A Bayesian approach on reconstructing multistate populations and education specific fertility rates

Dilek Yıldız[1,2] Guy Abel[2,3] Anne Goujon[1,2]

[1]Wittgenstein Centre (IIASA, VID/ÖAW, WU),
Vienna Institute of Demography/Austrian Academy of Sciences

[2]Wittgenstein Centre (IIASA, VID/ÖAW, WU),
International Institute for Applied Systems Analysis

[3]Asian Demographic Research Institute, Shanghai University

8 April 2020

Outline

Introduction

- Consistent time series for population by educational attainment is required to comprehensively assess the returns to investments in formal education at the national level as well as to know the impacts of educational status of the adult population on systemic changes.
- Previous research does not reconstruct fertility rates by education or provide uncertainty estimates around the results.

Introduction

- Two approaches to rebuild past populations: back projection and reconstruction.
- Back projections by educational attainment (Wrigley and Schofield 1982, Lee 1978, Barro and Lee 1993, Lutz et al. 2007, Goujon et al. 2016).
- Bayesian modelling for simultaneously estimating past population by age, fertility, mortality and net migration (Wheldon et al. 2013).

We propose a hierarchical Bayesian approach to reconstruct multi-state populations that simultaneously estimates population by age, gender and educational attainment, with education specific fertility proportions along with uncertainty around the estimates.

Data sources - Brazil (1980-2010)

- Population counts by age, gender and educational attainment:
  - IPUMS-International database for education distribution
  - The Brazilian Institute of Geography and Statistics (IBGE) reconstructed census counts
  - Precision: Size of 95% CI equals to 3%, 5% and 10% of the population
- TFR and ASFR:
  - IBGE reconstructed fertility rates
  - Precision: Highest standard deviation from DHS surveys and applied to all

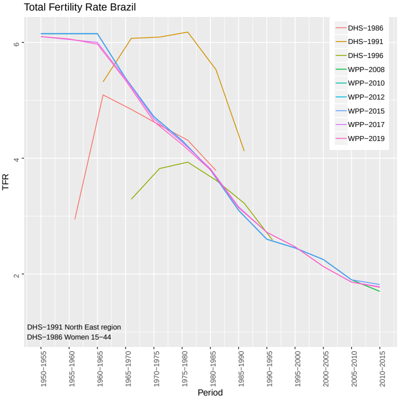
Data sources - Brazil (1980-2010)

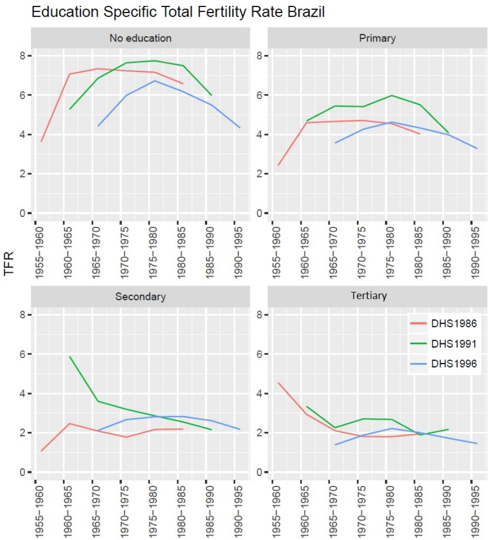
- TFR Education patterns (wESTFR):
  - Cohort Fertility Education Database (CFE)
  - CFE patterns calibrated using female population by education
  - Sum of ESTFR matched reconstructed TFR from IBGE
- ASFR and ASFR education patterns (wESASFR):
  - Currently ASFR and ESASFR priors are equal
  - Future work: select ASFR from different countries to match different education levels
- Survival proportions:
  - Life tables from WPP 2019
  - WIC education patterns for 2015-2019
  - One of the main areas to be investigated in the new project

Brazil population by educational attainment

Year	Sex	No education	Primary	Secondary	Tertiary
1980	Male	55.1%	33.0%	10.1%	1.7%
1990	Male	48.2%	33.7%	15.4%	2.8%
2000	Male	39.1%	34.7%	23.0%	3.3%
2010	Male	32.5%	28.2%	33.4%	6.0%
1980	Female	56.0%	32.0%	10.8%	1.2%
1990	Female	47.5%	33.0%	16.9%	2.6%
2000	Female	37.5%	33.1%	25.7%	3.6%
2010	Female	30.6%	25.8%	35.6%	6.0%

Available fertility data





Notation

$$\mathbf{n_{a,s,t+5,e}} = (n_{0,s,t+5,e},\ n_{5,s,t+5,e},\ \ldots,\ n_{A,S,t+5,E})'$$

a:	0-4, 5-9, ..., 80+, s: f,m
t:	1960, 1965, ..., 2010
e:	$e1$(No education), $e2$(Primary), $e3$(Secondary), $e4$(Tertiary)
f:	Age-specific fertility rate
TFR:	Total fertility rate
$ESTFR$:	Education-specific total fertility rate
$ESASFR$:	Education- and age-specific fertility rate
s:	Survival proportions
g:	Net migration proportions
$h_{a,\Theta}$:	Education transition proportions at age a between education levels Θ
Θ:	Transitions from one education level to another (1-2, 2-3, ...)

