On the Quantum of Fertility: A Bias Correction Approach Using the Slope Information

P. C. Roger Cheng [*]

A very preliminary version
Please do not cite

Abstract

Given the fact that a satisfactory estimate of cohort fertility depends crucially on an accurate prediction of the future trend of tempo-adjusted period quantum, this paper shows that one can utilize available fertility data to disclose some useful information about that trend so as to effectively correct the prediction bias, especially in the presence of a strong quantum change. Specifically, we extract clues about both the slope and the change of slope in current quantum movements and then regress uncorrected prediction errors on these related variables using a large number of experiments which come from the existing fertility data of Canada, the U.S., and 23 European countries. Empirical evidence demonstrates a surprising high goodness of fit of our regression models, and the prediction bias can thus be significantly corrected based on the revealed relationship.

Introduction

In a previous study, Cheng and Goldstein (2011, attached below) specified a simple but definite relationship between period and cohort fertility measures, i.e.

$$\mathrm{CFR}(c) = \sum_{a=15}^{44} f(a, c+a) = \sum_{t=c+15}^{c+44} w(t-c, t)\, \mathrm{BF}(t), \tag{1}$$

where $f(a,t)$ denotes the ASFR at age a in year t, $p(a,t) = f(a,t)/\mathrm{TFR}(t)$ denotes the period age-specific fertility proportion, $r(t)$ represents the change in the mean age of childbearing at time t, and $w(a,t) = \left[1 - r(t)\right] p(a,t)$. Mathematically, the

[*]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.

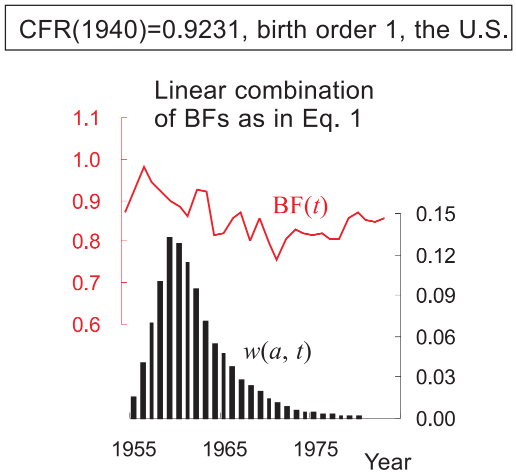

Figure 1: An Illustration of the Period-Cohort Relationship Specified in Eq. (1)

completed fertility of a cohort is equivalent to a linear combination of period fertility rates over the cohort's childbearing years, and Figure 1 illustrates this identity using the first-birth CFR of cohort 1940 (whose prime childbearing years were 1955–84) in the U.S. as an example. Furthermore, the cohort fertility rate (CFR hereafter) curve is actually a smoothed version of the BF (i.e. the tempo-adjusted measure proposed by Bongaarts and Feeney, 1998) curve.

To predict a cohort's lifetime fertility (C) one must encounter a situation that childbearing is unfinished (suppose it is truncated at age A), and C is accordingly divided into an observed past component (C_1) and a remaining future component (C_2):

$$\begin{aligned} C &= C_1 + C_2 \\ &= \sum_{a=15}^{A} f(a, c+a) + \sum_{a=A+1}^{44} f(a, c+a). \end{aligned} \quad (2)$$

Let $\widetilde{\mathrm{BF}}$ represent the weighted average of BFs over the unfinished years, C_2 can be expressed as

$$C_2 = \sum_{a=A+1}^{44} w(a, c+a)\,\mathrm{BF}(c+a) = \widetilde{\mathrm{BF}} \sum_{a=A+1}^{44} w(a, c+a) \quad (3)$$

by following the logic embedded in Equation (1). The prediction of the unfinished fertility is then transformed into the product of an (average) anticipation of future movements in the BF and an inference on the sum of unfinished coefficients.

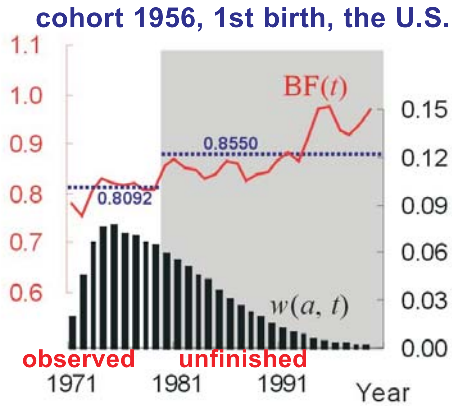

Figure 2: An Illustration of the Freeze BF-A Approach

Furthermore, if the shape of period fertility schedules is assumed invariant during the unfinished period (rather than the entire lifetime), one may derive

$$\sum_{a=A+1}^{44} w(a, c+a) = \sum_{a=A+1}^{44} p(a, T), \tag{4}$$

equalizing the sum of unfinished w's to the sum of observed p's in year T. By substituting this conversion into Equation (3), the cohort fertility is approximated to

$$\mathrm{CFR}(c) = C_1 + \widetilde{\mathrm{BF}} \sum_{a=A+1}^{44} p(a, T), \tag{5}$$

a framework focusing on the prediction of $\widetilde{\mathrm{BF}}$.

Cheng and Goldstein (2011) proposed a prediction approach, named Freeze BF-A, which uses $\overline{\mathrm{BF}}$ (the weighted average of BFs observed) as the estimator of $\widetilde{\mathrm{BF}}$. As illustrated by the example in Figure 2, $\widetilde{\mathrm{BF}} = 0.8550$ is the target to be estimated, whereas $\overline{\mathrm{BF}} = 0.8092$ is the estimate used by the Freeze BF-A method. The prediction error is mainly decided by the discrepancy between these two values. Empirical evidence across birth orders suggests that the Freeze BF-A predictor provides satisfactory estimates and outperforms the other selected methods for lower birth orders where the quantum change is mild, particularly when the observed cohort experience is truncated at a quite young age. In the presence of a strong quantum change, however, none of the prediction methods investigated in their paper demonstrate adequate performance. Another analysis across cohort subgroups within each birth order also identifies that the source of prediction error comes from the failure to estimate $\widetilde{\mathrm{BF}}$ accurately.

This paper is a follow-up research of Cheng and Goldstein (2011) and explores a possible solution to the quantum problem. The next section introduces the fertility data utilized in both studies and describe how experiments are designed to evaluate a predictor. In the section that follows, we introduce how BF slope information is connected to the prediction error of the Freeze BF-A method and how to use this relationship to correct the estimation bias so that the prediction errors can be reduced. Subsequently, empirical evidence from Canada, the U.S., and 23 European countries demonstrates how sigificant the improvement is. The final section summarizes and concludes.

Data and Experiment Design

Data

The data employed in this study are age-specific fertility rates by one-year period and by single-year age group, taken from the Human Fertility Database[1] and the Eurostat Database, downloaded in March, 2012. We restrict our analysis to the set of single-year age groups 15-44 and exclude countries whose data contain less than 10 completed cohorts, so not all countries listed in the databases are included. In addition, we use the separate schedules for East and West Germany, as well as for England/Wales and Scotland in the U.K., rather than their combined data. Thus there are 27 schedules in total, as listed in Table 1. Since non-parity-specific data are more readily available than parity-specific data, there are 907 completed cohorts from the former and 326 from the latter, respectively.

Experiment Design

For completed cohorts, ASFR data that span the entire childbearing years (and thus the actual CFR) are available. So, one can choose any integer between 16 and 43 to be the truncation age A as if the fertility were unfinished, and then use

[1] Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available at http://www.humanfertility.org.

Table 1: Data from the Human Fertility Database and the Eurostat

	all-birth-combined		parity-specific	
country	periods	completed cohorts	periods	completed cohorts
from the Human Fertility Database				
Austria	1951–2010	1936–1966 (31)	1984–2010	NA
Bulgaria	1947–2009	1932–1965 (34)	1947–2009	1932–1965 (34)
Canada	1921–2007	1906–1963 (58)	1944–2007	1929–1963 (35)
Czech Republic	1950–2009	1935–1965 (31)	1950–2009	1935–1965 (31)
Estonia	1959–2009	1944–1965 (22)	1959–2009	1944–1965 (22)
Finland	1939–2009	1924–1965 (42)	1982–2009	NA
France	1946–2009	1931–1965 (35)	NA	NA
Germany				
East	1956–2010	1941–1966 (26)	1956–1989	NI
West	1956–2010	1941–1966 (26)	NA	NA
Hungary	1950–2009	1935–1965 (31)	1952–2009	1937–1965 (29)
Lithuania	1959–2009	1944–1965 (22)	1970–2009	1955–1965 (11)
Netherlands	1950–2009	1935–1965 (31)	1950–2009	1935–1965 (31)
Portugal	1940–2009	1925–1965 (41)	1959–2009	1944–1965 (22)
Russia	1959–2009	1944–1965 (22)	1959–2009	1944–1965 (22)
Slovakia	1950–2009	1935–1965 (31)	1950–2009	1935–1965 (31)
Sweden	1891–2010	1876–1966 (91)	1970–2010	1955–1966 (12)
Switzerland	1932–2009	1917–1965 (49)	1998–2009	NA
U.K.				
England/Wales	1938–2009	1923–1965 (43)	NA	NA
Scotland	1945–2009	1930–1965 (36)	NA	NA
U.S.	1933–2007	1918–1963 (46)	1933–2007	1918–1963 (46)
from the Eurostat				
Belgium*	1954–2009	1939–1965 (27)	NA	NA
Denmark	1950–2010	1935–1966 (32)	NA	NA
Greece	1961–2010	1946–1966 (21)	NA	NA
Iceland	1963–2010	1948–1966 (19)	NA	NA
Italy	1952–2008	1937–1964 (28)	NA	NA
Norway	1961–2010	1946–1966 (21)	NA	NA
Spain	1971–2010	1956–1966 (11)	NA	NA

Note:
1. When a country is included in both databases, we prioritize data from the Human Fertility Database.
2. Countries whose data are insufficient to construct at least 10 completed cohorts (covering age groups 15–44) will be excluded from the analysis.
3. In parentheses are numbers of completed cohorts.
4. NA denotes that data are not available, while NI denotes that data are not included in the analysis because the number of completed cohorts is less than 10.

∗ Data in 2001 and 2002 are missing for Belgium. We use spline smoothing to interpolate these values by age.

the "observed" ASFRs to produce an estimated CFR; we denote this process as one experiment. There can be at most 28 experiments (each corresponding to a particular truncation age) for every completed cohort, and hence 25,396 and 9,128 experiments in total from non-parity- and parity-specific data, respectively.

To evaluate the prediction performance of a CFR predictor across experiments with various truncation ages, we employ the following prediction error index which measures how much proportion of the unfinished fertility has not been correctly estimated:

$$\text{Prediction Error} = \frac{\widehat{C} - C}{C - C_1} \times 100\% = \frac{\widehat{C}_2 - C_2}{C_2} \times 100\%. \tag{6}$$

For each birth order, we categorize experiments by completed proportion (defined as C_1/C),[2] construct a distribution of prediction error (or of error in absolute value) for each category, and then compare distributional statistics among approaches.

