The Future of Completed Cohort Fertility in Low Fertility Countries: A Comparison of Forecasting Methods

Yen-hsin Alice Cheng
Max Planck Institute for Demographic Research

Introduction

Starting in the early twentieth century, period total fertility rates (PTFR) have been declining in the Nordic countries. By the mid-twentieth century, many European countries have experienced sub-replacement fertility. Other parts of Europe followed such a trend at later points in the past century. By the turn of the twenty-first century, PTFR have dropped below replacement-level in nearly all European countries. In 2008, total fertility in Austria, Bulgaria, Russia, Spain, and Greece is 1.4 children, while TFR is only 1.3 in Germany, Poland, Italy, and Portugal (PRB, 2008). Today, three-quarters of Europe's population live in countries with TFRs below 1.6 births per woman. Demographers and policy makers are often worried about the long-term implications brought by such dramatic demographic changes. However, given that fertility declines in Europe are often accompanied by a strong postponement in fertility behaviors, it is believed that PTFR disguises the fact that a lot of the postponed births will eventually be recuperated at older ages. In turn, the drop in cohort total fertility rates (CTFR) should not be as drastic as what are observed in PTFR in many European countries. Thus, forecasting completed fertility is an important tool to gauge the future level of fertility in developed countries. This study aims to forecast the CTFR among recent cohorts of women (those born between 1965 and1975) across ten countries from Northern to Southern Europe and Japan, using the Gompertz Model along with linear extrapolation and rate-freezing methods. Forecasting errors will also be estimated using error simulation by truncating completed cohort rates between ages 30 and 45 to measure the precision and bias of each method.

Past Research on Cohort Fertility

Scholars in the past have sought to predict the future level of CTFR with various methods (Chandola, Coleman, and Hiorns, 1999; Frejka and Calot, 2001; Frejka and Sardon, 2004; Li and Wu, 2003; Schmertmann 2003; Peristera and Kostaki, 2007). While several of these studies seek to develop parametric models to fit curves that describe fertility schedules in low-fertility countries, Frejka and Sardon utilized a rate-freezing method and Li and Wu utilized the singular-value-decomposition (SVD) model to forecast completed fertility. A more recent attempt is the behavioral Gompertz Model advanced by Goldstein (2008) and later refined by Myrskylä and Goldstein (2010). They showed that the model fits very well in several low and moderate fertility countries with completed cohort data and seems to have good forecasting capability for several truncated cohorts.

One recent study that forecasted CTFR values for a set of developed countries indicates a low CTFR of 1.2-1.3 children in Italy and German-speaking countries to 1.5 children in the Netherlands and Japan (Frejka and Sardon, 2004). They applied frozen rate method to forecast cohorts that have only less than 15 percent of fertility left in their reproductive career. For cohorts of women who were born after 1970, 85 percent of the total completed fertility is usually attained around ages 34 or 35. We suspect that these low estimates for completed cohort fertility might be underestimations, since freezing the last observed set of rates does not fully capture the fertility recuperation pattern that are common at older ages in low-fertility countries.

Although most forecasts contain errors, forecasting methods that make good use of period rates with reasonable assumptions can provide us with valuable information on the future of completed fertility, such as the scope and pace of fertility postponement and whether there is also actual decrease in the level of fertility. Prior studies that forecast CTFR have not sought to compare forecast estimates using multiple methods, nor have forecast errors been systematically examined with age truncation experiments, using completed cohort schedules. This study aims to forecast the CTFR among recent cohorts of women (those born between 1965 and1975) across ten countries from Northern to Southern Europe and Japan, using the Gompertz Model along with linear extrapolation and frozen rate methods. A simulation of forecasting error will also be performed by arbitrarily truncating completed cohort fertility schedules between ages 30 and 45. These truncation points will also be converted into percentiles achieved CTFR to show the distribution of forecasting error in an alternative perspective. Preliminary findings show that the rate-freezing method consistently offers underestimated completed cohort fertility level for all truncation points. For many countries, the Gompertz model and the linear extrapolation method offer more precise and close forecasted CTFR values for the 1975 cohorts who were in their early 30s in 2007. The paper will end with a comparison of the forecasted CTFR values estimated by these three methods and also with CTFR estimated by a recent study by Frejka and Sardon (2004), followed by a general discussion about the merits and caveats of each method.

Study Design

Data

The data employed are long time-series of age-specific fertility rates obtained from the Human Fertility Database[1] and the Eurostat Database.[2] Data collection of period fertility rates for ages 16 to 50 up to year 2007 for Sweden (1950-2007), Switzerland (1950-2007), Austria (1951-2007), and the Netherlands (1950-2007) are utilized. Period fertility rates from the Eurostat Statistics Database were obtained for Denmark (1960-2007), Italy (1960-2007), Greece (1961-2008), and Portugal (1960-2007). In addition, we thank Drs. Michaela Kreyenfeld and Ryuichi Kaneko for providing us with the data from W. Germany (1952-2008) and Japan (1950-2008).

Analytical Plan

Three methods are adopted to forecast completed cohort fertility schedules in these ten countries: linear extrapolation, the Gompertz Model, and rate-freezing methods. To start out, the details of each method and its application will be described. Then, we conducted an experiment of truncating completed cohort fertility schedules at different ages (from ages 30 to 45) to gauge the precision and bias of the forecasting methods. Then each truncation age is converted to a corresponding percentile of achieved CTFR for that specific cohort in a country to examine the distribution of forecast error by truncation percentile. Finally, forecast errors of ten countries were merged into one scatterplot, separated by the method used, to see the overall performance of each method in making reliable completed fertility forecasts for younger cohorts of women. These checks on error terms provide us a ground to investigate the accuracy of our estimates. Next, we'll explain the details of each forecasting method.

I. Linear Extrapolation

Period fertility data obtained from Eurostat are rate matrices of age by year, with each column showing a series of age-specific fertility rates in a given year. Trend of fertility rates at a given age across years were plotted to see the general pattern of recent fertility level (see Figures 1 for an example of data trend in Switzerland). In this figure, an inflection point around 1975 is observed, indicating the start of the fertility postponement process in Switzerland. In all countries analyzed here, such an inflection point for age 30 to age 40 trend lines occur between the mid-1970s and the mid-1980s. In turn, the linear slope defined by

---

[1] http://www.humanfertility.org/

[2] http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

data points after the inflection point may serve as a good extrapolation support. For all ten countries the linear slope defined by observed data in the last five years (2002-2007 in this example) is used as the extrapolation support. Age-specific fertility data series were then forecasted into year 2020. Then, a cohort fertility matrix is constructed by taking the diagonal cells from the observed and extrapolated period fertility matrix, starting from the cohort of 1944 (for ages 16 to 50) in the case of Swtizerland. For cohorts who have not reached the end of their reproductive years in 2007, part of their fertility schedule now contains forecasted data. Adding up each column of data in this cohort matrix yields forecasted CTFR level.

