Bayesian Hierarchical Temporal Modeling and Targeted Learning with Application to Reproductive Health

Candidate: Herb Susmann
Committee: Leontine Alkema (Chair), Laura Balzer, Antoine Chambaz

June 27, 2022

UMass Amherst

Background

- The international community has set ambitious goals for improvement in global health.
- Where is improvement needed?
  - Chapters 1 and 2: contributions related to statistical estimation and projection of global health indicators, with a focus on family planning.
- What interventions are effective in improving health outcomes?
  - Chapter 3: methods for estimating the effect of interventions on family planning outcomes.

Outline

1. Chapter 1: Temporal models for demographic and global health outcomes in multiple populations
2. Chapter 2: Flexible Modeling of Transition Processes with B-Splines
3. Chapter 3: Automatic Bayesian Targeted Likelihood Estimation of Marginal Structural Models

Outline

1. Chapter 1: Temporal models for demographic and global health outcomes in multiple populations
2. Chapter 2: Flexible Modeling of Transition Processes with B-Splines
3. Chapter 3: Automatic Bayesian Targeted Likelihood Estimation of Marginal Structural Models

Background

- There is interest in modeling demographic and health indicators in order to measure progress towards international goals.
  - Example: Under-5 Mortality Rate
- Data availability and quality are varied.
  - Some countries have high quality U5MR data from vital registration systems, in other countries data may only come from surveys.
- Many statistical models have been created to provide estimates and projections.
- Comparing across models can be difficult.
- **This chapter:** an overarching model class called *Temporal Models for Multiple Populations* (TMMPs).

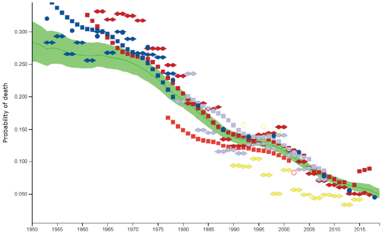
Background

- Published in *International Statistical Review*:
  - Susmann, Herbert, Monica Alexander, and Leontine Alkema. "Temporal Models for Demographic and Global Health Outcomes in Multiple Populations: Introducing a New Framework to Review and Standardise Documentation of Model Assumptions and Facilitate Model Comparison." *International Statistical Review* (2022).

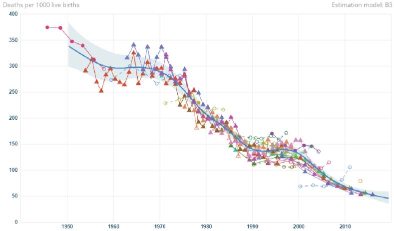
Under-5 Mortality Rate (U5MR) Models

Under-five Mortality Rate Estimates in Senegal, 1950-2019

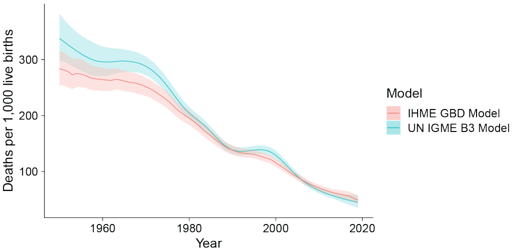
(A) IHME Data and Estimates



(B) UN IGME Data and Estimates



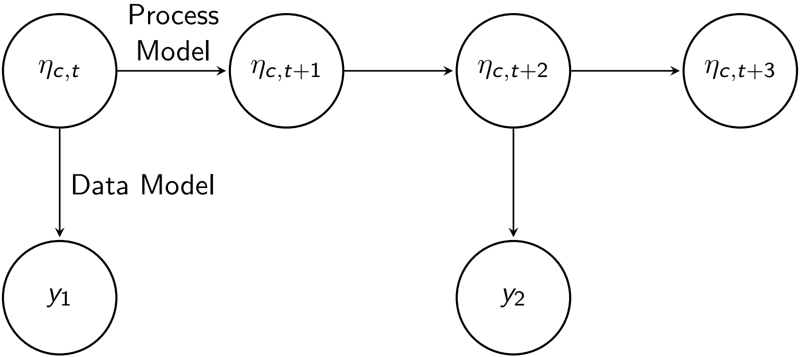
(C) Comparison of Estimates



Modeling Framework

- Let $\eta_{c,t}$ be the true value of the indicator in country c at time t $(c = 1, \ldots, C,\ t = 1, \ldots T)$.
- Observed data y_i, $i = 1, \ldots, n$ with associated properties $c[i]$, $t[i]$, ...
- *Process model* describes evolution of $\eta_{c,t}$.
  - Covariates
  - Systematic trends
- *Data model* describes relationship between y_i and $\eta_{c[i],t[i]}$.

Modeling Framework



Data Model Examples

Examples of data models:

- Normal:

$$y_i|\eta_{c[i],t[i]}, \sigma_i^2 \sim N(\eta_{c[i],t[i]}, \sigma_i^2)$$

  where $y_i \in \mathbb{R}$ and σ_i^2 is the sampling variance.

- Binomial:

$$y_i|\eta_{c[i],t[i]} \sim \mathrm{Binom}(n_i, \eta_{c[i],t[i]})$$

  where y_i, n_i are integers.

Process Model

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \boldsymbol{\beta}_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \boldsymbol{\alpha}_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

Covariate component

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \boldsymbol{\beta}_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \boldsymbol{\alpha}_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

- Regression function for incorporating covariates.
- Example: IHME U5MR

$$g_2(\boldsymbol{X}_{c,t}, \boldsymbol{\beta}_c) = \exp\left[\beta_{c,1} \cdot \log(X_{c,t}^{LDI}) + \beta_{2,c} \cdot X_{c,t}^{EDU} + \beta_{3,c}\right] + \beta_{4,c} x_{c,t}^{HIV}$$

Systematic component

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \beta_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \boldsymbol{\alpha}_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

- Parametric function for modeling systematic temporal trends.
- Example: The Family Planning Estimation Model (FPEM) models the rate of change in Modern Contraceptive Prevalence Rate as following logistic growth (Cahill et al., 2018).

Offset

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \beta_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \alpha_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

- The offset term incorporates external information, for example from a separate modeling step.
- Example: IHME U5MR model uses an offset derived from the smoothed residuals of a separate mixed-effects model.

Smoothing Component

$$g_1(\eta_{c,t}) = \underbrace{g_2(X_{c,t}, \beta_c)}_{\text{covariate}} + \underbrace{g_3(t, \eta_{c,s\neq t}, \alpha_c)}_{\text{systematic}} + \underbrace{a_{c,t}}_{\text{offset}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}}$$

- The smoothing component allows data-driven deviations from the other components, while still enforcing smoothness.
- Many choices B-splines, Gaussian processes, AR(p), RW(p), spatio-temporal smoothing, ...

