Scientific African 9 (2020) e00494

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

Modelling fertility levels in Nigeria using Generalized Poisson regression-based approach

Jecinta U. Ibeji*, Temesgen Zewotir, Delia North, Lateef Amusa

School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal, Durban, South Africa

A R T I C L E I N F O

Article history:
Received 21 May 2020
Revised 14 July 2020
Accepted 22 July 2020

Keywords:
Children ever born
Poisson regression
Negative Binomial regression
Generalized Poisson regression
interaction effect

A B S T R A C T

The rapid increase in total children ever born without a proportionate growth in the Nigerian economy has been a major concern. The total children ever born, being a count data, requires applying an appropriate regression model. Poisson distribution is the ideal distribution to describe this data, but it is deficient due to equality of variance and mean. This deficiency results in under/over-dispersion and the estimation of standard errors will be biased rendering the test statistics incorrect. This study aimed to model count data with the application of total children ever born using a Negative Binomial and Generalized Poisson regression. The Nigeria Demographic and Health Survey 2013 data of women within the age of 15–49 years were used. A comparison of the three models revealed that Generalized Poisson regression is the appropriate model to correct for under/over-dispersion with age of household head ($P < .0001$), age of respondent at the time of first birth ($P < .0001$), urban-rural status ($P < .0001$), and religion ($P < .0001$) being significantly associated with total children ever born. Early marriage, religious belief and uninformed nature of women who dwell in rural areas should be checked to control fertility levels in Nigeria.

Introduction

Fertility is known as one of the three primary components dynamics that determine the structure, size, and component of the population of any country [1]. According to [2], assuming all women lived to the end of their childbearing years and bore children according to a given fertility rate at each age, the mean number of children that would be born per woman is not just a direct measure of the level of fertility but also an indicator of the expected population change in the country. They further stressed that a rate of two children per woman is seen as the replacement rate for a population and this results in relative stability in terms of total numbers while rates above two children show populations growing in size and whose median age is declining. Although higher rates may be challenging for families, in some situations, to feed and educate their children and for women to enter the labor force, rates below two children bespeak a population decrease in size and growing older.

* Corresponding author.
E-mail address: jecintaanyaogu@gmail.com (J.U. Ibeji).

https://doi.org/10.1016/j.sciaf.2020.e00494

According to [3] report, 55 countries or areas of which 42 were in Africa, 8 in Asia and 5 in Oceania, the Total Fertility Rate averaged more than 3.5 births per woman from 2010 - 2015. Since 1970 - 1975, it is recorded that in 47 of these 55 “high-fertility” countries, fertility declines were smaller than the world median fertility decline of 3.2 births per women. Notwithstanding, there has been quick fertility declined over the past 40 years in the other 8 countries, namely, Jordan, Kenya, Mayotte, State of Palestine, Rwanda, Tajikistan, Yemen and Zimbabwe of which Nigeria is not among.

In terms of high and medium-high fertility except for Southern Africa with notably lower fertility, and selected countries of Asia and Oceania, more than 3.5 births per woman are concentrated in countries located in sub-Sahara Africa. It is on record that in descending order of population size, Nigeria has the highest population in the group of most populous countries followed by Ethiopia, the Democratic Republic of Congo and the United Republic of Tanzania [3].

Though culturally, a woman is said to be martially fulfilled in Nigeria by giving birth to more children [2]. But on the contrary, [4–6] cited the effect of this culture to Nigeria economy, stating that to ensure the economic growth of less developed countries like Nigeria, population growth must be held in check. For example, it is reported by [2,4,7] that the socio-economic, demographic, and environmental development of Ethiopia and other less developed countries has been adversely affected by uncontrolled fertility rate. Besides this, studies conducted in Nigeria and other African countries and some less developed countries have shown that food insecurity, high unemployment rate and environmental risk is closely linked to high rate of fertility [8–12]. The fertility rate is undoubtedly conditioned by political, health, cultural, demographic and socioeconomic setting [13–15]. Hence study the proximate and socio-demographic determinants is very paramount.

Count data is a type of statistical data in which the observations can take only non-negative integer values $\{0, 1, 2, \ldots\}$, which arise from counting rather than ranking [16]. Examples of statistical analyses that involve count data are simple counts that include the number of thunderstorms that occurred in one-year calendar, fatal vehicle accidents per day, customers arriving at the shopping mall per hour, with categorical data being the counts denote the number of items belonging to each category [17].

Count data can be applied to different fields, including medicine, agriculture, life science, business, social, behavioral science, and demographic and health survey data. For example, in medicine, [18] opined that researches have shown that the effect and worth of health information technology (HIT) frequently used on outcome measures that are counts of things, such as hospital admissions, the number of laboratory tests per patient, adverse drug events (ADEs), and rates of cardiac arrest. This kind of data gives several analytic challenges, such as a large and perhaps disproportionate number of zero values, slightly high frequency of small integer values, and non-constant variance (where the variance of the residuals differs for different ranges of independent variables) [19]. In agriculture, count data was used to describe the implementation of agricultural and natural resource management technologies by small farmers in Central America [20].

Count data models are most appropriate for a certain type of adoption data. They are applied in the investigation of the impact of key socio-economic, biophysical, and institutional factors on the implementation of integrated pest management, agroforestry, and soil conservation technologies among small farmers. Factors affecting farmers’ quantity decisions related to farming precision technology can also be determined using count data models [21]. Furthermore, [22] in their work opined that count data models can be applied in life science to predict the number of *C. caretta* hatchlings dying from exposure to the sun. Applying count data models in business centres focuses on the consumption of a product. [23] compared count data models with the application of daily consumption of cigarette by young people in Turkey, and found Zero-inflated negative binomial (ZINB) and Negative binomial hurdle (NBH) to be preferable for analysis. The outcome variable in certain financial studies is a count that takes a non-negative integer value. Examples include the number of takeover bids received by a target firm, unpaid credit instalments (for scoring credit), accidents or accident claims (for insurance premia determination) and mortgage loans prepaid (mortgage-backed securities pricing). [24] applied Poisson and Negative Binomial Poisson and Negative Binomial Models for count data which had prominence on the underlying count process and links to dual data on durations. Likewise, modelling count data is a common task in economics and social sciences. According to [25], Hurdle and Zero-inflated model are capable of handling over-dispersion and excess zeros (which are two problems that mostly occur in count data sets of economics and social sciences).

