

Male-Female Differences

Henrik Schubert

2023-06-05

Abstract

This project aims to analyze and compare male and female fertility rates in Mexico using various statistical tools and techniques. The research focuses on both temporal trends and national-level variations in fertility rates. The study employs graphical tools, distributional comparison measures, decomposition methods, and statistical tests to provide a comprehensive understanding of the differences between male and female fertility rates. By examining fertility rates over time, the project seeks to identify any significant changes and patterns that may have occurred. Furthermore, by conducting national-level comparisons, the study aims to shed light on potential disparities or similarities between male and female fertility rates across different regions of Mexico. The research utilizes graphical tools to visualize the data, enabling a visual exploration of the trends and patterns in male and female fertility rates. Distributional comparison measures are employed to quantitatively assess the similarities and differences between the two groups. Additionally, decomposition methods are used to understand the underlying factors contributing to the observed variations in fertility rates. Statistical tests will be conducted to determine the statistical significance of the observed differences between male and female fertility rates. These tests will provide a rigorous evaluation of the findings, enabling meaningful conclusions to be drawn from the data. The project's findings will contribute to the existing body of knowledge on fertility rates in Mexico and provide valuable insights into gender-specific patterns and variations.

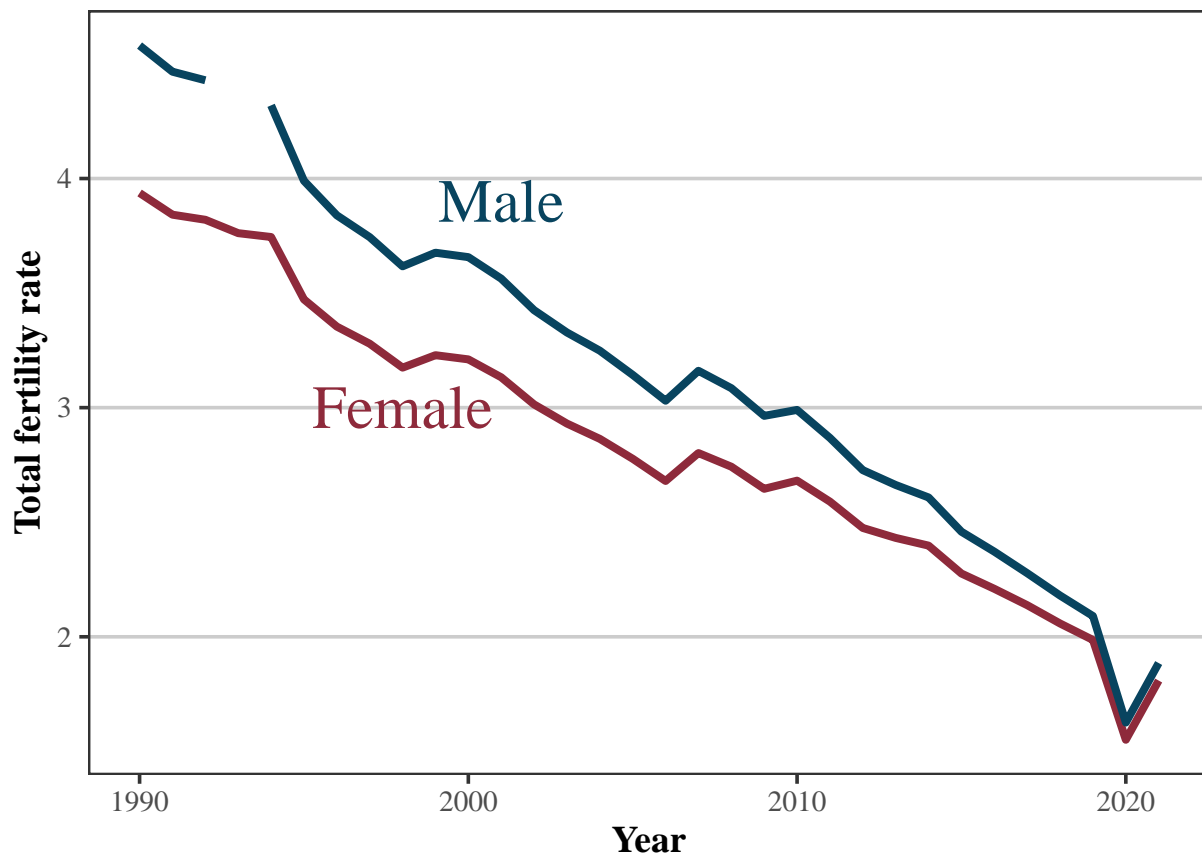
Contents

Time trend	1
Age-Specific Differences	2
Decomposing the difference in the crude fertility rate	5
Regression based methods	10
Generalized additive models (GAMs)	10
Demographic modeling	10
References	12

