# How Well can Period Fertility Measures Help to Forecast Cohort Fertility Measures? Cross-Country Evidence

P. C. Roger Cheng*
Department of Economics
National Central University
Taiwan

Eric S. Lin
Department of Economics
National Tsing Hua University
Taiwan

March 17, 2011

**ABSTRACT**

With non-parity and parity specific fertility data taken from 20 European countries, the U.S. and Japan, this paper carry out a comprehensive testing of predictive power for a few competing methods in forecasting the cohort total fertility rate. Two types of methods are included in this paper: conventional ones and period measures. We make full use of these data and construct distributions of forecast errors to determine the relative performances of these methods. The evidence strongly suggests that smoothed period measures do help to forecast cohort fertility ones, and they outperform conventional methods significantly for the first and second births. Some useful conclusions for future research can be derived from our experiment results.

## 1 Introduction

Eric wants to visit a city located to the south of his, in which direction should he go? The answer to this seemingly simple question can, in fact, be complicated. Fifty years ago, if a

*Correspondence: P. C. Roger Cheng, Department of Economics, National Central University, Taoyuan 320, Taiwan. Tel: 886-3-422-7151 ext. 66310; Fax: 886-3-422-2876; E-mail: paochih@mgt.ncu.edu.tw.

professor at Rutgers University (New Brunswick, NJ) wanted to attend a conference held in Washington D.C., she/he could drive or take a train, heading south, to the conference venue. But nowadays, there is an alternative: going north first to Newark Liberty International Airport (EWR) and then taking a plane to get to D.C. From this instance we learn that a seemingly indirect or roundabout way might be the best solution, depending on which criterion (say, travel time) is adopted.

If one wants to measure the cohort total fertility rate (CTFR) for women currently in the middle of their reproductive period, it seems reasonable to use information derived either from the completed part of the target cohort or from the experience of previous cohorts to forecast the uncompleted part that lies ahead. Examples of the former consist mainly of curve fitting models such as a modified version of the Coale-McNeil's (1972) double exponential model (e.g., Bloom 1982; Chen and Morgan 1991), the Hadwiger function (e.g., Chandola, Coleman, and Hiorns 1999), and the linearized Gompertz model (Myrskylä and Goldstein 2010). The latter examples include the Evans method (Evans 1986), the Li-Wu model (Li and Wu 2003), and the Willekens-Baydar approach (Willekens and Baydar 1984). All these methods make a particular assumption regarding the structure behind the existing data, estimate related parameters, and then produce their forecasts.

What if one proposes an extremely simplified method (denoted by the Naive method hereafter) assuming that the age specific fertility rates (ASFRs) at ages beyond the truncation point remain the same as those in the last year of the data range? Some researchers may turn their noses up at this approach, because summing up such age-specific incidence rates (also called "rates of the second kind") period-wise will lead to a result distorted by the parity distribution of the women who are at risk, especially when there is a strong change in fertility tempo between cohorts. Given this methodological point, they may conclude that it is meaningless to check how well the Naive method can provide estimates approaching the CTFRs. Likewise, adjusted period measures, such as the Bongaarts-Feeney (Bongaarts and Feeney 1998; BF) and the Kohler-Philipov (Kohler and Philipov 2001; KP), may be overruled for the same reason as they are regarded as variants of the period-sum method.

We, however, take a different perspective and share the view of Van Imhoff (2001:36) that "the justifiability of [period measures] can only be verified empirically".[1] In fact, there are quite a few researchers who compare the values of these period indicators with the CTFR (e.g., Kohler and Ortega 2002; Ryder 1990; Schoen 2004; Smallwood 2002; Sobotka 2003; Van Imhoff and Keilman 2000), explicitly or implicitly regarding them as forecasts of the completed fertility for cohorts who have not yet finished childbearing. As to these studies, some researchers (e.g., Kim and Schoen 2000; Li and Wu 2003; Schoen 2004; Van Imhoff and Keilman 2000) do not deny the usefulness of adjusted period measures based on the methodological ground but remain skeptical due to their assumptions regarding the shape of the fertility schedules and/or due to their fluctuating and occasionally absurd values.

In our previous study (Cheng and Lin 2010), we propose a *smoothed* version of the tempo-variance-adjusted KP meausre and treat it as a substitute for the actual CTFR. By smoothing the *primitive* KP values, the problem of random fluctuations and the occurrence of impossible values can be largely removed, and the approximation to the CTFR is consistently and significantly improved. According to our experiment (refer to ibid., Table 1), our proposed approach outperforms all the competing methods under investigation in terms of forecasting the U.S. CTFR of first birth. Also, the simple Naive method surprisingly outperforms most of the competing approaches.

With regard to our experiment evidence, some might argue the following:

[1]Suppose that today is Monday but Roger somehow thinks it is Sunday. Based on the wrong assumption, he may still obtain several correct conclusions such as "today is not Tuesday", "today is not Wednesday", and so on.

1. Cohort fertility measures and period ones are basically incomparable, so how can you infer the former from the latter?
2. Why don't you use the Kohler-Ortega method (Kohler and Ortega 2002; KO) which works with age- and parity-specific occurrence-exposure rates (also called "rates of the first kind") to replace the KP method which works with problematic incidence rates?
3. Your test of predictive power for various models is confined to the U.S. data of first birth and cohorts whose observed experiences are up to age 24. The results you obtain are thus limited. How well can your proposed method perform when it is applied to other countries and/or birth orders?
4. In addition to the KP, there are other period adjusted measures such as the BF and the newly developed Goldstein-Cassidy method (Goldstein and Cassidy 2010; denoted by TFR.dagger in their paper). Why do you focus on the KP only?
5. You utilize the kernel smoother to smooth values of period indicators, but users of your proposed method may prefer to have a simpler way to smooth these values. How about just using the simple moving average method?

The first question is just like "Since your destination is in the south, why do you go north?" It is true that simply summing up incidence rates period-wise is problematic. But if some kind of manipulation or adjustment can help to bridge the gap between cohort fertility measures and period ones, why should we refuse proceeding in this way to figure out the mechanism behind this bridge? The second question also casts doubt on the justification of using incidence rates, a point to which we have just responded. In addition, our focus on measures using incidence rates can be based on the practical reason that ASFR data are much more available for either developed or developing countries.

As to questions 3 – 5, we will respond to each of these as this paper proceeds with a comprehensive analysis. With all-birth-combined and parity-specific fertility data from 20 European countries, the U.S., and Japan, we treat every 25 consecutive years as a sample,[2] and carry out forecasting models to evaluate their predictive powers of the actual CTFR for cohorts covered in each sample period. Two types of models are included in this paper. The first one comprises a few conventional models, including the Naive, the Evans, the linearized Gompertz, and the linear extrapolation method.[3] The second consists of smoothed versions (with various degrees of smoothing) of the adjusted period measures, including the BF, the KP, and the TFR.dagger.[4] Experiment results will be summarized using box-whisker plots so that not only the central tendency but also the dispersion of forecast errors can be investigated.

The remainder of this paper is organized as follows: Section 2 provides an overview of the fertility data from 22 countries and describes the experiment design adopted in this paper. Taking countries with lengthy data spans as examples, we illustrate the (all-births-combined or parity-specific) CTFR forecasts derived from various methods and compare them with the actual ones in Section 3. Next, Section 4 presents the summary statistics across all 22 countries so that we can conclude which method performs better parity by parity. Furthermore, we will investigate what kinds of information can be revealed by ad-

[2]We choose 25 as the sample length in consideration of the applicability since a 25-year ASFR data set can be readily obtained for many countries.

[3]Other conventional approaches such as the Hadwiger, the Coale-McNeil, the Lee-Carter, and the Willekens-Baydar will not be investigated in this paper owing to their poor performances in our previous study.

The linear extrapolation method takes the observed ASFRs in the last five years to fit linear trends age by age, and then extrapolates the ASFRs in the following years. Taking the cells from the observed and extrapolated ASFRs diagonally leads to the forecasted CTFR.

[4]In previous studies, two popular ways of associating period and cohort fertility measures were adopted. The first is to compare the CTFR of a cohort with the period estimate of the year in which the cohort reaches its mean age at birth. The other is to compare the period estimate of a year with the CTFR for women who reach the mean age at birth in that year. We follow the second strategy in this paper.

justed period measures in Section 5, providing some 'materials' for possible future research. Section 6 summarizes and concludes.

## 2 Data and Experiment Design

The data employed in this study are ASFRs by year and age which are derived mainly from the Human Fertility Database[5] and the Eurostat Database.[6] In addition, we thank Dr. Joshua R. Goldstein and Dr. Michaela Kreyenfeld for providing us with the Japan data. Note that not all countries listed in the Human Fertility Database and the Eurostat Database are included in this study because their data ranges are not lengthy enough to derive actual CTFRs for at least 10 cohorts.