II. Gompertz Model

The second method is Goldstein's (2008) behavioral Gompertz Model. Curve fitting with this model can be done to observed data using non-linear least squares estimates. As mentioned in his paper, the model fits completed cohort fertility schedules in several low fertility countries very well, especially when biological infertility is taken into account. This is done by multiplying the Gompertz function $f(x) = Ka\exp\left[-\frac{a}{b}e^{-bx} - bx\right]$ with an infertility index G(x) to depress the over-estimated fertility level at ages over 35 caused by the exponential function assumed in the Gompertz Model (for a more detailed description on the setup and application of the model please see Goldstein, 2008). The improved function below is used to fit a complete fertility schedule for recent truncated cohorts in this study:

$$f^*(x) = Ka\exp\left[-\frac{a}{b}e^{-bx} - bx\right]G(x)$$

The index g(x) takes on values of 1 up to age 33 and then with a linear decline in values to 0 at age 45 and older. The Gompertz Model with infertility index fits completed cohort fertility schedules in France, Italy, and Japan very well (Goldstein, 2008). Myrskylä and Goldstein (2010) further developed a linearized Gompertz model as a refinement to the previous modeling procedure. The steps to acquire a forecasted CTFR using the Linearized Gompertz Model are: First, inflate the observed rates (divide the observed data series by the infertility index G(x)) to get estimated fertility schedule "in the absence" of infertility (see Figure 2a for an example of observed and inflated fertility schedule). Second, linearize the inflated cumulative fertility rates using the equation proposed in the paper (see Figure 2b for the linearized Gompertz):

$$\ln\left(\frac{d\ln Pt}{dt}\right) = \ln b + a - bt = gt$$

Third, forecast fertility rates at ages above truncation by utilizing the following two formulae (see hollowed red dots in Figure 2b):

$$g_t = g_0 + \delta_t + \sum_{i=1}^{t} \varepsilon_i$$

(where δ is the slope for the linearized Gompertz line)

$$P_{t+k} = \frac{P_{t+k-1}}{1 - \exp[0.5 * (g_{t+k} + g_{t+k-1})]}$$

Finally, incorporate infertility index back to the observed and forecasted data series to obtain predicted CTFR (see Figure 2c for final inflated and infertility adjusted schedules). In this study, the infertility-adjusted linearized Gompertz Model will be applied to truncated cohort fertility schedules in ten developed countries.

III. Frozen Rate Method

Finally, another way to forecast the completed fertility level of truncated cohorts is to freeze all age-specific fertility rates in the last observed year, which is year 2007 in this study. Then, the cohort fertility matrix is constructed by taking the diagonal cells from the period fertility matrix with frozen rates for years 2008 to 2020, based on the last observed values in 2007 or in 2008.

Preliminary Results

Figure 3 presents the distribution of forecast error for Switzerland. Forecasts obtained by using rate-freezing method constantly offer underestimated CTFR level across all truncation ages and truncation percentiles. Forecast errors from linear extrapolation and Gompertz model are similarly small and clustered around zero at higher truncation percentiles. At lower truncation percentiles, linear extrapolation method outperforms the Gompertz model in the case of Switzerland.

In Figure 4, color coded forecast errors are grouped into three scatterplots by forecasting method. Frozen rate method has a bigger downward bias at every truncation point than the other two methods. The Gompertz method has a smaller bias, yet a larger variance (especially among more truncated younger cohorts) than the other methods. Linear extrapolation method has a relatively small bias and little variance against the actual observed CTFR (for completed cohorts), making the forecasted CTFR level more reliable than the other two methods. To sum up, in most cases, forecast errors are within ±0.05 children for linear extrapolation and ±0.1 children for the Gompertz method.

Finally, Figure 5 shows the forecasted CTFR values obtained from the linearized Gompertz Model, linear extrapolation, and rate-freezing methods in all ten low-fertility countries. As can been seen in these graphs, the three methods yield quite close (sometimes nearly identical) CTFR values for cohorts born between 1965 and 1975. CTFR values from these three methods only diverge from each other for more than .05 births for cohorts born between 1970 and 1975. Overall, the graphs show that using the rate-freezing method to forecast CTFR tends to yield the lowest fertility level for recent truncated cohorts. The level of the Gompertz estimates sometimes resembles those from linear extrapolation, but sometimes they are different.

The methods used here seem to offer reliable forecasts for cohorts as recent as those born in 1975. As for the level of completed cohort fertility, we can expect it to be close to 2 children for Swedish and Danish women who are in their early thirties in 2007. In Switzerland and Austria, a CTFR of 1.6 can be expected for the same cohorts of women, whereas the CTFR for Western German women in these cohorts is about 1.55. The completed cohort fertility is higher in the Netherlands for the 1975 cohort, achieving a level of 1.75. In all three Southern Europe, women who are born in 1975 can expect to have 1.45 children in Italy, 1.55 in Portugal, and 1.6 in Greece. Finally, the 1975 cohort of Japanese women is estimated to have about 1.4 children. Overall, the picture forecasted by the linear extrapolation and the linearized Gompertz Model is more optimistic than the forecasts made by Frejka and Sardon (2004).

	Netherlands	**W. Germany**	**Switzerland**	**Austria**	**Italy**	**Greece**	**Japan**
Frejka & Sardon 2004	**1.5**	**1.3**	**1.3**	**1.3**	**1.2**	**1.4**	**1.5**
Our Forecast	**1.75**	**1.55**	**1.6**	**1.6**	**1.4**	**1.55**	**1.4-1.45**

For the cohorts born between 1965 and 1975, it seems that most of the countries are expecting a declining CTFR level, with the exception of Denmark and W. Germany where a reversal of cohort total fertility can be expected. Sweden is the only country that has a stabilized CTFR at the level of two children across all the cohorts examined.

References

Andersson, Gunnar, Marit Rønsen, Lisbeth B. Knudsen, Trude Lappegård, Gerda Neyer, Kari Skrede, Kathrin Teschner, and Andres Vikat. 2009. "Cohort Fertility Patterns in the Nordic Countries." *Cohort Fertility Patterns in the Nordic Countries* 20:313-352.

Chandola, T., D. A. Coleman, and R. W. Hiorns. 1999. "Recent European fertility patterns: Fitting curves to distorted distributions." *Population Studies* 53:317-329.

Frejka, T. and G. Calot. 2001. "Cohort reproductive patterns in low-fertility countries." *Population and Development Review*:103-132.

Frejka, T. and J. P. Sardon. 2004. *Childbearing Trends and Prospects in Low-Fertility countries: A cohort analysis*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Goldstein, Joshua R. (2008). A Behavioral Gompertz Model for Cohort Fertility Schedules in Low andModerate Fertility Populations. Paper presented at the Annual Meeting of the Population Association of America at New Orleans, LA, March, 2008.

Li, Nan and Zheng Wu (2003). Forecasting Cohort Incomplete Fertility: A method and an application. *Population Studies,* 57(3): 303-320.

Myrskylä, M. and J. R. Goldstein (2010). Prediction and Error Propagation in Innovation Diffusion Models, with Applications to Demographic Processes. MPIDR Working Paper WP-2010-013.

Peristera, P. and A. Kostaki. 2007. "Modeling fertility in modern populations." *Demographic Research* 16:141-194.

Schmertmann, C. P. 2003. "A system of model fertility schedules with graphically intuitive parameters." *Demographic Research* 9:82-110.

Van Imhoff, Evert (2001). On the Impossibility of Inferring Cohort Fertility Measures from Period Fertility Measures. *Demographic Research,* 5: 23-64.

