Jabaru, S. et al., Yemeni J Med Sci. 2025; 19: 1 - 12
https://doi.org/10.20428/yjms.v19i1.2479
https://journals.ust.edu/index.php/yjms

ORIGINAL ARTICLE OPEN ACCESS

Analysis of Fertility Determinants and Regional Disparities in Nigeria using Geo-Additive Regression

Saheed Olalekan. Jabaru[1*], Kamoru Jimoh[2], Waheed Babatunde. Yahya[3]

[1] *Department of Physical Sciences, Faculty of Natural and Applied Sciences, Al-Hikmah University, Ilorin, Nigeria*
[2] *Department of Physical Sciences, Faculty of Natural and Applied Sciences, Al-Hikmah University, Ilorin, Nigeria*
[3] *Department of Statistics, Faculty of Applied Science, University of Ilorin, Nigeria*

ABSTRACT

This study applied geo-additive regression modelling on high-dimensional data with metrical and categorical predictors, using data from the latest Nigeria Demographic and Health Survey (NDHS-6). The sampled data comprises ninety-three predictors of total children ever born by eighty women of age group between fifteen and forty-nine years. Three penalty-based variable selection criteria—Elastic Net, Smoothly Clipped Absolute Deviation (SCAD), and Minimax Concave Penalty (MCP) were employed to reduce the dimensionality of the data. Geo-additive regression model were thereafter applied on the selected predictors to determine the impact of metrical, categorical and spatial predictors on the response variable. Findings revealed that predictors such as contraceptive use, age at first birth, and marital status are major determinants of total children ever born by a woman in Nigeria. Furthermore, spatial analysis revealed regional disparities in fertility rates within Nigeria, with notably higher fertility rate in northeastern states. This study's findings have broad applications across disciplines. By providing robust methodologies for handling complex datasets, this research supports evidence-based decision-making in public health, agriculture, and environmental policy. Ultimately, these findings contribute to efforts aimed at promoting sustainable development and enhancing maternal and child health outcomes in Nigeria and similar contexts globally.

Keywords: Children, Variable selection, Geo-additive, Regression, Fertility

[*] ***Corresponding author: Saheed Olalekan. Jabaru – Email: saheedjabar@alhikmah.edu.ng***

Jabaru, S. et al., Yemeni J Med Sci. 2025; 19: 1 - 12
https://doi.org/10.20428/yjms.v19i1.2479
https://journals.ust.edu/index.php/yjms

Introduction

Fertility is a critical demographic factor that significantly impacts population growth and socioeconomic development, particularly in developing nations like Nigeria. A key measure of fertility is the total number of children ever born (CEB) to a woman, which offers insights into both individual reproductive behaviour and societal trends[1]. Understanding the determinants of CEB is essential for implementing effective population management strategies, family planning initiatives, and public health policies. Nigeria ranks among the countries with the highest fertility rates globally, with notable regional disparities driven by diverse cultural, economic, and geographic factors. These complexities necessitate sophisticated analytical methods to accurately capture fertility patterns and their determinants [2].

Geo-additive regression modeling has emerged as an innovative approach to fertility analysis, offering a more nuanced alternative to traditional linear models. This methodology simultaneously incorporates fixed effects—such as demographic, socioeconomic, and cultural variables—and spatial effects, which address geographical variations[3]. Such a framework is particularly relevant for Nigeria, where stark geographic disparities influence factors like education, income, healthcare access, and cultural norms. By integrating spatial data, geo-additive models provide region-specific insights that conventional models often fail to capture[4].

Fertility in Nigeria varies significantly across regions. For example, northern regions consistently exhibit higher fertility rates compared to southern regions[2]. These differences can be attributed to a range of factors, including cultural norms, religious beliefs, limited access to contraceptives, and lower levels of female education. Traditional regression models often overlook these spatial variations, resulting in incomplete or biased interpretations of fertility trends. Geo-additive regression models address this gap by incorporating geographical information, allowing for a more comprehensive analysis of the interplay between location and fertility determinants[5].

Education plays a particularly influential role in shaping fertility patterns. Studies show that higher levels of female education correlate with reduced fertility rates, as educated women tend to delay marriage and childbirth and are more likely to use contraceptives effectively. Similarly, access to healthcare services and family planning resources varies widely across Nigeria, with rural areas experiencing significant disadvantages. Geo-additive models effectively map these disparities, enabling policymakers to identify and address region-specific barriers to fertility regulation[6].

