Small area estimation of age-specific fertility rates in Nigeria

Thomas Hsiao

3/12/2019

Introduction

The development of fertility rates in Sub-Saharan Africa (SSA) is of particular interest to demographers and leaders in public policy, due to the region's importance to projections of total population in the future. Fertility has been shown to follow a reverse sigmoid curve and is often classified into three phases: a stable high fertility phase, a fertility decline or transition, and a stable low fertility phase. A majority of SSA countries are either still in the high fertility phase, or in the beginning of the fertility transition. The UN's World Population Projection 2017 results suggest that continued slow declines in fertility in SSA will be the main driver of explosive population growth up to 2100, where the total population is projected to be at 11.2 billion, compared to 7.2 billion now. However, great uncertainty remains as to the future development of SSA fertility.

Of all the SSA countries, Nigeria's demographic future is key to understanding how the region's population will develop. It accounts for approximately 15% of the total population on the African continent. TFR in Nigeria has observed a slow decline only falling from 6.11 in 2000 to 5.53 as of 2016 (World Bank). In contrast, Ethiopian TFR has declined from 6.53 to 4.20 in the same span of time. Heterogeneity of common health and demographic indicators along regional, ethnic, and religious divides has been well-documented in Nigeria. The Muslim Hausa-Fulani people occupy the north, while the predominantly Christian Yoruba and Igbo people reside in the southwest and southeast respectively. Understanding the variation within Nigeria can give a better idea as to why fertility remains elevated for the country as a whole.

The goal of this analysis is to apply small area estimation techniques used to estimate 5q0 in Mercer et al. (2015) to the age-specific fertility rates reported by Nigeria DHS and MIS survey data from 1995-2013. The Demographic Health Survey (DHS) series offers a robust source of fertility data in developing countries. The UN Population Division (UNPD) and the Institute for Health Metrics and Evaluation (IHME), two research groups that estimate global demographic trends worldwide, rely heavily on DHS data to inform fertility schedules in over 90 countries dating back to the 1990's. Smoothed fertility estimates will give a better understanding of the level of heterogeneity at the admin 2 level (37 total subnational units). Special attention is paid to what proportion of the variance is explained by spatial and temporal related effects, and looking at how these proportions change between age groups.

Data

Fertility is measured through the age-specific fertility rate (ASFR), which is defined as the number of births occurring during a reference period per 1000 women years of reproductive age for a given age group. The more commonly used aggregate measure of fertility, the total fertility rate (TFR), is defined as the average number of children born per woman if she were part of the hypothetical cohort that would experience the current schedule of ASFR's for a reference period. All the information required to compute ASFR can be found in the demographic health surveys (DHS) and malaria indicator surveys (MIS).

Survey data

The Nigerian 2003 DHS, 2008 DHS, 2010 MIS, and 2013 DHS were used in this study. Sample sizes vary greatly between the data, with over 30,00 women interviewed in the 2008 and 2013 DHS, and under 10,000 women in the 2003 DHS and 2010 MIS. The smaller sample sizes lead to instability and large standard errors at the age-specific admin 2 level. ASFR can be computed from DHS data in the form of complete birth histories (CBH). CBH data is a format of fertility data where the DOB of every child each woman respondent has reported given birth to is known. This is in contrast to the less informative summary birth history data (SBH) more common in censuses and older surveys, where only the children ever born (CEB) to a woman is known, with no information on timing of that birth.

Data extracted and geographic coverage

Since 2000, Nigeria has been composed of six geopolitical zones at the admin 1 level. Within those zones are 37 admin 2 level subnational units, which include 36 states and one federal capital territory - FCT Abuja. This analysis focuses on the admin 2 level. Extracted data included the DOB of each woman respondent, the DOB of each reported birth, each woman's individual sampling weight, the date of the interview, the GPS coordinates of each woman's interview, whether the sample was an ever-married sample and the corresponding all-woman factor. Since the survey was not an ever-married sample, no all-woman factor was required to adjust the data to apply toward the general population of all Nigerian women.

