Binomial (NB), Negative Binomial Hurdle (NBH), and Poisson Hurdle (PH) were fitted to the data using fifteen covariates, evaluated and interpretation made based on the best model selected at the modeling stage.

A. *Model Fitting and Selection*

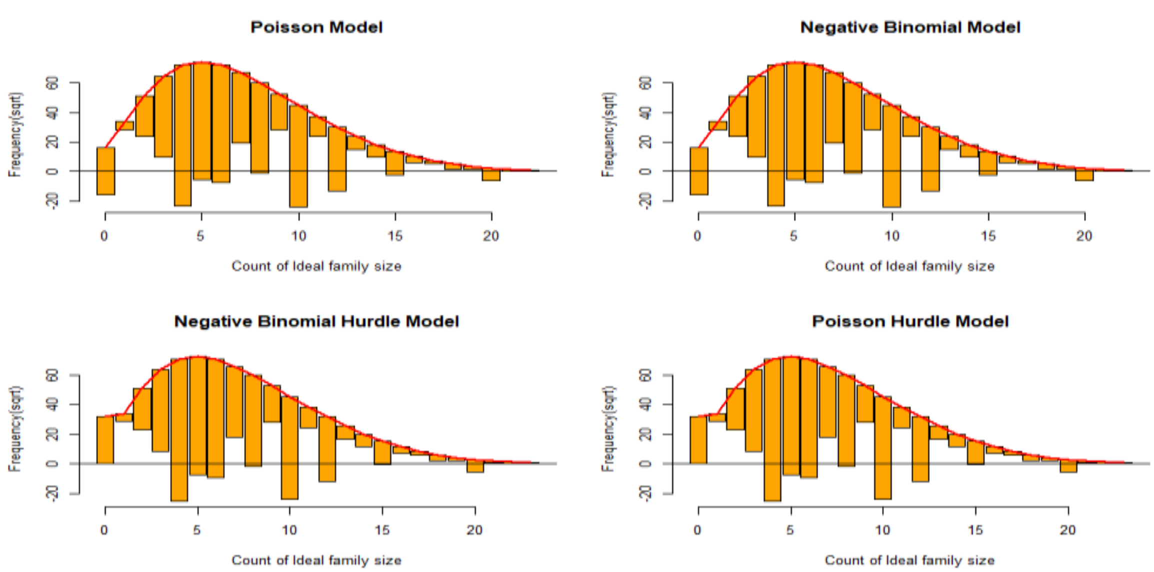

Fig. 3. The rootograms of all the four competing count models

The Fig. 3 is the hanging-style rootograms used for comparing the models visually. The closer the bars are to the horizontal axis, the better the model is and the red curved line is the theoretical model fit. A bar hanging below zero (0) shows underfitting whereas, a bar hanging above 0 shows overfitting. [5] discovered that models fit to count data using generalized Poisson and negative binomial distributions are always similar. It is obvious from the graphs that both Poisson and Negative binomial regression models show similar behavior and underfit the zeros count but Negative binomial hurdle and Poisson hurdle perfectly fit the zero count. However, it can be inferred that are evidence of some underestimate and overestimate since there exist underfitting and overfitting in the remaining positive counts. Both models (NBH and PH) show perfect fit for the counts 0, 8, and 15.

TABLE 3
LRT AND VUONG TEST RESULTS OF COMPETING MODELS

LRT		Vuong Non-nested Test	
Nested Models	Chi-square (χ^2)	Non Nested Models	z-statistic
P vs. NB		P vs. NBH	–22.9844 ($p < 0.01$)
		P vs. PH	–22.9844 ($p < 0.01$)
PH vs. NBH	0.0012 ($p = 0.9722$)	NB vs. NBH	–22.9871 ($p < 0.01$)
		NB vs. PH	–22.9871 ($p < 0.01$)
		PH vs. NBH	0.2435 ($p = 0.4038$)

Table 3 presents the LRT and Vuong test results of competing models. The Vuong test was used to make comparison between pairs of non-nested models. If the two models do not differ, the test statistic for Vuong would be asymptotically standard Normal and the p-value would be relatively large. The following pairs of non-nested models were compared: NBH against P, PH against P, NB against NBH and NB against PH; to test whether the overdispersion in the ideal number of children data was attributed to extra (excess) zeros. Further investigation was carried out to identify the structure (between both structural and sampling zeros, and only structural zeros) of excess zeros by comparing PH against NBH. Here, the output indicates that the hurdle models (Poisson hurdle and Negative binomial hurdle) are both better than P ($Z = –22.9844$, $p < 0.01$) and NB ($Z = –22.9871$, $p < 0.01$). The result further revealed that there was no difference between the PH and NBH models ($Z = 0.2435$, $p = 0.4038$). This shows that both Poisson hurdle and Negative binomial hurdle models handled the excessive zeros equally without making justification for overdispersion but Poisson hurdle was preferred since the test statistic is positive.