Cultural and religious influences also shape fertility outcomes. Predominantly Islamic northern Nigeria, for instance, exhibits fertility patterns distinct from the predominantly Christian southern regions due to differing cultural expectations regarding family size and contraceptive use. Geo-additive regression modeling captures these variations by incorporating spatially contextual data, offering a more detailed understanding of fertility dynamics across the country.

In conclusion, geo-additive regression modeling represents a powerful tool for analyzing fertility, particularly in contexts characterized by pronounced regional disparities. By accommodating both fixed and spatial effects, this method enables a nuanced exploration of fertility determinants, providing valuable insights for designing targeted interventions. Such approaches are critical for addressing high fertility rates in Nigeria and fostering sustainable socioeconomic development[7].

Beyond spatial factors, a range of socioeconomic variables also impacts fertility in Nigeria. These include age at first marriage, education level, employment status, and wealth index[8]. For example, women with higher educational attainment generally have fewer children, while those in poorer socioeconomic circumstances tend to exhibit higher

https://doi.org/10.20428/yjms.v19i1.2479
https://journals.ust.edu/index.php/yjms

fertility rates due to limited access to family planning services[2]. Geo-additive modeling facilitates the analysis of these socioeconomic factors while accounting for spatial dependencies and geographic variations[9].

Although prior studies have employed various statistical methods to model fertility, many of these focus solely on socioeconomic factors or spatial patterns[10]. Geo-additive regression modeling, however, provides a more holistic approach, capturing the complexities of fertility behaviour by integrating both spatial and non-spatial factors. This approach can assist policymakers and researchers in identifying general determinants as well as region-specific influences, supporting more targeted and effective interventions[11].

This study seeks to apply geo-additive regression modeling to analyze the total number of children ever born to selected Nigerian women. The goal is to identify the critical socioeconomic and geographic factors influencing fertility in Nigeria, with particular attention to regional disparities. In doing so, the study aims to offer insights into the spatial dimensions of fertility and provide recommendations for policies to reduce fertility rates and improve reproductive health outcomes in Nigeria.

Methods

Data for this study were obtained from the 6th Nigeria Demographic and Health Survey conducted in 2018. Access to the data was granted after registration and necessary approvals were secured from the DHS program website: www.DHSprogram.com/Data. The response variable is the total number of children ever born (TCEB) by women in each household (whether dead or alive) at the time of the 2018 NDHS survey.

A total of 93 (p = 93) metric and categorical predictors were identified in the survey data as potential predictors of TCEB. Eighty (80) women were randomly selected without replacement, creating a high-dimensional dataset due to the large number of predictors. The number of women selected was constrained by the number of available predictors. The TCEB for the 80 selected women ranged from 1 to 6, with a mean of 3.04 and a variance of 2.80. Descriptions of the metrical and categorical predictors are provided in Table 1 and Table 2, respectively.

Data analysis

The initial task in modeling high-dimensional data involves reducing its dimensionality using appropriate criteria. This study employed penalty-based criteria for dimensionality reduction, including Elastic Net[12], Smoothly Clipped Absolute Deviation (SCAD)[13], and Minimax Concave Penalty (MCP)[14]. Data analyses were performed using R statistical software (Version 4.1.0), with the glmnet package (version 4.1-1) for Elastic Net feature selection and the ncvreg package (version 3.13.0) for SCAD and MCP criteria.

The cv.glmnet and cv.ncvreg functions, which utilize cross-validation (CV) folds, were used to search for the optimal shrinkage parameter (λ). Cross-validation involves partitioning the original sample into subsets, with some subsets serving as the training set and others as the test set. For this study, 4-fold cross-validation was employed, randomly dividing the observations into four equal groups. Each group was treated as a validation set in turn, with the method fitted on the remaining three groups. This process was repeated four times to ensure robustness.