Uncorrected Prediction Error and the BF Slope

Since our experiments utilize fertility data of completed cohorts, we may have the actual value of C and then calculate an (uncorrected) prediction error (denoted by PE) for each Freeze BF-A prediction value of $\widehat{C}$. In addition, we may also construct a BF curve by parity for each country in our dataset and then derive related slope information — At any particular truncation time point T, a BF slope can be computed by

$$\text{FST}(T) = \text{BF}(T) - \text{BF}(T-1). \tag{7}$$

Note that the BF curve, in general, may not be smooth enough. Its slope curve, therefore, may become even more fluctuant and thus uninformative. To solve this problem, we use the G3GRID procedure in SAS code to create a smoothed Lexis surface from the original data and then construct smooth BF and FST curves. Figure 3 depicts the BF curve and its slope curve of the first parity in the U.S. before (thin black line) and after strong smoothing (thick red curve).

According to the logic mentioned in Introduction, a higher absolute value of FST tends to be associated with a larger discrepancy between $\widetilde{\text{BF}}$ and $\overline{\text{BF}}$, which

[2] Since the completed proportion by a particular truncation age may differ across countries, cohorts, and birth orders, categorizing experiments by completed proportion is more appropriate than by truncation age.

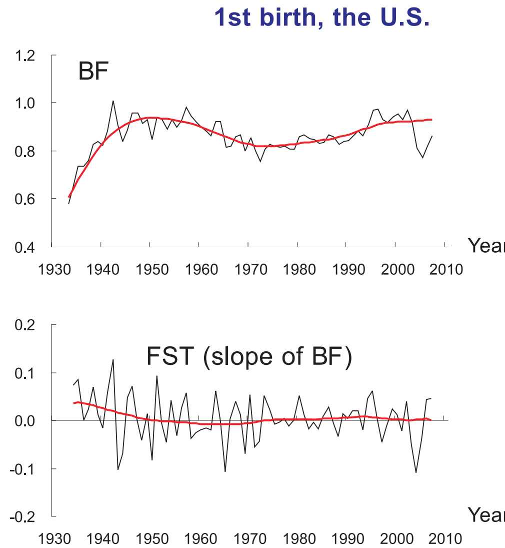

Figure 3: The BF Curve and its Slope Curve: Before and After Smoothing

corresponds to a larger absolute value of PE. In addition, when FST is positive (negative), $\overline{\mathrm{BF}}$ is more likely to be less (greater) than $\widetilde{\mathrm{BF}}$ such that the CFR is under (over) estimated and PE becomes negative (positive). In consequence, one may expect a negative correlation between PE and FST. Table 2 exhibits Pearson's correlation coefficient between PE and FST by birth order and degree of smoothing. Most coefficients in the table have an absolute value larger than 0.3, indicating a

Table 2: Pearson's Correlation Coefficient Between PE and FST

Birth Order smoothness	1	2	3+	all
NONE	−0.285	−0.342	−0.620	−0.354
MILD	−0.427	−0.577	−0.779	−0.601
MEDIUM	−0.497	−0.701	−0.861	−0.743
STRONG	−0.471	−0.748	−0.881	−0.810

moderate or high correlation of these two variables that we may utilize to correct the prediction bias. Figure 4 depicts a scatter plot to show the relationship between FST (strongly smoothed) and PE for all experiments of third birth and above with truncation percentage being between 10% and 85%.

Now let's turn to the bias correction procedure:

1. Establish the mechanism behind the uncorrected PE. For example, regress PE on FST

$$\mathrm{PE} = \alpha + \beta \cdot \mathrm{FST}$$

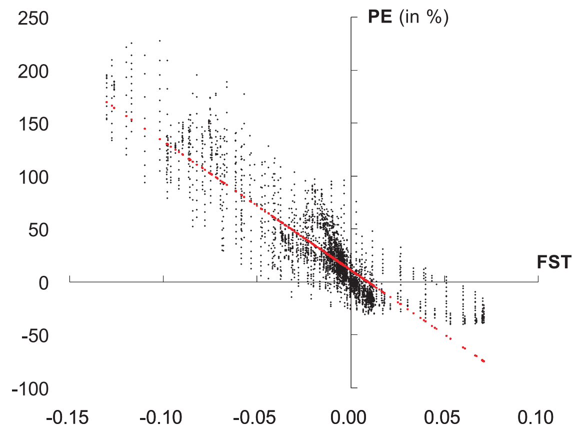

Figure 4: A Scatter Plot of FST and PE: Third Birth and Above

and obtain the estimates $\widehat{\alpha}$ and $\widehat{\beta}$.

2. Derive the fitted value of PE (denoted $\widehat{\mathrm{PE}}$) for a particular FST by

$$\widehat{\mathrm{PE}} = \widehat{\alpha} + \widehat{\beta} \cdot \mathrm{FST}.$$

3. Rearrange Eq. (6) and reestimate the CFR (denoted $\widetilde{C}$) by

$$\widetilde{C} = (\widehat{C} + C_1 \cdot \widehat{\mathrm{PE}})/(1 + \widehat{\mathrm{PE}}).$$

Thus, a bias corrected estimate can be obtained.

Besides FST, we may in fact construct more variables from the observed part to help capture the mechnism behind the uncorrected PE. For example, the difference of FST

$$\mathrm{SND}(T) = \mathrm{FST}(T) - \mathrm{FST}(T-1) \tag{8}$$

describes the concavity of the BF curve. In addition, a positively convex curve and a negatively concave curve both signify a strong likelihood of large discrepancies between $\widetilde{\mathrm{BF}}$ and $\overline{\mathrm{BF}}$. We thus define an indicator variable, STRONG, if the product of FST and SND is positive. Moreover, the sign of FST (POSITIVE=1 if FST> 0), the truncation age (denoted TAGE hereafter), the squared FST (denoted FST2), the squared SND (denoted SND2), and all interactions may matter. In the case of parity three and above, a full model using OLS gives an R-square as high as 0.9147 and most coefficient estimates are 1% significantly different from zero.

Table 3: Mean Absolute Prediction Error by Method, Completed Proportion, and Birth Order

Birth Order	Samples	Freeze Rates (1)	Freeze BF-A (2)	bias corrected Freeze BF-A: None (3)	Mild (4)	Medium (5)	Strong (6)
			completed proportion $\in [10\%, 30\%)$				
1	677	9.47	5.05	4.43	3.81	3.34	3.41
2	833	10.92	7.43	6.90	6.09	4.87	3.73
3+	998	31.29	34.23	19.16	16.23	11.27	8.83
all	2,759	17.03	12.74	11.25	9.52	7.23	4.90
			completed proportion $\in [30\%, 50\%)$				
1	597	10.76	5.28	4.83	4.05	3.51	3.32
2	687	12.52	8.02	7.01	6.01	4.40	3.15
3+	843	26.17	32.56	16.24	13.26	8.69	7.05
all	2,440	16.96	13.38	10.81	8.71	6.19	4.15
			completed proportion $\in [50\%, 65\%)$				
1	499	11.87	6.10	5.62	4.79	4.47	4.22
2	576	13.77	9.30	7.76	6.47	4.55	3.52
3+	673	23.42	34.12	15.56	12.48	8.24	6.79
all	1,999	17.15	14.69	11.24	8.58	5.76	4.11
			completed proportion $\in [65\%, 75\%)$				
1	415	12.23	6.67	6.26	5.59	5.21	4.83
2	451	14.35	10.61	8.50	7.07	5.03	4.25
3+	529	21.19	35.69	16.98	13.77	9.05	7.42
all	1,607	17.22	15.85	11.83	8.98	6.20	4.70
			completed proportion $\in [75\%, 85\%)$				
1	620	12.39	7.49	7.03	6.44	5.92	5.39
2	611	14.73	11.93	9.82	8.14	5.81	4.90
3+	679	18.66	37.32	17.23	14.09	10.43	8.52
all	2,119	17.63	17.98	12.99	10.31	7.69	5.64

Prediction Performance Without and With Bias Correction

Without considering the sign of prediction errors which indicates whether the CFR is over- or under-estimated, Table 3 summarizes the mean *absolute* prediction error of selected competing methods, with or without bias correction, across various categories of completed proportion and birth orders. Columns (1) and (2) replicate the results presented in Cheng and Goldstein (2011), and columns (3) through (6) are the corrected outcomes with various degree of smoothing. As can be seen, the prediction errors now drop significantly, especially for the third-and-above birth which is usually accompanied by a large quantum effect, even when no smoothing is implemented.

Conclusion

Although we have established an innovative procedure to correct the estimation bias due to quantum changes and obtain quite satisfactory results, there are still a few things unsolved yet. The most serious one can be the end-point problem caused by smoothing the BF curve. In this paper, we smooth over the entire data range so that the slope-related information in most experiments is quite correct. However, when one tries to employ the same procedure to get a corrected CFR prediction of the most recent cohort, the end-point problem is unavoidably encountered and will be more serious as the degree of smoothing gets higher. We therefore have to face a trade-off issue between reducing the prediction error and facing a higher risk of incorrect slope information. This is left to future research.

References

Bongaarts, John and Griffith Feeney (1998), "On the Quantum and Tempo of Fertility", *Population and Development Review*, 24(2), 271–291.

Cheng, P. C. Roger and Joshua R. Goldstein (2011), "On the Possibilities of Predicting Cohort Fertility Measures from Period Fertility Measures: Theory and Empirical Evidence", Period and Cohort Fertility Dynamics in the Developed World: The First Human Fertility Database Symposium, Rostock, Germany.

On the Possibilities of Predicting Cohort Fertility Measures from Period Fertility Measures: Theory and Empirical Evidence

P. C. Roger Cheng* Joshua R. Goldstein†

Abstract

This paper specifies a simple but definite relationship between period and cohort fertility measures. Following this nonstandard perspective, we derive a prediction framework which serves as a platform to compare seemingly unrelated methods, responds to the literature that casts doubts on the usefulness of period measures as cohort estimators, and provides a clear theoretical explanation on the major source of prediction errors. Empirical evidence from Canada, the U.S., and 23 European countries suggests that our predictor provides satisfactory estimates and outperforms some conventional methods when the quantum change is mild, particularly when the observed cohort experience is truncated at a very young age. In the presence of a strong quantum change, however, none of the prediction methods investigated in this paper demonstrate adequate performance.