Hierarchical Modeling

- Each component introduces many country specific parameters that need to be estimated.
- Hierarchical modeling is a way to share information between countries.
- Example: hierarchical model with one level of hierarchy for a country-specific parameter θ_c:

$$\theta_c \mid \theta_w, \sigma_\theta \sim N(\theta_w, \sigma_\theta^2)$$

Contributions

- A model class, Temporal Models for Multiple Populations (TMMPs), that encompasses many existing demographic and health models.
  - Model class makes a clear distinction between the *process model* and the *data model*.
  - Process model is split into building blocks: covariates, systematic trends, offsets, and smoothing components.
- Detailed description of six existing models using TMMP notation, and templates provided for documenting additional models as TMMPs.

Outline

1. Chapter 1: Temporal models for demographic and global health outcomes in multiple populations
2. Chapter 2: Flexible Modeling of Transition Processes with B-Splines
3. Chapter 3: Automatic Bayesian Targeted Likelihood Estimation of Marginal Structural Models

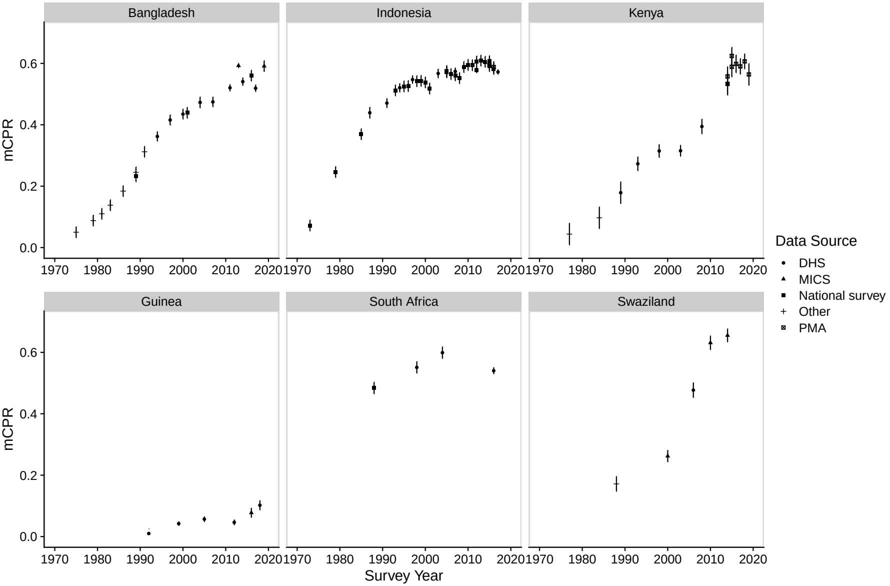
Background

- Some indicators have been observed to evolve similarly across populations.
  - They tend to follow a *transition* between stable states.
- Classic example: demographic transition.
  - Transition from high total fertility rate and high under-5 mortality to low fertility, low mortality.
- Existing statistical models for estimating and projecting trends in these indicators draw on these patterns.
- **This chapter:** We propose a new type of model, called *B-Spline Transition Models*, for flexibly estimating indicators that follow transitions.

Case Study

- Modern Contraceptive Prevalence Rate (mCPR) for married or in-union women: proportion of married or in-union women of reproductive age using (or with partner using) a modern contraceptive method.
- Transition: low to high mCPR.
- Existing model: Family Planning Estimation Model (FPEM, Cahill et al. 2018).
- Goal: estimate and project mCPR in countries from 1970-2030.
- Dataset aggregated by United Nations Population Division (UNPD) from surveys conducted by governments or international organizations.

Case Study



Bangladesh
Indonesia
Kenya
Guinea
South Africa
Swaziland
mCPR
0.0
0.2
0.4
0.6
1970
1980
1990
2000
2010
2020
Survey Year
Data Source
DHS
MICS
National survey
Other
PMA

Transition Models

- First, we define a model class for indicators that follow a transition.
- Core assumption: the rate of change of the indicator is a function of its level.
- *Transition Models* are have a process model given by

$$g_1(\eta_{c,t}) = \underbrace{g_3(t, \boldsymbol{\eta}_{c,s\neq t}, \boldsymbol{\alpha}_c)}_{\text{systematic}} + \underbrace{\epsilon_{c,t}}_{\text{smoothing}} \quad .$$

- The systematic component has the following form:

$$g_3(t, \boldsymbol{\eta}_{c,s\neq t}, \boldsymbol{\alpha}_c) = \begin{cases} \Omega_c, & t = t_c^*, \\ g_1(\eta_{c,t-1}) + f(\eta_{c,t-1}, \tilde{P}_c, \beta_c), & t > t_c^*, \\ g_1(\eta_{c,t+1}) - f(\eta_{c,t+1}, \tilde{P}_c, \beta_c), & t < t_c^*, \end{cases}$$

where $\boldsymbol{\alpha}_c = \{\Omega_c, \tilde{P}_c, \beta_c\}$.

- The function f is called the *transition function*.

FPEM Example

- FPEM is an example of a Transition Model.
- Because $\eta_{c,t} \in (0,1)$, FPEM process model uses a logit transform:

$$\text{logit}(\eta_{c,t}) = g_3(t, \boldsymbol{\eta}_{c,s\neq t}, \boldsymbol{\alpha}_c) + \epsilon_{c,t}.$$

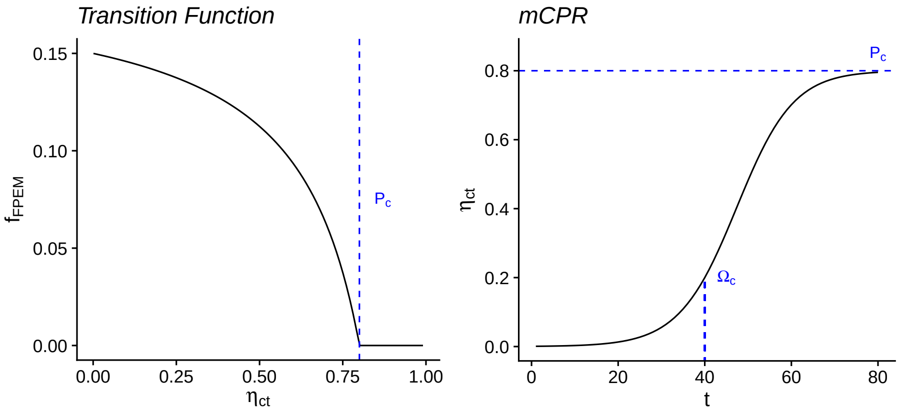
- The FPEM transition function was chosen such that mCPR follows a logistic growth curve:

$$f(\eta_{c,t-1}, \tilde{P}_c, \omega_c) = \begin{cases} \frac{(\eta_{c,t-1} - \tilde{P}_c)\omega_c}{\tilde{P}_c(\eta_{c,t-1} - 1)}, & \eta_{c,t-1} < \tilde{P}_c, \\ 0, & \text{otherwise.} \end{cases}$$

with

- ω_c: rate parameter,
- $\tilde{P}_c$: asymptote parameter.

