Geo-Additive Modelling of Family Size in Nigeria

Oyeronke Alaba O[1]* and Olaomi JO[2]

[1]*Department of Statistics, University of Ibadan, Nigeria and Department of Statistics, University of South Africa, Nigeria*
[2]*Department of Statistics, University of South Africa, Nigeria*

Abstract

We used the 2013 Nigeria Demographic Health Survey (NDHS) data to investigate the determinants of family size in Nigeria using the geo-additive model. The model was used to simultaneously measure the fixed, nonlinear, spatial and random effects. The fixed effect of categorical covariates were modeled 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 over dispersion 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. 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 inference; Geo-additive model; Family size; Nigeria; Negative binomial

Introduction

Before the 19th and 20th centuries, studies have shown that family size was related to ecological features and means of subsistence. The family size was an indicator of the wealth of a farmer (main occupation) in West Africa. Agricultural families are characterized by large extended families. However, because of economic changes and technology the traditional family systems are no longer totally dependent on agriculture [1]. Various sources support the contention that the family has changed as a result of the impact of industrialization and urbanization [2]. Demographers have shown great concern on how many children is ideal for an average family or individual to have [3]. 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 [4]. However, family size mechanism is undoubtedly conditioned by cultural, political and socio-economic setting [5]. Another line of thought is [6] that various factors influencing family-size desires are categorized into five: the costs and benefits of children; opportunity costs of childbearing; tastes and personal preferences; income and wealth; and childbearing itself. 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 [4]. 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 [7].

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 [8] used chi square on determinants of family size. Oppong [9] used ordinal scale to assess the degree of approval of closure of conjugal family. Snyder [10] employed the multiple linear regression and simultaneous equation model to investigate the economic determinants of family size. The pairity progression probabilities were utilized to examine how offspring affects family size [11]. McCarthy and Oni [6] used a Two-stage analysis to investigate the desired family size which they opined is affected by five categories [12-14], utilized the multiple linear regression using socio-economic determinants for ideal family size preference by men, observed linkages between religious composition of unions and fertility behaviour and age at family formation as a significant factor of family size using a cross-cultural comparison respectively. Murphy and Wang [15] employed simulation methods and multi-level model to investigate how successive generation affects the number of children born. Adsera [16] used a two-tailed test to explain the relationship between family size and factors associated with religion.

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?

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

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

Bayesian Prior distributions for covariate effects

For the continuous/metrical covariates, we assume Penalized Splines (P-spline) prior with second order random walk [17,18]

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

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 [19]

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

Where

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

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

The random effects b_g were modelled from exchangeable normal priors $b_{ij} \sim N(0, \tau_\varepsilon^2)$ where τ_b^2 is the variance that accounts for over dispersion 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) [18,20]. These values can be varied to know the sensitivity of the choices of hyperparameters to the inverse Gamma distribution.

Posterior distribution

Let $\alpha=(f, fspat)$ 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(α, τ, β/y) α P(y/α, β, τ) p(α) p(β) p(τ) (6)

where

p(y|α, τ, β) is the likelihood function of the data given the parameters of the model (based on the dependent variable)

p(α) p(β) p(τ) are the prior densities of all the parameters

The Deviance Information Criterion (DIC) [21] is employed for comparison of the models.

Given by

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

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 respectively. 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.

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 (Figure 1). The data are over dispersed [22]. Equidispersion is often a mirage in real life studies, inappropriate imposition of Poisson regression will underestimate and overstate the significance of regression parameters [23]. The negative binomial distribution has been suggested as an alternative to the Poisson regression when the data are overdispersed [24-26].

The socio- economic variables used as explanatory variables in explaining family size are educational attainment (EDUAT), body mass index (BMI), ethnicity (ETHNI), age at first birth (AGEFB), marital status (MARST), religion (RELIG), place of residence (RESID), wealth index (WEIND), family planning (FAMPL), number of daughters (DAUGH), number of dead children (CHDEA), method of delivery (DELIV), work status (WOKST), region (REGIO),state, community (COMUN), household (HHOLD) and partner education (PATED).