Introduction

Can period fertility measures help to predict cohort fertility? The answer, in fact, depends heavily on how the question is perceived. This paper responds to this question, which has been extensively studied in the literature and generated mostly negative answers, from a *nonstandard* perspective based on a simple but definite relationship between period and cohort measures. Adopting this perspective can, on the one hand, waive most (if not all) challenges raised by previous studies. On the other hand, the prediction framework so derived serves as a platform to compare seemingly unrelated methods and provides a clear theoretical explanation on the

*Correspondence: P. C. Roger Cheng, Department of Economics, National Central University, Taoyuan 32001, Taiwan; E-mail: paochih@mgt.ncu.edu.tw; Tel: 886-3-422-7151 ext. 66310; Fax: 886-3-422-2876;

†Max Planck Institute for Demographic Research, Germany

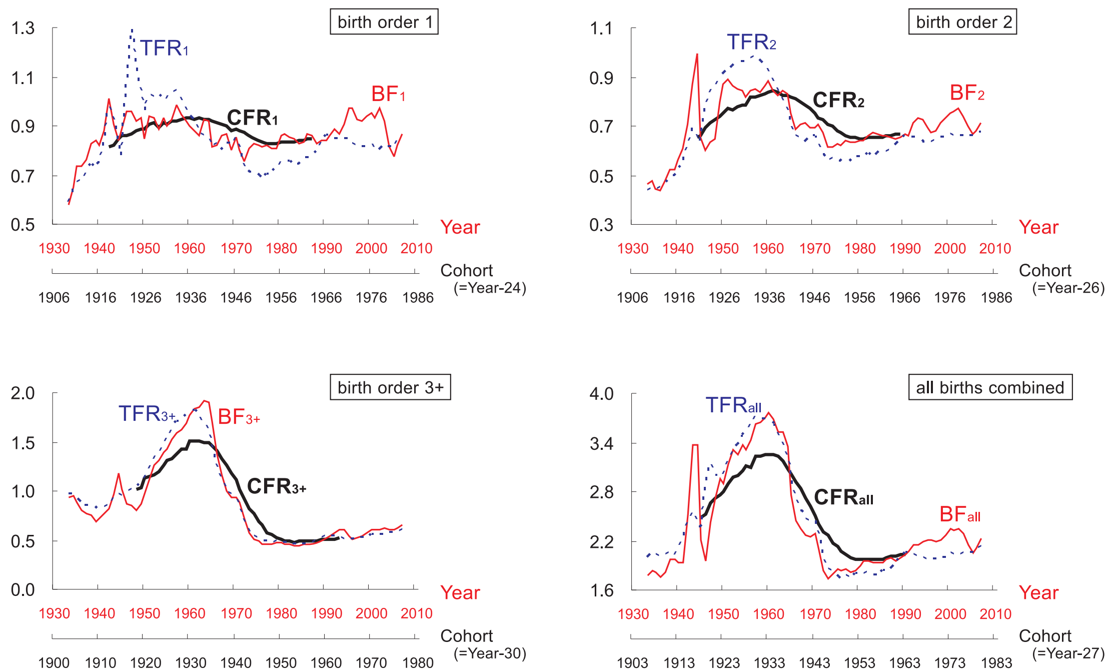

Figure 1: The Conventional Way in Comparing Period and Cohort Fertility Measures: the U.S. Example

major source of prediction errors, as predicting cohort fertility can be degenerated to predicting the average period quantum.

Conventionally, a period fertility measure, such as the total fertility rate (TFR) or any of its adjusted variant, in year t is related to the completed fertility rate (CFR) of the cohort born in year $t - \text{MAC}$, where MAC denotes the mean age at childbearing. Taking the fertility in the U.S. as an example, Figure 1 exhibits the time series of the CFR, the TFR, and the Bongaarts and Feeney (1998) measure (BF hereafter) by birth order, with the MAC being set at 24 (order 1), 26 (order 2), 30 (order 3+), and 27 (all-birth-combined) for convenience. Given this kind of *one-to-one* correspondence, period fertility measures tend to be considered as unreliable CFR predictors because of their fluctuant nature, occasionally absurd resulting values, and the obvious discrepancy from the CFR.

In contrast, this paper highlights a *many-to-one* relationship between period and cohort measures which regards the CFR as a *linear combination* of period rates over the period during which women of this cohort were in their prime childbearing ages, say 15–44. Specifically, one can derive the following identities for the cohort born

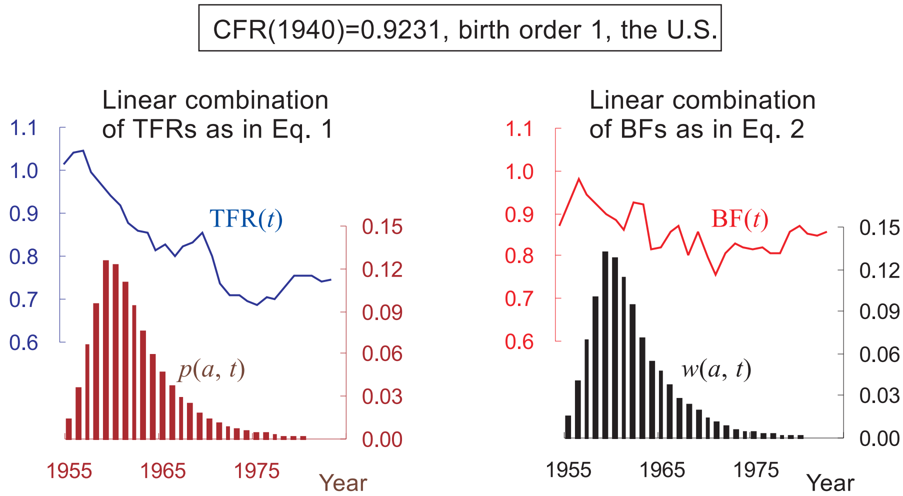

Figure 2: An Illustration of the Many-to-One Relationship in Equations (1) and (2)

in year c *without* assumptions:

$$\begin{aligned} \text{CFR}(c) &= \sum_{t=c+15}^{c+44} p(t-c,t)\,\text{TFR}(t) && (1) \\ &= \sum_{t=c+15}^{c+44} w(t-c,t)\,\text{BF}(t), && (2) \end{aligned}$$

where coefficients $p(a,t)$ and $w(a,t)$ may vary with age a and year t (refer to Appendix A for more mathematical details).[1] Figure 2 illustrates this many-to-one perspective using the first-birth CFR of cohort 1940 (whose prime childbearing years were 1955–84) in the U.S. as an example.

Some remarks regarding Equations (1) and (2) should be noted. First of all, subscripts denoting the birth order are suppressed, owing to the fact that these two identities always hold regardless of birth order and no matter whether the fertility is order or non-order specific. "[T]o get an improved reading of period fertility", Bongaarts and Feeney (1998) emphasized that the BF method should be applied to each birth order separately and then summed over all orders (see their Equation 4); it may otherwise give erroneous quantum adjustment when applied to non-parity-specific data directly. In contrast, our use of the BF method is for a completely

[1]In fact, Bongaarts and Feeney (1998, p.282–284) had proposed to compare the completed fertility of true cohorts with a weighted average of their adjusted total fertility rates, only that the weights described in their Footnote 7 differ from $w(a,t)$. We derived these equations independently, only later finding a similar proof had been provided in Bongaarts and Feeney (2006), but they did not proceed to develop a CFR predictor as we did.

different purpose: to construct a CFR predictor. Therefore, a non-parity-specific application of the BF formula is allowed in this paper and our prediction approach is thus valid even when fertility rates by birth order are not available.

Second, we use *linear combination* rather than *weighted average* in describing the many-to-one relationship. To see the distinction, one may rewrite Equations (1) and (2) into

$$\begin{aligned} \mathrm{CFR}(c) &= \left[\frac{\sum_{t=c+15}^{c+44} p(t-c,t)\,\mathrm{TFR}(t)}{\sum_{t=c+15}^{c+44} p(t-c,t)} \right] \cdot \sum_{t=c+15}^{c+44} p(t-c,t) && \text{(1a)} \\ &= \left[\frac{\sum_{t=c+15}^{c+44} w(t-c,t)\,\mathrm{BF}(t)}{\sum_{t=c+15}^{c+44} w(t-c,t)} \right] \cdot \sum_{t=c+15}^{c+44} w(t-c,t). && \text{(2a)} \end{aligned}$$

Note that the term in square parentheses in either Equations (1a) or (2a) represents a weighted average of TFRs or BFs, but neither $\sum_{t=c+15}^{c+44} p(t-c,t)$ nor $\sum_{t=c+15}^{c+44} w(t-c,t)$ necessarily equals one. However, with the same assumption regarding the shape of the age pattern of period fertility as adopted in Bongaarts and Feeney (1998), Appendix B shows that the CFR becomes equivalent to a weighted average of BFs.

Finally, one can always derive an identity similar to Equations (1) and (2):

$$\mathrm{CFR}(c) = \sum_{t=c+15}^{c+44} \mu(t-c,t)\,X(t), \tag{3}$$

where $X(t)$ denotes any period measure, such as the Kohler and Philipov (2001) method (KP hereafter), and $\mu(a,t) = p(a,t)\,\mathrm{TFR}(t)/X(t)$ the corresponding coefficient. Nevertheless, this paper can only provide a theoretical foundation for the use of Equation (2) in devising a feasible CFR predictor, as will be specified in later sections.

The next section revisits previous critiques that deny the use of period measures in inferring cohort fertility, and then examines for each in turn whether the many-to-one perspective can waive these challenges. In the three sections that follow, we introduce how a CFR predictor can be constructed based on the many-to-one relationship, describe how experiments are designed to evaluate a predictor, demonstrate the prediction performance with empirical evidence from Canada, the U.S., and 23 European countries, and identify the source of prediction errors. The final section summarizes and concludes.

Previous Critiques on Period Measures as a CFR Predictor

Throughout this paper, we focus on discussions regarding period fertility measures that work in terms of age-specific fertility rates (ASFRs), including the TFR (the sum of ASFRs during a particular year), the BF (an adjustment to remove tempo distortions of TFR), and the KP (a further adjustment that takes variance effects into account). Period measures of this type have been extensively used mainly because of the wide availability of ASFR data and their attractively simple formulae (except for the KP).

Nevertheless, whatever original objectives these period measures are constructed for, as a CFR predictor, they have been widely criticized for the following reasons: the exclusion of the past fertility process (Ní Bhrolcháin, 2011), the use of incidence rates for non-repeatable events (Keilman, 2000; van Imhoff, 2001; van Imhoff and Keilman, 2000), the fluctuant nature and the obvious discrepancy from the CFR (Kohler and Ortega, 2002; Ryder, 1990; Schoen, 2004; Smallwood, 2002; Sobotka, 2003; van Imhoff and Keilman, 2000), and the adoption of unrealistic assumptions (van Imhoff and Keilman, 2000; van Imhoff, 2001). These challenges are definitely valid toward the conventional one-to-one correspondence between period and cohort measures, but may not apply when one adopts the many-to-one perspective.

Exclusion of the Past Fertility Process

Information regarding the cumulated fertility of incomplete cohorts at the base period is readily available in general; simply summing up the past ASFRs diagonally across the Lexis surface. However, period measures rigidly apply a transformation of ASFRs only from a particular year (and/or the immediately adjacent years) to generate fertility rates for that year. As a consequence, when a lone period fertility is used to predict the CFR, the available information on the past fertility process of the cohort is just ignored. But when the CFR is regarded as a linear combination of period measures over the whole childbearing years, the entire past of the fertility process has been actually included in the derivation of our predictors.

Use of Incidence Rates for Non-repeatable Events

Some critiques on ASFR-based period measures emphasized that the use of incidence rates for non-repeatable events is erroneous. Specifically, a tempo shift affects not only the numerator but also the denominator of such rates. Any index derived by summing ASFRs over ages in a given year (and any variant of the period-sum method as well) cannot be interpreted as a proper quantum indicator, because extra tempo distortions can be introduced.

