# Probabilistic Subnational Population Projections

**Hana Ševčíková and Adrian Raftery**

University of Washington, hanas@uw.edu

Research supported by NICHD

Population Association of America 2025 Annual Meeting
April 10-13, 2025, Washington DC

# Probabilistic Population Projection

- Demographic balancing equation:

$$Pop_t = Pop_{t-1} + Births_t - Deaths_t + Net\ Migration_t$$

$\longrightarrow$ solved using Cohort Component Method (CCM)

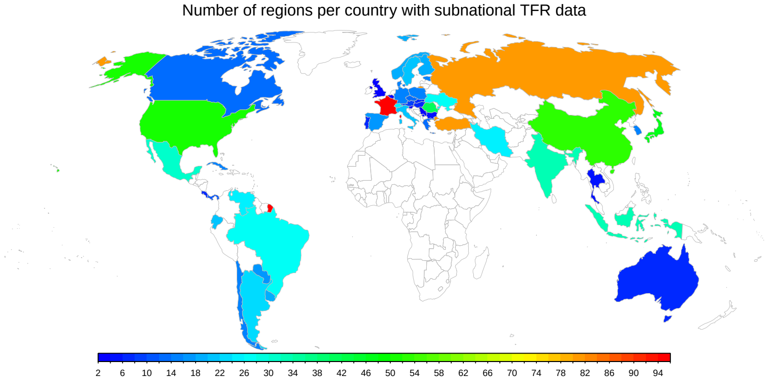
- **Probabilistic approach:**
  1. Using Bayesian hierarchical models, generate sets of trajectories for TFR, $e_0$ and net migration.
  2. Convert each trajectory to age-specific rates and counts.
  3. Apply the CCM to each trajectory of age-specific fertility and mortality rates and migration counts.
  4. Result: Future trajectories of age- and sex-specific population quantities.

# Subnational TFR and $e_0$: Objectives

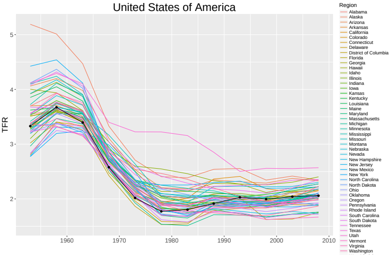
- There are established Bayesian Hierarchical Models (BHM) for projecting TFR (Alkema et al. 2011) and $e_0$ (Raftery et al. 2013) that the UN uses for projections on national level.
- **Objectives for subnational models:**
  - Probabilistic approach built on established national BHM;
  - Works well for all regions of all countries;
  - Yields correlation between regions similar to observed data;
  - Is easy to use, understand and implement.

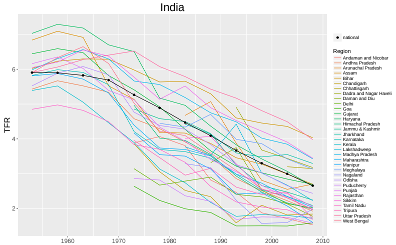
# Subnational TFR: Data

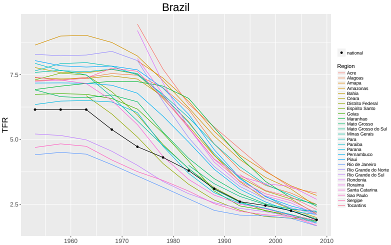
- Subnational TFR data collected by the UN, from 1950–2010: 47 countries, 1 092 regions.

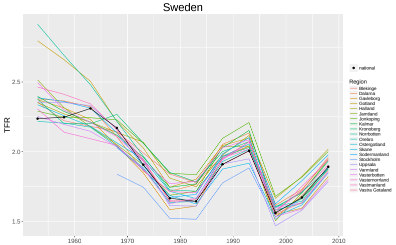


# Data Examples



# Methods for Subnational TFR

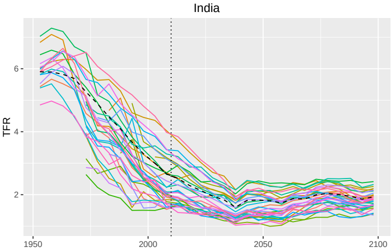
**Scale-AR(1)**

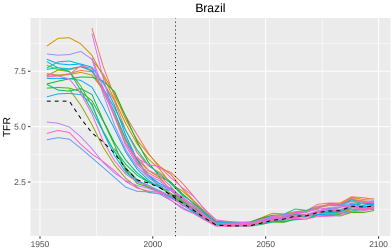
- Multiply each national trajectory by a region-specific scaling factor that changes slowly over time.

