Probabilistic projection of the sex ratio at birth and missing female births by State and Union Territory in India

Fengqing Chao[1] Christophe Z. Guilmoto[2] Samir K.C.[3,4] Hernando Ombao[1]

[1]CEMSE, KAUST [2]CEPED/IRD, Université de Paris [3]ADRI, Shanghai University [4]IIASA

KAUST King Abdullah University of Science and Technology

STATISTICS @KAUST

Introduction

Sex Ratio at Birth (SRB)

- SRB: ratio of male to female births;
- An important indicator:
  - For population estimation and projection;
  - To assess the pre-natal gender equality.

SRB imbalance

- The biological/natural SRB fluctuates within a narrow band between 103 to 107 male births per 100 female births;
- Since 1970, observed SRB in some Asian and Eastern European countries are much higher than the natural level.
- The imbalanced SRB is due to the coexistence of 3 main factors:
  1. Strong son preference at population level;
  2. Sex determine and abortion technology is accessible and affordable;
  3. Family size is getting smaller over time.
- On national level, my previous study (Chao, F. et al 2019a) reports SRB inflation in India among 11 other countries/areas.

Indian demography and SRB imbalance

- India has a huge population size: 1.38b (by 2030 will be the most populous countries, projected at 1.5b, UN WPP).
- Great heterogeneity of Indian demography;
- SRB imbalance in India began the earliest in mid 1970s;
- Studies have shown great diversity in the SRB trajectories across geographic locations before 2016 (Chao, F. et al 2019b);
- Important to produce SRB projection by Indian states/territories.

Objective

- Project SRB by State and Union Territory (UT) in India during 2017–2030;
- Project the female births deficits based on SRB.

Now on arXiv!

- Chao, F., Guilmoto, C. Z., K.C, S., Ombao, H. (2020). Probabilistic Projection of the Sex Ratio at Birth and Missing Female Births by State and Union Territory in India. *arXiv preprint arXiv:2004.02228.*
- About me: `https://www.fengqingchao.com/`