To fix this problem, a few studies (e.g. Bongaarts and Sobotka, 2012; Kohler and Ortega, 2002; Yamaguchi and Beppu, 2004) turned to construct adjusted period measures by using age and parity-specific occurrence-exposure rates (hazard rates). Nevertheless, research in this line confronts a tradeoff: Improved prediction results may be obtained,[2] but the range of application is more limited due to the availability of this kind of data.

Note that the use of incidence rates is problematic because ASFR-based period measures are utilized in the conventional one-to-one manner, as if one is comparing a synthetic cohort with a real cohort. In contrast, our use of ASFR-based period measures is built on a definite many-to-one relationship, which aims to work in terms of real cohorts. Such a strategy therefore avoids this problem and reserves the advantage of data availability at the same time.

Fluctuant Nature and Obvious Discrepancy from the CFR

As illustrated graphically in Figure 1, ASFR-based period measures tend to be considered unreliable because of their fluctuant nature, occasionally absurd resulting values, and the obvious discrepancy from the CFR. Although van Imhoff and Keilman (2000, p.552) indicated that most of these annual fluctuations persist even after *moderate* smoothing, Cheng and Lin (2010) showed that *strong* smoothing can remove fluctuations and then provide a good estimate of the CFR.[3]

Since smoothing a series of values is identical to taking a weighted average of

[2]For example, Bongaarts and Sobotka (2012) showed that their new measure (denoted as $TFRp^*$ in their paper), which controls for the parity composition of the female population, can generate less fluctuant estimates and provide a closer approximation to the CFR.

[3]Bongaarts and Sobotka (2012) also proposed that fluctuations can be minimized by smoothing time series of the BF, but they only used a simple 5-year moving average.

them, the many-to-one period-cohort relationship shown in Equations (1) and (2) in effect provides a theoretical justification for the use of strong smoothers whose density function resembles the distribution of coefficients $p(a,t)$ or $w(a,t)$ as depicted in Figure 1. However, no particular smoother is unconditionally recommended, because these coefficients does not sum to one necessarily and the coefficient distribution may vary with cohort and parity.

Adoption of Unrealistic Assumptions

That the rationale behind the BF measure assumes an invariant shape of period fertility schedules has incurred a few discussions. van Imhoff and Keilman (2000) and van Imhoff (2001) showed that this assumption is violated by the data from countries such as Italy, the Netherlands, and Norway, while Zeng and Land (2001) argued in empirical analyses using data from the United States and Taiwan that the adjusted fertility rate is in general insensitive to assumptions.

Indeed, these discussions help the reader to understand related properties and limitations of period measures. But, from a methodological point of view, falsifying the underlying assumtions cannot directly reject the validity of a measure as a predictor. As van Imhoff (2001, p.36) pointed out, "[a]ny procedure trying to estimate cohort quantum from period quantum is based on simplifying assumptions, the justifiability of which can only be verified empirically: by comparing the estimated cohort fertility with actual cohort fertility." This paper thus addresses the importance of fully utilizing the existing fertility data in evaluating the prediction performance of competing methods.

Constucting a CFR Predictor in the Many-to-One Framework

When completed data on childbearing are available, measuring cohort fertility is straightforward — simply summing up the ASFRs along the entire cohort fertility schedule, i.e.

$$\mathrm{CFR}(c) = \sum_{a=15}^{44} f(a, c+a), \tag{4}$$

where $f(a,t)$ denotes the ASFR at age a in year t. In this case, Equations (1) and (2) add no value to the CFR prediction, but only specify the relationship between period

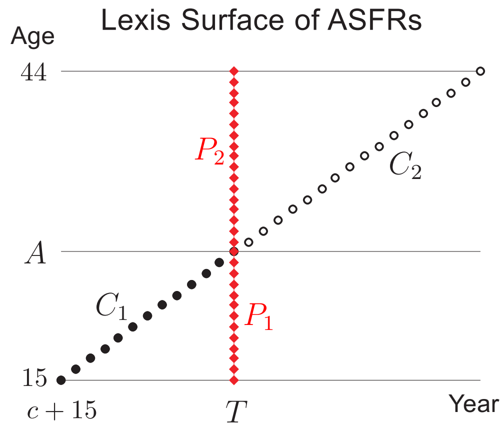

Figure 3: An Illustration of the Construction of a CFR Predictor

and cohort measures and provide an unconventional perspective of interpretation.

In contrast, when childbearing is unfinished and truncated at age A, as depicted in Figure 3, the cohort's lifetime fertility (C) is accordingly divided into an observed past component (C_1) and a remaining future component (C_2):

$$\begin{aligned} C &= \qquad C_1 \qquad + \qquad C_2 \\ &= \sum_{a=15}^{A} f(a, c+a) + \sum_{a=A+1}^{44} f(a, c+a). \end{aligned} \tag{5}$$

Analogously, the period TFR in the truncation year $T = c + A$ can be devided into a cumulative total through age A, $P_1 = \sum_{a=15}^{A} f(a, T)$, and a remainder, $P_2 = \sum_{a=A+1}^{44} f(a, T)$. Next, we introduce a few plain prediction approaches, so that the merits of adopting the many-to-one framework can be demonstrated by contrast.

Some Plain Prediction Methods

The CFR prediction problem is in effect to estimate the future component (C_2), and there are three obvious, simple strategies:

1. Freeze Rates $\widehat{C} = C_1 + P_2$,
2. Equal Ratio $\widehat{C} = C_1 + (C_1/P_1)P_2$,
3. Freeze Adjusted Rates $\widehat{C} = C_1 + [1 - r(T)]^{-1}P_2$,

where $\widehat{C}$ denotes the estimator of the CFR and $r(T)$ represents the change in the mean age of childbearing at time T. The first strategy simply uses the ASFRs in the latest year observed $f(a,T)$ as an imputation of future rates $f(a,c+a)$ for each age. The second strategy assumes that the cumulative fertility proportion through age A for cohort c (i.e. $\frac{C_1}{C_1+C_2}$) is identical to that in year T (i.e. $\frac{P_1}{P_1+P_2}$). The last strategy resembles the first except the BF adjustment fertility rates $[1-r(T)]^{-1}f(a,T)$ are used as an imputation. Note that among these plain strategies, a theoretical rationale in deciding which one is superior to another is unavailable; there is no obvious reason why the multiple of P_2 should be 1, C_1/P_1, or $[1-r(T)]^{-1}$ to accurately estimate C_2.

Prediction in the Many-to-One Framework

Let $\widetilde{\mathrm{BF}}$ represent the weighted average of BFs over the unfinished years. Following the many-to-one logic embedded in Equation (2), C_2 can be expressed as

$$C_2 = \sum_{a=A+1}^{44} w(a,c+a)\,\mathrm{BF}(c+a) = \widetilde{\mathrm{BF}} \sum_{a=A+1}^{44} w(a,c+a). \tag{6}$$

The prediction of the unfinished fertility is then transformed into the product of an (average) anticipation of future movements in the BF and an inference on the sum of unfinished coefficients.

Furthermore, if the shape of period fertility schedules is assumed invariant during the unfinished period (rather than the entire lifetime), one may derive

$$\sum_{a=A+1}^{44} w(a,c+a) = \sum_{a=A+1}^{44} p(a,T), \tag{7}$$

equalizing the sum of unfinished w's to the sum of observed p's in year T (refer to Appendix B for the detailed proof and the supporting empirical evidence). By substituting this conversion into Equation (6), the cohort fertility is approximated to

$$\mathrm{CFR}(c) = C_1 + \widetilde{\mathrm{BF}} \sum_{a=A+1}^{44} p(a,T), \tag{2b}$$

a framework focusing on the prediction of $\widetilde{\mathrm{BF}}$.[4]

[4] As for the sum of unfinished p's or μ's, we have not yet derived an analogous observed counterpart, which confines our discussion to the use of the BF, rather than other period measures, in predicting the CFR.

Both $\overline{\text{BF}}$ (the weighted average of BFs observed) and $\text{BF}(T)$ (the latest BF observed) are straightforward $\widetilde{\text{BF}}$ estimators, each corresponding to a CFR prediction strategy:

4. Freeze BF-A $\quad \widehat{C} = C_1 + \overline{\text{BF}} \sum_{a=A+1}^{44} p(a,T),$[5]

5. Freeze BF-L $\quad \widehat{C} = C_1 + \text{BF}(T) \sum_{a=A+1}^{44} p(a,T).$

Moreover, given that $P_2 = \text{TFR}(T)\sum_{a=A+1}^{44} p(a,T)$, the three aforementioned plain strategies can be rewritten

1. Freeze Rates $\quad \widehat{C} = C_1 + \text{TFR}(T) \sum_{a=A+1}^{44} p(a,T),$

2. Equal Ratio $\quad \widehat{C} = C_1 + (\frac{P_1}{P_1+P_2})^{-1} C_1 \sum_{a=A+1}^{44} p(a,T),$

3. Freeze Adjusted Rates $\quad \widehat{C} = C_1 + \text{BF}(T) \sum_{a=A+1}^{44} p(a,T).$

Not only is the Freeze Adjusted Rates method shown equivalent to the Freeze BF-L approach, Equation (2b) also serves as a platform to compare these seemingly unrelated methods with a definite criterion: As a $\widetilde{\text{BF}}$ estimator, which of $\text{TFR}(T)$, $(\frac{P_1}{P_1+P_2})^{-1}C_1$, $\text{BF}(T)$, and $\overline{\text{BF}}$ performs the best? Before proceeding to the empirical evidence, we provide some theoretical discussions as follows.

Some Notable Remarks

First, the TFR differs from the BF due to the tempo effect which exists in general, especially for lower birth orders. Therefore, using $\text{TFR}(T)$ to estimate $\widetilde{\text{BF}}$ (as in the Freeze Rates method) seems more indirect and thus less likely to produce an accurate prediction than using $\text{BF}(T)$ (as in the Freeze BF-L method).

Second, the Equal Ratio method and the Freeze BF-A strategy are in fact more closely related than they appear on the surface. As $(\frac{P_1}{P_1+P_2})^{-1}C_1$ in the formula of the Equal Ratio method can be interpreted as inflating the observed cohort fertility C_1 by a period cumulative proportion $P_1/(P_1+P_2) = \sum_{a=15}^{A} p(a,T)$, $\overline{\text{BF}}$ in the

[5] $\overline{\text{BF}} = \sum_{a=15}^{A} w(a,c+a)\,\text{BF}(c+a)/\sum_{a=15}^{A} w(a,c+a) = [\sum_{a=15}^{A} w(a,c+a)]^{-1}C_1.$

formula of the Freeze BF-A strategy inflates C_1 by $\sum_{a=15}^{A} w(a, c+a)$ (see footnote 5). These two inflation proportions coincide if the shape of period fertility schedules is invariant during the observed period (see the proposition in Appendix B).