Table 1 presents the date ranges and related information of all-birth-combined and parity-specific ASFRs for 22 countries. As can be seen, non-parity specific fertility data are more readily available than parity-specific ones. Our experiment results may thus be more robust for the former, but results for the latter also provide very useful information. In addition, the comparison of experiment results across birth orders can exhibit the advantages and disadvantages of each forecasting method.

Instead of depleting all of the observations once only, we follow the process adopted in our previous study to yield more robust results. Specifically, we treat any period of 25 consecutive years during the data range as a sample, and there will be various numbers of samples for each country, depending on the length of the data range. Within each 25-year sample, all cohort fertility schedules are incomplete; some lack their past records, some lack

[5] Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available at `http://www.humanfertility.org` (data downloaded on Dec. 28, 2010).

[6] Available at `http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search\_database` (data downloaded on Dec 28, 2010).

Table 1. Information on ASFR Data for 22 Countries

| | All-Birth-Combined | | | Parity-Specific | | |
|---|---|---|---|---|---|---|
| | Data range | | Complete cohorts in data range | Data range | | Complete cohorts in data range |
| Country | Year | Age | | Year | Age | |
| Austria | 1951–2008 | 14–49 | 1937–1958 | 1984–2008 | 14–49 | |
| Bulgaria | 1960–2008 | 15–49 | 1945–1958 | | | |
| Canada | 1921–2007 | 14–49 | 1907–1957 | 1944–2007 | 14–49 | 1930–1957∗ |
| Czech | 1950–2008 | 14–49 | 1936–1958 | 1950–2008 | 14–49 | 1936–1958 |
| Denmark | 1960–2008 | 15–49 | 1945–1958 | | | |
| Estonia | 1959–2007 | 14–49 | 1945–1957 | 1959–2007 | 14–49 | 1945–1957 |
| Finland | 1939–2008 | 14–49 | 1925–1958 | 1982–2008 | 14–49 | |
| France | 1946–2007 | 14–49 | 1934–1957 | | | |
| W. Germany | 1956–2008 | 14–49 | 1942–1958 | | | |
| Greece | 1961–2008 | 15–49 | 1946–1958 | | | |
| Hungary | 1960–2008 | 15–49 | 1945–1958 | | | |
| Iceland | 1963–2008 | 15–49 | 1948–1958 | | | |
| Italy | 1960–2007 | 15–49 | 1945–1957 | | | |
| Japan | 1950–2006 | 15–49 | 1935–1956 | 1968–2006 | 15–49 | 1953–1956 |
| Netherlands | 1950–2008 | 14–49 | 1936–1958 | 1950–2008 | 14–49 | 1936–1958 |
| Norway | 1961–2008 | 15–49 | 1946–1958 | | | |
| Portugal | 1960–2008 | 15–49 | 1945–1958 | | | |
| Russia | 1959–2008 | 14–49 | 1945–1958 | 1959–2008 | 14–49 | 1945–1958 |
| Slovakia | 1950–2008 | 14–49 | 1936–1958 | 1950–2008 | 14–49 | 1936–1958 |
| Sweden | 1891–2007 | 14–49 | 1877–1957∗ | 1970–2007 | 14–49 | 1956–1957 |
| Switzerland | 1944–2007 | 14–49 | 1930–1957 | | | |
| U.S. | 1917–2006 | 14–49 | 1903–1956∗ | 1917–2006 | 14–49 | 1903–1956∗ |

Source: Countries whose data come from the Human Fertility Database include Austria, Canada, Czech, Estonia, Finland, France, (West) Germany, Netherlands, Russia, Slovakia, Sweden, Switzerland, and the U.S. The Japan data are provided by Dr. Joshua R. Goldstein and Dr. Michaela Kreyenfeld. Data of the remaining countries are from the Eurostat Database.

Note: ∗ indicates the countries that are selected as examples in this paper for illustration purpose due to their lengthy data spans.

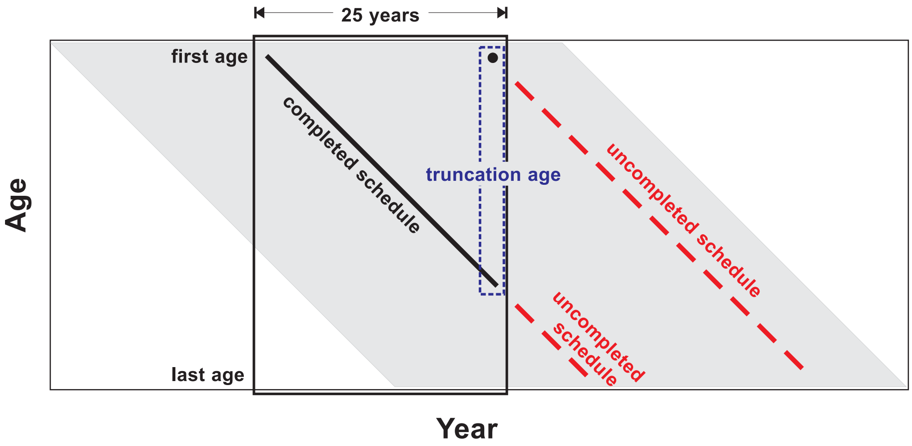



Figure 1. An illustration of how the data is utilized

Note: Given a 25-year sample, there are 25 cohorts who lack their future fertility schedules only and whose ages in the last year of the sample are denoted by *truncation ages*. The shaded parallelogram indicates the full information regarding cohort fertility schedules that can be derived from the entire data range.

their future experiences, and some lack both. In other words, no actual CTFRs can be derived from any single 25-year sample. Both conventional models (including the Naive, the Evans, the linearized Gompertz, and the linear extrapolation method) and smoothed period measures (including the BF, the KP, and the TFR.dagger) will be applied to all 25-year samples of ASFRs, and their CTFR forecasts can thus be evaluated, based on their approximation to the actual CTFR values.[7] To measure the extent to which a forecasted CTFR ($\hat{C}$) approximates to the corresponding actual value ($C$), we adopt the absolute percentage error (APE)

$$\text{APE} = \frac{|\hat{C} - C|}{C} \times 100\%$$

as the criterion.

Figure 1 illustrates how we utilize the entire data set (represented by the largest year-age rectangle) for each country listed in Table 1. The shaded parallelogram indicates the full information regarding complete cohort fertility schedules that can be derived from the

[7]We are not going to introduce these models here, for all of them have been described in detail either in their original papers or in Introduction.

entire data range, including actual CTFR values. As can be seen, within a particular 25-year rectangle there will be 25 cohorts whose early fertility experiences (represented by diagonal solid lines) are fully observed, while their subsequent schedules (represented by diagonal dashed lines) are not. We denote the age in the last year of the sample for a particular cohort by *truncation age*; a younger truncation age represents a larger proportion of fertility which has not been completed yet. All forecasting methods under investigation will be applied to this particular 25-year sample to derive their forecasted CTFR values (and the corresponding APEs if the actual CTFRs are available) for these 25 cohorts. As the 25-year sample moves along the year dimension, a few APE values for each truncation age can be collected to construct a distribution. Note that the number of APE values one can collect differs across truncation ages due to the availability of actual CTFRs.

## 3 Graphical Illustrations of Forecasted CTFR Values for Selected Countries

Before presenting the summary statistics of forecast errors, we first plot the forecast results by various methods for selected countries, and then compare these values with actual CTFRs graphically. To save space, we select Sweden and the U.S. as examples for non-parity specific data and Canada and the U.S. for parity-specific (from order 1 to order 3) data, due to their lengthy data spans. Consequently, there will be 8 figures (Figures 2–9) in total and 16 panels in each figure. One figure is for one selected country at a specific birth order, and each panel in a figure exhibits the forecast results derived by a conventional method or by a period measure with a specific degree of smoothing.[8] In a panel, there are numerous whiskers, each depicting the forecasts derived from a particular 25-year sample, and we use a curve marked with hollow squares to depict the actual CTFRs. Note that the minimum

[8]We adopt the kernel smoother and a higher bandwidth represents a higher degree of smoothing.