Notwithstanding different models used in modelling count data, Poisson regression has been reported as the main methods for count data modelling, Poisson distribution has been the most commonly used distribution for modelling count data, but it assumes equal variance with the mean which makes it less appropriate since count data usually show over or under dispersion. According to [26], Poisson regression has been reported as the main methods for count data modelling, but violates equi-dispersion hypothesis and confines its use in several real-world applications due to under/over-dispersion. This superfluous disparity could lead to inaccurate inferences in the standard errors, tests, confidence intervals and parameter estimates, over-dispersion frequently surfaces for several reasons as well as mechanisms that cause too much zero counts [27,28]. For this reason, developing and applying statistical models as a substitute for modelling over-dispersed data are important. Accordingly, in various areas, over-dispersed count data are common, which in turn has led to the development of a statistical methodology for modelling these over-dispersed data. Studies on dealing with under-dispersion and over-dispersion issues have been reported [29,30] tried to overcome over-dispersion by using Zero-inflated Poisson and Negative Binomial regression to analyze the death rate of patients infected with AIDs.

Negative Binomial regression and Generalized Poisson regression were used to model count data involving the number of cervical cases to overcome the problem of overdispersion [31,32] investigated the effects of demographic and socio-economic factors on the number of children ever born by married women of age 15 to 49 years in Ethiopia using the ‘quasi-likelihood’ in a generalized linear model to overcome the problem of over and under-dispersion. In some countries, especially where

Table 1
Description dependent variable and selected independent variables.

Variables	Description
Cheb	Total children ever born
Age1stmar	Age at first marriage or cohabitation
Kidsex	Sex of child
Religion	Religion
Pregterm	Ever had a pregnancy terminated via abortion, miscarriage or stillbirth
Age1stbirth	Age of respondent at the time of first birth
Kidtwin	A child is twin or single birth
Fp	Fertility preferences
Pregnant	Currently pregnant
Region	Region
Hhha	Age of household head
Urban	Urban-rural status
Curwor	Currently working
Woc	Woman's occupation
Educ	Educational level
Childdesire	Whether and when this child's pregnancy is wanted
Kidalive	Child is alive

marriage is recognized as a major medium for procreation, data on children ever born are only available for ever married or currently married women. In Nigeria, the 2015 average fertility rate per woman is 5.5 children compared with 5.7 children in 2003 and 2008. Residence and region are major determinants of fertility variation. In urban areas, women have 4.7 children on average, compared with 6.2 children per woman in rural areas. The North-West Zone has the highest fertility rate with an average of 6.7 children per woman, while the South-South Zone has the lowest fertility rate of 4.3 children on the average [33]. Fertility also varies with mother's educational level and economic status [33]. Some work has been reported on the application of Poisson and Negative Binomial on over-dispersed count data, but to the best of our knowledge, there is little work on the use of Negative Binomial and Generalized Poisson regression as an alternative in handling count data. Because of this, the appropriate model to describe the total number of children ever born in Nigeria is still a subject of study. Generalized Poisson regression is a useful model for fitting both over-dispersed and under-dispersed count data because it allows for more variability and it is more flexible in analyzing independent variables. In this work Negative, Binomial and Generalized Poisson regression are applied as an alternative for handling over-dispersed and under-dispersed count data considering the number of children ever born in Nigeria.

Materials and methods

This study used secondary data from the individual's questionnaire of Nigeria Demographic and Health Survey 2013, which covers all the regions in the country. In the 2013 Nigeria Demographic and Health Survey demographic, socioeconomic and health information were collected for both men and women. According to the NDHS 2013 report, 30977 households were selected, of which 30878 women who were still within childbearing age were interviewed. [34] defined children ever born as the number of children born alive by married women from age 15 years and above, this includes all the live birth that is living or dead from married women up to the time of collection of data. From previous literature [32,35–37], the independent variables were chosen based on some socioeconomic and demographic factors that have a significant relation with fertility levels. Since equidispersion is not obtainable in real life, Generalized Poisson regression was used to analyze the data because of overdispersion. Table 1 is the description of the selected variables from the data, while Fig. 1 is the schematic representation of the conceptual framework.

Poisson regression

Poisson regression is a generalized linear model (GLM) form of regression analysis with the assumption Mean = Variance. It is known as the reference line model for count data analysis [38–41]. Therefore, if Y is the discreet random variable that has Poisson distribution with mean μ, then Y has a probability distribution function in Eq. (1)

$$f(y) = \frac{e^{-\mu}\mu^{y}}{y!}, \quad y = 0, 1, \ldots \tag{1}$$

where $E(y) = Var(y) = \mu$.

The first four moments are

$$\text{Mean, } \mathrm{E}(Y) = \mu, \text{Variance, } \sigma^2 = \mu, \text{Skweness, } \alpha_3 = \frac{1}{\sqrt{\mu}} \text{ and Kurtosis, } \alpha_4 = 3 + \frac{1}{\mu}$$

Poisson distribution has a rigid assumption that the mean must equal variance which is not feasible in real life [42].

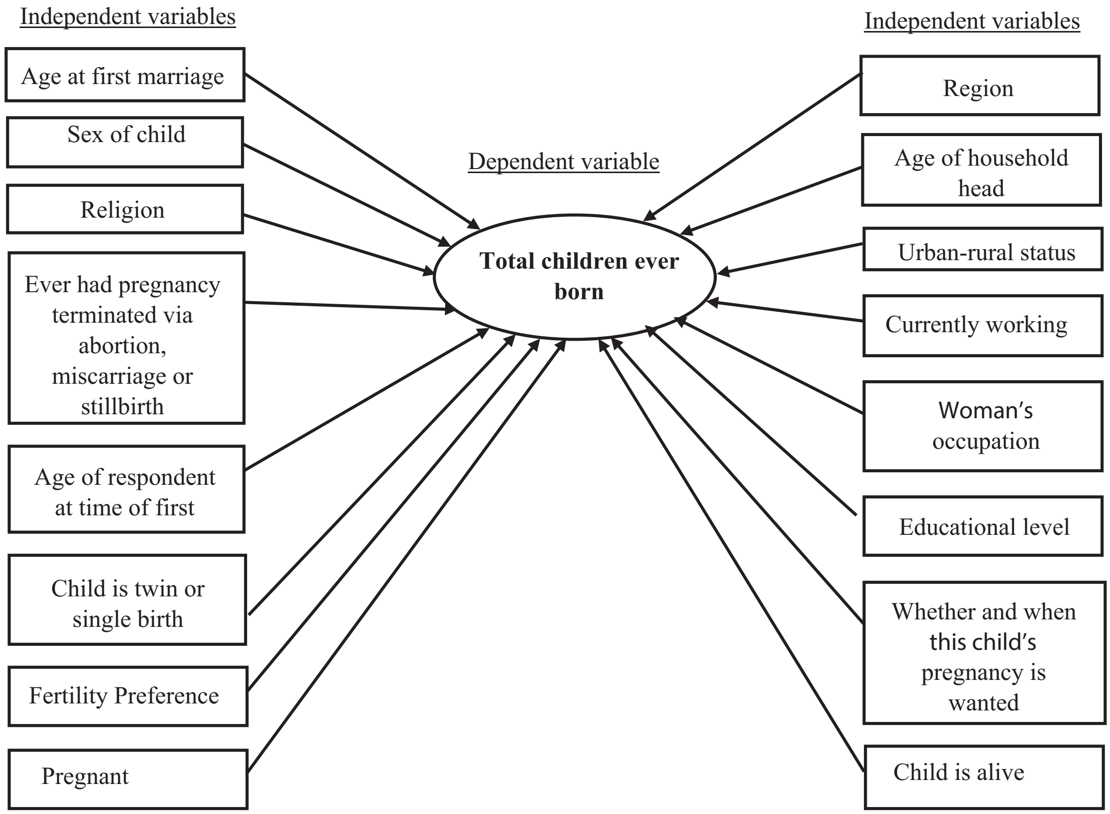


Fig. 1. The schematic representation of the conceptual framework.

