Future projections of child welfare outcomes in the United States using a Bayesian state-space model

Monica Alexander[1]

Christopher Wildeman[2]

Alexander Roehrkasse[2]

[1] University of Toronto

[2] Duke University

Paper presented at PAA 2022

Abstract

Anticipating demand for foster care is important for planning and policymaking. In this paper we propose a Bayesian state-space model for estimating and projecting child welfare outcomes by state in the United States. The model is formulated within a Bayesian hierarchical framework, incorporating information about the relationship between changes in foster care populations and changes in other factors such as incarceration, suicide and overdose mortality rates. The approach also enables states to formulate expectations of future trends by drawing on information from neighboring states. We project multiple outcomes for different race/ethnicity groups, including annual per capita entries and exits into foster care and investigations of child abuse and neglect. We also develop an interactive web-based application accompanying the model, which allows administrators and policy-makers to understand more easily the likely effects of changes in key parameters and the associated uncertainty in projections.

1 Introduction

Obtaining future projections of indicators related to the foster care system is important for planning and policymaking. From a planning point of view, projections are important to anticipate any resource deficiencies based on a likely increase in caseloads, or an increase those entering the foster care system, for example. From a policymaking perspective, understanding why indicators of interest are changing may help to improve outcomes through targeted programs or resource allocations. There are multiple outcomes of interest, including the likely number of investigations of child abuse and neglect in a given time frame, entries into the foster care system, and exits out of the system, both permanent and non-permanent. In the US context, state-level policymakers are also interested in trends of outcomes for different race/ethnicity groups.

At its heart, the problem of obtaining projections foster care outcomes could thought as a purely time series exercise. For planning purposes, we are interested in obtaining projections that are as accurate as possible for a particular outcome and population group, which in many cases is best achieved by modeling the set of outcomes using time series techniques, and choosing a method that produces the best metrics in terms of out-of-sample predictive performance. Focusing on this perspective reduces the problem of obtaining projections as a prediction exercise. However, in addition to forecasting the likely trajectory of these outcomes, policymakers are also interested in demographic, socioeconomic, health, and welfare indicators that may be associated with child welfare outcomes, particularly if such indicators are modifiable through interventions or other policies. From this perspective, the problem of obtaining projections now becomes one of inference. We are not solely interested in obtaining the most accurate forecasts as possible, but also in understanding why indicators are moving up or down.

In addition to these competing priorities, projecting child welfare outcomes quickly becomes a complex statistical problem for other reasons, particularly when considering projections across multiple geographies and race/ethnicity groups. Firstly, trends in outcomes differs substantially across indicators, geography and race/ethnicity. Trends are not linear, and can change direction rapidly in a short time frame. This suggests a suitably flexible statistical model is required. Secondly, population sizes in some race/ethnicity groups are relatively small, which causes trends in observed data to be erratic and uncertain. We need a statistical model that adequately accounts for differing amounts of uncertainty across groups and propagates this uncertainty through projections, and has the ability to smooth erratic trends over time.

Additionally, there are hundreds of potential covariates that may be associated with child welfare outcomes, and the association between the outcomes and a particular covariate may change over time, and is likely to be different across geographic space. In previous work, Swann and Sylvester (2006) argued that increases in female incarcerations and reductions in cash welfare benefits explained a large part of the growth in state-level foster care caseloads over the period 1985 to 2000. However, it is not clear whether these observations still hold for the most recent past, particularly in light of the current opioid epidemic, which has a strong presence in much of the United States. In addition, the opioid epidemic in the United States has affected states differently and has evolved over time with regards to the types of populations affected (Alexander, Kiang, and Barbieri 2018; Kiang et al. 2019), which suggests the association between, for example, entries into the foster care system and drug overdoses, may have changed over time.

Previous work has focused on understanding individual-level predictive factors of exposure to the foster care system (English, Thompson, and White 2015; Logan-Greene and Semanchin Jones 2018; Davidson et al. 2019). In contrast, for this project we were interested in state-level trends and projections to better inform child welfare policy. In particular, the goal of this paper is to project child welfare outcomes over the next 5 years at the state level in the United States. We are also interested in investigating associations between changes in outcomes and other demographic, socioeconomic, health and welfare measures to try and identify key points of intervention. We introduce a Bayesian state-space model to project multiple child welfare outcomes for different race/ethnicity groups. The model accounts for changes in key demographic, socioeconomic, health and welfare factors, and allows for this association to vary over space and time. The approach also enables states to formulate expectations of future trends by drawing on information from neighboring states. This application demonstrates the utility of state-space models in flexibly estimating and projecting demographic outcomes.

