

Modeling age patterns of childlessness: a semi-parametric approach

Benjamin-Samuel Schluter^{1,*}, Monica Alexander²

¹Department of Statistical Sciences, University of Toronto, Canada

²Departments of Statistical Sciences and Sociology, University of Toronto, Canada

Abstract

Trends and patterns in childlessness are important to measure and understand its relationship with fertility rates and fertility intentions. However, data on childlessness is often limited to survey data, which suffers from small sample sizes for key subpopulations. We propose a model estimating age-specific proportions of individuals being childless at a subnational level. The model consists of an ‘expected’ component – capturing the shape of childlessness over age using a logistic function – and deviations from the expected level, modeled with P-splines. The model estimates summary parameters which are useful to understand trends over time. The model is estimated in a Bayesian framework allowing temporal smoothing and pooling of the parameters. We apply the model to estimate the proportions of childless women by race/ethnicity in the U.S. The preliminary results are promising and the model could be expanded to model multiple parities and perform projection of childlessness.

1 Introduction

Childlessness is a demographic phenomenon that has gained increased attention, particularly in high-income countries characterized by low fertility rates. The share of childless American women aged 40 to 44 was 10% in 1976, reached a peak at 20% in 2005, and fall hereafter to 15% in 2014 (Livingston 2015). In line with previous results, Dye (2010) showed that among American women aged 45 in 2006, one fifth were childless. While prior studies focused on female childlessness, recent years also brought research on male childlessness (Gray, Evans, and Reimondos 2013; Nisén et al. 2014).

Trends and patterns in childlessness are important to understand, both in terms of the relationship with overall fertility, and whether trends match with trends in fertility intentions. This has motivated research trying to identify its causes, determinants and consequences (Dykstra and Wagner 2007; Gemmill 2019; Kendig et al. 2007; Kirmeyer and Hamilton 2011; Mynarska et al. 2015).

Despite the increasing attention received by childlessness, no modeling framework has been proposed to estimate the age-specific share of individuals not having a child. A statistical model is particularly relevant when obtaining age-specific childlessness proportions in small subpopulations. Indeed, small population sizes lead to high stochasticity in the raw data, rendering the underlying trend unclear. For example, it has been shown that the proportion of women childless varies across racial/ethnic groups (Livingston 2015; Lundquist, Budig, and Curtis 2009) but not much is known about the patterns of subpopulations defined by different combination of individual characteristics such as the

*benjamin.schluter@utoronto.ca.

state of residence, race/ethnicity, and education level.

In this paper we propose a model framework to estimate age-specific proportions of individuals being childless at a subnational and subpopulation level. The model is semi-parametric in the sense that we allow some deviations from an expected level parametrized by a logistic function. The logistic function estimates the entire age schedule proportions of childless individuals with three parameters that are smoothed over time. This semi-parametric framework is robust to zero counts while being flexible. We show preliminary results as applied to race-specific childlessness proportion in the U.S. Future work will expand the model to allow estimating population subgroups defined by a combination of factors (i.e. state of residence and race/ethnicity).

2 Modeling Framework

Define $y_{x,g,t}$ to be the number of childless persons of age x belonging to subpopulation group g at time t . Define $n_{x,g,t}$ to be the subpopulation group g aged x at time t . We assume the counts of parent to be Binomially distributed as follows,

$$y_{x,g,t} | \theta_{x,g,t} \sim \text{Binomial}(n_{x,g,t}, \theta_{x,g,t}).$$

Our aim is to estimate the proportion being childless $\theta_{x,g,t}$. We model these proportions on the logit scale as follows,

$$\text{logit}(\theta_{x,g,t}) = \text{logit}(1 - f(\zeta_{g,t}, \gamma_{g,t}, \delta_{g,t})) + \sum_k B_{x,k} \alpha_{k,g,t}$$

where

$$f(\zeta_{g,t}, \gamma_{g,t}, \delta_{g,t}) = \frac{\zeta_{g,t}}{1 + e^{-\gamma_{g,t}(x - \delta_{g,t})}}$$

is the logistic function with δ the function's midpoint, ζ the supremum of the values of the function, and γ the steepness of the curve. Deviations from the parametric function are modeled with P-splines (Eilers and Marx 1996). More precisely, \mathbf{B} is the cubic basis matrix of dimension $(A \times K)$ where $A = 44$ the upper age bound, and $K = 6$ the number of knots; $\boldsymbol{\alpha}_{g,t}$ is a cubic-splines parameter vector of dimension $(K \times 1)$ estimated in the model. We impose smoothness on the deviations by using a Random Walk 1 prior on $\boldsymbol{\alpha}_{g,t}$,

$$\alpha_{k,g,t} \sim \text{Normal}(\alpha_{k-1,g,t}, \sigma_{\alpha,t}^2)$$

where the variance parameter $\sigma_{\alpha,t}^2$ is shared across all subpopulation groups in each year.