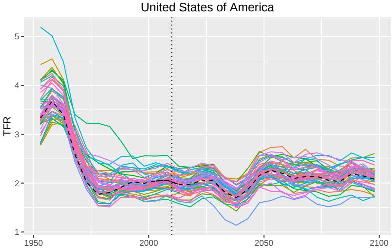
$$
\begin{aligned}
f^{(R)}_{r_c,t,i} &= \alpha_{r_c,t} f^{(C)}_{c,t,i} \\
\alpha_{r_c,t} - 1 &= \phi(\alpha_{r_c,t-1} - 1) + \varepsilon_{r_c,t}, \quad \text{with } \varepsilon_{r_c,t} \overset{iid}{\sim} N(0, \sigma^2_c)
\end{aligned}
$$

| | |
|---|---|
| $f^{(R)}_{r_c,t,i}$ | TFR for region $r_c$ of country $c$ at time $t$ in the $i$-th trajectory |
| $f^{(C)}_{c,t,i}$ | national TFR projection for country $c$ at time $t$ in the $i$-th trajectory (output of national BHM) |
| $\alpha_{r_c,t}$ | regional scaling factor; initialized as $\alpha_{r_c,P} = f^{(R)}_{r_c,P}/f^{(C)}_{c,P}$ ($P$ is the last observed time period) converges to a distribution centered around one |
| $\phi$ | determines the rate of convergence; $\hat{\phi} = 0.925$ (slow) |

# Results: 1 trajectory out of 2000

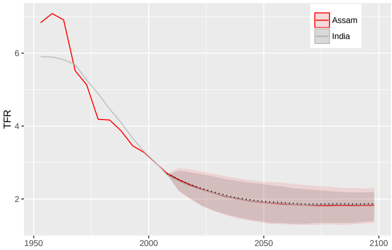


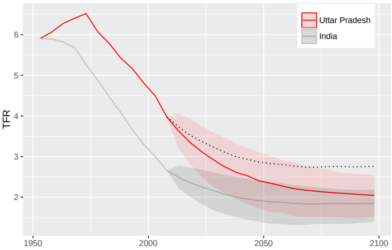



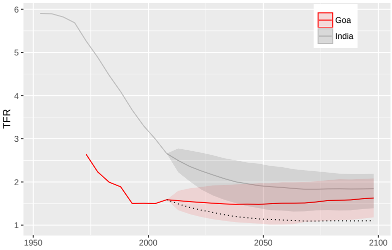



# Results: Projections for Scale-AR(1)

## Three regions of India



# Conclusions on Subnational TFR

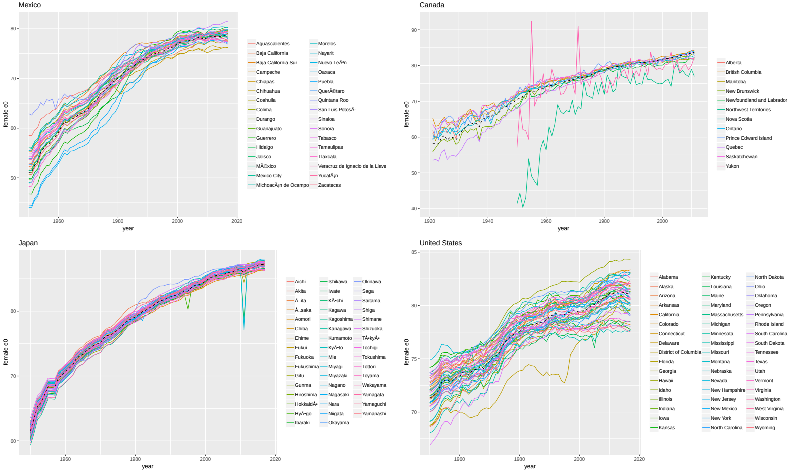
The **Scale-AR(1) method** for subnational total fertility projections:

- is a simple, probabilistic model that builds on the established national BHM method;
- yields correlation between regions similar to observed data;
- yields reasonable out-of-sample validation;
- is in line with hypothesis of Watkins (1990, 1991) that within-country TFR converges in response to country-specific factors (Europe 1870–1960);
- implemented in bayesTFR.

Published in Ševčíková et al. (*DR*, 2018).

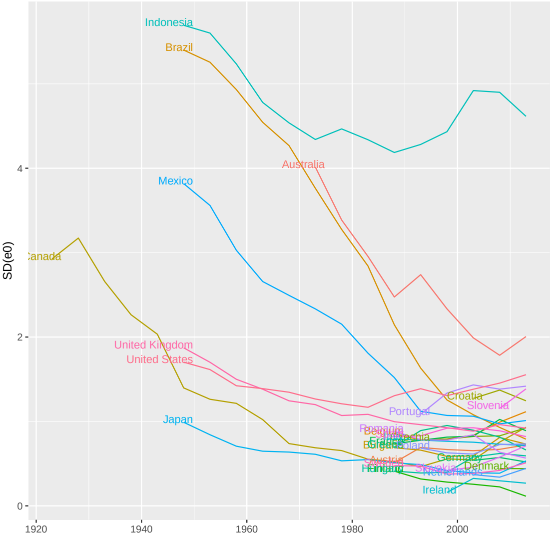
# Data on subnational life expectancy at birth ($e_0$)

Data from 28 countries, 447 regions



# Data on subnational $e_0$

## Within-country between-region standard deviation



# Models for subnational $e_0$

## Shift-AR(1) method

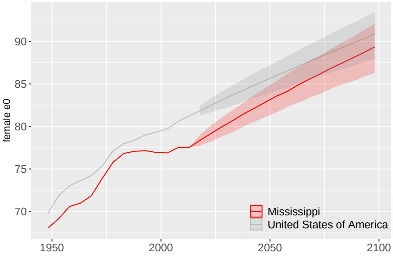
$$
\begin{aligned}
e^{(R)}_{r,c,t} &= e^{(C)}_{c,t} + \alpha_{r,c,t} \\
\alpha_{r,c,t} &= \rho \alpha_{r,c,t-1} + \varepsilon_{r,c,t}, \quad \text{with } \varepsilon_{r,c,t} \overset{\text{ind}}{\sim} N(0, \sigma^2_{c,t})
\end{aligned}
$$

- $\alpha_{r,c,t}$ is time dependent and additive, modeled via an AR(1)
- $\rho$ is constant across countries and regions
- Residual variance $\sigma^2_{c,t}$ varies between countries and is defined as

$$
\sigma^2_{c,t} = \begin{cases} a + b(e^{(C)}_{c,t} - U) & e^{(C)}_{c,t} < U \\ a & e^{(C)}_{c,t} \geq U \end{cases}
$$

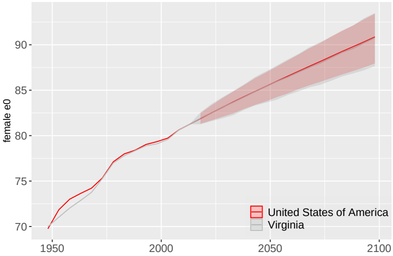
- Using observed data, estimation yields
  $\hat{\rho} = 0.95$, $\hat{U} = 82.5$, $\hat{a} = 0.0482$, and $\hat{b} = -0.0154$

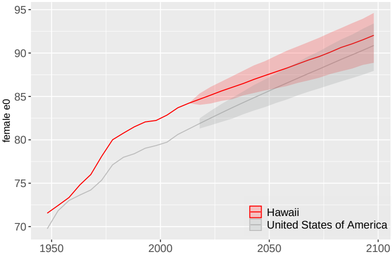
# Subnational predictive distribution of female $e_0$ (USA)