The tuning parameter (α) for the Elastic Net criterion ranges from 0 to 1[15]. For this study, α values of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7 were used. Tuning parameters greater than 2 are recommended for SCAD, while tuning parameters greater than 1 was recommended for MCP[16]; this study however compared the variable selection performances of both SCAD and MCP at tuning parameters 3, 4, 5, 10, 15, 20, and 30.

https://doi.org/10.20428/yjms.v19i1.2479
https://journals.ust.edu/index.php/yjms

All possible combinations of cross-validation folds and tuning parameters were considered in the analysis. After dimensionality reduction, geo-additive models were fitted to the selected predictors using a Bayesian approach, with appropriate priors assigned to the model parameters. Monte Carlo Markov Chain (MCMC) simulations, involving 10,000 iterations, were used for posterior estimation. To minimize autocorrelation, thinning was applied to the Markov Chain by storing every 10th sampled parameter. The penalized likelihood method was used for estimating the parameters of the geo-additive model, allowing for simultaneous variable and model selection. This method determines whether:

1. A predictor should be included in the model,
2. A continuous variable should enter the model linearly or nonlinearly,
3. A spatial variable should be included in the model.

In the geo-additive model, a map object in boundary format is required to estimate the weights associated with neighbouring regions. This auxiliary object contains the boundary information of geographical maps; in this study, a map object of Nigeria was used. The geo-additive model was fitted using the R2bayesx package in the R statistical software (Version 4.1.0).

The three geo-additive models fitted to the selected significant predictors of TCEB are as follows:

Model 1R

$$\begin{aligned}\eta_1 &= \beta_0 + \beta_1(Previous_{state}) \\ &+ \beta_2(Current\_contraceptive) \\ &+ \beta_3(Marital\_status) + f(Number\_in\_household) \\ &+ f(Children\_under\_five\_years) \\ &+ f(Spatial) \qquad (1)\end{aligned}$$

Model 2R

$$\begin{aligned}\eta_2 &= \beta_0 + \beta_1(Previous\_state) \\ &+ \beta_2(Current\_contraceptive) \\ &+ \beta_3(Marital\_status) + f(Age\_at\_first\_birth) \\ &+ f(Number\_in\_household) \\ &+ f(Children\_under\_five\_years) \\ &+ f(Spatial) \qquad (2)\end{aligned}$$

Model 3R

$$\begin{aligned}\eta_3 &= \beta_0 + \beta_1(Respondent\_Region) \\ &+ \beta_2(Education\_level) + \beta_3(Previous\_state) \\ &+ \beta_4(Current\_contraceptive) \\ &+ \beta_5(Marital\_status) \\ &+ \beta_6(Respondent\_healthcare) \\ &+ f(Birth\_in\_last\_five\_yrs) \\ &+ f(Age\_at\_first\_birth) \\ &+ f(Number\_in\_household) \\ &+ f(Children\_under\_five\_years) \\ &+ f(Spatial) \qquad (3)\end{aligned}$$

Model performance metric

Model selection in this study utilized the improved Akaike Information Criterion with bias correction, (AICc). The model with the lowest AICc value is preferred, as it indicates a better fit with fewer parameters. AICc is an enhancement of the traditional Akaike Information Criterion (AIC) aimed at addressing potential biases in model selection, especially in small samples or complex models[17]. Traditional AIC can be biased when sample sizes are small or when the number of parameters is large relative to the sample size. This bias can lead to overfitting, where a model with more parameters appears better simply because it fits the sample data more closely. The corrected version of AIC adjusts for finite sample size by adding a correction term to the traditional AIC formula. AICc is particularly useful in fields such as ecology and biology, where sample sizes may be limited, and the risk of overfitting is high. However, as sample size increases, the correction term becomes negligible, and AICc converges to AIC. Therefore, AICc is considered a

https://doi.org/10.20428/yjms.v19i1.2479
https://journals.ust.edu/index.php/yjms

more general and robust criterion for model selection.

$$AIC = 2k - 2\log(L) \quad (4)$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad (5)$$

where:
k is the number of estimated parameters in the model.
n is the sample size
L is the likelihood of the fitted model given the data.

Results

Table 1: Description of metrical predictors

Predictor	Description
Respondent_Age	Age of the respondent in years
Number_in_household	Number of household members
Children_under_five_years	Number of children that are 5 years and below in the household
Births_in_last_five_yrs	Number of births in the last five years
Age_at_first_birth	Age of respondent at first birth in years
Age_first_cohabitation	Age of respondent at first cohabitation in years
Years_first_cohabitation	Years since the first cohabitation
Age_first_sex	Age of respondent at first sex in years
Time_since_last_sex	Time since last sex in months
Months_of_abstinence	Months of abstinence from sex by respondent
Age_of_partner	Husband/Partner's age in years
Preceding_birth_interval	Preceding birth interval of last child in years
Succeeding_birth_interval	Succeeding birth interval of last child in years

