Childlessness Modeling: Updates and Future Work

Updates

- Corrected data: Significant drop in the percent of women aged 35-44 who are childless between 2010 and 2012. It has been shown by the U.S. Census Bureau to be related to changes in the data. I corrected the data for the years 2008 and 2010. Prior to 2008 the data does not contain the information required to perform the correction. Hence, the study period is now shorter: 2008-2020 (2022 data does not contain the info required to study childlessness).
- Previous model was too flexible (logistic function with splines): Janoschek sigmoidal function
- Coded the model in STAN: see fit.

Data

The Current Population Survey (CPS) has a Fertility and Marriage Supplement collected every two years in June. Women aged 15-44 years old are asked about their number of live birth. This information is recorded in the variable *frever*. A woman is considered to be a mother when $frever \neq 0$ (conversely, being a childless woman is defined by $frever == 0$).

Model

We assume the counts of mothers to be Binomially distributed as follows,

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

where $\theta_{x,g,t}$ is the proportion of women being mothers, $n_{x,g,t}$ is the number of women, x is age, g is the subpopulation considered (races/ethnicities, states), and t is the year.

We model $\theta_{x,g,t}$ with a **Janoschek sigmoidal function**:

$$\theta_{x,g,t} = \Theta_{Max_{g,t}} - (\Theta_{Max_{g,t}} - \Theta_{Min_{g,t}})e^{-\gamma_{g,t}x^{\delta_{g,t}}}$$

where

Θ_{Min} is the proportion of mother at age 15 years old, Θ_{Max} is the supremum of the values of the function, γ is a growth parameter for a fixed δ, and δ is below or above one for exponential or sigmoidal growth, respectively.

We smooth the Janoschek parameters by assuming that all parameters are distributed as Random Walk of order 2. The mean parameters are estimated for each subpopulation but the variance is assumed the same for all subgroups to enforce some pooling.

Aims for PAA (and paper)

Yearly extrapolation

The data is only available every two years. We allow for missing data by assuming that

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

It allows estimating $\theta_{x,g,t}$ for the missing years but the uncertainty does not reflect the fact that we don't have data on these years.

Age extrapolation

In 2012, the CPS surveyed women aged 15-50 years old (15-44 years old prior to 2012). The strategy proposed in the previous section could be used to obtain estimates for women aged 45-50 years old in the years 2008 and 2010.

Small-area estimation: state ∩ race

We set hierarchical priors on all Janoschek parameters of the form

$$\mu_{\Theta_{Max_{t,r}}} \sim \text{Normal}(2\mu_{\Theta_{Max_{t-1,r}}} - \mu_{\Theta_{Max_{t-2,r}}}, \sigma^2_{\mu_{\Theta_{Max}}})$$
$$\Theta_{Max_{t,r,s}} \sim \text{Normal}(\mu_{\Theta_{Max_{t,r}}}, \sigma^2_{\Theta_{Max}})$$

where r is race/ethnicity, and s is state. We assume the variances to be shared across race/ethnicity to obtain additional pooling.

Combining NSFG and CPS data

The National Survey of Family Growth (NSFG) asks the parity of surveyed women aged 15-44 years old (variable named *parity*). Each data release represents a two years period (i.e 2015-2017, 2017-2019). The NSFG is designed to be nationally representative, hence it can not be used to study childlessness at a state-level.

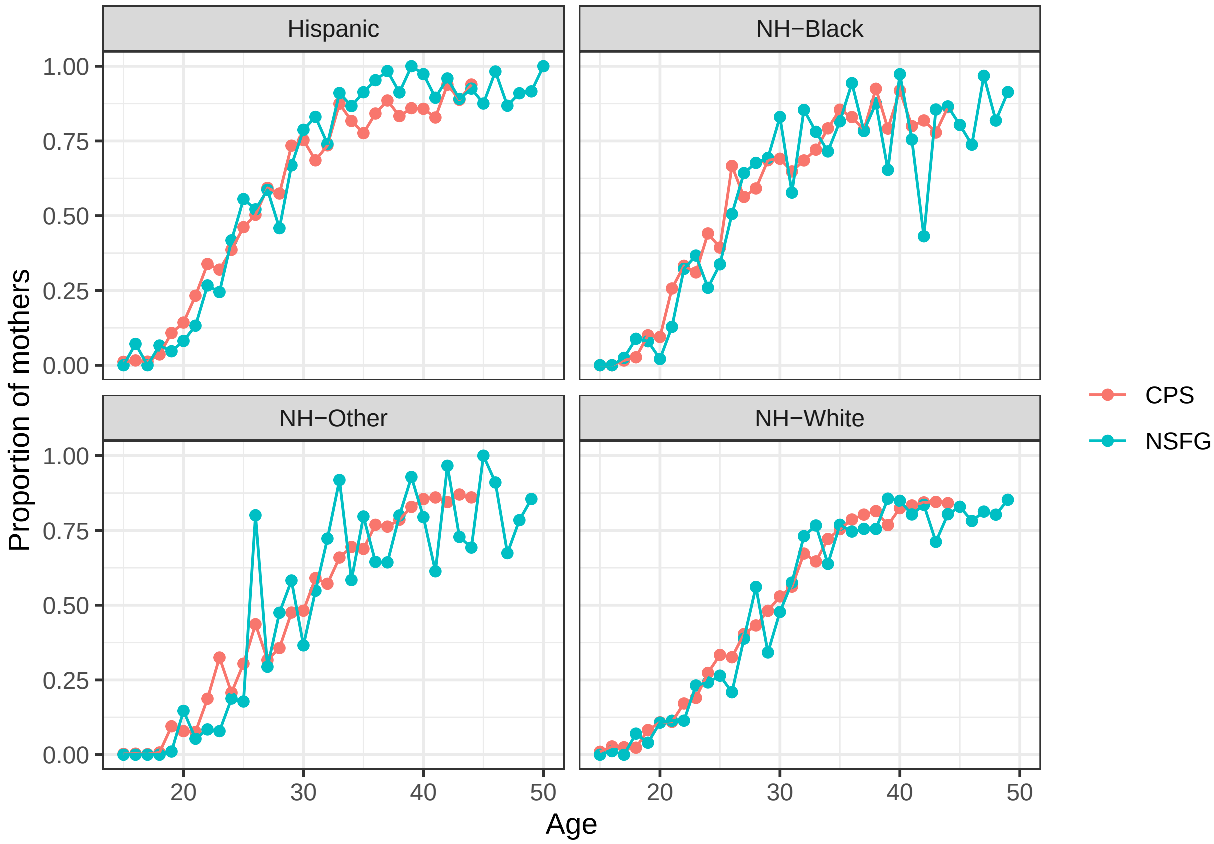

Figure 1: Proportion of mothers from CPS 2018 and NSFG 2017-2019

Bayesian modeling could combine the two data sources to estimate the proportion of childless women at the national level by race/ethnicity. The difficulty could be to combine one-year data with two-years data.

Childlessness of men

The *frever* or *parity* variables are not recorded for men. However, from the CPS (at least), we can obtain the proportion of men who are having a child in their household (with all the limitations related to divorce, older men having older children that might have left the household, . . .). This information might allow to develop a modeling strategy to estimate the proportion of men who are childless.