Last, the decision of whether the Freeze BF-L or the Freeze BF-A is more appropriate for use depends on how the unfinished BF is anticipated to change. When the BF series is anticipated to fluctuate around a certain level without trend, such that only mild difference is expected between the averages of BFs before and after the truncation year T, $\overline{\mathrm{BF}}$ could be a better estimator of $\widetilde{\mathrm{BF}}$. In contrast, when one anticipates that the unfinished BFs deviate from the observed ones in a significant trend, $\mathrm{BF}(T)$ may provide a more accurate prediction of $\widetilde{\mathrm{BF}}$ than $\overline{\mathrm{BF}}$ does.

Data, Experiment Design, and Prediction Performance

Data

The data employed in this study are age-specific fertility rates by one-year period and by single-year age group, taken from the Human Fertility Database[6] and the Eurostat Database, downloaded in January, 2012. We restrict our analysis to the set of single-year age groups 15-44 and exclude countries whose data contain less than 10 completed cohorts, so not all countries listed in the databases are included. In addition, we use the separate schedules for East and West Germany, as well as for England/Wales and Scotland in the U.K., rather than their combined data. Thus there are 27 schedules in total, as listed in Table 1. Since non-parity-specific data are more readily available than parity-specific data, there are 907 completed cohorts from the former and 326 from the latter, respectively.

Experiment Design

For completed cohorts, ASFR data that span the entire childbearing years (and thus the actual CFR) are available. So, one can choose any integer between 16 and 43 to be the truncation age A as if the fertility were unfinished, and then use

[6]Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available at `http://www.humanfertility.org`.

Table 1: Data from the Human Fertility Database and the Eurostat

	all-birth-combined		parity-specific	
country	periods	completed cohorts	periods	completed cohorts
from the Human Fertility Database				
Austria	1951–2010	1936–1966 (31)	1984–2010	NA
Bulgaria	1947–2009	1932–1965 (34)	1947–2009	1932–1965 (34)
Canada	1921–2007	1906–1963 (58)	1944–2007	1929–1963 (35)
Czech Republic	1950–2009	1935–1965 (31)	1950–2009	1935–1965 (31)
Estonia	1959–2009	1944–1965 (22)	1959–2009	1944–1965 (22)
Finland	1939–2009	1924–1965 (42)	1982–2009	NA
France	1946–2009	1931–1965 (35)	NA	NA
Germany				
East	1956–2010	1941–1966 (26)	1956–1989	NI
West	1956–2010	1941–1966 (26)	NA	NA
Hungary	1950–2009	1935–1965 (31)	1952–2009	1937–1965 (29)
Lithuania	1959–2009	1944–1965 (22)	1970–2009	1955–1965 (11)
Netherlands	1950–2009	1935–1965 (31)	1950–2009	1935–1965 (31)
Portugal	1940–2009	1925–1965 (41)	1959–2009	1944–1965 (22)
Russia	1959–2009	1944–1965 (22)	1959–2009	1944–1965 (22)
Slovakia	1950–2009	1935–1965 (31)	1950–2009	1935–1965 (31)
Sweden	1891–2010	1876–1966 (91)	1970–2010	1955–1966 (12)
Switzerland	1932–2009	1917–1965 (49)	1998–2009	NA
U.K.				
England/Wales	1938–2009	1923–1965 (43)	NA	NA
Scotland	1945–2009	1930–1965 (36)	NA	NA
U.S.	1933–2007	1918–1963 (46)	1933–2007	1918–1963 (46)
from the Eurostat				
Belgium*	1954–2009	1939–1965 (27)	NA	NA
Denmark	1950–2010	1935–1966 (32)	NA	NA
Greece	1961–2010	1946–1966 (21)	NA	NA
Iceland	1963–2010	1948–1966 (19)	NA	NA
Italy	1952–2008	1937–1964 (28)	NA	NA
Norway	1961–2010	1946–1966 (21)	NA	NA
Spain	1971–2010	1956–1966 (11)	NA	NA

Note:
1. When a country is included in both databases, we prioritize data from the Human Fertility Database.
2. Countries whose data are insufficient to construct at least 10 completed cohorts (covering age groups 15–44) will be excluded from the analysis.
3. In parentheses are numbers of completed cohorts.
4. NA denotes that data are not available, while NI denotes that data are not included in the analysis because the number of completed cohorts is less than 10.

∗ Data in 2001 and 2002 are missing for Belgium. We use spline smoothing to interpolate these values by age.

the "observed" ASFRs to produce an estimated CFR; we denote this process as one experiment. There can be at most 28 experiments (each corresponding to a particular truncation age) for every completed cohort, and hence 25,396 and 9,128 experiments in total from non-parity- and parity-specific data, respectively.

To evaluate the prediction performance of a CFR predictor across experiments with various truncation ages, we employ the following prediction error index which measures how much proportion of the unfinished fertility has not been correctly estimated:[7]

$$\text{Prediction Error} = \frac{\widehat{C} - C}{C - C_1} \times 100\% = \frac{\widehat{C}_2 - C_2}{C_2} \times 100\%.$$

For each birth order, we categorize experiments by completed proportion (defined as C_1/C),[8] construct a distribution of prediction error (or of error in absolute value) for each category, and then compare distributional statistics among approaches.

Note that conventional approaches other than the Freeze Rates and the Equal Ratio are not included in the comparison owing to the following considerations:[9] (1) This paper is focused on explaining how the many-to-one version of CFR as a weighted sum of period measures can help to predict cohort fertility. Experimental results are provided, as supporting evidence of the above theoretical remarks, to exhibit the strength and weakness of predictors of this unconventional type (i.e. the Freeze BF-A and the Freeze BF-L). (2) In reality, making a global comparison of prediction performance across approaches ever developed is unnecessary and impossible. But since the Freeze Rates method is quite easy to implement and its prediction performance shown in this paper can be replicated and verified without difficulty, interested researchers may regard this method as a common reference in

[7]For example, suppose that $C = 2.0$ and $\widehat{C} = 1.8$, the error is -16.67% when $C_1 = 0.8$, indicating that 83.33% of C_2 has been estimated. In contrast, if C and $\widehat{C}$ remain unchanged while C_1 increases to 1.2, then the error becomes -25.00%, indicating that only 75.00% of C_2 has been estimated.

[8]Since the completed proportion by a particular truncation age may differ across countries, cohorts, and birth orders, categorizing experiments by completed proportion is more appropriate than by truncation age.

[9]These approaches include the Evans method (Evans, 1986), the Li-Wu model (Li and Wu, 2003), the Willekens-Baydar approach (Willekens and Baydar, 1984), and various curve fitting models such as a modified version of the Coale and McNeil's (1972) double exponential model (e.g., Bloom, 1982; Chen and Morgan, 1991), the Hadwiger function (e.g., Chandola et al., 1999), and the linearized Gompertz model (Myrskylä and Goldstein, 2013).

comparison.[10]

In addition to comparing among prediction methods, we adopt another predetermined criterion to judge how well a particular method performs: a CFR estimate whose prediction error exceeds 20% is then considered *poor*. This 20% value is, of course, somewhat arbitrary but comparable with Sobotka (2003, Table 12)'s classification.[11]

Prediction Performance

Without considering the sign of prediction errors which indicates whether the CFR is over- or under-estimated, Table 2 summarizes the mean *absolute* prediction error of four competing methods across various categories of completed proportion and birth orders. Also included in the table are the sample size, the average truncation age, and the mean absolute conversion error.[12] In the following analysis, we focus on the parity-specific results to identify the crucial factor(s) affecting the prediction performance. The all-birth-combined results are listed merely for reference; they are aggregately influenced by factors that exert impacts on the fertility prediction of individual birth order. Before proceeding to the detailed discussion, we provide two pieces of background information for reference.

First of all, the conversion error (in the fourth column of Table 2) is found limited to a level far below the 20% standard and accounts for less than one half of the prediction error by all prediction methods in the table (except for the first birth). Besides, as the birth order gets higher, the conversion error descends slightly (except in the first category of completed proportion) while the prediction error rises significantly (gently from the first to the second birth and then dramatically from the second to the third-and-above). The contrast in magnitude and trend

[10] The Freeze Rates method has been shown to outperform several conventional approaches in predicting the CFR, although the evidence is limited to the first birth in the U.S.; see Cheng and Lin (2010, Table 1).

[11] Sobotka's error index (the percentage error) differs from ours by taking the true CFR, rather than the unfinished CFR, as the denominator. One can assume the completed proportion at mean age of birth to be, for example, 60% to compare these two indices.

[12] The conversion error is the prediction error made by a control treatment of Equation (2b) which adopts the true $\widetilde{\mathrm{BF}}$ and measures the isolated effect of using $\sum_{a=A+1}^{44} p(a,T)$ in replacement of $\sum_{a=A+1}^{44} w(a,c+a)$.

Table 2: Mean Absolute Prediction Error by Method, Completed Proportion, and Birth Order

Birth Order	Samples	Average Trunc. Age	Conversion Error†	Freeze Rates	Equal Ratio	Freeze BF-A	Freeze BF-L
			completed proportion $\in [10\%, 30\%)$				
1	677	19.10	2.34	9.47	7.40	**5.05**	6.39
2	833	21.79	2.25	10.92	8.60	**7.43**	9.07
3+	998	24.33	2.32	31.29	35.82	34.23	**29.64**
all	2,759	21.16	2.48	17.03	14.59	**12.74**	13.02
			completed proportion $\in [30\%, 50\%)$				
1	597	21.12	3.36	10.76	7.29	**5.28**	6.90
2	687	24.13	2.97	12.52	9.27	**8.02**	10.10
3+	843	27.11	2.61	26.17	37.49	32.56	**25.03**
all	2,440	24.04	2.98	16.96	14.94	13.38	**13.22**
			completed proportion $\in [50\%, 65\%)$				
1	499	22.83	4.42	11.87	7.33	**6.10**	7.84
2	576	26.07	3.70	13.77	10.47	**9.30**	11.05
3+	673	29.42	3.15	23.42	35.65	34.12	**22.79**
all	1,999	26.48	3.65	17.15	15.99	14.69	**13.53**
			completed proportion $\in [65\%, 75\%)$				
1	415	24.18	5.10	12.23	7.45	**6.67**	8.35
2	451	27.64	4.35	14.35	11.55	**10.61**	10.94
3+	529	31.23	3.63	21.19	37.30	35.69	**21.01**
all	1,607	28.44	4.41	17.22	16.90	15.85	**13.08**
			completed proportion $\in [75\%, 85\%)$				
1	620	25.74	5.58	12.39	7.97	**7.49**	8.84
2	611	29.21	5.26	14.73	12.63	11.93	**11.45**
3+	679	33.05	4.22	18.66	38.62	37.32	**18.54**
all	2,119	30.40	5.42	17.63	18.86	17.98	**13.49**

Note: Numbers in the last four columns are mean absolute prediction errors (in percentage), and the best performance (i.e. the least number) for each combination of birth order and completed proportion category is noted in bold.

† Coversion error represents the prediction error made by adopting the conversion as in Eq. 7 while using the true $\widetilde{\mathrm{BF}}$.

between the two types of error reveals the importance of accurately predicting $\widetilde{\text{BF}}$ and indicates that its difficulty correlates with the birth order.