Table 2: Description of categorical predictors, their levels and percentage across levels

Predictor	Description	% of women by category
Respondent_Region	Geo-political zone of the respondent	North-West (17%) North-East (20.6%) North-Central (31.3%) South-West (10.3%) South-South (9.8%) South East (11%)
Residence_type	Type of place of residence	Urban (34.6%) Rural (65.4%)
Previous_residence	Type of place of previous residence	City (6.8%) Town (14%) Countryside (79.2%)
Previous_state	State of previous residence	36 States of Nigeria + FCT
Education_level	Highest educational level of the respondent	No education (49.9%) Primary (19.8%) Secondary (24.1%) Higher (6.1%)
Wealth_index	Wealth index	Poorest (24.4%) Poorer (23.1%) Middle (21.3%) Richer (18.2%) Richest (13%)
Had_terminated_pregnancy	Whether the respondent ever had a terminated pregnancy	No (84.6%) Yes (15.4%)
Current_contraceptive	Current contraceptive method used by respondent	Not using (84.2%) Pill (1.4%) IUD (0.9%) Injection (3.6%) Male condom (1.1%) Female sterilization (0.3%) Periodic abstinence (1.3%) Withdrawal (1.9%) Other traditional method (0.6%) Implants/Norplants (3.6%) Lactational Amenorrhea (1.1%) Female condom (0.00011%) Emergency contraception (0.1%) Other modern method (0.038%) Standard days method (0.036%)

Last_birth_CS	Last birth was through operation (CS)	No (98.4%) Yes (1.6%)
Marital_status	Current marital status of respondent	Never in union (1.2%) Married (89.4%) Living with partner (2.4%) Widowed (4.4%) Divorced (1.2%) Separated (1.3%)

Table 3: Number of features selected by Elastic Net ($0 < \alpha < 1$), SCAD ($3 \leq \gamma \leq 30$) and MCP ($3 \leq \gamma \leq 30$) at 4-fold cross-validation

Elastic Net		SCAD		MCP	
α	features selected	γ	features selected	γ	features selected
0.1	19	3	7	3	0
0.2	11	4	7	4	0
0.3	11	5	7	5	0
0.4	11	10	7	10	7
0.5	11	15	7	15	7
0.6	11	20	7	20	7
0.7	11	30	7	30	7

Table 4: Summary of categorical and metrical features selected by Elastic Net, SCAD and MCP

7 features	**11 features**	**19 features**
	Metrical predictors	
Number_in_ household	Number_in_ household	Number_in_ household
Children_under_ five_years	Children_under_ five_years	Children_under_five_ years
	Age_at_first_birth	Age_at_first_birth
		Births_in_last_five_yrs
	Categorical predictors	
Previous_state (2 levels: Akwa-Ibom, Nasarawa) Ref: Abia	Previous_state (5 levels: Akwa-Ibom, Delta, Kwara, Lagos, Nasarawa) Ref: Abia	Previous_state (6 levels: Akwa-Ibom, Delta, Kwara, Lagos, Nasarawa, Oyo) Ref: Abia
Current_contraceptive (1 level other traditional method) Ref: Diaphragm	Current_contraceptive (1 level: other traditional method) Ref: Diaphragm	Current_contraceptive (2 levels: pill, other traditional method) Ref: Diaphragm
Marital_status (2 levels: never in union, widowed) Ref: Divorced	Marital_status (2 levels: never in union, widowed) Ref: Divorced	Marital_status (3 levels: married, never in union, widowed) Ref: Divorced
		Respondent_Region (1 level: South-East)
		Education_level (2 levels: no education, primary) Ref: higher education
		Respondent_healthcare (1 level: Respondent partner) Ref: Husband/partner alone

The significant predictors of Total Children Ever Born (TCEB) selected using penalty-based criteria are shown in Table 3 and Table 4. The results indicate that for the Elastic Net criterion, the number of predictors selected remains consistent for tuning parameter values between 0.2 and 0.7. For the SCAD criterion, the number of predictors selected remains the same across all tuning parameter values. For the MCP criterion, the number of predictors selected is consistent for tuning parameter values of 10, 15, 20, and 30. However, no predictors were selected by MCP at tuning parameters of 3, 4, and 5.