Computation of ASFR

ASFR was computed as described by the DHS Stats Guide. ASFR was calculated as a ratio of births to woman-years computed from the microdata and collapsed to the admin 2 level. Nigerian DHS surveys are conducted in a two-stage cluster sample, and as such report sampling weights for each respondent. The weight attached to each respondent reflects the sampling probability as well as a non-response adjustment. The sample-weighted estimates of ASFR correct for bias, and were then used as data input into the model. ASFR was calculated over 3 year periods within 15 years of recall since the date of interview.

Methods

A Bayesian spatial smoothing model was applied to each of the weighted age-specific fertility rates in the five-year age groups within 15-44. 45-49 was not included due to lack of data in many estimation periods, and fertility in the group tending to be relatively negligible. Each age group is fit in a separate model of the same form. A Poisson likelihood is assumed for the fertility rate, with births as the count and woman-years as the offset. Spatiotemporal smoothing in the form of ICAR and second-order random walk (RW) processes, as done in Mercer 2014, was implemented with the following model specification:

$$B|\beta_0, S_i, \epsilon_i, \gamma_t, \alpha_t, \delta_{it} \sim Pois(Ee^{\beta_0+S_i+\epsilon_i+\gamma_t+\alpha_t+\delta_{it}}) \tag{1}$$
$$S_i|\sigma_s^2 \sim ICAR(\sigma_s^2) \tag{2}$$
$$\epsilon_i|\sigma_\epsilon^2 \sim N(0,\sigma_\epsilon^2) \tag{3}$$
$$\gamma_t|\sigma_\gamma^2 \sim RW2(\sigma_\gamma^2) \tag{4}$$
$$\alpha_t|\sigma_\alpha^2 \sim N(0,\sigma_\alpha^2) \tag{5}$$
$$\delta_{it}|\sigma_\delta^2 \sim N(0,\sigma_\delta^2) \tag{6}$$
$$\beta_0 \sim N(0,\infty) \tag{7}$$
$$\sigma^2 \sim LogGamma(0.5, 0.001) \tag{8}$$

For the intercept and variance terms, the default INLA priors were used. B represents the number of births, E is women-years, i refers to a specific admin 2 state, S_i denotes the structured ICAR spatial random effect, ϵ_i represents an unstructured spatial effect, γ_t is the structured RW2 temporal random effect, α_t is the unstructured temporal random effect, $\delta_{i,t}$ is a random effect on spatiotemporal interaction, and β_0 is the overall mean or intercept. The space-time interaction term is included since time trends in fertility could easily vary between states. TFR was then computed from the smoothed estimates of ASFR, by taking the sum of the ASFR's within each period-state, and multiplying by five (assume the ASFR within each age in a five-year age group is the same).

The proportion of the total variance explained by each of the random effects was also computed. Complications arise from the variance parameter of the ICAR and RW random effects, which are conditional variances and cannot be directly compared to the marginal variances of the standard iid effects. The marginal variance was estimated empirically as detailed by the slides on Small Area Estimation. The posterior densities of the marginal distribution for the ICAR and RW1 effects were extracted from INLA, and then taking the sampling variance of the Monte Carlo simulations generated by inla.rmarginal.

Results and Discussion

Model results are presented in terms of subnational location-specific time series plots, as well as maps. Figures 1 and 2 display the model estimates against the survey data for states Plateau and Lagos. We see that despite noisy data of great uncertainty from the 2010 MIS and 2003 DHS, the model's smoothing capabilities result in much smaller standard errors, and tend to follow the more stable 2008 and 2013 DHS data. In this case, Plateau observes a consistent and slow decline in fertility across all age groups across the estimation period. Fertility in Lagos overall is much lower than Plateau, and has remained at around the same level since 1995.