FPEM Example



B-Spline Transition Model

- Core idea: estimate the transition function f while making weaker functional form assumptions.
- Approach: estimate f using B-splines.
- Define a transition function f_b as:

$$f_b(\eta_{c,t}, \tilde{P}_c, \beta_c) = \sum_{j=1}^{J} \underbrace{h_j(\beta_{c,j})}_{\text{coefficient}} \cdot \underbrace{B_j(\eta_{c,t}/\tilde{P}_c)}_{\text{basis function}},$$

  where $\tilde{P}_c$ is an asymptote parameter.
- Flexibility of f_b can be tuned through the spline degree and number and positioning of knots.

B-Spline Transition Model for mCPR

- Next, we tailor the B-Spline Transition Model for use in estimating mCPR in countries.
- Process model:

$$\text{logit}(\eta_{c,t}) = g_3(t, \boldsymbol{\eta}_{c,s\neq t}, \boldsymbol{\alpha}_c) + \epsilon_{c,t}.$$

- Systematic component:

$$g_3(t, \boldsymbol{\eta}_{c,s\neq t}, \boldsymbol{\alpha}_c) = \begin{cases} \Omega_c, & t = t_c^*, \\ \text{logit}(\eta_{c,t-1}) + f_b(\eta_{c,t-1}, \tilde{P}_c, \boldsymbol{\beta}_c), & t > t_c^*, \\ \text{logit}(\eta_{c,t+1}) - f_b(\eta_{c,t+1}, \tilde{P}_c, \boldsymbol{\beta}_c), & t < t_c^*. \end{cases}$$

- Smoothing model: autoregressive (AR(1)) process.

B-Spline Transition Model for mCPR

- Process model:

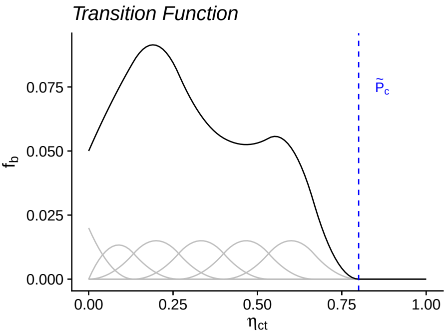
$$\mathrm{logit}(\eta_{c,t}) = g_3(t, \boldsymbol{\eta}_{c,s\neq t}, \boldsymbol{\alpha}_c) + \epsilon_{c,t}.$$

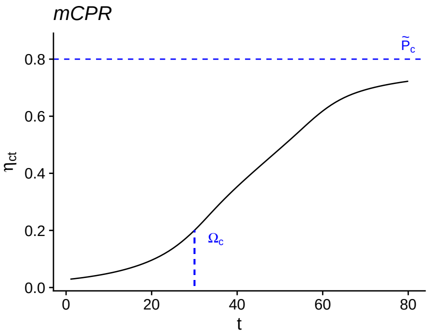
- B-spline transition function:

$$f_b(\eta_{c,t}, \tilde{P}_c, \boldsymbol{\beta}_c) = \sum_{j=1}^{J} h_k(\beta_{c,j}) B_j(\eta_{c,t}/\tilde{P}_c).$$

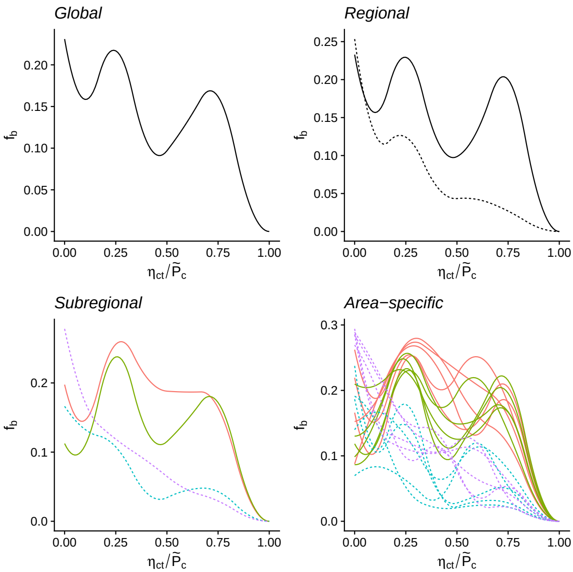
- B-Spline basis functions of order d. K knots incuding $J = K + d - 1$ basis functions.
- Constraints set up such that mCPR follows an S-shaped transition.
- AR(1) smoothing model.

Example B-Spline Transition Function





Sharing information on shape of transition function



Model Validation

- Validation exercise 1: hold out 20% of all observations at random.

			95% UI			Error	
	n		% Below	% Included	% Above	ME	MAE
All	126	B-spline	1.6%	95.2%	3.2%	0.0053	0.032
		Logistic	1.6%	94.4%	4.0%	0.0055	0.030
High	36	B-Spline	0%	94.4%	5.6%	0.027	0.039
		Logistic	0%	94.6%	5.4%	0.024	0.041
Middle	48	B-Spline	0%	100%	0%	-0.0044	0.034
		Logistic	0%	97.8%	2.2%	-0.0027	0.033
Low	42	B-Spline	4.8%	90.5%	4.8%	0.00022	0.027
		Logistic	4.7%	90.7%	4.7%	0.0011	0.027

95% UI: 95% uncertainty interval. ME: median error. MAE: median absolute error.

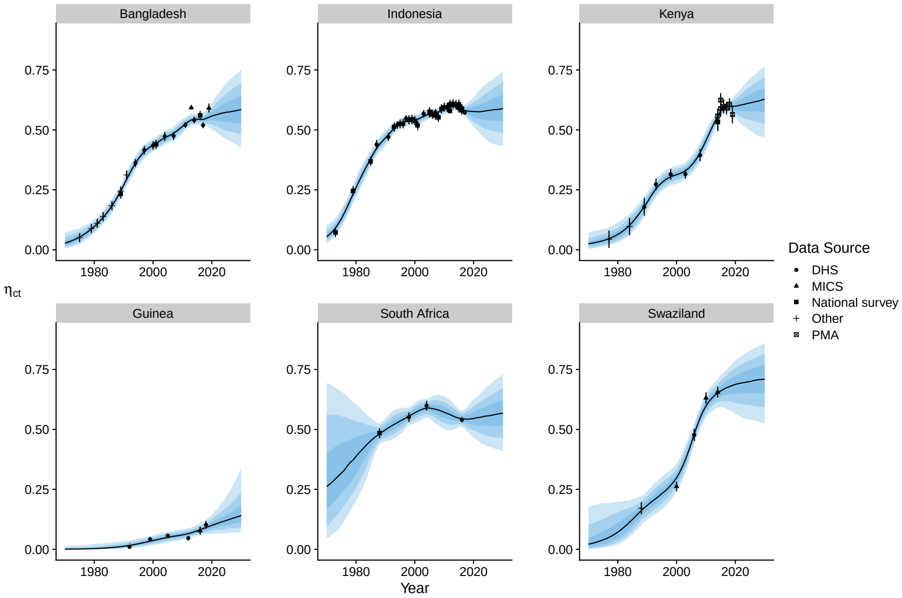
Model Validation (11/17)

- Validation exercises 2: hold out all observations after a cutoff year $L = 2010$.