Time trend

First, we compare the time-trend of male and female total fertility rates in Mexico. We use the estimates based on vital statistics data from `INEGI` and from `the demographic group`.

```
# What is the trend in the national tfr s
ggplot(tfr_nat, aes(year)) +
  geom_line(aes(y = tfr_f, colour = "female"), linewidth = 1.4) +
  geom_line(aes(y = tfr_m, colour = "male"), linewidth = 1.4) +
  annotate(geom = "text",
    x = 2001, y = 3.9,
    label = "Male",
    colour = MPIDRblue, size = 8, family = "serif") +
  annotate(geom = "text",
    x = 1998, y = 3,
    label = "Female",
    colour = MPIDRred, size = 8, family = "serif") +
  scale_colour_manual(values = c(MPIDRred, MPIDRblue)) +
  ylab("Total fertility rate") +
  xlab("Year") +
  guides(colour = "none")
```



The graph shows several time trends. First, both male and female total fertility rates declined between 1990 and 2021 monolithically. Second, the male TFR declined stronger than the female TFR, which led to a closing of the sex difference in total fertility rates.

Age-Specific Differences

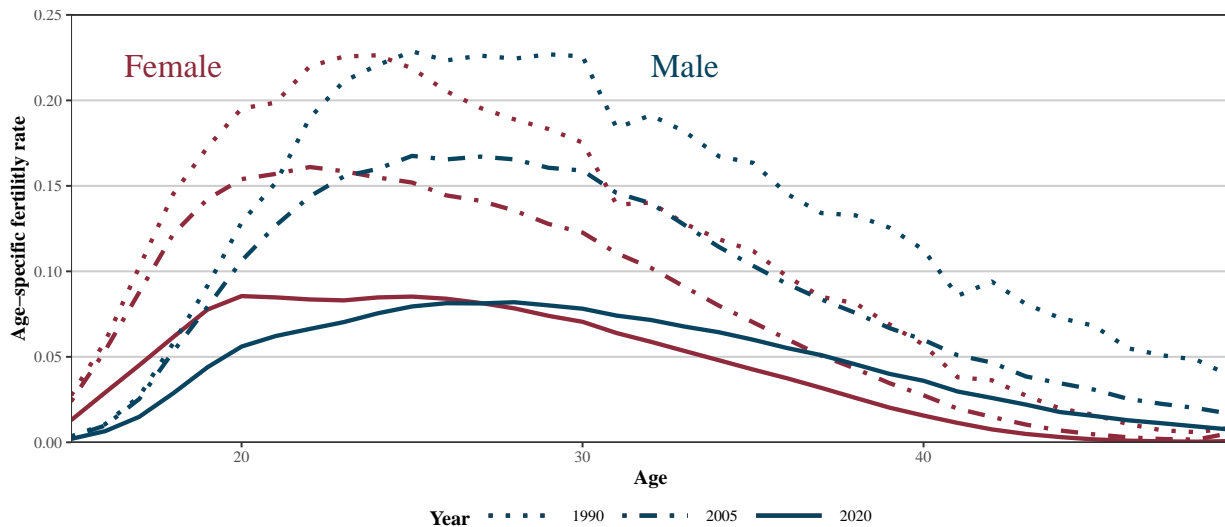
Now, we plot the age-specific differences in fertility between men and women.

```

# Load the national level data
load("Data/asfr_national_mexico.Rda")

# What is the trend in the national tfr s
ggplot(subset(asfr_nat, year %in% seq(1990, 2020, by = 15)), aes(x = age, linetype = as.factor(year))) +
  geom_line(aes(y = asfr_f, group = year),
            linewidth = 1.4, colour = MPIDRred) +
  geom_line(aes(y = asfr_m, group = year),
            linewidth = 1.4, colour = MPIDRblue) +
  annotate(geom = "text",
          x = 18, y = 0.22,
          label = "Female",
          colour = MPIDRred, size = 8, family = "serif") +
  annotate(geom = "text",
          x = 33, y = 0.22,
          label = "Male",
          colour = MPIDRblue, size = 8, family = "serif") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 0.25)) +
  scale_x_continuous(expand = c(0, 0)) +
  scale_linetype_manual(name = "Year", values = c("dotted", "dotdash", "solid")) +
  ylab("Age-specific fertility rate") +
  xlab("Age") +
  guides(colour = "none", alpha = "none")