The remainder of the paper is structured as follows. We first give a brief overview of state-space models in general, dynamic linear models in particular (which is what is used in this context), and show how these models can be formulated as a Bayesian hierarchical model. We then discuss the data sources used to obtain both the outcomes of interest and possible covariates that are included in the projection models. The broad modeling framework is then presented, and some results are illustrated. We also discuss the interactive web-based application that was developed to aid policymakers in using this projection model. The final section discusses limitations and possible extensions.

2 An overview of state-space models and dynamic linear models

At a very broad level, state-space models describe how a particular process or state x_t evolves over time, and how those states relate to data we observe, y_t. The idea is that the latent states x_t change over time through some process (which we don't see but can model), and this underlying process drives changes in our observed data, y_t. State-space models were initially proposed by Kalman (1960), in the space tracking setting, where the state equation defines the motion equations for the position or state of a spacecraft with location x_t and the data y_t reflect information that can be observed from a tracking device such as velocity. Models of this type are used extensively in ecology to track animal movement (Langrock et al. 2012), in macroeconomic modeling (Harvey and Koopman 2009), and other physical and engineering problems (Hamilton 1994).

2.1 Dynamic linear models

The linear Gaussian state-space model, also called a dynamic linear model, assumes Normal errors and can be written in a general form as

$$
\begin{aligned}
y_t &= F_t x_t + v_t, \quad v_t \sim N\left(0, V_t\right) \\
x_t &= G_t x_{t-1} + w_t, \quad w_t \sim N\left(0, W_t\right)
\end{aligned}
$$

Above y_t are the p observations at time t, with $t = 1, \ldots, n$. Vector x_t of length m contains the unobserved states of the system that are assumed to evolve in time according to a linear system operator G_t (a $m \times m$ matrix). We observe a linear combination of the states with noise and matrix F_t $(p \times m)$ is the observation operator that transforms the model states into observations. Both observation end system equations can have additive Gaussian errors with covariance matrices V_t and W_t.

The development of state-space models was in situations where the outcome of interest was the latent states x_t. In our case, we shift the viewpoint of the analysis to the observed outcomes y_t, with the goal of relating these to a set of k covariates $\mathbf{X}_t$ through regression coefficients $\boldsymbol{\beta}_t$. For example a simple dynamic linear regression would have the form

$$
\begin{aligned}
y_t &= \boldsymbol{X}_t' \boldsymbol{\beta}_t + \epsilon_t \\
\boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t \\
\epsilon_t &\sim N\left(0, \sigma_\epsilon^2\right) \\
\boldsymbol{\eta}_t &\sim N\left(\mathbf{0}, \Sigma_\eta\right)
\end{aligned}
$$

The first line here is our usual linear regression set-up, with the only difference being the regression coefficients $\boldsymbol{\beta}_t$ vary over time. In the model above in particular, we are assuming the regression coefficients evolve according to a random walk over time: the current value of β_t is the value from the previous period, plus some error. This model set up allows the association between the outcome of interest y_t and a set of covariates, and allows that association to vary over time.

2.2 Dynamic linear regression as a Bayesian hirerachical model

The dynamic linear model formulation can be seen as a special case of a general hierarchical statistical model with three levels: data, process and parameters. First, the observation uncertainty $p(y_t|x_t, \theta)$ described by the observation equation and forming the statistical likelihood function. For example, in the linear regression above, we have $p(y_t|\beta_t, \sigma_\epsilon^2)$. Second, the process uncertainty of the unknown states x_t and their evolution given by the process equations as $p(x_t|\theta)$ or $p(x_t|x_{t-1}, \theta)$; in our example above this is $p(\beta_t|\beta_{t-1}, \Sigma_\eta)$. And third, the unconditional prior uncertainty for the model parameters $p(\theta)$, which in our example is $p(\sigma_\epsilon^2, \Sigma_\eta)$. Using the Bayes formula, we can write the state and parameter posterior distributions as a product of the conditional distributions

$$p(x_t, \theta|y_t) \propto p(y_t|x_t, \theta)p(x_t|\theta)p(\theta)$$

which is the basis for full Bayesian estimation procedures. Indeed, one option for estimating dynamic linear models is using Kalman filters (Kalman 1960); however, considering the model as a Bayesian hierarchical model allows us to estimate parameters of interest using Markov Chain Monte Carlo (MCMC) techniques, which are readily fit using standard statistical tools such as Stan or JAGS.