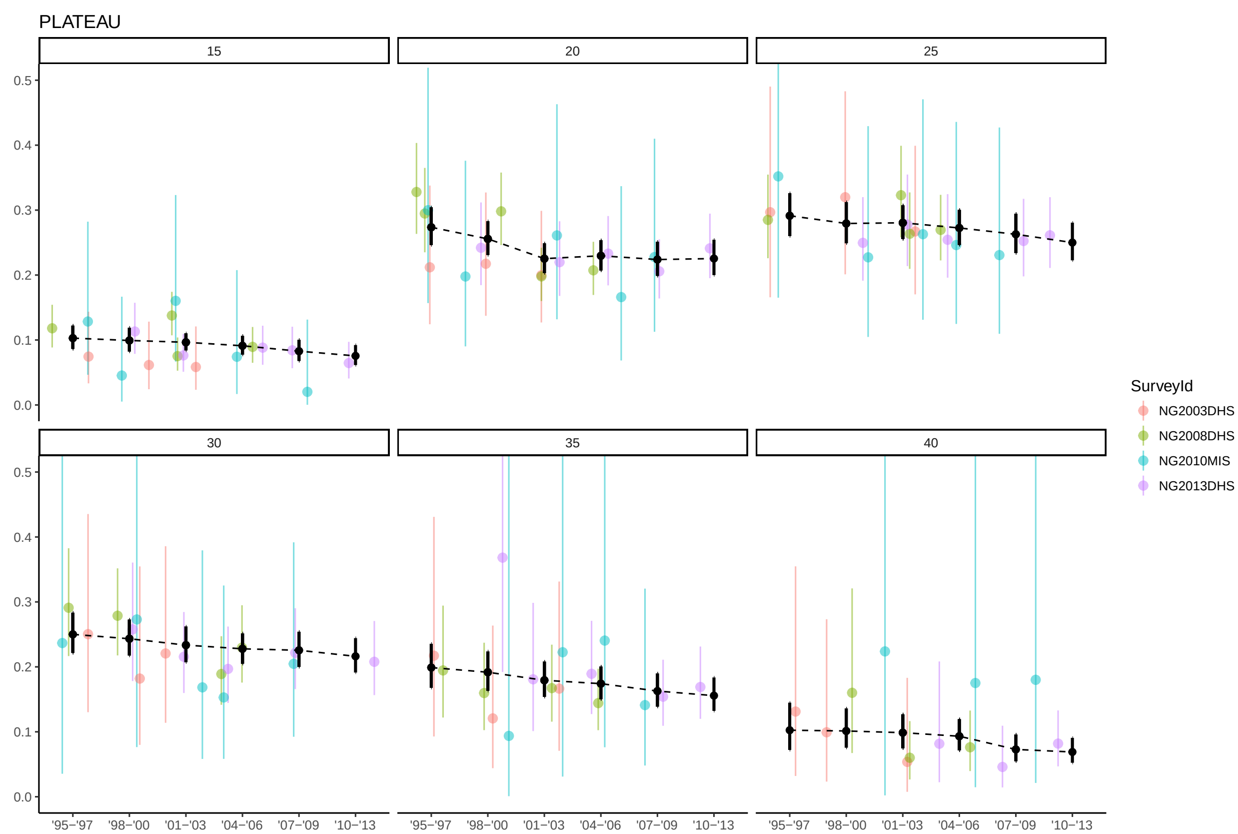

Figure 1: Plateau time series

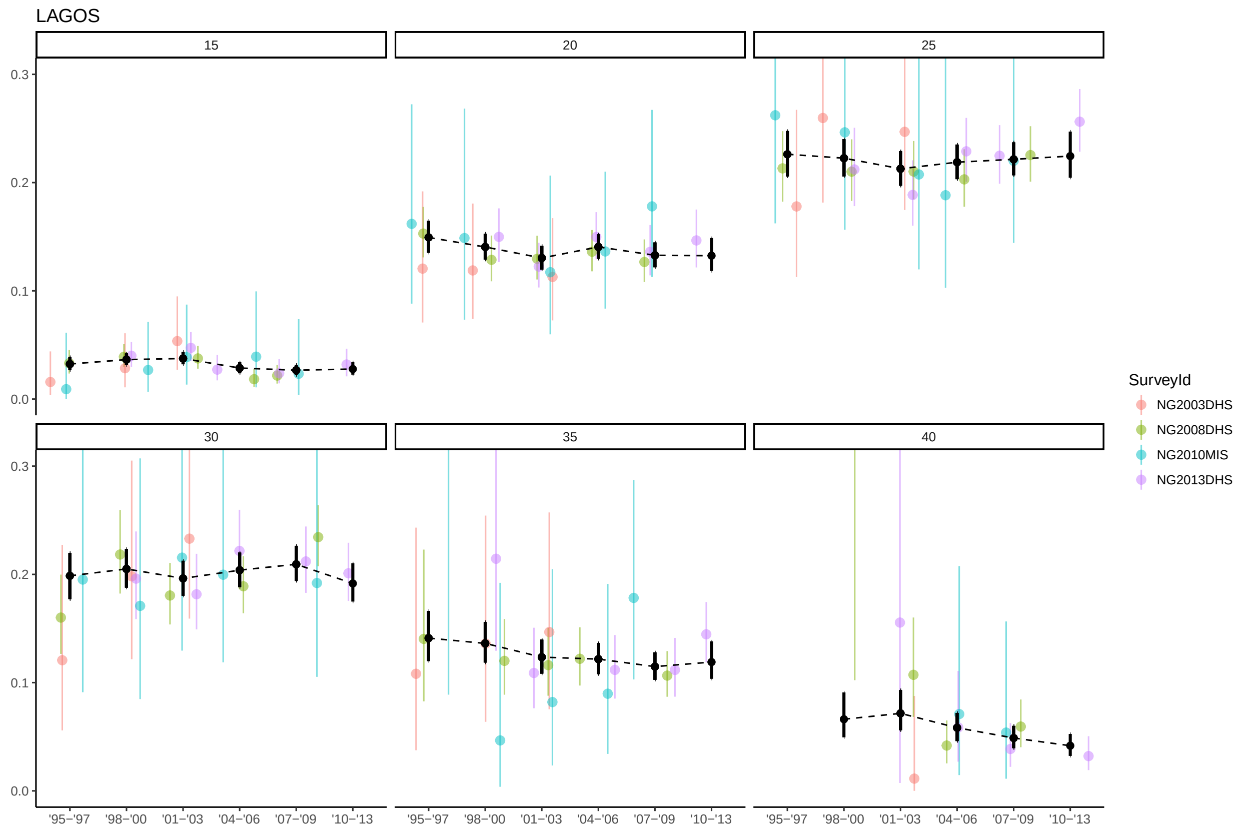

Figure 2: Lagos time series

The proportion of the total variance in the data that is explained by each of the random effects is reported in Table 1. The spatial ICAR effect is responsible for an overwhelming amount of the variance in each of the age-specific models, particulary in the 15-19 age group at over 90%. In all age specific models, the ICAR effect explains at least 50% of the total variance, which is also supported by the large difference in fertility levels between the northern and southern states. An interesting trend is that the proportion that is spatial structural effects decreases with age, but the proportion that is temporal structural effects increase. Temporal effects appear to be minimal, which support the previous observation that Nigerian fertility has been quite stagnant, even at the subnational level. There exists an extremely large spread in the adolescent and 20-24 age group fertility levels between states compared to the other older age groups, which have also experienced greater declines over time.

Table 1: Proportion of each random effect of the total variance, by age group. The spatial structural variance decreases with age, while the temporal structural variance increases with age.