Results

Covariate Effect on State-level SRB

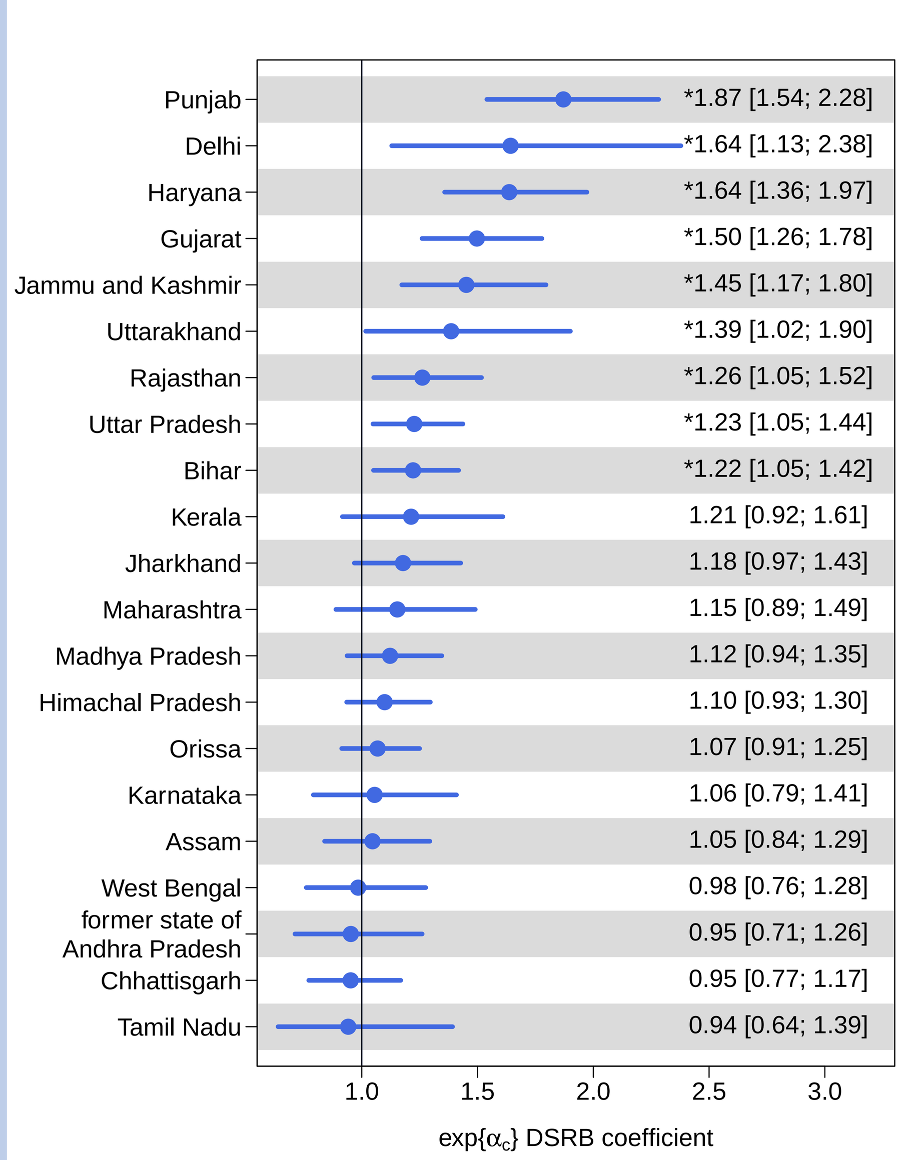

Figure 1: **DSRB effect on SRB by Indian State/UT.** Dots are median estimates. Horizontal line segments are 95% credible intervals. * indicates that the effect is statistically significantly different from 1. States/UTs are in descending order of the median estimates.