Data Analysis and Presentation of Results

Data analysis

Given

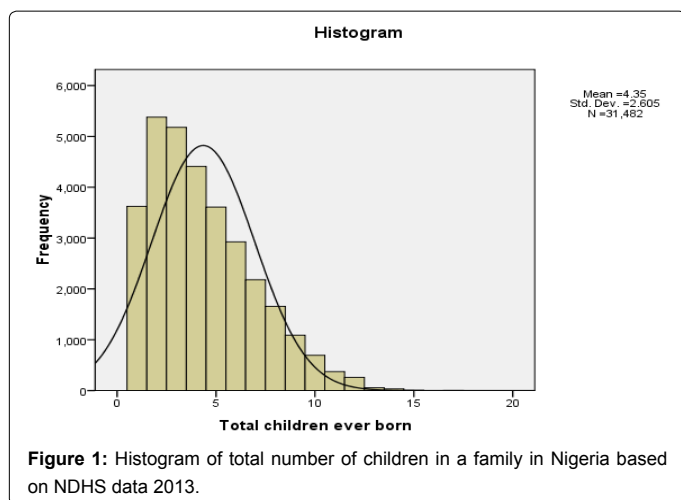$$y_{ijk} / \gamma, b_{ij}, b_i \sim NB(\mu_{ijk})$$

We fit

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

Where

$w'\gamma$ is the mean number of children ever born per woman

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

Figure 1: Histogram of total number of children in a family in Nigeria based on NDHS data 2013.

EDUAT, RELIG, MARST, RESID, WEIND,ETHNI, FAMPL, DAUGH, CHDEA, DELIV, WOKST, REGIO, PATED

$f(AGEFB), f(BMI)$ 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 respectively

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. The four models were implemented in Bayes X version 2.1 [27]. We carried out 15000 iterations with the first 2000 considered as a burn-in sample. We thinned every10th 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 [28].

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 numbers 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 [29,30]. 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 [12]. 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 [30]. 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 [5]. 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 [3,12]. 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 [29] 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 (-0.219, -0.208) 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.

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 respectively (Table 2). The posterior result showed that women who are obese (BMI >35.2) showed strong desire for large family size (Figure 2).

There is a decline in the desire for large family size as women grow older as depicted in (Figure 3). However, the 95% CI for BMI and age at first birth showed positive significant effect with (0.003, 0.004) and (0.001, 0.019) respectively. 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 (Figure 4). There is overall spatial effect on family size. Household and community effects were also positively significant in explaining family size.

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.