The nested models (NB against P and NBH against PH) were compared and tested whether the overdispersion parameter in the NB-type models was necessary using LRT. Neither of the LRTs ($\chi^2 = 0.3434$, $p = 0.5579$) for NB against P nor ($\chi^2 = 0.0012$, p = 0.9722) for PH against NBH was significant, it was revealed that Poisson-type models (P and PH) were more appropriate in handling of the overdispersion in the ideal number of children data than the NB-type models (NB and NBH).

TABLE 4
THE AIC, BIC AND LOG LIKELIHOOD RESULTS OF COMPETING MODELS

Competing Models	AIC	BIC	Log Likelihood
NB	179915.1572	180190.7007	–89925.5786
P	179912.8137	180179.7465	–89925.4069
NBH	173634.1329	174176.6091	–86754.0664
PH	173632.1317*	174165.9972*	–86754.0658*

*best fit model

Table 4 summarizes the AIC, BIC and log likelihood for the four competing models. According to both AIC and BIC of the four models, the Poisson hurdle models fitted the ideal number of children data best than the other models (Negative binomial, Poisson and Negative binomial hurdle), because the PH model had the smallest

AIC and BIC, while the Negative binomial model had the largest Information Criterion (IC). However, as a result of that, together with the Fig. 3 and Table 3; it is thereby suggested that Poisson hurdle regression model should be considered when modeling ideal number of children, that is, fertility preference data in Nigeria.

B. Model Validation

TABLE 5
FREQUENCY (PERCENTAGES IN PARENTHESIS) DISTRIBUTION OF ALL EXPLANATORY VARIABLES ACROSS IDEAL NUMBER OF CHILDREN

Competing Models	MSE	RMSE	MAE
NB	40370.3915	200.9239	1.5605
P	40370.3939	200.9239	1.5605
NBH	40190.5859	200.4759	1.5601
PH	40190.4944*	200.4757*	1.5601*

*outperformed model

The dataset was divided into training and testing sets in the ratio 4:1 respectively; meaning, large portion of the dataset randomly selected was used to estimate the parameters of the models 32,536 (80% of the data) while the left portion of the dataset used for model evaluation 8,134 (20% of the data). This was done so that the out-of-sample performance of the four models would be investigated by employing validation measures such as MSE, RSME and MAE. Table 5 presents the result of the model validation. According to the result, Poisson hurdle has the least MSE, RMSE and MAE values other than the rest models. However, both P and NB were approximately the same. Hence, PH model outperformed the rest models.