Figure 1: observed and extrapolated age-specific fertility rates for ages 27-41 in Switzerland

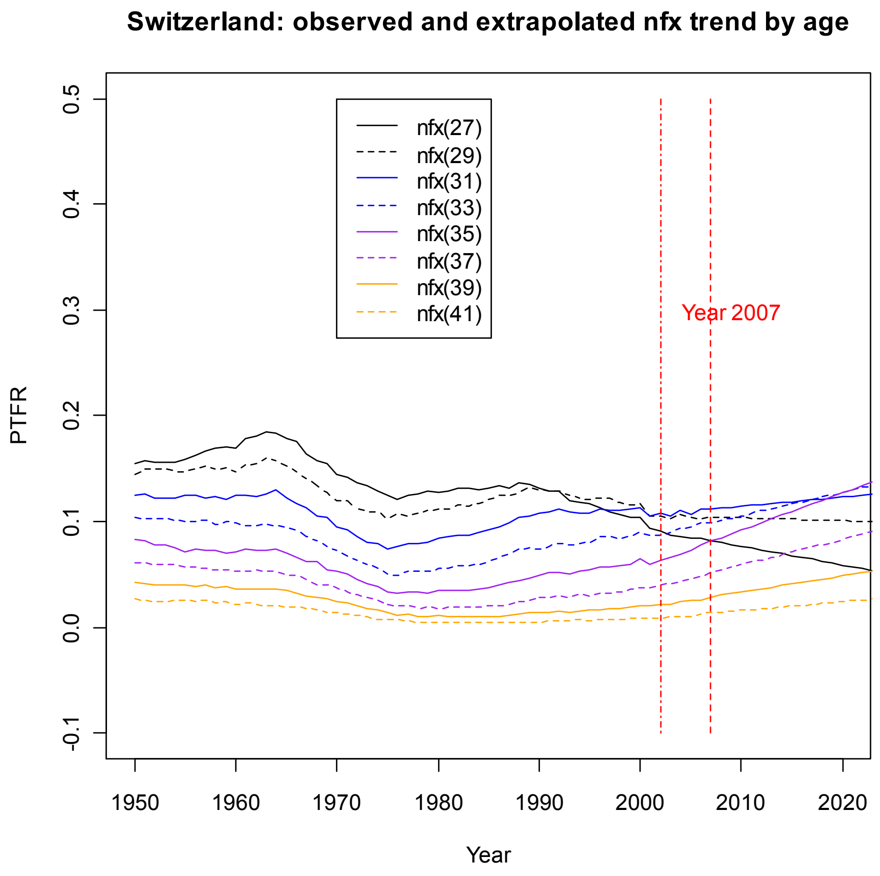


Figure 2a: Observed and Inflated Fertility Schedule 1970 Cohort in Switzerland

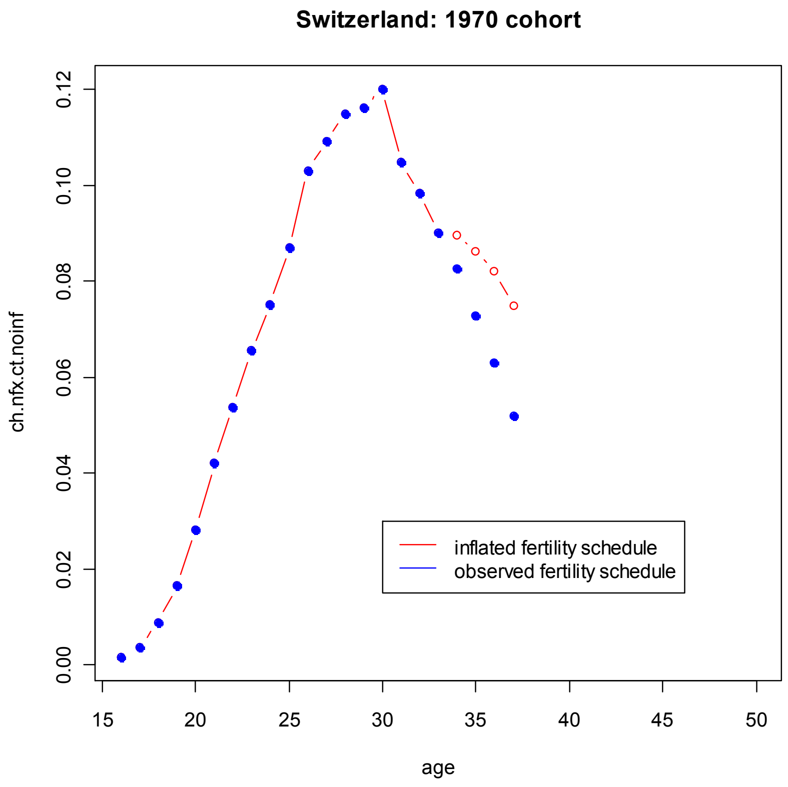


Figure 2b: Linearized Gompertz

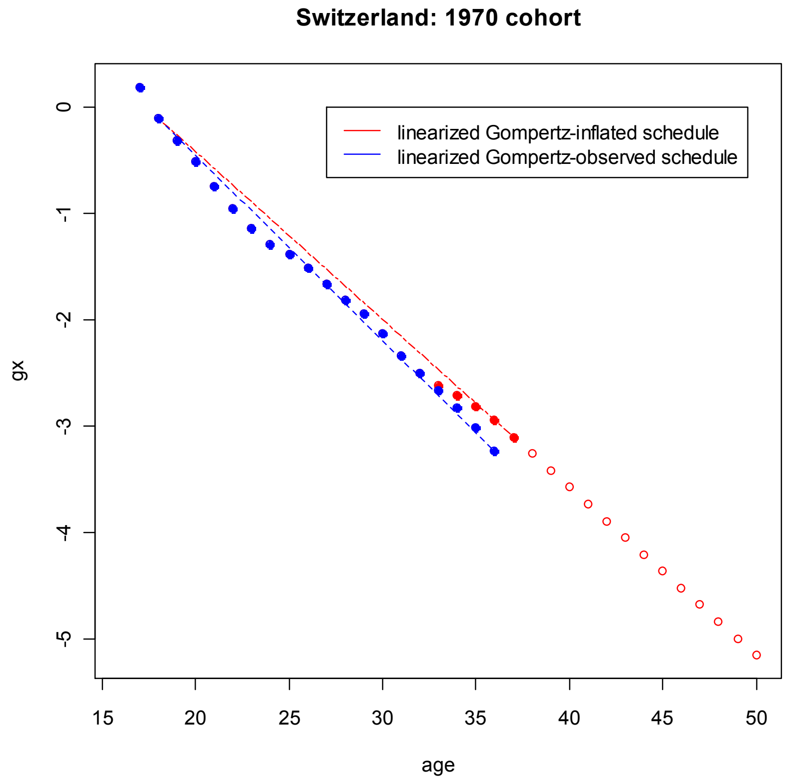


Figure 2c: Inflated and Infertility-Adjusted Cohort Fertility Schedules

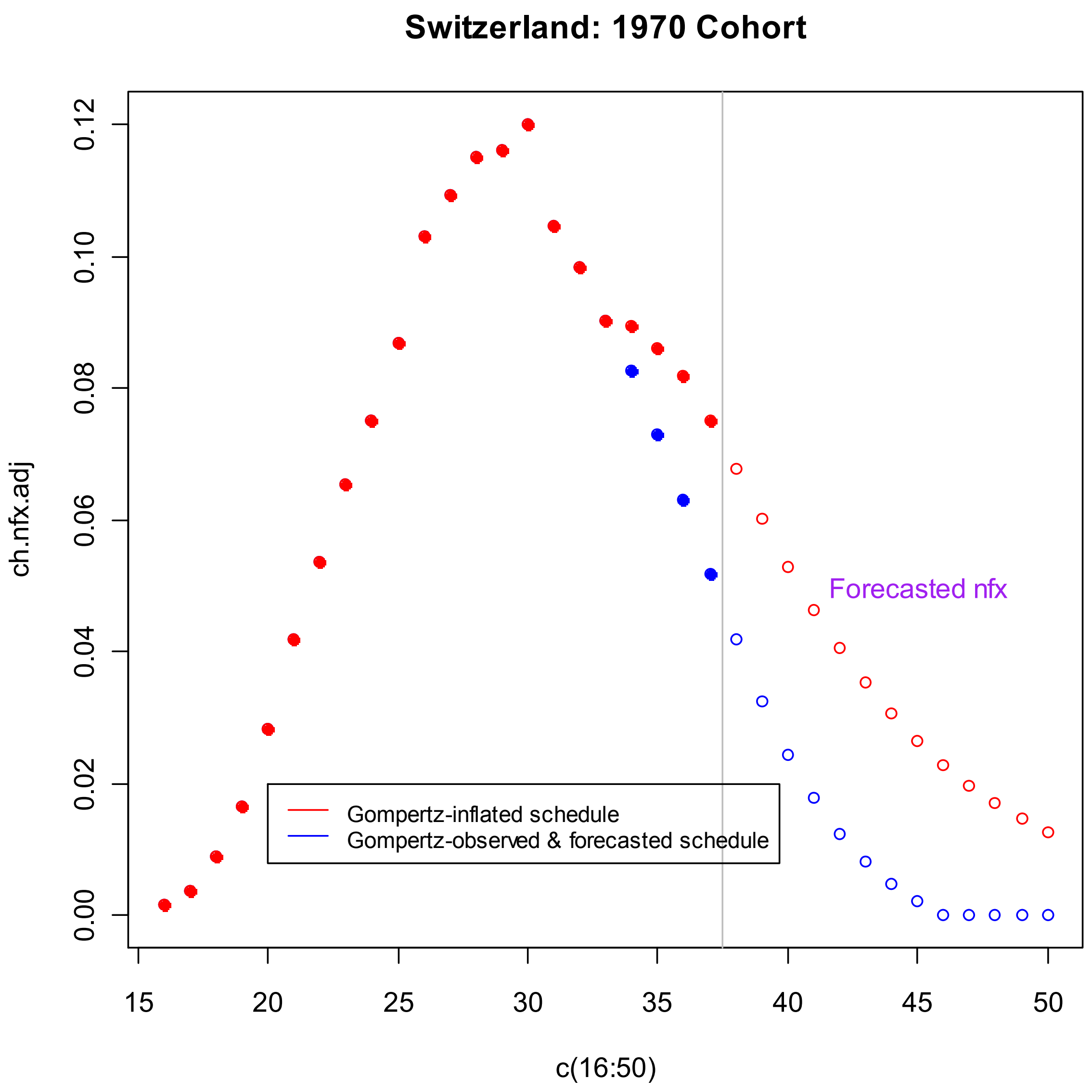


Figure 3: Forecast Errors by Truncation Age and Truncation Percentile (of achieved CTFR) for Switzerland

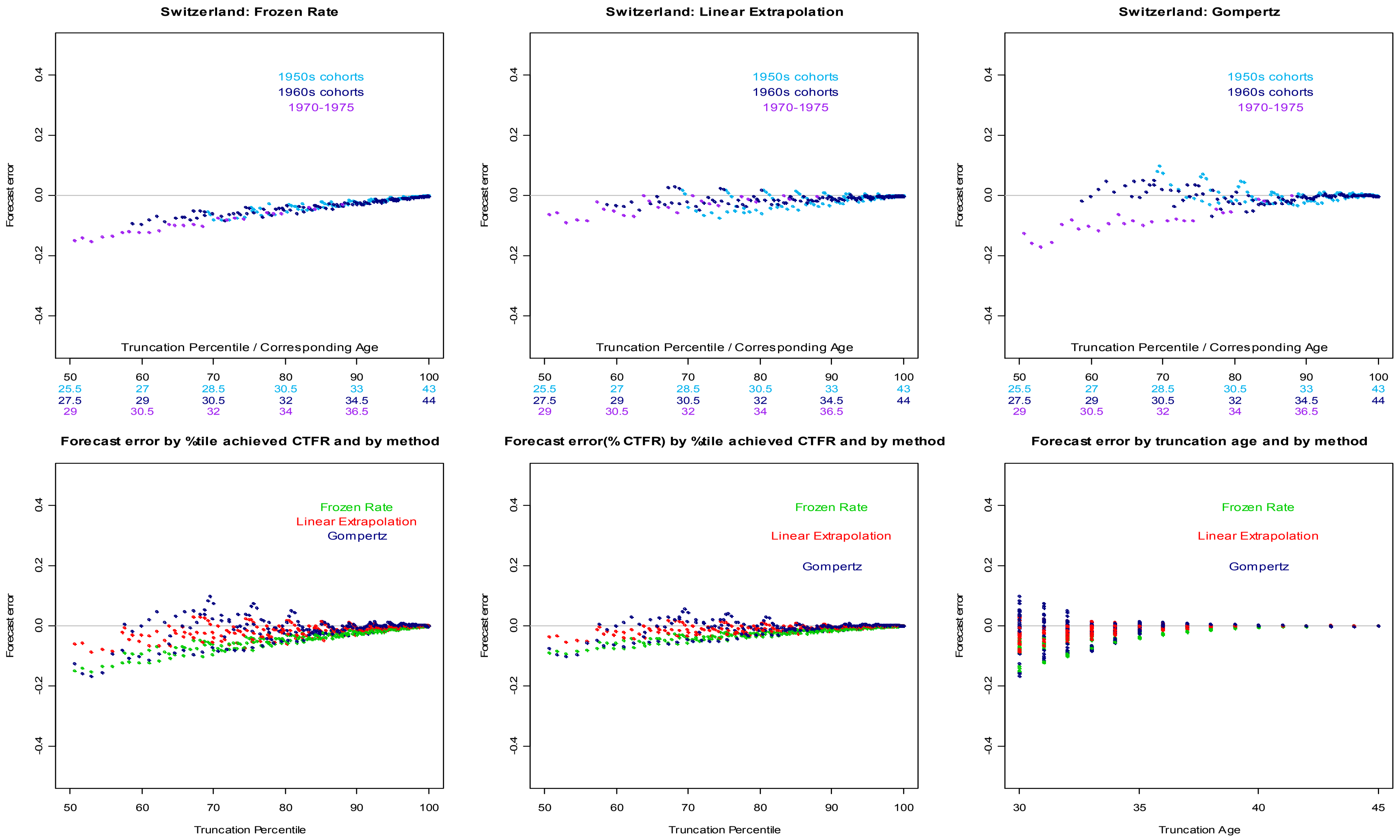


Figure 4: Forecast Errors for Ten Low-Fertility Countries, Grouped by Forecasting Method