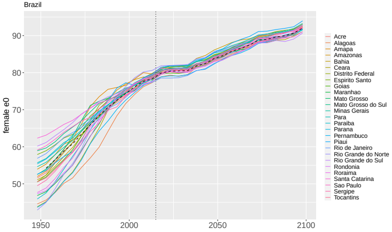

Red: Shift-AR(1) method

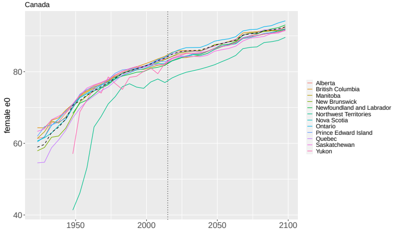
Shaded: 80% probability intervals

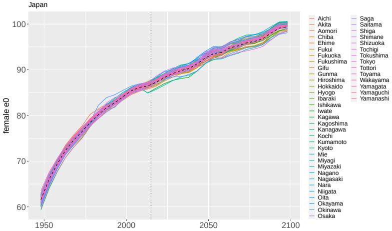


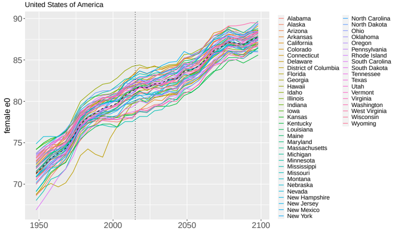



# Crossover patterns for Shift-AR(1): 1 trajectory out of 1000



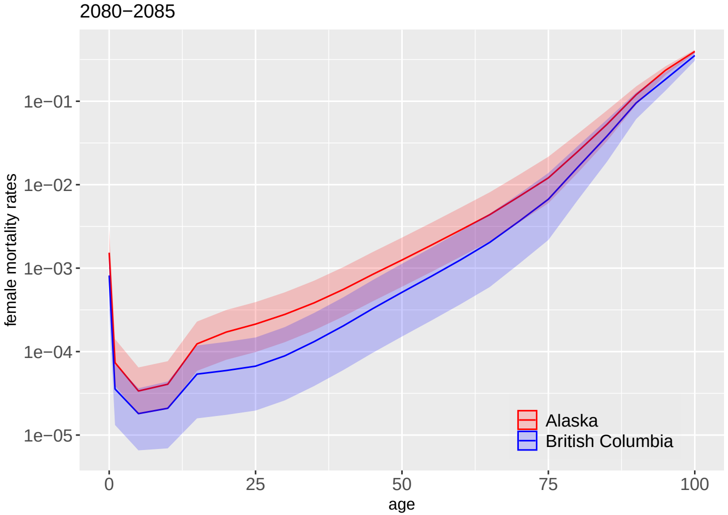



# Probabilistic subnational mortality rates

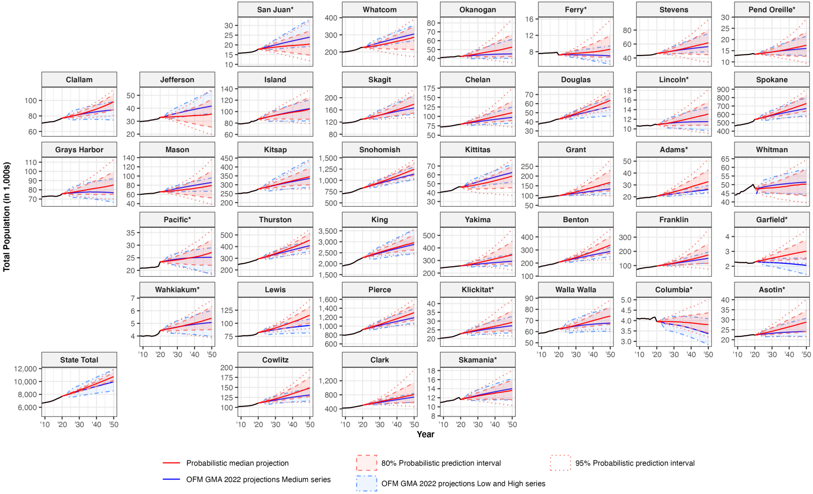
Each trajectory of $e_0$ converted into one trajectory of $m_x$ (Ševčíková et al 2016; MortCast R package) $\longrightarrow$ probabilistic distribution of $m_x$