C. Model Parameters Interpretation

TABLE 6
RESULTS OF POISSON HURDLE REGRESSION MODEL

Variables	Count Part		Logit Part	
	Estimates	OR (CI)	Estimates	OR (CI)
Intercept	2.2017	9.0399* (8.4863, 9.6297)	5.0411	154.6395* (51.6467, 463.0188)
Mother's current age	0.0086	1.0087* (1.0082, 1.0092)	–0.0421	0.9588* (0.9515, 0.9661)
Mother's Body Mass Index (BMI)	–0.0012	0.9988 (0.9976, 1.0001)	0.0025	1.0025 (0.9836, 1.0218)
Mother's age at first birth	–0.0076	0.9924* (0.9908, 0.9940)	0.0443	1.0453* (1.0201, 1.0712)
Mother's rank among partner's wives	–0.0503	0.9509* (0.9430, 0.9589)	–0.0471	0.9540 (0.8411, 1.0820)
Mother's age at first intercourse	–0.0009	0.9991* (0.9984, 0.9998)	0.0363	1.0369* (1.0231, 1.0509)
Mother's age at first cohabitation	–0.0041	0.9959* (0.9945, 0.9973)	–0.0082	0.9918 (0.9720, 1.0121)
Place of residence				
Rural (ref.)		1.0000		1.0000
Urban	–0.0254	0.9749* (0.9650, 0.9849)	–0.1787	0.8364* (0.7165, 0.9763)
Region of residence				
North-East (ref.)		1.0000		1.0000
North-Central	–0.2217	0.8012* (0.7902, 0.8123)	0.3986	1.4898* (1.1162, 1.9883)
North-West	–0.0264	0.9739* (0.9632, 0.9847)	–0.7737	0.4613* (0.3822, 0.5568)
South-East	–0.1471	0.8632* (0.8467, 0.8799)	–0.8296	0.4362* (0.3070, 0.6200)
South-South	–0.2615	0.7699* (0.7553, 0.7848)	–0.7871	0.4552* (0.3224, 0.6426)
South-West	–0.4341	0.6479* (0.6361, 0.6599)	0.2125	1.2368 (0.8701, 1.7579)
Educational level				
No Education (ref.)		1.0000		1.0000
Primary Education	–0.0534	0.9480* (0.9357, 0.9605)	–0.1567	0.8550 (0.7040, 1.0384)
Secondary Education	–0.1170	0.8896* (0.8778, 0.9015)	0.1602	1.1737 (0.9440, 1.4593)
Higher Education	–0.1921	0.8252* (0.8087, 0.8421)	0.2466	1.2797 (0.8997, 1.8203)
Religion				
Traditionalist/Others (ref.)		1.0000		1.0000
Catholic	–0.0893	0.9146* (0.8727, 0.9584)	–0.5391	0.5833 (0.2491, 1.3657)
Islam	0.0888	1.0929* (1.0444, 1.1436)	–1.0830	0.3386* (0.1460, 0.7850)
Other Christian	–0.0747	0.9280* (0.8872, 0.9706)	–0.2641	0.7679 (0.3365, 1.7525)
Wealth quintile				
Lowest (ref.)		1.0000		1.0000
Second	–0.0080	0.9921 (0.9804, 1.0039)	–0.1805	0.8348 (0.6910, 1.0087)
Middle	–0.0362	0.9645* (0.9518, 0.9773)	–0.3050	0.7371* (0.6006, 0.9048)
Fourth	–0.0752	0.9276* (0.9135, 0.9419)	–0.0042	0.9958 (0.7796, 1.2720)
Highest	–0.1220	0.8851* (0.8692, 0.9013)	0.2029	1.2249 (0.9068, 1.6547)
Contraceptive use				
Not using (ref.)		1.0000		1.0000
Using	–0.0415	0.9594* (0.9466, 0.9724)	0.4424	1.5564 (1.1908, 2.0343)
Number of other wives				
Has co-wife (ref.)		1.0000		1.0000
No co-wife	0.0559	1.0575* (1.0464, 1.0688)	0.0619	1.0638 (0.8944, 1.2653)

*significant at 0.05 alpha level

Having selected Poisson hurdle regression as the best model for fitting fertility preference count data; Table 6 presents the significant factors after employing backward selection technique, by starting from the fullest model and systematically removes terms that do not result in a statistically significant. This model allows an interpretation of different independent variables for those women who declared that; if they could go back to the time when they did not have any children and could choose exactly the number of children to have in their whole life, then they chose not to have a child "never exposed" (logit part) and those women who chose to have at least a child are "exposed" (count part). The logit part is to predict the probability of when the ideal number of children is zero (that is, no child) given the independent variables. While, the truncated Poisson (count part) is to predict the positive count of the ideal number of children given those same independent variables.

According to the result presented in Table 6 of the count part of the Poisson hurdle regression model, mother's current age, age at first birth, rank among partner's wives, age at first intercourse, age at first cohabitation, place of residence, region of residence, educational level, religion, at least middle wealth quintile category, contraceptive use and number of other wives were factors associated with the ideal number of children. The model is 2.2017 times more predicted one or more fertility preference (ideal number of children) for reproductive age women with all factors considered.