These selected predictors were used to develop three geo-additive models by including the spatial variable of Nigerian states. Model 1R includes five predictors that were commonly selected by all criteria: previous state lived in by the respondent, current contraceptive use, marital status, number of people in the household, and number of children under five in the household. Model 2R includes all predictors from Model 1R plus the age of the respondent at first birth. Model 3R includes all predictors from Model 2R plus the respondent's region, education level, person responsible for the respondent's healthcare needs, and number of births in the last five years.

Table 5 provides the parameter estimates for predictors modeled parametrically as fixed effects in the three models. Model 2R has the lowest AICc, indicating it is the preferred model, and

https://doi.org/10.20428/yjms.v19i1.2479
https://journals.ust.edu/index.php/yjms

interpretations are based on this model. It was observed that women who previously lived in Delta, Enugu, and Kaduna States have 91.04%, 220.11%, and 50.68% higher TCEB, respectively, compared to those who previously lived in Abia State. Additionally, the age of the respondent at first birth, initially assumed to have a non-linear effect on TCEB has a linear effect. An increase in the age of a woman at first birth by one year reduces her TCEB by 4.52%.

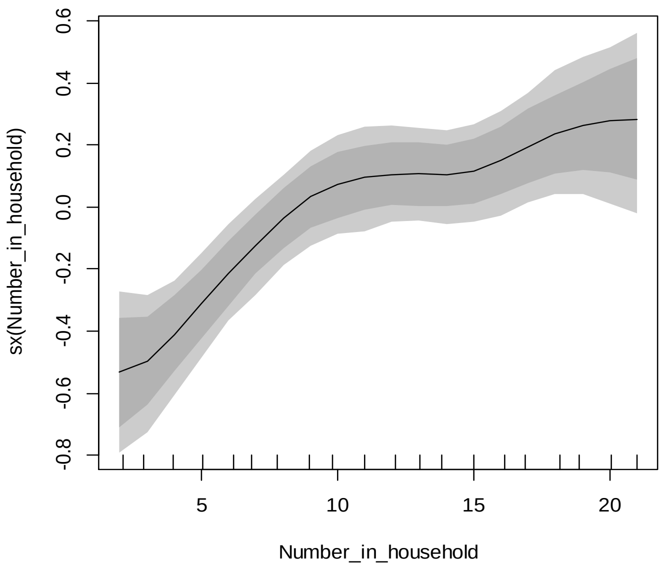

Figure 1: Non-linear effect of number of household members on TCEB in Model 2R

Figure 1 illustrates the non-linear effect of the metrical predictor "number of household members" on TCEB. It shows that TCEB increases with household size until it exceeds 15 members, after which a decrease is observed before it starts increasing again.

Table 5: Estimates of posterior means for fixed effects of Models 1R – 3R with 95% credible intervals

Predictors	Posterior Mean (β_i)	Standard Deviation	95% Credible Intervals: Lower	95% Credible Intervals: Upper	Incidence Rate (IR) =	% increase or decrease
MODEL 1R; AICc = -690.597						
Intercept	1.8291*	0.0532	1.7216	1.9256	6.2283	
Previous_state (Reference = Abia)	0				1	
Anambra	-1.2617	0.7553	-2.9463	0.0386	0.2832	-71.68
Delta	0.8803*	0.3654	0.1105	1.5902	2.4116	+141.16
Ekiti	-0.6904	0.3783	-1.5029	0.0050	0.5014	-49.86
Enugu	0.7119	0.3666	-0.0247	1.4171	2.0379	+103.79
Current_contraceptive *(Reference = Diaphragm)*	0				1	
Pill	-0.5779*	0.3125	-1.2961	-0.0337	0.5611	-43.89
MODEL 2R; AICc = -690.807						
Intercept	2.5726*	0.8702	2.0254	3.1573	13.0998	
Previous_state (Reference = Abia)	0				1	
Delta	0.6473	0.3567	-0.1213	1.3098	1.9104	+91.04
Enugu	1.1635*	0.3814	0.3455	1.8734	3.2011	+220.11
Kaduna	0.4100	0.2676	-0.1713	0.8902	1.5068	+50.68
Age_at_first_birth	-0.0463*	0.0162	-0.0809	-0.0147	0.9548	-4.52
MODEL 3R; AICc = -685.856						
Intercept	2.1615*	0.3216	1.5678	2.8323	8.6842	
Previous_state (Reference = Abia)	0				1	

https://doi.org/10.20428/yjms.v19i1.2479
https://journals.ust.edu/index.php/yjms

Delta	0.7273	0.3676	-0.0736	1.4093	2.0695	+106.95
Kaduna	0.4541	0.2695	-0.1194	0.9750	1.5748	+57.48
Education_level *(Reference = Higher)*	0				1	
No_education	0.2188	0.1137	-0.0024	0.4469	1.2446	+24.46
Age_at_first_birth	-0.0295	0.0162	-0.0625	0.0012	0.9709	-2.91

