

ORIGINAL ARTICLE

Open Access



Structured additive distributional hurdle Poisson modelling of individual fertility levels in Nigeria

Omodolapo Somo-Aina and Ezra Gayawan*

*Correspondence:

egayawan@futa.edu.ng;

ezrascribe@yahoo.com

Biostatistics and Spatial Statistics
Laboratory, Department of
Statistics, Federal University of
Technology, Akure, Nigeria

Abstract

Fertility is one of the dynamic components of population and has been modelled through children ever born per woman, which is a count variable that can be characterized with excessive zeros origination from women without any births. In order to examine the spatial variation across states of Nigeria, we proposed the use of hurdle models that classifies the data into a truncated count and point mass of zeros. We adopt distributional regression model that allows all parameters of the hurdle model to be linked to covariates of different types so as to allow for accessing the spatial variations and nonlinear forms of metrical covariates on the level of fertility and in the likelihood of having no child. Data was sourced from the 2013 Nigeria Demographic and Health Survey. Findings reveal the existence of north-south divide in the average level of fertility and in the likelihood of a woman not giving birth to any child. Women with higher level of education and those from richer or richest households have higher likelihood of having no child, but this is not the case for women with primary or secondary education, users of traditional or modern contraceptive, ever-married women and those working. There is therefore the need to strengthen family planning policies so that investment in contraceptive would yield the expected results in Nigeria.

Keywords: Fertility, Distributional regression, Hurdle Poisson, Spatial analysis, Nigeria

Background

Fertility rates in sub-Saharan African countries are still high when considered by what is obtainable in other developing countries. About 60% of the world population growth between 2017 and 2050 is projected to take place in Africa based on the observed annual growth rate of about 2.6% between 2010 and 2015. Nigeria is currently ranked the seventh most populous nation in the world and, based on the current total fertility rate of 5.5 and a rapid annual growth rate of 2.4%, it is projected that the country's human population would exceed that of the USA by 2050 and thus becoming the third most populous nation in the world (United-Nation 2017). However, population growth without corresponding expansion in per capita income and social amenities is usually characterized by food insecurity, high unemployment rate, poor economic growth, increased violence, prolonged disease prevalence and poverty, as currently being witnessed in some sub-Saharan African countries. Populations policies and measures to ensure reduction in

fertility levels are now being put in place in most sub-Saharan African countries including Nigeria (Bongaarts and Casterline 2013; NPC 2014).

Effective measures of reducing fertility rates must be based on appropriate identification and evaluation of the major determinants and the spatial distributions. Previous studies classified factors that determine fertility levels into background and proximate in nature (Davis and Blake 1956). Socioeconomic and demographic factors such as education, wealth index, type of residence, age of respondents and age at first marriage constitute background factors while the proximate factors feature biological and behavioural variables such as contraceptive use, postpartum amenorrhoea, postpartum abstinence, menopausal age, induced abortion, frequency of sexual intercourse and intrauterine mortality (Bongaarts et al. 1984; Bongaarts 2015).

Introduction

Analytically, children ever born (CEB) by each woman of reproductive age have been used as proxy for modelling fertility (Becker and Lewis 1973; Kazembe 2009). Being a count variable, CEB is normally modelled through appropriate count models such as the log-linear Poisson or negative binomial model. However, due to the presence of a large proportion of women who would not have given birth to any child, the variable is often characterized with excessive zeros and sometimes over- or under-dispersion, a situation where the mean is not equal to the variance. Hence, the basic assumptions of the standard count models would be violated thereby rendering them unsuitable for consideration. Hurdle count model, which is a discrete mixture of two components, a point mass at zero and a truncated count distribution for the non-zero observations, can be considered a suitable alternative to cater for the unique property of the fertility data (Hu et al. 2011). The Poisson hurdle model assumes that the zero observations are realized through structural source and follow binary distribution, while the non-zero observations originate from sampling sources and assumed to have truncated Poisson distribution (Neelon et al. 2013).

Fertility data collected through survey are usually realized through stratified cluster sampling techniques, giving rise to similarities in fertility behaviour among women within clusters, due to shared behavioural and cultural beliefs, but substantial variations could exist between clusters due to possible differences in fertility risk exposure. This important property of the data needs to be accounted for in any modelling framework to ensure, among others, proper parameter estimation and inference. Therefore, a latent variable that could account for these unobserved covariates as random effects needs to be introduced into the modelling process (Dunson 2008). It is also possible that geographical locations are spatially formulated such that there are similarities in the fertility behaviour of individuals from locations that share common boundaries or lie within the same neighbourhood, leading to spatial autocorrelation. To investigate this, a structured spatial component needs to be introduced into the analytical model. Furthermore, a large number of categorical and metrical covariates can be available for consideration as the determinants of fertility. The categorical covariates can be reasonably assumed to have linear effects on the response variable and thus, modelled parametrically, but nonlinear smooth functions need to be assumed for the metrical covariates in order to estimate them nonparametrically, more so studies have confirmed that these variables do not

always have linear relationships with demographic indicators (Cameron and Trivedi 2013; Gayawan and Adebayo 2013).

Investigations into the spatial distributions of fertility have revolved around modelling of the average fertility levels of women in Africa (Kazembe 2009; Alaba et al. 2017). However, information from other higher moments of the distribution assumed for the demographic indicator, which is not considered, could provide useful information that can assist in policy formulation. Within the concept of distributional regression (Klein et al. 2014; Rigby and Stasinopoulos 2005), it is possible to link different covariates to parameters of the response distribution beyond just the mean, as this would allow us to establish how the covariates impart on other moments of the distribution. In the case of Poisson hurdle model, one can link the available covariates to the zero-truncated and mean parameters of the model thereby allowing us to examine the likelihood of having zero birth and the average level of fertility. Further, the possibilities of extending the linear parameters of the regression model to structured additive predictors (Fahmeir et al. 2013) renders it possible to consider spatial structured and unstructured random effects, smooth functions for nonlinear effects of metrical covariates and the usual linear effects for categorical variables in a single framework. Thus, the goal of this study is to simultaneously investigate the spatial distributions, at a disaggregated level of state, of the likelihood of having no births and in the average level of fertility among women of reproductive age in Nigeria.

