

Flexibly Modeling Shocks to Demographic and Health Indicators with Bayesian Shrinkage Priors

Herb Susmann¹ Leontine Alkema²

¹NYU Grossman School of Medicine

²University of Massachusetts Amherst



UMassAmherst

PAA 2025

Washington, D.C., April 12, 2025

This work was supported by the Gates Foundation.

Slides:

herbsusmann.com/paa2025

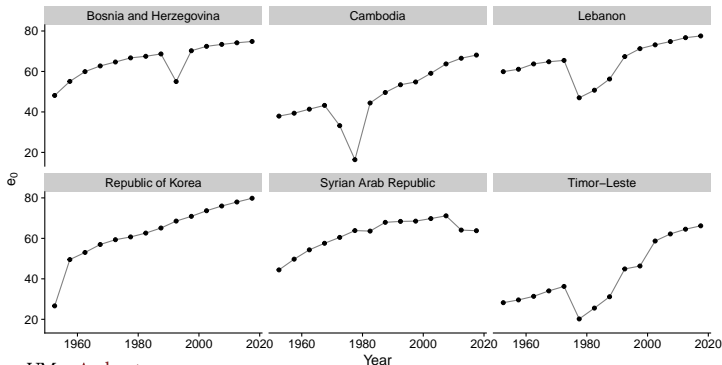
Preprint:

<https://arxiv.org/abs/2410.09217>



Motivation

- ▶ Goal: building statistical models for estimating and projecting demographic and health indicators.
- ▶ Many statistical models assume *smoothness* of the data.
- ▶ Statistical models that assume smoothness typically will not perform well when fit to data that exhibit shocks.
- ▶ **This talk:** we propose using Bayesian shrinkage priors as a practical way to build statistical models robust to shocks.



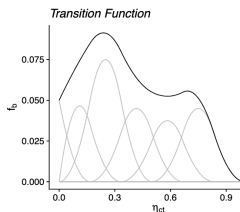
Smooth Transition Model

- ▶ Let $\eta_{c,t}$ be the true male period life expectancy at birth in country c and time t .
- ▶ Model change in $\eta_{c,t}$ as:

$$\eta_{c,t} = \eta_{c,t-1} + f(\eta_{c,t-1}, \beta_c) + \epsilon_{c,t},$$

where

- ▶ f : expected change with parameters β_c
 - ▶ $\epsilon_{c,t}$: deviations from expected change
- ▶ We model f using B-splines (Susmann & Alkema 2025 JRSS-C).
 - ▶ Takeaway: f is flexible, but assumes change in life expectancy is smooth.
 - ▶ Deviations typically modelled as ARIMA process; following Raftery et al. we use white noise for e_0 :



$$\epsilon_{c,t} | \tau_{\epsilon} \sim N(0, \tau_{\epsilon}^2).$$

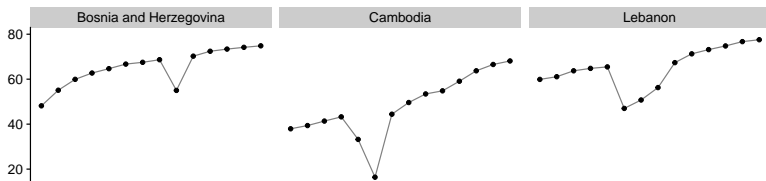
Transition Model with Shocks

- Proposal: add an additional term to the transition model to handle shocks.

$$\eta_{c,t} = \eta_{c,t-1} + f(\eta_{c,t-1}, \beta_c) - \delta_{c,t} + \epsilon_{c,t},$$

where $\delta_{c,t} > 0$.

- We call $\delta_{c,t}$ the *shock term*.
- A-priori we do not think that $\delta_{c,t}$ will be large for most country-years.