Level 1: Modelling census counts

$$\log\ n^*_{a,s,t,e} \sim \text{Normal}\left(\log\ n_{a,s,t,e}, \tau^n_{a,s,t,e}\right) t = 1970, 1980, 1990, 2000, 2010$$

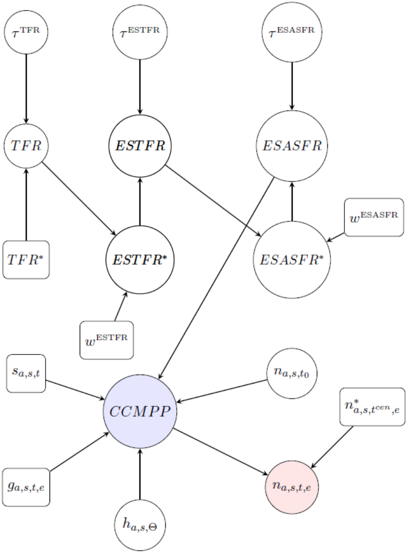
Level 2: Cohort component population projection method
Level 3: Modelling initial estimates

$$\log\ f_{a,t} \sim \text{Normal}\left(\log\ f^*_{a,t}, \tau^f_{a,t}\right)$$
$$\log\ \text{TFR}_t \sim \text{Normal}\left(\log\ \text{TFR}^*_t, {\tau_t}^{TFR}\right)$$
$$\log\ \text{ESTFR}_{t,e} \sim \text{Normal}\left(\log\ \text{ESTFR}^*_{t,e}, \tau^{ESTFR}_{t,e}\right)$$
$$\log\ \text{ESASFR}_{a,t,e} \sim \text{Normal}\left(\log\ \text{ESASFR}^*_{a,t,e}, \tau^{ESASFR}_{a,t,e}\right)$$

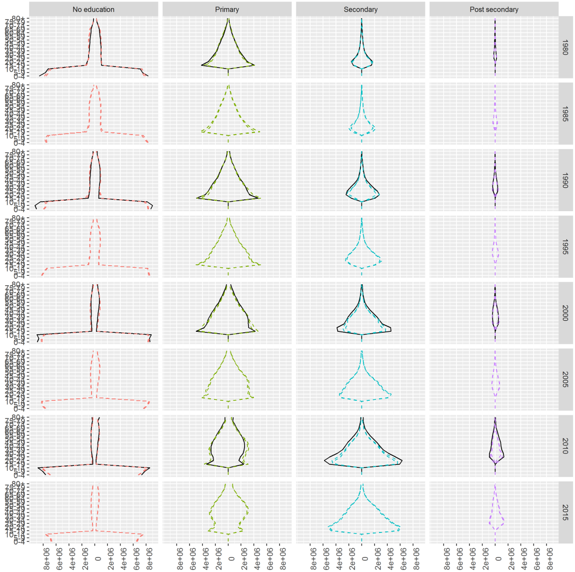
Level 3 continued:

$$
\begin{aligned}
\log\ n_{a,s,t_0,e} &\sim \text{Normal}\left(\log\ n^*_{a,s,t_0,e}, \tau^n_{a,t_0,e}\right) \\
h^*_{a,\Theta} &\sim \text{Normal}\left(h^*_{a,\Theta}, \tau^h_{a,\Theta}\right) \\
\theta &= (1-1, 1-2, 2-2, 1-3, 2-3, 3-4)
\end{aligned}
$$