Second, the reader may want to know the prediction performance of period fertility measures when they are used as a CFR predictor in the conventional one-to-one manner. In this case, there is one and only one experiment for each completed cohort since the truncation age has been prescribed to the mean age of childbearing at which the completed proportion falls between 50% (the median age) and 75% (the third quartile age) in general. Across the 326 parity-specific and the 907 non-parity-specific completed cohorts in our dataset, the TFR (the BF) yields a mean absolute prediction error of 23.63% (14.08%), 28.98% (17.51%), 34.93% (34.48%), and 29.32% (19.58%) for the first, second, third-and-above, and all-combined births, respectively. It concludes that the BF significantly outperforms the TFR for lower birth orders while no statistically reliable difference is found for higher ones.

Possibly the most striking result observed in Table 2 is that: In 677 trials to predict the first-birth CFR with the completed proportion being within $[10\%, 30\%)$, the Freeze BF-A method yields an average of absolute prediction errors as low as 5.05%. In other words, nearly 95% of the unfinished fertility has been correctly estimated. This performance is notable not only because the mean absolute error is smaller than its counterparts by the other methods and far below the 20% standard, but because it is obtained with relatively little information that truncates at about age 19 on average.

Throughout the five categories of completed proportion between 10% and 85%, the empirical results support the remarks proposed in the previous section that (1) the Freeze BF-L dominates the Freeze Rates, especially for low birth orders; (2) the Equal Ratio and the Freeze BF-A yield similar performance; and (3) which of the Freeze BF-A and the Freeze BF-L is more accurate varies with the birth order that is associated with how the unfinished BF changes. With few exceptions, the Freeze BF-A method retains its relative (as compared with the other methods) and absolute (as compared with the 20% standard) advantages in predicting the CFR of the first and the second births. In contrast, the Freeze BF-L method takes the lead in predicting the CFR of the third-and-above birth, but the mean absolute prediction error falls either beyond or just slightly below the 20% standard, signifying that no method investigated in this paper can provide satisfactory predictions for high birth orders.

By depicting distribution functions (i.e. cumulated density functions) in absolute prediction error, Figure 4 further provides a more comprehensive comparison of the four methods across categories of completed proportion (along each row) and across birth orders (down each column). In the top panels where fertility is of first birth, the distribution curve of the Freeze BF-A method lies entirely above the other three curves, exhibiting its statistical dominance — more than 99 percent of trials yield an absolute prediction error below the 20% absolute standard (depicted by a virtical broken line) when the completed proportion is between 10% and 75% (and 97 percent between 75% and 85%).

Essentially similar but slightly less accurate results are found when fertility is of second birth: the Freeze BF-A curve still lies (almost entirely) above the other three, while the percentage of trials whose absolute prediction error is below the 20% standard now drops to 90 percent when the completed proportion is between 10% and 75% (and to 83 percent between 75% and 85%). All curves lean further to the right compared with their counterparts in the first birth case.

When fertility is of third-and-above birth, however, no distribution curve lies entirely above all the others. More than 30 percent of trials yield an absolute prediction error exceeding the 20% standard (beyond which a prediction is categorized as poor) regardless of method or completed proportion category. Furthermore, all methods are accompanied by a certain proportion of trials whose absolute prediction error is larger than 70%, for no curve reaches 1.0 (the vertical maximum) within the (horizontal) range of absolute prediction error presented in the figure.

Identifying the Source of Prediction Error

With the conversion of coefficient sum as specified in Equation (7), Equation (2b) provides a framework that degenerates the prediction of CFR to the prediction of the average BF in the unfinished part (i.e. $\widetilde{\mathrm{BF}}$). Empirically, the conversion error has been shown limited, and the prediction performance of a particular method is thus mainly decided by how well $\widetilde{\mathrm{BF}}$ can be estimated.

Table 2 and Figure 4 exhibit a uniform pattern across birth orders: As the birth order gets higher, the prediction performance becomes poorer. This is not