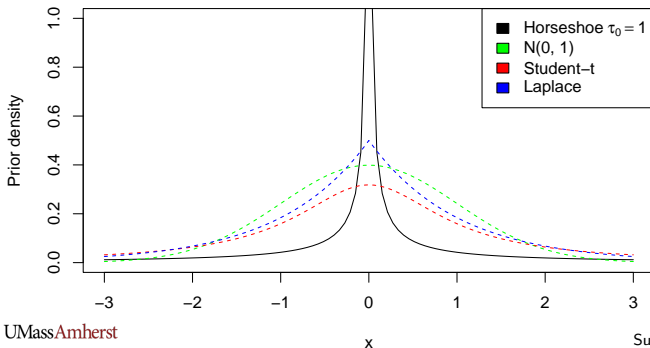


The horseshoe prior

- The *horseshoe prior* (Carvalho 2009 PMLR) is given by

$$\delta_{c,t} \mid \tau, \gamma_{c,t} \sim N(0, \tau^2 \gamma_{c,t}^2)$$
$$\gamma_{c,t} \sim C^+(0, 1).$$

- Global scale parameter $\tau > 0$ shrinks all shocks to zero.
- Local scale parameters $\gamma_{c,t} > 0$ allow some shocks to escape shrinkage.



Regularized Horseshoe

- ▶ The *regularized horseshoe prior* (Piironen 2017 EJS) is given by

$$\delta_{c,t} \mid \tau, \gamma_{c,t}, \vartheta \sim N(0, \tau^2 \tilde{\gamma}_{c,t}^2),$$

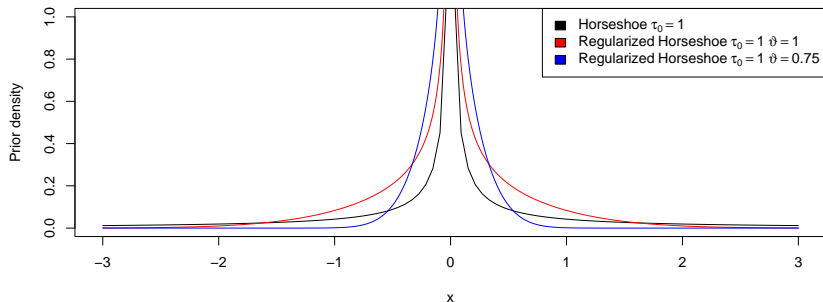
$$\tilde{\gamma}_{c,t}^2 = \frac{\vartheta^2 \gamma_{c,t}^2}{\vartheta^2 + \tau^2 \gamma_{c,t}^2},$$

$$\gamma_{c,t} \sim C^+(0, 1),$$

$$\tau \sim C^+(0, \tau_0^2).$$

- ▶ Global scale parameter $\tau > 0$ shrinks all shocks to zero.
- ▶ Local scale parameters $\gamma_{c,t}$ allow some shocks to escape shrinkage.
- ▶ “Slab scale” parameter $\vartheta > 0$ regularizes shocks that escape regularization.
 - ▶ For large $\delta_{c,t}$, prior approaches a Gaussian prior with variance ϑ^2

Regularized Horseshoe



Estimated shocks

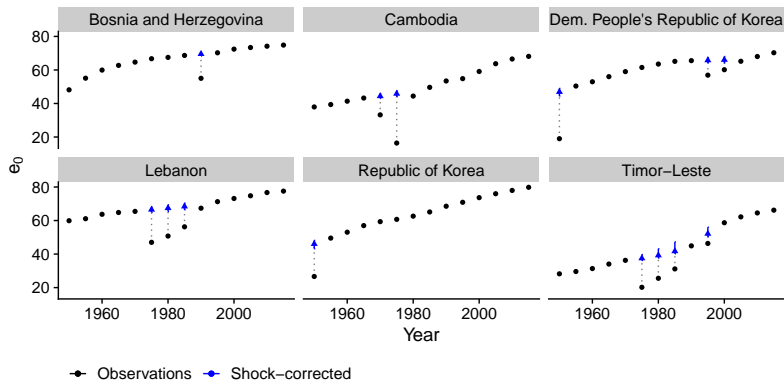


Figure: Six countries with the largest estimated detected shocks. Shocks are illustrated by plotting “shock-corrected” estimates, given by the observed e_0 minus the shock $\delta_{c,t}$

