A Bayesian Model to Estimate Male and Female Fertility Patterns at a Subnational Level

Riccardo Omenti [1] Monica Alexander [2,3] Nicola Barban [1]

[1]Department of Statistical Sciences, University of Bologna [2]Department of Statistical Sciences, University of Toronto [3]Department of Sociology, University of Toronto

Motivation

- Need for new methods to obtain indicators of **male fertility** in **data sparse contexts** and **small areas**
- High share of missing values in paternal childbearing ages
- Explore **heterogeneity** in **fertility behaviors** between **men** and **women** across space and over time within the US

Main contribution

- Extend Bayesian modelling framework by Schmertmann and Hauer (2019) to estimate **male** and **female fertility** at a **subnational** level without the **distribution** of **births** by **parental ages**
  - Include **men aged 15-59**
  - Include **subnational mortality estimates**
  - Allow for **spatial** and **temporal dependencies**
- Generate time series of **male** and **female total fertility rate** (TFR) estimates for **US counties** during the period **1982-2019**

Model summary

Data model

$C_{a,t}|K_{x,a,t} \sim \mathcal{P}\left(\sum_{x=15}^{55} K_{x,a,t} \cdot M_{x,a,t}\right)$

$C_{a,t}$: number of children under 5 in year t and area a
$M_{a,t}$: number of men in age class x, year t and area a

Data Source: US census

Expected child-man ratio

$K_{x,a,t} = TFR_{a,t} \cdot \frac{L_{0,a,t}}{5} \cdot \frac{1}{2} \cdot \left[\frac{L^M_{x-5,a,t}}{L^M_{x,a,t}} \cdot \phi_{x-5,a,t} + \phi_{x,a,t}\right]$

$TFR_{a,t}$: total fertility rate in year t and area a
$\phi_{x,a,t}$: total fertility proportion in age class x, year t, area a
$L^M_{x,a,t}$: male person-years lived in age class x, year t, area a
$L_{0,a,t}$: male person-years lived in age class $0-4$, year t, area a
$x = 15, \ldots, 55$

Prior distributions for TFR

$TFR_{a,t} \sim \mathcal{N}(TFR^{nat}_t, \sigma^2_{TFR})$

$\sigma_{TFR} \sim \mathcal{N}^+(0,1)$

TFR^{nat}_t: country-level TFR in year t

Data Source: Human Fertility Collection

Statistical model for $\phi_{x,a,t}$

$\phi_{x,a,t} = \frac{\exp(\gamma_{x,a,t})}{\sum_{x=15}^{55}\exp(\gamma_{x,a,t})}$

$\gamma_{x,a,t} = m + \beta_{1,a,t}Y_{1,x} + \beta_{1,a,t}Y_{2,x}$

m: mean of national age-specific fertility patterns

$Y_{1,x}$ and $Y_{2,x}$: first and second principal components from a set of standard age-specific fertility patterns

Prior distributions for $\beta_{p,a,t}$

Information Pooling

$\beta_{p,a,t} \sim \mathcal{N}(\mu_{p,t}, \sigma^2_{\beta_p})$

$\mu_{p,t}$: state-level mean

Smoothing over time

$\mu_{p,t} \sim \mathcal{N}(2\mu_{p,t-1} - \mu_{p,t-2}, \sigma^2_{\mu_p})$
$\mu_{p,1}, \mu_{p,2} \sim \mathcal{N}(0, \sigma^2_{\mu_p})$
$\sigma_{\beta_p}, \sigma_{\mu_p} \sim \mathcal{N}^+(0,1)$
$p = 1, 2$

Prior distributions for $L^M_{x,a,t}$

$L^M_{x,a,t} \sim \mathcal{N}(\hat{L}^M_{x,a,t}, \hat{\sigma}^2_{L_{x,a,t}})$
$L_{0,a,t} \sim \mathcal{N}(\hat{L}_{0,a,t}, \hat{\sigma}^2_{L_{0,a,t}})$

$\hat{L}^M_{x,a,t}$: estimated person-years in age class x, area a, year t
$\hat{L}_{0,a,t}$: estimated male person-years in age class $0-4$, area a, year t
$\hat{\sigma}_{L_{x,a,t}}$: corresponding standard error calculated via bootstrap

Data Source: United States Mortality DataBase

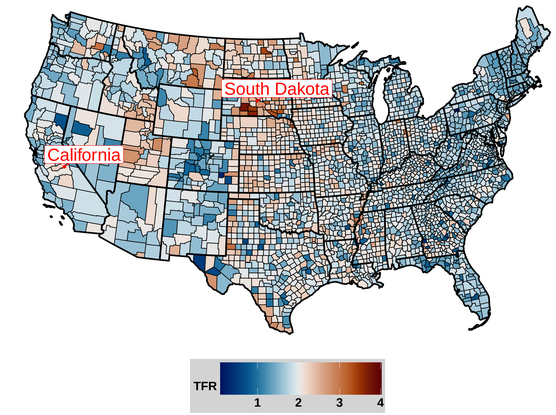
TFR interpretation

Average number of children per woman aged $15-49$ or per man aged $15-59$ throughout their lifetime based on current age-specific fertility rates

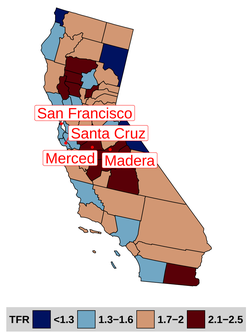
Estimation of TFR

- Draw Posterior samples using **Markov Chain Monte Carlo** algorithm in *stan*
- Focus on the samples from the **marginal posterior distribution** of the TFR parameters conditional on the observed data and the other parameters and employ their **median** as the final best TFR estimates.

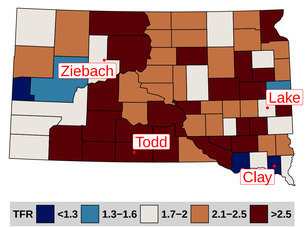
Subnational male TFR estimates in 2019



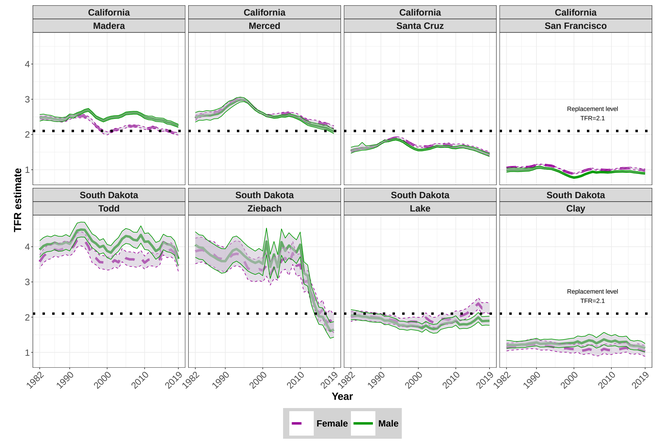
California



South Dakota



TFR estimates in selected counties during 1982-2019



Preliminary Conclusions

- Using **county-level population counts** by age and sex allows to derive **subnational** period TFR estimates without the need of information on parental ages
- No striking differences between **male** and **female fertility**
- Country-specific characteristics determine a high **spatial heterogeneity** and **distinct temporal trajectories**

Future developments

- Adjustment for **migration** in future model developments
- Accounting for **counties** with **unbalanced sex ratios** in **childbearing ages** such as those with a military base or a prison

References

[1] Carl P Schmertmann and Mathew E Hauer. Bayesian estimation of total fertility from a population's age–sex structure. *Statistical Modelling*, 19(3):225–247, 2019.