Male-Female Fertility Differences

A stable population approach

Henrik-Alexander Schubert
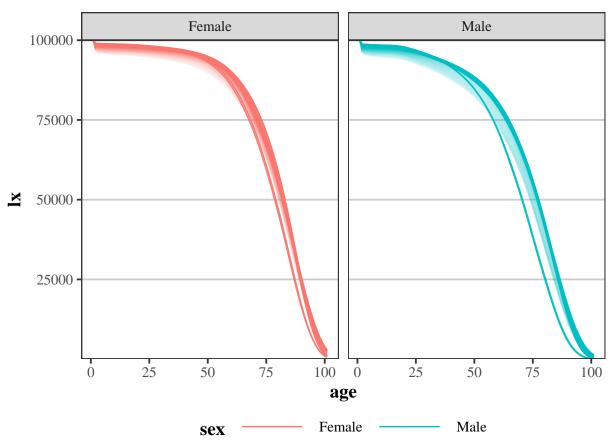
2023-06-02

Abstract

In 2019, a paper by Bruno Schoumaker was published on male fertility around the world. The key messages of the paper are that 1) the fertility of men can differ substantially from the fertility of women, 2) that the lower the fertility level, the more exceeds the male fertiltiy the levels for females, and 3) that the ratio of male-to-female TFR can be approximated in stable-population estimation. This file uses the stable populatin theory proposed by Schoumaker and examines the relationship between male and female fertiltiy for Mexico. Harnessing vital statistics data from INEGI and mortality data from the United Nations, this project analyses the stable-population approximation of the male-to-female fertilitiy ratio.

Data wrangling

First, the data needs to be loaded. Harnessing the United Nations World Population Prospects from 2023, which provides age-specific death rates. The data can be accessed under United Nations.

```r
# Load the mortality data
mortality <- fread("Raw/mortality_mexico_un.csv")

# Clean the names
mortality <- mortality %>%
  clean_names() %>%
  select(indicator_short_name, time, author, location, sex, age, value) %>%
  mutate(age = as.integer(str_remove_all(age, "\\+")))

# Split the data
mx <- split(mortality$value,list(mortality$time, mortality$sex))

# Estimate the lifetables
lifetables <- lapply(mx, lifetable, sex = "M")

# Combine the lifetables
lifetables <- bind_rows(lifetables, .id = "id")

# Get the names
lifetables <- lifetables %>%
  mutate(sex = as.factor(str_split(id, pattern = "\\.", simplify = T)[, 2]),
                         year  = as.integer(str_split(id, pattern = "\\.", simplify = T)[, 1])) %>%
  select(-id)


# Plot the lifetable survivors for men and women over time
ggplot(lifetables, aes(age, lx, alpha = year / max(year), colour = sex, group = year)) +
  geom_line() +
```

```
facet_wrap(~ sex) +
scale_y_continuous(expand = c(0, 0)) +
guides(alpha = "none")
```


Schoumaker paper

In his paper, Schoumaker (2019) describes male fertility around the world. The key messages of the paper are that 1) fertility of men differed from the fertility of women, 2) that the lower the fertility level, the larger the male fertility exceeds the levels for females, and 3) that the ratio of male-to-female TFR can be approximated in stable-population estimation.

In his paper, he approximates the male-to-female fertility ratio using *stable population theory*.

$$\frac{TFR_m}{TFR_f} = \frac{1}{SRB} \cdot \frac{p(MAC)_m}{p(MAC)_f} \cdot exp(r \cdot (T_m - T_f))$$

Where SRB is the sex ratio at birth, which can be approximated by 105, $p(MAC)$ is the probability to survive until the mean age of childbearing, r is the population growth rate, and T is the mean generation length. The subsequent sections show the estimation of the different components and describes also the data.

The estimation can be transformed into the following function.

```
# Schoumaker function
schoumaker <- function(SRB = 1.05, pMAC, pMAF, r = 1, TM, TF){
  tmp <- 1 / SRB * pMAC / pMAF * exp(r * (TM - TF))
  return(tmp)
}
```
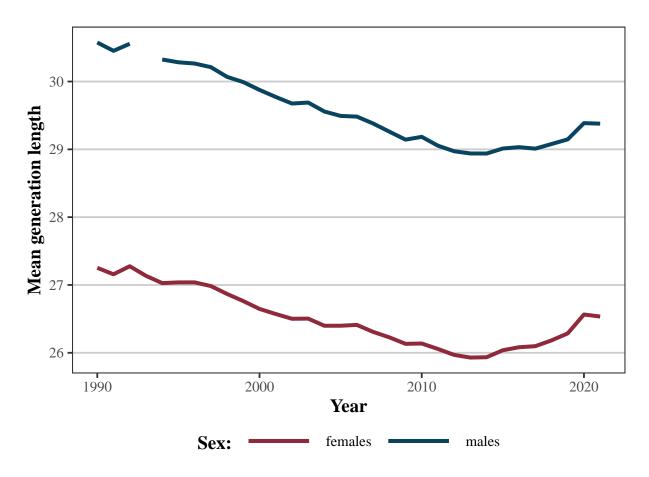
Estimate the mean generation length