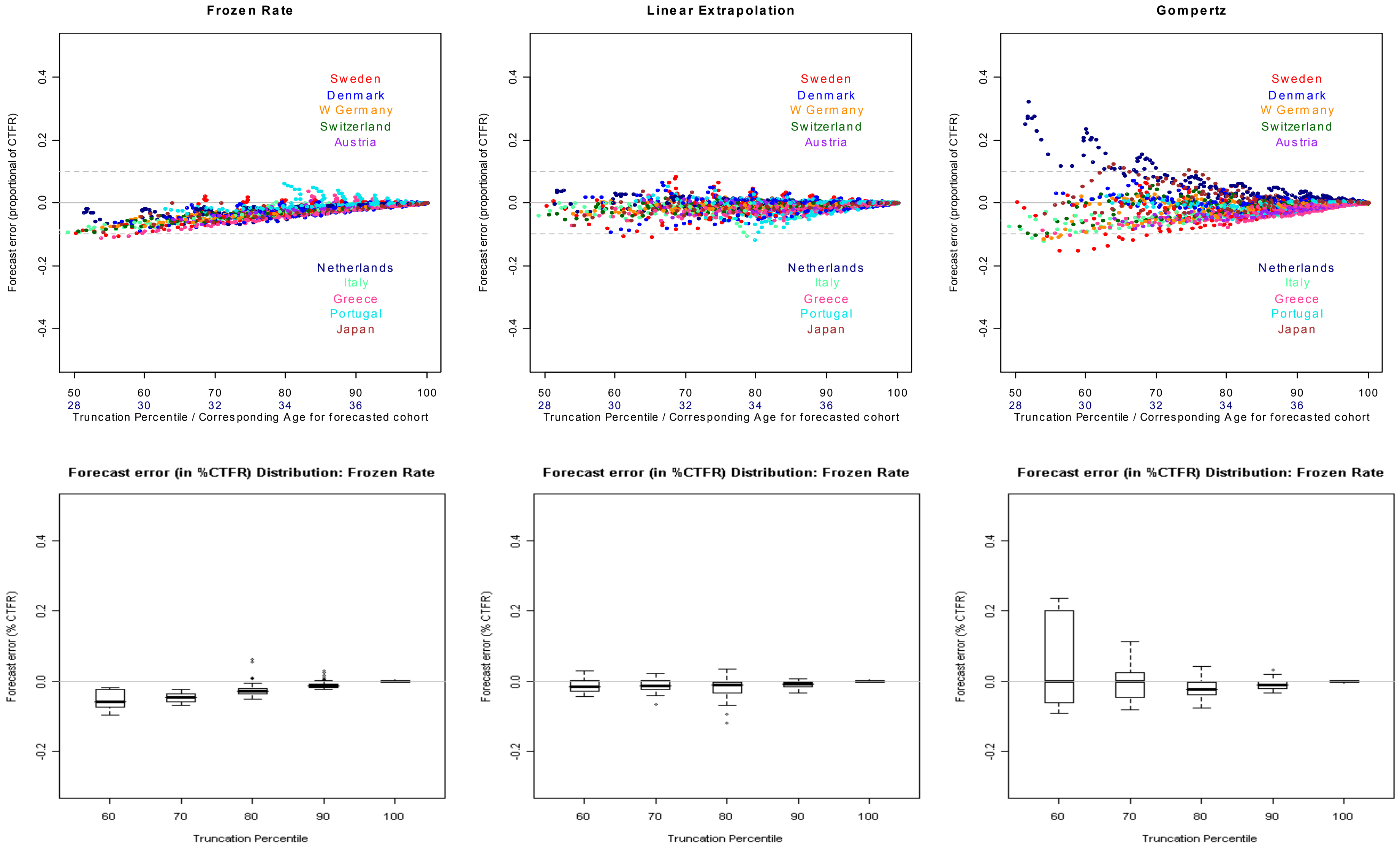
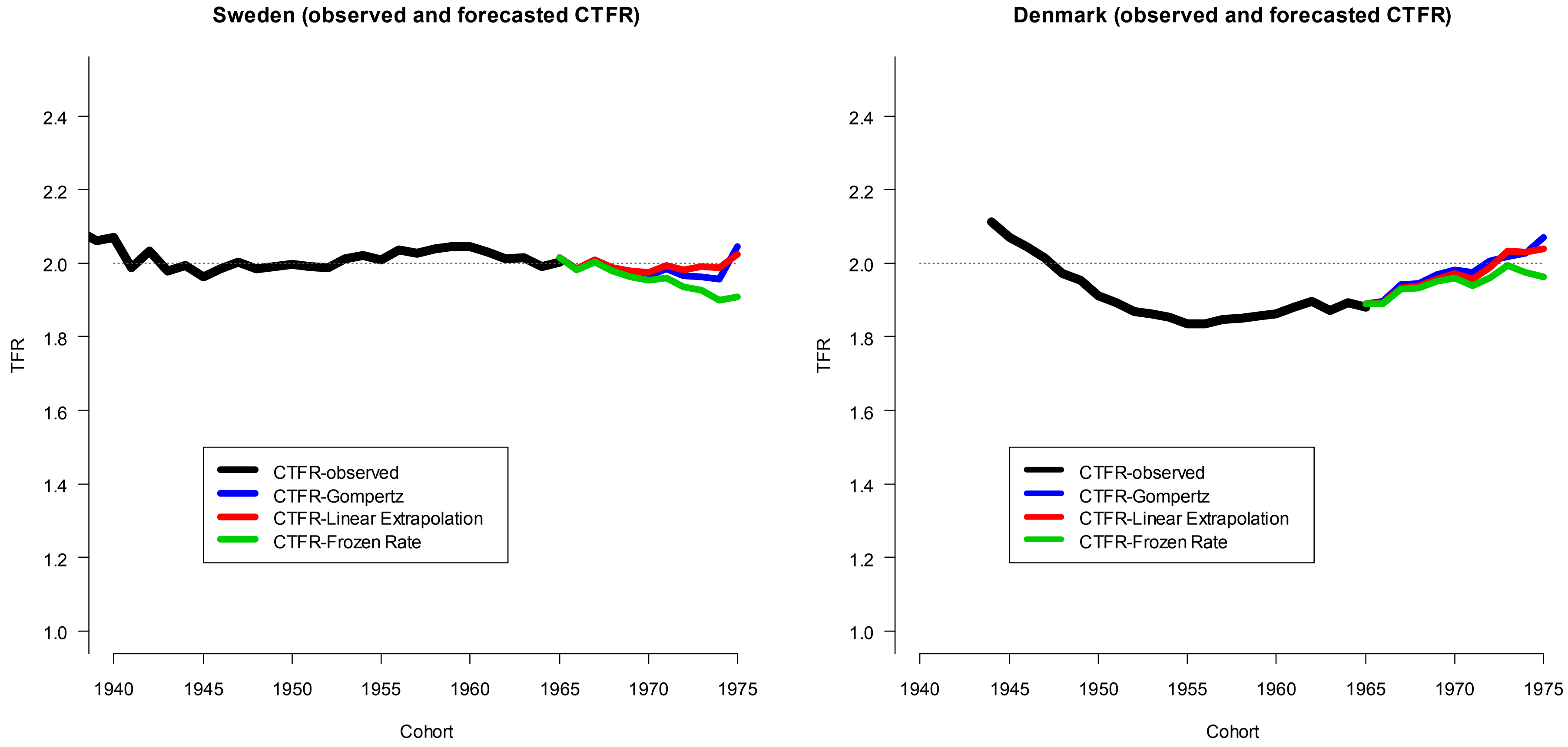
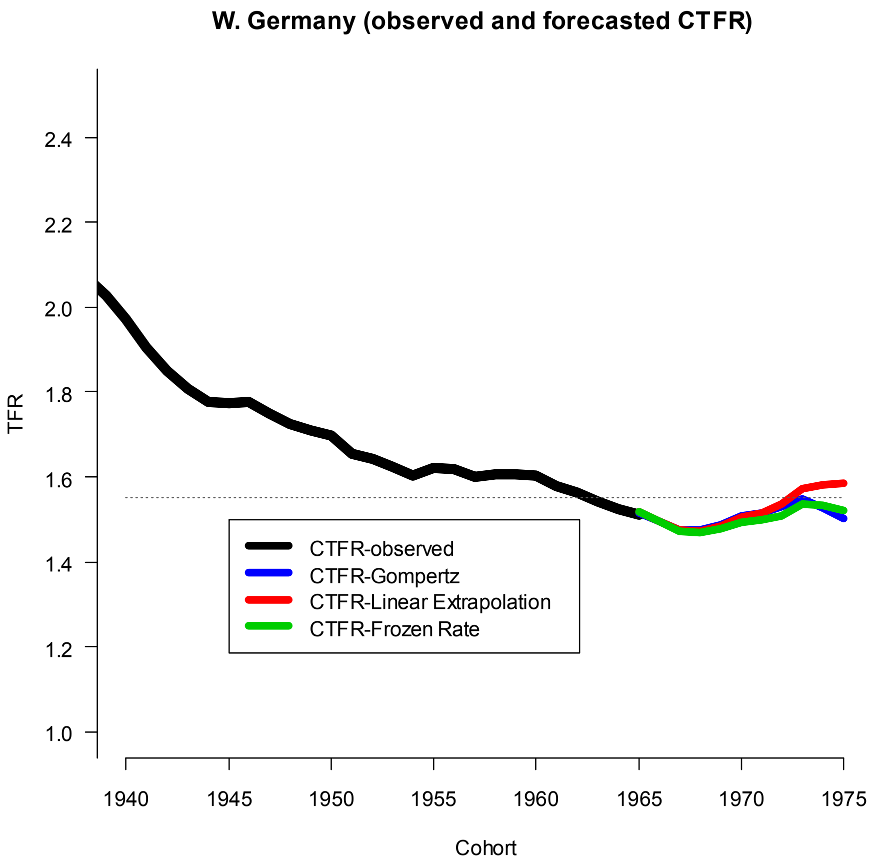