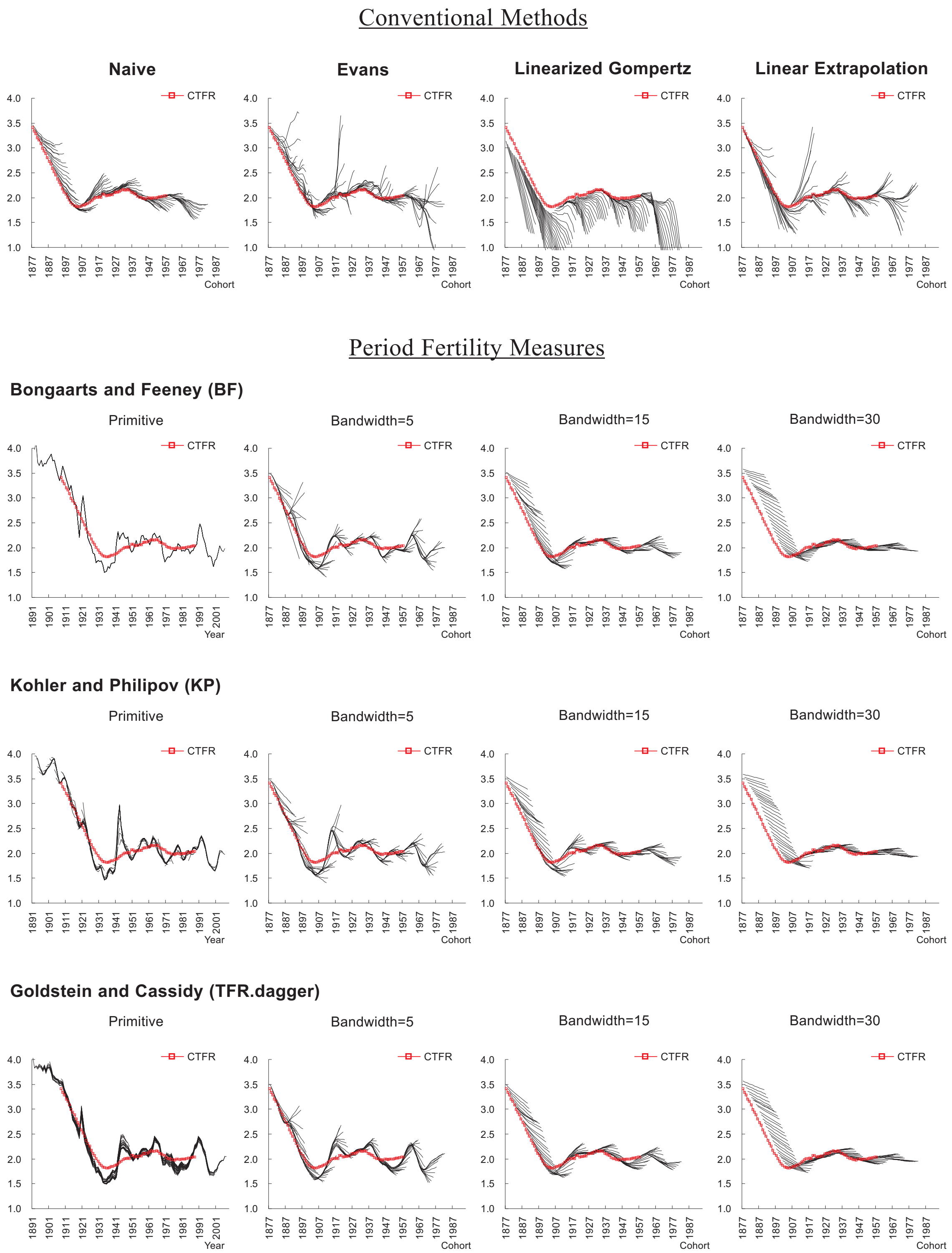



Figure 2. All-Birth-Combined CTFR Forecasts by Conventional Methods and Period Measures: Sweden

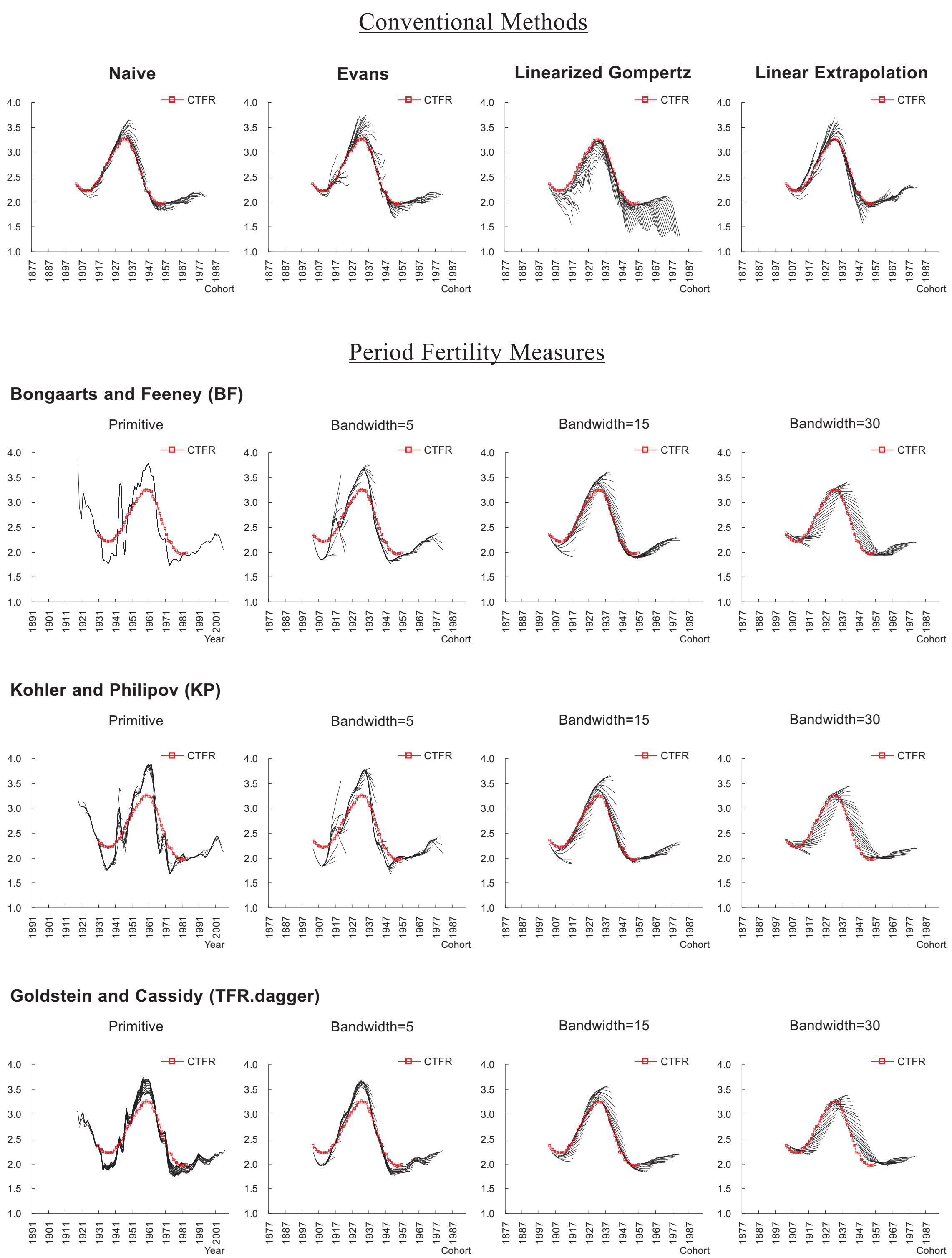



Figure 3. All-Birth-Combined CTFR Forecasts by Conventional Methods and Period Measures: the U.S.