One of the ways to model for over-dispersion is Negative Binomial regression model [43]. It assumes that the conditional variance of the outcome variable is greater than its conditional mean. Negative Binomial has an advantage over Poisson regression due to its ability to possess one extra parameter that helps to adjust the variance independently from the mean with probability mass function as in Eq. (2)

$$f(y) = \frac{\Gamma\left(y+\frac{1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \frac{(k\mu)^y}{(1+k\mu)^{y+1/k}} \text{ for } y = 0, 1, 2, \ldots \tag{2}$$

And the first four moments are

$$\text{Skweness, } \alpha^3 = \frac{2\mu+\frac{1}{k}}{\left(\mu+\frac{1}{k}\right)\sqrt{\left(\frac{1}{k}\right)\left(\frac{\mu}{\mu+\frac{1}{k}}\right)}} \text{ and}$$

$$\text{Mean, } E(y) = \mu, \text{ Variance, } Var(y) = \mu + \mu^2 k,$$

$$\text{Kurtosis, } \alpha^4 = 3 + \frac{1}{\mu+\frac{1}{k}} + 5k + \frac{1}{\frac{1}{k}\left(\frac{2/k+\mu}{\mu+1/k}\right)}$$

The relationship between Negative Binomial and Poisson distribution is seen when a gamma prior is used for a Poisson distribution. In other words, μ is distributed as a gamma distribution with shape= r and scale $\beta = \frac{(1-p)}{p}$,when μ is itself a random variable [44].

The generalized Poisson regression model is used as an extension of Poisson regression that accounts for over-dispersion ("dist.Generalized.Poisson function | R Documentation," [45]. The probability mass function of this distribution is

$$p(y;\omega,\theta) = \frac{\theta(\theta+\omega y)^{y-1}}{y!} e^{-\theta-\omega y} \quad y = 0, 1, 2, \ldots; 0 \le \omega \langle 1; \theta \rangle 0 \tag{3}$$

With θ and ϖ denoting the shape of the parameters.

Table 2
Summary statistics of continuous explanatory variables.

Variables	Mean	Standard Deviation	N
Hhha	41.41	12.03	31424
Age1stmar	17.62	4.45	30878
Age1stbirth	19.36	4.26	31482

The first four moments are

$$\text{Mean},\ \mu = \frac{\theta}{1-\omega}, \qquad \text{Variance},\ \sigma^2 = \frac{\theta}{(1-\omega)^3},$$

$$\text{Skewness},\ \alpha^3 = \frac{(1+2\omega)^2}{\theta(1-\omega)}$$

$$\text{Kurtosis},\ \alpha^4 = 3 + \frac{\left(1+8\omega+6\omega^2\right)}{\theta(1-\omega)}$$

It has been shown by Consul and Jain, [46] that the probability mass function of this distribution is a probability distribution since it has the property $\sum_{n=0}^{\infty} P_n = (\theta,\omega) = 1$. This is achieved by using the identity $\sum_{n=0}^{\infty} \frac{(\theta+\omega n)^n}{n!} e^{-\theta-\omega n} = \frac{1}{1-\omega},\quad for\ -\omega_0 < \omega < 1$ found in [47].

Data analysis

Descriptive statistics were used to summarize all the variables of this study. Mean and the standard deviation was used for continuous variable description while the box and whisker plot was used for categorical variables. In the presence of under/over-dispersion, generalized linear models were used to analyze the data by employing the Generalized Poisson regression model. The Pearson chi-square and the full log-likelihood of Poisson regression and Negative Binomial regression were compared, and Negative Binomial regression was selected as a better model. The extent to which the fit of the model was improved when extra variables were added to the model was measured using the deviance. Notwithstanding, the deviance for Negative Binomial regression was found to be under-dispersed. The generalized Poisson regression model was chosen as the appropriate model to check for under-dispersion. Cook's distance against observations was used to identify the influential observations and lastly the link function was derived from the exponential family.

All possible two-interactions among the explanatory variables were evaluated by including the interactions one at a time to the main effects model. In the first round, all the possible interactions level of significance was recorded, and the highly significant ones were entered into the model with the main effect. The same procedure was used for the second-round two-way interaction term entry to the model with main effects and the interaction selected in the previous round. The process continued until no more significant interaction effect was left to be included. *Fp* interacted with three variables: *age1stmar, region* and *kidtwin. Educ* interacted with *childdesire*, and *kidalive*. While *pregterm* interacted with *woc*. All the analysis was done using SAS Version 9.4.

Results and discussion

The descriptive statistics of *hhha, age1stmar* and *age1stbirth* are shown in Table 2 It was found that the mean *hhha* is about 41 years with standard deviation 12.03 and the mean *age1stmar* is about 28 years with a standard deviation of 4.45. Meanwhile, the mean *age1stbirth* is about 19 years with a standard deviation of 4.26.

The boxplots give a better understanding of the data by its distribution, outliers, mean, median and spread. For example, in *urban*, women in rural area have a higher spread of *cheb* than women in the urban area. This is obvious from the lengths of their whiskers and the position of the median line. Also, for *religion*, the box and whisker plot indicate that *Muslim* women have a higher number of *cheb* than Christian women. This is obvious from the lengths of their whiskers. North East was observed to be the highest number of *cheb* while South West had the least *cheb*. The box and whisker plot for *kidsex* shows that there is not much difference between the total number of male and female child given birth to by the women (Fig. 2).