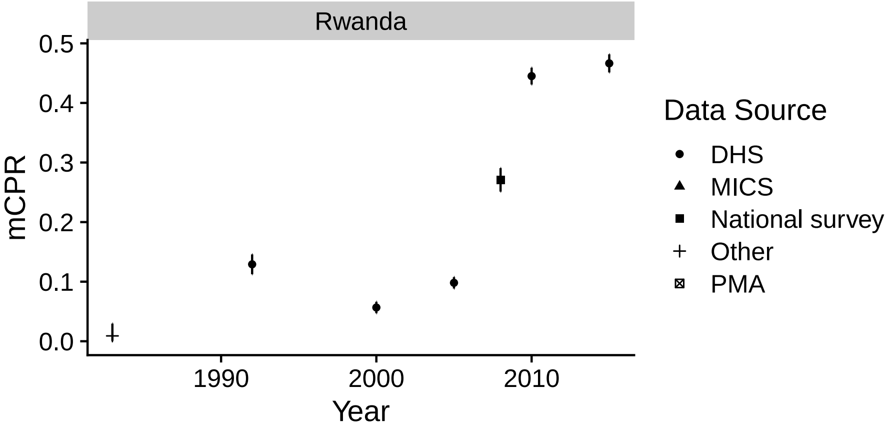
			95% UI			Error	
	n		% Below	% Included	% Above	ME	MAE
Model Check 2 ($L = 2010$)							
All	143	B-spline	6.99%	90.9%	2.1%	-0.0480	0.0646
		Logistic	5.59%	93%	1.4%	-0.0477	0.0586
High	34	B-Spline	2.94%	94.1%	2.94%	0.0107	0.0304
		Logistic	0%	97%	3.03%	0.0115	0.0350
Middle	46	B-Spline	2.17%	93.5%	4.35%	-0.0524	0.0723
		Logistic	2.13%	95.7%	2.13%	-0.0575	0.0631
Low	63	B-Spline	12.7%	87.3%	0%	-0.0812	0.0812
		Logistic	11.1%	88.9%	0%	-0.0658	0.0729

95% UI: 95% uncertainty interval. ME: median error. MAE: median absolute error.
Measures calculated using the last held-out observation within each area.

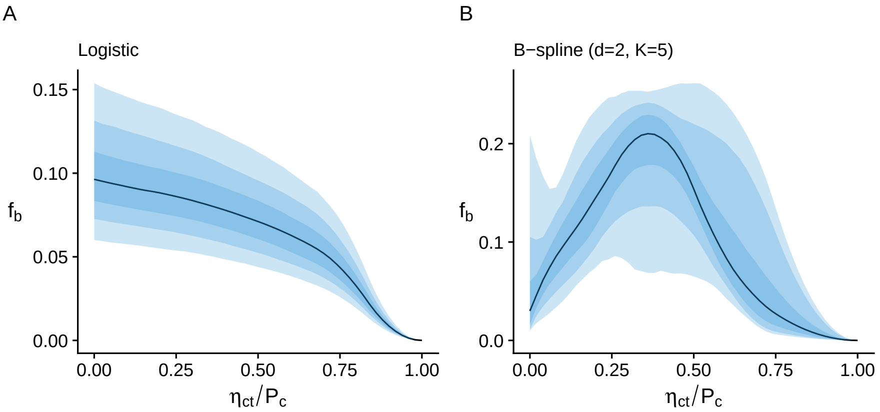
Illustrative Fits from B-spline Model



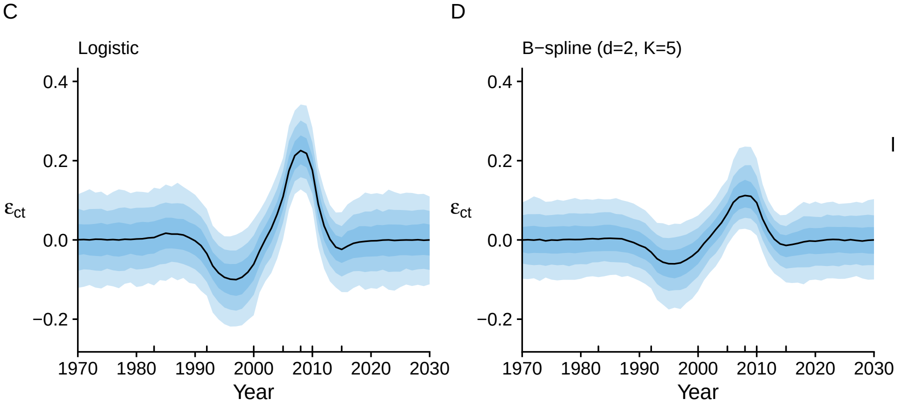
Detailed Example: Rwanda



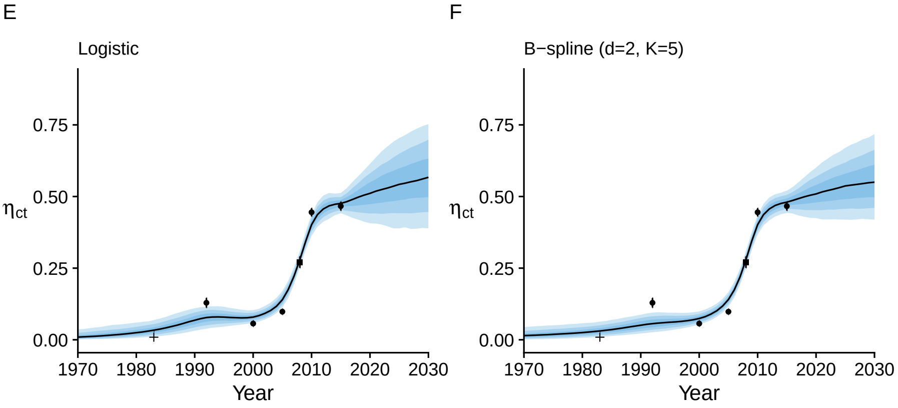
Transition Functions



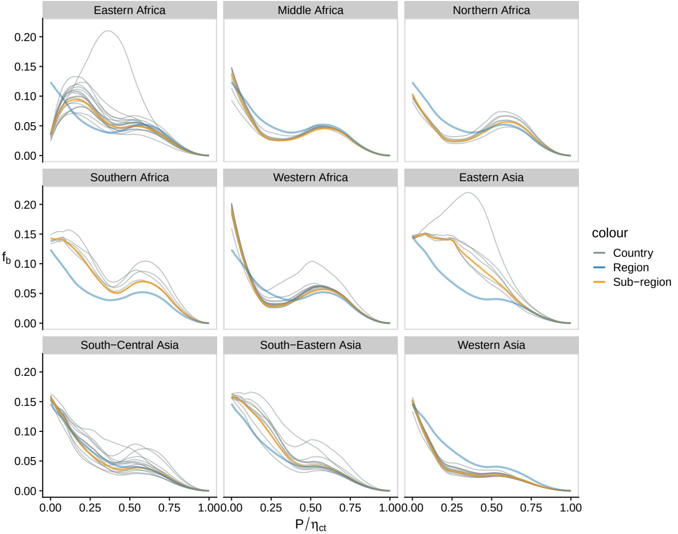
Smoothing component



Modern Contraceptive Prevalence Rate



Trends can be seen in regional and subregional transition functions



Contributions

- Subclass of *Transition Models* for indicators that follow transitions.
- B-spline Transition Model: flexible modelling approach based on B-splines.
- Generated estimations and projections of mCPR in countries from 1970-2030.
- Model implementation in `R` and `Stan`.