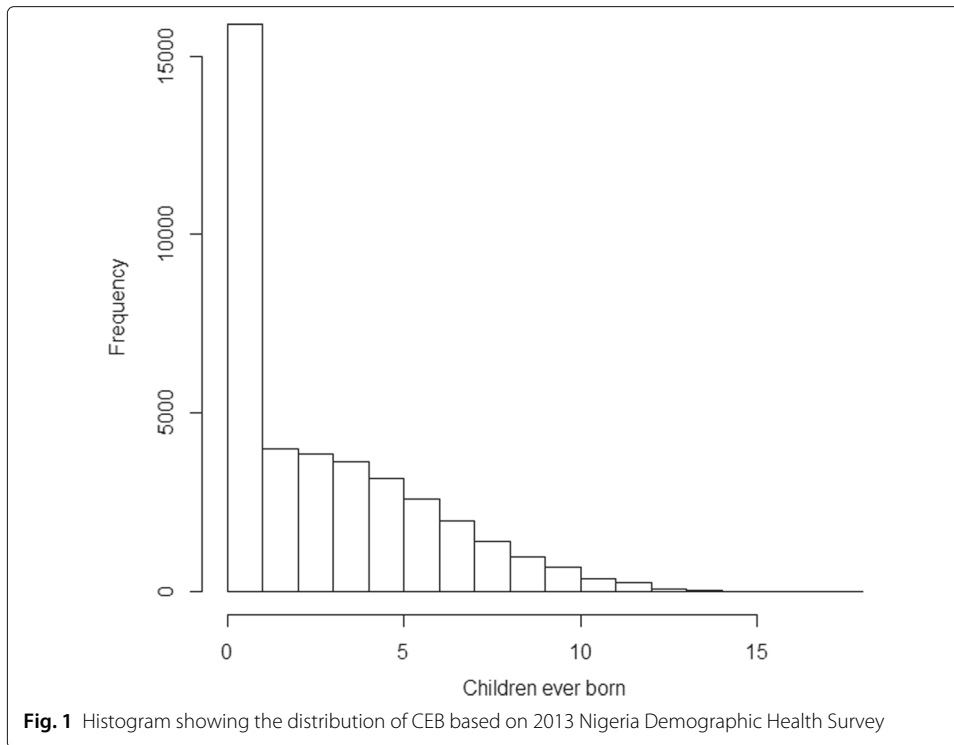
The rest of this article is structured as follows: the next session presents the data and statistical methodology followed by the results and discussion session. Lastly, we present some concluding remarks.

Methods

Data

Data for this study were obtained from the 2013 Nigeria Demographic and Health Survey (NDHS) available upon request at <https://dhsprogram.com/>. Nigeria is made up of 36 states and a Federal Capital Territory (FCT), Abuja. Each state is divided into Local Government Areas (LGAs) and the LGAs are subdivided into localities called enumeration areas (EAs). The EAs formed the basis on which the primary sampling units, known as clusters, were developed for the survey. A stratified triple-stage cluster design was adopted for the survey. Altogether, 904 clusters comprising of 372 in urban and 532 in rural areas were selected. The sampling frame was the list of all the non-institutional houses in each cluster. A sample comprising of 45 households was selected from each cluster and all women aged 15 – 49 years that resided permanently or came to visit on the night before the survey were eligible for interview. Out of the 40,320 households that were selected, 38,904 were found to be occupied. A total of 39,902 women were eligible for individual interview and 38,948 were successfully interviewed. Detailed report about the sampling scheme of the survey can be found in NPC (2014).

The survey was designed to collect information on the respondents' reproductive history, family planning behaviour, knowledge and attitude towards HIV/AIDS and utilization of healthcare services among others. The women respondents were asked to name all the children they had ever given birth to whether dead or alive. Figure 1 presents a histogram showing the frequency distribution of CEB per woman in Nigeria. The figure lends credence to the existence of massive zeros in the fertility data. Figure 2 a and b



shows the labelled maps of Nigeria indicating the distribution of average CEB by states and the percentage distribution of women aged 15 – 49 years without any child, respectively. Table 1 presents the percentage and other descriptive statistics of CEB for all the women respondents and percentage distribution for women without any child based on the socioeconomic variables included in the study.

Hurdle Poisson model

Hurdle Poisson model analyses count data with a larger than expected number of observed zeros as assumed under standard count distribution models (Neelon et al. 2013).

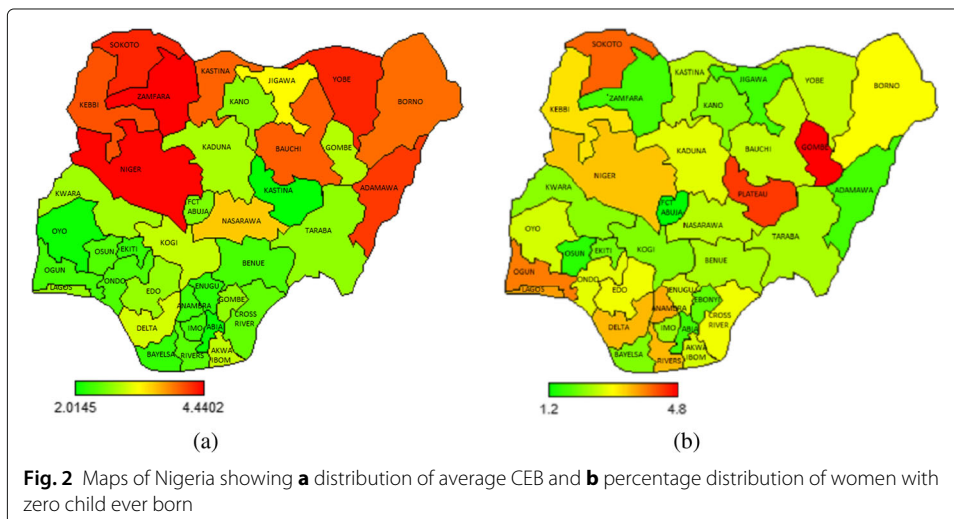


Table 1 Descriptive statistics of children ever born and women with no children

Covariates	CEB (all women)				Women with zero CEB	
	Respondents	Percent	Mean	SD	Respondents	Percent
Place of residence						
Rural	23403	60.1	3.44	3.14	5730	49.8
Urban	15545	39.9	2.50	2.75	5767	50.2
Education level						
No education	13740	35.3	4.42	3.25	1788	15.6
Primary	7104	18.2	3.93	2.91	1151	10.0
Secondary	14407	37.0	1.69	2.22	6932	60.3
Higher	3697	9.5	1.70	2.06	1626	14.1
Wealth index						
Poorest	6602	17.0	4.33	3.40	1059	9.2
Poorer	7515	19.3	3.64	3.19	1679	14.6
Middle	8001	20.5	3.07	3.01	2413	21.0
Richer	8450	21.7	2.62	2.76	2952	25.7
Richest	8380	21.5	1.99	2.27	3394	29.5
Use of contraceptives						
Ever used	9420	24.2	3.07	2.61	2204	19.2
Never used	29528	75.8	3.06	3.15	9293	80.8
Contraceptives by method						
No method	32723	84.0	3.09	3.10	9773	85.0
Traditional	1799	4.6	3.35	2.54	320	2.8
Modern	4426	11.4	2.79	2.66	1404	12.2
Marriage status						
Never married	9820	25.2	0.10	0.42	9076	78.9
Currently married	27274	70.0	4.07	2.88	2287	19.9
Formerly married	1854	4.8	4.02	2.69	134	1.2
Employment status						
Not employed	14942	38.4	1.91	2.73	7423	64.6
Employed	24006	61.6	3.78	2.98	4074	35.4