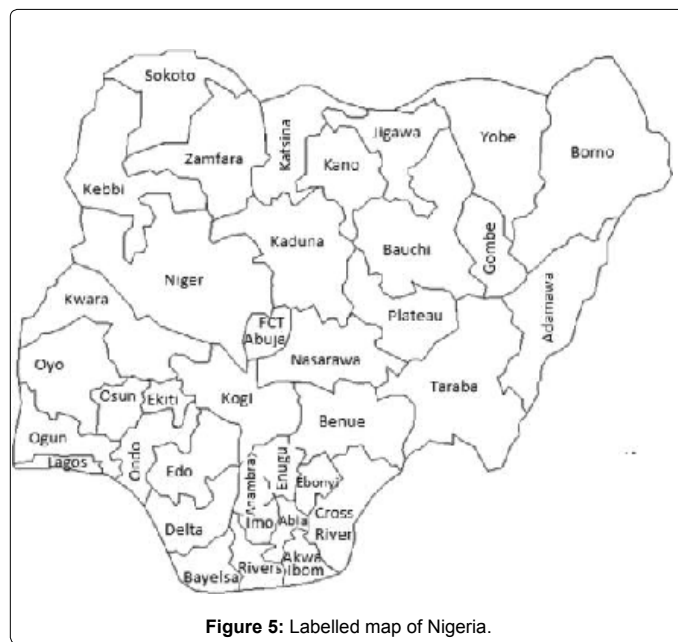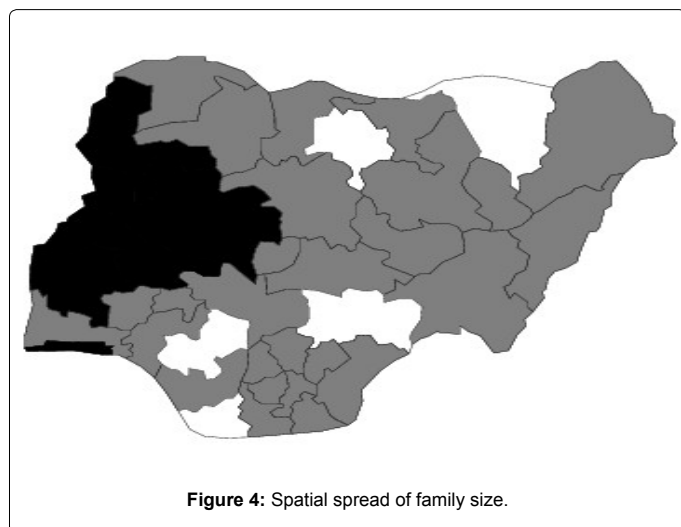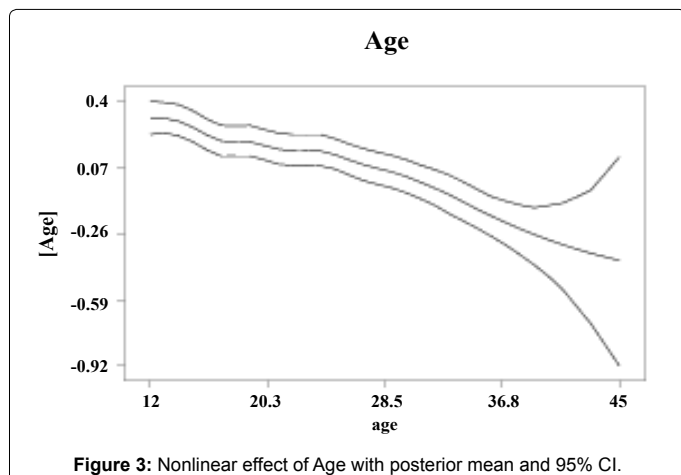| Model	$\bar{D}$	pD	DIC
M1: All variables fixed	21808.642	24.659	21857.960
M2: All variables fixed+spatial effect	21728.681	47.053	21822.787
M3: All variables fixed+spatial+community+ household effect	20622.962	425.056	21473.073
M4: All categorical fixed+nonlinear of continuous variable+spatial+community+hou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

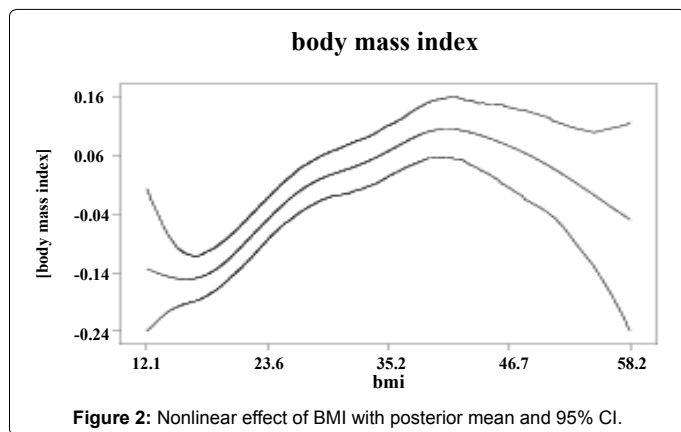**Table 1:** Summary of diagnostic accuracy of the four models.

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	0.01	(-0.051, -0.012)**
Islam	-0.013	0.011	(-0.034, 0.007)
Family Planning			
No method (ref.)	0		
Folkloric/Traditional/Modern	-0.04	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

Table 2: Posterior estimates of M4 within 95% credible interval (CI).

Figure 2: Nonlinear effect of BMI with posterior mean and 95% CI.

Figure 3: Nonlinear effect of Age with posterior mean and 95% CI.

Figure 4: Spatial spread of family size.

Figure 5: Labelled map of Nigeria.