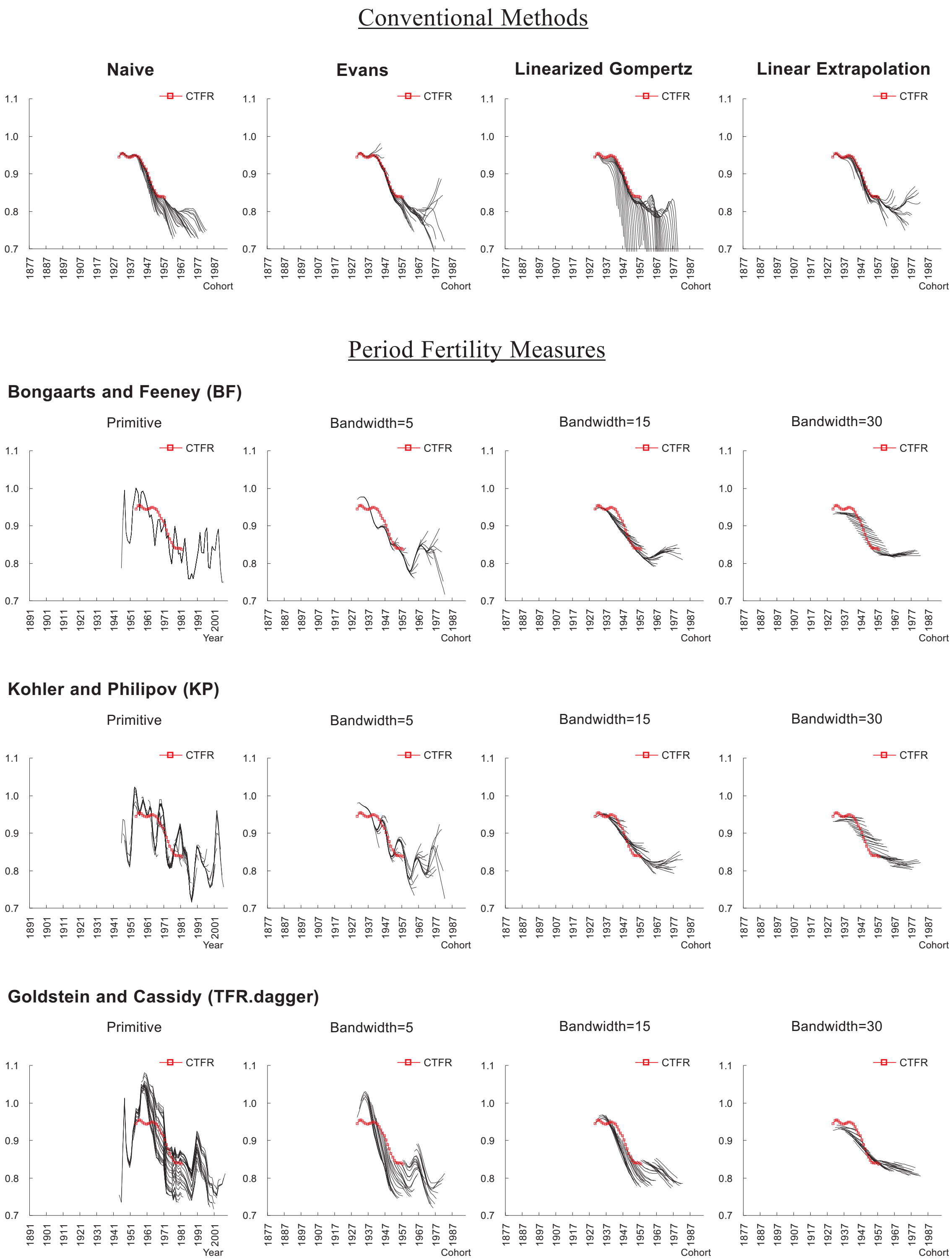



Figure 4. First-Birth CTFR Forecasts by Conventional Methods and Period Measures: Canada

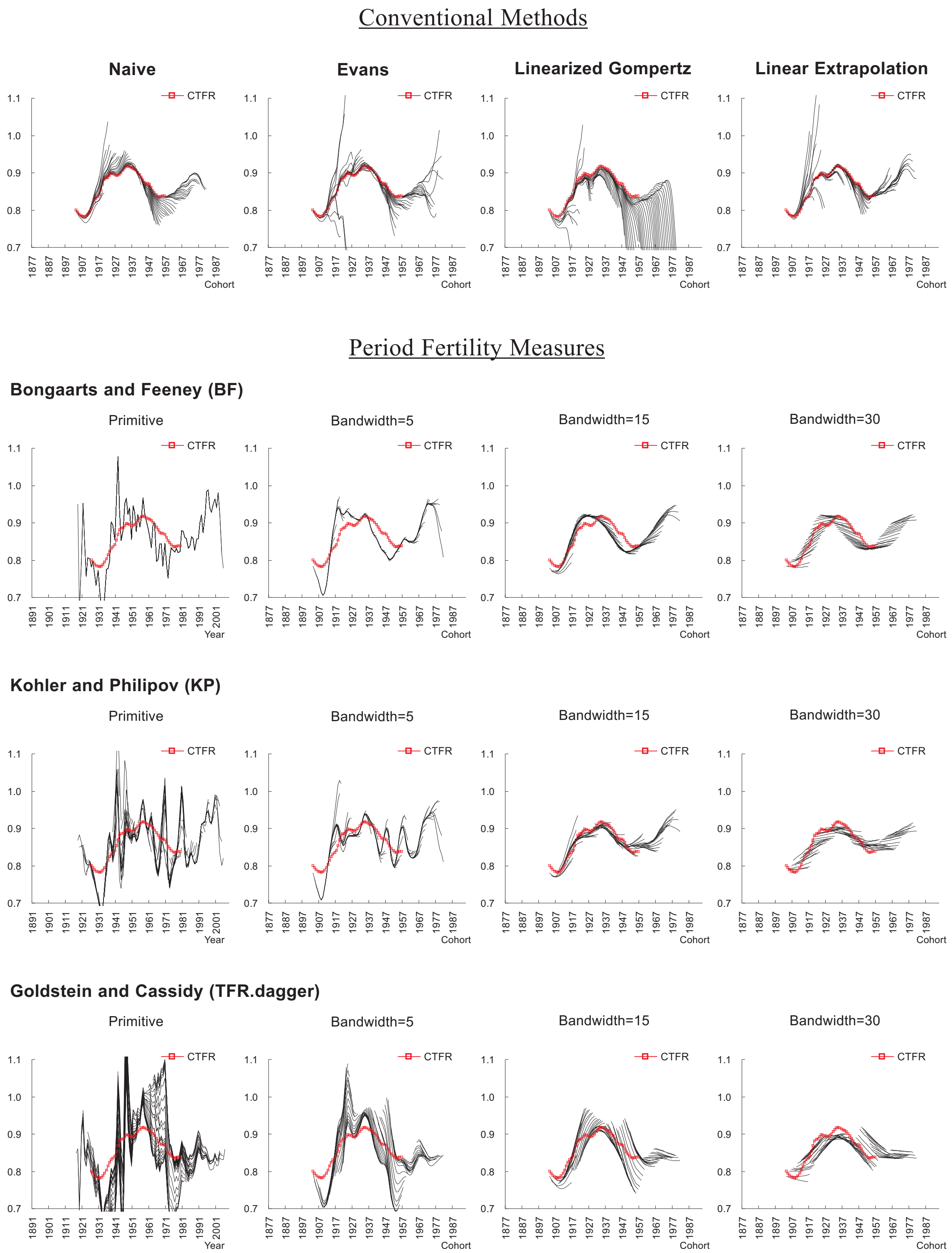



Figure 5. First-Birth CTFR Forecasts by Conventional Methods and Period Measures: the U.S.