It assumes that the zero observations come from structural sources. Its probability density function (pdf) is given as follows:

$$P(y | \pi, \lambda) = \begin{cases} \pi & y = 0 \\ \frac{(1-\pi)}{1-\exp(-\lambda)} \frac{\lambda^y \exp(-\lambda)}{y!} & y > 0 \end{cases} \tag{1}$$

where y is the response variable indicating CEB in this case, π is the probability for the binary component that indicates whether or not the CEB for a particular woman is zero and λ is the mean CEB. The model is therefore a discrete mixture of two components that comprise of a point mass at zero and a truncated Poisson distribution for the non-zero observations. Each of the two components can be expressed as follows:

$$Pr(Y_i = 0) = \pi \tag{2}$$

and

$$Pr(Y_i = k) = (1 - \pi) \frac{\lambda^k e^{-\lambda}}{k!(1 - e^{-\lambda})} \quad k = 1, 2, \dots \tag{3}$$

where Y_i denotes the response for woman i , $i = 1, \dots, 38,948$. The mean and variance are given as $E(Y_i) = \frac{(1-\pi)\lambda}{1-e^{-\lambda}}$ and $\text{Var}(Y_i) = \frac{(1-\pi)\lambda}{1-e^{-\lambda}} + \frac{\lambda^2(1-\pi)(\pi-e^{-\lambda})}{1-e^{-\lambda^2}}$, $0 < \pi < 1$ and $\lambda > 0$, respectively.

Given that π is the probability of zero-CEB, whenever $(1-\pi) > e^{(-\lambda)}$, the data set is said to be zero-inflated with respect to a standard Poisson distribution, but if $(1-\pi) < e^{(-\lambda)}$, then we have a zero-deflated issue. It is said to be equi-dispersed if $(1-\pi) = e^{(-\lambda)}$ (Hu et al. 2011; Neelon et al. 2013). In situations where $\pi = 1$, it signifies that every woman had zero-CEB, resulting to a point mass distribution, but when $\pi = 0$, every woman in the system is fertile, and the hurdle model reduces to a truncated Poisson distribution.

Assume that in addition to the response variable y_i , CEB, a generic vector of covariates of different types, \mathbf{v}_i , $i = 1, \dots, n$ has been made available. In structured additive distributional regression, each parameter of the response distribution ϑ_l where $\vartheta_1 = \pi$ and $\vartheta_2 = \lambda$ in the case of the hurdle Poisson is related to a semiparametric predictor $\eta_i^{\vartheta_l}$ defined in terms of the vector of covariates \mathbf{v}_i . A suitable response function is used to map the predictor to the parameter of interest, $\vartheta_{il} = h^{\vartheta_l}(\eta_i^{\vartheta_l})$. The superscript entails that the l^{th} parameters or functions of parameters of the distribution of the response variable, rather than just the mean, are linked to covariates. For the observation $i = 1, \dots, n$, a suitable structured additive predictor for the l^{th} parameter ϑ_l is given as follows:

$$\eta_i^{\vartheta_l} = f_j^{\vartheta_l}(z_{ij}) + \mathbf{x}'_i \beta^{\vartheta_l} + f_s^{\vartheta_l}(s) + u_i^{\vartheta_l} \tag{4}$$

where z_{ij} is the j^{th} metrical covariates such as the woman's age and years of marriage and $f_j^{\vartheta_l}$ is a smooth function assumed for the covariate; \mathbf{x}'_i is a vector of categorical covariates such as the woman's education and contraceptive use and β is a vector of regression coefficients; $f_s^{\vartheta_l}(s)$ is the spatial effect for state s where the woman resides, and $u_i^{\vartheta_l}$ accounts for unstructured random effects at cluster level.

The Bayesian method is the natural choice for estimation models of this nature. This is due to the capability of the technique to analyse complex statistical models with the aid of the computer-based Markov chain Monte Carlo (MCMC) technique. It assumes all parameters to be random variables upon which suitable prior distributions are assigned.

A flat, non-informative prior was assumed for the linear terms. This allows the estimation to be similar to the classical approach. For the nonlinear effects of the metrical covariates, Bayesian penalized splines (P-splines) with second order random walk was adopted (Eilers and Marx 1996; Brezger and Lang 2006). The prior allows for the non-parametric estimation of f_j as a linear combination of B-splines (basis splines)

$$p(z) = \sum_{g=1}^G \alpha_g B_g(z) \tag{5}$$

where $B_g(z)$ are B-spline basis functions that are constructed from piecewise polynomials of a certain degree upon an equidistant grid of knots and the coefficients α_g are further defined to follow a random walk of certain order. Relying on findings from previous studies (Brezger and Lang 2006; Lang and Brezger 2004), cubic splines, 20 equidistance knots and second order random walk, i.e $\alpha_g = 2\alpha_{g-1} - \alpha_{g-2} + \epsilon_g$ with identically distributed noise $\epsilon_g \sim N(0, \tau_{\epsilon_g}^2)$ have been considered appropriate. For the random effect component of the model, an exchangeable normal prior was considered

$$u_i \sim N(0, \tau_{u_i}^2) \tag{6}$$

where the variance component $\tau_{u_i}^2$ incorporates over-dispersion and heterogeneity. In the case of the spatial effects for state of residence, which is a surrogate for many unobserved factors, Markov random field (Besag et al. 1991) was considered for the distinct regression coefficients corresponding to the different states. The prior is based on neighbourhood structure that defines areas as neighbour if they share common boundary. It is defined as follows:

$$[\beta_s \mid \beta_r, s \neq r, \tau_d^2] \sim N\left(\frac{1}{N_s} \sum_{r \in \Delta_s} \beta_r, \frac{\tau_d^2}{N_s}\right) \tag{7}$$

where $N_s = |\Delta_s|$ is the number of neighbouring states of s and $r \in \Delta_s$ denotes that the state r is a neighbour to the state s and τ_d^2 is a spatial variance that controls for smoothness. For all the smoothing variances, inverse gamma hyperprior, e.g. $\tau_d^2 \sim IG(a, b)$ are assigned in order to obtain data-driven amount of smoothness. Possible choices for the hyperparameters are $a = b = 0.001$ or $a = 1$ and $b = 0.005$. In this study, we performed sensitivity analysis by varying the choice of the hyperprior, but the results turn out to have only minimal differences. Results reported are for $a = b = 0.001$.