```



We find in contrast to other countries, that the fertility decline is not largely driven by postponement, but rather by a decline in age-specific fertility rates across all age-groups. A fertility recuperation or a shift of childbearing at a later point is absent in Mexico. Regarding the differences between men and women, there are some noticeable differences. First, across the all years, childbearing for men takes place later in life. This is apparent in higher female age-specific fertility rates ($f(x)$) at younger ages and higher male $f(x)$ at higher ages. This life course behaviour produces a certain cross-over. Second, the spread of childbearing over the life course is wider. For women above age 45, the age-specific fertility rates are almost zero, while for men there is some childbearing at every age. This is consistent with findings from Schoumaker (2019) and Dudel and Klüsener (2019).

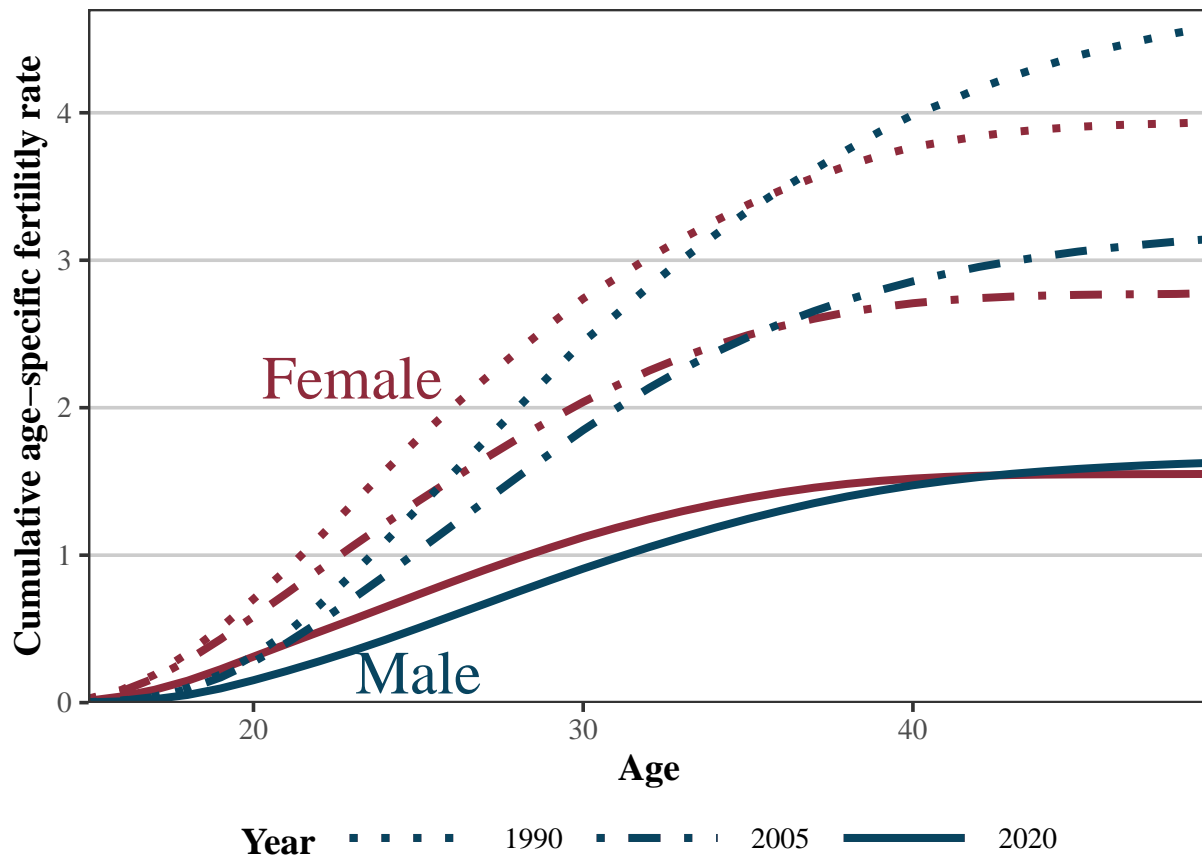
Cumulative distribution

Moreover, to understand better the postponement behaviour, we are going to estimate the cumulative age-specific fertility rate.

