Flexible Modeling of Transition Processes via Bayesian Spline Rate Models with Application to Estimating and Projecting Modern Contraceptive Use

Herbert Susmann[1] and Leontine Alkema[1]

[1]University of Massachusetts Amherst

Abstract

Many statistical models exist to generate probabilistic estimates and projections of demographic and health indicators. Some of these models, for indicators like the modern contraceptive use rate (mCPR), draw on the observation that these indicators follow a common pattern across countries characterized by a transition between stable states (for mCPR, this is the transition from low adoption of modern family planning to high adoption.) These models work by modeling the rate of change of the indicator as a function of its level. Commonly used approaches posit parametric functional forms for the relationship between rate and level that may not be true in practice. In this work, we relax this assumption by using B-splines with time-varying coefficients to model the rate vs. level relationship. We use this approach to construct estimates and projections of mCPR in Family Planning 2020 initiative countries, and compare its performance to existing models.

1 Introduction

Demographic and health indicators such as the child mortality rate and fertility rate among many others are used to identify disadvantaged populations, target resources, and track progress towards meeting international goals. Determining trends in these indicators is challenging, especially when data are sparse or when projections past the most recent data are needed. As such, statistical models are employed to generate probabilistic estimates and projections for indicators based on the available data.

Some indicators, like the total fertility rate (TFR) or modern contraceptive use rate (mCPR), have been observed to evolve similarly across countries, reflecting a transition between stable states. For the TFR, this is the transition from high to low fertility, called the demographic transition (Kirk, 1996); for the mCPR, this is the transition from low availability and adoption of modern family planning methods to high availability and adoption as modern contraceptive methods are introduced. Characteristic of these transitions is that the rate of change of an indicator is related to its level. For the mCPR case, the rate of change slows as modern contraceptive use increases and demand is saturated (Alkema et al., 2013; Cahill et al., 2018).

Several statistical models, which we refer to generally as *transition models* have been constructed that draw on this structure by modeling rate of change as a function of level. The Bayesian hierarchical model of TFR in countries of Alkema et al. (2011) assumes a double logistic functional form for the relationship between rate of change and level of TFR. The model of mCPR in countries of Cahill et al. (2018) posits that the rate of change in mCPR is related to its level by a logistic growth equation. The assumptions made by these models, that the rate vs. level relationships follow a particular functional form (double logistic or logistic), may not hold true in practice. In addition, commonly used parametrizations of such forms impose long-term memory of the pace of the transition at the population unit of interest. This assumption does not necessarily hold true in practice either. As such, in this work we seek to relax the functional form assumptions through flexible modeling techniques that learn the form of the rate vs. level relationship from the available data.

Our method, which we we refer to as a *Bayesian Spline Rate Model* (BSRM), models the relationship between the rate of change of an indicator using B-splines with time-varying coefficients. Our approach allows for incorporating prior knowledge on the shape of this relationship, such as monotonicity constraints, without having to specify a restrictive functional form. Rather, the specific shape of this relationship is learned from the available data, rather than posited as a modeling assumption. In this paper, we describe the Bayesian Spline Rate Model and use it to construct estimates and projections of mCPR in 68 countries targeted by the Family Planning 2020 (FP2020) initiative. We present preliminary comparisons between the model and a comparison model that makes logistic growth assumptions similar to existing models. In the final paper, we will present the proposed model in greater detail and include a comprehensive set of (model validation) results comparing the two approaches.

2 Methods

We adopt the general notation and model structure of Temporal Models for Multiple Populations (TMMPs), a general class of models for demographic and health indicators (Susmann et al., 2021). Basic to models in this class is the distinction between a *process model* and *data model*. The process model describes how the true, latent value of the indicator (in this case, mCPR) evolves over time, and the data model relates these latent values to the observed data. We start by introducing the novel process model we use in this work.