The complex nature of the likelihood functions of non-standard distributions utilized in distributional regression often lead to full conditionals for the unknown regression coefficients that are not analytically tractable. Thus, (Klein et al. 2014) developed a generic Metropolis-Hastings algorithm based on iteratively weighted least squares (IWLS) approximations for sampling from the full conditionals.

To analyse our data based on the above described distributional regression, three models of various specifications were considered, with intention of examining what could be gained by including variables of different types into the regression model. The first model was based on linking only categorical variables in linear form to the response variable. In the second, metrical covariates were added to the first and lastly, spatial and random components were included. The full model considered is expressed as follows:

$$\begin{cases} \eta^\pi = \beta_0^\pi + \text{urb}\beta_1^\pi + \dots + \text{empl}\beta_p^\pi + f_1^\pi(\text{age}) + f_2^\pi(\text{yrmarr}) + f_3^\pi(\text{agemarr}) + f_{sp}^\pi(s) + g(\text{cluster}) \\ \eta^\lambda = \beta_0^\lambda + \text{urb}\beta_1^\lambda + \dots + \text{empl}\beta_p^\lambda + f_1^\lambda(\text{age}) + f_2^\lambda(\text{yrmarr}) + f_3^\lambda(\text{agemarr}) + f_{sp}^\lambda(s) + g(\text{cluster}) \end{cases} \tag{8}$$

Each equation in (8) estimates the likelihood of zero birth and average level of fertility, respectively, among the women. The two equations were jointly estimated using the software BayesX version 3.0.2. For parameter estimation, MCMC simulations were based on 35,000 iterations with a burn-in sample of 5000 and thinning of every 30th observation yielding 1000 samples for parameter estimation. Convergence of the MCMC for each model was investigated through the trace plots and plots for posterior densities of the parameters. Model diagnostic was based on deviance information criterion (DIC), which is commonly used in Bayesian inference because of the ease of implementation from MCMC output (Klein et al. 2014; Spiegelhalter et al. 2002). It is expressed as

$$\text{DIC} = D(\bar{\theta}) + pD \tag{9}$$

where $D(\bar{\theta})$ is the posterior mean of the deviance and pD is the effective number of parameters, which is similar but not equal to degrees of freedom. Small values of $D(\bar{\theta})$ indicate a good fit while small values of pD indicate a parsimonious model, but small values of DIC imply a better model.

Results

Table 2 presents the estimated values of the model diagnostic criteria. Clearly, the model incorporating the linear, nonlinear, spatial and random effects outperformed the other two models considered. As expected, the diagnostic tool shows that model complexity increase with additional terms but yielding better fits. Presentations of results would therefore be based on the full model (M3).

Table 3 presents the estimates for the linear effects. The table presents the posterior means, standard errors and 95% credible intervals for the truncated Poisson and zero components of the hurdle model. Also included are the values of the variances of the random effects components. As expected, women who attained secondary or higher level of education have lower number of children compared with those without education, but the estimate for primary education is not significant. Women from richer and richest wealth quantile have fewer number of children relative to those of the poorest category but the results are not significant for those in the middle and poorer categories. Our findings show that current users of traditional or modern contraceptive methods have a higher number of children compared with those not using any method. Currently working women also have significantly higher number of children compared with their counterparts who were not working at the time of the survey, but results based on marital status and place of residence are not significant.

Results for the Bernoulli component show that the likelihood of having no child is significantly higher among women who attained higher level of education and those in the richer or richest wealth index. However, the likelihood of having no child is significantly lower among women with primary or secondary education, those currently using traditional or modern contraceptives and currently working women when compared with their other colleagues in their respective reference categories. The random components show higher estimates with wider credible intervals in the case of the Bernoulli component when compared with the truncated Poisson part. The community (cluster) random effect was 0.003 (CI 0.002, 0.005) and 0.036 (CI 0.146, 0.286) for the truncated Poisson and Bernoulli components, respectively, and 0.002 (CI 0.000, 0.005) and 0.015 (CI 0.001, 0.071), respectively, for the structured spatial effect.

Figure 3 a–c presents, for the truncated Poisson part, the nonlinear effects of age at first marriage/cohabitation, years spent in marriage and current age of the respondent, respectively. Presented are the posterior means (black) and 95% credible intervals (red). The results further confirm the existence of nonlinear relationship between demographic variables. Age at first marriage presents an approximately “U” relationship with mean CEB, having its base around age 33 years. The wide credible intervals between age 0 and 10 years indicate few respondents were available for these age group, as respondents who had not married were also considered in the study and assigned age 0. As for years of marriage, there is rapid rise in CEB in the first 7 years of marriage but this stabilizes

Table 2 Values of the deviance information criterion (DIC)

Model	Description	<i>D</i>	<i>pD</i>	DIC
M1	Linear effects	116,109.3	53.262	116,211.8
M2	Nonlinear + linear effects	109,782.3	91.677	109,965.7
M3	Nonlinear effects + linear + spatial effects	107,736.6	635.35	109,007.3

