Projection of state-level foster care outcomes by race/ethnicity

Summary of technical model details

Monica Alexander

December 11, 2020

1 Overview

The purpose of this document is to outline the technical details of the statistical model used to project state-level outcomes related to the foster care system. Broadly, the model can be described as a **Bayesian hierarchical state space model**. It has several key parts:

1. A model to capture the association between the outcome rate and demographic, socioeconomic, health and welfare variables
2. A time component that captures state-specific fluctuations over time
3. A hierarchical structure such that information about levels, trends and patterns can be shared across states within regions
4. A projection model for the covariates
5. A data model that accounts for the varying amounts of volatility in trends across states

1.1 Overview of web-based application

Current results can be viewed here: https://monica-alexander.shinyapps.io/foster_care/. There are three tabs, accessed through the menu on the left hand side:

- National overview, which shows broad trends in entries at the US national level. You can choose to display the estimates as either number of entries per 1,000 children (population aged 0-18) or as the number of entries.

- State projections: This tab shows a graph of the estimated number of entries over time (as well as an uncertainty interval shown in red). Again the results can either be displayed as entries per 1,000 or just the number of entries. The first table below the graph shows the estimated probability that the number of entries will increase from year to year, and will increase from the 2017 level. The second table shows the 'top 5' covariates that are associated with changes in the projections.
- Covariates: This tab allows you to visualize the estimate association between entries and various covariates that are included in the model. The covariates are listed on the left hand side, and checking the box next to each variable adds that variable to the graphs. The graphs show the estimated association between entries and the variable selected over time and across census divisions. You can select up to 9 variables at a time.

2 Outcomes estimated

There are four outcomes estimated and projected

- Entries into the foster care system
- Investigations
- Permanent exits out of the system
- Non-permanent exits out of the system

These outcomes are estimated for six race/ethnicity groups:

- Total
- Non-Hispanic White
- Non-Hispanic Black
- Non-Hispanic Asian/Pacific Islander
- Non-Hispanic American Indian/Alaska Native
- Hispanic

3 Model

Define the outcome of interest y for a particular state and race/ethnicity group s and year t to be rate of a particular outcome of interest per child population, i.e.

$$y_{s,t} = \frac{\text{Number of outcome}_{s,t}}{\text{Population aged 0-18}_{s,t}}$$

Where the outcome is entries, exits or investigations.

The goal is to model and project forward $y_{s,t}$ five years past the most recent observation (in 2017). The $y_{s,t}$ are modeled on the log scale and then transformed back to the natural scale to ensure the outcome is always positive. In particular, we assume

$$\log y_{s,t} \sim N(\mu_{s,t}, s_y^2)$$

where μ_{st} is the expected log rate, and s_y^2 is the **stochastic standard error** associated with the observations. Accounting for the stochastic error allows the model to take into account that rates are naturally more volatile in some states than others, because the population exposed to risk is smaller. In practice, s_y^2 is larger for smaller populations.

The expected log rate $\mu_{s,t}$ has the form

$$\mu_{s,t} = \alpha_s + \mathbf{X_{s,t}}'\beta_{r,t} + \delta_{s,t} \tag{1}$$

where

- α_s is a state-specific intercept
- $\mathbf{X_{s,t}}$ is a vector of K covariates for that particular state s and year t
- β_r, t is a matrix of length K the region- and year-specific effects of the covariates
- $\delta_{s,t}$ are state-year fluctuations

The following subsections explain in more detail:

- The hierarchical model for α_s
- The set of covariates considered $\mathbf{X}$
- The projection model(s) for the covariates $\mathbf{X}$
- The time series model the δ_{st}
- Steps of projection

3.1 Hierarchical structure

The natural hierarchical structure of the data (states within regions within the US), and the fact that some states are smaller and have more volatile patterns that other states, suggest that a hierarchical model would be appropriate. The state-specific intercepts α_s are modeled hierarchically within census division r such that

$$\alpha_s \sim N(\mu_\alpha[r], \sigma^2_\alpha[r])$$

This set-up assumes that the state-specific effects α_s are a draw from a region-level distribution with some common mean and an associated variance. In practice this allows for information about levels and trends to be shared across states within the same region. The smaller the population in a particular state, the more that state's estimates of α are influenced by the overall mean μ_α.