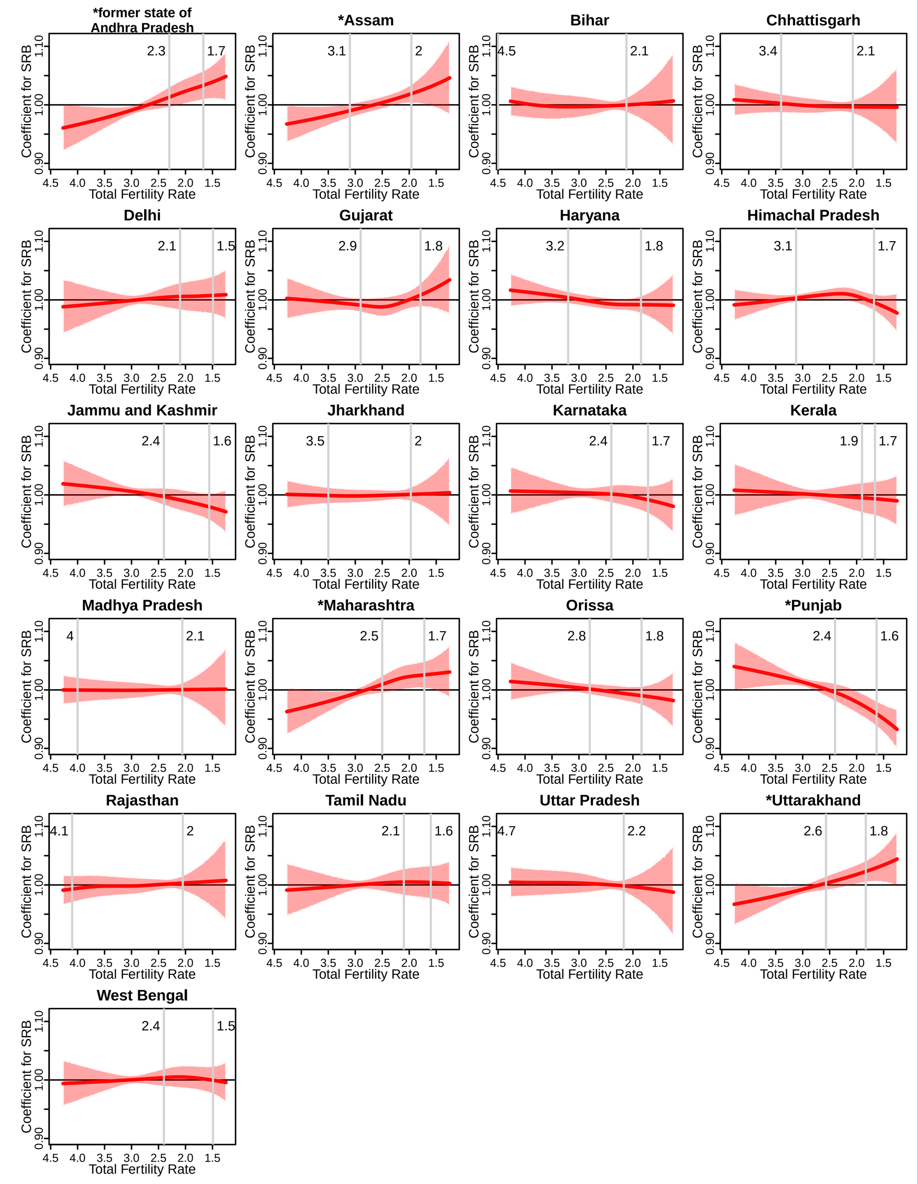

Figure 2: **TFR effect on SRB by Indian State/UT.** Curves are median estimates. Shaded areas are 95% credible intervals. The maximum and minimum TFR that is available for each state during 1990–2030 are indicated with vertical lines and values are shown by the lines. * indicates that the effect of TFR is statistically significantly different from 1 for at least one given value of TFR.