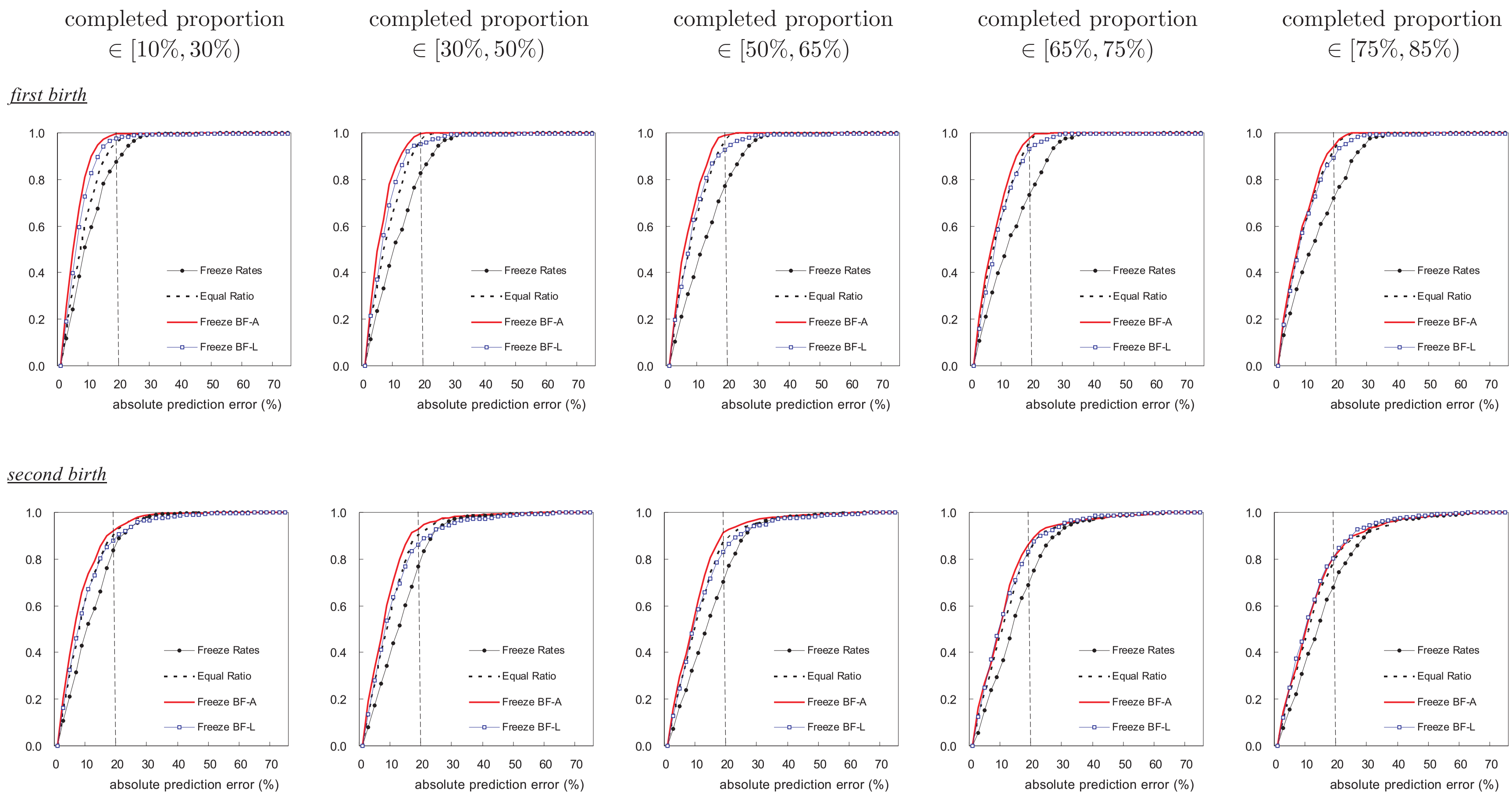

Figure 4: Cumulated Density Functions of Absolute Prediction Error by Method

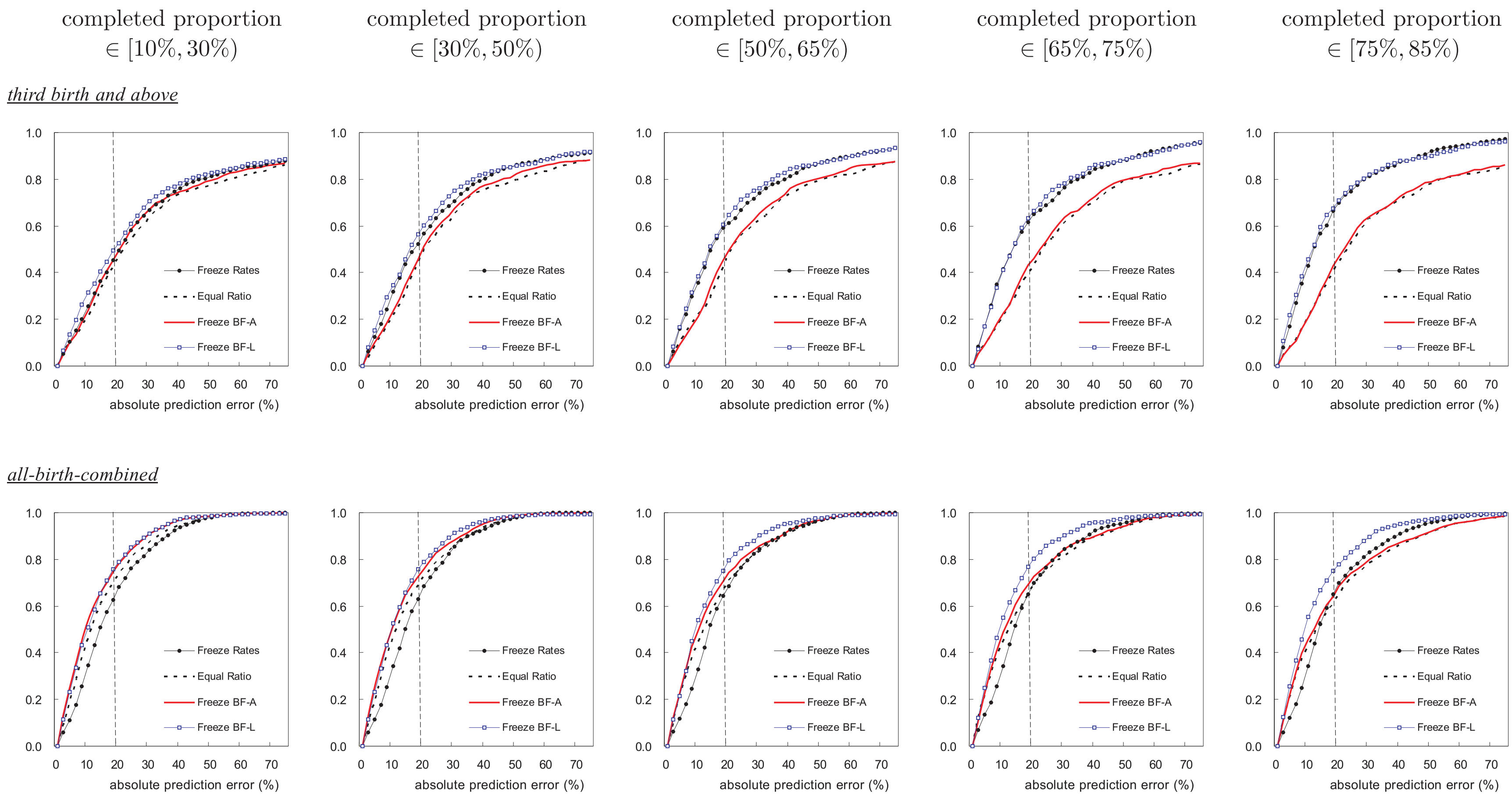

Figure 4: continued

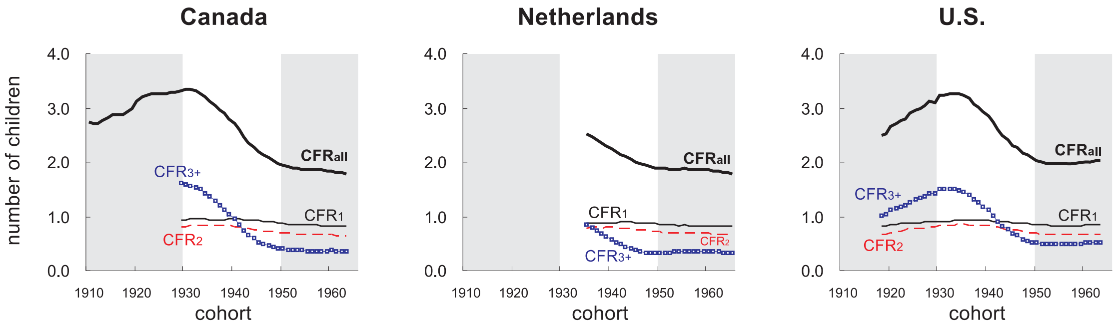

Figure 5: CFR Curves by Birth Order for Canada, the Netherlands, and the U.S.

unexpected since the BF of a higher birth order tends to vary within a greater range (refer to the U.S. example in Figure 1) such that the prediction accuracy of $\widetilde{\mathrm{BF}}$ (and thus of CFR) drops. In this section, we implement a further analysis across cohort subgroups within each birth order to identify the source of prediction error.

As specified in Equation (2), the CFR curve is identical to the *smoothed* version of the BF curve which uses the w distribution as its smoothing density function. In other words, the *original* BF curve fluctuates around the CFR curve (but not in the sense of the conventional one-to-one correspondence as depicted in Figure 1, for the w distribution is not fixed across cohorts), and a cohort locating in an upward (downward) interval of the CFR curve tends to have a lower (higher) observed $\overline{\mathrm{BF}}$ than the unfinished $\widetilde{\mathrm{BF}}$. Since the Freeze BF-A method employs $\overline{\mathrm{BF}}$ to estimate $\widetilde{\mathrm{BF}}$, one may therefore expect that the sign and the magnitude of its prediction error correlate with the direction and the slope of the CFR curve, respectively.

In the following analysis, we select Canada, the Netherlands, and the U.S. as example countries because of their wide ranges of parity-specific fertility data. Figure 5 depicts their CFR curves by birth order, showing that the extent of variation increases as the birth order gets higher. Cohorts are divided into three subgroups according to birth year: 1910–30, 1935–50, and 1950–65, corresponding to an upward, downward, and relatively flat interval of the CFR curve.[13] Females in these three subgroups experienced their main childbearing ages during the baby-boom, the baby-bust, and the post baby-bust periods, respectively. It is worth a mention that the U.S. cohorts 1935–50 experienced a much stronger quantum change (in

[13]Due to data availability, the first subgroup contains parity-specific cohort information from the U.S. only.

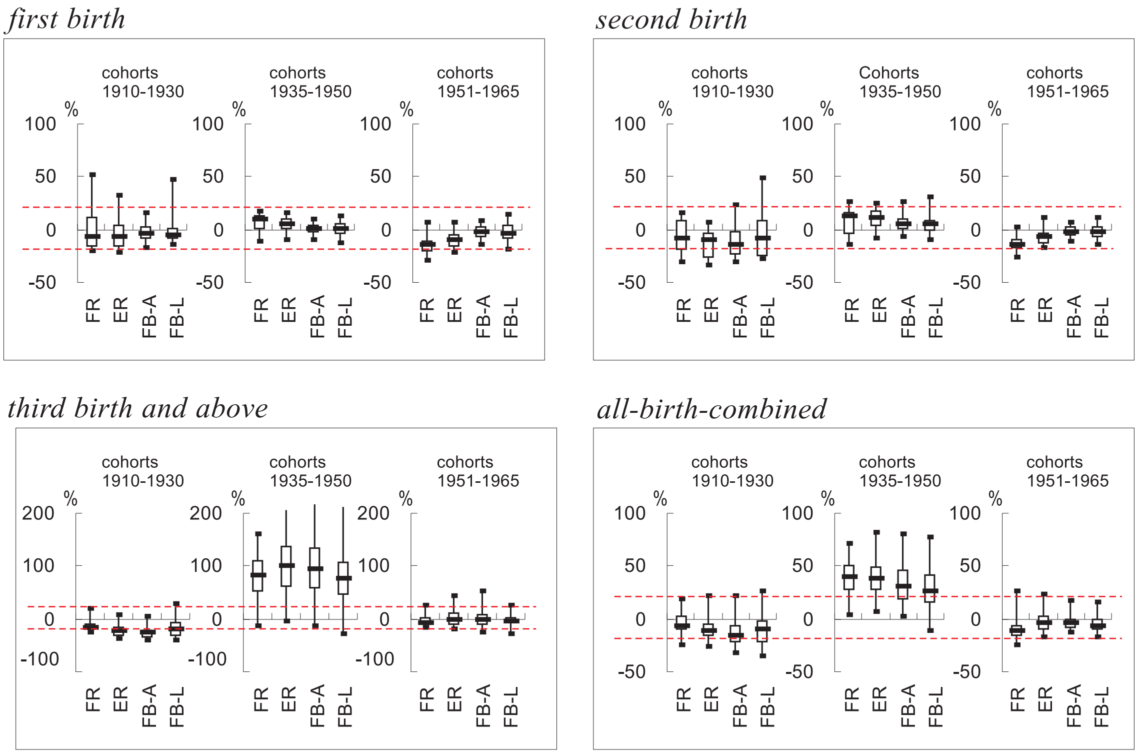

Figure 6: Prediction Error Distributions by Method and Cohort Subgroup

Note: "FR", "ER", "FB-A", and "FB-L" stand for the Freeze Rates, the Equal Ratio, the Freeze BF-A, and the Freeze BF-L methods, respectively. Each box-and-whisker plot provides five summary statistics of an error distribution: the bottom and top of the box are the lower and upper quartile, the band near the middle of the box is the median, and the ends of the whisker represent the minimum and maximum. In addition, broken lines in each panel represent the +20% and −20% absolute criteria.

absolute terms) of CFR_{3+} and CFR_{all} than did cohorts 1910–30,[14] and a similarly situation of CFR_{all} occurred in Canada.[15]

Using box-and-whisker plots to mark five distributional statistics,[16] Figure 6 displays prediction error distributions of four competing approaches for three cohort subgroups in example countries with the completed proportion being within $[10\%, 30\%)$.[17] Also depicted in each panel are two broken lines marking the +20%

[14]The linear regression slopes in CFR of birth order 1, 2, 3+, and all are 0.0076, 0.0122, 0.0394, and 0.0591, respectively, for cohort subgroup 1910–30, while they are −0.0056, −0.0121, −0.0690, and −0.0867, respectively, for cohort subgroup 1935–50.

[15]The linear regression slopes are 0.0365 and −0.0847 for cohort subgroups 1910–30 and 1935–50, respectively.

[16]The ends of the whisker represent the minimum and maximum, the bottom and top of the box are the lower and upper quartile, and the band near the middle of the box is the median.

[17]Figures regarding other completed proportion ranges are omitted due to the similarity in

and −20% thresholds beyond which a prediction is considered poor. Focusing on the prediction performance of the Freeze BF-A method, one may observe that (1) its CFR estimate tends to be biased downward (upward) for cohorts 1910–30 (cohorts 1935–50), but appears unbiased for cohorts 1951–65; (2) the degree of bias increases as the birth order gets higher only for the first two cohort subgroups; and (3) the contrast in the degree of bias between cohort subgroups 1910–30 and 1935–50 becomes especially significant for CFR_{3+} and $\mathrm{CFR}_{\mathrm{all}}$. These observations are consistent with our previous discussion and therefore support that the source of prediction error comes from the failure to estimate $\widetilde{\mathrm{BF}}$ accurately.

When the quantum change of BF is mild, such as in the cases of the first and the second births across all cohort subgroups or in the cases of cohorts 1951–65 across all birth orders, most trials by each method yield a prediction error falling within the ±20% thresholds. In contrast, when a strong quantum effect is present, such as in the cases of the third-and-above and the all-combined births for cohorts 1935–50, all the methods investigated in this paper fail to provide satisfactory predictions. As a result, the central issue in the prediction of cohort fertility is to detect a strong quantum effect and then devise an appropriate procedure to deal with it, which we leave to future research.

Conclusion

In this paper, we specify a simple but definite relationship between period and cohort fertility measures. Mathematically, the completed fertility of a cohort is identical to a linear combination of period fertility rates over the cohort's childbearing years. Graphically, the CFR curve is actually a smoothed version of the BF curve.

Following this many-to-one perspective, we derive a prediction framework which serves as a platform to compare seemingly unrelated methods and provides a clear theoretical explanation on the major source of prediction errors, as predicting cohort fertility can be degenerated to predicting the average period quantum. Our innovative CFR predictor, either the Freeze BF-A or the Freeze BF-L, is easy to implement, works in terms of age-specific fertility rates, and waives most (if not all) previous critiques that deny the use of period measures in inferring cohort fertility.

pattern, but available upon request from the authors.

We fully utilize the existing data from Canada, the U.S., and 23 European countries in designing a large number of experiments so as to evaluate the prediction performance of four competing approaches. Empirical evidence across birth orders suggests that our Freeze BF-A predictor provides satisfactory estimates and outperforms the other three methods for lower birth orders where the quantum change is mild, particularly when the observed cohort experience is truncated at a quite young age. In the presence of a strong quantum change, however, none of the prediction methods investigated in this paper demonstrate adequate performance. An analysis across cohort subgroups within each birth order also identifies that the source of prediction error comes from the failure to estimate $\widetilde{\mathrm{BF}}$ accurately.

In sum, this paper illuminates the importance of accurately anticipating the average period measure so as to generate a satisfactory estimate of cohort fertility, and acknowledges that the quantum effect severely confines the range of application of our Freeze BF-A method. In a passive sense, one can decide whether to employ our approach upon one's *subjective* judgment regarding the future trend of quantum. Or, one can actively search for some *objective* clues that disclose the future quantum trend, and then revise the prediction accordingly. We leave the study of which information is useful, how to derive this information, and thereby how to adjust the related estimate to future research.