The odds of women stating that one or more child is the ideal number of children is 0.87% more likely than the mother's current age with the Odd Ratio (OR) = 1.0087; and Credible Interval (CI) = (1.0082, 1.0092), but almost 100% lesser than the mother's BMI, age at first birth, rank among partner's wives, age at first intercourse and age at first cohabitation. Women who reside in urban and not exposed to one or more childbearing were 97% less likely than those that reside in rural with the CR (OR = 0.9749; 0.9650, 0.9849). Women from the North-East region of the country were nearly 22%, 3%, 15%, 26% and 43% less likely to be exposed to fertility preference of one or more children than those from the North-Central, North-West, South-East, South-South and South-West respectively. The odds of women with primary, secondary and higher education and exposed to at least a child were respectively nearly 95%, 89% and 83% smaller than those without formal education. Mothers who were practicing traditional/other religions were 9.29% more likely to state that they could choose not to have at least a child than those who practice Islam with CR (OR = 1.0929: 1.0444, 1.1436) but 91% and 93% less likely than those who were Catholic and other Christians. Women in the highest category of wealth quintile were 89% less likely exposed to one or more child than those in the lowest category, OR = 0.8851; CI = (0.8692, 0.9013). Also, lowest category women were 99%, 96% and 93% respectively more likely not to be exposed than second, middle and fourth category of wealth quintile. Women whose husband has another wife were nearly 94% less likely to state that they could choose to have one or more child than women whose husband has no other wife other than them with the CR (OR = 1.0575: 1.0464, 1.0688). Those who use contraceptives were 0.9594 times less likely exposed to one or more child than women who do not use contraceptives, OR = 0.9594; CI = (0.9466, 0.9724).

Similarly, the logit (zero) part of the hurdle model is 5.0411 times more predicted never exposed to having at least a child in their whole life for reproductive age women with the factors considered. Mother's current age, age at first birth, age at first intercourse, place of residence, region of residence exclude South-West, Islamic religion and middle category of wealth quintile were significant factors for women who declared that; if they could go back to the time when they did not have any children and could choose exactly the number of children to have in their whole life, then, they would chose to be childless.

The mothers' current ages were 0.9588 times less likely than never exposed to having at least a child when to choose the ideal number of children with the OR = 0.9588; CI = (0.9515, 0.9661) but women age at first birth and age at first intercourse were respectively 4.53% and 3.69% more likely. The odds of women that reside in the rural were 91.64% more probably not exposed to choosing zero as the ideal number of children than their urban counterparts. Mothers from the North-central region of the country were 1.4898 times higher than those from the North-East and never exposed to choosing at least a child as the ideal number of children, with the CR (OR = 1.4898; 1.1162, 1.9883), whereas, North-West, South-East and South-South were respectively 46%, 44% and 46% lesser. Muslim mothers were about 34% less likely never exposed to stating zero as the number of children to have in their life time than traditionalists/others. Mothers from the middle category of wealth quintile were about 74% less likely never exposed to one or more child than those in the lowest category, OR = 0.7371; CI = (0.6006, 0.9048).

V. DISCUSSION AND CONCLUSION

This study employed four different competing models for count data, a pair each from NB-type and Poisson-type. Each of these models was fit to the dataset extracted from 2018 Nigeria Demographic Health Survey such that ideal number of children is the response variable and fifteen other factors representing the explanatory variables. The response variable was presence with an overdispersion and significant extra zeros count.

It was revealed that 2.48% of the women in the dataset preferred that having no child is the ideal family size, but the poisson model predicts that only 0.66% would rather stay childless as the ideal number of children as shown in Fig. 2. Clearly, the model underestimates the probability of zero counts. Poisson regression would have been severely underfitting zero counts, because there are 1010 observed zero counts but Poisson only predicted 267.

This is where the hurdle model comes to play. The hurdle model is a two-stage model that specifies one process for zero counts and another process for positive counts. The poisson hurdle and negative binomial hurdle models both predicted that there are 1007 number of zeros in the observed data.

The results showed that there exists overdispersion, whereby Poisson regression is not suitable and NB was the model of choice that could explain for the overdispersion parameter but it couldn't account for overdispersion resulting from extra zeros. This springs the choice of employing hurdle models (Poisson and Negative binomial), since it could explain for the overdispersion resulting from extra zeros in the data. It was crystal clear from the rootograms that both Poisson and Negative binomial regression models show similar behavior and underfit the zeros count but Negative binomial hurdle and Poisson hurdle perfectly fit the zero count. The LRT revealed that Poisson-type models (P and PH) were more appropriate in handling of the overdispersion in the ideal number of children data than the NB-type models (NB and NBH). The Vuong test output indicates that the hurdle models (PH and NBH) are both better than Poisson and NB. The result further revealed that there was no difference between the PH and NBH models. This shows that both PH and NBH models handled the excessive zeros equally without making justification for overdispersion but PH was preferred since the test statistic is positive. The finding of this work indicated that PH model was the most flexible of the competing models in terms of the AIC and BIC. Poisson hurdle has the least MSE, RMSE and MAE values during the investigation of out-of-sample performance than the rest models. However, both P and NB were approximately the same. Hence, PH model outperformed the rest models.