$$cumulativef(x) = \sum_{12}^x f(x)$$

```
# Estimate the cumulative distribution
asfr_nat <- asfr_nat %>%
  group_by(year) %>%
  arrange(age) %>%
  mutate(cum_f = cumsum(asfr_f),
         cum_m = cumsum(asfr_m))

# What is the trend in the national tfr s
ggplot(subset(asfr_nat, year %in% seq(1990, 2020, by = 15)), aes(x = age, linetype = factor(year))) +
  geom_line(aes(y = cum_f, group = year),
            linewidth = 1.4, colour = MPIDRred) +
  geom_line(aes(y = cum_m, group = year),
            linewidth = 1.4, colour = MPIDRblue) +
  annotate(geom = "text",
          x = 23, y = 2.2,
          label = "Female",
          colour = MPIDRred, size = 8, family = "serif") +
  annotate(geom = "text",
          x = 25, y = 0.2,
          label = "Male",
          colour = MPIDRblue, size = 8, family = "serif") +
  scale_linetype_manual(name = "Year", values = c("dotted", "dotdash", "solid")) +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 4.7)) +
  scale_x_continuous(expand = c(0, 0)) +
  ylab("Cumulative age-specific fertility rate") +
  xlab("Age") +
  guides(colour = "none", alpha = "none")
```



Again, we see that fertility declined over time. The cumulative fertility rate in recent years is much lower than the rates for 2005 and 1990. Moreover, we see that in every year the cross-over behavior of male and female fertility. Moreover, the cross-over for male and female fertility rates occurs in all years. However, the cross-over shifted to later ages. In 1990, the male-female fertility cross-over occurred at age 36, while in 2020, 43.

Decomposing the difference in the crude fertility rate

Another tool to study the difference between the male and female is *decomposition*. The crude fertility rate can be decomposed into two components: *population structure* and *birth rates*. Because of the additive nature, it is feasible to use standard demographic decomposition methods (Preston, Heuveline, and Guillot (2008), p. 28)

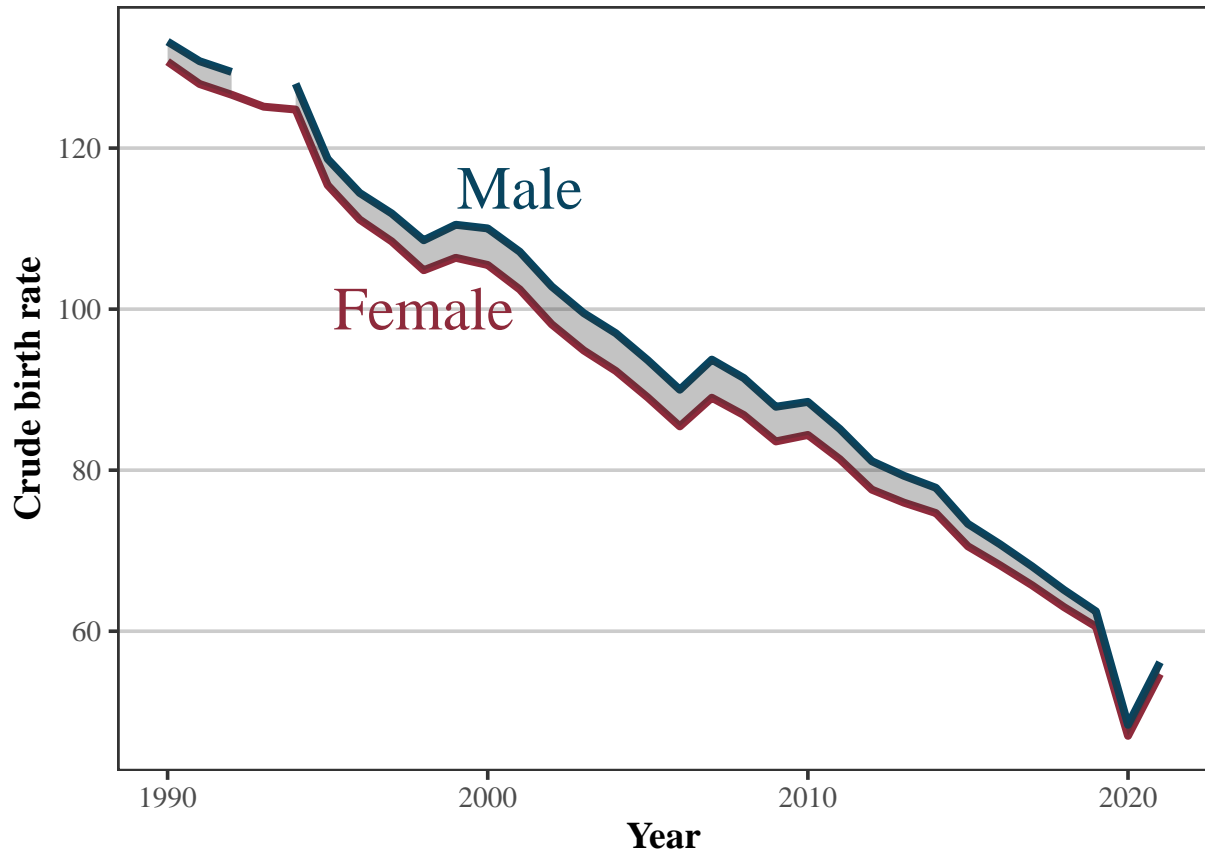
```
# Plot the crude birth rate for men and women
cbr_nat <- asfr_nat |>
  group_by(year) |>
  summarise(cbr_f = 1000 * sum(births_f) / sum(exposure_f),
            cbr_m = 1000 * sum(births_m) / sum(exposure_m))

# Plot the crude birth rates
ggplot(data = cbr_nat, aes(x = year)) +
  geom_line(aes(y = cbr_f, colour = "female"), linewidth = 1.4) +
  geom_line(aes(y = cbr_m, colour = "male"), linewidth = 1.4) +
  geom_ribbon(aes(ymin = cbr_f, ymax = cbr_m), alpha = 0.3) +
  annotate(geom = "text",
```

```

    x = 2001, y = 115,
    label = "Male",
    colour = MPIDRblue, size = 8, family = "serif") +
  annotate(geom = "text",
    x = 1998, y = 100,
    label = "Female",
    colour = MPIDRred, size = 8, family = "serif") +
  scale_colour_manual(values = c(MPIDRred, MPIDRblue)) +
  ylab("Crude birth rate") +
  xlab("Year") +
  guides(colour = "none")

