Report: male TFR

Henrik-Alexander Schubert

2025-10-03

Male Fertility approximation

In the paper by Keilman et al. (2014), the male fertility rates is approximated throught a regression model that uses the TFR for women and the sex ratio in the population aged 20 to 39 as predictors.

$$log(TFR_m) = \alpha + \beta_1 log(TFR_w) - \beta_2 log(SR_{20-39}) + \epsilon$$

The authors obtained the following expression after rearranging, because the two regression coefficients β_1 and β_2 were not statistically significant.

$$TFR_m = 0.971 \cdot \frac{TFR_w}{SR_{20-39}}$$

The re-estimation

I have re-estimated the male TFR approximation using male fertility data from the Human Fertility Collection (Dudel, 2021), Robert Schoen (1985), a survey-based estiamte by Schoumaker (2019), and my own subnational male fertility collection (Schubert, 2025). There are a few differences to those in Keilman et al. (2014): 1. The coefficients are statistically significant: This might be related to the larger sample size (`r`) and/or the better fit (R-squared=`r`). Therefore, we cannot use the reduced version, but need the complete regression equation for the approximation. 2. Better fit: As mentioned the before, the fit of the regression model bettern than in the original publication. Our R^2 is at 0.967, while the R^2 in the original publication was only 0.83.

The regression results are the following:

```
##
## Call:
## lm(formula = log(tfr_male) ~ log(tfr_female) + log(asr), data = fert_global)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18069 -0.04411 -0.00745  0.03433  1.67170
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.095978   0.002813  -34.12   <2e-16 ***
## log(tfr_female)  1.186982   0.003747  316.77   <2e-16 ***
## log(asr)        -0.887622   0.019814  -44.80   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06941 on 4392 degrees of freedom
## Multiple R-squared:  0.9666, Adjusted R-squared:  0.9666
```

```
## F-statistic: 6.363e+04 on 2 and 4392 DF,  p-value: < 2.2e-16
```

$$log(TFR_m) = -0.096 + 1.187 \cdot log(TFR_w) - 0.88 \cdot SR_{20-29}$$

The regression is visiualized in the figure below:

Standardization

Standardization is a classical demographic tool to emphasize the impact of a difference in population structure on aggregate demographic measures (see, Preston et al. 2001). In this case, we use the method to show the impact of the difference in the male population structure to the female population structure on fertility rates. The estimation looks as follows

$$TFR_m = \sum_{x=15}^{49} \frac{B_f(x)}{P_m(x)}$$

Thus, this standardization takes the distribution of births by age of mother and applies it to the male population structure. This implicitly assumes that men and women have the same fertility schedule, which is not true (see Paget et al. 1991, Schoumaker 2019). Normally, men have a fertility schedule that is shifted to later ages, have a wider reproductive period into older ages, and have a more gradual decline of fertility rates after the age mode of childbearing. *Nevertheless, the standardization reveals the impact of male skewed sex ratios at reproductive age. Below is the result measured in relative difference between male and female TFR estiamted by $diff = \frac{TFR_m - TFR_w}{TFR_w}$.

