Geo- Additive Modelling of Family Size in Nigeria

By
Oluwayemisi Oyeronke ALABA*
oluwayemisioyeronke@yahoo.com
Department of Statistics, University of Ibadan, Nigeria and Department of Statistics, University of South Africa
and
J. O. OLAOMI
olaomjo@unisa.ac.za
Department of Statistics, University of South Africa

Abstract

The household and family are the most fundamental socioeconomic institutions in human society. However, factors that affect the family size are affected by interrelated factors which vary across geo-political zones. The 2013 Nigerian Demographic Health Survey (NDHS) data was used to investigate the determinants of family size in Nigeria using the geo-additive model. The fixed effect of categorical covariates were modelled using the diffuse prior, P-spline with second-order random walk for the nonlinear effect of continuous variable, spatial effects followed Markov random field priors while the exchangeable normal priors were used for the random effects of the community and household. The Negative Binomial distribution was used to handle overdispersion of the dependent variable. Inference was fully Bayesian approach. Results showed a declining effect of secondary and higher education of mother, Yoruba tribe, Christianity, family planning, mother giving birth by caesarean section and having a partner who has secondary education on family size. Big family size is positively associated with age at first birth, number of daughters in a household, being gainfully employed, married and living with partner, community and household effects.

Keywords: Bayesian analysis; Negative binomial; posterior and prior distributions.

1. Introduction

Various sources support the contention that the family has changed as a result of the impact of industrialization and urbanization (Seward, 1974). Demographers have shown great concern on how many children is ideal for an average family or individual to have (Gustavus and Nam, 1970). Such information has been of great importance for trends in fertility. Considerable evidence from economically advanced countries has documented family size has a strategy to foster economic development and social well-being of the citizenry. The household and family are the most fundamental socioeconomic institutions in human society (Bongaarts, 2001). However, family size mechanism is undoubtedly conditioned by cultural, political and socio-economic setting (Anh et al., 1998, McCarthy and Oni, 1987). The dominant trend in most developed countries is a steady decline in household size from around 5 members in the middle of the 19th century to between 2 and 3 in 1990 (Bongaarts, 2001). From 1960 – 2013, the family size dropped from 3.67 to 3.12 in USA (www.statista.com). There is still a long way to go in Nigeria. In Nigeria, ideal numbers of children are 6.5 for all women and 7.1 for currently married women. Only 9% of women think three or less children is ideal (NDHS, 2013). Family size and total number of children ever born are used interchangeably in this work. However, family size pattern still remains a puzzle for demographers in the industrial world (Goldstein et al., 2003).
Model-based analyses are becoming important sources of global information, largely because of the absence of reliable national level empirical data in most sub- Saharan Africa countries. Family size has attracted researchers, some of these include: Keller (1973), Oppong (1974), Snyder (1974), Wood and Bean (1977), McCarthy and Oni (1987), Campbell (1993), Lehrer (1996) and Pillai (1984), Murphy and Wang (2001) and Adsera (2006). In-spite of the linear, nonlinear, spatial and random effect that exists among some variables, astonishingly such models are still lacking or scarce in literature to

simultaneously capture family size. It is therefore imperative to proffer solution to this question: what are the effects of fixed, nonlinear, spatial and unobserved heterogeneity on family size (a count variable) within the Bayesian context using a geo-additive model?

2. Geo- Additive Model

The model is given as

$$\eta_r = f_1(x_{r1}) + \ldots + f_k(x_{rk}) + f_{spat}(s_r) + u_r'\gamma + b_g \qquad (1)$$

Where

$f_{i,i=1,\ldots,k}$ is the nonlinear effect of metrical or continuous covariates x

$f(spat)$ is the spatially correlated effect of location s_r

u is the fixed effect of categorical variables γ

b_g $g \in \{1,\ldots.,G\}$ are uncorrelated (unstructured) random effects to model unobserved heterogeneity

2.1 Bayesian Prior distributions for covariate effects

For the continuous/metrical covariates, we assume Penalized Splines (P-spline) prior with second order random walk (Lang and Brezger, 2004; Fahmeir and Lang, 2001).