```



We see, that similar as for the total fertility rate, the gap between male and female crude fertility rates has diminished. The decomposition approach looks as follows:

$$\Delta CFR_{f,m} = \underbrace{\sum_i (C_i^f - C_i^m) \left[\frac{f_i^f + f_i^m}{2} \right]}_{\Delta Population} + \underbrace{\sum_i (f_i^f - f_i^m) \cdot \left[\frac{C_i^f + C_i^m}{2} \right]}_{\Delta Rate}$$

where f is the age-specific fertility rate, C is the population share, i indexes the different age-groups, and f and m are indicating the male or female part. The first term marks the weighted change in the rate ($\Delta Rate$) and the second term marks the weighted change in population ($\Delta Population$).

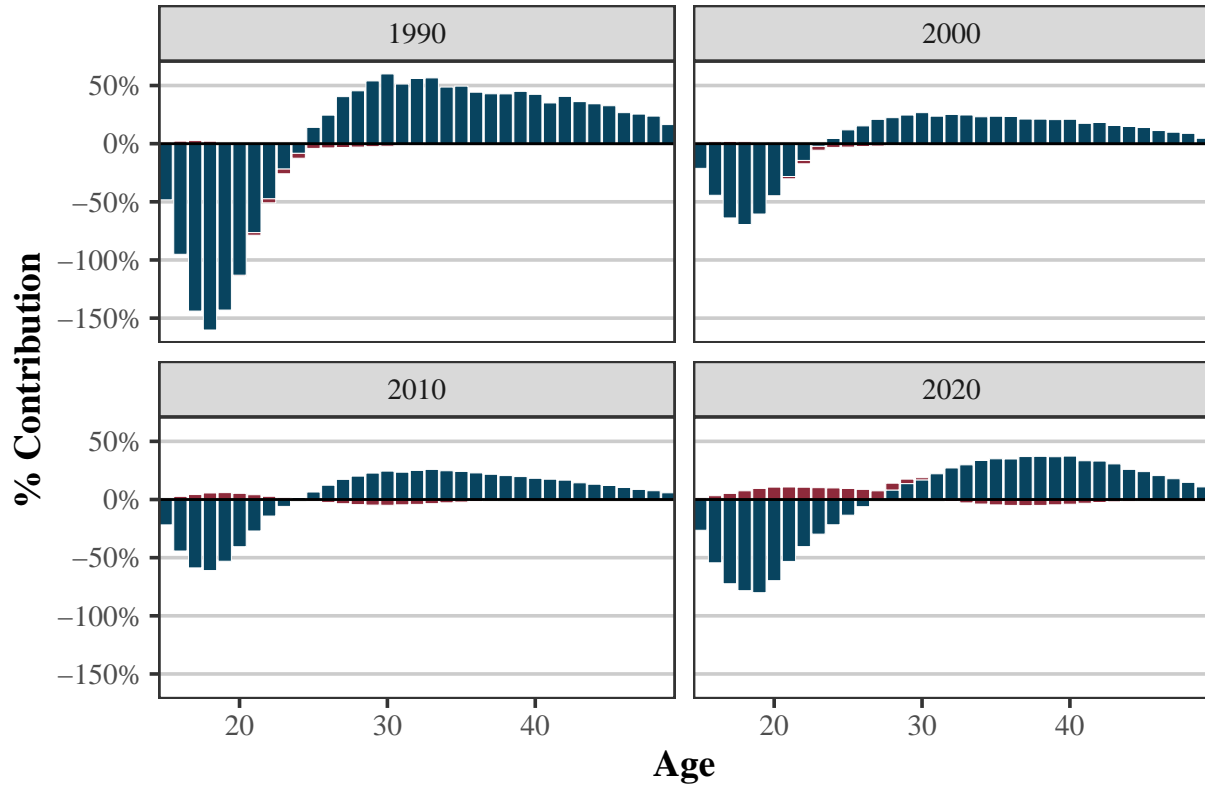
```

# Estimate the decomposition
comp_asfr <- asfr_nat |>
  group_by(year) |>
  mutate(across(starts_with("asfr"), ~ .x * 1000)) |>
  mutate(pop_share_f = pop_share(exposure_f),
         pop_share_m = pop_share(exposure_m),
         mean_pop    = averaging(pop_share_f, pop_share_m),
         mean_rate    = averaging(asfr_f, asfr_m),
         delta_rate    = (asfr_f - asfr_m) * mean_pop,
         delta_pop     = (pop_share_f - pop_share_m) * mean_rate,
         .groups       = "drop")

# Make the decomposition
decomp_nat <- comp_asfr %>%
  select(year, age, delta_rate, delta_pop) %>%
  pivot_longer(cols = starts_with("delta"),
               names_to = "component",
               values_to = "contribution") |>
  inner_join( cbr_nat, by = "year") |>
  mutate(contribution = contribution / (cbr_f - cbr_m))

# Plot the decomposition
decomp_nat %>%
  filter(year %in% c(1990, 2000, 2010, 2020)) %>%