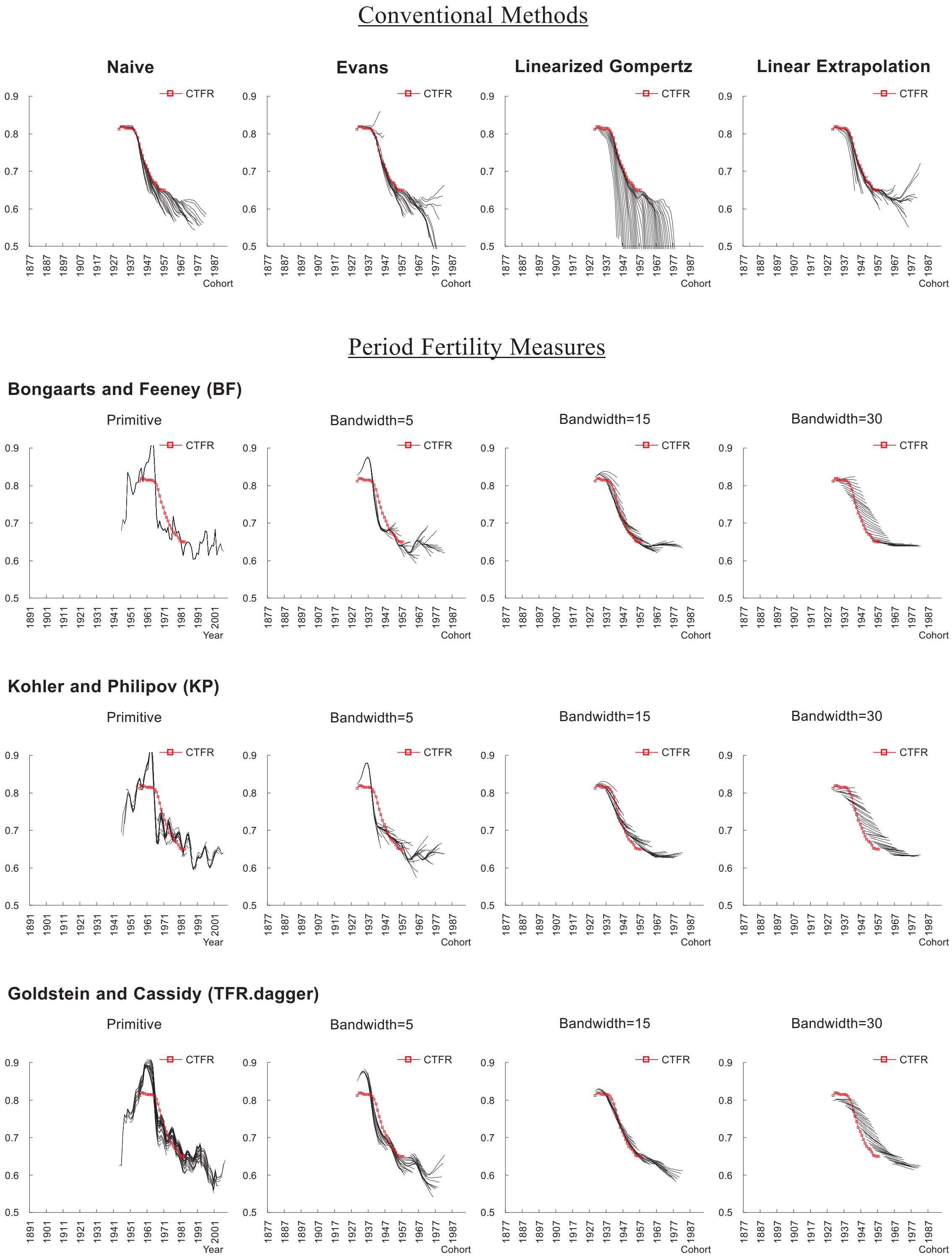



Figure 6. Second-Birth CTFR Forecasts by Conventional Methods and
Period Measures: Canada

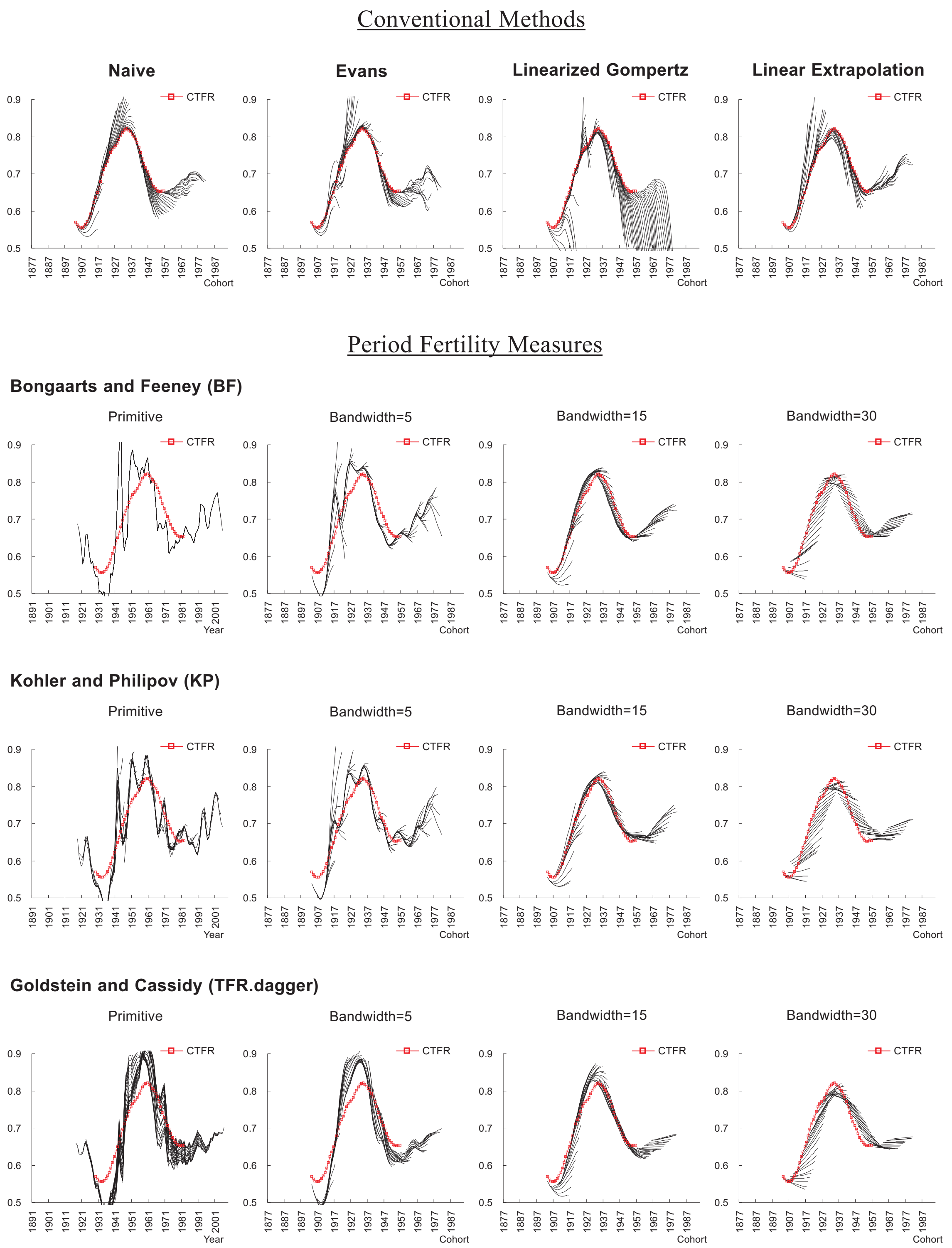



Figure 7. Second-Birth CTFR Forecasts by Conventional Methods and Period Measures: the U.S.


Figure 8. Third-Birth CTFR Forecasts by Conventional Methods and Period Measures: Canada

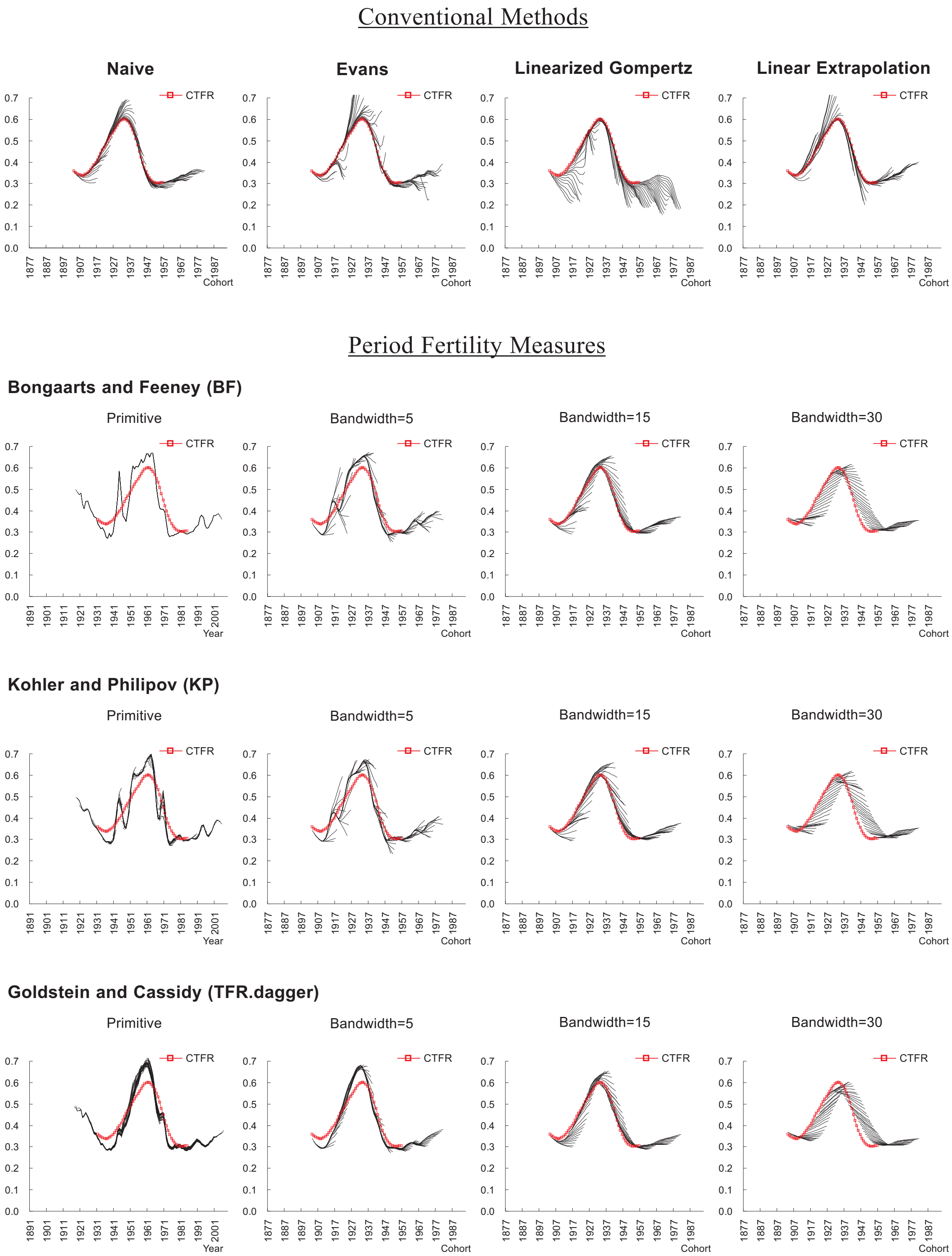



Figure 9. Third-Birth CTFR Forecasts by Conventional Methods and Period Measures: the U.S.

truncation age for cohorts in a whisker is set to 24.