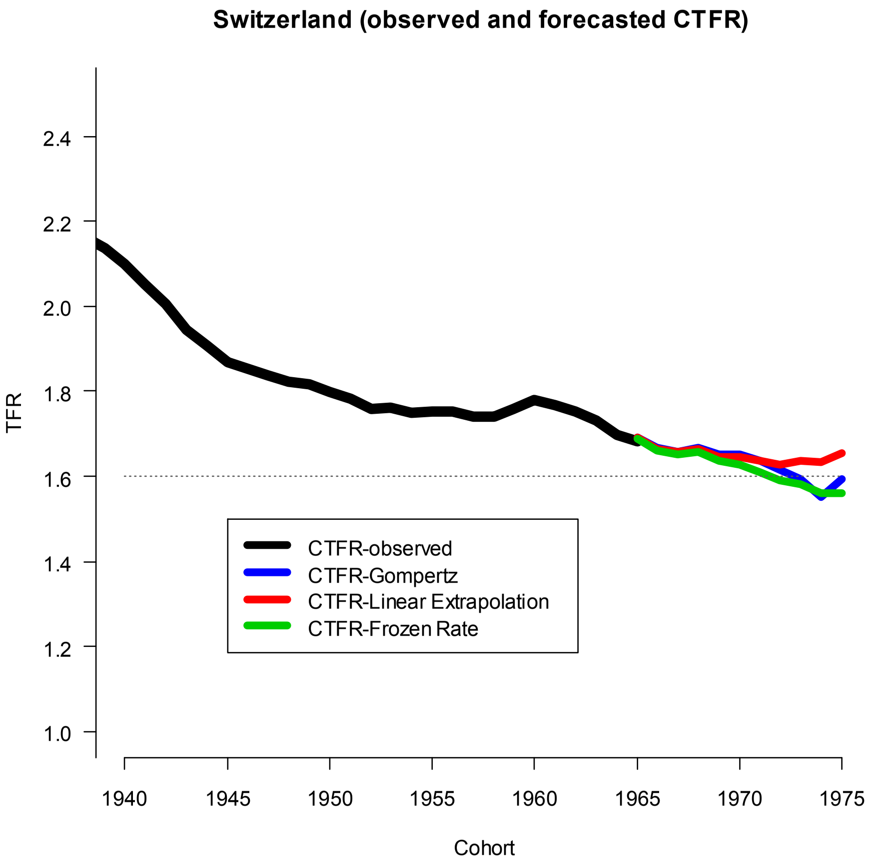
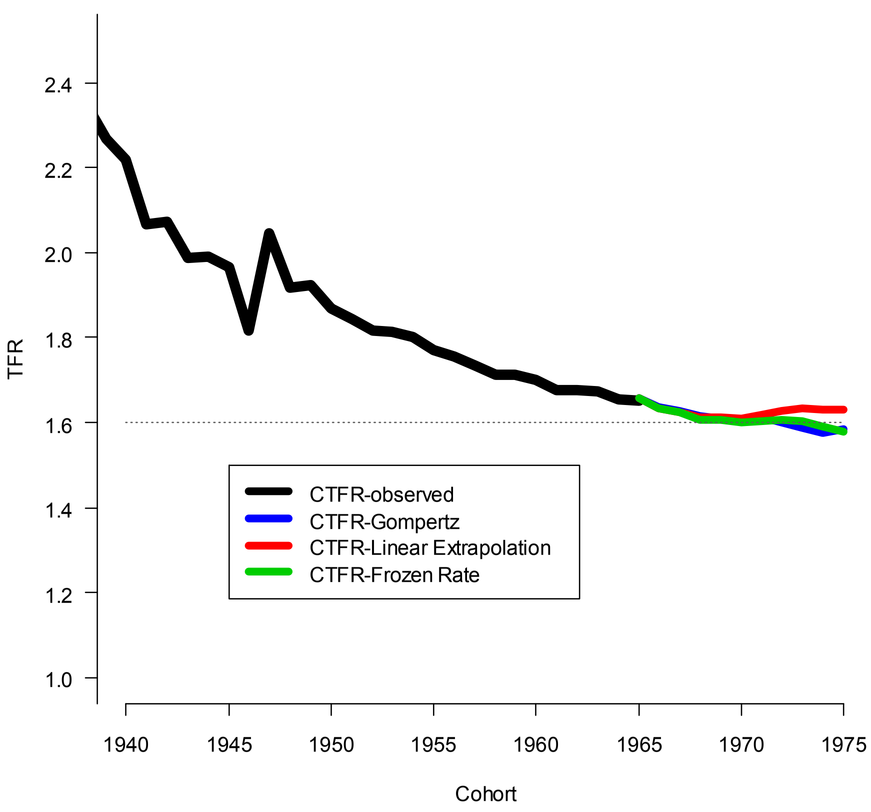