*Significant predictors at 5% significance level, whose 95% credible intervals do not contain zero

The spatial map in Figure 2a indicates that states in red have the highest TCEB, followed by states in grey, while states in blue have the lowest TCEB. Figure 2b displays the names of States in Nigeria with their respective geopolitical zone.

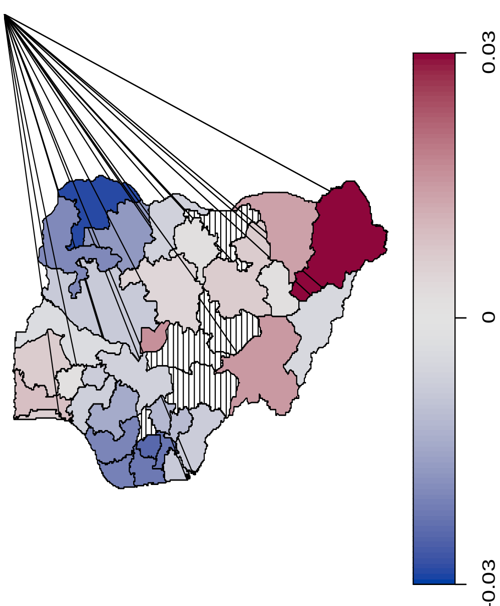

Figure 2a: Spatial effects of States in Nigeria on TCEB

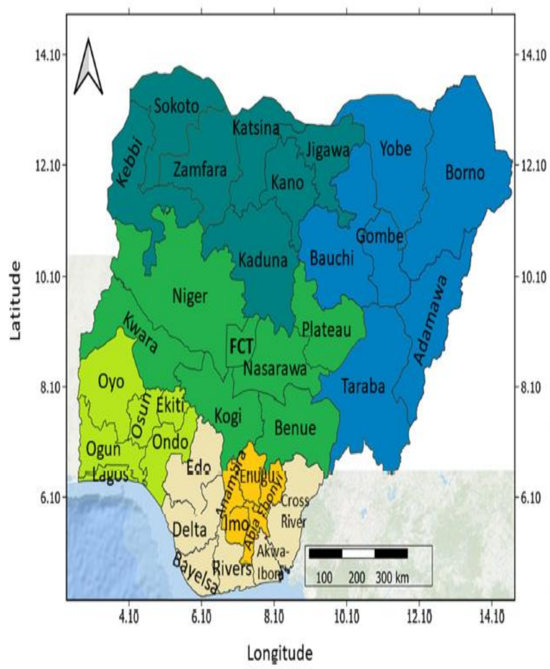

Figure 2b: Names of States in Nigeria

Discussion

Predictor Selection and Criteria Behavior

The Elastic Net criterion showed stability in the number of predictors selected across a wide range of tuning parameter values (0.2 to 0.7). Similarly, SCAD demonstrated consistent performance, selecting the same number of predictors across all tuning parameters. In contrast, MCP revealed a more variable pattern, with predictors selected only for higher tuning parameters (10, 15, 20, and 30), while failing to select any predictors at lower values (3, 4, and 5). This variability underlines the importance of tuning parameter selection in influencing model behaviour, as also emphasized by [16], who noted that tuning parameters significantly impact model sparsity and predictor stability.

Model Development

Three geo-additive models were developed by incorporating spatial variables of Nigerian states alongside the selected predictors:

- **Model 1R**: Focuses on five universally selected predictors, including previous state lived in, contraceptive use, marital status, household size, and number of children under five.

- **Model 2R**: Builds on Model 1R by adding the age of the respondent at first birth.
- **Model 3R**: Extends Model 2R by including region, education level, healthcare responsibility, and number of births in the last five years.

