A Systematic Assessment of National, Regional and Global Levels and Trends in the Sex Ratio at Birth and Identification of Countries with Outlying Levels

Fengqing Chao, Leontine Alkema, Patrick Gerland *

March 8, 2016

Extended Abstract for PAA 2016

Abstract

Under normal circumstances, the sex ratio at birth (SRB) has been reported to vary between 1.04 and 1.07. But gender discrimination and other factors associated with the SRB can result in different SRB levels. While SRB estimates have been published for various countries and periods, a systematic assessment of SRBs for all countries over time using all available data and reproducible estimation methods has not been reported to date. In this paper, we estimated country-specific SRBs for 212 countries and areas from 1990 to 2013 using a Bayesian hierarchical time series model and assessed the uncertainty in SRBs. The analysis is based on an extensive database with national-level data, including data from vital registration systems, international and national survey programs. We present results for countries and areas and identify country-years with SRBs which differ significantly from estimates that are typical for the country-specific region.

*Fengqing Chao: Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117546; email: ephchf@nus.edu.sg. Leontine Alkema: Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA 01003-9304, USA; email: lalkema@schoolph.umass.edu. Patrick Gerland: Demographic Estimates and Projections Section, United Nations Population Division, DESA, United Nations, New York, NY 10017, USA; email: gerland@un.org. The project described is solely the responsibility of the authors and does not necessarily represent the official views of the United Nations. The authors are grateful to Christophe Z. Guilmoto for helpful comments and discussions, and Vladimira Kantorova for assistance with the data. This work is funded by a research grant from the National University of Singapore R155000135646 *Missing Girls: A Comprehensive and Systematic Analysis of Trends in Pre- and Postnatal Gender Discrimination.*

1 Introduction

Under normal circumstances, the sex ratio at birth (SRB; refers to the number of male live births per 100 female live births) varies between 1.04 and 1.07 (1). However, the observed SRB from countries like India and China have been well above the biological level for decades (2; 3). While SRB estimates have been published for various countries and periods, to date, no assessment of SRBs for all countries over time has been carried out using all available data and reproducible estimation methods.

In this paper, we estimated country-specific SRBs for 212 countries and areas from 1990 (or the earliest year of data collection) using a Bayesian hierarchical time series model, accounting for the varying levels of uncertainty associated with different observations. We also assessed the uncertainty in SRBs.

2 Data

An overview of the data sources included in the database is in Table 1. There are 9,848 data points available from 192 countries and areas. In total, there are 14,866 country-years available in our database. On average, 70.1 country-years of data are available for each of the 212 countries and areas included in our analysis.

Data source type	Number of observations
Census Direct	56
DHS Direct	1,867
Other DHS Direct	793
Others Direct	155
VR	6,984

Table 1: **Distribution of observations by source type.** Observations are grouped by source types. "Direct" refers to observation obtained from full birth histories. "Indirect" refers to observations obtained from summary information and demographic methods. DHS: Demographic and Health Surveys; VR: Vital Registration.

3 Summary of estimation method

We modeled the true SRB, denoted by $R_{c,t}$ for country c, year t, as follows:

$$R_{c,t} \quad = \quad N_{r[c]} \cdot P_{c,t} + \alpha_{c,t}, \tag{1}$$

where $N_{r[c]}$ is the regional biological norm from country c in region $r[c]$ in the absence of prenatal gender discrimination and sex-selective abortion, $\alpha_{c,t}$ is the upward adjustment factor to capture higher SRB levels that may be due to sex-selective abortion, and country-specific multiplier $P_{c,t}$ captures the divergence from the regional natural level for the respective country-year. We model the biological norm on the regional level to capture racial variation in SRB (4; 5). We assume the regional biological norm $N_{r[c]}$ to be constant over time. The multiplier $P_{c,t}$ is estimated by a time series model. For countries without any data or with very limited information, the multiplier fluctuates around one, such that the estimated SRB without prenatal gender discrimination is given by $N_{r[c]}$. For countries where the data suggest different levels or trends, $P_{c,t}$ captures these deviations from $N_{r[c]}$.

$\alpha_{c,t}$ is the upward adjustment factor for country c year t to capture higher SRB levels that may be due to sex-selective abortion. It is modeled from 1970 onwards for a selected group of countries/areas where gender discrimination may be present and where a son preference may have led or lead to prenatal gender discrimination once fertility declines and sex selective technology becomes accessible (2). The adjustment factor $\alpha_{c,t}$ is set to be zero for all the other country-years. The selection criteria of countries/areas are based on: 1) high level of SRB data in our database and supporting literature; 2) high level of desired sex ratio at birth (DSRB) or high sex ratio of last birth (SRLB) suggested in Bongaarts 2013 (6); and 3) outlying level of excess female under-5 mortality rate (U5MR) suggested in Alkema et al. 2014 (7), which may be caused by postnatal gender discrimination. In our current model, we selected

21 countries/areas as listed in Table 2, which were among those countries that satisfied one or more of aforementioned criteria and in addition, had VR data. We also included China and India, which do satisfy the criteria but do not have VR data. For a future analyses, we plan to also include countries with with evidence of gender discrimination but non-VR data only.