The finding from this study was that mother's current age, age at first birth, age at first intercourse, place of residence, region of residence except South-West; middle wealth quintile category and Muslim women were found to be significant factors for mothers choosing no child and at least a child as the ideal number of children to have for their whole life in Nigeria.

One of the major strengths of this study is the ability to unfold the significant role fertility preferences in terms of the ideal number of children played to understand Nigeria's fertility. The shortcoming in this work is the inconsideration of the attitudes and pressure from family members, most especially the husband, whom may pose major influence on reproductive decisions. More so, fertility preference of reproductive age women is the pivotal of the study whereby cross-sectional data was employed. Due to the nature of the data used, some sensitive information (e.g., age at first intercourse and contraceptive use) which is amongst paramount variables to the study might not be accurately disclosed by the respondents during the survey. This has been another major limitation to this study.

The future research will be directed to addressing one of the deficits identified in this work by making necessary consideration for the influence of family members' (or husband's) attitudes and pressure on women fertility.

REFERENCES

[1] F. Famoye, J. T. Wulu, and K. P. Singh, "On the generalized poisson regression model with an application to accident data," *Journal of Data Science*, vol. 2, no. 3, pp. 287–295, 2004.

[2] G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright, "A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 54, no. 1, pp. 127–142, 2005. doi: https://doi.org/10.1111/j.1467-9876.2005.00474.x

[3] D. Lord, S. D. Guikema, and S. R. Geedipally, "Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1123–1134, 2008. doi: https://doi.org/10.1016/j.aap.2007.12.003

[4] D. Lord, S. R. Geedipally, and S. D. Guikema, "Extension of the application of Conway-Maxwell-Poisson models: Analyzing traffic crash data exhibiting underdispersion," *Risk Analysis: An International Journal*, vol. 30, no. 8, pp. 1268–1276, 2010. doi: https://doi.org/10.1111/j.1539-6924.2010.01417.x

[5] H. Joe and R. Zhu, "Generalized poisson distribution: The property of mixture of poisson and comparison with negative binomial distribution," *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, vol. 47, no. 2, pp. 219–229, 2005. doi: https://doi.org/10.1002/bimj.200410102

[6] D. Lord, S. P. Washington, and J. N. Ivan, "Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory," *Accident Analysis & Prevention*, vol. 37, no. 1, pp. 35–46, 2005. doi: https://doi.org/10.1016/j.aap.2004.02.004

[7] J. Hilbe. (2013) Re: Which is the most appropriate method to analyze counts? [Online]. Available: https://bit.ly/3u0gdJj

[8] E. E. Gbur, W. W. Stroup, K. S. McCarter, S. Durham, L. J. Young, M. Christman, M. West, and M. Kramer, "Generalized linear mixed models," in *Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences*. Madison, WI: American Society of Agronomy, 2012.

[9] K. C. Yip and K. K. Yau, "On modeling claim frequency data in general insurance with extra zeros," *Insurance: Mathematics and Economics*, vol. 36, no. 2, pp. 153–163, 2005. doi: https://doi.org/10.1016/j.insmatheco.2004.11.002

[10] A. Zeileis, C. Kleiber, and S. Jackman, "Regression models for count data in R," *Journal of Statistical Software*, vol. 27, no. 8, pp. 1–25, 2008.

[11] M. F. Tüzen and S. Erbaş, "A comparison of count data models with an application to daily cigarette consumption of young persons," *Communications in Statistics-Theory and Methods*, vol. 47, no. 23, pp. 5825–5844, 2018. doi: https://doi.org/10.1080/03610926.2017.1402050

[12] National Population Commission (NPC) [Nigeria] and ICF, *Nigeria Demographic and Health Survey 2018 Key Indicators Report*. Maryland, MD: NPC and ICF International, 2019.