Outline

1. Chapter 1: Temporal models for demographic and global health outcomes in multiple populations
2. Chapter 2: Flexible Modeling of Transition Processes with B-Splines
3. Chapter 3: Automatic Bayesian Targeted Likelihood Estimation of Marginal Structural Models

Background

- What interventions are effective in improving health outcomes?
- *Marginal Structural Models* are a way to summarize how the effect of an intervention on an outcome changes within subgroups.
- **This Chapter**: We introduce a novel targeted Bayesian estimator for the parameter of a Marginal Structural Model in a general setting.

Motivating Example

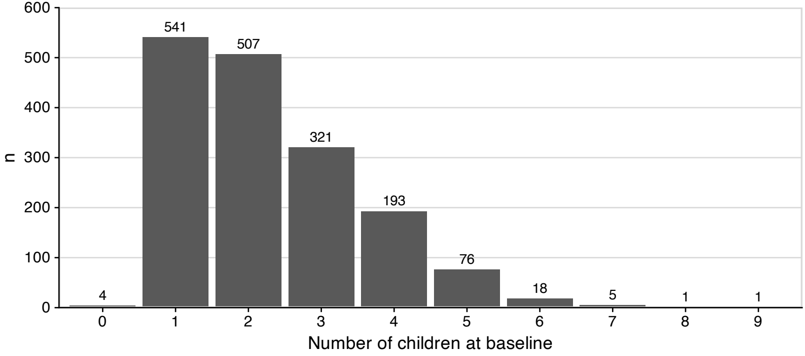
- Randomized field experiment conducted in Lilongwe, Malawi to investigate effect of broad-based family planning intervention on contraceptive use. (Karra et al., 2020, 2022)
- Intervention:
- Outcome:

Motivating Example

- For each participant, we have:
  - X: set of 11 covariates measured at baseline, including number of children X_c;
  - A: indicator of randomization into intervention group;
  - Y: indicator of contraceptive use at endline.
- Let $O_1, \ldots, O_n$ be n i.i.d. draws of the generic variable $O = (X, A, Y)$ from a law P_0.

Motivating Example

- Scientific question: does the treatment effect differ depending on number of children at baseline?



Conditional Average Treatment Effect

- *Conditional Average Treatment Effect* (CATE):

$$\begin{aligned}\Psi_P^{\mathrm{CATE}}(x) &= \bar{Q}_P^{(1)}(x) - \bar{Q}_P^{(0)}(x) \\ &= \mathbb{E}_P[Y \mid A = 1, X = x] - \mathbb{E}_P[Y \mid A = 0, X = x].\end{aligned}$$

- Causally identifiable under "standard causal assumptions" (consistency, positivity, no unmeasured confounders).

An example of a Marginal Structural Model

- Approach: summarize the relationship between *treatment effect modifiers* and conditional treatment effects using a user-supplied working model.
- Treatment effect modifier: number of children at baseline X_c.
- For instance, let $B(P) \in \mathbb{R}^2$ be the solution to the following optimisation problem:

$$B(P) = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \mathbb{E}_P \left[\left(\Psi_P^{\mathrm{CATE}}(X) - (1, X_c)\boldsymbol{\beta} \right)^2 \right].$$

An example of a Marginal Structural Model

$$B(P) = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \mathbb{E}_P \left[\left(\underbrace{\boxed{\Psi_P^{\mathrm{CATE}}(X)}}_{\text{conditional average treatment effect}} - (1, X_c)\boldsymbol{\beta} \right)^2 \right]$$

An example of a Marginal Structural Model

$$B(P) = \underset{\boldsymbol{\beta} \in \mathbb{R}^2}{\arg\min}\, \mathbb{E}_P \left[\left(\Psi_P^{\mathrm{CATE}}(X) - \underbrace{\boxed{(1, X_c)\boldsymbol{\beta}}}_{\text{linear working model}} \right)^2 \right]$$

An example of a Marginal Structural Model

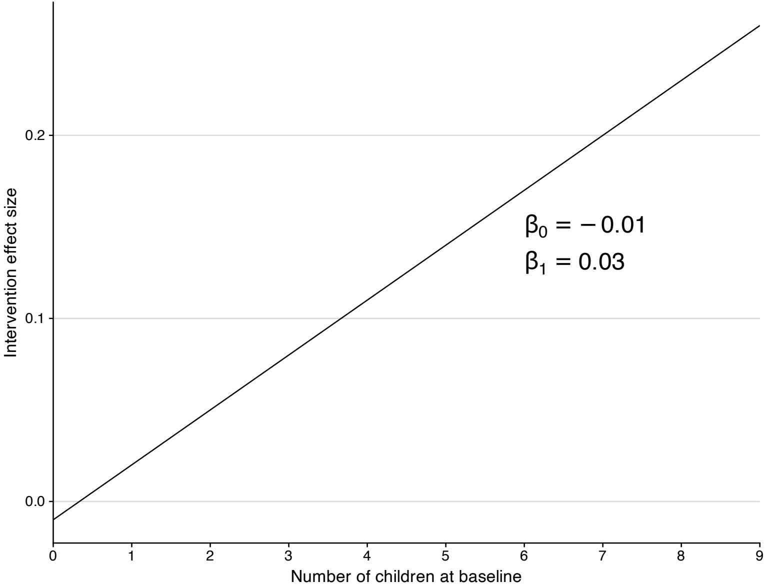
$$B(P) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^2} \underbrace{\boxed{\mathbb{E}_P\left[\left(\Psi_P^{\mathrm{CATE}}(X) - (1, X_c)\beta\right)^2\right]}}_{\text{squared-error risk}}$$

An example of a Marginal Structural Model

$$\boxed{B(P)} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \mathbb{E}_P \left[\left(\Psi_P^{\mathrm{CATE}}(X) - (1, X_c)\beta \right)^2 \right]$$

defined in terms of P

What a plot of the results will look like



General Setting for MSMs

- Observed data: $O_1, \ldots, O_n$ i.i.d. copies of a generic variable $O \sim P_0$.
- Assume that P_0 is in a non-parametric statistical model $\mathcal{M}$.
  - The more we know about the law P_0, the smaller the model $\mathcal{M}$.
- Assume that $O = (Z, X)$ for variables $Z \in \mathcal{Z}$, $X \in \mathcal{X}$.
- For all $P \in \mathcal{M}$, let $\Psi_P : \mathcal{X} \to \mathbb{R}$ be a functional summary of P with argument X.
- For the motivating example:
  - $O = (Y, A, X)$, $Z = (Y, A)$, $X = X$.
  - $\Psi_P(X) = \mathbb{E}_P[Y \mid A = 1, X] - \mathbb{E}_P[Y \mid A = 0, X]$, the CATE.