Country/area	**Indicator**	**Reference**
Albania	high SRB;	(2)
Armenia	high SRB; high DSRB; high SRLB;	(8; 9; 10; 6)
Azerbaijan	high SRB; high DSRB; high SRLB;	(8; 9; 10; 6)
China	high SRB; excess female U5MR;	(11; 12; 13; 7)
Egypt	high SRB; excess female U5MR;	(7)
Fiji	high SRB;	suggest by VR data
Georgia	high SRB;	(8; 9; 10)
Hong Kong SAR (China)	high SRB;	(14)
India	high SRB; excess female U5MR; high DSRB; high SRLB;	(11; 15; 2; 6; 7)
Iran	excess female U5MR;	(7)
Jordan	excess female U5MR; high DSRB; high SRLB;	(7; 6)
Korea	high SRB;	(16; 10)
Macao SAR (China)	high SRB;	suggested by VR data
Macedonia	high SRB;	suggested by VR data
Malaysia	high SRB;	(17)
Montenegro	high SRB;	(2)
Singapore	high SRB;	(18)
Taiwan province (China)	high SRB;	(19)
Tonga	high SRB;	suggested by VR data
Tunisia	high SRB;	suggested by VR data
Vietnam	high SRB;	(20)

Table 2: **Countries/areas selected to model the SRB inflation adjustment factor $\alpha_{\mathbf{c,t}}$.** Theses selected places are based on: 1) inflated SRB data; 2) excess female U5MR; 3) high DSRB and/or SRLB.

Adjustment factor $\alpha_{c,t}$ follows a shape of trapezoid (see Figure 1) and its parameters are estimated with a Bayesian hierarchical model (21; 22). E.g., start year parameter $t_{c,0}$ follows a truncated normal distribution with mean equal to the year that the total fertility rate (TFR) in country c decreased (or will decrease) to 3, i.e. $t_{c,tfr3}$:

$$t_{c,0} \quad \sim \quad N(t_{c,tfr3}, \sigma^2_{tfr3})T(1970, 2100).$$

Sampling and non-sampling errors are accounted for in the data model for SRB data. For the i-th observed SRB r_i, we assume:

$$\log(r_i) \quad \sim \quad N(\log(R_{c_i,t_i}), \sigma_i^2 + \omega^2_{s[i]}), \tag{2}$$

where σ_i^2 is the sampling variance (calculated using a Jackknife approach (23) for data from DHS and other DHS data series; and stochastic variance for VR data) and $\omega^2_{s[i]}$ is the non-sampling variance, where $s[i]$ refers to the source type s (shown in Table 1) for the i-th observation. For VR data, we assume their non-sampling variances are zero.

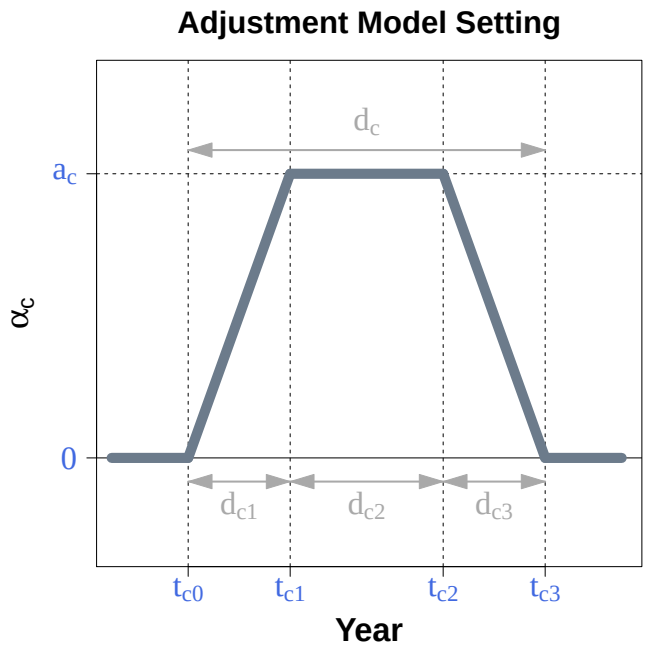

Figure 1: **Illustration for adjustment model setting.** $t_{c,0}$ and $t_{c,3}$ refer to the start and end year of SRB inflation period for country c. d_c is the total length of the SRB inflation period for country c. a_c is the maximum value that the adjustment value could reach for country c.