State-space models, and in particular, dynamic linear regression models, are useful in the context of modeling time series structurally; that is, trying to understand changes in an outcome of interest over time, and how those changes are related to other factors of interest. In this particular case, we would like the relationship between child welfare indicators and important covariates, such as incarceration rates and drug overdose mortality, to be able to vary overtime, to account for the changing social and economic conditions underlying these processes. Dynamic linear regression models, as conceived as Bayesian hierarchical models, are a useful tool to codify these temporal changes.

3 Data

3.1 Scope, units, categorization

Data collection took place between August 2019 and December 2020. Analysis included all 50 states and the District of Columbia, hereafter referred to as a state. The historical range of observed data was 2005-2018. The unit of analysis is the state-year.

Analysis was conducted for all races/ethnicities as well as separately by ethnoracial group. We group individuals into four single-race non-Hispanic/Latino groups—American Indian/Alaska Native, Asian American/Pacific Islander, Black/African American, and White—and one Hispanic/Latino group. The Hispanic group includes multiracial children; non-Hispanic multiracial children, who made up 7.6% of all children in foster care between 2000 and 2018, are excluded from group analyses.

3.2 Child welfare outcomes

Four different child welfare outcomes were measured as counts: investigations of child abuse or neglect; entries into the foster care system; permanent exits from the system; and non-permanent exits from the foster care system. Investigations were counted multiple times if they pertained to multiple children, and all events were counted multiple times if children experienced them more than once in a given year. Outcome measures were limited to individuals aged less than 18 years.

Data on child welfare outcomes were accessed through the National Data Archive on Child Abuse and Neglect (NDACAN, https://www.ndacan.acf.hhs.gov/datasets). Restricted-access microdata were used to create counts. Data on maltreatment investigations came from the National Child Abuse and Neglect Data System (NCANDS) child file. NCANDS is a nationwide voluntary reporting system through which states report investigations of child maltreatment carried out by local child protective services (CPS) agencies. Investigations correspond to "screened-in" reports, or the subset of all reports to CPS of suspected maltreatment that are determined credible and serious enough to warrant investigation. NCANDS includes only those investigations that receive a disposition, and therefore excludes investigations that were not concluded within the historical range of analysis. Child maltreatment investigations are a leading indicator of foster care placements because most foster placements result from a investigation.

Data on entries and exits to and from the foster care came from the Adoption and Foster Care Analysis and Reporting System (AFCARS) foster care file. AFCARS is a nationwide mandatory system through which states report on all children in foster care at any point over the fiscal year. Foster care is defined as out-of-home care in a foster family home (relative or non-relative), group home, institution, or supervised independent living. Children are removed from parental care and placed into foster care for a variety of reasons, including physical or sexual abuse or neglect of the child; drug or alcohol abuse by the parent or child; child disability or behavior problem; parental death or incarceration; or caretaker inability to cope, abandonment, relinquishment, or inadequate housing. Children exiting care are measured separately by destination. Children exiting care to permanency include children reunified with parents, adopted, or placed permanently with a guardian or relative. Children exiting care without permanency include children who have aged out of foster care, been transferred to another agency, run away, or died.

3.3 Covariates

A large number of covariates were considered in this analysis, and obtained from a number of different sources. A full summary of covariates considered and data sources is presented in the Appendix.

Among chosen covariates, a variety of demographic, economic, housing, and social support measures were calculated using 1-year American Community Survey (ACS) via IPUMS-USA (https://usa.ipums.org/usa/). These included measures of the educational attainment of the population aged 18-64; the ethnoracial composition, urbanization, nativity, geographic mobility, and family composition of the child population; family income of children; average persons per bedroom in households, average gross monthly rent, and median value of occupied housing units; welfare recipients per 1,000 population; percentage of school-aged children in school; and the unemployment rate.

Measures of social policy and social support came from the University of Kentucky Center for Poverty Research's National Welfare Data. These included: AFDC/TANF recipients per 1,000 population; total and reduced NSLP participants and reduced SBP participants per 1,000 school-aged children; and minimum wage (highest of state or federal). Measures of the median salary of social workers and of community and social services specialists came from the Bureau of Labor Statistics, and upper- and lower-bounded measures of child welfare caseworker turnover came from Edwards and Wildeman.