  ggplot(aes(age, contribution, fill = component)) +
  geom_col(colour = "white", linewidth = 0.01) +
  geom_hline(yintercept = 0, colour = "black") +
  facet_wrap(~ year) +
  scale_fill_manual(values = c(MPIDRred, MPIDRblue),
                    labels = c(expression(paste(Delta, "Population")),
                               expression(paste(Delta, "Rate"))),
                    name = "Component: ") +
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(labels = scales::percent) +
  ylab("% Contribution") + xlab("Age") +
  theme(legend.key.width = unit(0.3, "cm"),
        legend.key.height = unit(0.2, "cm"))

```



Component: ■ Δ Population ■ Δ Rate

In the figure, we see that the largest contribution stems from the age-specific rates. This is the reason, why the sex differences exist also for the *total fertility rate*, and not just the crude birth rate. Over time we see that contribution of the rates diminish in contribution to differences.

In order to better understand the size of the contribution, we create tables that display the components with the largest contribution.

```
# Estimate the relative contribution
relative_contribution <- decomp_nat |>
  filter(year == 1990) |>
  ungroup() |>
  arrange(contribution) |>
  mutate(relative_contribution = paste(round(100 * contribution, 2), "%"),
         diff = cbr_f - cbr_m)

# Create a table for the relative contributions
relative_contribution |>
  select(component, age, relative_contribution, cbr_f, cbr_m, diff) |>
  slice_head(n = 5) |>
  pander(col.names = c("Component", "Age", "Contribution (%)", "CBR (females)", "CBR (males)", "Difference"))
```