4 Preliminary Results

4.1 World and region results

Figure 2 illustrates estimates and 95% credible intervals of sex ratio at birth in 1990 and 2013 for the world and regions. Results are weighted by the number of live births in the respective region. Globally, sex ratio at birth is 1.051 (95% credible interval [1.046; 1.055]) in 1990, and 1.046 (95% credible interval [1.040; 1.053]) in 2013. Levels and trends varied across regions. In 1990, the sex ratio at birth is highest in South-eastern Asia in 1990 at 1.066 (95% credible interval [1.059; 1.073]), and decreases slightly to 1.064 (95% credible interval [1.054; 1.073]) in 2013, and still remains the highest among all the regions. Sex ratio at birth is lowest for Africa for both 1990 and 2013 at 1.039 ([1.034; 1.044]) and 1.029 ([1.023; 1.036]).

4.2 Country-level SRB estimates and projections

Figure 3 illustrates model estimates of SRB for the United States and Zambia, two countries that were not selected as countries with (a potential for) inflated SRBs. Bayesian model-based estimates are similar to the WPP 2012 estimates (added in blue) for the United States and higher for Zambia. For future years, the Bayesian projection converges to the regional norm for each country, which is lower for Zambia as compared to the United States, and higher overall for the Bayesian model as compared to the WPP projections.

Figure 4 and Figure 5 illustrate our model estimates of SRB for Albania, Armenia, China and Republic of Korea, which are examples of countries where we incorporated the adjustment factor to account for possible inflation of the SRB. The TFR estimates/projections for these four countries are overlaid on to the SRB estimates in the two figures. This is to illustrate the relationship between the starting year of SRB inflation period (red vertical line) and the fertility decline. For example, based on our model results, the start year of the SRB inflation period for Armenia is estimated to be 1991, which corresponds with the year in which the TFR has decreased to 2.5. For China, the start year is estimated to be 1975, which is quite close to the year when the TFR in China declined to 3, i.e. 1978. Our model suggests that the SRB for Albania will converge back to the regional expected SRB level in 2022, in 2028 for Armenia, in 2030 for China and the level reverted back to normal in the Republic of Korea in 2011. The uncertainty intervals seem

