Homework 04

Ihsan Kahveci

2022-05-13

Contents

0.1 *Q3* 2
0.1.1 *Q3.a* 2
0.1.2 *Q3.b* 4
0.1.3 *Q3.c* 5

0.1 *Q3*

Obtain the values of TFR for the Netherlands for 1950-2020 from the 2019 World Population Prospects.

```
# Question 3 ------------------------------------------------------------------
nld_tfr <- tfr_all %>% filter(name == "Netherlands") %>% select(-name)

knitr::kable(
  nld_tfr,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Period Start", "TFR"),
  caption = "Total fertility rates, Netherlands, 1950-2020")
```

Table 1: Total fertility rates, Netherlands, 1950-2020

Period Start	TFR
1950	3.052
1955	3.097
1960	3.166
1965	2.795
1970	2.100
1975	1.598
1980	1.515
1985	1.555
1990	1.592
1995	1.599
2000	1.740
2005	1.746
2010	1.732
2015	1.660

0.1.1 *Q3.a*

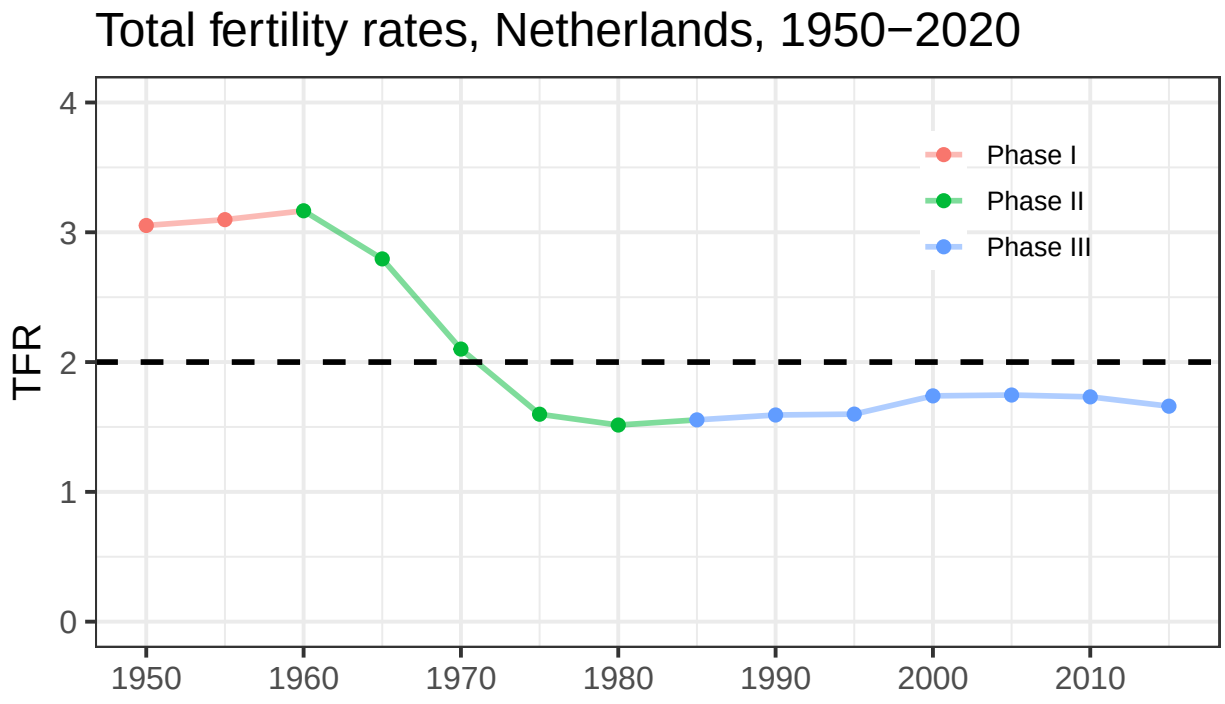
Identify the period in which the Netherlands entered Phase III of the fertility model.

```
# Question 3a -----------------------------------------------------------------

nld_phase3_year <- nld_tfr %>%
  arrange(year) %>%
  filter(tfr < 2) %>%
  mutate(
    year_diff = lead(year) - year,
    period_5 = year_diff == 5 & lag(year_diff) == 5,
    two_increases = tfr > lag(tfr, 1) & tfr < lead(tfr, 1)
  ) %>%
  filter(period_5 & two_increases) %>%
  slice(1) %>%
  pull(year)
```

> "The countries that have entered Phase III are defined as the countries in which two subsequent five-year increases below a TFR of 2 children have been observed." (Alkema et al. 2011)

Looking at TFR chart for the Netherlands, we see that Phase III starts with the period beginning in **1985**.

```
nld_tfr %>%
  mutate(phase = case_when(
    year < 1960 ~ "Phase I",
    year >= 1985 ~ "Phase III",
    TRUE ~ "Phase II")) %>%
  ggplot(aes(year, tfr, color = phase, group = 1)) +
  geom_line(size = 1, alpha = 0.5) +
  geom_point(size = 2) +
  geom_hline(yintercept = 2, linetype = "dashed", size = 1) +
  labs(title = "Total fertility rates, Netherlands, 1950-2020",
       x = " ", y = "TFR") +
  scale_x_continuous(n.breaks=8) +
  scale_y_continuous(n.breaks=6, limits = c(0, 4)) +
  theme_bw(base_size = 15) +
  theme(legend.title = element_blank(),
        legend.position = c(0.81, 0.8),
        legend.background = element_blank(),
        legend.text = element_text(size=10))
```