Public health data from the National Vital Statistics System included crude measures of: post-neonatal infant mortality rates; mortality rates for children aged 1-4; accident and non-transport accident mortality rates for children aged less than 18; and male and female suicide mortality rates for adults aged 18+. Measures of the percentages of the adult population with alcohol use disorder or illicit drug use disorder were from the National Study of Drug Use and Health.

Criminal justice data came from the National Prisoner Statistics, accessed via the ICPSR, and included measures of annual female prison admissions per 100,000 females and end-of-year counts of males in prison custody per 100,000 males. Measures of income inequality, namely the percentage of total income received by the top 1% of earners and 10% of earners, were sourced from Mark Frank's U.S. State-Level Income Inequality Data (https://www.shsu.edu/eco_mwf/inequality.html). All monetary variables were adjusted for inflation to 2018 U.S. dollars and adjusted for state differences in cost of living using regional price parities from the Bureau of Economic Analysis.

4 Model

Broadly, our modeling framework is 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

There are four outcomes estimated and projected: Entries into the foster care system; Investigations; Permanent exits out of the system; and 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; and Hispanic.

4.1 Model details

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

4.2 Hirerachical 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.

4.3 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. 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 domain experts (from Casey Foundation) on the suitability of covariates, access to data and modifiability;
- Model testing and evaluation.

Covariates included are detailed in Table 1.

4.4 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 $t-1$.

4.5 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.

Table 1: List of variables included in model

variable	category
Percentage of adult (18-64) population with associate's degree	Education
Percentage of adult (18-64) population with bachelor's degree	Education
Percentage of adult (18-64) population with graduate degree	Education
Percentage of adult (18-64) population with HS degree	Education
AFDC/TANF recipients per 1,000 population	Social support
Percentage of adults (18+) reporting alcohol disorder in the past year	Public health measures
Average monthly gross rent	Housing measures
Average persons per bedroom	Housing measures
Percentage of children foreign born	Demographic measures
Median real family income of children	Economic well-being
Percentage of children living in metropolitan area	Demographic measures
Percentage of children changing residence in prior year	Housing measures
Percentage of children with absent father	Demographic measures
Percentage of children white (only) non-Hispanic	Demographic measures
Percentage of children black (only) non-Hispanic	Demographic measures
Percentage of children Native American (only) non-Hispanic	Demographic measures
Percentage of children Hispanic	Demographic measures
Median salary of community and social service specialist	Child welfare measures
Share of income going to top 1% of earners	Economic well-being
Share of income going to top 10% of earners	Economic well-being
Median value of occupied housing units	Housing measures
Real federal minimum wage (2018USD)	Economic well-being
Crude mortality rate for children age 1 to 4 (deaths divided by 100,000 births)	Public health measures
Crude child mortality rate for accidents (deaths divided by 100,000 births)	Public health measures
Crude child mortality rate for non-transport accidents (deaths divided by 100,000 births)	Public health measures
Post-neonatal infant mortality crude rate (deaths divided by 1,000 births)	Public health measures
NSLP reduced participants per 1,000 school-aged children	Social support
NSLP total participants per 1,000 school-aged children	Social support
Crude death rate (deaths per 100K) by drug overdose for adult (18+) females	Public health measures
female prison admits per 100K female	Criminal justice
male prison population per 100K males	Criminal justice
SBP reduced participants per 1,000 school-aged children	Social support
Percentage of school-aged children in school	Education
Median salary of social worker	Child welfare measures
Crude death rate (deaths per 100K) by suicide for adult (18+) females	Public health measures
Crude death rate (deaths per 100K) by suicide for adult (18+) males	Public health measures
Annual caseworker turnover (lower-bound)	Child welfare measures
Annual caseworker turnover (upper-bound)	Child welfare measures
Unemployment rate	Economic well-being
Welfare recipiants per 1,000 population	Social support

4.6 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.

4.7 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:

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

4.8 Priors

This model was fit in a Bayesian framework and as such priors were placed on all hyper-parameters. Weakly informative priors were placed on all parameters; in particular half-standard Normal priors were used for all variance parameters and standard Normal priors were placed on the μ_α's.

4.9 Computation

Posterior samples were obtained using Hamiltonian Monte Carlo implemented in Stan via the `rstan` R package, with 4 parallel chains of 2000 warmup iterations and 4000 sampling iterations. Standard checks for $\hat{R}$ and effective sample size were performed. All code is available at: https://github.com/MJAlexander/child_welfare_projections.