Following exploratory data analyses, we smooth the logistic parameters by assuming that

$$\zeta_{g,t} \sim \text{Normal}(\zeta_{g,t-1}, \sigma_{\zeta}^2)$$

$$\gamma_{g,t} \sim \text{Normal}(\gamma_{g,t-1}, \sigma_{\gamma}^2)$$

and

$$\delta_{g,t} \sim \text{Normal}(2 \cdot \delta_{g,t-1} - \delta_{g,t-2}, \sigma_{\delta}^2).$$

The variance parameters are shared across all subpopulation groups to enforce some pooling of information. All variance parameters take $\text{Normal}(0, 1)^+$ priors (weakly informative priors).

Our set-up resembles the TOPALS (De Beer 2012; Gonzaga and Schmertmann 2016) and P-TOPALS (Dyrting 2020) models except that in our case, the expected level is estimated.

3 Preliminary Results

3.1 Data

We test our model on the proportion being childless by race/ethnicity using Current Population Surveys (CPS) data Fertility and Marriage June supplement. The survey has been conducted every two years¹ and we consider the period 1990-2020. We define childlessness with the variable “frever” which indicates the number of live birth a women ever had. A women of a given age was defined childless when “frever” was equal to zero. Women considered are aged 15 to 44 years old. The variables “race” and “hispan” were combined to construct the racial/ethnic categories: “non-Hispanic Black”, “non-Hispanic White”, and “Hispanic”. The obtained counts were adjsuted based on survey weights.

3.2 Preliminary results

Figure 1 shows the percent of women childless (y-axis) over the ages 15 to 44 years old (x-axis) for the three racial/ethnic groups considered (horizontal panels). In order to fully illustrate the model, the vertical panels show the fit of two different models. The top panels reflect the fit with only a logistic function. The bottom panels present the fit allowing deviations from the logistic function with P-splines. The points and lines reflect the raw data and the model fit, respectively. From the Figure, the logistic function is generally modeling the proportion of childless women accurately but lacks some flexibility at the youngest and oldest ages (especially visible for non-Hispanic Black and Hispanic populations). Looking at the bottom panel, adding the P-splines improve the fit at these two extremes. Smaller population groups have wider estimated 95% credible intervals.

¹Except for the year 1996 which was collected in 1995 instead.