Preferred Model and Parameter Estimates

Based on the Akaike Information Criterion corrected (AICc), Model 2R emerged as the preferred model, providing a balance between complexity and explanatory power. This result aligns with that of [17], who emphasized the utility of AICc in selecting optimal models in ecological studies.

In Model 2R, key findings include:

- **Impact of Previous State of Residence**: Women from Delta, Enugu, and Kaduna States exhibit significantly higher TCEB compared to those from Abia State, indicating regional differences in fertility behaviour. This observation aligns with findings of [19], which highlighted the influence of regional socio-cultural factors on fertility.
- **Effect of Age at First Birth**: Contrary to the initial assumption of non-linearity, age at first birth has a linear effect on TCEB. Each additional year delays first childbirth, reducing TCEB by 4.52%. This finding corroborates studies like those by [20], which identified age at first birth as a critical determinant of fertility outcomes.

The results of the dimensionality reduction analysis indicate that the SCAD and MCP methods performed similarly across most tuning parameters. These findings are consistent with that of [21], who also compared penalized regression methods using logistic regression for low-dimensional data. Additionally, [22] observed similar performance between MCP and SCAD in their Monte Carlo simulations for low-dimensional data with ten-fold cross-validation, noting that Elastic Net provided more conservative estimates.

However, our results contradict [23], who found significant differences in the number of covariates selected by MCP and SCAD in low-dimensional datasets with normally distributed responses. This discrepancy may be attributed to the different data types used in each study. Conversely, [24] found that SCAD and MCP perform equally well in their assessment of penalized regression methods using simulated data with a normally distributed response, a finding that aligns with our study's use of a non-Gaussian response.

While much research has focused on variable selection and geo-additive modeling with low and high-dimensional data, few studies have addressed mixed predictors with count responses. This study fills that gap. One major challenge in regression analysis is the dimensionality of the data. High-dimensional data, where the number of explanatory variables exceeds the sample size, is common in fields such as machine learning [25], genetics [26], and medicine [27]. Such data requires sparse techniques to identify significant variables and eliminate redundant ones.

This study identified the use of contraceptives and the respondent's age at first birth as significant factors affecting the number of children ever born (TCEB), aligning with the findings of [28]. Marital status also emerged as a crucial factor, echoing the results of [29] and [30]. It is expected that being married or cohabiting increases the likelihood of pregnancy and childbirth. The spatial analysis revealed higher TCEB in the northeastern states compared to the South-south, consistent with [31].

Conclusion

This study evaluated the performance of different penalty-based criteria for dimensionality reduction and developed geo-additive models to predict the number of children ever born (TCEB) based on various predictors. The analysis demonstrated that

SCAD and MCP methods yielded similar performances across most tuning parameters, corroborating findings from previous studies. However, these results contrast with some studies that reported significant differences between these methods, likely due to differences in data nature. The inclusion of contraceptive use, age at first birth, and marital status as significant predictors of TCEB is consistent with existing literature, highlighting their crucial roles in fertility studies. The spatial analysis revealed higher TCEB in the northeastern states compared to the South-south, aligning with regional demographic patterns.

This study bridges a significant gap in research by addressing mixed predictors with count response, a relatively unexplored area. The findings emphasize the importance of appropriate dimensionality reduction techniques in handling high-dimensional data common in various fields, including genetics, machine learning, and medicine. By using penalty-based methods like Elastic Net, SCAD, and MCP, researchers can effectively identify significant predictors and enhance model accuracy. This study's approach and findings offer valuable insights for future research and practical applications in demographic and health studies, demonstrating the utility of geo-additive models in analyzing complex, high-dimensional datasets.

Acknowledgment

The authors thank National Population Commission (NPC) [Nigeria] and ICF for giving us access to the NDHS dataset that was utilized for this study.

Conflict of interest

The authors declare that no conflict of interest.

References

[1] Caldwell JC, Caldwell P. The cultural context of high fertility in sub-Saharan Africa. Popul Dev Rev. 1987;13(3):409–37.

[2] National Population Commission (NPC). Nigeria Demographic and Health Survey 2018. Abuja: NPC and ICF; 2019.

[3] Kammann EE, Wand MP. Geoadditive models. J R Stat Soc Ser C. 2003;52(1):1–18.

[4] Bruckner TA, Hajat A. Integrating the social and natural environments in health research: A multilevel, geoadditive perspective. Soc Sci Med. 2014;117:153–61.