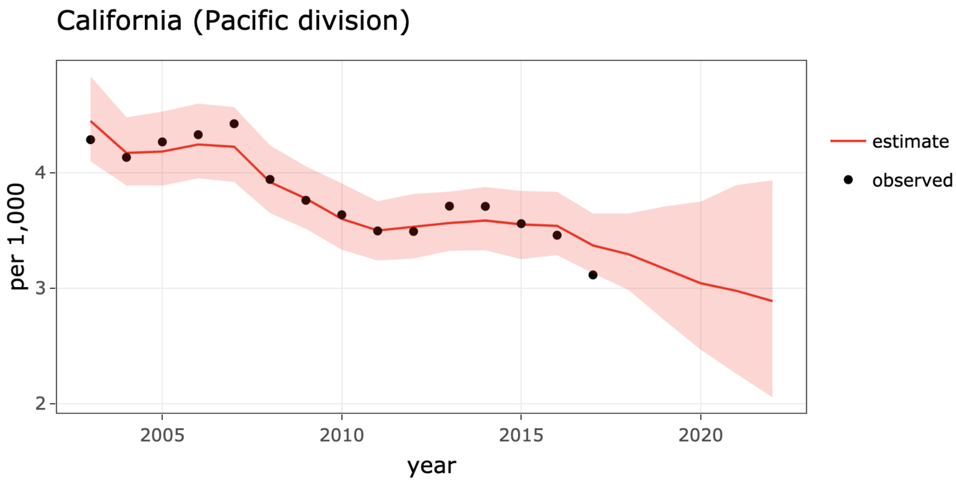
5 Results

Illustrative results for 3 example states and outcomes for the total population are shown in Figure 1. For Californian entries, there is a trend downwards, and the model estimates a probability of increase in 2022 at around 30%. Notable associated measures include alcohol disorder prevalence (which is positively associated) and household size (which is negatively associated). For investigations in DC, the probability of increase in 2022 is around 57%, and notable associations include salary of social worker (negative) and alcohol disorder prevalence (positive). For entries in Wyoming, the probability of increase is 84%,and notable associations are salary of social worker (negative) and alcohol disorder prevalence (positive).

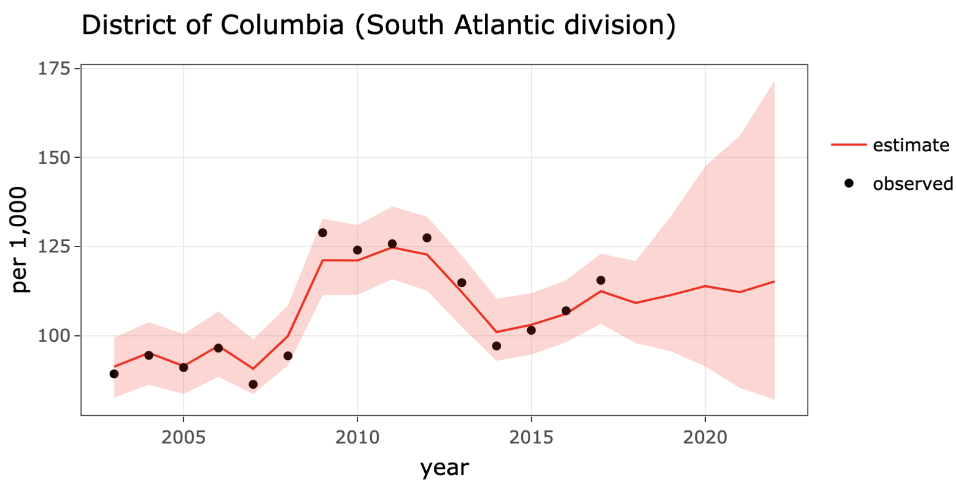
5.1 Shiny application

All current results can be viewed here: https://monica-alexander.shinyapps.io/foster_care/. There are three tabs, accessed through the menu on the top of the screen:

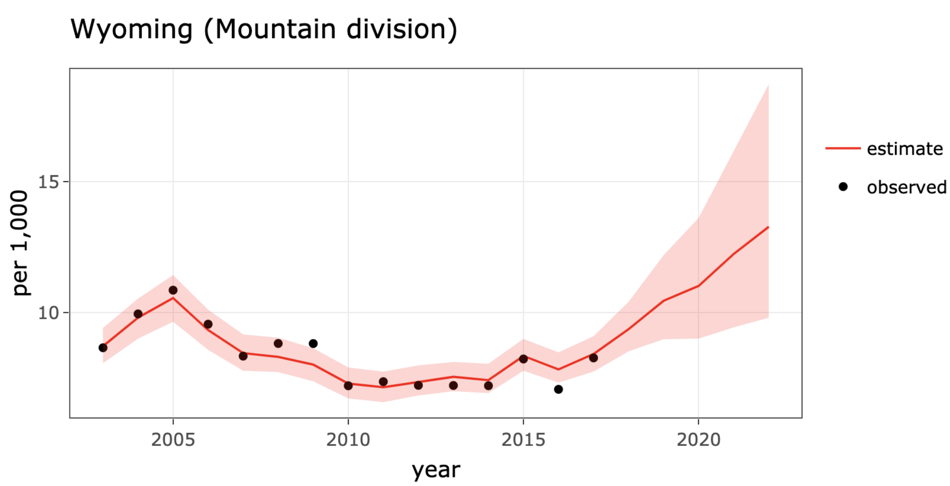
- National overview, which shows broad trends in entries at the US national level. You can choose to display the estimates as either number of entries per 1,000 children (population aged 0-18) or as the number of entries.



(a) California entries (total population)



(b) DC investigations (total population)



(c) Wyoming entries (total population)

Figure 1: Estimated and projected outcomes for three example populations

- 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.

6 Summary

In this paper we present a Bayesian state-space model to estimate and project key child welfare outcomes by state for different race/ethnicity groups. The model includes associations with numerous covariates, which are allowed to vary over time and across space. The model also accounts for varying uncertainty in different populations. Accompanying this work is a web-based Shiny application which allows the results to be disseminated in a understandable, interactive way. The model results and accompanying interactive tool have now been incorporated into policy planning dashboards used by the Casey Foundation, and are actively been reviewed and updated.

Beyond the application to child welfare outcomes, this work illustrates the utility of Bayesian hierarchical state-space models more generally in social science applications. Considering how an outcome of interest evolves over time, and how that evolution is related to changes in underlying driving factors, is an important problem that appears in many contexts, particularly in issues related to public policy. Increased computational power and availability of computational tools that make large-scale Bayesian models feasible to fit means that these class of models are now readily available for social scientists to consider.

A clear limitation of the projection model as it currently stands is the absence of any accounting for possible effects of the COVID-19 pandemic. It is difficult to forecast the flow-on effects of such a large, unprecedented event. From a modeling standpoint, the most effective approach may be scenario-based, where we consider a range of plausible scenarios of increase or decrease in important covariates, and then see how tha affects child welfare projections. Future work will focus on this aspect, particularly as new data become available.

Appendix

Table 2: Data source for all covariates considered

Variable label	Dataset(s)	Organization/Source	Years Available
Total population	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Number of children (<18)	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Number of school-age children (6-17)	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Number of persons 65 and older	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children white (only) non-Hispanic	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children black (only) non-Hispanic	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children Hispanic	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children Asian/Pacific Islander (only) non-Hispanic	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children Native American (only) non-Hispanic	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children with working mother	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Average hours worked by working mothers	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children with absent father	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children living in metropolitan area	IPUMS ACS 1-Year (2005-2018); DC 2000 1%	U.S. Census Bureau	2000-2018
Percentage of children foreign born	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of population children	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of population 65 and older	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Child/family/school social workers per 1,000 children	Occupational Employment Statistics	Bureau of Labor Statistics	1997-2018
Median real cost-of-living adjusted salary of child/family/school social workers	Occupational Employment Statistics	Bureau of Labor Statistics	1997-2018

:: Median salary of social worker	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Community and social service occupations per 1,000 children	Occupational Employment Statistics	Bureau of Labor Statistics	1997-2018
Median real salary of community and social service occupations	Occupational Employment Statistics	Bureau of Labor Statistics	1997-2018
:: Median salary of community and social service specialist	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Annual caseworker turnover (lower-bound)	NA	Edwards and Wildeman	2006-2015
Annual caseworker turnover (upper-bound)	NA	Edwards and Wildeman	2006-2015
Annual supervisor turnover	NA	Edwards and Wildeman	2006-2015
Percentage of children changing residence in prior year	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children living in public housing	CPS ASEC	U.S. Census Bureau	1976-2017
Percentage of housing units vacant	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Average persons per bedroom in household	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Average monthly gross rent	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Median rent as proportion of family income	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Median value of occupied housing units	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children living in owner-occupied dwellings	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Rent burden	Eviction Lab	Eviction Lab at Princeton University	2000-2016
Eviction rate	Eviction Lab	Eviction Lab at Princeton University	2000-2016
Eviction filing rate	Eviction Lab	Eviction Lab at Princeton University	2000-2016
Real GDP per capita	NA	Bureau of Economic Analysis	1977-2018
Median real cost-of-living-adjusted family income of children	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Unemployment rate	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018