In order to estimate the mean generation length, we use the formula from (Keyfitz and Caswell (2005), p. 264):

$$\mu_1 = \frac{\sum_x x(\prod_{j=1}^{x-1} P_j)F_x}{\sum_x(\prod_{j=1}^{x-1} P_j)F_x}$$

Which can be transformed into the following function `generation_length`:

```
# Estimate the mean lenght of generation
generation_length <- function(x, px, asfr){
  gen_len <- sum(x * px * asfr) / sum(px * asfr)
  return(gen_len)
}



# Estimate it for the lifetables
px <- lifetables %>% select(age, px, sex, year)

# Merge with the fertility rates
load("Data/asfr_national_mexico.Rda")
load("Data/tfr_national_mexico.Rda")

# Pivot wider
px <- pivot_wider(px, values_from = "px", names_from = "sex")

# JOin the data
mor_fer <- inner_join(px, asfr_national, by = c("year", "age" = "age_mot"))

# Estimate the generation length
gen_length <- mor_fer %>%
  group_by(year) %>%
  summarise(gen_len_m = sum(age * cumprod(Male) * asfr_m) / sum(cumprod(Male) * asfr_m),
            gen_len_f = sum(age * cumprod(Female) * asfr_f) / sum(cumprod(Female) * asfr_f))

# Plot the generation length
ggplot(gen_length, aes(x = year)) +
  geom_line(aes(y = gen_len_f, colour = "females"), linewidth = 1.4) +
  geom_line(aes(y = gen_len_m, colour = "males"), linewidth = 1.4) +
  scale_colour_manual(name = "Sex:", values = c(MPIDRred, MPIDRblue)) +
  ylab("Mean generation length") + xlab("Year")
```
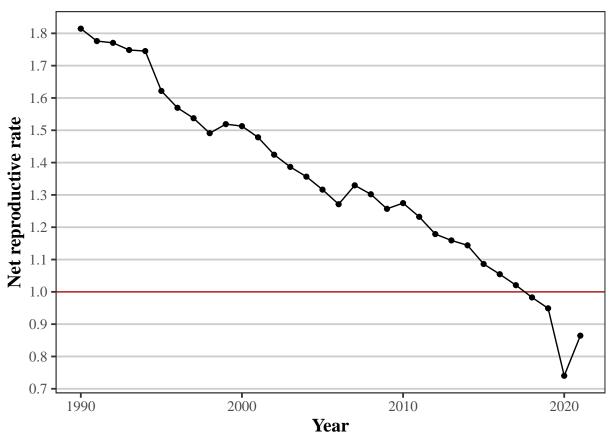
Estimate the growth rate

One part of the stable population theory is the population growth rate: r. In this case, we estimate the *Net Reproduction Rate*, which is defined as the number of offspring a person is producing during the lifetime. The formula is:

$$r = \sum_{x=1}^{50} f(x)_f \cdot \frac{L(x)}{l_0}$$

Here, $f(x)_f$ is the number of female births a women at age x produces on average. We approximate this value by multiplying the age-specific fertility rate $f(x)$ with the share of female births, which is 0.4886.

```
# Get the lifetables
lifetables_f <- lifetables %>% filter(sex == "Female")

# Join with fertility data
mor_fer <- inner_join(lifetables_f,asfr_national, by = c("year", "age" = "age_mot"))

# Estimate the growth rate
growth_rate <- mor_fer %>% filter(year %in% 1990:2021) %>%
  mutate(asfr_f = if_else(is.na(asfr_f), 0, asfr_f)) %>%
                                group_by(year) %>%
          summarise(r = sum((asfr_f*0.4886)*Lx/100000))

# Plot the growth rate
ggplot(growth_rate, aes(year, r)) +
```

```
geom_hline(yintercept = 1, colour = "firebrick") +
geom_line() +
geom_point() +
scale_y_continuous(n.breaks = 10) +
ylab("Net reproductive rate") + xlab("Year")
```
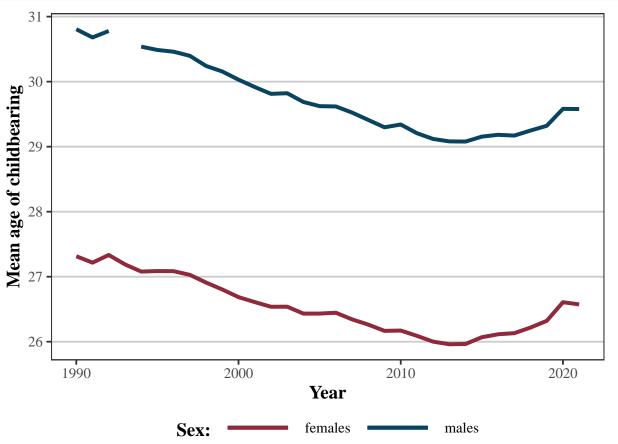

Mean age of childbearing

The last component is the *mean age of childbearing* for men and for women. According to Preston, Heuveline, and Guillot (2008), it can be estimated by:

$$A = \frac{\sum_{x=12}^{50} x \cdot f(x)}{\sum_{x=12}^{50} f(x)}$$

In R, we estimate the *mean age of childbearing* in the following way:

```
# Join the data
mean_age <- asfr_national %>% group_by(year) %>%
summarise(mac_f = sum(age_mot * asfr_f)/ sum(asfr_f),
          mac_m = sum(age_mot * asfr_m)/ sum(asfr_m),
          difference =  mac_m - mac_f)

# Plot the development of the mean age of childbering
ggplot(mean_age, aes(x = year)) +
  geom_line(aes(y = mac_f, colour = "females"), linewidth = 1.4) +
  geom_line(aes(y = mac_m, colour = "males"), linewidth = 1.4) +
```

```
  scale_colour_manual(name = "Sex:", values = c(MPIDRred, MPIDRblue)) +
  ylab("Mean age of childbearing") + xlab("Year")
```


We estimated the mean age of childbearing in order to obtain the probability of living until the that age. Thus, in the next step, we combine life table estimates with the mean age.

```
# Combine the data
surv_mean_age <- inner_join(lifetables, mean_age, by = c("year")) %>%
  mutate(across(c(mac_f, mac_m), round)) %>%
  filter((age == mac_f & sex == "Female") | (age == mac_f & sex == "Male")) %>%
  select(px, year, age, sex) %>% pivot_wider(names_from = "sex", values_from = "px", names_prefix =   "s
```

Estimation

Now, that we have all components to estimate the male-to-female TFR ratio in a stable population, we just insert all components. Figure (**figure-ratio?**) shows the time-series for the observed male-to–female TFR ratio against the value from stable population theory. We see that the trends are similar in terms of direction, as both curves decline with time. However, the level of the two curves is very different. The observed ratio starts in 1990 with a value of 1.11 and declines subsequently towards 1.01. Thus, throughout time, the male TFR exceeds the female TFR. In contrast, for the time series derived from stable population theory, the ratio is in year 1990 above 3 and declines almost to 0.

```
# Combine all the data
head(surv_mean_age) %>% pander()
```
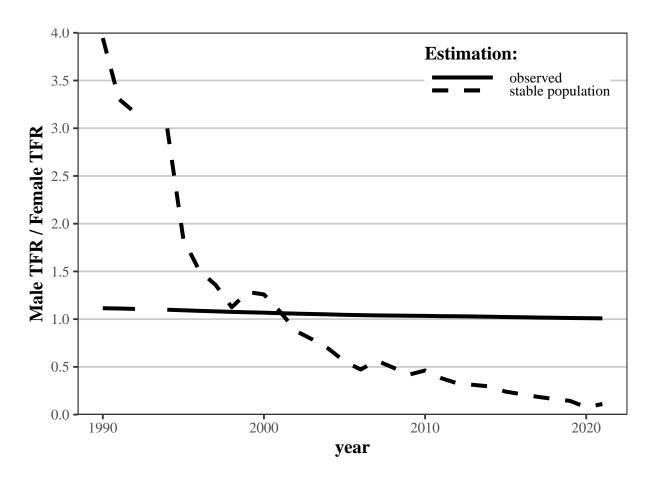
year	age	surv_Female	surv_Male
1990	27	0.999	0.9969
1991	27	0.999	0.9969
1992	27	0.9991	0.997
1993	27	0.9991	0.9971
1994	27	0.9992	0.9971
1995	27	0.9992	0.9972

```
head(growth_rate) %>% pander()
```