SRB median projections in 2030, by Indian State/UT

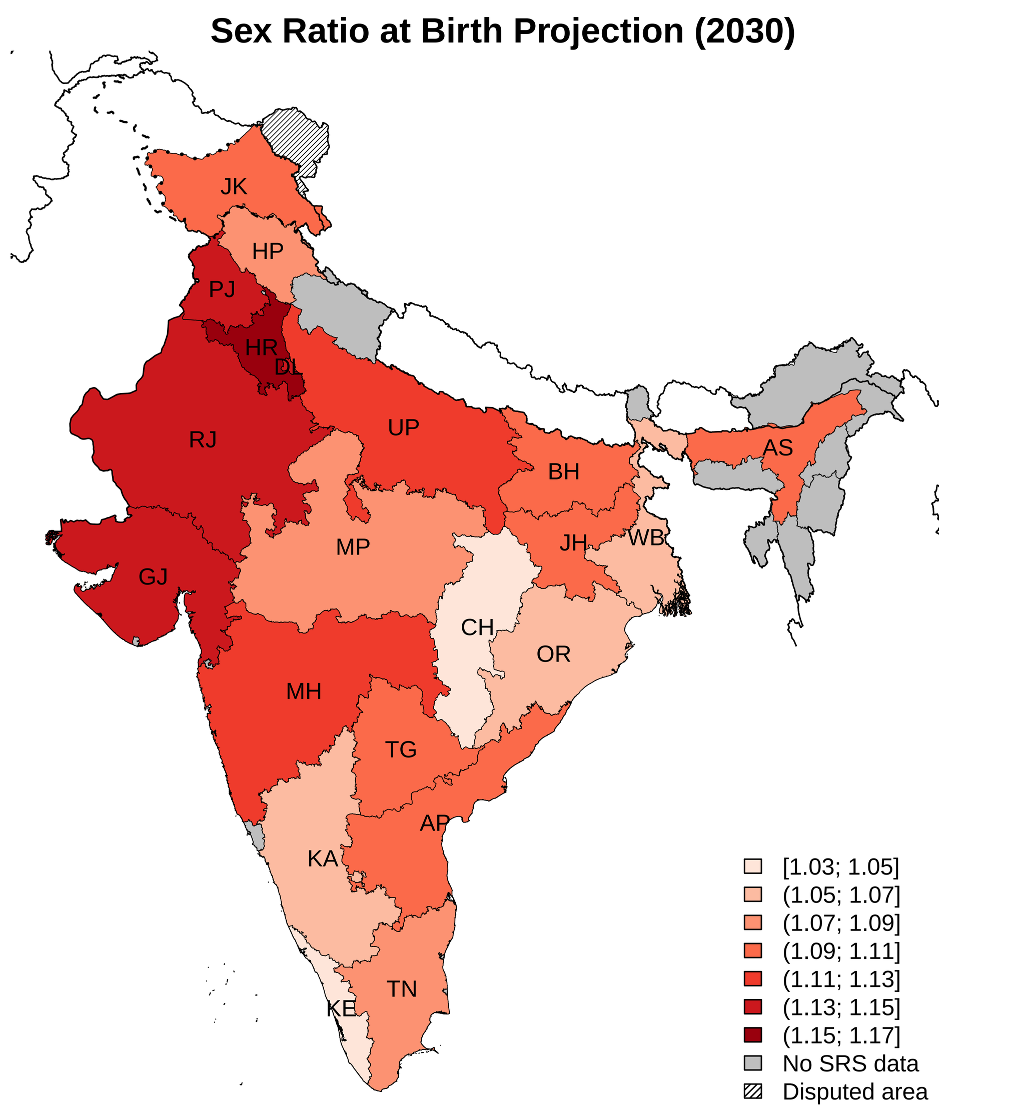

Figure 3: Results are shown for the 21 Indian States/UTs with SRS data. Values for Andhra Pradesh and Telangana are the same. State/UT names are: Andhra Pradesh (AP); Assam (AS); Bihar (BH); Chhattisgarh (CH); Delhi (DL); Gujarat (GJ); Haryana (HR); Himachal Pradesh (HP); Jammu and Kashmir (JK); Jharkhand (JH); Karnataka (KA); Kerala (KE); Madhya Pradesh (MP); Maharashtra (MH); Orissa (OR); Punjab (PJ); Rajasthan (RJ); Tamil Nadu (TN); Telangana (TG; same value as in AP); Uttar Pradesh (UP); Uttarakhand (UT); West Bengal (WB). The boundaries and names shown and the designations used on this map do not imply official endorsement. Dotted line represents approximately the Line of Control in Jammu and Kashmir agreed upon by India and Pakistan. The final status of Jammu and Kashmir has not yet been agreed upon by the parties.

Method

Model for state-level SRB

$R_{c,t}$ for Indian states $c = 1, \cdots, C$ and year $t = 1990, \cdots, 2030$ is modeled as:

$$R_{c,t} = N \cdot P_{c,t}.$$

$N = 1.053$ is the baseline SRB for India (Chao, F. 2019a).
$P_{c,t}$ accounts for the deviation of $R_{c,t}$ from the baseline N and follow a time series model with AR(1) structure, conditioning on country-year-specific mean $V_{c,t}$. For $c = 1, \cdots, C$:

$$P_{c,t}|V_{c,t} \sim \mathcal{N}(V_{c,t}, \sigma_\epsilon^2/(1-\rho_c^2)), \text{ for } t = 1,$$
$$P_{c,t}|P_{c,t-1}, V_{c,t} = V_{c,t} + \rho_c \cdot (P_{c,t-1} - V_{c,t}) + \epsilon_{c,t}, \text{ for } t = 2, \cdots, T,$$
$$\epsilon_{c,t} \sim \mathcal{N}(0, \tau_{\epsilon_c}^{-1}), \text{ for } t = 2, \cdots, T.$$

$V_{c,t}$ is modeled as a multivariate regression function:

$$V_{c,t} = \alpha_c \cdot D_{c,t+5} + f_c(F_{c,t}), \text{ for } \forall c \forall t.$$

- $D_{c,t+5}$: log of desired sex ratio at birth (as a proxy for son preference), is modeled as $\exp\{D_{c,t}\} = 1 + \Delta_{c,t}$ where
$$\Delta_{c,t} = \frac{\delta_c \cdot \exp\{\phi_c \cdot \log(t) + \zeta_c\}}{1 + \exp\{\phi_c \cdot \log(t) + \zeta_c\}}.$$
- $f_c(F_{c,t})$: state-specific second-order random walk (RW2) function as a continuous time process on the log-scaled TFR $F_{c,t}$:
$$f_c(F_{c,t}) = \Delta_{c,t}^2 = F_{c,t} - 2F_{c,t+1} + F_{c,t+2},$$
$$\Delta_{c,t}^2 \sim \mathcal{N}(0, \tau_c^{-1}).$$
- α_c follows hierarchical normal distributions: $\alpha_c|\tau_\alpha \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau_\alpha^{-1})$.

Data Model For the i-th observation r_i:

$$\log(r_i) \sim \mathcal{N}(\log(R_{c[i],t[i]}), 0.001^2), i = 1, \cdots, 528.$$

Computing We fit the model in the open source software `R 3.5.1` and `R`-package `R-INLA` (Rue H. et al 2009).

References

- Chao, F., Gerland, P., Cook, A. R., Alkema, L. (2019). Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. *Proceedings of the National Academy of Sciences, 116*(19), 9303-9311.
- Chao, F., Yadav, A. K. (2019). Levels and trends in the sex ratio at birth and missing female births for 29 states and union territories in India 1990–2016: A Bayesian modeling study. *Foundations of Data Science, 1*(2), 177-196.
- Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B, 71*(2), 319-392.

Population Association of America Annual Meeting (virtual), Apr 22-25 2020, Washington DC