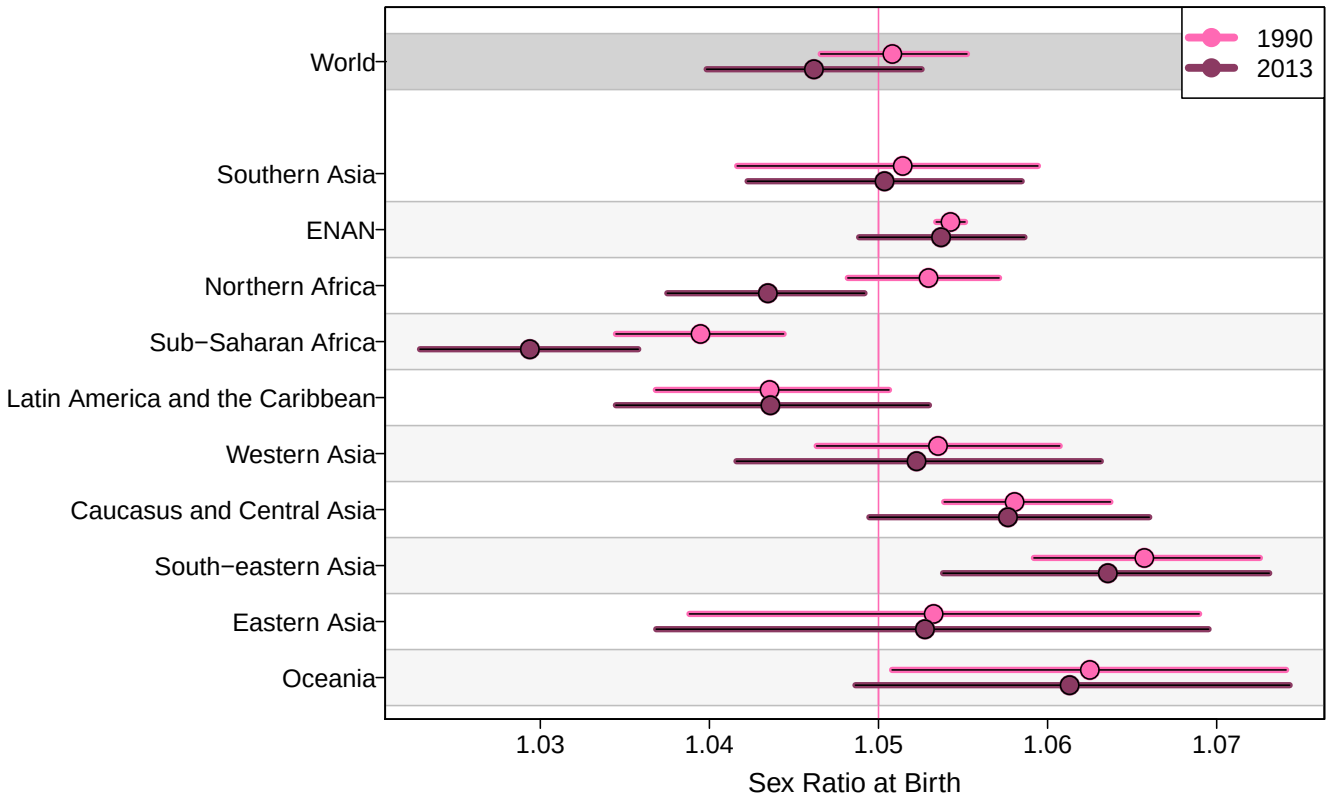

Figure 2: **Sex ratio at birth for the world and regions in 1990 and 2013.** Dots indicate median estimates, and horizontal lines refer to 95% credible intervals. The vertical line refers to the estimated natural level of SRB at 1.057. Region "ENAN" refers to the combination of countries in Europe, North America, Australia, and New Zealand.

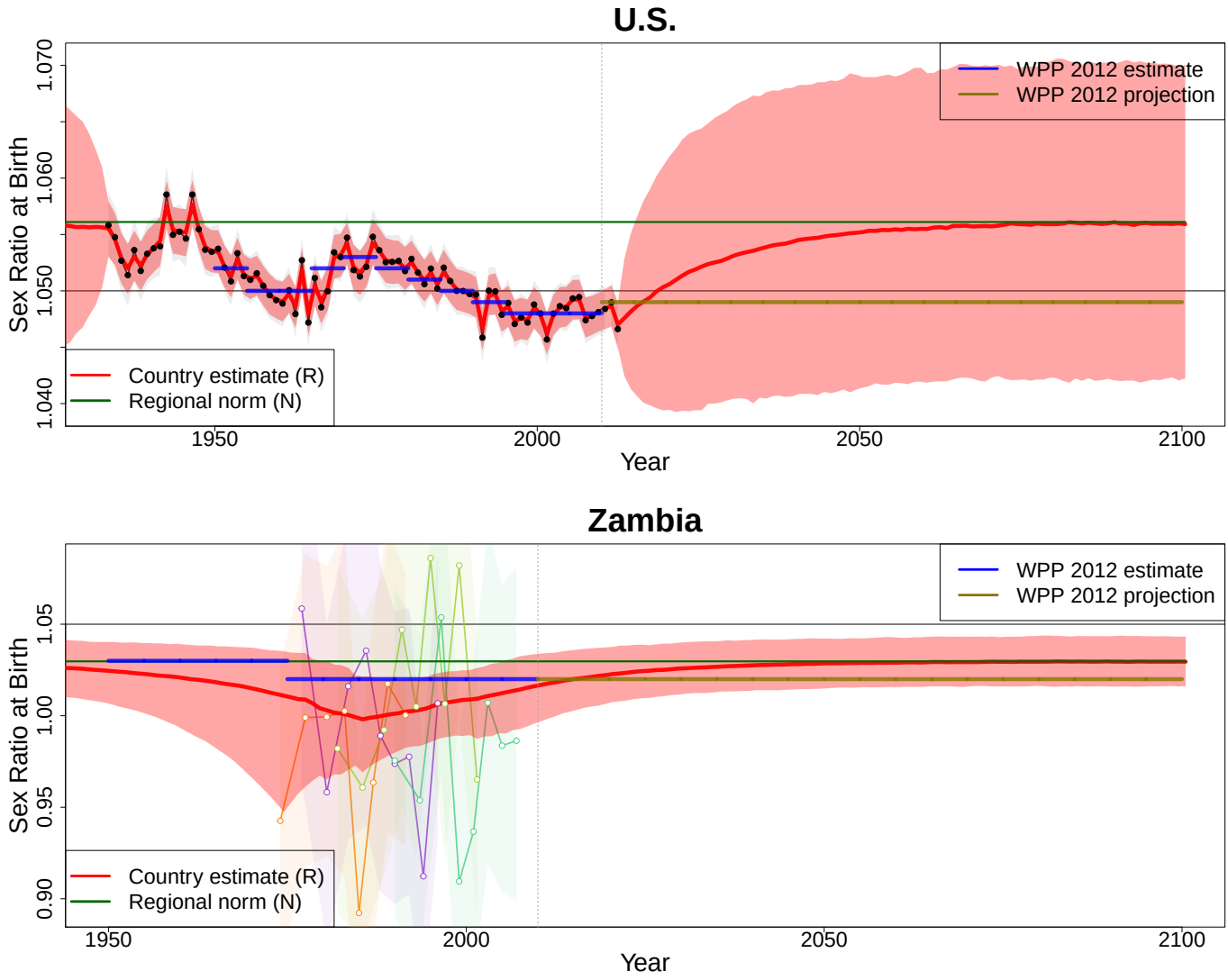

Figure 3: **SRB over time for the United States and Zambia with WPP 2012 SRB estimates.** Observations from different data series are differentiated by colors. The shaded areas surrounding observations represent the sampling or stochastic errors associated with the observations. The red lines and shaded areas are the country-specific SRB estimates and their 95% credible bounds. The green lines and shades are the regional estimates for the biological norm of SRB N_r and their 95% credible bounds. The blue (gold) lines are WPP 2012 estimates/projections.

quite narrow for the projection period, we are testing the out-of-sample performance of this model and will make modifications if performance is not satisfactory.