$$f(x) = \sum_{t=1}^{k} \alpha_t B_t(x) \qquad (2)$$

where

$B_t(x)$ are B-splines, α_t are defined to follow a first order or second order random walk prior. The second order random walk is given as

$$\alpha_t = 2\alpha_{t-1} - \alpha_{t-2} + \varepsilon_t \qquad (3)$$

with Gaussian errors $\varepsilon_t \sim N(0, \tau_\varepsilon^2)$ where τ_ε^2 controls the smoothness of f . This variance is estimated jointly with the coefficients of the basis function by assigning a weakly informative inverse Gamma prior with $\tau_\varepsilon^2 \sim IG(\varepsilon, \varepsilon)$. A suitable choice of diffuse prior is assumed for the fixed effect of categorical covariates given as

$$p(\gamma) \alpha\ const \qquad (4)$$

The spatial effects follow Markov random field priors (Besag *et al.*, 1991)

$$\left\{ f_{spat}(s_r) \mid f_{spat}(t); t \neq i, \tau_s^2 \right\} \sim N\left(\sum_{t \in \partial_i} \frac{f_{str}(t)}{N_i}, \frac{\tau_s^2}{N_i} \right) \qquad (5)$$

where

N_i is the sum of adjacent sites and ∂_i is the set of neighbours of site i

τ_s^2 is the spatial variance which controls the spatial smoothness

The random effects b_g were modelled from exchangeable normal priors, $b_{ij} \sim N(0, \tau_b^2)$

where τ_b^2 is the variance that accounts for overdispersion and heterogeneity. We assigned highly dispersed but proper prior for all variance components. An inverse Gamma distribution with hyperparameters a and b is chosen, such that $\tau^2 \sim IG(a,b)$. Standard choices of hyperparameters are a=1 and b=0.005 or a=b=0.001(which is close to Jeffrey's non-informative prior) (Fahmeir and Lang, 2001; Kazembe 2009). These values can be varied to know the sensitivity of the choices of hyperparameters to the inverse Gamma distribution.

2.2 Posterior distribution

Let α = (f, f_{spat}) and τ represent the vector of all variance components, and β is the vector of fixed effects parameters, then the posterior probability distribution is

$$p(\alpha,\tau,\beta / y) \;\; \alpha \;\; p(y/\alpha,\beta,\tau)\, p(\alpha)\, p(\beta)\, p(\tau) \tag{6}$$

where

$p(y|\alpha, \tau, \beta)$ is the likelihood function of the data given the parameters of the model (based on the dependent variable)

$p(\alpha)\,p(\beta)\,p(\tau)$ are the prior densities of all the parameters

The Deviance Information Criterion (DIC; Spiegelhalter *et. al*., 2002) is employed for comparison of the models.

Given by

$$DIC = \bar{D}(\theta) + pD \tag{7}$$

where

$\bar{D}$ is the posterior mean of the deviance

pD is the effective number of parameters (not equal to degrees of freedom)

Small values of $\bar{D}$ and pD indicate a better and parsimonious model. The model with the lowest DIC is the best. The Bayesian framework based on Markov Chain Monte Carlo (MCMC) simulation techniques from full conditional will be used for estimation of the unknown posterior distribution.

3. Data

The data used for this study were drawn from Nigeria Demographic and Health Survey (NDHS) for 2013 (www.measuredhs.com). The 2013 NDHS was conducted by the National Population Commission (NPC) with funding support from U.S Agency for International Development (USAID), the United Nations Population Fund (UNFPA), the United Kingdom Department for International Development (DFID). Technical support was provided by ICF International. The 2013 NDHS sample was selected using a three-stage stratified design consisting of 904 clusters, 372 urban areas and 532 in rural areas. In the 2013 NDHS dataset, 40,320 households were selected, out of which 38,522 were interviewed. In the interviewed households, 39,902 women in the childbearing age (15 – 49 years) and 18,229 men were found eligible for the interview. This represents a response rate of 99% for households, 98% for women and 95% for men. This study is based on the survey data with all participant identifiers removed. Although, different covariates on population and health issues in Nigeria were presented in the comprehensive and well detailed dataset, we focused on total number of children ever born as the dependent variable. The mean of the total children ever born is 4.35, variance = 6.786, skewness = 0.828, range=17. The data are over dispersed (Cameron and Trivedi, 1998). Equidispersion is often a mirage in real life studies, inappropriate imposition of Poisson regression will underestimate and overstate the significance of regression parameters (Ismail and Zamani, 2013).