:: Unemployment rate	National Welfare Data	UKCPR	1980-2017
Percentage of children with family income below 100% poverty line	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children with family income below 75% poverty line	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of children with family income below 50% poverty line	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Real minimum wage (2018 USD)	National Welfare Data	UKCPR	1980-2017
Gini coefficient	NA	Mark Frank	1917-2015
Percentage of total income received by top 10% of earners	NA	Mark Frank	1917-2015
Percentage of total income received by top 1% of earners	NA	Mark Frank	1917-2015
Percentage of total income received by top 0.1% of earners	NA	Mark Frank	1917-2015
Post-neonatal deaths per 1,000 live births	CDC Wonder	CDC	1999-2019
Mortality rate, 1-4 year olds	CDC Wonder	CDC	1999-2019
Mortality rate, 5-14 year olds	CDC Wonder	CDC	1999-2019
Child mortality rate, non-transport accidents	CDC Wonder	CDC	1999-2019
Child mortality rate, accidents	CDC Wonder	CDC	1999-2019
Child maltreatment fatality rate	NCANDS	NDACAN	2002-2018
Child mortality rate	CDC Wonder	CDC	1999-2019
Percentage of population marginally food insecure	National Welfare Data	UKCPR	1999-2019
Percentage of population food insecure	National Welfare Data	UKCPR	2001-2017
Percentage of population very low food secure	National Welfare Data	UKCPR	2001-2017
Percent adult population with alcohol use disorder	NSDUH	SAMHSA	2002-2017
Percent adult population with illicit drug use disorder	NSDUH	SAMHSA	2002-2017
Male adult substance abuse treatment admissions per 1,000 adult male population	TEDS-A	SAMHSA	2000-2017

Female adult substance abuse treatment admissions per 1,000 adult female population	TEDS-A	SAMHSA	2000-2017
Unintentional drug overdose deaths rate, adult men	CDC Wonder	CDC	1999-2019
Unintentional drug overdose deaths rate, adult women	CDC Wonder	CDC	1999-2019
Suicide rate, adult men	CDC Wonder	CDC	1999-2019
Suicide rate, adult women	CDC Wonder	CDC	1999-2019
Percentage of adult population meeeting criteria for heavy drinking	BRFSS	CDC	2000-2018
Percentage of adult population having fair or poor health	BRFSS	CDC	2000-2018
Percentage births low birthweight	CDC Wonder	CDC	1995-2018
Teen birth rate (births to minor women per 1,000 minor women)	CDC Wonder	CDC	1995-2018
Number of ER Visits per 1000 people, by ownership type (govt/non-profit/for-profit)	American Hospital Association (AHA) Annual Surveys	The Henry J. Kaiser Family Foundation	1999-2018
Welfare recipiants per 1,000 population	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Average welfare income among recipients	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Natural logarithm of real cost-of-living-adjusted maximum monthly AFDC/TANF benefit for 3-person family (2018 USD)	National Welfare Data	UKCPR	1980-2018
AFDC/TANF recipients per 1,000 population	National Welfare Data	UKCPR	1980-2018
Percentage of children participating in TANF	Characteristics and Financial Circumstances of TANF Recipients	Office of Family Assistance, ACF	2001-2018
Natural logarithm of real cost-of-living-adjusted maximum monthly Food Stamp/SNAP benefit for 3-person family (2018 USD)	National Welfare Data	UKCPR	1980-2018