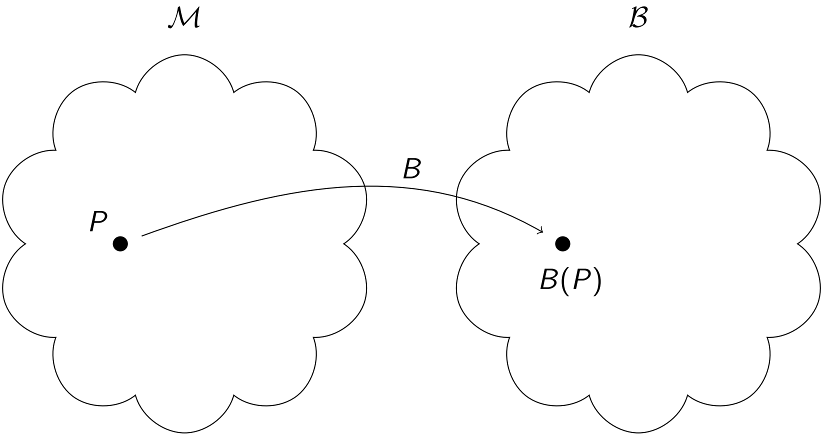
Marginal Structural Models

- Idea: approximate Ψ_P using a user-supplied working model.
- Working model: set $\{m_{\boldsymbol{\beta}} : \boldsymbol{\beta} \in \mathcal{B}\}$ of functions $m_{\boldsymbol{\beta}} : \mathcal{X} \to \mathbb{R}$ with $\mathcal{B}$ a parameter of dimension p.
- Loss function: $L_m(\Psi_P(X), \boldsymbol{\beta})(X)$.
- Define the parameter of interest B as the solution to the optimization problem:

$$B(P) = \underset{\boldsymbol{\beta} \in \mathcal{B}}{\arg\min}\, \mathbb{E}_P \left[L_m(\Psi_P(X), \boldsymbol{\beta})(X) \right].$$

- The combination of working model and loss function is called a *Marginal Structural Model* (Robins et al., 2000; Van der Laan and Rose, 2011).
- Causally identifiable under same assumptions as Ψ_P.

Marginal Structural Models



Danger! Infinite dimensional space visualized in two dimensions!

Semi-parametric inference

- Goal: estimate $\beta_0 := B(P_0)$.
- What is the semi-parametric efficiency bound for estimating β_0?
- We can write any regular estimator of β_0 as:

$$\hat{\beta}_n = \beta_0 + \frac{1}{n}\sum_{i=1}^{n} IC_{P_0}(O_i) + o_p(n^{-1/2}),$$

  where IC_{P_0} is called an *influence function* of the parameter β_0.
- The influence function with the smallest variance is called the *efficient influence function* (EIF), which we denote D^*.
- The semi-parametric efficiency bound for estimating β_0 is given by $\mathrm{var}_{P_0}(D^*(P_0)(O))$

Efficient Influence Function of $B(P)$

Theorem (Efficient Influence Function)

(Simplified) The target functional $P \mapsto B(P)$ is pathwise differentiable at every $P \in \mathcal{M}$, with an efficient influence function $D^*(P)$ given by

$$D^*(P)(O) = M^{-1}\left[D_1^*(P)(O) + D_2^*(P)(X)\right],$$

where $D_1^*(P), D_2^*(P) \in L_0^2(P)$ are given by

$$\begin{aligned} D_1^*(P)(O) &= \nabla \dot{L}(\Psi_P(X), B(P))(X) \times \Delta^*(P)(O), \\ D_2^*(P)(X) &= \dot{L}(\Psi_P(X), B(P))(X), \end{aligned}$$

and the normalizing matrix M is given by

$$M = -\mathbb{E}_P\left[\ddot{L}_m(\Psi_P(X), B(P))(X)\right].$$

Efficient Influence Function of $B(P)$

Theorem

(Simplified) The target functional $P \mapsto B(P)$ is pathwise differentiable at every $P \in \mathcal{M}$, with an efficient influence function $D^*(P)$ given by

$$D^*(P)(X, A, Y) = M^{-1} \left\{ D_1^*(P)(X, A, Y) + D_2^*(P)(X) \right\},$$

where

$$D_1^*(P)(X, A, Y) = \left\{ \frac{\mathbb{I}(A = 1)}{P[A = 1|X])} - \frac{\mathbb{I}(A = 0)}{P[A = 0|X]} \right\} (Y - \bar{Q}_P^{(A)}(X))(1, X)^\top,$$

$$D_2^*(P)(X) = (\Psi_P(X) - B(P)^\top (1, X)^\top)(1, X)^\top,$$

and the normalizing matrix M is given by

$$M = -\mathbb{E}_P \left[(1, X)^\top (1, X) \right].$$

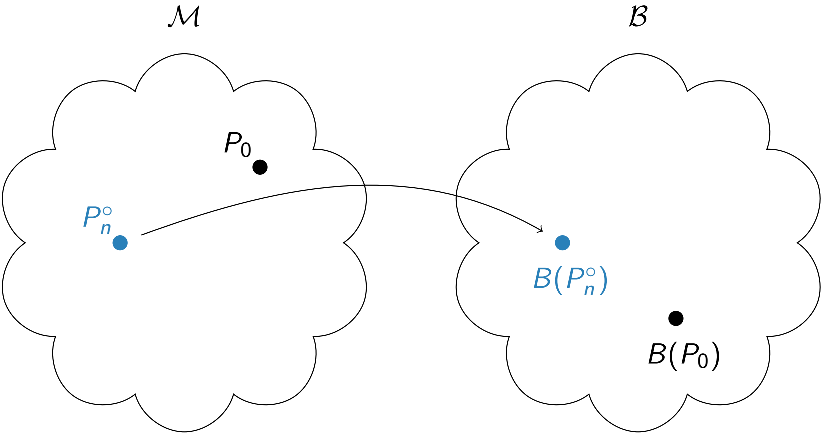
Targeted Minimum Loss-Based Estimation

- It turns out we can construct an estimator that achieves this efficiency bound!
  - *Targeted Minimum Loss-Based Estimation (TMLE)* (Van der Laan and Rose, 2011; van der Laan and Rose, 2018)
- Suppose we have an initial estimator P_n^0 of the pieces of P_0 relevant to β.
- We can then form a plugin estimator

$$\hat{\beta}^{\text{plugin}} = B(P_n^0). \tag{1}$$

- The plugin estimator will be biased!