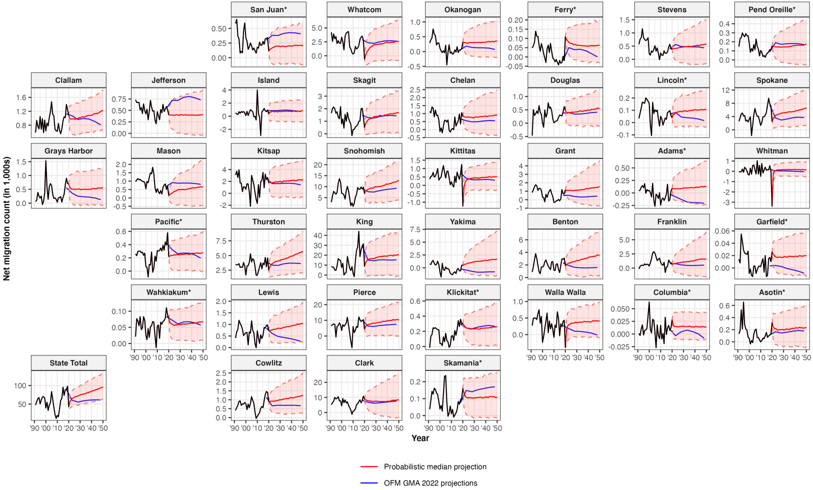
# Subnational Probabilistic Population Projection

- Methodology applied to 39 counties of Washington state (12 counties smaller than 25 000).
- For each county, TFR and $e_0$ projections generated to 2050.
- Special treatment of TFR in counties with large college population $\rightarrow$ projected noncollege TFR
- Net migration BHM excluded 12 small counties from influencing the global experience.
- Cohort-component method applied to each trajectory of the three components of population change.
- College population was excluded from the CCM and added back after advancing the remaining population.
- Yu et al. (2023), *Demography*
- Implemented in the bayesPop R package

# Probabilistic Population Projection for WA counties

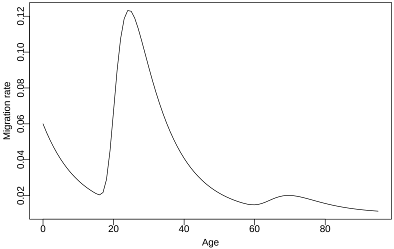


# Probabilistic Migration Projection for WA counties



# Age-specific Net Migration

- Practitioners often distribute total net migration (= difference between in- and out-mi gration) into ages via the Rogers-Castro model




- The Rogers-Castro model was developed for out-migration.
- For subnational population projections, applying Rogers-Castro to net migration can yield unrealistic future population age patterns.

# Age-specific Net Migration: Flow Difference Method (FDM)

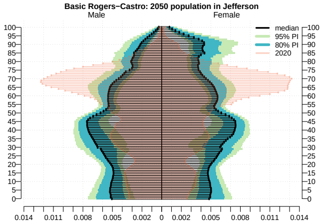
(Ševčíková, Raymer & Raftery, 2024, arXiv:2411.09878)

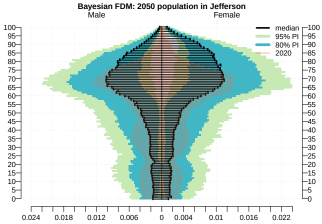
Idea:

1. Split total net migration into total in-migration and total out-migration.
2. Estimate Rogers-Castro parameters for each flow and distribute them into age-specific in- and out-migration.
3. Take their difference.

*More details in Session # 119, Saturday 8am*

# FDM: Use in Probabilistic Population Projections



# Probabilistic Population Projections: Workflow