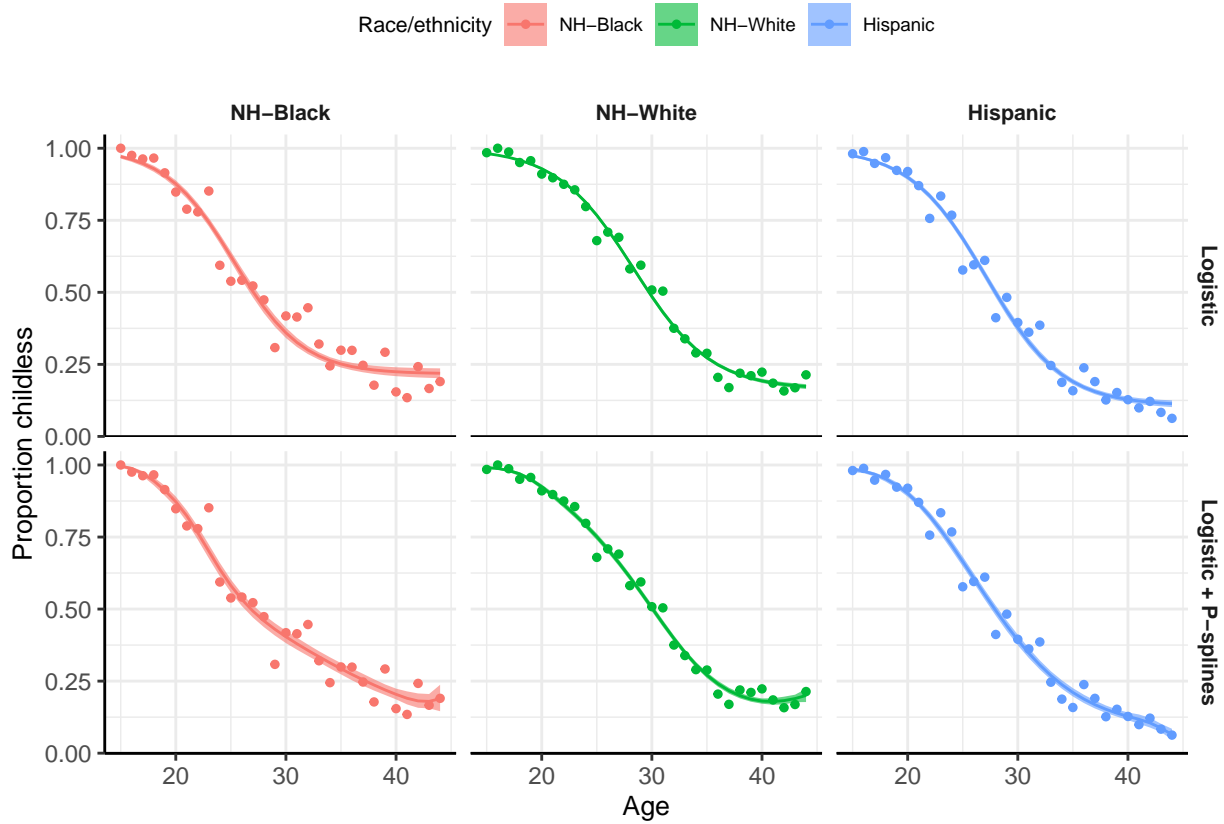


Figure 1: Proportion of women childless by race/ethnicity with and without deviations from a logistic function during the year 2020

Figure 2 shows the median posterior estimates of the three logistic parameters over the period 1990-2020. The right-panel reflecting the function's midpoint (δ) shows an upward trend, especially visible for Hispanic and non-Hispanic Black populations. This can be interpreted as a postponement of childbearing as it translates into the logistic function being shifted to the right (i.e to older ages). In line with the previous observation, the steepness of the curve (γ) shows a slight downward trend for all races/ethnicities. This illustrates that childbearing are spread over a wider age range. The middle panel showing the supremum of the function (ζ) does not reflect any particular pattern.