Age Group	Spatial Structure	Spatial Unstructure	Time Structure	Time Unstructure	Space-Time
15-19	0.9248395	0.0197048	0.0262272	0.0058513	0.0233773
20-24	0.8768080	0.0310648	0.0375339	0.0173687	0.0372247
25-29	0.6144263	0.0671541	0.1405265	0.0654278	0.1124653
30-34	0.5989215	0.0978328	0.1065975	0.0768433	0.1198048
35-39	0.6517733	0.0484117	0.1823407	0.0348574	0.0826168
40-44	0.6532411	0.0341222	0.2236686	0.0376422	0.0513258

Table 2 shows the posterior median estimates of each variance parameter. The 15-19 spatial structural effect variance is incredibly large, with fairly tight uncertainty, explaining the large proportion observed in table 1.

Table 2: Median and 95% credible interval for each marginal variance parameter.

Age Group	Spatial Structure	Spatial Unstructure	Time Structure	Time Unstructure	Space-Time
15-19	0.368 (0.336-0.404)	0.003 (0-0.048)	0.01 (0.004-0.018)	0.002 (0-0.009)	0.009 (0.005-0.014)
20-24	0.083 (0.073-0.094)	0.002 (0-0.012)	0.003 (0.001-0.007)	0.001 (0-0.005)	0.003 (0.002-0.006)
25-29	0.016 (0.013-0.02)	0.001 (0-0.005)	0.004 (0.001-0.008)	0.001 (0-0.006)	0.003 (0.002-0.005)
30-34	0.013 (0.01-0.017)	0.002 (0-0.006)	0.002 (0.001-0.006)	0.001 (0-0.006)	0.003 (0.001-0.005)
35-39	0.036 (0.029-0.044)	0.002 (0-0.009)	0.01 (0.005-0.016)	0.001 (0-0.006)	0.004 (0.002-0.009)
40-44	0.102 (0.08-0.126)	0.004 (0.001-0.026)	0.035 (0.018-0.063)	0.003 (0-0.04)	0.007 (0.001-0.02)

Figure 3 shows the age-specific fertility rate of 15-19 year olds. ASFR ranges from 28 births per 1000 women-years in Lagos to 231 births per 1000 women-years in Bauchi in the 2010-2013 estimation period. Large spatial heterogeneity exists between the northern and southern states, separated greatly by the southern border of Niger.