Acknowledgment

References

1. Georgas J (2003) Family variations and changes across cultures. Center for Cross-Cultural Research. Western Washington University, Bellingham, Washington, USA.

2. Seward RR (1974) Family size in the united states an exploratory study of trends. Kansas Journal of Sociology 10: 119-136.

3. Gustavus SO, Nam CB (1970) The formation and stability of ideal family size among young people. Demography 7: 43-51.

4. Bongaarts J (2001) Household size and composition in the developing world. Population Studies 55: 263-279.

5. Anh TS, Knodel J, Lam D, Friedman J (1998) Family size and children's education in vietnam. Demography 35: 57-70.

6. McCarthy J, Oni GA (1987) Desired family size and its determinants among urban nigerian women a two-stage analysis. Demography 24: 279-290.

7. Goldstein J, Lutz W, Testa MR (2003) The emergence of sub-replacement family size ideals in europe. Population Research and Policy Review 22: 479-496.

8. Keller AB (1973) Psychological determinants of family size in a mexican village. Merrill-Palmer Quarterly of Behavior and Development 19: 289-299.

9. Oppong C (1974) Attitudes toward family type and family size in west africa a study of norms among ghanaian student population. International Journal of Sociology of the Family 4: 170-178.

10. Synder DW (1974) Economic determinants of family size in west africa. Demography 11: 613-627.

11. Wood CH, Bean FD (1977) Offspring gender and family size implications from a comparison of mexican americans and anglo americans. Journal of Marriage and Family 39: 129-139.

12. Campbell EK (1993) Family Size preferences of men in the western area of sierra leone. Method and Determinants. Genus 49: 181-199.

13. Lehrer EL (1996) Religion as a determinant of marital fertility. Journal of Population Economics 9: 173-196.

14. Pillai VK (1984) Age-at-family formation and family size a cross-cultural comparison. Journal of Comparative Family Studies 15: 291-298.

15. Murphy M, Wang D (2001) Family-level continuities in childbearing in low fertility societies. European Journal of Population/Revue Européenne de Démographie 17: 75-96.

16. Adsera A (2006) Religion and changes in family-size norms in developed countries. Review of Religious Research 47: 271-286.

17. Lang S, Brezger A (2004) Bayesian P splines. Journal of Computational Graph Statistics 13: 183-212.

18. Fahmeir L, Lang S (2001) Bayesian inference for generalized additive mixed models based on markov random field priors. Journal of the Royal Statistical Society 50: 201-220.

19. 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- 20.

20. Kazembe LN (2009) Modelling individual fertility levels in malawian women a spatial semiparametric regression model. Statistical Methods and Applications 18: 237-255.

21. Spiegelhalter DJ, Best NG, Carlin BP, Vander Linde A (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society B 64: 583-639.

22. Cameron AC, Trivedi PK (1998) Regression analysis of count data. (1stedn), Cambridge University Press, Cambridge.

23. Ismail N, Zamani H (2013) Estimation of claim count data using negative binomial generalized poisson zero-inflated negative binomial and zero-inflated generalized poisson regression models. Casualty Actuarial Society E-Forum 1-28.

24. Paternoster R, Brame R (1997) Multiple routes to delinquency? a test of developmental and general theories of crime. Criminology 35: 45-84.

25. Osgood DW (2000) Poisson based regression analysis of aggregate crime rates. Journal of Quantitative Criminology 16: 21-43.

26. Ismail N, Jemain AA (2007) Handling overdispersion with negative binomial and generalized poisson regression models. Casualty Actuarial Society Forum 103-158.

27. Belitz C, Brezger A, Kneib T, Lang S, Umlauf N (2012) Bayes X software for bayesian inference in structured additive regression models.

28. Gayawan E, Adebayo SB (2014) Spatial pattern and determinants of age at marriage in nigeria using a geo-additive survival model. Mathematical Population Studies 21: 112-124.

29. Ali SM (1989) Determinants of family size preferences in pakistan. The Pakistan Development Review 28: 207-231.

30. Angele G, Guilkey DK, Mroz TA (2005) The effects of education and family planning programs on fertility in indonesia. Economic Development and Cultural Change 54: 165-201.