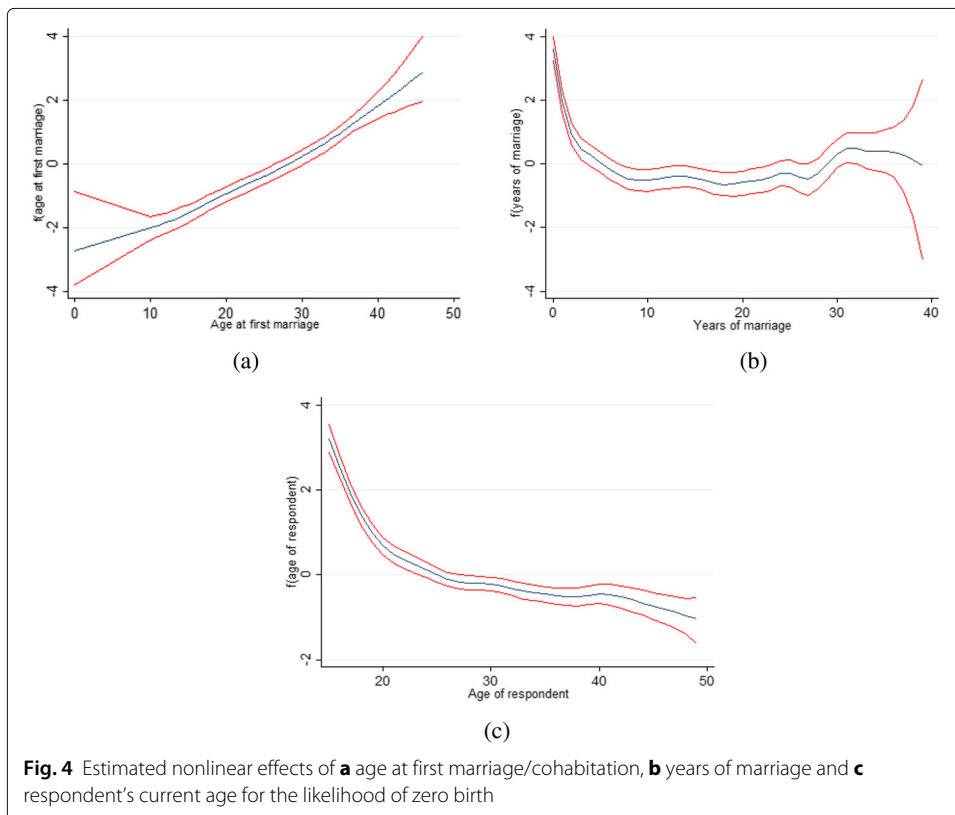
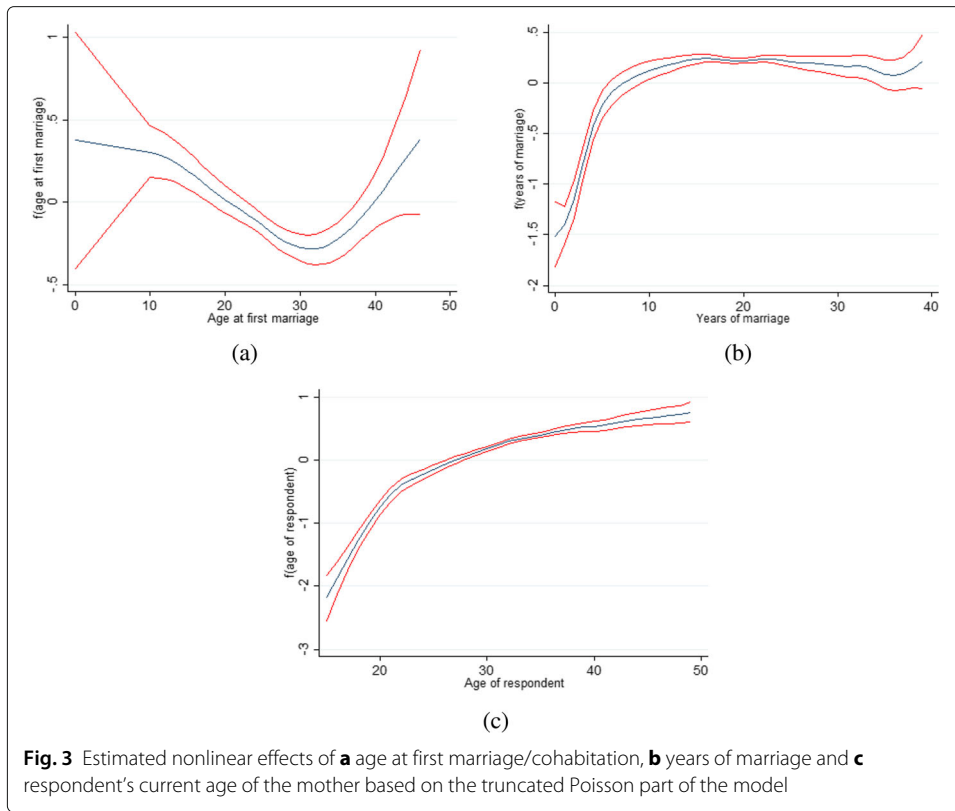
Table 3 Posterior summaries for hurdle Poisson modelling of the number of CEB per woman at 95% credible interval

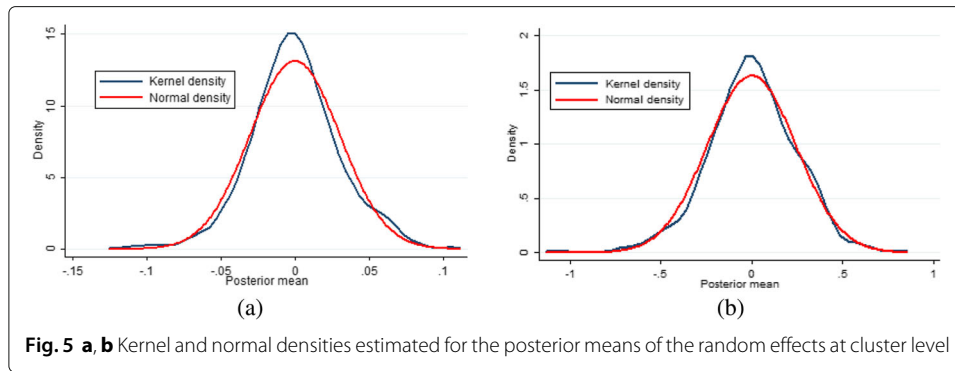
Covariates	Truncated-Poisson component: number of CEB per woman			Bernoulli component: probability of zero-CEB per woman		
	Mean	SD	95% CI	Mean	SD	95% CI
Constant	0.935	0.422	0.167, 1.818	0.183	1.777	- 0.533, 2.384
Place of residence						
Rural	0.000					
Urban	- 0.001	0.010	- 0.020, 0.020	0.010	0.075	- 0.164, 0.137
Education level						
No education	0.000					
Primary	- 0.001	0.009	- 0.017, 0.019	- 0.450	0.091	- 0.627, - 0.263
Secondary	- 0.077	0.012	- 0.100, - 0.055	- 0.296	0.092	- 0.475, - 0.121
Higher	- 0.208	0.018	- 0.244, - 0.172	0.315	0.116	0.091, 0.540
Wealth index						
Poorest	0.000					
Poorer	0.004	0.009	- 0.014, 0.022	0.151	0.086	- 0.012, 0.323
Middle	- 0.003	0.012	- 0.025, 0.020	0.134	0.097	- 0.063, 0.316
Richer	- 0.040	0.013	- 0.066, - 0.014	0.389	0.110	0.166, 0.593
Richest	- 0.131	0.017	- 0.163, - 0.098	0.691	0.121	0.455, 0.921
Contraceptives by method						
No method	0.000					
Traditional	0.101	0.015	0.073, 0.132	- 0.728	0.118	- 0.953, - 0.489
Modern	0.087	0.011	0.066, 0.109	- 0.511	0.080	- 0.663, - 0.353
Marriage status						
Never married	0.000					
Currently married	0.050	0.398	- 0.798, 0.780	- 3.527	0.786	- 4.683, - 1.609
Formerly married	- 0.149	0.398	- 0.991, 0.596	- 2.635	0.788	- 3.811, - 0.737
Working status						
No	0.000					
Yes	0.023	0.008	0.008, 0.037	- 0.179	0.051	- 0.277, - 0.080
Random effects: variance component						
State (unstructured)	0.011	0.005	0.004, 0.021	0.459	0.168	0.202, 0.845
State (structured)	0.002	0.001	0.000, 0.005	0.015	0.020	0.001, 0.070
Community effects	0.003	0.001	0.002, 0.004	0.214	0.036	0.146, 0.292

thereafter. Similarly, findings show evidence of rising CEB with current age of the women respondents.