Component	Age	Contribution (%)	CBR (females)	CBR (males)	Difference
delta_rate	18	-160.21 %	130.7	133.2	-2.483

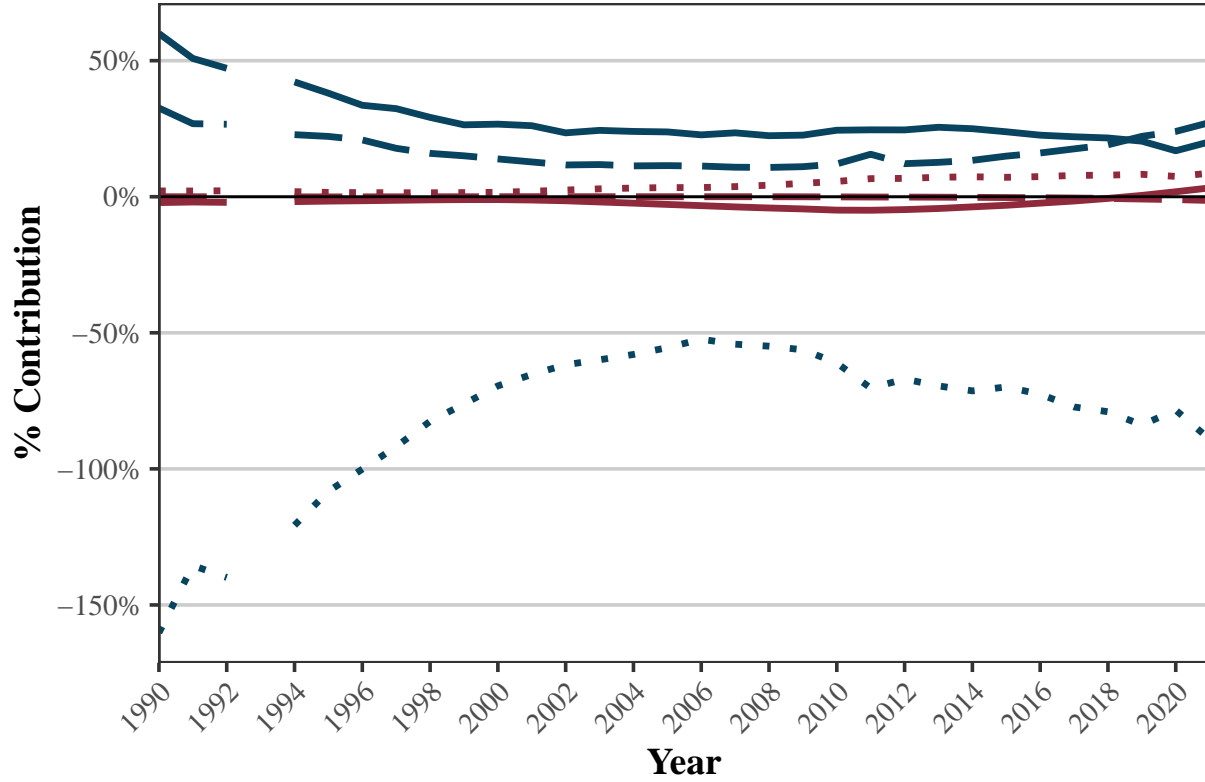
Component	Age	Contribution (%)	CBR (females)	CBR (males)	Difference
delta_rate	17	-143.93 %	130.7	133.2	-2.483
delta_rate	19	-143.07 %	130.7	133.2	-2.483
delta_rate	20	-113.2 %	130.7	133.2	-2.483
delta_rate	16	-95.11 %	130.7	133.2	-2.483

```
# Cumulative contribution of components
relative_contribution |>
  group_by(year, component) |>
  summarise(contribution = sum(contribution),
            cbr_m = unique(cbr_m),
            cbr_f = unique(cbr_f),
            diff = unique(diff),
            .groups = "drop") |>
  filter(year %in% seq(1990, 2020, by = 15)) |>
  pivot_wider(names_from = "component", values_from = "contribution") |>
  pander(col.names = c("Year", "CBR (men)", "CBR (women)", "Difference", "Population ", "Rate"))
```

Year	CBR (men)	CBR (women)	Difference	Population	Rate
1990	133.2	130.7	-2.483	-0.3095	1.309