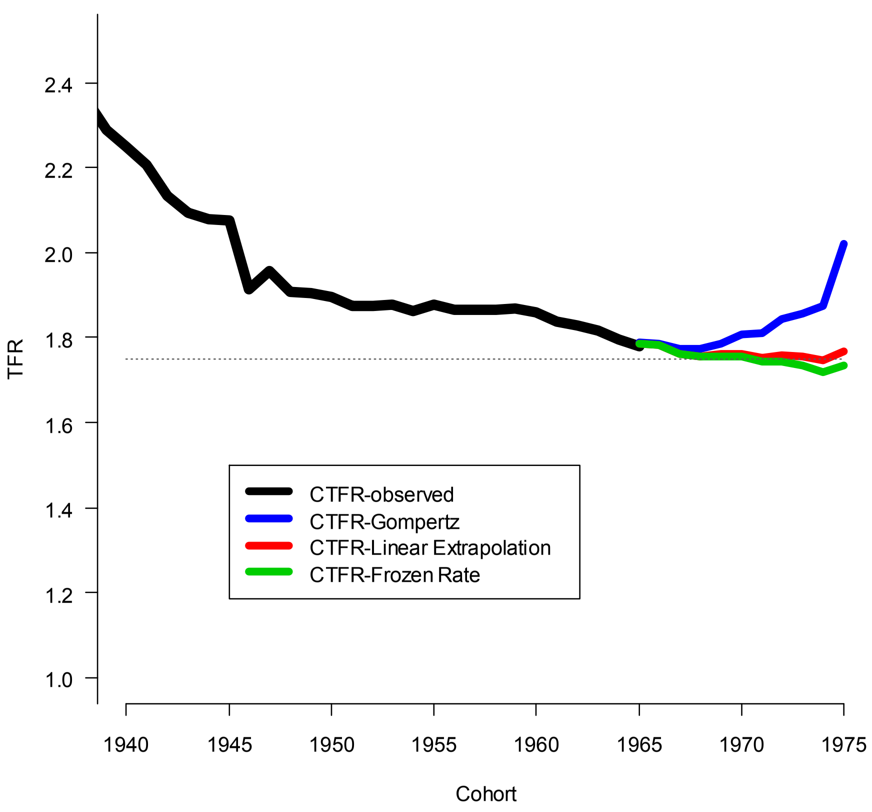
Figure 5: Level of CTFR by Cohort using Linear Extrapolation (with 5-year extrapolation support), Linearized Gompertz Model with infertility, and Rate-freezing methods