Appendix A: Regarding the Completed Fertility Rate as a Linear Combination of Period Fertility Measures

Distinguished from the demographic translation literature (Keilman, 1994, 2000; Ryder, 1964), this paper derives a simple but definite relationship between period and cohort measures as follows. Let $f(a,t)$ be an age- and time-specific fertility rate, with a and t representing age and time, respectively. Suppose that $f(a,t) = 0$ when a falls beyond the prime childbearing age range $\mathcal{R} \equiv [15, 44]$, then summing the rates over childbearing ages for cohort c and for calendar year t yields the cohort quantum $\mathrm{CFR}(c) = \sum_{a \in \mathcal{R}} f(a, c+a)$ and the period quantum $\mathrm{TFR}(t) = \sum_{a \in \mathcal{R}} f(a,t)$, respectively. We further define the period age-specific fertility proportion as $p(a,t) = f(a,t)/\mathrm{TFR}(t)$ so that $\sum_{a \in \mathcal{R}} p(a,t) = 1$ for any t.

Replacing $f(a, c+a)$ with $p(a, c+a)\,\mathrm{TFR}(c+a)$ leads to Equation (1) in the text:

$$\mathrm{CFR}(c) = \sum_{a=15}^{44} p(a, c+a)\,\mathrm{TFR}(c+a) = \sum_{t=c+15}^{c+44} p(t-c, t)\,\mathrm{TFR}(t),$$

showing that the CFR can be expressed as a linear combination of TFRs with $p(t-c,t)$ as the nonnegative coefficient for TFR at time t.

One may divide $p(t-c,t)$ and multiply $\mathrm{TFR}(t)$ by the same factor $1 - r(t)$ to derive Equation (2) in the text:

$$\mathrm{CFR}(c) = \sum_{t=c+15}^{c+44} w(t-c, t)\,\mathrm{BF}(t),$$

where $r(t)$ represents the change in the mean age of childbearing at time t, $\mathrm{BF}(t) = \mathrm{TFR}(t)/\big[1 - r(t)\big]$ denotes the Bongaarts-Feeney adjusted total fertility rate, and $w(t-c,t) = \big[1 - r(t)\big]\,p(t-c,t)$ the corresponding coefficient.

In the same way, the CFR can also be linked to any period measure $X(t)$ by

$$\mathrm{CFR}(c) = \sum_{t=c+15}^{c+44} \mu(t-c, t)\,X(t),$$

where $\mu(t-c,t) = p(t-c,t)\,\mathrm{TFR}(t)/X(t)$. Note that no assumption is required for all the inferences above.

Appendix B: Converting the Sum of Cohort w's to the Sum of Period p's

With a particular truncation age A between 15 and 44, one may partition the entire childbearing age range $\mathcal{R} \equiv [15, 44]$ into two parts — $\mathcal{R}_1 \equiv [15, A]$ and $\mathcal{R}_2 \equiv (A, 44]$ — and then define two time intervals accordingly — $\mathcal{I}_1 \equiv [c + 15, T]$ and $\mathcal{I}_2 \equiv (T, c + 44]$, where $T = c + A$ denotes the truncation year. This appendix provides the proof of the following proposition:

Proposition B1. *Assume that the shape of period fertility schedules is invariant within $\mathcal{I}_1$, we have*

$$\sum_{a \in \mathcal{R}_1} w(a, c + a) = \sum_{a \in \mathcal{R}_1} p(a, T). \tag{B1}$$

Similarly, if the shape is invariant within $\mathcal{I}_2$, then

$$\sum_{a \in \mathcal{R}_2} w(a, c + a) = \sum_{a \in \mathcal{R}_2} p(a, T). \tag{B2}$$

For convenience, the proof is done by using integral calculus, and we only prove the second part; the first one can be easily done by symmetry.

Proof. Let $\widetilde{A} > A$ be a finite number such that $p(a, t) = 0$ for any $a \geq \widetilde{A}$. By the assumption of a constant period pattern in $p(a, t)$ within the time interval between T and $c + \widetilde{A}$, one may have

$$p(a, t) = p(a - R(t), T) \quad \text{with} \quad R(t) = \int_T^t r(k)\, dk, \tag{B3}$$

where $r(k)$ denotes the instantaneous shift at time k. Equation (B3) states that $p(a, t)$ at time $t \in (T, c + \widetilde{A}]$ has the same shape as $p(a, T)$, but has shifted along the age axis by $R(t)$ years. Let

$$u = a - R(c + a) = a - \int_T^{c+a} r(k)\, dk \tag{B4}$$

so that $p(a, c + a) = p(u, T)$.

Differentiating both sides of Equation (B4) gives $du = \big[1 - r(c + a)\big]\, da$. Thus,

$$\begin{aligned} \int_A^{\widetilde{A}} w(a, c + a)\, da &= \int_A^{\widetilde{A}} \big[1 - r(c + a)\big]\, p(a, c + a)\, da \\ &= \int_A^{\widetilde{A} - R(c + \widetilde{A})} p(u, T)\, du. \end{aligned}$$

Table B1: The OLS Rgression Results of Equation (B6)

birth order	1	2	3+	all
$\widehat{\beta}_0$	-0.00017	-0.00104 ***	-0.00331 ***	-0.00383 ***
$\widehat{\beta}_1$	1.00879 ***	1.00474 ***	1.00231 ***	1.00435 ***
R^2	0.99835	0.99848	0.99754	0.99761
N	9,128	9,128	9,128	25,396

Note: *** indicates that the null hypothesis ($\beta_0 = 0$ or $\beta_1 = 1$) is rejected at one percent siginificance level.

Since $f(a,t)$ is supposed to be zero when a falls beyond the range $\mathcal{R}$ (refer to Appendix A), the above equation can be rewritten

$$\int_{\mathcal{R}_2} w(a, c+a)\, da = \int_{\mathcal{R}_2} p(u, T)\, du \tag{B5}$$

by choosing an appropriate value of $\widetilde{A}$ such that both $\widetilde{A}$ and $\widetilde{A} - R(c + \widetilde{A})$ are no less than 44. The discrete version of Equation (B5) concludes the proof. □

In the special case that A is set at 15 or 44 (i.e., when the constant shape is assumed within $\mathcal{I}_1 \cup \mathcal{I}_2$),

$$\sum_{a \in \mathcal{R}} w(a, c+a) = \sum_{a \in \mathcal{R}} p(a, T) = 1$$

such that the CFR can be regarded as a weighted average of BFs (refer to Equation (2a) in the text).

To investigate whether Equation (B2), or Equation (7) in the text, is supported by empirical data, one may regress its left-hand side (the sum of unfinished w's) on its right-hand counterpart (the sum of observed p's) using a linear model

$$\sum_{a \in \mathcal{R}_2} w(a, c+a) = \beta_0 + \beta_1 \sum_{a \in \mathcal{R}_2} p(a, T) + \varepsilon \tag{B6}$$

for all completed cohorts in our dataset with the truncation age A varying from 16 to 43, and then test null hypotheses such as $\beta_0 = 0$ and $\beta_1 = 1$. Table B1 presents the OLS regression results by birth order. On the one hand, all $\widehat{\beta}_0$ look quite close to 0, all $\widehat{\beta}_1$ appear very close to 1, and all R-square values are extremely high as well, signifying the high correlation between the two sums. On the other hand, however, both hypotheses are statistically rejected at one percent siginificance level for each birth, except that $\beta_0 = 0$ is accepted for the first birth. Given these mixed results, we further examine the isolated impact of adopting this conversion equation on prediction performance and find that the effect is quite limited (see Table 2 and the related discussion of the conversion error for details).

References

Bloom, David 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, John and Griffith Feeney (1998), "On the Quantum and Tempo of Fertility", *Population and Development Review*, 24(2), 271–291.

——— (2006), "The Quantum and Tempo of Life-Cycle Events", *Vienna Yearbook of Population Research 2006*, 115–151.

Bongaarts, John and Tomáš Sobotka (2012), "Demographic Explanation for the Recent Rise in European Fertility", *Population and Development Review*, forthcoming.

Chandola, T., D. A. Coleman, and R. W. Hiorns (1999), "Recent European Fertility Patterns: Fitting Curves to 'Distorted' Distribution", *Population Studies*, 53, 317–329.

Chen, Renbao and S. Philip Morgan (1991), "Recent Trends in the Timing of First Births in the United States", *Demography*, 28(4), 513–533.

Cheng, P. C. Roger and Eric S. Lin (2010), "Completing Incomplete Cohort Fertility Schedules", *Demographic Research*, 23(9), 223–256.

Coale, Ansley J. and Donald R. McNeil (1972), "The Distribution of Age at First Marriage in a Female Cohort", *Journal of the American Statistical Association*, 67, 743–749.

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.

Keilman, Nico (1994), "Translation Formulae for Non-Repeatable Events", *Population Studies*, 48(2), 341–357.

——— (2000), "Demographic Translation: From Period to Cohort Perspective and Back", Memorandum No. 06/2000, Department of Economics, University of Oslo.

Kohler, Hans-Peter and José Antonio Ortega (2002), "Tempo-Adjusted Period Parity Progression Measures, Fertility Postponement and Completed Cohort Fertility", *Demographic Research*, 6, 92–144.

Kohler, Hans-Peter and Dimiter Philipov (2001), "Variance Effects in the Bongaarts-Feeney Formula", *Demography*, 38(1), 1–16.

Li, Nan and Zheng Wu (2003), "Forecasting Cohort Incomplete Fertility: A Method and An Application", *Population Studies*, 57(3), 303–320.

Myrskylä, Mikko and Joshua R. Goldstein (2013), "Probabilistic Forecasting Using Stochastic Diffusion Models, With Applications to Cohort Processes of Marriage and Fertility", *Demography*, 50(1), 237–260.

Ní Bhrolcháin, Máire (2011), "Tempo and the TFR", *Demography*, 48, 841–861.

Ryder, Norman B. (1964), "The Process of Demographic Translation", *Demography*, 1(1), 74–82.

——— (1990), "What Is Going to Happen to American Fertility", *Population and Development Review*, 16(3), 433–454.

Schoen, Robert (2004), "Timing Effects and the Interpretation of Period Fertility", *Demography*, 41(4), 801–819.

Smallwood, Steve (2002), "The Effect of Changes in Timing of Childbearing on Measuring Fertility in England and Wales", *Population Trends*, 109, 36–45.

Sobotka, Tomáš (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, Evert (2001), "On the Impossibility of Inferring Cohort Fertility Measures from Period Fertility Measures", *Demographic Research*, 5, 23–64.

van Imhoff, Evert and Nico Keilman (2000), "On the Quantum and Tempo of Fertility: Comment", *Population and Development Review*, 26(3), 549–553.

Willekens, Frans and Nazli Baydar (1984), "Age-period-cohort models for forecasting fertility", *Working Paper No. 45, NIDI, The Hague.*

Yamaguchi, Kazuo and Motomi Beppu (2004), "Survival Probability Indices of Period Total Fertility Rate", Discussion Paper Series 2004-01, The Population Research Centre, NORC and The University of Chicago.

Zeng, Yi and Kenneth C. Land (2001), "A Sensitivity Analysis of the Bongaarts-Feeney Method for Adjusting Bias in Observed Period Total Fertility Rates", *Demography*, 38(1), 17–28.