4.3 Countries with outlying SRBs

Figure 6 illustrates countries and areas with substantially higher (left column in plot) or lower (right column in plot) SRB estimates in 1990 and/or 2013 as compared to the biological norm. We remind the reader of the fact that the current model does not yet include adjustment factors for all countries where SRB inflation may be or become present. Republic of Korea has the highest SRB estimate in 1990 at 1.164, and decreased significantly to 1.056 in 2013. Montenegro, Hong Kong SAR (China), Macao SAR (China), Fiji, Vietnam, Armenia, Georgia, and Azerbaijan do not have outlying SRB levels in 1990 but become outlying in 2013 (SRBs given by 1.089, 1.155, 1.097, 1.085, 1.123, 1.128, 1.083, and 1.154 respectively). Among the 16 countries and areas identified to have very high level of SRB in 1990 and/or 2013 (left panel), ten of them are from Asia (among the ten, five are from Eastern Asia). Among the 14 countries and areas identified to have very low SRB in 1990 and/or 2013 (right panel), 11 of them are from Sub-Saharan Africa.

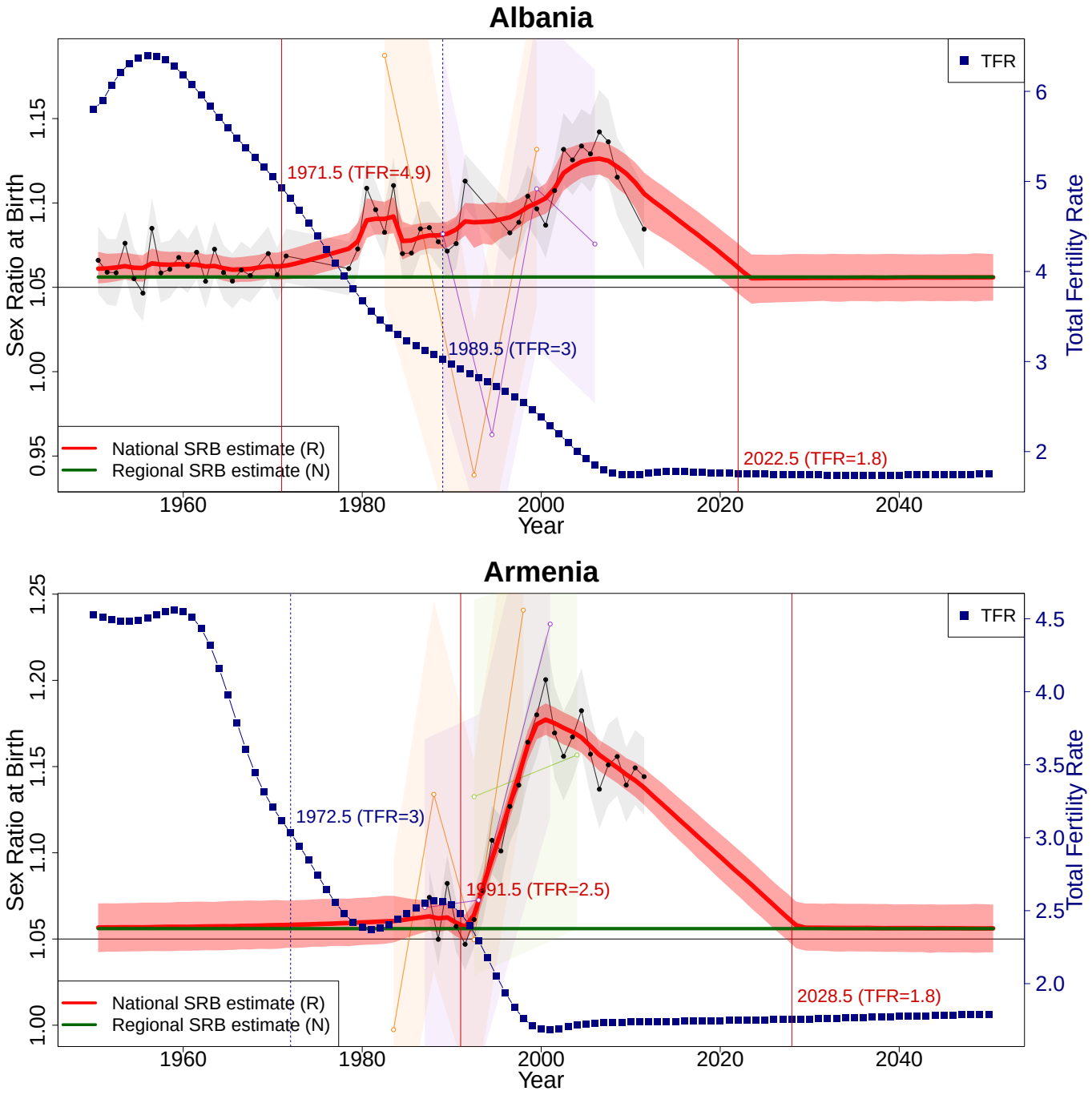

Figure 4: **SRB over time for China and India with WPP 2012 TFR estimates.** Observations from different data series are differentiated by colors. The shaded areas surrounding observations represent the sampling or stochastic errors associated with the observations. The red lines and shaded areas are the country-specific SRB estimates and their 95% credible bounds. The green lines and shades are the regional estimates for the biological norm of SRB N_r and their 95% credible bounds. The blue square dots are WPP 2012 total fertility rate estimates. The two red vertical lines indicate the median estimates for starting and ending years of SRB inflation period. The blue dashed vertical line indicates the first year when a country reached a TFR of 3.