At a first glance through these 8 figures, all panels look alike in the sense that the forecasts are expected to approach the actual CTFRs. Without knowing the exact method behind each panel, even the researchers who deny the usefulness of the Naive and period measures based on theoretical ground are very likely to be convinced that these forecasts must be generated by some acceptable model. If one is asked to judge which panel performs best by visual inspection only, the Naive and the smoothed period measures (bandwidth=15) are hardly possible to be excluded from the top few candidates.

Moreover, one may obtain from these figures the following observations:

1. Forecasts tend to departure from the actual CTFR curves around the turning points. This may conclude that predicting these turning points would be important and helpful to improve the forecasts.

2. Forecasts by the linearized Gompertz method tend to display downward and dramatic deviations. According to our experiment, the forecasts are found to be very sensitive to the value of a parameter (which is called the drift parameter in Myrskylä and Goldstein 2010). As we change the way in estimating this parameter, results may vary a lot but the dramatic deviations still remain.

3. As the 25-year sample moves, the Evans method is most likely to produce unstable and occasionally extreme deviations. By contrast, forecasts by the Naive and by the smoothed period measures with bandwidth higher than 15 tend to change in a smooth way.

4. As the degree of smoothing on period measures gets higher, the extent of fluctuations around the actual CTFR values is significantly alleviated, and all the whiskers are stretched closer to straight lines. In fact, the kernel smoother is quite strong when

the bandwidth parameter is set to 15; smoothers such as the simple moving average may have to be utilized recursively for many times to obtain the same effect.

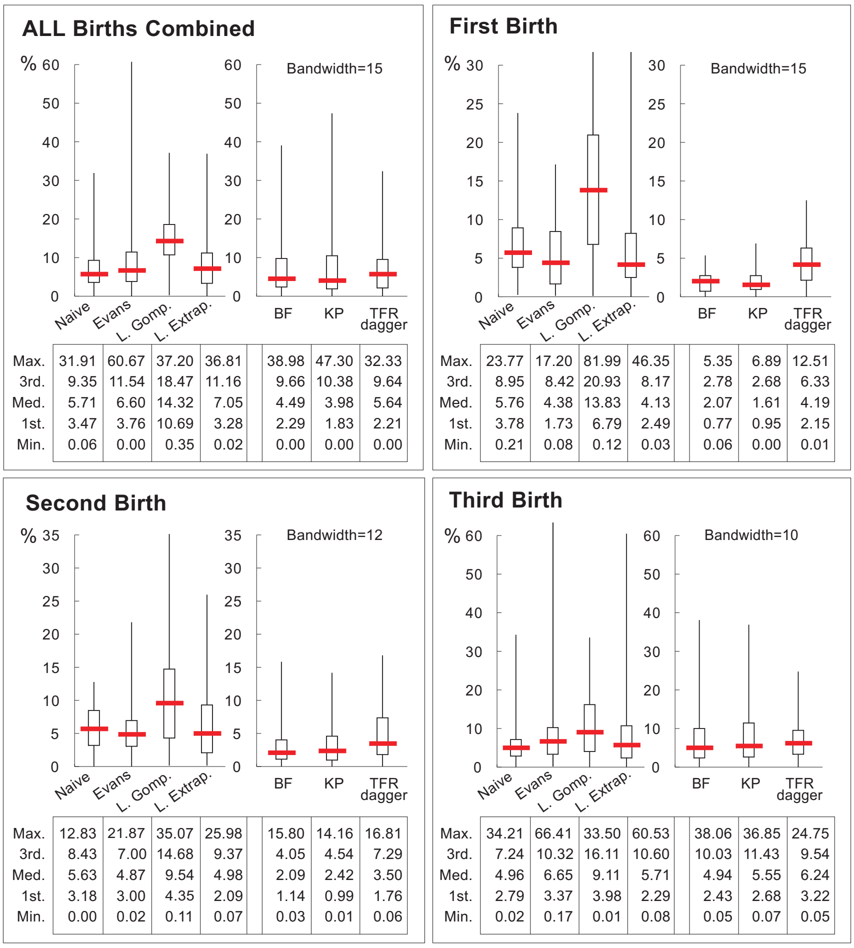
Given these observations, however, some questions still remain unanswered. Does there exist an optimal degree of smoothing which is common across countries so that the forecast errors can be minimized? Which period measure is the best option? Can the smoothed period measures outperform conventional ones in terms of approximating the actual CTFR values? Do the answers to these questions vary with birth orders? In the next section, we turn to summary statistics of forecast errors to see whether these questions can be well responded.

## 4 Distributions of Forecast Errors

As mentioned above, one may collect forecast errors measured by APE for each truncation age to construct a distribution. In addition, each truncation age must correspond to a percentile at which the fertility level of a cohort has been completed, while the truncation percentile associated to a specific truncation age varies by cohort and/or by country. In consequence, competing approaches investigated in this paper will be compared using the distribution of forecast errors by truncation percentile rather than by truncation age. More specifically, we will construct distributions of forecast errors for cohorts whose truncation percentiles fall within interval [50, 65).[9]

Since this paper adopts 7 methods and takes data of various birth orders from 22 countries, it would be cumbersome if we display all results of forecast errors country by country.

[9] Myrskylä and Goldstein (2010) indicated that "[e]xperiments with the Gompertz model ... suggested that the proportion should be close to 2/3 before reasonable fit can be expected". We thus select interval [65, 75) as a response to their proposed percentile for comparison, and select interval [50, 65) as a challenge to see how well the performances of these competing approaches can be. But in the main text we focus on results within interval [50, 65) and leave results within interval [65, 75) to the appendix.




**ALL Births Combined**

| | Naive | Evans | L. Gomp. | L. Extrap. | BF | KP | TFR dagger |
|---|---|---|---|---|---|---|---|
| Max. | 31.91 | 60.67 | 37.20 | 36.81 | 38.98 | 47.30 | 32.33 |
| 3rd. | 9.35 | 11.54 | 18.47 | 11.16 | 9.66 | 10.38 | 9.64 |
| Med. | 5.71 | 6.60 | 14.32 | 7.05 | 4.49 | 3.98 | 5.64 |
| 1st. | 3.47 | 3.76 | 10.69 | 3.28 | 2.29 | 1.83 | 2.21 |
| Min. | 0.06 | 0.00 | 0.35 | 0.02 | 0.00 | 0.00 | 0.00 |

**First Birth**

| | Naive | Evans | L. Gomp. | L. Extrap. | BF | KP | TFR dagger |
|---|---|---|---|---|---|---|---|
| Max. | 23.77 | 17.20 | 81.99 | 46.35 | 5.35 | 6.89 | 12.51 |
| 3rd. | 8.95 | 8.42 | 20.93 | 8.17 | 2.78 | 2.68 | 6.33 |
| Med. | 5.76 | 4.38 | 13.83 | 4.13 | 2.07 | 1.61 | 4.19 |
| 1st. | 3.78 | 1.73 | 6.79 | 2.49 | 0.77 | 0.95 | 2.15 |
| Min. | 0.21 | 0.08 | 0.12 | 0.03 | 0.06 | 0.00 | 0.01 |

**Second Birth**

| | Naive | Evans | L. Gomp. | L. Extrap. | BF | KP | TFR dagger |
|---|---|---|---|---|---|---|---|
| Max. | 12.83 | 21.87 | 35.07 | 25.98 | 15.80 | 14.16 | 16.81 |
| 3rd. | 8.43 | 7.00 | 14.68 | 9.37 | 4.05 | 4.54 | 7.29 |
| Med. | 5.63 | 4.87 | 9.54 | 4.98 | 2.09 | 2.42 | 3.50 |
| 1st. | 3.18 | 3.00 | 4.35 | 2.09 | 1.14 | 0.99 | 1.76 |
| Min. | 0.00 | 0.02 | 0.11 | 0.07 | 0.03 | 0.01 | 0.06 |

**Third Birth**

| | Naive | Evans | L. Gomp. | L. Extrap. | BF | KP | TFR dagger |
|---|---|---|---|---|---|---|---|
| Max. | 34.21 | 66.41 | 33.50 | 60.53 | 38.06 | 36.85 | 24.75 |
| 3rd. | 7.24 | 10.32 | 16.11 | 10.60 | 10.03 | 11.43 | 9.54 |
| Med. | 4.96 | 6.65 | 9.11 | 5.71 | 4.94 | 5.55 | 6.24 |
| 1st. | 2.79 | 3.37 | 3.98 | 2.29 | 2.43 | 2.68 | 3.22 |
| Min. | 0.02 | 0.17 | 0.01 | 0.08 | 0.05 | 0.07 | 0.05 |

Figure 10. Distributions of Forecast Errors by Conventional Methods and Period Measures: Truncation percentiles in [50, 65)

Related figures for each single country will be therefore omitted but available upon request from the authors. However, it is worthwhile mentioning that there is not a single method, conventional or period measure with a specific degree of smoothing, which can best applied to all countries. We thus turn to comparing the performances of these competing approaches without distinguishing countries from one another.