year	r
1990	1.815
1991	1.776
1992	1.771
1993	1.748
1994	1.745
1995	1.622

```
head(gen_length) %>% pander()
```
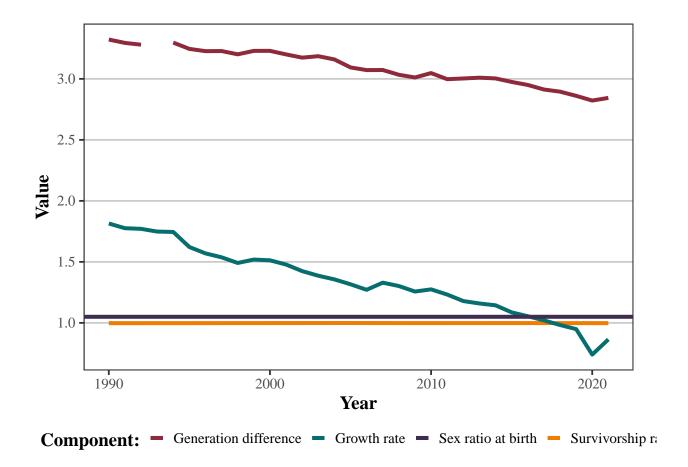
year	gen_len_m	gen_len_f
1990	30.58	27.25
1991	30.45	27.16
1992	30.56	27.28
1993	NA	27.14
1994	30.32	27.03
1995	30.28	27.04

```
# Merge the data
data <- inner_join(surv_mean_age, growth_rate, by = "year") %>%
  inner_join(., gen_length, by = "year")

# Estimate Schoumakers ratio
data <- data %>% mutate(survivor_ratio = surv_Male / surv_Female,
              gen_diff = gen_len_m - gen_len_f,
              male_female_tfr = 1 / 105 * survivor_ratio * exp(r * (gen_diff)))

# Plot the result
ggplot(data, aes(x = year, y = male_female_tfr)) +
  geom_line(aes(linetype = "stable population"), linewidth = 1.4) +
  geom_line(data = tfr_national, aes(y = TFR_m / TFR_f, linetype = "observed"), linewidth = 1.4) +
  scale_y_continuous(expand = c(0, 0), n.breaks = 10, limits = c(0, 4)) +
  scale_linetype_manual(name = "Estimation:", values = c("solid", "dashed")) +
  theme(legend.position = c(0.8, 0.9)) +
  ylab("Male TFR / Female TFR")
```

Contribution of components

Now, we are analyzing the different components behind the decline of the male to female TFR ratio. Figure (**figure-components?**) illustrates the different components of the stable population model of the *male-to-female TFR ratio*. We find that the trend is driven by the population growth rate and by the generation difference. The generation length difference between men and women is declining over time, which leads to a declining male TFR to female TFR ratio. Moreover, and even larger, the population growth rate is declining, which translates into a declining male TFR to female TFR ratio.

```
# Plot the components
ggplot(data, aes(x = year)) +
 geom_line(aes(y = gen_diff, colour = "Generation difference"), linewidth = 1.4) +
  geom_line(aes(y = survivor_ratio, colour = "Survivorship ratio"), linewidth = 1.4) +
  geom_line(aes(y = r, colour = "Growth rate"), linewidth = 1.4) +
  geom_hline(aes(yintercept = 1.05, colour = "Sex ratio at birth"), linewidth = 1.4) +
  scale_colour_manual(name = "Component:", values = c(MPIDRred, MPIDRgreen, MPIDRpurple,  MPIDRorange))
  theme(legend.key.width = unit(0.4, "cm")) +
  ylab("Value") + xlab("Year")
```

Component: — Generation difference — Growth rate — Sex ratio at birth — Survivorship r[...]

References

Keyfitz, Nathan, and Hal Caswell. 2005. *Applied Mathematical Demography.* 3. ed. Statistics for Biology and Health. New York, NY: Springer.

Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2008. *Demography: Measuring and Modeling Population Processes.* 9. [pr.]. Oxford: Blackwell.

Schoumaker, Bruno. 2019. "Male Fertility Around the World and Over Time: How Different Is It from Female Fertility?" *Population and Development Review* 45 (3): 459–87. https://doi.org/10.1111/padr.12273.