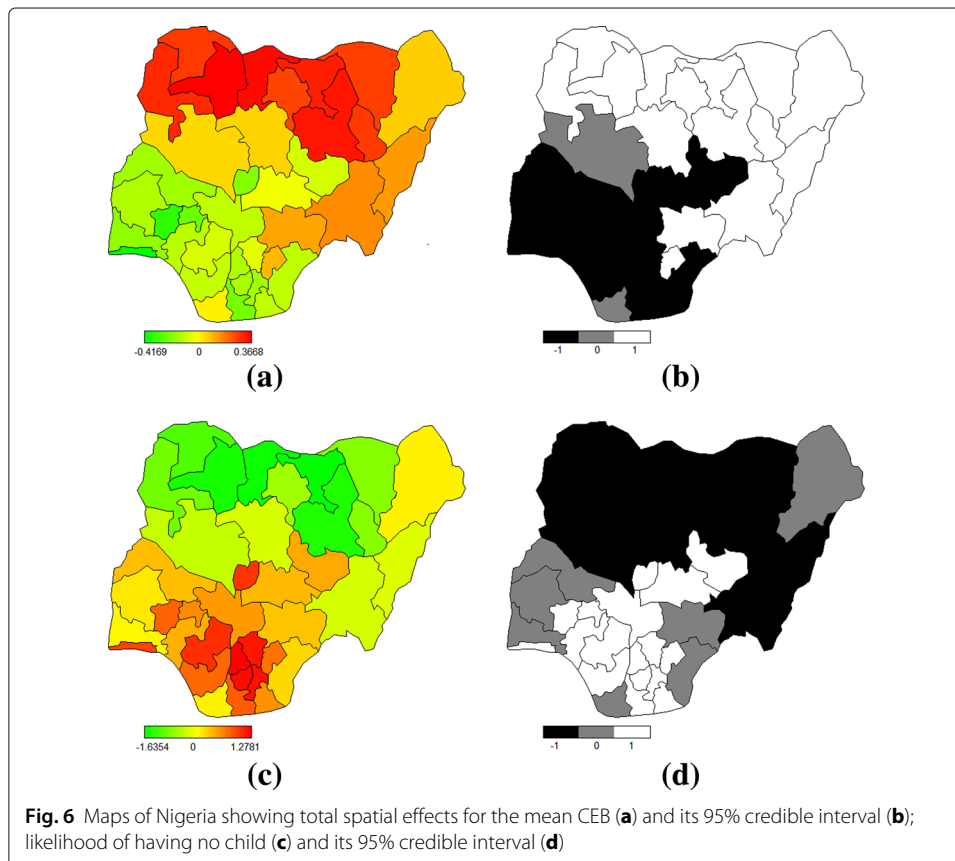
Figure 4 a–c shows the estimates based on the likelihood of having no children. Findings show that for every unit increase in age at first marriage, the likelihood of having no children similarly increases. However, in the case of years of marriage, the likelihood decreases sharply from 0 to around 8 years and then stabilizes for the rest of the years while, for age of the respondent, there is a sharp drop between age 15 and 20 followed by a gentle drop through age 49 years.

Figure 5 a and b presents, for the truncated Poisson and Bernoulli parts of the model, the kernel density and normal plots for the estimated random effects at community level. As evident from the plots, there is no departure from normality in the estimates of the random effects implying that variations in levels of CEB and in likelihood of a woman having zero birth are not strongly affected by variations among communities in Nigeria.





Results for the spatial components are presented in Fig. 6 a–d. The left panel (a and c) are the maps of the posterior mean estimates while the right panel (b and d) are those for the 95% credible intervals, which are used in determining the significance of the posterior estimates. From the maps of credible intervals, white (black) shading signifies states with strictly positive (negative) credible interval, while estimates for states in grey colour are not significant. The findings show evidence of north-south divide in the number of CEB and in the likelihood of a woman not giving birth to any child in Nigeria. Specifically, women from neighbouring Kebbi, Sokoto, Zamfara, Katsina, Kano, Jigawa, Kaduna, Bauchi, Gombe, Yobe Borno, Adamawa, Taraba, Benue, Abia and Ebonyi states have higher number of children but lower for those residing in FCT (Abuja), Kwara, Oyo, Osun,



Ekiti, Ogun Ondo, Lagos, Edo, Delta, Kogi, Nasarawa, Plateau, Enugu, Imo, Anambra, Rivers, Cross River and Akwa Ibom states. For the likelihood of zero CEB, as expected, states with high number of children are places that record significantly lower likelihood of zero CEB among the women. However, there are states where the number of CEB are significantly lower but estimates for the likelihood of zero birth are not significant. These states include Kwara, Oyo, Ogun and Cross River states.

Discussion

This study proposes the adoption of structured additive hurdle models for analysis of fertility data represented by CEB, which are often characterized by excessive zeros emanating from the structure of the data. The model formulation builds on existing methodological considerations in the analysis of the important demographic variable by simultaneously considering the non-zero part through a truncated Poisson model and then a zero point mass component through a Bernoulli distribution, thus allowing for the examination of available covariates on the two components. The choice method also allows for assessing the impact of different types of covariates both at individual and area levels, on the two parts of the fertility data. Results from the spatial components reveal some striking variations, indicating north-south divide, in fertility behaviour among the different states of Nigeria.

The interesting residual geographical differentials observed from the spatial effects, which persist after accounting for observed covariates, are surrogate of other unobserved factors which have exact strong spatial influence on fertility behaviour in Nigeria. For instance, the high fertility sustaining norms in parts of the country could be a major force. These could include some cultural and ethnic norms that inhibit women from implementing their desires for fewer number of children due to the regimental control of women by men in most northern states, a behaviour that is not rampant in other parts of the country (Health Communication 2015). Studies have equally reported that women in most northern Nigerian states deliberately give birth to many children to discourage their partners' tendency to divorce or to engage in plural marriage (Izugbara and Ezeh 2010). Insecurity and the huge disproportional level of poverty being experienced in Nigeria can also be a major driving force for the high differentials in the level of fertility across the country (Mbacké 2017). Further, the unequal distribution of reproductive health and family planning services across various locations within the country could as well breed lopsided fertility behaviour. In the northern states where CEB are observed to be high, there have been reported cases of strong resistance to family planning services. The prevalence of modern contraceptive use was estimated to be 3.2 and 4.3 for the states in North East and North West, respectively, as against 29.3 and 38 for South East and South West states, respectively (NPC 2014).