Figure 10 presents the overall distributions of forecast errors with box-whisker plots by method and by parity. It can be found that:

1. For non-parity specific data, more than half errors are smaller by smoothed period measures (bandwidth=15) than by conventional methods, but smoothed period

measures also produce a quarter of errors above 10%. In addition, the Naive seems to be the best method among conventional ones. It is not easy to judge which of the Naive, the BF, the KP, and the TFR.dagger performs best by simple visual inspection.

2. For first birth data, however, it is obvious that either the BF or the KP (bandwidth=15) performs better than other methods. But it is still not easy to determine the best one.
3. Similarly, the BF and the KP (bandwidth=12) perform better than others for second birth data, but they produce higher maximum errors than the Naive method does.
4. For third birth data, the Naive again seems to be the best among conventional methods, but it can produce errors above 30%.

One can further compare distributions of forecast errors with some reasonable criteria. Here we propose to apply the first- and the second-order stochastic dominance concepts in deciding which method performs best. More specifically, there is no doubt that distribution $A$ can be said to be better than distribution $B$ if $A$ is first-order stochastically dominated by $B$, i.e.,

$$B(x) \leq A(x) \text{ for every } x, \tag{FSD}$$

where $x$ denotes the forecast error. But sometimes such a clear-cut relationship between distributions may not be found. The second-order stochastic dominance criterion can then be applied: if

$$\int_0^x B(t)dt \leq \int_0^x A(t)dt \text{ for every } x, \tag{SSD}$$

then $A$ is said to be better than $B$. Based on these two criteria, we conclude from our calculations that:

1. For non-parity specific data, not a single method can be considered as the best using the FSD criterion. But by the SSD criterion, the KP (bandwidth=15) outperforms other methods, although the advantage is not much.
2. For first birth data, the BF (bandwidth=15) is first-order stochastically dominated by every other method except the KP (bandwidth=15). These two period measures are on a par with each other and they both outperform other methods significantly by the FSD and the SSD.
3. For second birth data, no best method can be determined by the FSD, but the BF (bandwidth=12) outperforms others by the SSD.
4. Strictly speaking, there is not a single method which performs best for third birth data according to the FSD and the SSD. Nevertheless, the Naive tends to be considered as the best one by the SSD criterion.

For first and second births data, the performances of the BF (and/or the KP) can be satisfactory since more than three quarters of forecast errors are below 3% and 5%, respectively. By contrast, however, neither conventional methods nor smoothed period measures perform sufficiently well when they are applied to all-birth-combined data since the largest forecast errors can be above 30%. This drawback is especially severe when considering the fact that many countries have non-parity specific data only but lack parity specific data. Can period measures provide some useful information that help to reduce forecast errors?

## 5 Further Information Contained in Period Measures

As aforementioned in Section 3, forecasts tend to departure from the actual CTFR curves around the turning points. Methods that can be utilized in predicting these turning points are thus called for. We are not going to propose any method of that sort in this paper, but instead will present some further information that can be helpful.

For each 25-year sample, one can compute the rate of change in forecast values (the slope) by any approach around the cohort which is the youngest in the first year of the sample. Connecting the slope values from all 25-year samples yields a curve, denoted by the *slope curve* hereafter, and we find the slope curves of the Naive method very zigzag.[10] By contrast, Figure 11 shows that 1) the slope curves of smoothed period measures look very smooth, and, most importantly, 2) the points at which these slope curves pass through the horizontal axis mostly coincide the turning points of the actual CTFR curves, which are marked by hollow squares in the figure.

Based on these two facts, one may utilize the slope information to predict the turning points. For example, if we have a data set spanning 35 years, we may subdivide them into eleven 25-year subperiods and thus derive eleven slope values. Fitting these values into a particular function, it can be possible to determine the points whose slope values are zero. We leave this work to our future research.

## 6 Summary and Conclusions

Based on a methodological point that comparing period measures with cohort ones is meaningless, one may not agree with the strategy we propose to forecast CTFR using adjusted period measures such as the BF, the KP, and the TFR.dagger. However, the empirical evidence provided in this paper is so convincing that there must be some useful information contained in the period measures; smoothing these measures is a possible way to extract information that helps to forecast CTFR values.

[10] The figure is not shown in this paper but available upon request from the authors.