The regions in which states were grouped were chosen to be Census divisions. These are a convenient choice; however, exploratory data analysis of patterns in the rates across states suggests that there are noticeable similarities across states within Census divisions, suggesting they are a reasonable grouping of states.

3.2 Set of covariates included in the model

There are many different factors (or covariates) that could potentially be associated with changes in rates over time, including demographic, socioeconomic, geographic, health and welfare factors. The model currently has a set of 28 covariates of each of these different types. Please see the web-based application 'Covariates' tab for the list of what is included. The decisions to include these covariates in the model was based on:

- Exploratory data analysis of the raw bivariate correlations between covariates and entries;
- Advice and input from the Casey team on suitability of covariates, access to data and modifiability;
- Model testing and evaluation.

3.3 Varying association between entries and covariates across geography and time

In the model, the relationship between each covariate is allowed to vary by Census division and by time. In addition, the relationship between each covariate within each division $\beta_{r,t}$ is modeled as a time series, in particular:

$$\beta_{r,t} \sim N(2 \cdot \beta_{r,t-1} - \beta_{r,t-2}, \sigma_\beta^2) \tag{2}$$

This model captures the fact that, while the relationship between foster care entries and a particular covariate might change over time, the association in particular year t is likely to be similar to the association in the previous year $t-1$.

3.4 Projection model for the covariates

Equation 1 suggests a direct relationship between the outcome $\log y_{s,t}$ and a set of covariates $\mathbf{X}_{s,t}$ at the same time point. This means that to obtain projections for y we also need projections for $\mathbf{X}$. Each covariate is currently projected forward assuming a three-year moving average.

3.5 State-time component

The final piece of Equation 1 is the $\delta_{s,t}$. This term aims to capture any fluctuations over time within each states that are not already explained by changes in the covariates. Note that in a more traditional regression set up, these $\delta_{s,t}$'s would usually be assumed to be independent and identically distributed, e.g. $\delta_{s,t} \sim N(0, \sigma^2)$. However, as this model deals with time, we model $\delta_{s,t}$ to take into consideration autocorrelation over time. In particular, the $\delta_{s,t}$'s are modeled as an auto-regressive process, i.e.

$$\delta_{s,t} \sim N(\rho_s \delta_{s,t-1}, \sigma_\delta^2) \tag{3}$$

where $\rho_s \in [0, 1]$ and with the first observation in each state as

$$\delta_{s,1} \sim N(0, \sigma_\delta^2)$$

This set-up assumes that the $\delta_{s,t}$ in a particular time period is correlated to the value in the previous time period. Values of δ_s, t can be projected forward using this equation. In terms of the projections, the value of $\delta_{s,t}$ will eventually converge to zero.

3.6 Steps of projection

The projection of the rate occurs through the projection of the covariates, coefficients on the covariates, and state-time components. To obtain a projection for the next time period $y_{s,T+1}$, the broad steps are

1. Project forward each covariate using a three-year moving average.
2. Project forward each covariate coefficient using Equation 2:

$$\beta_{r,T+1} \sim N(2 \cdot \beta_{r,T} - \beta_{r,T-1}, \sigma_\beta^2)$$

3. Project forward $\delta_{s,T+1}$ using Equation 3:

$$\delta_{s,T+1} \sim N(\rho_s \delta_{s,T}, \sigma_\delta^2)$$

4. Calculate projection of the expected rate, $\mu_{s,T+1}$ based on Equation 1:

$$\mu_{s,T+1} = \alpha_s + \mathbf{X_{s,T+1}}' \beta_{r,t} + \delta_{s,T+1}$$