A number of findings from the other covariates considered in the study can as well be elaborated. Educated women are found to have fewer numbers of children and were more likely to have no children when compared with women with lesser or no education. This has similarly been observed in previous studies and could be attributed to the fact that women, in pursuit of educational achievements, focus more on their career than producing offspring (Alaba et al. 2017; Kravdal 2002; Becker and Lewis 1973; Olopha and Aladeniyi 2015). Education has been identified as a vital determinant of fertility with several causal relationships from a theoretical point of view (Leone 2004). The pathway

through which it affects fertility include delayed marriage and first birth (Gayawan and Adebayo 2013; 2014), increase bargaining power with husband, low regard for high fertility sustaining norms and appropriate use of contraceptives (Adebayo et al. 2013). These attributes can also hold true for women in the high wealth quantile. The findings therefore further strengthens the calls for encouraging girl child education up to at least secondary level as this could have multiplying effects not only on fertility but also on other maternal health outcomes.

Contraceptive use should normally be seen to reduce fertility levels, but this study found the users to have higher number of children. Though there are evidence of wide spread knowledge of family planning methods in Nigeria, the knowledge has not been translated to appropriate usage (Otoide et al. 2001). Family planning services are still regarded as exclusive for married women but sexually active unmarried women turn out to be the major users of modern contraceptive in the country. These women use it to differ pregnancy until later in the future (Abiodun and Balogun 2009; Tsui et al. 2017). The majority of the married women who use contraceptives in sub-Saharan Africa do so to space their children rather than for limiting (Westoff and Koffman 2010). Until this pattern of usage is changed in the country, the target of reducing fertility levels through contraceptive use would be difficult to achieve. Our findings further show working women to have higher number of children than the non-working ones. This is similar to findings in Malawi by Kazembe (2009). Participation in labour force ought to be an opportunity cost to childbearing because of the possible incompatibility between child rearing and formal work setting, but African women do not seem to be seeing participation in labour market as a threat to childbearing. There are also evidence that demonstrate that rising wage of skilled workers sometimes leads to rise in fertility of such workers (Day 2016). Moreover, the fact that the survey whose data was used considers women who practise subsistence farming as working women could however be a reason for our findings because these women pay high premium on children due to the perceived assistance they derived on their farms.

Marriage is supposed to be the estimated period of regular sexual relationship, but due to adolescent fertility transformation in Africa that see considerable proportion of fertility taking place outside marriage (Orubuloye et al. 1991; Smith-Greenaway 2016); increase in de-facto polygamy, modernized and dynamic nature of legalizing union, marriage institution had been reduced to a mere event instead of a life process, thus making it difficult to define when marriage has actually been consummated (Bledsoe 1990; Bledsoe and Cohen 1993). Premarital childbearing has become rampant not only in Nigeria but in other sub-Saharan African countries and there are signs that this may become increasingly due to rise in women's age at first marriage (Smith-Greenaway 2016). These factors could account for the non-significant estimates obtained for marital status based on the average number of children. Our findings on zero birth which reveal significantly less likelihood among the ever-married women is expected because childbearing expectations for married women in all parts of Nigeria remains high and thus women are expected to begin having children moments after they have been married (Solanke 2015).

Using Bayesian smooth functions to estimate the relation between CEB and metrical covariates has rendered it possible to observe the form of relationships among these variables and how they vary as time evolve. For instance, women who marry early and those who have spent longer periods in marriage would have elongated coital frequency which

would enhance their number of CEB in the face of low contraceptive use as evident from Figs. 3 a, b and 4 a, b. Also, CEB increases proportionally with age while the likelihood of zero-CEB reduce. This is due to the early childbearing as well as sustained childbearing within much of the reproductive life span (Caldwell et al. 1992). These are evidence of marked difference between the fertility behaviour of Nigeria and some developed countries, where fertility levels are only higher among younger women or largely confined with marriage (Caldwell et al. 1992; Bongaarts and Casterline 2013). Findings from the plots of random effects estimated at community levels are indications that cultural and socio-economic factors of neighbouring geographical locations influence their collective behaviour such that whenever fertility transition begins within a location, neighbouring communities sharing similar language or culture adjust to the new fertility pattern, which could be a reason for the approximately normal distribution observed from the plots in Fig. 5 (Kazembe 2009; Bongaarts 2017).

It is worthy of pointing out that the least spatial unit that was considered in this research is the state of residence, which is naturally composed of many geographical compartments. Consequently, within state variability in fertility can be expected, and thus, an analysis that considers smaller units might be desirable. Hence, further study should go into examining the spatial variation using a continuous spatial approach that utilizes point reference spatial data that are capable to identifying the exact locations where data were collected. This would unravel the spatial variations at more subdivided geographical entity rather than area level analysis. Also, since fertility is a dynamic component of population (NPC 2014), there is a need to include time scales as part of covariates for further study.

Conclusion

Spatial modelling remains an appropriate approach that presents easy to understand and to measure demographic variables among women across different geographical locations. This may be intractable with other quantitative and qualitative data analysis. Findings from the present study suggest the need for contextual area effects and individual-level factors in tackling the high fertility levels in Nigeria. The maps generated from the study and results of the covariates could be useful to policymakers in further designing evidence-based strategy for cutting down the levels of fertility at various locations in the country. The modelling approach that saw the simultaneous consideration levels of CEB through a truncated Poisson component and possibility of having zero child via a point mass of zero can also be extended to other research situations where the data suffers from excessive zero and over-dispersion in which case a truncated negative binomial model would replace the Poisson component.

Acknowledgments

The authors' appreciation goes to the DHS Program that granted access to use the data analysed in the study.

Authors' contributions

EG conceived the idea. OS collated the data. EG and OS participated in data analysis and report writing. Both authors read and approved the final manuscript.