2.1 Process Model

First, define the mCPR as the proportion of women of reproductive age in a population who use a modern contraceptive method. Let $\eta_{c,t}$ be the latent, true value of the mCPR in country c and year t. We assume the following general form for the process model:

$$\text{logit}(\eta_{c,t}) = g(t, \boldsymbol{\eta}_c, \boldsymbol{\beta}_{c,t}) + \epsilon_{c,t}, \tag{1}$$

where the function g is called the "systematic component" and can depend on t, the other values of the indicator $\boldsymbol{\eta}_c = \{\eta_{c,1}, \dots, \eta_{c,T}\}$ and parameters $\boldsymbol{\beta}_{c,t}$ that can depend on the country and time point, and the random variable $\epsilon_{c,t}$ is referred to as the "stochastic smoothing component", and is intended to capture trends in the data not accounted for by the systematic component.

Systematic Component The systematic component has the following structure:

$$g(t, \boldsymbol{\eta}_c, \boldsymbol{\beta}_{c,t}) = \begin{cases} \Omega_c, & t = t^*, \\ \text{logit}(\eta_{c,t-1}) + f(\eta_{c,t-1}, \boldsymbol{\beta}_{c,t}), & t > t^*, \\ \text{logit}(\eta_{c,t+1}) - f(\eta_{c,t+1}, \boldsymbol{\beta}_{c,t}), & t < t^*. \end{cases} \tag{2}$$

We call the function f the *rate vs. level function*, t^* the *reference year*, and Ω_c the *level in the reference year*. Note that f takes as input the level of the indicator on the original scale, and returns the rate of change on the logit scale (therefore $f : [0, 1] \mapsto \mathbb{R}$.) The parameters $\boldsymbol{\beta}_{c,t}$ allow f to have country-time specific behavior. Figure 1 shows an example of a rate vs. level function f.

We allow the rate vs. level function f to be estimated flexibly using B-splines. We assume that the basic shape of this function is the same between all countries, but allow for country-time specific deviations. Let K be the number of spline knots, and $B_k(\eta)$ be the kth spline basis function of order 2 evaluated at η, with spline knots spaced evenly between 0 and 1. We set $K = 7$ for this analysis. Let $\delta_{c,t,k}$ be the kth spline coefficient for country c and year t. Then define f as:

$$f(\eta, \beta_{c,t}) = \sum_{k=1}^{K} \delta_{c,t,k} B_k(\eta) \tag{3}$$

We then break down $\delta_{c,t,k}$ as: $\delta_{c,t,k} = h_k(\beta_{c,t} + \gamma_k)$, using new parameters $\{\gamma_1, \dots, \gamma_K\}$, $k = 1, \dots, K$ to capture the global relationship between rate and level shared across all countries and time points. The functions h_k, $k = 1, \dots, K$ can be used to set constraints on the shape and scale of the function. Investigation of optimal choices is ongoing. In the current implementation, we constrain the first spline coefficient to be positive and the last coefficient to be negative. For computational reasons, we also constrain each spline coefficient to be bounded. The parameters $\beta_{c,t}$ allow for country-time specific deviation from the global