Example e_0 projections

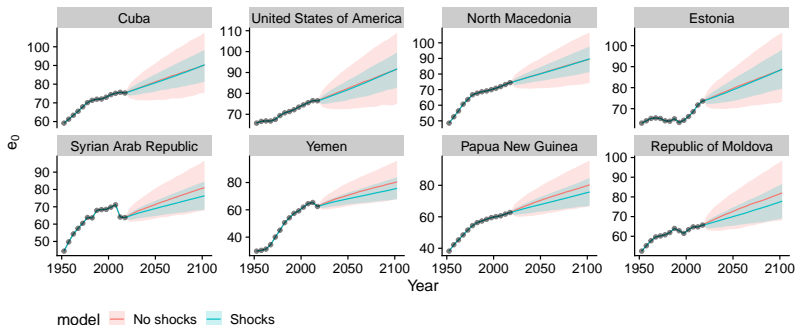
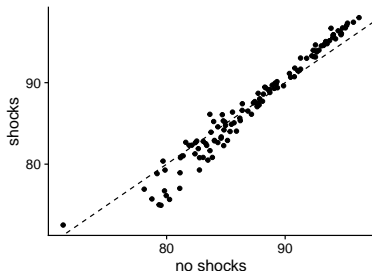


Figure: Projections of e_0 from the model with and without shocks, for countries with the smallest (top row) and largest (bottom row) differences in posterior median projected e_0 in 2095-2100.

Adding shocks has little effect on median projections, and reduces projection uncertainty

Male period life expectancy by country, 2095–2100

A posterior median



B 80% credible interval width

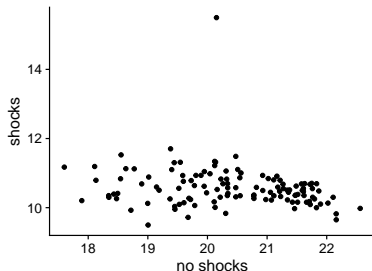


Figure: Posterior medians (A) and 80% projection interval widths (B) for male period life expectancy at birth by country in 2095-2100 for the model with and without shocks included.

Another application: adding shocks improves historical estimates

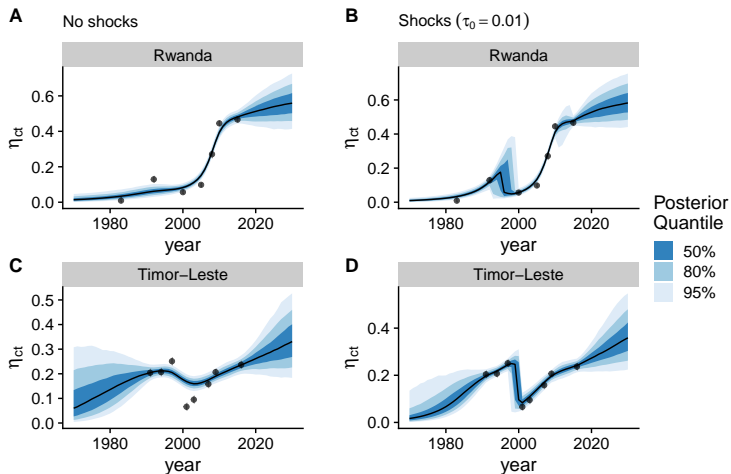


Figure: Modern Contraceptive Use Rate (mCPR) estimates with and without shocks.

Discussion

- ▶ Bayesian shrinkage priors provide a reasonable way to adapt models to handle demographic and health indicators that exhibit shocks.
- ▶ Important to carefully define the target indicator and goal of projections.
- ▶ Practically: regularized horseshoe prior available in `brms`, can be implemented in Stan.
- ▶ More details: <https://arxiv.org/abs/2410.09217>



Appendix: Tuning the prior

- Regularized horseshoe, full specification:

$$\delta_{c,t} \mid \gamma_{c,t}, \tau, \vartheta \sim N(0, \tau^2 \tilde{\gamma}_{c,t}^2),$$

$$\tilde{\gamma}_{c,t}^2 = \frac{\vartheta^2 \gamma_{c,t}^2}{\vartheta^2 + \tau^2 \gamma_{c,t}^2},$$

$$\gamma_{c,t} \sim C^+(0, 1),$$

$$\tau \sim C^+(0, \tau_0^2),$$

$$\vartheta^2 \sim \text{Inv-Gamma}(\nu/2, \nu s^2/2).$$

- Calibrate priors by assuming
 - shock cannot exceed 100 life-expectancy years,
 - a 10% probability of shock exceeding 20 life-expectancy years,
 - and $P(\delta_{c,t} > \delta^*) \approx 0.5\%$.