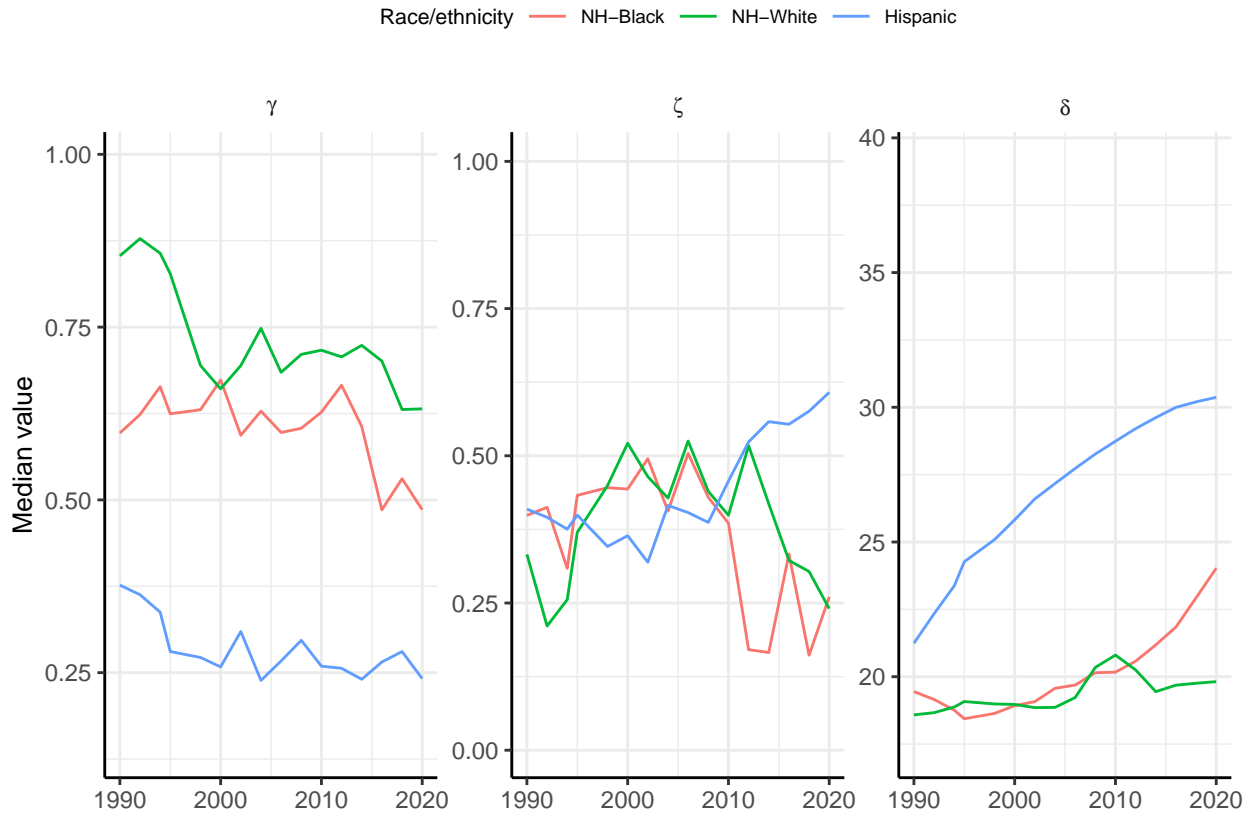


Figure 2: Median posterior estimates of the logistic parameters over the period 1990-2020

4 Next Steps & Extensions

In this abstract we proposed a modeling framework to estimate age-specific proportions of women being childless. We propose a semi-parametric model consisting of a logistic function plus some deviations modeled with P-splines. Preliminary results shown for childlessness by racial/ethnic groups are promising.

Future work will model childlessness by U.S. state and racial/ethnic group. This modeling framework can also be used to project childlessness proportions of different subpopulation groups in the future. Finally, this modeling framework could be expanded to estimate simultaneously different parities by using a Multinomial data model with a series of constrained (semi)-parametric functions.

References

- De Beer, J. (2012). Smoothing and projecting age-specific probabilities of death by TOPALS. *Demographic Research* 27:543–592.
- Dye, J. (2010). Fertility of american women: 2008. Current population reports, P20–563. *Washington, DC: US Census Bureau*.
- Dykstra, P.A. and Wagner, M. (2007). Pathways to childlessness and late-life outcomes. *Journal of Family Issues* 28(11):1487–1517.
- Dyrting, S. (2020). Smoothing migration intensities with p-TOPALS. *Demographic Research* 43:1607–1650.
- Eilers, P.H. and Marx, B.D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science* 11(2):89–121.
- Gemmill, A. (2019). From some to none? Fertility expectation dynamics of permanently childless women. *Demography* 56(1):129–149.
- Gonzaga, M.R. and Schmertmann, C.P. (2016). Estimating age-and sex-specific mortality rates for small areas with TOPALS regression: An application to brazil in 2010. *Revista Brasileira de Estudos de População* 33:629–652.
- Gray, E., Evans, A., and Reimondos, A. (2013). Childbearing desires of childless men and women: When are goals adjusted? *Advances in life course research* 18(2):141–149.
- Kendig, H., Dykstra, P.A., Gaalen, R.I. van, and Melkas, T. (2007). Health of aging parents and childless individuals. *Journal of Family Issues* 28(11):1457–1486.
- Kirmeyer, S. and Hamilton, B.E. (2011). *Childbearing Differences Among Three Generations of US Women*. Citeseer.
- Livingston, G. (2015). Childlessness falls, family size grows among highly educated women.
- Lundquist, J.H., Budig, M.J., and Curtis, A. (2009). Race and childlessness in america, 1988–2002. *Journal of Marriage and Family* 71(3):741–755.
- Mynarska, M., Matysiak, A., Rybińska, A., Tocchioni, V., and Vignoli, D. (2015). Diverse paths into childlessness over the life course. *Advances in life course research* 25:35–48.
- Nisén, J., Martikainen, P., Silventoinen, K., and Myrskylä, M. (2014). Age-specific fertility by educational level in the finnish male cohort born 1940–1950. *Demographic research* 31:119–136.