Funding

The research did not benefit from any funding source.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Received: 18 December 2018 Accepted: 12 August 2019

Published online: 14 September 2019

References

- Abiodun, O.M., & Balogun, O.R. (2009). Sexual activity and contraceptive use among young female students of tertiary educational institution in Ilorin, Nigeria. *Contraception*, *79*, 146–149.
- Adebayo, S., Gayawan, E., Ujuju, C., Ankomah, A. (2013). Modelling geographical variations and determinants of use of modern family planning methods among women of reproductive age in Nigeria. *Journal of Biological Science*, *45*, 57–77.
- Alaba, O., Olubosoye, O., Olaomi, J. (2017). Spatial patterns and determinants of fertility levels among women of child bearing age in Nigeria. *South African Family Practice*, *59*(4), 143–147.
- Becker, G., & Lewis, H. (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy*, *81*, 279–288.
- Besag, J., York, J., Mollie, A. (1991). Bayesian image restoration, with two application in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*, 1–20.
- Bledsoe, C. (1990). Transformation in Sub-saharan African marriage and fertility. *The Annals of the American Academy of Political and Social Science*, *510*(1), 115–125.
- Bledsoe, C., & Cohen, B. (1993). *Social Dynamics of Adolescent Fertility in Sub-Saharan Africa*. Washington, DC: National Academies Press(US).
- Bongaarts, J. (2015). Modeling the fertility impact of the proximate determinants: Time for a time-up. *Demography Research*, *33*, 535–560.
- Bongaarts, J. (2017). Africa's unique fertility transition. *Population and Development Review*, *43*, 39–58.
- Bongaarts, J., & Casterline, J. (2013). Fertility transition: Is sub-saharan africa different? *Population and Development Review Population and Development Review*, (Suppl 1), 153–168.
- Bongaarts, J., Frank, O., Lesthaeghe, R. (1984). The proximate determinant of fertility in sub-saharan africa. *Population and Development Review*, *10*, 511–537.
- Brezger, A., & Lang, S. (2006). Generalized structured additive regression based on bayesian p-splines. *Computational Statistics & Data Analysis*, *50*, 967–991.
- Caldwell, J., Orubuloye, I., Caldwell, P. (1992). Fertility in Africa: A new type of transition? *Population and Development Review*, *18*, 211–242.
- Cameron, A., & Trivedi, P. (2013). *Regression Analysis for Count Data Regression: Cambridge University Press*. Econometric Society Monograph, second edition.
- Davis, K., & Blake, J. (1956). Social structure and fertility: An analytic framework. *Economic Development and Cultural Change*, *4*(4), 211–235.
- Day, C. (2016). Fertility and economic growth: The role of workforce skill composition and child care prices. *Oxford Economic papers*, *68*, 546–565.
- Dunson, D. (2008). *Random Effects and Latent Model Selection*, vol. 192. New York: Springer. www.springer.com. Accessed 23 May 2018.
- Eilers, P., & Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, *11*, 89–121.
- Fahmeir, L., Kneib, T., Lang, S., Marx, B. (2013). *Regression Methods, Models and Application: Springer*. www.springer.com. Accessed 27 May 2018.
- Gayawan, E., & Adebayo, S.B. (2013). A Bayesian semiparametric multilevel survival modelling of age at first birth in Nigeria. *Demographic Research*, *28*, 1339–1372.
- Gayawan, E., & Adebayo, S.B. (2014). Spatial Pattern and determinants of age at marriage in Nigeria using a geo-additive survival model. *Mathematical Population Studies*, *21*, 112–124.
- Health Communication, CapacityCollaborative (2015). Assessment of family planning use in Bauchi and Sokoto States, Nigeria, revised, 1–56.
- Hu, M., Paulicova, M., Nunes, E. (2011). Zero-inflated and hurdle models of count data with extra zeros: Examples from an hiv-risk reduction intervention trial. *The American Journal of Drug and Alcohol Abuse*, *37*(5), 367–375.
- Izugbara, C.O., & Ezech, A.C. (2010). Women and high fertility in Islamic northern Nigeria. *Studies in Family Planning*, *41*, 193–204.
- Kazembe, L. (2009). Modelling individual fertility levels in malawian women: A spatial semiparametric regression model. *Statistical Methods and Applications*, *18*, 237–255.
- Klein, N., Kneib, T., Klasen, S., Lang, S. (2014). Bayesian structured additive distributional regression for multivariate responses. *Journal of the Royal Statistical Society Series C*, *64*, 569–591.
- Kravdal, O. (2002). Education and fertility in sub-saharan africa: Individual and community effects. *Demography*, *39*, 233–250.
- Lang, S., & Brezger, A. (2004). Bayesian p-splines. *Journal of Computational and Graphical Statistics*, *13*, 183–221.
- Leone, A. (2004). *The effect of education on fertility: evidence from compulsory schooling laws*. Pittsburgh: University of Pittsburgh Press.
- Mbacké, C. (2017). The persistence of high fertility in sub-Saharan Africa: A comment. *Population and Development Review*, *43*, 330–337.
- Neelon, B., Ghosh, P., Loeb, P. (2013). A spatial poisson hurdle model for exploring geographic variation in emergency department visits. *Journal of Royal Statistical Society: Series A*, *176*, 389–413.
- NPC, ICF (2014). Nigeria demographic and health survey 2013, In *National Population Commission [Nigeria] and ICF International*. Rockville.
- Olopha, P., & Aladeniyi, O. (2015). Demographic analysis of the effect of some determinants of fertility on fertility intentions - the rural and urban factor. *Journal of Natural Science Research*, *5*, 18.

- Orubuloye, I.O., Caldwell, J.C., Caldwell, P. (1991). Sexual networking in the Ekiti district of Nigeria. *Studies in Family Planning*, 22(2), 61–73. <https://doi.org/10.2307/1966777>.
- Otoide, V., Oronsaye, F., Okonofua, F.E. (2001). Why Nigerian adolescents seek abortion rather than contraceptives: Evidence from focus group discussions. *International Family Planning Perspective*, 27, 77–81.
- Rigby, A., & Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*, 54, 507–554.
- Smith-Greenaway, E. (2016). Premarital childbearing in sub-Saharan Africa: Can investing in women's education offset disadvantages for children? *SSM-Population Health*, 2, 164–174.
- Solanke, B.L. (2015). Marriage age, fertility behaviour and women's employment in Nigeria. *SAGE Open*, 5(4), 1–9.
- Spiegelhalter, D., Best, N., Carlin, B., Van der Linder, A. (2002). Bayesian measures of model complexity and fit. *Journal of Royal Statistical Society*, 64, 583–639.
- Tsui, A.O., Brown, W., Li, Q. (2017). Contraceptive practice in sub-Saharan Africa. *Population and Development Review*, 43, 166–191.
- United-Nation (2017). Department of Economic and Social Affairs, Population Division (2017). World Population Prospects 2017 Ú Data Booklet (ST/ESA/SER.A/401).
- Westoff, C.F., & Koffman, D. (2010). Birth spacing and limiting connections, DHS Analytical Studies No. 21; USAID Contract No. GPO-C-00-08-00008-00.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