Plug-in estimator may be biased



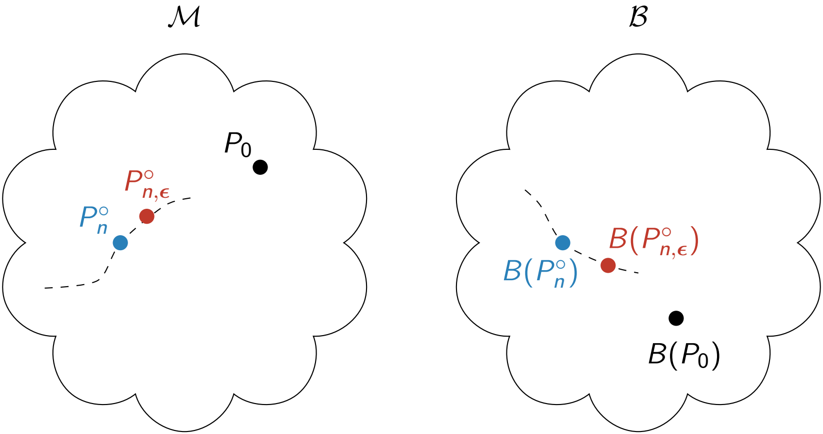
Targeted Minimum Loss-Based Estimation

- Targeted Minimum Loss-Based Estimation (TMLE): plug-in estimator of the form

$$\hat{\boldsymbol{\beta}}^{TMLE} = B(P_n^\circ(\boldsymbol{\epsilon}_n^*))$$

where $\{P_n^\circ(\boldsymbol{\epsilon}) : \boldsymbol{\epsilon} \in \mathbb{R}^p\}$ is a *fluctuation* of an initial estimator P_n° of the pieces of P_0 relevant to $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}_n^*$ is chosen by minimising the empirical risk of a loss function $\mathcal{L}$.

TMLE: update initial estimate in direction of truth



A glimpse at how TMLE works

- The fluctuation and loss function are chosen to satisfy (among other things) a key property:

$$D^*(P_n^\circ) \in \mathrm{Span}\left(\frac{\partial}{\partial \epsilon}\mathcal{L}\left(P_n^\circ(\epsilon)\right)\Big|_{\epsilon=0}\right).$$

- Importantly, the TMLE solves the EIF of the target parameter:

$$\mathbb{E}_{P_n}[D^*(P_n^\circ(\epsilon_n^*))(O)] = 0.$$

- Under certain conditions, $\hat{\beta}^{TMLE}$ is asymptotically normal and efficient.

Blueprint for fluctuation model

How do we choose the form of the fluctuation model $P_n^\circ(\epsilon)$? We propose a blueprint:

TMLE Blueprint. The following choice of loss functions and fluctuation model satisfy the conditions (L1), (L2), and (M1).

- For any $P \in \mathcal{M}$ with corresponding ψ_P, $\bar{Q}_P = \{\bar{Q}_P^{(1)}, \ldots, \bar{Q}_P^{(J)}\}$, Q_P, and η_P, define the parametric fluctuation model as

$$\bar{Q}_{P,\epsilon}^{(1)}(O) = \bar{Q}_P^{(1)}(O) + H_1(O)\boldsymbol{\epsilon}^\top \nabla \dot{L}(\psi_P(X), B(P))(X),$$
$$\vdots$$
$$\bar{Q}_{P,\epsilon}^{(J)}(O) = \bar{Q}_P^{(J)}(O) + H_J(O)\boldsymbol{\epsilon}^\top \nabla \dot{L}(\psi_P(X), B(P))(X),$$
$$Q_{P,\epsilon}(X) = C \exp\left(\boldsymbol{\epsilon}^\top \dot{L}(\psi_P(X), B(P))(X)\right) Q_P(X).$$

- Choose $\mathcal{L}_j$ and H_j for $j = 1, \ldots, J$ such that

$$\sum_{j=1}^{J} \dot{\mathcal{L}}_j(\bar{Q}_P^{(j)}(O), O) H_j(O) = \Delta^*(P)(O).$$

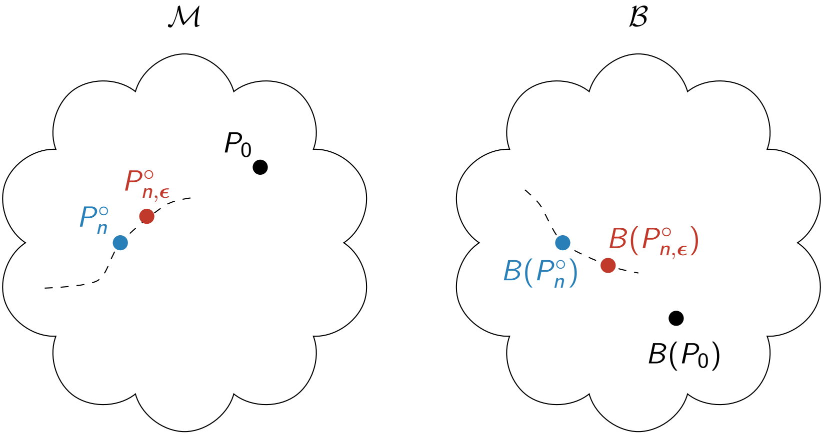
Bayesian TMLE

- Can we make this procedure Bayesian?
- Core idea: some choices of TMLE loss function can be interpreted as defining a likelihood for the data O conditional on the parameter ϵ under the fluctuation submodel.
- We can then use Bayesian inference to estimate ϵ! (Diaz et al., 2011; Díaz et al., 2020)
- Basic application of Bayes rule: posterior distribution of ϵ is given by

$$\Pi_\epsilon(\epsilon \mid O_1, \ldots, O_n) \propto \pi_\epsilon(\epsilon) \prod_{i=1}^{n} p_n^\circ(O_i \mid \epsilon)$$

  where π_ϵ is a prior distribution for ϵ and $p_n^\circ(O \mid \epsilon)$ is the likelihood of O under $P_n^\circ(\epsilon)$.
- Once we have a posterior distribution for ϵ we can map it to a posterior distribution for $\boldsymbol{\beta}$.

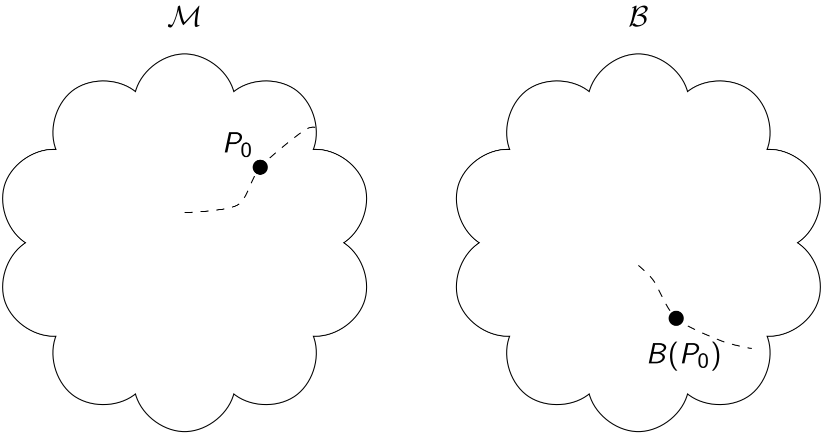
Bayesian TMLE



Bernstein von-Mises

- Desired result: the posterior distribution for β converges to a normal distribution centered on the frequentist TMLE with variance given by the variance of the efficient influence function.
- We prove an *oracular* version that provides conditions under which the posterior distribution based on fluctuation of P_0 will converge to the optimal distribution.