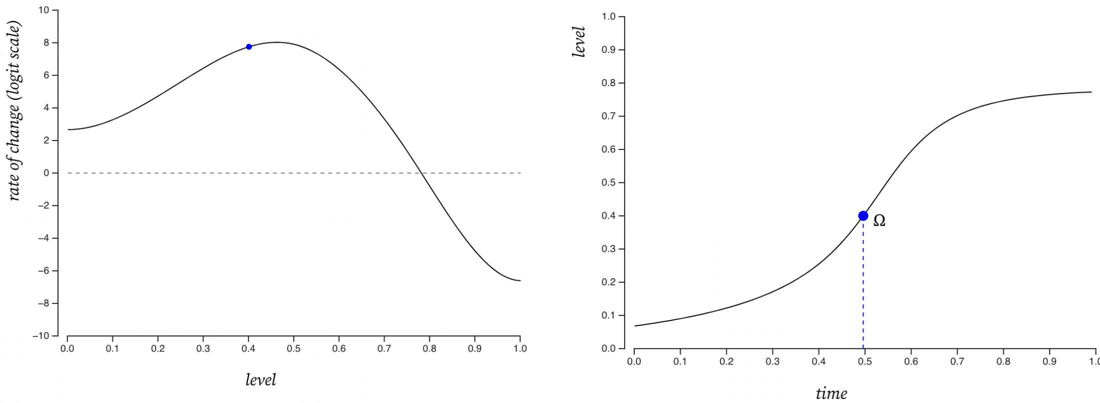

Figure 1: Left: an example of a rate vs. level function f that gives the rate of change of $\eta_{c,t}$ on the logit scale as a function of $\eta_{c,t}$. In our approach, the function is modeled via B-splines that allow the shape of the relationship to be estimated from the data. Right: the evolution of $\eta_{c,t}$ over time that follows from the pictured rate vs. level function.

relationship between rate and level given by γ_k. An AR(1) process is applied to the $\beta_{c,t}$s to regularize their values:

$$\beta_{c,t} | \rho_\beta, \tau_\beta^2 \sim N(\rho_\beta \cdot \beta_{c,t-1}, \tau_\beta^2). \tag{4}$$

We place a uniform prior on ρ_β between zero and one, and a standard half-normal prior on τ_β. The AR process will cause the $\beta_{c,t}$ to be centered around zero unless there are sufficient data from the country to justify deviations from the global relationship given by the γ_k. As such, in data sparse settings or in projections the rate will be informed by the global relationship learned from the trajectories of other countries.

Smoothing Component The smoothing component is designed to capture short-term deviations from the trend defined by the systematic component. The intent is to model stalls in adoption of modern contraceptives or short-term downturns in adoption. We model the $\epsilon_{c,t}$ with ARIMA processes.

2.1.1 Comparison: logistic growth

As a comparison, we consider a model is similar to the one given above, but with a functional form for f that induces a logistic growth curve in the indicator over time. This is similar to the approach used by the existing family planning model (Alkema et al., 2013; Cahill et al., 2018). Let $\boldsymbol{\beta}_{c,t} = \{\omega_c, \tilde{P}_c\}$, where ω_c is a rate parameter and $\tilde{P}_c$ is the country-specific asymptote of the logistic curve. Define f as:

$$f(\eta, \boldsymbol{\beta}_{c,t}) = \begin{cases} \frac{\omega_c(\eta - \tilde{P}_c)}{\tilde{P}_c(\eta-1)}, & \eta < \tilde{P}_c \\ \frac{\omega_c(\eta - \tilde{P}_c)}{\tilde{P}_c(2\tilde{P}_c - \eta - 1)}, & \text{otherwise.} \end{cases} \tag{5}$$

Hierarchical priors are set on the parameters to share information between countries.

2.2 Data and Data Model

We present preliminary results based on data collected by the UN Population Division for the mCPR in 68 countries targeted by the Family Planning 2020 initiative (UN 2017). These data include observations of the mCPR derived by preprocessing data from national and international surveys. As such, each observation also includes an estimate of the associated sampling error, calculated for example by design-based methods.

The main contribution of our approach is the form of the process model, but for completeness we describe the data model used to relate the latent values $\eta_{c,t}$ described by the process model to the observed data. Let y_i, $i = 1, \ldots, N$ be observations of the mCPR, with $c[i]$ and $t[i]$ giving the country and year of each observation. Each observation also has an estimate of the sampling error variance associated with it, which we denote s_i^2. We then define the following truncated-normal data model: $y_i \sim N_{(0,1)}(\eta_{c[i],t[i]}, s_i^2)$.