Moreover, since we have the additive components for single years, we can also investigate how the different components changed over time in order to better explain the changes in the gap between male and female fertility rate.

```
# Plot the time trend for populatoio estimates
decomp_nat %>%
  filter(age %in% c(18, 30, 45))|>
  ggplot(aes(x = year, contribution, colour = component, group = interaction(component, age), linetype = age)) +
  geom_line(linewidth = 1.2) +
  geom_hline(yintercept = 0, colour = "black") +
  scale_colour_manual(values = c(MPIDRred, MPIDRblue),
                     labels = c(expression(paste(Delta, "Population")),
                                expression(paste(Delta, "Rate"))),
                     name = "Component: ") +
  scale_y_continuous(labels = scales::percent) +
  scale_x_continuous(expand = c(0, 0), n.breaks = 20) +
  scale_linetype_manual(name = "Age", values = c("dotted", "solid", "longdash")) +
  theme(legend.key.width = unit(0.5, "cm"),
        legend.key.height = unit(0.2, "cm"),
        axis.text.x = element_text(angle = 45, hjust = 1)) +
  ylab("% Contribution") + xlab("Year")
```



Age · · 18 — 30 — 45 **Component:** — ΔPopulation — ΔRate

We find an U-shape pattern for the rate components. This implies that the effects on the sex-gap in the CBR are larger in the early and late years, compared to the period in the early 2000. For the population component

Regression based methods

Generalized additive models (GAMs)

Demographic modeling

([paget_relational_1994?](#)) developed a relational gompertz model for male fertility. It combines theoretical and empirical approaches. The model is based on a standard pattern of fertility. The observed fertility is assumed to be linearly related to the standard, rather than to age. The model contains two parameters α and β , which determine the location and dispersion of the fertility distribution.

This fact is particularly usefull, since male fertility is more dispersed than female fertility.

Estimation

1. A *stretch factor* is estimated. Which is the ratio of the male reproductive span to the female reproductive span.

$$stretch - factor = \frac{(Age_{max}^m - Age_{min}^m)}{(Age_{max}^f - Age_{min}^f)}$$

2. The female ages are transformed, while the the original $Y_s(x)$ remains attached.

$$Age^m = age^f \cdot stretch - factor$$

3. The stretched values of $Y_s(x)$ are then converted back to data for whole years by polynomial interpolation on the Gompertz scale.

```
# 1. Stretch the age factor
stretch_factor <- (59 - 15) / (49 - 15)

# 2. Stretch the ages for women
asfr_nat <- asfr_nat |>
  mutate(age_stretched = age * stretch_factor)

# 3. Fit the polynomial
model_data <- asfr_nat |>
  group_by(year) |>
  nest()

# Run the regressions
predictions <- model_data |>
  mutate(polyModel = map(data, ~ lm(data = ., asfr_f ~ age_stretched + age_stretched^2)),
    coefficients = map(polyModel, coefficients)) |>
  select(year, coefficients) |>
  unnest(cols = c(coefficients)) |>
  mutate(nr = row_number(), .groups = year) |>
  pivot_wider(names_from = "nr", values_from = "coefficients")

#
predict_funct <- function(coeff1, coeff2){
  tmp <- 15:59 * coeff1 + 15:59 *coeff2^2
  return(tmp)
}

# Estimate the predictions
predictions |>
  mutate(predictions = map(`1`, `2`, ~ predict_funct(coeff1 = .x, coeff2 = .y)))
```

```
## # A tibble: 32 x 5
## # Groups:   year [32]
##   year .groups   '1'      '2' predictions
##   <int> <int> <dbl>    <dbl> <named list>
## 1 1990   1990 0.278 -0.00399 <NULL>
## 2 1991   1991 0.275 -0.00398 <NULL>
## 3 1992   1992 0.269 -0.00386 <NULL>
## 4 1993   1993 0.270 -0.00392 <NULL>
## 5 1994   1994 0.272 -0.00399 <NULL>
## 6 1995   1995 0.252 -0.00369 <NULL>
## 7 1996   1996 0.244 -0.00357 <NULL>
## 8 1997   1997 0.240 -0.00353 <NULL>
```

```
## 9 1998 1998 0.236 -0.00350 <NULL>
## 10 1999 1999 0.243 -0.00363 <NULL>
## # i 22 more rows
```

References

- Dudel, Christian, and Sebastian Klüsener. 2019. “Estimating Men’s Fertility from Vital Registration Data with Missing Values.” *Population Studies* 73 (3): 439–49. <https://doi.org/10.1080/00324728.2018.1481992>.
- 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>.