Bayesian TMLE



Bernstein von-Mises

- Let $p_n^0(O \mid \epsilon)$ be the likelihood of the submodel fluctuating P_0.
- Key conditions:
  - The gradient satisfies

$$\frac{\partial}{\partial \boldsymbol{\epsilon}} \log p_n^0(O|\boldsymbol{\epsilon})\Big|_{\boldsymbol{\epsilon}=0} = D^*(P_0)(O).$$

  - The Hessian satisfies

$$P_0\left[\frac{\partial^2}{\partial \boldsymbol{\epsilon}^2} \log p_n^0(O|\boldsymbol{\epsilon})\Big|_{\boldsymbol{\epsilon}=0}\right] = P_0[D^*(P_0)D^*(P_0)^\top].$$

Theorem (Oracular Bernstein von-Mises)

(Simplified) Let $N(\mu, \Sigma)$ denote the multivariate normal distribution with mean vector μ and covariance matrix Σ. Then, under certain assumptions,

$$\|\Pi_{\boldsymbol{\beta}}^{0}(\cdot \mid O_1, \ldots, O_n) - N\left(\Delta_n^0, P_0[D^*(P_0)D^*(P_0)^\top]\right)\|_1 = o_P(1)$$

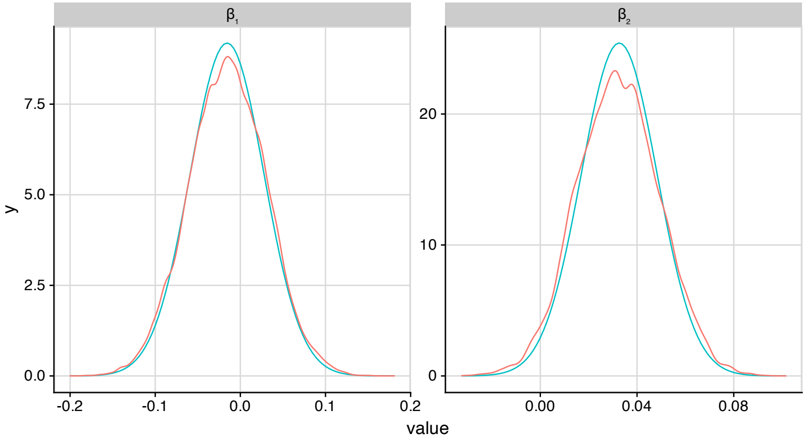
where

$$\Delta_n^0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} P_0[\lambda^*(P_0)]^{-1} D^*(P_0)(O_i). \tag{2}$$

Universal Algorithm

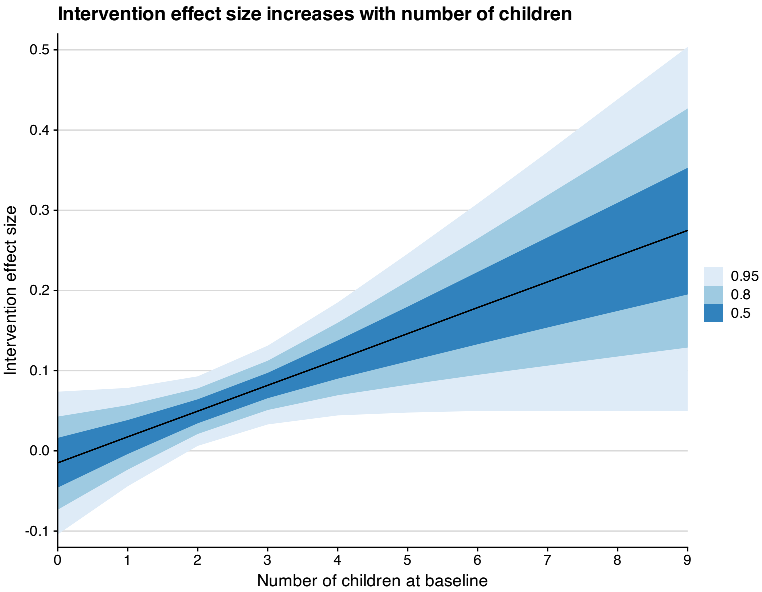
- In practice, users may want to try several working models and loss functions.
- We could choose several working models and loss functions and hand-code the required derivatives. But what if a user wants to use something we haven't implemented?
- An alternative is to use *automatic differentiation* to compute the required derivatives automatically.
- We implemented a universal algorithm in Julia that uses auto-differentiation to automatically adapt the fluctuation model and efficient influence function to arbitrary well-chosen working models and loss functions.

Motivating Example: Results



Posterior density (red) and a normal density (blue) centered on the frequentist MLE with variance given by the estimated variance of the efficient influence function.

Motivating Example: Results



Contributions

- Definition of MSMs in a general setting.
- Derivation of efficient influence function for general MSM parameters.
- Framework for TMLE for general MSMs.
- Novel Bayesian TMLE for MSMs.
- Universal algorithm implemented in Julia using autodifferentiation.
- Application to estimate relationship between effect of intervention on contraceptive use with number of children as an effect modifier in a randomized field experiment.

Summary

Where is improvement needed?

- Chapter 1: Temporal Models for Multiple Populations
- Chapter 2: B-Spline Transition Model

What interventions are effective in improving health outcomes?

- Chapter 3: Bayesian targeted learning for Marginal Structural Models

Acknowledgements

References I

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, Mar. 2018. ISSN 0140-6736. doi: 10.1016/S0140-6736(17)33104-5. URL http://www.sciencedirect.com/science/article/pii/S0140673617331045.

I. Diaz, A. E. Hubbard, and M. J. van der Laan. Targeted bayesian learning. In *Targeted Learning*, pages 475–493. Springer, 2011.

I. Díaz, O. Savenkov, and H. Kamel. Nonparametric targeted bayesian estimation of class proportions in unlabeled data. *Biostatistics*, 2020.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis (3rd ed)*. Chapman and Hall/CRC, 2013.

A. Gelman, A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P.-C. Bürkner, and M. Modrák. Bayesian workflow. *arXiv preprint arXiv:2011.01808*, 2020.

M. Karra, D. Canning, et al. The effect of improved access to family planning on postpartum women: protocol for a randomized controlled trial. *JMIR research protocols*, 9(8):e16697, 2020.

References II

M. Karra, D. Maggio, M. Guo, B. Ngwira, and D. Canning. The causal effect of a family planning intervention on women’s contraceptive use and birth spacing. *Proceedings of the National Academy of Sciences*, 119(22):e2200279119, 2022. doi: 10.1073/pnas.2200279119. URL https://www.pnas.org/doi/abs/10.1073/pnas.2200279119.

J. M. Robins, M. A. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology, 2000.

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.

M. J. Van der Laan and S. Rose. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media, 2011.

M. J. van der Laan and S. Rose. *Targeted learning in data science: causal inference for complex longitudinal studies*. Springer, 2018.