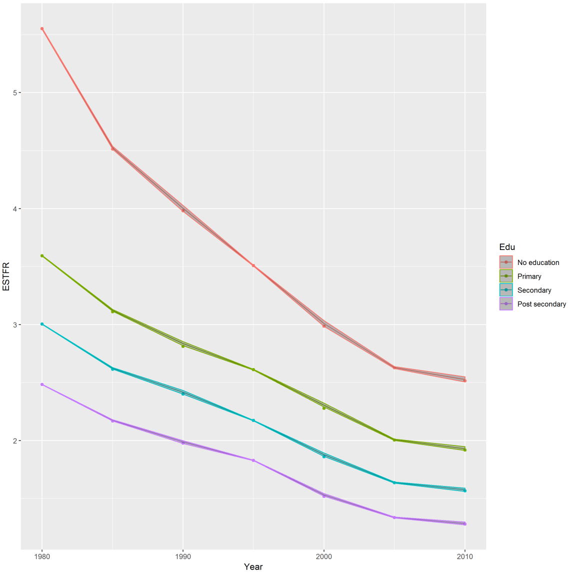
Multistate model



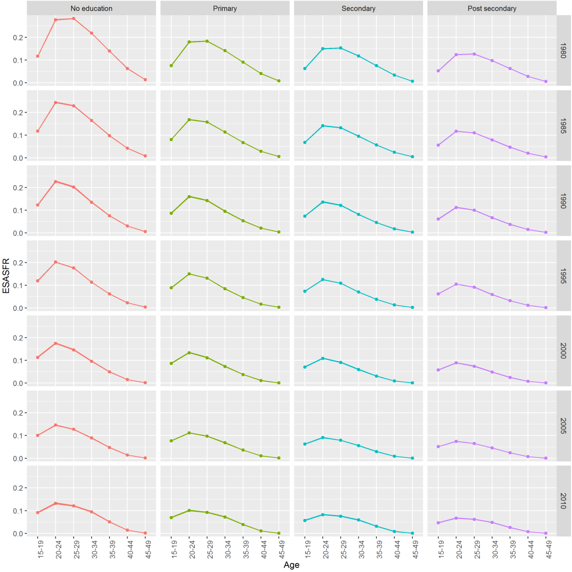
Multistate model - Brazil - population count estimates



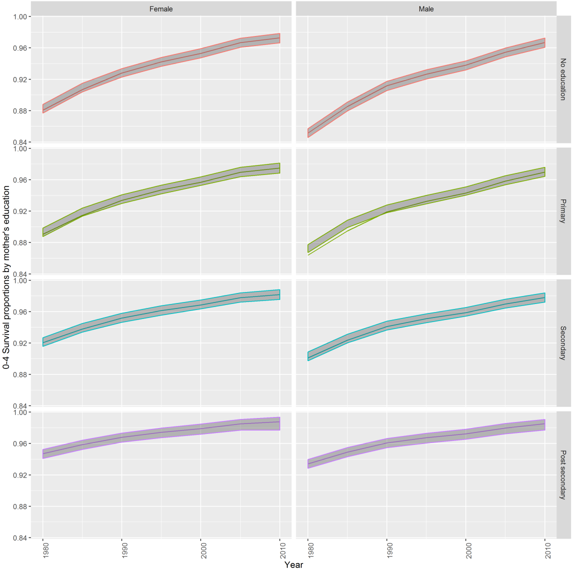
ESTFR


5
4
3
2
ESTFR
1980
1990
2000
2010
Year
Edu
No education
Primary
Secondary
Post secondary

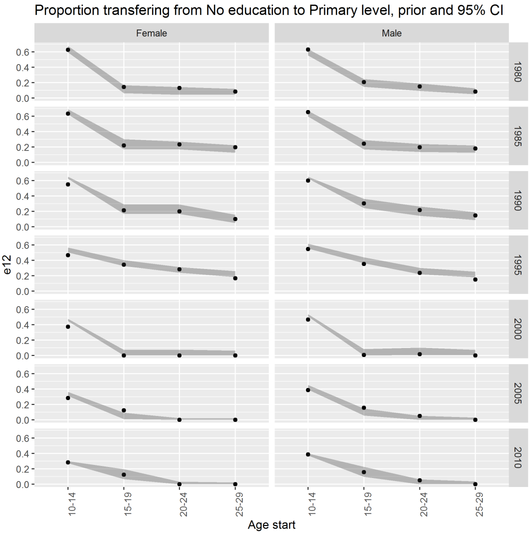
ESASFR


No education
Primary
Secondary
Post secondary
1980
1985
1990
1995
2000
2005
2010
ESASFR
0.0
0.1
0.2
Age
15-19
20-24
25-29
30-34
35-39
40-44
45-49

0-4 Survival proportions



Education transition - No education to Primary



Computational details

- Software: **R** and **JAGS** & **R** packages: **rjags**, **dclone** and **snow**
- **JAGS** is a program for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) simulation.
- **JAGS** allows us to parallelize, i.e., to use the full power of our computer.
- **rjags** enables to call **JAGS** from **R**.
- **dclone** allows us to do parallel updating of **JAGS** models.
- **snow** allows us to set up a cluster of subsidiary **R** programs, each one running on a different CPU of our computer.

Conclusion and future work

- Better precision values
- Sensitivity analysis
- Better prior transition proportions
- Net migration
- Expert opinion
- ASFR patterns
- Comparison with back projection estimates

Contact Information

Dilek Yıldız
Wittgenstein Centre (IIASA, VID/ÖAW, WU),
Vienna Institute of Demography/Austrian Academy of Sciences,
International Institute for Applied Systems Analysis

E-mail: dilek.yildiz@oeaw.ac.at

Wittgenstein Centre FOR DEMOGRAPHY AND GLOBAL HUMAN CAPITAL

Vienna Institute of Demography

ÖAW AUSTRIAN ACADEMY OF SCIENCES

IIASA International Institute for Applied Systems Analysis www.iiasa.ac.at