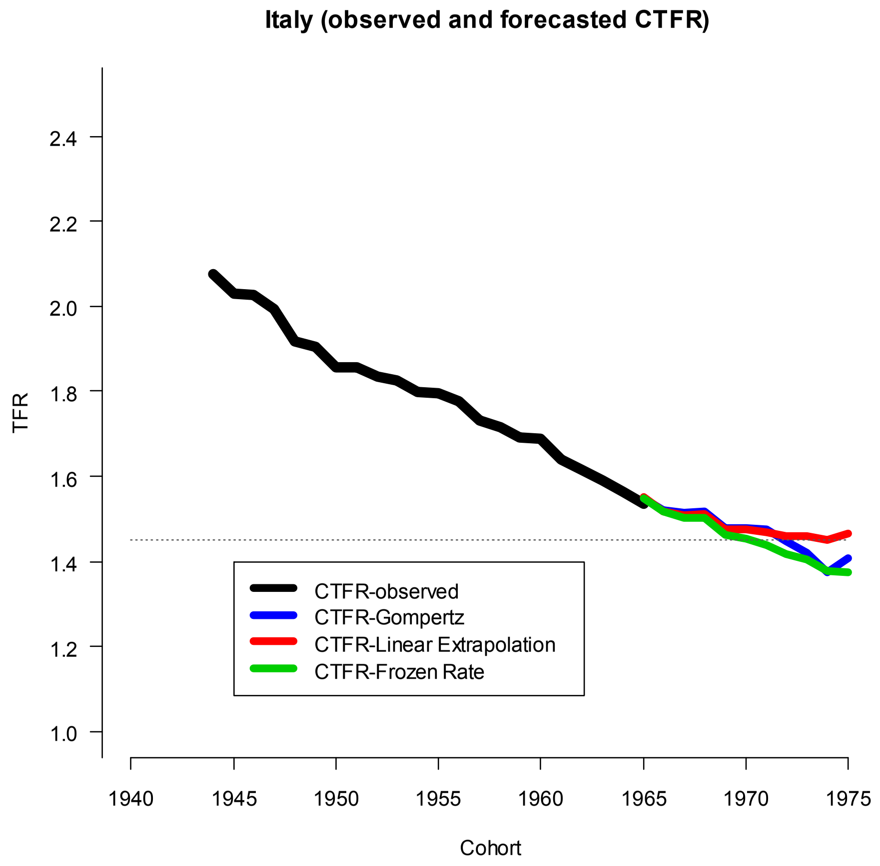


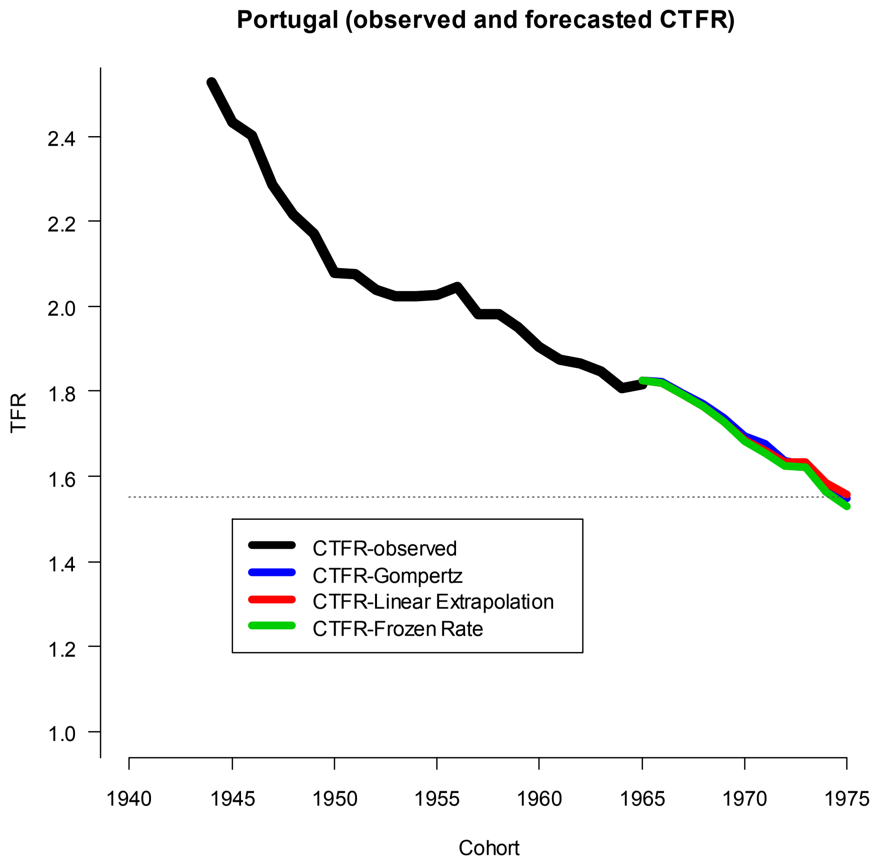

W. Germany (observed and forecasted CTFR)
TFR
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
1940
1945
1950
1955
1960
1965
1970
1975
Cohort
CTFR-observed
CTFR-Gompertz
CTFR-Linear Extrapolation
CTFR-Frozen Rate