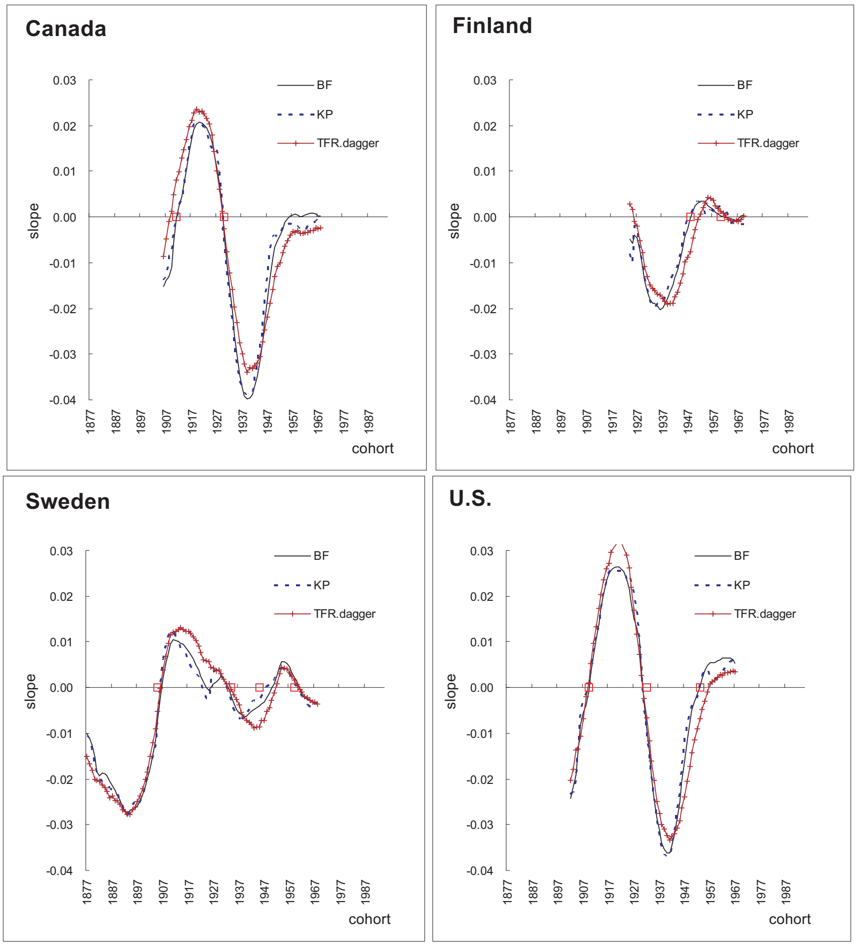


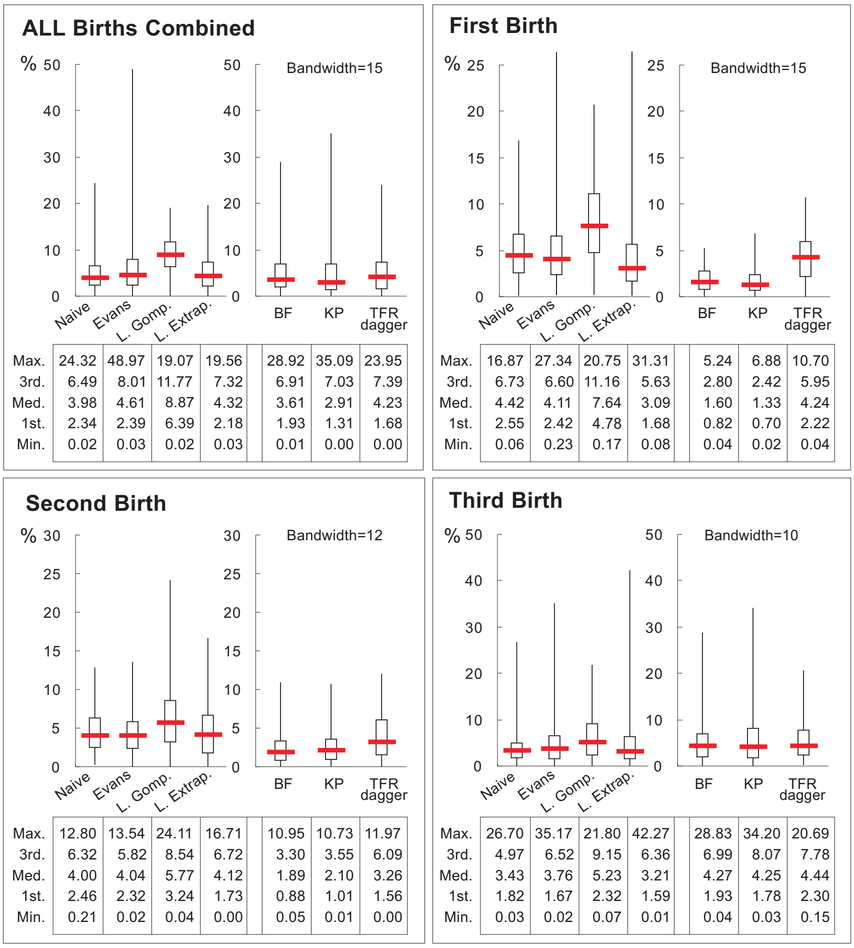

Figure 11. Slope Information of All-Birth-Combined Data Derived from Period Measures for Selected Countries

Note: Hollow squares in each panel represent the actual turning points of the actual CTFR curves and the slope curves are derived under the setting of bandwidth=30.

In this paper, we show that smoothed period measures, including the BF, the KP, and the newly developed TFR.dagger, do help to forecast CTFR values and outperform the four conventional methods significant for the first and the second births. Among the three period measures, we conclude that the BF can be a better choice than the other two because of its excellent performance and its easy computation property. Although the performances of period measures are not significantly better than that of conventional methods in forecasting the all-birth-combined and the third CTFRs, they do provide some essential information that may be utilized to improve the forecasting accuracy. It requires interested researchers to devoted more effort in this line of research.

This paper can be viewed as a starting point to consider the usefulness of adjusted period measures as an instrument for predicting CTFR values. A possible direction for future research is to establish a theretical framework in explaining the results revealed in this paper. Since smoothing a series of values is in effect to share information contained in these values with one another, it is thus important to figure out what kinds of information can be embedded in these period measures.

## Appendix




ALL Births Combined

| | Naive | Evans | L. Gomp. | L. Extrap. | BF | KP | TFR dagger |
|---|---|---|---|---|---|---|---|
| Max. | 24.32 | 48.97 | 19.07 | 19.56 | 28.92 | 35.09 | 23.95 |
| 3rd. | 6.49 | 8.01 | 11.77 | 7.32 | 6.91 | 7.03 | 7.39 |
| Med. | 3.98 | 4.61 | 8.87 | 4.32 | 3.61 | 2.91 | 4.23 |
| 1st. | 2.34 | 2.39 | 6.39 | 2.18 | 1.93 | 1.31 | 1.68 |
| Min. | 0.02 | 0.03 | 0.02 | 0.03 | 0.01 | 0.00 | 0.00 |

First Birth

| | Naive | Evans | L. Gomp. | L. Extrap. | BF | KP | TFR dagger |
|---|---|---|---|---|---|---|---|
| Max. | 16.87 | 27.34 | 20.75 | 31.31 | 5.24 | 6.88 | 10.70 |
| 3rd. | 6.73 | 6.60 | 11.16 | 5.63 | 2.80 | 2.42 | 5.95 |
| Med. | 4.42 | 4.11 | 7.64 | 3.09 | 1.60 | 1.33 | 4.24 |
| 1st. | 2.55 | 2.42 | 4.78 | 1.68 | 0.82 | 0.70 | 2.22 |
| Min. | 0.06 | 0.23 | 0.17 | 0.08 | 0.04 | 0.02 | 0.04 |

Second Birth

| | Naive | Evans | L. Gomp. | L. Extrap. | BF | KP | TFR dagger |
|---|---|---|---|---|---|---|---|
| Max. | 12.80 | 13.54 | 24.11 | 16.71 | 10.95 | 10.73 | 11.97 |
| 3rd. | 6.32 | 5.82 | 8.54 | 6.72 | 3.30 | 3.55 | 6.09 |
| Med. | 4.00 | 4.04 | 5.77 | 4.12 | 1.89 | 2.10 | 3.26 |
| 1st. | 2.46 | 2.32 | 3.24 | 1.73 | 0.88 | 1.01 | 1.56 |
| Min. | 0.21 | 0.02 | 0.04 | 0.00 | 0.05 | 0.01 | 0.00 |

Third Birth

| | Naive | Evans | L. Gomp. | L. Extrap. | BF | KP | TFR dagger |
|---|---|---|---|---|---|---|---|
| Max. | 26.70 | 35.17 | 21.80 | 42.27 | 28.83 | 34.20 | 20.69 |
| 3rd. | 4.97 | 6.52 | 9.15 | 6.36 | 6.99 | 8.07 | 7.78 |
| Med. | 3.43 | 3.76 | 5.23 | 3.21 | 4.27 | 4.25 | 4.44 |
| 1st. | 1.82 | 1.67 | 2.32 | 1.59 | 1.93 | 1.78 | 2.30 |
| Min. | 0.03 | 0.02 | 0.07 | 0.01 | 0.04 | 0.03 | 0.15 |

Figure 12. Distributions of Forecast Errors by Conventional Methods and Period Measures: Truncation percentiles in [65, 75)

## REFERENCES


Bloom, D.E. (1982). What's happening to the age at first birth in the United States? A study of recent cohorts. *Demography* 19(3): 351-370.

Bongaarts, J. and Feeney, G. (1998). On the quantum and tempo of fertility. *Population and Development Review* 24(2): 271-291.

Chandola, T., Coleman, D.A., and Hiorns, R.W. (1999). Recent European fertility patterns: Fitting curves to 'distorted' distributions. *Population Studies* 53: 317-329.

Chen, R. and Morgan, S.P. (1991). Recent trends in the timing of first births in the United States. *Demography* 28(4): 513-533.

Cheng, P.C.R. and Lin, E.S. (2010). Completing incomplete cohort fertility schedules. *Demographic Research* 23(9): 223-256.

Evans, M.D.R. (1986). American fertility patterns: A comparison of white and nonwhite cohorts born 1903–56. *Population and Development Review* 12(2): 267-293.

Goldstein, J.R. and Cassidy, T. (2010). Cohort postponement and period measures. MPIDR working paper WP-2010-015.

Kim, Y.J. and Schoen, R. (2000). On the quantum and tempo of fertility: Limits to the Bongaarts-Feeney adjustment. *Population and Development Review* 26(3): 554-559.

Kohler, H.-P. and Ortega, J. A. (2002). Tempo-adjusted period parity progression measures, fertility postponement and completed cohort fertility. *Demographic Research* 6: 92-144.

Kohler, H.-P. and Philipov, D. (2001). Variance effects in the Bongaarts-Feeney formula. *Demography* 38(1): 1-16.

Li, N. and Wu, Z. (2003). Forecasting cohort incomplete fertility: A method and an application. *Population Studies* 57(3): 303-320.

Myrskylä, M. and Goldstein, J.R. (2010). Prediction and error propagation in innovation diffusion models, with applications to demographic processes. MPIDR working paper WP-2010-013.

Ryder, N.B. (1990). What is going to happen to American fertility. *Population and Development Review* 16(3): 433-454.

Schoen, R. (2004). Timing effects and the interpretation of period fertility. *Demography* 41(4): 801-819.

Smallwood, S. (2002). The effect of changes in timing of childbearing on measuring fertility in England and Wales. *Population Trends* 109: 36-45.

Sobotka, T. (2003). Tempo-quantum and period-cohort interplay in fertility changes in Europe: Evidence from the Czech Republic, Italy, the Netherlands and Sweden. *Demographic Research* 8: 151-214.

Van Imhoff, E. and Keilman, N. (2000). On the quantum and tempo of fertility: Comment. *Population and Development Review* 26(3): 549-553.

Van Imhoff, E. (2001). On the impossibility of inferring cohort fertility measures from period fertility measures. *Demographic Research* 5: 23-64.

Willekens, F. and Baydar, N. (1984). Age-period-cohort models for forecasting fertility. Working Paper No. 45, NIDI, The Hague.