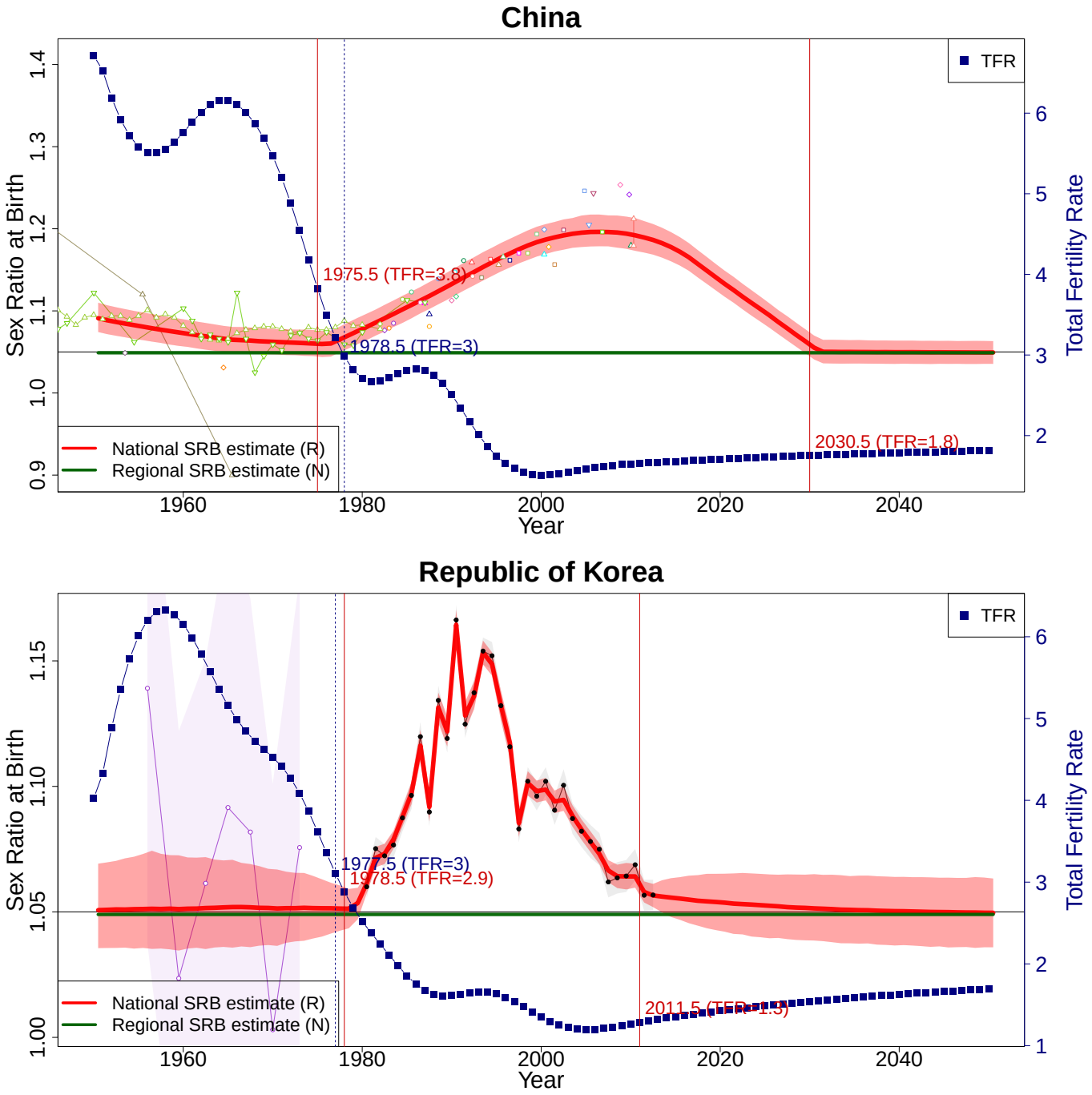

Figure 5: **SRB over time for Republic of Korea and Vietnam with WPP 2012 TFR estimates.** Observations from different data series are differentiated by colors. The shaded areas surrounding observations represent the sampling or stochastic errors associated with the observations. The red lines and shaded areas are the country-specific SRB estimates and their 95% credible bounds. The green lines and shades are the regional estimates for the biological norm of SRB N_r and their 95% credible bounds. The blue square dots are WPP 2012 total fertility rate estimates. The two red vertical lines indicate the median estimates for starting and ending years