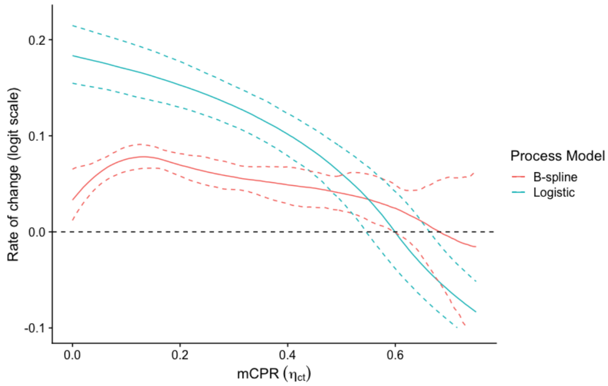

Figure 2: Posterior median and 95% credible intervals of the global rate vs. level functions f for B-spline and logistic models.

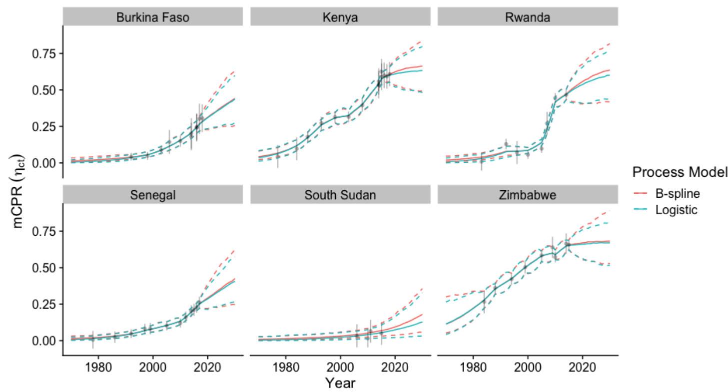

Figure 3: Posterior median and 95% credible intervals for the latent mCPR ($\eta_{c,t}$) in six countries from 1970 to 2030.

3 Preliminary Results

Both the B-spline and logistic models estimate a global relationship between the rate of change of mCPR and its level. For the B-spline model, the spline coefficient parameters γ_k, $k = 1, \ldots, K$ define a global rate vs. level relationship. For the logistic model, the hierarchical means of the logistic rate and asymptote (ω_w and $\tilde{P}_w$, respectively) define a overall logistic curve. We compare the estimated global rate vs. level functions from each model in Figure 2. The flexibility of the B-spline approach can be seen in the estimated function, which has a shape that is not possible to reproduce with the logistic approach. Figure 3 presents the model estimates (posterior medians and 95% credible intervals) for six countries. Both models yield similar estimates, illustrating that it is possible to relax strong assumptions about the rate vs. level function without compromising the quality of the estimates. In the final paper we will present summary measures of the model fits for all countries and a comprehensive set of model validation results comparing the logistic and B-spline modeling approaches, including leave-future-out cross-validation results.

References

L. Alkema, A. E. Raftery, P. Gerland, S. J. Clark, F. Pelletier, T. Buettner, and G. K. Heilig. Probabilistic projections of the total fertility rate for all countries. *Demography*, 48(3):815–839, 2011.

L. Alkema, V. Kantorova, C. Menozzi, and A. Biddlecom. National, regional, and global rates and trends in contraceptive prevalence and unmet need for family planning between 1990 and 2015: a systematic and comprehensive analysis. *The Lancet*, 381(9878):1642–1652, 2013.

N. Cahill, E. Sonneveldt, J. Stover, M. Weinberger, J. Williamson, C. Wei, W. Brown, and L. Alkema. Modern contraceptive use, unmet need, and demand satisfied among women of reproductive age who are married or in a union in the focus countries of the family planning 2020 initiative: a systematic analysis using the family planning estimation tool. *The Lancet*, 391(10123):870–882, 2018.

D. Kirk. Demographic transition theory. *Population studies*, 50(3):361–387, 1996.

H. Susmann, M. Alexander, and L. Alkema. Temporal models for demographic and global health outcomes in multiple populations: Introducing a new framework to review and standardize documentation of model assumptions and facilitate model comparison, 2021.

United Nations, Department of Economic and Social Affairs, Population Division. World contraceptive use by marital status and age 2017, 2017.