Table 3 reveals the result of the independent variables on *cheb* by the women. More than 5000 women gave birth to 2 or 3 children (Fig. 3). Categorically, though there was an increase in the frequency of women that gave birth to one child, the frequency started dropping as the number of children ever born increases. The estimated value, standard error, z value, estimated IRR and Pr>|z| are presented in the table. To prevent external influence, all the main effects were retained in the model. Furthermore, we examine the joint effects of some factors acting together over and above their main effects and evaluated all possible two-interactions among the explanatory variables by including the interactions one at a time to the main effects model. All the possible interactions level of significance was recorded, and the highly significant ($P < 0.01$) ones

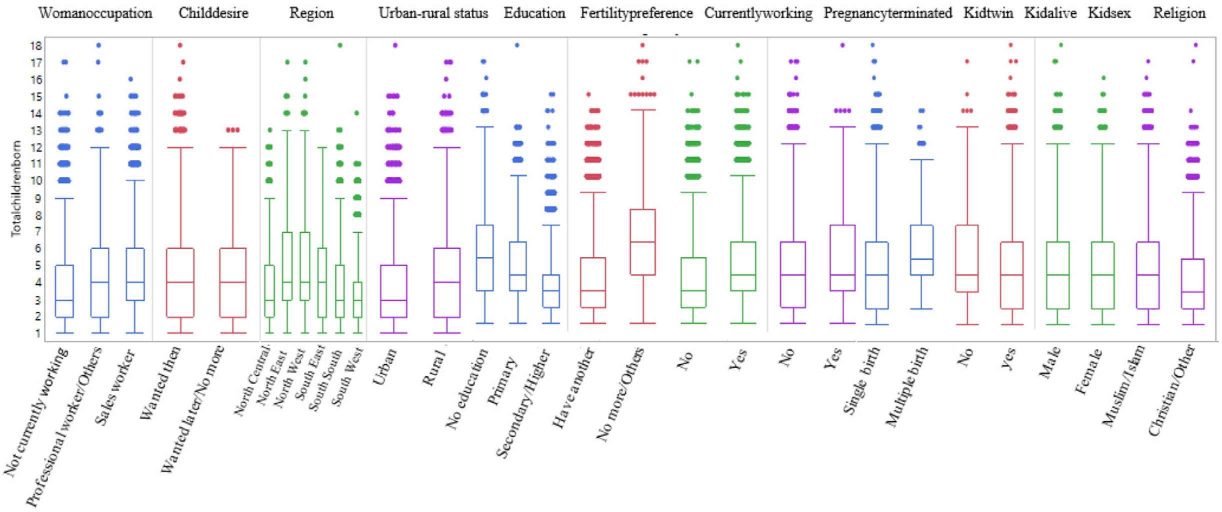


Fig. 2. Box and whisker plot for total children ever born by categorical explanatory variables.

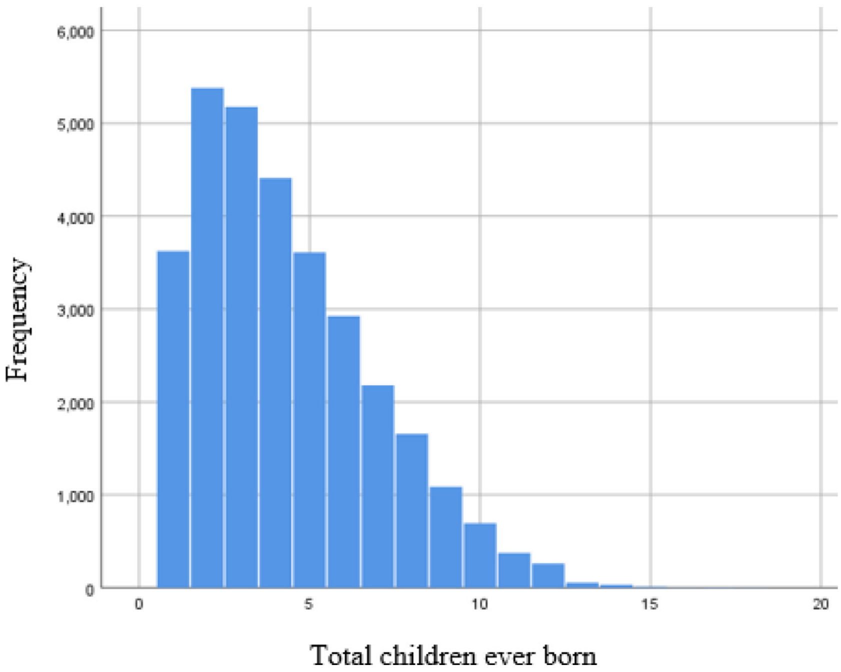


Fig. 3. Histogram of cheb based on NDHS 2013 data.

were entered into the model with the main effect. The process continued until no more significant interaction effect was left to be included. Hence, six interaction effects were gotten, they are *age1stmar* and *fp*, *region* and *fp*, *pregterm* and *woc*, *fp* and *kidtwin*, *childdesire* and *educ*, and *kidalive* and *educ*. Therefore, all the main effects and the six two-way interaction effects were included in the final model.

The analysis of maximum likelihood parameter estimates in Table 3 revealed that *urban*, *region*, *religion*, *woc*, *fp*, *childdesire*, *kidalive*, *educ*, *age1stmar*, *hhha*, *age1stbirth* and the interactions between *age1stmar* and *fp*, *region* and *fp*, *fp* and *kidtwin*, *childdesire* and *educ*, and *kidalive* and *educ* are significantly related to *cheb*.

The results of the main effect parameter estimate, the risk ratios and the z values are presented in Table 4 of this study. There is a strong relationship between *urban* and the mean of *cheb*. The risk ratio *cheb* by mothers who reside in rural area is 1.023 times the risk ratio of *cheb* by mothers who reside in the urban area. The risk ratio of *cheb* by Christian/Others women is 0.981 times the risk ratio of *cheb* by *Muslim* women. This means that Muslim women gave birth more than christian women and this is in line with previous studies in Nigeria [37,48,49]. Regarding *kidalive*, a woman whose child is not alive had 1.104 risk ratio compared to *cheb* by a woman whose child is alive. Invariably, women that have dead

Table 3
Analysis of main and interaction effects on the total children ever born.