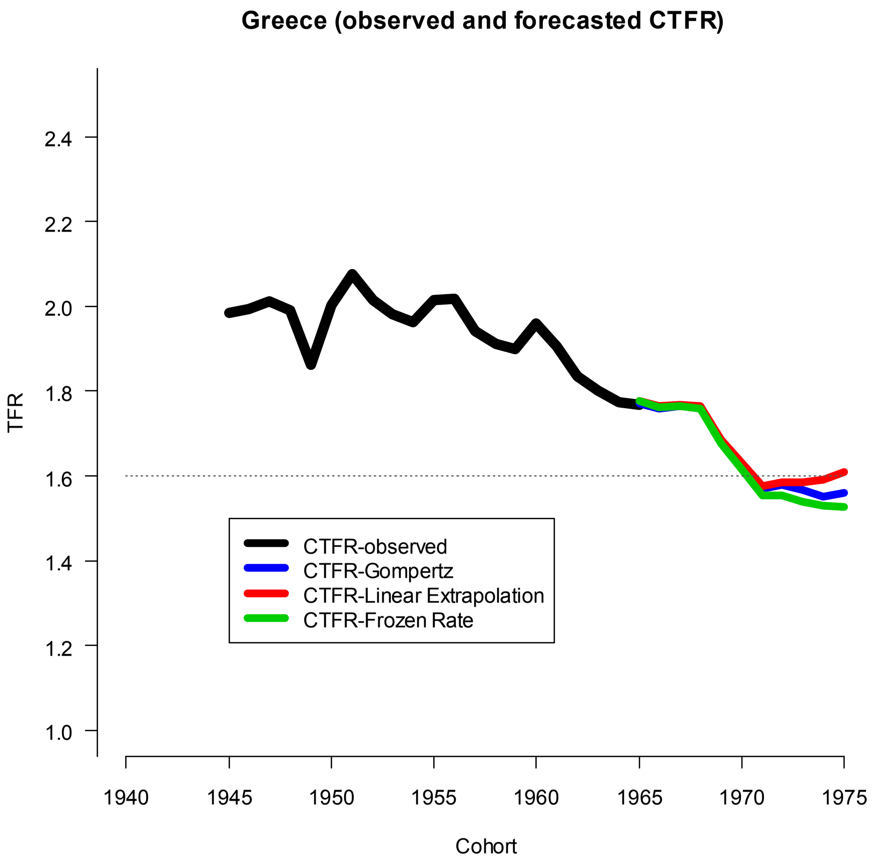

Switzerland (observed and forecasted CTFR)
TFR
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
1940
1945
1950
1955
1960
1965
1970
1975
Cohort
CTFR-observed
CTFR-Gompertz
CTFR-Linear Extrapolation
CTFR-Frozen Rate

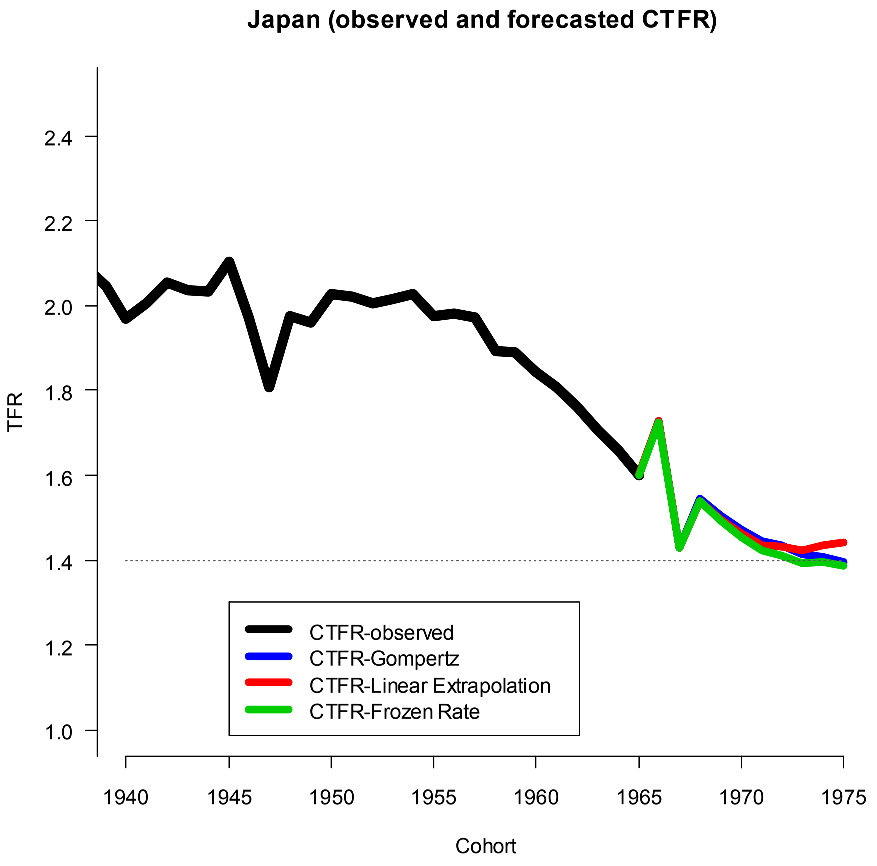

Austria (observed and forecasted CTFR)


The Netherlands (observed and forecasted CTFR)


Italy (observed and forecasted CTFR)
TFR
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
1940
1945
1950
1955
1960
1965
1970
1975
Cohort
CTFR-observed
CTFR-Gompertz
CTFR-Linear Extrapolation
CTFR-Frozen Rate


Portugal (observed and forecasted CTFR)
TFR
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
1940
1945
1950
1955
1960
1965
1970
1975
Cohort
CTFR-observed
CTFR-Gompertz
CTFR-Linear Extrapolation
CTFR-Frozen Rate


Greece (observed and forecasted CTFR)
TFR
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
1940
1945
1950
1955
1960
1965
1970
1975
Cohort
CTFR-observed
CTFR-Gompertz
CTFR-Linear Extrapolation
CTFR-Frozen Rate


Japan (observed and forecasted CTFR)
TFR
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
1940
1945
1950
1955
1960
1965
1970
1975
Cohort
CTFR-observed
CTFR-Gompertz
CTFR-Linear Extrapolation
CTFR-Frozen Rate