Appendix: Defining shocks

- ▶ The distribution of the deviations $\epsilon_{c,t}$ characterize “regular” fluctuations in life-expectancy.
- ▶ We propose analyzing the shocks $\delta_{c,t}$ relative to the distribution of $\epsilon_{c,t}$.
- ▶ Define a *shock* as when the value of δ is greater than twice the marginal standard deviation of ϵ : $\delta > 2\text{sd}(\epsilon) := \delta^*$

Appendix: Country with largest credible interval width

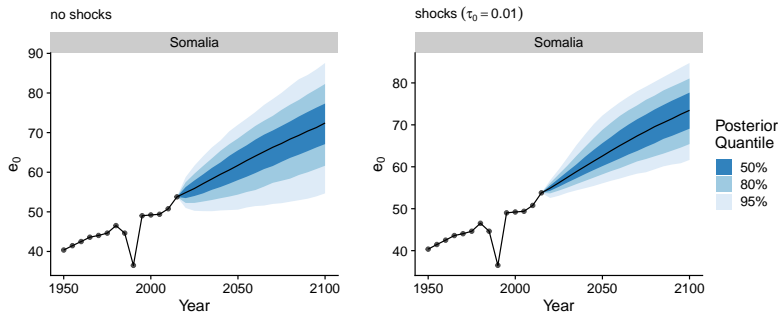


Figure: Projections of e_0 from the model with and without shocks for Somalia.

Appendix: Validations

- ▶ Out-of-sample model validations were used to compare predictive performance of the model with and without shocks.
- ▶ All observations after 2005-2010 held out, and fitted model used to predict observations from 2015-2020.

Appendix: Validation Results

Region	n	ME	MAE
<i>Shocks</i>			
Africa	11	0.49	1.14
Americas	25	-0.07	0.67
Asia & Oceania	50	0.62	0.86
Europe	35	1.10	1.10
<i>Overall</i>	<i>121</i>	<i>0.57</i>	<i>0.92</i>
<i>No shocks</i>			
Africa	11	0.42	1.27
Americas	25	-0.22	0.68
Asia & Oceania	50	0.67	1.02
Europe	35	0.76	0.78
<i>Overall</i>	<i>121</i>	<i>0.61</i>	<i>0.86</i>

Table: Validation results for the male period life expectancy at birth model with and without shocks. Included validation metrics are median error (ME), and median absolute error (MAE).

Appendix: Validation Results

Region	<i>n</i>	% Below	80% projection Interval		PI Width
			% Included	% Above	
<i>Shocks</i>					
Africa	11	9.1%	72.7%	18.2%	3.12
Americas	25	12.0%	88.0%	0.0%	3.12
Asia & Oceania	50	4.0%	72.0%	24.0%	3.13
Europe	35	0.0%	71.4%	28.6%	3.08
<i>Overall</i>	<i>121</i>	<i>5.0%</i>	<i>75.2%</i>	<i>19.8%</i>	<i>3.11</i>
<i>No shocks</i>					
Africa	11	0.0%	100.0%	0.0%	7.55
Americas	25	0.0%	100.0%	0.0%	7.61
Asia & Oceania	50	4.0%	92.0%	4.0%	7.56
Europe	35	0.0%	100.0%	0.0%	7.64
<i>Overall</i>	<i>121</i>	<i>1.7%</i>	<i>96.7%</i>	<i>1.7%</i>	<i>7.60</i>

Table: Validation results for the male period life expectancy at birth model with and without shocks. Included validation metrics are the % of observations below, above, and included within the 80% projection interval, and the mean width of the 80% projection interval.