Predictors	Categories	Estimated	Standard Error	z Value	Estimated IRR	Pr > |z|
Urban (Reference=Urban)	Rural	0.0255	0.0056	4.53	1.023	<.0001
Region (Reference=South West)	North Central	-0.0925	0.0186	-4.97	0.912	<.0001
	North East	0.0750	0.0187	4.02	1.078	<.0001
	North West	0.0169	0.0234	0.71	1.017	0.4807
	South East	0.1345	0.0228	5.88	1.144	<.0001
	South-South	0.2267	0.0166	13.68	1.254	<.0001
Religion (Reference= Muslim/Islam)	Christian/Others	-0.0888	0.0079	-11.19	0.915	<.0001
Pregterm (Reference=Yes)	No	-0.0009	0.0143	-0.07	0.999	0.9475
Woc (Reference=Sales worker)	Not currently working	-0.0823	0.0206	-4.00	0.921	<.0001
	Professional worker/Others	-0.0059	0.0182	-0.32	0.994	0.7485
Fp (Reference=Undecided/Others)	Have another	-0.6407	0.0333	-19.27	0.579	<.0001
Childdesire (Reference=Wanted then)	Wanted later /No more	-0.1152	0.0169	-6.82	0.891	<.0001
Kidsex (Reference=Male)	Female	0.0038	0.0025	1.54	1.004	0.1247
Kidalive (Reference=Yes)	No	0.0989	0.0224	4.41	1.104	<.0001
Kidtwin (Reference= Single birth)	Multiple birth	-0.0319	0.0236	-1.35	0.969	0.1771
Educ (Reference=Secondary/Higher)	No education	0.3255	0.0084	38.70	1.385	<.0001
	Primary	0.3443	0.0070	49.15	1.411	<.0001
Hhha		0.0096	0.0001	68.89	1.010	<.0001
Age1stmar		-0.0092	0.0013	-6.59	0.991	<.0001
Age1stbirth		-0.0195	0.0011	-17.02	0.981	<.0001
Age1stmar*Fp (Reference= Undecided/Others)	Have another	0.0115	0.0013	9.01	1.012	<.0001
Region (Reference= South West)*Fp (Reference=Undecided/Others)	North Central*Have another	-0.0345	0.0254	-1.36	1.005	0.1735
	North East*Have another	0.1294	0.0233	5.55	0.939	<.0001
	North West*Have another	0.1692	0.0280	6.04	0.889	<.0001
	South East*Have another	-0.0499	0.0306	-1.63	0.926	0.1029
	South-South*Have another	0.1889	0.0221	8.56	1.148	<.0001
Pregterm (Reference= Yes)*Woc (Reference=Sales worker)	No* Not currently working	-0.0060	0.0216	-0.28	0.994	0.7795
	No* Professional worker	0.0118	0.0194	0.61	1.012	0.5427
Fp (Reference=Undecided/Others)*Kidtwin (Reference=Single birth)	Have another*Multiple birth	0.1967	0.0305	6.45	1.217	<.0001
Chiddesire(Reference=Wanted later)*Educ(Reference=Secondary/Higher)						
Kidalive (Reference=Yes)*Educ (Reference=Secondary/Higher)	Wanted later/No more*No Education	0.1807	0.0232	7.78	1.198	<.0001
	Wantedlater/No more*Primary	-0.0218	0.0201	-1.08	0.978	0.2795
	No*No Education	-0.0833	0.0246	-3.38	0.909	0.0007
	No*Primary	0.1051	0.0247	4.26	1.037	<.0001

Table 4
The Generalized Poisson regression risk ratios extracted for main effects which not involved in the interaction.

Factors	Risk ratios	*Z*-Value
Hhha	1.010	68.89
Age1stbirth	0.981	-17.02
Urban (Reference=Urban)		
Rural	1.023	4.53
Religion (Reference=Islam)		
Christian/Others	0.915	-11.19
Kidsex (Reference=Male)		
Female	1.004	0.1247
Kidalive (Reference=Yes)		
No	1.104	4.41

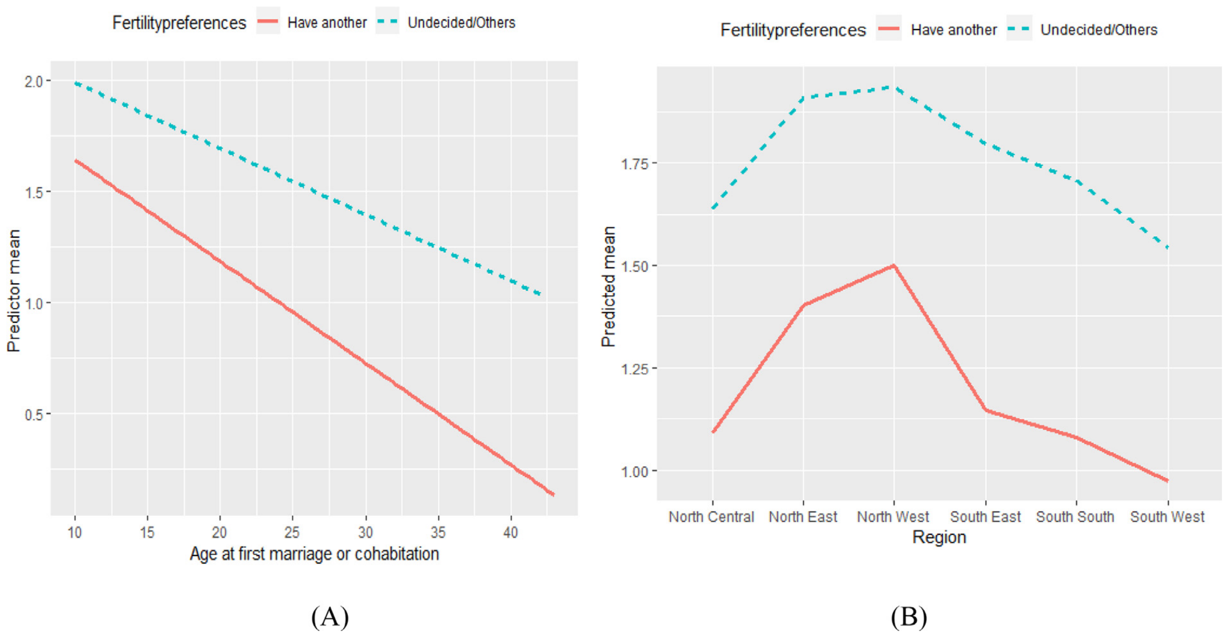


Fig. 4. The mean number of cheb by (A) Age at first birth and (B) Region.