Food Stamp/SNAP recipients per 1,000 population	National Welfare Data	UKCPR	1980-2018
Medicaid recipients per 1,000 population	National Welfare Data	UKCPR	1990-2018
WIC participants per 1,000 population	National Welfare Data	UKCPR	1989-2018
NSLP free participants per 1,000 school-aged children	National Welfare Data	UKCPR	1989-2018
NSLP reduced participants per 1,000 school-aged children	National Welfare Data	UKCPR	1989-2018
NSLP total participants per 1,000 school-aged children	National Welfare Data	UKCPR	1989-2018
SBP free participants per 1,000 school-aged children	National Welfare Data	UKCPR	1989-2018
SBP reduced participants per 1,000 school-aged children	National Welfare Data	UKCPR	1989-2018
SBP total participants per 1,000 school-aged children	National Welfare Data	UKCPR	1989-2018
State/local spending on education (% total personal income)	State Policy Database	Ruger & Sorens	1957-2017
State/local spending on health/hospitals (% total personal income)	State Policy Database	Ruger & Sorens	1957-2017
State/local spending on housing/com dev (% total personal income)	State Policy Database	Ruger & Sorens	1957-2017
State/local spending on public welfare (% total personal income)	State Policy Database	Ruger & Sorens	1957-2017
Governor is Republican	NA	IPPSR/Klarner through 2011; NCSL after	1937-2018
State senate is controlled by Republicans	NA	IPPSR/Klarner through 2011; NCSL after	1937-2018
State house is controlled by Republicans	NA	IPPSR/Klarner through 2011; NCSL after	1937-2018
State government is unified Republican	NA	IPPSR/Klarner through 2011; NCSL after	1937-2018
State government is unified Democrat	NA	IPPSR/Klarner through 2011; NCSL after	1937-2018

Male prison admissions per 100,000 males	National Prisoner Statistics	DOJ	1978-2018
Female prison admissions per 100,000 females	National Prisoner Statistics	DOJ	1978-2018
Male prison population per 100,000 males	National Prisoner Statistics	DOJ	1978-2018
Female prison population per 100,000 females	National Prisoner Statistics	DOJ	1978-2018
Violent crime rate	UCR	FBI	1960-2018
Property crime rate	UCR	FBI	1960-2018
Percentage of school-aged children in school	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of in-school children in private schools	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of adult (18-64) population with HS degree	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of adult (18-64) population with associate's degree	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of adult (18-64) population with bachelor's degree	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Percentage of adult (18-64) population with graduate degree	IPUMS ACS 1-Year	U.S. Census Bureau	2000-2018
Minimum compulsory school age	State Policy Database	Ruger & Sorens	2000-2017
Maximum compulsory school age	State Policy Database	Ruger & Sorens	2000-2017
CPI multiplier	FRED	Federal Reserve	1945-2018
COLA adjusted CPI	Berry-Fording-Hanson state COL index, BEA RPP	Berry-Fording/Bureau of Economic Analysis	2000-2018

References

Alexander, Monica J, Mathew V Kiang, and Magali Barbieri. 2018. “Trends in Black and White Opioid Mortality in the United States, 1979–2015.” *Epidemiology (Cambridge, Mass.)* 29 (5): 707.

Davidson, Ryan D, Claire S Tomlinson, Connie J Beck, and Anne M Bowen. 2019. “The Revolving Door of Families in the Child Welfare System: Risk and Protective Factors Associated with Families Returning.” *Children and Youth Services Review* 100: 468–79.

English, Diana J, Richard Thompson, and Catherine Roller White. 2015. “Predicting Risk of Entry into Foster Care from Early Childhood Experiences: A Survival Analysis Using LONGSCAN Data.” *Child Abuse & Neglect* 45: 57–67.

Hamilton, James D. 1994. “State-Space Models.” *Handbook of Econometrics* 4: 3039–80.

Harvey, Andrew, and Siem Jan Koopman. 2009. “Unobserved Components Models in Economics and Finance.” *IEEE Control Systems Magazine* 29 (6): 71–81.

Kalman, Rudolph Emil. 1960. “A New Approach to Linear Filtering and Prediction Problems.”

Kiang, Mathew V, Sanjay Basu, Jarvis Chen, and Monica J Alexander. 2019. “Assessment of Changes in the Geographical Distribution of Opioid-Related Mortality Across the United States by Opioid Type, 1999-2016.” *JAMA Network Open* 2 (2): e190040–40.

Langrock, Roland, Ruth King, Jason Matthiopoulos, Len Thomas, Daniel Fortin, and Juan M Morales. 2012. “Flexible and Practical Modeling of Animal Telemetry Data: Hidden Markov Models and Extensions.” *Ecology* 93 (11): 2336–42.

Logan-Greene, Patricia, and Annette Semanchin Jones. 2018. “Predicting Chronic Neglect: Understanding Risk and Protective Factors for CPS-Involved Families.” *Child & Family Social Work* 23 (2): 264–72.

Swann, Christopher A, and Michelle Sheran Sylvester. 2006. “The Foster Care Crisis: What Caused Caseloads to Grow?” *Demography* 43 (2): 309–35.