0.1.2 *Q3.b*

Fit a (non-Bayesian) AR(1) model to the Phase III data, estimating the long-term mean, autoregressive parameter, and error variance.

```
# Question 3b ------------------------------------------------------------------
nld_ts <- nld_tfr %>%
  filter(year >= nld_phase3_year) %>%
  pull(tfr) %>%
  ts()

nld_model = arima(nld_ts, order=c(1,0,0))
```

I fit an order 1 autoregressive model to the subset of Netherlands TFR data in Phase III, and extract some model parameters below. *Note that the AR(1) model was fit using the "mle" method.*

Table 2: Netherlands Phase III AR(1) model parameters

	Model 1
ar1	0.61 [0.03, 1.19]
intercept	1.64 [1.55, 1.74]
Num.Obs.	7
AIC	−13.3
BIC	−13.5
Log.Lik.	9.652

0.1.3 *Q3.c*

Find the predictive distribution of Netherlands TFR for 2020-2025 conditional on this model, analytically or by simulation. Plot the distribution and give its median and a 95% prediction interval.

$$(F_{t+1} - \mu) = \rho * (F_t - \mu) + \epsilon$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (\hat{y_i} - y_i)^2}{N}$$

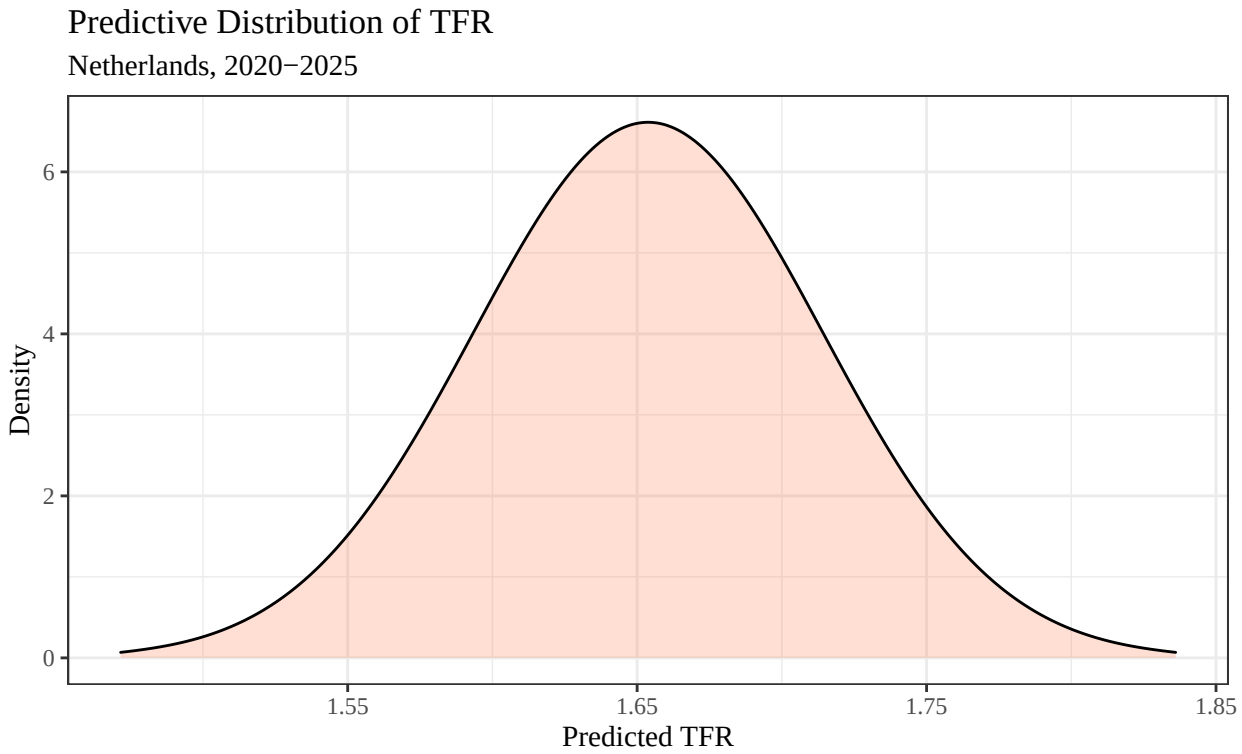
```
nld_tfr_last = nld_tfr$tfr[length(nld_tfr$tfr)]
nld_sd = sqrt(nld_model$sigma2)
nld_mean = nld_model$coef["intercept"]
nld_ar1 = nld_model$coef["ar1"]
nld_mean_pred = nld_ar1 * (nld_tfr_last - nld_mean) + nld_mean
nld_pred_dist <- qnorm(seq(.001, .999, .001), mean = nld_mean_pred, sd = nld_sd)

nld_pred_tbl <- tibble(
  Mean = mean(nld_pred_dist),
  Median = median(nld_pred_dist),
  `2.5% PI` = Mean - 1.96 * nld_sd,
  `97.5% PI` = Mean + 1.96 * nld_sd)
```

- $\sigma^2 = 0.0034781$
- $F_{2020} = 1.6538011$

```
knitr::kable(
  nld_pred_tbl,
  booktabs = TRUE,
  digits = 3,
  caption = "Predictive distribution of Netherlands TFR, 2020-2025") %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

Table 3: Predictive distribution of Netherlands TFR, 2020-2025

Mean	Median	2.5% PI	97.5% PI
1.654	1.654	1.538	1.769


Predictive Distribution of TFR
Netherlands, 2020−2025
Density
0
2
4
6
1.55
1.65
1.75
1.85
Predicted TFR