[5] Oyefara JL. Socio-economic consequences of high fertility among women in Muslim households in Nigeria. Soc Indic Res. 2012;107(3):361–78.

[6] Adewuyi A, Ogunjuyigbe P. The role of men in family planning: An examination of men's knowledge and attitude to contraceptive use among the Yorubas. Afr Popul Stud. 2003;18(1):35–49.

[7] Mberu BU, Reed HE. Understanding subgroup fertility differentials in Nigeria. Popul Rev. 2014;53(2):23.

[8] Odimegwu CO, Zerai A. Understanding the cultural context of fertility in sub-Saharan Africa: The influence of female education and contraceptive use. Afr Popul Stud. 1996;11(2):31–44.

[9] Fahrmeir L, Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. J R Stat Soc Ser C. 2001;50(2):201–20.

[10] Davis J, Hyndman R, Wang Y. Spatial variations in fertility rates: A Bayesian geo-additive model for studying regional differences. J R Stat Soc Ser A. 2019;182(2):383–404.

[11] Nwakeze NM, Kandala NB. The spatial distribution of factors associated with FGM/C in Nigeria: A geo-additive Bayesian discrete-time

survival modelling. J Biosoc Sci. 2011;43(1):1–17.

[12] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B. 2005;67(2):301–20.

[13] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc. 2001;96(456):1348–60.

[14] Zhang CH. Nearly unbiased variable selection under minimax concave penalty. Ann Stat. 2010;38(2):894–942.

[15] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.

[16] Breheny P, Huang J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. Ann Appl Stat. 2011;5(1):232.

[17] Burnham KP. Information and likelihood theory: A basis for model selection and inference. In: Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. Springer; 2002. p. 49–97.

[18] Zhao P, Yu B. On model selection consistency of LASSO. J Mach Learn Res. 2006;7:2541–63.

[19] Ezeh AC, Kodzi I, Emina J. Reaching the urban poor with family planning services. Stud Fam Plann. 2010;41(2):109–16.

[20] Bongaarts J, Westoff CF. The potential role of contraception in reducing abortion. Stud Fam Plann. 2000;31(3):193–202.

[21] Bonney DK. General penalized logistic regression for gene selection in high-dimensional microarray data classification [Master's thesis]. El Paso: The University of Texas; 2020.

[22] Zhang K, Yin F, Xiong S. Comparisons of penalized least squares methods by simulations. arXiv preprint arXiv:1405.1796; 2014.

[23] Lima E, Davies P, Kaler J, Lovatt F, Green M. Variable selection for inferential models with relatively high-dimensional data: Between method heterogeneity and covariate stability as adjuncts to robust selection. Sci Rep. 2020;10(1):8002.

[24] Lu M, Zhou J, Naylor C, Kirkpatrick BD, Haque R, Petri WA, et al. Application of penalized linear regression methods to the selection of environmental enteropathy biomarkers. Biomark Res. 2017;5:1–10.

[25] Ye YF, Shao YH, Deng NY, Li CN, Hua XY. Robust Lp-norm least squares support vector regression with feature selection. Appl Math Comput. 2017;305:32–52.

[26] Algamal ZY, Lee MH. A two-stage sparse logistic regression for optimal gene selection in high-dimensional microarray data classification. Adv Data Anal Classif. 2019;13(3):753–71.

[27] Dondelinger F, Mukherjee S, Alzheimer's Disease Neuroimaging Initiative. The joint lasso: high-dimensional regression for group structured data. Biostatistics. 2020;21(2):219–35.

[28] Rahman A, Hossain Z, Rahman ML, Kabir E. Determinants of children ever born among ever-married women in Bangladesh: Evidence from the Demographic and Health Survey 2017–2018. BMJ Open. 2022;12(6):e055223.

[29] Alaba OO, Olubusoye OE, Olaomi JO. Spatial patterns and determinants of fertility levels

among women of childbearing age in Nigeria. S Afr Fam Pract. 2017;59(4):143–7.

[30] Adebowale AS. Ethnic disparities in fertility and its determinants in Nigeria. Fertil Res Pract. 2019;5:1–16.

[31] Mashood LO, Ani CI, Balogun OS, Abdulazeez SA. A geo-additive model of fertility level on female education among women of childbearing age in Nigeria. In: Proceedings of the 1st Faculty of Science International Conference FSIC; 2022. p. 175–88.