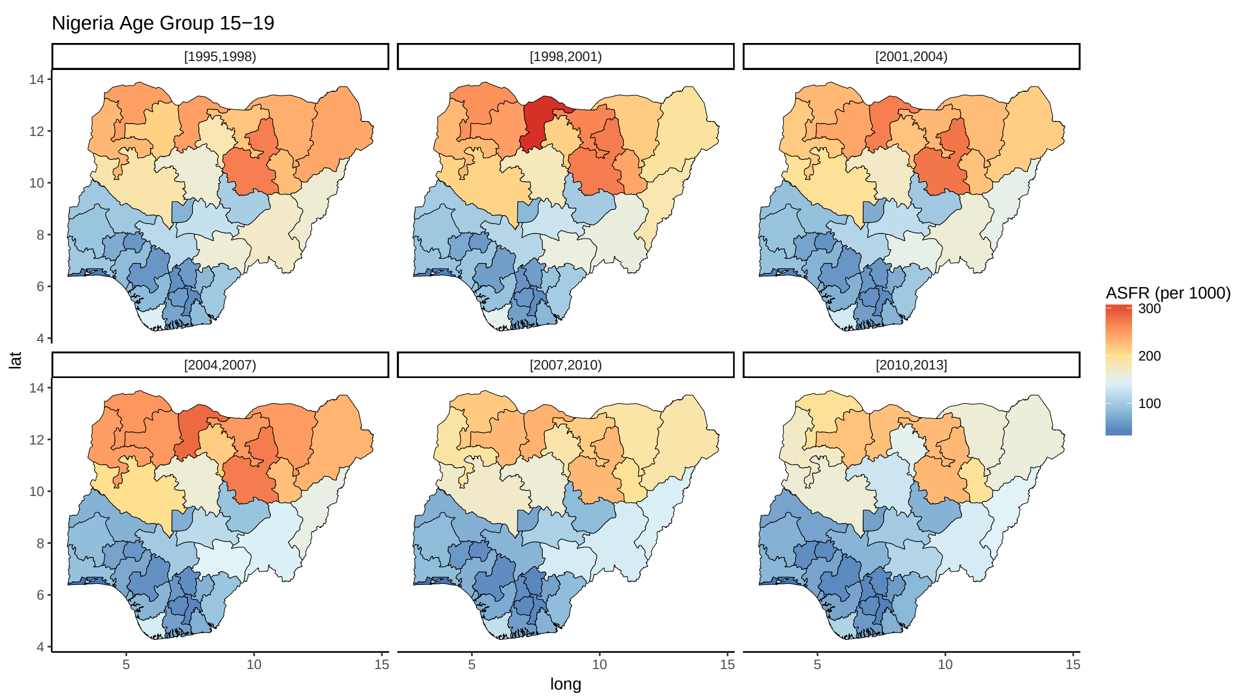

Figure 3: Smoothed estimates of 15-19 ASFR for Nigeria across all estimation periods.

Figure 4 shows the TFR in Nigeria, aggregated from model results. The TFR gives an even clearer picture of how different fertility levels are within Nigeria, and shows that the differences are not just limited to the 15-19 age group. In the 2010-2013 period, the TFR ranges from 3.75 in Lagos to 7.92 in Zamfara, with a standard deviation of 1.21. The weighted national estimate from the 2013 DHS survey for TFR is 5.5. For comparison, the World Bank reported that in 2016 Egypt had a TFR of 3.26 while Niger had the highest national TFR at 7.24. Based on these results, it is clear that the national TFR reported by the 2013 DHS of 5.5 masks huge structural variation between admin 2 states in Nigeria. While fertility rates in all states, notably the northern states have decreased, they have done so very slowly, with reductions of TFR on average of 0.85 children from 1995 to 2013. Northern states have witnessed the largest decreases, such as Borno reducing from 7.51 to 5.73, while southern states like Lagos have remained at their previous fertility level of 3.74.

While the slow decline from the high fertility phase in the northern states is similar to other SSA countries like Mozambique, the stableness of TFR in the southern Nigerian states does not seem to have much precedent. Once the fertility transition begins, most countries TFR tends to observe a decrease up to around 2 children per woman at the national level. However, the more developed southern states of Nigeria have remained at a TFR of around 4 for at least a decade, with no significant decreases, which is quite contrary to the national Western experience. Such large variation between states and obseved trends without much historical precedent contribute to the great uncertainty surrounding the demographic projections of Nigeria and other Sub-Saharan African countries.