of SRB inflation period. The blue dashed vertical line indicates the first year when a country reached a TFR of 3.

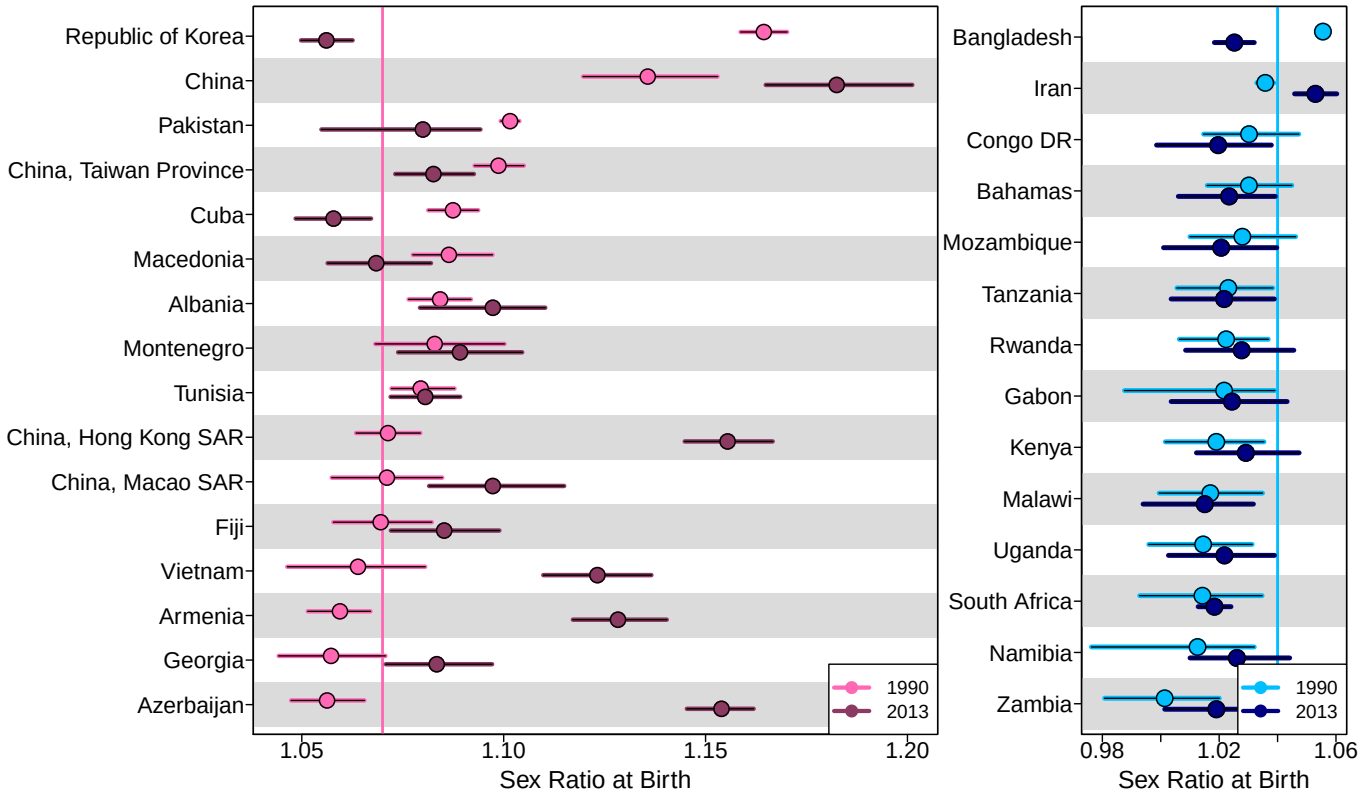

Figure 6: **Countries and areas with outlying SRB estimates in 1990 and 2013.** Countries and areas are ordered by decreasing point estimates for the year 1990. Dots are median estimates, and horizontal lines are 95% credible intervals. Left: Countries and areas with SRB credible intervals higher than 1.07, in 1990 and/or 2013. Right: Countries and areas with SRB credible intervals lower than 1.04, in 1990 and/or 2013. The vertical lines indicates are at 1.04 and 1.07 to identify outlying high and low levels of SRB respectively.