The negative binomial distribution has been suggested as an alternative to the Poisson regression when the data are overdispersed (Paternoster and Brame, 1997; Osgood, 2000; Ismail and Jemain, 2007).
The socio- economic variables used as explanatory variables in explaining family size are: Category A: educational attainment, ethnicity, marital status, religion, place of residence, wealth index, family planning, number of daughters, number of dead children, method of delivery, work status, region and partner education.
Category B: age at first birth, BMI

4. Data Analysis and Presentation of Results

4.1 Data Analysis

Given

$$y_{ijk} / \gamma, b_{ij}, b_i \sim NB(\mu_{ijk})$$

We fit

$$\log(\mu_{ijk}) = \eta_{ijk} = w'\gamma + f(AGEFB) + f(BMI) + f(spat) + b_{i1} + b_{i2} \quad (8)$$

Where

η_{ijk} is the mean number of children ever born per woman

$w'\gamma$ is the vector of fixed effect of the categorical covariates of category A

$f(AGEFB), f(BMI)$ are the vectors of unknown smooth functions for BMI and AGEFB that are continuous and nonlinear

$f(spat)$ is the spatial effect

b_{i1} *and* b_{i2} are the community and household effects

We considered four models to investigate the best approach to family size modelling of (8). The first model (M1), we fixed all the categorical variables, AGEFB and BMI, such that their effects were estimated linearly. We used effect coding for all the categorical variables. In the second model (M2), we included the spatial effect to determine the magnitude of family size across the states. In the third model (M3), we introduced unobserved random effects of household and community while in (M4) explains the linear effect of the categorical variables, the nonlinear effect of continuous variables, the spatial effect and the unobserved random community and household effect.

Table 1: Summary of Diagnostic Accuracy of the Four Models **(** $\bar{D}$ is the posterior mean of the deviance, pD is the effective number of parameters, DIC is the deviance information criterion)

Model	Deviance($\bar{D}$)	pD	DIC
M1: All variables fixed	21808.642	24.659	21857.960
M2: All variables fixed + spatial effect	21728.681	47.053	21822.787
M3: All variables fixed + spatial + community effect	20622.962	425.056	21473.073
M4: All categorical fixed + nonlinear of continuous variable + community effect	20291.970	440.535	21173.041

The four models were implemented in BayesX version 2.1 (Belitz et al., 2012). We carried out 15000 iterations with the first 2000 considered as a burn-in sample. We thinned every 10th iteration of the remaining 13000 used for parameter estimation. Convergence and mixing were monitored through plotting and estimation of sampling paths and autocorrelation. Sensitivity analysis was carried out by varying the hyperparameters. The different choices of hyperparameters considered were a=1 and *b*=0.005, *a*=*b*=0.005 and *a*=*b*=0.001 (default). We reported the latter as the results were less sensitive to variation of the choices of the parameters (Gayawan & Adebayo, 2014).