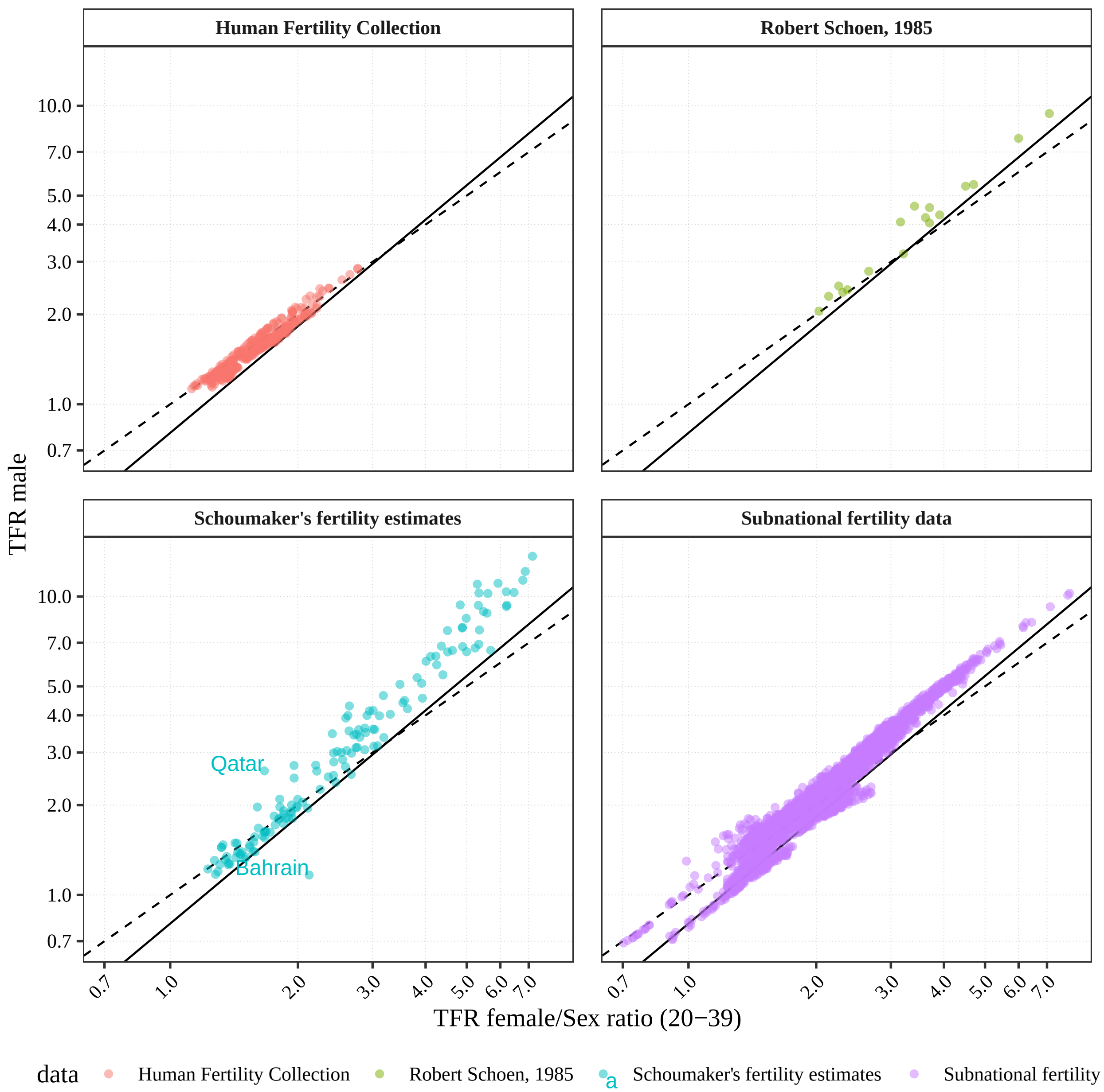

Figure 1: Relationship between female TFR and male TFR. Dashed line is the estimated regression line from the male TFR approximation. Solid line is a simple diagonal reflecting unity between the TFR for women and men.

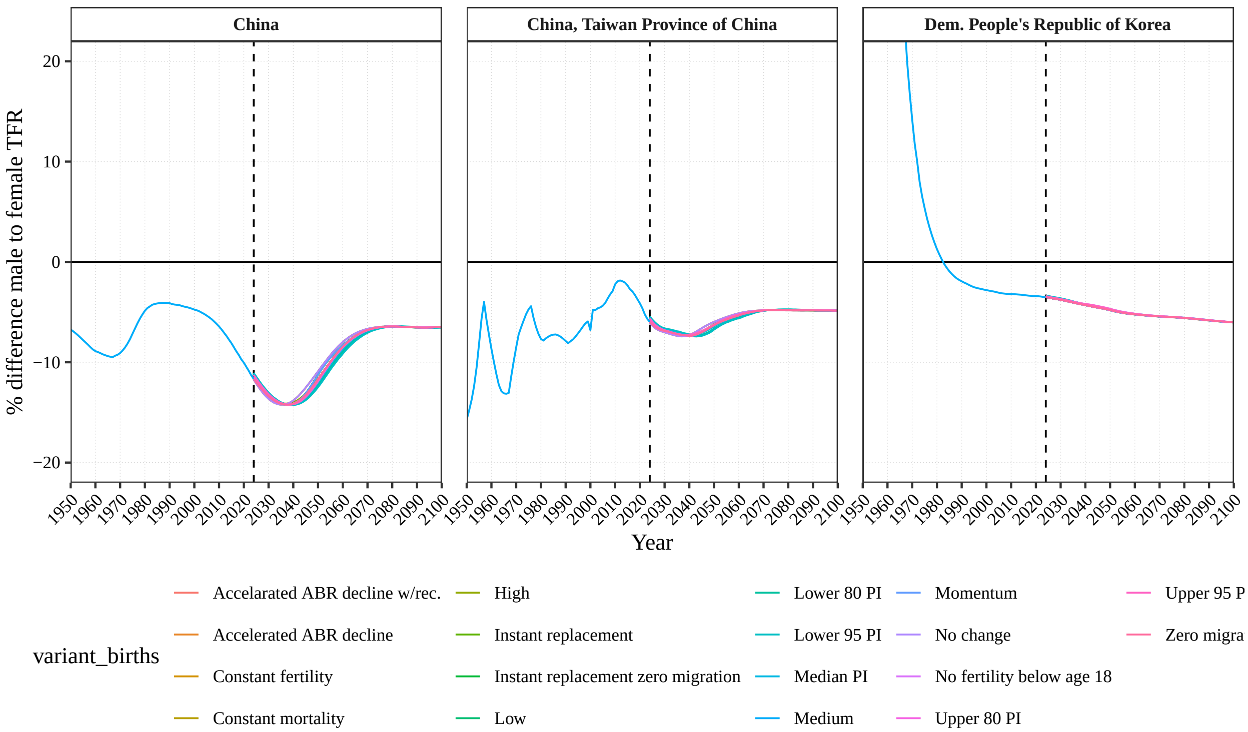

Figure 2: Relative difference in the male TFR to the female TFR over time using the standardization procedure.