[13] S. B. Adebayo and E. Gayawan, "Exploring spatial variations, trend and effect of exposure to media as an enhancer to uptake of modern family planning methods: Evidence from 2003 to 2018 Nigeria demographic health survey," *Spatial Demography*, pp. 1–26, 2021. doi: https://doi.org/10.1007/s40980-021-00080-z

[14] O. A. Paul, F. S. Adedamola, O. T. AO, B. M. Ardo, and F. M. Aderemi, "Bayesian semi-parametric modeling of infertility in Nigeria," *European Scientific Journal*, vol. 15, no. 27, p. 221–230, 2019. doi: http://dx.doi.org/10.19044/esj.2019.v15n27p221

[15] D. N. Ononokpono, O. G. Adebola, E. Gayawan, and A. F. Fagbamigbe, "Modelling determinants of geographical patterns in the marital statuses of women in Nigeria," *Spatial Demography* (Just Accepted), pp. 1–28, 2021. doi: https://doi.org/10.1007/s40980-020-00072-5

[16] C. B. Polis, C. M. Cox, Ö. Tunçalp, A. C. McLain, and M. E. Thoma, "Estimating infertility prevalence in low-to-middle-income countries: An application of a current duration approach to demographic and health survey data," *Human Reproduction*, vol. 32, no. 5, pp. 1064–1074, 2017. doi: https://doi.org/10.1093/humrep/dex025

[17] S. Babalola and O. Oyenubi, "Factors explaining the North-South differentials in contraceptive use in Nigeria: A nonlinear decomposition analysis," *Demographic Research*, vol. 38, pp. 287–308, 2018. doi: https://doi.org/10.4054/DemRes.2018.38.12

[18] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992. doi: https://doi.org/10.1080/00401706.1992.104852

[19] W. Liu and J. Cela, *Count Data Models in SAS*. Albuquerque, NM: SAS Global Forum, 2008.

[20] A. M. Garay, E. M. Hashimoto, E. M. Ortega, and V. H. Lachos, "On estimation and influence diagnostics for zero-inflated negative binomial regression models," *Computational Statistics & Data Analysis*, vol. 55, no. 3, pp. 1304–1318, 2011. doi: https://doi.org/10.1016/j.csda.2010.09.019

[21] J. G. Morel and N. K. Neerchal, *Overdispersion Models in SAS*. Cary, NC: SAS Institute, Inc., 2012.

[22] H. Wickham *et al.*, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, pp. 1–6, 2019. doi: https://doi.org/10.21105/joss.01686

[23] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S-PLUS*. New York, NY: Springer Science & Business Media, 2013.

[24] A. Zeileis and T. Hothorn. (2002) Diagnostic checking in regression relationships. [Online]. Available: https://bit.ly/3ABL2Xd

[25] A. Zeileis, "Object-oriented computation of sandwich estimators," *Journal of Statistical Software*, vol. 16, no. 1, pp. 1–16, 2006. doi: https://doi.org/10.18637/jss.v016.i09

[26] C. Kleiber and A. Zeileis, "Visualizing count data regressions using rootograms," *The American Statistician*, vol. 70, no. 3, pp. 296–303, 2016. doi: https://doi.org/10.1080/00031305.2016.1173590

[27] A. A. Yirga, S. F. Melesse, H. G. Mwambi, and D. G. Ayele, "Negative binomial mixed models for analyzing longitudinal CD4 count data," *Scientific Reports*, vol. 10, no. 1, pp. 1–15, 2020. doi: https://doi.org/10.1038/s41598-020-73883-7

[28] X. Liu, *Methods and Applications of Longitudinal Data Analysis*. New York, NY: Elsevier, 2015.

[29] K. P. Burnham and D. R. Anderson, "Multimodel inference: Understanding AIC and BIC in model selection," *Sociological Methods & Research*, vol. 33, no. 2, pp. 261–304, 2004. doi: https://doi.org/10.1177/0049124104268644

[30] J. M. Hilbe, *Negative Binomial Regression*. New York, NY: Cambridge University Press, 2011.

[31] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica:*

Journal of the Econometric Society, vol. 57, no. 2, pp. 307–333, 1989. doi: https://doi.org/10.2307/1912557

[32] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006. doi: https://doi.org/10.1016/j.ijforecast.2006.03.001

[33] Z. Pierri and B. A., "Muslims in Northern Nigeria: Between challenge and opportunity," in *Muslim minority-state relations, The modern Muslim world*, R. Mason, Ed. New York, NY: Palgrave Macmillan,, 2016, p. 133–153.

[34] R. Paternoster and R. Brame, "Multiple routes to delinquency? A test of developmental and general theories of crime," *Criminology*, vol. 35, no. 1, pp. 49–84, 1997. doi: https://doi.org/10.1111/j.1745-9125.1997.tb00870.x

[35] D. W. Osgood, "Poisson-based regression analysis of aggregate crime rates," *Journal of Quantitative Criminology*, vol. 16, no. 1, pp. 21–43, 2000. doi: https://doi.org/10.1023/A:1007521427059