4.2 Presentation and Discussion of Results

The primary outcomes of the four models were summarized in Table 1. Model 1 gave a parsimonious model of 24.659 effective number of parameters while the best model based on least DIC of 21173.041 for the Negative Binomial models is M4. The regression coefficients were almost similar in the other three models. Precision is enhanced in M4, therefore we present the results of M4 which gave the best fit. Results of the posterior negative binomial regression are given in Table 2. Regional differences are evident from the results, women from the North and South Eastern and South Southern parts tend to have more children. Women in the urban area have desire for large family size which actually negates documented literature. Education of mothers at higher level is inversely related to having a large family size with mean of -0.123. Low education (primary) showed desire for more children with mean of 0.0845 which supports the findings of Ali (1989) and Angeles et al. (2005). Women from the Ibo and Hausa ethnic groups tend to have more children than Yoruba women. The middle class wealth index showed desire for more children with mean of 0.006 while the richer and richest wealth index showed a reducing effect with mean of -0.009. Religion plays a significant role in family size, Christianity reduce the desire for a large family size which can be further explained by the fact that modern Christianity encourages monogamy which is further corroborated by Campbell (1993). Astonishingly, Islam which encourages polygamy showed a declining effect on family size with mean of -0.013. The negative effect of family planning on family size is well documented (Angeles et al., 2005). This study further supports the reducing effect of family planning on family size with mean of -0.039. One would not be surprised that a married woman who stays with her spouse will be at a higher risk of having more children as shown in our result with a mean value of 0.090 as opined also by Anh et. al.(1998). A positive relationship exists between partners education with only primary education and large family size while its negative for partners who have secondary school education which justifies the findings of Gustavus and Nam (1970) and Campbell (1993). However, from our results, partners' with higher education showed a positive association. This may be explained by the fact that higher education can be associated with higher income to cater for more children. Mothers who gave birth through caesarean sectioning or who have lost at least a child do not have desire for large family. The desire for more children is high for women who have only daughters. Infact, Ali (1989) concluded that until women have at least a son, the family size is incomplete. The negative significant results for the fixed effect at 95% Credible Interval (CI) are higher (-0.148, -0.097) and secondary education of mother, Yoruba tribe, Christianity, family planning, partner secondary education, caesarean section, child dead while the positive significant results are urban (0.003, 0.023), mother primary education , married and living with partner, mother is working and having daughters only.

Table 2: Posterior estimates of M4 within 95% Credible Interval (CI)

Variable	Mean	SD	95% CI
Constant	1.101	0.083	(0.913, 1.260)
Region			
North Central (ref.)	0		
North East	0.001	0.024	(-0.046, 0.050)
North West	-0.019	0.025	(-0.074, 0.029)
South East	0.045	0.027	(-0.004, 0.100)
South West	-0.004	0.026	(-0.054, 0.050)
South South	0.024	0.022	(-0.016, 0.070)
Place of Residence			
Rural (ref.)	0		
Urban	0.013	0.005	(0.003, 0.023)*
Mother's Educational Attainment			
No education (ref.)	0		
Primary	0.085	0.006	(0.072, 0.097)*
Secondary	-0.076	0.007	(-0.090, -0.063)**
Higher	-0.123	0.013	(-0.148, -0.097)**
Ethnicity			
Other ethnic groups (ref.)	0		
Yoruba	-0.039	0.014	(-0.065, -0.011)**
Ibo	0.012	0.017	(-0.023, 0.043)
Hausa	0.017	0.011	(-0.005, 0.038)
Wealth Index			
Poorest/Poorer (ref.)	0		
Middle Class	0.006	0.005	(-0.004, 0.016)
Richer/Richest	-0.009	0.007	(-0.023, 0.004)
Religion			
None/Traditional (ref.)	0		
Christianity	-0.031	0.010	(-0.051, -0.012)**
Islam	-0.013	0.011	(-0.034, 0.007)
Family Planning			
No method (ref.)	0		
Folkloric/Traditional/Modern	-0.040	0.004	(-0.048, -0.031)**
Marital Status			
Other (ref.)	0		
Married and living with partner	0.091	0.008	(0.076, 0.106)*
Partner's Education			
No education (ref.)	0		
Primary	0.002	0.006	(-0.009, 0.013)
Secondary	-0.048	0.006	(-0.059,-0.037)**
Higher	0.001	0.008	(-0.013, 0.017)
Mother's Working Status			
Not Working (ref.)	0		
Working	0.058	0.003	(0.052, 0.064)*
Mode of Delivery			
Normal delivery (ref.)	0		
Caesarean section	-0.049	0.012	(-0.072, -0.026)**
Sex of Children			
Boys (ref.)	0		
Daughters	0.277	0.003	(0.271, 0.283)*
Children Dead			
No (ref.)	0		
Yes	-0.214	0.003	(-0.219, -0.208)**
The continuous variables			
Age at first birth	0.006	0.005	(0.001, 0.019)*
Body mass index	0.001	0.001	(0.003, 0.004)*
The spatial variable			
States (36) and FCT	0.003	0.001	(0.001, 0.006)*
The Random effect			
Community	0.005	0.001	(0.004, 0.006)*
Household	0.001	0.001	(0.001, 0.002)*

** - Negatively significant *- Positively significant

The posterior nonlinear effect of BMI and age at first birth (in years) showed positive effect on family size with mean value of 0.001 and 0.006. However, the 95% CI for BMI is (0.003, 0.004) and age at first birth is (0.001, 0.019) which showed positive significant effect. Women in Yobe, Kano, Benue, Edo and Bayelsa have higher positive significant result of having more children while women in Kebbi, Niger, Kwara, Oyo, Osun, Ekiti and Lagos showed negative significant result of having a large family size (Fig.2). There is overall spatial effect on family size. Household and community effects were also positively significant in explaining family size.