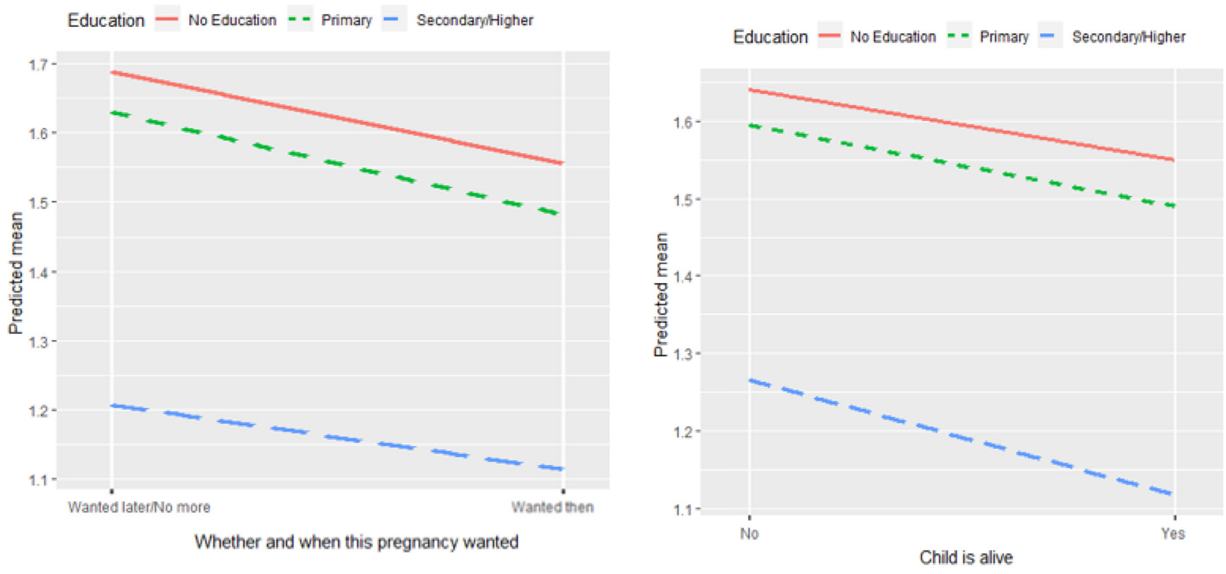
children gave birth more than those who children are alive. For *kidsex*, women that gave birth to female children had risk ratio of 1.004 compared to women that gave birth to male children. In other words, women with female children gave birth more than their contempolaries. This is because a male child has more preference than female child [50]. *Age1stbirth* is also positively related to *cheb*. It is recorded that women who gave birth at early age, have more children than women who gave birth late. This applies to hhha.

The summary of the interaction effects is shown in Table 3. Fig. 4 shows the mean number of *cheb* by *fp*, *age1stmar* and *region*. With respect to *age1stmar*, the increase in ages of women both in the group of have another and undecided/others decrease the mean number of *cheb*. On the other hand, the mean number of *cheb* is higher in rural areas than that of the urban areas in all regions of Nigeria.

Fig. 5 presents the relationship between *educ*, *childdesire*, and *kidalive*. Regarding *chiddesire* and *education*, women who wanted later/no more, the mean number of *cheb* is significantly higher than women who wanted then irrespective of their educational level. Also, in *kidalive*, there is a significant increase in the mean number of *cheb* by women with no education than their contemporaries. This means that as the level of education of women increases, the mean number of *cheb* decreases irrespective of whether the child is alive or not alive.

The interaction between *pregterm* and *woc* is presented in Fig. 6. It shows that women who had terminated pregnancy via abortion, miscarriage, or stillbirth and who work as sales worker have significantly more mean number of *cheb* than their contemporaries.

Finally, the interaction effects between *fp* and *kidtwin* as shown in Fig. 7 revealed that the mean number of *cheb* is significantly high for women with multiple birth who are undecided/others than women with single birth.



(A) (B)

Fig. 5. The mean number of cheb by (A) Whether and when pregnancy wanted and (B) Child is alive.

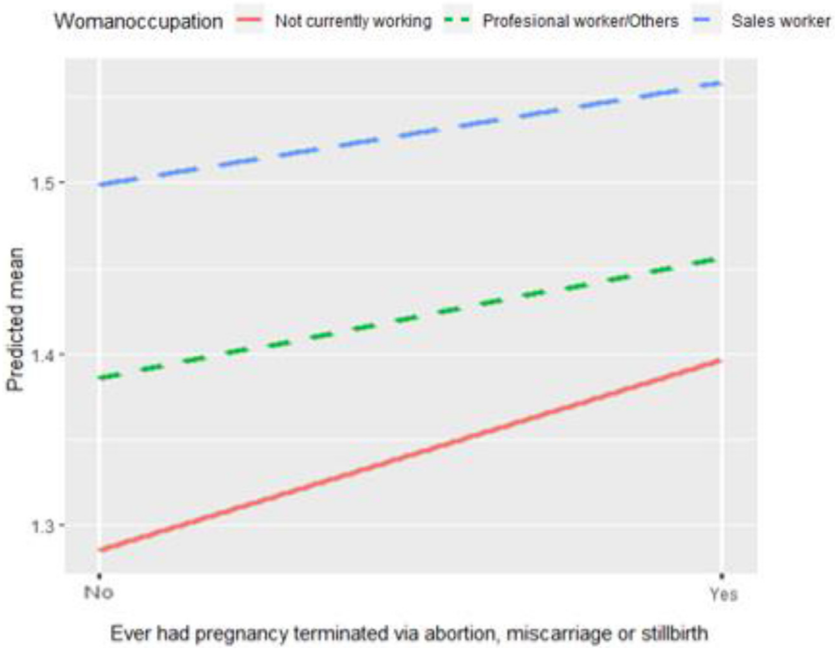


Fig. 6. The mean number of cheb by woc and pregterm.

This work is aimed at studying the socio-demographic factors associated with the fertility of married women in Nigeria. Generalized Poisson regression was used as an efficient method in modelling count data. It is derived based on Generalized Poisson distribution proposed by Famoye [51] and is implemented within the maximum-likelihood framework. The developed model has greater flexibility in modelling count data with the incorporation of the dispersed parameter. Especially its ability to accommodate both over-dispersion [52,53] and under-dispersion [54].

The Generalized Poisson regression performs the same as Poisson regression approach for count data when there is no dispersion but outperforms it when the underlying data are potentially dispersed. From the study, we note that the parameter estimates from GPR model are close to those from Poisson regression model when the dispersion parameter ω is very close to 0. Nevertheless, when ω is substantially different from 0, there is much difference in the parameter estimates.

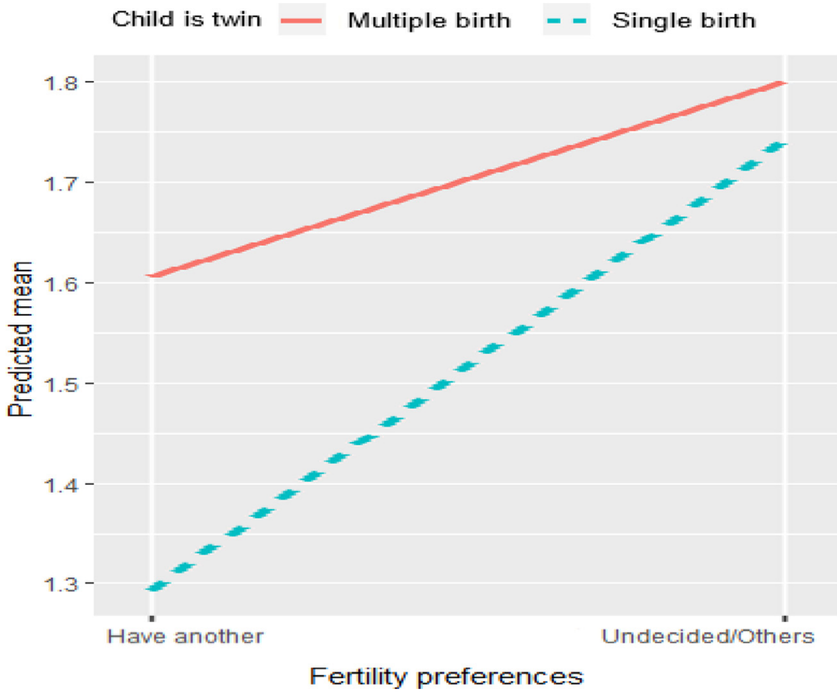


Fig. 7. The mean number of cheb by kidtwin and fertpref.