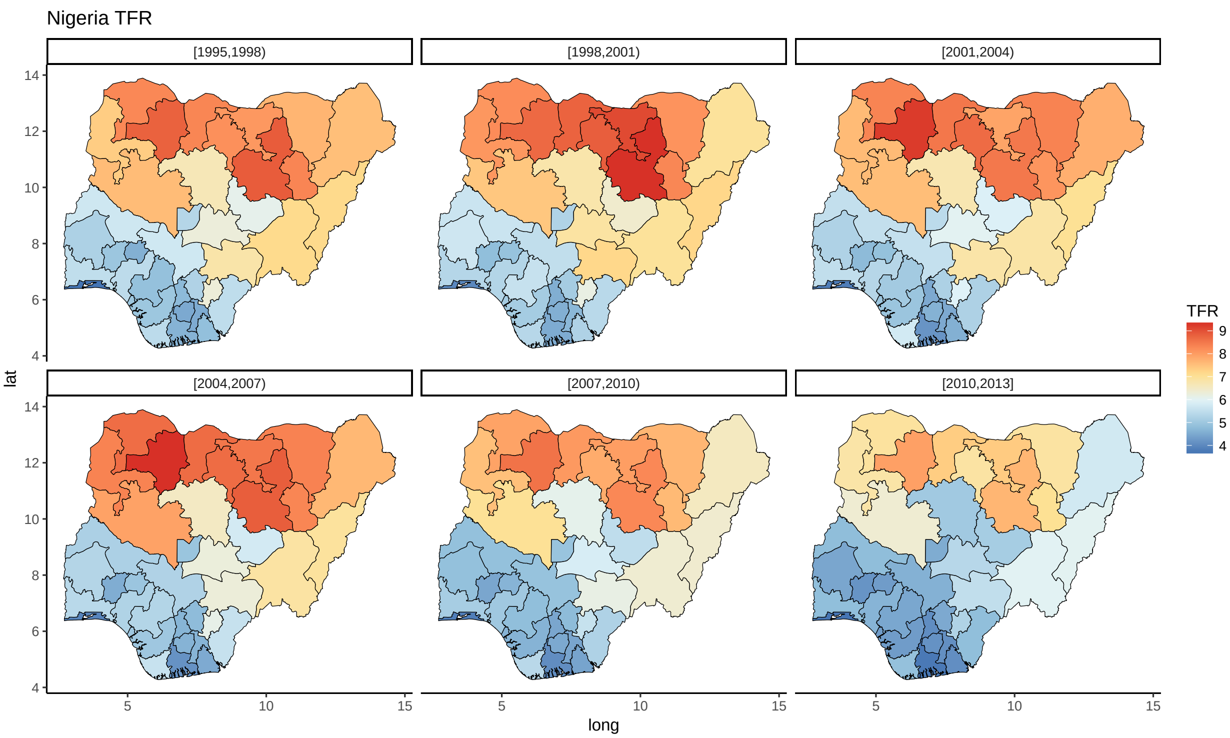

Figure 4: Smoothed estimates of TFR for Nigeria across all estimation periods.

Limitations

This analysis was focused on small area estimation through Bayesian spatiotemporal smoothing. DHS data contains many indicators including ethnicity, religion, education, female literacy, and female met need for contraception that can probably explain much of the observed spatial and temporal variation in the model. Model validation and coverage assessment through holdouts by subnational unit would be desirable to evaluate performance. A sensitivity analysis on the choice of prior for the variance parameters would also be ideal to get a sense of how volatile the model is to assumptions on prior distributions. A re-evaluation of the treatment of sampling weights may also be worthwhile.

Conclusion

The spatial-temporal Bayesian smoothing model gives a more reasonable set of fertility estimates compared to the weighted estimators from the DHS data, and the standard errors of the estimates shrunk accordingly. What is striking is that even after heavy spatiotemporal smoothing, the heterogeneity of fertility remains between the subnational units even in the most recent period, particular at the border between separating the north and south. The spatial effects are more pronounced at younger ages, while temporal effects gain prominence in older ages. However, overall, spatial effects dominate, with the structured ICAR spatial effect explaining over 60% of the total variance in each age group model. Maps of the estimates show a clear divide in fertility between north and south, particularly in the 15-19 and 20-24 age groups, implying that adolescent birth rates may be responsible for the slow decline of national fertility rates.

Code

Analytic code is attached in the code appendix. Totality of code, including data pre-processing can be found at https://github.com/tXiao95/mapping-TFR-Nigeria.

References

Mercer, Laina D, Jon Wakefield, Athena Pantazis, Angelina M Lutambi, Honorati Masanja, and Samuel Clark. 2015. "Space-Time Smoothing of Complex Survey Data: Small Area Estimation for Child Mortality." *The Annals of Applied Statistics*. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4959836/.