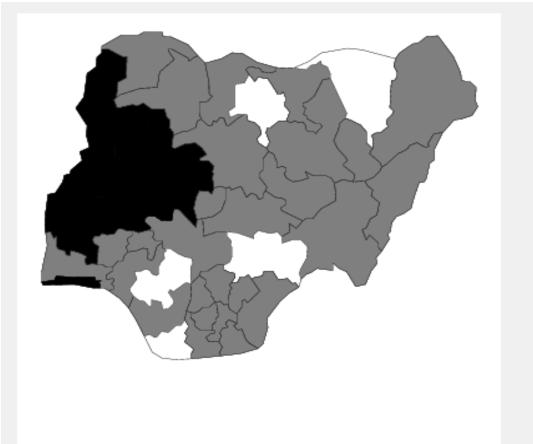

Fig 2: Spatial spread of family size in Nigeria

Fig 3: Labelled map of Nigeria

White denotes states with strictly positive CI (significant high risks), black denotes states with strictly negative CI (significant low risks) and grey denotes states with insignificant risk of unmet need of FP

5. Conclusion

This study revealed that education, ethnic group, religion, use of family planning, marrying a partner who is educated, loss of at least a child and giving birth by caesarean section explains low family size.

Acknowledgements: We appreciate the permission granted by www.measuredhs.com to use the Nigerian Demographic Health Survey (NDHS) 2013 data.

References

Adsera, A. (2006): Religion and Changes in Family-Size Norms in Developed Countries. *Review of Religious Research*. Vol. 47(3), pp. 271-286

Ali, S.M. (1989): Determinants of Family Size Preferences in Pakistan. *The Pakistan Development Review*, Vol. 28(3), pp. 207-231

Angeles, G., Guilkey, D.K. and Mroz, T.A. (2005): The Effects of Education and Family Planning Programs on Fertility in Indonesia. *Economic Development and Cultural Change*. Vol. 54(1) pp. 165-201

Anh, T. S., Knodel, J., Lam, D. and Friedman, J. (1998): Family Size and Children's Education in Vietnam. *Demography*. Vol. 35(1), pp 57 – 70

Belitz, C., Brezger, A., Kneib, T., Lang, S. Umlauf, N. (2012): BayesX Software for Bayesian Inference in Structured Additive Regression Models. Retrieved from www.stat.uni-muenchen.de/~bayesx

Besag, J., york, Y., Mollie, A. (1991): Bayesian Image Restoration with Two Applications in Spatial Statistics. *Annals of the Institute of Statistical Mathematics*, 43(1): 1- 59.

Bongaarts, J. (2001): Household Size and Composition in the Developing World. Policy Research Division, Population Council No. 144

Cameron, A.C. and Trivedi, P.K. (1998): Regression Analysis of Count Data, Cambridge Univeraity Press pp 4, 21

Campbell, E. K. (1993): Family Size Preferences of Men in the Western Area of Sierra Leone. Method and Determinants. *Genus*. Vol. 49 (1/2), pp. 181-199

:

:

:

:

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde A. (2002): Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society B,* 64(4): 1-34

Synder, D.W. (1974): Economic Determinants of Family Size in West Africa. *Demography*. Vol. 11(4), pp. 613-627

Murphy, M. and Wang, D. (2001): Family-Level Continuities in Childbearing in Low-Fertility Societies. *European Journal of Population / Revue Européenne de Démographie*. Vol. 17 (1). Special Issue on Lowest Low Fertility, pp. 75-96

Wood, C. H. and Bean, F.D. (1977): Offspring Gender and Family Size: Implications from a Comparison of Mexican Americans and Anglo Americans. *Journal of Marriage and Family*. Vol. 39(1), pp. 129-139