In this situation, the results obtained from the Poisson quasi-likelihood method may not be reliable as those from the GPR model.

It was found that all the selected socioeconomic and demographic factors considered in this study do not only possess the individual predictive capability but also strongly associated with fertility rate even when the direct determinants of fertility were adjusted for. Also, rural women were found to give birth more than urban women. This is in line with studies done previously in Nigeria and another part of the world [1,2,55,56].

Also, mother's educational level, religion, region to mention a few are among the major determinant of fertility rate. This confirms that the 2013 NDHS listed proximate determinants of fertility are strongly associated with fertility rate when studied individually, together and with the selected socio-demographic characteristics adjusted for. It was also found that when the socio-demographic characteristics were adjusted for, women who do not use any contraceptive were more likely to have increased fertility, while they were less likely to experience higher fertility when the socio-demographic characteristics were adjusted for.

Because the source of over-dispersion depends on many situations [57] when the excess of zero is the source of over-dispersion, Zero-inflated Poisson regression is preferred but otherwise, GPR model is most preferred. This is because Generalized Poisson regression it can accommodate both under-dispersion and over-dispersion when the data has little or no zero observation.

To study the outcome variable, the Generalized Poisson regression is preferred over Poisson regression and Negative Binomial because of its under-dispersion property (variance < mean).

Future studies need to focus on more explanatory variables that might be available from other sources. It will be useful in future research to investigate the motive and reason for childbearing, and to establish the present trend, as the increase in total children ever born could have an adverse effect on the Nigerian economy and security.

Conclusion

In this paper, we reported the effects of some determinant of fertility in Nigeria using the Generalized Poisson regression model. Notwithstanding three direct factors and information of over thirty thousand women were excluded due to missing values, the extracted data from the original data was enough for this study. It is obvious from our results that in Nigeria, societal factors have an immense effect on fertility rate. High fertility was independently associated with Hhha, Age1stbirth, Urban, Religion, Kidsex and Kidalive. Our understanding of the relationships among the risk factors was expanded by the inclusion of the interaction terms and this changes the interpretation of the main effect. age1stmar and fp, region and fp, childdesire and education, kidalive and education, pregterm and woc and fp and kidtwin are jointly associated with fertility status. We recommend that Nigeria population growth need to be controlled otherwise accompanied by rapid economic growth to alleviate the problem in society. The findings can be used for developing integrated support tools for the government, health policymakers and international agencies interested in infertility-related issues.

Declaration of Competing Interest

The authors do declare that there is no conflict of interest.

Acknowledgement

Authors are grateful to the University of KwaZulu-Natal, Durban for operational and infrastructural support and DHS for granting the author access to NDHS 2013 data.

This work was not funded.