References

[1] Chahnazarian A. Determinants of the sex ratio at birth: Review of recent literature. Biodemography and Social Biology. 1988;35(3-4):214–235. Available at http://www.tandfonline.com/doi/abs/10.1080/19485565.1988.9988703#.U7ogvpSSx1Z.

[2] Guilmoto C. Sex Imbalances at Birth: Current trends, consequences and policy implications. Bangkok, Thailand: UNFPA Asia and Pacific Regional Office; 2012. Available from: http://www.unfpa.org/webdav/site/global/shared/documents/publications/2012/SexImbalancesatBirth.PDFUNFPAAPROpublication2012.pdf.

[3] Goodkind D. Child underreporting, fertility, and sex ratio imbalance in China. Demography. 2011;48(1):291–316.

[4] James WH. The human sex ratio. Part 1: A review of the literature. Human Biology. 1987;p. 721–752. Available at http://www.jstor.org/stable/41464836.

[5] Mathews T, Hamilton BE, et al. Trend analysis of the sex ratio at birth in the United States. National vital statistics reports. 2005;53(20):1–17.

[6] Bongaarts J. The implementation of preferences for male offspring. Population and Development Review. 2013;39(2):185–208.

[7] Alkema L, Chao F, You D, Pedersen J, Sawyer CC. National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: a systematic assessment. The Lancet Global Health. 2014;2(9):e521–e530.

[8] Meslé F, Vallin J, Badurashvili I. A sharp increase in sex ratio at birth in the Caucasus. Why? How? Watering the Neighbour's Garden: The Growing Demographic Female Deficit in Asia, Paris: Committee for International Cooperation in National Research in Demography. 2007;p. 73–88.

[9] Duthé G, Meslé F, Vallin J, Badurashvili I, Kuyumjyan K. High sex ratios at birth in the Caucasus: modern technology to satisfy old desires. Population and Development Review. 2012;38(3):487–501.

[10] Guilmoto CZ. The sex ratio transition in Asia. Population and Development Review. 2009;35(3):519–549.

[11] Guilmoto CZ. Skewed sex ratios at birth and future marriage squeeze in China and India, 2005–2100. Demography. 2012;49(1):77–100.

[12] Guilmoto CZ, Ren Q. Socio-economic Differentials in Birth Masculinity in China. Development and change. 2011;42(5):1269–1296.

[13] Gupta MD, Chung W, Shuzhuo L. Evidence for an incipient decline in numbers of missing girls in China and India. Population and Development Review. 2009;35(2):401–416.

[14] Basten S, Verropoulou G. 'Maternity migration' and the increased sex ratio at birth in Hong Kong SAR. Population studies. 2013;67(3):323–334.

[15] Sharma O, Haub C. Sex ratio at birth begins to improve in India. Population Reference Bureau. 2008;Available at http://www.prb.org/Publications/Articles/2008/indiasexratio.aspx.

[16] Park CB, Cho NH. Consequences of son preference in a low-fertility society: imbalance of the sex ratio at birth in Korea. Population and Development Review. 1995;p. 59–84. Available at http://www.jstor.org/discover/10.2307/2137413?uid=3738992&uid=2&uid=4&sid=21104249494547.

[17] Chua SP. Sex Ratio at Birth among Chinese Malaysians, 1963–2003. SOJOURN: Journal of Social Issues in Southeast Asia. 2014;29(1):184–194.

[18] Graham E. Son preference, Female Deficit and Singapore's Fertility Transition. Watering the Neighbour's Garden: The Growing Demographic Female Deficit in Asia, Paris: Committee for International Cooperation in National Research in Demography. 2007;p. 89–106.

[19] Lin Tc. The decline of son preference and rise of gender indifference in Taiwan since 1990. Demographic research. 2009;20:377.

[20] Guilmoto CZ, Hoàng X, Van TN. Recent increase in sex ratio at birth in Viet Nam. PLoS One. 2009;4(2):e4624.

[21] Lindley DV, Smith AFM. Bayes Estimates for the Linear Model. 1972;34:1–41.

[22] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. 2nd ed. Boca Raton, Fl.: Chapman & Hall/CRC; 2004.

[23] ICF International. Demographic and Health Survey Sampling and Household Listing Manual. Calverton, Maryland, U.S.A.; 2012. ISBN 3-900051-07-0. Available from: `http://www.R-project.org/`.