References

[1] H.P. Upadhyay, K.R. Bhandari, Factors associated with children ever born: a case study of somadi village development committee of Palpa District, Nepal, Adv. J. Soc. Sci. 1 (1) (2017) 15–29.
[2] O.M. Akpa, O. Ikpotokin, Modeling the determinants of fertility among women of childbearing age in Nigeria: analysis using generalized linear modeling approach, Int. J. Humanit. Soc. Sci. 2 (18) (2012) 167–176.
[3] United NationsWorld Fertility Report 2015, 2015.
[4] J.C. Caldwell, P. Caldwell, Africa: the new family planning frontier, Stud. Fam. Plann. 33 (1) (2002) 76–86.
[5] T.J. Goliber, Population and reproductive health in Sub-Saharan Africa, Popul. Bull. 52 (4) (1997) 1–44.
[6] W. Mekonnen, A. Worku, Determinants of fertility in rural Ethiopia: the case of Butajira Demographic Surveillance System (DSS), BMC Public Health 11 (1) (2011) 782.
[7] G.D. Alene, A. Worku, Differentials of fertility in North and South Gondar zones, northwest Ethiopia: a comparative cross-sectional study, BMC Public Health 8 (1) (2008) 397.
[8] A. Cazzola, L. Pasquini, A. Angeli, The relationship between unemployment and fertility in Italy: a time-series analysis, Demogr. Res. 34 (2016) 1–38.
[9] O.A. Alo, Fertility regimentation of the rural Yoruba women of South-west Nigeria: the case of Ido and Isinbode, J. Soc. Sci. 26 (1) (2011) 57–65.
[10] J. Casterline, Wanted fertility, unwanted fertility, and fertility decline: A fresh assessment, Annual Meeting of the Population Association of America, 2010 April.
[11] E. Gurmu, R. Mace, Fertility decline driven by poverty: the case of Addis Ababa, Ethiopia, J. Biosoc. Sci. 40 (3) (2008) 339.
[12] N. Gupta, I. da Costa Leite, Adolescent fertility behavior: trends and determinants in northeastern Brazil, Int. Fam. Plan. Perspect. 25 (3) (1999) 125–130.
[13] P. Singh, in: An Introductory Text on Demography, 2, Dabco Investments Limited, Kaduna Nigeria, 2002, pp. 53–65.
[14] L.N. Kazembe, Modelling individual fertility levels in Malawian women: a spatial semiparametric regression model, Stat. Methods Appl. 18 (2) (2009) 237–255.
[15] V. Skirbekk, Trends in fertility by social status, Demogr. Res. 18 (5) (2008) 145–180.
[16] A.C. Cameron, P.K. Trivedi, Regression Analysis of Count Data, Cambridge University Press, 2013.
[17] R. Adhikari, Demographic, socio-economic, and cultural factors affecting fertility differentials in Nepal, BMC Pregnancy Childbirth 10 (1) (2010) 19.
[18] J. Du, Y.-T. Park, N. Theera-Ampornpunt, J.S. McCullough, S.M. Speedie, The use of count data models in biomedical informatics evaluation research, J. Am. Med. Inform. Assoc. 19 (1) (2011) 39–44.
[19] H. Zhou, P.Z. Siegel, J. Barile, R.S. Njai, W.W. Thompson, C. Kent, Y. Liao, Peer reviewed: models for count data with an application to healthy days measures: are you driving in screws with a hammer? Prev. Chronic Dis. 11 (2014).
[20] O.A. Ramirez, S.D. Shultz, Poisson count models to explain the adoption of agricultural and natural resource management technologies by small farmers in Central American countries, J. Agric. Appl. Econ. 32 (1) (2000) 21–33.
[21] T. Isgin, A. Bilgic, D.L. Forster, M.T. Batte, Using count data models to determine the factors affecting farmers' quantity decisions of precision farming technology adoption, Comput. Electron. Agric. 62 (2) (2008) 231–242.
[22] I. Ozmen, F. Famoye, Count regression models with an application to zoological data containing structural zeros, J. Data Sci. 5 (4) (2007) 491–502.
[23] M.F. Tüzen, S. Erbaş, A comparison of count data models with an application to daily cigarette consumption of young persons, Commun. Stat. 47 (23) (2018) 5825–5844.
[24] A.C. Cameron, P.K. Trivedi, 12 Count data models for financial data, Handbook Stat. 14 (1996) 363–391.
[25] A. Zeileis, C. Kleiber, S. Jackman, Regression models for count data in R, J. Stat. Softw. 27 (8) (2008) 1–25.
[26] G. Osuji, M. Obubu, H. Obiora-Ilouno, Uterine fibroid on women's fertility and pregnancy outcome in Delta State, Nigeria, J. Nat. Sci. Res. 6 (2) (2016) 27–33.
[27] S.D. Guikema, J.P. Goffelt, A flexible count data regression model for risk analysis, Risk Anal. 28 (1) (2008) 213–223.
[28] O. Maxwell, B.A. Mayowa, I.U. Chinedu, A.E. Peace, Modelling count data; a generalized linear model framework, Am. J. Math. Stat. 8 (6) (2018) 179–183.
[29] AA M.A, N.N. Naing, Comparison between Negative Binomial and Poisson death rate regression analysis: AIDS mortality co-infection patients, J. Math. (Wuhan) 3 (6) (2012) 34–38.
[30] J.-M. Kim, S. Jun, Zero-inflated poisson and negative binomial regressions for technology analysis, Int. J. Softw. Eng. Appl 10 (2016) 431–448.
[31] A. Melliana, Y. Setyorini, H. Eko, Purhadi: the comparison of generalized Poisson regression and negative binomial regression methods in overcoming overdispersion, IJSTR 2 (2013) 255–258.
[32] E.K. Muluneh, T. Zewotir, T.M. Yayeh, Fertility status of married women and its determinants in Ethiopia, J> Econ. Behav. Stud. 8 (4(J) (2016) 32–40.
[33] I. Macro, N.P. Commission, Nigeria demographic and health survey 2013, (2014).
[34] H. Kiser, M.D. Alamgir Hossain, Estimation of number of ever born children using zero truncated count model: evidence from Bangladesh Demographic and Health Survey, Health Inf. Sci. Syst. 7 (1) (2019) 3.
[35] N. Adhikari, A Study on Number of Child Ever Born: a Case Study of Dayanagar vdc of Rupendehi Distric, Central Department of Economics, Tribhuwan University Kritipur, Kathmandu, Nepal, 2014.
[36] H.P. Upadhyay, K.R. Bhandari, Factors associated with children ever born: a case study of somadi village development committee of Palpa District, Nepal, Adv. J. Soc. Sci. 1 (1) (2017) 15–29.
[37] O.O. Alaba, O.E. Olubusoye, J. Olaomi, Spatial patterns and determinants of fertility levels among women of childbearing age in Nigeria, S. Afr. Fam. Pract. 59 (4) (2017) 143–147.
[38] C. Ac, P.K. Trivedi, Regression Analysis of Count Data, Cambridge University Press, Cambridge, United Kingdom, 1998.
[39] C. Chatfield, J. Zidek, J. Lindsey, An Introduction to Generalized Linear Models, Chapman and Hall/CRC2010.
[40] S.J. Long, in: Regression Models for Categorical and Limited Dependent Variables, SAGE Publications, Inc., Thousand Oaks, California, 1997, p. 7.
[41] R. Winkelmann, Econometric Analysis of Count Data, Springer Science & Business Media, 2008.

[42] N. Ismail, A.A. Jemain, in: Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models, Casualty Actuarial Society Forum, Citeseer, 2007, pp. 103–158.
[43] A. Lindén, S. Mäntyniemi, Using the negative binomial distribution to model overdispersion in ecological count data, Ecology 92 (7) (2011) 1414–1421.
[44] J.D. Cook, Notes on the negative binomial distribution, Unknow, October 28 (2009) 2009.
[45] I. Ntzoufras, A. Katsis, D. Karlis, Bayesian assessment of the distribution of insurance claim counts using reversible jump MCMC, N. Am. Actuar. J. 9 (3) (2005) 90–108.
[46] P.C. Consul, G.C. Jain, A generalization of the Poisson distribution, Technometrics 15 (4) (1973) 791–799.
[47] J.L.W.V. Jensen, Sur une identité d'Abel et sur d'autres formules analogues, Acta Math. 26 (1) (1902) 307–318.
[48] E.E. Ebingha, J.S. Eni, J.T. Okpa, Population growth and socio-economic development of cross river state, Nigeria, Eur. J. Econ. Financ. Res. (2019).
[49] C.S. Iheyinwa, M. Oladosu, Socio-demographic factors, contraceptive use and fertility preference among married women in South-South Region of Nigeria, 3rd International Conference on African Development Issues, 2016.
[50] O. Fayehun, O. Omololu, U.C. Isiugo-Abanihe, Sex of preceding child and birth spacing among Nigerian ethnic groups, Afr. J. Reprod. Health 15 (2) (2011).
[51] F. Famoye, Restricted Generalized Poisson regression model, Commun. Stat. 22 (5) (1993) 1335–1354.
[52] H. Zhang, S.B. Pounds, L. Tang, Statistical methods for overdispersion in mRNA-seq count data, Open Bioinform. J. 7 (1) (2013) 34–40.
[53] J.F. Lawless, Negative binomial and mixed Poisson regression, Can. J. Stat./La Revue Canadienne de Statistique 15 (3) (1987) 209–225.
[54] P. Consul, F. Famoye, Generalized Poisson regression model, Commun. Stat. 21 (1) (1992) 89–109.
[55] E. Gurmu, R. Mace, Fertility decline driven by poverty: the case of Addis Ababa, Ethiopia, J. Biosoc. Sci. 40 (3) (2008) 339.
[56] W.O. Adebimpe, A.O. Esther, J.O. Bamidele, A. Olugbenga, A comparative study of socio-demographic determinants and fertility pattern among women in rural and urban communities in southwestern Nigeria, Cont. J. Med. Res. 5 (1) (2011) 32.
[57] Y. Mouatassim, E.H. Ezzahid, Poisson regression and Zero-inflated Poisson regression: application to private health insurance data, Eur. Actuar. J. 2 (2) (2012) 187–204.